



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

REVSTAT

Statistical Journal



REVSTAT

Statistical Journal

Catálogo Recomendada

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

CREDITS

- EDITOR-IN-CHIEF

- Isabel Fraga Alves

- CO-EDITOR

- Giovani L. Silva

- ASSOCIATE EDITORS

- Marília Antunes
- Barry ARNOLD (2019)
- Narayanaswamy Balakrishnan
- Jan Beirlant
- Graciela Boente (2019-2020)
- Paula Brito
- Vanda Inácio de Carvalho
- Arthur Charpentier
- Valérie Chavez-Demoulin
- David Conesa
- Charmaine Dean
- Jorge Milhazes Freitas
- Alan Gelfand
- Stéphane Girard
- Wenceslao Gonzalez-Manteiga
- Marie Kratz
- Victor Leiva
- Maria Nazaré Mendes-Lopes
- Fernando Moural
- John Nolan
- Paulo Eduardo Oliveira
- Pedro Oliveira
- Carlos Daniel Paulino (2019-2021)
- Arthur Pewsey
- Gilbert Saporta
- Alexandra M. Schmidt
- Julio Singer

- Manuel Scotto

- Lisete Sousa

- Milan Stehlik

- María Dolores Ugarte

- EXECUTIVE EDITOR

- José A. Pinto Martins

- FORMER EXECUTIVE EDITOR

- Maria José Carrilho

- Ferreira da Cunha

- SECRETARY

- Liliana Martins

- PUBLISHER

- Instituto Nacional de Estatística, I.P. (INE, I.P.)
Av. António José de Almeida, 2
1000-043 LISBOA
PORTUGAL
Tel.: + 351 21 842 61 00
Fax: + 351 21 845 40 84
Web site: <http://www.ine.pt>
Customer Support Service
+ 351 218 440 695

- COVER DESIGN

- Mário Bouçadas, designed on the stain glass
window at INE by the painter Abel Manta

- LAYOUT AND GRAPHIC DESIGN

- Carlos Perpétuo

- PRINTING

- Instituto Nacional de Estatística, I.P.

- EDITION

- 140 copies

- LEGAL DEPOSIT REGISTRATION

- N.º 191915/03

- PRICE [VAT included]

- € 9,00

Editorial Note: Letter from the Editor-in-Chief

It is my pleasure to welcome you to the first issue of REVSTAT for which I act as Editor.

First of all, a special thanks to my predecessor, Ivette Gomes. It is an honour to succeed her as editor and to help the journal evolve to continue to meet the needs of the applied and theoretical statisticians. Over the past 18 years, she improved the visibility and impact of REVSTAT, serving the science community with dedication and commitment, setting up and shaping the journal to ensure it has maintained its place among the international benchmark journals in the field of Statistics. With the regular publication of REVSTAT since 2003, she has acted with the utmost scientific integrity, which combined with her broad and deep understanding of statistical methodologies and applications, and a hard work capacity that makes her an example to be followed, but hard to match. Throughout this period Ivette has also promoted some special issues, spreading the scientific communications of thematic workshops or conferences. The importance of obtaining high quality reviews and timeliness in publish-decisions, have also been a major concern under Ivette's Editorship, assisted by the co-editor Antónia Amaral Turkman.

I also would like to thank those who served on the journal editorial board, some of the Associated Editors cooperating with REVSTAT since the first issue. Thanks are also due to reviewers, which constitute integrant part of this publishing process, providing the support and feedback necessary to find, develop and publish high-quality material. REVSTAT receives since 2010 the Journal Citation Reports impact factor, which to some extent reflects the dedication and expertise of our editors and reviewers. In 2017 the journal was given a 5-Year JCR Impact Factor of 1.238. Close to this, I also present my thanks to both readers and authors, who helped us by citing papers from our journal. To improve this result, we hope to receive high quality manuscripts from authors all over the world.

The new AE team will greatly contribute to the high standards of the Journal, and we are thankful for their committed participation in the respective field of

expertise, enabling the journal will continue to publish original high quality standards research in Statistics. We will encourage all authors to work to these standards. Peer review remains a prominent component of our assessment of submitted manuscripts. It is important we have a good balance of different article type within the journal, encompassing theoretical, methodological, real case studies applications, and promoting overviews or reviews of emerging subjects. Special Issues will continue to be published, not necessarily associated to scientific meetings.

Lastly, I must thank all our submitting authors, both current and future, who worked on the production of their research, and have chosen REVSTAT as the journal they would like to publish in. Unfortunately, due to the great volume of submissions, less than 20% of submissions are eventually accepted for publication, and inevitably many of those submitting authors will be disappointed by a negative decision of rejection.

I am very aware of the responsibilities that the editor's role entails, and I face my new role and challenge with both enthusiasm and some anxiety! At last, but not least, I am fortunate to be supported by a highly effective team from editorial office of Statistics Portugal, in particular by the executive editor Pinto Martins and Secretary Liliana Martins. It is their goal to adopt a refreshed look for the journal's website and in a near future an effective automatic submission editorial platform.

I look forward to working with all of you, aiming to make REVSTAT a success and we welcome your submissions, sincerely hoping you find the future issues of interest by benefiting from the articles appearing in this publication.

Sincerely,

Isabel Fraga Alves

INDEX

Some Monitoring Procedures Related to Asymmetry Parameter of Azzalini's Skew-Normal Model	
<i>Chenglong Li, Amitava Mukherjee, Qin Su and Min Xie</i>	1
On the Design Points for a Rotatable Orthogonal Central Composite Design	
<i>Christos P. Kitsos</i>	25
Random Environment INAR Models of Higher Order	
<i>Aleksandar S. Nastić, Petra N. Laketa and Miroslav M. Ristić</i>	35
Stochastic Inequalities for the Run Length of the EWMA Chart for Long-Memory Processes	
<i>Yarema Okhrin and Wolfgang Schmid</i>	67
Using Shrinkage Estimators to Reduce Bias and MSE in Estimation of Heavy Tails	
<i>Jan Beirlant, Gaonyalelwe Maribe and Andréhette Verster</i>	91
An Integrated Functional Weissman Estimator for Conditional Extreme Quantiles	
<i>Laurent Gardes and Gilles Stupfler</i>	109

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

SOME MONITORING PROCEDURES RELATED TO ASYMMETRY PARAMETER OF AZZALINI'S SKEW-NORMAL MODEL

Authors: CHENGLONG LI

– School of Management, Xi'an Jiaotong University, Xi'an, China
Department of Systems Engineering and Engineering Management,
City University of Hong Kong, Hong Kong, China
lchenglon2-c@my.cityu.edu.hk

AMITAVA MUKHERJEE

– Production, Operations and Decision Sciences Area,
XLRI-Xavier School of Management, Jamshedpur, India
amitmukh2@yahoo.co.in

QIN SU

– School of Management, Xi'an Jiaotong University, Xi'an, China
qinsu@mail.xjtu.edu.cn

MIN XIE

– Department of Systems Engineering and Engineering Management,
City University of Hong Kong, Hong Kong, China
minxie@cityu.edu.hk

Received: April 2016

Revised: December 2016

Accepted: December 2016

Abstract:

- In the real world, we often observe that the underlying distribution of some Gaussian processes tends to become skewed, when some undesirable assignable cause takes place in the process. Such phenomena are common in the field of manufacturing and in chemical industries, among others, where a process deviates from a normal model and becomes a skew-normal. The Azzalini's skew-normal (hereafter ASN) distribution is a well-known model for such processes. In other words, we assume that the in-control (hereafter IC) distribution of the process under consideration is normal, that is a special case of the ASN model with asymmetry parameter zero, whereas the out-of-control (hereafter OOC) process distribution is ASN with any non-zero asymmetry parameter. In the ASN model, a change in asymmetry parameter also induces shifts in both the mean and variance, even if, both the location and scale parameters remain invariant. Traditionally, researchers consider a shift either in the mean or in variance or in both the parameters of the normal distribution. Some inference and monitoring issues related to deviation from symmetry are essential problems that are largely overlooked in literature. To this end, we propose various test statistics and design for sequential monitoring schemes for the asymmetry parameter of the ASN model. We examine and compare the performance of various procedures based on an extensive Monte-Carlo experiment. We provide an illustration based on an interesting manufacturing case study. We also offer some concluding remarks and future research problems.

Key-Words:

- *disruption of symmetry; distance skewness; maximum likelihood estimator; Monte-Carlo simulations; skew-normal distribution; statistical process monitoring.*

AMS Subject Classification:

- 62F03, 62P30, 65C05, 90B99.

1. INTRODUCTION

In many of the practical applications, a univariate process characteristic, such as the warp length of semiconductor wafers, or the diameter of piston rings among others, is assumed to follow a normal distribution. A normal distribution can be completely specified by its mean (μ) and variance (σ^2), the two parameters of the distribution. In standard quality control literature, a number of control charts are developed and studied for detecting a shift in mean (also called location parameter), among them by Tsiamyrtzis and Hawkins [37], Ryu *et al.* [31], Khoo *et al.* [18], Peng *et al.* [26], and many others. Similarly, there are host of research articles for detection of a shift in variance or the scale parameter (σ), such as, Castagliola [6], Shu *et al.* [33], Zhang [40], Guo and Wang [14], among others. In the recent years, several researchers have also addressed the problem of jointly monitoring both the location and scale parameters of a normally distributed process. We recommend reading Hawkins and Deng [15], Wu *et al.* [39], Sheu *et al.* [32], McCracken *et al.* [24], Reynolds *et al.* [29], Knoth [19] and Li *et al.* [20], among others, for more details.

Despite a great progress of parametric testing of hypothesis and process monitoring, we, traditionally, assume that the parent population distribution of the process characteristic remains normal and only change takes place in the parameters of the distribution. Generally, we assume that the shift may occur either in its mean or variance or in both. Nevertheless, this assumption is more often very stringent. There are other ways in which a normally distributed process may change. Ross and Adams [30] stated that, in many real-life applications, it could be desirable to monitor for a change in the shape of the process distribution. Similar arguments can also be found in Zou and Tsung [41] and Li *et al.* [22]. Normal distribution is well known as a symmetric bell-shaped distribution and in consequence when a shift occurs in normal distribution, it may tend to become skewed or asymmetric.

This phenomenon, in fact, is quite common in practice, especially in physical, chemical or geological research field. Vincent and Walsh [38] indicated that the experimental intensity distributions in convergent beam electron diffraction patterns always exhibit deviations from ideal symmetry, attributable to the causes, such as, strain, inclined surfaces, incomplete unit cells and imperfections in the electron optics. Rahman and Hossain [27] showed another very relevant example, about the groundwater arsenic contamination in Bangladesh. They noted that the transmission of contaminants can affect the symmetric nature of the distribution of arsenic concentration, being positively skewed. Interested readers may also see Mukherjee *et al.* [25] for more details. In the context of statistical process monitoring, Figueiredo and Gomes [9] studied a real industrial example related to the diameters of cork stoppers produced by a manufacturing unit and

noted that the data nicely follows an ASN distribution. After a close examination, we find that the distribution of the diameters of the cork was actually normal in the initial phase and slowly it tends to become skew-normal. Figure 1 shows the histogram and density estimate of the first 200 diameter observations from the production data set of Figueiredo and Gomes [9], which has altogether 1000 observations. The p-value of the Shapiro-Wilk normality test for the first 200 observations is 0.9296 which strongly supports the normality assumption in the initial stage of production. Naturally, we can imagine that the process is shifted from a normal distribution to an ASN distribution in the later stage of production. We provide a detailed illustration with the cork stoppers' data later in Section 5.

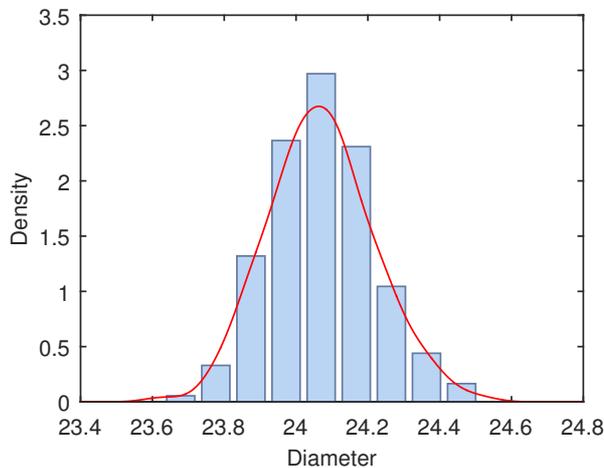


Figure 1: Histogram and density estimate of the first 200 diameter observations.

In the nice work, Ferreira and Steel [8] proposed a constructive representation of skewed distributions and provided three common methods of generating univariate skewed distributions, namely, hidden truncation, inverse scale factors, and order statistics. Among them, the skew-normal distribution of Azzalini [2] is probably the most common and most intensively studied one from diverse areas of application and is developed with the idea—hidden truncation. The statistical properties of ASN distribution and its variations have been discussed by several authors and many similarities with the ordinary normal distribution are observed, see for example, Azzalini [2, 3], Henze [16], Genton *et al.* [12], Arellano-Valle *et al.* [1], Chen *et al.* [7], Azzalini [4], Gómez *et al.* [13], Mamei and Musio [23], Su and Gupta [35]. Various researches established that the ASN family of distributions have rather important roles to play in the production practice, such as, modeling real datasets or simulating skewed data with different degrees of asymmetry and tail-weight. Interested readers may see, among others, Chen *et al.* [7], Bartoletti

and Loperfido [5], Fruhwirth-Schnatter and Pyne [11], Razzaghi [28], Figueiredo and Gomes [10]. Nevertheless, there are merely few research articles that have addressed the process monitoring issues with ASN distributions. The problems related to process monitoring are considered in Tsai [36], Figueiredo and Gomes [9], Su *et al.* [34], Li *et al.* [21]. The major theme of these researches is, however, the construction of control charts for skewed data, that includes detection of shifts in location and/or scale. Needless to say that the problem of monitoring and detecting departures from normality, i.e., from the normal to skew-normal, has not been considered yet. Such a distributional change might not be readily spotted by some traditional control charting schemes, such as the \bar{X} chart and S chart, because they are not designed for that purpose. Our current work aims at addressing this long-standing problem in the context of process monitoring and attempts to bridge the existing research gap.

The rest of this paper is organized as follows: Section 2 provides some information about ASN distribution and also introduces several competitive test statistics for the purpose of detecting the disruption of symmetry of a normally distributed process characteristic. The respective sequential monitoring procedures, as well as the determination of their design parameters are presented in Section 3. An extensive performance comparison and analysis is included in Section 4. Section 5 illustrates the real example based on the corks' diameter data from Figueiredo and Gomes [9]. Finally, we offer some concluding remarks and problems for future research in Section 6.

2. SOME STATISTICAL TESTS FOR ASYMMETRY PARAMETER

Let X be the continuous random variable (r.v.) denoting the process characteristic subject to testing or monitoring. The r.v. X is said to follow ASN distribution if its probability density function (pdf) is of the form:

$$f(x; \xi, \omega, \lambda) = \frac{2}{\omega} \phi\left(\frac{x - \xi}{\omega}\right) \Phi\left(\lambda \frac{x - \xi}{\omega}\right), \quad -\infty < x < \infty,$$

$$-\infty < \xi < \infty, \quad -\infty < \lambda < \infty, \quad \omega > 0,$$

where ξ is the location parameter, ω is the scale parameter, and λ is the asymmetry parameter, also called shape parameter; $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cumulative distribution function (cdf) of the standard normal distribution, respectively. In a standard notation, we express it as $X \sim ASN(\xi, \omega, \lambda)$.

The ASN distribution is positively skewed if $\lambda > 0$, and is negatively skewed if $\lambda < 0$. The critical parameter λ controls the skewness of the distribution (see Figure 2). Note that, when $\lambda = 0$, the ASN distribution boils down to a normal

distribution with mean $\mu = \xi$ and variance $\sigma^2 = \omega^2$, which implies that normal distribution is a special case of the ASN family of distributions. Accordingly, the ordinary normal distribution can also be denoted as $X \sim ASN(\xi, \omega, 0)$, instead of $X \sim N(\mu, \sigma^2)$.

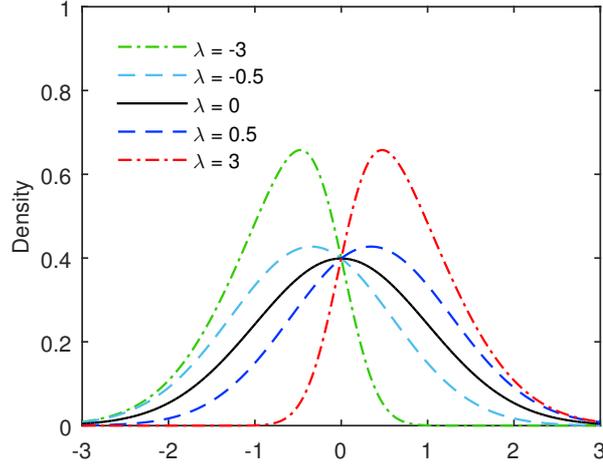


Figure 2: The pdf of Azzalini's SN distribution.

The variation of asymmetry parameter λ in ASN distribution automatically changes its mean and variance. If the distribution shifts from $ASN(\xi_0, \omega_0, 0)$ to $ASN(\xi_0, \omega_0, \lambda)$, it is easy to see that the changed mean $\mu_1 = \mu_0 + \lambda\sigma_0\sqrt{\frac{2}{\pi(1+\lambda^2)}}$ and variance $\sigma_1^2 = \left(1 - \frac{2\lambda^2}{\pi(1+\lambda^2)}\right)\sigma_0^2$ where we have $\mu_0 = \xi_0$ and $\sigma_0 = \omega_0$. In this context, if we apply a simultaneous testing or monitoring scheme, designed for the mean and variance of a normal model, we often get illusive results as the shifted model no longer follows normal distribution. A more practical problem is to test or monitor the asymmetry parameter of ASN distribution or all the parameters of the ASN distribution. In this paper, we only consider the inference and monitoring problems related to the asymmetry parameter λ of the ASN distribution and assume both the location and scale parameters remain invariant and known. In other words, we consider the problem of sequential monitoring of the ideal condition of $\lambda = 0$. Further research on simultaneous monitoring of all the parameters may be taken separately as a highly warranted research problem. In the present context, we assume that ξ_0 and ω_0 , the process location and scale or the mean and variance of the normally distributed IC process are known. This assumption is realistic as practitioners commonly have a fair idea about the process parameters either based on certain target set-up of the companies or based on the prior knowledge about the distribution of process characteristics.

We first consider the tests based on the likelihood ratio criterion, and the moment estimator for the asymmetry parameter λ . Noting that ad-hoc inference

is gaining more and more popularity, we also introduce several ad-hoc statistics for tracking skewness which is related to the value of λ . In the subsequent subsections, we introduce these statistics and the tests based on them.

2.1. Likelihood ratio test

Let $\mathbf{x}_n = (x_1, x_2, \dots, x_n)$ be the sample of size n drawn from the r.v. X with $X \sim ASN(\xi, \omega, \lambda)$. Given \mathbf{x}_n , the log-likelihood function for λ is given by

$$l(\lambda|\mathbf{x}_n, \mu_0, \sigma_0) = n \ln 2 - n \ln \sigma_0 + \sum_{i=1}^n \ln \phi\left(\frac{x_i - \mu_0}{\sigma_0}\right) + \sum_{i=1}^n \ln \Phi\left(\lambda \frac{x_i - \mu_0}{\sigma_0}\right).$$

Writing $z_i = \frac{x_i - \mu_0}{\sigma_0}$, we can obtain the maximum likelihood (ML) estimator of λ , say $\hat{\lambda}_{MLE}$, as the solution of

$$\frac{dl}{d\lambda} = \sum_{i=1}^n \frac{z_i \phi(\lambda z_i)}{\Phi(\lambda z_i)} = 0.$$

In the language of theory of testing of statistical hypothesis, if we consider the problem of testing $H_0: X \sim ASN(\xi_0, \omega_0, \lambda_0 = 0)$, that is, $X \sim N(\mu_0, \sigma_0^2)$ against $H_1: X \sim ASN(\xi_0, \omega_0, \lambda_1)$, a likelihood ratio criterion can be given by:

$$\Lambda(\mathbf{x}_n|\mu_0, \sigma_0) = \frac{1}{2^n \prod_{i=1}^n \Phi(\hat{\lambda}_{MLE} z_i)}.$$

We reject H_0 at a given level of significance if $\Lambda(\mathbf{x}_n|\mu_0, \sigma_0) < c_{LR}$, where c_{LR} is a pre-determined constant that satisfies the level criterion. Note that, $\Lambda(\mathbf{x}_n|\mu_0, \sigma_0) < c_{LR}$, indeed, is equivalent to $T = -2 \ln \Lambda(\mathbf{x}_n|\mu_0, \sigma_0) = 2n \ln 2 + 2 \sum_{i=1}^n \ln \Phi(\hat{\lambda}_{MLE} z_i) > -2 \ln c_{LR}$. Thus, writing $c_{LR}^* = -2 \ln c_{LR}$, the critical region of the test may be given by $T > c_{LR}^*$.

Using the nice analogy between theory of testing of hypothesis and statistical process control, we can easily develop a sequential monitoring procedure based on T . It is known that, under the null hypothesis, T follows a chi-squared distribution with 1 degree of freedom if n is large. However, we found that this approximation is useful if the test sample size n is at least 100. For small n , the approximation is not at all satisfactory. In the context of statistical process monitoring, test sample sizes are usually very small, say $n = 5$ or 10 or 25. A sample size of $n > 100$ is very rare in practice and in quality control literature. Therefore, we omit the asymptotic theory related to T in subsequent analysis and discussion. Instead, we choose to work with the simulated distribution of T traced via Monte-Carlo.

2.2. Test based on the moment estimators

2.2.1. Inadmissibility of test based on method of moments estimator

Unlike the ML estimation, the method of moments (MM) to estimate λ may be obtained more explicitly using the sample skewness, say $\hat{\gamma}$, by inverting the skewness equation given as:

$$|\delta| = \sqrt{\frac{\pi}{2} \frac{|\hat{\gamma}|^{\frac{2}{3}}}{|\hat{\gamma}|^{\frac{2}{3}} + \left(\frac{4-\pi}{2}\right)^{\frac{2}{3}}}}, \quad \delta = \frac{\lambda}{\sqrt{1+\lambda^2}},$$

where the sign of δ is the same as the sign of $\hat{\gamma}$ and thus, we can derive the MM estimator of λ , say $\hat{\lambda}_{MME} = \frac{\delta}{\sqrt{1-\delta^2}}$. Note that here theoretically the maximum skewness is obtained by setting $\delta = 1$, which gives $\hat{\gamma}$ approximately equal to 0.99527. Nevertheless, in practice, it may happen that the observed sample skewness is larger. In such situations, $\hat{\lambda}_{MME}$ cannot be obtained from the above equation. Admittedly, we may consider a trade-off by letting $|\hat{\gamma}| = 0.99527$ when the obtained $|\hat{\gamma}|$ is coincidentally greater than 0.99527.

Interestingly, we have found that the MM estimation, in the present context, is rather inefficient especially for small-to-moderate sample size. We observe with $n = 5$, when the process is IC, the probability that the sample skewness exceeds 0.99527 is about 12.2%. That is, we cannot construct a nontrivial exact test at 5% level in this context. If $n = 15, 25$ and 50 the probability of the same event becomes 5.9%, 2.7%, and 0.5% respectively. In the process monitoring context, with $n = 50$, we may construct a sequential inspection scheme that will allow us to achieve a maximum IC average run length: $IC-ARL = \frac{1}{0.005} = 200$. This is certainly undesirable and thus, we drop this statistic from further discussion.

2.2.2. Estimator based on L-moments

L-moments are a sequence of statistics used to summarize the shape of a probability distribution. They are linear combinations of the order statistics analogous to conventional moments. Let x_1, x_2, \dots, x_n be the sample and $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ be the ordered sample, and direct estimators for the first three L-moments in a finite sample of n observations are defined to be (see Hosking [17])

$$l_1 = n^{-1} \sum_i x_i,$$

$$l_2 = \frac{1}{2} \binom{n}{2}^{-1} \sum_i \sum_j (x_{(i)} - x_{(j)}), \quad \text{for } i > j,$$

$$l_3 = \frac{1}{3} \binom{n}{3}^{-1} \sum_i \sum_j \sum_k (x_{(i)} - 2x_{(j)} + x_{(k)}), \quad \text{for } i > j > k.$$

The L-skewness is estimated by $L = l_3/l_2$. The L-skewness is a coefficient that measures the degree of asymmetry and may take on positive or negative values. It is known that $0 \leq |L| < 1$, where $L = 0$ indicates a possible symmetry. Therefore, we can consider a test based on $|L|$ and reject H_0 at a given level of significance if $|L| > c_{LS}$, where c_{LS} is a pre-determined constant satisfies the level criterion.

2.3. Ad-hoc approaches

2.3.1. Test based on the sample skewness statistic

Instead of the MM estimation, the sample skewness, $\hat{\gamma}$, may be directly adopted to track the skewness of the process distribution and judge whether there is a shift from $\lambda = 0$. The form of sample skewness is given by

$$\hat{\gamma} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left[\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^{3/2}},$$

where \bar{x} denotes the mean of the sample of size n . In general, we expect that under symmetry, $\hat{\gamma}$ should be closer to 0. Under positive or negative skewness, we generally expect that $\hat{\gamma}$ is greater than or less than 0 respectively. If we are interested in detecting a general two-sided shift (both left or right skewness) from symmetry, we prefer a test based on $|\hat{\gamma}|$ and reject H_0 at a given level of significance if $|\hat{\gamma}| > c_{SS}$, where c_{SS} is a pre-determined constant satisfies the level criterion.

2.3.2. Test based on the distance skewness statistic

There is a simple consistent statistical test of diagonal symmetry based on the sample distance skewness:

$$D = 1 - \frac{\sum_{i,j} |z_i - z_j|}{\sum_{i,j} |z_i + z_j|},$$

where the z 's are the standardized observations in a sample. The sample distance skewness can be used as a way to decide whether there is a shift from $\lambda = 0$. Its value is always between 0 and 1, and in general, it is expected that under symmetry, $D = 0$ and under positive or negative skewness, D is expected to be greater than 0. Thus, the statistic D can be considered for two-sided test of $H_0: \lambda = 0$ versus $H_1: \lambda \neq 0$. We reject H_0 at a given level of significance if $D > c_{DS}$, where c_{DS} is a pre-determined constant satisfies the level criterion.

2.3.3. Test based on the median skewness statistic

The Pearson's median skewness, or second skewness coefficient, is defined by

$$M = 3 \frac{(\bar{x} - \tilde{x})}{s},$$

where \tilde{x} is the sample median and s is the sample standard deviation of size n . It is a simple multiple of the nonparametric skew. In general, it is expected that under symmetry, $\bar{x} = \tilde{x}$ and consequently $M = 0$. Under positive or negative skewness, in general, we expect M greater than or less than 0 respectively. Therefore, we may reject H_0 at a given level of significance if $|M| > c_{MS}$, where c_{MS} is a pre-determined constant satisfies the level criterion.

2.3.4. Quantile-based approach

Writing Q_i , $i = 1, 2, 3$ as the i^{th} quartile of the distribution, the Bowley's measure of skewness is given by

$$B = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}.$$

It is expected that for a normal distribution $B = 0$ and for an ASN distribution $B > 0$ or $B < 0$ according as $\lambda > 0$ or $\lambda < 0$. Therefore, for simplicity one can use B to verify whether the symmetry condition of the normal distribution remains valid or an asymmetric pattern creeps in. For two-sided monitoring, we may use $|B|$ as the monitoring statistic. We reject H_0 at a given level of significance if $|B| > c_{BS}$, where c_{BS} is a pre-determined constant satisfies the level criterion.

3. DESIGN AND IMPLEMENTATION OF SEQUENTIAL MONITORING OF ASYMMETRY PARAMETER

From production and manufacturing to various other sectors, often sequential monitoring and control of process parameter is of primary interest. In this section, we present six monitoring procedures based on the test statistics, respectively, introduced in Section 2. These statistics are

- (a) the likelihood ratio statistic T ,
- (b) the L-skewness statistic $|L|$,
- (c) the sample skewness statistic $|\hat{\gamma}|$,
- (d) the sample distance skewness statistic D ,
- (e) the median skewness statistic $|M|$, and
- (f) the Bowley's statistic $|B|$.

Thus, we consider the following six schemes (A-F) for sequential monitoring of asymmetry parameter.

- A: The NSN-LR chart based on likelihood ratio statistic as in Section 2.1;
- B: The NSN-LS chart based on L-skewness as in Section 2.2.2;
- C: The NSN-SS chart based on sample skewness as in Section 2.3.1;
- D: The NSN-DS chart based on sample distance skewness as in Section 2.3.2;
- E: The NSN-MS chart based on median skewness as in Section 2.3.3;
- F: The NSN-BS chart based on Bowley's measure of skewness as in Section 2.3.4.

The abbreviation NSN is used to highlight the purpose of detecting a shift from Normal(N) to Skew-Normal (SN). We first consider the sequential monitoring procedure based on the NSN-LR chart. The method of constructing a NSN-LR chart involves the following steps:

- Step-1: Collect $\mathbf{x}_{jn} = (x_{j1}, x_{j2}, \dots, x_{jn})$, the j^{th} test sample from the process for $j = 1, 2, \dots$. Clearly, n is the fixed sample size for the j^{th} test sample or the so called rational subgroup.
- Step-2: Compute the plotting statistic: $T_j = 2n \ln 2 + 2 \sum_{i=1}^n \ln \Phi(\hat{\lambda}_{MLE} z_{ji})$.
- Step-3: Plot T_j against an upper control limit (UCL) H_{LR} . The lower control limit (LCL) is by default 0, noting that $T_j \geq 0$ by definition as $\Lambda(\mathbf{x}_{jn} | \mu_0, \sigma_0)$ takes a value between 0 and 1.
- Step-4: If T_j exceeds H_{LR} , the process is declared OOC at the j^{th} test sample. If not, the process is considered to be IC, and testing continues to the next sample.

The sequential monitoring procedures based on other statistics are very similar except for the steps related to computing the plotting statistics and using corresponding control limits. Therefore, we omit the details for brevity.

It is easy to note that we are basically considering standard Phase-II Shewhart-type charts with standards known (Case-K). Consequently, the run-length distribution will be exactly geometric. Consider any statistic U and corresponding UCL as H_U . The expected IC run length can be expressed in terms of probabilities: $p_U(H_U) = P[U_j > H_U | IC]$. Let $F_U(\cdot)$ be the cdf of the plotting statistic under IC set-up. Then, we can also write $p_U(H_U) = 1 - F_U(H_U)$. In the present context, we identify U with T , $|L|$, $|\hat{\gamma}|$, D , $|M|$ and $|B|$, respectively, for the schemes A to F discussed above. Further, we identify H_U with H_{LR} , H_{LS} , H_{SS} , H_{DS} , H_{MS} and H_{BS} , respectively, for these six schemes.

Table 1: The UCL values for the NSN charts.

n	The NSN-LR chart: H_{LR}			The NSN-LS chart: H_{LS}		
	$IC-ARL$ = 250	$IC-ARL$ = 370	$IC-ARL$ = 500	$IC-ARL$ = 250	$IC-ARL$ = 370	$IC-ARL$ = 500
5	6.9315*	6.9315*	6.9315*	0.7980	0.8208	0.8383
10	10.1944	11.2063	12.2779	0.4754	0.4942	0.5086
15	9.0789	9.9548	10.6227	0.3615	0.3754	0.3854
20	8.8727	9.6034	10.2425	0.3090	0.3215	0.3292
25	8.6560	9.4490	9.9779	0.2689	0.2795	0.2896
30	8.5876	9.3119	9.8591	0.2442	0.2534	0.2610
50	8.4382	9.2416	9.7183	0.1831	0.1906	0.1957

n	The NSN-SS chart: H_{SS}			The NSN-DS chart: H_{DS}		
	$IC-ARL$ = 250	$IC-ARL$ = 370	$IC-ARL$ = 500	$IC-ARL$ = 250	$IC-ARL$ = 370	$IC-ARL$ = 500
5	1.4429	1.4557	1.4631	0.8185	0.8364	0.8496
10	1.7707	1.8493	1.9050	0.5760	0.6018	0.6197
15	1.6327	1.7181	1.7797	0.4355	0.4609	0.4793
20	1.5028	1.5848	1.6484	0.3501	0.3707	0.3854
25	1.3948	1.4670	1.5286	0.2925	0.3122	0.3259
30	1.2815	1.3587	1.4126	0.2530	0.2691	0.2815
50	1.0233	1.0797	1.1331	0.1638	0.1762	0.1857

n	The NSN-MS chart: H_{MS}			The NSN-BS chart: H_{BS}		
	$IC-ARL$ = 250	$IC-ARL$ = 370	$IC-ARL$ = 500	$IC-ARL$ = 250	$IC-ARL$ = 370	$IC-ARL$ = 500
5	1.9909	2.0165	2.0360	0.9459	0.9555	0.9623
10	1.5366	1.5842	1.6173	0.8638	0.8797	0.8929
15	1.5111	1.5614	1.6003	0.7998	0.8221	0.8350
20	1.2690	1.3147	1.3444	0.7011	0.7206	0.7367
25	1.2227	1.2652	1.3003	0.6722	0.6916	0.7061
30	1.0807	1.1178	1.1545	0.6242	0.6452	0.6607
50	0.8815	0.9154	0.9506	0.5149	0.5339	0.5464

Note: * indicates invalid UCL values.

In general, the charts are designed such that the appropriate UCL is found for a desired nominal *IC-ARL* or called ARL_0 . Now equating expected run length with the target *IC-ARL*, we have $IC-ARL = \frac{1}{p_U(H_U)}$ from which we can find expression for H_U in terms of the target *IC-ARL*. To this end, we use a Monte-Carlo simulation with adequate replicates (100,000 times) and acquire the appropriate quantile based on the empirical distribution function for realizing the target *IC-ARL*. Throughout the paper, we adopt this simulation technique, and in Table 1, we offer some UCL values for these aforementioned NSN charts for various test sample size n and for various nominal *IC-ARL* values.

From the UCL values of Table 1, we observe that H_{LR} and H_{SS} increase initially when n is small and then decrease gradually when n is relatively larger, while for the other charts UCL values decrease monotonically within the purview of range of n considered here, that is, $n \leq 50$. It is worth mentioning that it is difficult to obtain UCL values for the NSN-LR chart for some common *IC-ARL* context, when test sample size is small, say, $n < 10$. This is not surprising as Figueiredo and Gomes [9] noted that small n may often produces boundary estimates. The log-likelihood function will be an increasing (decreasing) function of λ if all observations are positive (negative). Nevertheless, overall performance of the NSN-LR chart is very encouraging in most cases, as long as the test sample size is not too small.

4. PERFORMANCE ANALYSIS FOR QUICKEST DETECTION

4.1. The performance comparisons between NSN charts

In the present paper, clearly the IC value of λ is $\lambda_0 = 0$. To compare these NSN charts thoroughly and for performance analysis, we choose the shifted (OOC) value of λ as $\lambda_1 = 0.3, 0.5, 1, 2, 3, 5, 10$, for drawing Phase-II samples. Without loss of generality, for both the IC or OOC situations, we consider $\mu_1 = \mu_0 = 0$ and $\sigma_1 = \sigma_0 = 1$. For specified n and *IC-ARL* ($= 370$), we compute the *ARL* and the standard deviation of the run length (*SDRL*). Our findings for $n = 5, 10, 15, 25$ are summarized in Table 2. For some other values of *IC-ARL* (say, 250 and 500), the results are comparable and consistent, and therefore, we omit the details for brevity.

First we notice that, for specified test sample size n , the *ARL* and the *SDRL* of all the NSN charts decrease steeply with the increasing shift in λ . Further, when n increases, in general, we see that for any NSN scheme, barring some sampling fluctuations, both the *ARL* and *SDRL* tend to decrease. Precisely, the larger the value of n is, the quicker the detection of a specified magnitude of shift will be.

For the NSN-BS chart, however, the rate of change of *OOC-ARL* with n is very slow.

Table 2: The *OOC* performance comparisons between NSN charts for various λ_1 and n when $IC-ARL = 370$.

λ_1	$n = 5$					
	NSN-LR chart	NSN-LS chart	NSN-SS chart	NSN-DS chart	NSN-MS chart	NSN-BS chart
0.3	Not Useful	365.94 (367.66)	361.33 (362.72)	230.85 (229.96)	360.92 (359.35)	354.16 (354.21)
0.5		365.94 (363.62)	360.13 (361.32)	145.69 (145.00)	360.37 (359.34)	352.48 (351.78)
1		358.65 (357.29)	353.72 (353.22)	60.77 (60.21)	361.97 (362.68)	351.79 (352.46)
2		314.71 (314.08)	310.02 (310.61)	30.19 (29.60)	352.62 (352.00)	337.77 (338.82)
3		268.01 (267.64)	263.68 (263.41)	24.70 (24.12)	329.34 (328.60)	311.48 (310.98)
5		214.00 (213.31)	210.57 (209.94)	22.45 (21.87)	288.18 (287.75)	271.05 (270.74)
10		171.76 (170.93)	170.43 (169.73)	21.97 (21.41)	244.67 (243.94)	228.48 (227.81)
λ_1	$n = 10$					
	NSN-LR chart	NSN-LS chart	NSN-SS chart	NSN-DS chart	NSN-MS chart	NSN-BS chart
0.3	129.81 (128.97)	360.69 (360.97)	362.55 (362.36)	123.92 (122.95)	350.36 (351.69)	375.60 (375.65)
0.5	55.69 (54.73)	357.54 (357.25)	360.44 (359.63)	51.78 (51.16)	350.22 (349.81)	374.47 (371.86)
1	13.06 (12.53)	325.34 (325.80)	322.95 (320.52)	12.04 (11.53)	350.59 (350.57)	373.06 (372.07)
2	3.78 (3.23)	187.36 (186.80)	178.54 (177.59)	3.84 (3.31)	302.18 (302.43)	353.01 (354.16)
3	2.31 (1.74)	111.74 (111.11)	111.15 (109.89)	2.61 (2.05)	229.49 (228.47)	319.94 (320.17)
5	1.55 (0.93)	64.90 (64.73)	73.35 (72.47)	2.01 (1.43)	149.87 (149.05)	279.57 (280.26)
10	1.18 (0.46)	44.29 (43.99)	57.45 (56.79)	1.75 (1.15)	103.43 (103.09)	253.59 (254.70)
λ_1	$n = 15$					
	NSN-LR chart	NSN-LS chart	NSN-SS chart	NSN-DS chart	NSN-MS chart	NSN-BS chart
0.3	73.57 (73.20)	329.54 (328.94)	338.05 (336.81)	74.68 (74.47)	354.19 (353.90)	380.89 (379.71)
0.5	24.61 (24.10)	326.13 (326.26)	334.20 (335.13)	24.98 (24.46)	353.86 (353.78)	378.44 (376.32)
1	4.56 (4.02)	279.58 (279.16)	269.27 (267.96)	4.80 (4.25)	353.81 (352.84)	375.84 (377.41)
2	1.46 (0.82)	115.69 (115.01)	112.15 (112.21)	1.62 (1.00)	282.86 (282.64)	345.41 (343.83)
3	1.11 (0.35)	56.16 (55.43)	62.81 (62.63)	1.23 (0.53)	190.95 (190.35)	294.50 (293.73)
5	1.01 (0.11)	28.17 (27.46)	39.39 (39.11)	1.07 (0.28)	108.10 (107.82)	235.77 (236.28)
10	1.00 (0.01)	17.96 (17.40)	30.27 (29.69)	1.02 (0.16)	69.06 (68.37)	204.56 (203.73)
λ_1	$n = 25$					
	NSN-LR chart	NSN-LS chart	NSN-SS chart	NSN-DS chart	NSN-MS chart	NSN-BS chart
0.3	35.52 (34.95)	367.13 (366.06)	370.12 (371.24)	36.60 (36.09)	345.62 (346.94)	357.06 (358.49)
0.5	9.49 (9.00)	361.87 (361.88)	361.69 (362.16)	9.99 (9.50)	346.00 (346.17)	355.45 (355.78)
1	1.80 (1.21)	276.97 (276.77)	250.80 (250.88)	1.93 (1.34)	333.34 (332.28)	350.34 (351.09)
2	1.02 (0.16)	70.21 (69.51)	69.69 (69.30)	1.05 (0.22)	198.51 (197.95)	294.49 (293.77)
3	1.00 (0.03)	26.42 (25.85)	33.83 (33.27)	1.00 (0.06)	99.94 (99.03)	218.99 (217.80)
5	1.00 (0.00)	11.43 (10.87)	19.29 (18.75)	1.00 (0.01)	45.51 (45.03)	153.77 (153.12)
10	1.00 (0.00)	6.90 (6.40)	14.31 (13.86)	1.00 (0.00)	28.00 (27.58)	130.94 (130.07)

For a given n and λ_1 , we compare the schemes in terms of *OOC-ARL*, and consider a scheme the best, if it offers the lowest *OOC-ARL*. The cells correspond to the best performing chart are shown in bold typeface in the tables. We further see from Table 2 that the NSN-LR chart and the NSN-DS chart are uniformly superior to the other four NSN charts. The NSN-DS chart is particularly well suited for the cases where n is small (e.g., $n = 5$), where the NSN-LR chart is inadmissible, as mentioned earlier. The two charts, namely, NSN-LR chart and

NSN-DS chart, perform rather similarly with a moderate-to-large test sample size (say $n \geq 10$), though the NSN-LR chart displays a slight advantage over the NSN-DS chart. Besides, both these charts have a rather low *OOC-ARL* value when n is large, even if, the shift in λ is relatively small. The rest four NSN charts perform poorly in almost all cases. The NSN-SS chart, the NSN-MS chart and the NSN-LS chart are very inefficient when n is small and shift size is also small, however, under large n , and for large shift size, the performance of these charts improves significantly. Nevertheless, even with $n = 50$ and $\lambda_1 = 10$, these schemes are inferior compared with the NSN-LR chart or the NSN-DS chart. Unfortunately, the NSN-BS chart is the worst and is practically useless.

Based on the results displayed in Table 2, we highly recommend the NSN-LR chart and the NSN-DS chart for detecting the shift from normal to skew-normal, especially the latter. The NSN-DS chart has a broader scope of application in practice than the NSN-LR chart as it is effective even if the test sample size is small where the NSN-LR chart is inadmissible. Further, we almost always see that the NSN-DS chart performs very close to the NSN-LR chart when the NSN-LR chart is the best in terms of *OOC-ARL* values. Therefore, the NSN-DS chart is very competitive, and moreover, it may be more preferable to the users taking into account the simplicity of implementation and its inherent ability to detect a deviation from symmetry. Nevertheless, from the performance perspective, we recommend both charts and the users can have a choice to adopt the NSN-LR chart or the NSN-DS chart according to their practical requirement. The rest four NSN charts based on common measures of skewness are much more inefficient and we suggest not to use them.

4.2. The comparisons to traditional charts for mean and/or variance

We have noted earlier that when the underlying process distribution deviates from normality and becomes skew-normal, as a result of a shift in the asymmetry parameter λ from 0, the process mean and variance also change. The mean and variance of the shifted process are given respectively by $\mu_1 = \mu_0 + \lambda_1 \sigma_0 \sqrt{\frac{2}{\pi(1+\lambda_1^2)}}$ and $\sigma_1^2 = \left(1 - \frac{2\lambda_1^2}{\pi(1+\lambda_1^2)}\right) \sigma_0^2$. Therefore, one may argue that it might be still meaningful to employ traditional process control schemes to monitor process mean, or process variance or both at the same time, without giving much importance to shift in the shape. To this end, it is worthy to compare some traditional monitoring procedures, such as, the \bar{X} chart for solely monitoring the process mean, the S chart for monitoring the process variance, as well as the charts based on ordinary max or distance statistic for jointly monitoring both the mean and variance, with the NSN-LR and NSN-DS charts. Such a comparative performance study will reflect whether the proposed schemes are really suitable in detecting an overall process shift quickly compared to traditional schemes.

Following are the plotting statistics of the existing schemes for monitoring the process mean, process variance or both, used for the comparative study:

\bar{X} chart:

$$Q_{\bar{X}}(\mathbf{X}_j) = \left| \frac{\bar{X}_j - \mu}{\sigma/\sqrt{n}} \right|;$$

S chart:

$$Q_S(\mathbf{X}_j) = \left| \Phi^{-1} \left\{ F_{\chi_{(n-1)}^2} \left(\frac{(n-1)S_j^2}{\sigma^2} \right) \right\} \right|;$$

max chart:

$$Q_M(\mathbf{X}_j) = \max \left\{ \left| \frac{\bar{X}_j - \mu}{\sigma/\sqrt{n}} \right|, \left| \Phi^{-1} \left\{ F_{\chi_{(n-1)}^2} \left(\frac{(n-1)S_j^2}{\sigma^2} \right) \right\} \right| \right\};$$

distance chart:

$$Q_D(\mathbf{X}_j) = \sqrt{\left(\frac{\bar{X}_j - \mu}{\sigma/\sqrt{n}} \right)^2 + \left(\Phi^{-1} \left\{ F_{\chi_{(n-1)}^2} \left(\frac{(n-1)S_j^2}{\sigma^2} \right) \right\} \right)^2}.$$

We compare the above four schemes with the proposed NSN-LR chart and NSN-DS chart under the similar OOC set-up used in Section 4.1. For a fair comparison, we only consider the standard Shewhart-type version of all the charts involved. In Table 3, we present the mean, the standard deviation and the skewness coefficient of the $ASN(\lambda)$ distribution for various values of λ_1 , considered in Section 4.1.

Table 3: Means, standard deviations and skewness coefficients of the ASN distribution under IC value and various OOC values of λ .

IC Situation			
λ_0	μ_0	σ_0	γ_0
0	0	1	0
OOC Situation			
λ_1	μ_1	σ_1	γ_1
0.3	0.2293	0.9734	0.0056
0.5	0.3568	0.9342	0.0239
1	0.5642	0.8256	0.1369
2	0.7136	0.7005	0.4538
3	0.7569	0.6535	0.6670
5	0.7824	0.6228	0.8510
10	0.7939	0.6080	0.9556
$+\infty$	0.7979	0.6028	0.9953

From Table 3, it is easy to see that when λ increases from 0 to $+\infty$, μ and γ increase, but σ decreases. One may check that when λ_1 decreases from 0 to $-\infty$, all three measures, the mean, the variance and the skewness coefficient decrease. To be precise, in our simulation set-up, if there is a shift from $\lambda = 0$ to $\lambda_1(-\lambda_1)$, the mean $\mu = 0$ will change to $\mu_1(-\mu_1)$, the standard deviation $\sigma = 0$ will change to $\sigma_1(\sigma_1)$, and $\gamma = 0$ will change to $\gamma_1(-\gamma_1)$. For brevity, we omit the case of decreasing shift in λ .

In Table 4, we summarize the result of performance comparisons among the NSN-LR chart, the NSN-DS chart and the four traditional alternatives (i.e.,

Table 4: The OOC performance comparisons among the NSN-LR chart, the NSN-DS chart and other alternatives for various λ_1 and n when $IC-ARL = 370$.

λ_1	$n = 5$					
	NSN-LR chart	NSN-DS chart	\bar{X} chart	S chart	max chart	distance chart
0.3		230.85 (229.96)	181.20 (180.84)	421.72 (422.96)	244.14 (243.51)	240.61 (239.75)
0.5		145.69 (145.00)	106.09 (105.12)	454.31 (453.81)	160.33 (158.94)	159.68 (159.26)
1	Not Useful	60.77 (60.21)	52.23 (51.62)	337.54 (337.87)	83.64 (82.75)	81.57 (80.57)
2		30.19 (29.60)	36.06 (35.38)	164.33 (163.60)	55.55 (54.98)	52.02 (51.27)
3		24.70 (24.12)	33.17 (32.55)	113.58 (113.55)	48.38 (47.73)	46.10 (45.50)
5		22.45 (21.87)	31.74 (31.18)	83.34 (83.27)	43.36 (42.79)	42.89 (42.42)
10		21.97 (21.41)	31.12 (30.60)	69.16 (69.13)	40.41 (39.78)	41.24 (40.82)

λ_1	$n = 10$					
	NSN-LR chart	NSN-DS chart	\bar{X} chart	S chart	max chart	distance chart
0.3	129.81 (128.97)	123.92 (122.95)	102.21 (101.85)	410.33 (409.76)	150.01 (150.07)	149.23 (147.96)
0.5	55.69 (54.73)	51.78 (51.16)	43.94 (43.37)	380.25 (378.66)	68.56 (68.16)	68.96 (67.96)
1	13.06 (12.53)	12.04 (11.53)	13.99 (13.45)	159.10 (159.37)	21.24 (20.76)	19.31 (18.78)
2	3.78 (3.23)	3.84 (3.31)	6.91 (6.38)	41.70 (41.34)	9.63 (9.10)	6.82 (6.28)
3	2.31 (1.74)	2.61 (2.05)	5.71 (5.18)	23.12 (22.59)	7.42 (6.91)	4.69 (4.15)
5	1.55 (0.93)	2.01 (1.43)	5.11 (4.58)	15.43 (14.98)	6.18 (5.64)	3.63 (3.09)
10	1.18 (0.46)	1.75 (1.15)	4.86 (4.33)	12.65 (12.18)	5.62 (5.08)	3.18 (2.62)

λ_1	$n = 15$					
	NSN-LR chart	NSN-DS chart	\bar{X} chart	S chart	max chart	distance chart
0.3	73.57 (73.20)	74.68 (74.47)	66.56 (66.18)	397.62 (396.70)	101.11 (100.47)	101.42 (100.85)
0.5	24.61 (24.10)	24.98 (24.46)	23.82 (23.17)	320.89 (321.72)	36.77 (36.22)	37.22 (36.61)
1	4.56 (4.02)	4.80 (4.25)	6.19 (5.67)	90.72 (89.86)	8.74 (8.27)	7.58 (7.06)
2	1.46 (0.82)	1.62 (1.00)	2.77 (2.21)	17.09 (16.60)	3.43 (2.88)	2.28 (1.69)
3	1.11 (0.35)	1.23 (0.53)	2.24 (1.66)	9.09 (8.55)	2.55 (1.98)	1.57 (0.95)
5	1.01 (0.11)	1.07 (0.28)	1.99 (1.40)	6.12 (5.58)	2.11 (1.52)	1.28 (0.59)
10	1.00 (0.01)	1.02 (0.16)	1.88 (1.29)	5.10 (4.56)	1.92 (1.32)	1.17 (0.45)

λ_1	$n = 25$					
	NSN-LR chart	NSN-DS chart	\bar{X} chart	S chart	max chart	distance chart
0.3	35.52 (34.95)	36.60 (36.09)	34.95 (34.50)	373.78 (375.17)	53.70 (53.19)	54.59 (53.98)
0.5	9.49 (9.00)	9.99 (9.50)	10.34 (9.88)	237.89 (238.12)	15.01 (14.54)	15.22 (14.70)
1	1.80 (1.21)	1.93 (1.34)	2.41 (1.85)	40.32 (39.85)	3.01 (2.47)	2.58 (2.02)
2	1.02 (0.16)	1.05 (0.22)	1.26 (0.58)	5.76 (5.20)	1.33 (0.66)	1.10 (0.33)
3	1.00 (0.03)	1.00 (0.06)	1.13 (0.38)	3.21 (2.64)	1.12 (0.36)	1.01 (0.11)
5	1.00 (0.00)	1.00 (0.01)	1.07 (0.28)	2.35 (1.77)	1.04 (0.20)	1.00 (0.03)
10	1.00 (0.00)	1.00 (0.00)	1.05 (0.23)	2.06 (1.46)	1.02 (0.13)	1.00 (0.01)

the \bar{X} chart, the S chart, the max chart, and the distance chart) in the cases of $\lambda_1 = 0.3, 0.5, 1, 2, 3, 5, 10$, and $n = 5, 10, 15, 25$. From Table 4, we see that our proposed NSN-LR chart and NSN-DS chart outperform the four traditional charts in most of the cases, except for very small shift in asymmetry parameter. To be precise, when λ_1 is very small, the \bar{X} chart is slightly more effective when sample size n is also small. Further, we observe that, as the test sample size n increases, our proposed NSN-LR and NSN-DS schemes become very competitive to the traditional \bar{X} chart even for small shift in asymmetry parameter. In general, among these traditional alternatives, the \bar{X} chart performs the best when λ_1 is small and then the distance chart supersedes the \bar{X} chart when λ_1 gets larger. The max chart performs similarly as the distance chart. The S chart performs the worst compared to the other schemes, probably due to the decreasing variance. The performance of these traditional charts are, however, better than the other four charts introduced in this paper based on various measures of skewness. We further notice that, when the test sample size is large enough and the shift in the asymmetry parameter is also very large, that is, both the values of n and λ_1 are relatively large, all four traditional schemes considered here display commanding performance similar to the NSN-LR chart or the NSN-DS chart.

In summary, we can conclude that our proposed NSN-LR chart and NSN-DS chart have some distinct advantages in detecting a shift when the process distribution deviates from normal to skew-normal, specially when λ_1 is moderate-to-large. Otherwise, one may simply apply the traditional alternative, like \bar{X} chart for detecting shifts in the process mean. Nevertheless, using \bar{X} chart may be misleading in practice as it is designed for capturing a shift in the mean of a normally distributed process. It may not reflect the actual phenomenon, that is, the shift has taken place in the distribution itself. It may not be realized that the assignable cause has actually led to a disruption of symmetry of the process distribution. The effect of shift in asymmetry parameter would be confounded if we use any of the traditional charts. This clarifies the motivation behind developing the NSN-type control charts.

5. APPLICATION TO A MANUFACTURING PROCESS

In this section, we revisit the real example of a cork stopper's process production presented by Figueiredo and Gomes [9]. Figueiredo and Gomes [9] considered a consecutive sample of size $n = 1000$, related to corks' diameters as well as some other measurements from the production process. They applied the Shapiro test for normality and the Kolmogorov-Smirnov (K-S) test for goodness of fit of the ASN distribution on the 1000 data points. They noted that the Shapiro test rejects the normality of the diameter data at 5% level of significance, but the K-S test accepts the ASN distribution as a decent model for the diameter data.

They concluded accordingly that the ASN distribution may be considered to model the diameter data instead of the normal distribution.

Nevertheless, we revisit the diameter data set, and observe that during the initial stage of production, the underlying data distribution appears to be normal. We note that for the first 200 observations on corks' diameter, the p-value of the Shapiro test is very high and is 0.9296. This finding strongly supports the normality assumption (also see Figure 1) for the initial stage of production. We also notice that the process distribution deviates from normality and gradually becomes skew-normal (the p-value of the Shapiro test for normality gradually becomes lower and soon becomes less than 1%) as the production continues, probably due to unobservable occurrence of one or more assignable cause(s) at some point of time. Hence, we may argue that the process distribution has deviated from the normality and tends to follow an ASN distribution with some non-zero asymmetry parameter.

In this context, we illustrate the implementation of the proposed Shewhart-type NSN-LR and NSN-DS charts for monitoring the diameter data observed from the cork stopper's process production. We take the first 200 observations related to corks' diameter as the IC sample which is also referred to as the Phase I observations in literature. We obtain the estimates for the mean value and the standard deviation as 24.0695 and 0.1459 respectively. We use these estimates to approximate the true process parameters. The following 800 observations may be regarded as the Phase II data that consists of $m = 40$ subgroups each of size $n = 20$. For $n = 20$ and a target $IC-ARL$ of 370, we see from Table 1, the control limits for the two charts are, respectively, $H_{LR} = 9.6034$ and $H_{DS} = 0.3707$. Hereafter, we compute the LR and DS statistics for these 40 subgroups and plot them in Figure 3 and 4, along with the respective UCL.

We see that the movement of the plotting statistics in these two charts are very similar in nature. We receive the first signal at the 17th test sample for both charts. Moreover, several points fall above the UCL in both these charts. We may consider this as a strong evidence of deviation of process distribution and therefore, may conclude that the initial assumption of normally distributed process is no longer valid and the process distribution becomes asymmetric. To be precise, the ASN distribution (with some non-zero asymmetry parameter) emerges as the new process distribution. Since in the whole data set, normally distributed IC data have been contaminated (mixed) with the shifted data which are generated from the OOC process, this easily leads to an illusion that the data inherently follows an ASN distribution.

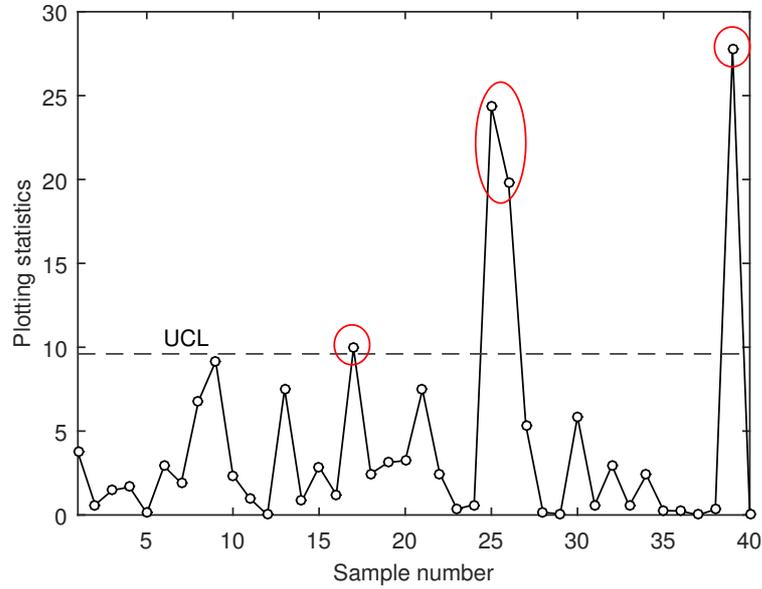


Figure 3: The NSN-LR chart for the corks' diameter data.

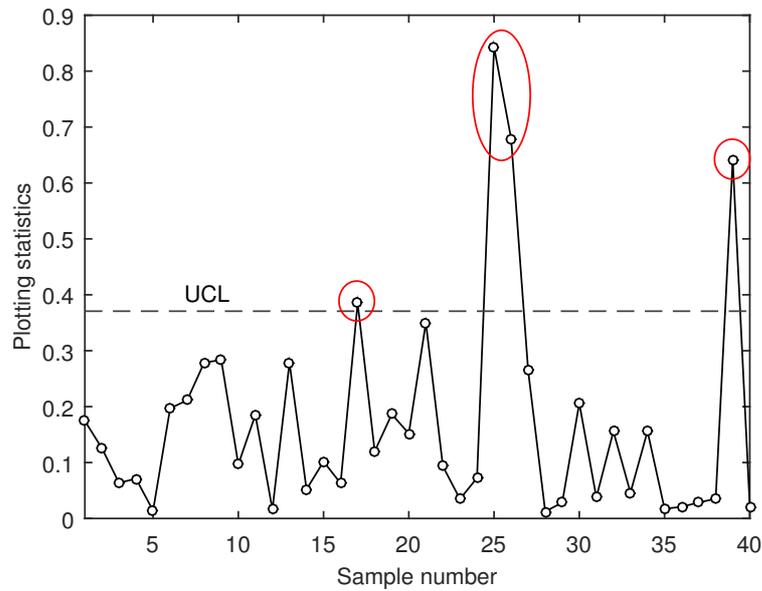


Figure 4: The NSN-DS chart for the corks' diameter data.

6. CONCLUDING REMARKS

In this paper, we study on the statistical process monitoring problem in regard to detecting a shift from normal to skew-normal. A class of possible test statistics are thoroughly examined and we find that the proposed monitoring procedures based on the likelihood ratio statistic and the sample distance skewness statistic operate most competitively, especially the latter. Therefore, these two approaches are expected to be very useful in practice to monitor the asymmetry parameter of an ASN distribution and to detect a process shift from normality to skew-normal with $\lambda \neq 0$.

In the present context, we study the performance of the standard Shewhart-type version for all charts. It is well-known that the Shewhart-type charts are usually good for detecting large and abrupt shifts in a process, however, the change in skewness actually is relatively small even if λ_1 gets very large for the ASN distribution. Thus, a straightforward extension of the proposed monitoring schemes under the EWMA or CUSUM set-up may be considered as a future research problem. Further, more researches on the economic and the economic-statistical design of the NSN-LR and NSN-DS charts are highly warranted in future.

In addition, as stated before, further research on simultaneous monitoring of all the parameters (location, scale and shape) of the ASN distribution needs to be studied in detail. It will also be an interesting future research problem to develop process monitoring schemes when the parameters are unknown and estimated from the reference sample. Clearly, the present work may lead to some interesting future research problems.

ACKNOWLEDGMENTS

The authors would like to thank Professors Fernanda Figueiredo of Universidade do Porto and M. Ivette Gomes of Universidade de Lisboa for their kind help on supporting the industrial process data. The authors are also grateful to the anonymous reviewer for providing helpful suggestions and comments for revising this paper. The collaborative work described in this paper was supported by Research Grant Council (G-CityU108/14) and University Grants Council of Hong Kong (GRF 11213116) and National Natural Science Foundation of China (No. 71371163, 71371151).

REFERENCES

- [1] ARELLANO-VALLE, R.B.; GÓMEZ, H.W. and QUINTANA, F.A. (2004). A new class of skew-normal distributions, *Communications in Statistics – Theory and Methods*, **33**, 1465–1480.
- [2] AZZALINI, A. (1985). A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics*, **12**, 171–178.
- [3] AZZALINI, A. (1986). Further results on a class of distributions which includes the normal ones, *Statistica*, **46**, 199–208.
- [4] AZZALINI, A. (2005). The skew-normal distribution and related multivariate families, *Scandinavian Journal of Statistics*, **32**, 159–188.
- [5] BARTOLETTI, S. and LOPERFIDO, N. (2009). Modelling air pollution data by the skew-normal distribution, *Stochastic Environmental Research and Risk Assessment*, **24**, 513–517.
- [6] CASTAGLIOLA, P. (2005). A new S2-EWMA control chart for monitoring the process variance, *Quality and Reliability Engineering International*, **21**, 781–794.
- [7] CHEN, J.T.; GUPTA, A.K. and NGUYEN, T.T. (2004). The density of the skew normal sample mean and its applications, *Journal of Statistical Computation and Simulation*, **74**, 487–494.
- [8] FERREIRA, J.T.A.S. and STEEL, M.F.J. (2006). A constructive representation of univariate skewed distributions, *Journal of the American Statistical Association*, **101**, 823–829.
- [9] FIGUEIREDO, F. and GOMES, M.I. (2013). The skew-normal distribution in SPC, *REVSTAT – Statistical Journal*, **11**, 83–104.
- [10] FIGUEIREDO, F. and GOMES, M.I. (2015). The role of asymmetric families of distributions in eliminating risk, *Theory and Practice of Risk Assessment*, **136**, 267–277.
- [11] FRUHWIRTH-SCHNATTER, S. and PYNE, S. (2010). Bayesian inference for finite mixtures of univariate and multivariate skew-normal and skew-t distributions, *Biostatistics*, **11**, 317–336.
- [12] GENTON, M.G.; HE, L. and LIU, X. (2001). Moments of skew-normal random vectors and their quadratic forms, *Statistics & Probability Letters*, **51**, 319–325.
- [13] GÓMEZ, H.W.; CASTRO, L.M.; SALINAS, H.S. and BOLFARINE, H. (2010). Properties and inference on the skew-curved-symmetric family of distributions, *Communications in Statistics – Theory and Methods*, **39**, 884–898.
- [14] GUO, B. and WANG, B.X. (2015). The design of the ARL-unbiased S2 chart when the in-control variance is estimated, *Quality and Reliability Engineering International*, **31**, 501–511.
- [15] HAWKINS, D.M. and DENG, Q. (2009). Combined charts for mean and variance information, *Journal of Quality Technology*, **41**, 415–425.
- [16] HENZE, N. (1986). A probabilistic representation of the ‘skew-normal’ distribution, *Scandinavian Journal of Statistics*, **13**, 271–275.

- [17] HOSKING, J.R.M. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics, *Journal of the Royal Statistical Society: Series B*, **52**, 105–124.
- [18] KHOO, M.B.C.; WONG, V.H.; WU, Z. and CASTAGLIOLA, P. (2012). Optimal design of the synthetic chart for the process mean based on median run length, *IIE Transactions*, **44**, 765–779.
- [19] KNOTH, S. (2015). Run length quantiles of EWMA control charts monitoring normal mean or/and variance, *International Journal of Production Research*, **53**, 4629–4647.
- [20] LI, C.; MUKHERJEE, A.; SU, Q. and XIE, M. (2016). Design and implementation of two CUSUM schemes for simultaneously monitoring the process mean and variance with unknown parameters, *Quality and Reliability Engineering International*, **32**, 2961–2975.
- [21] LI, C.-I.; SU, N.-C.; SU, P.-F. and SHYR, Y. (2014). The design of X-bar and R control charts for skew normal distributed data, *Communications in Statistics – Theory and Methods*, **43**, 4908–4924.
- [22] LI, Z.; ZOU, C.; WANG, Z. and HUWANG, L. (2013). A multivariate sign chart for monitoring process shape parameters, *Journal of Quality Technology*, **45**, 149–165.
- [23] MAMELI, V. and MUSIO, M. (2013). A generalization of the skew-normal distribution: The beta skew-normal, *Communications in Statistics – Theory and Methods*, **42**, 2229–2244.
- [24] MCCracken, A.K.; CHAKRABORTI, S. and MUKHERJEE, A. (2013). Control charts for simultaneous monitoring of unknown mean and variance of normally distributed processes, *Journal of Quality Technology*, **45**, 360–376.
- [25] MUKHERJEE, A.; ABD-ELFATTAH, A.M. and PUKAIT, B. (2013). A rule of thumb for testing symmetry about an unknown median against a long right tail, *Journal of Statistical Computation and Simulation*, **84**, 2138–2155.
- [26] PENG, Y.; XU, L. and REYNOLDS, M.R., JR. (2015). The design of the variable sampling interval generalized likelihood ratio chart for monitoring the process mean, *Quality and Reliability Engineering International*, **31**, 291–296.
- [27] RAHMAN, S. and HOSSAIN, F. (2008). A forensic look at groundwater arsenic contamination in Bangladesh, *Environ Forensics*, **9**, 364–374.
- [28] RAZZAGHI, M. (2014). A hierarchical model for the skew-normal distribution with application in developmental neurotoxicology, *Communications in Statistics – Theory and Methods*, **43**, 1859–1872.
- [29] REYNOLDS, M.R., JR.; LOU, J.; LEE, J. and WANG, S. (2013). The design of GLR control charts for monitoring the process mean and variance, *Journal of Quality Technology*, **45**, 34–60.
- [30] ROSS, G.J. and ADAMS, N.M. (2012). Two nonparametric control charts for detecting arbitrary distribution changes, *Journal of Quality Technology*, **44**, 102–116.
- [31] RYU, J.-H.; WAN, H. and KIM, S. (2010). Optimal design of a CUSUM chart for a mean shift of unknown size, *Journal of Quality Technology*, **42**, 311–326.
- [32] SHEU, S.-H.; HUANG, C.-J. and HSU, T.-S. (2012). Extended maximum generally weighted moving average control chart for monitoring process mean and variability, *Computers & Industrial Engineering*, **62**, 216–225.

- [33] SHU, L.; YEUNG, H.-F. and JIANG, W. (2010). An adaptive CUSUM procedure for signaling process variance changes of unknown sizes, *Journal of Quality Technology*, **42**, 69–85.
- [34] SU, N.-C.; CHIANG, J.-Y.; CHEN, S.-C.; TSAI, T.-R. and SHYR, Y. (2014). Economic design of two-stage control charts with skewed and dependent measurements, *The International Journal of Advanced Manufacturing Technology*, **73**, 1387–1397.
- [35] SU, N.-C. and GUPTA, A.K. (2015). On some sampling distributions for skew-normal population, *Journal of Statistical Computation and Simulation*, **85**, 3549–3559.
- [36] TSAI, T.-R. (2007). Skew normal distribution and the design of control charts for averages, *International Journal of Reliability, Quality and Safety Engineering*, **14**, 49–63.
- [37] TSIAMYRTZIS, P. and HAWKINS, D.M. (2005). A Bayesian scheme to detect changes in the mean of a short-run process, *Technometrics*, **47**, 446–456.
- [38] VINCENT, R. and WALSH, T.D. (1997). Quantitative measurement of symmetry in CBED patterns, *Ultramicroscopy*, **70**, 83–94.
- [39] WU, Z.; YANG, M.; KHOO, M.B.C. and YU, F.J. (2010). Optimization designs and performance comparison of two CUSUM schemes for monitoring process shifts in mean and variance, *European Journal of Operational Research*, **205**, 136–150.
- [40] ZHANG, G. (2014). Improved R and s control charts for monitoring the process variance, *Journal of Applied Statistics*, **41**, 1260–1273.
- [41] ZOU, C. and TSUNG, F. (2010). Likelihood ratio-based distribution-free EWMA control charts, *Journal of Quality Technology*, **42**, 174–196.

ON THE DESIGN POINTS FOR A ROTATABLE ORTHOGONAL CENTRAL COMPOSITE DESIGN

Author: CHRISTOS P. KITSOS
– Department of Informatics,
Technological Educational Institute of Athens,
Greece
xkitsos@teiath.gr

Received: September 2016 Revised: December 2016 Accepted: December 2016

Abstract:

- The aim of this paper is to adjust the existing believe on the condition that a Rotatable Orthogonal Central Composite Design can be created from an Orthogonal Central Composite Design. Therefore the appropriate introduction is discussed concerning the Rotatable Orthogonal Central Composite Design. Then, the main result is presented, i.e. The necessary and sufficient conditions in order an Orthogonal Central Composite Design to be Rotatable.

Key-Words:

- *rotatable design; orthogonal design; central design; composite design.*

AMS Subject Classification:

- 62K20, 62K10, 62K15.

1. INTRODUCTION

When we are constructing designs to explore the overall response surface, rather than the response to individual factors, the main request is for optimum position, i.e. the combination of factor levels for which the expected response is maximized. In principle, the design of an experiment to explore the response surface will cover — eventually — only a region of the unknown surface, for which a rather known center exists. Therefore observations at the center are important. Moreover, blocking and replication are always important, as for any experiment. Therefore, the Response Surface Method (RSM) is based on this framework. As an example, there is a nice experiment to investigate the effect of the levels of two additives on the quality of a cake production process in [7]. In addition to the factorial experiment portion as well as to the number of observations at the center, additional points are added to the design of each factor or, equivalently, to each axis, known as axial (or star).

Therefore, a Central Composite Design (CCD) is: an experimental fractional design which is supplemented by additional experimental points such as center points and star points. In principle, we need the main effects to be separately estimated. That is why the sense of Orthogonal design is essential: a design in which the given variables (or a linear combination of them) are regarded as statistical independent. In RSM it is important to choose the “path” towards the optimum. Thus, the steepest accent method is applied from the Numerical Analysis (see [5] among others), while the sequential principle of design is adopted. Therefore, the Rotatable design is a real need to be defined, as the one which has equal predictive power (predictiveness) in all directions from the center point and from the points that are at equidistant from the center.

The aim of this paper is to discuss, and eventually adjust, the necessary and sufficient conditions, for the number of design points, to be satisfied so that to estimate the design points for an Orthogonal Central Composite Design (OCCD) to be rotatable (ROCCD). The problem has been discussed extensively by [3], [7], [6], [8], where the conditions for a Central Composite Design (CCD) to be OCCD and ROCCD are reviewed and examined. We shall refer and extend/adjust the condition for a ROCCD, as appeared in [6, p. 304] and [7, p. 550]. The improvement is that the imposed already condition is now used for obtaining real solutions, when positive integer solution is actually needed. To the best of our knowledge, we have not see any attempts trying to adjust this conclusion. This adjustment is our contribution, so that the experimenter can work to Response Surface Methods, with a number of up to 14 input variables, as we are providing the appropriate calculations, based on the developed theory. For a compact form of the obtained calculations see Table 1. Through out this paper the standard notation for the Response Surface Methods is adopted; see [6] and [8].

2. CONDITIONS FOR A ROCCD

The Central Composite Design (CCD) appears an aesthetic appeal within the class of the second order response surface design. It was introduced in the pioneering paper of [4]. In principle the Central Composite Design (CCD) can always be constructed as a two block Orthogonal (OCCD). This is based in two blocks: the factorial portion and the star portion. The first block is based on N_F factorial points and N_{CF} center points. The second block is based on N_A axial points and N_{CA} center points for the star portion. It has been traditionally denoted by (a) the distance of the star points from the center of the design. For a factorial or fractional factorial experiment $N_F = 2^q$ or 2^{k-q} observations are needed. Consider $N_A = 2k$ points with k being the number of input variables and q such that $0 < k < q$.

For Orthogonal blocking, in two blocks in a CCD, the fractional of the total sum of squares, of each input variable contributed by every block, has to be equal to the fraction of the total observations allotted to the block. It is, for each block,

$$(2.1a) \quad \frac{N_F}{N_F + 2a^2} = \frac{N_F + N_{CF}}{N} \quad \text{and}$$

$$(2.1b) \quad \frac{2a^2}{N_F + 2a^2} = \frac{N_A + N_{CA}}{N},$$

respectively. The total number of observation is then $N = N_F + N_A + N_C$, with N_C the number of central points i.e. $N_C = N_{CF} + N_{CA}$. Then from (2.1a) and (2.1b) it is

$$(2.2) \quad a^2 = \frac{N_F(2k + N_{CA})}{2(N_F + N_{CF})}.$$

When the design is required to be also a rotatable one, then [6, p. 304],

$$(2.3) \quad a^2 = N_F^{1/2}.$$

From (2.2) and (2.3) the second degree equation

$$(2.4) \quad 2N_F - N_F^{1/2}(2k + N_{CA}) + 2N_{CF} = 0,$$

has to be satisfied, see also [7, p. 550]. Both [7] and [6] note that is not always possible to find a design that satisfies (2.4). Moreover, in [6] is provided as a necessary condition for the satisfaction of equation (2.4) the relation

$$(2.5) \quad D = (2k + N_{CA})^2 - 16N_{CF} \geq 0,$$

But the discriminant D positive means that (2.4) has real roots. And we are looking for positive integers as solution of (2.4). This is the crucial point. The necessary and sufficient conditions, so that the number of the design points N_F for a ROCCD needs more investigation and we are providing this investigation in section 3.

3. INTEGER SOLUTION FOR A ROCCD

Trivially the coefficient of the second order equation (2.4) are asked to be positive integers and not just real numbers. We state and prove in Appendix A the following Theorem which shall help us to develop the line of thought tackling the problem, see Theorem 3.1 and Proposition 3.1.

Theorem 3.1. *Consider the second order equation*

$$(3.1) \quad Ax^2 + Bx + C = 0,$$

with A , B and C integers. Then the roots of (3.1) are integers if and only if

1. A divides B , i.e. $A|B$,
2. A divides C , i.e. $A|C$, and
3. The discriminant D of (3.1) is a square, i.e. $D = \mu^2 \in \mathbb{Z}$.

Now, consider (2.4). The following theorem holds.

Theorem 3.2. *The necessary and sufficient conditions in order the equation*

$$(3.2) \quad 2N_F - N_F^{1/2}(2k + N_{CA}) + 2N_{CF} = 0,$$

to have positive integer solutions are:

$$(3.3a) \quad 2|N_{CA}, \text{ i.e. } 2 \text{ divides } N_{CA}, \text{ i.e. is even,}$$

$$(3.3b) \quad D = (2k + N_{CA})^2 - 16N_{CF} = \mu^2, \quad \mu \in \mathbb{Z}.$$

Proof: Trivially, if we let $x = N_F^{1/2}$, (3.2) is then reduced to (3.1) with $A = 2$, $B = -(2k + N_{CA})$, $C = 2N_{CF}$. Therefore condition (1) of Theorem 3.1 is reduced to (3.3a), (2) holds, and (3) is reduced to (3.3b). Thus, the roots are integer numbers. Moreover the sum of roots of (3.2) is $(2k + N_{CA}) > 0$ and the product of roots is $2N_{CF} > 0$. Therefore, the integer roots are positive. Eventually the conditions (3.3a) and (3.3b) are necessarily and sufficient to have positive integer roots. \square

Practically the factorial portion consists from a 2^k design. Usually k is not greater than 6, otherwise a portion 2^{k-p} is used. We investigate next the cases of k up to 14 in the following Proposition.

Proposition 3.1. Consider (3.2) and the case that $N_F = 2^k$. Then, an integer solution exists for k even. Moreover, for

$$(3.4) \quad k = 2, 4, 6, 8, 10, 12, 14,$$

the relation between N_{CF} and N_{CA} should be of the form

$$(3.5) \quad N_{CF} = 2^v(N_{CA} - K), \quad v = \frac{1}{2}k - 1,$$

and $K = K(k)$

$$(3.6) \quad K = 0, 0, 4, 16, 44, 104, 228,$$

respectively.

Proof: Trivially k has to be an even integer otherwise there is no integer solution. Thus, $2^{k/2}$ has to be integer. Now, from (3.2), we obtain:

$$\text{For } k = 2: \quad 2^3 - 2(4 + N_{CA}) + 2N_{CF} = 0, \quad \text{i.e. } N_{CF} = N_{CA} = 2^0(N_{CA} - 0).$$

$$\text{For } k = 4: \quad N_{CF} = 2N_{CA} = 2^1(N_{CA} - 0).$$

$$\text{For } k = 6: \quad N_{CF} = 4N_{CA} - 16 = 2^2(N_{CA} - 4).$$

In order that both N_{CF} and N_{CA} be positive integers, it is required to be $N_{CF} = 4p$, $p \in \mathbb{Z}^+$ as

$$\frac{1}{4}N_{CF} + 4 = N_{CA}.$$

Therefore for $k = 6$,

$$N_{CF} = 4p, \quad N_{CA} = p + 4.$$

For $k = 8$ it is from (3.2)

$$2^8 - 2^3(16 + N_{CA}) + N_{CF} = 0, \quad \text{i.e.}$$

$$N_{CF} = 8N_{CA} - 128 = 2^3(N_{CA} - 16).$$

Now, in order that both N_{CA} and N_{CF} are positive integers, N_{CA} should be an integer greater than 16, so $\frac{1}{8}N_{CF} + 16 = N_{CA}$ and therefore, $N_{CF} = 8p$, $N_{CA} = 8(p + 2)$. Thus the case $k = 8$ can rarely be of practical use.

For $k = 10$ it is from (3.2)

$$1024 - 320 - 16N_{CA} + N_{CF} = 0, \quad \text{i.e.}$$

$$N_{CF} = 2^4N_{CA} - 2^444 = 2^4(N_{CA} - 44).$$

There is no practical use for the case $k = 10$ as well as in order N_{CF} to be positive N_{CA} has to be greater than 44 i.e.

$$N_{CF} = 2^4p, \quad N_{CA} = p + 44.$$

Similar, for $k = 12$, it is

$$\begin{aligned} N_{\text{CF}} &= 32N_{\text{CA}} - 3328 = 2^5(N_{\text{CA}} - 104), \\ N_{\text{CF}} &= 2^5p, \quad N_{\text{CA}} = p + 104. \end{aligned}$$

For $k = 14$ it is then

$$\begin{aligned} N_{\text{CF}} &= 64N_{\text{CS}} - 14592 = 2^6(N_{\text{CA}} - 228), \quad \text{i.e.} \\ N_{\text{CF}} &= 2^6p, \quad N_{\text{CA}} = p + 228. \end{aligned} \quad \square$$

There is no practical use to investigate grater values, as the number of observations turns to be very large in such a case. From the above discussion it is easy to see that the following holds.

Corollary 3.1. *In principle,*

$$(3.7) \quad N_{\text{CF}} = 2^{k/2-1} \left[N_{\text{CA}} - (2^{k/2-1} - 2k) \right], \quad k = 2, 4, 6, \dots$$

Corollary 3.2. *The general form of required samples are:*

$$(3.8) \quad N_{\text{CF}} = 2^{k/2-1}p, \quad N_{\text{CA}} = p + 2^{k/2+1} - 2k, \quad p \in \mathbb{Z}^+.$$

Proposition 3.2. *For the equation (3.2) as in Theorem 3.2, considering $k = 2(2)14$ the corresponding pair of values $(N_{\text{CA}}, N_{\text{CF}})$ for a double root $x = N_{\text{F}}^{1/2} = 2^{k/2}$, are*

$$(4, 4), (8, 16), (20, 64), (48, 256), (108, 1024), (232, 4096), (484, 16384).$$

Proof: The proof is based on (3.3b) with $\nu = 0$ and on the results obtained in Proposition 3.1. Namely for:

- $k = 2$, $D = (4 - N_{\text{CA}})^2$, hence $N_{\text{CA}} = 4 = N_{\text{CF}}$;
- $k = 4$, $D = (8 + N_{\text{CA}})^2 - 32N_{\text{CA}} = (8 - N_{\text{CA}})^2$, hence $N_{\text{CA}} = 8$, $N_{\text{CF}} = 16$;
- $k = 6$, $D = (12 + N_{\text{CA}})^2 - 16(4N_{\text{CA}} - 16) = (20 - N_{\text{CA}})^2$,
so $N_{\text{CA}} = 20$, $N_{\text{CF}} = 64$;
- $k = 8$, $D = (16 + N_{\text{CA}})^2 - 16(8N_{\text{CA}} - 128) = (48 - N_{\text{CA}})^2$,
so $N_{\text{CA}} = 48$, $N_{\text{CF}} = 256$;
- $k = 10$, $D = (20 + N_{\text{CA}})^2 - 16(16N_{\text{CA}} - 704) = (108 - N_{\text{CA}})^2$,
so $N_{\text{CA}} = 108$, $N_{\text{CF}} = 1024$;
- $k = 12$, $D = (24 + N_{\text{CA}})^2 - 16(32N_{\text{CA}} - 3328) = (232 - N_{\text{CA}})^2$,
so $N_{\text{CA}} = 232$, $N_{\text{CF}} = 4096$;
- $k = 14$, $D = (28 + N_{\text{CA}})^2 - 16(64N_{\text{CA}} - 14592) = (484 - N_{\text{CA}})^2$,
so $N_{\text{CA}} = 484$, $N_{\text{CF}} = 16384$. □

It is clear that for k greater than 8 there is no practical use, as we have already comment, as the required observations are too many and it is not practical use of an experiment 2^8 .

4. DISCUSSION

The above provided analysis proves that the restriction $D \geq 0$ is not the appropriate one for an OCCD to be ROCCD. In Table 1 we summarize, for practical use values of k and the appropriate values of design points, according to the above-mentioned calculations. The appropriate necessary and sufficient condition was stated and proved, adjusting an old wrong result, with a rather “simple” approach. Some experimenters decide in advance, rather from experience or depending on how easy is to perform the experiment, the needed size of the experiment. But the investigation needs a deeper approach, we believe, with not such a difficult mathematical approach for the experimenter. We worked towards this direction: to keep it simple. Table 1 summarizes the results from the above discussion. In a future attempt, it would be interesting to construct, mainly from a theoretical point of view, the appropriate calculations with k larger than 14, in order to see the behavior of the discussed “system” for “large” values. It is also clear that the researcher working at EVOP designs (see the pioneering paper in [2]) can adopt the calculations performed here, for the initial design, as EVOP is based, briefly speaking, on a “factorial + centre” design. Therefore, despite its theoretical framework and background, the above proposed integer solution can be very helpful in practice as well.

Table 1: Design points needed for a ROCCD with $k = 2(2)8$, for a double root $x = N_F^{1/2}$.

k	N_F	N_A	N_{CF}	N_{CA}	N_C	N
4	4	4	2	2	4	12
			3	3	6	14
			4	4	8	16
4	16	8	4	2	6	30
			6	3	9	33
			8	4	12	36
			16	8	24	48
6	64	12	4	5	9	85
			8	6	14	90
			12	7	19	95
			64	20	84	160
6	256	16	8	17	25	297
			16	18	34	306
			24	19	34	306
			256	48	304	576

APPENDIX A

Proposition A.1. *Let $x^2 + px + q$, $p, q \in \mathbb{Z}$. Then its roots $x_1, x_2 \in \mathbb{Z}$ if and only if the discriminant $D = \mu^2$, $\mu \in \mathbb{Z}$, or $\mu = 0$.*

Proof: If $x_1, x_2 \in \mathbb{Z}$: $x^2 + px + q \Leftrightarrow x^2 - (x_1 + x_2)x + x_1x_2 = 0$.

Let $D = (x_1 + x_2)^2 - 4x_1x_2 = (x_1 - x_2)^2 = \mu^2$, with $\mu = x_1 - x_2 \in \mathbb{Z}$ as $x_1, x_2 \in \mathbb{Z}$. Now, let $D = p^2 - 4q = \mu^2$, $\mu \in \mathbb{Z}$. Then

$$x_1, x_2 = \frac{-p \pm \mu}{2} = -\frac{p \mp \mu}{2}.$$

But: $p^2 - 4q = \mu^2 \Rightarrow p^2 - \mu^2 = 4q \Leftrightarrow (p - \mu)(p + \mu) = 4q \Rightarrow p + \mu = 2n_1$, $n_1 \in \mathbb{Z}$ and $p - \mu = 2n_2$, $n_2 \in \mathbb{Z}$. Therefore $x_1 = n_1$ and $x_2 = -n_2$, i.e. x_1 and x_2 are integers. \square

Proof of Theorem 3.1: If $A|B$ and $A|C$ then:

$$Ax^2 + Bx + C = 0 \Leftrightarrow x^2 + \frac{B}{A}x + \frac{C}{A} = 0 \Leftrightarrow x^2 + px + q = 0, \quad p, q \in \mathbb{Z}.$$

It is also: $D_1 = p^2 - 4q = \mu_1^2$, $\mu_1 \in \mathbb{Z} \Leftrightarrow$

$$\frac{B^2}{A^2} - 4\frac{C}{A} = \mu_1^2 \Leftrightarrow \frac{B^2 - 4AC}{A^2} = \mu_1^2 \Rightarrow D = B^2 - 4AC = (\mu_1 A)^2 = \mu^2, \quad \mu \in \mathbb{Z}.$$

So $x^2 + px + q = 0$ has integer roots and so does $Ax^2 + Bx + C = 0$.

The inverse: Let $x_1, x_2 \in \mathbb{Z}$ be the roots of $Ax^2 + Bx + C = 0$. Then: $x_1 + x_2 = -\frac{B}{A}$ and $x_1x_2 = \frac{C}{A}$. Thus $x_1 + x_2 \in \mathbb{Z} \Rightarrow A|B$ and $x_1x_2 \in \mathbb{Z} \Rightarrow A|C$. Moreover:

$$Ax^2 + Bx + C = 0 \Rightarrow x^2 + \frac{B}{A}x + \frac{C}{A} = 0 \Rightarrow x^2 + px + q = 0,$$

has integer roots (Proposition A.1).

Let $D_1 = p^2 - 4q = \mu_1^2$, $\mu_1 \in \mathbb{Z}$, i.e.

$$\frac{B^2}{A^2} - 4\frac{C}{A} = \mu_1^2 \Rightarrow \frac{B^2 - 4AC}{A^2} = \mu_1^2 \Rightarrow D = B^2 - 4AC = (A\mu_1)^2 = \mu^2, \quad \mu \in \mathbb{Z}.$$

\square

ACKNOWLEDGMENTS

I would like to thank the referee for his/her valuable comments, which improve the final version of this paper.

REFERENCES

- [1] BAILEY, R.A. (1982). The decomposition of treatment degrees of freedom in quantitative factorial experiment, *JRSS B*, **44**, 63–70.
- [2] BOX, G.E.P. (1957). Evolutionary operation: a method for increasing industrial productivity, *Appl. Stat.*, **6**, 81–101.
- [3] BOX, G.E.P. and DRAPER, N.R. (1987). *Empirical Model – Building and Response Surfaces*, John Wiley, New York.
- [4] BOX, G.E.P. and WILSON, K.B. (1951). On the experimental attainment of optimum conditions, *JRSS B*, **13**, 1–38 (Discussion 38–45).
- [5] HAMMING, R.W. (1986). *Numerical Methods for Scientists and Engineers*, Dover.
- [6] KHURI, A.I. and CORNELL, J.A. (1996). *Response surfaces: Designs and Analyses*, Marcel Dekker.
- [7] MONTGOMERY, D.C. (1991). *Design and Analysis of Experiments*, John Wiley, Singapore.
- [8] MYERS, R.H. and MONTGOMERY, D.C. (1995). *Response Surface Methodology*, John Wiley.

RANDOM ENVIRONMENT INAR MODELS OF HIGHER ORDER

Authors: ALEKSANDAR S. NASTIĆ
– Department of Mathematics, University of Niš, Serbia
anastic78@gmail.com

PETRA N. LAKETA
– Department of Mathematics, University of Niš, Serbia
petra.laketa@pmf.edu.rs

MIROSLAV M. RISTIĆ
– Department of Mathematics, University of Niš, Serbia
miristic72@gmail.com

Received: February 2016 Revised: October 2016 Accepted: January 2017

Abstract:

- Two different random environment INAR models of higher order, $RrNGINAR_{\max}(p)$ and $RrNGINAR_1(p)$, are introduced. Both of them are of variable order, which is not a random variable. Each step order is defined using the random environment process. Properties of the defined models are analyzed in parallel. The conditional expectation and variance are calculated. The strongly consistent Yule-Walker estimators are defined and new modified estimators are given. Models are applied to the simulated and the real-life data and the results show benefits of the introduced models.

Key-Words:

- *random environment; INAR(p); RrNGINAR; negative binomial thinning; geometric marginals.*

AMS Subject Classification:

- 62M10.

1. INTRODUCTION

The integer-valued autoregressive (INAR) processes are introduced by McKenzie ([9]) and Al-Osh and Alzaid ([2]). They were the subject of research of many scientists, so there are a lot of different models which all intend to better describe the data obtained from some natural processes. The most of them are stationary, since this property gives some simplifications. Some of the models which investigate different thinning operators are given in [3], [8], [17, 18] and [14]. Models with various marginal distributions can be found in [10], [1], [5] and [6]. Weiß ([16]) and Nastić and Ristić ([11]) considered mixed processes. The first combined INAR(p) process is introduced in [16] and the combined process which is important for this paper is CGINAR(p) from [12]. It is combined in the sense that in every step recursive formula for an element of the process has one of the p possible forms (which match with formulas for INAR process) with some given probabilities.

We can say that stationary processes are rigid, because some of their properties are conserved in time. However, the real data are not often like that. One of the models which improves this weakness is RrNGINAR(1), defined in [13]. This is achieved by letting elements of the process to have varying distribution. Namely, quantitative properties observed from the nature depend on the environment. Since these values are represented by the elements of the process, it is natural to expect mentioned distribution to depend on the environment, too. It is supposed that environment conditions can be divided into r different types, which are called states, and each state is associated with a fixed distribution, so element of the process has the distribution of its state.

The main idea of this article is to make CGINAR(p) process more flexible, using the idea from RrNGINAR(1) process. Therefore, the aim is to construct a CGINAR(p) process with random states. So, in the second section of this article two different ways of constructing such a process which overcome problems that occur in classical (the most intuitive) way of defining this kind of model are discussed. Its correlation structure is analyzed in the third section. The fourth section is about Yule-Walker (YW) estimators of the parameters of the defined models. The quality of YW estimators is examined on the simulated data in Section 5. In the final section the introduced processes are applied to the real data and the results are compared for different models.

2. MODELS DEFINITIONS AND PROPERTIES

As mentioned earlier, our aim in this paper is to introduce the combined RrNGINAR process, where RrNGINAR process of order one is introduced in [13]. An attempt to construct this kind of combined process in the classical way, as it was done in [16], [12] and [15] will bring some difficulties, so new approaches will be used. In this section we will define two processes, which overcome this problem. Discussion about some of their properties will be given.

Let $E_r = \{1, 2, \dots, r\}$ be the set of all possible states, where $r \in \mathbb{N}$ and let $\{z_n\}$, $n \in \mathbb{N}_0$, be a realization of an r states random environment process $\{Z_n\}$ (we use Definition of the r random environment process given in [13]). For $i, j \in E_r$, let $\{\varepsilon_n(i, j)\}$, $n \in \mathbb{N}$, be sequences of independent identically distributed (i.i.d.) random variables. We will use notation $X_n(z_n)$ for an element of the new process, where z_n (which represents realized value of the random environment process in the moment $n \geq 0$) determines the distribution of that element. Let $\alpha*$ be the negative binomial thinning operator, for $\alpha \in (0, 1)$, with a counting sequence $\{U_i, i \geq 1\}$ of i.i.d. random variables with probability mass function (pmf) given by

$$P(U_i = u) = \frac{\alpha^u}{(1 + \alpha)^{u+1}}, \quad u = 0, 1, 2, \dots$$

As it was noted, it would be natural to introduce the combined random environment NGINAR process of order p in the following (classical) way

$$(2.1) \quad X_n(z_n) = \begin{cases} \alpha * X_{n-1}(z_{n-1}) + \varepsilon_n(z_{n-1}, z_n), & \text{w.p. } \phi_1, \\ \alpha * X_{n-2}(z_{n-2}) + \varepsilon_n(z_{n-2}, z_n), & \text{w.p. } \phi_2, \\ \vdots & \vdots \\ \alpha * X_{n-p}(z_{n-p}) + \varepsilon_n(z_{n-p}, z_n), & \text{w.p. } \phi_p, \end{cases}$$

for arbitrary $n \geq p$ and fixed $p \in \mathbb{N}$, where $\phi_i \geq 0$, $i \in \{1, 2, \dots, p\}$, $\sum_{i=1}^p \phi_i = 1$, where distribution of $X_n(z_n)$ is given by

$$P(X_n(z_n) = x) = \frac{\mu_{z_n}^x}{(1 + \mu_{z_n})^{x+1}}, \quad x = 0, 1, 2, \dots, \quad n = 0, 1, 2, \dots,$$

$\mu_{z_n} \in \{\mu_1, \mu_2, \dots, \mu_r\}$ is the parameter determined by the value z_n , $\mu_i > 0$, $i \in \{1, 2, \dots, r\}$ and where the next conditions are satisfied

- (A1) $\{Z_n\}$, $\{\varepsilon_n(1, 1)\}$, $\{\varepsilon_n(1, 2)\}$, ..., $\{\varepsilon_n(r, r)\}$, are mutually independent for all $n \geq 1$,
- (A2) $X_n(l)$ is independent of Z_m and $\varepsilon_m(i, j)$, for $0 \leq n < m$ and any $i, j, l \in E_r$.

If we try to derive the distribution of $\varepsilon_n(i, j)$, $i, j \in E_r$, using procedure similar to the one of CGINAR(p) process, it wouldn't be so easy. Actually, it

is not necessary that z_{n-j} for all $j = 1, 2, \dots, p$ are the same, so, consequently, $\varepsilon_n(z_{n-j}, z_n)$ do not have to be identically distributed for all $j = 1, 2, \dots, p$. This leads to a complex expression for the distribution of $\varepsilon_n(i, j)$, where i and j are arbitrary values from E_r .

The first method for avoiding this problem is to define $X_n(z_n)$ using (2.1), but substituting p with p_n , where p_n is the maximal number less or equal to the given value p ($p \in \mathbb{N}$ is a fixed number, not depending on n), which satisfies $z_{n-1} = \dots = z_{n-p_n}$. Then $\varepsilon_n(z_{n-j}, z_n)$ for $j = 1, 2, \dots, p_n$ are the same, and obviously all have the same distribution. Let's define this more precisely.

Definition 2.1. Let z_n be the realization of the random environment process $\{Z_n\}$ in the moment $n \geq 0$. We say that $\{X_n(z_n)\}_{n \in \mathbb{N}_0}$ is an INAR process with r -states random environment guided geometric marginals based on the negative binomial thinning operator of maximal order p (RrNGINARmax(p)), $p \in \mathbb{N}$, if the random variable $X_n(z_n)$ is defined as

$$(2.2) \quad X_n(z_n) = \begin{cases} \alpha * X_{n-1}(z_{n-1}) + \varepsilon_n(z_{n-1}, z_n), & \text{w.p. } \phi_1^{(p_n)}, \\ \alpha * X_{n-2}(z_{n-2}) + \varepsilon_n(z_{n-2}, z_n), & \text{w.p. } \phi_2^{(p_n)}, \\ \vdots & \vdots \\ \alpha * X_{n-p_n}(z_{n-p_n}) + \varepsilon_n(z_{n-p_n}, z_n), & \text{w.p. } \phi_{p_n}^{(p_n)}, \end{cases}$$

for $n \geq 1$, where

$$p_n = \begin{cases} p, & p_n^* \geq p, \\ p_n^*, & p_n^* < p, \end{cases}$$

$p_n^* = \max \{i \in \{1, 2, \dots, n\} : z_{n-1} = z_{n-2} = \dots = z_{n-i}\}$ and the following conditions are satisfied:

1. $\phi_i^{(p_n)} \geq 0$, $i \in \{1, 2, \dots, p_n\}$, $\sum_{i=1}^{p_n} \phi_i^{(p_n)} = 1$;
2. $\alpha \in (0, 1)$ and the counting sequence $\{U_i\}_{i \in \mathbb{N}}$ of the negative binomial thinning operator $\alpha*$ has pmf $P(U_i = u) = \frac{\alpha^u}{(1+\alpha)^{u+1}}$, $u \in \{0, 1, 2, \dots\}$;
3. $P(X_n(z_n) = x) = \frac{\mu_{z_n}^x}{(1+\mu_{z_n})^{x+1}}$, $x \in \{0, 1, 2, \dots\}$, where $\mu_{z_n} \in \{\mu_1, \mu_2, \dots, \mu_r\}$, $\mu_i > 0$, $i \in \{1, 2, \dots, r\}$ and $r \in \mathbb{N}$ is the number of states of the random environment process $\{Z_n\}$;
4. For fixed $i, j \in E_r = \{1, 2, \dots, r\}$, $\{\varepsilon_n(i, j)\}_{n \in \mathbb{N}}$ is a sequence of i.i.d. random variables;
5. $\{Z_n\}, \{\varepsilon_n(1, 1)\}, \{\varepsilon_n(1, 2)\}, \dots, \{\varepsilon_n(r, r)\}$ are mutually independent sequences of random variables;
6. $X_n(l)$ is independent of Z_m and $\varepsilon_m(i, j)$, for $0 \leq n < m$ and any $i, j, l \in E_r$.

We want to emphasize that this model contains p different sets of the probability parameters $\Psi_i = \{\phi_1^{(i)}, \phi_2^{(i)}, \dots, \phi_i^{(i)}\}$, for $i \in \{1, 2, \dots, p\}$. Set Ψ_i has i elements, so the total number of the probability parameters is $1 + 2 + \dots + p = \frac{p(p+1)}{2}$.

For each i there is a condition $\sum_{j=1}^i \phi_j^{(i)} = 1$, so there are $\frac{p(p+1)}{2} - p = \frac{p(p-1)}{2}$ unknown probability parameters. Specially, for $i = 1$, we have $\phi_1^{(1)} = 1$.

Remark 2.1. Important feature of the introduced process is a variable order. Actually, $\{X_n(z_n)\}$ is defined like a process of order p_n , where p_n is not fixed and depends on n . But, p_n is not a random variable due to the fact that it could be calculated for given $\{z_n\}$, using its building mechanism given in Definition 2.1 and the fact that our process is defined for the realized random environment process $\{z_n\}$. Once p_n reaches p , process takes shape of the model of fixed order, p , and this lasts as long as the state does not change. When it changes ($z_n \neq z_{n-1}$), then order (p_{n+1}) becomes equal 1. The order further continues to grow until the state changes again or until it reaches p . Therefore, we consider process which is mostly of order p , but it has some transitional periods of variable and ascending order, which begin when the state changes and end when the process reaches order p , or when state changes again.

This is similar to the idea of the Variable-Order Markov (VOM) model, which was investigated in [7]. As it is known a random variable in the Markov chain model depends on a fixed number of previous conditioning elements. However, in VOM models the number of conditioning random variables (which is called the context) depends on the specific observed realization and may vary over time.

Now, we will describe one more combined random environment NGINAR process. It is similar to the previous one, but differs during the transitional period (for $p_n^* < p$), where the process of variable order is replaced with the process of order one.

Definition 2.2. Let z_n be the realization of the random environment process $\{Z_n\}$ in the moment $n \geq 0$. We say that $\{X_n(z_n)\}_{n \in \mathbb{N}_0}$ is an INAR process with r -states random environment guided geometric marginals based on the negative binomial thinning operator of order p ($\text{RrNGINAR}_1(p)$) if random variable $X_n(z_n)$ is defined as

$$(2.3) \quad X_n(z_n) = \begin{cases} \alpha * X_{n-1}(z_{n-1}) + \varepsilon_n(z_{n-1}, z_n) & \text{w.p. } \phi_1^{(p_n)}, \\ \alpha * X_{n-2}(z_{n-2}) + \varepsilon_n(z_{n-2}, z_n), & \text{w.p. } \phi_2^{(p_n)}, \\ \vdots & \vdots \\ \alpha * X_{n-p_n}(z_{n-p_n}) + \varepsilon_n(z_{n-p_n}, z_n), & \text{w.p. } \phi_{p_n}^{(p_n)}, \end{cases}$$

for $n \geq 1$, where

$$p_n = \begin{cases} p, & p_n^* \geq p, \\ 1, & p_n^* < p, \end{cases}$$

$p_n^* = \max \{i \in \{1, 2, \dots, n\} : z_{n-1} = z_{n-2} = \dots = z_{n-i}\}$ and conditions 1–6 from Definition 2.1 are satisfied.

For $p_n = 1$ we have only one probability parameter $\phi_1^{(1)}$ and from the condition $\sum_{i=1}^{p_n} \phi_i^{(p_n)} = 1$ it follows $\phi_1^{(1)} = 1$, so only $\phi_1^{(p)}, \dots, \phi_p^{(p)}$ are unknown, but related via one equation. Therefore, it is sufficient to determine $p - 1$ probability parameters.

Remark 2.2. For the process given by Definition 2.2, p_n is not a random variable as well. Now, p_n takes one of the two possible values. Every time when state changes ($z_n \neq z_{n-1}$), order (p_{n+1}) becomes 1 and it remains the same until there is a series of enough (p) previous elements corresponding to the same state. So, we can divide this process into the series which can be represented as parts of the processes of order 1 and order p . In the series of order 1 state can be changed, so this series has the same form as RrNGINAR(1) process. For the series of order p it is necessary to stay in the same state, so they are whole in the one state and have the same form as CGINAR(p) process. There are as much different (by their marginals) CGINAR(p) processes as much different states we have.

Our next step is derivation of the distribution of the random variable $\varepsilon_n(i, j)$.

Theorem 2.1. *Let $\{X_n(z_n)\}$ be the RrNGINARmax(p) time series process or the RrNGINAR₁(p) process, and let $\mu_1 > 0, \mu_2 > 0, \dots, \mu_r > 0$. If $0 \leq \alpha \leq \min \left\{ \frac{\mu_l}{1+\mu_k}, k, l \in E_r \right\}$, then if $z_n = j$ and $z_{n-1} = i$, for $i, j \in E_r$, the distribution of the random variable $\varepsilon_n(i, j)$ can be written as a mixture of two geometric distributions*

$$(2.4) \quad \varepsilon_n(i, j) \stackrel{d}{=} \begin{cases} \text{Geom} \left(\frac{\mu_j}{1+\mu_j} \right), & \text{w.p. } 1 - \frac{\alpha\mu_i}{\mu_j - \alpha}, \\ \text{Geom} \left(\frac{\alpha}{1+\alpha} \right), & \text{w.p. } \frac{\alpha\mu_i}{\mu_j - \alpha}, \end{cases}$$

for $n \geq 1$.

Proof: Consider the probability generating function (pgf) of a random variable $X_n(z_n)$ in the case when $\{X_n(z_n)\}$ is the RrNGINARmax(p) process. Due to the properties of the pgf and the definition of the negative binomial thinning operator, it holds

$$\begin{aligned} \Phi_{X_n(z_n)}(s) &= \sum_{l=1}^{p_n} \phi_l^{(p_n)} E(s^{\alpha * X_{n-l}(z_{n-l})}) E(s^{\varepsilon_n(z_{n-l}, z_n)}) \\ &= \sum_{l=1}^{p_n} \phi_l^{(p_n)} \Phi_{X_{n-l}(z_{n-l})}(\Phi_U(s)) \Phi_{\varepsilon_n(z_{n-l}, z_n)}(s). \end{aligned}$$

We used notation Φ_U for Φ_{U_m} , $m \geq 1$, since U_m have all the same distribution. Because $z_{n-1} = z_{n-2} = \dots = z_{n-p_n} = i$, it holds $\Phi_{\varepsilon_n(z_{n-1}, z_n)}(s) = \Phi_{\varepsilon_n(z_{n-2}, z_n)}(s) =$

$\dots = \Phi_{\varepsilon_n(z_{n-p_n}, z_n)}(s) = \Phi_{\varepsilon_n(i, j)}(s)$ and $\Phi_{X_{n-1}(z_{n-1})}(s) = \Phi_{X_{n-2}(z_{n-2})}(s) = \dots = \Phi_{X_{n-p_n}(z_{n-p_n})}(s) = \Phi_{X_{n-1}(i)}(s)$, so

$$\Phi_{X_n(j)}(s) = \sum_{l=1}^{p_n} \phi_l^{(p_n)} \Phi_{X_{n-1}(i)}(\Phi_U(s)) \Phi_{\varepsilon_n(i, j)}(s) = \Phi_{X_{n-1}(i)}(\Phi_U(s)) \Phi_{\varepsilon_n(i, j)}(s).$$

The last equation is equivalent to

$$\frac{1}{1 + \mu_j - \mu_j s} = \Phi_{\varepsilon_n(i, j)}(s) \frac{1}{1 + \mu_i - \frac{\mu_i}{1 + \alpha - \alpha s}},$$

because $X_n(j)$ has $Geom\left(\frac{\mu_j}{1 + \mu_j}\right)$ distribution, $X_{n-1}(i)$ has $Geom\left(\frac{\mu_i}{1 + \mu_i}\right)$ distribution and U_m has $Geom\left(\frac{\alpha}{1 + \alpha}\right)$ distribution. Calculation of $\Phi_{\varepsilon_n(i, j)}(s)$ gives

$$\Phi_{\varepsilon_n(i, j)}(s) = \frac{\alpha \mu_i}{\mu_j - \alpha} \cdot \frac{1}{1 + \alpha - \alpha s} + \left(1 - \frac{\alpha \mu_i}{\mu_j - \alpha}\right) \cdot \frac{1}{1 + \mu_j - \mu_j s},$$

which implies (2.4).

Suppose now that $\{X_n(z_n)\}$ is the $RrNGINAR_1(p)$ process. Let fix $n \in \mathbb{N}$. If $p_n^* \geq p$, then $X_n(z_n)$ is generated in the same way as $RrNGINAR_{max}(p)$ process and it holds $p_n = p$. So, applying the same procedure as before, substituting p_n with p we get (2.4). If the previous condition doesn't hold, $X_n(z_n)$ has the form like in the $RrNGINAR(1)$ process, for which we know that $\varepsilon_n(i, j)$ has the required distribution. \square

Now, we derive conditional expectation and variance of the introduced processes.

Theorem 2.2. *Let $\{X_n(z_n)\}$ be $RrNGINAR_{max}(p)$ or $RrNGINAR_1(p)$ time series process, and let $\mu_1 > 0, \mu_2 > 0, \dots, \mu_r > 0$. If $0 \leq \alpha \leq \min\left\{\frac{\mu_l}{1 + \mu_k}, k, l \in E_r\right\}$, $z_{n+1} = j$ and $z_n = i$, for $i, j \in E_r$, then the conditional expectation and the conditional variance of this process are given by*

$$E(X_{n+1}|H_n) = \mu_j - \alpha \mu_i + \alpha \sum_{l=1}^{p_{n+1}} \phi_l^{(p_{n+1})} X_{n+1-l},$$

$$\begin{aligned} Var(X_{n+1}|H_n) &= \mu_j(\mu_j + 1) - \alpha \mu_i(1 + 2\alpha + \alpha \mu_i) + \alpha(1 + \alpha) \sum_{l=1}^{p_{n+1}} \phi_l^{(p_{n+1})} X_{n+1-l} \\ &\quad + \alpha^2 \sum_{l=1}^{p_{n+1}} \phi_l^{(p_{n+1})} X_{n+1-l}^2 - \alpha^2 \left(\sum_{l=1}^{p_{n+1}} \phi_l^{(p_{n+1})} X_{n+1-l} \right)^2, \end{aligned}$$

where $H_n = \sigma(X_n, X_{n-1}, \dots, X_{n-p_{n+1}})$ represents the σ -field generated by $\{X_n, X_{n-1}, \dots, X_{n-p_{n+1}}\}$.

Proof: For the simplicity of notation, we will use X_n instead of $X_n(z_n)$, for $n \geq 0$ and ε_n instead of $\varepsilon_n(z_{n-1}, z_n)$, for $n \geq 1$. From the definition of the negative binomial thinning and the properties of the conditional expectation, the conditional probability generating function is

$$\begin{aligned}\Phi_{X_{n+1}|H_n}(s) &\equiv E(s^{X_{n+1}}|H_n) = \Phi_{\varepsilon_{n+1}}(s) \sum_{l=1}^{p_{n+1}} \phi_l^{(p_{n+1})} E(s^{\alpha * X_{n+1-l}}|H_n) \\ &= \Phi_{\varepsilon_{n+1}}(s) \sum_{l=1}^{p_{n+1}} \phi_l^{(p_{n+1})} \Phi_U^{X_{n+1-l}}(s),\end{aligned}$$

where $\Phi_U(s) = \frac{1}{1+\alpha-\alpha s}$. It holds

$$E(X_{n+1}|H_n) = \Phi'_{X_{n+1}|H_n}(1)$$

and

$$Var(X_{n+1}|H_n) = \Phi''_{X_{n+1}|H_n}(1) + \Phi'_{X_{n+1}|H_n}(1) - (\Phi'_{X_{n+1}|H_n}(1))^2.$$

Derivating function $\Phi_{X_{n+1}|H_n}(s)$ with respect to s and using results

$$\Phi_U(1) = 1, \quad \Phi'_U(1) = \alpha, \quad \Phi''_U(1) = 2\alpha^2$$

and

$$\Phi_{\varepsilon_{n+1}}(1) = 1, \quad \Phi'_{\varepsilon_{n+1}}(1) = \mu_j - \alpha\mu_i, \quad \Phi''_{\varepsilon_{n+1}}(1) = 2\mu_j^2 - 2\alpha\mu_i(\mu_j + \alpha)$$

gives

$$\Phi'_{X_{n+1}|H_n}(1) = \mu_j - \alpha\mu_i + \alpha \sum_{l=1}^{p_{n+1}} \phi_l^{(p_{n+1})} X_{n+1-l}$$

and

$$\begin{aligned}\Phi''_{X_{n+1}|H_n}(1) &= 2\mu_j^2 - 2\alpha\mu_i(\mu_j + \alpha) + \alpha^2 \sum_{l=1}^{p_{n+1}} \phi_l^{(p_{n+1})} X_{n+1-l}^2 \\ &\quad + \alpha(2\mu_j - 2\alpha\mu_i + \alpha) \sum_{l=1}^{p_{n+1}} \phi_l^{(p_{n+1})} X_{n+1-l}.\end{aligned}$$

The requested formulas directly follow from here. \square

The conditional expectation and the conditional variance of higher order can be calculated using following recurrent relations:

$$E(X_{n+k}|H_n) = \mu_j - \alpha\mu_i + \alpha \left[\sum_{l=1}^{k-1} \phi_l^{(p_{n+1})} E(X_{n+k-l}|H_n) + \sum_{l=k}^{p_{n+k}} \phi_l^{(p_{n+1})} X_{n+k-l} \right]$$

for $2 \leq k \leq p_{n+k}$,

$$E(X_{n+k}|H_n) = \mu_j - \alpha\mu_i + \alpha \sum_{l=1}^{p_{n+k}} \phi_l^{(p_{n+1})} E(X_{n+k-l}|H_n), \quad k > p_{n+k},$$

$$\begin{aligned}
\text{Var}(X_{n+k}|H_n) &= \mu_j(\mu_j + 1) - \alpha\mu_i(1 + 2\alpha + \alpha\mu_i) \\
&+ \alpha^2 \sum_{l=1}^{k-1} \phi_l^{(p_{n+1})} \text{Var}(X_{n+k-l}|H_n) \\
&+ \alpha(1 + \alpha) \left(\sum_{l=1}^{k-1} \phi_l^{(p_{n+1})} E(X_{n+k-l}|H_n) + \sum_{l=k}^{p_{n+k}} \phi_l^{(p_{n+1})} X_{n+k-l} \right) \\
&+ \alpha^2 \left(\sum_{l=1}^{k-1} \phi_l^{(p_{n+1})} [E(X_{n+k-l}|H_n)]^2 + \sum_{l=1}^{p_{n+k}} \phi_l^{(p_{n+1})} X_{n+k-l}^2 \right) \\
&- \alpha^2 \left(\sum_{l=1}^{k-1} \phi_l^{(p_{n+1})} E(X_{n+k-l}|H_n) + \sum_{l=k}^{p_{n+k}} \phi_l^{(p_{n+1})} X_{n+k-l} \right)^2
\end{aligned}$$

for $2 \leq k \leq p_{n+k}$,

$$\begin{aligned}
\text{Var}(X_{n+k}|H_n) &= \mu_j(\mu_j + 1) - \alpha\mu_i(1 + 2\alpha + \alpha\mu_i) \\
&+ \alpha^2 \sum_{l=1}^{p_{n+k}} \phi_l^{(p_{n+1})} \text{Var}(X_{n+k-l}|H_n) \\
&+ \alpha(1 + \alpha) \left(\sum_{l=1}^{p_{n+k}} \phi_l^{(p_{n+1})} E(X_{n+k-l}|H_n) \right) \\
&+ \alpha^2 \sum_{l=1}^{p_{n+k}} \phi_l^{(p_{n+1})} [E(X_{n+k-l}|H_n)]^2 \\
&- \alpha^2 \left(\sum_{l=1}^{p_{n+k}} \phi_l^{(p_{n+1})} E(X_{n+k-l}|H_n) \right)^2
\end{aligned}$$

for $k > p_{n+k}$.

3. CORRELATION STRUCTURE

Let's now investigate the correlation structure of the defined processes. First, we will examine the $\text{RrNGINARmax}(p)$ process. From (2.2), it follows

$$\begin{aligned}
\text{Cov}(X_n, X_{n-1}) &= \alpha \sum_{i=1}^{p_n} \phi_i^{(p_n)} \text{Cov}(X_{n-i}, X_{n-1}), \\
\text{Cov}(X_n, X_{n-2}) &= \alpha \sum_{i=1}^{p_n} \phi_i^{(p_n)} \text{Cov}(X_{n-i}, X_{n-2}), \\
&\vdots \\
\text{Cov}(X_n, X_{n-p_n}) &= \alpha \sum_{i=1}^{p_n} \phi_i^{(p_n)} \text{Cov}(X_{n-i}, X_{n-p_n}),
\end{aligned}$$

where we have used X_n instead of $X_n(z_n)$ for the simplicity of notation. Denote $Cov(X_n, X_{n-h})$ as $\gamma_n^{(h)}$, for $h \geq 0$. This system can be represented in the matrix form

$$(3.1) \quad \begin{bmatrix} \gamma_n^{(1)} \\ \gamma_n^{(2)} \\ \vdots \\ \gamma_n^{(p_n)} \end{bmatrix} = \begin{bmatrix} \gamma_{n-1}^{(0)} & \gamma_{n-1}^{(1)} & \cdots & \gamma_{n-1}^{(p_n-1)} \\ \gamma_{n-1}^{(1)} & \gamma_{n-2}^{(0)} & \cdots & \gamma_{n-2}^{(p_n-2)} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{n-1}^{(p_n-1)} & \gamma_{n-2}^{(p_n-2)} & \cdots & \gamma_{n-p_n}^{(0)} \end{bmatrix} \cdot \begin{bmatrix} \theta_1^{(p_n)} \\ \theta_2^{(p_n)} \\ \vdots \\ \theta_{p_n}^{(p_n)} \end{bmatrix},$$

where $\theta_i^{(p_n)} = \alpha \phi_i^{(p_n)}$, for $i \in \{1, 2, \dots, p_n\}$. More simple form is of course obtained by matrix notation, i.e.

$$\gamma_n = \mathbf{\Gamma}_n \cdot \boldsymbol{\theta}_n,$$

where we denoted the corresponding vectors with γ_n and $\boldsymbol{\theta}_n$, and $\mathbf{\Gamma}_n$ is the covariance matrix of the vector $(X_{n-p_n}, X_{n-p_n-1}, \dots, X_{n-1})'$.

In accordance with the definition of p_n random variables $X_{n-1}, X_{n-2}, \dots, X_{n-p_n}$ have the same distribution, so it holds

$$Cov(X_{n-1}, X_{n-1}) = Cov(X_{n-2}, X_{n-2}) = \cdots = Cov(X_{n-p_n}, X_{n-p_n}) = \sigma_{X_{n-1}}^2.$$

Now, it is possible to divide covariance matrix $\mathbf{\Gamma}_n$ with $\sigma_{X_{n-1}}^2$ and the result is the correlation matrix

$$(3.2) \quad \mathbf{R}_{p_n \times p_n}^{(n-1)} = \begin{bmatrix} 1 & \rho_{n-1}^{(1)} & \cdots & \rho_{n-1}^{(p_n-1)} \\ \rho_{n-1}^{(1)} & 1 & \cdots & \rho_{n-2}^{(p_n-2)} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1}^{(p_n-1)} & \rho_{n-2}^{(p_n-2)} & \cdots & 1 \end{bmatrix}.$$

However, dividing the left side of the equation (3.1) with $\sigma_{X_{n-1}}^2$ will not give the vector of the correlations, because $z_n \neq z_{n-1}$ in general. Actually, the equation which is satisfied by the correlation matrix is

$$(3.3) \quad \begin{bmatrix} 1 & \rho_{n-1}^{(1)} & \cdots & \rho_{n-1}^{(p_n-1)} \\ \rho_{n-1}^{(1)} & 1 & \cdots & \rho_{n-2}^{(p_n-2)} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1}^{(p_n-1)} & \rho_{n-2}^{(p_n-2)} & \cdots & 1 \end{bmatrix} \cdot \begin{bmatrix} \theta_1^{(p_n)} \\ \theta_2^{(p_n)} \\ \vdots \\ \theta_{p_n}^{(p_n)} \end{bmatrix} = \frac{\sigma_{X_n}}{\sigma_{X_{n-1}}} \begin{bmatrix} \rho_n^{(1)} \\ \rho_n^{(2)} \\ \vdots \\ \rho_n^{(p_n)} \end{bmatrix}.$$

Remark 3.1. In the special case when $z_n = z_{n-1}$, we have that $\sigma_{X_n} = \sigma_{X_{n-1}}$, so the equation for the correlation matrix takes the same form as the equation for the covariance matrix. It is important to notice that the subsample $X_{n-1}, X_{n-2}, \dots, X_{n-p_n}$ of the $\text{RrNGINARmax}(p)$ process cannot be seen as a subsample of a stationary process in general. Really, it is possible to be $p_i \neq p_j$, for $i, j \in \{n-1, n-2, \dots, n-p_n\}$ (for example, if $z_{n-p_n-1} \neq z_{n-p_n}$, then

$p_{n-p_n+1} = 1, p_{n-p_n+2} = 2, \dots, p_{n-1} = p_n - 1$), so we deal with a process of variable order, which does not have to be stationary. However, if $p_{n-p_n} = p$ and $z_{n-p_n} = z_{n-p_n-1}$ then elements of the subsample $X_{n-1}, X_{n-2}, \dots, X_{n-p_n}$ all have the same distribution and are defined based on the p previous elements, so it is possible to consider this subsample as a subsample of CGINAR(p) process, which is stationary. Really, from the definition of p_m it holds $z_{m-1} = z_{m-2} = \dots = z_{m-p_m}$ for arbitrary m . For $m = n$ we have $z_{n-1} = z_{n-2} = \dots = z_{n-p_n}$ and for $m = n - p_n$ it holds $z_{n-p_n-1} = z_{n-p_n-2} = \dots = z_{n-p_n-p}$, where we used relation $p_{n-p_n} = p$. Combining these results with equation $z_{n-p_n} = z_{n-p_n-1}$ gives $z_{n-p_n-\max p} = \dots = z_{n-p_n-1} = z_{n-p_n} = \dots = z_{n-1}$ and consequently, $p_{n-p_n+1} = p_{n-p_n+2} = \dots = p_n = p$. If $z_n = z_{n-1}$, additionally, then the same conclusion holds for the subsample $X_n, X_{n-1}, \dots, X_{n-p_n}$.

Now, let's consider the RrNGINAR₁(p) process. It can be partitioned into samples of CGINAR(p) or RrNGINAR(1) processes. So, correlation structure is determined by the correlation structure of the mentioned processes. For $p_n = 1$ we have

$$\gamma_n^{(1)} = \alpha \gamma_{n-1}^{(0)},$$

and for $p_n = p$ it holds

$$\begin{bmatrix} \gamma_n^{(1)} \\ \gamma_n^{(2)} \\ \vdots \\ \gamma_n^{(p)} \end{bmatrix} = \begin{bmatrix} \gamma_{n-1}^{(0)} & \gamma_{n-1}^{(1)} & \dots & \gamma_{n-1}^{(p-1)} \\ \gamma_{n-1}^{(1)} & \gamma_{n-2}^{(0)} & \dots & \gamma_{n-2}^{(p-2)} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{n-1}^{(p-1)} & \gamma_{n-2}^{(p-2)} & \dots & \gamma_{n-p}^{(0)} \end{bmatrix} \cdot \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{bmatrix},$$

where $\theta_i = \alpha \phi_i$, for $i \in \{1, 2, \dots, p\}$. These equations can be represented by (3.1), substituting p_n with 1 and p . It also holds (3.2) and (3.3). RrNGINAR_{max}(p) cannot have series parts with two or more successive elements of order one in the same state. However, for RrNGINAR₁(p) process the maximal length of such a series is p . The Theorem 3, from [13], holds for n and k which satisfy $z_n = \dots = z_{n-k}$ and $p_n = 1$. Therefore, based on this theorem, the maximal value that k can take is p .

Remark 3.2. For RrNGINAR₁(p) process, subsample $X_{n-1}, X_{n-2}, \dots, X_{n-p_n}$ can always be viewed as a sample of a stationary process. The case $p_n = 1$ is trivial since it gives subsample of only one element. If $p_n = p$ and $p_{n-p_n} = p$, then directly from the definition of this process it follows that $X_{n-1}, X_{n-2}, \dots, X_{n-p_n}$ are all in the same state and of the same order. If $z_n = z_{n-1}$ then the same holds for $X_n, X_{n-1}, \dots, X_{n-p_n}$.

4. YULE-WALKER ESTIMATION OF THE PARAMETERS

In the proof of Theorem 5 from [13] stationarity of the processes attached to the maximal subsamples provided the strong consistency of the estimators. If we want here to prove the same, it would be useful to define estimators only on the part of the process to which we can attach a stationary process.

In accordance with Remark 3.1 let $p_n = p_{n-p_n} = p$, $z_{n-p_n} = z_{n-p_n-1}$ and $z_n = z_{n-1} = k \in E_r$ for $\text{RrNGINARmax}(p)$ process. Then $p_i = p$ and $z_i = k$, for all $i \in \{n - p_n, n - p_n + 1, \dots, n\}$. Because of stationarity it is possible to write $\gamma_{|i-j|}$ instead of $\gamma_j^{(i)}$, for $i \in \{0, 1, \dots, p_n\}, j \in \{n - p_n, n - p_n + 1, \dots, n\}$ without loss of generality. Then, we introduce $\gamma_{|i-j|}^{(k)}$ as a more informative and adequate notation, where index k indicates the random state. Applying this to the system (3.1), using analogously $\theta_{p,i}$ instead of $\theta_i^{(p_n)}$ we obtain

$$(4.1) \quad \begin{bmatrix} \gamma_1^{(k)} \\ \gamma_2^{(k)} \\ \vdots \\ \gamma_p^{(k)} \end{bmatrix} = \begin{bmatrix} \gamma_0^{(k)} & \gamma_1^{(k)} & \cdots & \gamma_{p-1}^{(k)} \\ \gamma_1^{(k)} & \gamma_0^{(k)} & \cdots & \gamma_{p-2}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{p-1}^{(k)} & \gamma_{p-2}^{(k)} & \cdots & \gamma_0^{(k)} \end{bmatrix} \cdot \begin{bmatrix} \theta_{p,1} \\ \theta_{p,2} \\ \vdots \\ \theta_{p,p} \end{bmatrix}.$$

Notice that p_n equals p , as it is assumed above. We estimate μ_k and $\gamma_h^{(k)}$, $h \in \{0, 1, 2, \dots, p-1\}$ as in [13], but only based on the part of a sample. Precisely, estimators for state $k \in \{1, 2, \dots, r\}$ are based on the sets $V_{0,p}^{(k)} = \{i \in \{1, 2, \dots, N\} | z_i = k, p_i = p\}$ and $V_{h,p}^{(k)} = \{i \in V_{0,p}^{(k)} | i + h \in V_{0,p}^{(k)}\}$, for $h \geq 1$, where N is the size of the sample, and are given by

$$(4.2) \quad \hat{\mu}_k = \frac{1}{n_{0,p}^{(k)}} \sum_{i \in V_{0,p}^{(k)}} X_i(k), \quad \hat{\gamma}_{h,p}^{(k)} = \frac{1}{n_{h,p}^{(k)}} \sum_{i \in V_{h,p}^{(k)}} (X_{i+h}(k) - \hat{\mu}_k)(X_i(k) - \hat{\mu}_k),$$

for $h \geq 0$, $k \in \{1, 2, \dots, r\}$ and where $n_{h,p}^{(k)} = |V_{h,p}^{(k)}|$, for $h \geq 0$.

Substituting the theoretical moments in (4.1) with the empirical ones and then expressing the vector of the unknown parameters, we get

$$\begin{bmatrix} \hat{\theta}_{p,1}^{(k)} \\ \hat{\theta}_{p,2}^{(k)} \\ \vdots \\ \hat{\theta}_{p,p}^{(k)} \end{bmatrix} = \begin{bmatrix} \hat{\gamma}_{0,p}^{(k)} & \hat{\gamma}_{1,p}^{(k)} & \cdots & \hat{\gamma}_{p-1,p}^{(k)} \\ \hat{\gamma}_{1,p}^{(k)} & \hat{\gamma}_{0,p}^{(k)} & \cdots & \hat{\gamma}_{p-2,p}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}_{p-1,p}^{(k)} & \hat{\gamma}_{p-2,p}^{(k)} & \cdots & \hat{\gamma}_{0,p}^{(k)} \end{bmatrix}^{-1} \begin{bmatrix} \hat{\gamma}_{1,p}^{(k)} \\ \hat{\gamma}_{2,p}^{(k)} \\ \vdots \\ \hat{\gamma}_{p,p}^{(k)} \end{bmatrix}.$$

Now, it is possible to estimate α and $\phi_i^{(p)}$. First we obtain, respectively

$$\hat{\alpha}^{(k)} = \sum_{i=1}^p \hat{\theta}_{p,i}^{(k)}, \quad \hat{\phi}_{p,i}^{(k)} = \frac{\hat{\theta}_{p,i}^{(k)}}{\hat{\alpha}^{(k)}}, \quad i \in \{1, 2, \dots, p\}, \quad k \in \{1, 2, \dots, r\},$$

where (k) , as above, indicates that the estimators are based on the subsample with state equal k . At last, taking into account all states and their frequencies of occurrence, the final Yule-Walker estimators are

$$(4.3) \quad \widehat{\alpha}^{YW} = \sum_{k=1}^r \frac{n^{(k)}}{N} \widehat{\alpha}^{(k)}, \quad \widehat{\phi}_{p,i}^{YW} = \sum_{k=1}^r \frac{n^{(k)}}{N} \widehat{\phi}_{p,i}^{(k)}, \quad \widehat{\mu}_k^{YW} = \widehat{\mu}_k.$$

Now, we move our attention to $\text{RrNGINAR}_1(p)$ process. Let $z_n = z_{n-1}$. Estimators of the process covariances are based on the maximal union (in the sense of number of elements) of the samples which can be treated as the samples of stationary processes. Estimators are given by

$$(4.4) \quad \widehat{\mu}_{k,j} = \frac{1}{n_{0,j}^{(k)}} \sum_{i \in V_{0,j}^{(k)}} X_i(k), \quad \widehat{\gamma}_{h,j}^{(k)} = \frac{1}{n_{h,j}^{(k)}} \sum_{i \in V_{h,j}^{(k)}} (X_{i+h}(k) - \widehat{\mu}_{k,j})(X_i(k) - \widehat{\mu}_{k,j}),$$

where $h \geq 0$, $k \in \{1, 2, \dots, r\}$, $j \in \{1, p\}$, and they are based on the sets $V_{0,1}^{(k)} = \{i \in \{1, 2, \dots, N\} \mid z_i = k, p_i = 1\}$, $V_{h,1}^{(k)} = \{i \in V_{0,1}^{(k)} \mid i+h \in V_{0,1}^{(k)}\}$, $n_{h,1}^{(k)} = |V_{h,1}^{(k)}|$, for $h \geq 1$ and $V_{h,p}^{(k)}$ and $n_{h,p}^{(k)}$, for $h \geq 0$ are defined as before. Let $\alpha_j^{(k)}$ represents the autocorrelation parameter corresponding to the process subsamples of state k and order j , where $j \in \{1, p\}$.

Similarly as for $\text{RrNGINARmax}(p)$ process we get

$$\widehat{\alpha}_1^{(k)} = \frac{\widehat{\gamma}_{1,1}^{(k)}}{\widehat{\gamma}_{0,1}^{(k)}}$$

and

$$(4.5) \quad \begin{bmatrix} \widehat{\theta}_1^{(k)} \\ \widehat{\theta}_2^{(k)} \\ \vdots \\ \widehat{\theta}_p^{(k)} \end{bmatrix} = \begin{bmatrix} \widehat{\gamma}_{0,p}^{(k)} & \widehat{\gamma}_{1,p}^{(k)} & \cdots & \widehat{\gamma}_{p-1,p}^{(k)} \\ \widehat{\gamma}_{1,p}^{(k)} & \widehat{\gamma}_{0,p}^{(k)} & \cdots & \widehat{\gamma}_{p-2,p}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\gamma}_{p-1,p}^{(k)} & \widehat{\gamma}_{p-2,p}^{(k)} & \cdots & \widehat{\gamma}_{0,p}^{(k)} \end{bmatrix}^{-1} \begin{bmatrix} \widehat{\gamma}_{1,p}^{(k)} \\ \widehat{\gamma}_{2,p}^{(k)} \\ \vdots \\ \widehat{\gamma}_{p,p}^{(k)} \end{bmatrix}.$$

From (4.5) we obtain the estimators as

$$\widehat{\alpha}_p^{(k)} = \sum_{i=1}^p \widehat{\theta}_i^{(k)}, \quad \widehat{\phi}_i^{(k)} = \frac{\widehat{\theta}_i^{(k)}}{\widehat{\alpha}_p^{(k)}}, \quad i \in \{1, 2, \dots, p\}, \quad k \in \{1, 2, \dots, r\}.$$

Now, we get

$$\widehat{\alpha}^{(k)} = \frac{n_1^{(k)} \widehat{\alpha}_1^{(k)} + n_p^{(k)} \widehat{\alpha}_p^{(k)}}{n_1^{(k)} + n_p^{(k)}}, \quad \widehat{\mu}_k = \frac{n_1^{(k)} \widehat{\mu}_{k,1} + n_p^{(k)} \widehat{\mu}_{k,p}}{n_1^{(k)} + n_p^{(k)}},$$

and finally, using preceding results for all states, we obtain YW estimators as

$$(4.6) \quad \hat{\alpha}^{YW} = \sum_{k=1}^r \frac{n_1^{(k)} + n_p^{(k)}}{N} \hat{\alpha}^{(k)}, \quad \hat{\phi}_i^{YW} = \sum_{k=1}^r \frac{n_p^{(k)}}{N} \hat{\phi}_i^{(k)}, \quad \hat{\mu}_k^{YW} = \hat{\mu}_k.$$

Theorem 4.1. *Estimators given by (4.2) and (4.4) are strongly consistent.*

Proof: The general idea is to divide the process subsample, indexed, i.e. determined by $V_{0,p}^{(k)}$, into maximal subsamples and then use the proof of Theorem 5 from [13]. It is easy to notice that this theorem can be expanded so that applies for $h > 1$. Really, if in the expression for γ_1 we replace $i + 1$ by $i + h$, it becomes γ_h . Further procedure is the same. If X_i, X_{i+1}, \dots, X_j is a subsample such that $\{i, i + 1, \dots, j\} \subseteq V_{0,p}^{(k)}$, we say that it is maximal if $z_i = z_{i+1} = \dots = z_j = k$, $z_{j+1} \neq k$, $p_i = p$ and $p_{i-1} \neq p$. Based on Remark 3.1, X_i, X_{i+1}, \dots, X_j represents a sample of CGINAR(p) process, so it is stationary. The rest of the proof is the same as the proof of Theorem 5. The same procedure is applied to the RrNGINAR $_1(p)$ process. \square

Since the quotient of linear combinations of the strongly consistent statistics is also strongly consistent, we have the following corollary.

Corollary 4.1. *Estimators given in (4.3) and (4.6) are strongly consistent.*

If we want to estimate the probabilities $\phi_i^{(p_n)}$, where $1 < p_n < p$ and $1 \leq i \leq p_n$, given in (2.2) for the definition of RrNGINARmax(p) process, preceding approach cannot be used. The problem is that elements X_n , for $1 < p_n < p$ are isolated in the sense that the both of their neighbors have different order, so it is impossible to form the subsample containing X_n (to define the estimators) with two or more successive elements of the same order.

This problem is worked out by defining new modified YW estimators which have less restrictive conditions in using the corresponding subsamples of the process. These modified estimators are obtained from the strongly consistent YW estimators, discussed above, by substituting their corresponding sets $V_{h,p}^{(k)}$, for $h \geq 1$, with $V_{0,p}^{(k)}$. In other words, if the corresponding sets of the modified YW estimators are denoted by $\tilde{V}_{h,p}^{(k)}$, then $\tilde{V}_{h,p}^{(k)} = V_{0,p}^{(k)}$, for $h \geq 1$. Note that $\tilde{V}_{h,p}^{(k)} \supseteq V_{h,p}^{(k)}$. However, because of these modifications, we cannot claim the modified YW strong consistence, but their goodness may be verified in the application on the simulated process values. In this regard, the results obtained in the next section show that the corresponding estimates gradually converge towards parameter values when the size of the sample increases.

5. SIMULATIONS

In this section we investigate the correctness of the modified Yule-Walker estimators. For this purpose we have simulated realizations of the processes $\text{RrNGINAR}_{\max}(p)$ and $\text{RrNGINAR}_1(p)$, and estimated unknown parameters in both cases. There are 100 replicates, each of size 10000. Both of the processes are considered in parallel. We choose the parameters $\alpha, p, r, \boldsymbol{\mu}, \mathbf{p}_{mat}$ and $\boldsymbol{\phi}$. The random environment process transition probability matrix is noted by \mathbf{p}_{mat} , and $\boldsymbol{\mu}$ is a vector of means. In the case of $\text{RrNGINAR}_{\max}(p)$ process, the p_n th row, $p_n \in \{2, \dots, p\}$, of the matrix $\boldsymbol{\phi}$ contains (up to the p_n th column) probabilities $\phi_i^{(p_n)}$, $i \in \{1, 2, \dots, p_n\}$, from (2.2). In the case of $\text{RrNGINAR}_1(p)$ process, the last row represents probabilities in (2.3). Matrix \mathbf{p}_{mat} controls changing of the states, where diagonal elements represents the probabilities of staying in the same state. When its diagonal values are high, it is expected for the random environment process to stay in the same state more often than to change the state. This is a preferable situation since it makes the sets $V_{h,p}^{(k)}$, for $h \geq 1$, to be bigger and in this way relatively less different from $\tilde{V}_{h,p}^{(k)}$. Consequently, modified estimators became approximately equal (at least being close) to the strongly consistent YW estimators.

The simulation of the random environment process $\{Z_n\}$ represents the first step in the simulation of the defined processes. After the generation of the observed values $\{z_n\}$, we can easily evaluate the process of orders $\{p_n\}$ for both processes by using their definitions. Finally, we can simulate the values of both defined processes by using the observed values $\{z_n\}$ and $\{p_n\}$, and definitions of the defined processes.

We considered six different cases. In each case we obtained the modified YW estimators of the unknown parameters for both of the processes. All the results are given in the appropriate tables. Comparison of the results is based on the relative errors, since values of the parameters are different.

- 1) In the first case vector of means is $\boldsymbol{\mu} = (1, 2)$. For these values maximal value for α is $1/3$ and we chose $\alpha = 0.3$. The random state process transition probability matrix we used is $\mathbf{p}_{mat} = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$. Diagonal elements are equal 0.8, so, based on the discussion above, good estimates are expected. Matrix of probabilities is $\boldsymbol{\phi} = \begin{bmatrix} 1 & 0 \\ 0.6 & 0.4 \end{bmatrix}$.
- 2) Here we investigate what happens when α , from case 1), reduces to $\alpha = 0.15$, where $\boldsymbol{\phi} = \begin{bmatrix} 1 & 0 \\ 0.5 & 0.5 \end{bmatrix}$. Since lower value of α contributes to less correlation, it is natural to expect that estimates for $\boldsymbol{\phi}$ are worse than in the case 1 and this is confirmed by the results. Estimates for α and $\boldsymbol{\mu}$ are almost the same, but slightly better.

- 3) This case differs from the first by the probabilities of changing state. We used $\mathbf{p}_{mat} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$. Now, there is equal probability to stay in the same state as it is to change it. As we discussed earlier, this is not favorably, so worse estimates for ϕ are expected and this conclusion is confirmed by the results. However, our sample is big enough, so difference is very small. Estimates for α and μ are again slightly better than the estimates of case 1).
- 4) Vector of means is now $\mu = (4, 5)$ and probabilities on the diagonal of the transition matrix of random states are 0.7, i.e. $\mathbf{p}_{mat} = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}$. Estimates are almost the same as in the first case. Thereby, $\text{RrNGINAR}_{\max}(p)$ process provides better results, while the $\text{RrNGINAR}_1(p)$ process gives worse results than in the first case.

These preceding four cases refer to the processes based on the environment process with two random states. The corresponding results are presented in Table 1.

- 5) Here we consider what happens when, in case 1), maximal order p increases to 3. The estimates of α are slightly better, but the estimates of ϕ are worse, because the probabilities for the order 2 (in case of $\text{RrNGINAR}_{\max}(p)$ process) are estimated using very small sample. However, they are significantly improved when sample size increases to 10000. These results are presented in Table 2.
- 6) In the last case we have simulated process with three possible states ($r = 3$). Parameters, as well as the results, are given in Table 3. The greater number of states contributes to the smaller probability of staying in the same state, so for the small sample sizes, estimates are not so good, but increasing the size of the simulated sample gives much better results. Estimates for $\phi_1^{(2)}$ and $\phi_2^{(2)}$ are not so good, which is reasonable because elements x_n of the simulated sample, such that $p_n = 2$, have neighbors with order different from 2.

In the cases 1), 5) and 6) the $\text{RrNGINAR}_1(p)$ process provides better estimates, while in cases 2) and 4) the $\text{RrNGINAR}_{\max}(p)$ process is better choice. In the third case they are almost equally good. It is important to notice that in each case results are better when the size of the sample increases.

Table 1: Estimates for $p = 2, r = 2$.

n_1	$\hat{\mu}_1^{YW}$	$\hat{\mu}_2^{YW}$	$\hat{\alpha}^{YW}$	$\hat{\phi}_{2,1}^{YW}$	$\hat{\phi}_{2,2}^{YW}$	$\hat{\alpha}^{YW}$	$\hat{\phi}_{2,1}^{YW}$	$\hat{\phi}_{2,2}^{YW}$
1) True values $\boldsymbol{\mu} = (1, 2), \alpha = 0.3, \boldsymbol{\phi} = \begin{bmatrix} 1 & 0 \\ 0.6 & 0.4 \end{bmatrix}, \mathbf{P}_{mat} = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$								
500	0.9924	1.9782	0.3226	0.6389	0.3611	0.318	0.624	0.376
SE	0.1200	0.2097	0.1474	0.2009	0.2009	0.1477	0.2338	0.2338
1000	0.9912	2.0064	0.3185	0.6198	0.3802	0.3188	0.6084	0.3916
SE	0.0872	0.1487	0.1071	0.1371	0.1371	0.1066	0.1282	0.1282
5000	0.9953	1.9952	0.3048	0.6074	0.3926	0.2984	0.6089	0.3911
SE	0.0375	0.0539	0.0525	0.0578	0.0578	0.0466	0.057	0.057
10000	0.9978	1.9999	0.3049	0.6038	0.3962	0.2993	0.5971	0.4029
SE	0.0288	0.0407	0.0388	0.0381	0.0381	0.0318	0.0409	0.0409
2) True values $\boldsymbol{\mu} = (1, 2), \alpha = 0.15, \boldsymbol{\phi} = \begin{bmatrix} 1 & 0 \\ 0.5 & 0.5 \end{bmatrix}, \mathbf{P}_{mat} = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$								
500	0.9988	2.0047	0.1568	0.688	0.312	0.1781	2.2577	-1.2577
SE	0.0939	0.1965	0.13	0.8714	0.8714	0.1438	15.5205	15.5205
1000	0.9933	2.0136	0.1532	0.4913	0.5087	0.1636	0.4655	0.5345
SE	0.0745	0.1312	0.0983	1.2527	1.2527	0.1038	0.7767	0.7767
5000	1.0011	1.9999	0.1526	0.5068	0.4932	0.1547	0.4921	0.5079
SE	0.0349	0.0562	0.0442	0.0984	0.0984	0.0459	0.0992	0.0992
10000	1.0031	1.9995	0.1543	0.5008	0.4992	0.1523	0.4951	0.5049
SE	0.0252	0.0368	0.0284	0.0649	0.0649	0.0309	0.0647	0.0647
3) True values $\boldsymbol{\mu} = (1, 2), \alpha = 0.3, \boldsymbol{\phi} = \begin{bmatrix} 1 & 0 \\ 0.6 & 0.4 \end{bmatrix}, \mathbf{P}_{mat} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$								
500	1.0044	1.9879	0.3201	0.5478	0.4522	0.3279	0.5864	0.4136
SE	0.1082	0.2065	0.1306	0.9084	0.9084	0.1501	1.2909	1.2909
1000	1.0069	1.9916	0.3132	0.7188	0.2812	0.3095	0.6959	0.3041
SE	0.0783	0.1413	0.0888	0.7523	0.7523	0.0955	0.6136	0.6136
5000	0.9995	1.9948	0.3047	0.5965	0.4035	0.3024	0.5946	0.4054
SE	0.0293	0.0618	0.043	0.0892	0.0892	0.0415	0.0926	0.0926
10000	0.9997	1.9922	0.3032	0.5847	0.4153	0.3018	0.5991	0.4009
SE	0.0197	0.0430	0.0283	0.0624	0.0624	0.0274	0.0657	0.0657
4) True values $\boldsymbol{\mu} = (4, 5), \alpha = 0.5, \boldsymbol{\phi} = \begin{bmatrix} 1 & 0 \\ 0.6 & 0.4 \end{bmatrix}, \mathbf{P}_{mat} = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}$								
500	4.0397	4.9501	0.5506	0.3952	0.6048	0.5466	0.4286	0.5714
SE	0.4132	0.4852	0.1469	0.1352	0.1352	0.1489	0.2029	0.2029
1000	4.0295	4.9785	0.5259	0.4237	0.5763	0.526	0.4309	0.5691
SE	0.3016	0.3590	0.105	0.0819	0.7523	0.1162	0.0945	0.0945
5000	4.0046	5.0008	0.5037	0.4163	0.5837	0.514	0.411	0.589
SE	0.1286	0.1601	0.0522	0.0408	0.0408	0.0492	0.0396	0.0396
10000	3.9947	4.9978	0.4982	0.4132	0.5868	0.5054	0.4135	0.5865
SE	0.0900	0.1173	0.0366	0.0278	0.0278	0.0354	0.0279	0.0279

Table 2: Estimates for $p = 3, r = 2$.

n_1	$\hat{\mu}_1^{YW}$	$\hat{\mu}_2^{YW}$	$\hat{\alpha}^{YW}$	$\hat{\phi}_{2,1}^{YW}$	$\hat{\phi}_{2,2}^{YW}$	$\hat{\phi}_{3,1}^{YW}$	$\hat{\phi}_{3,2}^{YW}$	$\hat{\phi}_{3,3}^{YW}$	$\hat{\alpha}^{YW}$	$\hat{\phi}_{3,1}^{YW}$	$\hat{\phi}_{3,2}^{YW}$	$\hat{\phi}_{3,3}^{YW}$
5) True values $\boldsymbol{\mu} = (1, 2), \alpha = 0.3, \boldsymbol{\phi} = \begin{bmatrix} 1 & 0 & 0 \\ 0.6 & 0.4 & 0 \\ 0.5 & 0.3 & 0.2 \end{bmatrix}, \mathbf{P}_{mat} = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$												
500	1.0167	1.9948	0.3185	2.1388	-1.1388	0.4742	0.2559	0.2699	0.2949	0.4969	0.3206	0.1825
SE	0.1181	0.2203	0.1656	8.6452	8.6452	0.9294	0.4239	0.6798	0.109	1.136	0.572	0.6893
1000	1.0076	1.9915	0.2954	0.6158	0.3842	0.5373	0.2636	0.199	0.3052	0.4992	0.2923	0.2085
SE	0.0813	0.1452	0.1052	1.0074	1.0074	0.2578	0.2562	0.2199	0.0853	0.1857	0.1793	0.1791
5000	0.9994	2.0023	0.2956	0.6182	0.3818	0.4983	0.2893	0.2124	0.2966	0.491	0.3073	0.2017
SE	0.0416	0.0717	0.0435	0.1311	0.1311	0.0659	0.0651	0.0686	0.0378	0.0809	0.0661	0.0735
10000	0.9978	1.9983	0.2956	0.6115	0.3885	0.5001	0.2964	0.2035	0.2998	0.4916	0.3121	0.1963
SE	0.0254	0.0509	0.0345	0.0935	0.0935	0.0437	0.0413	0.0489	0.0241	0.0522	0.0496	0.0482

Table 3: Estimates for $p = 3, r = 3$.

n_1	$\hat{\mu}_1^{YW}$	$\hat{\mu}_2^{YW}$	$\hat{\mu}_3^{YW}$	$\hat{\alpha}^{YW}$	$\hat{\phi}_{3,1}^{YW}$	$\hat{\phi}_{3,2}^{YW}$	$\hat{\phi}_{3,3}^{YW}$	$\hat{\alpha}^{YW}$	$\hat{\phi}_{3,1}^{YW}$	$\hat{\phi}_{3,2}^{YW}$	$\hat{\phi}_{3,3}^{YW}$	
6) True values $\boldsymbol{\mu} = (1, 1.5, 2), \alpha = 0.3, \boldsymbol{\phi} = \begin{bmatrix} 1 & 0 & 0 \\ 0.6 & 0.4 & 0 \\ 0.5 & 0.3 & 0.2 \end{bmatrix}, \mathbf{P}_{mat} = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.1 & 0.7 & 0.2 \\ 0.1 & 0.2 & 0.7 \end{bmatrix}$												
500	1.0068	1.4991	1.9849	0.3228	4.9236	-3.9236	-2.3274	5.6962	-2.3688	0.3205	0.4554	0.1098
SE	0.1617	0.1734	0.2385	0.1181	25.8252	25.8252	26.8363	53.7332	26.9268	0.1042	0.7894	1.2224
1000	1.0084	1.5031	1.9742	0.3156	0.6707	0.3293	0.5876	0.2652	0.1472	0.3096	0.5513	0.1157
SE	0.1130	0.1321	0.1704	0.0946	1.447	1.447	0.4793	0.3186	0.3543	0.0804	0.3914	0.4626
5000	0.9976	1.5018	1.9981	0.3092	0.6306	0.3694	0.4833	0.302	0.2147	0.2957	0.5002	0.1875
SE	0.0465	0.0544	0.0679	0.0435	0.1296	0.1296	0.0884	0.098	0.0854	0.0322	0.0951	0.0918
10000	0.9970	1.5024	1.9983	0.3058	0.5994	0.4006	0.4743	0.3106	0.2151	0.2955	0.4876	0.1967
SE	0.0329	0.0328	0.0510	0.0297	0.0796	0.0796	0.0492	0.0538	0.054	0.0232	0.0705	0.0639

6. APPLICATION

Quality of the processes introduced in this paper will be investigated by comparing the results obtained in the application of the various models to the same data. For this purpose, here we use two data sets of counts. Since the processes introduced in this article are not stationary, we expect them to perform well on the data chosen in [13]. So, in the first case, we choose this time series, which was created by counting drug offenses per month registered in the 27th police car beat in Pittsburg from January 1990 to December 2001. It has a length of 144 realizations and is downloaded from a website Forecasting Principles (<http://www.forecastingprinciples.com>). The plots of the given series and its autocorrelation and partial autocorrelation function are given in Figures 1, 3 and 5. Here, we might have used two different approaches for choosing the model order p . The first one, which is more intuitive, is based on choosing p as the number of first p significant values of the partial autocorrelation functions observed from the diagram (in this case Figure 5). The other approach is defining p as the value from $\{1, 2, \dots, q\}$ for which the smallest RMS value is obtained (RMS is the quality criterion explained later in this paragraph), where q is some reasonably large integer value. However, to make things easier to follow, we have decided to use the compromise of these two approaches. Namely, we choose the maximal considered model order p as the larger of the two numbers obtained by the first (intuitive) approach used for both data sets, increased by one. Since, these values for both data sets are 2 (for the 27th police car station) and 3 (for the second data set considered later in this section), we choose 4 as a maximal order of the INAR models considered for both of the observed counting time series. Therefore, INAR(p) models, for $p \in \{1, 2, 3, 4\}$, might be the reasonable choice. Considering the referent models of order 1, we chose INAR(1) model with Poisson marginals (PoINAR(1)) given in [2], quasi-binomial INAR(1) model with generalized Poisson marginals (GPQINAR(1)) from [5], geometric INAR(1) model (GINAR(1)) introduced in [4], new geometric INAR(1) (NGINAR(1)) defined in [14], negative binomial INAR(1) (NBINAR(1)) introduced in [19, 20], iterated INAR(1) model (NBIINAR(1)) with negative binomial marginals given in [1] and random coefficient INAR(1) model with negative binomial marginals (NBRCINAR(1)) constructed in [18]. Since our models, which quality we want to verify, are combinations of the RrNGINAR(1) process from [13] and CGINAR(p) process from [12] in some way, it is natural to include them in consideration, too. For this purpose we used R2NGINAR(1), R3NGINAR(1), CGINAR(2), CGINAR(3) and CGINAR(4) models. Another process of higher order which is included in this section, because of the completeness of the comparison, is PoINAR(p) ([16]), precisely, PoINAR(2), PoINAR(3) and PoINAR(4). The root mean squares (RMS) of differences between the observations and predicted values (using maximum likelihood estimation) are calculated and all the results are given in Table 4.

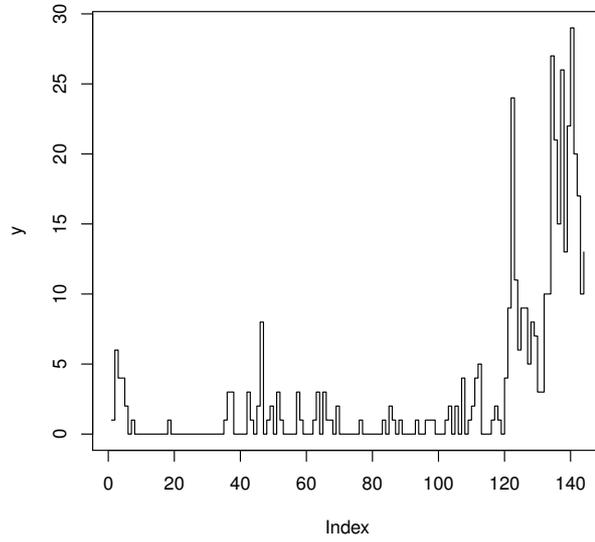


Figure 1: Drugs data from the 27th police station.

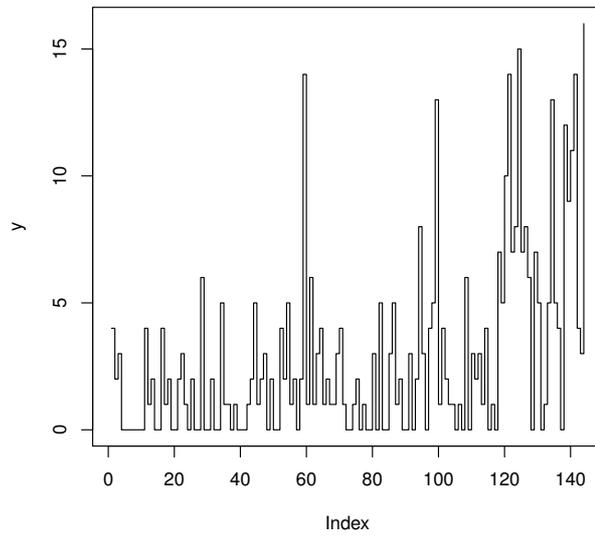


Figure 2: Drugs data from the 58th police station.

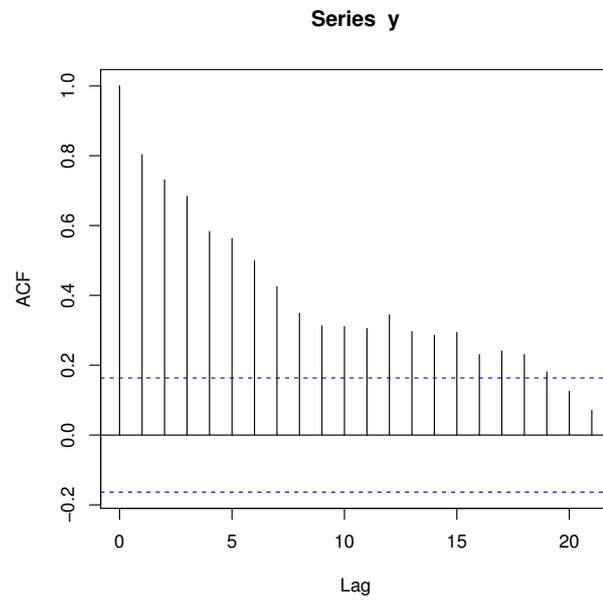


Figure 3: ACF for the data from the 27th police station.

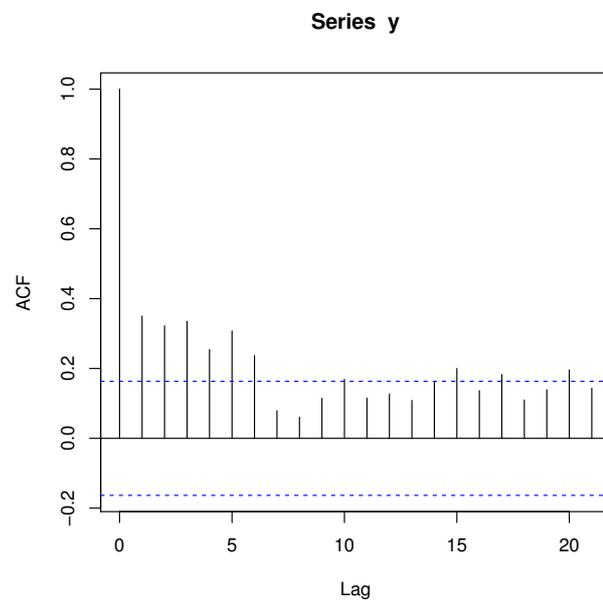


Figure 4: ACF for the data from the 58th police station.

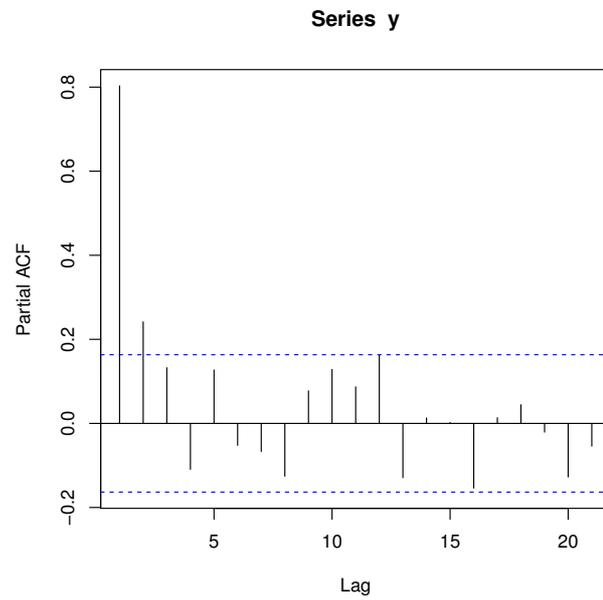


Figure 5: PACF for the data from the 27th police station.

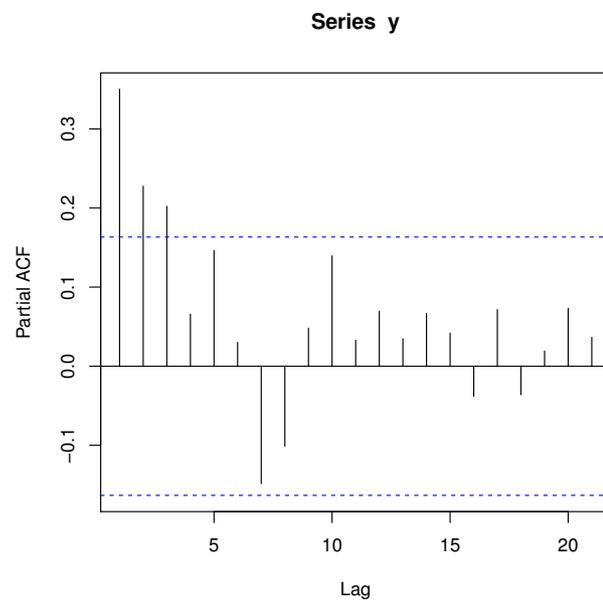


Figure 6: PACF for the data from the 58th police station.

Table 4: ML parameter estimates and RMS for different models for the data from the 27th police station.

Model	MLE	RMS
PoINAR(1)	$\hat{\lambda} = 1.237$ $\hat{\alpha} = 0.5948$	3.6613
GPQINAR(1)	$\hat{\lambda} = 0.5505$ $\hat{\theta} = 0.6108$ $\hat{\rho} = 0.392$	4.3398
GINAR(1)	$\hat{q} = 0.7596$ $\hat{\alpha} = 0.4809$	3.9456
NGINAR(1)	$\hat{\mu} = 3.3014$ $\hat{\alpha} = 0.7308$	3.4595
NBINAR(1)	$\hat{q} = 0.2173$ $\hat{\theta} = 0.834$ $\hat{\alpha} = 0.4563$	4.0185
NBIINAR(1)	$\hat{n} = 0.323$ $\hat{p} = 0.5335$ $\hat{\rho} = 0.8107$	3.4211
NBRCINAR(1)	$\hat{n} = 0.5435$ $\hat{p} = 0.1854$ $\hat{\rho} = 0.46$	4.0232
R2NGINAR(1)	$\hat{\mu}_1 = 1.1085$ $\hat{\mu}_2 = 12.9138$ $\hat{\alpha} = 0.052$	3.1090
R3NGINAR(1)	$\hat{\mu}_1 = 9.5906$ $\hat{\mu}_2 = 0.821$ $\hat{\mu}_3 = 23.249$ $\hat{\alpha} = 0.028$	1.6628
CGINAR(2)	$\hat{\mu} = 3.2042$ $\hat{\alpha} = 0.7473$	3.3801
CGINAR(3)	$\hat{\mu} = 3.1122$ $\hat{\alpha} = 0.745$	3.3749
CGINAR(4)	$\hat{\mu} = 3.124$ $\hat{\alpha} = 0.7464$	3.398
PoINAR(2)	$\hat{\lambda} = 0.9903$ $\hat{\alpha} = 0.687$	3.4822
PoINAR(3)	$\hat{\lambda} = 0.8616$ $\hat{\alpha} = 0.7449$	3.4253
PoINAR(4)	$\hat{\lambda} = 0.84$ $\hat{\alpha} = 0.7682$	3.4803

The lower values of RMS indicate the better and more appropriate models. The values of the maximum likelihood parameter estimates and RMS statistics are also calculated in case of application of our processes $RrNGINAR_{\max}(p)$ and $RrNGINAR_1(p)$ of appropriate orders 2, 3 and 4, taking into account the cases with two or three possible random states. (see Tables 5 and 6).

Table 5: ML parameter estimates and RMS for R2NGINARmax(p) and R3NGINARmax(p) process for the data from the 27th police station.

r	2	2	2
p	2	3	4
$\hat{\alpha}$	0.0368	0.0141	0.0556
$\hat{\mu}$	(12.2357, 0.6728)	(15.8556, 0.6226)	(19.8312, 1.3864)
$\hat{\phi}$	$\begin{bmatrix} 1.0000 & 0.0000 \\ 0.4601 & 0.5310 \end{bmatrix}$	$\begin{bmatrix} 1.0000 & 0.0000 & 0.0000 \\ 1.0000 & 0.0000 & 0.0000 \\ 0.3395 & 0.3670 & 0.2935 \end{bmatrix}$	$\begin{bmatrix} 1.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 1.0000 & 0.0000 & 0.0000 \\ 0.3310 & 0.3315 & 0.3375 & 0.0000 \\ 0.2384 & 0.2157 & 0.2493 & 0.2966 \end{bmatrix}$
RMS	3.2154	2.9089	2.8400
r	3	3	3
p	2	3	4
$\hat{\alpha}$	0.0296	0.0124	0.0241
$\hat{\mu}$	(23.2410, 9.0586, 0.7170)	(23.2500, 9.0570, 0.4382)	(23.2500, 9.0585, 0.7009)
$\hat{\phi}$	$\begin{bmatrix} 1.0000 & 0.0000 \\ 0.0432 & 0.9568 \end{bmatrix}$	$\begin{bmatrix} 1.0000 & 0.0000 & 0.0000 \\ 0.0000 & 1.0000 & 0.0000 \\ 0.3391 & 0.3226 & 0.3382 \end{bmatrix}$	$\begin{bmatrix} 1.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 1.0000 & 0.0000 & 0.0000 \\ 0.3300 & 0.3300 & 0.3401 & 0.0000 \\ 0.2477 & 0.1993 & 0.2383 & 0.3146 \end{bmatrix}$
RMS	1.6528	1.6532	1.6175

Table 6: ML parameter estimates and RMS for R2NGINAR₁(p) and R3NGINAR₁(p) process for the data from the 27th police station.

r	2	2	2
p	2	3	4
$\hat{\alpha}$	0.0368	0.0276	0.0226
$\hat{\mu}$	(12.2357, 0.6728)	(13.1349, 0.6075)	(13.6046, 0.5570)
$\hat{\phi}$	(0.4601, 0.5400)	(0.3958, 0.4267, 0.1775)	(0.3399, 0.3956, 0.1292, 0.1353)
RMS	3.2154	3.1079	3.0595
r	3	3	3
p	2	3	4
$\hat{\alpha}$	0.0296	0.0090	0.0093
$\hat{\mu}$	(23.2450, 9.0586, 0.7170)	(22.8019, 8.0055, 0.2286)	(22.7873, 7.9381, 0.2277)
$\hat{\phi}$	(0.0432, 0.9568)	(0.3828, 0.4293, 0.1878)	(0.3617, 0.4204, 0.1618, 0.0560)
RMS	1.6498	1.7446	1.7271

The same procedure of comparison of our processes to all the INAR models used above is also conducted in the second case of counting time series, i.e. on the drugs offenses counting data which were registered in the 58th police car beat in Pittsburg. The corresponding results are given by Figures 2, 4 and 6 and Tables 7, 8 and 9.

Table 7: ML parameter estimates and RMS for different models for the data from the 58th police station.

Model	MLE	RMS
PoINAR(1)	$\hat{\lambda} = 2.2349$ $\hat{\alpha} = 0.2189$	3.4011
GPQINAR(1)	$\hat{\lambda} = 1.0578$ $\hat{\theta} = 0.541$ $\hat{\rho} = 0.17$	3.4624
GINAR(1)	$\hat{q} = 0.7449$ $\hat{\alpha} = 0.1342$	3.4629
NGINAR(1)	$\hat{\mu} = 2.9157$ $\hat{\alpha} = 0.1734$	3.4315
NBINAR(1)	$\hat{q} = 0.2188$ $\hat{\theta} = 0.8033$ $\hat{\alpha} = 0.1155$	3.4789
NBIINAR(1)	$\hat{n} = 1$ $\hat{p} = 0.5$ $\hat{\rho} = 0.5$	3.4184
NBRCINAR(1)	$\hat{n} = 0.8442$ $\hat{p} = 0.2327$ $\hat{\rho} = 0.1827$	3.4247
R2NGINAR(1)	$\hat{\mu}_1 = 1.5485$ $\hat{\mu}_2 = 9.1053$ $\hat{\alpha} = 0.0521$	2.0096
R3NGINAR(1)	$\hat{\mu}_1 = 0.8719$ $\hat{\mu}_2 = 6.0089$ $\hat{\mu}_3 = 14.3936$ $\hat{\alpha} = 0.3012$	1.3361
CGINAR(2)	$\hat{\mu} = 2.9524$ $\hat{\alpha} = 0.3232$	3.3403
CGINAR(3)	$\hat{\mu} = 2.9326$ $\hat{\alpha} = 0.395$	3.2912
CGINAR(4)	$\hat{\mu} = 2.9517$ $\hat{\alpha} = 0.4241$	3.2769
PoINAR(2)	$\hat{\lambda} = 1.8966$ $\hat{\alpha} = 0.3534$	3.3152
PoINAR(3)	$\hat{\lambda} = 1.5812$ $\hat{\alpha} = 0.4779$	3.2438
PoINAR(4)	$\hat{\lambda} = 1.4452$ $\hat{\alpha} = 0.5521$	3.2236

In both cases of the observed offenses data the realization of the random environment process $\{z_n\}$ is determined in the same way as in [13], by clustering the data. Each cluster is assigned to a state. Then the corresponding sequence $\{p_n\}$ is calculated using the definitions of the models. The plots of the clusterings (Figures 7, 8, 9 and 10) show big difference between the data recorded by the police stations in the way of the environment state changing. For the data from

Table 8: ML parameter estimates and RMS for R2NGINARmax(p) and R3NGINARmax(p) process for the data from the 58th police station.

r	2	2	2
p	2	3	4
$\hat{\alpha}$	0.1328	0.1322	0.1317
$\hat{\mu}$	(8.3430, 1.2408)	(8.3422, 1.2347)	(8.3420, 1.2307)
$\hat{\phi}$	$\begin{bmatrix} 1.0000 & 0.0000 \\ 0.0967 & 0.9033 \end{bmatrix}$	$\begin{bmatrix} 1.0000 & 0.0000 & 0.0000 \\ 0.0511 & 0.9489 & 0.0000 \\ 0.3366 & 0.3414 & 0.3220 \end{bmatrix}$	$\begin{bmatrix} 1.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0574 & 0.9426 & 0.0000 & 0.0000 \\ 0.3290 & 0.3340 & 0.3371 & 0.0000 \\ 0.2690 & 0.2194 & 0.2450 & 0.2667 \end{bmatrix}$
RMS	2.0759	2.0785	2.0821
r	3	3	3
p	2	3	4
$\hat{\alpha}$	0.0488	0.0488	0.0488
$\hat{\mu}$	(4.6800, 12.8180, 0.6746)	(4.6798, 12.8181, 0.6747)	(4.6798, 12.8181, 0.6748)
$\hat{\phi}$	$\begin{bmatrix} 1.0000 & 0.0000 \\ 0.0483 & 0.9517 \end{bmatrix}$	$\begin{bmatrix} 1.0000 & 0.0000 & 0.0000 \\ 0.0154 & 0.9846 & 0.0000 \\ 0.3388 & 0.3363 & 0.3250 \end{bmatrix}$	$\begin{bmatrix} 1.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0070 & 0.9930 & 0.0000 & 0.0000 \\ 0.3341 & 0.3295 & 0.3364 & 0.0000 \\ 0.2557 & 0.2094 & 0.2404 & 0.2945 \end{bmatrix}$
RMS	1.1813	1.1764	1.1795

Table 9: ML parameter estimates and RMS for R2NGINAR₁(p) and R3NGINAR₁(p) process for the data from the 58th police station.

r	2	2	2
p	2	3	4
$\hat{\alpha}$	0.1328	0.1321	0.1321
$\hat{\mu}$	(8.3430, 1.2408)	(8.3425, 1.2345)	(8.3421, 1.2340)
$\hat{\phi}$	(0.0967, 0.9033)	(0.3363, 0.4347, 0.2290)	(0.3283, 0.3309, 0.2725, 0.0683)
RMS	2.0536	2.0732	2.0784
r	3	3	3
p	2	3	4
$\hat{\alpha}$	0.0488	0.0491	0.0491
$\hat{\mu}$	(4.6800, 12.8180, 0.6746)	(4.6801, 12.8181, 0.6782)	(4.6804, 12.8181, 0.6789)
$\hat{\phi}$	(0.0483, 0.9517)	(0.3193, 0.4168, 0.2640)	(0.2732, 0.2719, 0.25489, 0.2001)
RMS	1.1889	1.1760	1.1793

the 27th police station probability of staying in the same state is much higher than for the observations from the 58th station. Analysis of the results lead us to a conclusion that in both cases of the selected data, the models introduced in this paper are better than the others which we applied. Namely, in the case of the 27th car beat drug offenses, R3NGINARmax(4) is the most appropriate model, while in the case of the data recorded by the 58th police station, R3NGINAR₁(3) process shows the best performance. It is interesting to note that the optimal order 3 obtained for the 58th police station data is in accordance with the value obtained by the graphical intuitive (the first one) approach for choosing order p .

However, in the case of the 27th police station this is not the case (where graphical approach gives $p = 2$), which justify the usage of our compromise approach for choosing order of the model.

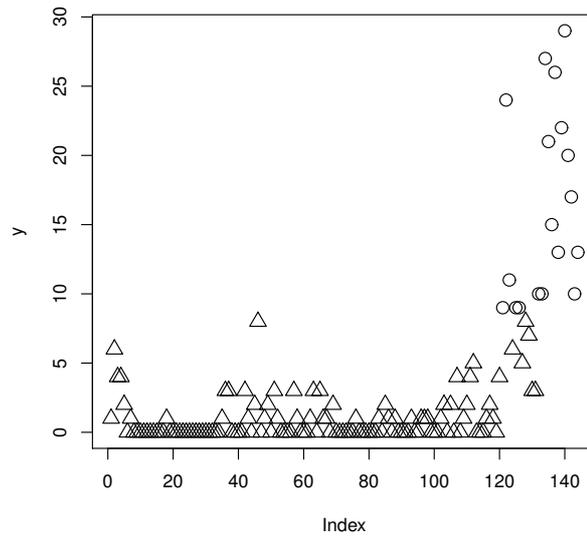


Figure 7: Clusters for two states for the 27th police station.

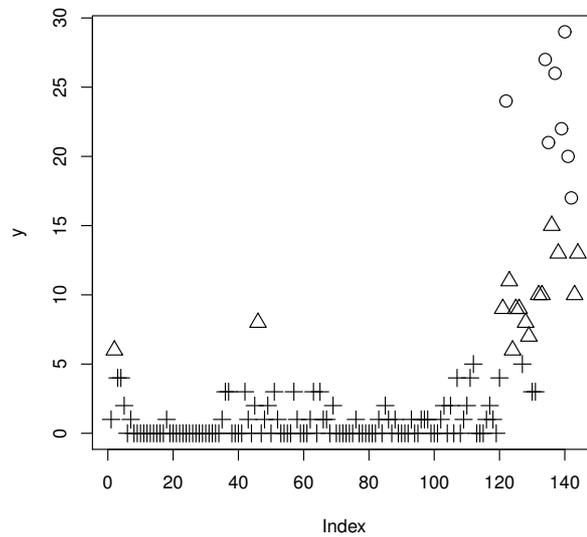


Figure 8: Clusters for three states for the 27th police station.

Successful performance of our Random Environment INAR models in both of the cases show that they might be very appropriate for the processes which quite often change their marginal distribution (58th car beat data counting),

as well as for the processes which are, on the other hand, much more passive (27th police car beat data), i.e. which only rarely shift from one set of environment circumstances to another. Increase of the number of random states contributes to the improving of the results for both models in each case of the observed data. However, optimal order depends on the data. Thus, for the 27th station the more appropriate Random Environment INAR models are mostly the ones with the higher order, while for the 58th police station we have the opposite situation.

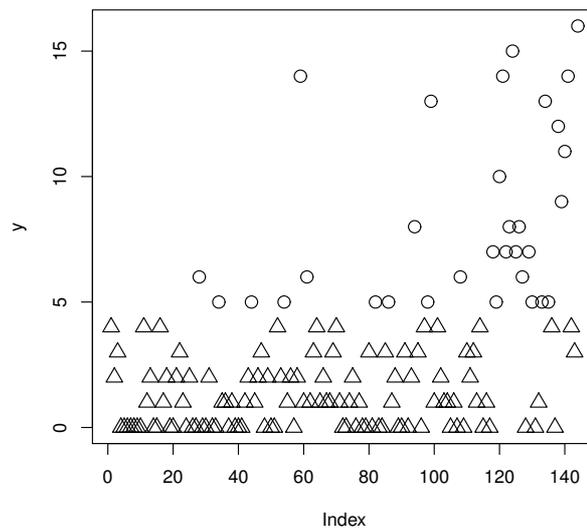


Figure 9: Clusters for two states for the 58th police station.

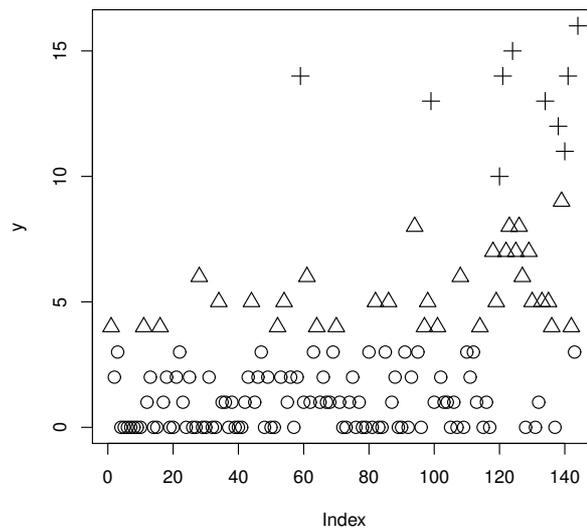


Figure 10: Clusters for three states for the 58th police station.

REFERENCES

- [1] AL-OSH, M.A. and ALY, E.E.A.A. (1992). First order autoregressive time series with negative binomial and geometric marginals, *Commun. Statist. Theory Meth.*, **21**, 2483–2492.
- [2] AL-OSH, M.A. and ALZAID, A.A. (1987). First-order integer-valued autoregressive (INAR(1)) process, *J. Time Ser. Anal.*, **8**, 261–275.
- [3] ALY, E.E.A.A. and BOUZAR, N. (1994). On some integer-valued autoregressive moving average models, *Journal of Multivariate Analysis*, **50**, 132–151.
- [4] ALZAID, A.A. and AL-OSH, M.A. (1988). First-order integer-valued autoregressive (INAR(1)) process: distributional and regression properties, *Statistica Neerlandica*, **42**, 53–61.
- [5] ALZAID, A.A. and AL-OSH, M.A. (1993). Some autoregressive moving average processes with generalized Poisson marginal distributions, *Ann. Inst. Statist. Math.*, **45**, 223–232.
- [6] BAKOUCH, H.S. and RISTIĆ, M.M. (2010). Zero truncated Poisson integer valued AR(1) model, *Metrika*, **72**(2), 265–280.
- [7] BEGLEITER, R.; EL-YANIV, R. and YONA, G. (2004). On prediction using variable order Markov models, *Journal of Artificial Intelligence Research*, **22**, 385–421.
- [8] LATOUR, A. (1998). Existence and stochastic structure of a non-negative integer-valued autoregressive process, *J. Time Ser. Anal.*, **19**, 439–455.
- [9] MCKENZIE, E. (1985). Some simple models for discrete variate time series, *Water Resour. Bull.*, **21**, 645–650.
- [10] MCKENZIE, E. (1986). Autoregressive moving-average processes with negative binomial and geometric distributions, *Adv. Appl. Prob.*, **18**, 679–705.
- [11] NASTIĆ, A.S. and RISTIĆ, M.M. (2012). Some geometric mixed integer-valued autoregressive (INAR) models, *Statist. Probab. Lett.*, **82**, 805–811.
- [12] NASTIĆ, A.S.; RISTIĆ, M.M. and BAKOUCH, H.S. (2012). A combined geometric INAR(p) model based on negative binomial thinning, *Mathematical and Computer Modelling*, **55**, 1665–1672.
- [13] NASTIĆ, A.S.; LAKETA, P.N. and RISTIĆ, M.M. (2016). Random environment integer-valued autoregressive process, *J. Time Ser. Anal.*, **37**, 267–287.
- [14] RISTIĆ, M.M.; BAKOUCH, H.S. and NASTIĆ, A.S. (2009). A new geometric first-order integer-valued autoregressive (NGINAR(1)) process, *J. Stat. Plann. Inf.*, **139**, 2218–2226.
- [15] RISTIĆ, M.M. and NASTIĆ, A.S. (2012). A mixed INAR(p) model, *J. Time Ser. Anal.*, **33**, 903–915.
- [16] WEISS, C.H. (2008). The combined INAR(p) models for time series of counts, *Statist. Probab. Lett.*, **72**, 1817–1822.
- [17] ZHENG, H.; BASAWA, I.V. and DATTA, S. (2006). Inference for p th-order random coefficient integer-valued autoregressive processes, *J. Time Ser. Anal.*, **27**, 411–440.

- [18] ZHENG, H.; BASAWA, I.V. and DATTA, S. (2007). First-order random coefficient integer-valued autoregressive processes, *J. Stat. Plann. Inf.*, **137**, 212–229.
- [19] ZHU, R. and JOE, H. (2006). Modelling count data time series with Markov processes based on binomial thinning, *J. Time Ser. Anal.*, **27**, 727–738.
- [20] ZHU, R. and JOE, H. (2010). Negative binomial time series models based on expectation thinning operators, *J. Stat. Plann. Inf.*, **140**, 1874–1888.

STOCHASTIC INEQUALITIES FOR THE RUN LENGTH OF THE EWMA CHART FOR LONG-MEMORY PROCESSES

Authors: YAREMA OKHRIN
– Department of Statistics, University of Augsburg,
Germany
yarema.okhrin@wiwi.uni-augsburg.de

WOLFGANG SCHMID
– Department of Statistics, European University Viadrina,
Frankfurt (Oder), Germany
schmid@europa-uni.de

Received: November 2015 Revised: February 2017 Accepted: February 2017

Abstract:

- In this paper the properties of the modified EWMA control chart for detecting changes in the mean of an ARFIMA process are discussed. The central question is related to the false alarm probability and its behavior for different autocorrelation structures and parameters of the underlying process. It is shown under which conditions the false alarm probability of an ARFIMA(p, d, q) process is larger than that of the pure ARFIMA($0, d, 0$) process. Furthermore, it is shown that the false alarm probability for ARFIMA($0, d, 0$) and ARFIMA($1, d, 1$) is monotonic in d for common parameter values of the processes.

Key-Words:

- *statistical process control; EWMA control chart; long-memory process; ARFIMA process.*

AMS Subject Classification:

- 62L10, 62M10.

1. INTRODUCTION

Detection of structural changes in a time series is an important issue and is actively tackled in the current literature. The objective is to detect the change as soon as possible after it has occurred. The change is an indication that something important has happened and the characteristics of the process have drifted from the original values. This type of questions is of particular relevance in engineering, public health, finance, environmental sciences (see, e.g., Montgomery ([18]), Lawson and Kleinman ([15]), Frisén ([7])).

The most frequently applied surveillance technique is a control chart (e.g., Montgomery ([18])), with Shewhart, EWMA and CUSUM charts being the most popular ones. Initially they were developed for monitoring independent processes. Since in many applications the process of interest appears to be time dependent, several approaches evolved to extend the above schemes to time series. One approach relies on monitoring the residuals of the fitted time series process (e.g., Alwan and Roberts ([1]), Montgomery and Mastrangelo ([19]), Wardell *et al.* ([32]) and ([33]), Lu and Reynolds ([16])). This makes, however, the interpretation of the signal given by the control scheme difficult. Furthermore, the estimated residuals are not independent after a change, implying that the use of the classical charts is still erroneous (cf. Knoth and Schmid ([14])). Alternatively, we can adjust the monitoring schemes to reflect the dependence structure of the analyzed processes. This type of charts are called modified charts. The modified CUSUM charts and the related generalized likelihood ratio tests were discussed, for example, in Nikiforov ([21]), Yashchin ([34]), Schmid ([28]), Knoth and Schmid ([14]), Capizzi and Masarotto ([5]), Knoth and Frisén ([13]). The extension of the EWMA chart to time series data was suggested by Schmid ([27]). Note that the derivation of modified schemes is technically tedious, since the autocorrelation structure of the process should be explicitly taken into account while determining the design parameters of the monitoring procedure. Furthermore, most of the literature with just a few exceptions considers the ARMA processes. These processes are of great importance in practice, but assume inherently a short memory in the underlying data. Recently, Rabyk and Schmid ([25]) considered several control charts for long memory processes and compared these schemes within an extensive Monte Carlo study.

It is desirable for any control scheme to give a signal as soon as possible after the change has occurred, i.e. the process is out-of-control, and to give a signal as rarely as possible if no change occurred, i.e. the process is in-control. False alarms deteriorate the surveillance procedures and lead to potentially misleading inferences by practitioners. The performance of the chart in the in-control state can be quantified by the false alarm probability up to a given time point or, equivalently, the probability that the run length of the chart is longer than a

given time span. Of particular importance is the impact of the process parameters on this probability. For a general family of linear processes and particularly for ARMA processes several results on stochastic ordering of the run length can be found in Schmid and Okhrin ([29]) and Morais *et al.* ([20]). In these papers the authors derive constraints on the autocorrelations of the observed in-control process to guarantee stochastic monotonicity of the EWMA or more general monitoring schemes. The case of nonlinear time series was treated by Pawlak and Schmid ([24]) and Gonçalves *et al.* ([9]).

The subject of the analysis in this paper is the one-sided EWMA chart aimed to detect an increase of the mean. The one-sided problem is of key importance in many fields, such as engineering (loading capacity, tear strength, etc.), environmental sciences (high tide, concentration of particulate matter, ozone level, etc), economics and finance (riskiness of financial assets, interest and unemployment rates). In all these examples we are interested only in increases (or only in decreases) of the quantity of interest. Its dynamics can be assessed with the tools considered here.

In this paper we discuss the stochastic ordering for the false alarm probability of modified one-sided EWMA control charts aimed to detect a shift in the mean of an ARFIMA process. First, we show that for an arbitrary ARFIMA(p, d, q) process the probability of a false signal is always larger than this probability for an i.i.d. or an ARFIMA($0, d, 0$) processes. To guarantee this it suffices to assume that the autocorrelation of the underlying ARMA process is always non-negative. Second, we extend the above results by showing that the false alarm probabilities are non-decreasing functions in d for ARFIMA($0, d, 0$) and ARFIMA($1, d, 1$) under specific assumptions on the process parameters. These results are of great importance, since it is well known that the parameter of fractional differencing is difficult to estimate. Thus we indicate the consequences of under- or overestimation of d for monitoring procedures.

The paper is structured as follows. Section 2 summarizes relevant results on modified EWMA control charts and on ARFIMA processes. The main results together with numerical examples and counterexamples are given in Section 3. The proofs of some results are given in the appendix.

2. THE MODIFIED EWMA CHART FOR ARFIMA PROCESSES

The aim of statistical process control is to detect structural deviations in a process over time. We assume that at each time point one observation is available. The given observations x_1, x_2, \dots are considered to be a realization of the actual (observed) process. The underlying target process is denoted by $\{Y_t\}$.

The objective of a monitoring procedure is to give a signal if the target and the observed processes differ in their characteristics. A good procedure should give a signal as soon as possible if the processes differ and give a signal as rarely as possible if the processes coincide. In the following we analyze the behavior of the modified EWMA control chart for the mean in the in-control case, i.e. if no change is present. The underlying target process is assumed to be a long-memory process, which frequently encounters in applications, for example stock market risk in finance or environmental data (see Andersen *et al.* ([2]), Pan and Chen ([23])). The objective of the paper is to analyze the performance of in-control EWMA charts for different memory patterns of the long memory processes. First we introduce the control scheme and subsequently discuss the process and its features in detail.

2.1. The modified EWMA chart

The exponentially weighted moving average (EWMA) chart was introduced by Roberts ([26]). Contrary to the Shewhart chart all previous observations are taken into account for the decision rule. It turns out to perform better than the Shewhart control chart for detecting small and moderate shifts, namely, in the process mean (Lucas and Saccucci ([17])) or in the process variance (Crowder and Hamilton ([6])) of an independent output. An extension of the EWMA control chart to time series was given by Schmid ([27]).

The EWMA chart for monitoring the process mean is based on the statistic

$$(2.1) \quad Z_t = (1 - \lambda)Z_{t-1} + \lambda X_t, \quad t \geq 1.$$

Z_0 is the starting value. Here we choose it equal to the mean of the target process, i.e. $Z_0 = E(Y_t) = \mu_0$. The process starts in zero state, a head-start is not considered. The parameter $\lambda \in (0, 1]$ is a smoothing constant determining the influence of past observations.

The quantity Z_t can be written as weighted average

$$(2.2) \quad Z_t = \lambda \sum_{i=0}^{t-1} (1 - \lambda)^i X_{t-i} + (1 - \lambda)^t Z_0, \quad t = 1, 2, \dots,$$

whose weights decrease geometrically. This shows that if λ is close to one then we have a short memory EWMA chart while for λ close to zero the preceding values get a larger weight. For $\lambda = 1$ the EWMA chart reduces to the Shewhart control chart.

Further $\{Y_t\}$ is assumed to be a stationary process with mean μ_0 and au-

tocovariance function $\gamma(k)$. Then (see Schmid ([27])) $E(Z_t) = \mu_0$ and

$$\text{Var}(Z_t) = \lambda^2 \sum_{|i| \leq t-1} \gamma(i) \sum_{j=\max\{0, -i\}}^{\min\{t-1, t-1-i\}} (1-\lambda)^{2j+i} = \sigma_{e,t}^2.$$

In this paper we consider a one-sided EWMA chart. Our aim is to detect an increase of the mean. The process is concluded to be out of control at time point t if

$$Z_t > \mu_0 + c\sqrt{\text{Var}(Z_t)}$$

with $c > 0$. The run length of the EWMA control chart is given by

$$N_e = \inf \left\{ t \in \mathbb{N} : Z_t > \mu_0 + c\sqrt{\text{Var}(Z_t)} \right\}.$$

Stochastic inequalities for the modified EWMA chart have been given in Schmid and Schöne ([30]) and Schöne *et al.* ([31]). Assuming that $\{Y_t\}$ is a stationary Gaussian process with non-negative autocovariances Schmid and Schöne ([30]) showed that the probability of a false signal up to a certain time point k , i.e. $P(N_e > k)$ is greater or equal to that in the i.i.d. case. Thus the dependence structure leads to an increase of the false alarm probability. Demanding further assumptions on the autocovariances Schöne *et al.* ([31]) proved that the false alarm probability is an increasing function in the autocorrelations provided that they satisfy a certain monotonicity condition.

2.2. The target process

Throughout this paper the target process is assumed to be a stationary autoregressive fractionally integrated moving average (ARFIMA) process. In many applications we are faced with processes having a long memory. The frequently applied autoregressive moving average (ARMA) modeling is not suitable in such a situation as its autocorrelation structure is geometrically decreasing.

Let L denote the lag operator, i.e. $LY_t = Y_{t-1}$ and let $\Delta = 1 - L$ be the difference operator, i.e. $\Delta Y_t = Y_t - Y_{t-1}$. In the study of non-stationary time series more generalized ARIMA(p,d,q) models are often used (cf. Box *et al.* ([3])). They make use of a d -multiple difference operator Δ to the original time series Y_t where d is a non-negative integer. In the approach of Granger and Joyeux ([10]) and Granger ([11]), however, d is a real number.

Let $d > -1$ then Granger and Joyeux ([10]) and Granger ([11]) define Δ^d using the binomial expansion as

$$\Delta^d = (1 - L)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k L^k = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)} L^k,$$

where $\Gamma(\cdot)$ is the gamma function. Let

$$A(L) = 1 - \alpha_1 L - \dots - \alpha_p L^p, \quad B(L) = 1 + \beta_1 L + \dots + \beta_q L^q$$

and $\{\varepsilon_t\}$ be a white noise process, i.e.

$$E(\varepsilon_t) = 0, \quad \text{Var}(\varepsilon_t) = \sigma^2, \quad \text{Cov}(\varepsilon_t, \varepsilon_s) = 0 \quad \forall t \neq s.$$

Now $\{Y_t\}$ is said to be an autoregressive fractionally integrated moving average process of order (p, d, q) (ARFIMA(p,d,q)) if $\{Y_t\}$ is stationary and satisfies the equation

$$(2.3) \quad A(L)\Delta^d(Y_t - \mu_0) = B(L)\varepsilon_t$$

for $d \in (-0.5, 0.5)$.

The condition on the existence and uniqueness of a stationary solution of an ARFIMA process is given in the following theorem.

Theorem 2.1. *Suppose that $\{Y_t\}$ is an ARFIMA(p,d,q) process as defined in (3). Let $d \in (-0.5, 0.5)$ and $A(\cdot)$ and $B(\cdot)$ have no common zeroes.*

- a) *If $A(z) \neq 0$ for $|z| = 1$ then there is a unique purely nondeterministic stationary solution of (3) given by*

$$(2.4) \quad Y_t = \mu_0 + \sum_{j=-\infty}^{\infty} \psi_j \Delta^{-d} \varepsilon_{t-j},$$

$$\text{where } \psi(z) = \sum_{j=-\infty}^{\infty} \psi_j z^j = B(z)/A(z).$$

- b) *The solution $\{Y_t\}$ is causal if and only if $A(z) \neq 0$ for $|z| \leq 1$.*

Proof: Brockwell and Davis ([4]). □

The parameter of fractional differencing d determines the strength of the process memory. Since $\rho(k) \sim ck^{2d-1}$ as $k \rightarrow \infty$ with $c \neq 0$ ARFIMA processes have a long memory for $d \in (0, 0.5)$. For $d \in (-0.5, 0)$ the process is called to have an intermediate memory. ARMA processes are referred to as short memory processes since $|\rho(k)| \leq Cr^{-k}$ for $k = 0, 1, \dots$ with $C > 0$ and $0 < r < 1$.

The knowledge of the autocovariance function of an ARFIMA(p,d,q) process is crucial for the application of the monitoring techniques discussed in this paper and their theoretical properties. Nevertheless, it is difficult to obtain explicit formulas for the autocovariance and the autocorrelation functions. The

autocovariance function of an ARFIMA(0,d,0) process was derived by Hosking ([12]). It holds that (see, e.g., Brockwell and Davis ([4], Theorem 13.2.1))

$$(2.5) \quad \gamma_d(0) = \sigma^2 \frac{\Gamma(1-2d)}{(\Gamma(1-d))^2}, \quad \gamma_d(k) = \gamma_d(0) \rho_d(k) \quad k \in \mathbb{Z}$$

where

$$(2.6) \quad \begin{aligned} \rho_d(k) &= \frac{\Gamma(k+d)\Gamma(1-d)}{\Gamma(k-d+1)\Gamma(d)} \\ &= \prod_{i=1}^k \frac{i-1+d}{i-d} = \prod_{i=1}^k \left(1 - \frac{1-2d}{i-d}\right), \quad k = 1, 2, \dots \end{aligned}$$

and $\rho_d(-k) = \rho_d(k)$.

To determine the autocovariance function of a general ARFIMA process it is convenient to deploy the splitting method. This method is based on the decomposition of the ARFIMA model into its ARMA and its fractionally integrated parts. Let $\gamma_{ARMA}(\cdot)$ be the autocovariance of the ARMA component which has a unit variance white noise and let $\gamma_d(\cdot)$ denote the autocovariance of the ARFIMA(0,d,0) process given by (2.5) and (2.6). If the conditions of Theorem 2.1a) are satisfied then the autocovariance of the corresponding ARFIMA process is given by the convolution of these two functions (see, e.g., Palma ([22]), Brockwell and Davis ([4], p.525, (13.2.19))

$$(2.7) \quad \gamma(k) = \sum_{i=-\infty}^{\infty} \gamma_d(i) \gamma_{ARMA}(k-i).$$

This result is obvious since $\{Y_t\}$ is an ARFIMA(p,d,q) process if and only if $\Delta^d(Y_t - \mu_0)$ is an ARMA(p,q) process.

In the following we shall frequently consider processes with $\gamma_{ARMA}(i) \geq 0$ for all $i \geq 1$, $d \in (0, 0.5)$ and $\sigma > 0$. Then it holds that

$$(2.8) \quad \gamma(k) \geq \gamma_d(k) \gamma_{ARMA}(0) > 0$$

since $\gamma_d(k) > 0$ and $\gamma_{ARMA}(0) > 0$. Thus the autocovariance function of a general ARFIMA process is strictly positive for all k , if the autocovariances of the underlying ARMA process are non-negative.

Next we consider the special case of an ARFIMA(1,d,1) process and use the simplified notation $\alpha = \alpha_1$ and $\beta = \beta_1$. It holds that its autocovariance function is

$$(2.9) \quad \gamma(k) = \sigma^2 \sum_{i=-\infty}^{\infty} \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} \frac{\Gamma(i+d)}{\Gamma(1+i-d)} \gamma_{ARMA}(k-i),$$

where $\gamma_{ARMA}(k)$ is the autocovariance of the ARMA(1,1) process, i.e.

$$\gamma_{ARMA}(k) = \begin{cases} \frac{1+2\alpha\beta+\beta^2}{1-\alpha^2} & \text{for } k = 0 \\ \frac{(1+\alpha\beta)(\alpha+\beta)}{1-\alpha^2} \alpha^{|k|-1} & \text{for } k \neq 0 \end{cases}$$

3. MONOTONICITY RESULTS FOR THE MODIFIED EWMA MEAN CHART

In the following we consider the probability of a false signal assuming that the underlying process is an ARFIMA process. It is always assumed that the process is in control. We use the notation $P_{(p,d,q)}$ to denote that the probability is calculated assuming that the underlying process is an ARFIMA(p,d,q) process. Note that for $P_{(0,0,0)}$ the in-control process is assumed to be independent and identically distributed random sequence and for $P_{(0,d,0)}$ it is a pure ARFIMA(0,d,0) process.

In this section it is always demanded that the variance of the white noise σ^2 is positive. In the case $\sigma = 0$ the process $\{Y_t\}$ is deterministic, $N_e = \infty$ and Theorem 3.1, Lemma 3.1, Theorem 3.2 and Theorem 3.3 hold without any further assumption.

3.1. Influence of the ARMA structure on the false alarm probability

First, it is proved that for an ARFIMA(p,d,q) process the probability of a false signal up to a fixed time point k is always greater equal to the corresponding probability for an independent random process. This result is an immediate consequence of Schmid and Schöne ([30]).

Theorem 3.1. *Let $\{Y_t\}$ be an ARFIMA(p,d,q) process as defined in (3) with $d \in [0, 0.5)$ and let $\{\varepsilon_t\}$ be a Gaussian white noise with $\sigma > 0$. Let $A(z) \neq 0$ for $|z| = 1$ and let $\gamma_{ARMA}(v) \geq 0$ for all $v \in \mathbb{Z}$. Then*

$$P_{(p,d,q)}(N_e(c) > k) \geq P_{(0,0,0)}(N_e(c) > k), \quad k = 0, 1, 2, \dots$$

Proof: Because of (2.7) we get that $\gamma(v) \geq 0$ for all v . Since $\{Y_t\}$ is a Gaussian process the result is an immediate consequence of Theorem 1 of Schmid and Schöne ([30]). \square

This result gives a lower bound for the probability of a false signal. The bound itself is the probability of a false signal for an i.i.d. random sequence.

One of the crucial assumptions of Theorem 1 of Schmid and Schöne ([30]) is that all autocovariances are non-negative. Here we illustrate that the inequality may not hold for negative autocovariances. The left picture in Figure 1 shows the autocorrelations up to lag 5 of an ARFIMA(0,d,0) process. We see that for positive d 's the autocorrelations are large and positive, but they become negative

and small for negative d 's. The right-hand side picture shows the probability of no signal up to the time point k . The probabilities are computed by numerical integration of the $k + 1$ -dimensional normal density with the covariance matrix determined using Lemma 1 of Schmid and Schöne ([30]). The integration utilizes the Genz-Bretz algorithm (see Genz ([8])). The solid black line stands for the case $d = 0$ and thus for the i.i.d. process. Positive d 's (grey lines) induce positive autocovariances, fulfill the assumptions of Theorem 3.1 and lead to probabilities larger than for the i.i.d. case. However, the negative autocorrelations for $d < 0$ (black lines) lead to smaller probabilities of no signal. Thus we have a counterexample for the case if the assumptions of Theorem 3.1 are not satisfied.

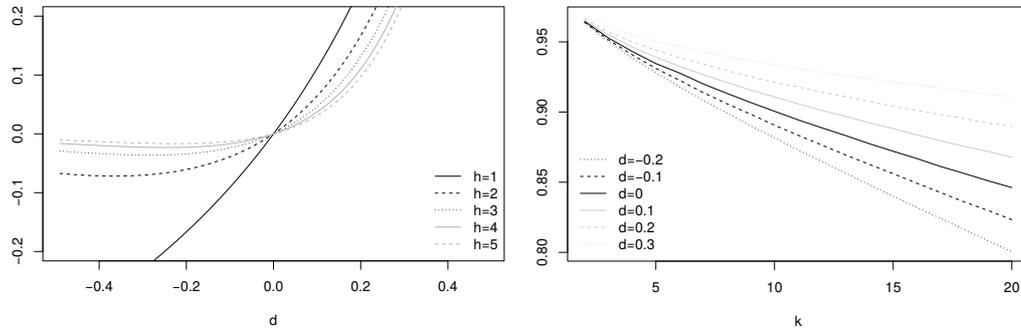


Figure 1: The first five autocorrelations $\rho_d(h)$ (left side) and the probability of no signal up to the time point k (right side) for the modified EWMA chart with $\lambda = 0.1$ and $c = 2.04$ applied to an ARFIMA(0, d ,0) process.

It is important to note that in the case $\gamma_{ARMA}(v) \leq 0$ for all $v \in \mathbb{Z}$ the inequality in Theorem 3.1 does not with the reversed inequality. This is illustrated in Figure 2. It is shown that for an ARFIMA(1, d ,0) process with $d = 0.2$ and $\alpha = -0.5$ the probability of no signal up to time point k is sometimes larger and sometimes smaller than that of the i.i.d. case depending on the choice of k .

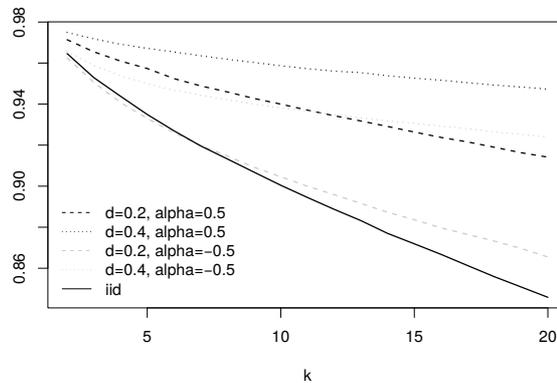


Figure 2: The probability of no signal up to the time point k for the modified EWMA chart with $\lambda = 0.1$ and $c = 2.04$ applied to an ARFIMA(1, d ,0) process.

Next, we try to improve the lower bound. It is analyzed for which ARFIMA(p,d,q) processes it can be replaced by the probability of a false signal of an ARFIMA(0,d,0) process.

Let $\{\rho_{ARMA}(h)\}$ denote the autocorrelation function of an ARMA(p,q) process and let $\{\rho_d(h)\}$ denote the autocorrelation function of an ARFIMA(0,d,0) process.

Let $k \in \mathbb{N}$, $1 \leq v \leq k-1$ and

$$\begin{aligned} I_v &= \sum_{i=1}^{v-1} i \rho_{ARMA}(i) \left(\frac{\rho_d(v-1+i)}{v+i-d} - \frac{\rho_d(v-1-i)}{v-i-d} \right), \\ II_v &= v \rho_{ARMA}(v) \left(\frac{\rho_d(2v-1)}{2v-d} + \frac{1}{1-d} \right), \\ III_v &= \sum_{i=v+1}^{\infty} i \rho_{ARMA}(i) \left(\frac{\rho_d(i-v)}{i-v+1-d} + \frac{\rho_d(i+v-1)}{v+i-d} \right). \end{aligned}$$

Lemma 3.1. *Let $k \in \mathbb{N}$ and let $\{Y_t\}$ be an ARFIMA(p,d,q) process as defined in (3) with $d \in [0, 0.5)$ and let $\{\varepsilon_t\}$ be a Gaussian white noise with $\sigma > 0$. Let $A(z) \neq 0$ for $|z| = 1$ and $\gamma_{ARMA}(v) \geq 0$ for all v . If additionally*

$$(3.1) \quad I_v + II_v + III_v \geq 0, \quad v = 1, \dots, k-1$$

then $P_{(p,d,q)}(N_e(c) > k) \geq P_{(0,d,0)}(N_e(c) > k)$.

Proof: See Appendix. □

Keeping in mind the conditions of Lemma 3.1, it can be seen that II_v and III_v are non-negative while I_v is non-positive because

$$\frac{\rho_d(v-1+i)}{v+i-d} \leq \frac{\rho_d(v-1+i)}{v-i-d} \leq \frac{\rho_d(v-1-i)}{v-i-d}$$

using Lemma A.1c of the Appendix. Thus it is not clear for which processes this condition is fulfilled at all. Next the condition (3.1) is analyzed for various processes.

Lemma 3.2. *Suppose that the conditions of Lemma 3.1 are fulfilled. Then it holds that:*

- a) For $k = 2$ condition (3.1) is always fulfilled.
- b) For $k = 3$ condition (3.1) is satisfied if $2\rho_{ARMA}(2) \geq \rho_{ARMA}(1)$.
- c) Let $k \geq 2$. If $\{Y_t\}$ is an ARFIMA(1,d,0) process with autoregressive coefficient $\alpha \in [0, 1)$ and $\alpha \geq (k-2)/(k-1)$ then condition (3.1) is satisfied.

Proof: See Appendix. □

Note that Lemma 3.1 is an extension of Theorem 3.1. It shows that the presence of the ARMA part with specific parameters (see Lemma 3.2) leads to an increase of the probability of no signal. The condition in b) does not hold for an ARFIMA(0,d,1) process with positive coefficient. As it is shown in the left picture in Figure 3 for $k = 3$ the probabilities of no signal are larger for the ARFIMA(0,d,1) process than for ARFIMA(0,d,0) for negative β 's and smaller for positive β 's. The same holds, however, also for higher values of k too, i.e. no signals for longer time intervals.

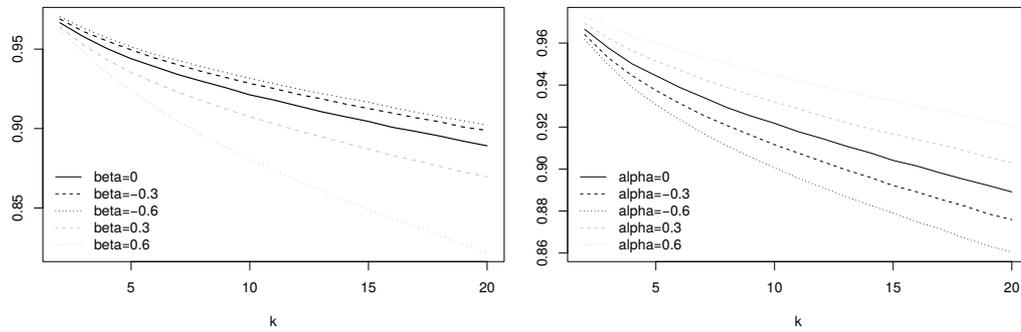


Figure 3: The probability of no signal up to the time point k for the EWMA control chart with $\lambda = 0.1$ and $c = 2.04$. The target process is an ARFIMA(0,d,1) process with $d = 0.2$ (left side) and an ARFIMA(1,d,0) process with $d = 0.2$ (right side), respectively.

The case of ARFIMA(1,d,0) is particularly important from practical perspective. The right picture of Figure 3 reveals a similar pattern as we observed for ARFIMA(0,d,1). The probabilities of no signal are larger for the ARFIMA(1,d,0) process than for ARFIMA(0,d,0) for positive α 's and smaller for negative α 's.

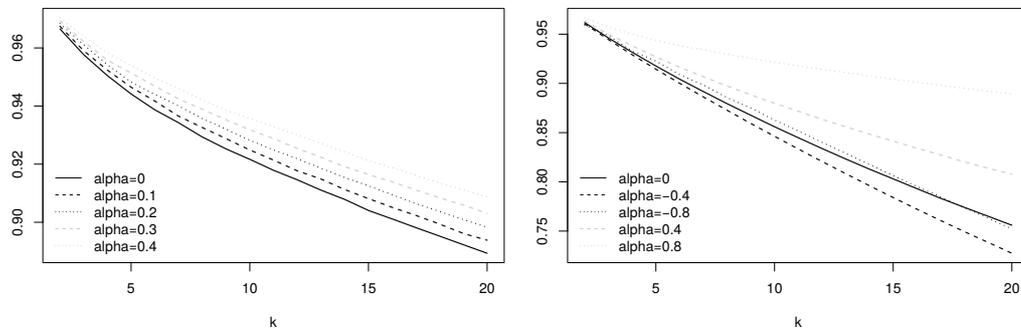


Figure 4: The probability of no signal up to the time point k for the EWMA control chart with $\lambda = 0.1$ and $c = 2.04$. The target process is an ARFIMA(1,d,0) process with $d = 0.2$ (left side) and an ARFIMA(1,d,1) process with $d = 0.2$ and $\beta = -0.8$ (right side), respectively.

However, part c) of Lemma 3.2 contains an additional constraint which makes the set where (3.1) holds very small. It stems from a statement about the magnitude of a hypergeometric function in α which is hard to obtain. Nevertheless, numerically we can argue that the monotonicity also holds for $0 \leq \alpha \leq (k-2)/(k-1)$. For the left picture in Figure 4 the selected α 's are small and satisfy $\alpha \leq (k-2)/(k-1)$ for $k \geq 3$. Despite the condition in part c) of Lemma 3.2 is not fulfilled, we observe that $P_{(1,d,0)}(N_e(c) > k) \geq P_{(0,d,0)}(N_e(c) > k)$ still holds. As a counterexample consider an ARFIMA(1,d,1) process with $d = 2$ and $\beta = -0.8$ and the right picture in Figure 4. The probability of no signal up to time point k is sometimes larger for $\alpha = 0.6$ than for $\alpha = 0.0$, sometimes smaller. This depends on the value of k . The probabilities for the discussed figures are determined by numerical integration as above.

3.2. Behavior of the false alarm probability as a function of the fractional parameter

In the previous subsection the probability of a false signal for an ARFIMA process was compared with that of an independent random process and an ARFIMA(0,d,0) process, respectively. In this subsection we want to analyze how the probability of a false signal behaves as a function of the fractional parameter d .

In a first stage we consider the simplest case, an ARFIMA(0,d,0) process.

Theorem 3.2. *Let $\{Y_t\}$ be an ARFIMA(0,d,0) process as defined in (3) with $d \in [0, 0.5)$ and let $\{\varepsilon_t\}$ be a Gaussian white noise with $\sigma > 0$. Then $P_{(0,d,0)}(N_e(c) > k)$ is a non-decreasing function in d .*

Proof: First we observe that the autocovariances of an ARFIMA(0,d,0) process can be easily recursively calculated. It holds that $\gamma_d(k) = \frac{k+d-1}{k-d}\gamma_d(k-1)$ for $k \geq 1$ (cf. Lemma A.1b of the appendix).

Next let $0 < d_1 < d_2 < 1/2$. Then it holds for $k \geq 1$ that

$$\frac{\gamma_{d_2}(k)}{\gamma_{d_2}(k-1)} = \frac{k+d_2-1}{k-d_2} \geq \frac{k+d_1-1}{k-d_1} = \frac{\gamma_{d_1}(k)}{\gamma_{d_1}(k-1)}.$$

The result follows with Theorem 1 of Schöne *et al.* ([31]).

If $d_1 = 0$ the result is a special case of Schmid and Schöne ([30]). □

Figure 5 illustrates the result of Theorem 3.2. It shows that the probabilities of no signal up to the time point k are increasing in d .

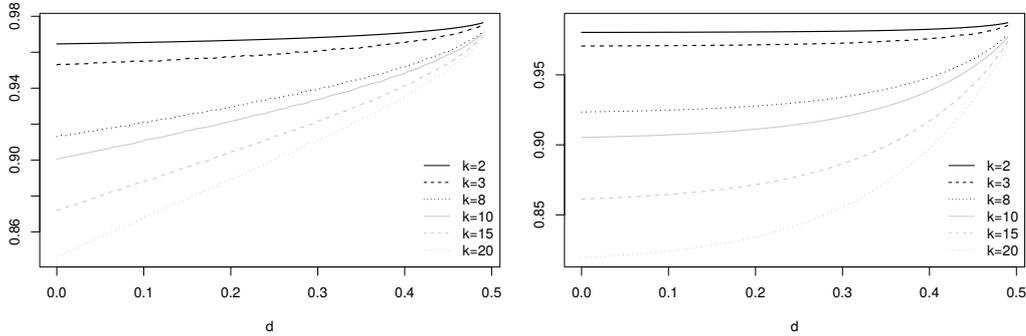


Figure 5: Probabilities of no signal up to the time point k as a function of d . The target process is an ARFIMA(0, d ,0) process. The EWMA parameters are $c = 2.04$, $\lambda = 0.1$ (left), and $c = 2.33$, $\lambda = 1.0$ (right).

Next we want to study the behavior of the false alarm probability for an ARFIMA(1, d ,1) process.

Theorem 3.3. *Let $\{Y_t\}$ be an ARFIMA(1, d ,1) process as defined in (3) with $d \in [0, 0.5)$ and let $\{\varepsilon_t\}$ be a Gaussian white noise with $\sigma > 0$. Suppose that $0 \leq \alpha < 1$ and $\beta \geq 0$. Then it follows that for all $k \in \mathbb{N}$ the quantity $P_{(1,d,1)}(N_e(c) > k)$ is a non-decreasing function in d .*

Proof: See Appendix. □

This result is quite remarkable. It says that the probability of a false signal is increasing with the fractional parameter d for positive parameters α and β . In the left plot of Figure 6 we visualize this effect for $\alpha = 0.4$ and several values of the MA parameter. On the right hand side figure we show a counterexample of nonmonotonicity if the assumptions of the theorem are not fulfilled. Here α is set equal to -0.8 .

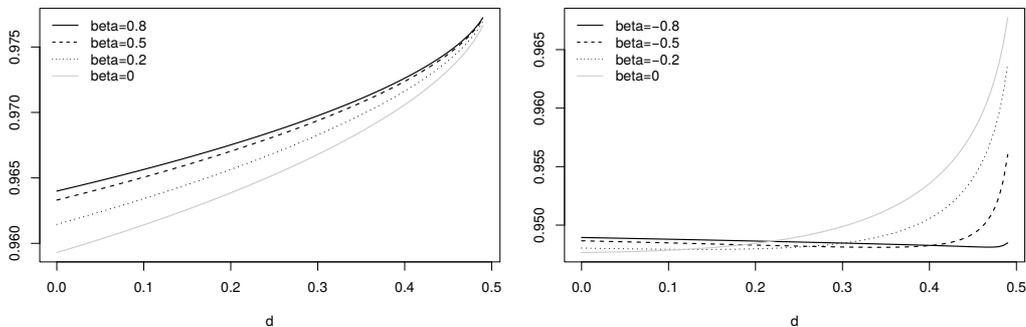


Figure 6: Probabilities of no signal up to the time point k as a function of d for ARFIMA(1, d ,1) processes. We set $\alpha = 0.4$ for the left figure, $\alpha = -0.8$ for the right one and choose $c = 2.04$, $\lambda = 0.1$.

4. SUMMARY

In this paper we consider the stochastic properties of the run length of an EWMA monitoring scheme if applied to an ARFIMA process. Particularly we are interested in the monotonic behavior of the probability of no signal up to an arbitrary time point as a function of the fractional differencing parameter d . We compare the probability of no signal for an ARFIMA(p,d,q) process with non-negative autocovariances with the probability of no signal for a sequence of i.i.d. variables and for an ARFIMA($0,d,0$) process, respectively. It is analyzed under what conditions the probability of no signal of an ARFIMA(p,d,q) process is greater or equal to that of an i.i.d. sequence and of an ARFIMA($0,d,0$) process. Furthermore, we prove that for ARFIMA($0,d,0$) and ARFIMA($1,d,1$) with positive parameters the probability of no signal is increasing in d .

APPENDIX

In the following lemma some useful properties on the behavior of the auto-correlation function $\rho_d(k)$ of an ARFIMA(0,d,0) process are summarized which will be used in the proofs.

Lemma A.1. *Suppose that $\{Y_t\}$ is an ARFIMA(0,d,0) process with $d \in (-0.5, 0.5)$.*

a) *Let $k \in \mathbb{N}$. Then $\rho_d(k) > 0$ for $d > 0$, $\rho_d(k) = 0$ for $d = 0$, and $\rho_d(k) < 0$ for $d < 0$.*

b)

$$(A.1) \quad \rho_d(k) = \frac{k-1+d}{k-d} \rho_d(k-1) = \left(1 - \frac{1-2d}{k-d}\right) \rho_d(k-1), \quad k = 1, 2, \dots$$

c) *Let $0 \leq d < 0.5$. Then it holds that $\rho_d(k)$ is a non-increasing function in k .*

d) *Let $k \in \mathbb{N} \cup \{0\}$. Then*

$$\rho'_d(k) = \rho_d(k) A_d(k)$$

with

$$A_d(k) = \sum_{i=1}^k \left(\frac{1}{i-1+d} + \frac{1}{i-d} \right).$$

e) *Let $0 \leq d < 0.5$. Then $\rho_d(k)$ is a non-decreasing function in d .*

Proof: Parts a) and b) are obvious.

c) The statement follows from the fact that $\frac{1-2d}{k-d} \geq 0$, since $d \in [0, 0.5)$ and $k \geq 1$.

d) Since

$$\log(\rho_d(k)) = \sum_{i=1}^k (\log(i-1+d) - \log(i-d))$$

the result follows by building the derivative with respect to d .

e) Follows from d). □

Proof of Lemma 3.1: Note that for $d = 0$ Lemma 3.1 reduces to Theorem 3.1 which was already proved. Thus it can be assumed that $d > 0$ in the following.

In order to prove Lemma 3.1 we apply Theorem 1 of Schöne *et al.* ([31]) and the comment after the theorem. In order to do that we need that $\gamma(v) > 0$ for $v = 1, \dots, k-1$ but this was already proved in (2.8).

Following Theorem 1 of Schöne *et al.* ([31]) it holds that $P_{(p,d,q)}(N_e(c) > k) \geq P_{(0,d,0)}(N_e(c) > k)$ if all autocovariances are positive and if for all $1 \leq v \leq k-1$ it holds that

$$\gamma_d(v-1) \sum_{i=-\infty}^{\infty} \gamma_d(v-i) \gamma_{ARMA}(i) \geq \gamma_d(v) \sum_{i=-\infty}^{\infty} \gamma_d(v-1-i) \gamma_{ARMA}(i).$$

This condition is equivalent to

$$\sum_{i=-\infty}^{\infty} \rho_{ARMA}(i) (\rho_d(v-i) \rho_d(v-1) - \rho_d(v) \rho_d(v-1-i)) \geq 0, \quad v = 1, \dots, k-1$$

and

$$\sum_{i=1}^{\infty} \rho_{ARMA}(i) ((\rho_d(v-i) + \rho_d(v+i)) \rho_d(v-1) - (\rho_d(v-1-i) + \rho_d(v-1+i)) \rho_d(v)) \geq 0,$$

for $v = 1, \dots, k-1$ and

$$\begin{aligned} & \sum_{i=1}^{v-1} \rho_{ARMA}(i) [(\rho_d(v-i) + \rho_d(v+i)) \rho_d(v-1) - (\rho_d(v-1-i) + \rho_d(v-1+i)) \rho_d(v)] \\ & + \rho_{ARMA}(v) (1 + \rho_d(2v)) \rho_d(v-1) - [\rho_d(1) + \rho_d(2v-1)] \rho_d(v) \\ & + \sum_{i=v+1}^{\infty} \rho_{ARMA}(i) [(\rho_d(i-v) + \rho_d(i+v)) \rho_d(v-1) - (\rho_d(i-v+1) + \rho_d(i+v-1)) \rho_d(v)] \geq 0 \end{aligned}$$

for $v = 1, \dots, k-1$ respectively. Using the recursion property of the autocorrelations of an ARFIMA(0,d,0) process (cf. (Lemma A.1b of the Appendix)) the last condition can be rewritten as follows

$$(A.2) \quad \frac{\rho_d(v-1)(1-2d)}{v-d} [I_v + II_v + III_v] \geq 0, \quad v = 1, \dots, k-1$$

with

$$\begin{aligned} I_v &= \sum_{i=1}^{v-1} i \rho_{ARMA}(i) \left(\frac{\rho_d(v-1+i)}{v+i-d} - \frac{\rho_d(v-1-i)}{v-i-d} \right), \\ II_v &= v \rho_{ARMA}(v) \left(\frac{\rho_d(2v-1)}{2v-d} + \frac{1}{1-d} \right), \\ III_v &= \sum_{i=v+1}^{\infty} i \rho_{ARMA}(i) \left(\frac{\rho_d(i-v)}{i-v+1-d} + \frac{\rho_d(i+v-1)}{v+i-d} \right). \quad \square \end{aligned}$$

Proof of Lemma 3.2: a) The proof is obvious.

b) Since

$$I_2 = -\rho_{ARMA}(1) \frac{6}{(3-d)(2-d)}, \quad II_2 = 24\rho_{ARMA}(2) \frac{d^2 - 2d + 2}{(4-d)(3-d)(2-d)(1-d)}$$

it holds that $I_2 + II_2 \geq 0$ if $\rho_{ARMA}(1) \leq 2\rho_{ARMA}(2)$.

c) For an ARFIMA(1,d,0) process with coefficient α it holds that

$$\begin{aligned} I_v &= \sum_{i=v}^{2v-2} \frac{\rho_d(i)}{i+1-d} (i-v+1) \alpha^{i-v+1} - \sum_{i=0}^{v-2} \frac{\rho_d(i)}{i+1-d} (v-1-i) \alpha^{v-1-i}, \\ III_v &= \sum_{i=0}^{v-2} \frac{\rho_d(i)}{i+1-d} (i+v) \alpha^{i+v} + \sum_{i=v-1}^{\infty} \frac{\rho_d(i)}{i+1-d} (i+v) \alpha^{i+v} \\ &\quad + \sum_{i=2v}^{\infty} \frac{\rho_d(i)}{i+1-d} (i-v+1) \alpha^{i-v+1}. \end{aligned}$$

Consequently

$$\begin{aligned} I_v + II_v + III_v &= \alpha \left[\sum_{i=0}^{v-2} \frac{\rho_d(i)}{i+1-d} ((i+v) \alpha^{i+v-1} - (v-1-i) \alpha^{v-2-i}) \right. \\ &\quad + \sum_{i=v}^{2v-1} \frac{\rho_d(i)}{i+1-d} (i-v+1) \alpha^{i-v} + \sum_{i=2v}^{\infty} \frac{\rho_d(i)}{i+1-d} (i-v+1) \alpha^{i-v} \\ &\quad \left. + \sum_{i=v-1}^{\infty} \frac{\rho_d(i)}{i+1-d} (i+v) \alpha^{i+v-1} \right]. \end{aligned}$$

This quantity is non-negative if $(i+v) \alpha^{i+v-1} - (v-1-i) \alpha^{v-2-i} \geq 0$ for $i = 0, \dots, v-2$. This is fulfilled if $\alpha \geq \left(1 - \frac{2i+1}{i+v}\right)^{1/(2i+1)}$ for all $i = 0, \dots, v-2$ since $\left(1 - \frac{2i+1}{i+v}\right)^{1/(2i+1)} \leq 1 - \frac{1}{v}$. Using mathematical induction we shall prove that $\left(1 - \frac{1}{v}\right)^{2i+1} \geq 1 - \frac{2i+1}{i+v}$ for all $i = 0, \dots, v-2$. For $i = 0$ it is obvious. Next we consider the induction step. Note that

$$\begin{aligned} \left(1 - \frac{1}{v}\right)^{2i+3} &\geq \left(1 - \frac{2i+1}{i+v}\right) \left(1 - \frac{1}{v}\right)^2 \\ &= 1 - \left(\frac{2i+1}{i+v} + \frac{2v-1}{v^2} - \frac{(2v-1)(2i+1)}{v^2(i+v)}\right) \\ &\geq 1 - \frac{2i+3}{i+1+v} \end{aligned}$$

since after some calculations it can be seen that the last inequality is equivalent to $i^2 + 2i + 1 \geq 0$.

Since $v \leq k-1$ we finally get that $\alpha \geq (k-2)/(k-1)$ for $k \geq 2$. \square

Proof of Theorem 3.3: Note that

$$\rho_{ARMA}(k) = \alpha^{k-1}\rho_1, \quad k \geq 1, \quad \rho_1 = \frac{(1 + \alpha\beta)(\alpha + \beta)}{1 + 2\alpha\beta + \beta^2}.$$

Since $0 \leq \alpha < 1$ and $\beta \geq 0$ it follows that $\rho_1 \geq \alpha$.

In Theorem 3.1 it was proved that for an ARFIMA process the in-control probability of a false signal up to a given time point is greater or equal than for an independent random sequence. Thus we may assume in the following that $d > 0$. As shown in the proof of Lemma 3.1 this implies that $\gamma(v) > 0$.

a) Let $\{\gamma(h)\}$ denote the autocovariance function of an ARFIMA(1,d,1) process. In order to prove the result we make use of the comment after Theorem 1 of Schöne *et al.* ([31]) which says that it is sufficient to show that $\gamma(k)/\gamma(k-1)$ is a non-decreasing function in d . Suppose that $\alpha > 0$. Let $\rho^* = \rho_1/\alpha$. Then

$$\begin{aligned} \frac{\gamma(k)}{\gamma_{ARMA}(0)\gamma_d(0)} &= \sum_{i=-\infty}^{\infty} \rho_d(i)\rho_{ARMA}(k-i) = a_d(k) + b_d(k) \\ &= (1 - \rho^*)\rho_d(k) + \alpha a_d^*(k-1) + \frac{1}{\alpha}b_d(k-1). \end{aligned}$$

with

$$\begin{aligned} a_d(k) &= \sum_{i=-\infty}^k \rho_d(i)\rho_{ARMA}(k-i) = (1 - \rho^*)\rho_d(k) + a_d^*(k) = \rho_d(k) + \alpha a_d^*(k-1), \\ b_d(k) &= \sum_{i=k+1}^{\infty} \rho_d(i)\rho_{ARMA}(k-i) = \rho^* \sum_{i=k+1}^{\infty} \rho_d(i)\alpha^{i-k} = \frac{1}{\alpha}b_d(k-1) - \rho^*\rho_d(k), \\ a_d^*(k) &= \rho^* \sum_{i=-\infty}^k \rho_d(i)\alpha^{k-i}. \end{aligned}$$

b) The numerator of the derivative of $\gamma(k)/\gamma(k-1)$ with respect to d is equal to

$$\begin{aligned} &\left(\frac{1}{\alpha} - \alpha\right) [a_d^*(k-1)b_d'(k-1) - a_d^*(k-1)b_d(k-1)] \\ &+ (1 - \rho^*) a_d^*(k-1) [\alpha\rho_d(k-1) - \rho_d(k)] + (1 - \rho^*) a_d^*(k-1) [-\alpha\rho_d'(k-1) + \rho_d'(k)] \\ &+ (1 - \rho^*) b_d'(k-1) \left[\frac{1}{\alpha}\rho_d(k-1) - \rho_d(k)\right] + (1 - \rho^*)b_d(k-1)\left[-\frac{1}{\alpha}\rho_d'(k-1) + \rho_d'(k)\right] \\ &+ (1 - \rho^*)^2[\rho_d(k-1)\rho_d'(k) - \rho_d'(k-1)\rho_d(k)]. \end{aligned}$$

It is sufficient to prove that this quantity is not negative.

Let $\gamma = (1 - \rho^*)/(1/\alpha - \alpha)$. Note that $\gamma \leq 0$. An equivalent representation is

$$\begin{aligned}
& [a_d^*(k-1)b'_d(k-1) - a_d^{*'}(k-1)b_d(k-1)] \\
& + \gamma \alpha [\rho_d(k-1)a_d^{*'}(k-1) - \rho_d'(k-1)a_d^*(k-1)] \\
& + \gamma [\rho_d'(k)a_d^*(k-1) - \rho_d(k)a_d^{*'}(k-1)] + \frac{\gamma}{\alpha} [\rho_d(k-1)b'_d(k-1) - \rho_d'(k-1)b_d(k-1)] \\
& + \gamma [\rho_d'(k)b_d(k-1) - \rho_d(k)b'_d(k-1)] + \gamma(1 - \rho^*)[\rho_d(k-1)\rho_d'(k) - \rho_d'(k-1)\rho_d(k)] \\
& = I + II + III + IV + V + VI.
\end{aligned}$$

c) Next we apply Lemma A.1d. Defining $A_d(-h) = A_d(h)$ for $h \geq 1$ we get that $\rho_d'(h) = \rho_d(h)A_d(h)$ for all $h \in \mathbb{Z}$.

It holds that

$$I/\rho^{*2} = \sum_{i=-\infty}^{k-1} \sum_{j=k}^{\infty} \rho_d(i)\rho_d(j)\alpha^{j-i}(A_d(j) - A_d(i)) = I_1 + I_2$$

with

$$\begin{aligned}
I_1 &= \sum_{j=k}^{\infty} \sum_{i=-\infty}^{-j-1} \rho_d(i)\rho_d(j)\alpha^{j-i}(A_d(j) - A_d(-i)), \\
I_2 &= \sum_{j=k}^{\infty} \sum_{i=-j}^{k-1} \rho_d(i)\rho_d(j)\alpha^{j-i}(A_d(j) - A_d(|i|)).
\end{aligned}$$

Now

$$\begin{aligned}
I_1 &= - \sum_{j=k}^{\infty} \sum_{i=j+1}^{\infty} \rho_d(i)\rho_d(j)\alpha^{j+i}(A_d(i) - A_d(j)) \\
&= - \sum_{j=k+1}^{\infty} \sum_{i=k}^{j-1} \rho_d(i)\rho_d(j)\alpha^{j+i}(A_d(j) - A_d(i)), \\
I_2 &= \sum_{j=k}^{\infty} \sum_{i=-k+1}^{k-1} \rho_d(i)\rho_d(j)\alpha^{j-i}(A_d(j) - A_d(|i|)) \\
&\quad + \sum_{j=k}^{\infty} \sum_{i=-j}^{-k} \rho_d(i)\rho_d(j)\alpha^{j-i}(A_d(j) - A_d(|i|)) \\
&= \sum_{j=k}^{\infty} \sum_{i=-k+1}^{k-1} \rho_d(i)\rho_d(j)\alpha^{j-i}(A_d(j) - A_d(|i|)) - I_1.
\end{aligned}$$

Thus

$$(A.3) \quad I/\rho^{*2} = \sum_{j=k}^{\infty} \sum_{i=-k+1}^{k-1} \rho_d(i)\rho_d(j)\alpha^{j-i}(A_d(j) - A_d(|i|)).$$

d) Since

$$\begin{aligned}
I &= \rho^{*2} \left(\sum_{j=k}^{\infty} \sum_{i=-k+1}^{k-1} \rho_d(i) \rho_d(j) \alpha^{j-i} (A_d(j) - A_d(|i|)) \right) \\
&= \rho^{*2} \left(\sum_{j=k+1}^{\infty} \sum_{i=-k+2}^{k-2} \rho_d(i) \rho_d(j) \alpha^{j-i} (A_d(j) - A_d(|i|)) \right. \\
&\quad + \rho_d(k) \sum_{i=-k+1}^{k-1} \rho_d(i) \alpha^{k-i} (A_d(k) - A_d(|i|)) \\
&\quad + \rho_d(k-1) \sum_{j=k+1}^{\infty} \rho_d(j) \alpha^{j-k+1} (A_d(j) - A_d(k-1)) \\
&\quad \left. + \rho_d(k-1) \sum_{j=k+1}^{\infty} \rho_d(j) \alpha^{j+k-1} (A_d(j) - A_d(k-1)) \right) \\
&= I_3 + I_4 + I_5 + I_6, \quad I_i \geq 0, i = 3, \dots, 6,
\end{aligned}$$

$$\begin{aligned}
II &= -\gamma \alpha \rho^* \rho_d(k-1) \sum_{i=-k+2}^{k-2} \rho_d(i) \alpha^{k-1-i} (A_d(k-1) - A_d(|i|)) \\
&\quad + \gamma \alpha \rho^* \rho_d(k-1) \sum_{j=k}^{\infty} \rho_d(j) \alpha^{k-1+j} (A_d(j) - A_d(k-1)) = II_1 + II_2, \\
&\text{with } II_1 \geq 0, II_2 \leq 0,
\end{aligned}$$

$$\begin{aligned}
III &= \gamma \rho^* \rho_d(k) \sum_{i=-k+1}^{k-1} \rho_d(i) \alpha^{k-1-i} (A_d(k) - A_d(|i|)) \\
&\quad + \gamma \rho^* \rho_d(k) \sum_{j=k+1}^{\infty} \rho_d(j) \alpha^{k-1+j} (A_d(k) - A_d(j)) \\
&= III_1 + III_2, \quad III_1 \leq 0, III_2 \geq 0,
\end{aligned}$$

$$IV = \frac{\gamma \rho^*}{\alpha} \rho_d(k-1) \sum_{j=k}^{\infty} \rho_d(j) \alpha^{j-k+1} (A_d(j) - A_d(k-1)) \leq 0$$

we get that

$$\begin{aligned}
VII &= I_4 + III_1 = (\rho^{*2} + \gamma \rho^* / \alpha) \rho_d(k) \sum_{i=-k+1}^{k-1} \rho_d(i) \alpha^{k-i} (A_d(k) - A_d(|i|)) \geq 0, \\
I_6 + II_2 + VI &= (\rho^{*2} + \gamma \alpha \rho^*) \rho_d(k-1) \sum_{j=k+1}^{\infty} \rho_d(j) \alpha^{j+k-1} (A_d(j) - A_d(k-1)) \\
&\quad + (\gamma \rho^* \alpha^{2k} + \gamma(1 - \rho^*) \rho_d(k-1) \rho_d(k)) (A_d(k) - A_d(k-1)) \\
&= VIII_1 + VIII_2, \quad \text{with } VIII_1 \geq 0,
\end{aligned}$$

$$\begin{aligned}
I_5 + IV &= (\rho^{*2} + \gamma\rho^*/\alpha)\rho_d(k-1) \sum_{j=k+1}^{\infty} \rho_d(j)\alpha^{j-k+1}(A_d(j) - A_d(k-1)) \\
&\quad + \gamma\rho^*\rho_d(k-1)\rho_d(k)(A_d(k) - A_d(k-1)) = IX_1 + IX_2, \\
&\qquad\qquad\qquad \text{with } IX_1 \geq 0, IX_2 \leq 0,
\end{aligned}$$

$$VIII_2 + IX_2 = \gamma(1 + \rho^*\alpha^{2k})\rho_d(k-1)\rho_d(k)(A_d(k) - A_d(k-1)) = X.$$

Now

$$\begin{aligned}
VII &\geq (\rho^{*2} + \frac{\gamma\rho^*}{\alpha})(\alpha + \alpha^{2k-1})\rho_d(k-1)\rho_d(k)(A_d(k) - A_d(k-1)), \\
&= \frac{\rho_1^2 + \gamma\rho_1}{\alpha}(1 + \alpha^{2k-2})\rho_d(k-1)\rho_d(k)(A_d(k) - A_d(k-1))
\end{aligned}$$

and

$$\begin{aligned}
VII + X &\geq (\gamma(1 + \rho^*\alpha^{2k}) + \frac{\rho_1^2 + \gamma\rho_1}{\alpha}(1 + \alpha^{2k-2})) \\
&\quad \cdot \rho_d(k-1)\rho_d(k)(A_d(k) - A_d(k-1)).
\end{aligned}$$

Since

$$\begin{aligned}
&\gamma(1 + \rho^*\alpha^{2k}) + \frac{\rho_1^2 + \gamma\rho_1}{\alpha}(1 + \alpha^{2k-2}) \\
&= \gamma(1 + \rho_1\alpha^{2k-1}) + \frac{\rho_1^2 + \gamma\rho_1}{\alpha}(1 + \alpha^{2k-2}) \\
&\geq \frac{\rho_1^2 + \gamma\rho_1}{\alpha}(1 + \alpha^{2k-1}) + \gamma(1 + \alpha^{2k-1}) \\
&= \frac{1 + \alpha^{2k-1}}{\alpha}(\rho_1^2 + \gamma\rho_1 + \gamma\alpha) \\
&= \frac{1 + \alpha^{2k-1}}{\alpha(1 - \alpha^2)}(\rho_1^2(1 - \alpha^2) + \rho_1(\alpha - \rho_1) + \alpha(\alpha - \rho_1)) \\
&= \frac{\alpha(1 - \rho_1^2)(1 + \alpha^{2k-1})}{1 - \alpha^2} \geq 0
\end{aligned}$$

it holds that $VII + X \geq 0$.

Moreover, we get that

$$\begin{aligned}
V &= \gamma\rho^* \sum_{i=k}^{\infty} \rho_d(k)\rho_d(i)\alpha^{i-k+1}(A_d(k) - A_d(i)) \geq 0, \\
VI &= \gamma(1 - \rho^*)\rho_d(k-1)\rho_d(k)(A_d(k) - A_d(k-1)) \geq 0.
\end{aligned}$$

Thus the result is proved. \square

REFERENCES

- [1] ALWAN, L. and ROBERTS, H. (1988). Time series modelling for statistical process control, *Journal of Business and Economic Statistics*, **6**, 87–95.
- [2] ANDERSEN, T.G.; BOLLERSLEV, T. and LANGE, S. (1999). Forecasting financial market volatility: Sample frequency vis-a-vis forecast horizon, *Journal of Empirical Finance*, **6**, 457–477.
- [3] BOX, G.; JENKINS, G. and REINSEL, G. (1994). *Time Series Analysis – Forecasting and Control*, Prentice-Hall, Englewood Cliffs.
- [4] BROCKWELL, P. and DAVIS, R. (1991). *Time series: theory and methods*, Springer.
- [5] CAPIZZI, G. and MASAROTTO G. (2008). Practical design of generalized likelihood ratio control charts for autocorrelated data, *Technometrics*, **50**, 357–370.
- [6] CROWDER, S. and HAMILTON, M. (1992). An EWMA for monitoring a process standard deviation, *Journal of Quality Technology*, **24**, 12–21.
- [7] FRISÉN, M. (2008). *Financial Surveillance*, Wiley.
- [8] GENZ, A. (1992). Numerical computation of multivariate normal probabilities, *Journal of Computational and Graphical Statistics*, **1**, 141–150.
- [9] GONÇALVES, E.; LEITE, J. and MENDES-LOPES, N. (2013). The ARL of modified Shewhart control charts for conditionally heteroskedastic models, *Statistical Papers*, **54**, 1–19.
- [10] GRANGER, C. and JOYEUX, R. (1980). An introduction to long-range time series models and fractional differencing, *Journal of Time Series Analysis*, **1**, 15–30.
- [11] GRANGER, C. (1981). Some properties of time series data and their use in econometric model specification, *Journal of Econometrics*, **16**, 121–130.
- [12] HOSKING, J. (1981). Fractional differencing, *Biometrika*, **68**, 165–176.
- [13] KNOTH, S. and FRISÉN, M. (2012). Minimax Optimality of CUSUM for an Autoregressive Model, *Statistica Neerlandica*, **66**, 357–379.
- [14] KNOTH, S. and SCHMID, W. (2004). Control charts for time series: A review, *Frontiers in Statistical Quality Control* (H.-J. Lenz and P.-T. Wilrich, Eds.), Physica, Vol. **7**, 210–236.
- [15] LAWSON, A.B. and KLEINMAN, K. (2005). *Spatial & Syndromic Surveillance*, Wiley.
- [16] LU, C.-W. and REYNOLDS, JR., M. (1999). EWMA control charts for monitoring the mean of autocorrelated processes, *Journal of Quality Technology*, **31**, 166–188.
- [17] LUCAS, J. and SACCUCCI, M. (1990). Exponentially weighted moving average control schemes: properties and enhancements, *Technometrics*, **32**, 1–12.
- [18] MONTGOMERY, D. (2009). *Introduction to Statistical Quality Control (Sixth Edition)*, John Wiley & Sons, New York.
- [19] MONTGOMERY, D. and MASTRANGELO, C. (1991). Some statistical process control methods for autocorrelated data, *Journal of Quality Technology*, **23**, 179–204.

- [20] MORAIS, M.; OKHRIN, Y.; PACHECO, A. and SCHMID, W. (2006). On the stochastic behaviour of the run length of EWMA control schemes for the mean of correlated output in the presence of shifts in sigma, *Statistics & Decisions*, **24**, 397–413.
- [21] NIKIFOROV, I. (1975). Sequential analysis applied to autoregressive processes, *Automation and Remote Control*, **36**, 1365–1368.
- [22] PALMA, W. (2007). *Long-Memory Time Series: Theory and Methods*, John Wiley & Sons, New York.
- [23] PAN, J.-N. and CHEN, S.-T. (2008). Monitoring long-memory air quality data using ARFIMA model, *Environmetrics*, **19**, 209–219.
- [24] PAWLAK, M. and SCHMID, W. (2001). On distributional properties of GARCH processes, *Journal of Time Series Analysis*, **22**, 339–352.
- [25] RABYK, L. and SCHMID, W. (2016). EWMA control charts for detecting changes in the mean of a long-memory process, *Metrika*, **79**, 267–301.
- [26] ROBERTS, S. (1959). Control charts tests based on geometric moving averages, *Technometrics*, **1**, 239–250.
- [27] SCHMID, W. (1997a). On EWMA charts for time series, *Frontiers in Statistical Quality Control* (H.-J. Lenz and P.-T. Wilrich, Eds.), Physica, Vol. **5**, 115–137.
- [28] SCHMID, W. (1997b). CUSUM control schemes for Gaussian processes, *Statistical Papers*, **38**, 191–217.
- [29] SCHMID, W. and OKHRIN, Y. (2003). Tail behaviour of a general family of control charts, *Statistics & Decisions*, **21**, 77–90.
- [30] SCHMID, W. and SCHÖNE, A. (1997). Some properties of the EWMA control chart in the presence of data correlation, *Annals of Statistics*, **25**, 1277–1283.
- [31] SCHÖNE, A.; SCHMID, W. and KNOTH, S. (1999). On the run length of the EWMA scheme — a monotonicity result for normal variables, *Journal of Statistical Planning and Inference*, **79**, 289–297.
- [32] WARDELL, D.; MOSKOWITZ, H. and PLANTE, R. (1994a). Run length distributions of residual control charts for autocorrelated processes, *Journal of Quality Technology*, **26**, 308–317.
- [33] WARDELL, D.; MOSKOWITZ, H. and PLANTE, R. (1994b). Run length distributions of special-cause control charts for correlated processes (with discussion), *Technometrics*, **36**, 3–27.
- [34] YASHCHIN, E. (1993). Performance of CUSUM control schemes for serially correlated observations, *Technometrics*, **35**, 37–52.

USING SHRINKAGE ESTIMATORS TO REDUCE BIAS AND MSE IN ESTIMATION OF HEAVY TAILS

Authors: JAN BEIRLANT

- Department of Mathematics, KU Leuven, Belgium, and
Department of Mathematical Statistics and Actuarial Science,
University of the Free State South Africa
`jan.beirlant@kuleuven.be`

GAONYALELWE MARIBE

- Department of Mathematical Statistics and Actuarial Science,
University of the Free State, South Africa
`maribeg@ufs.ac.za`

ANDRÉHETTE VERSTER

- Department of Mathematical Statistics and Actuarial Science,
University of the Free State, South Africa
`VersterA@ufs.ac.za`

Received: June 2016

Revised: December 2016

Accepted: February 2017

Abstract:

- Bias reduction in tail estimation has received considerable interest in extreme value analysis. Estimation methods that minimize the bias while keeping the mean squared error (MSE) under control, are especially useful when applying classical methods such as the Hill (1975) estimator. In the case of heavy tailed distributions, Caeiro *et al.* (2005) proposed minimum variance reduced bias estimators of the extreme value index, where the bias is reduced without increasing the variance with respect to the Hill estimator. This method is based on adequate external estimation of a pair of parameters of second order slow variation under a third order condition. Here we revisit this problem exploiting the mathematical fact that the bias tends to 0 with increasing threshold. This leads to shrinkage estimation for the extreme value index, which allows for a penalized likelihood and a Bayesian implementation. This new approach is applied starting from the approximation to excesses over a high threshold using the extended Pareto distribution, as developed in Beirlant *et al.* (2009). We present asymptotic results for the resulting shrinkage penalized likelihood estimator of the extreme value index. Finite sample simulation results are proposed both for the penalized likelihood and Bayesian implementation. We then compare with the minimum variance reduced bias estimators.

Key-Words:

- *extreme value index; tail estimation; extended Pareto distribution; shrinkage estimators.*

AMS Subject Classification:

- 62G32, 62F10, 62F15, 62J07.

1. INTRODUCTION

In this paper we consider the estimation of the extreme value index ξ and tail probabilities $P(X > x)$ for x large, on the basis of independent and identically distributed observations X_1, X_2, \dots, X_n which follow a Pareto-type distribution with right tail function (RTF) given by

$$(1.1) \quad \bar{F}(x) = 1 - F(x) = P(X > x) = x^{-1/\xi} \ell(x)$$

where ℓ is a slowly varying function at infinity, i.e.

$$\frac{\ell(ty)}{\ell(t)} \rightarrow 1, \quad \text{as } t \rightarrow \infty, \quad \text{for every } y > 1.$$

The most famous estimator of ξ was first derived by Hill (1975) as a maximum likelihood (ML) estimator approximating the RTF of the excesses $\frac{X}{t} | X > t$ over a large threshold t by a simple Pareto distribution with RTF $y^{-1/\xi}$:

$$(1.2) \quad \bar{F}(ty)/\bar{F}(t) \approx y^{-1/\xi}, \quad t \text{ large.}$$

When setting $t = X_{n-k,n}$ where $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$ the ML estimator is given by

$$(1.3) \quad H_{k,n} = \frac{1}{k} \sum_{j=1}^k \log \frac{X_{n-j+1,n}}{X_{n-k,n}}.$$

A simple estimator of a tail probability $P(X > x)$ with x large, introduced in Weissman (1978), is then obtained from (1.2) setting $ty = x$ and estimating $P(X > t)$ by the empirical proportion k/n :

$$(1.4) \quad \hat{p}_{x,k} = \frac{k}{n} \left(\frac{x}{X_{n-k,n}} \right)^{-1/H_{k,n}}.$$

In practice, a way to verify the validity of model (1.1) is to check whether the Hill estimates are stable as a function of k . However in most cases the stability is not visible, which can be explained by slow convergence in (1.2). For this reason bias reduced estimators have been proposed which lead to plots that are much more horizontal in k which facilitates the analysis of a practical case to a great extent. Here we can refer to Peng (1998), Beirlant *et al.* (1999, 2008), Feuerverger and Hall (1999), Caeiro *et al.* (2005, 2009) and Gomes *et al.* (2000, 2007) for bias-reduced estimators based on functions of the top k order statistics. Several of these methods focus on the distribution of log-spacings of high order statistics.

Beirlant *et al.* (2009) proposed a more flexible model capable of capturing the deviation between the true excess RTF $\bar{F}(ty)/\bar{F}(t)$ and the asymptotic Pareto model. For a heavy tailed distribution (1.1), this deviation can be parametrized

using a power series expansion (Hall, 1982), or more generally via second-order slow variation (Bingham *et al.*, 1987). More specifically in Beirlant *et al.* (2009) the subclass $\mathcal{F}(\xi, \tau)$ of the Pareto-type tails (1.1) was considered satisfying

$$(1.5) \quad \bar{F}(x) = Cx^{-1/\xi} (1 + \xi^{-1}\delta(x)),$$

with $\delta(x)$ eventually nonzero and of constant sign such that $|\delta(x)| = x^\tau \ell_\delta(x)$ with $\tau < 0$ and ℓ_δ slowly varying. It was shown that under $\mathcal{F}(\xi, \tau)$ as $t \rightarrow \infty$

$$\sup_{y \geq 1} \left| \frac{\bar{F}(ty)}{\bar{F}(t)} - \bar{G}_{\xi, \delta, \tau}(y) \right| = o(|\delta(t)|)$$

with $\bar{G}_{\xi, \delta, \tau}$ the RTF of the extended Pareto distribution (EPD)

$$(1.6) \quad \bar{G}_{\xi, \delta, \tau}(y) = \{y(1 + \delta - \delta y^\tau)\}^{-1/\xi}, \quad y > 1,$$

with $\tau < 0 < \xi$ and $\delta > \max(-1, 1/\tau)$. This shows that the EPD improves the approximation (1.2) with an order of magnitude. Then ML estimation of the parameters (ξ, δ) based on a set of excesses $(Y_{j,k} := X_{n-j+1,n}/X_{n-k,n}, j = 1, \dots, k)$ was used to obtain a bias reduced estimator $\hat{\xi}_{k,n}^{ML}$ of ξ . Bias reduction of the Weissman estimator of tail probabilities can analogously be obtained using

$$(1.7) \quad \hat{p}_{x,k}^{EP} = \frac{k}{n} \bar{G}_{\hat{\xi}_k, \hat{\delta}_k, \hat{\tau}} \left(\frac{x}{X_{n-k,n}} \right),$$

where $(\hat{\xi}_k, \hat{\delta}_k)$ denote the ML estimators based on the EPD model, and where $\hat{\tau}$ is a consistent estimator of τ , to be specified below, which was shown not to affect the asymptotic distribution of (ξ, δ) .

If F satisfies $\mathcal{F}(\xi, \tau)$, it is shown in Beirlant *et al.* (2009) that $U(x) := Q(1 - x^{-1})$ ($x > 1$), with $Q(p) = \inf\{x : F(x) \geq p\}$ ($p \in (0, 1)$), satisfies

$$(1.8) \quad U(x) = C^\xi x^\xi (1 + a(x))$$

with $a(x) = \delta(Q(1 - x^{-1}))\{1 + o(1)\} = \delta(C^\xi x^\xi)\{1 + o(1)\}$ as $x \rightarrow \infty$. In particular a is eventually nonzero and of constant sign and $|a(x)| = x^\rho \ell_a(x)$ with ℓ_a slowly varying and $\rho = \xi\tau$. Here we assume $|\ell_a(x)| = C_a(1 + o(1))$ as $x \rightarrow \infty$ for some constant $C_a > 0$.

The following asymptotic results have been derived for $H_{k,n}$ and $\hat{\xi}_{k,n}^{ML}$ assuming that F satisfies $\mathcal{F}(\xi, \tau)$, and $\sqrt{k}a(n/k) \rightarrow \lambda \in \mathbb{R}$ and $\hat{\rho}_{k,n} = \rho + o_p(1)$ as $k, n \rightarrow \infty$ and $k/n \rightarrow 0$:

$$(1.9) \quad \sqrt{k}(H_{k,n} - \xi) \rightarrow_d \mathcal{N}\left(\lambda \frac{\rho}{1 - \rho}, \xi^2\right),$$

$$(1.10) \quad \sqrt{k}(\hat{\xi}_{k,n}^{ML} - \xi) \rightarrow_d \mathcal{N}\left(0, \xi^2 \left(\frac{1 - \rho}{\rho}\right)^2\right).$$

An estimator $\hat{\rho}_{k,n}$ of ρ can be taken from Fraga Alves *et al.* (2003) using $k = k_1 = \lfloor n^{1-\epsilon} \rfloor$ for some $\epsilon > 0$. The required consistency for $\hat{\rho}_{k,n}$ was obtained under (1.8).

Asymptotic results of the type (1.9) and (1.10) are typical for bias reduced estimators when both ξ and $a(n/k)$ or δ are jointly estimated at every k value: for larger values of k corresponding to $\sqrt{k}a(n/k) \rightarrow \lambda \neq 0$, bias reduced estimators still have asymptotic bias 0 in contrast to the Hill estimator, but their variance is increased by a factor $((1-\rho)/\rho)^2$ compared to $H_{k,n}$. In a pioneering paper, Caeiro *et al.* (2005) proposed to estimate $(n/k)^{-\rho}a(n/k)$ at a high level $k = k_1 = \lfloor n^{1-\epsilon} \rfloor$, leading to a corrected Hill estimator (denoted below by $CH_{k,n}$) with asymptotic variance ξ^2 and excellent bias and MSE characteristics. To obtain the normal asymptotic behaviour of such minimum variance reduced bias estimators one needs a third-order slow variation condition which is more restrictive than (1.8) or condition $\mathcal{F}(\xi, \tau)$.

Up to now, to the best of our knowledge, the fact that $\delta(t) \rightarrow 0$ as $t \rightarrow \infty$, or $a(n/k) \rightarrow 0$ as $n/k \rightarrow \infty$ has not been exploited in the literature. However, this calls for shrinkage estimators. Such shrinkage approach can be implemented by putting a penalty on δ in an ML procedure, leading to penalized ML. Alternatively a penalty on δ can be naturally introduced in a Bayesian approach putting an appropriate prior on this parameter. Here we investigate the use of shrinkage estimation when modelling the distribution of the vector of excesses $\mathbf{Y}_k := (Y_{j,k}, j = 1, \dots, k)$ with an EPD. In section 2 we show that a quadratic penalty, or equivalently a normal prior, on δ with zero mean and variance $\sigma_{k,n}^2$, depending in an appropriate way on k and n , leads to interesting asymptotic MSE results for ξ . In section 3 we consider the finite sample behaviour of the penalized likelihood and Bayes approach, and make a comparison with the minimum variance reduced bias estimator, and consider a practical case.

2. SHRINKAGE ESTIMATORS OF THE EPD PARAMETERS

2.1. Penalized likelihood and Bayesian interpretation

ML estimation of the EPD parameters (ξ, δ) , given a value of τ , follows by maximizing the log-likelihood

$$\begin{aligned}
 \frac{1}{k} l_{EP}(\xi, \delta | \mathbf{y}) &= -\log \xi - \left(\frac{1}{\xi} + 1 \right) \frac{1}{k} \sum_{j=1}^k [\log y_{j,k} + \log(1 + \delta \{1 - y_{j,k}^\tau\})] \\
 (2.1) \qquad &+ \frac{1}{k} \sum_{j=1}^k \log(1 + \delta \{1 - (1 + \tau)y_{j,k}^\tau\}).
 \end{aligned}$$

Shrinkage estimators are then obtained by putting a penalty on δ . Below it will be shown that a quadratic penalty is appropriate in view of the asymptotic results for the penalized maximum likelihood (PML) estimators $(\hat{\xi}_{k,n}^P, \hat{\delta}_{k,n}^P)$. These estimators are then obtained by optimizing the log-likelihood

$$(2.2) \quad \frac{1}{k} l_{pen}(\xi, \delta | \mathbf{y}) = \frac{1}{k} l_{EP}(\xi, \delta | \mathbf{y}) - \omega \frac{\delta^2}{2k\sigma_{k,n}^2},$$

where $\omega > 0$ serves as a tuning constant regulating the amount of penalty, and $\sigma_{k,n}^2$ indicating the penalty rate as a function of k . From the asymptotic analysis below, it follows that $\sigma_{k,n}^2 = (k/n)^{-2\rho}$ is appropriate.

Alternatively, from a Bayesian perspective, a shrinkage estimator is obtained by considering the posterior mode estimators $(\hat{\xi}_k^B, \hat{\delta}_k^B)$ of the log-posterior

$$(2.3) \quad \frac{1}{k} \log p(\xi, \delta | \mathbf{y}) = \frac{1}{k} l_{EP}(\xi, \delta | \mathbf{y}) + \frac{1}{k} \log \pi(\xi, \delta),$$

where $\pi(\xi, \delta)$ denotes the prior density on (ξ, δ) . Following an objective Bayesian point of view, we assign a maximal data information (MDI) prior to ξ , which for a general parameter θ is defined as $\pi(\theta) \propto \exp(E(\log f(\mathbf{Y}|\theta)))$. The concept of MDI priors was introduced in Zellner (1971) in order to maximize the information contributed by the data density, relative to that of the prior density. Beirlant *et al.* (2004) derived that the MDI for a Pareto distribution is given by

$$(2.4) \quad \pi(\xi) \propto \frac{e^{-\xi}}{\xi}.$$

Next, in correspondance with the choice for the penalized log-likelihood (2.2), we here choose a normal prior on δ with mean 0 and variance $\sigma_{k,n}^2$. We also truncate it from the left in order to comply with the restriction $\delta > \max(-1, 1/\tau)$:

$$(2.5) \quad \pi(\delta) = \frac{1}{\sqrt{2\pi}\sigma_{k,n}} e^{-\frac{1}{2}\frac{\delta^2}{\sigma_{k,n}^2}} / (1 - \Phi(\max(-1, \tau^{-1})/\sigma_{k,n})).$$

2.2. Asymptotic results for the penalized ML estimator $\hat{\xi}_k^P$

In the Appendix we derive that the first order approximations $(\hat{\xi}_k^P, \hat{\delta}_k^P)$ of the penalized ML estimators are given by

$$\begin{aligned} \hat{\xi}_k^P &= H_{k,n} + \hat{\delta}_k^P (1 - E_{k,n}(\tau)), \\ \hat{\delta}_k^P &= \frac{1 - H_{k,n}\tau}{D_{k,n}^P} \left(E_{k,n}(\tau) - \frac{1}{H_{k,n}\tau} \right) \end{aligned}$$

where

$$E_{k,n}(s) = \frac{1}{k} \sum_{j=1}^k Y_{j,k}^s, \quad s < 0$$

and

$$D_{k,n}^P = \frac{\omega \hat{\xi}_k^P}{k\sigma_{k,n}^2} - \left(1 - 2(1 - \hat{\xi}_k^P \tau)E_{k,n}(\tau) + (1 - 2\hat{\xi}_k^P \tau - \hat{\xi}_k^P \tau^2)E_{k,n}(2\tau) \right. \\ \left. - \tau(1 - E_{k,n}(\tau))E_{k,n}(\tau) \right).$$

These expressions are identical to the asymptotic EPD-ML estimators derived in Beirlant *et al.* (2009) except for the extra term $\frac{\hat{\xi}_k^P}{k\sigma_{k,n}^2}$ in the expression of $D_{k,n}^P$. As an external estimator of τ we use $\hat{\tau} = \hat{\rho}_{k,n}/H_{k,n}$ with $\hat{\rho}_{k,n}$ taken from Fraga Alves *et al.* (2003). Moreover we set $\zeta = \xi^2(1 - 2\rho)(1 - \rho)^2$. The following result is derived in the Appendix.

Theorem. Let $F \in \mathcal{F}(\xi, \tau)$ with $|a(x)| = x^\rho C_a(1 + o(1))$ as $x \rightarrow \infty$. Assume that $\sqrt{ka}(n/k) \rightarrow \lambda$ as $k, n \rightarrow \infty$, $k/n \rightarrow 0$. Setting $\sigma_{k,n}^2 = (k/n)^{-2\rho}$, it follows that $\Xi_{k,n} := \sqrt{k}(\hat{\xi}_k^P - \xi)$ is asymptotically normal with asymptotic mean and variance given by

$$(2.6) \quad E_\infty(\Xi_{k,n}) = \frac{\lambda\rho}{1 - \rho} \frac{\zeta C_a^2 \omega}{\zeta C_a^2 \omega + \rho^4 \lambda^2},$$

$$(2.7) \quad Var_\infty(\Xi_{k,n}) = \frac{\xi^2 \rho^8 \lambda^4}{(\rho^4 \lambda^2 + \zeta C_a^2 \omega)^2} \left(\left(\frac{1 - \rho}{\rho} \right)^2 + \frac{\zeta^2 C_a^4 \omega^2}{\rho^8 \lambda^4} + 2 \frac{\zeta C_a^2 \omega}{\rho^4 \lambda^2} \right).$$

Minimizing $MSE_\infty(\Xi_{k,n}) = E_\infty^2(\Xi_{k,n}) + Var_\infty(\Xi_{k,n})$ with respect to ω , after some lengthy calculations, leads to the asymptotically optimal value

$$\omega_{opt} = C_a^{-2}.$$

One then obtains from (2.6) and (2.7) that

$$E_\infty^{opt}(\Xi_{k,n}) = \frac{\lambda\rho}{1 - \rho} \frac{\zeta}{\zeta + \lambda^2 \rho^4}, \\ Var_\infty^{opt}(\Xi_{k,n}) = \frac{\xi^2}{(\lambda^2 \rho^4 + \zeta)^2} \{ (1 - \rho)^2 \rho^6 \lambda^4 + \zeta^2 + 2\zeta \rho^4 \lambda^2 \},$$

from which

$$(2.8) \quad MSE_\infty^{opt}(\Xi_{k,n}) = \xi^2 + \frac{\lambda^2 \rho^2 \xi^2 (1 - 2\rho)}{\xi^2 (1 - 2\rho)(1 - \rho)^2 + \rho^4 \lambda^2}.$$

Since the right hand side of (2.8) is an increasing function in λ^2 it follows that

$$MSE_\infty^{opt}(\Xi_{k,n}) \leq \lim_{\lambda \rightarrow \infty} MSE_\infty^{opt}(\Xi_{k,n}) = MSE_\infty \left(\sqrt{k}(\hat{\xi}_{k,n}^{ML} - \xi) \right) = \xi^2 \left(\frac{1 - \rho}{\rho} \right)^2.$$

Also, expanding the right hand side of (2.8) for $\lambda^2 \rightarrow 0$ leads to

$$MSE_\infty^{opt}(\Xi_{k,n}) = \xi^2 + \lambda^2 \frac{\rho^2}{(1 - \rho)^2} (1 + o(1)).$$

We can conclude that the asymptotic MSE of the optimal penalized estimator is uniformly smaller than the MSE of the EPD-ML estimator as given in (1.10), while for smaller λ this asymptotic MSE follows the asymptotic MSE of the Hill estimator, given in (1.9), up to terms of order λ^2 . Hence with the penalty $\omega/\sigma_{k,n}^2 = C_a^{-2}(k/n)^{2\rho} = a^{-2}(n/k)$ in (2.2), the penalized ML estimator asymptotically follows the better of the two existing estimators as a function of λ or k .

Replacing $(\hat{\xi}_k, \hat{\delta}_k)$ by $(\hat{\xi}_k^P, \hat{\delta}_k^P)$ in $\hat{p}_{x,k}^{EP}$, it follows from the proof of Theorem 5.2 in Beirlant *et al.* (2009) that the resulting tail probability estimator $\hat{p}_{x,k}^P$ satisfies the following asymptotic result under the conditions of the Theorem:

When $p_n = P(X > x_n)$ satisfies $np_n/k \rightarrow 0$ and $\log(np_n)/\sqrt{k} \rightarrow 0$, then

$$\frac{\sqrt{k}}{\log(k/(np_n))} \xi \left(\frac{\hat{p}_{x_n,k}^P}{p_n} - 1 \right)$$

is asymptotically normal with the same limit distribution as in the Theorem.

Hence the asymptotic MSE behaviour for the tail probability estimator has the same characteristics as the tail index estimator.

From the simulations it will follow that the choice $\omega = 1$ and the use of estimator of ρ taken from Fraga Alves (2003) yields good results. However, in order to alleviate the problem of choosing the number of top order statistics k that are used in the estimation procedure, one can choose ω adaptively with each sample aiming for a plot of $\hat{\xi}_k^P$ as a function of k which is as horizontal as possible. Setting $\hat{\xi}_k^P = \hat{\xi}_k^P(\omega)$ in order to emphasize the dependence of the penalized ML estimator on ω , a possible choice of ω is obtained by minimizing the variance of the resulting estimators for $k = 1, \dots, n$:

$$(2.9) \quad \omega_{mv} = \operatorname{argmin}_{\omega} s_n^2 \left(\hat{\xi}^P(\omega) \right),$$

$$\text{with } s_n^2(\hat{\xi}^P(\omega)) = \frac{1}{n-1} \sum_{k=1}^n \left(\hat{\xi}_k^P(\omega) - \bar{\xi}^P \right)^2.$$

3. SIMULATIONS AND PRACTICAL CASE STUDIES

Both the Bayes maximum a posteriori probability estimator and the penalized maximum likelihood estimator are implemented in **R** using the general **optim** function with default parameters.

We performed a simulation study, taking 1000 repetitions of samples of size $n = 200, 500, 1000$ studying the finite sample behaviour of $\hat{\xi}_{k,n}^P(\omega)$ for different distributions. The bias and RMSE are plotted as a function of k .

The following distributions are used:

- The extreme value distribution (EV) with $F(x) = \exp(-(1 + \xi x)^{-1/\xi})$ ($1 + \xi x > 0$) taking $\xi = 0.25$ in which case $\rho = -0.25$ and $C_a = 1$.
- The Fréchet distribution with $\bar{F}(x) = 1 - \exp(-x^{-1/\xi})$ taking $\xi = 0.5$ in which case $\rho = -1$ and $C_a = 0.25$.
- The Burr distribution with $\bar{F}(x) = (1 + x)^{-4/3}$ so that $\xi = 0.75$ and $\rho = -0.75$ and $C_a = 1$.
- The loggamma distribution with $\bar{F}(x) \sim \text{constant} \times x^{-2}(\log x)^3$ so that $\xi = 0.5$, which does not belong to the class $\mathcal{F}(\xi, \tau)$.

First, in Figures 1-4 we plotted the bias and the RMSE of the Hill estimator H_k , the EPD-ML estimator $\hat{\xi}_k^{ML}$, the penalized ML estimator $\hat{\xi}_k^P(1)$ with $\omega = 1$, the Bayesian estimator $\hat{\xi}_k^B(1)$ with $\omega = 1$, and the minimum variance reduced bias estimator CH_k from Caeiro *et al.* (2005) given by

$$CH_k = H_{k,n} \left(1 - \frac{\hat{\beta}_{k_1}(\hat{\rho}_{k_1})}{1 - \hat{\rho}_{k_1}} \left(\frac{n}{k} \right)^{\hat{\rho}_{k_1}} \right),$$

with

$$\hat{\beta}_k(\rho) = \frac{\left(\frac{k}{n} \right)^\rho \left\{ \left(\frac{1}{k} \sum_{j=1}^k \left(\frac{j}{k} \right)^{-\rho} \right) \left(\frac{1}{k} \sum_{j=1}^k Z_j \right) - \left(\frac{1}{k} \sum_{j=1}^k \left(\frac{j}{k} \right)^{-\rho} Z_j \right) \right\}}{\left(\frac{1}{k} \sum_{j=1}^k \left(\frac{j}{k} \right)^{-\rho} \right) \left(\frac{1}{k} \sum_{j=1}^k \left(\frac{j}{k} \right)^{-\rho} Z_j \right) - \left(\frac{1}{k} \sum_{j=1}^k \left(\frac{j}{k} \right)^{-2\rho} Z_j \right)},$$

where $Z_j := j(\log X_{n-j+1,n} - \log X_{n-j,n})$ ($j = 1, 2, \dots$), and $k_1 = \lfloor n^{0.99} \rfloor$.

In Figure 5 we briefly report on the effect of the choice of ω using $\omega = 1$ and $\omega = \omega_{mv}$ and compare these with the optimal asymptotic RMSE expression from (2.8).

We conclude from the simulations that the finite sample behaviour of the proposed estimators follows the characteristics predicted by the asymptotic analysis to a great extent: for small k the shrinkage estimators $\hat{\xi}_k^P$ and $\hat{\xi}_k^B$ show a similar behaviour as the Hill estimator, while for larger k the proposed estimators tend to follow the characteristics of the bias reduced EPD-ML estimator. In between these two k -regions the shrinkage estimators make a transition from the EPD-ML to the Hill RMSE curve. Only in the Fréchet case the Hill estimator shows a smaller RMSE than the shrinkage estimators for small k , while the shrinkage estimators then still show a much smaller RMSE than the EPD-ML estimator.

The Bayesian implementation shows a smaller RMSE than the penalized ML estimator, except for the Fréchet distribution where both RMSEs are comparable. In the latter case $\hat{\xi}_k^B$ shows a negative bias. Also note that the difference between both the Bayesian and penalized likelihood implementation decreases as n increases.

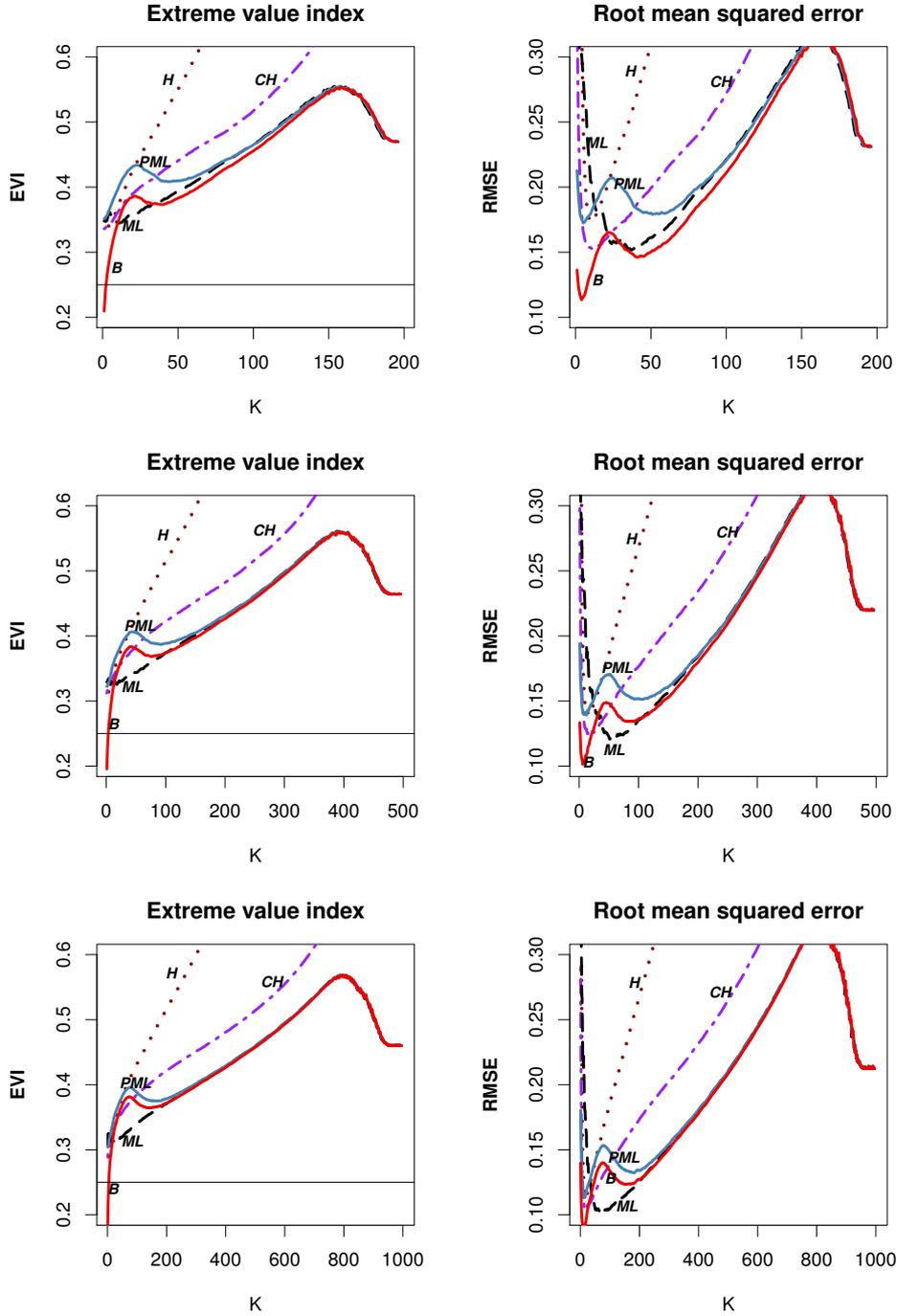


Figure 1: Bias (left) and root mean squared error (right) in case of the **EV distribution** with $\xi = 0.25$ for sample sizes $n = 200$ (top), $n = 500$ (middle) and $n = 1000$ (bottom) for the Hill estimator (H), the EPD-ML estimator $\hat{\xi}_k^{ML}$ (ML), the penalized ML estimator $\hat{\xi}_k^P(1)$ with $\omega = 1$ (PML), the Bayesian estimator $\hat{\xi}_k^B(1)$ with $\omega = 1$ (B), and the minimum variance reduced bias estimator CH_k (CH).

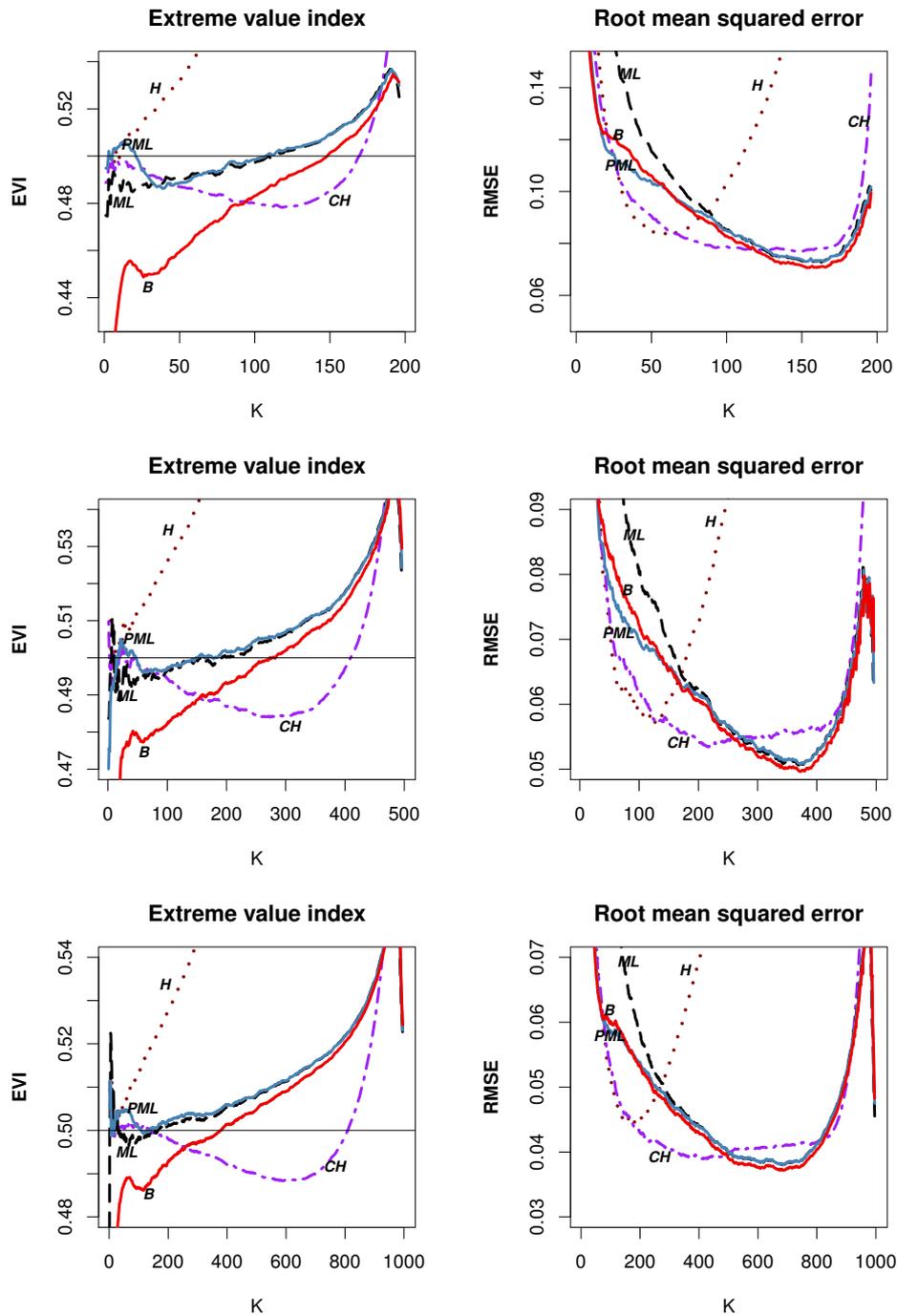


Figure 2: Bias (left) and root mean squared error (right) in case of the **Fréchet distribution** with $\xi = 0.5$ for sample sizes $n = 200$ (top), $n = 500$ (middle) and $n = 1000$ (bottom) for the Hill estimator (H), the EPD-ML estimator $\hat{\xi}_k^{ML}$ (ML), the penalized ML estimator $\hat{\xi}_k^P(1)$ with $\omega = 1$ (PML), the Bayesian estimator $\hat{\xi}_k^B(1)$ with $\omega = 1$ (B), and the minimum variance reduced bias estimator CH_k (CH).

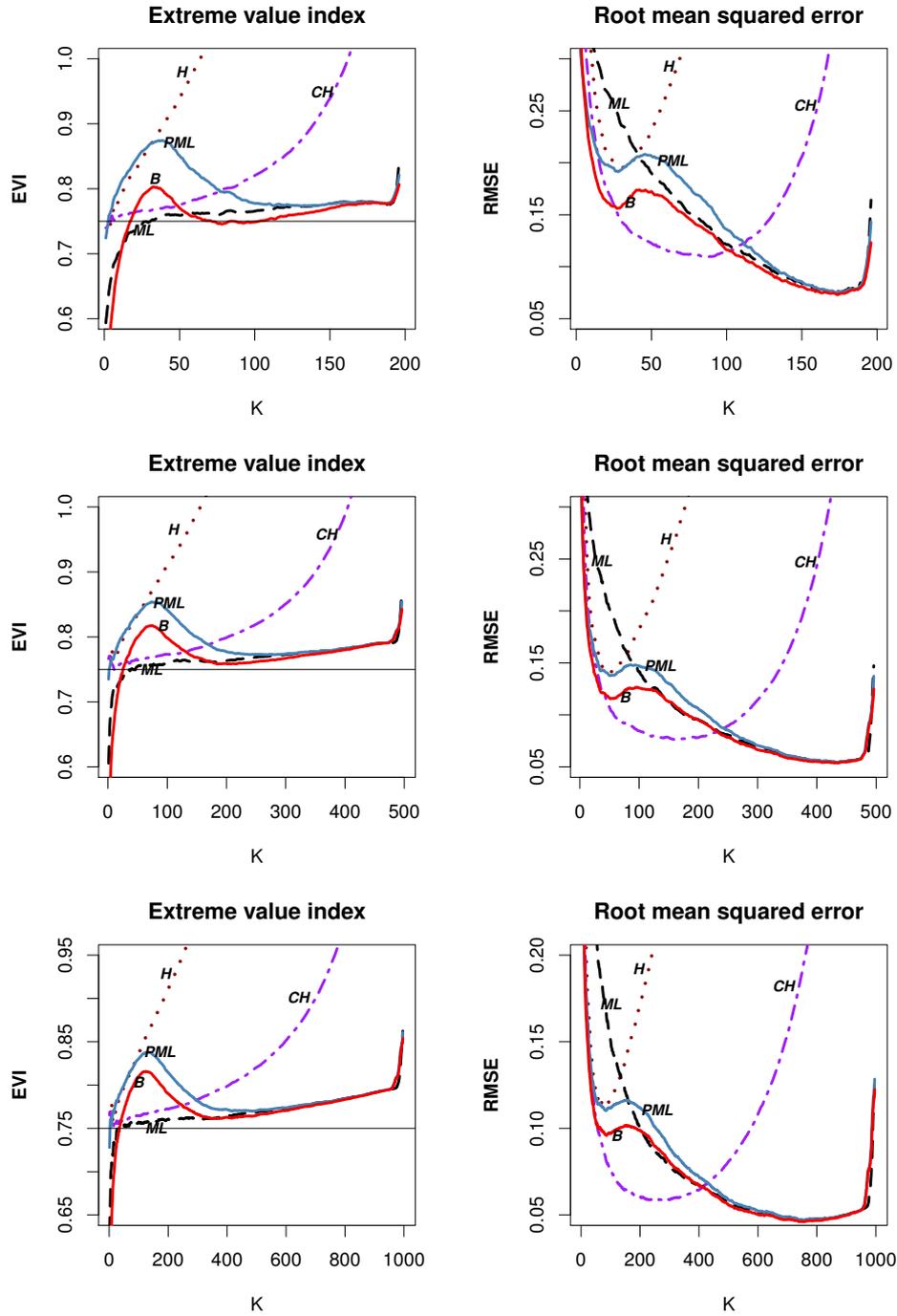


Figure 3: Bias (left) and root mean squared error (right) in case of the **Burr distribution** with $\xi = 0.75$ for sample sizes $n = 200$ (top), $n = 500$ (middle) and $n = 1000$ (bottom) for the Hill estimator (H), the EPD-ML estimator $\hat{\xi}_k^{ML}$ (ML), the penalized ML estimator $\hat{\xi}_k^P(1)$ with $\omega = 1$ (PML), the Bayesian estimator $\hat{\xi}_k^B(1)$ with $\omega = 1$ (B), and the minimum variance reduced bias estimator CH_k (CH).

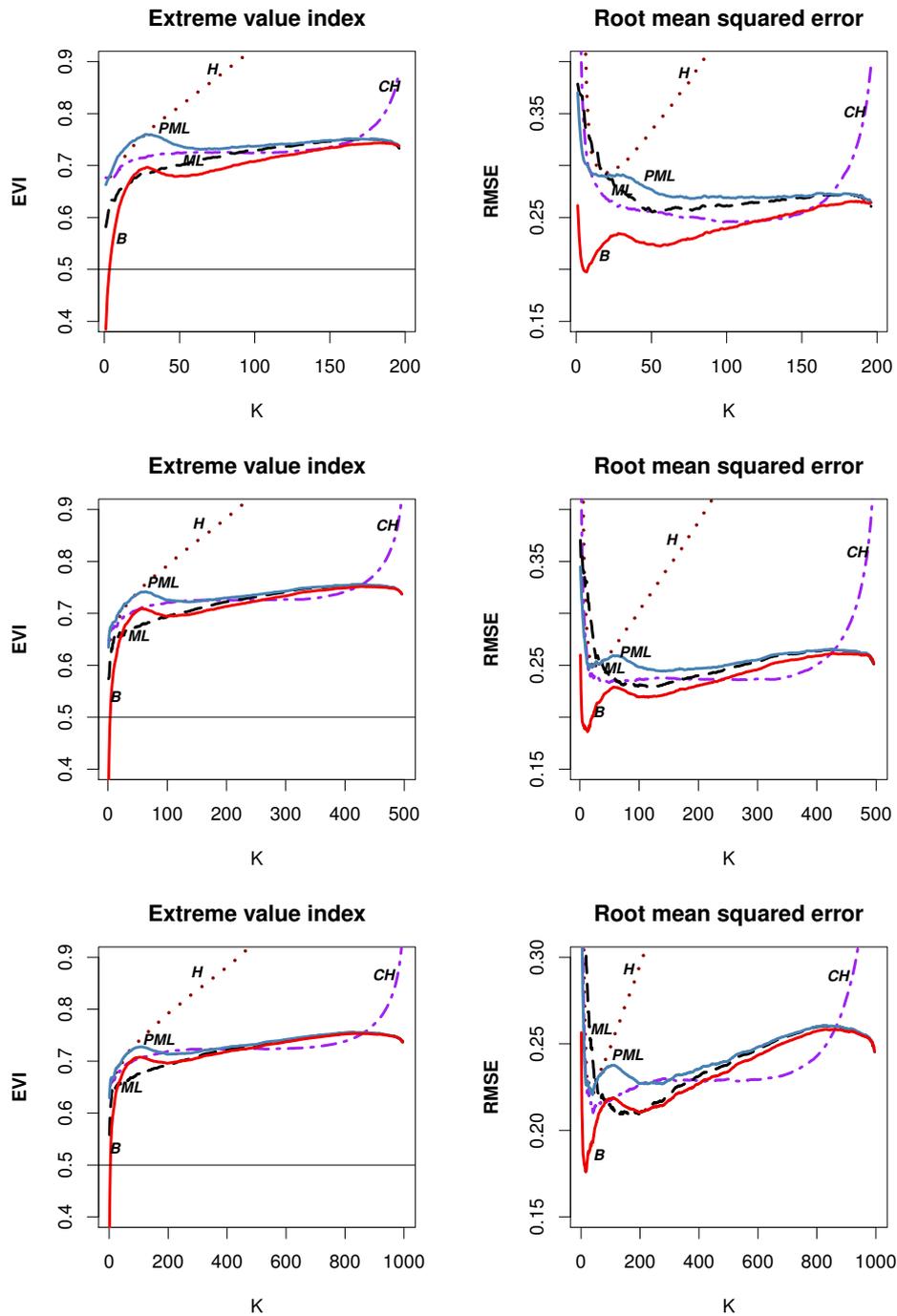


Figure 4: Bias (left) and root mean squared error (right) in case of the **loggamma distribution** with $\xi = 0.5$ for sample sizes $n = 200$ (top), $n = 500$ (middle) and $n = 1000$ (bottom) for the Hill estimator (H), the EPD-ML estimator $\hat{\xi}_k^{ML}$ (ML), the penalized ML estimator $\hat{\xi}_k^P(1)$ with $\omega = 1$ (PML), the Bayesian estimator $\hat{\xi}_k^B(1)$ with $\omega = 1$ (B), and the minimum variance reduced bias estimator CH_k (CH).

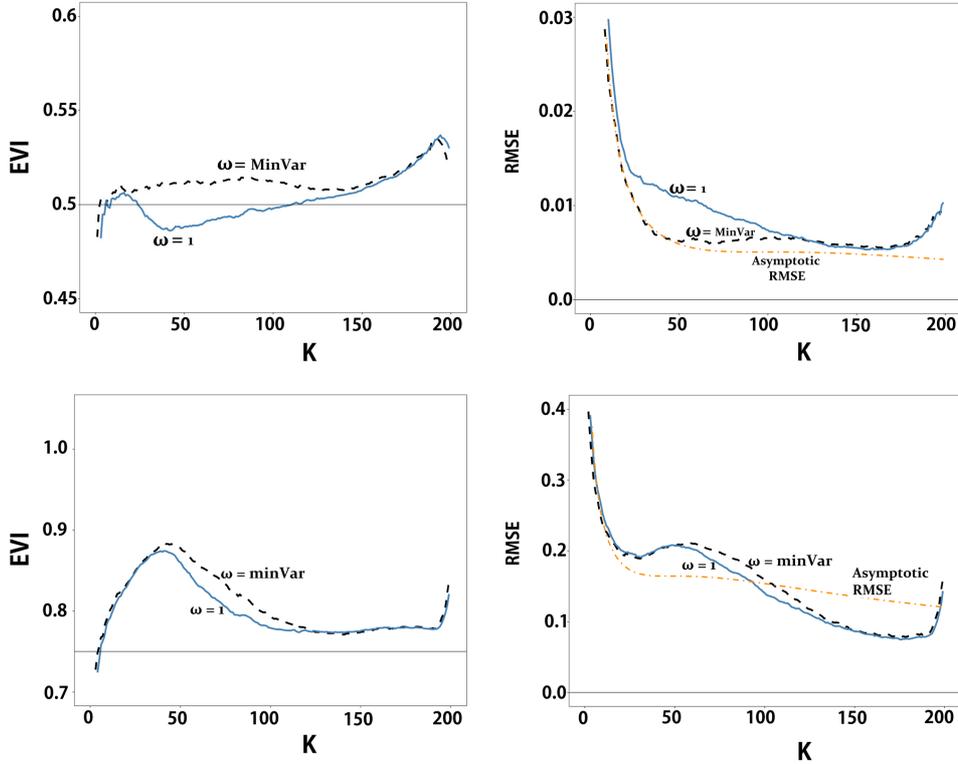


Figure 5: Bias (left) and root mean squared error (right) in case of the **Fréchet distribution** with $\xi = 0.5$ (top) and Burr distribution with $\xi = 0.75$ (bottom) for sample size $n = 200$ comparing the penalized ML estimator $\hat{\xi}_k^P(1)$ with $\omega = 1$, $\omega = \omega_{mv}$ from (2.9), and the optimal asymptotic RMSE from (2.8) replacing λ by $C_a \sqrt{k}(k/n)^{-\rho}$.

The results in case of the loggamma distribution are quite good. Hence it appears that the proposed method exhibits some robustness against deviations from the underlying model.

When the plots of the shrinkage estimators are not systematically increasing with increasing k as in the case of the Fréchet and the Burr distribution, it is useful to use the choice $\omega = \omega_{mv}$ when using the penalized ML estimator. In the case of the Fréchet distribution with $\omega_{opt} = 16$, this adaptive choice of ω leads to a clear RMSE improvement in the transition zone (in k) between the Hill and EPD-ML RMSE behaviour (see Figure 5, top). In the Burr case (see Figure 5, bottom) where $C_a = 1$ and hence $\omega_{opt} = 1$ the choice $\omega = 1$ is best, but the adaptive minimum variance choice $\omega = \omega_{mv}$ is almost as good in RMSE behaviour.

Overall, the proposed shrinkage estimators are competitive with respect to the minimum variance reduced bias estimator CH_k .

In order to illustrate the use of the proposed method we consider the Secura Belgian Re data introduced in section 6.2 in Beirlant *et al.* (2004). For $k \leq 100$ the penalized ML estimator $\hat{\xi}_k^P(1)$ is quite constant and follows the Hill estimator quite closely. This is in contrast with the EPD-ML estimates which vary a lot in that region. The Bayesian estimates $\hat{\xi}_k^B(1)$ and CH estimates show somewhat lower estimates. Beirlant *et al.* (2004) concluded that the Hill estimate in this k -region is an appropriate choice and the adaptive choice $\hat{k} = 98$ was proposed as one of the largest k -values in this region. This proposal is also supported by the present analysis, leading to an estimate $\hat{\xi}^P(1) = 0.28$.

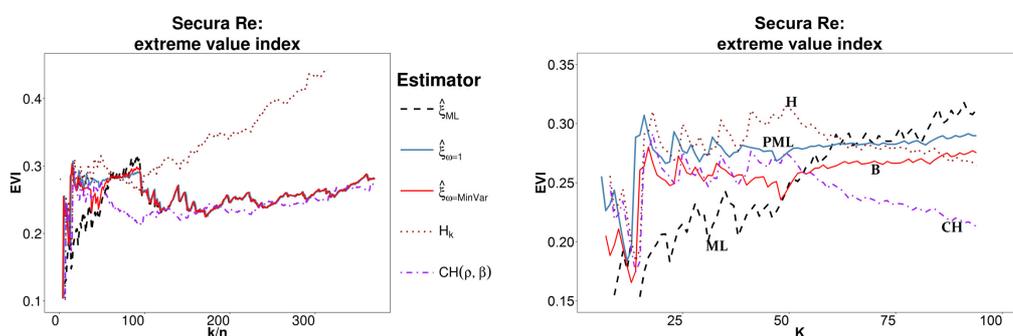


Figure 6: Estimates of ξ for Secura Belgian Re data set: results for the Hill estimator (H), the EPD-ML estimator $\hat{\xi}_k^{ML}$ (ML), the penalized ML estimator $\hat{\xi}_k^P(1)$ with $\omega = 1$ (PML), the Bayesian estimator $\hat{\xi}_k^B(1)$ with $\omega = 1$ (B), and the minimum variance reduced bias estimator CH_k (CH) (left), focused plot for $k = 1, \dots, 100$ (right).

4. CONCLUSION

We introduced the use of shrinkage estimators in tail estimation, in order to obtain bias reduction jointly with good MSE behaviour. Shrinkage estimators can be obtained through a penalized ML approach, or through a Bayesian implementation. For larger thresholds the proposed estimators follow the behaviour of the classical Hill estimator with small bias and minimal variance, while the new estimators are never worse than the corresponding bias reduced ML estimators without penalization. The simulated MSE results are competitive with those of other bias reduced estimators. In contrast to existing minimum variance bias reduced estimators we only use second order slow variation conditions.

APPENDIX

Derivation of the expressions of $(\hat{\xi}_k^P, \hat{\delta}_k^P)$. First consider the asymptotic approximations of the penalize ML estimator of ξ based on maximization of (2.2). From (2.1)–(2.2) using expansions in $\delta \rightarrow 0$ we obtain

$$\begin{aligned} \frac{1}{k} \log l_{pen}(\xi, \delta | \mathbf{y}) &= -(1 + \frac{1}{k}) \log \xi - \frac{1}{k}(1 + \xi) - (\frac{1}{\xi} + 1) \frac{1}{k} \sum_{j=1}^k \log y_{j,k} \\ &\quad - \frac{\delta}{1 + \xi} \frac{1}{k} \sum_{j=1}^k (1 - y_{j,k}^\tau) + \delta \frac{1}{k} \sum_{j=1}^k (1 - (1 + \tau) y_{j,k}^\tau) \\ &\quad - \frac{\omega \delta^2}{2k\sigma_{k,n}^2} + \frac{\delta^2}{2(1 + \xi)} \frac{1}{k} \sum_{j=1}^k (1 - y_{j,k}^\tau)^2 - \frac{\delta^2}{2} \frac{1}{k} \sum_{j=1}^k (1 - (1 + \tau) y_{j,k}^\tau)^2 \\ &\quad + O(\delta^3) + c, \end{aligned}$$

where c is a constant only depending on $\sigma_{k,n}^2$ and τ . Note that $\frac{1}{k} \sum_{j=1}^k \log y_{j,k} = H_{k,n}$. Then the score functions admit the following expansions in $\delta \downarrow 0$ for $j = 1, \dots, k$:

$$\begin{aligned} \frac{\partial}{\partial \xi} \log l_{pen}(\xi, \delta | y_{j,k}) &= -\frac{1}{\xi} + \frac{1}{\xi^2} \log y_{j,k} + \frac{\delta}{\xi^2} (1 - y_{j,k}^\tau) + O(\delta^2), \\ \frac{\partial}{\partial \delta} \log l_{pen}(\xi, \delta | y_{j,k}) &= -\frac{1}{\xi} (1 - (1 - \xi\tau) y_{j,k}^\tau) - \frac{\omega \delta}{k\sigma_{k,n}^2} \\ &\quad + \frac{\delta}{\xi} (1 - 2(1 - \xi\tau) y_{j,k}^\tau + (1 - 2\xi\tau - \xi\tau^2) y_{j,k}^{2\tau}) + O(\delta^2). \end{aligned}$$

Derivation of Theorem. Note that as $k, n \rightarrow \infty, k/n \rightarrow 0$ and $\sqrt{k} a(n/k) \rightarrow \lambda$, we also have $k\sigma_{k,n}^2 \rightarrow \lambda^2 C_a^{-2}$. Also as $\sqrt{k} a(n/k) \rightarrow \lambda$ we find using $E_{k,n}(s) \rightarrow 1/(1 - \xi s)$ (see Theorem A.1 in Beirlant *et al.*, 2009) that

$$D_{k,n}^P = -\frac{\xi C_a^2}{\lambda^2} + \frac{\rho^4}{\xi(1 - 2\rho)(1 - \rho)^2} + o_p(1).$$

Then, proceeding as in the proof of Theorem 3.1 in Beirlant *et al.* (2009), we obtain with $\Gamma_{k,n} = \sqrt{k}(H_{k,n} - \xi)$, $\mathbb{E}_{k,n}(s) = \sqrt{k}(E_{k,n}(s) - \frac{1}{1 - \xi s})$ ($s < 0$), that

$$\begin{aligned} \sqrt{k} (\hat{\xi}_k^P - \xi) &= \sqrt{k} \left(H_{k,n} - \xi - \hat{\delta}_k^P \frac{\rho}{1 - \rho} \right) \\ &= \Gamma_{k,n} - \frac{\rho}{1 - \rho} \sqrt{k} \hat{\delta}_k^P \\ &= \Gamma_{k,n} \left(1 + \frac{\rho^2}{\xi(1 - \rho^2)} \frac{1}{\xi C_a^2 / \lambda^2 + \rho^4 / \xi(1 - 2\rho)(1 - \rho)^2} \right) \\ &\quad - \frac{\rho}{\xi C_a^2 / \lambda^2 + \rho^4 / \xi(1 - 2\rho)(1 - \rho)^2} \mathbb{E}_{k,n}(\hat{\tau}) + o_p(1) \\ &= \Gamma_{k,n} \left(1 + \frac{\rho^2(1 - 2\rho)}{\zeta + \rho^4} \right) + \mathbb{E}_{k,n}(\hat{\tau}) \left(\frac{(-\rho)\xi(1 - 2\rho)(1 - \rho)^2}{\rho^4 + \zeta} \right) + o_p(1). \end{aligned}$$

Using Theorem A.1 in Beirlant *et al.* (2009), (2.6) and (2.7) follow under $\sqrt{k} a(n/k) \rightarrow \lambda$.

ACKNOWLEDGMENTS

This work is based on the research supported wholly/in part by the National Research Foundation of South Africa (Grant Number 102628). The Grantholder acknowledges that opinions, findings and conclusions or recommendations expressed in any publication generated by the NRF supported research is that of the author(s), and that the NRF accepts no liability whatsoever in this regard.

The authors take pleasure in thanking S. van der Merwe for his valuable advice concerning the Bayesian implementations.

Thanks to the referee for the constructive comments which improved the presentation of the paper significantly.

REFERENCES

- [1] BEIRLANT, J.; DIERCKX, G.; GOEGBEUR, Y. and MATTHYS, G. (1999). Tail index estimation and an exponential regression model, *Extremes*, **2**(2), 177–200.
- [2] BEIRLANT, J.; GOEGBEUR, Y.; SEGERS, J. and TEUGELS, J. (2004). *Statistics of Extremes: Theory and Applications*, Wiley, Chichester.
- [3] BEIRLANT, J.; FIGUEIREDO, F.; GOMES, M.I. and VANDEWALLE, B. (2008). Improved reduced-bias tail index and quantile estimators, *Journal of Statistical Planning and Inference*, **138**(6), 1851–1870.
- [4] BEIRLANT, J.; JOOSSENS, E. and SEGERS, J. (2009). Second-order refined peaks-over-threshold modelling for heavy-tailed distribution, *Journal of Statistical Planning and Inference*, **139**, 2800–2815.
- [5] BINGHAM, N.H.; GOLDIE, C.M. and TEUGELS, J.L. (1987). *Regular Variation*, Cambridge University Press.
- [6] CAEIRO, F.; GOMES, M.I. and PESTANA, D. (2005). Direct reduction of bias of the classical Hill estimator, *Revstat*, **3**(2), 113–136.
- [7] CAEIRO, F.; GOMES, M.I. and RODRIGUES, L.H. (2009). Reduced-bias tail index estimators under a third-order framework, *Communications in Statistics, Theory and Methods*, **38**(7), 1019–1040.
- [8] FEUERVERGER, A. and HALL, P. (1999). Estimating a tail exponent by modelling departure from a Pareto distribution, *The Annals of Statistics*, **27**(2), 760–781.
- [9] FRAGA ALVES, M.I.; GOMES, M.I. and DE HAAN, L. (2003). A new class of semi-parametric estimators of the second order parameter, *Portugaliae Mathematica*, **60**(2), 193–214.
- [10] GOMES, M.I.; MARTINS, M.J. and NEVES, M. (2000). Alternatives to a semi-parametric estimator of parameters of rare events — The jackknife methodology, *Extremes*, **3**(3), 207–229.

- [11] HALL, P. (1982). On some simple estimates of an exponent of regular variation, *Journal of the Royal Statistical Society. Series B (Methodological)*, 37–42.
- [12] HILL, B.M. (1975). A simple general approach to inference about the tail of a distribution, *The Annals of Statistics*, **3**(5), 1163–1174.
- [13] PENG, L. (1998). Asymptotically unbiased estimators for the extreme-value index, *Statistics & Probability Letters*, **38**(2), 107–115.
- [14] 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.
- [15] ZELLNER, A. (1971). *An Introduction to Bayesian Inference in Econometrics*, Wiley.

AN INTEGRATED FUNCTIONAL WEISSMAN ESTIMATOR FOR CONDITIONAL EXTREME QUANTILES

Authors: LAURENT GARDES
– University of Strasbourg, CNRS, IRMA 7501,
F-67000 Strasbourg, France
gardes@unistra.fr

GILLES STUPFLER
– School of Mathematical Sciences, The University of Nottingham,
University Park, Nottingham NG7 2RD, United Kingdom
Gilles.Stupfler@nottingham.ac.uk

Received: September 2016 Revised: January 2017 Accepted: February 2017

Abstract:

- It is well-known that estimating extreme quantiles, namely, quantiles lying beyond the range of the available data, is a nontrivial problem that involves the analysis of tail behavior through the estimation of the extreme-value index. For heavy-tailed distributions, on which this paper focuses, the extreme-value index is often called the tail index and extreme quantile estimation typically involves an extrapolation procedure. Besides, in various applications, the random variable of interest can be linked to a random covariate. In such a situation, extreme quantiles and the tail index are functions of the covariate and are referred to as conditional extreme quantiles and the conditional tail index, respectively. The goal of this paper is to provide classes of estimators of these quantities when there is a functional (*i.e.* possibly infinite-dimensional) covariate. Our estimators are obtained by combining regression techniques with a generalization of a classical extrapolation formula. We analyze the asymptotic properties of these estimators, and we illustrate the finite-sample performance of our conditional extreme quantile estimator on a simulation study and on a real chemometric data set.

Key-Words:

- *heavy-tailed distribution; functional random covariate; extreme quantile; tail index; asymptotic normality.*

AMS Subject Classification:

- 62G05, 62G20, 62G30, 62G32.

1. INTRODUCTION

Studying extreme events is relevant in numerous fields of statistical applications. In hydrology for example, it is of interest to estimate the maximum level reached by seawater along a coast over a given period, or to study extreme rainfall at a given location; in actuarial science, a major problem for an insurance firm is to estimate the probability that a claim so large that it represents a threat to its solvency is filed. When analyzing the extremes of a random variable, a central issue is that the straightforward empirical estimator of the quantile function is not consistent at extreme levels; in other words, direct estimation of a quantile exceeding the range covered by the available data is impossible, and this is of course an obstacle to meaningful estimation results in practice.

In many of the aforementioned applications, the problem can be accurately modeled using univariate heavy-tailed distributions, thus providing an extrapolation method to estimate extreme quantiles. Roughly speaking, a distribution is said to be heavy-tailed if and only if its related survival function decays like a power function with negative exponent at infinity; its so-called tail index γ is then the parameter which controls its rate of convergence to 0 at infinity. If Q denotes the underlying quantile function, this translates into: $Q(\delta) \approx [(1 - \beta)/(1 - \delta)]^\gamma Q(\beta)$ when β and δ are close to 1. The quantile function at an arbitrarily high extreme level can then be consistently deduced from its value at a typically much smaller level provided γ can be consistently estimated. This procedure, suggested by Weissman [42], is one of the simplest and most popular devices as far as extreme quantile estimation is concerned.

The estimation of the tail index γ , an excellent overview of which is given in the recent monographs by Beirlant *et al.* [2] and de Haan and Ferreira [27], is therefore a crucial step to gain understanding of the extremes of a random variable whose distribution is heavy-tailed. In practical applications, the variable of interest Y can often be linked to a covariate X . For instance, the value of rainfall at a given location depends on its geographical coordinates; in actuarial science, the claim size depends on the sum insured by the policy. In this situation, the tail index and quantiles of the random variable Y given $X = x$ are functions of x to which we shall refer as the conditional tail index and conditional quantile functions. Their estimation has been considered first in the “fixed design” case, namely when the covariates are nonrandom. Smith [36] and Davison and Smith [12] considered a regression model while Hall and Tajvidi [28] used a semi-parametric approach to estimate the conditional tail index. Fully nonparametric methods have been developed using splines (see Chavez-Demoulin and Davison [6]), local polynomials (see Davison and Ramesh [11]), a moving window approach (see Gardes and Girard [19]) and a nearest neighbor approach (see Gardes and Girard [20]), among others.

Despite the great interest in practice, the study of the random covariate case has been initiated only recently. We refer to the works of Wang and Tsai [41], based on a maximum likelihood approach, Daouia *et al.* [9] who used a fixed number of non parametric conditional quantile estimators to estimate the conditional tail index, later generalized in Daouia *et al.* [10] to a regression context with conditional response distributions belonging to the general max-domain of attraction, Gardes and Girard [21] who introduced a local generalized Pickands-type estimator (see Pickands [33]), Goegebeur *et al.* [25], who studied a non-parametric regression estimator whose strong uniform properties are examined in Goegebeur *et al.* [26]. Some generalizations of the popular moment estimator of Dekkers *et al.* [13] have been proposed by Gardes [18], Goegebeur *et al.* [23, 24] and Stupfler [37, 38]. In an attempt to obtain an estimator behaving better in finite-sample situations, Gardes and Stupfler [22] worked on a smoothed local Hill estimator (see Hill [29]) related to the work of Resnick and Stărică [34]. A different approach, that has been successful in recent years, is to combine extreme value theory and quantile regression: the pioneering paper is Chernozhukov [7], and we also refer to the subsequent papers by Chernozhukov and Du [8], Wang *et al.* [39] and Wang and Li [40].

The goal of this paper is to introduce integrated estimators of conditional extreme quantiles and of the conditional tail index for random, possibly infinite-dimensional, covariates. In particular, our estimator of the conditional tail index, based on the integration of a conditional log-quantile estimator, is somewhat related to the one of Gardes and Girard [19]. Our aim is to examine the asymptotic properties of our estimators, as well as to examine the applicability of our conditional extreme quantile estimator on numerical examples and on real data. Our paper is organized as follows: we define our estimators in Section 2. Their asymptotic properties are stated in Section 3. A simulation study is provided in Section 4 and we revisit a set of real chemometric data in Section 5. All the auxiliary results and proofs are deferred to the Appendix.

2. FUNCTIONAL EXTREME QUANTILE: DEFINITION AND ESTIMATION

Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be n independent copies of a random pair (X, Y) taking its values in $\mathcal{E} \times \mathbb{R}_+$ where (\mathcal{E}, d) is a (not necessarily finite-dimensional) Polish space endowed with a semi-metric d . For instance, \mathcal{E} can be the standard p -dimensional space \mathbb{R}^p , a space of continuous functions over a compact metric space, or a Lebesgue space $L^p(\mathbb{R})$, to name a few. For $y > 0$, we denote by $S(y|X)$ a regular version of the conditional probability $\mathbb{P}(Y > y|X)$. Note that since \mathcal{E} is a Polish space, such conditional probabilities always exist, see Jiřina [30].

In this paper, we focus on the situation where the conditional distribution of Y given X is heavy-tailed. More precisely, we assume that there exists a positive function $\gamma(\cdot)$, called the conditional tail index, such that

$$(2.1) \quad \lim_{y \rightarrow \infty} \frac{S(\lambda y|x)}{S(y|x)} = \lambda^{-1/\gamma(x)},$$

for all $x \in \mathcal{E}$ and all $\lambda > 0$. This is the adaptation of the standard extreme-value framework of heavy-tailed distributions to the case when there is a covariate. The conditional quantile function of Y given $X = x$ is then defined for $x \in \mathcal{E}$ by $Q(\alpha|x) := \inf \{y > 0 \mid S(y|x) \leq \alpha\}$. If $x \in \mathcal{E}$ is fixed, our final aim is to estimate the conditional extreme quantile $Q(\beta_n|x)$ of order $\beta_n \rightarrow 0$. As we will show below, this does in fact require estimating the conditional tail index $\gamma(x)$ first.

2.1. Estimation of a functional extreme quantile

Recall that we are interested in the estimation of $Q(\beta_n|x)$ when $\beta_n \rightarrow 0$ as the sample size increases. The natural empirical estimator of this quantity is given by

$$(2.2) \quad \widehat{Q}_n(\beta_n|x) := \inf \left\{ y > 0 \mid \widehat{S}_n(y|x) \leq \beta_n \right\},$$

where

$$\widehat{S}_n(y|x) = \frac{\sum_{i=1}^n \mathbb{I}\{Y_i > y\} \mathbb{I}\{d(x, X_i) \leq h\}}{\sum_{i=1}^n \mathbb{I}\{d(x, X_i) \leq h\}}$$

and where $h = h(n)$ is a nonrandom sequence converging to 0 as $n \rightarrow \infty$. Unfortunately, denoting by $m_x(h) := n\mathbb{P}(d(x, X) \leq h)$ the average number of observations whose covariates belong to the ball $B(x, h) = \{x' \in \mathcal{E} \mid d(x, x') \leq h\}$ with center x and radius h , it can be shown (see Proposition 6.1) that the condition $m_x(h)\beta_n \rightarrow \infty$ is required to obtain the consistency of $\widehat{Q}_n(\beta_n|x)$. This means that at the same time, sufficiently many observations should belong to the ball $B(x, h)$ and β_n should be so small that the quantile $Q(\beta_n|x)$ is covered by the range of this data, and therefore the order β_n of the functional extreme quantile cannot be chosen as small as we would like. We thus need to propose another estimator adapted to this case. To this end, we start by remarking (see Bingham *et al.* [4, Theorem 1.5.12]) that (2.1) is equivalent to

$$(2.3) \quad \lim_{\alpha \rightarrow 0} \frac{Q(\lambda\alpha|x)}{Q(\alpha|x)} = \lambda^{-\gamma(x)},$$

for all $\lambda > 0$. Hence, for $0 < \beta < \alpha$ with α small enough, we obtain the extrapolation formula $Q(\beta|x) \approx Q(\alpha|x)(\alpha/\beta)^{\gamma(x)}$ which is at the heart of Weissman's extrapolation method [42]. In order to borrow more strength from the available

information in the sample, we note that, if μ is a probability measure on the interval $[0, 1]$, another similar, heuristic approximation holds:

$$Q(\beta|x) \approx \int_{[0,1]} Q(\alpha|x) \left(\frac{\alpha}{\beta}\right)^{\gamma(x)} \mu(d\alpha).$$

If we have at our disposal a consistent estimator $\widehat{\gamma}_n(x)$ of $\gamma(x)$ (an example of such an estimator is given in Section 2.2), an idea is to estimate $Q(\beta_n|x)$ by:

$$(2.4) \quad \check{Q}_n(\beta_n|x) = \int_{[0,1]} \widehat{Q}_n(\alpha|x) \left(\frac{\alpha}{\beta_n}\right)^{\widehat{\gamma}_n(x)} \mu(d\alpha).$$

In order to obtain a consistent estimator of the extreme conditional quantile, the support of the measure μ , denoted by $\text{supp}(\mu)$, should be located around 0. To be more specific, we assume in what follows that $\text{supp}(\mu) \subset [\tau u, u]$ for some $\tau \in (0, 1]$ and $u \in (0, 1)$ small enough. For instance, taking μ to be the Dirac measure at u leads to $\check{Q}_n(\beta_n|x) = \widehat{Q}_n(u|x) (u/\beta_n)^{\widehat{\gamma}_n(x)}$, which is a straightforward adaptation to our conditional setting of the classical Weissman estimator [42]. If on the contrary μ is absolutely continuous, estimator (2.4) is a properly integrated and weighted version of Weissman's estimator. Due to the fact that it takes more of the available data into account, we can expect such an estimator to perform better than the simple adaptation of Weissman's estimator, a claim we investigate in our finite-sample study in Section 4.

2.2. Estimation of the functional tail index

To provide an estimator of the functional tail index $\gamma(x)$, we note that equation (2.3) warrants the approximation $\gamma(x) \approx \log[Q(\alpha|x)/Q(u|x)]/\log(u/\alpha)$ for $0 < \alpha < u$ when u is small enough. Let $\Psi(\cdot, u)$ be a measurable function defined on $(0, u)$ such that $0 < |\int_0^u \log(u/\alpha)\Psi(\alpha, u)d\alpha| < \infty$. Multiplying the aforementioned approximation by $\Psi(\cdot, u)$, integrating between 0 and 1 and replacing $Q(\cdot|x)$ by the classical estimator $\widehat{Q}_n(\cdot|x)$ defined in (2.2) leads to the estimator:

$$(2.5) \quad \widehat{\gamma}_n(x, u) := \int_0^u \Psi(\alpha, u) \log \frac{\widehat{Q}_n(\alpha|x)}{\widehat{Q}_n(u|x)} d\alpha \Big/ \int_0^u \log(u/\alpha)\Psi(\alpha, u)d\alpha.$$

Without loss of generality, we shall assume in what follows that

$$\int_0^u \log(u/\alpha)\Psi(\alpha, u)d\alpha = 1.$$

Particular choices of the function $\Psi(\cdot, u)$ actually yield generalizations of some well-known tail index estimators to the conditional framework. Let $k_x := uM_x(h)$, where $M_x(h)$ is the total number of covariates whose distance to x is not greater than h :

$$M_x(h) = \sum_{i=1}^n \mathbb{I}\{d(x, X_i) \leq h\}.$$

The choice $\Psi(\cdot, u) = 1/u$ leads to the estimator:

$$(2.6) \quad \widehat{\gamma}_n^H(x) = \frac{1}{k_x} \sum_{i=1}^{\lfloor k_x \rfloor} \log \frac{\widehat{Q}_n((i-1)/M_x(h)|x)}{\widehat{Q}_n(k_x/M_x(h)|x)},$$

which is the straightforward conditional adaptation of the classical Hill estimator (see Hill [29]). Now, taking $\Psi(\cdot, u) = u^{-1}(\log(u/\cdot) - 1)$ leads, after some algebra, to the estimator:

$$\widehat{\gamma}_n^Z(x) = \frac{1}{k_x} \sum_{i=1}^{\lfloor k_x \rfloor} i \log \left(\frac{k_x}{i} \right) \log \frac{\widehat{Q}_n((i-1)/M_x(h)|x)}{\widehat{Q}_n(i/M_x(h)|x)}.$$

This estimator can be seen as a generalization of the Zipf estimator (see Kratz and Resnick [31], Schultze and Steinebach [35]).

3. MAIN RESULTS

Our aim is now to establish asymptotic results for our estimators. We assume in all what follows that $Q(\cdot|x)$ is continuous and decreasing. Particular consequences of this condition include that $S(Q(\alpha|x)|x) = \alpha$ for any $\alpha \in (0, 1)$ and that given $X = x$, Y has an absolutely continuous distribution with probability density function $f(\cdot|x)$.

Recall that under (2.1), or equivalently (2.3), the conditional quantile function may be written for all $t > 1$ as follows:

$$Q(t^{-1}|x) = c(t|x) \exp \left(\int_1^t \frac{\Delta(v|x) - \gamma(x)}{v} dv \right),$$

where $c(\cdot|x)$ is a positive function converging to a positive constant at infinity and $\Delta(\cdot|x)$ is a measurable function converging to 0 at infinity, see Bingham *et al.* [4, Theorem 1.3.1]. We assume in what follows that

- (H_{SO}) $c(\cdot|x)$ is a constant function equal to $c(x) > 0$, the function $\Delta(\cdot|x)$ has ultimately constant sign at infinity and there exists $\rho(x) < 0$ such that for all $\lambda > 0$,

$$\lim_{y \rightarrow \infty} \left| \frac{\Delta(\lambda y|x)}{\Delta(y|x)} \right| = \lambda^{\rho(x)}.$$

The constant $\rho(x)$ is called the conditional second-order parameter of the distribution. These conditions on the function $\Delta(\cdot|x)$ are commonly used when studying tail index estimators and make it possible to control the error term in convergence (2.3). In particular, it is straightforward to see that for all $z > 0$,

$$(3.1) \quad \lim_{t \rightarrow \infty} \frac{1}{\Delta(t|x)} \left(\frac{Q((tz)^{-1}|x)}{Q(t^{-1}|x)} - z^{\gamma(x)} \right) = z^{\gamma(x)} \frac{z^{\rho(x)} - 1}{\rho(x)},$$

which is the conditional analogue of the second-order condition of de Haan and Ferreira [27] for heavy-tailed distributions, see Theorem 2.3.9 therein.

Finally, for $0 < \alpha_1 < \alpha_2 < 1$, we introduce the quantity:

$$\omega(\alpha_1, \alpha_2, x, h) = \sup_{\alpha \in [\alpha_1, \alpha_2]} \sup_{x' \in B(x, h)} \left| \log \frac{Q(\alpha|x')}{Q(\alpha|x)} \right|,$$

which is the uniform oscillation of the log-quantile function in its second argument. Such a quantity is also studied in Gardes and Stupfler [22], for instance. It acts as a measure of how close conditional distributions are for two neighboring values of the covariate.

These elements make it possible to state an asymptotic result for our conditional extreme quantile estimator:

Theorem 3.1. *Assume that conditions (2.3) and (H_{SO}) are satisfied and let $u_{n,x} \in (0, 1)$ be a sequence converging to 0 and such that $\text{supp}(\mu) \subset [\tau u_{n,x}, u_{n,x}]$ with $\tau \in (0, 1]$. Assume also that $m_x(h) \rightarrow \infty$ and that there exists $a(x) \in (0, 1)$ such that:*

$$(3.2) \quad c_1 \leq \liminf_{n \rightarrow \infty} u_{n,x} [m_x(h)]^{a(x)} \leq \limsup_{n \rightarrow \infty} u_{n,x} [m_x(h)]^{a(x)} \leq c_2$$

for some constants $0 < c_1 \leq c_2$, $z^{1-a(x)} \Delta^2(z^{a(x)}|x) \rightarrow \lambda(x) \in \mathbb{R}$ as $z \rightarrow \infty$ and

$$(3.3) \quad [m_x(h)]^{1-a(x)} \omega^2 \left([m_x(h)]^{-1-\delta}, 1 - [m_x(h)]^{-1-\delta}, x, h \right) \rightarrow 0$$

for some $\delta > 0$. If moreover $[m_x(h)]^{(1-a(x))/2} (\hat{\gamma}_n(x) - \gamma(x)) \xrightarrow{d} \Gamma$ with Γ a non-degenerate distribution, then, provided we have that $\beta_n [m_x(h)]^{a(x)} \rightarrow 0$ and $[m_x(h)]^{a(x)-1} \log^2([m_x(h)]^{-a(x)}/\beta_n) \rightarrow 0$, it holds that

$$\frac{[m_x(h)]^{(1-a(x))/2}}{\log([m_x(h)]^{-a(x)}/\beta_n)} \left(\frac{\check{Q}_n(\beta_n|x)}{Q(\beta_n|x)} - 1 \right) \xrightarrow{d} \Gamma.$$

Note that $[m_x(h)]^{1-a(x)} \rightarrow \infty$ depends on the average number of available data points that can be used to compute the estimator. More precisely, under condition (3.2), this quantity is essentially proportional to $u_{n,x} m_x(h)$, which is the average number of data points actually used in the estimation. In particular, the conditions in Theorem 3.1 are analogues of the classical hypotheses in the estimation of an extreme quantile. Besides, condition (3.3) ensures that the distribution of Y given $X = x'$ is close enough to that of Y given $X = x$ when x' is in a sufficiently small neighborhood of x . Finally, taking μ to be the Dirac measure at $u_{n,x}$ makes it possible to obtain the asymptotic properties of the functional adaptation of the standard Weissman extreme quantile estimator. In particular, as in the unconditional univariate case, the asymptotic distribution of the conditional extrapolated estimator depends crucially on the asymptotic properties of the conditional tail index estimator used.

We proceed by stating the asymptotic normality of the estimator $\widehat{\gamma}_n(x, u)$ in (2.5). To this end, an additional hypothesis on the weighting function $\Psi(\cdot, u)$ is required.

(H_Ψ) The function $\Psi(\cdot, u)$ satisfies for all $u \in (0, 1]$ and $\beta \in (0, u]$:

$$\frac{u}{\beta} \int_0^\beta \Psi(\alpha, u) d\alpha = \Phi(\beta/u) \quad \text{and} \quad \sup_{0 < v \leq 1/2} \int_0^v |\Psi(\alpha, v)| d\alpha < \infty,$$

where Φ is a nonincreasing probability density function on $(0, 1)$ such that $\Phi^{2+\kappa}$ is integrable for some $\kappa > 0$. In addition, there exists a positive continuous function g defined on $(0, 1)$ such that for any $k > 1$ and $i \in \{1, 2, \dots, k\}$,

$$(3.4) \quad |i\Phi(i/k) - (i-1)\Phi((i-1)/k)| \leq g(i/(k+1)),$$

and the function $g(\cdot) \max(\log(1/\cdot), 1)$ is integrable on $(0, 1)$.

Note that for all $t \in (0, 1)$, $0 \leq t\Phi(t) \leq \int_0^{t/2} |\Psi(\alpha, 1/2)| d\alpha$. Since the right-hand side converges to 0 as $t \downarrow 0$, we may extend the definition of the map $t \mapsto t\Phi(t)$ by saying it is 0 at $t = 0$. Hence, inequality (3.4) is meaningful even when $i = 1$.

Condition (H_Ψ) on the weighting function $\Psi(\cdot, u)$ is similar in spirit to a condition introduced in Beirlant *et al.* [1]. This condition is satisfied for instance by the functions $\Psi(\cdot, u) = u^{-1}$ and $\Psi(\cdot, u) = u^{-1}(\log(u/\cdot) - 1)$ with $g(\cdot) = 1$ for the first one and, for the second one, $g(\cdot) = 1 - \log(\cdot)$. In particular, our results shall then hold for the adaptations of the Hill and Zipf estimators mentioned at the end of Section 2.2.

The asymptotic normality of our family of estimators of $\gamma(x)$ is established in the following theorem.

Theorem 3.2. *Assume that conditions (2.3), (H_{SO}) and (H_Ψ) are satisfied, that $m_x(h) \rightarrow \infty$ and $u = u_{n,x} \rightarrow 0$. Assume that there exists $a(x) \in (0, 1)$ such that $z^{1-a(x)} \Delta^2(z^{a(x)}|x) \rightarrow \lambda(x) \in \mathbb{R}$ as $z \rightarrow \infty$, condition (3.3) holds and that there are two ultimately decreasing functions $\varphi_1 \leq \varphi_2$ such that $z^{1-a(x)} \varphi_2^2(z) \rightarrow 0$ as $z \rightarrow \infty$ and $\varphi_1(m_x(h)) \leq u_{n,x} [m_x(h)]^{a(x)} - 1 \leq \varphi_2(m_x(h))$. Then, $[m_x(h)]^{(1-a(x))/2} (\widehat{\gamma}_n(x, u_{n,x}) - \gamma(x))$ converges in distribution to*

$$\mathcal{N} \left(\lambda(x) \int_0^1 \Phi(\alpha) \alpha^{-\rho(x)} d\alpha, \gamma^2(x) \int_0^1 \Phi^2(\alpha) d\alpha \right).$$

Our asymptotic normality result thus holds under generalizations of the common hypotheses on the standard univariate model, provided the conditional distributions of Y at two neighboring points are sufficiently close. We close this section by pointing out that our main results are also similar in spirit to results obtained in the literature for other conditional tail index or conditional extreme-value index estimators, see e.g. Gardes and Stupfler [22] and Stupfler [37, 38].

4. SIMULATION STUDY

4.1. Hyperparameters selection

The aim of this paragraph is to propose a selection procedure of the hyperparameters involved in the estimator $\check{Q}_n(\beta_n|x)$ of the extreme conditional quantile and in the estimator $\hat{\gamma}_n(x, u)$ of the functional tail index. Assuming that the measure μ used in (2.4) is such that $\text{supp}(\mu) \subset [\tau u, u]$ for some $\tau \in (0, 1)$ fixed by the user (a discussion of the performance of the estimator as a function of τ is included in Section 4.2 below), these hyperparameters are: the bandwidth h controlling the smoothness of the estimators and the value $u \in (0, 1)$ which selects the part of the tail distribution considered in the estimation procedure. The criterion used in our selection procedure is based on the following remark: for any positive and integrable weight function $W : [0, 1] \rightarrow [0, \infty)$,

$$E_W := \mathbb{E} \left[\int_0^1 W(\alpha) (\mathbb{I}\{Y > Q(\alpha|X)\} - \alpha)^2 d\alpha \right] = \int_0^1 W(\alpha) \alpha(1 - \alpha) d\alpha.$$

The sample analogue of E_W is given by

$$\frac{1}{n} \sum_{i=1}^n \int_0^1 W(\alpha) (\mathbb{I}\{Y_i > Q(\alpha|X_i)\} - \alpha)^2 d\alpha,$$

and for a good choice of h and u , this quantity should of course be close to the known quantity E_W . Let then $W_n^{(1)}$ and $W_n^{(2)}$ be two positive and integrable weight functions. Replacing the unobserved variable $Q(\alpha|X_i)$ by the statistic $\hat{Q}_{n,i}(\alpha|X_i)$ which is the estimator (2.2) computed without the observation (X_i, Y_i) leads to the following estimator of $E_{W_n^{(1)}}$:

$$\hat{E}_{W_n^{(1)}}^{(1)}(h) := \frac{1}{n} \sum_{i=1}^n \int_0^1 W_n^{(1)}(\alpha) \left(\mathbb{I}\{Y_i > \hat{Q}_{n,i}(\alpha|X_i)\} - \alpha \right)^2 d\alpha.$$

Note that $\hat{E}_{W_n^{(1)}}^{(1)}(h)$ only depends on the hyperparameter h . In the same way, one can also replace $Q(\alpha|X_i)$ by the statistic $\check{Q}_{n,i}(\alpha|X_i)$ which is the estimator (2.4) computed without the observation (X_i, Y_i) . An estimator of $E_{W_n^{(2)}}$ is then given by:

$$\hat{E}_{W_n^{(2)}}^{(2)}(u, h) := \frac{1}{n} \sum_{i=1}^n \int_0^1 W_n^{(2)}(\alpha) \left(\mathbb{I}\{Y_i > \check{Q}_{n,i}(\alpha|X_i)\} - \alpha \right)^2 d\alpha.$$

Obviously, this last quantity depends both on u and h . We propose the following two-stage procedure to choose the hyperparameters u and h . First, we compute

our selected bandwidth h^{opt} by minimizing with respect to h the function

$$\text{CV}^{(1)}(h) := \left[\widehat{E}_{W_n^{(1)}}^{(1)}(h) - \int_0^1 W_n^{(1)}(\alpha)\alpha(1-\alpha)d\alpha \right]^2.$$

Next, our selected sample fraction u^{opt} is obtained by minimizing with respect to u the function $\text{CV}^{(2)}(u, h^{\text{opt}})$ where

$$\text{CV}^{(2)}(u, h) := \left[\widehat{E}_{W_n^{(2)}}^{(2)}(u, h) - \int_0^1 W_n^{(2)}(\alpha)\alpha(1-\alpha)d\alpha \right]^2.$$

Note that the functions $\text{CV}^{(1)}$ and $\text{CV}^{(2)}$ can be seen as adaptations to the problem of conditional extreme quantile estimation of the cross-validation function introduced in Li *et al.* [32].

4.2. Results

The behavior of the extreme conditional quantile estimator (2.4), when the estimator (2.5) of the functional tail index is used together with our selection procedure of the hyperparameters, is tested on some random pairs $(X, Y) \in \mathcal{C}^1[-1, 1] \times (0, \infty)$, where $\mathcal{C}^1[-1, 1]$ is the space of continuously differentiable functions on $[-1, 1]$. We generate $n = 1000$ independent copies $(X_1, Y_1), \dots, (X_n, Y_n)$ of (X, Y) where X is the random curve defined for all $t \in [-1, 1]$ by $X(t) := \sin[2\pi tU] + (V + 2\pi)t + W$, where U, V and W are independent random variables drawn from a standard uniform distribution. Note that this random covariate was used for instance in Ferraty *et al.* [16]. Regarding the conditional distribution of Y given $X = x$, $x \in \mathcal{C}^1[-1, 1]$, two distributions are considered. The first one is the Fréchet distribution, for which the conditional quantile is given for all $\alpha \in (0, 1)$ by $Q(\alpha|x) = [-\log(1-\alpha)]^{-\gamma(x)}$. The second one is the Burr distribution with parameter $r > 0$, for which $Q(\alpha|x) = (\alpha^{-r\gamma(x)} - 1)^{1/r}$. For these distributions, letting x' be the first derivative of x and

$$z(x) = \frac{2}{3} \left[\int_{-1}^1 x'(t)[1 - \cos(\pi t)]dt - \frac{23}{2} \right],$$

the functional tail index is given by

$$\gamma(x) = \exp \left[-\frac{\log(3)}{9} z^2(x) \right] \mathbb{I}\{|z(x)| < 3\} + \frac{1}{3} \mathbb{I}\{|z(x)| \geq 3\}.$$

In this setup, it is straightforward to show that $z(x) \in [-3.14, 3.07]$ approximately, and therefore the range of values of $\gamma(x)$ is the full interval $[1/3, 1]$. Let us also mention that the second order parameter $\rho(x)$ appearing in condition (H_{SO}) is then $\rho(x) = -1$ for the Fréchet distribution and $\rho(x) = -r\gamma(x)$ for the Burr distribution; in the latter case, the range of values of $\rho(x)$ is therefore $[-r, -r/3]$.

The space $\mathcal{C}^1[-1, 1]$ is endowed with the semi-metric d given for all x_1, x_2 by

$$d(x_1, x_2) = \left[\int_{-1}^1 (x_1'(t) - x_2'(t))^2 dt \right]^{1/2},$$

i.e. the L^2 -distance between first derivatives. To compute $\hat{\gamma}_n(x, u)$, we use the weight function $\Psi(\cdot, u) = u^{-1}(\log(u/\cdot) - 1)$, and the measure μ used in the integrated conditional quantile estimator is assumed to be absolutely continuous with respect to the Lebesgue measure, with density

$$p_{\tau, u}(\alpha) = \frac{1}{u(1-\tau)} \mathbb{I}\{\alpha \in [\tau u, u]\}.$$

In what follows, this estimator is referred to as the Integrated Weissman Estimator (IWE). Other absolutely continuous measures μ , with different densities with respect to the Lebesgue measure, have been tested, with different values of τ . It appears that the impact of the choice of the parameter τ is more important than the one of the measure μ . We thus decided to present in this simulation study the results for the aforementioned value of the measure μ only, but with different tested values for τ .

The hyperparameters are selected using the procedure described in Section 4.1. Since we are interested in the tail of the conditional distribution, the supports of the weight functions $W_n^{(1)}$ and $W_n^{(2)}$ should be located around 0. More specifically, for $i \in \{1, 2\}$, we take

$$W_n^{(i)}(\alpha) := \log \left(\frac{\alpha}{\beta_{n,1}^{(i)}} \right) \mathbb{I}\{\alpha \in [\beta_{n,1}^{(i)}, \beta_{n,2}^{(i)}]\},$$

where $\beta_{n,1}^{(1)} = \lfloor 2\sqrt{n} \log n \rfloor / n$, $\beta_{n,2}^{(1)} = \lfloor 3\sqrt{n} \log n \rfloor / n$, $\beta_{n,1}^{(2)} = \lfloor 5 \log n \rfloor / n$ and $\beta_{n,2}^{(2)} = \lfloor 10 \log n \rfloor / n$. The cross-validation function $\text{CV}^{(1)}(h)$ is minimized over a grid \mathcal{H} of 20 points evenly spaced between 1/2 and 10 to obtain the optimal value h^{opt} , while the value u^{opt} is obtained by minimizing over a grid \mathcal{U} of 26 points evenly spaced between 0.005 and 0.255 the function $\text{CV}^{(2)}(u, h^{\text{opt}})$.

For the Fréchet distribution and two Burr distributions (one with $r = 2$ and one with $r = 1/20$), the conditional extreme quantile estimator (2.4) is computed with the values u^{opt} and h^{opt} obtained by our selection procedure. The quality of the estimator is measured by the Integrated Squared Error given by:

$$ISE := \frac{1}{n} \sum_{i=1}^n \int_{\beta_{n,1}^{(2)}}^{\beta_{n,2}^{(2)}} \log^2 \frac{\check{Q}_{n,i}(\alpha | X_i)}{Q(\alpha | X_i)} d\alpha.$$

This procedure is repeated $N = 100$ times. To give a graphical idea of the behavior of our estimator (2.4), we first depict, in Figure 1, boxplots of the N obtained replications of this estimator, computed with $\tau = 9/10$, for the Fréchet distribution and for some values of the quantile order β_n and of the covariate x .

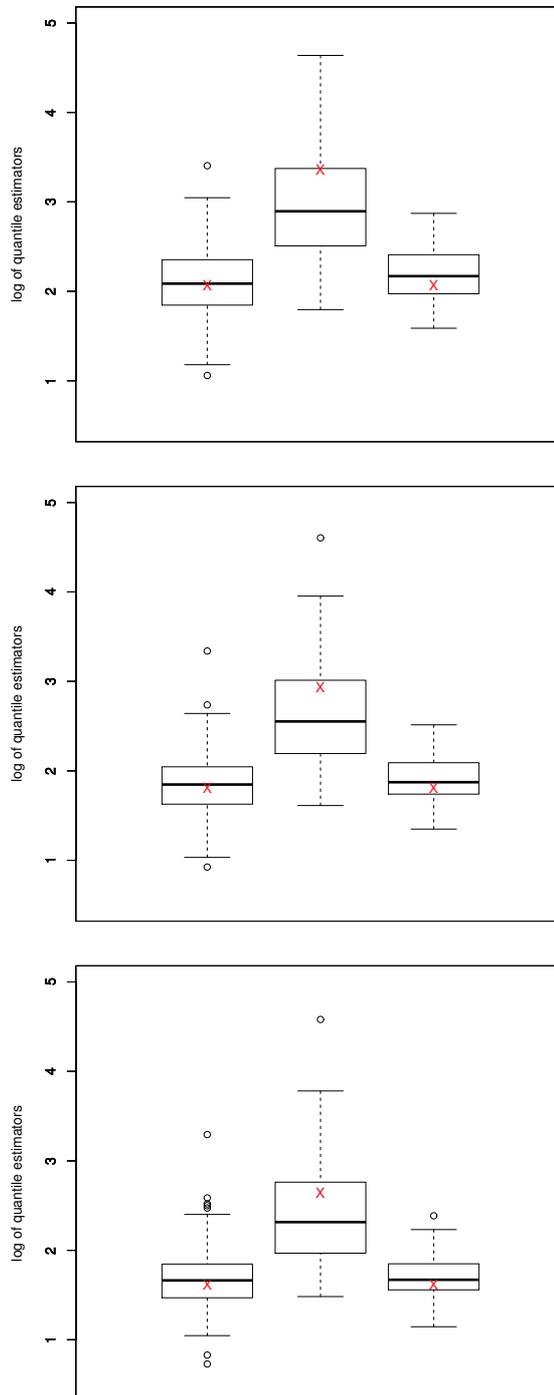


Figure 1: For the Fréchet distribution, boxplots of (the logarithm of) estimator (2.4) for $\beta_n = \beta_{n,1}^{(2)}$ (top), $\beta_n = (\beta_{n,1}^{(2)} + \beta_{n,2}^{(2)})/2$ (middle) and $\beta_n = \beta_{n,2}^{(2)}$ (bottom). In each picture, the covariate x is respectively (from left to right) such that $z(x) = -2$ ($\gamma(x) \approx 0.64$), $z(x) = 0$ ($\gamma(x) = 1$) and $z(x) = 2$ ($\gamma(x) \approx 0.64$). In each case, the true value of the conditional quantile to be estimated is represented by a cross.

More precisely, we take here $\beta_n \in \{\beta_{n,1}^{(2)}, (\beta_{n,1}^{(2)} + \beta_{n,2}^{(2)})/2, \beta_{n,2}^{(2)}\}$ and three values of the covariate are considered: $x = x_1$ with $z(x_1) = -2$ (and then $\gamma(x_1) \approx 0.64$), $x = x_2$ with $z(x_2) = 0$ (giving $\gamma(x_2) = 1$) and $x = x_3$ with $z(x_3) = 2$ (which entails $\gamma(x_3) \approx 0.64$). As expected, the quality of the estimation is strongly impacted by the quantile order β_n but also by the actual position of the covariate and, of course, by the value of the true conditional tail index $\gamma(x)$.

Next, the median and the first and third quartiles of the N values of the Integrated Squared Error are gathered in Table 1. The proposed estimator is compared to the adaptation of the Weissman estimator obtained by taking for the measure μ in (2.4) the Dirac measure at u . This estimator is referred to as the Weissman Estimator (WE) in Table 1. In the WE estimator, the functional tail index $\gamma(x)$ is estimated either by (2.6) or by the generalized Hill-type estimator of Gardes and Girard [21]: for $J \geq 2$, this estimator is given by

$$\widehat{\gamma}^{\text{GG}}(x, u) = \frac{\sum_{j=1}^J (\log \widehat{Q}_n(u/j^2|x) - \log \widehat{Q}_n(u|x))}{\sum_{j=1}^J \log(j^2)}.$$

Following their advice, we set $J = 10$. Again, the median and the first and third quartiles of the N values of the Integrated Squared Error of these two estimators are given in Table 1. In this Table, optimal median errors among the five tested estimators are marked in boldface characters. It appears that the IWEs outperform the two WEs in the case of the Fréchet and Burr (with $r = 1/20$) distributions.

Table 1: Comparison of the Integrated Squared Errors of the following extreme conditional quantile estimators: IWE with $\tau \in \{1/10, 1/2, 9/10\}$ (lines 1 to 3), WE when $\gamma(x)$ is estimated by (2.6) (line 4) and WE when $\gamma(x)$ is estimated by the Hill-type estimator (line 5). Results are given in the following form: [first quartile median third quartile].

	Fréchet dist.	Burr dist. ($r = 2$)
IWE ($\tau = 1/10$)	[0.0060 0.0077 0.0132]	[0.0063 0.0099 0.0147]
IWE ($\tau = 1/2$)	[0.0060 0.0077 0.0112]	[0.0058 0.0095 0.0128]
IWE ($\tau = 9/10$)	[0.0058 0.0076 0.0107]	[0.0059 0.0093 0.0119]
WE (with (2.6))	[0.0054 0.0078 0.0115]	[0.0054 0.0088 0.0137]
WE (Hill-type)	[0.0068 0.0094 0.0120]	[0.0071 0.0103 0.0137]
	Burr dist. ($r = 1/20$)	
IWE ($\tau = 1/10$)	[0.6427 0.9504 1.3982]	
IWE ($\tau = 1/2$)	[0.6040 0.8343 1.2018]	
IWE ($\tau = 9/10$)	[0.8010 1.0870 1.2725]	
WE (with (2.6))	[0.5848 0.8909 1.3372]	
WE (Hill-type)	[0.7679 1.1314 1.4599]	

It also seems that the choice of τ has some influence on the quality of the estimator but, unfortunately, an optimal choice of τ apparently depends on the unknown underlying distribution. It is interesting though to note that the optimal IWE estimator among the three tested here always enjoys a smaller variability than the WE estimator: for instance, in the case of the Burr distribution with $r = 2$, even though the IWE with $\tau = 9/10$ does not outperform the WE (with $\gamma(x)$ estimated by (2.6)) in terms of median ISE, the interquartile range of the ISE is 27.7% lower for the IWE compared to what it is for the WE. Finally, as expected, the value of $\rho(x)$ has a strong impact on the estimation procedure: a value of $\rho(x)$ close to 0 leads to large values of the Integrated Squared Error.

5. REAL DATA EXAMPLE

In this section, we showcase our extreme quantile Integrated Weissman Estimator on functional chemometric data. This data, obtained by considering $n = 215$ pieces of finely chopped meat, consists of pairs of observations (x_n, z_n) , where x_i is the absorbance curve of the i th piece of meat, obtained at 100 regularly spaced wavelengths between 850 and 1050 nanometers (this is also called the spectrometric curve), and z_i is the percentage of fat content in this piece of meat. The data, openly available at <http://lib.stat.cmu.edu/datasets/teacator>, is for instance considered in Ferraty and Vieu [14, 15]. Figure 2 is a graph of all 215 absorbance curves.

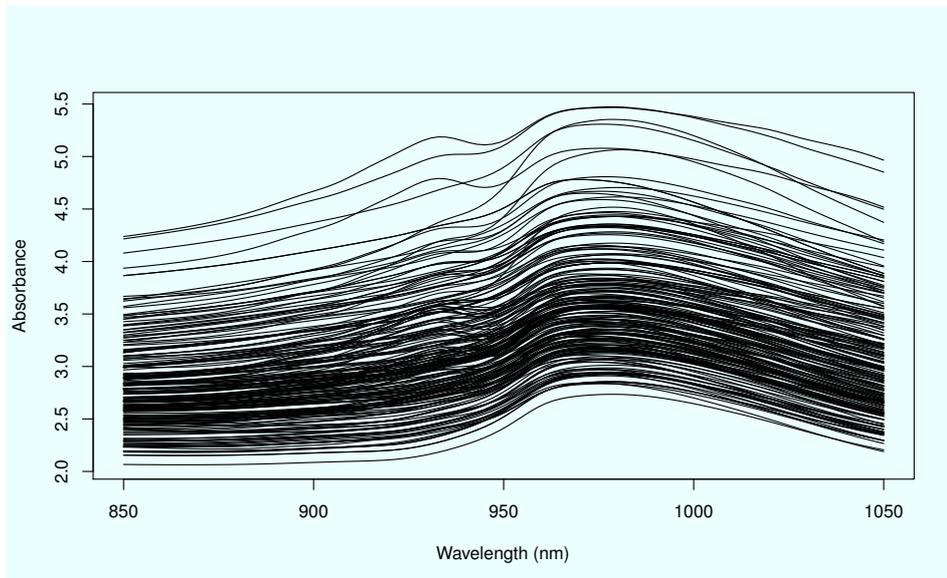


Figure 2: Spectrometric curves for the data.

Because the percentage of fat content z_i obviously belongs to $[0, 100]$, it has a finite-right endpoint and therefore cannot be conditionally heavy-tailed as

required by model (2.1). We thus consider the “inverse fat content” $y_i = 100/z_i$ in this analysis. The top panel of Figure 3 shows the Hill plot of the sample (y_1, \dots, y_n) without integrating covariate information. It can be seen in this figure that the Hill plot seems to be stabilizing near the value 0.4 for a sizeable portion of the left of the graph, thus indicating the plausible presence of a heavy right tail in the data (y_1, \dots, y_n) , see for instance Theorem 3.2.4 in de Haan and Ferreira [27]. The other panels in Figure 3 show exponential QQ-plots for the log-data points whose covariates lie in a fixed-size neighborhood of certain pre-specified points in the covariate space. It is seen in these subfigures that these plots are indeed roughly linear towards their right ends, which supports our conditional heavy tails assumption.

On these grounds, we therefore would like to analyze the influence of the covariate information, which is the absorbance curve, upon the inverse fat content. While of course the absorbance curves obtained are in reality made of discrete data because of the discretization of this curve, the precision of this discretization arguably makes it possible to consider our data as in fact functional. This, in our opinion, fully warrants the use of our estimator in this case.

Because the covariate space is functional, one has to wonder about how to measure the influence of the covariate and then about how to represent the results. A nice account of the problem of how to represent results when considering functional data is given in Ferraty and Vieu [15]. Here, we look at the variation of extreme quantile estimates in two different directions of the covariate space. To this end, we consider the semi-metric

$$d(x_1, x_2) = \left[\int_{850}^{1050} (x_1''(t) - x_2''(t))^2 dt \right]^{1/2},$$

also advised by Ferraty and Vieu [14], and we compute:

- A typical pair of covariates, *i.e.* a pair $(x_1^{\text{med}}, x_2^{\text{med}})$ such that

$$d(x_1^{\text{med}}, x_2^{\text{med}}) = \text{median}\{d(x_i, x_j), 1 \leq i, j \leq n, i \neq j\};$$

- A pair of covariates farthest from each other, *i.e.* a pair $(x_1^{\text{max}}, x_2^{\text{max}})$ such that

$$d(x_1^{\text{max}}, x_2^{\text{max}}) = \max\{d(x_i, x_j), 1 \leq i, j \leq n, i \neq j\}.$$

For the purpose of comparison, we also compute the “average covariate” $\bar{x} = n^{-1} \sum_{i=1}^n x_i$. In particular, we represent on Figure 4 our two pairs of covariates together with the average covariate, the same scale being used on the y -axis in both figures. Recall that since the semi-metric d is the L^2 -distance between second-order derivatives, it acts as a measure of how much the shapes of two covariate curves are different, rather than measuring how far apart they are.

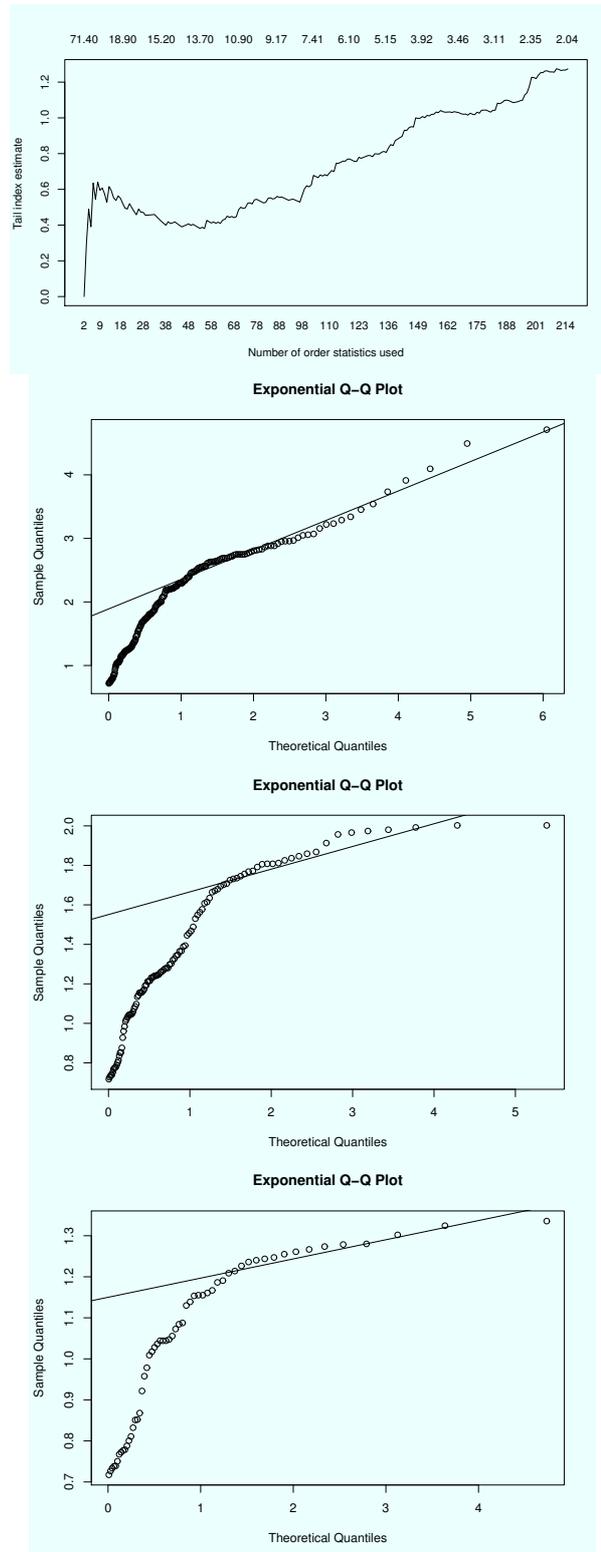


Figure 3: Top panel: Hill plot for the sample (y_1, \dots, y_n) . On the x -axis at the top of the panel is the value of the lower threshold for the computation of the Hill estimator, *i.e.* the lowest order statistic. Other panels: local exponential QQ-plots for the log-data points whose covariates belong to a neighborhood of certain pre-specified points in the covariate space.

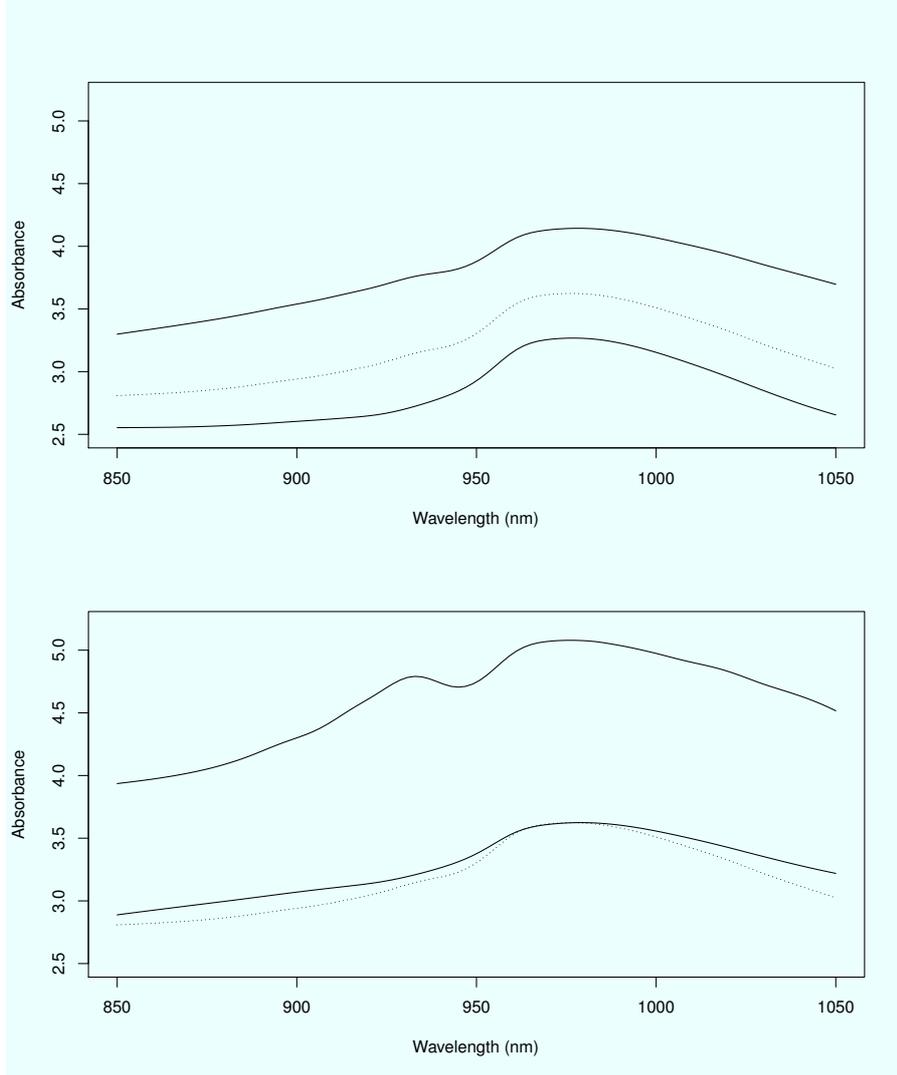


Figure 4: Top picture, solid lines: a pair of typical covariates. Bottom picture, solid lines: the pair of covariates farthest from each other. In both pictures the dotted line is the average covariate.

We compute our conditional extreme quantile estimator at the levels $5/n$ and $1/n$, using the methodology given in Section 4.2. In particular, the selection parameters $\beta_{n,1}^{(1)}$, $\beta_{n,2}^{(1)}$, $\beta_{n,1}^{(2)}$ and $\beta_{n,2}^{(2)}$ used in the cross-validation methodology were the exact same ones used in the simulation study, namely 0.437, 0.655, 0.035 and 0.069, respectively. The bandwidth h is selected in the interval $[0.00316, 0.0116]$, the lower bound in this interval corresponding to the median of all distances $d(x_i, x_j)$ ($i \neq j$) and the upper bound corresponding to 90% of the maximum of all distances $d(x_i, x_j)$, for a final selected value of 0.00717. The value of the parameter u is selected exactly as in the simulation study, and the selection procedure gives the value 0.185. Finally, we set $\tau = 0.9$ in our Integrated Weissman Estimator.

Results are given in Figure 5; namely, we compute the extreme quantile estimates $\hat{Q}_n(\beta|x)$, for $\beta \in \{5/n, 1/n\}$, and x belonging to either the line $[x_1^{\text{med}}, x_2^{\text{med}}] = \{tx_1^{\text{med}} + (1-t)x_2^{\text{med}}, t \in [0, 1]\}$ or to the line $[x_1^{\text{max}}, x_2^{\text{max}}]$. It can be seen in these figures that the estimates in the direction of a typical pair of covariates are remarkably stable; they are actually essentially indistinguishable from the estimates at the average covariate, which are 42.41 for $\beta = 5/n$ and 93.86 for $\beta = 1/n$.

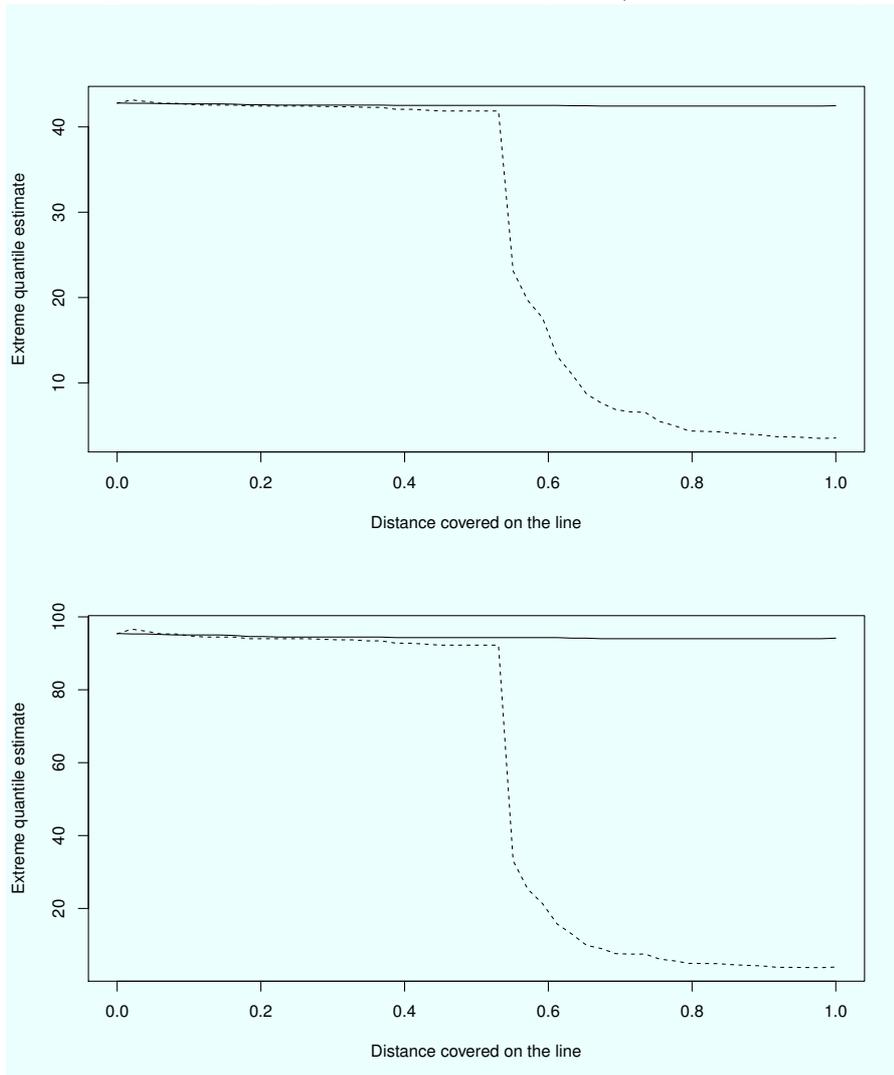


Figure 5: Solid line: extreme quantile estimate in the direction of a typical pair of covariates, dashed line: extreme quantile estimate in the direction of a pair of covariates farthest from each other. Top picture: case $\beta = 5/n$, bottom picture: $\beta = 1/n$.

By contrast, the estimates on the line $[x_1^{\text{max}}, x_2^{\text{max}}]$, while roughly stable for 60% of the line and approximately equal to the value of the estimated quantiles at the average covariate, very sharply drop afterwards, the reduction factor being close to 10 from the beginning of the line to its end in the case $\beta = 5/n$.

This conclusion suggests that while in typical directions of the covariate space the tail behavior of the fat content is very stable, there may be certain directions in which this is not the case. In particular, there appear to be certain values of the covariate for which thresholds for the detection of unusual levels of fat should differ from those of more standard cases.

6. PROOFS OF THE MAIN RESULTS

Before proving the main results, we recall two useful facts. The first one is a classical equivalent of

$$M_x(h) := \sum_{i=1}^n \mathbb{I}\{d(X_i, x) \leq h\}.$$

If $m_x(h) \rightarrow \infty$ as $n \rightarrow \infty$ then, for any $\delta \in (0, 1)$:

$$(6.1) \quad [m_x(h)]^{(1-\delta)/2} \left| \frac{M_x(h)}{m_x(h)} - 1 \right| \xrightarrow{\mathbb{P}} 0 \quad \text{as } n \rightarrow \infty,$$

see Lemma 1 in Stupfler [37]. For the second one, let $\{Y_i^*, i = 1, \dots, M_x(h)\}$ be the response variables whose associated covariates $\{X_i^*, i = 1, \dots, M_x(h)\}$ are such that $d(X_i^*, x) \leq h$. Lemma 4 in Gardes and Stupfler [22] shows that the random variables $V_i = 1 - F(Y_i^* | X_i^*)$ are such that, for all $u_1, \dots, u_p \in [0, 1]$,

$$(6.2) \quad \mathbb{P} \left(\bigcap_{i=1}^p \{V_i \leq u_i\} | M_x(h) = p \right) = u_1 \dots u_p,$$

i.e. they are independent standard uniform random variables given $M_x(h)$.

6.1. Proof of Theorem 3.1

The following proposition is a uniform consistency result for the estimator $\widehat{Q}_n(\beta_n | x)$ when β_n goes to 0 at a moderate rate.

Proposition 6.1. *Assume that conditions (2.3), (H_{SO}) , (3.2) and (3.3) are satisfied. If $m_x(h) \rightarrow \infty$, then*

$$\sup_{\alpha \in [\tau u_{n,x}, u_{n,x}]} \left| \frac{\widehat{Q}_n(\alpha | x)}{Q(\alpha | x)} - 1 \right| = \mathcal{O}_{\mathbb{P}} \left([m_x(h)]^{(a(x)-1)/2} \right).$$

Proof: Let $M_n := M_x(h)$, $\{U_i, i \geq 1\}$ be independent standard uniform random variables, $V_i := S(Y_i^*|X_i^*)$ and

$$Z_n(x) := \sup_{\alpha \in [\tau u_{n,x}, u_{n,x}]} \left| \frac{\widehat{Q}_n(\alpha|x)}{Q(\alpha|x)} - 1 \right|.$$

We start with the following inequality: $Z_n(x) \leq T_n(x) + R_n^{(Q)}(x)$, with

$$(6.3) \quad T_n(x) := \sup_{\alpha \in [\tau u_{n,x}, u_{n,x}]} \left| \frac{Q(V_{\lfloor \alpha M_n \rfloor + 1, M_n} | x)}{Q(\alpha|x)} - 1 \right|$$

$$(6.4) \quad \text{and } R_n^{(Q)}(x) := \sup_{\alpha \in [\tau u_{n,x}, u_{n,x}]} \left| \frac{\widehat{Q}_n(\alpha|x) - Q(V_{\lfloor \alpha M_n \rfloor + 1, M_n} | x)}{Q(\alpha|x)} \right|.$$

Let us first focus on the term $T_n(x)$. For any $t > 0$,

$$\mathbb{P}(v_{n,x} T_n(x) > t) = \sum_{j=0}^n \mathbb{P}(v_{n,x} T_n(x) > t | M_n = j) \mathbb{P}(M_n = j),$$

where $v_{n,x} := [m_x(h)]^{(1-a(x))/2}$. From (6.1), letting

$$(6.5) \quad I_n := [m_x(h)(1 - [m_x(h)]^{[a(x)/4]-1/2}), m_x(h)(1 + [m_x(h)]^{[a(x)/4]-1/2})],$$

one has $\mathbb{P}(M_n \notin I_n) \rightarrow 0$ as $n \rightarrow \infty$. Hence,

$$\mathbb{P}(v_{n,x} T_n(x) > t) \leq \sup_{p \in I_n} \mathbb{P}(v_{n,x} T_n(x) > t | M_n = p) + o(1).$$

Using Lemma A.1,

$$\sup_{p \in I_n} \mathbb{P}(v_{n,x} T_n(x) > t | M_n = p) = \sup_{p \in I_n} \mathbb{P}(v_{n,x} \underline{T}_p(x) > t),$$

where

$$\underline{T}_p(x) := \sup_{\alpha \in [\tau u_{n,x}, u_{n,x}]} \left| \frac{Q(U_{\lfloor p\alpha \rfloor + 1, p} | x)}{Q(\alpha|x)} - 1 \right|.$$

Using condition (3.2), it is clear that there are constants $d_1, d_2 > 0$ with $d_1 < d_2$ such that for n large enough, we have for all $p \in I_n$:

$$\underline{T}_p(x) \leq \sup_{\alpha \in [d_1 p^{-a(x)}, d_2 p^{-a(x)}]} \left| \frac{Q(U_{\lfloor p\alpha \rfloor + 1, p} | x)}{Q(\alpha|x)} - 1 \right|.$$

Thus, for all $t > 0$, $\mathbb{P}(v_{n,x} T_n(x) > t)$ is bounded above by

$$\sup_{p \in I_n} \mathbb{P} \left(v_{n,x} \sup_{\alpha \in [d_1 p^{-a(x)}, d_2 p^{-a(x)}]} \left| \frac{Q(U_{\lfloor p\alpha \rfloor + 1, p} | x)}{Q(\alpha|x)} - 1 \right| > t \right) + o(1).$$

Furthermore, for n large enough, there exists $\kappa > 0$ such that for all $p \in I_n$, $v_{n,x} \leq \kappa p^{(1-a(x))/2}$ and thus, for all $t > 0$, $\mathbb{P}(v_{n,x} T_n(x) > t)$ is bounded above by

$$\sup_{p \in I_n} \mathbb{P} \left(\kappa p^{(1-a(x))/2} \sup_{\alpha \in [d_1 p^{-a(x)}, d_2 p^{-a(x)}]} \left| \frac{Q(U_{\lfloor p\alpha \rfloor + 1, p} | x)}{Q(\alpha|x)} - 1 \right| > t \right) + o(1).$$

Since

$$p^{(1-a(x))/2} \sup_{\alpha \in [d_1 p^{-a(x)}, d_2 p^{-a(x)}]} \left| \frac{Q(U_{\lfloor p\alpha \rfloor + 1, p} | x)}{Q(\alpha | x)} - 1 \right| = \mathcal{O}_{\mathbb{P}}(1),$$

(see Lemma A.2 for a proof), it now becomes clear that $T_n(x) = \mathcal{O}_{\mathbb{P}}(v_{n,x}^{-1})$.

Let us now focus on the term $R_n^{(Q)}(x)$. As before, one can show that for all $t > 0$,

$$\mathbb{P}(v_{n,x} R_n^{(Q)}(x) > t) \leq \sup_{p \in I_n} \mathbb{P}(v_{n,x} R_n^{(Q)}(x) > t | M_n = p) + o(1).$$

Lemma A.1 and condition (3.3) yield for any $t > 0$ and n large enough:

$$\begin{aligned} & \sup_{p \in I_n} \mathbb{P}(v_{n,x} R_n^{(Q)}(x) > t | M_n = p) \\ & \leq \sup_{p \in I_n} \mathbb{P}(v_{n,x} \omega(U_{1,p}, U_{p,p}, x, h) \exp(\omega(U_{1,p}, U_{p,p}, x, h)) (1 + \underline{T}_p(x)) > t) \\ & \leq \sup_{p \in I_n} \mathbb{P}(p^{(1-a(x))/2} \omega(U_{1,p}, U_{p,p}, x, h) \exp(\omega(U_{1,p}, U_{p,p}, x, h)) (1 + \underline{T}_p(x)) > t/\kappa) \\ & \leq \sup_{p \in I_n} \left[\mathbb{P}(U_{1,p} < [m_x(h)]^{-1-\delta}) + \mathbb{P}(U_{p,p} > 1 - [m_x(h)]^{-1-\delta}) \right]. \end{aligned}$$

Since for n large enough

$$\begin{aligned} (6.6) \quad & \sup_{p \in I_n} \left[\mathbb{P}(U_{1,p} < [m_x(h)]^{-1-\delta}) + \mathbb{P}(U_{p,p} > 1 - [m_x(h)]^{-1-\delta}) \right] \\ & = 2 \sup_{p \in I_n} \left[1 - [1 - [m_x(h)]^{-1-\delta}]^p \right] \leq 2 \left(1 - [1 - [m_x(h)]^{-1-\delta}]^{2m_x(h)} \right) \rightarrow 0, \end{aligned}$$

we thus have proven that $R_n^{(Q)}(x) = o_{\mathbb{P}}(v_{n,x}^{-1})$ and the proof is complete. \square

Proof of Theorem 3.1: The key point is to write

$$\tilde{Q}_n(\beta_n | x) = \int_{\tau u_{n,x}}^{u_{n,x}} Q(\alpha | x) \left(\frac{\alpha}{\beta_n} \right)^{\gamma(x)} \left\{ \frac{\hat{Q}_n(\alpha | x)}{Q(\alpha | x)} \left(\frac{\alpha}{\beta_n} \right)^{\hat{\gamma}_n(x) - \gamma(x)} \right\} \mu(d\alpha).$$

Now, by assumption $v_{n,x}(\hat{\gamma}_n(x) - \gamma(x)) \xrightarrow{d} \Gamma$ where $v_{n,x} := [m_x(h)]^{(1-a(x))/2}$. Since $\beta_n/u_{n,x}$ is asymptotically bounded from below and above by sequences proportional to $\beta_n [m_x(h)]^{a(x)} \rightarrow 0$, one has for n large enough that

$$\sup_{\alpha \in [\tau u_{n,x}, u_{n,x}]} \left| \log \left[\left(\frac{\alpha}{\beta_n} \right)^{\hat{\gamma}_n(x) - \gamma(x)} \right] \right| \leq |\hat{\gamma}_n(x) - \gamma(x)| \log \left(\frac{u_{n,x}}{\beta_n} \right) = o_{\mathbb{P}}(1),$$

since by assumption $v_{n,x}^{-1} \log(u_{n,x}/\beta_n) \rightarrow 0$. A Taylor expansion for the exponential function thus yields

$$\left(\frac{\alpha}{\beta_n} \right)^{\hat{\gamma}_n(x) - \gamma(x)} - 1 - \log(\alpha/\beta_n)(\hat{\gamma}_n(x) - \gamma(x)) = \mathcal{O}_{\mathbb{P}}(v_{n,x}^{-1} \log^2(u_{n,x}/\beta_n)),$$

uniformly in $\alpha \in [\tau u_{n,x}, u_{n,x}]$. We then obtain

$$\check{Q}_n(\beta_n|x) = \int_{\tau u_{n,x}}^{u_{n,x}} Q(\alpha|x) \left(\frac{\alpha}{\beta_n}\right)^{\gamma(x)} G_{n,x}(\alpha) \mu(d\alpha)$$

where

$$G_{n,x}(\alpha) := \frac{\widehat{Q}_n(\alpha|x)}{Q(\alpha|x)} [1 + \log(\alpha/\beta_n)(\widehat{\gamma}_n(x) - \gamma(x)) + \mathcal{O}_{\mathbb{P}}(v_{n,x}^{-1} \log^2(u_{n,x}/\beta_n))].$$

By Proposition 6.1,

$$\sup_{\alpha \in [\tau u_{n,x}, u_{n,x}]} \left| \frac{\widehat{Q}_n(\alpha|x)}{Q(\alpha|x)} - 1 \right| = \mathcal{O}_{\mathbb{P}}(v_{n,x}^{-1}),$$

and therefore:

$$(6.7) \quad G_{n,x}(\alpha) = 1 + \log(\alpha/\beta_n)(\widehat{\gamma}_n(x) - \gamma(x)) + \mathcal{O}_{\mathbb{P}}(v_{n,x}^{-1} \log^2(u_{n,x}/\beta_n)).$$

By Lemma A.3,

$$(6.8) \quad \sup_{\alpha \in [\tau u_{n,x}, u_{n,x}]} \left| \frac{Q(\alpha|x)}{Q(\beta_n|x)} \left(\frac{\alpha}{\beta_n}\right)^{\gamma(x)} - 1 \right| = \mathcal{O}(\Delta(u_{n,x}^{-1}|x)),$$

and thus, (6.7) and (6.8) lead to

$$\begin{aligned} \frac{\check{Q}(\beta_n|x)}{Q(\beta_n|x)} - 1 &= (\widehat{\gamma}_n(x) - \gamma(x)) \int_{\tau u_{n,x}}^{u_{n,x}} \log(\alpha/\beta_n) \mu(d\alpha) [1 + \mathcal{O}(\Delta(u_{n,x}^{-1}|x))] \\ &\quad + \mathcal{O}(\Delta(u_{n,x}^{-1}|x)) + \mathcal{O}_{\mathbb{P}}(v_{n,x}^{-1} \log^2(u_{n,x}/\beta_n)). \end{aligned}$$

Since $u_{n,x}/\beta_n \rightarrow 0$ and $\mu([\tau u_{n,x}, u_{n,x}]) = 1$, one has

$$\begin{aligned} \int_{\tau u_{n,x}}^{u_{n,x}} \log(\alpha/\beta_n) \mu(d\alpha) &= \int_{\tau u_{n,x}}^{u_{n,x}} [\log(u_{n,x}/\beta_n) + \log(\alpha/u_{n,x})] \mu(d\alpha) \\ &= \log(u_{n,x}/\beta_n)(1 + o(1)), \end{aligned}$$

and thus

$$\begin{aligned} \frac{\check{Q}(\beta_n|x)}{Q(\beta_n|x)} - 1 &= (\widehat{\gamma}_n(x) - \gamma(x)) \log(u_{n,x}/\beta_n) [1 + o(1)] \\ &\quad + \mathcal{O}(\Delta(u_{n,x}^{-1}|x)) + \mathcal{O}_{\mathbb{P}}(v_{n,x}^{-1} \log^2(u_{n,x}/\beta_n)). \end{aligned}$$

Using the convergence in distribution of $\widehat{\gamma}_n(x)$ completes the proof. \square

6.2. Proof of Theorem 3.2

For the sake of brevity, let $v_{n,x} = [m_x(h)]^{(1-a(x))/2}$, $M_n = M_x(h)$ and $K_n = u_{n,x} M_n$. The cumulative distribution function of a normal distribution with mean $\lambda(x) \int_0^1 \Phi(\alpha) \alpha^{-\rho(x)} d\alpha$ and variance $\gamma^2(x) \int_0^1 \Phi^2(\alpha) d\alpha$ is denoted by H_x in what

follows. Let $t \in \mathbb{R}$ and $\varepsilon > 0$. Denoting by $E_n(t)$ the event $\{v_{n,x}(\widehat{\gamma}_n(x, u_{n,x}) - \gamma(x)) \leq t\}$, one has

$$|\mathbb{P}[E_n(t)] - H_x(t)| \leq \sum_{p=0}^n \mathbb{P}(M_n = p) |\mathbb{P}[E_n(t)|M_n = p] - H_x(t)|.$$

Recall that from (6.1), $\mathbb{P}(M_n \notin I_n) \rightarrow 0$ as $n \rightarrow \infty$ where I_n is defined in (6.5). Hence, for n large enough,

$$(6.9) \quad |\mathbb{P}[E_n(t)] - H_x(t)| \leq \sup_{p \in I_n} |\mathbb{P}[E_n(t)|M_n = p] - H_x(t)| + \frac{\varepsilon}{8}.$$

Now, using the notation $V_i := S(Y_i^* | X_i^*)$ for $i = 1, \dots, M_n$, let us introduce the statistics:

$$(6.10) \quad \widetilde{\gamma}_n(x, u_{n,x}) := \sum_{i=1}^{\lfloor K_n \rfloor} W_{i,n}(u_{n,x}, M_n) \log \frac{Q(V_{i,M_n} | x)}{Q(V_{\lfloor K_n \rfloor + 1, M_n} | x)}$$

and $R_n^{(\gamma)}(x) := \widehat{\gamma}_n(x, u_{n,x}) - \widetilde{\gamma}_n(x, u_{n,x})$,

where

$$(6.11) \quad W_{i,n}(u_{n,x}, M_n) := \int_{(i-1)/M_n}^{i/M_n} \Psi(\alpha, u_{n,x}) d\alpha.$$

It is straightforward that for all $\kappa > 0$,

$$(6.12) \quad \sup_{p \in I_n} |\mathbb{P}[E_n(t)|M_n = p] - H_x(t)| \leq T_{n,x}^{(1)} + T_{n,x}^{(2)},$$

where

$$T_{n,x}^{(1)} := \sup_{p \in I_n} \left| \mathbb{P} \left[E_n(t) \cap \left\{ v_{n,x} | R_n^{(\gamma)}(x) | \leq \kappa \right\} | M_n = p \right] - H_x(t) \right|$$

and $T_{n,x}^{(2)} := \sup_{p \in I_n} \mathbb{P} \left[v_{n,x} | R_n^{(\gamma)}(x) | > \kappa | M_n = p \right]$.

Let us first focus on the term $T_{n,x}^{(1)}$. Let $\widetilde{E}_n(t) := \{v_{n,x}(\widetilde{\gamma}_n(x, u_{n,x}) - \gamma(x)) \leq t\}$. For all $p \in I_n$, $\mathbb{P}[E_n(t) \cap \{v_{n,x} | R_n^{(\gamma)}(x) | \leq \kappa\} | M_n = p] \leq \mathbb{P}[\widetilde{E}_n(t + \kappa) | M_n = p]$ and

$$\begin{aligned} & \mathbb{P} \left[E_n(t) \cap \left\{ v_{n,x} | R_n^{(\gamma)}(x) | \leq \kappa \right\} | M_n = p \right] \\ & \geq \mathbb{P} \left[\widetilde{E}_n(t - \kappa) \cap \left\{ v_{n,x} | R_n^{(\gamma)}(x) | \leq \kappa \right\} | M_n = p \right] \\ & \geq \mathbb{P} \left[\widetilde{E}_n(t - \kappa) | M_n = p \right] - \mathbb{P} \left[v_{n,x} | R_n^{(\gamma)}(x) | > \kappa | M_n = p \right]. \end{aligned}$$

Using the inequality $|x| \leq |a| + |b|$ which holds for all $x \in [a, b]$, it is then clear that for all $\kappa > 0$,

$$\begin{aligned} T_{n,x}^{(1)} & \leq \sup_{p \in I_n} \left| \mathbb{P} \left[\widetilde{E}_n(t + \kappa) | M_n = p \right] - H_x(t + \kappa) \right| \\ & \quad + \sup_{p \in I_n} \left| \mathbb{P} \left[\widetilde{E}_n(t - \kappa) | M_n = p \right] - H_x(t - \kappa) \right| \\ & \quad + |H_x(t) - H_x(t + \kappa)| + |H_x(t) - H_x(t - \kappa)| + T_{n,x}^{(2)}. \end{aligned}$$

Since H_x is continuous, we can actually choose $\kappa > 0$ so small that

$$|H_x(t) - H_x(t + \kappa)| \leq \frac{\varepsilon}{8} \text{ and } |H_x(t) - H_x(t - \kappa)| \leq \frac{\varepsilon}{8}$$

and therefore

$$(6.13) \quad T_{n,x}^{(1)} \leq \sup_{p \in I_n} \left| \mathbb{P} \left[\tilde{E}_n(t + \kappa) | M_n = p \right] - H_x(t + \kappa) \right| \\ + \sup_{p \in I_n} \left| \mathbb{P} \left[\tilde{E}_n(t - \kappa) | M_n = p \right] - H_x(t - \kappa) \right| + T_{n,x}^{(2)} + \frac{\varepsilon}{4}.$$

We now focus on the two first terms in the left-hand side of the previous inequality. From Lemma A.4, the distribution of $\tilde{\gamma}_n(x, u_{n,x})$ given $M_n = p$ is that of

$$\bar{\gamma}_p(x, u_{n,x}) := \frac{1}{pu_{n,x}} \sum_{i=1}^{\lfloor pu_{n,x} \rfloor} \Phi \left(\frac{i}{pu_{n,x}} \right) i \log \frac{Q(U_{i,p}|x)}{Q(U_{i+1,p}|x)}.$$

Hence, for all $s \in \mathbb{R}$ and $p \in I_n$, $\mathbb{P}[\tilde{E}_n(s) | M_n = p] = \mathbb{P}[v_{n,x}(\bar{\gamma}_p(x, u_{n,x}) - \gamma(x)) \leq s]$. Furthermore, for n large enough we have

$$p/2 \leq \frac{p}{1 + [m_x(h)]^{\lfloor a(x)/4 \rfloor - 1/2}} \leq m_x(h) \leq \frac{p}{1 - [m_x(h)]^{\lfloor a(x)/4 \rfloor - 1/2}} \leq 2p$$

for all $p \in I_n$, so that for n large enough:

$$(6.14) \quad \xi^{(+)}(p) \leq m_x(h) \leq \xi^{(-)}(p),$$

with $\xi^{(+)}(p) := p[1 + (2p)^{\lfloor a(x)/4 \rfloor - 1/2}]^{-1}$ and $\xi^{(-)}(p) := p[1 - (p/2)^{\lfloor a(x)/4 \rfloor - 1/2}]^{-1}$. Under our assumptions on the sequence $u_{n,x}$, the previous inequalities lead to $k_1(p) \leq pu_{n,x} \leq k_2(p)$ where $k_1(p) := p[\xi^{(-)}(p)]^{-a(x)}[1 + \varphi_1(\xi^{(-)}(p))]$ and $k_2(p) := p[\xi^{(+)}(p)]^{-a(x)}[1 + \varphi_2(\xi^{(+)}(p))]$. Since Φ is a nonincreasing function on $(0, 1)$, we then get that:

$$\bar{\gamma}_p(x, u_{n,x}) \leq \frac{1}{k_1(p)} \sum_{i=1}^{\lfloor pu_{n,x} \rfloor} \Phi \left(\frac{i}{\lfloor k_2(p) \rfloor + 1} \right) i \log \frac{Q(U_{i,p}|x)}{Q(U_{i+1,p}|x)} \\ \leq \frac{1}{k_1(p)} \sum_{i=1}^{\lfloor k_2(p) \rfloor + 1} \Phi \left(\frac{i}{\lfloor k_2(p) \rfloor + 1} \right) i \log \frac{Q(U_{i,p}|x)}{Q(U_{i+1,p}|x)} \\ = \check{\gamma}_p(x, k_1(p), k_2(p))$$

with

$$(6.15) \quad \check{\gamma}_p(x, k, k') := \frac{1}{k} \sum_{i=1}^{\lfloor k' \rfloor} \Phi \left(\frac{i}{\lfloor k' \rfloor + 1} \right) i \log \frac{Q(U_{i,p}|x)}{Q(U_{i+1,p}|x)}.$$

A similar lower bound applies and thus $\check{\gamma}_p(x, k_2(p), k_1(p) - 1) \leq \bar{\gamma}_p(x, u_{n,x}) \leq \check{\gamma}_p(x, k_1(p), k_2(p))$ for all $p \in I_n$. As a first conclusion, using the inequality $|x| \leq$

$|a| + |b|$ which holds for all $x \in [a, b]$, we have shown that for all $s \in \mathbb{R}$,

$$\begin{aligned} & \sup_{p \in I_n} \left| \mathbb{P} \left[\tilde{E}_n(s) | M_n = p \right] - H_x(s) \right| \\ & \leq \sup_{p \in I_n} \left| \mathbb{P} \left[v_{n,x}(\check{\gamma}_p(x, k_1(p), k_2(p)) - \gamma(x)) \leq s \right] - H_x(s) \right| \\ & \quad + \sup_{p \in I_n} \left| \mathbb{P} \left[v_{n,x}(\check{\gamma}_p(x, k_2(p), k_1(p) - 1) - \gamma(x)) \leq s \right] - H_x(s) \right|. \end{aligned}$$

Since from (6.14), $[\xi^{(+)}(p)]^{(1-a(x))/2} \leq v_{n,x} \leq [\xi^{(-)}(p)]^{(1-a(x))/2}$ for all $p \in I_n$ and since by assumption on the φ_i ,

$$\begin{aligned} \frac{k_1(p)}{k_2(p)} &= 1 + \mathcal{O} \left(p^{[a(x)/4]-1/2} \right) + \mathcal{O} \left(\varphi_1(\xi^{(+)}(p)) \right) + \mathcal{O} \left(\varphi_2(\xi^{(+)}(p)) \right) \\ &= 1 + o(p^{(a(x)-1)/2}), \end{aligned}$$

one can apply Lemmas A.6 and A.7 to show that for n large enough

$$(6.16) \quad \begin{aligned} & \sup_{p \in I_n} \left| \mathbb{P} \left[\tilde{E}_n(t + \kappa) | M_n = p \right] - H_x(t + \kappa) \right| \\ & + \sup_{p \in I_n} \left| \mathbb{P} \left[\tilde{E}_n(t - \kappa) | M_n = p \right] - H_x(t - \kappa) \right| \leq \frac{\varepsilon}{2}. \end{aligned}$$

It remains to study the term $T_{n,x}^{(2)}$. Lemma A.4 entails that

$$T_{n,x}^{(2)} \leq \sup_{p \in I_n} \mathbb{P} \left[2v_{n,x}\omega(U_{1,p}, U_{p,p}, x, h) \int_0^{u_{n,x}} |\Psi(\alpha, u_{n,x})| d\alpha > \kappa \right].$$

From condition (H_Ψ) ,

$$\limsup_{u \downarrow 0} \int_0^u |\Psi(\alpha, u)| d\alpha = C < \infty$$

and thus for n large enough, using (6.6):

$$(6.17) \quad \begin{aligned} T_{n,x}^{(2)} &\leq \sup_{p \in I_n} \mathbb{P} \left[v_{n,x}\omega(U_{1,p}, U_{p,p}, x, h) > \frac{\kappa}{4C} \right] \\ &\leq 2 \left(1 - [1 - [m_x(h)]^{-1-\delta}]^{2m_x(h)} \right) \leq \frac{\varepsilon}{8}. \end{aligned}$$

Collecting (6.9), (6.12), (6.13), (6.16) and (6.17) concludes the proof. \square

APPENDIX

The first lemma is dedicated to the statistics $T_n(x)$ and $R_n^{(Q)}(x)$ defined in the proof of Proposition 6.1, equations (6.3) and (6.4).

Lemma A.1. *Let $\{U_i, i \geq 1\}$ be independent standard uniform random variables. For $x \in \mathcal{E}$ such that $m_x(h) > 0$, the conditional distribution of $T_n(x)$ given $M_x(h) = p$ is that of*

$$\underline{T}_p(x) := \sup_{\alpha \in [\tau u_{n,x}, u_{n,x}]} \left| \frac{Q(U_{\lfloor p\alpha \rfloor + 1, p} | x)}{Q(\alpha | x)} - 1 \right|$$

and, given $M_x(h) = p$, $R_n^{(Q)}(x)$ is bounded from above by

$$\omega(U_{1,p}, U_{p,p}, x, h) \exp[\omega(U_{1,p}, U_{p,p}, x, h)] (1 + \underline{T}_p(x)).$$

Proof: Recall the notation $M_n := M_x(h)$ and $V_i := S(Y_i^* | X_i^*)$. First, given $M_n = p$, equation (6.2) entails that $\{V_i, 1 \leq i \leq M_n\} | \{M_n = p\} \stackrel{d}{=} \{U_i, 1 \leq i \leq p\}$ where U_1, \dots, U_p are independent standard uniform variables. It thus holds that

$$\{Q(V_{\lfloor \alpha M_n \rfloor + 1, M_n} | x), \alpha \in [0, 1]\} | \{M_n = p\} \stackrel{d}{=} \{Q(U_{\lfloor p\alpha \rfloor + 1, p} | x), \alpha \in [0, 1]\}.$$

As a direct consequence

$$(A.1) \quad T_n(x) | \{M_n = p\} \stackrel{d}{=} \underline{T}_p(x).$$

Let us now focus on the term $R_n^{(Q)}(x)$. Since $Q(\cdot | x)$ is continuous and decreasing, one has, for $i = 1, \dots, M_n$,

$$\begin{aligned} \log Q(V_i | x) - \omega(V_{1, M_n}, V_{M_n, M_n}, x, h) &\leq \log Y_i^* = \log Q(V_i | X_i^*) \\ &\leq \log Q(V_i | x) + \omega(V_{1, M_n}, V_{M_n, M_n}, x, h). \end{aligned}$$

It follows from Lemma 1 in Gardes and Stupfler [22] that for all $i \in \{1, \dots, M_n\}$,

$$(A.2) \quad \left| \log Y_{M_n - i + 1, M_n}^* - \log Q(V_{i, M_n} | x) \right| \leq \omega(V_{1, M_n}, V_{M_n, M_n}, x, h).$$

Since $\widehat{Q}_n(\alpha | x) = Y_{M_n - i + 1, M_n}^*$ for all $\alpha \in [(i - 1)/M_n, i/M_n)$, the mean value theorem leads to

$$\begin{aligned} &\sup_{\alpha \in [\tau u_{n,x}, u_{n,x}]} \left| \frac{\widehat{Q}_n(\alpha | x)}{Q(V_{\lfloor \alpha M_n \rfloor + 1, M_n} | x)} - 1 \right| \\ &\leq \omega(V_{1, M_n}, V_{M_n, M_n}, x, h) \exp[\omega(V_{1, M_n}, V_{M_n, M_n}, x, h)]. \end{aligned}$$

Hence,

$$\begin{aligned} R_n^{(Q)}(x) &= \sup_{\alpha \in [\tau u_{n,x}, u_{n,x}]} \left| \frac{\widehat{Q}_n(\alpha|x)}{Q(V_{\lfloor \alpha M_n \rfloor + 1, M_n}|x)} - 1 \right| \left| \frac{Q(V_{\lfloor \alpha M_n \rfloor + 1, M_n}|x)}{Q(\alpha|x)} \right| \\ &\leq \omega(V_{1, M_n}, V_{M_n, M_n}, x, h) \exp[\omega(V_{1, M_n}, V_{M_n, M_n}, x, h)] (1 + T_n(x)). \end{aligned}$$

Use finally (6.2) and (A.1) to complete the proof. \square

The next lemma examines the convergence of $T_n(x)$, defined in the above lemma, given $M_x(h)$.

Lemma A.2. *Let U_1, \dots, U_p be independent standard uniform variables. Assume that (2.3) and (H_{SO}) hold. If $a(x) \in (0, 1)$ is such that $p^{1-a(x)} \Delta^2(p^{a(x)}|x) \rightarrow \lambda \in \mathbb{R}$ as $p \rightarrow \infty$ then, for all $d_1, d_2 > 0$ with $d_1 < d_2$, we have:*

$$p^{(1-a(x))/2} \sup_{\alpha \in [d_1 p^{-a(x)}, d_2 p^{-a(x)}]} \left| \frac{Q(U_{\lfloor p\alpha \rfloor + 1, p}|x)}{Q(\alpha|x)} - 1 \right| = \mathcal{O}_{\mathbb{P}}(1).$$

Proof: Recall that (H_{SO}) entails that (3.1) holds. Then, one can apply [27, Theorem 2.4.8] to the independent random variables $\{Q(U_i|x), i = 1, \dots, p\}$ distributed from the conditional survival function $S(\cdot|x)$: because

$$\inf_{\alpha \in [d_1 p^{-a(x)}, d_2 p^{-a(x)}]} \frac{\alpha}{d_2 p^{-a(x)}} = \frac{d_1}{d_2} > 0,$$

it holds that

$$(A.3) \quad p^{(1-a(x))/2} \sup_{\alpha \in [d_1 p^{-a(x)}, d_2 p^{-a(x)}]} \left| \frac{Q(U_{\lfloor p\alpha \rfloor + 1, p}|x)}{Q(d_2 p^{-a(x)}|x)} - \left(\frac{\alpha p^{a(x)}}{d_2} \right)^{-\gamma(x)} \right| = \mathcal{O}_{\mathbb{P}}(1).$$

Since (3.1) must in fact hold locally uniformly in $z > 0$ (see [27, Theorem B.2.9]) and $[d_1, d_2]$ is a compact interval, it is clear that

$$(A.4) \quad p^{(1-a(x))/2} \sup_{\alpha \in [d_1 p^{-a(x)}, d_2 p^{-a(x)}]} \left| \frac{Q(\alpha|x)}{Q(d_2 p^{-a(x)}|x)} - \left(\frac{\alpha p^{a(x)}}{d_2} \right)^{-\gamma(x)} \right| = \mathcal{O}(1).$$

Combine (A.3) and (A.4) to conclude the proof. \square

Lemma A.3 below controls a bias term appearing in the proof of Theorem 3.1.

Lemma A.3. *Assume that conditions (2.3) and (H_{SO}) are satisfied. If $m_x(h) \rightarrow \infty$ and $\beta_n/u_{n,x} \rightarrow 0$ we have that:*

$$\sup_{\alpha \in [\tau u_{n,x}, u_{n,x}]} \left| \frac{Q(\alpha|x)}{Q(\beta_n|x)} \left(\frac{\alpha}{\beta_n} \right)^{\gamma(x)} - 1 \right| = \mathcal{O}_{\mathbb{P}}(\Delta(u_{n,x}^{-1}|x)).$$

Proof: Recall

$$\alpha^{\gamma(x)} Q(\alpha|x) = c(x) \exp \left(\int_1^{\alpha^{-1}} \frac{\Delta(v|x)}{v} dv \right),$$

and therefore

$$\frac{Q(\alpha|x)}{Q(\beta_n|x)} \left(\frac{\alpha}{\beta_n} \right)^{\gamma(x)} = \exp \left(\int_{\beta_n^{-1}}^{\alpha^{-1}} \frac{\Delta(v|x)}{v} dv \right).$$

Furthermore, since $\alpha \leq u_{n,x}$,

$$\left| \int_{\beta_n^{-1}}^{\alpha^{-1}} \frac{\Delta(v|x)}{v} dv \right| \leq |\Delta(u_{n,x}^{-1}|x)| \int_1^\infty \left| \frac{\Delta(yu_{n,x}^{-1}|x)}{\Delta(u_{n,x}^{-1}|x)} \right| \frac{dy}{y}.$$

As the function $y \mapsto y^{-1}\Delta(y|x)$ is regularly varying with index $\rho(x) - 1 < -1$, we may write, according to [4, Theorem 1.5.2],

$$\left| \int_{\beta_n^{-1}}^{\alpha^{-1}} \frac{\Delta(v|x)}{v} dv \right| \leq 2|\Delta(u_{n,x}^{-1}|x)| \int_1^\infty y^{\rho(x)-1} dy = -\frac{2}{\rho(x)} |\Delta(u_{n,x}^{-1}|x)|.$$

Since the right-hand side converges to 0 and does not depend on α , it follows by a Taylor expansion of the exponential function that

$$\sup_{\alpha \in (\tau u_{n,x}, u_{n,x}]} \left| \frac{Q(\alpha|x)}{Q(\beta_n|x)} \left(\frac{\alpha}{\beta_n} \right)^{\gamma(x)} - 1 \right| = \mathcal{O}_{\mathbb{P}} \left(\Delta(u_{n,x}^{-1}|x) \right),$$

which is the required conclusion. \square

The next result is dedicated to the statistics $\tilde{\gamma}_n(x, u_{n,x})$ and $R_n^{(\gamma)}(x)$ introduced in the proof of Theorem 3.2, equation (6.10).

Lemma A.4. *Let U_i , $i \geq 1$ be independent standard uniform random variables. For any $x \in \mathcal{E}$ such that $m_x(h) > 0$, the conditional distribution of $\tilde{\gamma}_n(x, u_{n,x})$ given $M_x(h) = p$ is that of*

$$\bar{\gamma}_p(x, u_{n,x}) = \frac{1}{pu_{n,x}} \sum_{i=1}^{\lfloor pu_{n,x} \rfloor} \Phi \left(\frac{i}{pu_{n,x}} \right) i \log \frac{Q(U_{i,p}|x)}{Q(U_{i+1,p}|x)},$$

and given $M_x(h) = p$, $R_n^{(\gamma)}(x)$ is bounded from above by

$$2\omega(U_{1,p}, U_{p,p}, x, h) \int_0^{u_{n,x}} |\Psi(\alpha, u_{n,x})| d\alpha.$$

Proof: Set again $M_n = M_x(h)$. Equation (6.2) entails that the conditional distribution of $\widehat{\gamma}_n(x, u_{n,x})$ given $M_n = p$ is that of

$$\begin{aligned} & \sum_{i=1}^{\lfloor pu_{n,x} \rfloor} W_{i,n}(u_{n,x}, p) \log \frac{Q(U_{i,p}|x)}{Q(U_{\lfloor pu_{n,x} \rfloor+1,p}|x)} \\ &= \sum_{i=1}^{\lfloor pu_{n,x} \rfloor} W_{i,n}(u_{n,x}, p) \sum_{j=i}^{\lfloor pu_{n,x} \rfloor} \log \frac{Q(U_{j,p}|x)}{Q(U_{j+1,p}|x)}, \end{aligned}$$

where $\{U_i, i \geq 1\}$ are independent standard uniform random variables, and this is equal to $\bar{\gamma}_p(x, u_{n,x})$ by switching the summation order and using assumption (H_Ψ) . Now, since $\widehat{Q}_n(\alpha|x) = Y_{M_n-i+1, M_n}^*$ for all $\alpha \in [(i-1)/M_n, i/M_n)$, one has

$$\widehat{\gamma}_n(x, u_{n,x}) = \sum_{i=1}^{\lfloor u_{n,x} M_n \rfloor} W_{i,n}(u_{n,x}, M_n) \log \frac{Y_{M_n-i+1, M_n}^*}{Y_{M_n-\lfloor u_{n,x} M_n \rfloor, M_n}^*},$$

where $W_{i,n}(u_{n,x}, M_n)$ was defined in (6.11). Hence the identity

$$R_n^{(\gamma)}(x) = \sum_{i=1}^{\lfloor u_{n,x} M_n \rfloor} W_{i,n}(u_{n,x}, M_n) \log \left[\frac{Q(V_{\lfloor u_{n,x} M_n \rfloor+1, M_n}|x)}{Q(V_{i, M_n}|x)} \frac{Y_{M_n-i+1, M_n}^*}{Y_{M_n-\lfloor u_{n,x} M_n \rfloor, M_n}^*} \right].$$

Using the bound (A.2) yields to

$$\begin{aligned} R_n^{(\gamma)}(x) &\leq 2\omega(V_{1, M_n}, V_{M_n, M_n}, x, h) \sum_{i=1}^{\lfloor u_{n,x} M_n \rfloor} |W_{i,n}(u_{n,x}, M_n)| \\ &\leq 2\omega(V_{1, M_n}, V_{M_n, M_n}, x, h) \int_0^{u_{n,x}} |\Psi(\alpha, u_{n,x})| d\alpha. \end{aligned}$$

Using equation (6.2) completes the proof. \square

Our next result studies some particular Riemann sums. It shall prove useful when examining the convergence of $\widetilde{\gamma}_n(x, u_{n,x})$ given $M_x(h)$, see Lemma A.6.

Lemma A.5. *Let f be an integrable function on $(0, 1)$. Assume that f is nonnegative and nonincreasing. For any nonnegative continuous function g on $[0, 1]$ we have that:*

$$\lim_{m \rightarrow \infty} \frac{1}{m-1} \sum_{i=1}^{m-1} f\left(\frac{i}{m}\right) g\left(\frac{i}{m}\right) = \int_0^1 f(t)g(t)dt.$$

If moreover f is square-integrable then:

$$\lim_{m \rightarrow \infty} \sqrt{m} \left| \frac{1}{m-1} \sum_{i=1}^{m-1} f\left(\frac{i}{m}\right) - \int_0^1 f(t)dt \right| = 0.$$

Proof: To prove the first statement, it suffices to show that $|S_m(f, g) - S(f, g)| \rightarrow 0$ as $m \rightarrow \infty$ where

$$S_m(f, g) := \frac{1}{m} \sum_{i=1}^{m-1} f\left(\frac{i}{m}\right) g\left(\frac{i}{m}\right) \text{ and } S(f, g) := \int_0^1 f(t)g(t)dt.$$

Note first that:

$$\begin{aligned} |S(f, g) - S_m(f, g)| &\leq \sum_{i=1}^{m-1} \int_{(i-1)/m}^{i/m} \left| f(t)g(t) - f\left(\frac{i}{m}\right) g\left(\frac{i}{m}\right) \right| dt \\ &\quad + \int_{(m-1)/m}^1 f(t)g(t)dt. \end{aligned}$$

Since g is nonnegative on $[0, 1]$ and f is nonincreasing, it is straightforward that for all $t \in [(i-1)/m, i/m)$

$$\begin{aligned} |f(t)g(t) - f(i/m)g(i/m)| &\leq f(t) \sup_{|s-s'|\leq 1/m} |g(s) - g(s')| \\ &\quad + \|g\|_\infty (f(t) - f(i/m)), \end{aligned}$$

where $\|g\|_\infty$ is the finite supremum of g on $[0, 1]$. The fact that f is nonincreasing yields $f(t) - f(i/m) \leq f((i-1)/m) - f(i/m)$ for all $i = 2, \dots, m$ and thus the previous inequality leads to

$$\begin{aligned} |S(f, g) - S_m(f, g)| &\leq \int_0^1 f(t)dt \sup_{|s-s'|\leq 1/m} |g(s) - g(s')| \\ &\quad + \|g\|_\infty \left(\int_0^{1/m} f(t)dt - \frac{f(1)}{m} \right) \\ \text{(A.5)} \quad &\quad + \|g\|_\infty \int_{(m-1)/m}^1 f(t)dt \rightarrow 0 \end{aligned}$$

by the uniform continuity of g on $[0, 1]$ and the fact that f is an integrable function. This proves the first statement of the result. To prove the second one, remark that:

$$\sqrt{m} \left| \frac{1}{m-1} \sum_{i=1}^{m-1} f\left(\frac{i}{m}\right) - \int_0^1 f(t)dt \right| \leq \frac{\sqrt{m}}{m-1} S_m(f, 1) + \sqrt{m} |S(f, 1) - S_m(f, 1)|.$$

Using the first statement with $g = 1$ entails that the first term of the left-hand side converges to 0 as $m \rightarrow \infty$. Now, taking $g = 1$ in (A.5) leads to

$$\sqrt{m} |S(f, 1) - S_m(f, 1)| \leq \sqrt{m} \int_0^{1/m} f(t)dt + \sqrt{m} \int_{(m-1)/m}^1 f(t)dt.$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned} \sqrt{m} \int_0^{1/m} f(t)dt &\leq \left(\int_0^{1/m} f^2(t)dt \right)^{1/2} \rightarrow 0 \\ \text{and } \sqrt{m} \int_{(m-1)/m}^1 f(t)dt &\leq \left(\int_{(m-1)/m}^1 f^2(t)dt \right)^{1/2} \rightarrow 0 \end{aligned}$$

since f^2 is integrable on $(0, 1)$. The proof is complete. \square

The next lemma establishes the asymptotic normality of the random variable $\bar{\gamma}_p(x, k, k')$ introduced in the proof of Theorem 3.2, equation (6.15).

Lemma A.6. *Assume that conditions (2.3), (H_{SO}) and (H_Ψ) are satisfied. Let $k(p)$ and $k'(p)$ be two sequences satisfying, for some $a(x) \in (0, 1)$, $p^{a(x)-1}k(p) \rightarrow 1$ and $p^{(1-a(x))/2}[k(p)/k'(p) - 1] \rightarrow 0$ as $p \rightarrow \infty$. Let U_1, \dots, U_p be independent standard uniform random variables. If $p^{1-a(x)}\Delta^2(p^{a(x)}|x) \rightarrow \lambda(x) \in \mathbb{R}$, then the random variable*

$$\tilde{\gamma}_p(x, k(p), k'(p)) := \frac{1}{k(p)} \sum_{i=1}^{\lfloor k'(p) \rfloor} \Phi\left(\frac{i}{\lfloor k'(p) \rfloor + 1}\right) i \log \frac{Q(U_{i,p}|x)}{Q(U_{i+1,p}|x)}$$

is such that $p^{(1-a(x))/2}(\tilde{\gamma}_p(x, k(p), k'(p)) - \gamma(x))$ converges in distribution to a normal distribution with mean $\lambda(x) \int_0^1 \Phi(\alpha) \alpha^{-\rho(x)} d\alpha$ and variance $\gamma^2(x) \int_0^1 \Phi^2(\alpha) d\alpha$.

Proof: For the sake of brevity, let $\tilde{\gamma}_p(x) := \tilde{\gamma}_p(x, k(p), k'(p))$. Let $v_p := p^{(1-a(x))/2}$ and for $j \in \{1, \dots, k'(p)\}$

$$\tilde{\Delta}_j(p|x) := \Delta\left(\frac{p+1}{\lfloor k'(p) \rfloor + 1} \middle| x\right) \left(\frac{j}{\lfloor k'(p) \rfloor + 1}\right)^{-\rho(x)}$$

Under conditions (2.3), (H_{SO}) and (H_Ψ) , one can apply Theorem 3.1 in Beirlant et al. [1] to prove that

$$v_p \left\{ \frac{k(p)}{\lfloor k'(p) \rfloor} \tilde{\gamma}_p(x) - \frac{1}{\lfloor k'(p) \rfloor} \sum_{j=1}^{\lfloor k'(p) \rfloor} \Phi\left(\frac{j}{\lfloor k'(p) \rfloor + 1}\right) [\gamma(x) + \tilde{\Delta}_j(p|x)] \right\}$$

converges to a centered normal distribution with variance $\sigma_\Phi^2 := \gamma^2(x) \int_0^1 \Phi^2(\alpha) d\alpha$. As a direct consequence of Lemma A.5, the previous convergence can be rewritten

$$(A.6) \quad v_p \left[\frac{k(p)}{\lfloor k'(p) \rfloor} \tilde{\gamma}_p(x) - \gamma(x) \right] \xrightarrow{d} \mathcal{N}\left(\lambda(x) \int_0^1 \Phi(\alpha) \alpha^{-\rho(x)} d\alpha, \sigma_\Phi^2\right).$$

Finally, since

$$\begin{aligned} v_p [\tilde{\gamma}_p(x) - \gamma(x)] &= v_p \left(\frac{\lfloor k'(p) \rfloor}{k(p)} - 1 \right) \frac{k(p)}{\lfloor k'(p) \rfloor} \tilde{\gamma}_p(x) \\ &\quad + v_p \left[\frac{k(p)}{\lfloor k'(p) \rfloor} \tilde{\gamma}_p(x) - \gamma(x) \right], \end{aligned}$$

a combination of convergence (A.6) and of the fact that $v_p[k(p)/k'(p) - 1] \rightarrow 0$ as $p \rightarrow \infty$ concludes the proof. \square

The final lemma is a technical tool we shall need to bridge the gap between the convergence of our estimators and that of their conditional versions.

Lemma A.7. *Let $\{Z_p, p \in \mathbb{N}\}$ be a sequence of random variables such that for all $t \in \mathbb{R}$, $\mathbb{P}(Z_p \leq t) \rightarrow H(t)$ where H is a continuous cumulative distribution function. For $n \in \mathbb{N}$, let $I_n := [u_n, v_n]$ where $u_n \rightarrow \infty$ as $n \rightarrow \infty$ and let (a_n) be a sequence such that there exist two functions ξ_1 and ξ_2 converging to 1 at infinity with*

$$\sup_{p \in I_n} \frac{\xi_1(p)}{a_n} \leq 1 \leq \inf_{p \in I_n} \frac{\xi_2(p)}{a_n}.$$

Then, for all $t \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} \sup_{p \in I_n} |\mathbb{P}(a_n Z_p \leq t) - H(t)| = 0.$$

Proof: We start by remarking that for all $\kappa > 0$,

$$\sup_{p \in I_n} |\mathbb{P}(a_n Z_p \leq t) - H(t)| \leq D_{n,p} + \sup_{p \in I_n} \mathbb{P}(|(a_n - 1)Z_p| > \kappa),$$

where

$$D_{n,p} := \sup_{p \in I_n} |\mathbb{P}(\{a_n Z_p \leq t\} \cap \{|(a_n - 1)Z_p| \leq \kappa\}) - H(t)|.$$

Now, since H is continuous, there exists $\kappa > 0$ such that for n large enough,

$$|H(t) - H(t + \kappa)| \leq \frac{\varepsilon}{6} \text{ and } |H(t) - H(t - \kappa)| \leq \frac{\varepsilon}{6}.$$

Furthermore, since $\xi_1(p) \leq a_n \leq \xi_2(p)$ for any $p \in I_n$, using the inequality $|x| \leq |a| + |b|$ which holds for all $x \in [a, b]$, one has for all $p \in I_n$ that $|a_n - 1| \leq |\xi_1(p) - 1| + |\xi_2(p) - 1|$; besides, since $Z_p = \mathcal{O}_{\mathbb{P}}(1)$ and ξ_1, ξ_2 converge to 1 at infinity, we have $|\xi_1(p) - 1|Z_p = o_{\mathbb{P}}(1)$ and $|\xi_2(p) - 1|Z_p = o_{\mathbb{P}}(1)$. Therefore, for all $\varepsilon > 0$,

$$\sup_{p \in I_n} \mathbb{P}(|(a_n - 1)Z_p| > \kappa) \leq \sup_{p \in I_n} \mathbb{P}(|\xi_1(p) - 1||Z_p| + |\xi_2(p) - 1||Z_p| > \kappa) \leq \frac{\varepsilon}{6}$$

for n large enough. Now remark that for all $p \in I_n$, $\mathbb{P}(\{a_n Z_p \leq t\} \cap \{|(a_n - 1)Z_p| \leq \kappa\}) \leq \mathbb{P}(Z_p \leq t + \kappa)$ and that

$$\begin{aligned} \mathbb{P}(\{a_n Z_p \leq t\} \cap \{|(a_n - 1)Z_p| \leq \kappa\}) &\geq \mathbb{P}(\{Z_p \leq t - \kappa\} \cap \{|(a_n - 1)Z_p| \leq \kappa\}) \\ &\geq \mathbb{P}(Z_p \leq t - \kappa) - \mathbb{P}(|(a_n - 1)Z_p| > \kappa). \end{aligned}$$

Hence, for all $\kappa > 0$, the inequality:

$$\begin{aligned} D_{n,p} &\leq \sup_{p \in I_n} |\mathbb{P}(Z_p \leq t + \kappa) - H(t + \kappa)| + \sup_{p \in I_n} |\mathbb{P}(Z_p \leq t - \kappa) - H(t - \kappa)| \\ &\quad + |H(t) - H(t + \kappa)| + |H(t) - H(t - \kappa)| + \frac{\varepsilon}{6} \\ &\leq \sup_{p \in I_n} |\mathbb{P}(Z_p \leq t + \kappa) - H(t + \kappa)| + \sup_{p \in I_n} |\mathbb{P}(Z_p \leq t - \kappa) - H(t - \kappa)| + \frac{\varepsilon}{2}. \end{aligned}$$

By assumption, for n large enough:

$$\sup_{p \in I_n} |\mathbb{P}(Z_p \leq t + \kappa) - H(t + \kappa)| \leq \frac{\varepsilon}{6} \text{ and } \sup_{p \in I_n} |\mathbb{P}(Z_p \leq t - \kappa) - H(t - \kappa)| \leq \frac{\varepsilon}{6}.$$

It is now straightforward to conclude the proof. \square

ACKNOWLEDGMENTS

We acknowledge the valuable suggestions from Associate Editor and referee.

REFERENCES

- [1] BEIRLANT, J.; DIERCKX, G.; GUILLOU, A. and STĂRICĂ, C. (2002). On exponential representations of log-spacings of extreme order statistics, *Extremes*, **5**(2), 157–180.
- [2] BEIRLANT, J.; GOEGBEUR, Y.; SEGERS, J. and TEUGELS, J.L. (2004). *Statistics of extremes – Theory and applications*, Wiley Series in Probability and Statistics.
- [3] BILLINGSLEY, P. (1979). *Probability and measure*, John Wiley and Sons.
- [4] BINGHAM, N.H.; GOLDIE, C.M. and TEUGELS, J.L. (1987). *Regular Variation*, Cambridge, U.K., Cambridge University Press.
- [5] CAI, J.-J.; DE HAAN, L. and ZHOU, C. (2013). Bias correction in extreme value statistics with index around zero, *Extremes*, **16**, 173–201.
- [6] CHAVEZ-DEMOULIN, V. and DAVISON, A.C. (2005). Generalized additive modelling of sample extremes, *Journal of the Royal Statistical Society, series C*, **54**, 207–222.
- [7] CHERNOZHUKOV, V. (2005). Extremal quantile regression, *The Annals of Statistics*, **33**(2), 806–839.
- [8] CHERNOZHUKOV, V. and DU, S. (2008). Extremal quantiles and Value-at-Risk, *The New Palgrave Dictionary of Economics*, second edition.
- [9] DAOUIA, A.; GARDES, L.; GIRARD, S. and LEKINA, A. (2011). Kernel estimators of extreme level curves, *Test*, **20**(2), 311–333.
- [10] DAOUIA, A.; GARDES, L. and GIRARD, S. (2013). On kernel smoothing for extremal quantile regression, *Bernoulli*, **19** 5B, 2557–2589.
- [11] DAVISON, A.C. and RAMESH, N.I. (2000). Local likelihood smoothing of sample extremes, *Journal of the Royal Statistical Society, series B*, **62**, 191–208.
- [12] DAVISON, A.C. and SMITH, R.L. (1990). Models for exceedances over high thresholds, *Journal of the Royal Statistical Society, series B*, **52**, 393–442.
- [13] DEKKERS, A.L.M.; EINMAHL, J.H.J. and DE HAAN, L. (1989). A moment estimator for the index of an extreme-value distribution, *Annals of Statistics*, **17**(4), 1833–1855.
- [14] FERRATY, F. and VIEU, P. (2002). The functional nonparametric model and application to spectrometric data, *Computational Statistics*, **17**, 545–564.
- [15] FERRATY, F. and VIEU, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*, Springer Series in Statistics, Springer.

- [16] FERRATY, F.; MAS, A. and VIEU, P. (2007). Nonparametric regression on functional data: inference and practical aspects, *Australian and New Zealand Journal of Statistics*, **49**(3), 267–286.
- [17] FRAGA ALVES, M.I.; GOMES, M.I. and DE HAAN, L. (2003). A new class of semi-parametric estimators of the second order parameter, *Portugaliae Mathematica*, **60**, 193–213.
- [18] GARDES, L. (2015). A general estimator for the extreme value index: applications to conditional and heteroscedastic extremes, *Extremes*, **18**(3), 479–510.
- [19] GARDES, L. and GIRARD, S. (2008). A moving window approach for nonparametric estimation of the conditional tail index, *Journal of Multivariate Analysis*, **99**, 2368–2388.
- [20] GARDES, L. and GIRARD, S. (2010). Conditional extremes from heavy-tailed distributions: an application to the estimation of extreme rainfall return levels, *Extremes*, **13**, 177–204.
- [21] GARDES, L. and GIRARD, S. (2012). Functional kernel estimators of large conditional quantiles, *Electronic Journal of Statistics*, **6**, 1715–1744.
- [22] GARDES, L. and STUPFLER, G. (2014). Estimation of the conditional tail index using a smoothed local Hill estimator, *Extremes*, **17**(1), 45–75.
- [23] GOEGBEUR, Y.; GUILLOU, A. and OSMANN, M. (2014). A local moment type estimator for the extreme value index in regression with random covariates, *Canadian Journal of Statistics*, **42**, 487–507.
- [24] GOEGBEUR, Y.; GUILLOU, A. and OSMANN, M. (2017). A local moment type estimator for an extreme quantile in regression with random covariates, *Communications in Statistics – Theory and Methods*, **46**(1), 319–343.
- [25] GOEGBEUR, Y.; GUILLOU, A. and SCHORGEN, A. (2014). Nonparametric regression estimation of conditional tails — the random covariate case, *Statistics*, **48**(4), 732–755.
- [26] GOEGBEUR, Y.; GUILLOU, A. and STUPFLER, G. (2015). Uniform asymptotic properties of a nonparametric regression estimator of conditional tails, *Annales de l'Institut Henri Poincaré (B): Probability and Statistics*, **51**(3), 1190–1213.
- [27] DE HAAN, L. and FERREIRA, A. (2006). *Extreme value theory: An introduction*, Springer.
- [28] HALL, P. and TAJVIDI, N. (2000). Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data, *Statistical Science*, **15**, 153–167.
- [29] HILL, B.M. (1975). A simple general approach to inference about the tail of a distribution, *Annals of Statistics*, **3**, 1163–1174.
- [30] JIŘINA, M. (1959). On regular conditional probabilities, *Czechoslovak Mathematical Journal*, **9**, 445–451.
- [31] KRATZ, M. and RESNICK, S. (1996). The QQ-estimator and heavy tails, *Stochastic Models*, **12**, 699–724.
- [32] LI, Q.; LIN, J. and RACINE, J.S. (2013). Optimal bandwidth selection for nonparametric conditional distribution and quantile functions, *Journal of Business & Economic Statistics*, **31**(1), 57–65.

- [33] PICKANDS, J. (1975). Statistical inference using extreme order statistics, *Annals of Statistics*, **3**, 119–131.
- [34] RESNICK, S. and STĂRICĂ, C. (1997). Smoothing the Hill estimator, *Advances in Applied Probability*, **29**, 271–293.
- [35] SCHULTZE, J. and STEINEBACH, J. (1996). On least squares estimates of an exponential tail coefficient, *Statistics and Decisions*, **14**, 353–372.
- [36] SMITH, R.L. (1989). Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone (with discussion), *Statistical Science*, **4**, 367–393.
- [37] STUPFLER, G. (2013). A moment estimator for the conditional extreme-value index, *Electronic Journal of Statistics*, **7**, 2298–2343.
- [38] STUPFLER, G. (2016). Estimating the conditional extreme-value index under random right-censoring, *Journal of Multivariate Analysis*, **144**, 1–24.
- [39] WANG, H.J.; LI, D. and HE, X. (2012). Estimation of high conditional quantiles for heavy-tailed distributions, *Journal of the American Statistical Association*, **107**(500), 1453–1464.
- [40] WANG, H.J. and LI, D. (2013). Estimation of extreme conditional quantiles through power transformation, *Journal of the American Statistical Association*, **108**(503), 1062–1074.
- [41] WANG, H. and TSAI, C.L. (2009). Tail index regression, *Journal of the American Statistical Association*, **104**(487), 1233–1240.
- [42] WEISSMAN, I. (1978). Estimation of parameters and large quantiles based on the k largest observations, *Journal of the American Statistical Association*, **73**, 812–815.

REVSTAT – STATISTICAL JOURNAL

Background

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

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

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

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

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

- The editorial policy will focus on publishing research articles at the highest level in the domains of Probability and Statistics with emphasis on the originality and importance of the research.
- All research articles will be refereed by at least two persons, one from the Editorial Board and another external.

— The only working language allowed will be English. — Four volumes are scheduled for publication, one in January, one in April, one in July and the other in October.

Aims and Scope

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

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

Abstract and Indexing Services

The REVSTAT is covered by the following abstracting/indexing services:

- Current Index to Statistics
- Google Scholar
- Mathematical Reviews
- Science Citation Index Expanded
- Zentralblatt für Mathematic

Instructions to Authors, special-issue editors and publishers

The articles should be written in English and may be submitted in two different ways:

- By sending the paper in PDF format to the Executive Editor (revstat@ine.pt) and to one of the two Editors or Associate Editors, whose opinion the author wants to be taken into account, together to the following e-mail address: revstat@fc.ul.pt

- By sending the paper in PDF format to the Executive Editor (revstat@ine.pt), together with the corresponding PDF or PostScript file to the following e-mail address: revstat@fc.ul.pt.

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

Manuscripts (text, tables and figures) should be typed only in black on one side, in double-spacing, with a left margin of at least 3 cm and with less than 30 pages. The first page should include the name, institution and address of the author(s) and a summary of less than one hundred words, followed by a maximum of six key words and the AMS 2000 subject classification.

Authors are obliged to write the final version of accepted papers using LaTeX, in the REVSTAT style. This style (REVSTAT.sty), and examples file (REVSTAT.tex), which may be download to PC Windows System (Zip format), Macintosh, Linux and Solaris Systems (StuffIt format), and Mackintosh System (BinHex Format), are available in the REVSTAT link of the Statistics Portugal Website: <http://www.ine.pt/revstat/inicio.html>

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

Accepted papers

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

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

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

Copyright

Upon acceptance of an article, the author(s) will be asked to transfer copyright of the article to the publisher, Statistics Portugal, in order to ensure the widest possible dissemination of information, namely through the Statistics Portugal website (<http://www.ine.pt>).

After assigning copyright, authors may use their own material in other publications provided that REVSTAT is acknowledged as the original place of publication. The Executive Editor of the Journal must be notified in writing in advance.

Editorial Board

Editor-in-Chief

Isabel Fraga Alves, University of Lisbon, Portugal

Co-Editor

Giovani L. Silva, University of Lisbon, Portugal

Associate Editors

Marília Antunes, University of Lisbon, Portugal

Barry Arnold (2019), University of California, USA

Narayanaswamy Balakrishnan, McMaster University, Canada

Jan Beirlant, Katholieke Universiteit Leuven, Belgium

Graciela Boente (2019-2020), University of Buenos Aires, Argentina

Paula Brito, University of Porto, Portugal

Vanda Inácio de Carvalho, University of Edinburgh, UK

Arthur Charpentier, Université du Québec à Montréal, Canada

Valérie Chavez-Demoulin, University of Lausanne, Switzerland

David Conesa, University of Valencia, Spain

Charmaine Dean, University of Waterloo, Canada

Jorge Milhazes Freitas, University of Porto, Portugal

Alan Gelfand, Duke University, USA

Stéphane Girard, Inria Grenoble Rhône-Alpes, France

Wenceslao Gonzalez-Manteiga, University of Santiago de Compostela, Spain

Marie Kratz, ESSEC Business School, France

Victor Leiva, Pontificia Universidad Católica de Valparaíso, Chile

Maria Nazaré Mendes-Lopes, University of Coimbra, Portugal

Fernando Moura, Federal University of Rio de Janeiro, Brazil

John Nolan, American University, USA

Paulo Eduardo Oliveira, University of Coimbra, Portugal

Pedro Oliveira, University of Porto, Portugal

Carlos Daniel Paulino (2019-2021), University of Lisbon, Portugal

Arthur Pewsey, University of Extremadura, Spain

Gilbert Saporta, Conservatoire National des Arts et Métiers, France

Alexandra M. Schmidt, McGill University, Canada

Julio Singer, University of Sao Paulo, Brazil

Manuel Scotto, University of Lisbon, Portugal

Lisete Sousa, University of Lisbon, Portugal

Milan Stehlík, University of Valparaíso, Chile and LIT-JK University Linz, Austria

María Dolores Ugarte, Public University of Navarre, Spain

Executive Editor

José A. Pinto Martins, Statistics Portugal

Former Executive Editors

Maria José Carrilho, Statistics Portugal (2005-2015)

Ferreira da Cunha, Statistics Portugal (2003–2005)

Secretary

Liliana Martins, Statistics Portugal