TECHNICKÁ UNIVERZITA V LIBERCI
Fakulta přírodovědně-humanitní
a pedagogická

Statistická inference pomocí L-momentů

Disertační práce

Studijní program: P1103 – Aplikovaná matematika
Studijní obor: 1103V035 – Matematické modely a jejich aplikace

Autor práce: **Mgr. Tereza Šimková**
Vedoucí práce: prof. RNDr. Jan Picek, CSc.

Liberec 2018

TECHNICAL UNIVERSITY OF LIBEREC
Faculty of Science, Humanities
and Education

Statistical Inference Using L-Moments

Dissertation

Study programme: P1103 – Applied mathematics
Study branch: 1103V035 – Mathematical models and their applications

Author: **Mgr. Tereza Šimková**
Supervisor: prof. RNDr. Jan Picek, CSc.

Liberec 2018

Prohlášení

Byla jsem seznámena s tím, že na mou disertační práci se plně vztahuje zákon č. 121/2000 Sb., o právu autorském, zejména § 60 – školní dílo.

Beru na vědomí, že Technická univerzita v Liberci (TUL) nezasahuje do mých autorských práv užitím mé disertační práce pro vnitřní potřebu TUL.

Užiji-li disertační práci nebo poskytnu-li licenci k jejímu využití, jsem si vědoma povinnosti informovat o této skutečnosti TUL; v tomto případě má TUL právo ode mne požadovat úhradu nákladů, které vynaložila na vytvoření díla, až do jejich skutečné výše.

Disertační práci jsem vypracovala samostatně s použitím uvedené literatury a na základě konzultací s vedoucím mé disertační práce a konzultantem.

Současně čestně prohlašuji, že tištěná verze práce se shoduje s elektronickou verzí, vloženou do IS STAG.

Datum:

Podpis:

Statistická inference pomocí L-momentů

Anotace: Rozdělení s těžšími chvosty, než má normální rozdělení, se vyskytují v oblastech, ve kterých jsou pozorovány extrémy, jako například v hydrologii, meteorologii nebo také v ekonomii. Použití konvenčních momentů v analýze náhodné veličiny s rozdělením s těžšími chvosty však není vhodné z důvodu předpokladu existence momentů vyšších řádů. Jednorozměrné L-momenty, které jsou alternativou ke konvenčním momentům, jsou definovány jako střední hodnota jisté lineární kombinace pořádkových statistik a to pouze za předpokladu konečné střední hodnoty. Podobně jako je tomu v jednorozměrném případě, mnohorozměrná analýza zahrnující zejména vektor středních hodnot a kovarianční nebo korelační matice je založena na předpokladu existence vyšších momentů. Rozšíření jednorozměrných L-momentů do mnohorozměrného případu umožňuje na rozdíl od těchto charakteristik popsat mnohorozměrné rozdělení pouze za předpokladu konečné střední hodnoty. Cílem práce je poskytnout komplexní přehled o L-momentech a jejich použití ve statistické inferenci a vypořádat se rovněž s problémy, které se objevily při jejich studiu. Kromě obecné teorie L-momentů se zaměřením na určité vlastnosti a metodologii jejich použití k odhadu parametrů pravděpodobnostních rozdělení a v regionální frekvenční analýze představuje práce první čtyři L-, LQ- a TL-momenty tříparametrického zobecněného Paretova rozdělení a rozdělení extrémních hodnot a odhady jejich parametrů založené na těchto momentech. Rovněž uvádí asymptotické L-momentové intervaly spolehlivosti parametrů a kvantilů těchto rozdělení. Dále přináší podrobný postup, jak provést testování homogenity v trojrozměrné regionální frekvenční analýze. Nakonec je představeno vylepšení dvourozměrného L-momentového testu homogenity pro případ prostorově korelovaných dat.

Klíčová slova: L-moment, rozdělení s těžšími chvosty, odhad parametrů a kvantilů, regionální frekvenční analýza, kopula

Statistical Inference Using L-Moments

Annotation: Distributions with heavier tails than has the normal distribution appear in many fields in which extremes are observed, such as climatology, hydrology, meteorology, or economics as well. However, in analysis of a random variable having a probability distribution with heavier tails, the traditionally used conventional moments are not sufficient due to the moment assumptions of higher orders. Univariate L-moments as an alternative to the conventional moments are defined as an expectation of certain linear combinations of order statistics under only first order moment assumptions. Analogously to the univariate framework, multivariate analysis of a random vector mainly including the mean vector and covariance or correlation matrices is based on the assumptions of second and higher order moments. In comparison to these characteristics, the extension of univariate L-moments to the multivariate case enables to describe a multivariate probability distribution under only finite mean assumptions. The aim of the thesis is to present a comprehensive overview of L-moments and their application in statistical inference, and to deal with issues that appeared in study of them as well. Among a general theory of L-moments with focus on their specific properties and methodology how to employ them in estimating parameters of a probability distribution and in regional frequency analysis, the thesis presents expressions of the first four L-, LQ-, and TL-moments of the three-parametric generalized Pareto and generalized extreme-value distributions and estimators of their parameters based on these quantities. The L-moments' asymptotic confidence intervals of parameters of these distributions are presented as well. Further, a detailed procedure how to perform L-moment homogeneity testing in trivariate regional frequency analysis is introduced. Finally, the improvement of the bivariate L-moment homogeneity test for the case of cross-correlated data is proposed.

Keywords: L-moment, distribution with heavier tails, parameter and quantile estimation, regional frequency analysis, copula

Acknowledgements

I would like to express my sincere gratitude to my advisor Prof. Jan Picek, for his support during my Ph.D. study and related research, patience and knowledge that broaden my horizons and without them would not this thesis arise.

Contents

List of Figures

List of Tables

Introduction

Moments, such as mean, variance, skewness, and kurtosis, are traditionally used to describe features of a univariate probability distribution. Hosking [55] unified and popularised an alternative approach to conventional moments which is now used both in descriptive statistics and statistical inference. This approach uses quantities called L-moments which are based on linear combinations of order statistics and from there also their name comes from. The main L-moments' advantage in comparison to the conventional moments is their existence of all orders under only a finite mean assumption. However, this limitation has also been removed. First, Mudholkar and Hutson [84] introduced analogs of Hosking's [55] L-moments, labeled LQ-moments, by replacing the expectation in the definition of L-moments by other measures of location of the distribution of the order statistics. Elamir and Seheult [38] then generalized L-moments by replacing the expectation by its trimmed version. Both the LQ- and TL-moments always exist which makes them an appropriate tool for analysis of probability distributions that do not have (finite) mean.

When describing a multivariate probability distribution, the situation is very similar. The mean vector, covariance, correlation, coskewness, and cokurtosis matrices with elements the covariance, correlation, coskewness, and cokurtosis are the characteristics usually used to summarize features of a multivariate probability distribution. However, central comoments are defined under finiteness of central moments. To avoid this drawback, Serfling and Xiao [97] generalized univariate L-moments and proposed multivariate L-moments in the matrix form with elements the L-comoments as analogues to the central comoments, however, without giving assumptions to finiteness of second and higher order central moments.

L-moments, being measures of shape of a probability distribution, may be used for summarizing data drawn from both the univariate and multivariate distributions.

They also play an important role in inference statistics. Traditional techniques for parameter estimation are the moments and maximum likelihood methods. L-moments are used for parameter estimation as an alternative to the traditional estimation methods, mainly in hydrology, climatology, and meteorology [70, 89], but also in socioeconomics, economics, and quality control [12, 47, 103, 110], or aerodynamics [41]. The estimators based upon L-moments are obtained in a similar way as it is in the moments method, which means the population L-moments are equated to their corresponding sample counterparts. Hosking [55] gave the parameter estimators of some common univariate distributions and prefers L-moments to the conventional moments, because they are more robust to the presence of outliers and, therefore, less subject to bias in parameter estimation in small samples. Several other studies have shown that the L-moments method in some cases outperforms also the maximum likelihood method [57, 59, 77].

L-moments have been at most often employed in regional frequency analysis (RFA), which yields more reliable estimates of high quantiles of extreme events than do the at-site approaches. RFA is based on pooling the data from the sites that have similar probability distributions, that overcomes the problem of small number of observations in single sites for estimating high quantiles and related unreliable estimates. Hence, in RFA quantile estimation is preceded by an important hypothesis testing step whether the sites have similar probability distributions apart from a site-specific scale factor. To check the homogeneity condition, L-moments are used as well. A univariate approach based on L-moments introduced by Hosking and Wallis [54] has been routinely used in areas such as hydrology, climatology, and meteorology, among others [24, 69, 70, 71, 86, 113]. After that Serfling and Xiao [97] proposed multivariate L-moments, it did not take long time and Chebana and Ouarda [22] and Chebana and Ouarda [23] generalized completely the Hosking and Wallis [54] RFA approach to the multivariate framework.

This dissertation thesis aims to summarize theory of both univariate and multivariate L-moments, to be a guide for statistical inference using them, and in particular, to propose some developments of L-moments' application as well. The

thesis is divided into six chapters. Two chapters, specifically the first and fifth, summarize theory of L-moments and their applications, however, it cannot be fully exhausted, and provide foundations for the entire work, while the rest are crucial as they include the own results. The first one introduces the theory of univariate L-moments: basic definitions, main properties and possible generalizations. The second chapter deals with derivation of parameters and quantiles of univariate distributions that are often used in modeling extreme events based upon L-moments and their generalizations. Specifically, they are the generalized Pareto and generalized extreme-value distributions. These estimation methods are compared with one another and they are also compared to the traditional maximum likelihood method via computer simulations. Since studies involving parameter estimation using L-moments have continued to be focused on point estimation, the next chapter presents asymptotic confidence intervals for parameters and quantiles estimators of the generalized Pareto and generalized-extreme value distributions based on L-moments and focuses on their comparison to another estimation techniques as well. The fourth chapter introduces generalization of univariate L-moments to the multivariate framework and presents their application in RFA, that is illustrated on the bivariate precipitation data. Methodology of trivariate L-moment homogeneity testing is presented in the fifth chapter, in which practical aspects of trivariate L-moment homogeneity testing are investigated for extreme precipitation events. In the last chapter, the multivariate L-moment homogeneity test is developed for spatially cross-correlated data using D-vine copulas and it is shown that the proposed modified L-moment homogeneity test faces the presence of cross-correlation better than the original one.

This dissertation thesis is based on the following articles published or prepared during my Ph.D. study:

- Šimková, T. and Picek, J. "A Comparison of L-, LQ-, TL-Moment and Maximum Likelihood High Quantile Estimates of the GPD and GEV Distribution". In: *Communications in Statistics - Simulation and Computation* 46.8 (2017), pp. 5991–6010. DOI: 10.1080/03610918.2016.1188206.

- Šimková, T. "Homogeneity Testing for Spatially Correlated Data in Multivariate Regional Frequency Analysis". In: *Water Resources Research* 53.8 (2017), pp. 7012–7028. DOI: 10.1002/2016WR020295.

- Šimková, T. "Statistical Inference Based on L-Moments". In: *Statistika: Statistics and Economy Journal* 97.1 (2017), pp. 44–58.

- Šimková, T. "L-Moment Homogeneity Test in Trivariate Regional Frequency Analysis of Extreme Precipitation Events". In: *Meteorological Applications* 25.1 (2018), pp. 11–22. DOI: 10.1002/met.1664.

- Šimková, T. "Asymptotic Confidence Inetrvals of Quantile Estimates for the GP and GEV Distributions". In preparation.

The research, on which these articles and thesis are based, was financially supported by the Project Klimatext CZ.1.07/2.3.00/20.0086, the Czech Science Foundation under the projects 14-18675S and 18-01137S, and the Student Grant Competition at the Technical University of Liberec under the projects 21116 and 21256, which I would like to thank.

All data analysis was carried out using R software [91]. Graphics were created using R and GeoGebra softwares [52, 91].

1 Univariate L-Moments

Contents

1.1 Introduction

Traditional tools for describing a univariate probability distribution of a random variable X include the mean

$$\mu = \mathrm{E}X,$$

central moments

$$\mu_r = \mathrm{E}(X-\mu)^r, r \geq 2,$$

and their scale-free versions

$$\alpha_r = \frac{\mu_r}{\sqrt{\mu_2^r}}, r \geq 3.$$

In particular, they are the variance

$$\mu_2 = \sigma^2 = \mathrm{E}(X - \mu)^2,$$

the coefficient of skewness

$$\alpha_3 = \frac{\mu_3}{\mu_2^{3/2}},$$

and the coefficient of kurtosis

$$\alpha_4 = \frac{\mu_4}{\mu_2^2}.$$

The shape of a probability distribution may be also described by the coefficient of variation

$$c_v = \frac{\sigma}{\mu}.$$

However, the central moments are not sufficient for analysis of distributions with heavier tails due to the moment assumptions of second and higher orders. To overcome this limitation, L-moments which are an alternative more robust system to the conventional moments may be used to describe a probability distribution under just first moment assumptions. Although univariate L-moments were formally introduced by Hosking [55], they had first appeared in the work of Sillitto [101] in the context of the quantile function's approximation by polynomials. They were not yet termed L-moments at that time. Hosking [55] pulled together the earlier findings mainly of Gini [46], Sillitto [101, 102], Downton [34], Chan [20], Konheim [65], Mallows [75], and Greenwood et al. [48], then assembled these into a unified whole that uses order statistics for analysis of a univariate probability distribution. L-moments can be interpreted as are conventional moments, because they are measures of location, scale and shape of the probability distribution. They also offer several advantages over them. First, the L-moment of any order exists if and only if the distribution has a finite mean. Moreover, if the mean of the distribution exists, then the L-moments uniquely define the distribution, because no two distributions are described by the same series of L-moments. Another useful property is the algebraic

boundedness of the L-moment ratios. The essential difference between conventional moments and L-moments is that the L-moments give smaller weights to the tails of the distribution. Sample L-moments give smaller weights to extreme observations, and this may lead also to more accurate parameter estimates based upon them.

The problem of use of L-moments occurs when the probability distribution has no or non-finite mean. For example, this happens for a Cauchy distribution. Therefore, some generalizations of L-moments were proposed. They are trimmed L-moments, termed TL-moments, and LQ-moments, which exist even if L-moments do not exist. Mudholkar and Hutson [84] introduced LQ-moments obtained by replacing the expectation in the definition of L-moments by quick estimators, such as medians, trimeans, or Gastwirth's location estimators. Elamir and Seheult [38] proposed robust modification of L-moments, so-called TL-moments, by assigning zero weights to the extremes.

The aim of this chapter is to review univariate L-moments, because they form together with multivariate L-moments the basis of the entire work. The chapter is organized as follows: The definitions and main features of the univariate L-moments are presented in Sections 1.2–1.3. The attention is then given to the L-moments method and asymptotic distribution of the sample L-moments, becuse the focus of Chapters 2 and 3 is on the quantile estimates and their confidence intervals of selected univariate distributions. In Section 1.6, univariate L-moments are generalized for a random variable with no or non-finite mean. This chapter relies mainly on the articles [117, 118].

1.2 Population L-Moments

Hosking [55] defined the population L-moment of the rth order as a linear combination of the order statistics $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ of a random sample of size n drawn from a univariate probability distribution of a random variable X with the cumulative distribution function F

$$\lambda_r = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} \mathrm{E} X_{r-k:r}, r = 1, 2, \ldots \tag{1.1}$$

When using the mean of the order statistic

$$\mathrm{E}X_{j:r} = \frac{r!}{(j-1)!(r-j)!} \int Q(u) u^{j-1} (1-u)^{r-j} \, \mathrm{d}u$$

and the $(r-1)$th shifted Legendre polynomial

$$P^*_{r-1}(u) = \sum_{i=0}^{r-1} (-1)^{r-i-1} \binom{r-1}{i} \binom{r+i-1}{i} u^i, r = 1, 2, \ldots, \tag{1.2}$$

formula (1.1) may be rewritten to a form

$$\lambda_r = \int_0^1 Q(u) P^*_{r-1}(u) \, \mathrm{d}u, r = 1, 2, \ldots, \tag{1.3}$$

that is useful particularly for computating L-moments of a specific probability distribution with the quantile function $Q(u) = \inf\{x \in \mathbb{R} \text{ s.t. } F(x) \geq u\}, 0 < u < 1$.

The first four L-moments are then in the form

$$\begin{aligned}
\lambda_1 &= \mathrm{E}\, X = \int_0^1 Q(u) \, \mathrm{d}u, \\
\lambda_2 &= \frac{1}{2} \mathrm{E}\, (X_{2:2} - X_{1:2}) = \int_0^1 Q(u)(2u-1) \, \mathrm{d}u, \\
\lambda_3 &= \frac{1}{3} \mathrm{E}\, (X_{3:3} - 2X_{2:3} + X_{1:3}) = \int_0^1 Q(u)(6u^2 - 6u + 1) \, \mathrm{d}u, \\
\lambda_4 &= \frac{1}{4} \mathrm{E}\, (X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4}) = \int_0^1 Q(u)(20u^3 - 30u^2 + 12u - 1) \, \mathrm{d}u.
\end{aligned}$$

The first L-moment λ_1 is just the ordinary mean. Serfling and Xiao [97] also presented the second and higher order L-moments in the covariance representation as

$$\lambda_r = \mathrm{cov}(X, P^*_{r-1}(F(X))), r \geq 2. \tag{1.4}$$

Note that L-moments are related to the notion of the probability-weighted moment (PWM). The PWM of a probability distribution with the cumulative distribution function F was proposed by Greenwood et al. [48] as

$$M_{p,r,s} = \mathrm{E}\, \{[X^p [F(X)]^r [1 - F(X)]^s\}. \tag{1.5}$$

A special case of PWM is obtained when $p = 1$ and $s = 0$ are substitued in Equation (1.5), that gives

$$\beta_r = \int_{-\infty}^{\infty} x[F(x)]^r \, \mathrm{d}F(x) = \int_0^1 x(u)u^r \mathrm{d}u, r = 0, 1, \ldots \tag{1.6}$$

L-moments were derived from PWMs, and thus L-moments may be written as certain linear combinations of PWMs as follows

$$\lambda_{r+1} = \sum_{k=0}^{r} (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} \beta_k, r = 0, 1, 2, \ldots,$$

where β_k are given by (1.6). All the procedures based on L-moments and PWMs are therefore equivalent, but the notion of L-moment is in practice more popular than PWM. The first four L-moments in terms of PWM are

$$\lambda_1 = \beta_0,$$
$$\lambda_2 = 2\beta_1 - \beta_0,$$
$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0,$$
$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0.$$

L-moments of the second and higher orders are standardized to be scale-free quantities

$$\tau = \frac{\lambda_2}{\lambda_1} \text{ (L-CV)},$$
$$\tau_r = \frac{\lambda_r}{\lambda_2}, r \geq 3 \text{ (L-moment ratios)}.$$

In contrast with the conventional moments, L-moment ratios are boundeded as follows

$$|\tau_r| < 1, r \geq 3.$$

Tighter bounds can be obtained for specific L-moment ratios. For example, Hosking and Wallis [54] have shown that the following holds

$$\frac{1}{4}(5\tau_3^2 - 1) \leq \tau_4 < 1.$$

Hosking [55] justified that the L-moments λ_1 and λ_2 and L-moment ratios τ_3 and τ_4 are measures of location, scale, skewness, and kurtosis, respectively, by considering linear combinations of the observations in a sample of data that are arranged in ascending order. Hence, they are useful quantities for summarizing a univariate probability distribution. Hosking [55] presented the first four L-moments of some common univariate distributions, including the uniform, exponential, normal, logistic, Gumbel, generalized Pareto, and generalized extreme-value distributions, which may be simply derived using Equation (1.3). See Table 1.1 for the selected ones.

1.3 Sample L-Moments

Population L-moments are in practice estimated from a finite observed random sample drawn from an unknown univariate probability distribution. The rth sample L-moment, being an L-estimator of the population L-moment λ_r, was defined by Hosking [55] as a linear combination of the order sample $x_{1:n} \leq x_{2:n} \leq \cdots \leq x_{n:n}$ of size n

$$l_r = \binom{n}{r}^{-1} \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} \cdots \sum r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} x_{i_{r-k}:n}, r = 1, 2, \ldots \tag{1.7}$$

Hosking [55] showed that l_r given by (1.7) may be expressed in the simpler form

$$l_{r+1} = \sum_{k=0}^{r-1} (-1)^{r-k-1} \binom{r-1}{k} \binom{r+k-1}{k} b_k, r = 0, 1, \ldots, n-1, \tag{1.8}$$

where

$$b_k = n^{-1} \binom{n-1}{r} \sum_{j=1}^{n} \binom{j-1}{r} x_{j:n}. \tag{1.9}$$

Note that b_k given by Equation (1.9) are unbiased estimators of β_k defined by Equation (1.6).

Theorem 1. *The sample L-moment l_r given by (1.8) is an unbiased estimator of the population L-moment $\lambda_r, r = 1, 2, \ldots$*

Proof. See Hosking [55], p. 114. □

Table 1.1: L-moments of selected univariate probability distributions

Distribution	Quantile function $Q(u)$	L-moments
Uniform $U(\alpha, \beta)$ $\alpha, \beta \in \mathbb{R}$, $\alpha < \beta$	$\alpha + (\beta - \alpha)u$	$\lambda_1 = \frac{1}{2}(\alpha + \beta)$ $\lambda_2 = \frac{1}{6}(\beta - \alpha)$ $\lambda_3 = 0$ $\lambda_4 = 0$
Exponential $Exp(\xi, \sigma)$ $\xi \in \mathbb{R}, \sigma > 0$	$\xi - \sigma \log(1 - u)$	$\lambda_1 = \xi + \sigma$ $\lambda_2 = \frac{1}{2}\sigma$ $\lambda_3 = \frac{1}{6}\sigma$ $\lambda_4 = \frac{1}{12}\sigma$
Normal $\mathcal{N}(\mu, \sigma)$ $\mu \in \mathbb{R}, \sigma > 0$	no explicit form, approximation used $Q(u) \approx \mu + 5.063\sigma \cdot$ $\cdot[u^{0.135} - (1 - u)^{0.135}]$	$\lambda_1 = \mu$ $\lambda_2 = \frac{\sigma}{\sqrt{\pi}}$ $\lambda_3 = 0$ $\lambda_4 = 0.0702\sigma$
Logistic $Logi(\xi, \sigma)$ $\xi \in \mathbb{R}, \sigma > 0$	$\xi + \sigma \log\left(\frac{1-u}{u}\right)$	$\lambda_1 = \xi$ $\lambda_2 = \sigma$ $\lambda_3 = 0$ $\lambda_4 = \frac{1}{6}\sigma$
Generalized Pareto $GP(\xi, \sigma, k)$ $\xi \in \mathbb{R}, \sigma > 0, k \in \mathbb{R}$	$\xi + \frac{\sigma}{k}[1 - (1 - u)^k]$	$\lambda_1 = \xi + \frac{\sigma}{k+1}$ $\lambda_2 = \frac{\sigma}{(k+1)(k+2)}$ $\lambda_3 = \frac{\sigma(1-k)}{(k+1)(k+2)(k+3)}$ $\lambda_4 = \frac{\sigma(k-1)(k-2)}{(k+1)(k+2)(k+3)(k+4)}$
Generalized extreme-value $GEV(\xi, \sigma, k)$ $\xi \in \mathbb{R}, \sigma > 0, k \in \mathbb{R}$	$\xi + \frac{\sigma}{k}[1 - (-\log u)^k]$	$\lambda_1 = \xi + \frac{\sigma}{k}[1 - \Gamma(k+1)]$ $\lambda_2 = \frac{\sigma}{k}\Gamma(k+1)(1 - 2^{-k})$ $\lambda_3 = \frac{\sigma}{k}\Gamma(k+1)(2 \cdot 3^{-k} +$ $+ 3 \cdot 2^{-k} - 1)$ $\lambda_4 = \frac{\sigma}{k}\Gamma(k+1)(-5 \cdot 4^{-k} +$ $+ 10 \cdot 3^{-k} - 6 \cdot 2^{-k} + 1)$

$\Gamma(\cdot)$ denotes the gamma function.

Wang [120] proposed another L-estimator of λ_r in a form

$$l_r^D = r^{-1}\binom{n}{r}^{-1} \sum_{k=1}^{n} \sum_{j=0}^{r-1} (-1)^j \binom{r-1}{j}\binom{k-1}{r-j-1}\binom{n-k}{j} x_{k:n}, \tag{1.10}$$

and called it "direct sample estimator" of the L-moment. Wang [120] stated that both l_r given by Equation (1.8) and l_r^D given by Equation (1.10) give the same

numerical values. However, he did not prove that these quantities are equal. Hosking and Balakrishnan [56] proved that if two L-statistics have the same mean, then they are identical. From this results that l_r given by Equation (1.8) and l_r^D given by Equation (1.10) are identical.

Thus, the first four sample L-moments are in the form

$$l_1 = \frac{1}{n}\sum_{k=1}^{n} x_{k:n},$$

$$l_2 = \frac{1}{n(n-1)}\sum_{k=1}^{n}(2k-n-1)x_{k:n},$$

$$l_3 = \frac{1}{n(n-1)(n-2)}\sum_{k=1}^{n}(6k^2-6k-6nk+n^2+3n+3)x_{k:n},$$

$$l_4 = \frac{1}{n(n-1)(n-2)(n-3)}\sum_{k=1}^{n}(20k^3-30k^2n-30k^2+12kn^2+30kn+22k- \\ -n^3-6n^2-11n-6)x_{k:n}.$$

The first sample L-moment, l_1, is just the sample mean and the second sample L-moment, l_2, is half of the Gini's mean difference statistic [46, 109].

The sample L-CV and L-moment ratios are obtained analogously to their population counterparts

$$t = \frac{l_2}{l_1},$$

$$t_r = \frac{l_r}{l_2}, r \geq 3,$$

respectively. Although the estimators t and t_r are not unbiased, Hosking [55] and Hosking and Wallis [54] showed that their biases are negligible for arbitrary sizes. Hosking and Wallis [54] estimated bias for small samples and several selected distributions, such as the generalized extreme-value and kappa distributions, from the simulations. For large samples, the biases of L-moment ratios may be evaluated using asymptotic theory. For example, Hosking [55] calculated the asymptotic bias of t_4 for the normal distribution with L-kurtosis τ_4 as $0.03n^{-1}$, where n is the sample size.

Observed data may alternatively be summarized by the sample L-mean, l_1; L-scale, l_2; L-skewness, t_3; and L-kurtosis, t_4. Hosking [55] recommends using sample

L-moments over conventional moments, because they are linear combinations of data and therefore are less prone to sampling variability or errors in data. That means they may result in better and more robust estimates of the distribution's characteristics and parameters.

1.4 Method of L-Moments

Following the same idea as in the case of the moments method, L-moments provide parameter estimators. Let X be a random variable with a probability density function $f(x;\theta_1,\theta_2,\ldots,\theta_k)$, which depends on some unknown parameter vector $\boldsymbol{\theta}=(\theta_1,\theta_2,\ldots,\theta_k)^T$. It may be estimated by solving the system of equations which arises from matching the first k population L-moments to their corresponding sample quantities, i.e.,

$$\lambda_i = l_i, \quad i=1,2,\ldots,k.$$

Hosking [55] and Hosking and Wallis [54] presented L-moments parameter estimators of some commonly used probability distributions. Parameter estimators of selected univariate distributions are shown in Table 1.2.

Table 1.2: Parameter estimators of selected univariate probability distributions

Distribution	Parameter estimators
Uniform	$\widehat{\alpha}=l_1-3l_2, \widehat{\beta}=2l_1-\widehat{\alpha}$
Exponential	$\widehat{\sigma}=2l_2, \widehat{\xi}=l_1-\widehat{\sigma}$
Normal	$\widehat{\mu}=l_1, \widehat{\sigma}=\sqrt{\pi}l_2$
Logistic	$\widehat{\xi}=l_1, \widehat{\sigma}=l_2$
Generalized Pareto	$\widehat{k}=\frac{1-3t_3}{1+t_3}, \widehat{\sigma}=l_2(\widehat{k}+1)(\widehat{k}+2), \widehat{\xi}=l_1-\frac{\widehat{\sigma}}{\widehat{k}+1}$
Generalized extreme-value	$\widehat{k}\approx 7.859z+2.9554z^2$, where $z=\frac{2}{3+t_3}-\log_3 2$, $\widehat{\sigma}=\frac{l_2}{\Gamma(\widehat{k})(1-2^{-\widehat{k}})}, \widehat{\xi}=l_1-\frac{\widehat{\sigma}}{\widehat{k}}[1-\Gamma(\widehat{k}+1)]$

$\Gamma(\cdot)$ denotes the gamma function.

1.5 Asymptotic Properties of L-Moments

It is not easy to derive exact distributions of L-moments. Therefore, Hosking [55] used the asymptotic theory for linear combinations of order statistics developed by Chernoff, Gastwirth and Johns [25], Moore [83] and Stigler [108], and applied it to sample L-moments and L-moment ratios to prove that they are asymptotically normal.

Theorem 2. *Let X be a real-valued random variable with a cumulative distribution function F, L-moments λ_r, and finite variance. Let $l_r, r = 1, 2, \ldots, m$, be sample L-moments calculated from a random sample of size n drawn from the distribution of X. Then, as $n \to \infty$:*

1. *the vector $\sqrt{n}[(l_1 - \lambda_1), (l_2 - \lambda_2), \ldots, (l_m - \lambda_m)]^T$ converges in distribution to the multivariate normal distribution $\mathcal{N}_m(\mathbf{0}, \mathbf{\Lambda})$, where $\mathbf{0}$ is the m-dimensional null vector and the elements $\Lambda_{rs}, r, s = 1, 2, \ldots, m$, of the covariance matrix $\mathbf{\Lambda}$ are given by*

$$\Lambda_{rs} = \int\int_{x<y} \{P^*_{r-1}[F(x)]P^*_{s-1}[F(y)] + P^*_{s-1}[F(x)]P^*_{r-1}[F(y)]\} \cdot \cdot F(x)[1 - F(y)]\,\mathrm{d}x\,\mathrm{d}y, \tag{1.11}$$

2. *the vector $\sqrt{n}[(l_1 - \lambda_1), (l_2 - \lambda_2), (t_3 - \tau_3), (t_4 - \tau_4), \ldots, (t_m - \tau_m)]^T$ converges in distribution to the multivariate normal distribution $\mathcal{N}_m(\mathbf{0}, \mathbf{T})$, where $\mathbf{0}$ is the m-dimensional null vector and the elements $T_{rs}, r, s = 1, 2, \ldots, m$, of the covariance matrix $\mathbf{T}$ are given by*

$$T_{rs} = \begin{cases} \Lambda_{rs} & \text{if } r \le 2, s \le 2, \\ (\Lambda_{rs} - \tau_r\Lambda_{2s})/\lambda_2 & \text{if } r \ge 3, s \le 2, \\ (\Lambda_{rs} - \tau_r\Lambda_{2s} - \tau_s\Lambda_{2r} + \tau_r\tau_s\Lambda_{22})/\lambda_2^2 & \text{if } r \ge 3, s \ge 3. \end{cases}$$

Proof. See Hosking [55], p. 116. □

When the random variable X is continuous and its quantile function is differentiable, then the elements Λ_{rs} of the covariance matrix $\mathbf{\Lambda}$ given by (1.11) may be rewritten to the form

$$\Lambda_{rs} = \int\int_{0<u<v<1} [P^*_{r-1}(u)P^*_{s-1}(v) + P^*_{s-1}(u)P^*_{r-1}(v)]u(1-v)Q'(u)Q'(v)\,\mathrm{d}u\,\mathrm{d}v. \tag{1.12}$$

1.6 Generalizations of L-Moments

1.6.1 LQ-Moments

Mudholkar and Hutson [84] proposed LQ-moments as a more robust analog to L-moments, which always exist even if a mean of a random variable does not exist. Hence, they are also already applicable for heavy-tail distributions. LQ-moments are obtained by replacing the expectation in Equation (1.1) by a quick measure of the location of the distribution of the order statistic $X_{r-k:r}$

$$\zeta_r = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} \tau_{p,\alpha}(X_{r-k:r}), \quad r = 1, 2, \ldots, \tag{1.13}$$

where $0 \leq p \leq \frac{1}{2}, 0 \leq \alpha \leq \frac{1}{2}$, and

$$\tau_{p,\alpha}(X_{r-k:r}) = pQ_{X_{r-k:r}}(\alpha) + (1-2p)Q_{X_{r-k:r}}(1/2) + pQ_{X_{r-k:r}}(1-\alpha) =$$
$$= pQ[B^{-1}_{r-k:r}(\alpha)] + (1-2p)Q[B^{-1}_{r-k:r}(1/2)] + pQ[B^{-1}_{r-k:r}(1-\alpha)],$$

Q is the quantile function of a random variable X and $B^{-1}_{r-k:r}(\alpha)$ is the α-quantile of a beta distribution function with parameters $(r-k)$ and $(k+1)$. The skewness and kurtosis measures $\eta_3 = \zeta_3/\zeta_2$ and $\eta_4 = \zeta_4/\zeta_2$ are called LQ-skewness and LQ-kurtosis, respectively.

Let us have order sample $x_{1:n} \leq x_{2:n} \leq \cdots \leq x_{n:n}$ of size n. Then the rth sample LQ-moments is given by

$$\widehat{\zeta}_r = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} \hat{\tau}_{p,\alpha}(x_{r-k:r}),\ r = 1, 2, \ldots, n,$$

where $0 \leq p \leq \frac{1}{2}, 0 \leq \alpha \leq \frac{1}{2}$, and

$$\widehat{\tau}_{p,\alpha}(x_{r-k:r}) = p\widehat{Q}[B^{-1}_{r-k:r}(\alpha)] + (1-2p)\widehat{Q}[B^{-1}_{r-k:r}(1/2)] + p\widehat{Q}[B^{-1}_{r-k:r}(1-\alpha)]$$

is the quick estimator of location of the distribution of the order statistic $X_{r-k:r}$, where $\widehat{Q}(u) = (1-\varepsilon)x_{\lfloor (n+1)u \rfloor:n} + \varepsilon x_{(\lfloor (n+1)u \rfloor+1):n}$ is the quantile estimator of the quantile function Q of a random variable X, $\varepsilon = (n+1)u - \lfloor (n+1)u \rfloor$ ($\lfloor \cdot \rfloor$ denotes the floor function).

Using large sample theory of the linear functions of order statistics, Mudholkar and Hutson [84] showed the asymptotic normality of the sample LQ-moments. See Mudholkar and Hutson [84, p. 201] for details about the covariance matrix.

Solving the system

$$\zeta_i = \widehat{\zeta}_i, \quad i = 1, 2, \ldots, k,$$

where k is the number of unknown parameters of a univariate probability distribution with the probability density function $f(x; \theta_1, \theta_2, \ldots, \theta_k)$, the parameter estimators based on LQ-moments are obtained.

1.6.2 TL-Moments

Elamir and Seheult [38] introduced trimmed L-moments as a more robust generalization of L-moments. Equally as in the case of LQ-moments, TL-moments always exist. On the top of it in comparison to L-moments, TL-moments are more resistant to outliers due to assigning zero weights to extreme observations. The rth population TL-moment with trimmed the t_1 smallest and the t_2 largest ordered statistics is defined as follows

$$\lambda_r^{(t_1,t_2)} = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} \mathrm{E}X_{r+t_1-k:r+t_1+t_2}, \quad r = 1, 2, \ldots$$

When $t_1 = t_2 = 0$, usual L-moments defined by Hosking [55] are obtained. Here, the focus is only on the symmetric case $t_1 = t_2 = t$

$$\lambda_r^{(t)} = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} \mathrm{E}X_{r+t-k:r+2t}, \quad r = 1, 2, \ldots \tag{1.14}$$

The formula in (1.14) may be rewritten to the form

$$\begin{aligned} \lambda_r^{(t)} = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} \frac{(r+2t)!}{(r+t-k-1)!(t+k)!} \cdot \\ \cdot \int_0^1 Q(u) u^{r+t-k-1} (1-u)^{t+k} \, \mathrm{d}u. \end{aligned} \tag{1.15}$$

Analogously to L-moments, the first two TL-moments $\lambda_1^{(t)}$ and $\lambda_2^{(t)}$ are measures of location and scale, the TL-skewness $\tau_3^{(t)} = \lambda_3^{(t)}/\lambda_2^{(t)}$ is a measure of skewness and the TL-kurtosis $\tau_4^{(t)} = \lambda_4^{(t)}/\lambda_2^{(t)}$ is a measure of kurtosis.

An unbiased estimator of the rth population TL-moment for symmetric case $\lambda_r^{(t)}$ is a linear combination of the ordered sample $x_{t+1:n} \leq x_{t+2:n} \leq \cdots \leq x_{n-t:n}$ of size $n - 2t$ in the form

$$l_r^{(t)} = r^{-1} \sum_{i=t+1}^{n-t} \left[\frac{\sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} \binom{i-1}{r+t-1-k} \binom{n-i}{t+k}}{\binom{n}{r+2t}} \right] x_{i:n},\ r = 1, 2, \ldots, n.$$

In the same way as in the L- and LQ-moments methods, the estimators of k unknown parameters of a univariate probability distribution with the probability density function $f(x; \theta_1, \theta_2, \ldots, \theta_k)$ based upon TL-moments are the solutions of the system

$$\lambda_i^{(t)} = l_i^{(t)}, \quad i = 1, 2, \ldots, k.$$

2

Comparison of L-, LQ-, TL-Moments and Maximum Likelihood Quantile Estimates of the GP and GEV Distributions

Contents

2.1 Introduction

Knowledge of the probability distribution, which specifies how frequently the possible values of a random variable occur, is of great importance in many fields. The focus is often on estimating high quantiles, and, hence, the parameter estimates

of a probability distribution are required. Usually, the maximum likelihood and moments methods are used. After that Hosking [55] introduced an analogy to the conventional moments, called L-moments, the method based on them has been used as the convenient alternative to these methods in the last years. Hosking [55] and Hosking and Wallis [54] presented also parameters estimators derived using the L-moments for many common distributions, including the generalized Pareto (GP) and generalized extreme-value (GEV) distributions. Due to the fact that the maximum likelihood method is based on large sample theory, the parameters estimates obtained by this technique for small samples may be unreliable. Therefore, Hosking, Wallis and Wood [59] and Hosking and Wallis [57] focused on the properties of estimates of the GP and GEV distributions based upon L-moments for small samples via computer simulations. They concluded that the L-moments method is the preferable one when the sample size is small to moderate and the distribution has heavier tails (i.e., $n < 100$ and $k < 0$). On the top of it, the estimators based upon L-moments are easily computable for many distributions in comparison to the maximum likelihood method.

Shortly after, L-moments were generalized by Mudholkar and Hutson [84] and Elamir and Seheult [38], and other alternative estimation techniques based on these statistics appeared. The LQ-moments parameters estimators have been already derived for probability distributions such as a GEV [84], extreme-value type I [98], kappa [99], generalized logistic [95], and exponentiated Pareto [5] distributions. Elamir and Seheult [38] presented as an example the TL-moments based parameters estimators of a normal, logistic, Cauchy, and exponential distributions. Ariff [4] derived the symmetric TL-moments for the other well-known distributions, such as generalized logistic, extreme-value type I, GEV, and GP distributions. The TL-moments method has also been applied to the generalized lambda [6], Dagum [100], exponentiated Pareto [5], and exponentiated generalized extreme-value [37] distributions.

The GP and GEV distributions have been used in modeling of extreme events in hydrology, climatology, meteorology, and other areas [42, 53, 64, 70, 88]. It is well

known that it does not exist finite mean for certain values of the shape parameter of these two distributions. Thus, neither moments nor L-moments can be well used exactly in analysis of extreme events in such cases. According to the definition of LQ- and TL-moments, it seems that reliable parameters estimates employing the LQ- and TL-moments methods would be obtained.

The aim of this chapter is to compare L-, LQ-, and symmetric TL-moments and maximum likelihood estimators of high quantiles of the GP and GEV distributions depending on various factors, such as sample size, probability, and values of the shape parameter. Using the formulas (1.3), (1.13), and (1.15), the population L-moments, LQ-moments, and symmetric TL-moments with the trimming parameters $t = 1$ and 2 up to the fourth order are computed by standard methods, including integration by parts. Matching the first two population L-, LQ-, and TL-moments and L-, LQ-, and TL-skewness with their corresponding sample counterparts, the estimators of three unknown parameters of the GP and GEV distributions are derived. The well-known maximum likelihood-based estimators are also given. The chapter is organized as follows: In Sections 2.2 and 2.3 the first four population L-moments, LQ-moments and symmetric TL-moments, and estimators of the GP and GEV distributions parameters based on these moments are presented. In Section 2.4, a simulation is performed. Applying the software R [91], parameters of the GP and GEV distributions are estimated by four methods: L-moments (LM), LQ-moments (LQM) with optimal values of parameters α and p, symmetric TL-moments with the trimming parameters $t = 1$ (TLM1) and $t = 2$ (TLM2), and maximum likelihood (ML) methods. The quantile estimates for probabilities $90\%, 99\%$, and 99.9% are compared with one another according to their sample mean squared errors. The results of the simulation study and recommendations are summarized as well. Conclusions are presented in the last section. This chapter relies on the article [118].

2.2 Parameters Estimators for the GP Distribution

The cumulative distribution function of the GP distribution with parameters $\xi \in \mathbb{R}$ (location), $\sigma > 0$ (scale), and $k \in \mathbb{R}$ (shape) is

$$F(x) = \begin{cases} 1 - \left[1 - \frac{k(x-\xi)}{\sigma}\right]^{\frac{1}{k}}, & k \neq 0, \\ 1 - \exp\left(-\frac{x-\xi}{\sigma}\right), & k = 0, \end{cases} \tag{2.1}$$

$\xi \leq x \leq \xi + \frac{\alpha}{k}$ if $k > 0$, $\xi \leq x < \infty$ if $k \leq 0$. L-moments of all orders exist when $k > -1$, while LQ- and TL-moments of all orders exist for an arbitrary value of k.

2.2.1 L-Moments Method

Theorem 3. *Let $X \sim GP(\xi, \sigma, k)$. Then, if $k > -1, k \neq 0$,*

$$\begin{aligned} \lambda_1 &= \xi + \frac{\sigma}{k+1}, \\ \lambda_2 &= \frac{\sigma}{(k+1)(k+2)}, \\ \lambda_3 &= \frac{\sigma(1-k)}{(k+1)(k+2)(k+3)}, \\ \lambda_4 &= \frac{\sigma(k-1)(k-2)}{(k+1)(k+2)(k+3)(k+4)}, \end{aligned}$$

and

$$\begin{aligned} \widehat{k} &= \frac{1-3t_3}{1+t_3}, \\ \widehat{\sigma} &= l_2(\hat{k}+1)(\hat{k}+2), \\ \widehat{\xi} &= l_1 - \frac{\hat{\sigma}}{\hat{k}+1}. \end{aligned}$$

Proof. Substituting the quantile function

$$Q(u) = \xi + \frac{\sigma}{k}[1 - (1-u)^k], k \neq 0, \tag{2.2}$$

into the formula (1.3) and integrating by parts gives the expressions of the first four population L-moments. The L-moments parameters estimators are obtained

by solving the following system of equations

$$\begin{aligned}\xi + \frac{\sigma}{k+1} &= l_1,\\ \frac{\sigma}{(k+1)(k+2)} &= l_2,\\ \frac{1-k}{k+3} &= t_3,\end{aligned}$$

which arises from matching the L-mean, λ_1, L-scale, λ_2, and the L-skewness, τ_3, to their corresponding counterparts l_1, l_2, and t_3. □

2.2.2 LQ-Moments Method

Theorem 4. *Let $X \sim GP(\xi, \sigma, k)$. Then, if $k \neq 0$,*

$$\zeta_1 = \xi + \frac{\sigma}{k} - \frac{\sigma}{k}\{p[1 - B_{1:1}^{-1}(\alpha)]^k + (1-2p)[1 - B_{1:1}^{-1}(1/2)]^k + p[1 - B_{1:1}^{-1}(1-\alpha)]^k\},$$

$$\begin{aligned}\zeta_2 = \frac{\sigma}{2k}\{&-p[1 - B_{2:2}^{-1}(\alpha)]^k - (1-2p)[1 - B_{2:2}^{-1}(1/2)]^k - p[1 - B_{2:2}^{-1}(1-\alpha)]^k +\\ &+p[1 - B_{1:2}^{-1}(\alpha)]^k + (1-2p)[1 - B_{1:2}^{-1}(1/2)]^k + p[1 - B_{1:2}^{-1}(1-\alpha)]^k\},\end{aligned}$$

$$\begin{aligned}\zeta_3 = \frac{\sigma}{3k}\{&-p[1 - B_{3:3}^{-1}(\alpha)]^k - (1-2p)[1 - B_{3:3}^{-1}(1/2)]^k - p[1 - B_{3:3}^{-1}(1-\alpha)]^k +\\ &+2p[1 - B_{2:3}^{-1}(\alpha)]^k + 2(1-2p)[1 - B_{2:3}^{-1}(1/2)]^k + 2p[1 - B_{2:3}^{-1}(1-\alpha)]^k -\\ &-p[1 - B_{1:3}^{-1}(\alpha)]^k - (1-2p)[1 - B_{1:3}^{-1}(1/2)]^k - p[1 - B_{1:3}^{-1}(1-\alpha)]^k\},\end{aligned}$$

$$\begin{aligned}\zeta_4 = \frac{\sigma}{4k}\{&-p[1 - B_{4:4}^{-1}(\alpha)]^k - (1-2p)[1 - B_{4:4}^{-1}(1/2)]^k - p[1 - B_{4:4}^{-1}(1-\alpha)]^k +\\ &+3p[1 - B_{3:4}^{-1}(\alpha)]^k + 3(1-2p)[1 - B_{3:4}^{-1}(1/2)]^k + 3p[1 - B_{3:4}^{-1}(1-\alpha)]^k -\\ &-3p[1 - B_{2:4}^{-1}(\alpha)]^k - 3(1-2p)[1 - B_{2:4}^{-1}(1/2)]^k - 3p[1 - B_{2:4}^{-1}(1-\alpha)]^k +\\ &+p[1 - B_{1:4}^{-1}(\alpha)]^k + (1-2p)[1 - B_{1:4}^{-1}(1/2)]^k + p[1 - B_{1:4}^{-1}(1-\alpha)]^k\},\end{aligned}$$

and

$$\widehat{\eta}_3 = \frac{2}{3}\{-p[1-B_{3:3}^{-1}(\alpha)]^k - (1-2p)[1-B_{3:3}^{-1}(1/2)]^k - p[1-B_{3:3}^{-1}(1-\alpha)]^k +$$
$$+2p[1-B_{2:3}^{-1}(\alpha)]^k + 2(1-2p)[1-B_{2:3}^{-1}(1/2)]^k + 2p[1-B_{2:3}^{-1}(1-\alpha)]^k -$$
$$-p[1-B_{1:3}^{-1}(\alpha)]^k - (1-2p)[1-B_{1:3}^{-1}(1/2)]^k - p[1-B_{1:3}^{-1}(1-\alpha)]^k\}/$$
$$/\{-p[1-B_{2:2}^{-1}(\alpha)]^k - (1-2p)[1-B_{2:2}^{-1}(1/2)]^k - p[1-B_{2:2}^{-1}(1-\alpha)]^k +$$
$$+p[1-B_{1:2}^{-1}(\alpha)]^k + (1-2p)[1-B_{1:2}^{-1}(1/2)]^k + p[1-B_{1:2}^{-1}(1-\alpha)]^k\},$$
$$\widehat{\sigma} = 2\widehat{\zeta}_2\widehat{k}/\{-p[1-B_{2:2}^{-1}(\alpha)]^{\widehat{k}} - (1-2p)[1-B_{2:2}^{-1}(1/2)]^{\widehat{k}} - p[1-B_{2:2}^{-1}(1-\alpha)]^{\widehat{k}} +$$
$$+p[1-B_{1:2}^{-1}(\alpha)]^{\widehat{k}} + (1-2p)[1-B_{1:2}^{-1}(1/2)]^{\widehat{k}} + p[1-B_{1:2}^{-1}(1-\alpha)]^{\widehat{k}}\},$$
$$\widehat{\xi} = \widehat{\zeta}_1 - \frac{\widehat{\sigma}}{\widehat{k}} + \frac{\widehat{\sigma}}{4\widehat{k}\{p[1-B_{1:1}^{-1}(\alpha)]^{\widehat{k}} + (1-2p)[1-B_{1:1}^{-1}(1/2)]^{\widehat{k}} + p[1-B_{1:1}^{-1}(1-\alpha)]^{\widehat{k}}\}}.$$

Proof. Substituting the quantile function given by (2.2) into the formula (1.13) gives the expressions of the first four population LQ-moments. The LQ-moments parameters estimators are obtained by solving the following system of equations

$$\begin{aligned} \zeta_1 &= \widehat{\zeta}_1, \\ \zeta_2 &= \widehat{\zeta}_2, \\ \eta_3 &= \widehat{\eta}_3. \end{aligned} \tag{2.3}$$

First the shape parameter k must be estimated numerically by solving the equation (2.3), then the estimators of ξ and σ are easily obtained. □

2.2.3 TL-Moments Method

Theorem 5. *Let $X \sim GP(\xi, \sigma, k)$. Then, if $k \neq 0$,*

$$\lambda_1^{(1)} = \xi + \frac{\sigma(k+5)}{(k+2)(k+3)},$$
$$\lambda_2^{(1)} = \frac{6\sigma}{(k+2)(k+3)(k+4)},$$
$$\lambda_3^{(1)} = \frac{20\sigma(1-k)}{3(k+2)(k+3)(k+4)(k+5)},$$
$$\lambda_4^{(1)} = \frac{15\sigma(k-1)(k-2)}{2(k+2)(k+3)(k+4)(k+5)(k+6)},$$

and

$$\widehat{k} = \frac{10 - 45t_3^{(1)}}{10 + 9t_3^{(1)}},$$
$$\widehat{\sigma} = \frac{1}{6} l_2^{(1)} (\widehat{k}+2)(\widehat{k}+3)(\widehat{k}+4),$$
$$\widehat{\xi} = l_1^{(1)} - \frac{\widehat{\sigma}(\widehat{k}+5)}{(\widehat{k}+2)(\widehat{k}+3)}.$$

If $k \neq 0$, *then*

$$\lambda_1^{(2)} = \xi + \frac{\sigma(k^2+12k+47)}{(k+3)(k+4)(k+5)},$$
$$\lambda_2^{(2)} = \frac{60\sigma}{(k+3)(k+4)(k+5)(k+6)},$$
$$\lambda_3^{(2)} = \frac{70\sigma(1-k)}{(k+3)(k+4)(k+5)(k+6)(k+7)},$$
$$\lambda_4^{(2)} = \frac{84\sigma(k-1)(k-2)}{(k+3)(k+4)(k+5)(k+6)(k+7)(k+8)},$$

and

$$\widehat{k} = \frac{7 - 42t_3^{(2)}}{7 + 6t_3^{(2)}},$$
$$\widehat{\sigma} = \frac{1}{60} l_2^{(2)} (\widehat{k}+3)(\widehat{k}+4)(\widehat{k}+5)(\widehat{k}+6),$$
$$\widehat{\xi} = l_1^{(2)} - \frac{\widehat{\sigma}(\widehat{k}^2+12\widehat{k}+47)}{(\widehat{k}+3)(\widehat{k}+4)(\widehat{k}+5)}.$$

Proof. Substituting the quantile function given by (2.2) into the formula (1.15) for $t = 1$ and 2, and integrating by parts gives the expressions of the first four population symmetric TL-moments. The TL-moments parameters estimators for $t = 1$ and 2 are obtained by solving the following systems of equations

$$\xi + \frac{\sigma(k+5)}{(k+2)(k+3)} = l_1^{(1)},$$
$$\frac{6\sigma}{(k+2)(k+3)(k+4)} = l_2^{(1)},$$
$$\frac{10(1-k)}{9(k+5)} = t_3^{(1)},$$

and

$$\xi + \frac{\sigma(k^2+12k+47)}{(k+3)(k+4)(k+5)} = l_1^{(2)},$$
$$\frac{60\sigma}{(k+3)(k+4)(k+5)(k+6)} = l_2^{(2)},$$
$$\frac{7(1-k)}{6(k+7)} = t_3^{(2)},$$

respectively. □

2.2.4 Maximum Likelihood Method

The log-likelihood function for a sample $\boldsymbol{x} = \{x_1, x_2, \ldots, x_n\}$ is

$$\log L(\boldsymbol{x}; \xi, \sigma, k) = -n \log \sigma + \frac{1-k}{k} \sum_{i=1}^{n} \log \left[1 - \frac{k(x_i - \xi)}{\sigma}\right]. \tag{2.4}$$

However, the location parameter ξ cannot be obtained by differentiating (2.4), because the log-likelihood function is not bounded with respect to ξ. Therefore, the minimum value of the sample data is used as its estimator [104]. The estimators of σ and k are achieved numerically by solving the equations

$$\sum_{i=1}^{n} \frac{x_i - \xi}{\sigma - k(x_i - \xi)} = \frac{n}{1-k},$$
$$\sum_{i=1}^{n} \log \left[1 - \frac{k(x_i - \xi)}{\sigma}\right] = -nk.$$

2.3 Parameters Estimators for the GEV Distribution

The cumulative distribution function of the GEV distribution with parameters $\xi \in \mathbb{R}$ (location), $\sigma > 0$ (scale) and $k \in \mathbb{R}$ (shape) is

$$F(x) = \begin{cases} \exp\left\{-\left[1 - \frac{k(x-\xi)}{\sigma}\right]^{\frac{1}{k}}\right\}, & k \neq 0, \\ \exp\left[\exp\left(-\frac{x-\xi}{\sigma}\right)\right], & k = 0, \end{cases} \tag{2.5}$$

$-\infty < x \leq \xi + \frac{\alpha}{k}$ if $k > 0$, $-\infty < x < \infty$ if $k = 0$, $\xi + \frac{\alpha}{k} \leq x < \infty$ if $k < 0$. L-moments of all orders exist when $k > -1$, while LQ- and TL-moments of all orders exist for an arbitrary value of k.

2.3.1 L-Moments Method

Theorem 6. *Let $X \sim GEV(\xi, \sigma, k)$. Then, if $k > -1, k \neq 0$,*

$$\begin{aligned}
\lambda_1 &= \xi + \frac{\sigma}{k}[1 - \Gamma(k+1)], \\
\lambda_2 &= \frac{\sigma}{k}\Gamma(k+1)(1 - 2^{-k}), \\
\lambda_3 &= \frac{\sigma}{k}\Gamma(k+1)(-2 \cdot 3^{-k} + 3 \cdot 2^{-k} - 1), \\
\lambda_4 &= \frac{\sigma}{k}\Gamma(k+1)(-5 \cdot 4^{-k} + 10 \cdot 3^{-k} - 6 \cdot 2^{-k} + 1),
\end{aligned}$$

and

$$\begin{aligned}
t_3 &= \frac{-2 \cdot 3^{-k} + 3 \cdot 2^{-k} - 1}{1 - 2^{-k}}, \\
\widehat{\sigma} &= \frac{l_2 \widehat{k}}{\Gamma(\widehat{k}+1)(1 - 2^{-\widehat{k}})}, \\
\widehat{\xi} &= l_1 - \frac{\widehat{\sigma}}{\widehat{k}}[1 - \Gamma(\widehat{k}+1)].
\end{aligned}$$

Proof. The proof of this theorem is similar to the proof of Theorem 3. □

Remark. *The shape parameter k may be approximated by $7.8590z + 2.9554z^2$, where $z = \frac{2}{3+t_3} - log_3 2$ [55].*

2.3.2 LQ-Moments Method

Theorem 7. *Let $X \sim GEV(\xi, \sigma, k)$. Then, if $k \neq 0$,*

$$\begin{aligned}
\zeta_1 &= \xi + \frac{\sigma}{k} - \frac{\sigma}{k}\{p[-\log B_{1:1}^{-1}(\alpha)]^k + (1-2p)[-\log B_{1:1}^{-1}(1/2)]^k + p[-\log B_{1:1}^{-1}(1-\alpha)]^k\}, \\
\zeta_2 &= \frac{\sigma}{2k}\{-p[-\log B_{2:2}^{-1}(\alpha)]^k - (1-2p)[-\log B_{2:2}^{-1}(1/2)]^k - p[-\log B_{2:2}^{-1}(1-\alpha)]^k + \\
&\quad + p[-\log B_{1:2}^{-1}(\alpha)]^k + (1-2p)[-\log B_{1:2}^{-1}(1/2)]^k + p[-\log B_{1:2}^{-1}(1-\alpha)]^k\}, \\
\zeta_3 &= \frac{\sigma}{3k}\{-p[-\log B_{3:3}^{-1}(\alpha)]^k - (1-2p)[-\log B_{3:3}^{-1}(1/2)]^k - p[-\log B_{3:3}^{-1}(1-\alpha)]^k + \\
&\quad + 2p[-\log B_{2:3}^{-1}(\alpha)]^k + 2(1-2p)[-\log B_{2:3}^{-1}(1/2)]^k + 2p[-\log B_{2:3}^{-1}(1-\alpha)]^k - \\
&\quad - p[-\log B_{1:3}^{-1}(\alpha)]^k - (1-2p)[-\log B_{1:3}^{-1}(1/2)]^k - p[-\log B_{1:3}^{-1}(1-\alpha)]^k\}, \\
\zeta_4 &= \frac{\sigma}{4k}\{-p[-\log B_{4:4}^{-1}(\alpha)]^k - (1-2p)[-\log B_{4:4}^{-1}(1/2)]^k - p[-\log B_{4:4}^{-1}(1-\alpha)]^k + \\
&\quad + 3p[-\log B_{3:4}^{-1}(\alpha)]^k + 3(1-2p)[-\log B_{3:4}^{-1}(1/2)]^k + 3p[-\log B_{3:4}^{-1}(1-\alpha)]^k - \\
&\quad - 3p[-\log B_{2:4}^{-1}(\alpha)]^k - 3(1-2p)[-\log B_{2:4}^{-1}(1/2)]^k - 3p[-\log B_{2:4}^{-1}(1-\alpha)]^k + \\
&\quad + p[-\log B_{1:4}^{-1}(\alpha)]^k + (1-2p)[-\log B_{1:4}^{-1}(1/2)]^k + p[-\log B_{1:4}^{-1}(1-\alpha)]^k\},
\end{aligned}$$

and

$$\widehat{\eta}_3 = \frac{2}{3}\{-p[-\log B_{3:3}^{-1}(\alpha)]^k - (1-2p)[-\log B_{3:3}^{-1}(1/2)]^k - p[-\log B_{3:3}^{-1}(1-\alpha)]^k +$$
$$+2p[-\log B_{2:3}^{-1}(\alpha)]^k + 2(1-2p)[-\log B_{2:3}^{-1}(1/2)]^k + 2p[-\log B_{2:3}^{-1}(1-\alpha)]^k -$$
$$-p[-\log B_{1:3}^{-1}(\alpha)]^k - (1-2p)[-\log B_{1:3}^{-1}(1/2)]^k - p[-\log B_{1:3}^{-1}(1-\alpha)]^k\}/$$
$$/\{-p[-\log B_{2:2}^{-1}(\alpha)]^k - (1-2p)[-\log B_{2:2}^{-1}(1/2)]^k - p[-\log B_{2:2}^{-1}(1-\alpha)]^k +$$
$$+p[-\log B_{1:2}^{-1}(\alpha)]^k + (1-2p)[-\log B_{1:2}^{-1}(1/2)]^k + p[-\log B_{1:2}^{-1}(1-\alpha)]^k\},$$
$$\widehat{\sigma} = 2\widehat{\zeta}_2\widehat{k}/\{-p[-\log B_{2:2}^{-1}(\alpha)]^{\widehat{k}} - (1-2p)[-\log B_{2:2}^{-1}(1/2)]^{\widehat{k}} -$$
$$-p[-\log B_{2:2}^{-1}(1-\alpha)]^{\widehat{k}} + p[-\log B_{1:2}^{-1}(\alpha)]^{\widehat{k}} +$$
$$-(1-2p)[-\log B_{1:2}^{-1}(1/2)]^{\widehat{k}} + p[-\log B_{1:2}^{-1}(1-\alpha)]^{\widehat{k}}\},$$
$$\widehat{\xi} = \widehat{\zeta}_1 - \frac{\widehat{\sigma}}{\widehat{k}} + \frac{\widehat{\sigma}}{4\widehat{k}}\{p[-\log B_{1:1}^{-1}(\alpha)]^{\widehat{k}} + (1-2p)[-\log B_{1:1}^{-1}(1/2)]^{\widehat{k}} +$$
$$+p[-\log B_{1:1}^{-1}(1-\alpha)]^{\widehat{k}}\}.$$

Proof. The proof of this theorem is similar to the proof of Theorem 4. □

2.3.3 TL-Moments Method

Theorem 8. *Let $X \sim GEV(\xi, \sigma, k)$. Then, if $k \neq 0$,*

$$\lambda_1^{(1)} = \xi + \frac{\sigma}{k}[1 - \Gamma(k+1)(2\cdot 3^{-k} - 3\cdot 2^{-k})],$$
$$\lambda_2^{(1)} = \frac{3\sigma}{k}\Gamma(k+1)(4^{-k} - 2\cdot 3^{-k} + 2^{-k}),$$
$$\lambda_3^{(1)} = \frac{10\sigma}{3k}\Gamma(k+1)(2\cdot 5^{-k} - 5\cdot 4^{-k} + 4\cdot 3^{-k} - 2^{-k}),$$
$$\lambda_4^{(1)} = \frac{5\sigma}{4k}\Gamma(k+1)(14\cdot 6^{-k} - 42\cdot 5^{-k} + 45\cdot 4^{-k} - 20\cdot 3^{-k} + 3\cdot 2^{-k}),$$

and

$$t_3^{(1)} = \frac{10(2\cdot 5^{-k} - 5\cdot 4^{-k} + 4\cdot 3^{-k} - 2^{-k})}{9(4^{-k} - 2\cdot 3^{-k} + 2^{-k})},$$
$$\widehat{\sigma} = \frac{l_2^{(1)}\widehat{k}}{3\Gamma(\widehat{k}+1)(4^{-\widehat{k}} - 2\cdot 3^{-\widehat{k}} + 2^{-\widehat{k}})},$$
$$\widehat{\xi} = l_1^{(1)} - \frac{\widehat{\sigma}}{\widehat{k}}[1 - \Gamma(\widehat{k}+1)(-2\cdot 3^{-\widehat{k}} + 3\cdot 2^{-\widehat{k}})].$$

If $k \neq 0$, then

$$\lambda_1^{(2)} = \xi + \frac{\sigma}{k}[1 - \Gamma(k+1)(6 \cdot 5^{-k} - 15 \cdot 4^{-k} + 10 \cdot 3^{-k})],$$
$$\lambda_2^{(2)} = \frac{10\sigma}{k}\Gamma(k+1)(-6^{-k} + 3 \cdot 5^{-k} - 3 \cdot 4^{-k} + 3^{-k}),$$
$$\lambda_3^{(2)} = \frac{35\sigma}{3k}\Gamma(k+1)(-2 \cdot 7^{-k} + 7 \cdot 6^{-k} - 9 \cdot 5^{-k} + 5 \cdot 4^{-k} - 3^{-k}),$$
$$\lambda_4^{(2)} = \frac{7\sigma}{k}\Gamma(k+1)(-9 \cdot 8^{-k} + 36 \cdot 7^{-k} - 56 \cdot 6^{-k} + 42 \cdot 5^{-k} - 15 \cdot 4^{-k} + 2 \cdot 3^{-k}),$$

and

$$t_3^{(2)} = \frac{7(-2 \cdot 7^{-k} + 7 \cdot 6^{-k} - 9 \cdot 5^{-k} + 5 \cdot 4^{-k} - 3^{-k})}{6(-6^{-k} + 3 \cdot 5^{-k} - 3 \cdot 4^{-k} + 3^{-k})},$$
$$\widehat{\sigma} = \frac{l_2^{(2)}\widehat{k}}{10\Gamma(\widehat{k}+1)(-6^{-\widehat{k}} + 3 \cdot 5^{-\widehat{k}} - 3 \cdot 4^{-\widehat{k}} + 3^{-\widehat{k}})},$$
$$\widehat{\xi} = l_1^{(2)} - \frac{\widehat{\sigma}}{\widehat{k}}[1 - \Gamma(\widehat{k}+1)(6 \cdot 5^{-\widehat{k}} - 15 \cdot 4^{-\widehat{k}} + 10 \cdot 3^{-\widehat{k}})].$$

Proof. The proof of this theorem is similar to the proof of Theorem 5. □

Remark. *For the estimators of k, Ariff [4] used the method of regression to approximate them as follows*

$$\widehat{k} \approx 0.291922291 - 2.89036313t_3^{(1)} + 1.291839815(t_3^{(1)})^2 - 0.403498762(t_3^{(1)})^3 -$$
$$-0.707333631(t_3^{(1)})^4 - 1.728715237(t_3^{(1)})^5 + 4.076511188(t_3^{(1)})^6 +$$
$$+2.525801801(t_3^{(1)})^7 - 5.225208913(t_3^{(1)})^8 - 1.910928577(t_3^{(1)})^9 +$$
$$+2.856823577(t_3^{(1)})^{10},$$
$$\widehat{k} \approx 0.300983183 - 3.911819242t_3^{(2)} + 1.38875248(t_3^{(2)})^2 - 1.11298955(t_3^{(2)})^3 +$$
$$+1.326160015(t_3^{(2)})^4 + 0.578634686(t_3^{(2)})^5 - 1.462068119(t_3^{(2)})^6 -$$
$$-1.103598046(t_3^{(2)})^7 + 1.366381534(t_3^{(2)})^8.$$

2.3.4 Maximum Likelihood Method

The log-likelihood function for a sample $\boldsymbol{x} = \{x_1, x_2, \ldots, x_n\}$ is

$$\log L(\boldsymbol{x}; \xi, \sigma, k) = -n \log \sigma + \sum_{i=1}^{n} \left(\frac{1-k}{k} \log y_i - y_i^{1/k} \right),$$

where $y_i = 1 - \frac{k(x_i-\xi)}{\sigma}$. The estimators of ξ, σ, and k are achieved by solving the system of equations

$$\frac{1}{\sigma}\sum_{i=1}^{n}\frac{1-k-y_i^{1/k}}{y_i} = 0,$$
$$-\frac{n}{\sigma} + \frac{1}{\sigma k}\sum_{i=1}^{n}(1-k-y_i^{1/k})\left(\frac{1}{y_i}-1\right) = 0,$$
$$-\frac{1}{k^2}\sum_{i=1}^{n}\left[(1-k-y_i^{1/k})\log y_i + (1-k-y_i^{1/k})\left(\frac{1}{y_i}-1\right)\right] = 0.$$

2.4 Simulation Study and Results

A simulation study was performed to compare four estimation methods: the L-moments, LQ-moments, TL-moments (with the trimming parameters $t = 1$ and 2), and maximum likelihood methods. The samples of three different size $n = 20, 50$, and 100 were simulated. The values for the location and scale parameter were fixed ($\xi = 0, \sigma = 1$), because different values have negligible effect on the results. From hydrological studies results that the range $-0.4 \leq k \leq 0.4$ covers the estimated shape parameter k of the GP and GEV distributions for many datasets, and, hence, nine values of the shape parameter k are considered: $k \in \{-0.4, -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4\}$. For each combination of n and k a sample drawn from the GP and GEV distributions was simulated 10 000 times. Using Theorems 3–8, L-, LQ-, and TL-moments estimates of the unknown parameters ξ, σ, and k were computed. The parameter estimates using LQ-moments were calculated considering $p \in [0, 0.5]$ for all sample sizes, $\alpha \in [0.15, 0.5]$ when $n = 20$, $\alpha \in [0.075, 0.5]$ when $n = 50$, and $\alpha \in [0.05, 0.5]$ when $n = 100$ with step size 0.025. Since in hydrology, climatology, and meteorology the interest is focused on quantile estimation, the parameter estimates were then substitued into the quantile functions derived from the distribution functions given by (2.1) and (2.5) for probabilities $u = 0.9, 0.99$, and 0.999 to estimate high quantiles. The estimation methods were compared with one another according to the sample mean squared

error (MSE) of a quantile Q

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(\widehat{Q}_i(u) - Q(u))^2.$$

In the simulation study, it was considered that the best the estimator is the one with the smallest MSE.

Tables 2.1 and 2.3 compare the performances of the L-, LQ-, and TL-moments, and maximum likelihood methods. The first one corresponds to the GP distribution and the second one to the GEV distribution. For each combination of the method, sample size n, shape parameter k, and probability u two numbers are displayed: the mean over 10 000 simulations is in the first row, the sample MSE is in the second gray row. The minimum value of MSE is reported in bold. In Tables 2.2 and 2.4 optimal pairs (p, α) for estimating high quantiles by LQ-moments are displayed. Figures 2.1 and 2.3 show the interaction plots, while Figures 2.2 and 2.4 show the main effects plots. The following recommendations are deduced from Tables 2.1 and 2.3, and Figures 2.1 to 2.4.

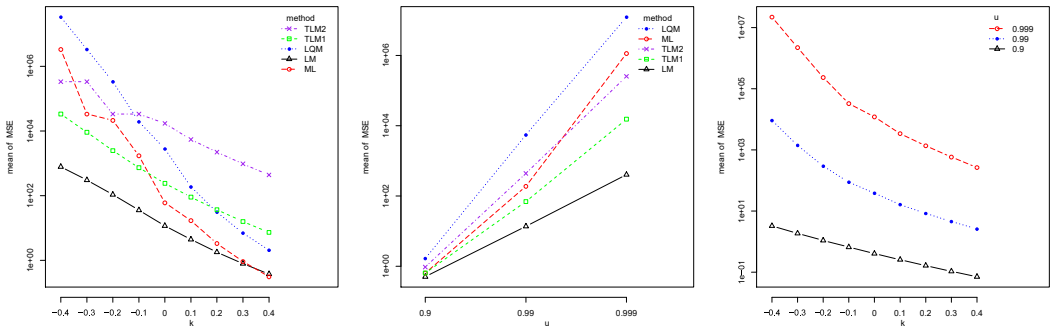
Let us start with the GP distribution - see Table 2.1, and Figures 2.1 and 2.2. The L-moments method provides definitely the smallest MSE when estimating the 99.9% quantile for small samples ($n = 20$) and all values of k, except $k = 0.4$. When the sample size is small and the 90% quantile is estimated, the L-moments method is the best one for $k \leq 0$. When $k \geq 0.1$, the maximum likelihood method is recommended. For the 99% quantile and $k \leq 0.2$ the L-moments method also provides the smallest MSE, otherwise the maximum likelihood estimates are the best. When the sample size is moderate ($n = 50$) and the 99% or 99.9% quantile is estimated, the L-moments method is preferred for $k \leq 0$, while the maximum likelihood method is recommended for $k \geq 0.1$. The L-moments method outperforms other estimation methods when estimating the 90% quantile and $k \leq -0.1$. The L-moments estimators are recommended for the 90% or 99.9% quantile and $k \leq -0.1$, and for the 99% quantile and $k \leq 0$ for sample sizes about 100. In other cases, the maximum likelihood estimators are the best. When the sample size n increases,

Table 2.1: Estimated high quantiles of the GP distribution by LM, TLM1, TLM2, LQM and ML methods (mean over 10 000 simulations in the white row, sample MSE in the gray row)

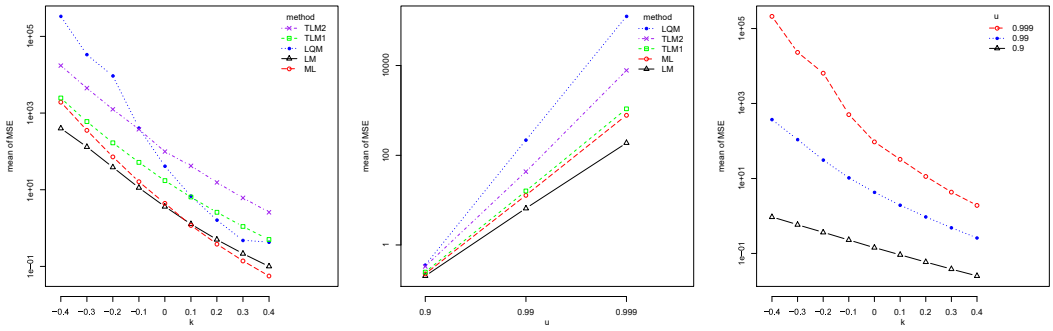
u	k	$Q(u)$	$n = 20$					$n = 50$					$n = 100$				
			LM	TLM1	TLM2	ML	LQM	LM	TLM1	TLM2	ML	LQM	LM	TLM1	TLM2	ML	LQM
0.9	-0.4	3.7797	3.6099	3.6887	3.7898	3.9177	4.9669	3.6922	3.7505	3.7916	3.8339	4.1526	3.7215	3.7616	3.7819	3.8025	3.9630
			1.6399	2.0445	2.8908	2.4016	7.3374	**0.6720**	0.8089	1.0467	0.8098	1.3473	**0.3445**	0.4042	0.5058	0.3844	0.5434
	-0.3	3.3175	3.1984	3.2537	3.3381	3.3866	4.1463	3.2624	3.2985	3.3310	3.3448	3.5901	3.2825	3.3052	3.3211	3.3278	3.4503
			1.0550	1.3291	1.9143	1.3755	3.6349	**0.4319**	0.5207	0.6844	0.4889	0.7857	**0.2203**	0.2594	0.3299	0.2357	0.3269
	-0.2	2.9245	2.8416	2.8801	2.9513	2.9491	3.5012	2.8900	2.9120	2.9385	2.9339	3.1299	2.9031	2.9161	2.9287	2.9264	3.0204
			0.6808	0.8652	1.2752	0.8004	1.8629	**0.2770**	0.3358	0.4497	0.2967	0.4613	**0.1402**	0.1669	0.2162	0.1448	0.1969
	-0.1	2.5893	2.5314	2.5585	2.6196	2.5867	2.9882	2.5674	2.5813	2.6032	2.5874	2.7493	2.5757	2.5835	2.5938	2.5859	2.6577
			0.4391	0.5647	0.8554	0.4738	0.9808	**0.1770**	0.2172	0.2972	0.1812	0.2696	**0.0889**	0.1078	0.1424	0.0892	0.1191
	0	2.3026	2.2651	2.2908	2.3511	2.2866	2.5850	2.2876	2.2976	2.3159	2.2936	2.4115	2.2940	2.2998	2.3087	2.2965	2.3479
			0.2764	0.3679	0.5915	0.2807	0.5137	0.1114	0.1398	0.1967	**0.1106**	0.1544	0.0557	0.0682	0.0920	**0.0549**	0.0705
	0.1	2.0567	2.0264	2.0424	2.0890	2.0320	2.2467	2.0466	2.0538	2.0696	2.0453	2.1264	2.0503	2.0541	2.0614	2.0491	2.0879
			0.1819	0.2441	0.3945	**0.1759**	0.2856	0.0719	0.0920	0.1323	**0.0692**	0.0925	0.0358	0.0454	0.0629	**0.0344**	0.0440
	0.2	1.8452	1.8222	1.8359	1.8771	1.8191	1.9732	1.8378	1.8437	1.8573	1.8332	1.8915	1.8404	1.8435	1.8497	1.8376	1.8658
			0.1170	0.1620	0.2717	**0.1104**	0.1607	0.0458	0.0604	0.0892	**0.0434**	0.0552	0.0228	0.0297	0.0422	**0.0215**	0.0270
	0.3	1.6627	1.6446	1.6571	1.6936	1.6376	1.7475	1.6570	1.6622	1.6740	1.6519	1.6921	1.6590	1.6617	1.6670	1.6560	1.6758
			0.0755	0.1084	0.1890	**0.0707**	0.0929	0.0293	0.0399	0.0607	**0.0275**	0.0333	0.0145	0.0196	0.0286	**0.0136**	0.0166
	0.4	1.5047	1.4902	1.5018	1.5345	1.4826	1.5594	1.5002	1.5050	1.5153	1.4962	1.5227	1.5018	1.5042	1.5089	1.4995	1.5123
			0.0489	0.0732	0.1329	**0.0461**	0.0553	0.0188	0.0265	0.0416	**0.0177**	0.0203	0.0093	0.0130	0.0195	**0.0087**	0.0103
0.99	-0.4	13.2739	12.186	16.684	25.361	16.897	60.094	12.659	14.492	16.514	14.167	24.401	12.895	13.868	14.710	13.654	17.674
			63.385	338.09	2085.7	1 290.0	42 291	**32.226**	79.404	210.11	73.555	1 479.7	**19.495**	32.645	63.303	26.960	157.08
	-0.3	9.9369	9.5440	12.397	18.445	11.352	31.505	9.7439	10.779	12.158	10.241	16.051	9.8293	10.345	10.915	10.055	12.363
			31.396	150.07	941.09	300.53	5 643.0	**14.893**	34.417	93.878	26.126	376.70	**8.4304**	14.349	28.868	10.416	46.972
	-0.2	7.5594	7.4802	9.3326	13.616	7.9353	17.899	7.5276	8.1477	9.1006	7.5755	11.229	7.5432	7.8422	8.2341	7.5466	8.8970
			15.072	68.129	442.80	75.597	874.26	**6.5658**	15.484	43.341	9.5384	80.821	**3.5003**	6.4956	13.531	4.0660	14.674
	-0.1	5.8489	5.8912	7.1352	10.208	5.7662	10.916	5.8685	6.2654	6.9318	5.7369	7.7149	5.8587	6.0479	6.3204	5.7776	6.5984
			7.0425	32.057	217.35	20.800	164.77	**2.8420**	7.2220	20.669	3.5867	17.906	**1.4589**	3.0339	6.5194	1.6040	4.9244
	0	4.6052	4.6762	5.5898	7.9786	4.3242	7.2756	4.6313	4.9030	5.3724	4.4442	5.5810	4.6195	4.7534	4.9542	4.5149	5.0305
			3.1676	15.880	123.27	5.2451	44.794	**1.2557**	3.5219	10.117	1.4197	5.0898	**0.6331**	1.4927	3.4027	0.6505	1.7353
	0.1	3.6904	3.7711	4.3915	6.0274	3.3887	5.1261	3.7211	3.9095	4.2461	3.5300	4.2362	3.7053	3.7940	3.9298	3.5984	3.9324
			1.5521	8.0703	58.862	2.1822	11.217	0.5697	1.7490	5.1686	**0.5613**	1.6773	0.2829	0.7296	1.6439	**0.2590**	0.6682
	0.2	3.0095	3.0825	3.5379	4.7499	2.7244	3.8186	3.0365	3.1723	3.4152	2.8643	3.3154	3.0226	3.0861	3.1835	2.9263	3.1845
			0.7640	4.3157	32.263	0.8521	3.7363	0.2713	0.9070	2.7052	**0.2352**	0.6162	0.1332	0.3757	0.8597	**0.1067**	0.2614
	0.3	2.4960	2.5594	2.8998	3.8063	2.2500	2.9666	2.5192	2.6190	2.7960	2.3730	2.6701	2.5074	2.5537	2.6241	2.4254	2.5919
			0.3933	2.3981	18.227	**0.3698**	1.4293	0.1353	0.4862	1.4570	**0.1027**	0.2435	0.0657	0.1996	0.4616	**0.0450**	0.1067
	0.4	2.1038	2.1582	2.4163	3.1003	1.9012	2.3875	2.1236	2.1980	2.3282	2.0040	2.2080	2.1135	2.1478	2.1992	2.0463	2.1557
			0.2122	1.3792	10.583	**0.1725**	0.6017	0.0706	0.2689	0.8064	**0.0467**	0.1040	0.0339	0.1093	0.2542	**0.0195**	0.0471
0.999	-0.4	37.1223	39.527	105.35	387.75	155.46	2 523.2	38.473	55.564	85.381	50.375	290.27	38.161	45.322	54.586	42.364	87.101
			2 277.4	$> 10^5$	$> 10^6$	$> 10^7$	$> 10^8$	**1 162.3**	7296.9	51604	5 709.4	$> 10^6$	**698.08**	1493.8	4302.7	1 033.9	38 394
	-0.3	23.1443	26.469	61.306	218.62	62.085	657.33	24.827	33.045	49.452	28.103	114.65	24.199	27.558	32.721	25.122	45.331
			881.86	27216	$> 10^6$	$> 10^5$	$> 10^7$	**376.89**	1766.4	13297	1 043.3	$> 10^5$	**197.96**	415.56	1245.1	242.85	5 542.2
	-0.2	14.9054	17.603	36.568	126.91	27.619	179.46	16.100	20.379	29.585	16.534	50.518	15.576	17.342	20.268	15.523	23.707
			307.10	7352.5	$> 10^5$	64 043	$> 10^6$	**108.70**	485.83	3713.2	207.04	27 919	**51.774**	123.97	382.92	59.508	772.42
	-0.1	9.9526	11.799	22.546	75.871	13.855	60.048	10.685	13.069	18.336	10.302	21.010	10.338	11.336	13.021	10.037	13.755
			99.757	2175.0	$> 10^5$	5 126.5	57 654	**30.892**	147.52	1110.1	45.060	1 204.5	**14.159**	39.676	124.82	15.293	123.45
	0	6.9078	8.028	14.368	49.002	7.5610	25.767	7.3303	8.7256	11.740	6.8044	11.073	7.1273	7.7443	8.8069	6.8143	8.5948
			31.679	705.02	5144 8	175.68	8 331.9	**9.5354**	48.291	286.74	11.739	118.44	4.3930	14.537	52.075	**4.3371**	23.294
	0.1	4.9881	5.7498	9.6240	29.573	4.9973	11.973	5.2553	6.0910	7.9216	4.7705	6.8031	5.1217	5.4737	6.0626	4.8432	5.7763
			11.592	263.42	16276	48.667	549.99	3.1882	17.564	120.56	**2.9082**	18.732	1.4408	4.9892	15.704	**1.1921**	5.5304
	0.2	3.7441	4.2420	6.6796	19.277	3.4987	6.8052	3.9139	4.4293	5.5412	3.5252	4.5866	3.8280	4.0445	4.4020	3.6078	4.1300
			4.4985	105.41	6645.3	9.0440	89.291	1.1868	6.7647	43.736	**0.8693**	4.2011	0.5309	1.9503	6.0485	**0.3660**	1.4987
	0.3	2.9137	3.2490	4.8221	12.934	2.6384	4.4153	3.0261	3.3514	4.0405	2.7300	3.3290	2.9689	3.1049	3.3261	2.8019	3.1100
			1.9074	45.207	2871.0	2.3227	19.409	0.4822	2.7788	16.883	**0.2884**	1.1633	0.2129	0.8103	2.4581	**0.1204**	0.4613
	0.4	2.3423	2.5749	3.6135	8.9337	2.1016	3.1231	2.4194	2.6294	3.0651	2.1993	3.3290	2.3800	2.4673	2.6069	2.2561	2.4472
			0.8740	20.503	1304.0	**0.7187**	5.5332	0.2116	1.2116	6.9165	**0.1050**	1.1633	0.0921	0.3567	1.0530	**0.0425**	0.1634

Table 2.2: Optimal combination of parameters (p, α) for estimation high quantiles of the GP distribution by LQ-moments

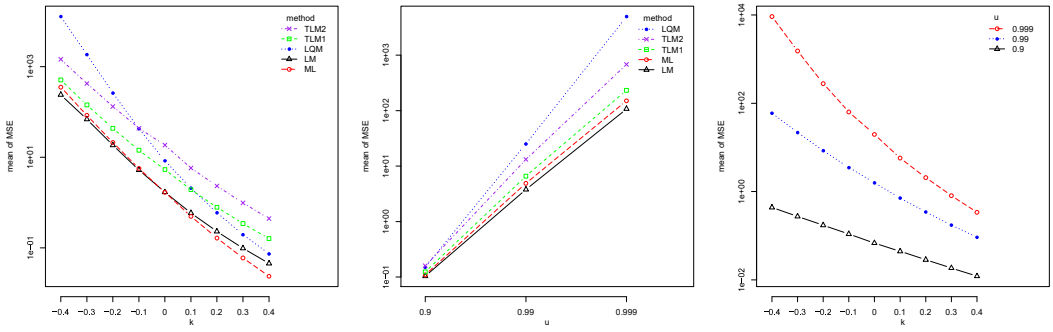
k	$u = 0.9$	$u = 0.99$	$u = 0.999$
		$n = 20$	
-0.4	(0.275, 0.175)	(0.125, 0.15)	(0.125, 0.15)
-0.3	(0.275, 0.175)	(0.15, 0.15)	(0.15, 0.15)
-0.2	(0.275, 0.175)	(0.175, 0.15)	(0.2, 0.15)
-0.1	(0.275, 0.175)	(0.225, 0.15)	(0.225, 0.15)
0	(0.25, 0.15)	(0.275, 0.15)	(0.275, 0.15)
0.1	(0.25, 0.15)	(0.275, 0.15)	(0.3, 0.15)
0.2	(0.275, 0.15)	(0.3, 0.15)	(0.3, 0.15)
0.3	(0.3, 0.15)	(0.3, 0.15)	(0.325, 0.15)
0.4	(0.325, 0.15)	(0.325, 0.15)	(0.325, 0.15)
		$n = 50$	
-0.4	(0.225,0.15)	(0.225, 0.25)	(0.2, 0.25)
-0.3	(0.225,0.15)	(0.25, 0.25)	(0.225, 0.25)
-0.2	(0.175,0.125)	(0.125, 0.1)	(0.1, 0.1)
-0.1	(0.175,0.1)	(0.175, 0.1)	(0.15, 0.1)
0	(0.2,0.1)	(0.225, 0.1)	(0.2, 0.1)
0.1	(0.175,0.075)	(0.25, 0.1)	(0.225, 0.1)
0.2	(0.2,0.075)	(0.275, 0.1)	(0.25, 0.1)
0.3	(0.2,0.075)	(0.275, 0.1)	(0.275, 0.1)
0.4	(0.225,0.075)	(0.3, 0.075)	(0.275, 0.1)
		$n = 100$	
-0.4	(0.125, 0.1)	(0.1, 0.1)	(0.275, 0.275)
-0.3	(0.15, 0.1)	(0.125, 0.1)	(0.1, 0.1)
-0.2	(0.15, 0.1)	(0.175, 0.1)	(0.15, 0.1)
-0.1	(0.175, 0.1)	(0.2, 0.1)	(0.175, 0.1)
0	(0.15, 0.075)	(0.25, 0.1)	(0.225, 0.1)
0.1	(0.175, 0.075)	(0.25, 0.1)	(0.25, 0.1)
0.2	(0.2, 0.075)	(0.325, 0.05)	(0.275, 0.1)
0.3	(0.225, 0.075)	(0.35, 0.05)	(0.275, 0.1)
0.4	(0.225, 0.075)	(0.375, 0.05)	(0.3, 0.1)



(a) $n = 20$

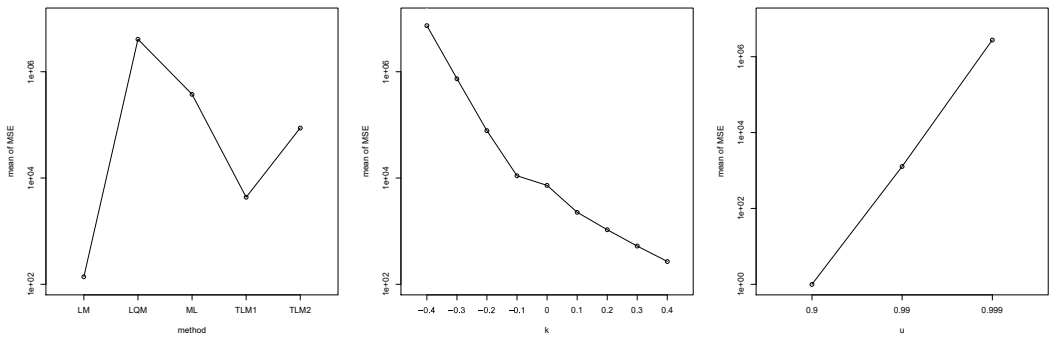


(b) $n = 50$

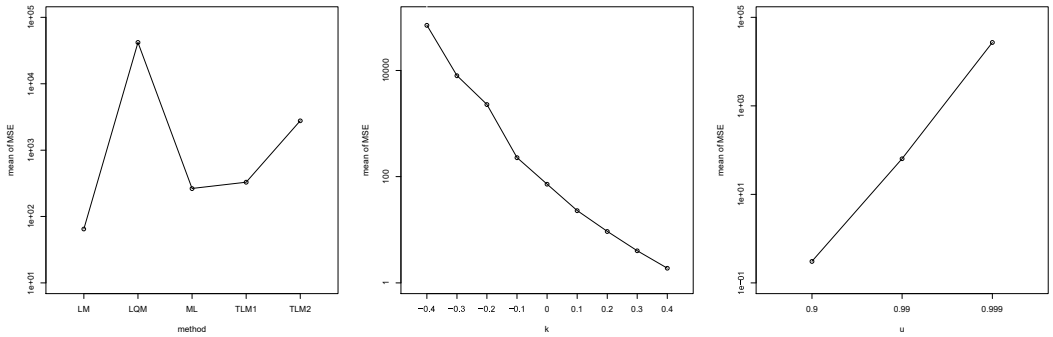


(c) $n = 100$

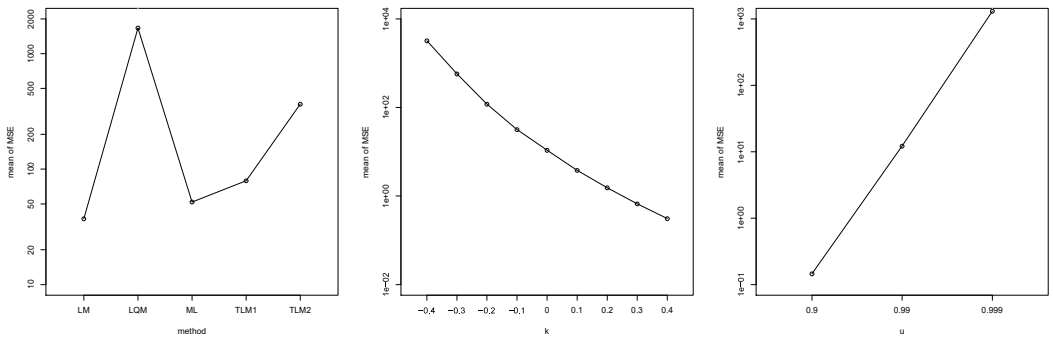
Figure 2.1: Interaction plots of MSEs for the GP distribution



(a) $n = 20$



(b) $n = 50$



(c) $n = 100$

Figure 2.2: Main effects plots of MSEs for the GP distribution

the MSE decreases. The bigger is the parameter k, the smaller is the MSE. The MSE increases with increasing probability u.

Finally, let us deal with the results for the GEV distribution – see Table 2.3, and Figures 2.3 and 2.4. The L-moments method provides the smallest MSE for small samples ($n = 20$), all values of k, and probabilities u. When the sample size is moderate ($n = 50$) and the 90% quantile is estimated, the L-moments method is preferred when $k \leq 0$, while the maximum likelihood method is recommended in other cases. When the 99% and 99.9% quantiles are estimated, the L-moments method yields the smallest MSE when $k \leq 0.1$ and the maximum likelihood method when $k \geq 0.2$. If a larger sample is available ($n = 100$), the L-moments method is the best one when $k \leq 0$ and the maximum likelihood when $k \geq 0.1$. When the sample size n increases, the MSE decreases. The bigger is the parameter k, the smaller is the MSE. The MSE increases with increasing probability u.

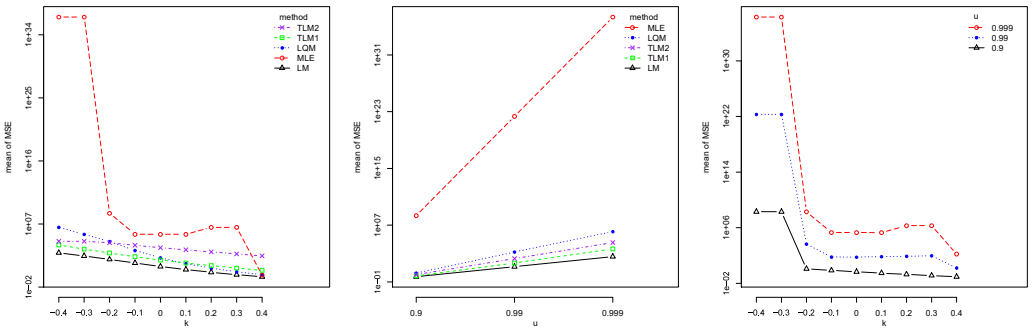
To summarize the obtained results it can be said that there are small differences when estimating high quantiles of the GP and GEV distributions. It was revealed that the L-moments and maximum likelihood methods outperform the LQ- and TL-moments methods. The L-moments method is preferred for distributions with heavier tails ($k < 0$), while the maximum likelihood method is recommended for distributions with lighter tails ($k \geq 0$). From Tables 2.2 and 2.4 it is observed that the common quick estimators such as the median $\left(p = 0, 0 \leq \alpha \leq \frac{1}{2}\right)$, trimean $\left(p = \alpha = \frac{1}{4}\right)$, and Gastwirth's $\left(p = \frac{3}{10}, \alpha = \frac{1}{3}\right)$ estimators are completely unsuitable for estimation high quantiles of the GP and GEV distributions. Comparing the MSEs of TL-moments methods with trimmed the one or two smallest and largest observations, it is observed that the TL-moments method with the trimming parameter $t = 1$ yields better estimates. The LQ-moments and TL-moments methods with the trimming parameter $t = 2$ provide definitely the worst estimates of high quantiles and therefore are not recommended.

Table 2.3: Estimated high quantiles of the GEV distribution by LM, TLM1, TLM2, LQM and ML methods (mean over 10 000 simulations in the white row, sample MSE in the gray row)

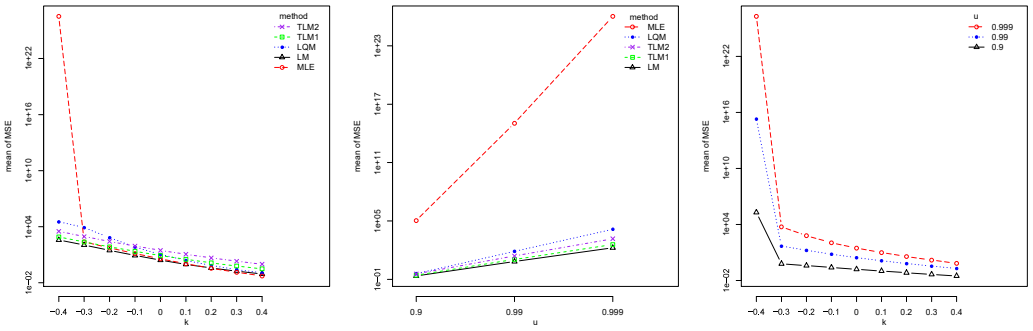
u	k	$Q(u)$	$n = 20$					$n = 50$					$n = 100$				
			LM	TLM1	TLM2	ML	LQM	LM	TLM1	TLM2	ML	LQM	LM	TLM1	TLM2	ML	LQM
0.9	-0.4	3.6499	3.4646	3.5800	3.6845	2 370.8	4.8778	3.5527	3.6197	3.6704	21.213	4.0561	3.5937	3.6290	3.6488	3.6779	3.8484
			1.7028	2.2586	3.1184	$> 10^9$	7.5124	**0.7165**	0.8287	1.0908	$> 10^6$	1.4406	**0.3874**	0.4325	0.5408	0.4371	0.5964
	-0.3	3.2142	3.0957	3.1698	3.2520	381.88	4.0766	3.1588	3.1966	3.2354	3.2291	3.5071	3.1854	3.2005	3.2141	3.2222	3.3588
			1.1208	1.4781	2.0852	$> 10^9$	3.8244	**0.4664**	0.5435	0.7258	0.5518	0.8460	**0.2488**	0.2808	0.3576	0.2701	0.3629
	-0.2	2.8421	2.7685	2.8149	2.8809	2.8386	3.4456	2.8115	2.8333	2.8632	2.8306	3.0608	2.8278	2.8338	2.8428	2.8366	2.9521
			0.7357	0.9700	1.4052	1.0149	1.9824	**0.3007**	0.3580	0.4858	0.3347	0.4987	**0.1579**	0.1833	0.2378	0.1667	0.2203
	-0.1	2.5237	2.4782	2.5080	2.5623	2.4718	2.9333	2.5066	2.5208	2.5442	2.4952	2.6790	2.5165	2.5192	2.5250	2.5093	2.6025
			0.4798	0.6402	0.9558	0.5653	1.0494	**0.1924**	0.2370	0.3274	0.2042	0.2955	**0.0998**	0.1204	0.1593	0.1029	0.1344
	0	2.2504	2.2217	2.2426	2.2883	2.1722	2.5397	2.2405	2.2514	2.2701	2.2121	2.3731	2.2466	2.2485	2.2524	2.2306	2.3054
			0.3111	0.4260	0.6573	0.3391	0.5710	**0.1228**	0.1579	0.2224	0.1255	0.1757	**0.0631**	0.0796	0.1076	0.0636	0.0826
	0.1	2.1051	1.9963	2.0129	2.0523	1.9251	2.2121	2.0091	2.0186	2.0341	1.9723	2.0997	2.0130	2.0149	2.0178	1.9927	2.0536
			0.2013	0.2863	0.4577	0.2118	0.3199	0.0786	0.1060	0.1526	**0.0779**	0.1050	0.0401	0.0531	0.0734	**0.0394**	0.0511
	0.2	1.8121	1.7991	1.8137	1.8484	1.7199	1.9454	1.8081	1.8170	1.8304	1.7687	1.8697	1.8108	1.8129	1.8154	1.7891	1.8398
			0.1307	0.1948	0.3233	0.1368	0.1866	0.0506	0.0718	0.1058	**0.0490**	0.0640	0.0257	0.0357	0.0506	**0.0246**	0.0319
	0.3	1.6363	1.6269	1.6408	1.6719	1.5489	1.7261	1.6335	1.6419	1.6540	1.5952	1.6731	1.6355	1.6375	1.6403	1.6145	1.6547
			0.0854	0.1344	0.2319	0.0939	0.1141	0.0328	0.0491	0.0742	**0.0312**	0.0399	0.0166	0.0243	0.0353	**0.0155**	0.0203
	0.4	1.4837	1.4768	1.4904	1.5186	1.4038	1.5437	1.4817	1.4896	1.5008	1.4471	1.5068	1.4831	1.4850	1.4882	1.4642	1.4941
			0.0563	0.0942	0.1691	**0.0562**	0.0742	0.0214	0.0340	0.0527	**0.0202**	0.0259	0.0108	0.0167	0.0249	**0.0099**	0.0135
0.99	-0.4	13.2423	12.547	15.765	20.840	$> 10^{10}$	46.237	12.888	14.006	15.473	$> 10^6$	20.444	13.024	13.555	14.127	14.023	15.951
			65.440	227.02	927.97	$> 10^{23}$	12 482	**33.841**	54.044	121.39	$> 10^{16}$	504.17	**20.152**	23.815	41.278	22.148	77.357
	-0.3	9.9169	9.7983	11.677	15.167	$> 10^9$	25.004	9.8821	10.462	11.427	10.824	13.724	9.8969	10.139	10.494	10.312	11.593
			32.190	97.837	424.85	$> 10^{23}$	1 917.4	**15.058**	23.928	54.485	24.565	126.88	**8.2899**	10.454	18.672	8.8705	26.485
	-0.2	7.5468	7.6375	8.7893	11.221	10.851	14.598	7.6019	7.9449	8.5843	7.9666	9.7023	7.5749	7.7116	7.9309	7.7238	8.4727
			14.982	43.670	202.34	21 936	336.08	**6.3431**	10.940	25.253	9.2018	34.013	**3.2910**	4.7553	8.6981	3.5844	8.8448
	-0.1	5.8410	5.9796	6.7352	8.4524	6.7391	9.3192	5.9066	6.1370	6.5666	5.9953	6.9923	5.8731	5.9664	6.1027	5.8973	6.3544
			6.7346	20.422	101.15	111.72	69.085	**2.6536**	5.1876	12.127	3.5377	9.7291	**1.3333**	2.2462	4.1845	1.4596	3.1636
	0	4.6001	4.7309	5.2587	6.4880	5.0311	6.3644	4.6563	4.8235	5.1177	4.6154	5.2148	4.6268	4.6965	4.7836	4.5924	4.9099
			3.0375	10.064	53.267	212.86	17.298	**1.1467**	2.5580	6.0456	1.3956	3.0923	**0.5666**	1.1032	2.0833	0.5986	1.1968
	0.1	3.6873	3.7979	4.1835	5.0767	3.8964	4.6147	3.7322	3.8577	4.0642	3.6351	4.0540	3.7082	3.7612	3.8197	3.6482	3.8543
			1.4223	5.2327	29.495	307.04	5.1998	**0.5220**	1.3131	3.1308	0.5660	1.0917	0.2549	0.5638	1.0749	**0.2475**	0.4775
	0.2	3.0075	3.0995	3.3896	4.0488	3.1210	3.5121	3.0435	3.1387	3.2877	2.9279	3.1959	3.0241	3.0640	3.1060	2.9562	3.0972
			0.7054	2.8644	17.080	380.27	1.8703	0.2524	0.7017	1.6836	**0.2369**	0.4279	0.1219	0.2998	0.5750	**0.1035**	0.2056
	0.3	2.4948	2.5725	2.7950	3.2889	2.5824	2.7759	2.5245	2.5967	2.7074	2.4094	2.5869	2.5084	2.5379	2.5702	2.4420	2.5413
			0.3736	1.6442	10.275	456.66	0.7872	0.1302	0.3901	0.9392	**0.1030**	0.1873	0.0622	0.1658	0.3189	**0.0441**	0.0974
	0.4	2.1030	2.1702	2.3431	2.7188	1.9683	2.2667	2.1282	2.1830	2.2674	2.0230	2.1441	2.1146	2.1360	2.1622	2.0545	2.1260
			0.2117	0.9853	6.3824	0.2272	0.3882	0.0717	0.2253	0.5426	**0.0467**	0.0938	0.0338	0.0953	0.1833	**0.0193**	0.0518
0.999	-0.4	37.1144	40.666	77.517	179.34	$> 10^{17}$	933.56	39.267	47.495	63.335	$> 10^{11}$	143.5204	38.423	41.431	46.535	43.108	59.984
			2 226.9	29 981	$> 10^5$	$> 10^{37}$	$> 10^7$	**1 102.4**	2 406.7	9 593.2	$> 10^{27}$	$> 10^5$	623.82	696.18	1 562.2	**622.06**	5 825.0
	-0.3	23.1403	26.695	44.693	102.04	$> 10^{16}$	265.14	24.956	28.744	37.230	29.779	58.684	24.137	25.445	28.116	25.796	33.257
			793.68	7 735.6	$> 10^5$	$> 10^{37}$	$> 10^6$	**320.01**	673.67	2 621.9	704.63	24 446	**158.32**	198.98	455.39	158.84	1 075.2
	-0.2	14.9034	17.395	26.714	59.819	814.71	81.564	16.012	18.036	22.665	17.701	27.778	15.466	16.195	17.598	16.024	19.443
			247.65	2 113.9	63 111	$> 10^9$	92 704	**83.816**	200.39	768.06	153.70	1 990.0	**38.585**	60.902	140.32	41.923	199.20
	-0.1	9.9516	11.491	16.666	36.302	27.833	29.946	10.574	11.767	14.347	11.055	15.451	10.253	10.706	11.455	10.377	11.93
			72.853	630.42	26 934	$> 10^6$	4 928.8	**22.482**	64.053	242.80	36.626	192.59	**10.218**	20.022	45.968	11.478	39.237
	0	6.9073	7.8337	10.883	22.864	31.433	14.071	7.2597	7.9960	9.4730	7.2759	9.0981	7.0732	7.3629	7.7737	7.0258	7.8427
			22.386	208.39	12 546	$> 10^6$	431.30	**6.6835**	22.124	82.983	9.4401	29.466	**3.0449**	7.0817	16.083	3.2736	8.9011
	0.1	4.9879	5.5619	7.4406	14.957	32.570	7.8082	5.1981	5.6622	6.5335	5.0499	5.9081	5.0852	5.2707	5.5057	4.9788	5.4086
			7.6655	76.367	6 147.7	$> 10^6$	60.207	**2.2459**	8.2659	30.609	2.6289	6.2158	1.0214	2.6961	6.0273	**0.9778**	2.3705
	0.2	3.7439	4.1186	5.3162	10.157	36.014	5.0020	3.8776	4.1744	4.7055	3.6888	4.1420	3.8053	3.9236	4.0653	3.6887	3.9377
			2.9717	30.731	3 100.1	$> 10^7$	12.701	0.8509	3.3342	12.137	**0.7946**	1.6786	0.3840	1.1041	2.4227	**0.3078**	0.7424
	0.3	2.9136	3.1722	3.9577	7.1502	36.520	3.5249	3.0043	3.1965	3.5316	2.8232	3.0867	2.9551	3.0303	3.1208	2.8483	3.0045
			1.2946	13.415	1 590.7	$> 10^7$	3.6301	0.3603	1.4470	5.1492	**0.2624**	0.5552	0.1608	0.4855	1.0442	**0.1030**	0.2742
	0.4	2.3422	2.5299	3.0583	5.2093	2.2627	2.6677	2.4073	2.5353	2.7524	2.2506	2.4164	2.3720	2.4197	2.4810	2.2816	2.3857
			0.6249	6.2816	826.01	1.2779	1.3270	0.1687	0.6729	2.3266	**0.0950**	0.2244	0.0743	0.2287	0.4819	**0.0369**	0.1193

Table 2.4: Optimal combination of parameters (p, α) for estimation high quantiles of the GEV distribution by LQ-moments

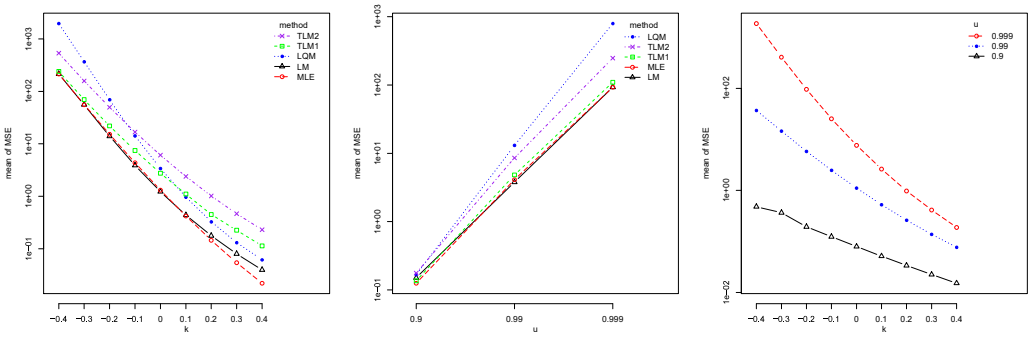
k	$u = 0.9$	$u = 0.99$	$u = 0.999$
		$n = 20$	
-0.4	(0.325, 0.175)	(0.1, 0.15)	(0.125, 0.15)
-0.3	(0.325, 0.175)	(0.125, 0.15)	(0.125, 0.15)
-0.2	(0.25, 0.175)	(0.175, 0.15)	(0.15, 0.15)
-0.1	(0.25, 0.175)	(0.2, 0.15)	(0.2, 0.15)
0	(0.225, 0.15)	(0.25, 0.15)	(0.225, 0.15)
0.1	(0.25, 0.15)	(0.275, 0.15)	(0.275, 0.15)
0.2	(0.275, 0.15)	(0.3, 0.15)	(0.3, 0.15)
0.3	(0.3, 0.15)	(0.3, 0.15)	(0.3, 0.15)
0.4	(0.325, 0.15)	(0.3, 0.15)	(0.3, 0.15)
		$n = 50$	
-0.4	(0.2,0.15)	(0.175, 0.2)	(0.125, 0.2)
-0.3	(0.2,0.15)	(0.225, 0.2)	(0.15, 0.175)
-0.2	(0.175,0.125)	(0.2, 0.15)	(0.2, 0.175)
-0.1	(0.2,0.125)	(0.2, 0.125)	(0.175, 0.1)
0	(0.2,0.1)	(0.25, 0.125)	(0.225, 0.1)
0.1	(0.225,0.1)	(0.25, 0.1)	(0.25, 0.1)
0.2	(0.25,0.1)	(0.275, 0.1)	(0.275, 0.1)
0.3	(0.25,0.1)	(0.275, 0.1)	(0.275, 0.1)
0.4	(0.275,0.1)	(0.275, 0.1)	(0.275, 0.1)
		$n = 100$	
-0.4	(0.175, 0.125)	(0.225, 0.2)	(0.3, 0.275)
-0.3	(0.2, 0.125)	(0.15, 0.125)	(0.25, 0.2)
-0.2	(0.175, 0.1)	(0.2, 0.125)	(0.175, 0.125)
-0.1	(0.2, 0.1)	(0.225, 0.125)	(0.2, 0.125)
0	(0.2, 0.1)	(0.25, 0.1)	(0.225, 0.1)
0.1	(0.225, 0.1)	(0.25, 0.1)	(0.25, 0.1)
0.2	(0.225, 0.075)	(0.275, 0.1)	(0.275, 0.1)
0.3	(0.25, 0.075)	(0.275, 0.1)	(0.275, 0.1)
0.4	(0.25, 0.075)	(0.275, 0.1)	(0.275, 0.1)



(a) $n = 20$

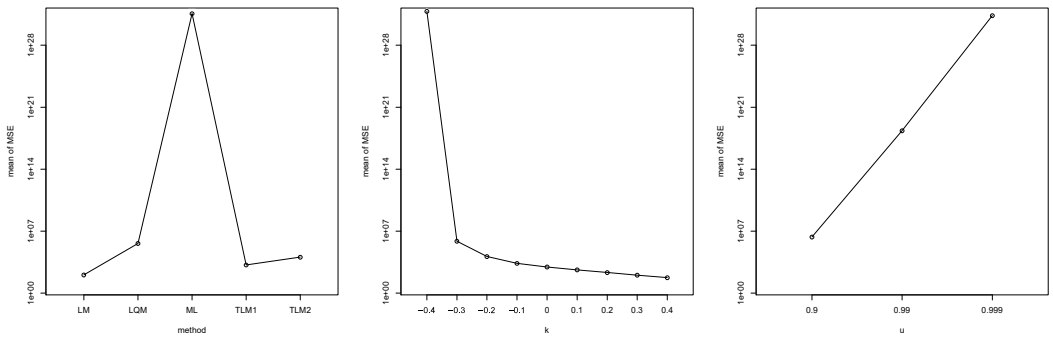


(b) $n = 50$

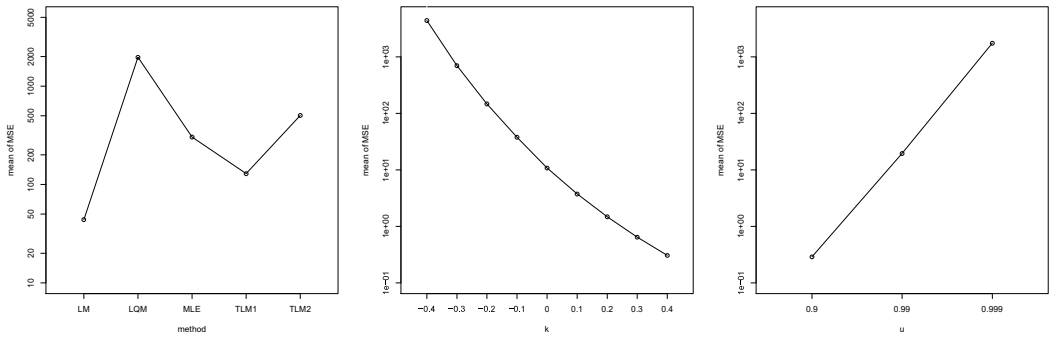


(c) $n = 100$

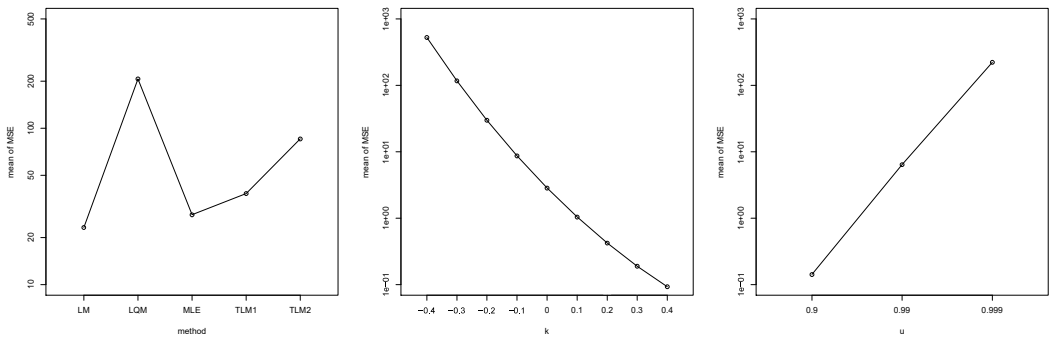
Figure 2.3: Interaction plots of MSEs for the GEV distribution



(a) $n = 20$



(b) $n = 50$



(c) $n = 100$

Figure 2.4: Main effects plots of MSEs for the GEV distribution

2.5 Conclusion

The chapter presents expressions of the first four L-, LQ-, and TL-moments of the GP and GEV distributions and estimators of their parameters based on these moments including the well-known maximum likelihood estimators as well. In the simulation study, the performance of the usually used maximum likelihood method was compared to the alternative considered estimation methods – L-, LQ-, and TL-moments methods – to estimate high quantiles of the GP and GEV distributions considering various sample sizes, values of the shape parameter and probabilities. The simulations revealed that L-moments and maximum likelihood methods outperform definitely other considered estimation methods when estimating high quantiles of these distributions, although the optimal combinations of parameters (p, α) for estimation LQ-moments based quantiles were identified, and the one or two largest observations were trimmed for estimation quantiles based upon TL-moments. It was found out that the common quick estimators such as the median, trimean, and Gastwirth's estimators highlighted by Mudholkar and Hutson [84] are completely unsuitable when estimating high quantiles of these distributions by LQ-moments. Although the results for symmetric TL-moments with the trimming parameter $t = 1$ and 2 are displayed, simulations were performed also for asymmetric TL-moments with trimming parameters $t_1 = 0, t_2 = 1$ and $t_1 = 0, t_2 = 2$. Using asymmetric TL-moments, smaller MSEs can be achieved in comparison with symmetric TL-moments for certain combinations of the probability, sample size and value of the shape parameter (particularly for very high quantiles and distributions with heavier tails), nevertheless the L-moments and maximum likelihood methods still provide the best high quantile estimates of the GP and GEV distributions.

3

Asymptotic Confidence Intervals of Quantile Estimates for the GP and GEV Distributions

Contents

3.1 Introduction

Traditional techniques for estimating parameters of a univariate probability distribution are the moments and maximum likelihood methods. After that Hosking [55] introduced L-moments, the estimation method based upon them began to be popular, especially in such areas as hydrology, climatology, and meteorology [41, 70, 89], but also in economics [12, 13, 110]. It has already been stated that the L-moments method is in some cases preferred over the traditional methods for quantile estimation, specifically when such distribution with heavier tails than

has the normal distribution are fitted to small sample data, as has been shown in several studies using large computer simulations [57, 59, 77, 118]. Studies involving parameter estimation using L-moments have continued to be focused on point estimation, even though the confidence interval for an estimate is preferable because it is more informative than a point estimator.

Prescott and Walden [90] derived the asymptotic variance of the maximum likelihood-based parameter estimators for the GEV distribution. Smith [106] presented the asymptotic variance of the maximum likelihood-based parameters estimators for the two-parametric GP distribution. Note that the maximum likelihood estimator of the location parameter of the three-parametric GP distribution does not exist, because the log-likelihood function is not bounded with respect to this parameter. In general, and as Hosking [55] had stated, it is very difficult to obtain the exact distributions of the L-moments parameter estimators of a univariate probability distribution. The asymptotic distributions of moments and L-moments parameter estimators may nevertheless be derived by using the delta method, because both the sample moments and L-moments have asymptotically a normal distribution [55, 92]. This approach was used by Hosking [57, 59] to obtain parameters and quantile estimators of the GP and GEV distributions that are based on PWMs. Hosking [59] showed that the variance of PWM estimators of both parameters and quantiles of the GEV distribution is well approximated by the asymptotic theory even for small sample sizes (about 50). Hosking [57] also concluded that the asymptotic PWM-based confidence intervals for quantile estimates of the GP distribution give the best results in terms of empirical coverage probability in comparison with the maximum likelihood and moments methods. Several other studies have dealt with the construction of asymptotic confidence intervals based on PWMs of different probability distributions and their comparison to the intervals obtained by other estimation methods: Dupius and Field [35] considered asymptotic and bootstrap confidence intervals for the GEV distribution, the work by Heo and Salas [49] was devoted to the log-Gumbel distribution, Heo, Salas and Kim [50]

presented estimation of confidence intervals for the Weibull distribution, Mahdi and Ashkar [74] investigated the two-parametric Weibull distribution, etc.

Although L-moments are in practice more popular than PWMs and are used in many studies, no attention has been given to estimation of confidence intervals for parameter and quantile estimates of the GP and GEV distributions based on L-moments. It is expected that, as in the case of PWMs, asymptotic confidence intervals for the GP and GEV parameter and quantile estimates based on L-moments will outperform those obtained by the moments and maximum likelihood methods.

This chapter deals with estimating confidence intervals for quantiles of the GP and GEV distributions and assessing of their accuracy. In Section 3.2 asymptotic distributions of parameter estimators of the GP and GEV distributions based on moments, L-moments, and maximum likelihood are presented. A simulation study, showing how the approximate confidence intervals based on various estimation methods behave for particular values of the parameters and sample size is performed in Section 3.3. In Section 3.4, the derived confidence intervals are applied to real meteorological data. The conclusions thus obtained are summarized and discussed in the final section. This chapter relies on the article [114].

3.2 Asymptotic Confidence Intervals

The asymptotic distribution of moments and L-moments estimators may be derived by the delta method (see, e.g., [92]).

Theorem 9. *Let us assume a random sample $x_1, x_2, \ldots, x_n$ whose assumed probability distribution depends on some unknown k-variate parameter vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)^T$. Let $\boldsymbol{s}_n = (s_1, s_2, \ldots, s_m)^T$ be an m-dimensional statistic with the asymptotic normal distribution having mean $(S_1, S_2, \ldots, S_m)^T$ and covariance matrix $n^{-1}\boldsymbol{\Sigma}$. Further, let $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_k)^T = \boldsymbol{g}(s_1, s_2, \ldots, s_m) = (g_1(s_1, s_2, \ldots, s_m), g_2(s_1, s_2, \ldots, s_m), \ldots, g_k(s_1, s_2, \ldots, s_m))^T$ and g_i be totally differentiable for each $i = 1, 2, \ldots, k$. Then, as $n \to \infty$*

$$\widehat{\boldsymbol{\theta}} \xrightarrow{d} \mathcal{N}_m((\theta_1, \theta_2, \ldots, \theta_k)^T, n^{-1}\mathbf{G}\boldsymbol{\Sigma}\mathbf{G}^T),$$

where

$$G_{ij} = \frac{\partial g_i}{\partial s_j}, i = 1, 2, \ldots, k, j = 1, 2, \ldots, m, \tag{3.1}$$

and $\mathcal{N}_m$ *denotes the* m*-variate normal distribution.*

Proof. See [92], p. 388. □

Corollary 1. *Let us assume a random sample* $x_1, x_2, \ldots, x_n$ *whose assumed probability distribution depends on some unknown* k*-variate parameter vector* $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)^T$*. Further, let the asymptotic distribution of* $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_k)^T$ *be normal with mean* $(\theta_1, \theta_2, \ldots, \theta_k)^T$ *and covariance matrix*

$$\begin{pmatrix} \operatorname{var}(\widehat{\theta}_1) & \operatorname{cov}(\widehat{\theta}_1, \widehat{\theta}_2) & \ldots & \operatorname{cov}(\widehat{\theta}_1, \widehat{\theta}_k) \\ \operatorname{cov}(\widehat{\theta}_2, \widehat{\theta}_1) & \operatorname{var}(\widehat{\theta}_2) & \ldots & \operatorname{cov}(\widehat{\theta}_2, \widehat{\theta}_k) \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{cov}(\widehat{\theta}_k, \widehat{\theta}_1) & \operatorname{cov}(\widehat{\theta}_k, \widehat{\theta}_2) & \ldots & \operatorname{var}(\widehat{\theta}_k) \end{pmatrix}.$$

Then, for $0 < u < 1$*,*

$$\widehat{Q}(u) \xrightarrow{d} \mathcal{N}(Q(u), \operatorname{var}(\widehat{Q}(u)),$$

where

$$\operatorname{var}(\widehat{Q}(u)) = \sum_{i=1}^{k} \left(\frac{\partial Q}{\partial \theta_i} \right)^2 \operatorname{var}(\widehat{\theta}_i) + 2 \sum_{\substack{i=1 \\ i \neq j}}^{k} \sum_{j=1}^{k} \left(\frac{\partial Q}{\partial \theta_i} \right) \left(\frac{\partial Q}{\partial \theta_j} \right) \operatorname{cov}(\widehat{\theta}_i, \widehat{\theta}_j),$$

and $\mathcal{N}$ *denotes the univariate normal distribution. Hence,* $1 - \alpha$ *asymptotic central confidence interval for a quantile* $Q(u), 0 < u < 1$*, is in the form*

$$\left(\widehat{Q}(u) - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \sqrt{\operatorname{var}(\widehat{Q}(u))}; \widehat{Q}(u) + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \sqrt{\operatorname{var}(\widehat{Q}(u))} \right),$$

where $\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$ *is the* $\left(1 - \frac{\alpha}{2}\right)$*-quantile of the standard normal distribution.*

Proof. Straightforward application of Theorem 9 gives these results. □

3.2.1 Asymptotic Confidence Intervals for the GP Distribution

According to the cumulative distribution function of the GP distribution given by Equation (2.1), its quantile function is

$$Q(u) = \begin{cases} \xi + \frac{\sigma}{k}[1-(1-u)^k], & k \neq 0, \\ \xi - \sigma \log(1-u), & k = 0. \end{cases}$$

Here, the focus is only on the case $k \neq 0$. With respect to Corollary 1, the asymptotic variance of the quantile estimator $\hat{Q}(u)$, if $k \neq 0$, is

$$\begin{aligned} \operatorname{var}(\widehat{Q}(u)) = \operatorname{var}(\widehat{\xi}) + \frac{[1-(1-u)^k]^2}{k^2}\operatorname{var}(\widehat{\sigma}) + \\ + \frac{\sigma^2[-(1-u)^k + k(1-u)^k \log(1-u) + 1]^2}{k^4}\operatorname{var}(\widehat{k}) + \\ + \frac{2[1-(1-u)^k]}{k}\operatorname{cov}(\widehat{\xi}, \widehat{\sigma}) - \\ - \frac{\sigma[-(1-u)^k + k(1-u)^k \log(1-u) + 1]}{k^2}\operatorname{cov}(\widehat{\xi}, \widehat{k}) - \\ - \frac{\sigma[1-(1-u)^k][-(1-u)^k + k(1-u)^k \log(1-u) + 1]}{k^3}\operatorname{cov}(\widehat{\sigma}, \widehat{k}), \end{aligned}$$

where the variances $\operatorname{var}(\widehat{\xi}), \operatorname{var}(\widehat{\sigma})$, and $\operatorname{var}(\widehat{k})$, and covariances $\operatorname{cov}(\widehat{\xi}, \widehat{\sigma}), \operatorname{cov}(\widehat{\xi}, \widehat{k})$, and $\operatorname{cov}(\widehat{\sigma}, \widehat{k})$ are given as follows with respect to the chosen estimation method.

Moments Method

Theorem 10. *Let $X \sim GP(\xi, \sigma, k)$. Then, if $k > -\frac{1}{6}$,*

$$\begin{aligned} \operatorname{var}(\widehat{\xi}) &= p_{11}^2\sigma_{11} + 2p_{11}p_{12}\sigma_{12} + 2p_{11}p_{13}\sigma_{13} + p_{12}^2\sigma_{22} + 2p_{12}p_{13}\sigma_{23} + p_{13}^2\sigma_{33}, \\ \operatorname{var}(\widehat{\sigma}) &= p_{21}^2\sigma_{11} + 2p_{21}p_{22}\sigma_{12} + 2p_{21}p_{23}\sigma_{13} + p_{22}^2\sigma_{22} + 2p_{22}p_{23}\sigma_{23} + p_{23}^2\sigma_{33}, \\ \operatorname{var}(\widehat{k}) &= p_{31}^2\sigma_{11} + 2p_{31}p_{32}\sigma_{12} + 2p_{31}p_{33}\sigma_{13} + p_{32}^2\sigma_{22} + 2p_{32}p_{33}\sigma_{21} + p_{33}^2\sigma_{33}, \\ \operatorname{cov}(\widehat{\xi}, \widehat{\sigma}) &= p_{11}p_{21}\sigma_{11} + p_{12}p_{21}\sigma_{12} + p_{11}p_{22}\sigma_{12} + p_{13}p_{21}\sigma_{13} + p_{11}p_{23}\sigma_{13} + \\ &\quad + p_{12}p_{22}\sigma_{22} + p_{13}p_{22}\sigma_{23} + p_{12}p_{23}\sigma_{23} + p_{13}p_{23}\sigma_{33}, \\ \operatorname{cov}(\widehat{\xi}, \widehat{k}) &= p_{11}p_{31}\sigma_{11} + p_{12}p_{31}\sigma_{12} + p_{11}p_{32}\sigma_{12} + p_{13}p_{31}\sigma_{13} + p_{11}p_{33}\sigma_{13} + \\ &\quad + p_{12}p_{32}\sigma_{22} + p_{13}p_{32}\sigma_{23} + p_{12}p_{33}\sigma_{23} + p_{13}p_{33}\sigma_{33}, \\ \operatorname{cov}(\widehat{\sigma}, \widehat{k}) &= p_{21}p_{31}\sigma_{11} + p_{22}p_{31}\sigma_{12} + p_{21}p_{32}\sigma_{12} + p_{23}p_{31}\sigma_{13} + p_{21}p_{33}\sigma_{13} + \\ &\quad + p_{22}p_{32}\sigma_{22} + p_{23}p_{32}\sigma_{23} + p_{22}p_{33}\sigma_{23} + p_{23}p_{33}\sigma_{33}, \end{aligned}$$

where

$$p_{11} = 1 + m_1\sqrt{\frac{1+2k}{m_2 - m_1^2}} + \frac{m_1 m_3 - m_2^2}{(m_2 - m_1^2)^2}\left(1 + \frac{2k}{1+k}\right)^2,$$

$$p_{12} = -\frac{1}{2}\sqrt{\frac{1+2k}{m_2 - m_1^2}} + \frac{m_1 m_2 - m_3}{4(m_2 - m_1^2)^2}\left(1 + \frac{2k}{1+k}\right)^2,$$

$$p_{13} = \frac{1}{6(m_2 - m_1^2)}\left(1 + \frac{2k}{1+k}\right)^2,$$

$$p_{21} = -m_1(1+k)\sqrt{\frac{1+2k}{m_2 - m_1^2}} - \frac{m_1 m_3 - m_2^2}{(m_2 - m_1^2)^2}(3k+2)\left(1 + \frac{2k}{1+k}\right)^2,$$

$$p_{22} = \frac{1+k}{2}\sqrt{\frac{1+2k}{m_2 - m_1^2}} - \frac{m_1 m_2 - m_3}{4(m_2 - m_1^2)^2}(3k+2)\left(1 + \frac{2k}{1+k}\right),$$

$$p_{23} = -\frac{1}{6(m_2 - m_1^2)}(3k+2)\left(1 + \frac{2k}{1+k}\right)^2,$$

$$p_{31} = -\frac{m_1 m_3 - m_2^2}{2(m_2 - m_1^2)^{5/2}}\sqrt{1+2k}\left(1 + \frac{2k}{1+k}\right)^2,$$

$$p_{32} = -\frac{m_1 m_2 - m_3}{4(m_2 - m_1^2)^{5/2}}\sqrt{1+2k}\left(1 + \frac{2k}{1+k}\right)^2,$$

$$p_{33} = -\frac{1}{6(m_2 - m_1^2)^{3/2}}\sqrt{1+2k}\left(1 + \frac{2k}{1+k}\right)^2,$$

$$\sigma_{11} = \frac{\sigma^2}{(1+k)^2(1+2k)},$$

$$\sigma_{12} = \frac{2\sigma^2(\xi + 2\sigma + 3\xi k)}{(1+k)^2(1+2k)(1+3k)},$$

$$\sigma_{13} = \frac{\sigma}{(1+k)^2(1+2k)(1+3k)(1+4k)}[18\sigma^3 + 4\xi\sigma^2(4 + 22k + 35k^2 + 50k^3 +$$

$$+24k^4) + 3\xi^2\sigma(1 + 7k + 12k^2) - 4\xi^3(1 + 10k + 35k^2 + 50k^3 + 24k^4)],$$

$$\sigma_{22} = \frac{4\sigma}{(1+k)^2(1+2k)^2(1+3k)(1+4k)}[\sigma^3(5 + 11k) + \xi\sigma^2(5 + 36k + 87k^2 +$$

$$+120k^3 + 124k^4 + 48k^5) + \xi^2\sigma(1 + 9k + 26k^2 + 24k^3) -$$

$$-\xi^3(1+2k)^2(1 + 8k + 19k^2 + 12k^3)],$$

$$\sigma_{23} = \frac{6\sigma^2}{(1+k)^2(1+2k)^2(1+3k)(1+4k)(1+5k)}[6\sigma^3(3 + 7k) +$$

$$+2\xi\sigma^2(8 + 57k + 85k^2) - 3\xi^2\sigma(1 + 14k + 71k^2 + 154k^3 + 120k^4) +$$

$$+\xi^3(1 + 14k + 71k^2 + 154k^3 + 120k^4)],$$

$$\sigma_{33} = \frac{9\sigma^2}{(1+k)^2(1+2k)^2(1+3k)^2(1+4k)(1+5k)(1+6k)}[4\sigma^4(19+105k+146k^2)+$$
$$+24\xi\sigma^3(3+34k+117k^2+126k^3)+4\xi^2\sigma^2(8+129k+155k^2+80k^3+1044k^4+$$
$$+720k^5)+8\xi^3\sigma(1+20k+155k^2+580k^3+1044k^4+720k^5)+\xi^4(1+3k)^2(1+$$
$$+17k+104k^2+268k^3+240k^4)].$$

Proof. The point estimators of parameters ξ and σ obtained by the moments method are

$$\widehat{\xi} = m_1 - \sqrt{(m_2 - m_1^2)(1+2\widehat{k})},$$
$$\widehat{\sigma} = (\widehat{k}+1)\sqrt{(m_2 - m_1^2)(1+2\widehat{k})},$$

while the estimator of the shape parameter k must be obtained numerically from the following equation

$$\frac{2(1-k)\sqrt{1+2k}}{1+3k} = \frac{m_3 - 3m_1m_2 + 2m_1^3}{(m_2 - m_1^2)^{3/2}}, \tag{3.2}$$

where

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad m_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2, \quad m_3 = \frac{1}{n}\sum_{i=1}^{n} x_i^3$$

are the first three sample raw moments. Note that the first three population moments exist when $k > -\frac{1}{3}$. If $k > -\frac{1}{6}$, then the asymptotic distribution of the vector of the first three sample raw moments $(m_1, m_2, m_3)^T$ is normal with mean $(\mu_1, \mu_2, \mu_3)^T$ and covariance matrix as follows [92]

$$n^{-1}\mathbf{\Sigma} = n^{-1}\begin{pmatrix} \mu_2 - \mu_1^2 & \mu_3 - \mu_1\mu_2 & \mu_4 - \mu_1\mu_3 \\ \mu_3 - \mu_1\mu_2 & \mu_4 - \mu_2^2 & \mu_5 - \mu_2\mu_3 \\ \mu_4 - \mu_1\mu_3 & \mu_5 - \mu_2\mu_3 & \mu_6 - \mu_3^2 \end{pmatrix},$$

where $\mu_i = \mathrm{E}X^i, i = 1, 2, \ldots, 6$, is the ith population raw moment of the variable

X. Specifically,

$$\mu_1 = \xi + \frac{\sigma}{1+k},$$

$$\mu_2 = \xi^2 + \frac{2\xi\sigma}{1+k} + \frac{2\sigma^2}{(1+k)(1+2k)},$$

$$\mu_3 = \xi^3 + \frac{3\xi^2\sigma}{1+k} + \frac{6\xi\sigma^2}{(1+k)(1+2k)} + \frac{6\sigma^3}{(1+k)(1+2k)(1+3k)},$$

$$\mu_4 = \xi^4 + \frac{4\xi^3\sigma}{1+k} + \frac{12\xi^2\sigma^2}{(1+k)(1+2k)} + \frac{24\xi\sigma^3}{(1+k)(1+2k)(1+3k)} + \\ + \frac{24\sigma^4}{(1+k)(1+2k)(1+3k)(1+4k)},$$

$$\mu_5 = \xi^5 + \frac{5\xi^4\sigma}{1+k} + \frac{20\xi^3\sigma^2}{(1+k)(1+2k)} + \frac{6\xi^2\sigma^3}{(1+k)(1+2k)(1+3k)} + \\ + \frac{120\xi\sigma^4}{(1+k)(1+2k)(1+3k)(1+4k)} + \frac{120\sigma^5}{(1+k)(1+2k)(1+3k)(1+4k)(1+5k)},$$

$$\mu_6 = \xi^6 + \frac{6\xi^5\sigma}{1+k} + \frac{30\xi^4\sigma^2}{(1+k)(1+2k)} + \frac{120\xi^3\sigma^3}{(1+k)(1+2k)(1+3k)} + \\ + \frac{360\xi^2\sigma^4}{(1+k)(1+2k)(1+3k)(1+4k)} + \frac{720\xi\sigma^5}{(1+k)(1+2k)(1+3k)(1+4k)(1+5k)} + \\ + \frac{720\sigma^6}{(1+k)(1+2k)(1+3k)(1+4k)(1+5k)(1+6k)}.$$

However, the Jacobi matrix $\mathbf{G}$ of the transformation

$$\begin{Bmatrix} m_1 \\ m_2 \\ m_3 \end{Bmatrix} \rightarrow \begin{Bmatrix} \xi \\ \sigma \\ k \end{Bmatrix}$$

that is used in Theorem 9 cannot be obtained directly, because the estimator of the shape parameter k is not expressed as an explicit function of the sample raw moments m_1, m_2, and m_3 (see Equation (3.2)). The covariance matrix of the vector of the parameters estimators $(\widehat{\xi}, \widehat{\sigma}, \widehat{k})^T$ may be then obtained by using the following system of transformations

$$\begin{Bmatrix} m_1 \\ m_2 \\ m_3 \end{Bmatrix} \overset{\text{1st transformation}}{\longrightarrow} \begin{Bmatrix} m_1 \\ m_2 \\ m_3 \\ R \end{Bmatrix} \overset{\text{2nd transformation}}{\longrightarrow} \begin{Bmatrix} m_1 \\ m_2 \\ m_3 \\ k \end{Bmatrix} \overset{\text{3rd transformation}}{\longrightarrow} \begin{Bmatrix} \xi \\ \sigma \\ k \end{Bmatrix}$$

as seen in the studies by Heo and Salas [49] and by Heo, Salas and Kim [50]. In the 1st transformation, R is the right-hand side of Equation (3.2), while in the 2nd transformation it is the left-hand side of Equation (3.2).

Hence, the Jacobi matrix of the first, second and third transformations are

$$\mathbf{J}_1 = \begin{pmatrix} \frac{\partial m_1}{\partial m_1} & \frac{\partial m_1}{\partial m_2} & \frac{\partial m_1}{\partial m_3} \\ \frac{\partial m_2}{\partial m_1} & \frac{\partial m_2}{\partial m_2} & \frac{\partial m_2}{\partial m_3} \\ \frac{\partial m_3}{\partial m_1} & \frac{\partial m_3}{\partial m_2} & \frac{\partial m_3}{\partial m_3} \\ \frac{\partial R}{\partial m_1} & \frac{\partial R}{\partial m_2} & \frac{\partial R}{\partial m_3} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{3(m_1 m_3 - m_2^2)}{(m_2 - m_1^2)^{5/2}} & \frac{3(m_1 m_2 - m_3)}{2(m_2 - m_1^2)^{5/2}} & \frac{1}{(m_2 - m_1^2)^{3/2}} \end{pmatrix},$$

$$\mathbf{J}_2 = \begin{pmatrix} \frac{\partial m_1}{\partial m_1} & \frac{\partial m_1}{\partial m_2} & \frac{\partial m_1}{\partial m_3} & \frac{\partial m_1}{\partial R} \\ \frac{\partial m_2}{\partial m_1} & \frac{\partial m_2}{\partial m_2} & \frac{\partial m_2}{\partial m_3} & \frac{\partial m_2}{\partial R} \\ \frac{\partial m_3}{\partial m_1} & \frac{\partial m_3}{\partial m_2} & \frac{\partial m_3}{\partial m_3} & \frac{\partial m_3}{\partial R} \\ \frac{\partial k}{\partial m_1} & \frac{\partial k}{\partial m_2} & \frac{\partial k}{\partial m_3} & \frac{\partial k}{\partial R} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -\frac{1}{6}\sqrt{1+2k}\left(1 + \frac{2k}{1+k}\right)^2 \end{pmatrix},$$

$$\mathbf{J}_3 = \begin{pmatrix} \frac{\partial \xi}{\partial m_1} & \frac{\partial \xi}{\partial m_2} & \frac{\partial \xi}{\partial m_3} & \frac{\partial \xi}{\partial k} \\ \frac{\partial \sigma}{\partial m_1} & \frac{\partial \sigma}{\partial m_2} & \frac{\partial \sigma}{\partial m_3} & \frac{\partial \sigma}{\partial k} \\ \frac{\partial k}{\partial m_1} & \frac{\partial k}{\partial m_2} & \frac{\partial k}{\partial m_3} & \frac{\partial k}{\partial k} \end{pmatrix} =$$

$$= \begin{pmatrix} 1 + m_1\sqrt{\frac{1+2k}{m_2 - m_1^2}} & -\frac{1}{2}\sqrt{\frac{1+2k}{m_2 - m_1^2}} & 0 & -\sqrt{\frac{m_2 - m_1^2}{1+2k}} \\ -m_1(1+k)\sqrt{\frac{1+2k}{m_2 - m_1^2}} & \frac{1+k}{2}\sqrt{\frac{1+2k}{m_2 - m_1^2}} & 0 & (3k+2)\sqrt{\frac{m_2 - m_1^2}{1+2k}} \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

respectively. Let us denote

$$\mathbf{P} = \mathbf{J}_3\mathbf{J}_2\mathbf{J}_1 = \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix}$$

and

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix}.$$

Then the covariance matrix of $(\widehat{\xi}, \widehat{\sigma}, \widehat{k})^T$ is in the form

$$\mathrm{cov}(\widehat{\xi}, \widehat{\sigma}, \widehat{k}) = n^{-1}\mathbf{P}\mathbf{\Sigma}\mathbf{P}^T.$$

□

L-Moments Method

Theorem 11. *Let $X \sim GP(\xi, \sigma, k)$. Then, if $k > -\frac{1}{2}$,*

$$\begin{aligned}
\operatorname{var}(\widehat{\xi}) &= \Lambda_{11} + 2g_{12}\Lambda_{12} + 2g_{13}\Lambda_{13} + g_{12}^2\Lambda_{22} + 2g_{12}g_{13}\Lambda_{23} + g_{13}^2\Lambda_{33},\\
\operatorname{var}(\widehat{\sigma}) &= g_{22}^2\Lambda_{22} + 2g_{22}g_{23}\Lambda_{23} + g_{23}^2\Lambda_{33},\\
\operatorname{var}(\widehat{k}) &= g_{32}^2\Lambda_{22} + 2g_{32}g_{33}\Lambda_{23} + g_{33}^2\Lambda_{33},\\
\operatorname{cov}(\widehat{\xi}, \widehat{\sigma}) &= g_{22}\Lambda_{12} + g_{23}\Lambda_{13} + g_{12}g_{22}\Lambda_{22} + g_{13}g_{22}\Lambda_{23} + g_{12}g_{23}\Lambda_{23} + g_{13}g_{23}\Lambda_{33},\\
\operatorname{cov}(\widehat{\xi}, \widehat{k}) &= g_{32}\Lambda_{12} + g_{33}\Lambda_{13} + g_{12}g_{32}\Lambda_{22} + g_{13}g_{32}\Lambda_{23} + g_{12}g_{33}\Lambda_{23} + g_{13}g_{33}\Lambda_{33},\\
\operatorname{cov}(\widehat{\sigma}, \widehat{k}) &= g_{22}g_{32}\Lambda_{22} + g_{23}g_{32}\Lambda_{23} + g_{22}g_{33}\Lambda_{23} + g_{23}g_{33}\Lambda_{33},
\end{aligned}$$

where

$$\begin{aligned}
\Lambda_{11} &= \frac{\sigma^2}{(k+1)^2(2k+1)},\\
\Lambda_{12} = \Lambda_{21} &= \frac{\sigma^2(1-k)}{(k+1)^2(k+2)(2k+1)},\\
\Lambda_{13} = \Lambda_{31} &= \frac{\sigma^2(2k^3 - 5k^2 - 12k + 3)}{(k+1)^2(k+2)(k+3)(2k+1)(2k+3)},\\
\Lambda_{22} &= \frac{\sigma^2(2k^3 - k^2 - 2k + 4)}{(k+1)^2(k+2)^2(2k+1)(2k+3)},\\
\Lambda_{23} = \Lambda_{32} &= \frac{\sigma^2(1-k)(2k^3 - 5k^2 - 6k + 6)}{(k+1)^2(k+2)^2(k+3)(2k+1)(2k+3)},\\
\Lambda_{33} &= \frac{\sigma^2(4k^6 - 16k^5 - 13k^4 + 184k^3 + 123k^2 - 66k + 72)}{(k+1)^2(k+2)^2(k+3)^2(2k+1)(2k+3)(2k+5)},\\
g_{12} &= \frac{-3l_2^2 - 6l_2l_3 + l_3^2}{(l_2+l_3)^2},\\
g_{13} &= \frac{4l_2^2}{(l_2+l_3)^2},\\
g_{22} &= \frac{2(3l_2^3 + 9l_2^2l_3 - 9l_2l_3^2 + l_3^3)}{(l_2+l_3)^2},\\
g_{23} &= \frac{4(3l_2^2l_3 - 5l_2^3)}{(l_2+l_3)^2},\\
g_{32} &= \frac{4l_3}{(l_2+l_3)^2},\\
g_{33} &= \frac{-4l_2}{(l_2+l_3)^2}.
\end{aligned}$$

Proof. The parameters estimators based upon L-moments are

$$\widehat{\xi} = l_1 - \frac{3l_2^2 - l_2 l_3}{l_2 + l_3},$$
$$\widehat{\sigma} = \frac{2l_2(l_2 - l_3)(3l_2 - l_3)}{(l_2 + l_3)^2},$$
$$\widehat{k} = \frac{l_2 - 3l_3}{l_2 + l_3}.$$

According to Theorem 2, if $k > -\frac{1}{2}$, the asymptotic distribution of the vector of the first three sample L-moments $(l_1, l_2, l_3)^T$ is normal with mean $(\lambda_1, \lambda_2, \lambda_3)^T$ and covariance matrix

$$n^{-1}\mathbf{\Lambda} = n^{-1}\begin{pmatrix} \Lambda_{11} & \Lambda_{12} & \Lambda_{13} \\ \Lambda_{21} & \Lambda_{22} & \Lambda_{23} \\ \Lambda_{31} & \Lambda_{32} & \Lambda_{33} \end{pmatrix},$$

where its elements are derived by using Equation (1.12), i.e.,

$$\Lambda_{11} = \frac{\sigma^2}{(k+1)^2(2k+1)},$$
$$\Lambda_{12} = \Lambda_{21} = \frac{\sigma^2(1-k)}{(k+1)^2(k+2)(2k+1)},$$
$$\Lambda_{13} = \Lambda_{31} = \frac{\sigma^2(2k^3 - 5k^2 - 12k + 3)}{(k+1)^2(k+2)(k+3)(2k+1)(2k+3)},$$
$$\Lambda_{22} = \frac{\sigma^2(2k^3 - k^2 - 2k + 4)}{(k+1)^2(k+2)^2(2k+1)(2k+3)},$$
$$\Lambda_{23} = \Lambda_{32} = \frac{\sigma^2(1-k)(2k^3 - 5k^2 - 6k + 6)}{(k+1)^2(k+2)^2(k+3)(2k+1)(2k+3)},$$
$$\Lambda_{33} = \frac{\sigma^2(4k^6 - 16k^5 - 13k^4 + 184k^3 + 123k^2 - 66k + 72)}{(k+1)^2(k+2)^2(k+3)^2(2k+1)(2k+3)(2k+5)}.$$

In comparison to the moments method, it is easy to determine the Jacobi matrix $\mathbf{G}$ for the L-moments method

$$\mathbf{G} = \begin{pmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{pmatrix} = \begin{pmatrix} \frac{\partial \xi}{\partial l_1} & \frac{\partial \xi}{\partial l_2} & \frac{\partial \xi}{\partial l_3} \\ \frac{\partial \sigma}{\partial l_1} & \frac{\partial \sigma}{\partial l_2} & \frac{\partial \sigma}{\partial l_3} \\ \frac{\partial k}{\partial l_1} & \frac{\partial k}{\partial l_2} & \frac{\partial k}{\partial l_3} \end{pmatrix} =$$
$$= \frac{1}{(l_2 + l_3)^2}\begin{pmatrix} (l_2 + l_3)^2 & -3l_2^2 - 6l_2 l_3 + l_3^2 & 4l_2^2 \\ 0 & 2(3l_2^3 + 9l_2^2 l_3 - 9l_2 l_3^2 + l_3^3) & 4(3l_2^2 l_3 - 5l_2^3) \\ 0 & 4l_3 & -4l_2 \end{pmatrix}.$$

The covariance matrix of the parameter estimators $(\widehat{\xi}, \widehat{\sigma}, \widehat{k})^T$ is then

$$\mathrm{cov}(\widehat{\xi}, \widehat{\sigma}, \widehat{k}) = n^{-1}\mathbf{G}\mathbf{\Lambda}\mathbf{G}^T.$$

□

3.2.2 Asymptotic Confidence Intervals for the GEV Distribution

The quantile function of the GEV distribution derived from the cumulative distribution function given by Equation (2.5) is

$$Q(u) = \begin{cases} \xi + \frac{\sigma}{k}[1 - (-\log u)^k], & k \neq 0, \\ \xi - \sigma \log(\log u), & k = 0. \end{cases}$$

Here, the focus is again only on the case $k \neq 0$. With respect to Corollary 1 the asymptotic variance of the quantile estimator $\widehat{Q}(u)$, if $k \neq 0$, is

$$\begin{aligned} \operatorname{var}(\widehat{Q}(u)) = \operatorname{var}(\widehat{\xi}) + \frac{[1-(-\log u)^k]^2}{k^2}\operatorname{var}(\widehat{\sigma}) + \\ + \frac{\sigma^2[1-(-\log u)^k + k(-\log u)^k \log(-\log u)]^2}{k^4}\operatorname{var}(\widehat{k}) + \\ + \frac{2[1-(-\log u)^k]}{k}\operatorname{cov}(\hat{\xi}, \widehat{\sigma}) - \\ - \frac{2\sigma[1-(-\log u)^k + k(-\log u)^k \log(-\log u)]}{k^2}\operatorname{cov}(\widehat{\xi}, \widehat{k}) - \\ - \frac{\sigma[1-(-\log u)^k][1-(-\log u)^k + k(-\log u)^k \log(-\log u)]}{k^3}\operatorname{cov}(\widehat{\sigma}, \widehat{k}), \end{aligned}$$

where $\operatorname{var}(\widehat{\xi}), \operatorname{var}(\widehat{\sigma})$, and $\operatorname{var}(\widehat{k})$, and $\operatorname{cov}(\widehat{\xi}, \widehat{\sigma}), \operatorname{cov}(\widehat{\xi}, \widehat{k})$, and $\operatorname{cov}(\widehat{\sigma}, \widehat{k})$ are given as follows with respect to the chosen estimation method.

Moments Method

Theorem 12. *Let $X \sim GEV(\xi, \sigma, k)$. Then, if $k > -\frac{1}{6}$,*

$$\begin{aligned} \operatorname{var}(\widehat{\xi}) &= p_{11}^2\sigma_{11} + 2p_{11}p_{12}\sigma_{12} + 2p_{11}p_{13}\sigma_{13} + p_{12}^2\sigma_{22} + 2p_{12}p_{13}\sigma_{23} + p_{13}^2\sigma_{33}, \\ \operatorname{var}(\widehat{\sigma}) &= p_{21}^2\sigma_{11} + 2p_{21}p_{22}\sigma_{12} + 2p_{21}p_{23}\sigma_{13} + p_{22}^2\sigma_{22} + 2p_{22}p_{23}\sigma_{23} + p_{23}^2\sigma_{33}, \\ \operatorname{var}(\widehat{k}) &= p_{31}^2\sigma_{11} + 2p_{31}p_{32}\sigma_{12} + 2p_{31}p_{33}\sigma_{13} + p_{32}^2\sigma_{22} + 2p_{32}p_{33}\sigma_{21} + p_{33}^2\sigma_{33}, \\ \operatorname{cov}(\widehat{\xi}, \widehat{\sigma}) &= p_{11}p_{21}\sigma_{11} + p_{12}p_{21}\sigma_{12} + p_{11}p_{22}\sigma_{12} + p_{13}p_{21}\sigma_{13} + p_{11}p_{23}\sigma_{13} + \\ &\quad + p_{12}p_{22}\sigma_{22} + p_{13}p_{22}\sigma_{23} + p_{12}p_{23}\sigma_{23} + p_{13}p_{23}\sigma_{33}, \\ \operatorname{cov}(\widehat{\xi}, \widehat{k}) &= p_{11}p_{31}\sigma_{11} + p_{12}p_{31}\sigma_{12} + p_{11}p_{32}\sigma_{12} + p_{13}p_{31}\sigma_{13} + p_{11}p_{33}\sigma_{13} + \\ &\quad + p_{12}p_{32}\sigma_{22} + p_{13}p_{32}\sigma_{23} + p_{12}p_{33}\sigma_{23} + p_{13}p_{33}\sigma_{33}, \\ \operatorname{cov}(\widehat{\sigma}, \widehat{k}) &= p_{21}p_{31}\sigma_{11} + p_{22}p_{31}\sigma_{12} + p_{21}p_{32}\sigma_{12} + p_{23}p_{31}\sigma_{13} + p_{21}p_{33}\sigma_{13} + \\ &\quad + p_{22}p_{32}\sigma_{22} + p_{23}p_{32}\sigma_{23} + p_{22}p_{33}\sigma_{23} + p_{23}p_{33}\sigma_{33}, \end{aligned}$$

where

$$p_{11} = 1 + \frac{m_1[1-\Gamma(1+k)]}{\sqrt{(m_2-m_1^2)[\Gamma(1+2k)-\Gamma^2(1+k)]}} + \frac{m_1 m_3 - m_2^2}{(m_2-m_1^2)^2}\cdot$$
$$\cdot[\Gamma(1+2k)-\Gamma^2(1+k)]\cdot\{\{[\Gamma(1+2k)-\Gamma^2(1+k)]\psi(1+3k)-$$
$$-\psi(1+2k)\Gamma(1+2k)+\psi(1+k)\Gamma^2(1+k)\}\cdot\Gamma(1+3k)+$$
$$+[\Gamma(1+k)\psi(1+2k)-\psi(1+k)\Gamma(1+k)]\Gamma^2(1+2k)\}^{-1}\cdot$$
$$\cdot\{\Gamma^2(1+k)\psi(1+k)+\Gamma(1+k)\Gamma(1+2k)[\psi(1+2k)-\psi(1+k)]-$$
$$-\Gamma(1+2k)\psi(1+2k)\},$$
$$p_{12} = \frac{\Gamma(1+k)-1}{2\sqrt{(m_2-m_1^2)[\Gamma(1+2k)-\Gamma^2(1+k)]}} + \frac{m_1 m_2 - m_3}{2(m_2-m_1^2)^2}\cdot$$
$$\cdot[\Gamma(1+2k)-\Gamma^2(1+k)]\cdot\{\{[\Gamma(1+2k)-\Gamma^2(1+k)]\psi(1+3k)-$$
$$-\psi(1+2k)\Gamma(1+2k)+\psi(1+k)\Gamma^2(1+k)\}\cdot\Gamma(1+3k)+$$
$$+[\Gamma(1+k)\psi(1+2k)-\psi(1+k)\Gamma(1+k)]\Gamma^2(1+2k)\}^{-1}\cdot$$
$$\cdot\{\Gamma^2(1+k)\psi(1+k)+\Gamma(1+k)\Gamma(1+2k)[\psi(1+2k)-\psi(1+k)]-$$
$$-\Gamma(1+2k)\psi(1+2k)\},$$
$$p_{13} = \frac{1}{3(m_2-m_1^2)}\cdot[\Gamma(1+2k)-\Gamma^2(1+k)]\cdot$$
$$\cdot\{\{[\Gamma(1+2k)-\Gamma^2(1+k)]\psi(1+3k)-$$
$$-\psi(1+2k)\Gamma(1+2k)+\psi(1+k)\Gamma^2(1+k)\}\cdot\Gamma(1+3k)+$$
$$+[\Gamma(1+k)\psi(1+2k)-\psi(1+k)\Gamma(1+k)]\Gamma^2(1+2k)\}^{-1}\cdot$$
$$\cdot\{\Gamma^2(1+k)\psi(1+k)+\Gamma(1+k)\Gamma(1+2k)[\psi(1+2k)-\psi(1+k)]-$$
$$-\Gamma(1+2k)\psi(1+2k)\},$$
$$p_{21} = \frac{-m_1|k|}{\sqrt{(m_2-m_1^2)[\Gamma(1+2k)-\Gamma^2(1+k)]}} + \frac{m_1 m_3 - m_2^2}{(m_2-m_1^2)^2}\cdot$$
$$\cdot[\Gamma(1+2k)-\Gamma^2(1+k)]\cdot\{\{[\Gamma(1+2k)-\Gamma^2(1+k)]\psi(1+3k)-$$
$$-\psi(1+2k)\Gamma(1+2k)+\psi(1+k)\Gamma^2(1+k)\}\cdot\Gamma(1+3k)+$$
$$+[\Gamma(1+k)\psi(1+2k)-\psi(1+k)\Gamma(1+k)]\Gamma^2(1+2k)\}^{-1}\cdot$$
$$\cdot k\cdot|k|^{-1}\cdot\{\Gamma^2(1+k)[k\psi(1+k)-1]+\Gamma(1+2k)[1-\psi(1+2k)]\},$$

$$p_{22} = \frac{|k|}{2\sqrt{(m_2 - m_1^2)[\Gamma(1+2k) - \Gamma^2(1+k)]}} - \frac{m_1 m_2 - m_3}{2(m_2 - m_1^2)^2} \cdot$$
$$\cdot[\Gamma(1+2k) - \Gamma^2(1+k)] \cdot \{\{[\Gamma(1+2k) - \Gamma^2(1+k)]\psi(1+3k) -$$
$$-\psi(1+2k)\Gamma(1+2k) + \psi(1+k)\Gamma^2(1+k)\} \cdot \Gamma(1+3k) +$$
$$+[\Gamma(1+k)\psi(1+2k) - \psi(1+k)\Gamma(1+k)]\Gamma^2(1+2k)\}^{-1} \cdot$$
$$\cdot k \cdot |k|^{-1} \cdot \{\Gamma^2(1+k)[k\psi(1+k) - 1] + \Gamma(1+2k)[1 - \psi(1+2k)]\},$$
$$p_{23} = -\frac{1}{3(m_2 - m_1^2)} \cdot [\Gamma(1+2k) - \Gamma^2(1+k)] \cdot$$
$$\cdot\{\{[\Gamma(1+2k) - \Gamma^2(1+k)]\psi(1+3k) -$$
$$-\psi(1+2k)\Gamma(1+2k) + \psi(1+k)\Gamma^2(1+k)\} \cdot \Gamma(1+3k) +$$
$$+[\Gamma(1+k)\psi(1+2k) - \psi(1+k)\Gamma(1+k)]\Gamma^2(1+2k)\}^{-1} \cdot$$
$$\cdot k \cdot |k|^{-1} \cdot \{\Gamma^2(1+k)[k\psi(1+k) - 1] + \Gamma(1+2k)[1 - \psi(1+2k)]\},$$
$$p_{31} = -\frac{m_1 m_3 - m_2^2}{(m_2 - m_1^2)^{5/2}}[\Gamma(1+2k) - \Gamma^2(1+k)] \cdot$$
$$\cdot\{\{[\Gamma(1+2k) - \Gamma^2(1+k)]\psi(1+3k) -$$
$$-\psi(1+2k)\Gamma(1+2k) + \psi(1+k)\Gamma^2(1+k)\} \cdot \Gamma(1+3k) +$$
$$+[\Gamma(1+k)\psi(1+2k) - \psi(1+k)\Gamma(1+k)]\Gamma^2(1+2k)\}^{-1},$$
$$p_{32} = -\frac{m_1 m_2 - m_3}{2(m_2 - m_1^2)^{5/2}}\Gamma(1+2k) - \Gamma^2(1+k)] \cdot$$
$$\cdot\{\{[\Gamma(1+2k) - \Gamma^2(1+k)]\psi(1+3k) -$$
$$-\psi(1+2k)\Gamma(1+2k) + \psi(1+k)\Gamma^2(1+k)\} \cdot \Gamma(1+3k) +$$
$$+[\Gamma(1+k)\psi(1+2k) - \psi(1+k)\Gamma(1+k)]\Gamma^2(1+2k)\}^{-1},$$
$$p_{33} = -\frac{1}{3(m_2 - m_1^2)^{3/2}}\Gamma(1+2k) - \Gamma^2(1+k)] \cdot$$
$$\cdot\{\{[\Gamma(1+2k) - \Gamma^2(1+k)]\psi(1+3k) -$$
$$-\psi(1+2k)\Gamma(1+2k) + \psi(1+k)\Gamma^2(1+k)\} \cdot \Gamma(1+3k) +$$
$$+[\Gamma(1+k)\psi(1+2k) - \psi(1+k)\Gamma(1+k)]\Gamma^2(1+2k)\}^{-1},$$

$$\sigma_{11} = \mu_2 - \mu_1^2,$$

$$\sigma_{12} = \mu_3 - \mu_1\mu_2,$$

$$\sigma_{13} = \mu_4 - \mu_1\mu_3,$$

$$\sigma_{22} = \mu_4 - \mu_2^2,$$

$$\sigma_{23} = \mu_5 - \mu_2\mu_3,$$

$$\sigma_{33} = \mu_6 - \mu_3^2,$$

$$\mu_1 = \xi + \frac{\sigma}{k}[1 - \Gamma(1+k)],$$

$$\mu_2 = \xi^2 + 2\xi\frac{\sigma}{k}[1 - \Gamma(1+k)] + \frac{\sigma^2}{k^2}[1 - 2\Gamma(1+k) + \Gamma(1+2k)],$$

$$\mu_3 = \xi^3 + 3\xi^2\frac{\sigma}{k}[1 - \Gamma(1+k)] + 3\xi\frac{\sigma^2}{k^2}[1 - 2\Gamma(1+k) + \Gamma(1+2k)] +$$
$$+\frac{\sigma^3}{k^3}[1 - 3\Gamma(1+k) + 3\Gamma(1+2k) - \Gamma(1+3k)],$$

$$\mu_4 = \xi^4 + 4\xi^3\frac{\sigma}{k}[1 - \Gamma(1+k)] + 6\xi^2\frac{\sigma^2}{k^2}[1 - 2\Gamma(1+k) + \Gamma(1+2k)] +$$
$$+4\xi\frac{\sigma^3}{k^3}[1 - 3\Gamma(1+k) + 3\Gamma(1+2k) - \Gamma(1+3k)] +$$
$$+\frac{\sigma^4}{k^4}[1 - 4\Gamma(1+k) + \Gamma(1+2k) - 4\Gamma(1+3k) + \Gamma(1+4k)],$$

$$\mu_5 = \xi^5 + 5\xi^4\frac{\sigma}{k}[1 - \Gamma(1+k)] + 10\xi^3\frac{\sigma^2}{k^2}[1 - 2\Gamma(1+k) + \Gamma(1+2k)] +$$
$$+10\xi^2\frac{\sigma^3}{k^3}[1 - 3\Gamma(1+k) + 3\Gamma(1+2k) - \Gamma(1+3k)] +$$
$$+5\xi\frac{\sigma^4}{k^4}[1 - 4\Gamma(1+k) + 6\Gamma(1+2k) - 4\Gamma(1+3k) + \Gamma(1+4k)] +$$
$$+\frac{\sigma^5}{k^5}[1 - 5\Gamma(1+k) + 10\Gamma(1+2k) - 10\Gamma(1+3k) + 5\Gamma(1+4k) - \Gamma(1+5k)],$$

$$\mu_6 = \xi^6 + 6\xi^5\frac{\sigma}{k}[1 - \Gamma(1+k)] + 15\xi^4\frac{\sigma^2}{k^2}[1 - 2\Gamma(1+k) + \Gamma(1+2k)] +$$
$$+20\xi^3\frac{\sigma^3}{k^3}[1 - 3\Gamma(1+k) + 3\Gamma(1+2k) - \Gamma(1+3k)] +$$
$$+15\xi^2\frac{\sigma^4}{k^4}[1 - 4\Gamma(1+k) + 6\Gamma(1+2k) - 4\Gamma(1+3k) + \Gamma(1+4k)] +$$
$$+6\xi\frac{\sigma^5}{k^5}[1 - 5\Gamma(1+k) + 10\Gamma(1+2k) - 10\Gamma(1+3k) + 5\Gamma(1+4k) - \Gamma(1+5k)] +$$
$$+\frac{\sigma^6}{k^6}[1 - 6\Gamma(1+k) + 15\Gamma(1+2k) - 20\Gamma(1+3k) + 15\Gamma(1+4k) - 6\Gamma(1+5k) +$$
$$+\Gamma(1+6k)],$$

and $\Gamma(\cdot)$ *is the gamma function and* $\psi(r) = \frac{\mathrm{d}\log\Gamma(r)}{\mathrm{d}r}$ *is the digamma function.*

Proof. The proof of this theorem is similar to the proof of Theorem 10.

□

L-Moments Method

Theorem 13. *Let* $X \sim GEV(\xi, \sigma, k)$*. Then, if* $k > 0$*,*

$$\mathrm{var}(\widehat{\xi}) = \Lambda_{11} + 2p_{12}\Lambda_{12} + 2p_{13}\Lambda_{13} + p_{12}^2\Lambda_{22} + 2p_{12}p_{13}\Lambda_{23} + p_{13}^2\Lambda_{33},$$

$$\mathrm{var}(\widehat{\sigma}) = p_{22}^2\Lambda_{22} + 2p_{22}p_{23}\Lambda_{23} + p_{23}^2\Lambda_{33},$$

$$\mathrm{var}(\widehat{k}) = p_{32}^2\Lambda_{22} + 2p_{32}p_{33}\Lambda_{23}p_{33}^2\Lambda_{33},$$

$$\mathrm{cov}(\widehat{\xi}, \widehat{\sigma}) = p_{22}\Lambda_{12} + p_{23}\Lambda_{13} + p_{12}p_{22}\Lambda_{22} + p_{13}p_{22}\Lambda_{23} + p_{12}p_{23}\Lambda_{23} + p_{13}p_{23}\Lambda_{33},$$

$$\mathrm{cov}(\widehat{\xi}, \widehat{k}) = p_{32}\Lambda_{12} + p_{33}\Lambda_{13} + p_{12}p_{32}\Lambda_{22} + p_{13}p_{32}\Lambda_{23} + p_{12}p_{33}\Lambda_{23} + p_{13}p_{33}\Lambda_{33},$$

$$\mathrm{cov}(\widehat{\sigma}, \widehat{k}) = p_{22}p_{32}\Lambda_{22} + p_{23}p_{32}\Lambda_{23} + p_{22}p_{33}\Lambda_{23} + p_{23}p_{33}\Lambda_{33},$$

where

$$p_{12} = \frac{\Gamma(1+k)-1}{(1-2^{-k})\Gamma(1+k)} - \frac{3^k l_3\{[1-\Gamma(1+k)]\log 2 + (2^k-1)\psi(1+k)\}}{\Gamma(1+k)[(2^k-1)\log 3 - (3^k-1)\log 2]},$$

$$p_{13} = \frac{3^k\{[1-\Gamma(1+k)]\log 2 + (2^k-1)\psi(1+k)\}}{2\Gamma(1+k)[(2^k-1)\log 3 - (3^k-1)\log 2]},$$

$$p_{22} = \frac{k}{(1-2^{-k}\Gamma(1+k))} + \frac{3^k l_3[1-2^k+k\cdot\log 2 + k(2^k-1)\psi(1+k)]}{2l_2\Gamma(1+k)[(2^k-1)\log 3 - (3^k-1)\log 2]},$$

$$p_{23} = -\frac{3^k[1-2^k+k\cdot\log 2 + k(2^k-1)\psi(1+k)]}{2\Gamma(1+k)[(2^k-1)\log 3 - (3^k-1)\log 2]},$$

$$p_{32} = -\frac{3^k l_3(2^k-1)^2}{2^{k+1}} l_2^2[(2^k-1)\log 3 - (3^k-1)\log 2],$$

$$p_{33} = \frac{3^k l_3(2^k-1)^2}{2^{k+1}} l_2[(2^k-1)\log 3 - (3^k-1)\log 2],$$

$$\Lambda_{11} = \frac{\sigma^2}{k}[2\Gamma(2k) - k\Gamma^2(k)],$$

$$\Lambda_{12} = \Lambda_{21} = -\frac{\sigma^2}{k} \cdot 2^{1-2k} \cdot 9^{-k} \cdot \Gamma(2k)[-2^{2k+1} + 9^k + 36^k + 2^{2k+1} \cdot 9^k \cdot H(-2) -$$
$$-4^{k+1} \cdot 9^k \cdot H(-1) + 2 \cdot 9^k \cdot H(-1/2)],$$

$$\Lambda_{13} = \Lambda_{31} = \frac{2\sigma^2}{k}\Gamma(2k)[1 + 3 \cdot 2^{1-4k} - 2 \cdot 3^{1-2k} + 4^{-k} - 6 \cdot H(-3) + 12 \cdot H(-2) -$$
$$-8 \cdot H(-1) + 3 \cdot 2^{1-2k} \cdot H(-1/2) - 2 \cdot 3^{1-2k} \cdot H(-1/3)],$$

$$\Lambda_{22} = \frac{\sigma^2}{k} \cdot 2^{1-4k} \cdot 9^{-k}\Gamma(2k)[-2^{4k+3} + 4 \cdot 9^k + 36^k +$$
$$+144^k + 4^{2k+1} \cdot 9^k \cdot H(-2) - 2^{2k+1} \cdot 9^k \cdot (4 + 3 \cdot 4^k) \cdot H(-1) + 3^{2k+1} \cdot 4^{k+1} \cdot H(-1/2)],$$

$$\Lambda_{23} = \Lambda_{32} = \frac{\sigma^2}{k} \cdot 2^{1-4k} \cdot 9^{-k}\Gamma(2k)[-103^{1+2k} + 5 \cdot 4^{1+2k} + 12^{1+2k} \cdot 25^{-k} - 36^k -$$
$$-144^k + 2^{1+4k} \cdot 3^{1+2k} \cdot H(-3) - 7 \cdot 2^{1+4k} \cdot 9^k \cdot H(-2) - 3^{1+2k} \cdot 4^{1+k} \cdot H(-3/2) +$$
$$+5 \cdot 2^{1+4k} \cdot 9^k \cdot H(-1) + 36^{1+k} \cdot H(-1) - 3 \cdot 4^{1+2k} \cdot H(-2/3) -$$
$$-2^{5+2k} \cdot 9^k \cdot H(-1/2) + 9 \cdot 2^{1+4k} \cdot H(-1/3)],$$

$$\Lambda_{33} = \frac{2\sigma^2}{k}\Gamma(2k)[1 + 15 \cdot 2^{3-4k} - 16 \cdot 3^{1-2k} + 4^{-k} - 108 \cdot 5^{-2k} + 6^{2-2k} -$$
$$-12 \cdot H(-3) + 24 \cdot H(-2) + 9 \cdot 2^{3-2k} \cdot H(-3/2) - 14 \cdot H(-1) - 9 \cdot 4^{2-k} \cdot H(-1) -$$
$$-8 \cdot 9^{1-k} \cdot H(-1) + 16 \cdot 9^{1-k} \cdot H(-2/3) + 21 \cdot 4^{1-k} \cdot H(-1/2) - 28 \cdot 3^{1-2k} \cdot H(-1/3)],$$

and $\Gamma(\cdot)$ *is the gamma function,* $\psi(r) = \frac{\mathrm{d}\log\Gamma(r)}{\mathrm{d}r}$ *is the digamma function and* $H(z) = {}_2F_1(k, 1+2k; 1+k; z)$ *is the hypergeometric function.*

Proof. The proof of this theorem is similar to the proofs of Theorems 10 and 11. □

Maximum Likelihood Method

The point estimators of ξ, σ, and k are achieved by solving the following system of equations

$$\frac{1}{\sigma}\sum_{i=1}^{n}\frac{1-k-y_i^{1/k}}{y_i} = 0,$$
$$-\frac{n}{\sigma} + \frac{1}{\sigma k}\sum_{i=1}^{n}(1-k-y_i^{1/k})\left(\frac{1}{y_i} - 1\right) = 0,$$
$$-\frac{1}{k^2}\sum_{i=1}^{n}\left\{[1-k-y_i^{1/k}]\log y_i + (1-k-y_i^{1/k})\left(\frac{1}{y_i} - 1\right)\right\} = 0,$$

where $y_i = 1 - \frac{k(x_i-\xi)}{\sigma}$.

Prescott and Walden [90] presented the asymptotic variances and covariances of the maximum likelihood estimators. They are given by the elements of the inversion of the information matrix $\mathbf{M}$ having the elements

$$
\begin{aligned}
M_{11} &= \frac{na}{\sigma^2}, \\
M_{12} = M_{21} &= \frac{n}{\sigma^2 k}[a - \Gamma(2-k)], \\
M_{13} = M_{31} &= -\frac{n}{\sigma k}\left(b + \frac{a}{k}\right), \\
M_{22} &= \frac{n}{\sigma^2 k^2}[1 - 2\Gamma(2-k) + a], \\
M_{23} = M_{32} &= \frac{n}{\sigma^2 k}\left[1 - \gamma - \frac{1-\Gamma(2-k)}{k} - b - \frac{a}{k}\right], \\
M_{33} &= \frac{n}{k^2}\left[\frac{\pi^2}{6} + \left(1 - \gamma - \frac{1}{k}\right)^2 + \frac{2b}{k} + \frac{a}{k^2}\right],
\end{aligned}
$$

where γ denotes the Euler's constant ($\gamma \approx 0.5772157$), $a = (1-k)^2\Gamma(1-2k), b = \Gamma(2-k)\left[\psi(1-k) - \frac{1}{k} + 1\right], \Gamma(\cdot)$ is the gamma function and $\psi(r) = \frac{\mathrm{d}\log\Gamma(r)}{\mathrm{d}r}$ is the digamma function. When $k < \frac{1}{2}$, the estimators are consistent and have asymptotic normal distribution [90].

3.3 Simulation Study and Results

A simulation study was performed to find out how well the approximate confidence intervals for the GP and GEV distributions based on the moments, L-moments, and maximum likelihood methods work for different choices of parameters' values and sample sizes. Samples of five sizes, $n = 20, 50, 100, 250$, and 500 were drawn from the GP and GEV distributions with fixed location and scale parameters ($\xi = 0, \sigma = 1$) and varying values of the shape parameter k, that variation ranging from -0.4 to 0.4 with step size 0.1. For each combination of n and k a sample was simulated 10 000 times. First, the unknown parameters ξ, σ, and k were estimated. Note that the shape parameter k for both the GP and GEV distributions is always evaluated numerically, except in the case of the L-moments method for the GP distribution. The estimators $\widehat{\xi}, \widehat{\sigma}$, and $\widehat{k}$ are then substitued into the corresponding quantile functions. Here, the focus was only on estimating the 95% population quantiles

of both distributions. Using variances of population quantiles derived in Sections 3.2.1 and 3.2.2, the approximate asymptotic 95% confidence intervals based on moments, L-moments, and maximum likelihood are computed.

The estimation methods were compared with one another with respect to the empirical coverage probability and median length of the approximate 95% confidence intervals for an estimate of the 95% population quantile of the GP and GEV distributions. The empirical coverage probability is defined to be the ratio of the number of confidence intervals that contain the true value to the number of simulated samples, which is equal to 10 000. Ideally, that probability should be close to 95%. Moreover, the median length should be short.

First, we will discuss the results obtained for the GP distribution. Let us recall that only approximate asymptotic confidence intervals for the three-parametric GP distribution based on L-moments and moments may be computed. In the case of the two-parametric GP distribution (with unknown scale and shape parameters), the maximum likelihood-based asymptotic confidence intervals may also be derived. The empirical coverage probabilities are shown in Figure 3.1. The L-moments method outperforms with absolutely no ambiguity the moments method: The moments-based confidence intervals can be found only when $k > -\frac{1}{6}$, while the confidence intervals based upon L-moments can be estimated for $k > -\frac{1}{2}$. The latter covers the estimated shape parameter in many practical situations. Moreover, the empirical coverage probabilities of the confidence intervals based upon L-moments are always higher than for the moments method. These two observations clearly favour use of the L-moments method over the moments method. The difference between empirical coverage probabilities of the two methods diminishes with increasing sample size n and the value of the shape parameter k. The estimation methods have comparable performance when $n \geq 100$ and $k \geq 0.1$. As expected, asymptotic theory works well for large sample sizes. It is also observed that the accuracy depends not only on the sample size n but also on the value of the shape parameter k. Good accuracy (about 90%) also can be achieved for smaller sizes: let us say about 50 for the L-moments method when $k \geq 0$, and in the case of moments method when $n \geq 100$

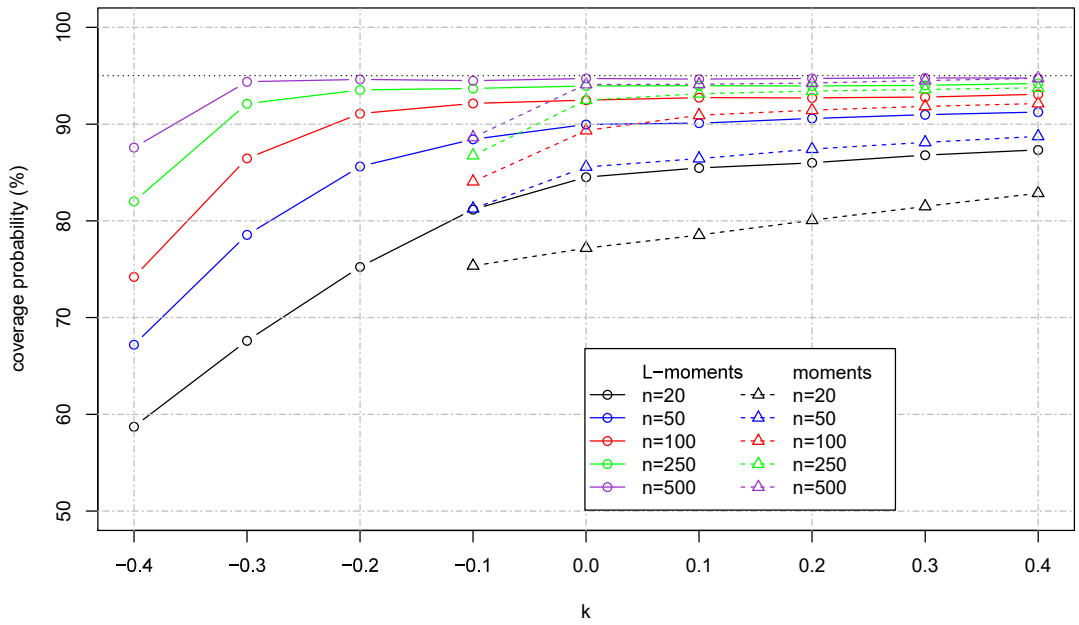


Figure 3.1: Empirical coverage probabilities of 95% confidence intervals for 95% quantile estimates of the GP distribution

and $k \geq 0$. Figure 3.2 shows that the median length of confidence intervals is always shorter for the moments method, but the median lengths are comparable with increasing sample size n and value of the shape parameter k. In both methods, the median length decreases with increasing n and k.

Figures 3.3 and 3.4 summarize the simulation results obtained for the GEV distribution. Here, the choice of best estimation method depends on the value of the shape parameter k. If one has no a priori information about the value of k, then the maximum likelihood method seems to be a good choice, because the variance of the maximum likelihood quantile estimator is defined for all $k < \frac{1}{2}$. In any case, the maximum likelihood method outperforms the moments and L-moments methods for $k \leq 0.1$ and all sample sizes, except the case when $n = 20$. For $k = 0.2$ all methods have comparable performances, except that in the case $n = 20$ the moments method surpasses the other two. If it can be assumed that the shape parameter k is greater than 0.2, then the moments and L-moments methods are preferred, because they show significantly better accuracy than does the maximum likelihood method. It

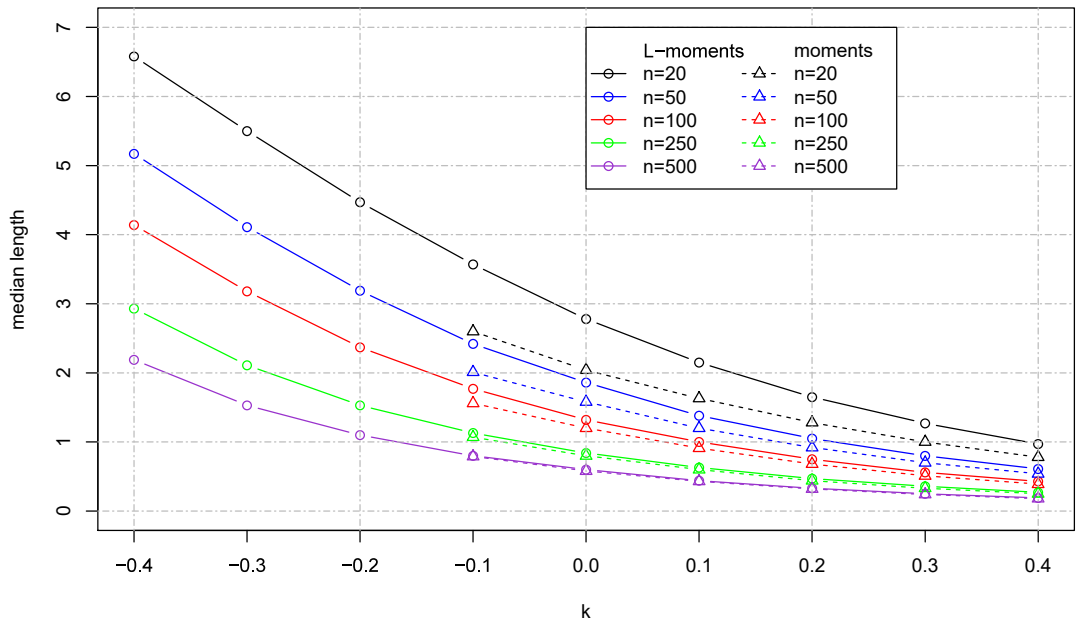


Figure 3.2: Median lengths of 95% confidence intervals for 95% quantile estimates of the GP distribution

can be concluded, however, that use of the L-moments method will probably remain uncommon in practice due to its very restrictive requirement that $k > 0$. It is also important to point out that in the case of the maximum likelihood method the empirical coverage probabilities are affected by the fact that in some cases the local maxima of the corresponding maximum likelihood function do not exist. Hence, the empirical coverage probabilities would probably be even higher if such samples would be omitted from the simulation study. Figure 3.4 shows that all methods provide confidence intervals of similar lengths when $k \geq 0.1$. For the case $-0.1 \leq k \leq 0$, it is obvious that the confidence intervals based on moments method are narrower than are those obtained by the maximum likelihood method. Of course, the median length decreases with increasing n and k.

3.4 Case Study

To show that the asymptotic L-moments derived on the basis of L-moments provide reasonable results comparable to those obtained by the standard techniques,

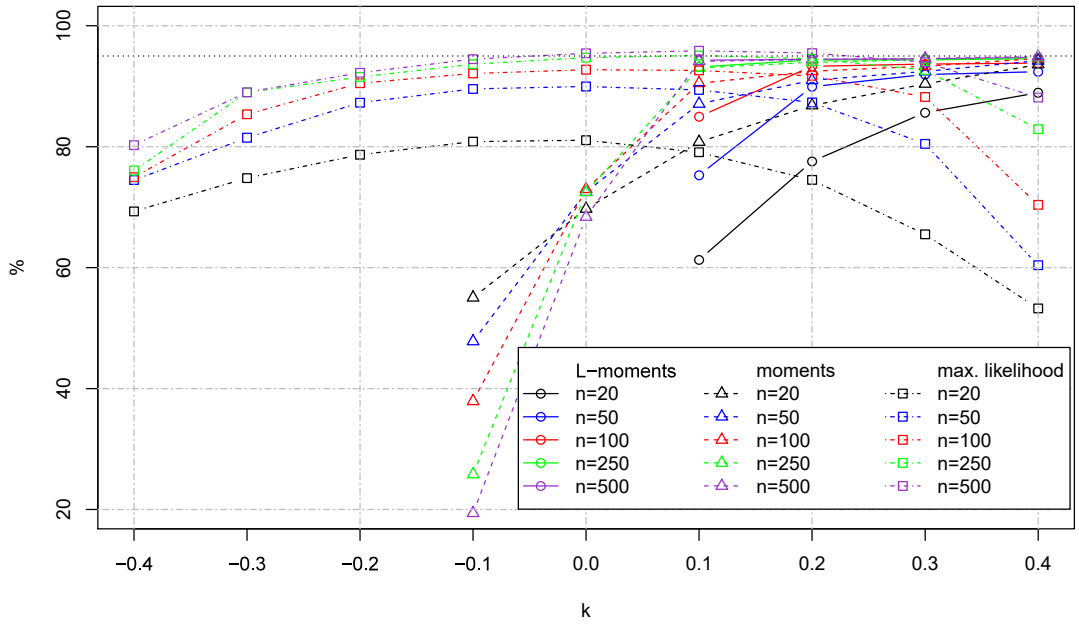


Figure 3.3: Empirical coverage probabilities of 95% confidence intervals for 95% quantile estimates of the GEV distribution

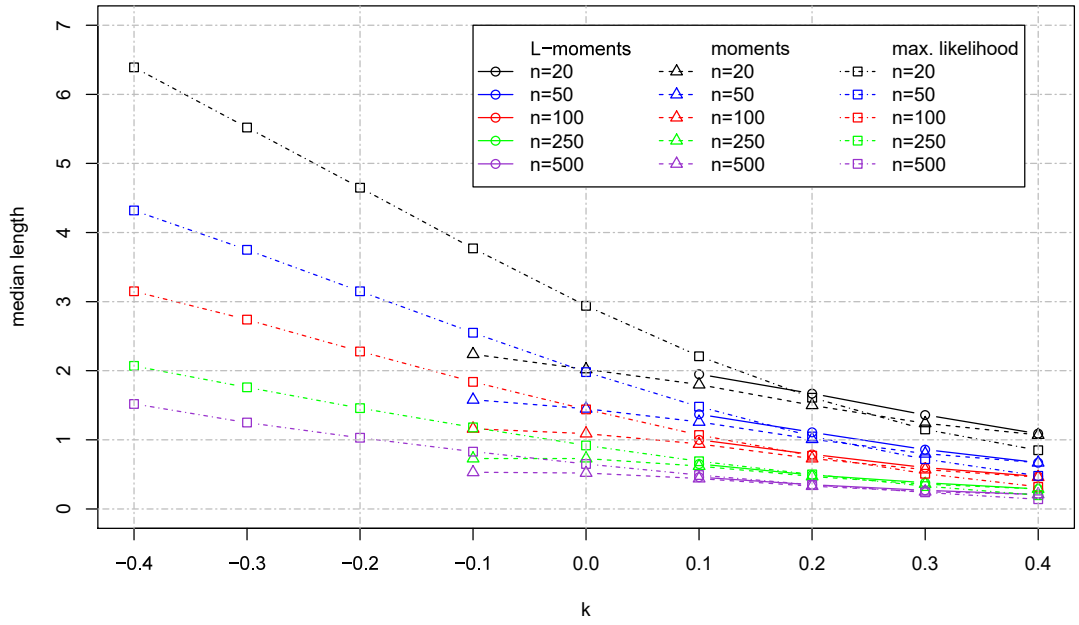


Figure 3.4: Median lengths of 95% confidence intervals for 95% quantile estimates of the GEV distribution

they are applied here to real meteorological data. The maximum annual 1-day precipitation totals for the period from 1961 to 2012 measured at the meteorological station Hustopeče located in the southeast part of the Czech Republic are used to calculate the $80\%, 90\%, 95\%$, and 99% approximate asymptotic confidence intervals of the parameters, and $90\%, 95\%$, and 99% quantile estimates. Basic information concerning the data set is summarized in Table 3.1. Data is displayed in Figure 3.5. According to the results of Kyselý and Picek [70], the GEV distribution seems to be the most suitable for modeling 1- to 7-day precipitation totals in this station.

Table 3.1: Basic information on maximum annual 1-day precipitation totals measured at Hustopeče

Record length (years)	52
Mean (mm)	32
Median (mm)	30.3
Lower quartile (mm)	25.2
Upper quartile (mm)	40.6
Minimum (mm)	14.3
Maximum (mm)	69.2
Variance (mm^2)	103.9
Standard deviation (mm)	10.2

Figure 3.6 shows the point estimates of parameters, and $90\%, 95\%$, and 99% quantile and their confidence intervals obtained by the moments, L-moments, and maximum likelihood methods. Neither point nor interval estimates obtained by the three methods differ significantly. Let us notice that the maximum likelihood-based confidence intervals are the narrowest for the parameters estimators, while they are the widest for the quantile estimates. It was found out that the L-moments method is inapplicable in analysis of daily and multidaily precipitation totals measured in many stations in the Czech Republic due to a very restrictive requirement that $k > 0$, when estimating the confidence intervals. For example, in almost 83% of stations the estimated shape parameter k of the GEV distribution fitted to the maximum annual 1-day precipitation totals was less than 0, that does not allow

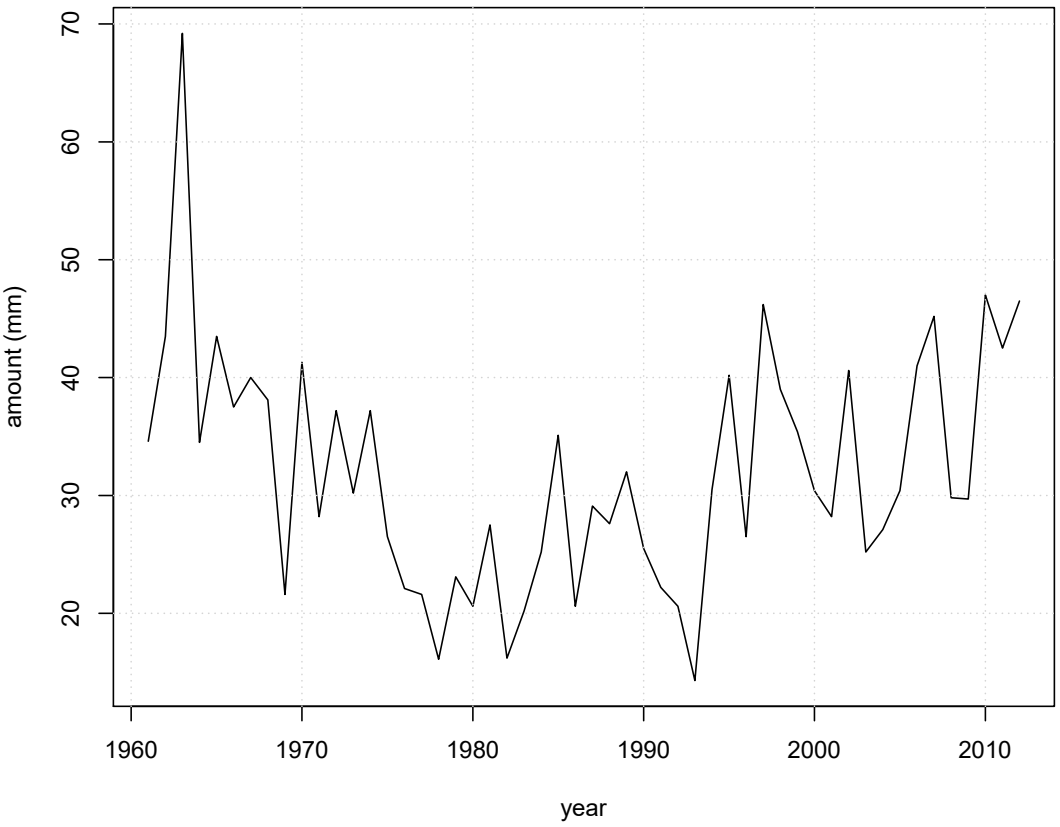


Figure 3.5: Maximum annual 1-day precipitation totals measured at Hustopeče for the period from 1961 to 2012

to use the L-moments method. In comparison to the moments method, in almost 42% of stations the estimated shape parameter k was less than $-\frac{1}{6}$. In contrast, the maximum likelihood method was always applicable.

3.5 Conclusion

This chapter presents asymptotic variances and covariances of parameter and quantile estimators based upon moments, L-moments, and maximum likelihood. They are then used to construct asymptotic confidence intervals of both parameter and quantile estimators. The results obtained by the simulation study performed show that the asymptotic confidence intervals for quantile estimates based on L-moments work well even for small sample size of 50 depending on the value of

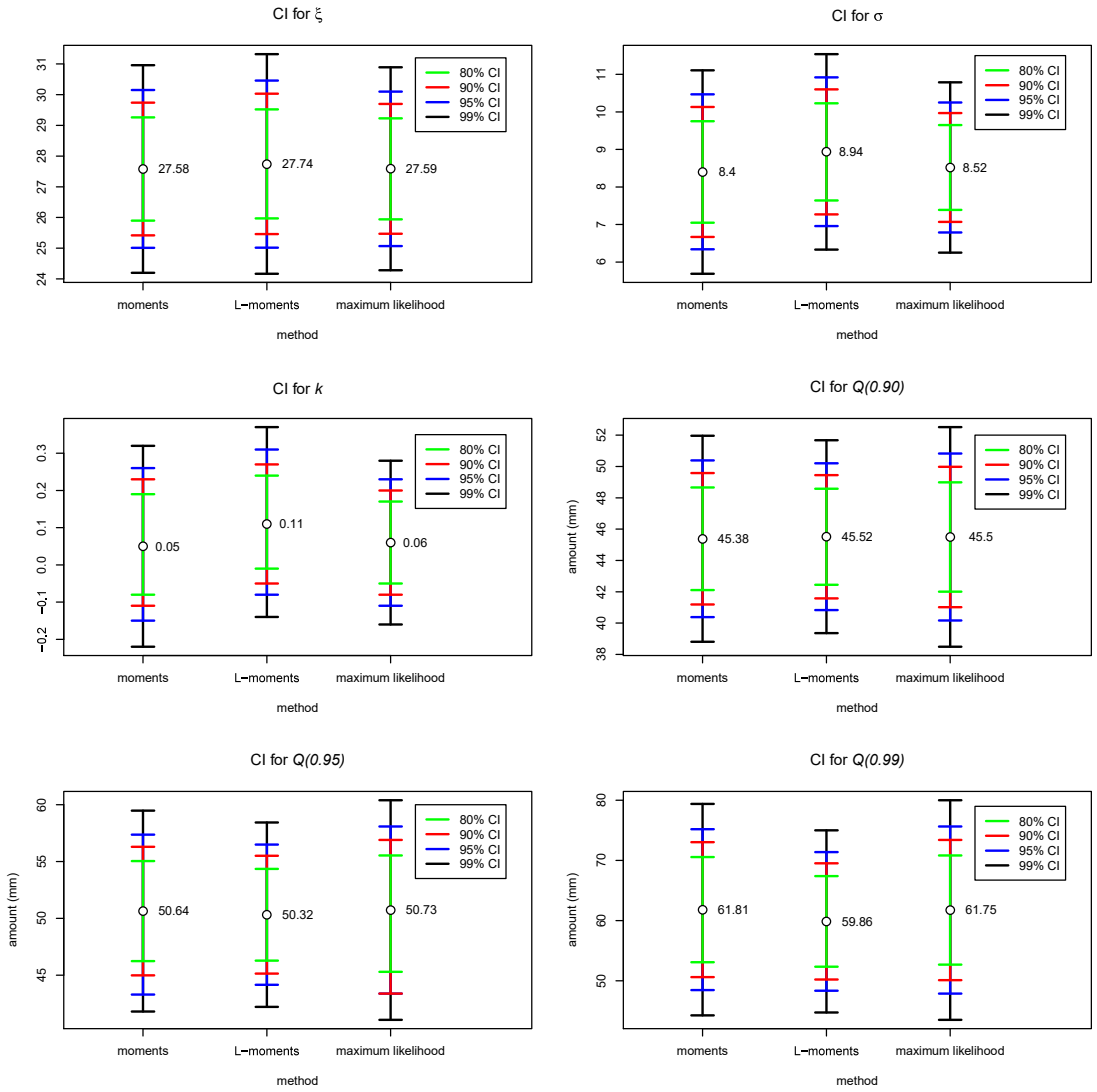


Figure 3.6: Approximate asymptotic confidence intervals for parameters and 90%, 95%, and 99% quantiles of the GEV distribution fitted to maximum annual 1-day precipitation totals measured at Hustopeče

the shape parameter k for both the GP and GEV distributions. In the case of the GP distribution, the confidence intervals based upon L-moments outperform those based on moments. It is difficult to recommend a single best method for estimating confidence intervals of quantiles for the GEV distribution, because the performances depend on the numerical values of the shape parameter k and sample size n. First, although it probably will be almost always possible to calculate the maximum likelihood-based confidence intervals in practice (the estimated values of the shape parameter k lie within $\left(-\frac{1}{2}, \frac{1}{2}\right)$ for many hydrological and meteorological

data sets), their performance is poor for higher numerical values of the shape parameter ($k \geq 0.2$) and smaller sample sizes ($n = 20, 50$). The moments method works very well when $k \geq 0.2$ for all samples, while the L-moments method is a little bit poorer. Of course, all the confidence intervals are accurate for large samples. It seems that, in practice, use of confidence intervals based both on moments and L-moments will be limited in quite restrictive requirements where $k > -\frac{1}{6}$ and $k > 0$, respectively. This was also confirmed in the case study performed, because the confidence intervals based upon moments and L-moments could not be calculated for many precipitation data sets. In relation to this problem there remains an issue for future consideration concerning comparison of the asymptotic confidence intervals to those confidence intervals obtained by bootstrap techniques. The use of bootstrap will eliminate the impracticability of the asymptotic confidence intervals using moments and L-moments, and it could lead to more accurate results in cases of small finite samples.

4

Index-Flood Based Bivariate RFA

Contents

4.1 Introduction

Considering the extremes of natural phenomena observed in many parts of the world over the past few decades, it is of great importance for planning and design engineering to have sufficient knowledge to construct a model providing adequate estimates of a given event. Usually, such interest is in the extreme event with a high return period T. To reliably estimate a quantile of a return period T, the observed data of length $n \geq T$ are generally required. In practice, this inequality is usually not satisfied, since the record length of annual data is typically less than 50 and often the return period is equal to 100 or even 1 000 [54]. However, the same variable is usually measured at many different sites. Due to the lack of observations in many environmental applications, regional frequency analysis (RFA) is useful, because it is an approach providing more accurate estimates of extreme hydrological and other environmental events compared to at-site approaches by taking into account data from a set of sites, which have probability distributions similar to that site of interest. The origins of RFA are traced to Dalrymple [31], who introduced the index-flood model, which is a way of pooling statistics of a set of different data samples. The method is based on the assumption that the sites have the same probability distributions apart from a site-specific scale factor termed the index-flood. Regions that meet the homogeneity condition are termed homogeneous, otherwise they are termed to be heterogeneous.

Let us have N sites in which data are available. The quantile function $Q_i(p), 0 < p < 1$, at site i may be estimated as

$$\widehat{Q}_i(p) = \widehat{\mu}_i \widehat{q}(p), \tag{4.1}$$

where $\widehat{\mu}_i$ is an estimate of the index-flood at site i (it is usually estimated by the at-site mean or median) and $\widehat{q}(p)$ is an estimate of the regional growth curve for probability p, which is a dimensionless quantile function of the probability distribution that is common to all sites in the region. The estimated regional quantile growth curve is obtained from the all rescaled data samples available

in the region. Several assumptions should be respected when using the index-flood based RFA, particularly that observations are identically distributed and serially independent at each site, observations are independent between sites, etc. Generally, RFA involves four main steps:

1. screening of the data,

2. identification of homogeneous regions,

3. choice and estimation of the regional probability distribution,

4. estimation of quantiles.

In the univariate framework, Hosking and Wallis [58] introduced index-flood based RFA using L-moments, which has become popular among practitioners. Several natural events may be described by multivariate characteristics which are not independent. Therefore it is important to jointly consider these characteristics to obtain the best information about the specific event. First attention to multivariate RFA was given recently in work of Chebana and Ouarda [22] dealing with generalization of the univariate Hosking and Wallis [58] L-moment homogeneity and discordancy tests using copulas [61, 85, 105] to assess the dependence structure between the variables of interest and multivariate L-moments [97] to construct the test statistics. The power of the proposed methodology to detect heterogeneity was illustrated in the bivariate case of flood events described by volume and peak. Chebana et al. [23] also studied practical aspects of the proposed L-moment homogeneity and discordancy tests on bivariate data corresponding to sites from a region in Quebec, Canada. Recently, Masselot, Chebana and Ouarda [78] introduced nonparametric procedures in the L-moment homogeneity test, which overcome drawbacks of the parametric test consisting of fitting a four-parametric kappa distribution and copula to the data, and also a rejection threshold based on simulations. Hence, these tests differ in the way of generating synthetic homogeneous regions and decision about homogeneity. Masselot, Chebana and Ouarda [78] proposed three nonparametric alternatives using the permutation

method, bootstrapping, and Pólya resampling in comparison to the parametric test, which uses the four-parametric kappa distribution and copula model when generating synthetic homogeneous regions. Better performance of the nonparametric version was achieved, because higher powers were obtained by Monte Carlo simulations. In particular, the nonparametric test based on the permutation method is the most powerfull amongst the rest of the generating methods. The quantile estimation step in the multivariate context was treated by Chebana and Ouarda [21] by adopting the multivariate quantile curves of Belzunce et al. [10], so the index-flood based univariate RFA using L-moments was completely generalized and simulations were carried out to evaluate the performance of the multivariate index-flood model. The bivariate RFA based upon the index-flood model was first applied using real data by Ben Aissia et al. [11]. Requena, Mediera and Garrotel [93] have recently presented a comprehensive stepwise procedure for multivariate index-flood model application whilst focusing on a bivariate case study situated in Spain.

This chapter is devoted to a comprehensive overview of procedure how to estimate high quantiles using index-flood based bivariate RFA. First, multivariate L-moments which play a key role in multivariate RFA are introduced in Section 4.2. Then, the methodology for constructing a bivariate distribution function using copula approach is presented in Section 4.3. The next four sections (Sections 4.4–4.7) show in details the main steps of the index-flood based RFA. Sections 4.4 and 4.5 deal with the L-moment discordancy, and parametric and nonparametric homogeneity tests, while the focus of Section 4.6 is on the selection and estimation of the regional distribution function that is common to all sites in the region except a site-specific scale factor. This is followed by estimating the quantile curve for a given probability or return period at the target site (Section 4.7). The chapter end with illustration of bivariate RFA applied to real meteorological data. This chapter relies mainly on the articles [116, 117].

4.2 Multivariate L-Moments

In the multivariate analysis of a d-variate random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)^T$, the two entities, which are generalizations of the univariate mean and variance, are usually employed. They are the mean vector

$$\mathrm{E}\boldsymbol{X} = (\mathrm{E}X_1, \mathrm{E}X_2, \ldots, \mathrm{E}X_d)^T = (\mu^{(1)}, \mu^{(2)}, \ldots, \mu^{(d)})^T$$

which consists of the means of each variable, and the variance-covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \mathrm{cov}(X_1, X_2) & \mathrm{cov}(X_1, X_3) & \ldots & \mathrm{cov}(X_1, X_d) \\ \mathrm{cov}(X_2, X_1) & \sigma_2^2 & \mathrm{cov}(X_2, X_3) & \ldots & \mathrm{cov}(X_2, X_d) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(X_d, X_1) & \mathrm{cov}(X_d, X_2) & \mathrm{cov}(X_d, X_3) & \ldots & \sigma_d^2 \end{pmatrix},$$

where σ_i^2 is the variance of the ith variable $X_i, i = 1, 2, \ldots, d$, and $\mathrm{cov}(X_i, X_j) = \mathrm{E}[(X_i - \mathrm{E}X_i)(X_j - \mathrm{E}X_j)]$ is the covariance of the variables X_i and $X_j, i, j = 1, 2, \ldots, d, i \neq j$. The central moments of third and higher orders were developed into the notion of central comoments, which are the elements of so-called comoment matrices, and they are sometimes employed in financial risk analysis (see, e.g., [26]). The rth central comoment of variable X_i with respect to X_j (in this order) is defined as

$$\xi_{r[ij]} = \mathrm{cov}(X_i, (X_j - \mu^{(j)})^{r-1}), r \geq 2,$$

where $\mu^{(j)}$ is the mean of the variable X_j. The second central comoment

$$\xi_{2[ij]} = \mathrm{cov}(X_i, X_j - \mu^{(j)}) = \mathrm{E}[(X_i - \mu^{(i)})((X_j - \mu^{(j)}) - 0)] = \mathrm{cov}(X_i, X_j)$$

is the usual covariance. The scale-free quantities are given by

$$\psi_{r[ij]} = \frac{\xi_{r[ij]}}{\sqrt{\sigma_i^2 (\sigma_j^2)^{r-1}}}, r \geq 2,$$

where σ_i^2 and σ_j^2 are the variances of the random variables X_i and X_j, respectively. The second, third, and fourth central rescaled comoments $\psi_{2[ij]}, \psi_{3[ij]}, \psi_{4[ij]}$ are called correlation, coskewness, and cokurtosis coefficients, respectively. The comoment

and comoment coefficients matrices include the corresponding central comoments and comoments coefficients

$$\Xi_r = \left(\xi_{r[ij]}\right)_{d\times d}, \quad \Psi_r = \left(\psi_{r[ij]}\right)_{d\times d}, \quad r \geq 2.$$

Increasing interest in modeling multivariate events by distributions with heavier tails has called for development of classical multivariate statistical analysis, because the classical approach is limited by moments assumptions of second and higher orders. Serfling and Xiao [97] were inspired by the idea of Hosking [55] that it would be promising to use the concomitants of order statistics to extend L-moments into the multivariate case. Serfling and Xiao [97] introduced multivariate L-moments as matrices with elements so-called L-comoments. L-comoments which measure association between two random variables are defined for all orders under only finite mean assumptions. They are robust analogues of central comoments and extension of univariate L-moments in the covariance representation given by (1.4) as well. In comparison to central comoments, L-comoments possess special features. This includes the already mentioned only first moment assumptions. Moreover, when the variables meet certain conditions, particularly when the variables are jointly distributed with affinely equivalent marginal distributions and one variable has linear regression on the other one, the L-comoments reduce to scalar multiple of univariate L-moments (see Section 4.2.1 for details). L-comoments and L-comoment coefficients are also bounded by their corresponding univariate L-moments and L-moment coefficients.

4.2.1 Population L-Comoments

Let us have a d-variate random vector $\mathbf{X} = (X_1, X_2, \ldots, X_d)^T$ and bivariate random vector $(X_i, X_j)^T, 1 \leq i, j \leq d, i \neq j$. Further, let $(X_i, X_j)^T$ has joint cumulative distribution function F, marginal distribution functions F_i, F_j, and finite means $\mu^{(i)}, \mu^{(j)}$. The rth L-comoment of variable X_i with respect to variable X_j (in this order) defined Serfling and Xiao [97] as

$$\lambda_{r[ij]} = \mathrm{cov}(X_i, P^*_{r-1}(F_j(X_j))), r \geq 2 \tag{4.2}$$

(the version $\lambda_{r[ji]}$ is defined similarly just with reversed order of the indices). Hence,

$$\begin{aligned}\lambda_{2[ij]} &= 2\mathrm{cov}\left(X_i, F_j(X_j)\right),\\ \lambda_{3[ij]} &= 6\mathrm{cov}\left(X_i, F_j^2(X_j) - F_j(X_j) + \frac{1}{6}\right),\\ \lambda_{4[ij]} &= \mathrm{cov}\left(X_i, 20F_j^3(X_j) - 30F_j^2(X_j) + 12F_j(X_j) - 1\right).\end{aligned}$$

Generally, $\lambda_{r[ij]}$ and $\lambda_{r[ji]}$ are not equal. The second to the fourth L-comoments, called L-covariance, L-coskewness, and L-cokurtosis, are robust alternatives to classical covariance, coskewness, and cokurtosis, respectively. In the case $i = j$, L-comoments reduce to univariate L-moments. When the variables X_i and X_j are independent, the L-comoments of second and higher orders are equal to zero.

Here, $\{\lambda_1^{(i)}, \lambda_2^{(i)}, \tau_3^{(i)}, \tau_4^{(i)}, \dots\}$ and $\{\lambda_1^{(j)}, \lambda_2^{(j)}, \tau_3^{(j)}, \tau_4^{(j)}, \dots\}$ are the sequences of population L-moments and L-moments ratios of the variables X_i and X_j, respectively. The L-covariance coefficient and L-comoment coefficients are given by

$$\begin{aligned}\tau_{2[ij]} &= \frac{\lambda_{2[ij]}}{\lambda_1^{(i)}},\\ \tau_{r[ij]} &= \frac{\lambda_{r[ij]}}{\lambda_2^{(i)}}, r \geq 3,\end{aligned}$$

respectively.

Serfling and Xiao [97] proved that in special cases L-comoments reduces to scalar multiple of univariate L-moments, that can be used in derivation of L-moments of some multivariate probability distribution, for example those presented in Table 4.1.

Theorem 14. *Let us have a bivariate random vector $(X_i, X_j)^T$ with joint cumulative distribution function F and marginal distribution functions F_i and F_j. If X_i has finite mean and linear regression on X_j, i.e., $\mathrm{E}(X_i|X_j) = a + bX_j$, then*

$$\lambda_{r[ij]} = b\lambda_r^{(j)}, r \geq 2. \tag{4.3}$$

Assuming also the marginal distribution functions F_i and F_j are affinely equivalent, i.e., for some constants θ and $\eta \neq 0$ holds $F_j(x) = F_i(\eta^{-1}(x - \theta))$, then

$$\tau_{r[ij]} = \rho_{[ij]}\tau_r^{(i)}, r \geq 3, \tag{4.4}$$

where $\rho_{[ij]} = \frac{\lambda_{2[ij]}}{\lambda_2^{(i)}}$ is L-correlation.

Proof. See [97], p. 15. □

Serfling and Xiao [97] also presented an important result about L-correlation $\rho_{[ij]}$: Its value lies between ± 1 as it is in the case of the classical Pearson correlation coefficient. Applying this result with the equality in (4.4), it is straightforward that

$$|\tau_{r[ij]}| \leq |\tau_r^{(i)}| < 1, r \geq 3.$$

An equivalent way how to define L-comoments is through a notion of concomitants. Let us have a sample $\{X_l^{(i)}, X_l^{(j)}, 1 \leq l \leq n\}$ from a bivariate distribution with joint cumulative distribution function F and marginal distribution functions F_i and F_j. When the variables $X^{(j)}$ are ordered in non-decreasing sequence $X_{1:n}^{(j)} \leq X_{2:n}^{(j)} \leq \cdots \leq X_{n:n}^{(j)}$, then the element of $\{X_1^{(i)}, X_2^{(i)}, \ldots, X_n^{(i)}\}$ that is paired with the element $X_{r:n}^{(j)}$ is called the concomitant of $X_{r:n}^{(j)}$ and it is denoted $X_{[r:n]}^{(ij)}$. Serfling and Xiao [97] defined L-comoments in terms of expected values of concomitants as follows

$$\lambda_{r[ij]} = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} \mathrm{E} X_{[r-k:r]}^{(ij)} = n^{-1} \sum_{k=1}^{n} w_{k:n}^{(r)} \mathrm{E} X_{[k:n]}^{(ij)}, r \geq 2,$$

where the weights $w_{k:n}^{(r)}$ are given by

$$w_{k:n}^{(r)} = \sum_{l=0}^{min\{k-1, r-1\}} (-1)^{r-1-l} \binom{r-1}{l} \binom{r-1+l}{l} \binom{n-1}{l}^{(-1)} \binom{k-1}{l}. \tag{4.5}$$

4.2.2 Sample L-Comoments

The unbiased estimator of the rth population L-comoment $\lambda_{r[ij]}$ is defined as a linear combination of concomitants

$$\widehat{\lambda}_{r[ij]} = n^{-1} \sum_{k=1}^{n} w_{k:n}^{(r)} x_{[k:n]}^{(ij)},$$

where the weights are defined by (4.5).

4.2.3 Multivariate L-Moments as L-Comoment Matrices

The multivariate L-moment of a d-variate random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)^T$ of the first order is just the vector mean

$$\boldsymbol{\Lambda}_1 = \mathrm{E}(X_1, X_2, \ldots, X_d)^T,$$

while the second and higher orders multivariate L-moments are defined in a matrix form with the elements being the corresponding L-comoments of variables $X_i, X_j, 1 \leq i, j \leq d$,

$$\boldsymbol{\Lambda}_r = \left(\lambda_{r[ij]}\right)_{d\times d}.$$

The second, third, and fourth multivariate L-moments $\boldsymbol{\Lambda}_2, \boldsymbol{\Lambda}_3$, and $\boldsymbol{\Lambda}_4$ are termed L-covariance, L-coskewness, and L-cokurtosis matrices, respectively. Scale-free versions of L-comoment matrices $\boldsymbol{\Lambda}_r, r \geq 2$, labelled as L-comoment coefficient matrices $\boldsymbol{\Lambda}_r^*$ consist of the corresponding L-comoments coefficients $\tau_{r[ij]}, 1 \leq i, j \leq d$,

$$\boldsymbol{\Lambda}_r^* = \left(\tau_{r[ij]}\right)_{d\times d}.$$

Particularly, for $r = 2$ the L-covariance coefficient matrix $\boldsymbol{\Lambda}_2^*$ is obtained. The diagonal elements of matrices $\boldsymbol{\Lambda}_r$ and $\boldsymbol{\Lambda}_r^*$ are obviously the univariate L-moments and L-moment ratios, respectively. Estimators of $\boldsymbol{\Lambda}_r$ and $\boldsymbol{\Lambda}_r^*$ are defined analogously with elements the L-comoment and L-comoment coefficient estimators.

Specifically, the first four L-moments of a bivariate random vector $\boldsymbol{X} = (X_1, X_2)^T$ are in the form

$$\begin{aligned}
\Lambda_1 &= (\mu^{(1)}, \mu^{(2)})^T, \\
\Lambda_2 &= \begin{pmatrix} \lambda_2^{(1)} & \lambda_{2[12]} \\ \lambda_{2[21]} & \lambda_2^{(2)} \end{pmatrix}, \\
\Lambda_3^* &= \begin{pmatrix} \tau_3^{(1)} & \tau_{3[12]} \\ \tau_{3[21]} & \tau_3^{(2)} \end{pmatrix}, \\
\Lambda_4^* &= \begin{pmatrix} \tau_4^{(1)} & \tau_{4[12]} \\ \tau_{4[21]} & \tau_4^{(2)} \end{pmatrix}.
\end{aligned}$$

Table 4.1 presents the first four L-moments of the three selected bivariate distributions computed using Theorem 14.

Table 4.1: L-moments of selected bivariate probability distributions

Distribution	Joint density function and L-moments
Normal	$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\cdot$ $\cdot \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right\},$ $x, y, \mu_1, \mu_2 \in \mathbb{R}; \sigma_1, \sigma_2 > 0, \rho \in (-1,1)$ $\mathbf{\Lambda}_1 = (\mu_1, \mu_2)^T, \mathbf{\Lambda}_2 = \frac{1}{\sqrt{\pi}}\mathbf{M}, \mathbf{\Lambda}_3 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \mathbf{\Lambda}_4 = 0.0702\mathbf{M},$ where $\mathbf{M} = \begin{pmatrix} \sigma_1 & \rho\sigma_1 \\ \rho\sigma_2 & \sigma_2 \end{pmatrix}$
Pareto type I	$f(x,y) = \frac{\alpha(\alpha+1)}{\sigma_1\sigma_2}\left(\frac{x}{\sigma_1} + \frac{y}{\sigma_2} - 1\right)^{-\alpha-2},$ $x \geq \sigma_1 > 0, y \geq \sigma_2 > 0, \alpha > 0$ $\mathbf{\Lambda}_1 = \frac{\alpha}{\alpha-1}(\sigma_1, \sigma_2)^T, \mathbf{\Lambda}_2 = \frac{1}{(\alpha-1)(2\alpha-1)}\mathbf{M},$ $\mathbf{\Lambda}_3 = \frac{\alpha+1}{(\alpha-1)(2\alpha-1)(3\alpha-1)}\mathbf{M}, \mathbf{\Lambda}_4 = \frac{(\alpha+1)(2\alpha+1)}{(\alpha-1)(2\alpha-1)(3\alpha-1)(4\alpha-1)}\mathbf{M},$ where $\mathbf{M} = \begin{pmatrix} \alpha\sigma_1 & \sigma_1 \\ \sigma_2 & \alpha\sigma_2 \end{pmatrix}$
Pareto type II	$f(x,y) = \frac{\alpha(\alpha+1)}{\sigma_1\sigma_2}\left(\frac{x-\mu_1}{\sigma_1} + \frac{y-\mu_2}{\sigma_2} + 1\right)^{-\alpha-2},$ $x > \mu_1, \mu_1 \in \mathbb{R}, \sigma_1 > 0, y > \mu_2, \mu_2 \in \mathbb{R}, \sigma_2 > 0, \alpha > 0$ $\mathbf{\Lambda}_1 = (\mu_1 + \frac{\sigma_1}{\alpha-1}, \mu + 2 + \frac{\sigma_2}{\alpha-1})^T, \mathbf{\Lambda}_2 = \frac{1}{(\alpha-1)(2\alpha-1)}\mathbf{M},$ $\mathbf{\Lambda}_3 = \frac{\alpha+1}{(\alpha-1)(2\alpha-1)(3\alpha-1)}\mathbf{M}, \mathbf{\Lambda}_4 = \frac{(\alpha+1)(2\alpha+1)}{(\alpha-1)(2\alpha-1)(3\alpha-1)(4\alpha-1)}\mathbf{M},$ where $\mathbf{M} = \begin{pmatrix} \alpha\sigma_1 & \sigma_1 \\ \sigma_2 & \alpha\sigma_2 \end{pmatrix}$

It is worth mentioning that second and higher orders L-moments proposed by Serfling and Xiao [97] in the matrix form with elements the L-moments given by (4.2) are a special case of multivariate L-moments introduced by Decurninge [33]. He proposed definition of multivariate L-moments under only finite mean assumptions similar to that one of univariate L-moments expressed as projections of the quantile function Q onto the shifted Legendre polynomials (see the definition (1.3)). The multivariate L-moment $\mathbf{\Lambda}_\alpha \in \mathbb{R}^d$ of multi-index α associated to the

transport $Q : [0,1]^d \to \mathbb{R}^d$ between the uniform distribution on $[0,1]^d$ and any measure ν on $\mathbb{R}^d$ defined by Decurninge [33] is in the form

$$\mathbf{\Lambda}_\alpha = \int\limits_{[0,1]^d} Q(t_1, t_2, \ldots, t_d) P^*_\alpha(t_1, t_2, \ldots, t_d) \mathrm{d}t_1 \mathrm{d}t_2, \ldots \mathrm{d}t_d, \tag{4.6}$$

where $\alpha = (i_1, i_2, \ldots, i_d) \in \mathbb{N}^d$ is a multi-index and $P^*_\alpha(u_1, u_2, \ldots, u_d) = \prod_{k=1}^d P^*_{i_k - 1}(u_k)$ is the multivariate Legendre polynomial ($P^*_{i_k-1}$ is the univariate Legendre polynomial given by (1.2)). However, there exist many ways how to transport a measure onto another one. One of them is Rosenblatt transport which is given by the successive conditional distributions $X_i | X_1 = x_1, X_2 = x_2, \ldots, X_{i-1} = x_{i-1}$ (see Rosenblatt [94] for the detailed transformation). The Serfling and Xiao [97] multivariate L-moments are obtained just when the quantiles based on the Rosenblatt transports are considered in (4.6). For a comprehensive theory of multivariate L-moments as collections of orthogonal projections of a multivariate quantile functions see Decurninge [33].

4.3 Bivariate Modeling

Identification and estimation of the multivariate distribution is a key task in RFA. Modeling the dependence structure between two random variables using a bivariate joint distribution such as a normal, Student's t, lognormal, exponential or Gumbel distribution is limited inasmuch as the margins must belong to the same family, but that is not usual in practice. Copula models first employed by Sklar [105] can overcome this limitation by coupling any marginal distribution functions to the joint distribution function. Hence, the multivariate distribution consists of two components – a copula and marginal distribution functions – which are selected independently of each other. This property of copula-based approach offers the possibility of constructing a large variety of joint distribution functions. For a comprehensive introduction to copulas, see monographs by Joe [60] and Nelsen [85].

4.3.1 Copulas

A copula is defined as a multivariate distribution function whose one-dimensional margins are uniform on the interval $[0, 1]$. In the particular case of a bivariate copula, it is a function C from $[0, 1]^2$ to $[0, 1]$ with the following properties:

1. $\forall u, v \in [0, 1]$: $C(u, 0) = C(0, v) = 0, C(u, 1) = u$ and $C(1, v) = v$,
2. $\forall u_1, v_1, u_2, v_2 \in [0, 1], u_1 \leq u_2, v_1 \leq v_2 : C(u_2, v_2) + C(u_1, v_1) \geq C(u_2, v_1) + C(u_1, v_2)$.

The well-known Sklar's theorem [105] states that the relationship between the bivariate cumulative distribution function F, univariate margins F_1 and F_2 and copula C is in the form

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2)),$$

where $F_1(X_1), F_2(X_2)$ are uniform on the interval $[0, 1]$.

There exists a large variety of copula families, usually categorized into four classes: Archimedean, extreme-value, elliptical, and other miscellaneous. Copulas may also be categorized by the number of parameters controlling the strength of dependence between variables. Although a wide range of families has been proposed in the literature, the Archimedean (Gumbel, Frank, Clayton) and elliptical (normal, Student's t) copula families are the most frequently used for modeling the joint distribution in hydrology and climatology [28, 39, 43, 81, 111, 122]. Let us discuss some important members of Archimedean and elliptical class of copula families.

Normal copula The normal (or Gaussian) copula is given by

$$\begin{aligned} C_\rho(u, v) &= \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v)) \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \exp\left[-\frac{s^2 - 2\rho st + t^2}{2(1-\rho^2)}\right] \mathrm{d}s\, \mathrm{d}t, \end{aligned}$$

where $\rho \in (-1, 1)$ is the dependence parameter, Φ^{-1} denotes the inverse of the standard univariate normal distribution function and Φ_ρ denotes the standard

bivariate normal distribution function with Pearson's correlation coefficient ρ. It is symmetric copula (i.e., $C(u, v) = C(v, u)\ \forall u, v \in [0, 1]$) and belongs to the elliptical class. Note that the normal copula cannot model tail dependence (see Section 4.3.3 for explanation). The scatterplots of 500 sample points drawn from the normal copula family are presented in Figure 4.1.

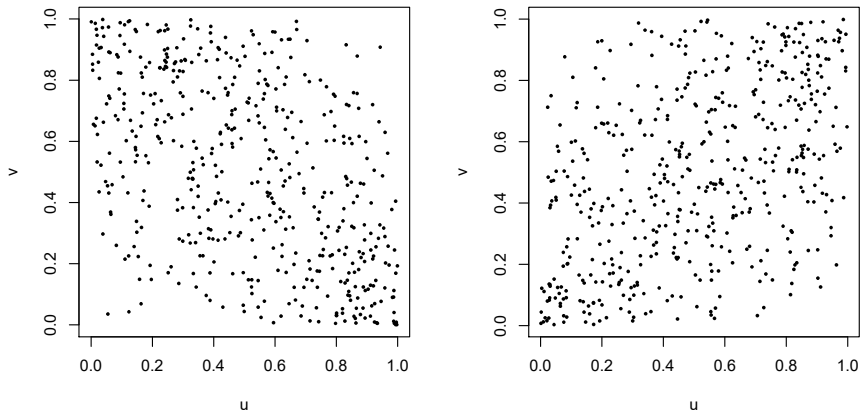


Figure 4.1: Normal copula, $\theta = -0.5$ (left) and $\theta = 0.5$ (right)

Student's t The Student's t copula that is closely related to the normal copula is given by

$$\begin{aligned} C_{\rho,\nu}(u, v) &= t_{\rho,\nu}(t_\theta^{-1}(u), t_\theta^{-1}(v)) \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{t_\nu^{-1}(u)} \int_{-\infty}^{t_\nu^{-1}(v)} \left[1 + \frac{s^2 - 2\rho st + t^2}{\nu(1-\rho^2)}\right] \mathrm{d}s\, \mathrm{d}t, \end{aligned}$$

where $\rho \in (-1, 1)$ is the dependence parameter, $\nu > 0$ denotes degrees of freedom, t_ν^{-1} denotes the inverse of the standard univariate Student's t distribution function and $t_{\rho,\nu}$ denotes the standard bivariate Student's t distribution function. It is symmetric copula and belongs to the elliptical class as well. In comparison to the normal copula the Student's t copula allows to model tail dependence. If $\nu \to \infty$, the Student's t copula converges to the normal copula. The scatterplots of 500 sample points drawn from the Student's t copula family are presented in Figure 4.2.

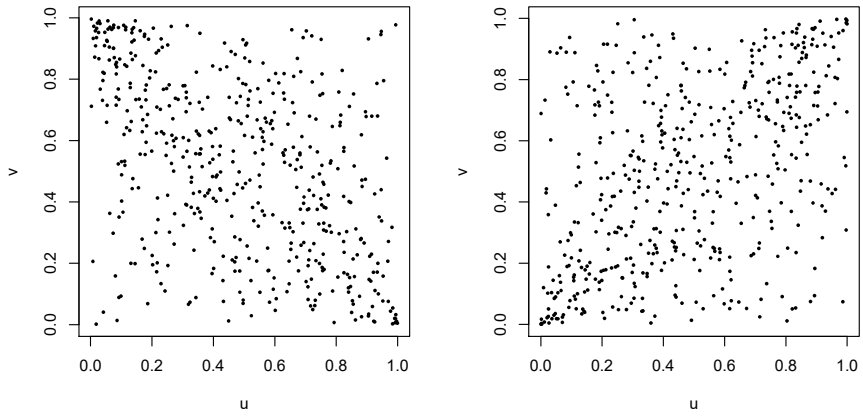


Figure 4.2: Student's t copula, $\theta = -0.5, \nu = 3$ (left) and $\theta = 0.5, \nu = 3$ (right)

Gumbel copula The Gumbel copula which is an extreme-value copula is given by

$$C_\theta(u, v) = \exp\left\{-\left\{\left[(-\log u)^\theta + (-\log v)^\theta\right]\right\}^{1/\theta}\right\},$$

where $\theta \in [1, \infty)$ is the dependence parameter. It is symmetric copula and belongs to the Archimedean class. The Gumbel copula can model only the upper-tail dependence. The scatterplots of 500 sample points drawn from the Gumbel copula family are presented in Figure 4.3.

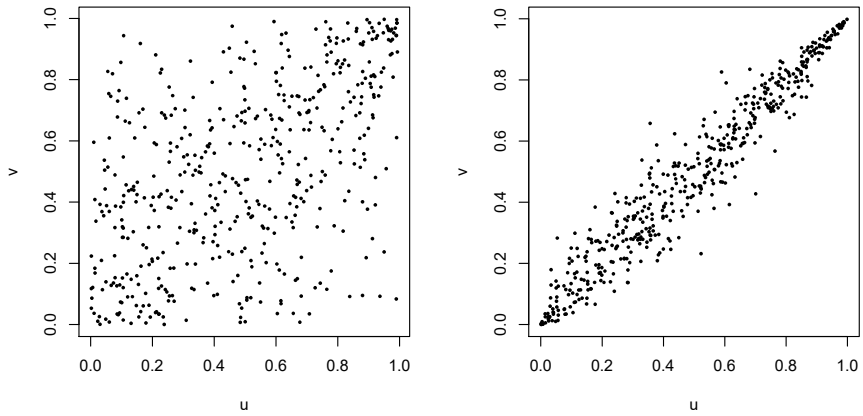


Figure 4.3: Gumbel copula, $\theta = 1.5$ (left) and $\theta = 7$ (right)

Frank copula The Frank copula is given by

$$C_\theta(u,v) = -\frac{1}{\theta}\log\left\{1 + \frac{[\exp(-\theta u)-1][\exp(-\theta v)-1]}{\exp(-\theta)-1}\right\},$$

where $\theta \in (-\infty,\infty)\setminus\{0\}$ is the dependence parameter. It is symmetric copula and belongs to the Archimedean class. The Frank copula cannot model tail dependence. The scatterplots of 500 sample points drawn from the Frank copula family are presented in Figure 4.4.

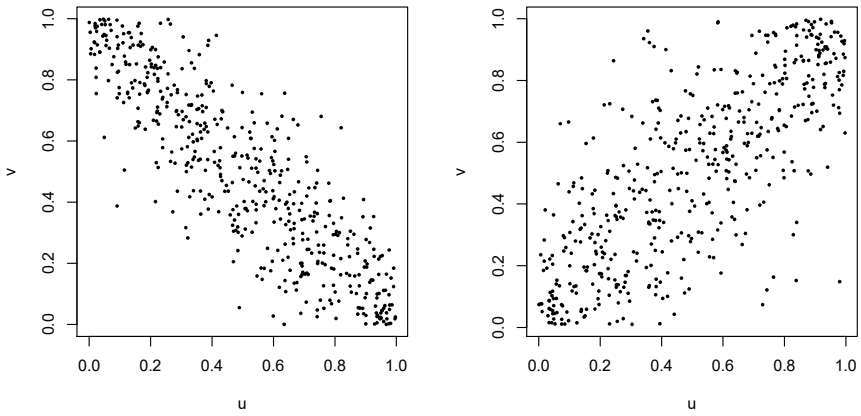


Figure 4.4: Frank copula, $\theta = -10$ (left) and $\theta = 7$ (right)

Clayton copula The Clayton copula is given by

$$C_\theta(u,v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta},$$

where $\theta \geq 0$ is the dependence parameter. It is symmetric copula and belongs to the Archimedean class. The Clayton copula can model only the lower-tail dependence. The scatterplots of 500 sample points drawn from the Clayton copula family are presented in Figure 4.5.

4.3.2 Copula's Parameter Estimation

When using copulas for modeling the dependence structure between two random variables, the problems of estimation of their parameter(s) appear. The parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^p, p \geq 1$, of the copula family $C_{\boldsymbol{\theta}}$ can be estimated using a

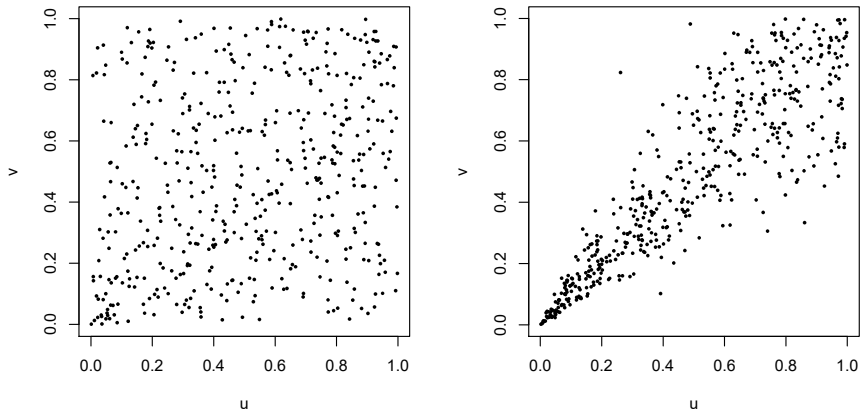


Figure 4.5: Clayton copula, $\theta = 0.5$ (left) and $\theta = 5$ (right)

nonparametric, semiparametric, or fully parametric approach. Several estimation methods have been developed, for instance the moment method based on the inversion of dependence measures such as the Spearman's ρ and Kendall's τ [85], maximum likelihood [60], maximum pseudo-likelihood [44], minimum-distance [112], and inference function for margins [61] methods, or the method based on bivariate L-moments [14]. Here, the focus is only on the maximum pseudo-likelihood method.

Let us have n independent copies $(X_1^{(1)}, X_1^{(2)})^T, (X_2^{(1)}, X_2^{(2)})^T, \ldots, (X_n^{(1)}, X_n^{(2)})^T$ of a bivariate vector $\boldsymbol{X} = (X^{(1)}, X^{(2)})^T$. The maximum pseudo-likelihood (MPL) estimator of $\boldsymbol{\theta}$ is determined by maximizing the log pseudo-likelihood function

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log c_{\boldsymbol{\theta}}(\widehat{U}_i, \widehat{V}_i), \tag{4.7}$$

where $\widehat{U}_i = R_i/(n+1)$ and $\widehat{V}_i = S_i/(n+1)$ are pseudo-observations (R_i is the rank of $X_i^{(1)}$ among $(X_1^{(1)}, X_2^{(1)}, \ldots, X_n^{(1)})^T$, and S_i is the rank of $X_i^{(2)}$ among $(X_1^{(2)}, X_2^{(2)}, \ldots, X_n^{(2)})^T$), and $c_{\boldsymbol{\theta}}(u,v) = \frac{\partial C_{\boldsymbol{\theta}}(u,v)}{\partial u \partial v}$ is the corresponding copula density that is absolutely continuous. Genest, Ghoudi and Rivest [44] showed that this estimator is consistent and asymptotically normal under certain regularity conditions.

4.3.3 Identification of Copula

There are variety of options to compare copula models, but there is no unique best way to do so. The selection approaches may be divided into numerical and graphical

methods. One of the numerical method are goodness-of-fit tests, which test the null hypothesis $H_0 : C \in \{C_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ against the alternative $H_1 : C \notin \{C_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$. Genest, Rémillard and Beaudoin [45] present and compare a variety of goodness-of-fit tests and recommend to use the test based on the Cramér-von Mises statistic

$$S_n = \int_{[0,1]^2} \mathbb{C}_n^2(u, v)\, \mathrm{d}C_n(u, v), \tag{4.8}$$

where $\mathbb{C}_n = \sqrt{n}(C_n - C_{\hat{\boldsymbol{\theta}}_n})$. The test statistic S_n compares the distance between the "empirical" copula C_n and copula $C_{\hat{\boldsymbol{\theta}}_n}$ obtained under the assumption of validity of H_0, which means that the true copula model belongs to a given copula model $C_{\boldsymbol{\theta}}$. See [45] for more details of this technique, in particular for a parametric bootstrap procedure to obtain approximate p-values.

Selection criteria, such as the Akaike (AIC) and Bayesian information criteria (BIC) are standard techniques for selecting the copula model which best fits the data among a set of possible models [3, 96]. They are not statistical tests in the sense of testing a null hypothesis in contrast with the goodness-of-fit tests. However, they enable to select from a group of several copula models the one which best fits the data. The AIC and BIC are defined as

$$\begin{aligned} AIC &= -2l(\widehat{\boldsymbol{\theta}}_n) + 2k, \\ BIC &= -2l(\widehat{\boldsymbol{\theta}}_n) + k \log n, \end{aligned} \tag{4.9}$$

where $l(\widehat{\boldsymbol{\theta}}_n)$ is the maximized value of the log pseudo-likelihood function for a given copula, k is the number of estimated parameters in the copula model and n is the number of observations. Usually, for all candidates AIC (or BIC) values are computed with the same set of observations and the best model is considered to be the one with the lowest AIC (or BIC) value. However, Burnham and Anderson [16] recommend to compute the AIC (or BIC) differences Δ_i for all candidate models as follows

$$\Delta_i = AIC_i - AIC_{min}$$

and to compare them on the basis of these differences. They also give guidelines for model support: Models with $0 \leq \Delta_i \leq 2$ have substantial support, models with $4 \leq \Delta_i \leq 7$ have considerably less support, and models with $\Delta_i > 10$ have no essential support.

The likelihood ratio tests, such as the Vuong test [119], are useful when comparing two competing copula models as well. The null hypothesis H_0 in the Vuong test may be interpreted as stating that the two competing models are equally distant from the true model. The two-sided alternative hypothesis then expresses that one of the models is closer to the true one. If A and B are two copula models with densities c_A and c_B and estimated parameters $\widehat{\boldsymbol{\theta}}_A$ and $\widehat{\boldsymbol{\theta}}_B$, respectively, then the Vuong test statistic

$$\nu = \frac{\sum_{i=1}^{n} m_i}{\sqrt{\sum_{i=1}^{n} (m_i - \bar{m})^2}},$$

where

$$m_i = \log \frac{c_A(\widehat{U}_i, \widehat{V}_i | \widehat{\boldsymbol{\theta}}_A)}{c_B(\widehat{U}_i, \widehat{V}_i | \widehat{\boldsymbol{\theta}}_B)}$$

and

$$\bar{m} = n^{-1} \sum_{i=1}^{n} m_i,$$

has asymptotically normal distribution under the assumption of the validity of H_0. Corrections corresponding to the penalty terms in the AIC and BIC are also possible when the copula models have different numbers of parameters.

When choosing an appropriate copula model, the attention should be also given to analysis of the copula behavior in extremes, i.e., in the upper-right and lower-left quadrant of the unit square. The tail dependence coefficients (TDC) depend on the copula dependence parameter $\boldsymbol{\theta}$ and they can be easily evaluated for many copula families, see for example [85]. As the interest in RFA is usually

in extreme high values, the focus is only on the upper TDC. The upper TDC is defined by the limit (if it exists)

$$\lambda_U = \lim_{t \to 1^-} \mathrm{P}[X_2 > F_2^{-1}(t) | X_1 > F_1^{-1}(t)], \tag{4.10}$$

where F_1 and F_2 are the distribution functions of variables X_1 and X_2, respectively. If C is the copula, Equation (4.10) may be rewritten to

$$\lambda_U = 2 - \lim_{t \to 1^-} \frac{1 - C(t,t)}{1 - t}. \tag{4.11}$$

Variables X_1 and X_2 are upper-tail dependent, if $\lambda_U \in (0, 1]$. The nonparametric estimator of λ_U proposed by Frahm, Junker and Schmidt [40] is

$$\widehat{\lambda}_U = 2 - 2\exp\left\{ n^{-1} \sum_{i=1}^{n} \log\left\{ \frac{\sqrt{\log \widehat{U}_i \cdot \log \widehat{V}_i}}{-2\log(\max\{\widehat{U}_i, \widehat{V}_i\})} \right\} \right\}. \tag{4.12}$$

To analyze the dependence in extremes, the estimator of the upper TDC $\widehat{\lambda}_U$ computed for the set of observations is compared to the upper TDC λ_U for each candidate copula family using an estimate of the parameter $\boldsymbol{\theta}$.

4.3.4 Marginal Distributions' Parameter Estimation

The parameters of the univariate distributions may be estimated by several methods, such as the L-moments method, that has been already reviewed in Section 1.4, and the traditional well-known moments and maximum likelihood methods already used in previous sections, that do not need to be discussed.

4.3.5 Identification of Marginal Distributions

In identification of marginal distributions the goodness-of-fit tests may be first employed to reduce a set of possible candidates of univariate distributions. There are various tests available for testing the goodness-of-fit of the observed data to a specific distribution. The most popular tests are the empirical distribution function tests, which are based on comparison between the hypothetical distribution function

$F_{\boldsymbol{\theta}}$ and empirical distribution function F_n, such as the Cramér-von Mises and Anderson-Darling tests with test statistics

$$W^2 = n \int_{-\infty}^{\infty} [F_n(x) - F_{\boldsymbol{\theta}}(x)]^2 \, \mathrm{d}F(x),$$

$$A^2 = n \int_{-\infty}^{\infty} \frac{[F_n(x) - F_{\boldsymbol{\theta}}(x)]^2}{F_{\boldsymbol{\theta}}(x)[1 - F_{\boldsymbol{\theta}}(x)]} \, \mathrm{d}F(x),$$

respectively. When the parameter $\boldsymbol{\theta}$ is partially or completely unspecified, it must be replaced by its estimator $\widehat{\boldsymbol{\theta}}$. The distributions of the test statistics A^2 and W^2 are then dependent on the specific distribution $F_{\boldsymbol{\theta}}$ that is tested. The estimation method, the presence of the shape parameter, and the sample size affect the distribution of the test statistic as well. Laio [72] presented a complete testing procedure for a set of probability models commonly used in extreme value analysis, including the normal, lognormal, and GEV distributions, among others.

Two or more probability distributions can be also compared by using the AIC and BIC already introduced in Section 4.3.3, which may be written for the marginal models as

$$AIC = -2l(\widehat{\boldsymbol{\theta}}_n) + 2k,$$

$$BIC = -2l(\widehat{\boldsymbol{\theta}}_n) + k \log n,$$

where $l(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(x_1, x_2, \ldots, x_n; \boldsymbol{\theta})$ is the log-likelihood function, $f(x_1, x_2, \ldots, x_n; \boldsymbol{\theta})$ is the probability density function of a random variable X, $\widehat{\boldsymbol{\theta}}_n$ is the maximum likelihood estimator of $\boldsymbol{\theta}$, and n is the number of observations. The best marginal distribution is selected by applying the Burnham and Anderson's [16] guidelines for AIC (or BIC) Δ_i values as it has been already discussed in Section 4.3.3.

4.4 Discordancy Test

The first step in any data analysis is to check that the data are suitable for the analysis. Sample L-moments and L-moment ratios may be used to reveal incorrect data values and outliers of the sample. The aim of the L-moment discordancy test is to detect those sites which are discordant with the group of sites as a

whole. The multivariate discordancy test is just an extension of the univariate Hosking and Wallis [58] discordancy test.

Let us have a group with N sites and the sample L-CV, $\mathbf{\Lambda}_2^{*(i)}$, L-skewness, $\mathbf{\Lambda}_3^{*(i)}$, and L-kurtosis, $\mathbf{\Lambda}_3^{*(i)}$, coefficient matrices for site i. The discordancy measure for site i is

$$\mathbf{D}_i = \frac{1}{3}(\mathbf{U}_i - \bar{\mathbf{U}})^T \mathbf{S}^{-1} (\mathbf{U}_i - \bar{\mathbf{U}}),$$

where

$$\mathbf{U}_i = [\mathbf{\Lambda}_2^{*(i)} \mathbf{\Lambda}_3^{*(i)} \mathbf{\Lambda}_4^{*(i)}]^T, \quad \bar{\mathbf{U}} = N^{-1} \sum_{i=1}^{N} \mathbf{U}_i,$$

and

$$\mathbf{S} = (N-1)^{-1} \sum_{i=1}^{N} (\mathbf{U}_i - \bar{\mathbf{U}})(\mathbf{U}_i - \bar{\mathbf{U}})^T.$$

The site i is discordant if $||\mathbf{D}_i||$ exceeds the critical value, which is equal to the constant 2.6049 for large regions [22]. $||\cdot||$ denotes an arbitrary matrix norm, however, the spectral matrix norm $||\mathbf{A}|| = \sqrt{\text{maximum eigenvalue of } \mathbf{A}^T\mathbf{A}}$ is the recommended one [22]. Chebana and Ouarda [22] draw attention to the critical value for small regions which should be obtained for example by bootstrap techniques. The sites flagged as discordant should be further checked out for errors. When there is an error, the discordant site must be removed from the group. But sometimes the site is discordant due to real observed outliers, and, hence, it should not be excluded from the further analysis.

4.5 Homogeneity Tests

Here, the steps of both parametric and nonparametric procedures introduced in [22] and [78] are presented. In the nonparametric case, the focus is only on the permutation test, because it is the most powerful among all compared procedures [78]. For each test suppose that the region has N sites, with site i having record length n_i, the sample L-moment ratios $t_2^{(i)}, t_3^{(i)}, t_4^{(i)}$ and L-covariance coefficient matrix $\mathbf{\Lambda}_2^{*(i)}$.

4.5.1 Parametric Test

1. Compute the statistic

$$V_{||\cdot||} = \left(\frac{\sum_{i=1}^{N} n_i ||\boldsymbol{\Lambda}_2^{*(i)} - \bar{\boldsymbol{\Lambda}}_2^{*}||^2}{\sum_{i=1}^{N} n_i} \right)^{1/2}, \tag{4.13}$$

where $\bar{\boldsymbol{\Lambda}}_2^{*} = \left(\sum_{i=1}^{N} n_i \boldsymbol{\Lambda}_2^{*(i)} \right) / \sum_{i=1}^{N} n_i$ is the sample regional L-covariance coefficient matrix.

2. Determine a bivariate copula family, that is common for all sites in the region. At each site a suitable copula is chosen by a combination of techniques described in Section 4.3.3, and the copula which is suitable for the most of sites is used for generating homogeneous synthetic regions. Estimate their parameters $\boldsymbol{\theta}_i$ for each site $i, i = 1, \cdots, N$, by the MPL method presented in Section 4.3.2.

3. Obtain the regional model parameters of the regional bivariate distribution that is given by the regional copula family determined by step 2 and the four-parametric kappa distribution with the cumulative distribution function given by

$$F(x) = \left\{ 1 - h \left[1 - \frac{k(x-\xi)}{\sigma} \right]^{1/k} \right\}^{1/h}$$

with lower bound $\xi + \frac{\sigma}{k(1-h^{-k})}$ if $h > 0$, $\xi + \frac{\sigma}{k}$ if $h \leq 0$ and $k < 0$, or $-\infty$ if $h \leq 0$ and $k \geq 0$, and upper bound $\xi + \frac{\sigma}{k}$ if $k > 0$, or ∞ if $k \leq 0$.

   (a) The regional copula's parameters are estimated as the weighted mean of the at-site estimates obtained by step 2, i.e., $\widehat{\boldsymbol{\theta}}^R = \sum_{i=1}^{N} n_i \widehat{\boldsymbol{\theta}}_i / \sum_{i=1}^{N} n_i$.

   (b) The regional parameters of the four-parametric kappa distribution are estimated using the L-moment method proposed by Hosking [55] by fitting the four-parametric kappa distribution to the regional L-moment ratios $(1, t_2^R, t_3^R, t_4^R)$, where t_k^R is a weighted mean of the at-site L-moment ratios for $k = 2, 3, 4$, i.e., $t_k^R = \sum_{i=1}^{N} n_i t_k^{(i)} / \sum_{i=1}^{N} n_i$.

4. Generate a large number N_{sim} of homogeneous regions (500 regions is enough [22]) with N sites, each having the same record length as its real-world counterpart. To get a sample with uniform margins use the regional copula given by step 2 with parameters estimated in step 3 (a), and to get the desired sample use the quantile function of a four-parametric kappa distribution with parameters estimated in step 3 (b).

5. Compute the statistic $V_{||\cdot||}^{(j)}$ defined by formula (4.13) on each of the simulated homogeneous region, $j = 1, 2, \ldots, N_{sim}$. Standardize $V_{||\cdot||}$ computed on the observed data in step 1 by the sample mean μ and standard deviation σ of the computed values of $V_{||\cdot||}^{(j)}$ for a large number of simulated regions, i.e.,

$$H_{||\cdot||} = \frac{V_{||\cdot||} - \mu}{\sigma}. \tag{4.14}$$

6. Categorize the region: The region is declared to be homogeneous if $H_{||\cdot||} < 1$, acceptably homogeneous if $1 \leq H_{||\cdot||} < 2$, and definitely heterogeneous if $H_{||\cdot||} \geq 2$.

Note that other measures used by Hosking and Wallis [58] may also be considered in the multivariate case to detect heterogeneity.

4.5.2 Nonparametric Permutation Test

1. Choose a significance level $\alpha \in (0, 1)$.

2. Calculate $V_{||\cdot||}$ defined by (4.13) on the observed data as in the first step of the parametric test.

3. Generate a large number N_{sim} of homogeneous regions, which means to reassign randomly the pooled data between N sites while preserving the real-world at-site record lengths.

4. Compute the statistic $V_{||\cdot||}^{(j)}$ defined by (4.13) on each of the simulated homogeneous regions, $j = 1, \ldots, N_{sim}$.

5. Compute the p-value given by

$$p - value = \frac{1}{N_{sim}} \#\{V^{(j)} > V_{||\cdot||}\}. \tag{4.15}$$

The null hypothesis of homogeneity is rejected if $p - value < \alpha$.

4.6 Choice and Estimation of Regional Distribution

The regional distribution consists of the regional copula C^R, which best describes the dependence between the variables, and regional marginal distribution functions, F_1^R and F_2^R, which best fit the univariate data, in the entire homogeneous region. Before selection of the regional copula and regional marginal distribution functions the observations are first standardised by the index-flood vector, which is usually estimated by the vector of the at-site means or medians. Here, the at-site means are used. Hence, the standardised data series at site i are

$$(\tilde{x}_{ij}^{(1)}, \tilde{x}_{ij}^{(2)})^T = \left(\frac{x_{ij}^{(1)}}{\mu_i^{(1)}}, \frac{x_{ij}^{(2)}}{\mu_i^{(2)}} \right)^T, i = 1, 2, \ldots, N, j = 1, 2, \ldots, n_i,$$

where $\mu_i^{(1)}$ and $\mu_i^{(2)}$ are the sample means of variables X_1 and X_2 at site i, respectively. Then the standardised data series are pooled that they create one data series $(\tilde{x}_k^{(1)}, \tilde{x}_k^{(2)}), k = 1, 2, \ldots, m$, of length $m = \sum_{i=1}^{N} n_i$. The bivariate regional distribution is obtained on the pooled vector by using the techniques presented in Sections 4.3.3 and 4.3.5. Finally, the parameter of the bivariate regional distribution is estimated as

$$\widehat{\boldsymbol{\theta}}^R = \frac{\sum_{i=1}^{N} n_i \widehat{\boldsymbol{\theta}}_i}{\sum_{i=1}^{N} n_i},$$

where $\widehat{\boldsymbol{\theta}}_i$ is the estimated parameter at site i obtained by the estimation methods presented in Sections 4.3.2 and 4.3.4.

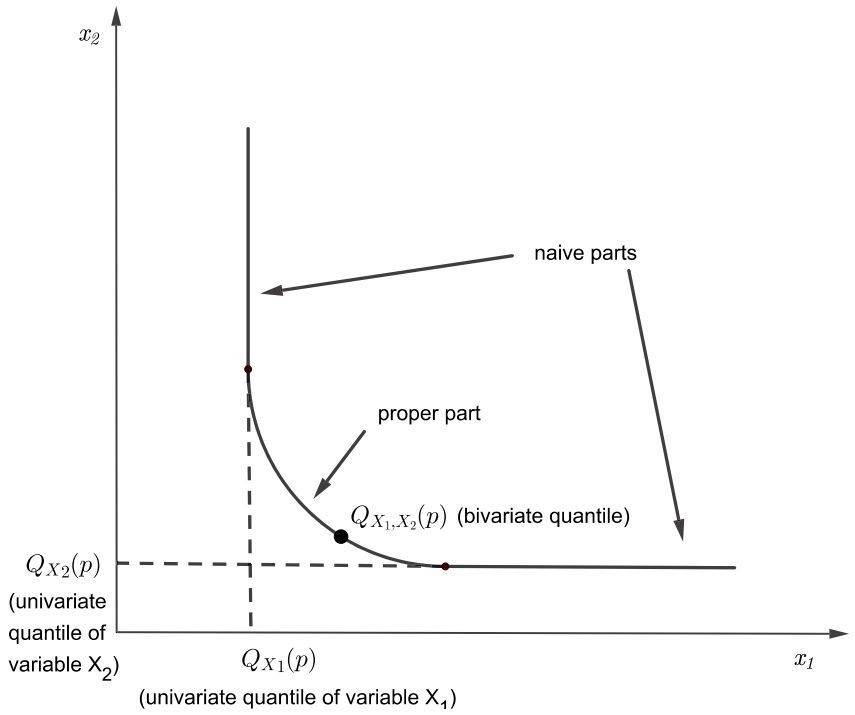


Figure 4.6: Illustration of bivariate quantile curve

4.7 Estimation of Quantile Curves

For the event of simultaneous non-exceedance $\{X_1 \leq x_1 \wedge X_2 \leq x_2\}$, Belzunce et al. [10] expressed a bivariate quantile curve for a given probability $p \in (0,1)$ as follows

$$\begin{aligned} \boldsymbol{Q}_{X_1,X_2}(p) = \{(x_1,x_2) \in \mathbb{R}^2 \text{ s.t. } x_1 = F_1^{-1}(u), x_2 = F_2^{-1}(v); \\ u,v \in [0,1] : C(u,v) = p\}, \end{aligned} \tag{4.16}$$

where C is the copula and F_1 and F_2 are the marginal distribution functions of variables X_1 and X_2, respectively. The bivariate quantile curve is an infinite set of pairs of values x_1, x_2 leading to the same probability p, and it is composed of two parts: the proper part (central part of the curve) and naive part (constant tails), see Figure 4.6. The constant values correspond to the univariate quantiles. This version of bivariate quantile was adopted by Chebana and Ouarda [21] to modify the univariate index-flood model defined by (4.1).

The bivariate index-flood model is in the form

$$\left(\widehat{\boldsymbol{Q}}_{X_1,X_2}(p)\right)_i = \begin{pmatrix}\widehat{\mu^{(1)}}\\ \widehat{\mu^{(2)}}\end{pmatrix}_i \widehat{\boldsymbol{q}}_{X_1,X_2}(p), i = 1, 2, \ldots, N, \tag{4.17}$$

where $\widehat{\boldsymbol{q}}_{X_1,X_2}$ is the regional quantile curve, that is dimensionless and it is common to all sites in the region, and $\left(\widehat{\mu^{(1)}}, \widehat{\mu^{(2)}}\right)_i^T$ is the index-flood vector [21]. The regional quantile curve $\widehat{\boldsymbol{q}}_{X_1,X_2}(p)$ for a specific probability p is obtained by considering the regional bivariate distribution consisting of C^R, F_1^R, and F_2^R with the regional parameter $\hat{\boldsymbol{\theta}}^R$ in the formula (4.16). The index-flood vector $\left(\widehat{\mu^{(1)}}, \widehat{\mu^{(2)}}\right)_i^T$ is estimated by the vector of at-site means $(\mu_i^{(1)}, \mu_i^{(2)})^T$. The index-flood vector at ungaged sites may be estimated by multivariate multiple linear regression model with regressors the selected physiographic characteristics of the sites [11].

Note that the bivariate quantile curve used in the index-flood model given by (4.17) is associated to the notion of the OR return period $T^{OR}_{X_1,X_2}$ of a given event, because

$$\begin{aligned} T^{OR}_{X_1,X_2} &= \frac{1}{\mathrm{P}\{X_1 > x_1 \vee X_2 > x_2\}} = \frac{1}{1 - \mathrm{P}\{X_1 \le x_1 \wedge X_2 \le x_2\}} = \\ &= \frac{1}{1 - C(u,v)} = \frac{1}{1-p}. \end{aligned}$$

For example, the 95% quantile curve corresponds to the joint 20-year OR return value curve.

4.8 Illustration of Application of Bivariate RFA

For illustration, the methodology previously described is applied to model bivariate extreme precipitation events characterized by 1- and 5-day maximum annual precipitation totals.

4.8.1 Study Regions

Maximum annual 1- and 5-day precipitation totals measured for the periods of 33–47 years at 210 stations covering the area of the Czech Republic were used as the input dataset (a large majority of the station records includes the whole

period 1961–2007, i.e., 47 years, but some station records are shorter). The data were provided by the Czech Hydrometeorological Institute (CHMI), where they underwent basic quality checking (the methodology used for checking for gross errors and missing readings is described in [29]). The vast majority of the data was also thoroughly checked for missing readings by Kyselý [68].

The delineation of the area shown in Figure 4.7 is based on the delineation obtained by cluster analysis presented in [69, 71]. Originally, the area of the Czech Republic was divided into four regions [71]. However, the regionalization was redefined by Kyselý [69] after data from more stations and longer periods were included in the study, because some regions did not meet the homogeneity condition. Here, small regions 1a, 1b, 1c, 2a, and 2b presented in [69] have been united back together as it was in [71]. This step is justified, because the stations have similar site characteristics. Hence, six regions were obtained:

Region 1 Seventy-five stations in the lowlands (along the Elbe, Morava, Dyje and Svratka rivers); elevations ranging from 150 to 400 m above sea level (a.s.l.).

Region 2 Seventy-nine stations at higher altitudes in the western and central parts of the Czech Republic; elevations ranging from 410 to 1 118 m a.s.l.

Region 3 Thirty-three stations in the northeastern part of Moravia and Silesia (near the Jeseníky and Beskydy Mountains); elevations ranging from from 220 to 1 490 m a.s.l.

Region 4 Sixteen stations in the northeastern part of Bohemia (near the Krkonoše and Orlické Mountains); elevations ranging from 255 to 572 m a.s.l.

Region 5a Four stations in the northernmost part of the Czech Republic (near the Lužické Mountains); elevations ranging from 315 to 440 m a.s.l.

Region 5b Three stations in the northernmost part of the Czech Republic (near the Jizerské Mountains); elevations ranging from 398 to 778 m a.s.l.

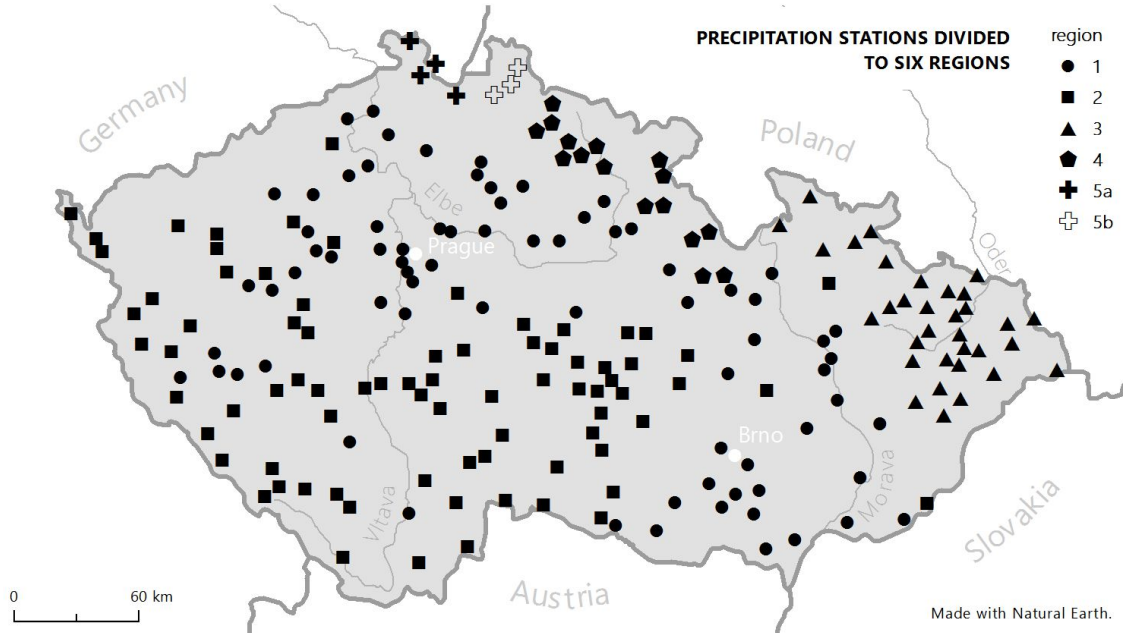


Figure 4.7: Location of stations and delineation of six regions

Locations of stations and formation of six regions are shown in Figure 4.7. Basic information concerning the datasets for each region is summarized in Table 4.2: the number of stations N, overall record length, average, minimal and maximal record length, average altitude, range of the sample Pearson's ρ_p, Spearman's ρ_s and Kendall's τ correlation coefficients between maximum annual 1- and 5-day precipitation totals. Naturally, the studied variables are positively correlated. The sample Kendall's τ correlation coefficient for each station lies within $[0.266, 0.712]$, most of them (about 92%) exceed the value 0.4. Hence, bivariate RFA should be preferred to the univariate ones.

4.8.2 Discordancy Test

First, the datasets are checked for occurrence of discordant sites. In every region, except regions 3, 5a, and 5b, at least one site is flagged as discordant (see Table 4.3). Since regions 5a and 5b are very small, the discordancy measure $||\mathbf{D}_i||$ does not give any information: In region 5b there are only 3 stations and the matrix $\mathbf{S}$ is singular, hence, the discordancy measure $||\mathbf{D}_i||$ cannot be calculated. In region 5a there are 4 stations, the matrix $\mathbf{S}$ is already regular, but for each i the matrix $\mathbf{D}_i$ is nearly the

Table 4.2: Basic information on the input datasets

Region	1	2	3	4	5a	5b
No. of stations	75	79	33	16	4	3
Overall record length (years)	3 438	3 633	1 508	719	188	141
Minimal record length (years)	33	37	33	36	47	47
Maximal record length (years)	47	47	47	47	47	47
Average record length (years)	45.8	46	45.7	44.9	47	47
Average altitude (m a.s.l.)	270.1	550.3	411.1	412.6	361	523.3
Sample Pearson's ρ_P range	[0.494, 0.872]	[0.500, 0.899]	[0.584, 0.885]	[0.516, 0.825]	[0.671,0.775]	[0.781, 0.826]
Sample Spearman's ρ_S range	[0.386, 0.887]	[0.477, 0.878]	[0.513, 0.806]	[0.372, 0.873]	[0.485,0.626]	[0.773,0.860]
Sample Kendall's τ range	[0.268, 0.712]	[0.339, 0.706]	[0.362, 0.621]	[0.266, 0.595]	[0.346, 0.468]	[0.587, 0.671]

identity matrix and $||\mathbf{D}_i||$ is equal to a constant regardless of the data values. None of the sites recognized as discordant was excluded from the further analysis, since all values of the discordancy measure larger than a critical value at the significance level 5% have originated from real observed outliers. Note that the spectral matrix norm is used for evaluating both the discordancy and heterogeneity measures.

Table 4.3: Number of discordant stations

Region	1	2	3	4	5a	5b
No. of discordant stations	4	7	0	1	–	–

4.8.3 Copula Selection

To model the dependence between 1- and 5-day maximum annual precipitation totals, 28 one-parameter bivariate copula families are considered:

- 22 one-parameter Archimedean copula families($C\#1 - C\#22$) described by Nelsen [85] including Clayton (C), Frank (F), Joe (J), and Ali-Mikhail-Haq (AMH) copula families,
- 3 extreme-value copula families: Gumbel (GH), Galambos (G), and Hüsler-Reiss (HR) [60],
- 2 meta-elliptical copula families: normal (N) and Student's t with 4 degrees of freedom (S) [60, 111],
- 2 other miscellaneous copula families: Farlie-Gumbel-Morgenstern (FGM) and Plackett (P) [60].

The above-mentioned copula family's functions, their parameter and Kendall's τ ranges are listed in [82]. The test space is constructed on the dependence ranges for the sample Kendall's τ correlation coefficient. The sample Kendall's τ correlation coefficient for all stations lies within $[0.266, 0.712]$, therefore some families may be eliminated. First, 5 Archimedean families ($C\#7, C\#9, C\#10, C\#11$, and $C\#22$)

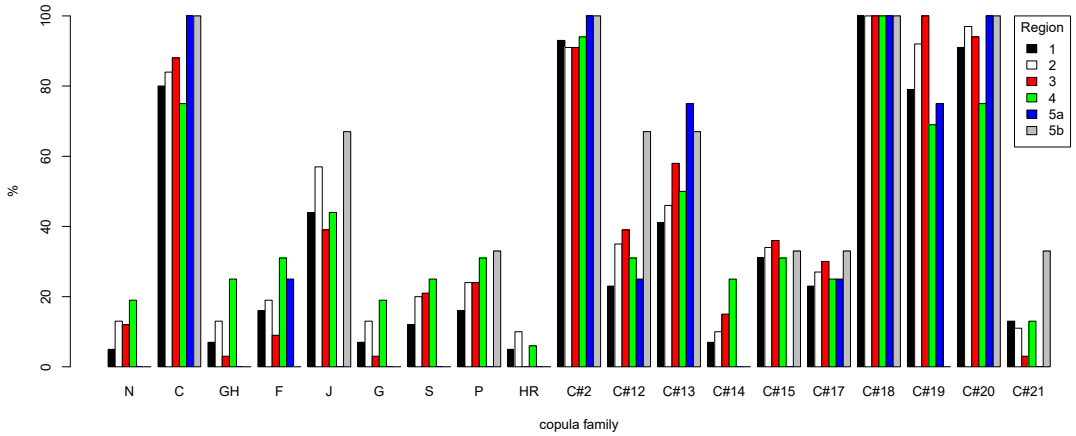


Figure 4.8: Relative frequency of rejecting the null hypothesis that the unknown bivariate copula family belongs to a given copula family at the 5% significance level

are excluded due to the fact that they are just negatively ordered. Since only for 4 stations the sample Kendall's τ is equal or smaller than $\frac{1}{3}$, further 3 Archimedean (Ali-Mikhail-Haq, $C\#8$, and $C\#16$) and the Farlie-Gumbel-Morgenstern copula families are excluded (the majority of computed values of the sample Kendall's τ are beyond the prescribed ranges of these copulas: $\left[\frac{1}{3}(5-8\log 2), \frac{1}{3}\right]$ for Ali-Mikhail-Haq, $\left[-1, \frac{1}{3}\right]$ for $C\#8$ and $C\#16$, and $\left[-\frac{2}{9}, \frac{2}{9}\right]$ for Farlie-Gumbel-Morgenstern copula families). Finally, the test space consists of 19 copula families – see the horizontal axis in Figure 4.8.

As a first step towards dependence model selection the formal goodness-of-fit test is carried out by evaluating the Cramér-von Mises goodness-of-fit statistic given by (4.8). A parametric bootstrap with the number of bootstrap samples $M = 1\ 000$ described by Genest, Rémillard and Beaudoin [45] was performed to obtain an approximate p-value for the test based on S_n. Each column of Figure 4.8 shows the percentage of rejection of the null hypothesis H_0 for a given copula family and region. The results provide sufficient evidence for rejecting the Clayton, $C\#2, C\#18, C\#19$ and $C\#20$ copula families.

In the second stage, the log pseudo-likelihood function given by (4.7) is maximized for the remaining candidates, and the AIC value is calculated using (4.9). Each column of Figure 4.9 shows the percentage of stations in which a given copula

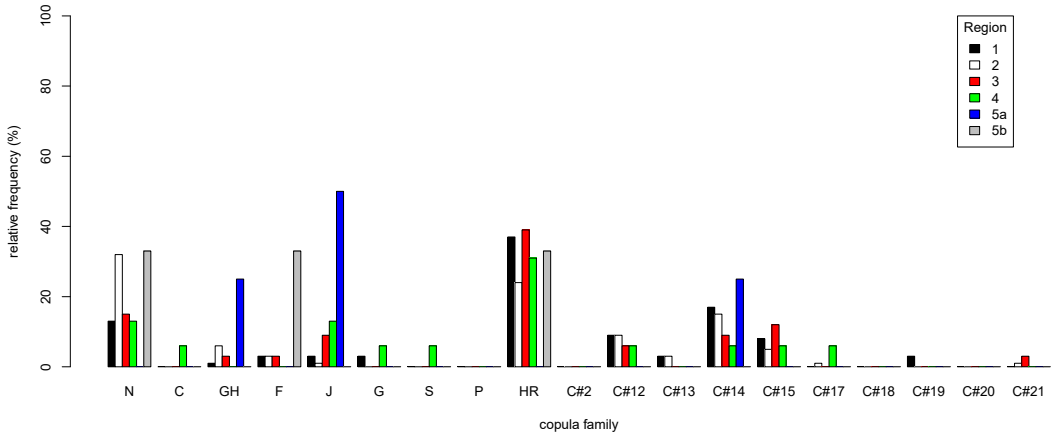


Figure 4.9: Relative frequency of stations in which a given bivariate copula achieves the minimum AIC value

family achieves the minimum value. Generally, it is very difficult to identify in practice only one distribution (both univariate or multivariate), which fits well the data for each site. From Figure 4.9 results that the copula family with the highest percentage (about 30%) in regions 1, 3, and 4 is the Hüsler-Reiss, and in region 2 it is the normal copula family. The copula family with the second highest percentage (about 15%) in regions 1, 3, and 4 is the normal, and in region 2 the Hüsler-Reiss copula family. Selection of the best fitted copula family is much more complicated when the regions are very small. The Joe copula family fits the data best at two of the four stations in region 5a, while the Gumbel and $C\#14$ copula families seem to be the best for data at the remaining stations. The normal copula family, as well as the Hüsler-Reiss and Frank families, fit the data in the region 5b best.

The last step of the copula selection consists of analyzing behaviour in the upper tail. There is apparently a positive dependence in the upper tail of the distribution of the studied variables as expected due to the nature of the data: The estimated upper TDC for each station lies within $[0.382, 0.787]$. The normal copula family, as well as the Clayton, Frank, Plackett, $C\#13, C\#17, C\#19$, and $C\#20$ copula families are upper-tail independent. On the other hand, the upper TDC is nonzero for extreme-value (Gumbel, Galambos, and Hüsler-Reiss), Joe, Student's t, $C\#2, C\#12, C\#14, C\#15, C\#18$, and $C\#21$ copula families. When

Table 4.4: Description of the most suitable copula families for modeling the dependence structure between 1- and 5-day precipitation totals

Copula	Properties (by rows: copula function; parameter range; Kendall's τ range; lower and upper TDC)
Joe	$1-[(1-u)^{\theta}+(1-v)^{\theta}-(1-u)^{\theta}(1-v)^{\theta}]^{1/\theta}$
	$\theta \in [1,\infty)$
	$\tau \in [-1,1]$
	$0; 2-2^{1/\theta}$
Hüsler-Reiss	$\exp\left\{\log u \cdot \Phi\left(\frac{1}{\theta}+\frac{\theta}{2}\log\frac{\log u}{\log v}\right)+\log v \cdot \Phi\left(\frac{1}{\theta}+\frac{\theta}{2}\log\frac{\log v}{\log u}\right)\right\}$
	$\theta \in [0,\infty)$
	$\tau \in [0,1]$
	$0; 2(1-\Phi\left(\frac{1}{\theta}\right))$
Gumbel	$\exp(-\{[(-\log u)^{\theta}+(-\log v)^{\theta}]\}^{1/\theta})$
	$\theta \in [1,\infty)$
	$\tau \in [0,1]$
	$0; 2-2^{1/\theta}$

comparing the estimated upper TDC computed using (4.12) to the upper TDC for each of the upper-tail dependent copula families obtained as the limit given by (4.11) with the MPL estimator of θ, the Gumbel, Galambos, and Hüsler-Reiss (therefore extreme-value), Joe and $C\#15$ (therefore Archimedean) copula families indicate the smallest difference.

Finally, including results of the formal goodness-of-fit test, AIC, and estimation of the upper TDC in the copula selection, the Hüsler-Reiss copula family was selected as the best model in regions 1, 2, 3, 4, and 5b, whereas the Gumbel and Joe copula families in region 5a. Summary of the copula families (their function, parameter range, and upper TDC) that are used to model dependence between 1- and 5-day maximum annual precipitation amounts is listed in Table 4.4.

4.8.4 L-Moment Homogeneity Tests

To construct the parametric L-moment homogeneity test statistic, 500 homogeneous regions were generated in accordance with the model defined by the copula family displayed in Table 4.4 and the four-parametric kappa distribution for the margins (the copula parameter is estimated by the MPL method, while the marginal parameters by the L-moments method). To estimate p-values given by (4.15), 500 synthetic regions were generated by permuting bivariate data between sites. The nonparametric L-moment homogeneity test is performed on the significance level $\alpha = 5\%$. Table 4.5 compares the results obtained by parametric and nonparametric homogeneity tests. The parametric one gives evidence about homogeneity of all regions, because values of the heterogeneity measure $H_{||\cdot||}$ are less than 2, moreover some values are even negative. The nonparametric version rejects the null hypothesis of homogeneity only for region 1 on the chosen significance level. Note that the unification of regions 5a and 5b would lead to a strong infraction of homogeneity.

Table 4.5: Results of bivariate homogeneity testing

Region	Parametric test		Nonparametric test	
	$H_{\|\|\cdot\|\|}$	Decision	P-value	Decision
1	1.4884	Possibly homogeneous	0.006	Reject H_0
2	1.1316	Possibly homogeneous	0.226	Do not reject H_0
3	-1.3857	Homogeneous	0.994	Do not reject H_0
4	0.7111	Homogeneous	0.090	Do not reject H_0
5a	-1.5404 (J) -1.5305 (G)	Homogeneous	0.976	Do not reject H_0
5b	-0.7635	Homogeneous	0.770	Do not reject H_0

4.8.5 Estimation of Regional Distributions and Quantile Curves

Since the homogeneity condition has been satisfied for all regions, the bivariate index-flood model defined by (4.17) may be employed to estimate the at-site quantile curves, sometimes called the joint T-year OR return value curves due to the relationship between the notions quantile and return period. For these purposes, the regional bivariate distributions (copulas and marginal cumulative distribution functions) have to be identified. Here, the three-parametric GEV and generalized logistic (GLO) distributions with location ξ, scale σ, and shape k parameters are used for estimating cumulative distribution functions of 1- and 5-day precipitation totals, because Kyselý and Picek [70] identified the GEV distribution as the most suitable distribution for modeling maximum annual 1- to 7-day precipitation totals according to the goodness-of-fit test based on L-kurtosis and L-moment ratio diagram. Only in the northeast region the GLO distribution was preferred. Hence, the focus is not on identification of the marginal distributions, because this step has been already carried out by Kyselý and Picek [70]. For estimation of quantile curves, the same copula families as those employed in parametric L-moment homogeneity testing are used. This is justified by the fact that RFA assumes that all sites within a region have a similar probability distribution (apart from a site-specific scale factor). Moreover, problems with identification of a single copula which fits all the standardized data from the entire region may appear, as discussed by Ben Aissia et al. [11]. Consequently, the regional bivariate distributions are defined by one of the Hüsler-Reiss, Joe, or Gumbel copula families, and the GEV or GLO distributions for the margins using the regional parameters given in Table 4.6 depending on the specific region.

The estimated regional quantile curves given by (4.16) for probabilities $p = 0.9, 0.95, 0.99, 0.995$, and 0.999 are illustrated in Figure 4.10. Regions 3 and 5b have definitely the highest upper-tail quantiles as compared to the other regions. With increasing probability p the length of the proper part shortens. For example, Figure 4.11 shows the simultaneous non-exceedance event for region 3 and $p = 0.95$. The

Table 4.6: Parameters estimates of the bivariate regional distributions

Region	Marginal parameters						Copula parameter
	1-day totals			5-day totals			
	ξ	σ	k	ξ	σ	k	
1	0.8178	0.2535	-0.1195	0.8271	0.2378	-0.1234	2.0091
2	0.8122	0.2584	-0.1214	0.8292	0.2461	-0.0978	2.0934
3	0.9189	0.1670	-0.2645	0.8770	0.1688	-0.3710	1.9231
4	0.8312	0.2376	-0.1094	0.8660	0.2018	-0.0776	1.6921
5a	0.8453	0.2133	-0.1304	0.8435	0.2086	-0.1506	2.2789 (J) 1.8442 (G)
5b	0.7642	0.2924	-0.1899	0.7752	0.2742	-0.1961	2.4369

combinations $(2.33, 1.78)$ and $(1.66, 2.95)$ correspond to approximate coordinates of the first point in which the quantile curve is constant, while the values 1.66 and 1.78 correspond to the univariate regional quantiles. When taking this combination of univariate regional quantile values, the smaller probability $p' = 0.93$ is obtained. Hence, the univariate approach overestimates the probability of the non-exceedance simultaneous event, and, generally, univariate RFA leads to biased results. Since in region 5a the Joe and Gumbel copula families were identified as the best model to describe the dependence structure, two quantile curves are obtained for each probability p. The differences between the quantile curves in the proper part are not at all significant, moreover the difference decreases with increasing probability p as it is shown in Figure 4.10. Confidence regions of the quantile curves would be of interest, however, no attention has been given to techniques to assess the error in their estimation for bivariate RFA. Nevertheless, the approach of Coblenz, Dyckerhoff and Grothe [27] seems to be probably developed to obtained them.

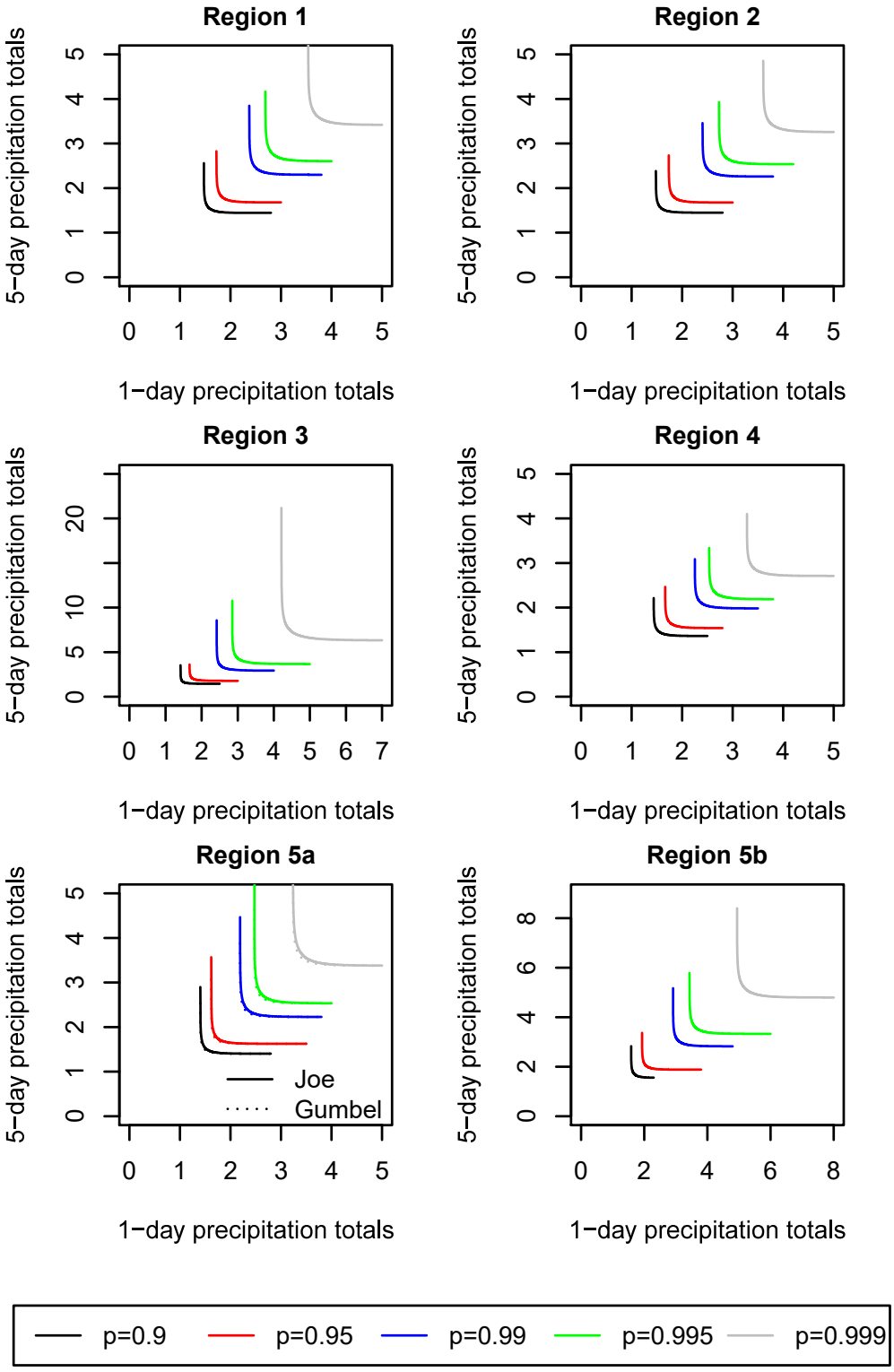


Figure 4.10: Estimated regional quantile curves for probabilities $p = 0.9$, 0.95, 0.99, 0.995, and 0.999

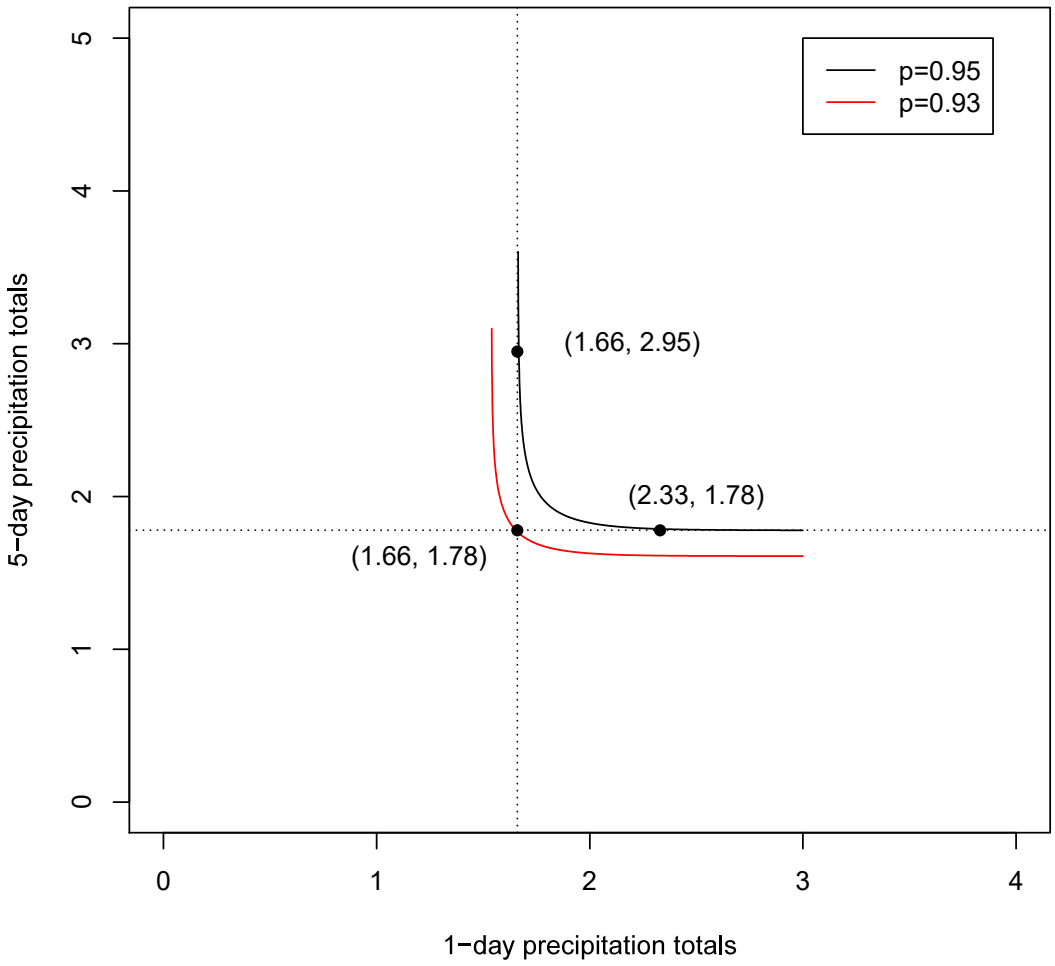


Figure 4.11: Comparison of regional quantiles obtained by univariate and bivariate RFA for region 3

5

L-Moment Homogeneity Test in Trivariate RFA of Extreme Precipitation Events

Contents

5.1 Introduction

All studies dealing with multivariate RFA have focused only on the bivariate case, however, because the study of multivariate copulas becomes more complicated in higher dimensions and, apart from the multivariate elliptical families (Gaussian, Student's t), their use is in applications time consuming. Under certain conditions,

the Archimedean class of bivariate copulas can be extended to d dimensions termed symmetric or exchangeable Archimedean copulas (EACs) [60]. This multivariate copula is specified by only one generator and therefore provides a little flexibility inasmuch as the associations between all pairings of d variables are identical. A generalization of multivariate EACs, termed nested Archimedean copulas, first discussed by Joe [60], allows for modeling correlation by bivariate Archimedean copulas between $d-1$ pairs, the remainder being given implicitly. Pair-copula construction, also known as vine copulas, is clearly the most flexible tool for constructing a multivariate copula model. The vine methodology was also originally discussed by Joe [60], and studied in detail by Aas and Berg [1], Bedford and Cooke [8, 9], Kurowicka and Cooke [66], Kurowicka [67] and Czado, Frigessi and Bakken [30], among others. Vine copulas enable multivariate dependence modeling to be enhanced by specifying $d(d-1)/2$ bivariate copulas that need to belong to neither the same class nor the same family.

In this chapter the practical aspects of trivariate L-moment homogeneity tests for extreme precipitation events in the Czech Republic are investigated. In the case of parametric L-moment homogeneity testing, a large number of synthetic homogeneous regions must be generated using copulas and the four-parametric kappa distribution to construct a test statistic. Hence, a special attention should be given to identification of an appropriate probability distribution of the considered trivariate vector with the components representing 1-, 3-, and 7-day maximum annual precipitation amounts that is common for all sites within the region apart from a site-specific scale factor. Although homogeneity testing may be simplified by using the nonparametric test, because in this case homogenous regions are generated just by permuting the pooled data between sites in a region, identification of an appropriate multivariate distribution still plays an essential role in RFA to obtain reliable quantile estimates. Moreover, the parametric approach is usually the preferred one, because when the model is well determined the information about data and their features is taken into account in the procedure through model. This chapter is organized as follows. First, the input dataset is described in Section 5.2.

The copula models used are briefly presented in Section 5.3. Section 5.4 provides results obtained in the case study, mainly choice of the appropriate trivariate copula model, which best represents the dependence structure between the variables, and checking of the homogeneity condition. The chapter ends with a summary section. This chapter relies on the article [116].

5.2 Input Dataset

Maximum annual 1-, 3-, and 7-day precipitation totals (variables X_1, X_3, and X_7) measured at stations covering the area of the Czech Republic were used as the input dataset. The number of stations and their delineation to the regions is the same as in bivariate analysis presented in Section 4.8, while the period of measurements increased from 35–47 to 35–52 years. Basic information concerning the datasets for each region is summarized in Table 5.1: the number of stations, overall record length, average, minimum and maximum record lengths, as well as the range of the sample Kendall's τ between all pairs of variables. Naturally, each pair of variables is positively correlated.

5.3 Multivariate Copula Models

As it has been already written the copula "couples" marginal distribution functions into a multivariate distribution function. This is explained by the Sklar's theorem [105] in d dimensions

$$F(x_1, x_2, \ldots, x_d) = C(F_1(x_1), F_2(x_2), \ldots, F_d(x_d)),$$

where F is a d-dimensional cumulative distribution function, $F_1, F_2, \ldots, F_d$ are univariate marginal distribution functions, and C is a copula.

Widely used copulas both bivariate and multivariate, and particularly in actuarial and finance applications, are elliptical. The most well-known members are the normal and Student's t copula families already discussed in Section 4.3.1. However, these families lack the flexibility. Here, the copula models for modeling higher-dimensional dependence are reviewed.

Table 5.1: Basic information on the input dataset

Region	1	2	3	4	5a	5b
No. of stations	75	79	33	16	4	3
Overall record length (years)	3775	4022	1663	782	208	156
Minimal record length (years)	38	42	36	35	52	52
Maximal record length (years)	52	52	52	52	52	52
Average record length (years)	50.3	50.9	50.4	48.9	52	52
Sample Kendall's τ_{13} range	[0.340, 0.715]	[0.399, 0.710]	[0.416, 0.668]	[0.390, 0.626]	[0.488, 0.574]	[0.697, 0.743]
Sample Kendall's τ_{17} range	[0.273, 0.662]	[0.217, 0.702]	[0.325, 0.602]	[0.188, 0.578]	[0.328, 0.436]	[0.551, 0.637]
Sample Kendall's τ_{37} range	[0.493, 0.765]	[0.459, 0.772]	[0.571, 0.807]	[0.447, 0.669]	[0.574,0.610]	[0.687, 0.803]
No. of discordant stations	14	9	5	1	–	–

5.3.1 Exchangeable Archimedean Copulas

The first important class of multivariate copula models consists of the EACs. An exchangeable Archimedean d-copula is an extension of the well-known bivariate Archimedean copula to d dimensions

$$C(u_1, u_2, \ldots, u_d) = \phi^{-1}(\phi(u_1) + \phi(u_2) + \cdots + \phi(u_d)),$$

where ϕ, termed a generator of the copula C, is a continuous, strictly decreasing function from $[0, 1]$ to $[0, \infty)$ such that $\lim_{t\to 0} \phi(t) = \infty$ and $\phi(1) = 0$, and its inversion ϕ^{-1} is completely monotonic. Table 5.2 presents some common bivariate Archimedean copulas, their probability and generator functions, which may be extended to higher dimension. Despite the EACs' simple construction, these copulas are greatly limited in their usefulness for assessing dependence structure. This is because the correlation between pairs of variables is identical, and that is not usual in practical applications. The structure for the 3-dimensional case is shown in Figure 5.1a.

The parameter of the EAC may be estimated by the usual MPL method. The algorithm for sampling a d-variate EAC with generator ϕ is as follows [76]:

1. Sample $V \sim F = \mathcal{LS}^{-1}(\phi)$, where $\mathcal{LS}^{-1}(\phi)$ denotes the inverse Laplace-Stieltjes transform of ϕ (the Laplace transform of the cumulative distribution function F is defined as $\mathcal{LS}[F](t) = \int_0^\infty \exp(-tx)\, \mathrm{d}F(x), t \in [0, \infty)$).

2. Sample $X_i \sim U(0, 1)$ for each $i = 1, 2, \ldots, d$.

3. Return $(U_1, U_2, \ldots, U_d)^T$, where $U_i = \phi(-\log(X_i)/V), i = 1, 2, \ldots, d$.

5.3.2 Fully Nested Archimedean Copulas

More flexibility in dependence modeling is provided by another generalization of the bivariate Archimedean copulas discussed by Joe [60] in the form

$$C(u_1, u_2, \ldots, u_d) = C_1(u_d, C_2[u_{d-1}, C_3\{u_{d-2}, \ldots, C_{d-1}(u_2, u_1) \ldots\}])$$

Table 5.2: Some common bivariate Archimedean copula families

Copula	Probability function $C_\theta(u,v)$	Generator function $\phi(t)$
Ali-Mikhail-Haq	$\frac{uv}{1-\theta(1-u)(1-v)}, \theta \in [-1,1)$	$\log\left(\frac{1-\theta(1-t)}{t}\right)$
Clayton	$(u^{-\theta}+v^{-\theta}-1)^{-1/\theta}, \theta \in [-1,\infty)\setminus\{0\}$	$\frac{1}{\theta}(t^{-\theta}-1)$
Frank	$-\frac{1}{\theta}\log\left(1+\frac{[\exp(-\theta u)-1][\exp(-\theta v)-1]}{\exp(-\theta)-1}\right), \theta \in (-\infty,\infty)\setminus\{0\}$	$-\log\left(\frac{\exp(-\theta t)-1}{\exp(\theta)-1}\right)$
Gumbel	$\exp(-\{[(-\log u)^\theta+(-\log v)^\theta]\}^{1/\theta}), \theta \in [1,\infty)$	$(-\log t)^\theta$

This is termed a fully nested Archimedean copula (FNAC), inasmuch as the first variables u_1 and u_2 are coupled by copula C_{d-1}, then variable u_3 and copula C_{d-1} are coupled by copula C_{d-2}, variable u_4 and copula C_{d-2} are coupled by copula C_{d-3}, and so on. The generator ϕ_i and its inversion ϕ_i^{-1} of copula C_i must satisfy the same conditions as for EACs. Moreover, $\phi_i \circ \phi_i^{-1}$ must belong to the class of infinitely differentiable increasing functions from $[0, \infty)$ to $[0, \infty)$ with alternating signs for the derivatives of all orders for all $i = 1, 2, \ldots, d-1$. Note that the generators ϕ_i can come from different generator families. The structure $s = ((12)3)$ for the 3-dimensional case is shown in Figure 5.1b.

The bivariate copulas may be selected and their parameters may be estimated by a recursive way described by Okhrin and Ristig [87]:

1. The couple of variables with the strongest dependency is selected. The copula of these two variables is selected and its parameters are estimated by the usual MPL method.

2. The selected couple define a new pseudo-variable.

3. The remaining variables and the pseudo-variable create a set of variables.

4. Iterate.

The sampling procedure was presented by Hofert and Maechler [51] as follows: Let C be a FNAC with root copula C_0 generated by ϕ_0. Let $\boldsymbol{U}$ be a vector of the same dimension as C.

1. Sample $V_0 \sim F_0 = \mathcal{LS}^{-1}(\phi_0)$.

2. For all components u of C_0 that are nested Archimedean copulas do:

   (a) Set C_1 with the generator ϕ_1 to the nested Archimedean copula u.

   (b) Sample $V_{01} \sim F_{01} = \mathcal{LS}^{-1}(\phi_{01}(\cdot; V_0))$.

   (c) Set $C_0 = C_1, \phi_0 = \phi_1$, $V_0 = V_{01}$, and continue with step 2.

3. For all other components u of C_0 do:

   (a) Sample $R \sim Exp(1)$.

   (b) Set the component of $\boldsymbol{U}$ corresponding u to $\phi(R/V_0)$

4. Return $\boldsymbol{U}$.

5.3.3 Vines

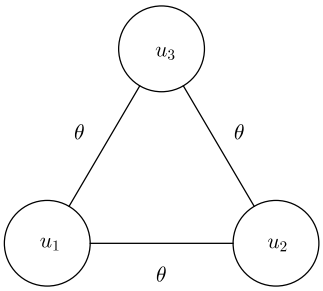
Vines apparently comprise the most flexible tool for modeling dependence structure in higher dimensions. The vine methodology is based on decomposition of the d-variate density of a random vector into a product of d univariate densities and $d(d-1)/2$ bivariate copula densities ($d-1$ of which are unconditional while the rest are conditional). These bivariate copulas may belong to different classes. Two main types thoroughly discussed in the literature are C-vines and D-vines [1, 8, 9, 30, 66]. Note that there are three C-vine and D-vine structures in the 3-dimensional case but they are the same. Hence, only three different decompositions can be obtained by permuting variables u_1, u_2 and u_3 in Figure 5.1c.

For the purposes of Chapters 5 (C-vines and D-vines in a 3-dimensional case are the same) and 6 (see Section 6.2 for explanation) the focus is only on D-vine copulas.

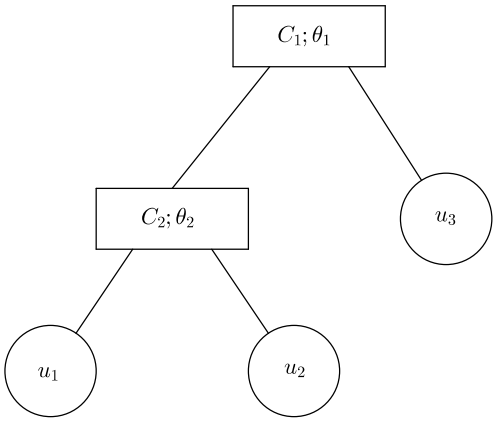
A d-dimensional D-vine copula may be graphically represented by $(d-1)$ trees $T_i, i = 1, \ldots, d-1$, consisting of $(d-i+1)$ nodes and $(d-i)$ edges. Each edge corresponds to a bivariate copula density (unconditional in the first tree, conditional in the rest of the trees). Bedford and Cooke [8] stated the density of a d-dimensional random vector in term of a D-vine as

$$f(x_1, x_2, \ldots, x_d) = \prod_{k=1}^{d} f_k(x_k) \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{j,j+i|1,\ldots,j-1}(F(x_j|x_1, x_2, \ldots, x_{j-1}), F(x_{j+i}|x_1, x_2, \ldots, x_{j-1})),$$

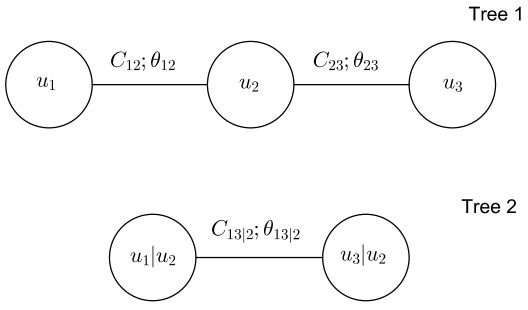
where $f_k, k = 1, \ldots, d$, denotes the univariate density, $c_{j,j+i|1,\ldots,j-1}$ bivariate copula density with the parameter $\theta_{j,j+i|1,\ldots,j-1}$, and $F(x_j|x_1, \ldots, x_{j-1})$ and $F(x_{j+i}|x_1, \ldots, x_{j-1})$ marginal conditional distributions [$F(x|\mathbf{v}) = \frac{\partial C_{x,v_j|\mathbf{v}_{-j}}(F(x|\mathbf{v}_{-j}),F(v_j|\mathbf{v}_{-j}))}{\partial F(v_j|\mathbf{v}_{-j})}$, where $C_{x,v_j|\mathbf{v}_{-j}}$ is a bivariate copula distribution of variables x and v_j, v_j is an arbitrarily chosen



(a) EAC



(b) FNAC



(c) Vine copula

Figure 5.1: 3-dimensional EAC, FNAC and vine copula structures

component of the vector $\mathbf{v}$, and $\mathbf{v}_{-j}$ denotes the vector $\mathbf{v}$ excluding the component v_j]. For example, a four-dimensional density $f(x_1, x_2, x_3, x_4)$ may be expressed in term of a D-vine as

$$\begin{aligned} f(x_1, x_2, x_3, x_4) &= f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot f_4(x_4) \cdot c_{12}(F(x_1), F(x_2)) \\ &\quad \cdot c_{23}(F(x_2), F(x_3)) \cdot c_{34}(F(x_3), F(x_4)) \\ &\quad \cdot c_{13|2}(F(x_1|x_2), F(x_3|x_2) \cdot c_{24|3}(F(x_2|x_3), F(x_4|x_3)) \\ &\quad \cdot c_{14|23}(F(x_1|x_2, x_3), F(x_4|x_2, x_3)). \end{aligned}$$

Parameters of the D-vine copula model may be estimated by the MPL method. First, parameters of all $d(d-1)/2$ pair-copulas are estimated by a sequential procedure and then the log psuedo-likelihood function is numerically maximized using the parameters obtained by the sequential procedure as starting values. All pair-copulas may be sequentially selected and estimated by the following algorithm [30]:

1. Determine pair-copulas in the first tree from the original data using selection techniques described in Section 4.3.3.

2. Estimate parameters of the selected copulas in the first tree.

3. Compute observations for the second tree using the estimated parameters in the first tree and the appropriate h-functions (the h-function is the conditional distribution function $F(x|\mathbf{v})$).

4. Determine pair-copulas in the second tree from the observations obtained by step 3.

5. Estimate parameters of the selected copulas in the second tree.

6. Compute observations for the third tree using the estimated parameters in the second tree and the appropriate h-functions.

7. Determine pair-copulas in the third tree from the observations obtained by step 6.

8. Estimate parameters of the selected copulas in the third tree.

9. Continue to the $(d-1)$th tree.

Czado, Frigess and Bakken [30] presented an algorithm for sampling from the D-vine as well.

5.4 Results

5.4.1 Discordancy Test

First, regions shown in Figure 4.7 were checked for the occurence of discordant sites. In every region except regions 5a and 5b at least one site was flagged as discordant (see Table 5.1). The problem of determining discordant sites in very small regions has been already discussed in Section 4.8.2. Figure 5.2 compares spectral norms of sample L-covariance, L-coskewness and L-cokurtosis coefficient matrices for each site in the regions. Discordant sites are the black points, while the spectral norms of the regional L-comoment coefficient matrices are marked by +. The sites flagged as discordant often have smaller norms of L-comoment coefficient matrices in comparison with the centre of a group marked by +. Hence, the dataset at sites flagged as discordant was further investigated. Because the data used were checked for gross errors and missing readings by the CHMI and Kyselý [68], there is no reason to doubt the reliability and validity of the data. Additional screening of datasets has shown that some 1-, 3-, and 7-day maximum annual precipitation totals observed at discordant sites are smaller or larger than values measured at other sites. These small and large valid values, reflecting the natural variability of the observed data, apparently cause discordancy. Hence, none of the sites recognized as discordant was excluded from the subsequent analysis (copula selection, homogeneity testing) inasmuch as all values of the discordancy measure larger than a critical value at the 5% significance level had originated from real, observed outliers and not from errors. Moreover, Table 5.6 presents measures of heterogeneity $H_{||\cdot||}$ and p-values for regions without discordant stations (marked by an asterisk), and it is shown that retention of discordant sites in regions does not harm their homogeneity

(except region 1 in the sense of nonparametric testing). The spectral matrix norm was used for evaluating both the discordancy and heterogeneity measures.

5.4.2 Trivariate Model Test Space Construction

Here, the interest was focused in estimating the dependence structure in the stations. Hence, the original datasets were transformed to approximately uniform variables using the empirical distribution functions before further analysis ($X_1 \rightarrow \tilde{U}_1, X_3 \rightarrow \tilde{U}_3, X_7 \rightarrow \tilde{U}_7$).

The trivariate model test space was partly constructed on the results of the bivariate model selection. The Hüsler-Reiss, Galambos, and Gumbel extreme-value copula families appropriately describe the dependence structure between daily and multidaily maximum annual precipitation amounts measured at stations in all regions. According to these results, exchangeable Gumbel, fully nested Gumbel, and mixed C-vine copula models, given by the bivariate Gumbel copula for the two unconditional pair-copulas, were considered to model the dependence between the variables $\tilde{U}_1, \tilde{U}_3,$ and $\tilde{U}_7$.

In the 3-dimensional case, there exist three different fully nested Archimedean and C-vine copula structures obtained by permuting variables in Figures 5.1b and 5.1c. The appropriate ordering of variables of the fully nested Gumbel copula was determined by the fact that the parameters must decrease from the highest to the lowest level in the structure [80], i.e., $\theta_2 \geq \theta_1 \geq 1$ in Figure 5.1b. As shown in Table 5.1, the pair with the strongest dependency is that of X_3 and X_7. The ordering of variables of the mixed C-vine copula model in Tree 1 was specified by two largest sample Kendall's τ values [30]. Hence, according to the values displayed in Table 5.1 the middle node in Figure 5.1c corresponds to the variable X_3, respectively $\tilde{U}_3$ after transforming to the approximately uniform variable.

To establish the C-vine model, the appropriate bivariate copula for the conditional pair-copula $C_{17|3}$ had to be determined. Here, the bivariate model test space for modeling dependence behaviour between marginal conditional distribution functions $F(\tilde{u}_1|\tilde{u}_3)$ and $F(\tilde{u}_7|\tilde{u}_3)$ consisted of the Gumbel, Frank, normal, Student's

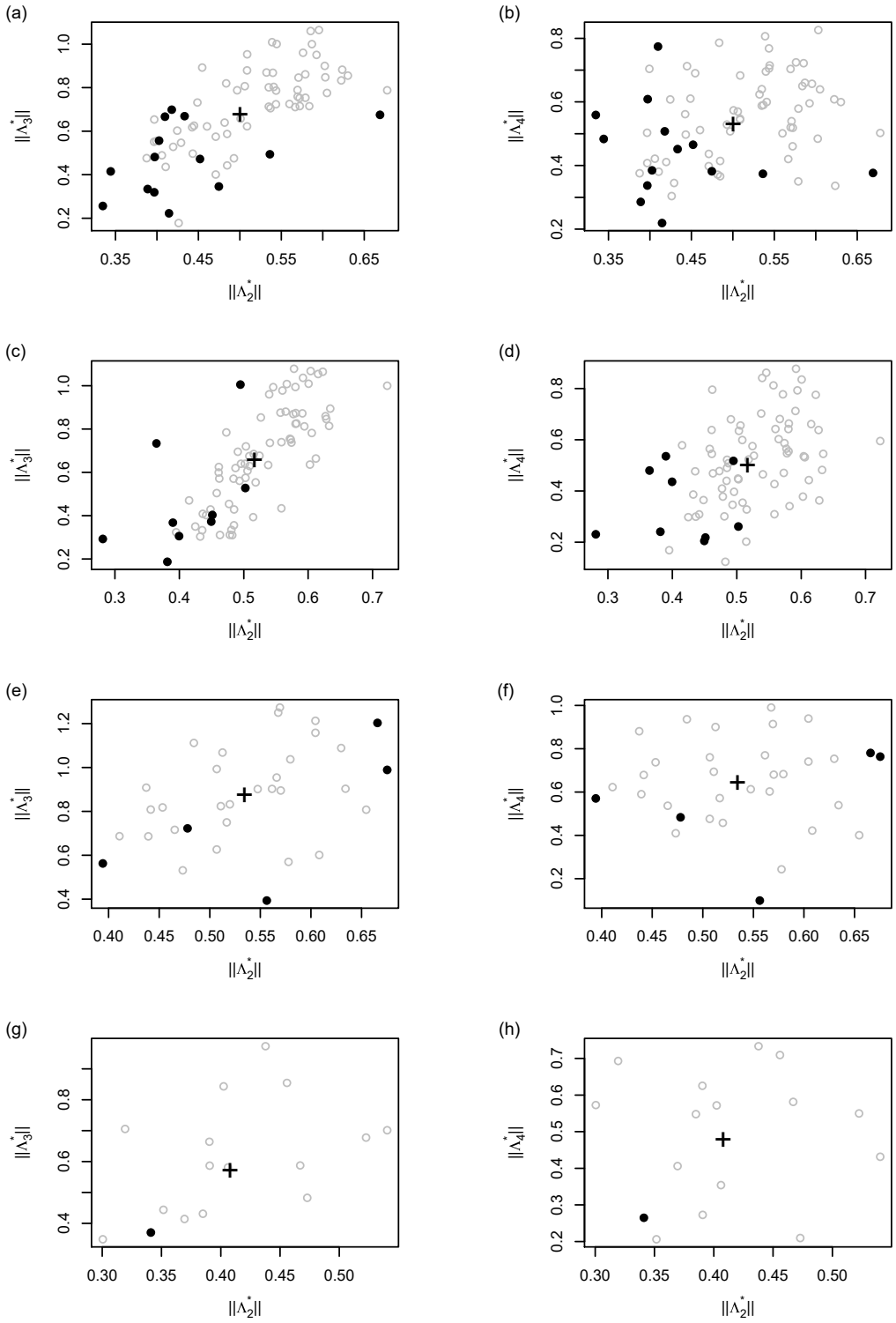


Figure 5.2: Spectral norms of L-comoment coefficient matrices (discordant sites are marked by black points, while the centre of a group is marked by +) for (a–b) region 1, (c–d) region 2, (e–f) region 3, (g–h) region 4

t, and Plackett families. Table 5.3 gives their copula family functions, parameter(s) and Kendall's τ ranges, as well as lower and upper TDCs. The best bivariate model was determined by the goodness-of-fit test based on the Cramér-von Mises statistic and AIC statistic. Based on MPL estimation, the estimated parameters were first obtained. Because the number of degrees of the Student's t copula was greater than 30 at many stations, the Student's t copula family converges to the normal family and, hence, the Student's t copula family was eliminated.

A bootstrap version based on the empirical copula was employed for testing the goodness-of-fit for the remaining copulas of interest. The Cramér-von Mises statistics and approximate p-values were obtained via parametric bootstrap procedure with the number of bootstrap samples fixed at $M = 1\ 000$ [45]. Figure 5.3 reports the results of goodness-of-fit testing using S_n. Each column in Figure 5.3 shows the percentage of rejecting the null hypothesis at the 5% significance level for a given copula family and region. In that phase of bivariate copula selection none of the families was eliminated, and the AIC was employed to choose the best one.

As displayed in Figure 5.4, the normal or Frank copula families achieve the minimum values for more than 38% of stations in each region. Nevertheless, the AIC statistic provides no evidence for one copula family over the other, because AIC differences Δ_i are less than 2 at 99% of stations (differences not presented). Analysis of copula behaviour in extremes by estimating the TDCs did not help to identify the optimal choice for the conditional pair-copula $C_{17|3}$. That is because the copulas of interest by definition have both lower and upper TDCs of zero (see Table 5.3). From the aforementioned results, it was concluded that normal, Frank, and Plackett copula families are essentially indistinguishable, and therefore three mixed C-vine copulas models, given by the bivariate Gumbel copula for the two unconditional pair-copulas and normal, Frank or Plackett copulas for the one conditional pair-copula were included in the trivariate copula model test space. Table 5.4 provides descriptions of the copula models involved in the final trivariate test space.

Table 5.3: Description of copula families employed in bivariate copula modeling

Copula	Properties (by rows: copula function; parameter(s) range; Kendall's τ range; lower and upper TDCs)
Normal	$\Phi_\theta(\Phi^{-1}(u), \Phi^{-1}(v))$ $\theta \in (-1, 1)$ $\tau \in (-1, 1)$ $0;\ 0$
Student's t	$t_{\theta,\nu}(t_\theta^{-1}(u), t_\theta^{-1}(v))$ $\theta \in (-1, 1), \nu > 0$ $\tau \in (-1, 1)$ $2t_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\theta)}{1+\theta}}\right); 2t_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\theta)}{1+\theta}}\right)$
Frank	$-\frac{1}{\theta}\log\left(1 + \frac{[\exp(-\theta u)-1][\exp(-\theta v)-1]}{\exp(-\theta)-1}\right)$ $\theta \in (-\infty, \infty) \setminus \{0\}$ $\tau \in [-1, 1]$ $0;\ 0$
Plackett	$\frac{[1+(\theta-1)(u+v)]-\sqrt{[1+(\theta-1)(u+v)]^2-4\theta(\theta-1)uv}}{2(\theta-1)}$ $\theta \in (0, \infty)$ $\tau \in [-1, 1]$ $0;\ 0$
Gumbel	$\exp(-\{[(-\log u)^\theta + (-\log v)^\theta]\}^{1/\theta})$ $\theta \in [1, \infty)$ $\tau \in [0, 1]$ $0; 2 - 2^{1/\theta}$

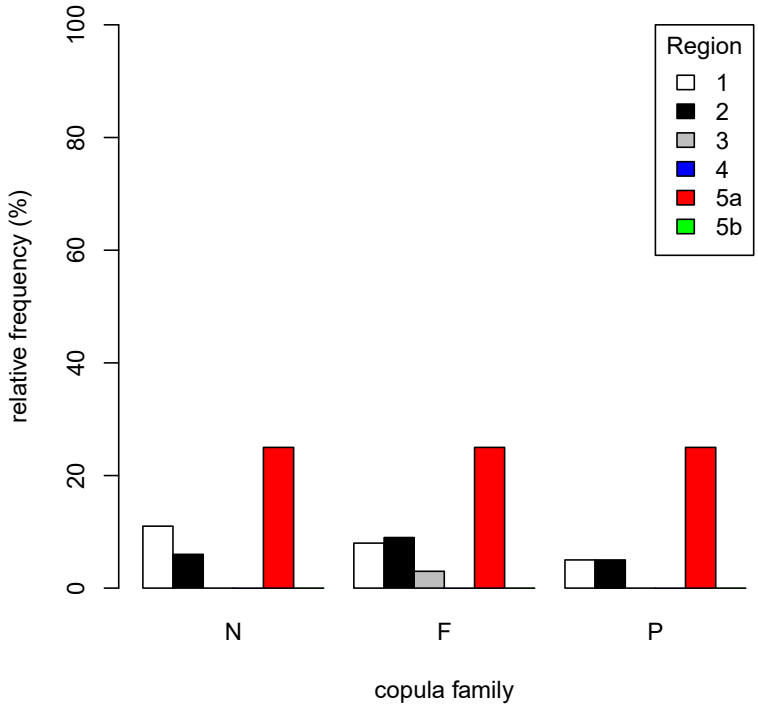


Figure 5.3: Relative frequency of rejecting the null hypothesis that the unknown bivariate copula family belongs to a given copula family at the 5% significance level

5.4.3 Trivariate Model Selection

The first step toward dependence model selection consisted of performing the goodness-of-fit test. For each model displayed in Table 5.4 the Cramér-von Mises statistic S_n was evaluated and the approximate p-values were obtained by parametric bootstrap with the number of bootstrap samples set to $M = 1\,000$. Each column in Figure 5.5 represents the percentage rejection of the null hypothesis at the 5% significance level for a given copula model and region. The results do not allow for reducing the set of feasible copula models, because all models are similarly rejected. However, M3 is at least rejected.

In the second stage, AIC and BIC are evaluated to compare the models. Each column in Figures 5.6 and 5.7 shows the percentage of stations in which a given copula model achieves the minimum value. From Figure 5.6 it follows that the copula model with the highest percentage (more than 33%) in regions 1, 2, 3, 5a,

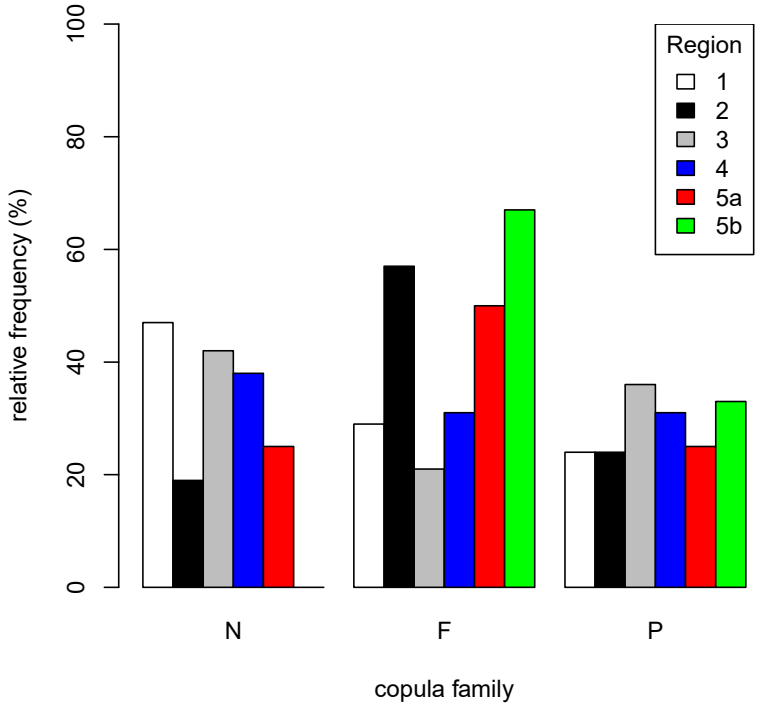


Figure 5.4: Relative frequency of stations in which a given bivariate copula achieves the minimum AIC value

Table 5.4: Description of considered trivariate copula models (designation of copulas and their parameters correspond to Figure 5.1c with variable ordering $U_1 = \tilde{U}_7, U_2 = \tilde{U}_3, U_3 = \tilde{U}_1$)

Model	Model type	Bivariate/trivariate copula(s)	Parameter constraints
M1	EAC	C Gumbel	$\theta \geq 1$
M2	FNAC	C_1, C_2 Gumbel	$\theta_2 \geq \theta_1 \geq 1$
M3	Mixed vine	C_{12}, C_{23} Gumbel; $C_{13\|2}$ normal	$\theta_{12}, \theta_{23} \geq 1; \theta_{13\|2} \in (-1, 1)$
M4	Mixed vine	C_{12}, C_{23} Gumbel; $C_{13\|2}$ Frank	$\theta_{12}, \theta_{23} \geq 1; \theta_{13\|2} \in R - \{0\}$
M5	Mixed vine	C_{12}, C_{23} Gumbel; $C_{13\|2}$ Plackett	$\theta_{12}, \theta_{23} \geq 1; \theta_{13\|2} \in (0, \infty)$

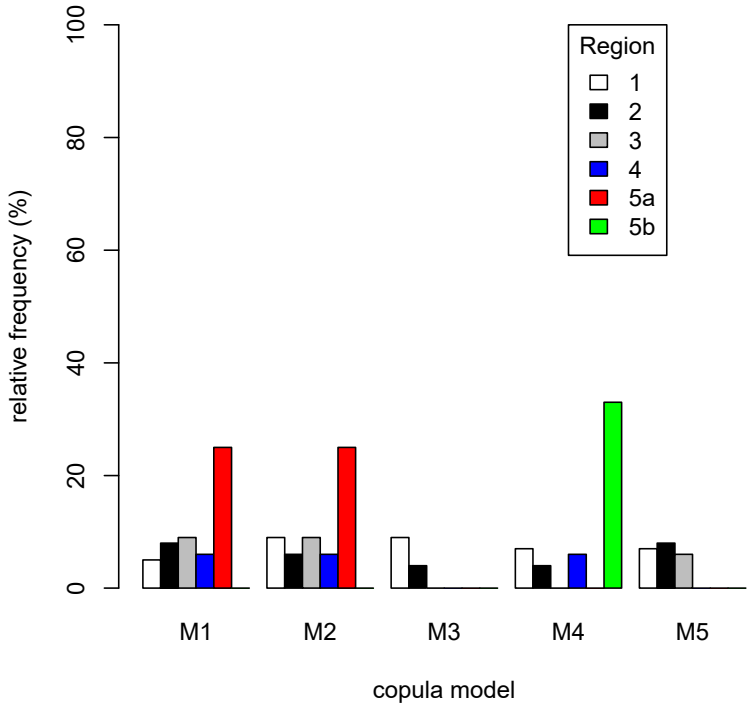


Figure 5.5: Relative frequency of rejecting the null hypothesis that the unknown copula model belongs to a given copula model at the 5% significance level

and 5b is the mixed C-vine normal, and in region 4 it is the Frank copula family. It was not possible, however, to distinguish between the three mixed C-vine models, since the values of the AIC statistic are very similar, and AIC differences Δ_i do not exceed the value of 2 in many cases. By contrast, the results surely do not support the Gumbel EAC and FNAC as appropriately fitting the data due to their providing less flexibility in dependence modeling. This is as has been expected in consideration of their properties (see Section 5.3). Similar conclusions are obtained using BIC (BIC just supports the mixed C-vine copula models given by the Frank and Plackett conditional pair-copulas less, and supports EAC and FNAC more than AIC in the sense of minimum achieved value).

To confirm indistinguishability of models M3, M4, and M5 also statistically, the Vuong test corrected for the number of model parameters analogous to the AIC was also performed. As shown in Table 5.5, which displays the percentage of

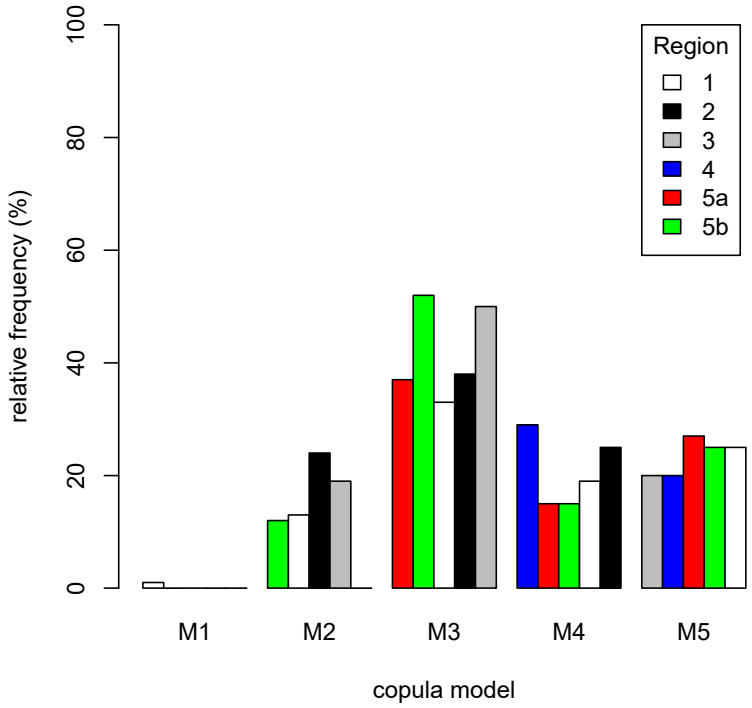


Figure 5.6: Relative frequency of stations in which a given copula model achieves the minimum AIC value

preference for the first, second and none of given copula models and region, none of the non-nested pairs can be distinguished statistically at the 5% level because almost in 100% of stations each pair of models is equally distant from the true one. Hence, the Vuong test also supports models M3, M4, and M5.

5.4.4 L-Moment Homogeneity Testing

To construct the heterogeneity measure $H_{||\cdot||}$, 500 synthetic homogeneous regions were generated in accordance with the three mixed C-vine copula models, designated M3, M4, and M5, and the four-parametric kappa distribution for the margins (the copula model parameters were estimated by the MPL method, while the marginal parameters were estimated by the L-moment method). The results show that regions may be regarded as homogeneous with respect to all considered models, because the values of the heterogeneity measure are less than 2. Moreover, some

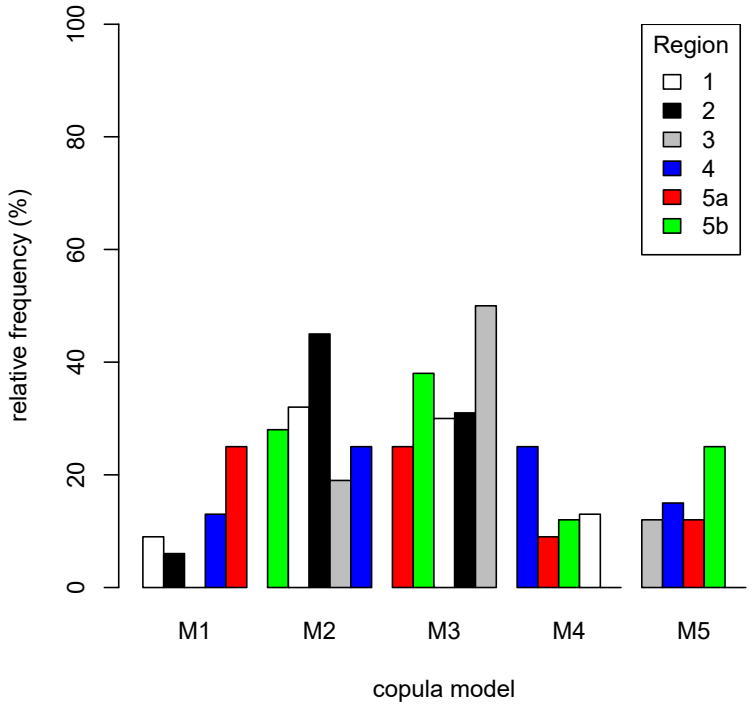


Figure 5.7: Relative frequency of stations in which a given copula model achieves the minimum BIC value

Table 5.5: Vuong test results

Models compared	Variant	1	2	3	4	5a	5b
M3 versus M4	M3 closer	0	0	3	0	0	0
	Equally distant	100	100	97	100	100	100
	M4 closer	0	0	0	0	0	0
M3 versus M5	M3 closer	1	0	3	0	0	0
	Equally distant	99	100	97	100	100	100
	M5 closer	0	0	0	0	0	0
M4 versus M5	M4 closer	0	1	0	0	0	0
	Equally distant	100	99	100	100	100	100
	M5 closer	0	0	0	0	0	0

Table 5.6: L-moment homogeneity test results (* denotes results when the discordant sites were removed from regions)

Region	Parametric test			Nonparametric test
	M3	M4	M5	P-value
1	0.8841	0.9672	1.1754	0.018
	-0.9091*	-0.9041*	-0.8108*	0.504*
1a	-2.7704	-2.9466	-3.0847	0.448
1b	0.0180	-0.0727	0.1684	0.278
1c	-1.1358	-1.1647	-1.0909	0.896
2	0.9322	0.7705	1.0357	0.270
	-0.9437*	-1.0000*	-0.9219*	0.936*
3	-1.1333	-1.2115	-1.2015	0.948
	-1.5291*	-1.6294*	-1.6690*	0.986*
4	0.5303	0.5075	0.6697	0.168
	0.3566*	0.3533*	0.4380*	0.148*
5a	-1.9931	-1.8548	-1.8742	1.000
5b	-0.6821	-0.6361	-0.6438	0.694

values are even negative. Note that the unification of regions 5a and 5b would lead to an infraction of the homogeneity with respect to all considered models. The permutation nonparametric test was also employed to check homogeneity. The p-values obtained are presented in Table 5.6 as well. Except region 1 for which p-value is equal to 0.018, the rest may be regarded as homogeneous at the 5% significance level. Hence, region 1 should be redefined. It was divided into three smaller regions 1a, 1b, and 1c as it was in the study of Kyselý et al. [69] and their homogeneity has been finally achieved (see Table 5.6).

5.4.5 Comparison of Models for Another Climate Zone

To determine whether the mixed C-vine models M3, M4, and M5 may be also suitable for modeling 1-, 3-, and 7-day maximum annual precipitation totals from another place and another climate, five meteorological stations located in Spain in the central Iberian Peninsula were chosen. Daily amounts measured at stations

in Madrid, Navacerrada, Zamora, Salamanca and Daroca from 1950 to 1978 were acquired online from http://www.ecad.eu. First, the interest was focused on finding whether extreme-value copulas (Gumbel, Galambos, Hüsler-Reiss) describe well the dependence structures between 1- and 3-day, and 3- and 7-day amounts, as it is indicated in the study performed for data from the Czech Republic. Therefore standard techniques, including the Cramér-von Mises goodness-of-fit test, AIC and estimating the upper TDC, were employed in bivariate copula selection considering a large copula test space. The copula test space consisted of 18 one-parameter copula families: 13 Archimedean families (Clayton, Frank, Joe, and C#2, C#12–C#15, C#17–C#21 described in [85]), 3 extreme-value families (Gumbel, Galambos, Hüsler-Reiss), 1 elliptical family (normal), and 1 miscellaneous family (Plackett). P-values (the first column), AIC differences Δ_i (the second column) and estimated upper TDCs $\hat{\lambda}_U$ obtained are presented in Tables 5.7 and 5.8. The results clearly show that the extreme-value copulas are suitable for modeling the dependence structure between 1- and 3-day in all stations, except Daroca. In the case of modeling the dependence between 3- and 7-day amounts, the extreme-value families fit the data well only in Madrid and Daroca. However, these families have less support in Zamora and Salamanca with AIC differences from 3 to 4.5, and in Navacerrada these families have no support. Results presented in Table 5.9 indicate that models M3, M4, and M5 seem to be suitable to model the complex dependence structure between 1-, 3-, and 7-day amounts in three stations of five (namely in Madrid, Zamora, and Salamanca). However, generally, it is not possible to claim that resulting models M3, M4 and M5 are the most suitable to model the trivariate distribution of 1-, 3-, and 7-day amounts in all precipitation regimes. The results of RFA of extreme precipitation events is only based on meteorological data from the Czech Republic, and, hence, the most suitable copula model may change from one area to another due to different precipitation regimes. It is important to analyse data and find the best copula model that is sufficiently common to all stations in the region just studied. However, the approach introduced presents a general

Table 5.7: Bivariate copula selection results for 1- and 3-day totals at Spanish stations (p-value in the first column, AIC differences in the second column)

Copula	Madrid		Navacerrada		Zamora		Salamanca		Daroca	
	$\widehat{\lambda} = 0.58$		$\widehat{\lambda} = 0.74$		$\widehat{\lambda} = 0.63$		$\widehat{\lambda} = 0.64$		$\widehat{\lambda} = 0.55$	
Normal	0.63	3.9	0.29	3.8	0.61	0.0	0.48	1.9	0.29	1.1
Clayton	0.07	8.7	0.01	14.4	0.07	5.2	0.04	8.8	0.04	3.2
Gumbel	0.97	0.9	0.64	0.3	0.33	2.9	0.46	1.3	0.04	5.2
Frank	0.36	6.3	0.51	3.9	0.22	5.1	0.19	5.3	0.53	0.4
Joe	0.91	0.0	0.25	0.7	0.06	6.8	0.12	3.0	0.98	9.6
Galambos	0.98	1.0	0.60	0.6	0.37	2.2	0.52	0.8	0.04	4.9
Plackett	0.56	5.6	0.53	4.1	0.11	6.5	0.12	6.1	0.21	2.5
Hüsler-Reiss	0.96	1.0	0.46	1.2	0.53	0.9	0.67	0.0	0.03	4.5
C#2	0.00	20.0	0.01	13.6	0.01	26.3	0.00	20.9	0.00	35.1
C#12	0.00	6.1	0.01	6.0	0.00	1.3	0.00	3.5	0.00	2.3
C#13	0.00	7.0	0.03	8.2	0.02	3.8	0.00	5.9	0.00	0.6
C#14	0.00	3.9	0.00	2.2	0.00	1.1	0.00	1.7	0.00	2.4
C#15	0.00	0.3	0.07	0.0	0.11	5.9	0.02	2.4	0.00	8.4
C#17	0.00	7.0	0.04	5.5	0.02	5.0	0.00	5.8	0.00	0.0
C#18	0.00	132.0	0.01	91.9	0.01	294.8	0.00	228.9	0.00	423.2
C#19	0.00	8.8	0.01	17.5	0.03	6.0	0.01	10.9	0.00	4.8
C#20	0.00	9.6	0.01	19.3	0.04	6.7	0.01	11.8	0.00	6.0
C#21	0.00	1.6	0.03	0.8	0.10	12.0	0.01	4.7	0.00	14.2

procedure how to deal with homogeneity testing in trivariate RFA and it may be applied in studies using data from other parts of the world.

5.5 Conclusion

This chapter deals with homogeneity testing in three dimensions and investigates its practical aspects. The testing procedures were applied for extreme precipitation data in the Czech Republic for the past three to five decades. For this purpose, quite a lot of attention was given to selecting the joint cumulative distribution function of the three-dimensional random vector of the interest. Several conclusions can be drawn from this study. Standard approaches involving goodness-of-fit tests

Table 5.8: Bivariate copula selection results for 3- and 7-day totals at Spanish stations (p-value in the first column, AIC differences in the second column)

Copula	Madrid		Navacerrada		Zamora		Salamanca		Daroca	
	$\widehat{\lambda} = 0.61$		$\widehat{\lambda} = 0.76$		$\widehat{\lambda} = 0.82$		$\widehat{\lambda} = 0.77$		$\widehat{\lambda} = 0.62$	
Normal	0.88	1.0	0.26	4.3	0.34	1.7	0.27	3.4	0.07	1.4
Clayton	0.14	5.5	0.40	0.0	0.01	18.1	0.07	5.2	0.00	8.0
Gumbel	0.88	0.6	0.03	10.7	0.25	3.1	0.26	4.5	0.18	0.0
Frank	0.38	5.7	0.61	5.5	0.58	0.0	0.03	10.8	0.26	0.2
Joe	0.35	1.9	0.00	20.5	0.04	10.6	0.04	10.5	0.07	1.3
Galambos	0.90	0.4	0.03	10.8	0.27	3.0	0.23	4.4	0.16	0.1
Plackett	0.38	6.2	0.46	6.0	0.27	2.7	0.29	6.1	0.21	0.7
Hüsler-Reiss	0.90	0.0	0.02	11.3	0.24	3.2	0.27	4.0	0.14	0.0
C#2	0.02	21.0	0.00	36.6	0.06	20.3	0.01	24.2	0.06	20.7
C#12	0.70	1.9	0.26	1.9	0.17	4.9	0.33	0.0	0.04	4.2
C#13	0.31	4.9	0.69	0.3	0.14	6.0	0.08	6.1	0.03	3.5
C#14	0.93	0.6	0.06	6.5	0.29	2.6	0.31	2.0	0.11	1.8
C#15	0.26	2.7	0.07	14.9	0.25	4.3	0.16	7.1	0.08	0.3
C#17	0.37	5.9	0.43	4.2	0.04	6.8	0.12	10.2	0.13	1.4
C#18	0.02	115.6	0.00	209.7	0.00	235.8	0.00	217.1	0.02	196.4
C#19	0.13	4.5	0.04	5.2	0.00	33.8	0.03	7.3	0.00	8.6
C#20	0.14	5.4	0.06	5.0	0.00	33.3	0.04	6.8	0.00	10.5
C#21	0.62	7.3	0.02	23.1	0.04	10.9	0.06	12.2	0.29	2.1

Table 5.9: Copula model selection results for 1-, 3-, and 7-day totals at Spanish stations (p-value in the first column, AIC differences in the second column)

Model	Madrid		Zamora		Salamanca	
M1	0.62	0.4	0.07	10.7	0.62	6.6
M2	0.87	1.0	0.72	0.0	0.60	5.6
M3	0.26	0.0	0.66	1.7	0.03	2.5
M4	0.06	0.3	0.27	2.6	0.24	1.8
M5	0.58	0.4	0.53	2.6	0.06	0.0

based on the Cramér-von Mises and Vuong statistics and AIC/BIC were used for selection of the appropriate trivariate copula model. Considering EACs, FNACs and C-vine copulas in the test space, selection techniques showed that three considered mixed C-vine copulas appropriately describe the dependence structure between 1-, 3-, and 7-day maximum annual precipitation totals in the Czech Republic. The decomposed models consisting of bivariate copulas are clearly preferable to EACs and FNACs, as they are more flexible in data fitting. Although only four bivariate copula families (normal, Student's t, Frank, and Plackett) were considered in the test space for the conditional pair-copula selection in the pair-copula construction, more simultaneously positively and negatively ordered families may be involved in the testing procedure, see for example the list of Archimedean copula families in [85]. The three aforementioned indistinguishable mixed C-vine copulas were employed to obtain the heterogeneity measures for regions formed in the Czech Republic. Regions may be regarded as homogeneous with respect to all considered mixed C-vine models. Because those regions with small numbers of stations are not suitable for RFA, it may be generally possible to redefine small regions and unify them into a region with more stations while maintaining the homogeneity. This was not possible for regions 5a and 5b, however, due to their absolutely different precipitation patterns. Some negative values of the heterogeneity measure appeared as a result of homogeneity testing. These values probably occur due to the presence of cross-correlation between sites. An appropriate modification of the Chebana and Ouarda [22] multivariate L-moment homogeneity test in order to overcome spatial dependence is a subject of interest in the next chapter. A small analysis of daily and multidaily maximum annual precipitation totals measured at selected meteorological stations in Spain has showed that the best copula model may change from one place to another. Hence, the resulting models M3, M4, and M5 are mainly valid for regions formed in the Czech Republic. However, the study presented gives a methodology on how to perform multivariate L-moment homogeneity testing including identifying the best copula model.

6

Homogeneity Testing for Spatially Correlated Data in Multivariate RFA

Contents

6.1 Introduction

Due to the nature of the simulated data such that cross-correlation between sites was omitted, the test powers presented by Chebana and Ouarda [22] and by Masselot, Chebana and Ouarda [78] would be accurate only if the observations were independent between sites. Unfortunately, in practice this assumption is not always valid (see, e.g., [2, 17, 19, 79, 107]). Several studies have already dealt with the problem of dependence between sites for the univariate Hosking and Wallis [58] L-moment homogeneity test [18, 54, 73]. Hosking and Wallis [54] themselves pointed out the effect of cross-correlation: If negative values and values less than

-2 of the heterogeneity measure H are observed, then there may be positive correlation between the data at different sites in the proposed region. Castellarin, Burn and Brath [18] achieved important results on the impact of cross-correlation when they investigated limitations of the test in heterogeneity detection for cross-correlated regions through a series of Monte Carlo experiments assuming that the joint distribution for all sites in the region to be a multivariate normal. They found out that cross-correlation may result in lower values of the heterogeneity measure than would be for uncross-correlated region, that may lead to miscategorization of region. Hence, the test detects heterogeneity less often and the power of the L-moment homogeneity test may be reduced when cross-correlation is present. Consequently, they proposed an empirical corrector of the test that provides an approximate categorization of the real degree of heterogeneity of a studied region. Recently, Lilienthal, Fried and Schumann [73] continued the results of Castellarin, Burn and Brath [18] about the negative impact of cross-correlation and improved the Hosking and Wallis [58] procedure by using a flexible multivariate copula model when generating cross-correlated data which features the same dependence structure as does the observed data. Hence, due to the fact that the presence of cross-correlation reduces the power of the univariate Hosking and Wallis [58] test, the impact of cross-correlation on the multivariate tests of Chebana and Ouarda [22] and Masselot, Chebana and Ouarda [78] should be also investigated.

The idea was motivated by results of the case studies conducted in the Czech Republic, in which relatively large negative values of heterogeneity measure have been received as results of the L-moment homogeneity test carried out as a first step in bivariate and trivariate RFA of extreme precipitation events (see Tables 4.5 and 5.6). It is assumed that as occurs in the univariate case and it is stated in the above paragraph, cross-correlation between observed data values leads to a reduction of the tests' powers also in the multivariate framework.

This chapter shows the negative impact of cross-correlation on the bivariate parametric and nonparametric L-moment homogeneity tests, and proposes a generalization of the Chebana and Ouarda [22] procedure to overcome the problem

of misspecification of region due to the presence of cross-correlation. The chapter is organized as follows: In Section 6.2 a generalization of the original bivariate parametric testing procedure of Chebana and Ouarda [22] using D-vine copulas is proposed. A series of Monte Carlo simulations is performed to show that the cross-correlation diminishes the power of the multivariate L-moment homogeneity tests and how the use of the modification improves the heterogeneity detection (Section 6.3). Results of the simulation study are summarized and discussed in Section 6.4. In Section 6.5 the proposed generalized test is applied to a real meteorological dataset to demonstrate its usefulness in practice. The chapter closes with a summary section. This chapter relies on the article [115].

6.2 Generalization of the Parametric L-Moment Homogeneity Test

In the bivariate parametric homogeneity test of Chebana and Ouarda [22], it is necessary to determine the regional bivariate copula that is common to all sites. However, in the presented approach, which deals with the problem of cross-correlations between sites, it is essential to control also the spatial dependencies. Hence, it is necessary to generate homogeneous regions of data with predefined regional bivariate copula at each site and also the approximate spatial dependencies. Based on the graphical representations of D- and C-vines copulas (see Figure 6.1 for examples of 6-dimensional C- and D-vine trees that could model dependencies between two variables measured at three sites), it is obvious that only D-vine copulas allow this type of control.

This section presents the solution of the Chebana and Ouarda [22] L-moment homogeneity test to overcome the problem of presence of cross-correlated data. The modification of the Chebana and Ouarda [22] test consists in the replacement of homogeneous uncross-correlated regions by homogeneous cross-correlated regions in the Monte Carlo simulation step. As discussed in the previous paragraph, data of synthetic cross-correlated homogeneous regions may be generated using D-vine copulas.

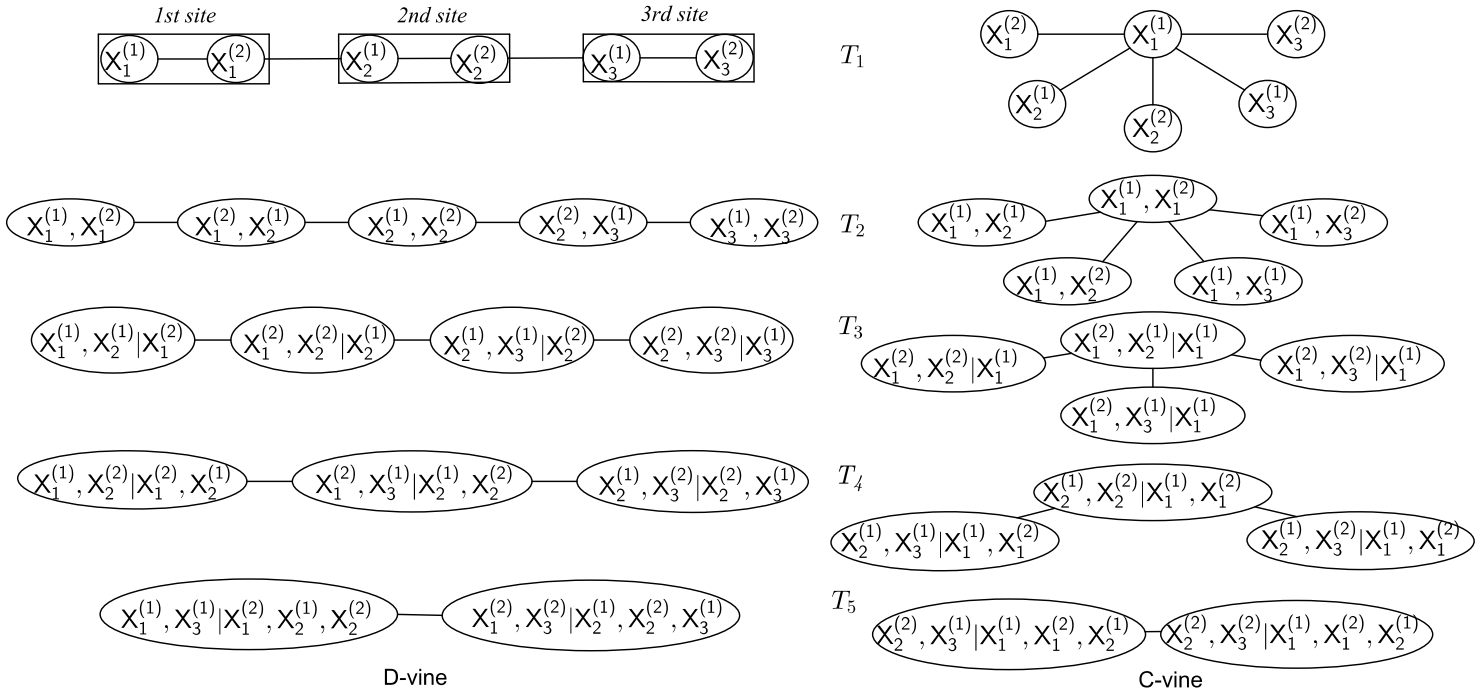


Figure 6.1: Examples of 6-dimensional D- and C-vine trees

The use of D-vine copulas requires data series of the same length, however, in the reality the number of measurements may differ from site to site due to different beginning of records. If the number of missing measurements at some sites is small, they may be omitted also at the remaining sites and the same number of measurements available at each site may be applied to perform the test. In the case of a very small number of the missing values, these values may be infilled. Bárdossy and Pegram [7] presented several methods used for infilling the missing data, such as nearest neighbor, simple or multiple linear regression, kriging, or copula based estimation, and their comparison. Note that infilled values are only estimates and the use of any infilling method brings a source of uncertainity into further analysis (copulas model selection and their parameters estimation). The omission of a large number of measurements at each site may lead to a very small number of common measurements, which may lead to inacurrate analysis. Hence, if the number of the missing values at some site is very large compared to the number of measurements at other sites, this site should be eliminated from analysis.

First, the variables in the tree T_1 must be ordered before pair-copulas selection and estimation (the following trees are then completely determined). For a $2N$-dimensional D-vine copula model in our case of N sites, there exist $2^{N-1} \cdot N!$ choices for how to order variables in the first tree. Unfortunately, it is technically impossible to construct all D-vine models in the sense of the variables ordering and compare them on the basis of resulting log pseudo-likelihood functions. However, in the first tree the variables should be order such that the copulas fitted are those corresponding to $2N - 1$ largest empirical Kendall's τ computed for each pair of variables. N pairs of variables are clearly given, they are those which are measured at the sites. These N pairs must be connected together to create the first tree. This is done such that the sum of empirical Kendall's τ in the first tree is maximized: The next feasible link is determined by the largest Kendall's τ, which has not been used yet.

All pair-copulas are then selected and estimated by the sequential procedure (see Section 5.3.3), while the pair-copulas at each site in the first tree must be the same (the regional bivariate copula parameters are computed as weighted

mean of the at-site estimates). Because the careful pair-copula selection might be time consuming in larger dimensions, it is possible to truncate the D-vine copula model. Brechmann, Czado and Aas [15] treated the problem of determining whether the copula may be truncated at certain level (that means the pair-copulas in all following trees are replaced by independence copulas) by model selection methods, such as the AIC, BIC, and Vuong test.

The synthetic homogeneous cross-correlated regions are simulated by the appropriate D-vine copula and the four-parametric kappa distribution. The heterogeneity measure $H_{||\cdot||}$ computing the degree of heterogeneity in a region and the heterogeneity criteria are the same as in the original test of Chebana and Ouarda [22]. Here, the heterogeneity meausure obtained by the modified parametric test is denoted by an asterisk to differ it from the original one. The whole modified procedure is summarized below.

Suppose that the region has N sites, with site i having record length n_i and the sample L-covariance coefficient matrix $\mathbf{\Lambda}_2^{*(i)}$.

1. Calculate $V_{||\cdot||}$ given by (4.13) on the observed data as in the first step of the original parametric test.

2. Transform data to approximately uniform variables using the empirical distribution functions.

3. Compute empirical Kendall's τ for all pairs and order variables in the first tree.

   (a) Choose the largest Kendall's τ and check that the corresponding variables can be linked. If it is not possible, choose the next largest Kendall's τ and check the possible link.

   (b) Iterate until the first tree is created.

4. Identify an appropriate D-vine copula model for simulation uniform data of synthetic homogeneous cross-correlated regions.

(a) Select a regional bivariate copula family, that is common for all sites in the region by a combination of techniques described in Section 4.3.3, and estimate the regional bivariate copula parameter as $\widehat{\boldsymbol{\theta}}^R = \sum_{i=1}^{N} n_i \widehat{\boldsymbol{\theta}}_i / \sum_{i=1}^{N} n_i$, where $\widehat{\boldsymbol{\theta}}_i, i = 1, 2, \ldots, N$, are the at-site MPL estimates.

(b) Select and estimate the remaining unconditional pair-copulas in the first tree of the D-vine model.

(c) Select and estimate all the conditional pair-copulas by the sequential procedure, while the pair-copulas in the first tree are selected and estimated in step 4(a) and 4(b).

5. Generate a large number N_{sim} of homogeneous cross-correlated regions (500 regions is again enough) with N sites, each having the same record length as its real-world counterpart. To get a sample with uniform margins use D-vine copula model given by step 4, and to get the desired sample use the quantile function of a four-parametric kappa distribution. The regional parameters of the kappa distribution are estimated using the L-moment method proposed by Hosking [55] by fitting this distribution to the regional L-moment ratios $(1, t_2^R, t_3^R, t_4^R)$, where t_k^R is a weighted mean of the at-site L-moment ratios for $k = 2, 3, 4$.

6. Compute the statistic $V_{||\cdot||}^{(j)}$ defined by equation (4.13) on each of the simulated cross-correlated homogeneous regions, $j = 1, 2, \ldots, N_{sim}$, and its mean μ and standard deviance σ over all replications.

7. Compute the heterogeneity measure

$$H_{||\cdot||}^{*} = \frac{V_{||\cdot||} - \mu}{\sigma}. \tag{6.1}$$

8. Categorize the region on the basis of the heterogeneity measure $H_{||\cdot||}^{*}$ given by formula (6.1) as homogeneous if $H_{||\cdot||}^{*} < 1$, acceptably homogeneous if $1 \leq H_{||\cdot||}^{*} < 2$, and definitely heterogeneous if $H_{||\cdot||}^{*} \geq 2$.

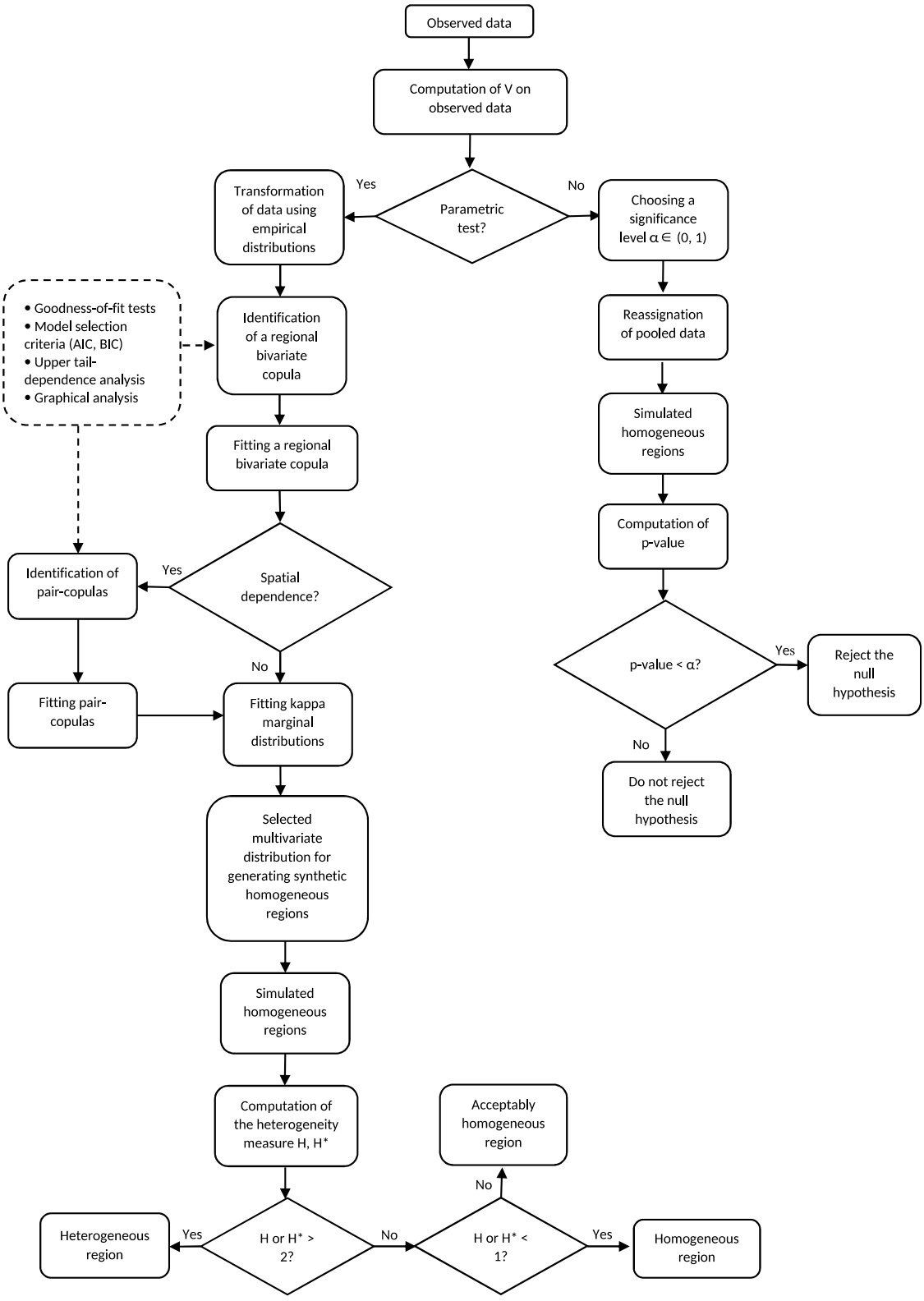


Figure 6.2: Diagram explaining the major differences of the homogeneity tests

6.3 Simulation Study

Monte Carlo simulation experiments were employed to assess the effect of cross-correlation on the L-moment homogeneity tests, and to evaluate the performance of the modification proposed in the previous section.

The simulation section was based on the simulation study carried out by Chebana and Ouarda [22], in which the focus was on flood events characterized by volume V and peak Q. The uncross-correlated homogeneous regions with N sites were generated by the joint distribution of a random vector $(V, Q)^T$ given by the Gumbel copula $C_{\theta_i}^G$ to model the dependence between V and Q, and the Gumbel distribution with the cumulative distribution function

$$F_i(x) = \exp\left\{-\exp\left(-\frac{x-\alpha_i}{\beta_i}\right)\right\}, x \in \mathbb{R}, \alpha_i \in \mathbb{R}, \beta_i > 0, i = 1, 2, \ldots, N,$$

for margins. The selected parameters of the joint distribution at each site were those presented in [121]:

$$\alpha_i^V = 1240, \;\; \beta_i^V = 300, \;\; \alpha_i^Q = 52, \;\; \beta_i^Q = 16, \;\; \theta_i = 1.41, i = 1, 2, \ldots, N. \tag{6.2}$$

Chebana and Ouarda [22] selected several types of representative regions:

Homogeneous Parameters given by (6.2) are the same for all sites in the region.

50% completely heterogeneous All parameters given by (6.2) (except the parameters of location insomuch as $V_{||\cdot||}$ is location invariant) increase linearly from the 1st to the Nth station in a 50% range centered around the homogeneous region parameters, which means $\beta_i^V \in [225, 375], \beta_i^Q \in [12, 20], \theta_i \in [1.0575, 1.7625], i = 1, 2, \ldots, N$.

50% heterogeneous on the marginal parameters The dependence parameter is fixed and equal to the homogeneous region parameter and the marginal parameters increase linearly as in the completely heterogeneous region, which means $\beta_i^V \in [225, 375], \beta_i^Q \in [12, 20], \theta_i = 1.41, i = 1, 2, \ldots, N$.

50% heterogeneous on the dependence parameter The marginal parameters are fixed and equal to the homogeneous region parameters and the dependence parameter increases linearly as in the completely heterogeneous region, which means $\beta_i^V = 300, \beta_i^Q = 16, \theta_i \in [1.0575, 1.7625], i = 1, 2, \ldots, N$.

30% completely bimodal The stations in the region are divided into two disjoint groups: The parameters of the first group are equal to the 85% value of the corresponding homogeneous region parameters, which means $\beta_i^V = 255, \beta_i^Q = 13.6, \theta_i = 1.1985$ for $i = 1, 2, \ldots, \left\lfloor \frac{N}{2} \right\rfloor$, and the parameters of the second group are equal to the 115% value of the corresponding homogeneous region parameters, i.e. $\beta_i^V = 345, \beta_i^Q = 18.4, \theta_i = 1.6215$ for $i = \left\lceil \frac{N}{2} \right\rceil, \ldots, N$, where $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ denote the floor and ceiling functions, respectively.

30% bimodal on the marginal parameters The dependence parameter is fixed and equal to the homogeneous region parameter and the marginal parameters are the same as in the completely bimodal region, which means $\beta_i^V = 255, \beta_i^Q = 13.6$ for $i = 1, 2, \ldots, \left\lfloor \frac{N}{2} \right\rfloor$, $\beta_i^V = 345, \beta_i^Q = 18.4$ for $i = \left\lceil \frac{N}{2} \right\rceil, \ldots, N$, and $\theta_i = 1.41$ for $i = 1, 2, \ldots, N$.

30% bimodal on the dependence parameter The marginal parameters are fixed and equal to the homogeneous region parameters and the dependence parameter is the same as in the completely bimodal region, which means $\beta_i^V = 300, \beta_i^Q = 16$ for $i = 1, 2, \ldots, N$, $\theta_i = 1.1985$ for $i = 1, 2, \ldots, \left\lfloor \frac{N}{2} \right\rfloor$, and $\theta_i = 1.6215$ for $i = \left\lceil \frac{N}{2} \right\rceil, \ldots, N$.

In this simulation study these kinds of regions were adopted with the difference that they were spatially correlated. Although other multivariate copulas, such as elliptical copulas, could be considered for simulation of cross-correlated regions, here representative cross-correlated homogeneous and heterogeneous regions were generated using D-vine copulas. This was because we did not have to fit data to the D-vine copula model, that is required for the simulation step in the modified test, as it was known. Hence, the errors were reduced. Each synthetic region

consisted of $2N$ spatially correlated series of the same length, the dependence between variables V and Q at each site was modeled by the Gumbel copula family, while the rest of pair-copulas in the D-vine model were modeled by the normal copula family C_{ρ}^{No} representing the spatial dependence. Description of these two bivariate copulas (their copula functions, parameter and Kendall's τ ranges, and the lower and upper TDCs) used as pair-copulas of D-vine model are displayed in Table 5.3. For simplicity the parameter ρ of the normal copula families C_{ρ}^{No} was considered to be constant across all trees. Five correlation levels in terms of Kendall's τ were selected for spatial dependence: $\tau = 0$ (uncross-correlated case), $0.13, 0.26, 0.41$, and 0.59, that corresponds to the normal copulas's parameter $\rho = 0, 0.2, 0.4, 0.6$, and 0.8. The structure of the $2N$-dimensional D-vine copula model is shown graphically in Figure 6.3.

For the simulation purposes, 100 regions were generated with N sites ($N = 10, 15, 20$, and 30), fixed record length $n_i = 30$ for each site, given kind of heterogeneity (see the list above) and spatial dependence ρ also given above.

The value of heterogeneity measures $H_{||\cdot||}$, $H^*_{||\cdot||}$ and p-values were computed for each replication. Means of $H_{||\cdot||}$ and $H^*_{||\cdot||}$ values and rejection rates were computed as well (the rejection rate is the ratio of the number of samples in which the value of $H_{||\cdot||}$ or $H^*_{||\cdot||}$ satisfies the condition $H_{||\cdot||} \geq 2$ or $H^*_{||\cdot||} \geq 2$ [the parametric and modified parametric tests], or the number of samples where the null hypothesis is rejected at the significance level $\alpha = 5\%$ [the nonparametric test] to the total number of generated regions, which is equal to 100). The computed rejection rates enabled to estimate the empirical first type error and power of the tests: The empirical first type error of a test is the rejection rate when the region is homogeneous, while the power of a test is the rejection rate when the region is heterogeneous.

6.4 Simulation Results

Figures 6.4 to 6.8 illustrate results of the simulation experiments. Figure 6.4 shows the relationship between average $H_{||\cdot||}$ values obtained by the original parametric homogeneity test for uncross-correlated and cross-correlated regions with the same

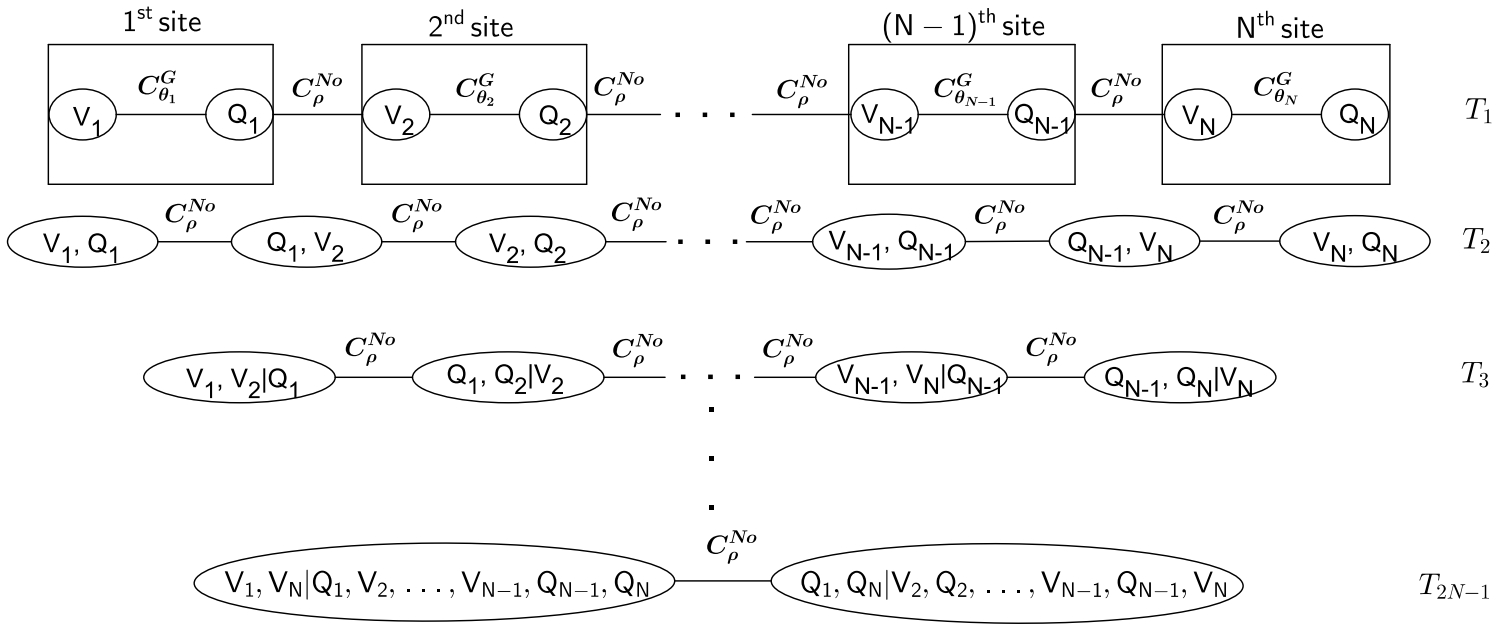


Figure 6.3: $2N$-dimensional D-vine structure employed in the simulation study for generating cross-correlated homogeneous regions

degree of heterogeneity. Figures 6.5 to 6.7 report empirical first type errors and powers for all tests. Figure 6.8 compares average $H_{||\cdot||}$ and $H^*_{||\cdot||}$ values for cross-correlated regions obtained by the original and modified parametric homogeneity tests. It is important to note that in simulation experiments copulas and their parameter values were known and there was no goodness-of-fit testing and parameter estimation. Hence, sources of errors were reduced.

Several issues for consideration were raised by the performed simulation experiments. First, since the simulated homogeneous regions in the parametric test of Chebana and Ouarda [22] are both serially and spatially uncross-correlated by definition, positive cross-correlation causes that the value of the heterogeneity measure $H_{||\cdot||}$ to be lower than the value of $H_{||\cdot||}$ for the uncorrelated region and it is even negative. This diminution may lead to the following miscategorizations of the region: 1) an acceptably homogeneous cross-correlated region being categorized as homogeneous, 2) a heterogeneous cross-correlated region being categorized as acceptably homogeneous, and 3) a heterogeneous cross-correlated region being categorized as homogeneous. Errors of the second and third types are definitely the most serious. Figure 6.4 presents a comparison of average $H_{||\cdot||}$ values obtained by the original parametric homogeneity test for uncross-correlated and cross-correlated regions with the same degree of heterogeneity. Most often, the first error occurs such that an acceptably homogeneous region is categorized as homogeneous. Errors of the second and third types occur much less frequently. This figure indicates that the number of heterogeneous regions which are identified to be homogeneous by the procedure of Chebana and Ouarda [22] rises with increasing cross-correlation, and thus cross-correlation may significantly affect the result of the hypothesis testing procedure in the sense of miscategorizating the region. These results agree with those obtained by Castellarin, Burn and Brath [18], who tested homogeneity in the univariate case, while using the multivariate normal distribution to generate cross-correlated synthetic regions. Moreover, they observed a significant linearity between the heterogeneity measure values for cross-correlated and uncross-correlated regions. Note that although certain linear dependence may be also observed in

Figure 6.4, the approach of Castellarin, Burn and Brath [18] can not be used for bivariate homogeneity testing. This is becuse their method is based on assumption that a region consists of a number of spatially normally distributed series with constant population mean and variance, which is not definitely this case.

The focus now is directed to the ability of the modified parametric procedure to detect heterogeneity. Figure 6.5 illustrates comparison of empirical first type errors of tests for regions with various number of sites and spatial dependence. It is observed that the empirical first type error is up to 6% for the original parametric test, up to 9% for the nonparametric test, and finally up to 7% for the modified parametric test.

Figure 6.6 illustrates empirical powers tests depending on the number of sites and spatial dependence for completely, marginally, and dependence heterogeneous regions. Completely heterogeneous regions are definitely well detected, with the test power ranging between 70% and 97%. The ability of the modification to detect heterogeneity when the regions are marginally or dependence heterogeneous is very similar, but it differs from the completely heterogeneous case. Values of the tests powers are considerably lower than the powers for completely heterogeneous regions in the cases of marginally heterogeneous (ranging between 25% and 65%), and dependence heterogeneous (between 17% and 70%) regions. Hence, the test gives false results more often for regions that are marginally or dependence heterogeneous. Despite the relatively low power of the modification, especially for the marginally and dependence heterogeneous regions, the modified procedure performs clearly better than both the original parametric and nonparametric tests when cross-correlation is present.

The powers of the original and modified parametric tests are similar in the independence case ($\tau = 0$), but they differ significantly when there is nonzero cross-correlation in the region. The power of the modified procedure is always higher. This indicates that the proposed modification is really a suitable generalization of the original test. When comparing the modified test to the nonparametric one, the nonparametric test is more powerful for the independence case which is in practice less likely, particularly when the sites are not far from each other. The

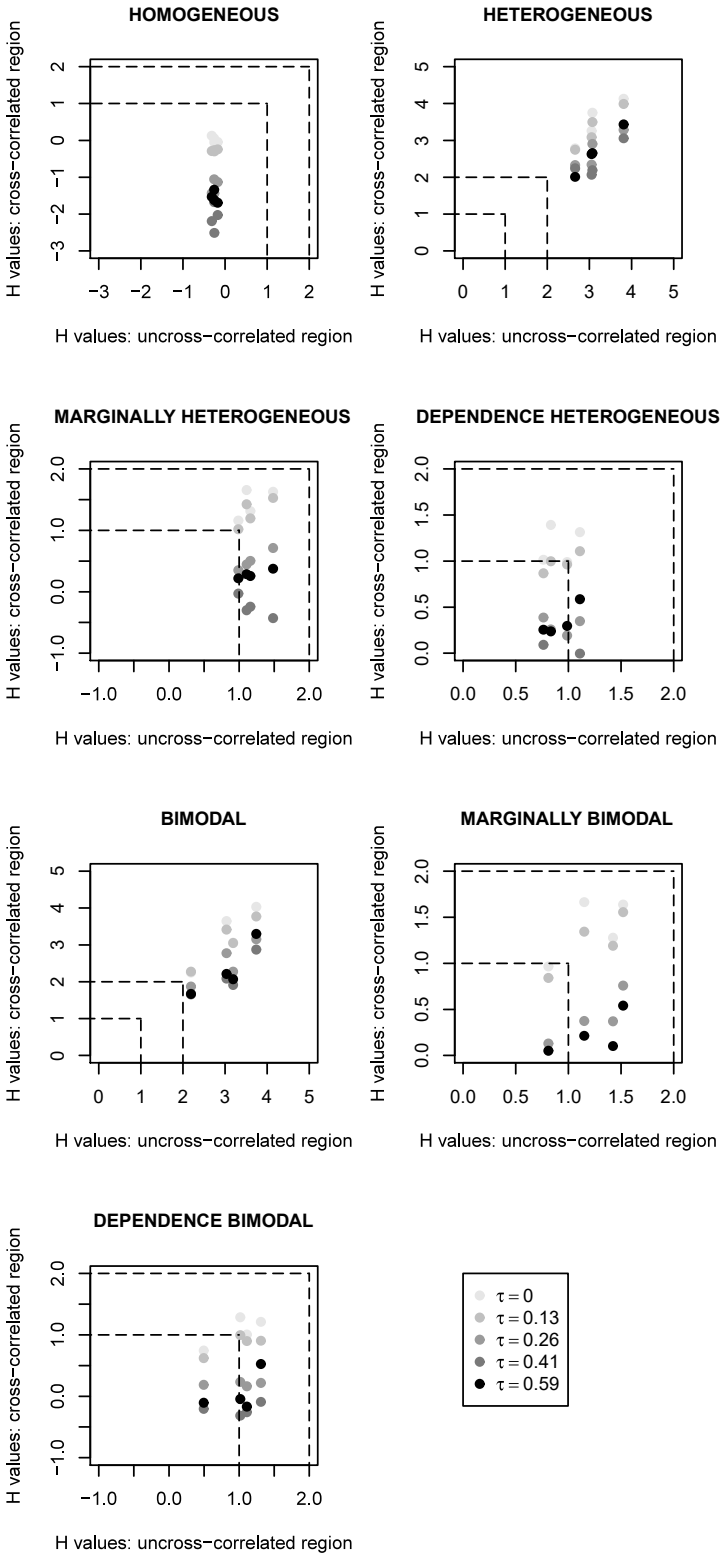


Figure 6.4: Average $H_{||\cdot||}$ values for uncross-correlated vs. cross-correlated regions obtained by the original parametric test

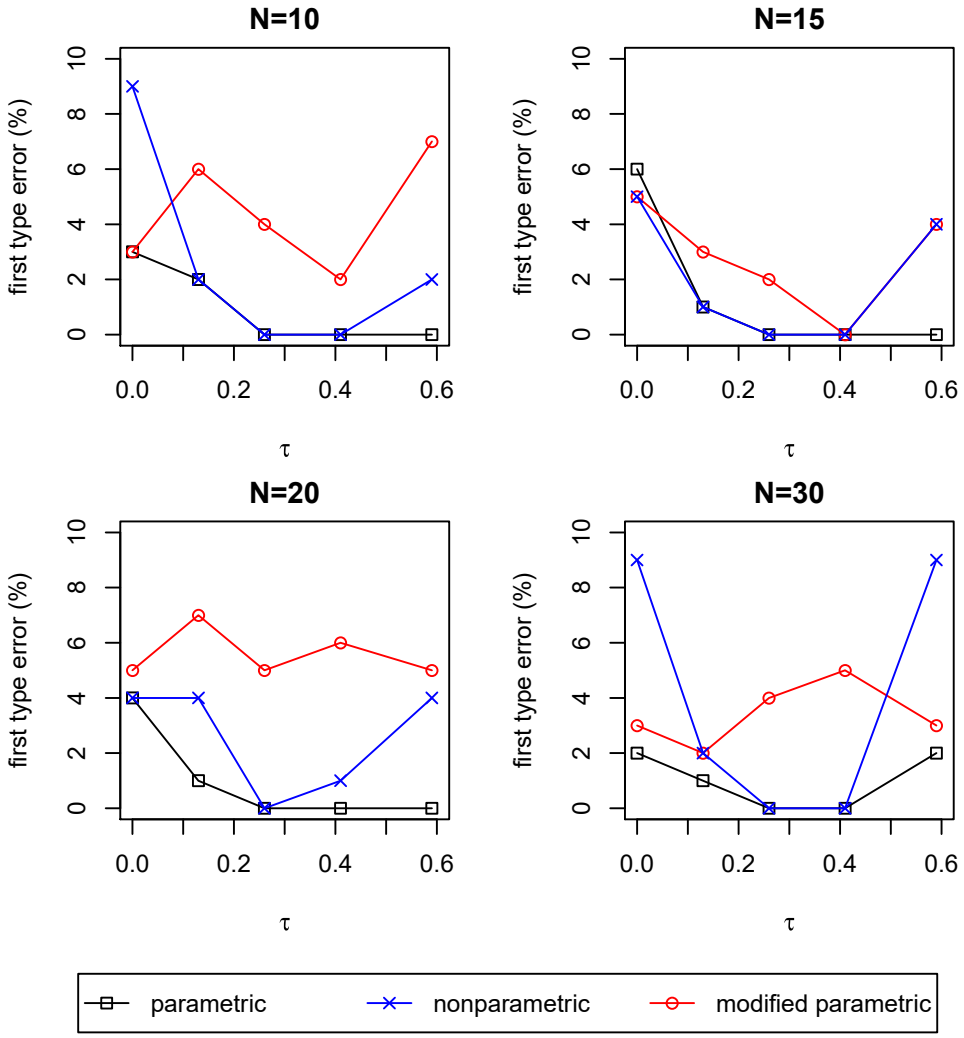


Figure 6.5: Comparison of empirical test first type errors for different cross-correlations and number of sites

nonparametric test is not able to detect heterogeneity when cross-correlation is present, and hence the modified test deals better with cross-correlation. Usually, the power of the modification increases up to $\tau \approx 0.4$ and then decreases, but it is still larger than for the other tests. This shows that the modified test does not reveal real degree of heterogeneity when the regions are very high cross-correlated. These results do not coincide with those obtained in the univariate simulation study performed by Lilienthal, Fried and Schumann [73]. They also utilized the D-vine copula model to generate cross-correlated synthetic regions, however, their

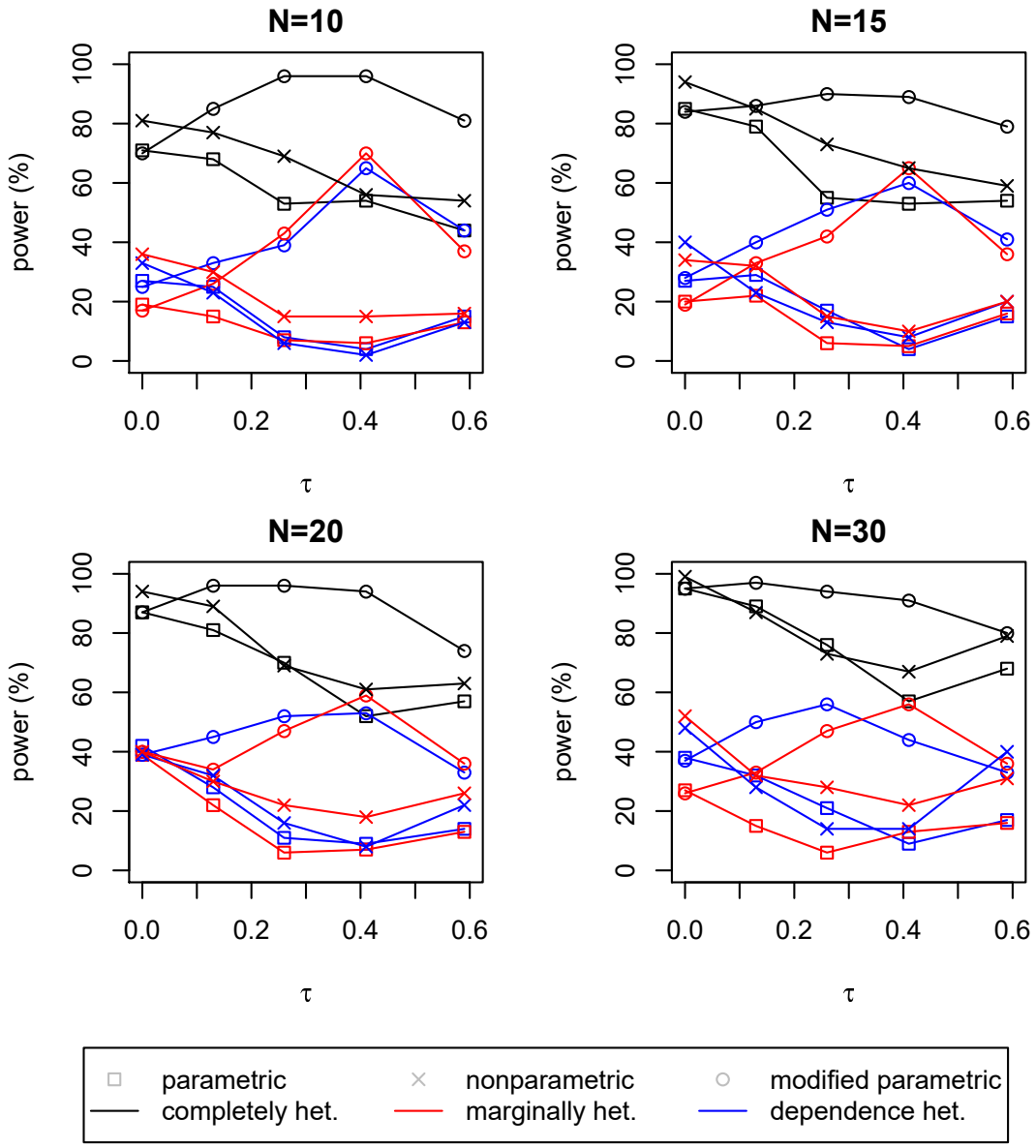


Figure 6.6: Comparison of empirical test powers for completely (black), marginally (blue) and dependence (red) heterogeneous regions with different cross-correlations and number of sites

modification of the Hosking and Wallis test [58] is able to increase the power with increasing cross-correlation, that does not apply to the proposed modification of the bivariate Chebana and Ouarda [22] test.

Analogous observations for heterogeneous regions summarized in the two previous paragraphs hold also for bimodal regions (see Figure 6.7).

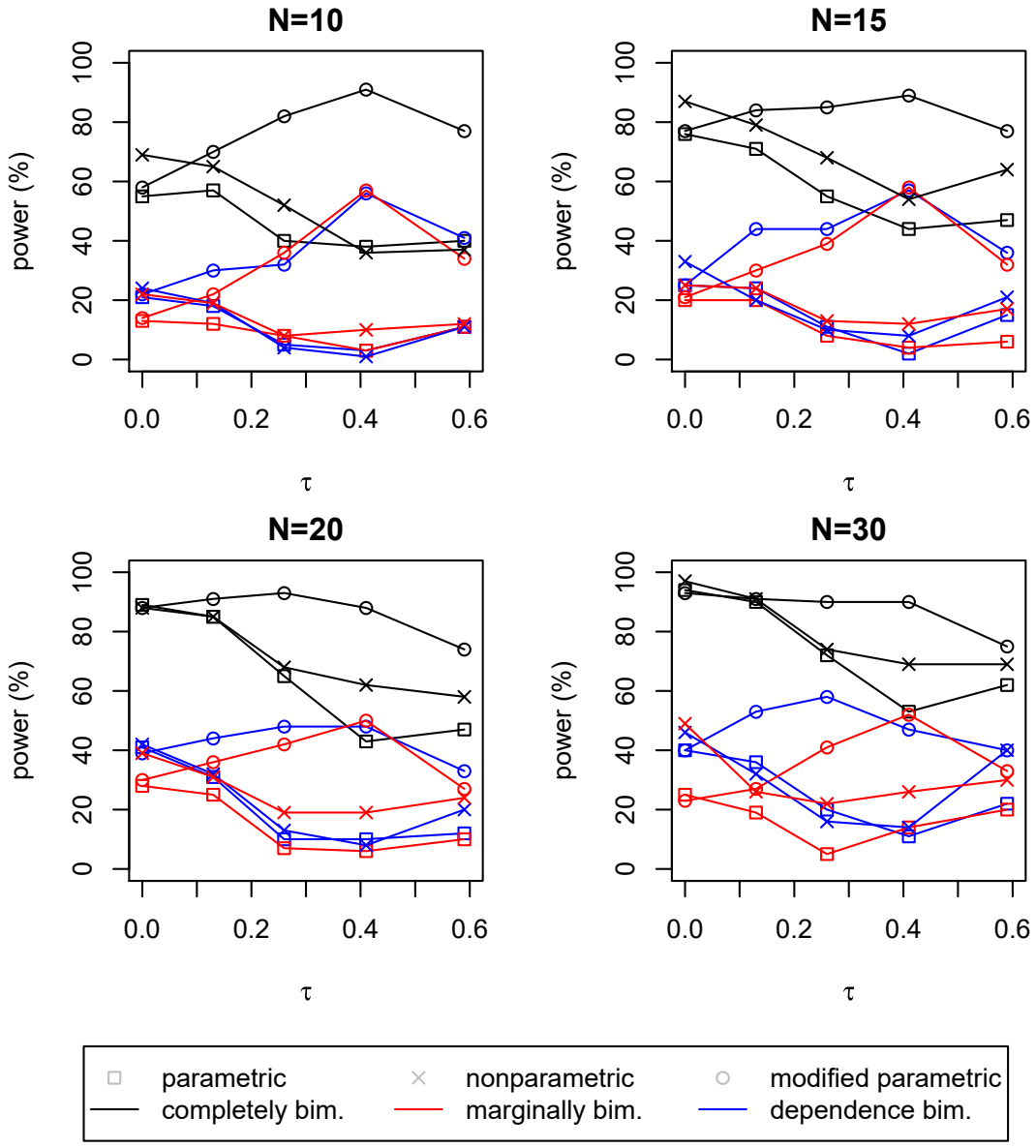


Figure 6.7: Comparison of empirical test powers for completely (black), marginally (blue) and dependence (red) bimodal regions with different cross-correlations and number of sites

Figure 6.8 presents a comparison of average $H_{||\cdot||}$ and $H^*_{||\cdot||}$ values for cross-correlated regions obtained by the original and modified parametric homogeneity tests. It clearly shows that use of the modified test provides approximately the real degree of heterogeneity and may avoid miscategorization of the region in contrast with the original one.

A small simulation study was also performed to find out the performance of the modified parametric test for uncross-correlated regions. It can be observed that also in originally uncross-correlated regions there are nonzero cross-correlations between data at different sites. This fact might raise a question how to find out whether the studied region is "originally" uncross-corellated or cross-correlated, and which test is the best one, because the original parametric and nonparametric tests work well for uncross-correlated regions, while the modified parametric should be used for cross-correlated regions according to the results presented in the above paragraphs. The results of the simulation experiment carried out for 100 homogeneous and completely heterogeneous regions show that the performance of the original parametric, nonparametric, and modified parametric tests are very similar. For this reason the modified parametric test is recommended, because if there is significant spatial dependence in region, the modified parametric test might capture it and take it into account when generating synthetic homogeneous regions.

6.5 Case Study

Here, the modified procedure is applied to real meteorological data to illustrate its usefulness in practice. For this, homogeneity of region 3 consisting of 33 meteorological stations located in the northeastern part of Moravia and Silesia near the Jeseníky and Beskydy mountains (see Figure 4.7) is checked also by the bivariate modified parametric L-moment homogeneity test. The result about homogeneity is compared to those obtained by the original parametric and nonparametric tests.

Maximum annual 1- and 5-day precipitation totals were used as the input dataset. The minimum record length is 36 years, so at least 36 records are available

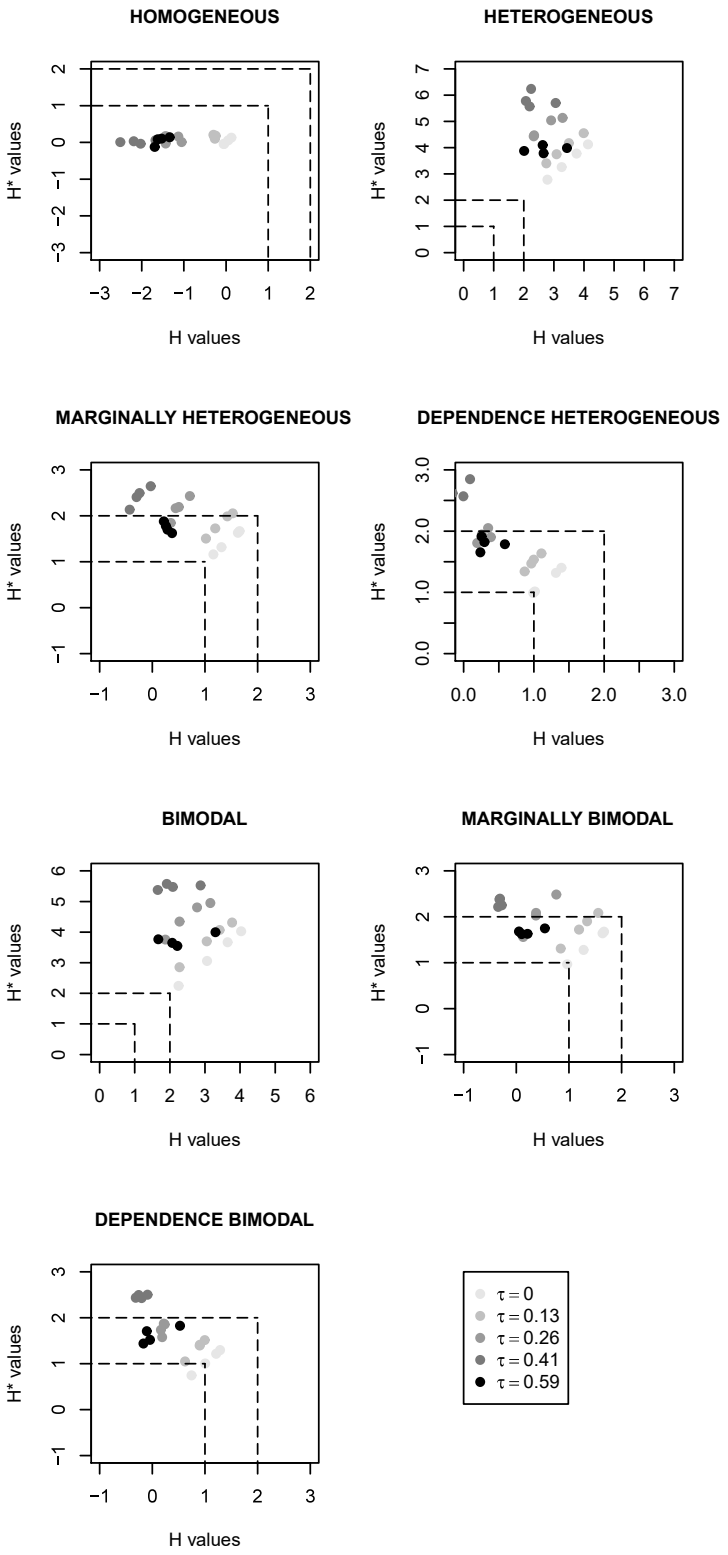


Figure 6.8: Average $H_{||\cdot||}$ and $H^*_{||\cdot||}$ values for cross-correlated regions obtained by the original vs. modified parametric tests

at each station. Note that the elimination of 2 stations in the region would increase the number of measurements to 43.

Although the region may be regarded as homogeneous according to the results of the parametric and nonparametric tests, because $H_{||\cdot||} = -1.3857$ and the p-value is 0.994, there exists correlation between data series (empirical Kendall's τ ranges from -0.45 to 0.87), which could devaluate these results. For this reason, the proposed modified homogeneity test should be preferred over the others. Because two correlated variables are measured at each of 33 stations, the 66-dimensional D-vine copula model is fitted to the data to assess both the dependence between the studied variables at each station and the intersite dependence structure. Hence, 2145 pair-copulas have to be determined and their parameters estimated, but these results are not presented due to the space limitation. Note that the Gumbel copula family was used as the regional bivariate copula. The proposed modified parametric test returns $H^*_{||\cdot||} = 1.5395$, which means that the region is acceptably homogeneous. Hence, the region is not such homogeneous as the other tests indicate.

6.6 Conclusion

This chapter introduces a modification of the bivariate L-moment homogeneity test proposed by Chebana and Ouarda [22] to overcome the problem of miscategorization of region when the data are spatially correlated. D-vine copula are used to preserve the approximate spatial dependence structure. A simulation study has been carried out to analyze how the presence of cross-correlation impacts upon the results of homogeneity testing and to evaluate the performance of the proposed modification. The tests were compared on the basis of the empirical first type errors and powers. As expected according to results of similar studies conducted in the univariate case, the simulation results clearly show that the proposed modification deals best with cross-correlation. The size of the modification may be slightly higher than the value of 5% usually required, but it does not exceed 7%. The emipirical power ranges from 17% to 97% depending on several factors (the degree of heterogeneity, number of sites, strength of cross-correlation). Although the lower limit is quite low, the

power of the modified parametric test is almost higher than for the other tests (except the uncross-correlated case). The modified parametric test is recommended to be used in practice, because it might capture spatial dependence, which impacts the results of the original parametric and nonparametric tests.

The study presents only a first attempt how to deal with cross-correlation in bivariate homogeneity testing in RFA. Hence, some issues should be addressed in future. The impact of various factors, which were not included in the study performed and may affect the performance of the proposed test should be investigated. Such factors are for example different at-site record lengths, difficulties with specification of a D-vine copula model (pair-copula selection and estimation), or misspecification of pair-copulas. Different at-site record lengths represent an important limitation, because the beginnings of records often differ from site to site in practice. This problem may be solved by omitting some measurements at other sites, infilling the missing values, or eliminating the site with a large number of the missing values, however these steps may affect the test performance. Moreover, with increasing number of sites in the region the specification of an appropriate D-vine copula model becomes more time consuming and technically complicated. The D-vine copula model may be simplified by truncation at certain level, which means the use of independence copula for pair-copulas. With this issue, misspecification of pair-copulas, which are unknown in practice, is related. Masselot, Chebana and Ouarda [78] shown that the both univariate and multivariate nonparametric tests are better than the parametric tests in performance as well as in implementation. Here, the focus was only on a generalization of the parametric Hosking and Wallis [58] homogeneity test as a first attempt to deal with cross-correlation. Moreover, it seems that the nonparametric test can not be simply generalized for cross-correlated. It would probably require construction of a new type of a statistical test. However, it would be interesting to do some research also in the nonparametric area and explore, if a nonparametric alternative could ever be proposed. Although the test procedure has been thoroughly described and applied in a case study using real meteorological

data in the study presented, it would be useful to investigate its practical aspects and summarize comprehensive guidelines and recommendations for routine practice.

Conclusions

Central moments and comoments matrices are traditionally used to characterize a probability distribution of a random variable or vector, respectively. However, their drawbacks lie in higher orders moments assumptions, that is unsufficient for analysis of distributions with heavier tails, to which the attention has recently been given not only in environmental fields, such as hydrology, meteorology, and climatology, but also in economics and finance. On the other hand, L-moments proposed by Hosking [55] and Serfling and Xiao [97] as alternative systems to the conventional moments and comoments yield a tool for analysis of a random variable or vector under just first moment assumptions.

In this thesis, both univariate and multivariate L-moments and their generalizations have been used to develop parameter and quantile estimation, as well as hypothesis testing. The thesis was divided into six chapters of which four bring new results. The first one presented the definition of univariate L-moments, their main features, and generalizations. The focus was on the L-moments method which provides parameter estimators of a probability distribution in the same way as it is in the usual moments method, however, they are more robust thanks to the less weights given to the extreme observations. They provide reliable estimates even in small to moderate samples and they are often easy to calculate as well. The highlighted useful feature of sample L-moments based on the asymptotic theory for linear combinations of order statistics is their asymptotic normality [55].

With generalization of L-moments into LQ- and TL-moments [38, 84], new alternatives to the traditional estimation methods appeared. Hence, there exists a relatively large number of methods for estimating unknown parameters of a univariate probability distribution, however, each of them has its strengths and

weakness. Hence, the second chapter dealt with derivation of L-, LQ-, and TL-moments, and parameter and quantile estimators based on these moments of selected univariate probability distributions, and their comparison. The considered distributions were the three-parametric generalized Pareto and generalized extreme-value distributions, which have been frequently used in the modeling extreme events [42, 53, 64, 70, 88], and, therefore, use of L-, LQ-, and TL-moments seem to provide reliable parameter and quantile estimates with respect to their features. The focus was on estimating high quantiles from samples drawn from these distributions. The simulation study performed revealed that the traditional maximum likelihood and L-moments methods surpass completely the LQ- and TL-moments methods with respect to the mean squared error of quantile estimates. Let us note that the LQ- and TL-moments are too robust for the considered range of the shape parameter k, however, they would be reasonable with k approaching to the value -1. It is worth mentioning that the LQ-moments based estimates are much more biased than the estimates obtained by the TL-moments method, moreover their computation is quite difficult and it also includes finding of optimal combination of parameters (p, α).

The L-moments method is a suitable alternative to the maximum likelihood method, and it is typically employed for estimation of unknown parameters of a probability distribution with heavier tails in comparison to the normal distribution when smaller sample is available. Much attention to the point estimation based on L-moments have been given in the literature, but only a little to the confidence intervals. Except for a few univariate probability distributions, such as the uniform or logistic distributions, it is practically impossible to construct exact confidence intervals based on L-moments for parameter or quantile estimates. The only convenient attempt to construct confidence intervals may be based on the asymptotic theory via the delta method. The third chapter presented approximate confidence intervals of parameters and quantile estimates for the three-parametric generalized Pareto and generalized extreme-value distributions obtained using the delta method [92]. Their performances were compared to performances of asymptotic confidence intervals derived for more traditional moments and maximum likelihood methods in terms of

empirical coverage probabilities and median lengths. From results of the performed simulation study can be concluded that the asymptotic confidence intervals based on L-moments for quantile estimates work well even for small sample size about 50 depending on the value of the shape parameter k. In both cases, with increasing value of the shape parameter k and sample size n the empirical coverage probability increases and median lenght decreases. However, the use and accuracy of the L-moments based confidence intervals for the generalized extreme-value distribution heavily depends on the value of the shape parameter k.

Application of L-moments in estimating high quantiles, when there is a lack of available data in a single site but the same variable is measured at many different sites, was introduced in the fourth chapter. This part of the work just presented an approach termed as index-flood based regional frequency analysis in the bivariate context [22, 23, 93]. All the main steps of regional frequency analysis were thouroughly described and special attention was given to bivariate modeling using copulas. The chapter ended by a detailed illustration of the described methodology to the extreme precipitation events characterized by 1- and 5-day maximum annual precipitation totals.

There exist many practical studies dealing with regional frequency analysis, however, all them are focused on univariate, at most bivariate cases. This is due to the fact that multivariate analysis becomes much more complicated and time consuming. Parametric L-moment homogeneity testing includes construction of a multivariate distribution, which is generally not an easy task. The aim of the fifth chapter was to present a detailed procedure of L-moment homogeneity testing in the trivariate framework as the first step of regional frequency analysis, and to investigate its practical aspects. The methodology presented combines recent developments in construction of higher dimensional models, such as the exchangeable Archimedean, fully nested Archimedean, and vine copulas, with discordancy and homogeneity tests proposed by Chebana and Ouarda [22]. Particular attention was devoted to estimation and simulation methods for the considered higher dimensional dependence models. The entire testing process was applied to the three-dimensional

precipitation dataset. Although, the methodology was illustrated on precipitation, it provides general guidelines how to deal with homogeneity testing for higher-dimensional data and may be adapted for various kinds of data.

Development of the parameteric L-moment homogeneity test proposed in the last chapter was motivated by obtaining large negative values of the heterogeneity measure in regional frequency analysis of extreme precipitation events presented in Chapters 4 and 5. Hence, there was a suspicion of presence of cross-correlation between sites as Hosking [54] pointed out in the univariate case. Simulation study was performed to illustrate how cross-correlation between sites may negatively impact results of the parametric and nonparametric L-moment homogeneity tests, both of which do not preserve eventual spatial dependence structure of data when generating homogeneous regions. D-vine copulas were employed to generalize the original parametric L-moment homogenity of Chebana and Ouarda [22] to overcome this problem. The simulation study considering various types of regions, number of sites, and strength of spatial dependence has showed that the proposed modification deals best with presence of cross-correlation, i.e., categorization of region is much more reliable. Although, the empirical first type error for the modified test is slightly higher (about 7%), the empirical power is almost always higher in comparison to the original parametric and nonparametric tests. The improved test was also applied to real meteorological data.

For future research, there are still many issues related to the topics discussed in this thesis. For example, in asymptotic theory it is assumed that sample size n grows infinitely, but in practical applications only finite samples are available of course. Moreover, samples are often of small size and approximation by the normal distribution in these cases may lead to inaccurate estimates. Issue for future work may be then comparison of the asymptotic L-moments based confidence intervals to those obtained by bootstrap techniques [36]. Further, no attention has been given to methods for estimating the confidence regions of the quantile curves in multivariate regional frequency analysis, although the precision of estimated quantile curves should be of interest. The approach of Coblenz, Dyckerhoff and

Grothe [27] to estimate confidence regions for multivariate quantiles based on copulas could be adopted. Since only the parametric L-moment homogeneity test has been developed to overcome the problem of presence of cross-correlation, it is a challenge to investigate the possibilities for improvement also the nonparametric version. However, no approach is immediately apparent. Further, although the modified L-moment homogeneity testing procedure has been in details described and illustrated in the case study, the practical aspects could be thoroughly investigated and recommendations how to overcome some difficulties encountered in routine practice would be summarized to be a guideline for practitioners. Since the L-moment is an L-estimator, the theory of L-estimators [32, 62, 63] may be used to obtain new results valid also for L-moments. Regional frequency analysis is nowadays popular in hydrology, meteorology, and climatology, among others, but in economics and finance, as fields in which extremes also often appear, it is not common. Hence, this approach may be also adopted for these kinds of data.

References

[1] K. Aas and D. Berg. "Models for Construction of Multivariate Dependence – A Comparison Study". In: *The European Journal of Finance* 15.7–8 (2009), pp. 639–659.

[2] P. Ailliot, C. Thompson, and P. Thompson. "Space-Time Modelling of Precipitation by Using a Hidden Markov Model and Censored Gaussian Distributions". In: *Journal of Royal Statistical Society, Series C* 58 (2009), pp. 405–426. DOI: 10.1111/j.1467-9876.2008.00654.x.

[3] H. Akaike. "A New Look at the Statistical Model Identification". In: *IEEE Transactions on Automatic Control* 19 (1974), pp. 716–723.

[4] N.B.M. Ariff. "Regional Frequency Analysis of Maximum Daily Rainfalls Using TL-moment Approach". PhD thesis. Universiti Teknologi Malaysia, 2009.

[5] S.K. Ashour, A.A. El-Sheik, and N.A.T.A. El-Magd. "TL-moments and LQ-moments of the Exponentiated Pareto Distribution". In: *Journal of Scientific Research and Reports* 4.4 (2015), pp. 328–347.

[6] W.H. Asquith. "L-moments and TL-moments of the Generalized Lambda Distribution". In: *Computational Statistics and Data Analysis* 51.9 (2007), pp. 4484–4496.

[7] A. Bárdossy and G. Pegram. "Infilling Missing Precipitation Records – A Comparison of a New Copula-Based Method with Other Techniques". In: *Journal of Hydrology* 519 (2014), pp. 1162–1170. DOI: 10.1016/j.jhydrol.2014.08.025.

[8] T.J. Bedford and R.M. Cooke. "Probability Density Decomposition for Conditionally Dependent Random Variables Modeled by Vines". In: *Annals of Mathematics and Artificial Intelligence* 32 (2001), pp. 245–268.

[9] T.J. Bedford and R.M. Cooke. "Vines - A New Graphical Model for Dependent Random Variables". In: *Annals of Statistics* 30 (2002), pp. 1031–1068. DOI: 10.1214/aos/1031689016.

[10] F. Belzunce et al. "Quantile Curves and Dependence Structure for Bivariate Distributions". In: *Computational Statistics and Data Analysis* 51 (2007), pp. 5112–5129.

[11] M.A. Ben Aissia et al. "Bivariate Index Flood Model for a Northern Case Study". In: *Hydrological Sciences Journal* 60.2 (2015), pp. 247–268.

[12] D. Bílková. "Alternative Means of Statistical Data Analysis: L-moments and TL-moments of Probability Distributions". In: *Statistika: Statistics and Economy Journal* 2 (2014), pp. 77–94.

[13] D. Bílková. "Estimating Parameters of Lognormal Distribution Using the Method of L-Moments". In: *Research Journal of Economics, Business and ICT* 4.1 (2011), pp. 4–9.

[14] B. Brahimi, F. Chebana, and A. Necir. "Copula Representation of Bivariate L-moments: A New Estimation Method for Multiparameter 2-Dimentional Copula Models". In: *Statistics: A Journal of Theoretical and Applied Statistics* 49.3 (2014), pp. 497–521. DOI: 10.1080/02331888.2014.932792.

[15] E.C. Brechmann, C. Czado, and K. Aas. "Truncated Regular Vines in High Dimensions with Application to Financial Data". In: *The Canadian Journal of Statistics* 40.1 (2012), pp. 68–85. DOI: 10.1002/cjs.10141.

[16] K.P. Burnham and D.R. Anderson. *Model Selection and Multimodial Inference: A Practical Information-Theoretic Approach*. Springer-Verlag: New York, 2002.

[17] A. Burton et al. "Models of Daily Rainfall Cross-Correlation for the United Kingdom". In: *Environmental Modelling & Software* 49 (2013), pp. 22–33. DOI: 10.1016/j.envsoft.2013.06.001.

[18] A. Castellarin, D.H. Burn, and A. Brath. "Homogeneity Testing: How Homogeneous Do Heterogeneous Cross-Correlated Regions Seem?" In: *Journal of Hydrology* 360 (2008), pp. 67–76.

[19] A. Castellarin, R.M. Vogel, and N.C. Matalas. "Multivariate Probabilistic Regional Envelopes of Extreme Floods". In: *Journal of Hydrology* 336 (2007), pp. 377–390.

[20] L.K. Chan. "On a Characterization of Distributions by Expected Values of Extreme Order Statistics". In: *American Mathematical Monthly* 74 (1967), pp. 950–951.

[21] F. Chebana and T.B.M.J. Ouarda. "Index-Flood Based Multivariate Regional Frequency Analysis". In: *Water Resources Research* 45.10 (2009). DOI: 10.1029/2008WR007490.

[22] F. Chebana and T.B.M.J. Ouarda. "Multivariate L-moment Homogeneity Test". In: *Water Resources Research* 43.8 (2007). DOI: 10.1029/2006WR005639.

[23] F. Chebana et al. "Multivariate Homogeneity Testing in a Northern Case Study in the Province of Quebec, Canada". In: *Hydrological Processes* 23.12 (2009), pp. 1690–1700.

[24] Y.D. Chen et al. "Regional Analysis of Low Flow using L-moments for Dongjiang Basin, South China". In: *Hydrological Sciences Journal* 51.6 (2006), pp. 1051–1064.

[25] H. Chernoff, J.L. Gastwirth, and M.V. Johns. "Asymptotic Distribution of Linear Combinations of Functions of Order Statistics with Applications to Estimation". In: *The Annals of Mathematical Statistics* 38 (1967), pp. 52–72.

[26] R. Christie-David and M. Chaudhry. “Coskewness and Cokurtosis in Futures Markets”. In: *Journal of Empirical Finance* 8 (2001), pp. 55–81.

[27] M. Coblenz, R. Dyckerhoff, and O. Grothe. “Confidence Regions for Multivariate Quantiles”. In: *Water* 10.8 (2018). DOI: 10.3390/w10080996.

[28] R.G. Cong and M. Brady. “The Interdependence Between Rainfall and Temperature: Copula Analyses”. In: *The Scientific World Journal* 2012.ID 405675 (2012).

[29] L. Coufal, P. Langová, and T. Míková. *Meteorological Data in the Czech Republic in the Period 1961–1990*. Tech. rep. Czech Hydrometeorological Institute, 1992.

[30] C. Czado, A. Frigessi, and H. Bakken. “Pair-Copula Constructions of Multiple Dependence”. In: *Insurance: Mathematics and Economics* 44.2 (2009), pp. 182–198. DOI: 10.1016/j.insmatheco.2007.02.001.

[31] T. Dalrymple. “Flood Frequency Analyses”. In: *US Geological Survey. Water Supply Paper* 1543 A (1960).

[32] C. Daniel. “Observations weighted according to order”. In: *American Journal of Mathematics* 42 (1920).

[33] A. Decurninge. *Multivariate Quantiles and Multivariate L-moments*. 2015. arXiv: 1409.6013.

[34] F. Downton. “Linear Estimates with Polynomial Coefficients”. In: *Biometrika* 53 (1966), pp. 129–141.

[35] D.J. Dupius and C. Field. “A Comparison of Confidence Intervals for Generalized Extreme-Value Distributions”. In: *Journal of Statistical Computation and Simulation* 61.4 (1998), pp. 341–360. DOI: 10.1080/00949659808811918.

[36] B. Efron and R.J. Tibshirani. *An Introduction to the Bootstrap*. Chapmann & Hall, 1994.

[37] N.A.T.A. El-Magd. “TL-moments of Exponentiated Generalized Extreme-Value Distribution”. In: *Journal of Advanced Research* 1.4 (2010), pp. 351–359.

[38] E.A.H. Elamir and A.H. Seheult. “Trimmed L-moments”. In: *Computational Statistics and Data Analysis* 43.3 (2003), pp. 299–314.

[39] A.-C. Favre et al. “Multivariate Hydrological Frequency Analysis Using Copulas”. In: *Water Resources Research* 40 (2004). DOI: 10.1029/2003WR002456.

[40] G. Frahm, M. Junker, and R. Schmidt. “Estimating the Tail-Dependence Coefficient: Properties and Pitfalls”. In: *Insurance: Mathematics and Economics* 37 (2005), pp. 80–100.

[41] P.H.A.J.M. van Gelder and M.D. Pandey. “An Analysis of Drag Forces Based on L-moments”. In: *Advances in Safety and Reliability - Kolowrocki (ed.)* Delft University of Technology. 2005.

[42] D. Gellens. “Combining Regional Approach and Data Extension Procedure for Assessing GEV Distribution of Extreme Precipitation in Belgium”. In: *Journal of Hydrology* 268.1–4 (2002), pp. 113–126.

[43] C. Genest and A.-C. Favre. “Everything You Always Wanted to Know About Copula Modeling but Were Afraid to Ask”. In: *Journal of Hydrologic Engineering* 12 (2007), pp. 347–368. DOI: 10.161/(ASCE)1084-0699(2007)12:4(347).

[44] C. Genest, K. Ghoudi, and L.-P. Rivest. “A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions”. In: *Biometrika* 82 (1995), pp. 543–552.

[45] C. Genest, B. Rémillard, and D. Beaudoin. “Goodness-of-Fit Tests for Copulas: A Review and a Power Study”. In: *Insurance: Mathematics and Economics* 44 (2009), pp. 199–213.

[46] C. Gini. “Variabilita e Mutabilita. Contributio allo studio delle distribuzioni e delle relazione statistiche”. In: *Studi Economico-Giuridici della Reale Universita di Cagliari* 3 (1912), pp. 3–159.

[47] C. Gourieroux and J. Jasiak. “Dynamic Quantile Models”. In: *Journal of Econometrics* 147 (2008), pp. 198–205.

[48] J.A. Greenwood et al. “Probability Weighted Moments: Definition and Relation to Parameters of Several Distributions Expressable in Inverse Form”. In: *Water Resources Research* 15 (1979), pp. 1049–1054.

[49] J.-H. Heo and J.D. Salas. “Estimation of Quantiles and Confidence Intervals for the Log-Gumbel Distribution”. In: *Stochastic Hydrology and Hydraulics* 10 (1996), pp. 187–207.

[50] J.-H. Heo, J.D. Salas, and K.-D. Kim. “Estimation of Confidence Intervals of Quantiles for the Weibull Distribution”. In: *Stochastic Environmental Research and Risk Assessment* 15 (2001), pp. 284–309.

[51] M. Hofert and M. Maechler. “Nested Archimedean Copulas Meet R: The nacopula Package”. In: *Journal of Statistical Software* 39.9 (2011), pp. 1–20.

[52] M. Hohenwarter et al. *Geogebra*. Version 5.0.512.0-d. Sept. 4, 2018. URL: https://wwwgeogebra.org.

[53] J.D. Holmes and W.W. Moriarty. “Application of the Generalized Pareto Distribution to Extreme Value Analysis in Wind Engineering”. In: *Journal of Wind Engineering and Industrial Aerodynamics* 83.1–3 (1999), pp. 1–10.

[54] J.M.R. Hosking and J.R. Wallis. *Regional Frequency Analysis: An Approach Based on L-moments*. Cambridge University Press, 1997.

[55] J.R.M. Hosking. “L-moments: Analysis and Estimation of Distributions using Linear Combinations of Order Statistics”. In: *Journal of the Royal Statistical Society, Series B* 52 (1990), pp. 105–124.

[56] J.R.M. Hosking and N. Balakrishnan. “A Uniqueness Result for L-Estimators, with Applications to L-moments”. In: *Statistical Methodology* 24 (2015), pp. 69–80.

[57] J.R.M Hosking and J.R. Wallis. “Parameter and Quantile Estimation for the Generalized Pareto Distribution”. In: *Technometrics* 29.3 (1987), pp. 339–349.

[58] J.R.M. Hosking and J.R. Wallis. “Some Statistics Useful in Regional Frequency Analysis”. In: *Water Resources Research* 29.2 (1993), pp. 271–281.

[59] J.R.M. Hosking, J.R. Wallis, and E.F. Wood. “Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments”. In: *Technometrics* 27.3 (1985), pp. 251–261.

[60] H. Joe. *Multivariate Models and Dependence Concepts*. Chapman & Hall, 1997.

[61] H. Joe and J.J. Xu. *The Estimation Method of Inference Functions for Margins for Multivariate Models*. Tech. rep. 166. The University of British Columbia, 1996.

[62] J. Jurečková and J. Picek. *Robust Statistical Methods with R*. Chapman & Hall, 2006.

[63] J. Jurečková and P.K. Sen. *Robust Statistical Procedures: Asymptotics and Interrelations*. Wiley-Interscience, 1996.

[64] R.W. Katz, M.B. Parlange, and P. Naveau. “Statistics of Extremes in Hydrology”. In: *Advances in Water Resources* 25 (2002), pp. 1287–1304.

[65] A.G. Konheim. “A Note on Order Statistics”. In: *American Mathematical Monthly* 78 (1971), p. 524.

[66] D. Kurowicka and R.M. Cooke. “Sampling Algorithms for Generating Joint Uniform Distributions Using Vine-Copula Method”. In: *Computational Statistics and Data Analysis* 51 (2007), pp. 2889–2906. DOI: 10.1016/j.csda.2006.11.043.

[67] D. Kurowicka and H. Joe. *Dependence Modeling: Vine Copula Handbook*. World Scientific, 2011.

[68] J. Kyselý. “Trend in Heavy Precipitation in the Czech Republic over 1961–2005”. In: *International Journal of Climatology* 29 (2009), pp. 1745–1758.

[69] J. Kyselý, L. Gaál, and J. Picek. “Comparison of Regional and At-site Approaches to Modelling Probabilities of Heavy Precipitation”. In: *International Journal of Climatology* 31 (2011), pp. 1457–1472.

[70] J. Kyselý and J. Picek. “Regional Growth Curves and Improved Design Value Estimates of Extreme Precipitation Events in the Czech Republic”. In: *Climate Research* 33 (2007), pp. 243–255.

[71] J. Kyselý, J. Picek, and R. Huth. “Formation of Homogeneous Regions for Regional Frequency Analysis of Extreme Precipitation Events in the Czech Republic”. In: *Studia Geophysica et Geodaetica* 51 (2006), pp. 327–334.

[72] F. Laio. “Cramer-von Mises and Anderson-Darling Goodness of Fit Tests for Extreme Value Distributions with Unknown Parameters”. In: *Water Resources Research* 40.9 (2004). DOI: 10.1029/2004WR003204.

[73] J. Lilienthal, R. Fried, and A.H. Schumann. “Homogenity Testing for Skewed and Cross-Correlated Data in Regional Flood Frequency Analysis”. In: *Journal of Hydrology* 556 (2018), pp. 557–571. DOI: 10.1016/j.jhydrol.2017.10.056.

[74] S. Mahdi and F. Ashkar. “Exploring Generalized Probability Weighted Moments, Generalized Moments and Maximum Likelihood Estimating Methods in Two-Parameter Weibull Model”. In: *Journal of Hydrology* 285.1–4 (2004), pp. 62–75.

[75] C.L. Mallows. “Bounds on Distribution Functions in Terms of Expectations of Order Statistics”. In: *Annals of Probability* 1 (1973), pp. 297–303.

[76] A.W. Marshall and I. Olkin. “Families of Multivariate Distributions”. In: *Journal of the American Statistical Association* 83 (1988), pp. 834–841.

[77] E.S. Martins and J.R. Stedinger. “Generalized Maximum-Likelihood Generalized Extreme-Value Quantile Estimators for Hydrologic Data”. In: *Water Resources Research* 36.3 (2000), pp. 737–744.

[78] P. Masselot, F. Chebana, and T.B.M.J. Ouarda. “Fast and Direct Nonparametric Procedures in the L-moment Homogeneity Test”. In: *Stochastic Environmental Research and Risk Assessment* (2016). DOI: 10.1007/s00477-016-1248-0.

[79] N.C. Matalas and W.B. Langbein. “Information Content of the Mean”. In: *Journal of Geophysical Research* 67.9 (1962), pp. 3441–3448.

[80] A.J. McNeil. “Sampling Nested Archimedean Copulas”. In: *Journal of Statistical Computation and Simulation* 78 (2008), pp. 567–581.

[81] C. de Michele and G. Salvadori. “A Generalized Pareto Intensity-Duration Model of Storm Rainfall Exploiting 2-Copulas”. In: *Journal of Geophysical Research* 108.D2 (2003). DOI: 10.1029/2002JD002534.

[82] F. Michiels and A. de Schepper. “A Copula Test Space Model How to Avoid the Wrong Copula Choice”. In: *Kybernetika (Prague)* 44 (2008), pp. 864–878.

[83] D.S. Moore. “An Elementary Proof of Asymptotic Normality of Linear Functions of Order Statistics”. In: *The Annals of Mathematical Statistics* 39 (1968), pp. 263–265.

[84] G.S. Mudholkar and A.D. Hutson. “LQ-moments: Analogs of L-moments”. In: *Journal of Statistical Planning and Inference* 71 (1998), pp. 191–208.

[85] R.B. Nelsen. *An Introduction to Copulas*. Springer-Verlag New York, 2006.

[86] L.V. Noto and G. La Loggia. “Use of L-moments Approach for Regional Flood Frequency Analysis in Sicily, Italy”. In: *Water Resources Management* 23 (2009), pp. 2207–2229.

[87] O. Okhrin and A. Ristig. “Hierarchical Archimedean Copulae: The HAC Package”. In: *Journal of Statistical Software* 58.4 (2014), pp. 1–20.

[88] M.M. de Oliveira et al. “Generalized Extreme Wind Speed Distributions in South America over the Atlantic Ocean Region”. In: *Theoretical and Applied Climatology* 104 (2011), pp. 377–385.

[89] C.P. Pearson. “New Zealand Regional Flood Frequency Analysis Using L-moments”. In: *Journal of Hydrology (New Zealand)* 30.2 (1991), pp. 53–64.

[90] P. Prescott and A.T. Walden. “Maximum Likelihood Estimation of the Parameters of the Generalized Extreme-Value Distribution”. In: *Biometrika* 67.3 (1980), pp. 723–724.

[91] R Core Team. *A Language and Environment for Statistical Computing*. Version 3.3.1. Sept. 4, 2018. URL: `https://www.r-project.org`.

[92] C.R. Rao. *Linear Statistical Inference and Its Applications*. John Wiley & Sons, 1973.

[93] A.I. Requena, L. Mediero, and L. Garrotel. “A Complete Procedure for Multivariate Index-Flood Model Application”. In: *Journal of Hydrology* 535 (2016), pp. 559–580.

[94] M. Rosenblatt. “Remarks on a Multivariate Transformation”. In: *Annals of Multivariate Statistics* 23.2 (1952), pp. 470–472.

[95] A. Sahbri and A.A. Jemain. “Fitting the Generalized Logistic Distribution by LQ-moments”. In: *Applied Mathematical Sciences* 5.54 (2011), pp. 2663–2676.

[96] G.A. Schwarz. “Estimating the Dimension of a Model”. In: *The Annals of Statistics* 6 (1978), pp. 461–464.

[97] R. Serfling and P. Xiao. “A Contribution to Multivariate L-comoments: L-comoment Matrices”. In: *Journal of Multivariate Analysis* 98 (2007), pp. 1765–1781.

[98] A. Shabri and A.A. Jemain. “Estimation of the Extreme Value Type I Distribution by the Method of LQ-moments”. In: *Journal of Mathematics and Statistics* 5.4 (2009), pp. 298–304.

[99] A. Shabri and A.A. Jemain. “LQ-moments: Parameter Estimation for Kappa Distribution”. In: *Sains Malaysiana* 39.5 (2010), pp. 845–850.

[100] M.N. Shahzad and Z. Asghar. “Comparing TL-moments, L-moments and Conventional Moments of Dagum Distribution by Simulated Data”. In: *Revista Colombiana Estadistica* 36.1 (2013), pp. 79–93.

[101] G.P. Sillitto. “Derivation of Approximants to the Inverse Distribution Function of a Continuous Univariate Population from the Order Statistics of a Sample”. In: *Biometrika* 56 (1969), pp. 641–650.

[102] G.P. Sillitto. “Interrelations Between Certain Linear Systematic Statistics of Samples from Any Continuous Population”. In: *Biometrika* 38 (1951), pp. 377–382.

[103] A.L.C. da Silva and B.V. de Melo Mendes. “Value-at-Risk and Extreme Returns in Asian Stock Markets”. In: *International Journal of Business* 8.1 (2003), pp. 17–40.

[104] V.P. Singh and H. Guo. “Parameter Estimation for 3-Parameter Generalized Pareto Distribution by the Principle of Maximum Entropy (POME)”. In: *Hydrological Sciences Journal* 40.2 (1995), pp. 165–181.

[105] A. Sklar. “Fonctions de Répartition à n Dimensions et Leurs Marges”. In: *Publications de l’Institut de Statistique de l’Universite de Paris* 8 (1959), pp. 229–231.

[106] R.L. Smith. “Threshold Methods for Sample Extremes”. In: *Statistical Extremes and Applications (ed. J.T. de Oliveira)* (1984), pp. 621–638.

[107] J.R. Stedinger. “Estimating a Regional Flood Frequency Distribution”. In: *Water Resources Research* 19.2 (1983), pp. 503–510.

[108] S.M. Stigler. “Linear Functions of Order Statistics with Smooth Weight Functions”. In: *The Annals of Statistics* 2 (1974), pp. 676–693.

[109] A. Stuart and J.K. Ord. *Kendall’s Advanced Theory of Statistics*. Charles Griffin & Company, 1987.

[110] K. Tolikas. “Value-at-Risk and Extreme-Value Distributions for Financial Returns”. In: *Journal of Risk* 10.3 (2008), pp. 31–77.

[111] P.K. Trivedi and D.M. Zimmer. “Copula Modeling: An Introduction for Practitioners”. In: *Foundations and Trends in Econometrics* 1.1 (2005), pp. 1–111.

[112] H. Tsukahara. “Semiparametric Estimation in Copula Models”. In: *The Canadian Journal of Statistics* 33 (2005), pp. 357–375.

[113] A. Viglione, F. Laio, and P.A. Claps. “A Comparison of Homogeneity Tests for Regional Frequency Analysis”. In: *Water Resources Research* 43.3 (2006). DOI: 10.1029/2006WR005095.

[114] T. Šimková. “Asymptotic Confidence Intervals of Qunatile Estimates for the GP and GEV Distributions”. In preparation.

[115] T. Šimková. “Homogeneity Testing for Spatially Cross-Correlated Data in Multivariate Regional Frequency Analysis”. In: *Water Resources Research* 53.8 (2017), pp. 7012–7028.

[116] T. Šimková. “L-Moment Homogeneity Test in Trivariate Regional Frequency Analysis of Extreme Precipitation Events”. In: *Meteorological Applications* 25.1 (2018), pp. 11–22. DOI: 10.1002/met.1664.

[117] T. Šimková. “Statistical Inference Based on L-Moments”. In: *Statistika: Statistics and Economy Journal* 97.1 (2017), pp. 44–58.

[118] T. Šimková and J. Picek. “A Comparison of L-, LQ- and TL-Moments and Maximum Likelihood High Quantile Estimates of the GP and GEV Distribution”. In: *Communications in Statistics - Simulation and Computation* 46.8 (2017), pp. 5991–6010.

[119] Q.H. Vuong. “Likelihood Ratio Tests for Model Selection and Non-Nested Hypothesis”. In: *Econometrica* 57 (1989), pp. 307–333.

[120] Q.J. Wang. “Direct Sample Estimators of L-Moments”. In: *Water Resources Research* 32 (1996), pp. 3617–3619.

[121] S. Yue and P.F. Rasmussen. “Bivariate Frequency Analysis: Discussion of Some Useful Concepts in Hydrological Applications”. In: *Hydrological Processes* 16 (2002), pp. 2881–2898. DOI: 10.1002/hyp.1185.

[122] L. Zhang and V.P. Singh. “Bivariate Rainfall Frequency Distributions Using Archimedean Copulas”. In: *Journal of Hydrology* 332 (2007), pp. 93–109. DOI: 10.1016/j.jhydrol.2006.06.033.