Thèse n° 8168

Inference on the Angular Distribution of Extremes

Présentée le 25 novembre 2020

à la Faculté des sciences de base
Chaire de statistique
Programme doctoral en mathématiques

pour l'obtention du grade de Docteur ès Sciences

par

Claudio Andri SEMADENI

Acceptée sur proposition du jury

Prof. V. Panaretos, président du jury
Prof. A. C. Davison, directeur de thèse
Prof. V. Chavez-Demoulin, rapporteuse
Prof. S. Padoan, rapporteur
Prof. S. Morgenthaler, rapporteur

École
polytechnique
fédérale
de Lausanne

2020

Whether you believe you can do
a thing or not, you are right.
— Henry Ford

To Paloma

Acknowledgements

I wish to express my sincere appreciation to my supervisor, Anthony C. Davison, who gave me the opportunity to undertake this thesis. Without his constant availability and valuable guidance, and his support, patience and understanding when progress was slow, the goal of this project would not have been realized.

I would like to thank the president of the jury, Victor Panaretos, and the members of the jury, Valérie Chavez-Demoulin, Stephan Morgenthaler, and Simone Padoan, for sitting on the thesis jury and for their careful reading of the thesis and detailed comments.

I would also like to thank Valérie Chavez-Demoulin who made me discover the field of extreme-value theory through her book recommendation before I started my thesis, and Christophe Osinski and Mariá Süveges for their insights on completing a PhD thesis whilst I was considering undertaking doctoral studies after a decade working in the finance industry.

I'm grateful to Jonathan Tawn for valuable comments on the coefficient of extremal asymmetry at a very early stage.

Over the course of my work at EPFL, I shared offices with many colleagues. I would like to thank them all, and in particular Jacques Ferrez, Miguel de Carvalho, Linda Mhalla, and Léo Belzile for our numerous and valuable exchanges on our respective research and discussions on more casual topics. I would also like to thank all my colleagues from the STAT, STAP and SMAT chairs who contributed over the years making the work environment friendly and stimulating.

I would like to thank all my friends and I'm in particular very grateful to Grégoire, Yvan, Fréd, Franck, and Christian for their friendship and support over the years, Alessandra and Talia for their presence and support, and Gwenol, Rodrigue and Sacha for their insights on numerical optimization problems.

Finally, I would like to thank my parents, Marie-Joëlle and Andri, and my sister, Sandra, for their great kindness and love and their continuous support, especially during the difficult times of this undertaking, and my wonderful daughter, Paloma, who provides unending inspiration and is the pride of my life.

Lausanne, 20 July 2020 C. S.

Abstract

The spectral distribution plays a key role in the statistical modelling of multivariate extremes, as it defines the dependence structure of multivariate extreme-value distributions and characterizes the limiting distribution of the relative sizes of the components of large multivariate observations. No parametric family captures all possible types of multivariate dependence, and numerous parametric models have been proposed.

Inference on the spectral distribution is typically based on the pseudo-angles of 'large' observations under the assumption that they follow the spectral distribution. There has been little attention on studying the impact of this approximation on inference, and it turns out that it can yield significantly biased estimates. We provide a characterization of the angular distribution of excesses corresponding to the distribution of pseudo-angles of 'large' observations that improves direct inference on the spectral distribution in the bivariate setting.

Extremal dependence is at the heart of extreme value modelling and numerous measures to quantify it have been proposed in the literature. In many applications, datasets seem to exhibit asymmetry in the dependence between the variables. Many parametric multivariate extreme-value models can accommodate asymmetry in the sense that the spectral density can be asymmetric, resulting in a non-exchangeable dependence structure. There has been little attention paid to quantifying asymmetry at extreme levels, which can be useful for diagnosis and model checking. We propose a coefficient of extremal asymmetry that quantifies the asymmetry at extreme levels for pairs of variables. We also propose two non-parametric estimators of the coefficient of extremal asymmetry and compare their properties through numerical simulation. The two estimators have diametrically opposed bias-variance trade-offs. The estimator based on maximum empirical likelihood performs well and is nearly unbiased.

Key words: Angular distribution of extremes; Bivariate extreme-value modelling; Coefficient of extremal asymmetry; Pickands dependence function; Spectral distribution.

Résumé

La distribution spectrale joue un rôle clé dans la modélisation statistique des valeurs extrêmes multivariées car elle défini la structure de dépendance des distributions des valeurs extrêmes multivariées et elle caractérise la distribution limite des tailles relatives des composantes des observations extrêmes. Aucune famille de distribution paramétrique ne couvre l'ensemble des types de dépendances multivariées possibles et de multiples familles de modèles paramétriques ont été proposés dans la littérature scientifique.

L'inférence sur la distribution spectrale est typiquement effectuée par le biais des pseudo-angles des plus grandes observations de l'échantillon sous l'hypothèse qu'ils proviennent de la distribution spectrale elle-même. Peu d'attention a été portée sur l'impact de cette approximation sur l'inférence et il se trouve qu'elle peut engendrer des estimateurs fortement biaisés. Nous proposons une caractérisation de la distribution angulaires des excès de seuils qui correspond à la distribution des pseudo-angles des plus grandes observations et qui améliore l'inférence directe sur la distribution spectrale dans la cas bivarié.

La dépendance extrémale est au coeur de la modélisation des valeurs extrêmes et de multiples measures pour quantifier cette dépendance ont été proposées dans la littérature scientifique. Dans les applications, de nombreux jeux de données semblent présenter une structure de dépendance entre les variables qui soit assymétrique. De nombreux modèles paramétriques pour les valeurs extrêmes peuvent tenir compte de cette assymétrie dans le sens où la distribution spectrale peut également être assymétrique conduisant à des structures de dépendance dont les variables aléatoires ne sont pas échangeables. Peu d'attention a été portée sur la quantification de l'assymétrie de la dépendance entre les variables aux niveaux extrêmes alors que cela pourrait être utile comme moyen de diagnostique et dans la validation de modèles.

Nous proposons un coefficient d'assymétrie extrémale qui quantifie l'assymétrie entre les réalisations extrêmes d'une paire de variables aléatoires. Nous proposons également deux estimateurs non paramétriques de ce coefficient d'assymétrie extrémale et nous comparons leurs propriétés par le biais de simulations numériques. Ces deux estimateurs ont des caractéristiques de biais et de variance diamétralement opposées. L'estimateur basé sur la maximisation de la vraisemblance empirique est meilleur et peu biaisé.

Mots clefs : Coefficient d'assymétrie extrémale ; Distribution angulaire de valeurs extrêmes ; Distribution spectrale ; Fonction de dépendance de Pickands ; Modélisation de valeurs extrêmes bivariées.

Contents

List of Figures

List of Tables

Abbreviations and notation

Abbreviations

CDF	cumulative distribution function
EVT	extreme value theory
GEV	generalized extreme value (distribution)
GP	generalized Pareto (distribution)
i.i.d.	independent and identically distributed
PDF	probability density function
MEV	multivariate extreme value (distribution)
MISE	mean integrated squared error
RMSE	root mean squared error

Notation

Throughout the thesis, vectors are denoted by bold fonts, whilst scalars are denoted by normal fonts. For example, $\boldsymbol{a} = (a_1, \ldots, a_d)$ denotes a vector in $\mathbb{R}^d$. The transpose of a vector $\boldsymbol{a}$ is denoted $\boldsymbol{a}^\mathrm{T}$. Operations on vectors are defined componentwise. For example, if $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{x} \in \mathbb{R}^d$, then

$$\begin{aligned}
\max(\boldsymbol{a}) &= \max(a_1, \ldots, a_d), \\
[\boldsymbol{a}, \boldsymbol{b}] &= [a_1, b_1] \times \cdots \times [a_d, b_d], \\
(\boldsymbol{x} - \boldsymbol{b}) / \boldsymbol{a} &= ((x_1 - b_1)/a_1, \ldots, (x_d - b_d)/a_d), \\
\boldsymbol{a} < \boldsymbol{b} &\Longleftrightarrow a_1 < b_1, \ldots, a_d < b_d.
\end{aligned}$$

Random variables and random vectors are denoted with capital letters. The convergence in distribution of a sequence of random variables $\{X_n\}_{n \geq 1}$ to a random variable X as $n \to \infty$ is denoted by $X_n \xrightarrow{\mathrm{D}} X$.

1 A review of extreme value theory

Extreme value theory (EVT) is the branch of probability theory focussing on the study of the asymptotic distribution of extreme events, that is, observations that are rare in frequency and large in magnitude compared to the bulk of observations. Statistics of extremes is the branch of statistical science concerned with inference on extremes and rare events. It is based on probability models that allow extrapolation into the tail of distributions, often beyond the largest observed data.

This chapter provides a survey of extreme value theory and the statistics of extremes in the finite-dimensional case. The aim is to provide the background for the rest of the thesis, and its content is well-established in the literature. An accessible introduction to EVT is Coles (2001). Early monographs and texts include Gumbel (1958), Resnick (1987), and Galambos (1987). An important contribution in the univariate setting is Embrechts *et al.* (1997). Leadbetter *et al.* (1983) is a standard reference to the literature on extremes of time series. More recent contributions covering also multivariate EVT include Kotz and Nadarajah (2000), Beirlant *et al.* (2004), de Haan and Ferreira (2006), and Resnick (2007). References with a focus on applications in fields such as environmental sciences, telecommunications, finance, and insurance include Finkenstädt and Rootzén (2004), Castillo *et al.* (2005), and Reiss and Thomas (2007). Recent edited volumes include Dey and Yan (2016) and Longin (2017), the latter being focused on applications of EVT in finance. Extreme value methodology is being increasingly used by practitioners from a wide range of fields. Reviews of software for the statistical modelling of extreme events include Gilleland *et al.* (2012) and Gilleland (2016).

Models for univariate extreme values are described in Section 1.1, and models for multivariate extremes discussed in Section 1.2. Some topics which are less relevant for the rest of the thesis are not discussed in this chapter. These include the modelling of spatial extremes, which has attracted a lot of attention in the extreme value community over the past few years; see Davison *et al.* (2012), Cooley *et al.* (2012), and Ribatet *et al.* (2016) for reviews and discussion of recent developments.

1.1 Univariate extremes

The analysis of extreme values dates back to the 1920s, with the characterization of the asymptotic behaviour of the largest value of a univariate sample. There are many results describing the stochastic behaviour of sample maxima and minima, upper order statistics such as the r largest values in a sample, and sample excesses over a high threshold. The two main kinds of models for exteme values are the block maximum model and the threshold excesses model.

The block maximum model described in Section 1.1.1 is for the stochastic behaviour of the largest observation from large samples of independent and identically distributed random variables. The threshold excesses model described in Section 1.1.2 is for the stochastic behaviour of observations that exceed a high level, and it is considered more useful for practical applications as it makes a more efficient use of the data. The threshold excesses model and the model for the r largest values are both special cases of a point process representation that simultaneously models the timing and the magnitude of excesses over a high threshold. This model described in Section 1.1.3. Dependent and non-stationary series are discussed in Section 1.1.4.

1.1.1 Maxima

Let $X_1, \dots, X_n$ be independent and identically distributed (i.i.d.) random variables distributed according to a distribution F. In applications, the X_i usually represent the values of a process measured on a regular time-scale (e.g., daily rainfall at some fixed location). For risk assessment purposes, the focus is often on the stochastic behaviour of the largest of these n variates, denoted by $M_n = \max(X_1, \dots, X_n)$. If n is the number of observations in a year, then M_n represents the annual maximum. However, in some applications such as system failure models, where the n variates might correspond to the lifetime of the components of the system, the focus is rather on the minimum of these n variates. Results for minima can easily be obtained from those for maxima since $\min(X_1, \dots, X_n) = -\max(-X_1, \dots, -X_n)$.

1.1.1.1 Asymptotic distribution of maxima

The cumulative distribution function (CDF) of M_n is

$$\Pr(M_n \leq x) = \Pr(X_1 \leq x, \dots, X_n \leq x) = \Pr(X_1 \leq x) \times \cdots \times \Pr(X_n \leq x) = F^n(x).$$

In applications, F is unknown, so one aims at approximating F^n by some appropriate limit distribution. However, F^n converges to a degenerate distribution, putting unit mass at the upper end of the support of F, that is $M_n \xrightarrow{\text{D}} x_F$ as $n \to \infty$, where $x_F = \sup\{x : F(x) < 1\}$.

It is often possible to stabilize the stochastic behaviour of M_n with a location and scale transformation, giving rise to max-stable distributions, which are the only possible non-degenerate limiting distributions of maxima, similarly to (sum-)stable distributions which are the only possible non-degenerate limiting distributions of linearly transformed sums of random variables.

Definition 1.1 (Equality in type). Two distributions F_1 and F_1 are said to be of the same type if there exist constants $a > 0$ and $b \in \mathbb{R}$ such that $F_1(ax+b) = F_2(x)$ for all x.

Definition 1.2 (Max-stability). A distribution F is said to be max-stable if $F^k(x) = F(a_k x + b_k)$ for all $k \in \mathbb{N}$ and for some constants a_k and b_k. A random variable X is said to be max-stable if its distribution is max-stable.

Max-stability implies that taking the maximum of a collection of i.i.d. random variables leads to a variable with the same distribution up to changes in location and scale; the distribution of the maximum is of the same type as that of the initial random variables. The following result states that if a limiting distribution for rescaled maxima exists, it must be max-stable, and conversely, that all max-stable distributions are limit laws of maxima of i.i.d. random variates.

Theorem 1.1 (Limit property of max-stable laws, Embrechts *et al.* (1997, pp.121)). *The class of max-stable distributions coincides with the class of all possible (non-degenerate) limit laws for (properly normalised) maxima of i.i.d. random variables.*

The forms of the limiting distribution were first studied by Fisher and Tippett (1928), whose result was subsequently formalized and unified by von Mises (1936), Gnedenko (1943) and Jenkinson (1955). The following result is the cornerstone of extreme value theory. It states that if the maximum of i.i.d. random variables can be linearly renormalized in such a way that it converges in distribution to a non-degenerate limit, then this limiting random variable must follow the Gumbel distribution, the Fréchet distribution, or the (negative) Weibull distribution.

Theorem 1.2 (extremal types, Embrechts *et al.* (1997, pp.121)). *Let* $(X_j)_{j\geq 1}$ *be a sequence of i.i.d. random variables and let* $M_n = \max(X_1,\ldots,X_n)$. *If there exist norming constants* $a_n > 0$, $b_n \in \mathbb{R}$, *and some non-degenerate probability distribution* G *such that*

$$\Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) \to G(x),$$

as $n \to \infty$ *and for every number* $x \in \mathbb{R}$ *at which* G *is continuous, then* G *must be of the same type as one of the following three distributions:*
Type I (Gumbel):

$$\Lambda(x) = \exp\left\{(-\exp(-x)\right\}, \quad -\infty < x < \infty;$$

Type II (Fréchet):

$$\Phi_\alpha(x) = \begin{cases} 0, & x \leq 0, \\ \exp\left(-x^{-\alpha}\right), & x > 0,\ \alpha > 0; \end{cases}$$

Type III (Weibull):

$$\Psi_\alpha(x) = \begin{cases} \exp\left(-(-x)^\alpha\right), & x < 0, \alpha > 0, \\ 1, & x \geq 0. \end{cases}$$

Each of the three types of distributions presented in Theorem 1.2 may appear as a limit for the distribution of the rescaled maximum, and does so when Λ, Φ_α, or Ψ_α is itself the distribution of the X_js. These three types of limiting distributions have distinct forms of behaviour, corresponding to different types of behaviour of the tail of the CDF of the X_is. The Gumbel distribution has unbounded support and the upper tail of the density decays exponentially. The Fréchet distribution has a finite lower end of the support and the upper tail of the density decays polynomially. The Weibull distribution has a finite upper end to its support.

Definition 1.3 (Extreme value distribution)**.** The distributions Λ, Φ_α, or Ψ_α presented in Theorem 1.2 are collectively called the standard extreme value distributions. Distributions of the type of Λ, Φ_α, or Ψ_α are called extreme value distributions.

By Theorem 1.2, the extreme value distributions are precisely the max-stable distributions. It is inconvenient to have to work with three possible limiting families, but they can be unified in a single parametric family of models called the generalized extreme value distribution.

Definition 1.4 (Generalized extreme value distribution)**.** The generalized extreme value (GEV) distribution with location parameter $\eta \in \mathbb{R}$, scale parameter $\tau > 0$, and shape parameter $\xi \in \mathbb{R}$, denoted GEV(η, τ, ξ), has CDF

$$H_{\eta,\tau,\xi}(x) = \begin{cases} \exp\left[-\left\{1 + \xi\left(\frac{x-\eta}{\tau}\right)\right\}_+^{-1/\xi}\right], & \xi \neq 0, \\ \exp\left[-\exp\left\{-\left(\frac{x-\eta}{\tau}\right)\right\}\right], & \xi = 0, \end{cases} \tag{1.1}$$

where $z_+ = \max(0, z)$.

The GEV distribution with PDF h has support $\text{supp}(h) = \{x \in \mathbb{R} : 1 + \xi(x-\eta)\tau > 0\}$. For $\xi > 0$, $\xi = 0$, and $\xi < 0$, the GEV distribution respectively reduces to the heavy-tailed Fréchet, light-tailed Gumbel, and short-tailed (negative) Weibull distributions. Specifically,

$$\begin{aligned} \Phi(x;\alpha) &\equiv H_{0,1,1/\alpha}\{\alpha(x-1)\}, \\ \Lambda(x;\alpha) &\equiv H_{0,1,\alpha}(x), \\ \Psi(x;\alpha) &\equiv H_{0,1,-1/\alpha}\{\alpha(1+x)\}. \end{aligned}$$

For fixed x, we have $\lim_{\xi\to 0} H_{\eta,\tau,\xi}(x) = H_{\eta,\tau,0}(x)$, so the GEV distribution is continuous in ξ.

1.1.1.2 Domains of attraction

Theorem 1.2 identifies the three possible limiting distributions for renormalized maxima of i.i.d. random variables. The class of distributions that are attracted to a particular limiting distribution is defined as its maximum domain of attraction.

Definition 1.5 (Maximum domain of attraction)**.** A distribution F belongs to the maximum domain of attraction of the exteme value distribution H if there exist constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that the limiting distribution of $(M_n - b_n)/a_n$ as $n \to \infty$ is H, where $M_n = \max(X_1, \ldots, X_n)$, and $X_1, \ldots, X_n$ are i.i.d. with distribution F. We write $F \in \text{MDA}(H)$.

A characterization of the maximum domain of attraction and necessary and sufficient conditions for F to lie in the different maximum domains of attraction are discussed for example in Embrechts *et al.* (1997, Sec. 3.3), Beirlant *et al.* (2004, Chap. 2), and de Haan and Ferreira (2006, Sec. 1.1.5 and 1.2).

There are distributions that don't belong to any maximum domain of attraction, as the existence of the normalizing constants a_n and b_n is not guaranteed. Such distributions include the Poisson, geometric, and negative binomial distributions (see e.g., Embrechts *et al.*, 1997, pp. 118–119). However, essentially all the common continuous probability distributions of statistics are in $\text{MDA}(H_{0,1,\xi})$ for some value of ξ.

The rate of convergence of the distribution of renormalized maxima to its limiting distribution, when the latter exists, can vary, and it can depend on the particular choice of the normalizing constants a_n and b_n; see for example Embrechts *et al.* (1997, pp. 150–151) and references therein. For example, the rate of convergence is considered to be fast in the case of the exponential distribution and to be slow in the case of the normal distribution. Smith (1982) and Balkema and de Haan (1990) derived rates of convergence for the convergence in distribution of renormalised sample maxima to the appropriate extreme-value distribution.

1.1.1.3 Inference

Motivated by the extremal types theorem (Theorem 1.2), the GEV distribution is asymptotically justified as a model for extreme values of a phenomenon of interest. Its application consists in grouping the data into blocks of equal length, and fitting the GEV distribution to maxima of non-overlapping blocks of consecutive observations, yielding the so-called block maximum approach. Let $x_1, \ldots, x_N$ be a series of observations of a phenomenon of interest at regular time points, e.g., the daily rainfall at a specific location. Supposing that the data can be grouped into m blocks of n consecutive observations, that is $N = mn$, with $n, m \in \mathbb{N}$, one forms m block maxima $y_1 = \max(x_1, \ldots, x_n), \ldots, y_m = \max(x_{n(m-1)+1}, \ldots, x_N)$. The block maxima $y_1, \ldots, y_m$ are typically assumed to be i.i.d. observations from the GEV distribution whose parameters are to be estimated, although in practice observations are only approximated GEV distributed. If the observations are independent, then the block maxima are also independent. Independence of the block maxima is likely to be a reasonable approximation even if the observations come from a dependent series, provided that the block size n is sufficiently large.

Maximum likelihood inference is widely used for the GEV distribution (see e.g., Embrechts *et al.*, 1997, Sec. 6.3.1). If the block size n is sufficiently large so that the components of the vector of block maxima $\boldsymbol{y} = (y_1, \ldots, y_m)$ can be regarded as mutually independent, then the log-likelihood derived from (1.1) for $\xi \neq 0$ is

$$l(\eta, \tau, \xi; \boldsymbol{y}) = -m \log \tau - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{m} \log \left\{1 + \xi \left(\frac{y_i - \eta}{\tau}\right)\right\}_+ - \sum_{i=1}^{m} \left\{1 + \xi \left(\frac{y_i - \eta}{\tau}\right)\right\}_+^{-1/\xi}, \tag{1.2}$$

which must be maximized subject to the constraint $\tau > 0$. In practice, this expression can be maximized numerically with standard optimization routines.

The log-likelihood (1.2) yields a non-regular likelihood problem, since the support of the GEV density depends on the parameter values; the log-likelihood equals $-\infty$ if $\xi(y_i - \eta)/\tau < -1$ for any i. Smith (1985) showed that the usual properties of consistency, asymptotic efficiency and asymptotic normality of maximum likelihood estimators hold when $\xi > -1/2$. Prescott and Walden (1980) provide explicit expressions for the information matrix of the GEV distribution. The value of ξ depends on the nature of the problem. In most environmental problems one finds $\hat{\xi} \approx 0$, and in financial and insurance applications, the data are heavy-tailed and one can find $\hat{\xi} \geq 1$, so, maximum likelihood estimation is in principle well-behaved.

Standard errors for parameter estimates can be computed from the inverse of the observed information matrix, which is also obtained numerically. Confidence intervals are preferably computed using profile likelihood-based methods (see e.g., Coles, 2001, Sec. 3.3.4); they provide asymmetric intervals which reflect the right-skewness of the likelihood function often observed for the shape parameter ξ.

The choice of the block length n is crucial because it corresponds to a trade-off between bias and variance: blocks that are too small can result in a poor approximation of the limiting GEV model, leading to bias in estimates and extrapolations; blocks that are too large result in few block maxima, leading to a large estimation variance. In practice, the blocks are often pragmatically chosen to be natural periods such as years or months. So, in environmental applications $n = 365$ or $n = 30$ are common choices, whilst in finance it is often $n = 250$ or $n = 20$ as trading occurs in general only on weekdays. Diagnostics such as quantile-quantile plots or parameter stability plots can be used to determine if the block size is appropriate.

Several other techniques have been proposed for inference on the parameters of the GEV distribution. Graphical techniques such as quantile-quantile plots are useful for data exploration, model-checking and presenting conclusions. Moment-based techniques in which functions of model moments are equated with their empirical equivalents are usually inefficient for extremes as moments may not exist. Probability-weighted moments have proven to be useful because of their computational simplicity and good small-sample properties, though they are relatively difficult to extend to more complex data. See for example Hosking *et al.*

(1985), Hosking (1990), and Diebolt *et al.* (2008). Bayesian inference provides an integrated approach to modelling uncertainty and can also be applied using Markov chain Monte Carlo (MCMC) algorithms and other types of stochastic computation algorithms; Coles and Powell (1996) is an important contribution, and Coles (2001, Sec. 9.1) and Beirlant *et al.* (2004, Chap. 11) provide broad discussions. However, the Bayesian approach requires the user to specify prior knowledge for the parameters in terms of probability distributions and the tuning of hyper-parameters is often tricky in practice.

Non-parametric estimators of the shape parameter ξ have also been proposed, such as the well-known Hill estimator (Hill, 1975), the Pickands estimator (Pickands, 1975), and the moment estimator introduced by Dekkers *et al.* (1989) as a generalization of the Hill estimator. A significant effort has been devoted to studying and improving these estimators, as the shape parameter ξ is key; it determines the tail weight of the GEV distribution and the predicted sizes of future extreme events. For example, see Li *et al.* (2008) and references therein.

1.1.1.4 Generalization with the *r* largest order statistics

The block maximum approach uses a very limited amount of data (one observation per block), so that parameter estimates in general have large variances. An extension to the block maximum model which doesn't use only the block maxima is the r largest order statistics model developed by Smith (1986) for the case $\xi = 0$, and by Tawn (1988b) for the general case, which consists in fitting the joint distribution of the r largest order statistics to the r largest observations in each fixed time period (Coles, 2001, Sec. 3.5). The precision of parameter estimates should be increased due to the inclusion of extra information, but the result is more vulnerable to departures from the i.i.d. assumption. The r-largest order model is a special case of the point process representation (see Section 1.1.3).

1.1.2 Threshold excesses

The block maximum approach discussed in Section 1.1.1 only uses the largest observation in each block; it can be inefficient when more than fixed time period maxima alone are available. It seems more efficient to use all extreme observations, those that exceed some high threshold, and to avoid blocking. For this reason, the block maximum approach has been largely superseded in practice by methods based on threshold excesses (Davison and Smith, 1990), where all the observations exceeding a high threshold are used.

As in Section 1.1.1, let $X_1, \dots, X_n$ be independent and identically distributed (i.i.d.) random variables distributed according to a distribution F. Here, the focus is on the stochastic behaviour of the random N_u observations that exceed a high threshold u; denote these observations by $\tilde{X}_1, \dots, \tilde{X}_{N_u}$. For each of these exceedances, let $Y_i = \tilde{X}_i - u$ be the sizes of the excesses.

1.1.2.1 Distribution of threshold excesses

The generalized Pareto distribution was used by Pickands (1975) to describe the distribution of the excesses over high thresholds.

Definition 1.6 (Generalized Pareto distribution)**.** The generalized Pareto (GP) distribution with scale parameter $\beta > 0$ and shape parameter $\xi \in \mathbb{R}$, denoted GP(β, ξ), has CDF

$$G_{\beta,\xi}(x) = \begin{cases} 1 - (1 + \xi x/\beta)_+^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp(-x/\beta)_+, & \xi = 0, \end{cases} \tag{1.3}$$

where $z_+ = \max(0, z)$.

The GP distribution has upper endpoint $x_F = -\beta/\xi$ if $\xi < 0$, and otherwise $x_F = \infty$. For $\xi > 0$, $\xi = 0$, and $\xi < 0$, the GP distribution respectively reduces to the ordinary Pareto, exponential, and Pareto type II distributions. For fixed x, we have $\lim_{\xi \to 0} G_{\beta,\xi}(x) = G_{\beta,0}(x)$, so the GP distribution is continuous in ξ, like the GEV distribution. The GP distribution is in the max domain of attraction of the GEV distribution with the same shape parameter ξ, that is $G_{\beta,\xi} \in$ MDA$(H_{0,1,\xi})$. Similarly to the max-stability of the GEV distribution, the GP distribution is threshold-stable, that is, if $X \sim$ GP(β, ξ) and $0 < v < x_F$, then $X - v \mid X > v \sim$ GP(β_v, ξ) with $\beta_v = \beta + \xi v$ and the same shape parameter ξ. The mean of the GP distribution exists provided $\xi < 1$ and is $\mathrm{E}(X) = \beta/(1 - \xi)$.

The excess distribution over a high threshold, and the corresponding mean excess function play an important role both in theory and applications.

Definition 1.7 (Excess distribution over a threshold)**.** Let X be a random variable with CDF F. The excess distribution over the threshold u has CDF

$$F_u(x) = \Pr(X - u \le x | X > u) = \frac{F(x+u) - F(u)}{1 - F(u)}, \quad 0 \le x < x_F - u,$$

where x_F denotes the right endpoint of the support of F.

Definition 1.8 (Mean excess function)**.** The mean excess function of a random variable with finite mean is

$$e(u) = \mathrm{E}(X - u | X > u).$$

The distribution of excesses F_u describes the distribution of the excess over the threshold u, given that the threshold u is exceeded. The mean excess function $e(u)$, also known as the mean residual life function, gives the mean of F_u as a function of u. If the random variable X has CDF $F = G_{\beta,\xi}$, then its excess distribution is

$$F_u(x) = G_{\beta(u),\xi}(x), \quad \beta(u) = \beta + \xi u,$$

defined for $0 \leq x < \infty$ if $\xi \geq 0$, and $0 \leq x < -\beta/\xi - u$ if $\xi < 0$. The excess distribution of a GP distribution remains GP with the same shape parameter ξ, and a scale parameter $\beta(u)$ that grows or shrinks linearly with the threshold u. The mean excess function of the GP distribution is

$$e(u) = \frac{\beta + \xi u}{1 - \xi},$$

defined for $0 \leq u < \infty$ if $0 \leq \xi < 1$, and $0 \leq u \leq -\beta/\xi$ if $\xi < 0$. The mean excess function is not defined if $\xi \geq 1$, as the GP distribution doesn't have a finite mean in this case. The function $e(u)$ is linear in the threshold u, a property of the GP distribution that is used in applications to help determining an appropriate threshold for inference.

The following result states that the set of distributions for which the excess distribution converges to the GP distribution as the threshold u is raised is equivalent to the set of distributions for which the rescaled maxima converge to a GEV distribution.

Theorem 1.3 (Pickands–Balkema–de Haan, Pickands (1975), Balkema and de Haan (1974))**.** *Let F be a CDF with right endpoint x_F. There exists a positive, measurable function $\beta(u)$ such that*

$$\lim_{u \to x_F} \sup_{0 \leq x < x_F - u} |F_u(x) - G_{\beta(u),\xi}(x)| = 0$$

if and only if $F \in \mathrm{MDA}(H_{\eta,\tau,\xi})$ for $\xi \in \mathbb{R}$.

Remarkably, the shape parameter of the limiting GP distribution for excesses is equal to the shape parameter of the GEV limiting distributions for the maxima. Theorem 1.3 is widely applicable and says that the GP distribution is the canonical distribution for modelling excesses over a high threshold, since essentially all common continuous probability distributions lie in $\mathrm{MDA}(H_{0,1,\xi})$ for some value of ξ.

1.1.2.2 Inference

Motivated by the Pickands–Balkema–de Haan theorem (Theorem 1.3), the GP distribution is asymptotically justified to model the excesses over a high threshold for a phenomenon of interest. Its application consists in selecting the observations that are above a high threshold, and fitting the GP distribution to the amount of excesses for these observations, yielding the so-called peaks over threshold approach. Let $x_1, \ldots, x_n$ be a series of observations of a phenomenon of interest at regular time points, and denote the N_u observations that exceed a high threshold u by $\tilde{x}_1, \ldots, \tilde{x}_{N_u}$. In the peaks over threshold approach one fits the GP distribution to the N_u excess amounts $y_1 = \tilde{x}_1 - u, \ldots, y_{N_u} = \tilde{x}_{N_u} - u$.

The excesses $y_1, \ldots, y_{N_u}$ are assumed to be i.i.d. observations from the GP distribution whose parameters are to be estimated, although in practice the observations are only approximately distributed as a GP distribution. Statistical properties of the threshold approach were investigated in Davison and Smith (1990). Maximum likelihood estimation is widely used for inference using the GP distribution (see e.g., Embrechts *et al.*, 1997, Sec. 6.5.1).

The log-likelihood function for a sample $\boldsymbol{y} = (y_1, \dots, y_{N_u})$ of i.i.d. random variables with GP distribution derived from (1.3) when $\xi \neq 0$ is

$$l(\beta, \xi; \boldsymbol{y}) = -N_u \log \beta - \left(1 + \frac{1}{\xi}\right) + \sum_{i=1}^{N_u} \log \left(1 + \xi \frac{y_i}{\beta}\right)_+,$$

which must be maximized subject to the constraint $\beta > 0$. In the case $\xi = 0$, the log-likelihood derived from (1.3) is

$$l(\beta; \boldsymbol{y}) = -N_u \log \beta - \frac{1}{\beta} \sum_{i=1}^{N_u} y_i,$$

In practice, these expressions can be maximized numerically with standard optimization routines. The GP distribution is regular for likelihood inference under the same conditions on ξ as for the GEV distribution.

Other approaches have been proposed for inference on the parameters of the GP distribution. Hosking and Wallis (1987) introduced the method of moments and the method of probability-weighted moments estimators for the GP distribution. The elemental percentile method introduced by Castillo and Hadi (1997) overcomes some of the difficulties associated with maximum likelihood and probability-weighted moment estimation.

A key element in inference is the selection of the threshold u, which entails a trade-off between bias and variance, similarly to the choice of block size in the block maximum approach. Taking u too low introduces bias in estimates and predictions because the approximation of the distribution of excesses with the GP distribution will be poor. Conversely, taking u too high increases the variance of estimates and predictions because too few observations are used.

Threshold selection is commonly based on graphical procedures that exploit the threshold-stability property of the GP distribution. The standard practice is to select the lowest possible threshold, provided that the limiting model provides a reasonable approximation. Two methods are available for this: one is an exploratory technique carried out prior to estimation based on the mean residual life plot introduced by Davison and Smith (1990); the other requires to fit the GP distribution for a grid of possible thresholds and to look for stability of parameter estimates (e.g., see Coles, 2001, Sec. 4.3.4). However, these graphical procedures are subjective and cannot be automated. Lots of effort has been devoted to finding good robust threshold estimators; see Scarrott and MacDonald (2012) and Caeiro and Gomes (2015) for reviews of some historical threshold estimation approaches and recent developments.

1.1.3 Point process representation

The threshold excesses model discussed in Section 1.1.2 considers only the magnitude of threshold excesses. In the point process approach, the times at which high-threshold exceedances occur and the excess values over the threshold are combined into one process on a two-dimensional plot which, under suitable normalisation, behaves like a non-homogeneous Poisson process. The block maximum model, r-largest order statistics model discussed in Section 1.1.1, and threshold excesses model are all special cases of the point process representation introduced by Pickands (1971), and illustrated in statistical applications by Smith (1989).

The following result, which is stated as given in the reference cited, enables the modelling of extremes with the point process framework.

Theorem 1.4 (Coles (2001, Sec. 7.5)). *Let $X_1, \ldots, X_n$ be a series of i.i.d.random variables. Then, on regions of the form $(0,1) \times [u, \infty)$ and for sufficiently large u, the point process*

$$N_n = \left\{ \left(\frac{i}{n+1}, X_i \right) : i = 1, \ldots, n \right\}$$

is approximately a Poisson process with intensity measure

$$\Lambda \{[t_1, t_2] \times (x, \infty)\} = n_y (t_2 - t_1) \left(1 + \xi \frac{x - \eta}{\tau} \right)_+^{-1/\xi}, \quad 0 \leq t_1 < t_2 \leq 1, x \geq u,$$

where n_y is the number of years of observations, and $z_+ = \max(0, z)$.

The term n_y in the expression for Λ is an adjustment that allows one to express extreme value limits in terms of approximate distributions of annual maxima. The parameters (η, τ, ξ) are the same, and have the same interpretations, as for the GEV distribution for annual maxima.

.

For a region of the form $\mathcal{A}_v = [0,1] \times [v, \infty)$ for $v > u$, and containing the points $x_1, \ldots, x_{N_v}$, the likelihood is

$$L(\eta, \tau, \xi; x_1, \ldots, x_{N_v}) \propto \tau^{-N_v} \exp \left\{ -n_y \left(1 + \xi \frac{v - \eta}{\tau} \right)_+^{-1/\xi} \right\} \prod_{i=1}^{N_v} \left(1 + \xi \frac{v - \eta}{\tau} \right)_+^{-1/\xi - 1},$$

which is to be maximized with the constraint $\tau > 0$. Standard errors and approximate confidence intervals can be obtained in the usual way. Refer to Coles (2001, Chap. 7) for a broader overview, and for example Embrechts *et al.* (1997, Chap. 5) for a detailed, mathematically complete, treatment of the point process representation.

1.1.4 Dependent and non-stationary series

The block maximum approach, the threshold excesses and the point process representation discussed above require the sequence of random variables to be i.i.d. However, temporal dependence or non-stationarity often arise in practice, possibly due to autocorrelation, seasonality, long-term trend, regime changes and dependence to external factors. The modeling of extreme events for dependant series and non-stationary series is briefly discussed in this section. Leadbetter *et al.* (1983) is a standard reference to the literature on extremes of time series. Other references include Beirlant *et al.* (2004, Chap. 10) and Coles (2001, Chap. 5–6). Chavez-Demoulin and Davison (2012) and Reich and Shaby (2016) provide more recent reviews.

1.1.4.1 Non-stationarity

Non-stationary series have characteristics that change systematically through time. Standard parametric and nonparemetric modelling techniques can be used to deal with non-stationarity of real data. The typical approach is to model the extreme value parameters using for example linear regression (e.g., Smith, 1989; Katz *et al.*, 2002), semi-parametric models based on local likelihood (e.g., Ramesh and Davison, 2002) or splines (e.g., Chavez-Demoulin and Davison, 2005; Padoan and Wand, 2008). Alternative approaches include to fit a non-stationary extremal model to the original data (Maraun *et al.*, 2009), or to use the full dataset to detect and estimate non-stationarities, and then to apply methods for stationary extreme-value modelling to the resulting residuals. Examples of the latter include Eastoe and Tawn (2009) who apply the Box–Cox transformation to the original time series where the power transformation depends linearly on covariates and apply a fixed threshold to the residual sa covariate dependent linear model, and McNeil and Frey (2000) who filter the returns of financial series with an AR-GARCH model and use the GPD to fit the tails of the filtered series. Chavez-Demoulin and Davison (2012, Sec. 4) provide an extensive review of methods dealing with non-stationarity.

1.1.4.2 Dependent series

Extreme value theory for dependent stochastic processes has been extensively developed (e.g., Leadbetter *et al.*, 1983). The most natural generalization of a sequence of independent random variables is to strictly stationary series, which correspond to series having stochastic properties that are homogeneous through time. A time series is strictly stationary if the joint distribution of every collection of values $\{X_{t_1}, \ldots, X_{t_k}\}$ is identical to that of the time-shifted set $\{X_{t_1+h}, \ldots, X_{t_k+h}\}$ (e.g., Shumway and Stoffer, 2011, Sec 1.5).

With applications in mind, it is common to consider time series with limited long-range dependence at extreme levels. The $D(u_n)$ condition defines the notion of near-independence of extreme events that are sufficiently distant in time.

Definition 1.9 ($D(u_n)$ condition)**.** A strictly stationary sequence $\{X_i\}$ having marginal distribution F with upper endpoint $x_F = \sup\{x : F(x) < 1\}$, is said to satisfy the $D(u_n)$ condition if, for any integers $i_1 < \cdots < i_p < j_1 < \cdots < j_q$ with $j_1 - i_p > l$,

$$\begin{aligned}\big|\Pr\big(X_{i_1} \le u_n, \ldots, X_{i_p} \le u_n, X_{j_1} \le u_n, \ldots, X_{j_q} \le u_n\big) \\ - \Pr\big(X_{i_1} \le u_n, \ldots, X_{i_p} \le u_n\big)\Pr\big(X_{j_1} \le u_n, \ldots, X_{j_q} \le u_n\big)\big| \le \alpha(n, l),\end{aligned}$$

where $\alpha(n, l_n) \to 0$ for some sequence $l_n = o(n)$, and $u_n \to x_F$ as $n \to \infty$.

The following result shows that if the $D(u_n)$ condition is satisfied, then the GEV distribution is the limiting distribution of maxima of dependent data, thereby justifying the use of the block maximum approach for most stationary time series.

Theorem 1.5 (Leadbetter (1974))**.** *Let $\{X_i\}$ be a strictly stationary process for which there exists sequences of normalizing constants $\{a_n > 0\}$ and $\{b_n\}$ and a non-degenerate distribution H such that $M_n = \max(X_1, \ldots, X_n)$ satisfies*

$$\Pr\left(\frac{M_n - b_n}{a_n}\right) \to H(z) \quad n \to \infty.$$

If the $D(u_n)$ condition holds with $u_n = a_n z + b_n$ for each z for which $H(z) > 0$, then H is a GEV distribution.

This remarkable result shows that maxima of stationary series satisfying the $D(u_n)$ condition also have a GEV distribution. However, the parameters of the limiting distribution are affected by the dependence in the series. The relation between the maxima of a dependent sequence and those of a corresponding independent sequence is summarised in the following result.

Theorem 1.6 (Leadbetter (1983))**.** *Let $\{X_i\}$ be a stationary process and let $\{X_i^*\}$s be a sequence of independent random variables with the same marginal distribution. Set $M_n = \max(X_1, \ldots, X_n)$ and $\tilde{M}_n = \max(\tilde{X}_1, \ldots, \tilde{X}_n)$. Under suitable regularity conditions,*

$$\Pr\left(\frac{\tilde{X}_n - b_n}{a_n} \le z\right) \to \tilde{H}(z), \quad n \to \infty,$$

for sequences of normalizing constants $\{a_n > 0\}$ and $\{b_n\}$, where $\tilde{H}$ is a non-degenerate distribution function, if and only if

$$\Pr\left(\frac{X_n - b_n}{a_n} \le z\right) \to H(z), \quad n \to \infty,$$

where $H(z) = \tilde{H}^{\theta}$ for some constant $\theta \in [0, 1]$.

Since the extremal types theorem (Theorem 1.2) implies that the only possible non-degenerate limit $\tilde{H}$ is the GEV distribution, then H is also a GEV distribution, except in the pathological case $\theta = 0$. If H is a $GEV(\eta, \tau, \xi)$ distribution, then $\tilde{H}$ is a $GEV\left(\tilde{\eta}, \tilde{\tau}, \tilde{\xi}\right)$ distribution with

$$\tilde{\eta} = \eta - \frac{\tau}{\xi}\left(1 - \theta^{-\xi}\right), \quad \tilde{\tau} = \theta^{\xi}\tau, \quad \tilde{\xi} = \xi,$$

so the shape parameter is not affected by the temporal dependence. So, M_n is stochastically smaller than $\tilde{M}_n$, implying that serial dependence tends to reduce the sizes of extremes. The constant θ is called the extremal index and determines by how much M_n is stochastically larger than $\tilde{M}_n$. Another consequence of serial dependence is that extremes tend to occur in clusters. The extremal index is linked to the size of clusters as θ^{-1} is the limiting mean cluster size of clusters of exceedances of increasingly high thresholds, and also to the probability that an exceedance over a high threshold is the last exceedance of a cluster.

Several estimators of the extremal index θ have been proposed. The classical block and runs estimators (Beirlant *et al.*, 2004, Sec. 10.3.4) are empirical counterparts of the mean cluster size and the probability of last exceedence of a block interpretations of the extremal index. Other estimators include the two-threshold estimator proposed by Laurini and Tawn (2003), and the intervals estimator proposed by Süveges (2007).

A common approach to circumvent the difficulties caused by temporal dependence is to use a declustering scheme to filter out a set of approximately independent threshold excesses. A less wasteful approach which uses all threshold excesses was proposed by Fawcett and Walshaw (2007, 2012). Within-cluster modelling can be done with first-order Markov chains using multivariate extreme-value dependence structures (e.g., Smith *et al.*, 1997; Bortot and Coles, 2003); flexible models for higher order chains are currently lacking.

1.2 Multivariate extremes

Many applied problems are essentially multivariate. However, in two or more dimensions it is less obvious than in the univariate case what 'extreme' means; depending on the application, one might want to consider an event as extreme if at least one of the component is large, or if all components are large, or if a function of the components is large, just to mention a few possibilities. In addition, different types of dependence between the components may arise at extreme levels, which requires flexible tools to quantify and model extremal dependence for suitable assessment and extrapolation of the risks of the phenomenon of interest.

The componentwise maximum approach discussed in Section 1.2.1 is the analogue of the block maximum approach in the univariate case. Section 1.2.2 describes a point process approach in the multivariate case. The concept of asymptotic independence is discussed in Section 1.2.3, and some measures of extremal dependence are presented in Section 1.2.4.

1.2.1 Componentwise maximum approach

1.2.1.1 Multivariate extreme value distributions and max-stability

Let $\boldsymbol{X} = (X_1, \dots, X_d)$ be a d-dimensional random vector with joint CDF F and marginal CDFs $F_1, \dots, F_d$. Let $\{\boldsymbol{X}_i\}_{i \geq 1}$ with $\boldsymbol{X}_i = (X_{i,1}, \dots, X_{i,d})$ be an i.i.d. sequence of random vectors distributed as $\boldsymbol{X}$, and let $\boldsymbol{M}_n = (M_{n,1}, \dots, M_{n,d})$ with $M_{n,i} = \max(X_{1,i}, \dots, X_{n,i})$ be the vector of componentwise maxima of $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$. The vector $\boldsymbol{M}_n$ does not always correspond to an element of the sequence $\{\boldsymbol{X}_i\}$ as the component maxima $M_{n,1}, \dots, M_{n,d}$ may correspond to different elements of the sequence. The aim is to characterize the family of possible asymptotic distribution for $\boldsymbol{M}_n$ as $n \to \infty$, after suitable renormalization.

Definition 1.10 (multivariate extreme value distribution)**.** If there exists sequences $\{\boldsymbol{a}_n\}_{n \geq 1}$ with $\boldsymbol{a}_n = (a_{n,1}, \dots, a_{n,d}) > \boldsymbol{0}$, and $\{\boldsymbol{b}_n\}_{n \geq 1}$ with $\boldsymbol{b}_n = (b_{n,1}, \dots, b_{n,d})$ such that

$$\Pr\left(\frac{\boldsymbol{M}_n - \boldsymbol{b}_n}{\boldsymbol{a}_n} \leq \boldsymbol{x} \right) \to G(\boldsymbol{x}), \quad n \to \infty,$$

where G is a d-dimensional CDF with non-degenerate margins, then G is a multivariate extreme value (MEV) distribution.

The characterisation of MEV distributions reduces to that of multivariate max-stable distributions with non-degenerate margins. From univariate extreme value theory and the extremal types theorem (Theorem 1.2), we know that the margins G_i must be GEV if they are non-degenerate, that is $(M_{n,i} - b_{n,i})/a_{n,i} \to Z_i \sim \mathrm{GEV}(\eta_i, \tau_i, \xi_i)$ as $n \to \infty$ for $j = 1, \dots, d$. So, the margins are max-stable. Max-stability in the multivariate setting is defined similarly to in the univariate case.

Definition 1.11 (Max-stability)**.** A d-dimensional distribution F is said to be max-stable if $F^k(\boldsymbol{z}) = F(\boldsymbol{a}_k \boldsymbol{z} + \boldsymbol{b}_k)$ for all $k \in \mathbb{N}$ and for some sequences constants $\{\boldsymbol{a}_k\} \in \mathbb{R}^d_+$ and $\{\boldsymbol{b}_k\} \in \mathbb{R}^d$. A random variable X is said to be max-stable if its distribution is max-stable.

Multivariate max-stability requires the max-stability of both the margins and the dependence structure. A common way to separate the dependence structure from the marginal distribution is using copulas.

1.2.1.2 Copulas

Copulas are broadly used to model the dependence structure in a multivariate setting; see Nelsen (2006) or Joe (2015) for an introduction. Copulas have gained a lot of attention in the extreme-value community; see for example Capéraà *et al.* (1997, 2000) and Coles *et al.* (1999) for early contributions, Heffernan (2000) for extremal properties of a large collection of copulas, Demarta and McNeil (2005) who introduce the skew-t copula, and Mikosch (2006) for a critical discussion of the use of copulas in extreme value modelling. More recent contributions include a detailed discussion of extreme-value copulas by Gudendorf and Segers (2010) and by Davison *et al.* (2012) for copula modelling of spatial extemes.

Definition 1.12 (Copula)**.** A d-dimensional copula is a distribution on $[0,1]^d$ with standard uniform margins.

Theorem 1.7 (Sklar (1959))**.** *Let F be the CDF of a joint distribution in $\mathbb{R}^d$ with margins $F_1,\ldots,F_d$. Then there exists a copula C such that for all $\boldsymbol{x} = (x_1,\ldots,x_d) \in \mathbb{R}^d$,*

$$F(\boldsymbol{x}) = C\{F_1(x_1),\ldots,F_d(x_d)\}. \tag{1.4}$$

If the margins are continuous, then C is unique. Conversely, if C is a copula and $F_1,\ldots,F_d$ are univariate CDFs, then the function F defined in (1.4) is a joint distribution function with margins $F_1,\ldots,F_d$.

The copula underlying some joint distribution F may be written as

$$C(\boldsymbol{u}) = F\left\{F_1^{-1}(u_1),\ldots,F_d^{-1}(u_d)\right\},$$

where $\boldsymbol{u} = (u_1,\ldots,u_d) \in [0,1]^d$, and F_i^{-1} denotes the (generalized) inverse of the margin F_i, that is $F_i^{-1}(u_i) = \inf\{x : F_i(x) \geq u_i\}$.

There exists an extensive catalogue of copulas, including the Archimedean copulas.

Definition 1.13 (Archimedean copula generator)**.** A decreasing, continuous, convex function $\psi : [0,\infty) \to [0,1]$ satisfying $\psi(0) = 1$ and $\lim_{t\to\infty}\psi(t) = 0$ is called an Archimedean copula generator.

Definition 1.14 (Archimedean copula)**.** A d-variate copula C is Archimedean if there exists a function $\psi : [0,\infty) \to [0,1]$ such that

$$C(\boldsymbol{u}) = \psi\left\{\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d)\right\}. \tag{1.5}$$

In the bivariate case, the construction (1.5) using Archimedean copula generators yields valid (Archimedean) copulas, but in higher dimensions the Archimedean copula generator must be completely monotone to yield a valid copula (Nelsen, 2006, Theorem 4.6.2).

An example of bivariate Archimedean copula is the Gumbel copula with parameter $\theta \geq 1$, which has generator

$$\psi_{\mathrm{G}}(t;\theta) = \exp\left(-t^{1/\theta}\right), \quad t \geq 1,$$

yielding copula

$$C_{\mathrm{G}}(u,v;\theta) = \exp\left[-\left\{(-\log u)^\theta + (-\log v)^\theta\right\}^{1/\theta}\right], \quad 0 \leq u,v \leq 1.$$

Another large class is the class of so-called extreme-value copulas, which result from max-stable distributions. The max-stability of a d-variate distribution F imples that its underlying copula C_F satisfies $C_F(u_1^{1/n}, \dots, u_d^{1/n}) = C_F(u_1, \dots, u_d)^{1/n}$ for all n.

Definition 1.15 (Extreme-value copula)**.** A d-variate copula C is an extreme-value copula if it satisfies

$$C(u_1^t, \dots, u_d^t) = C(u_1, \dots, u_d)^t, \quad 0 \le u_1, \dots, u_d \le 1, t > 0.$$

Genest and Rivest (1989) showed that the max-stable attractor of an Archimedean copula is also an Archimedean copula and that the Gumbel copula is the only max-stable copula that is Archimedean, so the only possible attractor of an Archimedean copula is the Gumbel copula. The following result gives a necessary and sufficient condition for a Gumbel copula to be the max-stable attractor of an Archimedean copula.

Theorem 1.8 (Max-domain of attraction of Archimedean copulas, Genest and Rivest (1989))**.** *An Archimedean copula with generator ψ and inverse generator $\phi = \psi^{-1}$ is in the max-domain of attraction of the Gumbel copula with parameter $\theta \ge 1$ if and only if*

$$\lim_{t \uparrow 1} \frac{\mathrm{d}}{\mathrm{d}t} \frac{\phi(t)}{\phi'(t)} = \theta^{-1} > 0. \tag{1.6}$$

Construction principles

1.2.1.3 Marginal standardization

A common approach to modelling multivariate extremes is to proceed in two stages: first, one estimates the margins, and second, one handles the multivariate dependence of the data transformed to a common scale. The marginal distributions are typically estimated using univariate extreme value methodology by fitting the GEV distribution to block maxima or the GP distribution to threshold excedances. Then, using the probability integral transform, the data are converted to a common scale. Standard uniform margins seem a natural choice given the extensive literature on copulas. However, due to a mathematically more elegant treatment of extreme value theory, the data are often transformed to have unit Fréchet margins. The choice of marginal distribution essentially makes no difference to inference results. Specifically, if the random vector $\boldsymbol{X} = (X_1, \dots, X_d)$ has joint distribution F and margins F_i, then the transformed random vector $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \dots, \tilde{X}_d)$, where $\tilde{X}_i = t_i(X_i)$ with the maps $t_i(\cdot) = -1/\log F_i(\cdot)$, has distribution $\tilde{F}$ with the same dependence structure as F, that is $G(z_1, \dots, z_d) = \tilde{G}\{t_1(z_1), \dots, t_d(z_d)\}$, and margins $\tilde{F}_i$ which are unit Fréchet, that is, $\tilde{F}_i(z) = e^{-1/z}$, $z > 0$.

1.2.1.4 Spectral representation

Theorem 1.9 (Characterization of multivariate extreme value distributions, Pickands (1981)). *If the vector of renormalized maxima is such that $(\boldsymbol{M}_n - \boldsymbol{b}_n)/\boldsymbol{a}_n \xrightarrow{D} \boldsymbol{Z} \sim G$, where G is a non-degenerate distribution, then*

$$G(\boldsymbol{z}) = \exp\{-V(\boldsymbol{z})\}, \quad \boldsymbol{z} > \boldsymbol{0}, \tag{1.7}$$

where

$$V(\boldsymbol{z}) = d \int_{S_d} \max\left(\frac{\boldsymbol{w}}{\boldsymbol{z}}\right) \mathrm{d}H(\boldsymbol{w}), \tag{1.8}$$

and H is a probability measure on the $(d-1)$-dimensional simplex $S_d = \left\{\boldsymbol{w} \in \mathbb{R}_+^d : \sum_{i=1}^d w_i = 1\right\}$, satisfying the mean constraints $\int_{S_d} w_i \,\mathrm{d}H(w_i) = 1/d$, $i = 1, \ldots, d$.

The function V, which we call the exponent function, is homogeneous of order -1, that is $V(k\boldsymbol{z}) = k^{-1}V(\boldsymbol{z})$ for $k > 0$, $\boldsymbol{z} > \boldsymbol{0}$. We call the measure H the spectral distribution since it integrates to one, and to differentiate it from the spectral measure used by some authors and which has total mass d. The term spectral is widely used in mathematics for many different things, and while calling H the spectral measure was unfortunate in the first place, it's too embedded in the literature to be changed now. If H is differentiable, then $\mathrm{d}H(\boldsymbol{w}) = h(\boldsymbol{w})\,\mathrm{d}\boldsymbol{w}$, and the function h is called the spectral density. When $d = 2$, H is a probability distribution on the interval $[0, 1]$, subject to the mean constraint $\int_0^1 w \,\mathrm{d}H(w) = 1/2$.

Equation (1.8) expresses the exponent function V in terms of the spectral distribution H. Coles and Tawn (1991) found a way to compute the spectral densities h from the partial derivatives of the exponent function V. In the bivariate case, the spectral density, if it exists, is

$$h(w) = -\frac{1}{2} \left.\frac{\partial^2 V(x, y)}{\partial x \partial y}\right|_{x=w, y=1-w}, \quad 0 < w < 1, \tag{1.9}$$

and the point masses of H, if any, are

$$H(\{0\}) = -\frac{y^2}{2} \lim_{x \to 0} \frac{\partial}{\partial y} V(x, y), \quad H(\{1\}) = -\frac{x^2}{2} \lim_{y \to 0} \frac{\partial}{\partial x} V(x, y). \tag{1.10}$$

Two important special cases are independence and perfect dependence. When the spectral distribution H puts masses $1/d$ on each of the d vertex $\boldsymbol{e}_i$ of the $(d-1)$-dimensional simplex S_d, then

$$G(\boldsymbol{z}) = \exp\left\{-\left(z_1^{-1} + \cdots + z_d^{-2}\right)\right\}, \quad \boldsymbol{z} > \boldsymbol{0},$$

which is the CDF d independent unit Fréchet variables. When the spectral distribution puts mass 1 on $\boldsymbol{d}^{-1}$, then

$$G(\boldsymbol{z}) = \exp\left\{-\max\left(\boldsymbol{z}^{-1}\right)\right\}, \quad \boldsymbol{z} > \boldsymbol{0},$$

which is the CDF of variables that are marginally unit Fréchet, but which are perfectly dependent, that is $Z_1 = \cdots = Z_d$ with probability one.

Unlike the univariate case, where the extremal types theorem (Theorem 1.2) imples that one parametric family covers all possible limiting distributions, the class of MEV distributions cannot be fully described by a finite number of parameters, since any spectral distribution satisfying the mean constraint yields a valid MEV distribution. So, when it comes to modelling and inference, one must rely on flexible parametric models or nonparametric techniques.

1.2.1.5 Parametric models

The specification of a parametric model for the limiting distribution G in (1.7) can be achieved for example by specifying a parametric form for the exponent function V in (1.8), or equivalently the spectral distribution H. A parametric model for G restricts the dependence to a particular structure, so it is important to build dependence models that are flexible but also parsimonious and interpretable.

Several parametric families, mainly bivariate ones, have been proposed; for a review see Kotz and Nadarajah (2000, Sec. 3.4–3.5) or Beirlant *et al.* (2004, Sec. 9.2.2). Flexible models in large dimensions are an area of research and new families are still being constructed and discussed, as in Stephenson (2009), Cooley *et al.* (2010), Padoan (2011), Ballani and Schlather (2011), and Segers (2012).

We briefly discuss the four classical bivariate parametric extreme value distributions used in the simulation study in Chapter 2, namely the logistic, Hüsler–Reiss, Coles–Tawn, and asymmetric logistic models. The logistic model is a special case of the asymmetric logistic model but we treat it separately, as the former is the only case of the latter which doesn't have point masses at the edge of its support.

Logistic model The oldest parametric family of bivariate extreme value distribution is the logistic model introduced by Gumbel (1960b). Its exponent function is

$$V(x, y) = \left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha}, \quad x, y > 0,$$

where the parameter $0 < \alpha \leq 1$ measures the strength of dependence between the two coordinates. The limiting case $\alpha = 1$ corresponds to independence, whilst the case $\alpha \to 0$ corresponds to complete dependence. Ledford and Tawn (1998) showed that in a random sample from this model, the probability that the maxima of the two coordinates occur for the same pair of observations converges to $1 - \alpha$ as the sample size tends to infinity. It is easy to see that the copula of the logistic model is the Gumbel copula with parameter $\theta = 1/\alpha$.

From equations (1.9) and (1.10) one can see that the spectral distribution has no point masses at 0 or 1, and when $\alpha \in (0,1)$, its spectral density is

$$h(w) = \frac{1-\alpha}{2\alpha}\{w(1-w)\}^{1/\alpha-2}\left\{(1-w)^{1/\alpha}+w^{1/\alpha}\right\}^{\alpha-2}, \quad 0<w<1.$$

The logistic model can be extended to higher dimensions in a straightforward way (e.g., see Gumbel, 1960a). This simple model is popular but lacks flexibility, as it is symmetric in its variables.

Asymmetric logistic model The asymmetric logistic model introduced by Tawn (1988a) is an extension of the logistic model. Its exponent function is

$$V(x,y) = \frac{1-\psi_1}{x} + \frac{1-\psi_2}{y} + \left\{\left(\frac{\psi_1}{x}\right)^{1/\alpha} + \left(\frac{\psi_2}{y}\right)^{1/\alpha}\right\}^{\alpha}, \quad x,y>0,$$

with dependence parameter $0<\alpha\le 1$ and asymmetry parameters $0 \le \psi_i \le 1$ for $j = 1,2$. When $\psi_1 = \psi_2$ this model is symmetric and corresponds to a mixture of independence and the logistic model, which arises for $\psi_1 = \psi_2 = 1$ (see Section 1.2.1.5). The cases $\psi_1 = \psi_2 = 0$ or $\alpha = 1$ correspond to independence. The strength of the dependence is determined by the dependence parameter α and the asymmetry parameters ψ_1, ψ_2, since the extremal coefficient, which is a measure of extremal dependence (Section 1.2.4), is $\theta_2 = 2-\psi_1-\psi_2+\left(\psi_1^{1/\alpha}+\psi_2^{1/\alpha}\right)^{\alpha}$.

From equation (1.10) one can see that the spectral distribution H has point masses $H(\{0\}) = (1-\psi_2)/2$ and $H(\{1\}) = (1-\psi_1)/2$, so the case $\psi_1 = \psi_2 = 1$ corresponding to the logistic model is the only case where the spectral distribution has no point masses. When $\alpha \in (0,1)$, by equation (1.9) the spectral density is

$$h(w) = \frac{1-\alpha}{2\alpha}\left(\psi_1\psi_2\right)^{1/\alpha}\{w(1-w)\}^{1/\alpha-2}\left[\{\psi_1(1-w)\}^{1/\alpha}+(\psi_2 w)^{1/\alpha}\right]^{\alpha-2}, \quad 0<w<1.$$

Hüsler–Reiss model The Hüsler–Reiss model is a symmetric model based on the normal distribution and introduced by Hüsler and Reiss (1989). Its exponent function is

$$V(x,y) = x^{-1}\Phi\left\{\frac{\alpha}{2}+\alpha^{-1}\log\left(\frac{y}{x}\right)\right\} + y^{-1}\Phi\left\{\frac{\alpha}{2}+\alpha^{-1}\log\left(\frac{x}{y}\right)\right\}, \quad x,y>0,$$

with dependence parameter $\alpha>0$, and where Φ denotes the standard normal CDF. Independence is obtained in the limit as $\alpha \to \infty$. Complete dependence is obtained as $\alpha \to 0$.

From equations (1.9) and (1.10) one can see that the spectral distribution has no point masses, and the spectral density is

$$h(w) = \frac{e^{-\alpha/8}}{2\alpha\{w(1-w)\}^{3/2}}\phi\left\{a^{-1}\log\left(\frac{w}{1-w}\right)\right\}, \quad 0 < w < 1,$$

where ϕ denotes the standard normal PDF.

Dirichlet model The Dirichlet model introduced by Coles and Tawn (1991) allows for asymmetry. Its exponent function is

$$V(x,y) = \frac{1}{x}\{1 - \mathrm{Be}(v_{xy}; \alpha+1, \beta)\} + \frac{1}{y}\mathrm{Be}(v_{xy}; \alpha, \beta+1), \quad x, y > 0,$$

with parameters $\alpha, \beta > 0$, and where $\mathrm{Be}(v_{xy}; a, b)$ is the CDF of the beta distribution with parameters a and b evaluated at $v_{xy} = \alpha x/(\alpha x + \beta y)$. The Dirichlet model is symmetric when $\alpha = \beta$. Complete dependence is obtained in the limit as $\alpha = \beta$ tends to infinity. Independence is obtained as $\alpha = \beta$ approaches zero, and when one of α, β is fixed and the other approaches zero.

From equations (1.9) and (1.10) one can see that the spectral distribution doesn't have point masses on 0 or 1, and when $\alpha \in (0,1)$, its spectral density is

$$h(w) = \frac{\alpha^\alpha \beta^\beta \Gamma(\alpha+\beta+1)\, w^{\alpha-1}(1-w)^{\beta-1}}{2\Gamma(\alpha)\Gamma(\beta)\{\alpha w + \beta(1-w)\}^{\alpha+\beta+1}}, \quad 0 < w < 1,$$

where Γ denotes the Gamma function.

1.2.1.6 Pickands' dependence function

An alternative representation of (1.8) introduced by Pickands (1981) was the so-called Pickands dependence function, which is denoted $A(\boldsymbol{w})$. In the bivariate case, A is defined on $[0,1]$ and determined by

$$V(z_1, z_2) = \left(\frac{1}{z_1} + \frac{1}{z_2}\right) A\left(\frac{z_1}{z_1+z_2}\right), \quad z_1, z_2 > 0, \tag{1.11}$$

and satisfies i) $\max(t, 1-t) \le A(t) \le 1$ for $t \in [0,1]$, and ii) A is convex.

The function A lies between the two bounding cases of independence when $A(t) \equiv 1$ and complete dependence when $A(t) = \max(w, 1-t)$. The scalar $A(w)$ quantifies the strength of the dependence between the random variables Z_1 and Z_2, in the 'direction' w, where $w = z_1/(z_1+z_2)$ is the so-called pseudo-angle of (z_1, z_2) in pseudo-polar coordinates.

The Pickands dependence function may be expressed in terms of the exponent function.

Setting $z_1 = (1-t)^{-1}$ and $z_2 = t^{-1}$ in (1.11) yields

$$A(t) = V\left(\frac{1}{1-t}, \frac{1}{t}\right), \quad 0 \le t \le 1, \tag{1.12}$$

and in terms of the spectral distribution H as

$$A(t) = 1 - t + 2\int_0^t H([0,w])\,\mathrm{d}w, \quad 0 \le t \le 1. \tag{1.13}$$

Conversely, the spectral distribution is linked to Pickands' dependence function through

$$H([0,w]) = \begin{cases} \left\{1 + A'(w)\right\}/2, & w \in [0,1), \\ 1, & w = 1, \end{cases} \tag{1.14}$$

where A' is the right-hand derivative of A, and the point masses of H at 0 and 1 are

$$H(\{0\}) = \frac{1 + A'(0)}{2}, \quad H(\{1\}) = \frac{1 - A'(1)}{2},$$

where $A'(1) = \sup_{0\le t<1} A'(t)$. If A' is absolutely continuous, then H is absolutely continuous on the interior of the unit interval with density $h = A''/2$.

Similarly to Archimedean copula generators, which generate Archimedean copulas (see Section 1.2.1.1), Pickands dependence functions generate extreme-value copulas, since a bivariate copula may be written in terms of the Pickands dependence function as

$$C(u,v) = \exp\left[\log(uv) A\left\{\frac{\log v}{\log(uv)}\right\}\right], \quad 0 \le u, v \le 1. \tag{1.15}$$

The Pickands dependence function may be expressed in terms of the copula of a (bivariate) max-stable distribution. Setting $u = e^{t-1}$ and $v = e^{-t}$ in (1.15) yields

$$A(t) = -\log\left\{C\left(e^{t-1}, e^{-t}\right)\right\}, \quad 0 \le t \le 1.$$

The Pickands dependence functions and derivatives for the four models presented in Section 1.2.1.5 are

$$A(t) = \left\{(1-t)^{1/\alpha} + t^{1/\alpha}\right\}^{\alpha}, \quad A'(t) = A(t)^{1-1/\alpha}\left\{t^{1/\alpha-1} - (1-t)^{1/\alpha-1}\right\}, \quad 0 \le t \le 1,$$

for the logistic model,

$$\begin{aligned} A(t) &= (1-\psi_1)(1-t) + (1-\psi_2)t + \left[\{\psi_1(1-t)\}^{1/\alpha} + (\psi_2 t)^{1/\alpha}\right]^{\alpha}, \\ A'(t) &= \psi_1 - \psi_2 + \left\{\psi_2^{1/\alpha} t^{1/\alpha-1} - \psi_1^{1/\alpha}(1-t)^{1/\alpha-1}\right\} \\ &\quad \times \left[\{\psi_1(1-t)\}^{1/\alpha} + (\psi_2 t)^{1/\alpha}\right]^{\alpha-1}, \quad 0 \le t \le 1, \end{aligned}$$

for the asymmetric logistic model,

$$
\begin{aligned}
A(t) &= (1-t)\Phi\left\{\frac{\alpha}{2}-\frac{1}{\alpha}\log\left(\frac{t}{1-t}\right)\right\}+t\Phi\left\{\frac{\alpha}{2}-\frac{1}{\alpha}\log\left(\frac{1-t}{t}\right)\right\},\\
A'(t) &= -\Phi\left\{\frac{\alpha}{2}-\frac{1}{\alpha}\log\left(\frac{t}{1-t}\right)\right\}-\frac{1}{\alpha t}\phi\left\{\frac{\alpha}{2}-\frac{1}{\alpha}\log\left(\frac{t}{1-t}\right)\right\}\\
&\quad+\Phi\left\{\frac{\alpha}{2}-\frac{1}{\alpha}\log\left(\frac{1-t}{t}\right)\right\}+\frac{1}{\alpha(1-t)}\phi\left\{\frac{\alpha}{2}-\frac{1}{\alpha}\log\left(\frac{1-t}{t}\right)\right\},\quad 0\le t\le 1,
\end{aligned}
$$

for the Hüsler–Reiss model, and

$$
\begin{aligned}
A(t) &= (1-t)\left\{1-\mathrm{Be}(\upsilon_t;\alpha+1,\beta)\right\}+t\mathrm{Be}(\upsilon_t;\alpha,\beta+1),\\
A'(t) &= \mathrm{Be}(\upsilon_t;\alpha+1,\beta)+\mathrm{Be}(\upsilon_t;\alpha,\beta+1)-1,\quad 0\le t\le 1,
\end{aligned}
$$

where $\upsilon_t = \alpha t/\{\alpha t+\beta(1-t)\}$ for the Dirichlet model.

Figure 1.1 shows the Pickands dependence functions of these four models for different parameter values.

1.2.1.7 Inference

Motivated by the characterization of multivariate extreme value distributions (Theorem 1.9), MEV distributions are asymptotically justified to model extreme values of a multivariate phenomenon of interest. Similarly to the block maximum approach in univariate case, its application consists in grouping the data into blocks of equal length, and fitting a MEV distribution to componentwise maxima of non-overlapping blocks of consecutive observations.

Let $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N$ be a series of d-dimensional observations of a phenomenon of interest at regular time points, e.g., the daily rainfall at a set of specific locations. Supposing that the data can be decomposed into m blocks of n independent observations, that is $N = mn$, with $n, m \in \mathbb{N}$, one can fit a MEV distribution, either parametrically (e.g., by maximum likelihood) or nonparametrically, to the m componentwise block maxima $\boldsymbol{y}_1 = \max(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n),\ldots,\boldsymbol{y}_m = \max(\boldsymbol{x}_{n(m-1)+1},\ldots,\boldsymbol{x}_N)$.

The fitting procedures described in this section are based on the assumption that the asymptotic extreme value model is a reasonable approximation for a finite number of observations. However, the limiting models lie in a big class of distributions, because the spectral distribution H is nonparametric. Furthermore, limiting models may not be good approximations for finite samples as the convergence to the limiting distribution can be slow, inducing substantial bias in estimation and subsequent prediction. Omey and Rachev (1991) show that the rate of converge depends on that of the corresponding dependence structure and that of the margins (see Section 1.1.1.2).

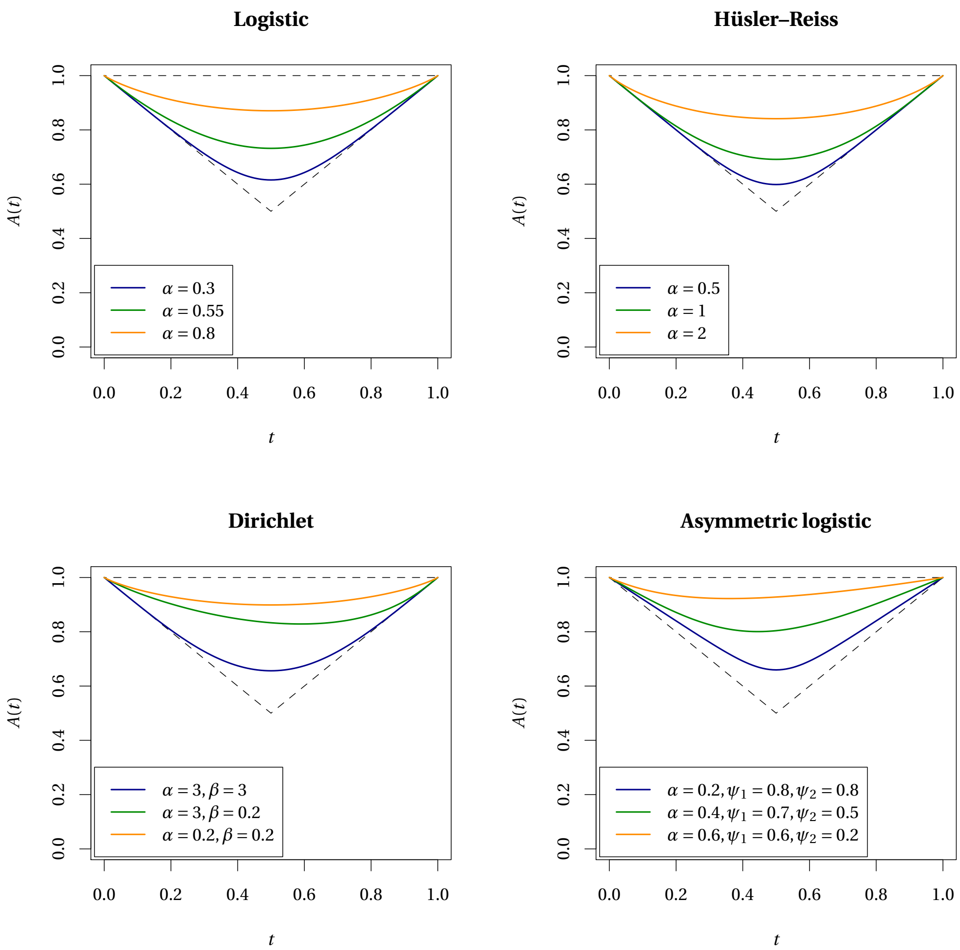

Figure 1.1 – *Pickands dependence function of the logistic, Hüsler–Reiss, Dirichlet, and asymmetric logistic models for different parameter values. The horizontal dashed line corresponds to independence, and the v-shaped dashed line corresponds to perfect dependence (comonotonicity).*

Maximum likelihood estimation Parametric estimation requires one to choose a suitable parametric model for G from the large collection of existing MEV distributions. Let $G(\cdot;\boldsymbol{\psi})$ denote the CDF of the chosen parametric model and $g(\boldsymbol{z};\boldsymbol{\psi}) = \partial^d G(\boldsymbol{z};\boldsymbol{\psi})/(\partial z_1 \cdots \partial z_d)$ its PDF, where $\boldsymbol{\psi}$ denotes the (finite-dimensional) vector of unknown model parameters which need to be estimated from the data. Let $\boldsymbol{z}_1,\ldots,\boldsymbol{z}_m$ be the m componentwise maxima transformed to have unif Fréchet margins based on univariate estimation of the GEV distribution (Section 1.1.1.3) and marginal transformations (Section 1.2.1.3). The likelihood function is

$$L(\boldsymbol{\psi};\boldsymbol{z}_1,\ldots,\boldsymbol{z}_n) = \prod_{i=1}^{n} g(\boldsymbol{z}_i;\boldsymbol{\psi}),$$

which is to be maximized with respect to $\boldsymbol{\psi}$, yielding the parameter estimate $\hat{\boldsymbol{\psi}}$. The variance of $\hat{\boldsymbol{\psi}}$ can be estimated using the observed information matrix. In practice, maximization of the (log-)likelihood and computation of the observed information matrix are performed numerically.

In dimension $d = 2$, the joint density function is

$$g(z_1,z_2;\boldsymbol{\psi}) = \{V_1(z_1,z_2)V_2(z_1,z_2) - V_{12}(z_1,z_2)\} \times \exp\{-V(z_1,z_2)\}, \quad z_1,z_2 > 0,$$

where $V_1(z_1,z_2) = \partial V(z_1,z_2)/\partial z_1$, and similarly for V_2 and V_{12}. The number of terms in the expression for the joint density grows extremely fast as the number of dimension d increases, quickly yielding an intractable likelihood, so alternative methods are needed for inference in high dimensions. Using the times of occurrence of extreme events, if available, simplifies the likelihood and also inference; see, e.g., Stephenson and Tawn (2005) and Wadsworth and Tawn (2014). An alternative is to use a surrogate for the full likelihood such as a composite likelihood, which is an inference function derived by multiplying a collection of valid likelihood objects usually related to small subsets of data, e.g., bivariate subsets in the case of pairwise likelihood; see e.g., Varin and Vidoni (2005), Varin (2008) and Varin *et al.* (2011). Under mild conditions, maximum pairwise likelihood estimators are strongly consistent and asymptotically normal. Using pairwise likelihood seems reliable and avoids the intractability of the likelihood function, as only bivariate margins need to be specified, provided the parameter $\boldsymbol{\psi}$ is identifiable from them.

A gain of efficiency may occur if the GEV margins and the dependence model are estimated together. The number of parameters to be estimated is increased by $3d$, which makes numerical optimization harder, especially in larger dimensions, but doesn't yield intractable (composite) likelihoods. However, jointly modelling the margins and the dependence structure may not always be desirable. Dupuis and Tawn (2001) show that misspecifying the dependence structure may have large adverse effects on the estimates of margin parameters. Goodness-of-fit tests may cast doubt on the hypothesis of extreme value margins, although a MEV distribution may seem valid for the dependence structure.

Non-parametric inference Several nonparametric estimators of the Pickands dependence function have been proposed. Pickands (1981) proposed an estimator that is conceptually simple and easy to compute, but the resulting estimates may not be valid Pickands dependence functions. Deheuvels (1991) and Hall and Tajvidi (2000) proposed improved estimators which satisfy some of the necessary constraints but still fail to be convex. Capéraà *et al.* (1997) proposed another estimator which also fails to satisfy some of the necessary constraints. Marcon *et al.* (2017) proposed a valid (multivariate) estimator using Bernstein polynomials.

1.2.2 Point process approach

The componentwise block maximum approach discussed in Section 1.2.1 only uses information equivalent to one observation in each block, though the componentwise maxima may not be actual observations. A less data-wasteful alternative is based on a point process characterization of extremes introduced by de Haan and Resnick (1977). The theory is detailed in Resnick (1987, Chap. 3 and 5), and summarized in Kotz and Nadarajah (2000, Sec. 3.2) and Fougères (2003). The main result is an extension of the point process characterization in the univariate case (Theorem 1.4).

As in Section 1.2.1.1, let $\boldsymbol{X} = (X_1, \dots, X_d)$ be a d-dimensional random vector with joint CDF F and marginal CDFs $F_1, \dots, F_d$. Let $\{\boldsymbol{X}_i\}_{i\geq 1}$ with $\boldsymbol{X}_i = (X_{i,1}, \dots, X_{i,d})$ be an i.i.d. sequence of random vectors distributed as $\boldsymbol{X}$, and let $\boldsymbol{M}_n = (M_{n,1}, \dots, M_{n,d})$ with $M_{n,i} = \max(X_{1,i}, \dots, X_{n,i})$ be the vector of componentwise maxima of $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$. Without loss of generality, suppose that the margins F_i are unit Fréchet, that is $F_i(z) = \exp(-1/z)$, $z > 0$.

Theorem 1.10 (Convergence of the point process of rescaled observations Resnick (1987, Prop. 5.11))**.** *If the renormalized vector maxima* $\boldsymbol{M}_n/n \xrightarrow{D} \boldsymbol{Z} \sim G$*, where* G *is a non-degenerate distribution, then the sequence of point processes* $\mathcal{P}_n = \{\boldsymbol{X}_i/n\}_{i=1}^n$*, for* $n \geq 1$*, converges to a non-homogeneous Poisson process* $\mathcal{P}$ *on* $(\boldsymbol{0}, \infty)$ *with measure* μ*.*

The scaling factor n in Theorem 1.10 corresponds to the normalizing constants $\boldsymbol{a}_n = \boldsymbol{n}$ and $\boldsymbol{b}_n = \boldsymbol{0}$ for unit Fréchet random variables in Theorem 1.9.

There is a strong connection between the characterization of MEV distributions (Theorem 1.9) and the convergence of the point process of rescaled observations (Theorem 1.10). Let $\boldsymbol{z} = (z_1, \dots, z_d) > \boldsymbol{0}$, and consider $\mathcal{A}_{\boldsymbol{z}} = \{\boldsymbol{x} \in \mathbb{R}_+^d : x_1 > z_1 \text{ or } \cdots \text{ or } x_d > z_d\}$. It follows from Theorem 1.10 and the Poisson property that

$$\Pr(\mathcal{P}_n \subset \mathcal{A}_{\boldsymbol{z}}^c) = \Pr(\mathcal{P}_n \cap \mathcal{A}_{\boldsymbol{z}} = \emptyset) \to \exp\{-\mu(\mathcal{A}_{\boldsymbol{z}})\}, \quad n \to \infty.$$

From Theorem 1.9 and under the assumptions of Theorem 1.10, we have

$$\Pr\left(\frac{\boldsymbol{M}_n}{n} \leq \boldsymbol{z}\right) \to G(\boldsymbol{z}) = \exp\{-V(\boldsymbol{z})\}, \quad n \to \infty,$$

where V is the exponent function defined in (1.8). Noting that $\Pr(\mathcal{P}_n \subset \mathcal{A}_{\boldsymbol{z}}^c) = \Pr(\boldsymbol{M}_n/n \leq \boldsymbol{z})$

yields $\exp\{-\mu(\mathcal{A}_{\boldsymbol{z}})\} = \exp(-V\{\boldsymbol{z})\}$, so the limiting Poisson process has measure $\mu(\mathcal{A}_{\boldsymbol{z}}) = V\{\boldsymbol{z}\}$, for $\boldsymbol{z} > \mathbf{0}$ or equivalently $\mathcal{A}_{\boldsymbol{z}} \subset (\mathbf{0}, \infty)$.

A transformation of $\boldsymbol{X}$ from Cartesian to pseudo-polar coordinates turns out to be useful. Let

$$R = \sum_{i=1}^{d} X_i, \quad \boldsymbol{W} = \frac{\boldsymbol{X}}{R},$$

respectively be the pseudo-radius and the vector of pseudo-angles of $\boldsymbol{X}$. Clearly, $R > 0$, $W_i \in [0,1]$, for $i = 1,\dots,d$, and $\sum_{i=1}^{d} W_i = 1$. The components of $\boldsymbol{W}$ correspond to the relative sizes of the component of $\boldsymbol{X}$.

The intensity measure $\mathrm{d}\mu$ of the limiting Poisson process with the pseudo-polar parametrization, if it exists, is

$$\mathrm{d}\mu(r, \boldsymbol{w}) = d\frac{\mathrm{d}r}{r^2}\,\mathrm{d}H(\boldsymbol{w}), \tag{1.16}$$

where H is the spectral distribution introduced in Theorem 1.9 and describing the structure of dependence of MEV distributions. Expression (1.16) implies that the intensity measure of the limiting process $\mathcal{P}$ in Theorem 1.10 factorizes across radial and angular components, so the radial distance is independent from the angular spread, which is determined by the spectral distribution H, and this remarkable property can be used for inference.

1.2.3 Asymptotic independence

Many multivariate distributions, including the non-degenerate multivariate normal distribution, lie in the maximum domain of attraction of the boundary case of independence. These distributions, which are often relevant in applications, need special treatment as independence of extreme events of the components is unrealistic.

Definition 1.16 (Asymptotic independence and asymptotic dependence)**.** A d-dimensional multivariate distribution F which is in the maximum domain of attraction of some MEV distribution G having margins $G_1,\dots,G_d$ is called asymptotically independent if, for all $\boldsymbol{x} = (x_1,\dots,x_d)$,

$$G(\boldsymbol{x}) = G_1(x_1) \times \cdots \times G_d(x_d).$$

A d-dimensional random variable is said to be asymptotically independent if its distribution is asymptotically independent. A d-dimensional distribution or random variable is said to be asymptotically dependent if it is not asymptotically independent.

Berman (1961) showed that a random vector $\boldsymbol{X} = (X_1,\dots,X_d)$ is asymptotically independent if all pairs (X_i, X_j) with $i \neq j$ are asymptotically independent. Sibuya (1960) showed that a necessary and sufficient condition for the asymptotic independence of a pair of random variables (X, Y) respectively having CDFs F_X and F_y is

$$\lim_{u \uparrow 1} \frac{\Pr\{F_Y(Y) > u, F_X(X) > u\}}{1-u} = 0. \tag{1.17}$$

Coles *et al.* (1999) illustrate that asymptotically independent random variables may exhibit (strong) upper joint tail dependence above high quantiles, that is $\Pr\{F_Y(Y) > u, F_X(X) > u\}$ may be far from zero even for values of u very close to one. Therefore, testing for asymptotic independence is difficult in practice, as a powerful test requires information very far in the tail. Several approaches to test bivariate asymptotic independence have been proposed; see e.g., Ledford and Tawn (1996, 1997), Coles *et al.* (1999), Ramos and Ledford (2005), Falk and Michel (2006), and the references therein. Bacro *et al.* (2010) proposed the so-called madogram test. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a sequence of i.i.d. pairs of random variable. Under the hypothesis of asymptotic independence, the statistic

$$\hat{\nu}_W = \frac{1}{2n} \sum_{i=1}^{n} \left| \hat{F}_X(X_i) - \hat{F}_Y(Y_i) \right|,$$

where $\hat{F}_X$ and $\hat{F}_Y$ respectively are the empirical CDFs of the X_is and Y_is, is asymptotically normal.

1.2.3.1 Models for asymptotic independence

Several models have been proposed in the literature to model asymptotic independence, including a characterization of bivariate joint tails based on slowly-varying functions proposed by Ledford and Tawn (1996), with the subsequent generalizations and refinements proposed by Ledford and Tawn (1997, 1998) and Ramos and Ledford (2009); inverted max-stable distributions proposed in the bivariate case by Ledford and Tawn (1997) and extended to the multivariate case by Heffernan and Tawn (2004); the semi-parametric conditional approach proposed by Heffernan and Tawn (2004); and the alternative limiting point process proposed Ramos and Ledford (2011). de Carvalho and Ramos (2012) provide a review of statistical modelling of asymptotically independent data.

The Ledford–Tawn and Ramos–Ledford models provide broader approaches, but only work for dimensions $d = 2, 3$. The Heffernan–Tawn model provides a broad approach to modelling in larger dimensions, but the approach is not fully satisfactory as conditioning on different variables yields incompatible results.

Ledford–Tawn model Let (Z_1, Z_2) be a pair of non-negatively associated random variables with joint distribution F and Fréchet margins. The marginal survival functions of Z_1 and Z_2 are $1 - e^{-1/z} \approx z^{-1}$ for large z. For the two bounding cases of independence and perfect dependence, the joint survival function of Z_1 and Z_2 is $\Pr(Z_1 > z, Z_2 > z) = e^{-\alpha} \approx z^{-\alpha}$ for large z, with $\alpha = 2$ for independence, and $\alpha = 1$ for perfect dependence. Ledford and Tawn (1996) proposed to model the joint survival function of (Z_1, Z_2) with a model that links these two bounding cases, and suggested

$$\Pr(Z_1 > z, Z_2 > z) \approx \mathcal{L}(z) z^{-1/\eta}, \tag{1.18}$$

as $z \to \infty$, where $1/2 \leq \eta \leq 1$ is a constant called the coefficient of tail dependence, and $\mathcal{L}$ is a slowly varying function, that is, $\mathcal{L}$ is such that $\mathcal{L}(tr)/\mathcal{L}(r) \to 1$ as $r \to \infty$ for any finite $t > 0$. Under this model, the tail dependence of (Z_1, Z_2) is characterised by both the constant η, and the slowly varying function $\mathcal{L}$. The rate of decay in (1.18) is primarily determined by η. The degree of dependence of large values Z_1 and Z_2 is determined by η, with larger values corresponding to stronger association. For a given level η, the relative strength of dependence is characterized by $\mathcal{L}$.

Distributions satisfying (1.18) can be separated into the following cases:

- Asymptotic dependence if $\eta = 1$ and $\lim_{z\to\infty} \mathcal{L}(z) = c$ for some constant $0 < c \leq 1$. The degree of dependence is quantified by c. This class comprises all non-independent bivariate extreme value distributions. The bounding case of perfect dependence has $\mathcal{L}(z) = 1$.

- Asymptotic independence with positive association if either $1/2 < \eta < 1$ or $\lim_{z\to\infty} \mathcal{L}(z) = 0$. This class contains for example the bivariate Gaussian distribution with correlation coefficient $0 < \rho < 1$, for which $\eta = (1+\rho)/2$.

- Asymptotic independence and near independence if $\eta = 1/2$. This class comprises bivariate distributions that are exactly independent in the limit but have some weak dependence at subasymptotic levels due to the fluctuations caused by $\mathcal{L}$. The bounding case of independence has $\mathcal{L}(z) = 1$.

The class of distributions considered by Ledford and Tawn (1996) can be extended to negatively associated distributions, for which $0 < \eta < 1/2$. This class contains for example the bivariate Gaussian distribution with correlation coefficient $\rho < 0$, for which $\eta = (1+\rho)/2$. Heffernan (2000) provides the coefficients of tail dependence for a collection of bivariate copulas. Schlather (2001) shows by a counterexample that (1.18) neither implies nor is implied by the domain-of-attraction condition.

Ledford and Tawn (1996) proposed an estimator of the coefficient of tail dependence η which relies on the fact that $\Pr(Z_1 > z, Z_2 > z) = \Pr\{\min(Z_1, Z_2) > z\}$, and they show that η can be estimated as the tail index of the (univariate) variable $T = \min(Z_1, Z_2)$.

Ledford and Tawn (1997, 1998) proposed an extension of model (1.18) with the more flexible joint tail asymptotic expansion

$$\Pr(Z_1 > z_1, Z_2 > z_2) \approx \mathcal{L}_2(z_1, z_1) z_1^{-1/c_1} z_2^{-1/c_2}, \tag{1.19}$$

where $c_1 + c_2 = \eta$, and $\mathcal{L}_2$ is a bivariate slowly varying function. Ramos and Ledford (2009) considered the special case of (1.19) where $c_1 = c_2$.

1.2.4 Measures of extremal dependence

Measures of extremal dependence are useful to summarize the level of extremal dependence for diagnostic and model checking purposes, and discriminate between asymptotic dependence and asymptotic independence. Several measures of extremal dependence have been proposed in the literature, including the extremal coefficient attributed to Smith (1990), the coefficient of tail dependence introduced by Ledford and Tawn (1996), and the coefficients χ and $\overline{\chi}$) introduced by Coles *et al.* (1999); see de Carvalho and Ramos (2012) for a review.

Extremal coefficient $\boldsymbol{\theta}_d$ Let $\boldsymbol{Z} = (Z_1, \ldots, Z_d)$ be a random vector with MEV distribution G and Fréchet margins, so $G(\boldsymbol{z}) = \exp\{-V(\boldsymbol{z})\}$, $\boldsymbol{z} > \mathbf{0}$. The structure of (extremal) dependence of $\boldsymbol{Z}$ is entirely determined by the exponent function V, which is homogeneous of order -1. A simple summary of dependence is the extremal coefficient

$$\theta_d = V(1, \ldots, 1) \in [1, d],$$

which satisfies $\theta_d = 1$ for perfectly dependent data, and $\theta_d = d$ for independent data. Since

$$\Pr\{\max(Z_1, \ldots, Z_d) \leq z\} = \exp\{-V(z, \ldots, z\} = \{\exp(-1/z)\}^{\theta_d}, \quad z > 0,$$

which correponds to the distribution of θ_d independent Fréchet variables, the extemal coefficient can be loosely interpreted as the number of independent components of $\boldsymbol{Z}$ contributing to $\max(Z_1, \ldots, Z_d)$. The extremal coefficient is only defined for MEV distributions. It can be expressed in terms of the Pickands dependence function A. In dimension $d = 2$, equation (1.11) yields $\theta_2 = 2A(1/2)$. Estimators of the extremal were proposed by Schlather and Tawn (2003) and Naveau *et al.* (2009), among others.

The coefficient is for MEV but one may want to extend it to it to any type of distribution. A natural generalization is $\theta_d = \lim_{u\uparrow 1} \log C(u, \ldots, u)/\log u$ as for MEV distributions F we have $C(u, \ldots, u) = F\{F_1^{-1}(u), \ldots, F_d^{-1}(u)\} = \exp[V\{-1/\log u, \ldots, -1/\log u\}] = \exp\{V(1, \ldots, 1)\log u\} = u^{V(1,\ldots,1)} = u^{\theta_d}$.

Coefficient of tail dependence η The coefficient of tail dependence η introduced by Ledford and Tawn (1996) characterizes the tail decay of the survival joint distribution of a pair of Fréchet variables (see Section 1.2.3.1). It allows discrimination between asymptotic dependence ($\eta = 1$)

and asymptotic independence ($0 < \eta < 1$), and measures the strength of extremal dependence within the class of asymptotically independendent models. Heffernan (2000) provides the coefficients of tail dependence for a collection of bivariate copulas.

The coefficient η corresponds to the shape parameter ξ of the GEV distribution of the univariate variable $T = \min(Z_1, Z_2)$ where Z_1 and Z_2 are Fréchet random variables, so that univariate threshold methods can be used for inference (Ledford and Tawn, 1996). Alternative estimators are described in Beirlant *et al.* (2004, Sec. 9.5.2).

Coefficients χ and $\overline{\chi}$ Coles *et al.* (1999) discuss two complementary quantities, χ and $\overline{\chi}$, that measure different aspects of extremal dependence for pairs of random variables. Both are needed to obtain a summary of extremal dependence that is informative for variables that may be either asymptotically dependent or asymptotically independent.

A way to quantify the extremal dependence of two random variables is to consider the probability that one variable is large given that the other is large. Let (X, Y) be a pair of random variables with joint distribution F, underlying copula C, and marginal distributions F_1 and F_2. The tail dependence index discussed by Coles *et al.* (1999) is

$$\chi = \lim_{u \uparrow 1} \Pr\{F_2(X_2) > u \mid F_1(X_1) > u\} = \lim_{u \uparrow 1} 2 - \frac{1 - C(u, u)}{1 - u},$$

provided the limit exists. The coefficient χ lies in $[0, 1]$ and is not restricted to bivariate MEV distributions. It quantifies the level of dependence that remains in the limit, allowing one to assess the level of asymptotic dependence. Given condition (1.17), the random variables X and Y are asymptotically independent if $\chi = 0$, and asymptotically dependent otherwise. If $\chi > 0$, χ quantifies the strength of asymptotic dependence, with the intuitive intrepretation that the larger χ is, the stronger the extremal dependence.

To quantify the level of subasymptotic upper tail dependence at any quantile level u, Coles *et al.* (1999) introduced the function

$$\chi(u) = 2 - \frac{\log C(u, u)}{\log u}, \quad 0 < u < 1, \tag{1.20}$$

and one has $\lim_{u \uparrow 1} \chi(u) = \chi$. In general, the function $\chi(u)$ is bounded from below and above by

$$2 - \frac{\log\{\max(2u - 1, 0)\}}{\log u} \leq \chi(u) \leq 1, \quad 0 < u < 1,$$

which follows from the Fréchet bounds for copulas (e.g., Nelsen, 2006, p. 11). The sign of $\chi(u)$ corresponds to the sign of the association between X and Y at level u, with $\chi(u) = 0$ corresponding to independence, since the variables are by definition positively quadrant dependent if $C(u, u) > u^2$ (Nelsen, 2006, Sec. 5.2.1).

The only situation corresponding to asymptotic independence is the bounding case $\chi = 0$, leading the measure to fail to discriminate between the degrees of relative strength of dependence for asymptotically independent variables. To measure extremal dependence under asymptotic independence, and by analogy to (1.20), Coles *et al.* (1999) proposed the complementary dependence function

$$\overline{\chi}(u) = \frac{2\log(1-u)}{\log\overline{C}(u,u)} - 1, \quad 0 < u < 1, \tag{1.21}$$

where

$$\overline{C}(u,v) = \Pr\{F_X(X) > u, F_Y(Y) > v\} = 1 - u - v - C(u,v), \quad 0 \le u, v \le 1,$$

is the survival copula. Fréchet bounds for copulas yield

$$\frac{2\log(1-u)}{\log\{\max(1-2u,0)\}} - 1 \le \overline{\chi}(u) \le 1, \quad -1 < u < 1.$$

The sign of $\overline{\chi}(u)$ corresponds to the sign of the association between X and Y at level u, with $\overline{\chi}(u) = 0$ corresponding to independence. The corresponding dependence coefficient is

$$\overline{\chi} = \lim_{u\uparrow 1} \overline{\chi}(u),$$

provided it exists, with $\overline{\chi} \in [-1,1]$.

The coefficients χ and $\overline{\chi}$ are linked to the extremal coefficient θ_2 and the tail dependence index η. Indeed, if the distribution F is a MEV distribution with Pickands dependence function A, then by (1.15) we have $C(u,u) = u^{\theta_2}$, where $\theta_2 = 2A(1/2)$ is the extremal coefficient. Thus $\chi(u) = 2 - \theta_2$, $0 < u < 1$, and thus $\chi = 2 - \theta_2$. In addition, expressing (1.18) in terms of the survival copula and setting $u = e^{-1/z}$ gives

$$\overline{C}(u,u) \approx \mathcal{L}(-(1-u)^{-1})(1-u)^{1/\eta},$$

and substituting into (1.21) yields

$$\chi = \lim_{u\uparrow 1} \overline{\chi}(u) = \lim_{u\uparrow 1} \frac{2\log(1-u)}{\log\mathcal{L}\left\{(1-c)^{-1}\right\} + \eta^{-1}\log(1-u)} - 1 = 2\eta - 1.$$

In summary, the complete pair $(\chi, \overline{\chi})$ with $\chi \in [0,1]$ and $\overline{\chi} \in [-1,1]$ is required to measure extremal dependence of a pair of random variables: $(\chi > 0, \overline{\chi} = 1)$ corresponds to asymptotic dependence and χ measures the strength of extremal dependence in this class; alternatively, $(\chi = 0, \overline{\chi} < 1)$ corresponds to asymptotic independence and $\overline{\chi}$ measures the strength of extremal dependence in this class.

Coles *et al.* (1999) suggest estimation of $\chi(u)$ and $\overline{\chi}(u)$ by using empirical estimates of C

and $\overline{C}$.

2 Spectral distributions and the angular distribution of excesses

The spectral distribution plays a key role in the statistical modelling of multivariate extremes, as it defines the dependence structure of multivariate extreme-value distributions and characterizes the limiting distribution of the relative sizes of the components of large multivariate observations. However, no parametric family captures all possible types of multivariate dependence. Numerous parametric models have been proposed, including those of Gumbel (1961), Tawn (1988a), Hüsler and Reiss (1989), Coles and Tawn (1991), Demarta and McNeil (2005), and Cooley *et al.* (2010), and construction principles have been suggested, such as that in Segers (2012).

Inference on the spectral distribution is typically based on the pseudo-angles of 'large' observations under the assumption that their distribution is equal to the spectral distribution. There has been little if any attention on studying the impact of this approximation on inference, and it turns out that it can yield significantly biased estimates.

The aim of this chapter is to characterize the angular distribution of excesses corresponding to the distribution of pseudo-angles of 'large' observations, in order to improve direct inference on the spectral distribution in the bivariate setting. Section 2.1 contains definitions and sets some notation. In Section 2.2, we illustrate some pitfalls of direct inference on the spectral distribution using observed pseudo-angles of 'large' observations. In Section 2.3, we give a characterization of the angular distribution of excesses and illustrate it with examples. In Section 2.4, we describe how to perform maximum likelihood-based parametric inference using the angular distribution of excesses, and we compare this approach with other classical parametric approaches through numerical simulations. Section 2.5 concludes the chapter with a summary.

2.1 Definitions and notation

Let $(X,Y)^{\mathrm{T}}$ be a random vector in $\mathbb{R}^2$ which follows a bivariate extreme value distribution denoted by $\mathcal{G}$ having joint CDF G with unit Fréchet margins, that is $F_X(z) = F_Y(z) = e^{-1/z}$,

$z > 0$. Recall from Section 1.2.1.4 that the joint CDF of $\mathcal{G}$ can be expressed as

$$G(x, y) = \exp\{-V(x, y)\}, \quad x, y > 0,$$

where

$$V(x, y) = 2\int_0^1 \max\left(\frac{t}{x}, \frac{1-t}{y}\right) \mathrm{d}H(t),$$

and H is the CDF of some probability distribution $\mathcal{H}$ defined on $[0, 1]$ and satisfying the mean constraint

$$\int_0^1 t\,\mathrm{d}H(t) = 1/2. \tag{2.1}$$

The function V is often called the exponent function, and we say that the distribution $\mathcal{H}$ is the spectral distribution associated with the extreme value distribution $\mathcal{G}$.

Definition 2.1. Let $\mathbb{G}_2$ denote the set of bivariate extreme value distributions with Fréchet margins that have a differentiable CDF on $\mathbb{R}_+^2$.

The set $\mathbb{G}_2$ corresponds to the collection of all bivariate extreme value distributions that have a PDF. For a distribution $\mathcal{G} \in \mathbb{G}_2$, let G denote its CDF and g its PDF.

Definition 2.2. A univariate spectral distribution $\mathcal{H}$ is called 01-continuous if its CDF is continuous on the open interval $(0, 1)$. Let $\mathbb{H}_2$ denote the set of all 01-continuous univariate spectral distributions.

The set $\mathbb{H}_2$ corresponds to the collection of all univariate spectral distributions that have no point masses in $(0, 1)$. For $\mathcal{H} \in \mathbb{H}_2$, let H denote its CDF. Let δ_0 and δ_1 denote potential point masses respectively at 0 and 1, and denote the generalized density function of $\mathcal{H}$ by

$$h(t) = \delta_0\delta(t) + \tilde{h}(t) + \delta_1\delta(t-1), \quad 0 \le t \le 1, \tag{2.2}$$

where $\delta(\cdot)$ denotes the Dirac delta function and $\tilde{h}(t) = \mathrm{d}H(t)/\mathrm{d}t$ on $(0, 1)$ extended by continuity at 0 and 1. Therefore,

$$\int_0^1 \tilde{h}(t)\,\mathrm{d}t = 1 - \delta_0 - \delta_1,$$

showing that $\tilde{h}$ itself is a valid PDF if and only if $\delta_0 = \delta_1 = 0$, in which case $h \equiv \tilde{h}$. The mean constraint (2.1) implies that

$$\int_0^1 t\tilde{h}(t)\,\mathrm{d}t = \frac{1}{2} - \delta_1. \tag{2.3}$$

Pseudo-polar coordinates are useful to characterize dependence at extreme levels. Let the random vector $(R, W)^{\mathrm{T}}$ denote the transformation of $(X, Y)^{\mathrm{T}}$ in pseudo-polar coordinates, where the pseudo-radius $R = X + Y$ corresponds to the L_1 distance to the origin and the pseudo-angle $W = X/R$ corresponds to the pseudo-angle with the y-axis. Clearly, $R > 0$ and $0 < W < 1$, since both X and Y follow unit Fréchet distributions.

The spectral distribution $\mathcal{H}$ can be interpreted as the limiting distribution of the pseudo-angle W as $R \to \infty$, giving the basis for classical direct inference on H. We use the term angular distribution for the distribution of the pseudo-angle W for finite values of R to distinguish it from the spectral distribution $\mathcal{H}$.

2.2 Pitfalls in inference on the spectral distribution

In this section we illustrate how direct inference on the spectral distribution can go wrong when using a natural approach based on the convergence of the angular distribution of 'large' observations, as the threshold z tends to infinity.

In statistics, it is common to rely on asymptotic theory for inference, assuming that the characteristics of a quantity of interest based on a finite set of observations are (approximately) equal to the limiting characteristics as the sample size tends to infinity. A classical example is the use of the Central Limit Theorem in some hypothesis testing procedures. A natural approach to inference on the spectral distribution is based on this paradigm, relying on the fact that the spectral distribution $\mathcal{H}$ is the limit distribution of the pseudo-angle W when $R > z$ as $z \to \infty$. This approach consists in fitting a model for $\mathcal{H}$ to the observed pseudo-angles of 'large' observations, which are the subset of observations having a pseudo-radius larger than some high threshold.

Let $(x_1, y_1), \dots, (x_N, y_N)$ be a sample from a bivariate max-stable distribution with unit Fréchet margins. The sample is assumed to come from a max-stable distribution, and not only from a distribution that is in the maximum domain of attraction of a max-stable distribution, in order to illustrate some specific behaviour of the estimator in the best case scenario. In practice, data can be transformed to have unit Fréchet margins (see Section 1.2.1.3). Expressed in pseudo-polar coordinates these observations are (r_i, w_i), where $r_i = x_i + y_i$ and $w_i = x_i/r_i$ for $i = 1, \dots, N$. Let $z > 0$ be a large threshold, and let n be the number of 'large' observations such that $r_i > z$. Without loss of generality and to simplify notation, suppose that the observations are ordered in decreasing order of pseudo-radius, that is $r_1 \geq \cdots \geq r_N$, so the set of pseudo-angles for 'large' observations is $\{w_1, \dots, w_n\}$.

Figure 2.1 shows scatter plots of pseudo-random samples from the logistic, Hüsler–Reiss, Dirichlet, and asymmetric logistic models with Fréchet margins both in Euclidean and pseudo-polar coordinates as well as histograms of pseudo-angles of 'large' observations. These histograms seem quite close to the spectral distributions associated with the bivariate extreme value distributions from which the data were sampled, illustrating the natural approach of using pseudo-angles of 'large' observations to directly estimate the spectral distribution. However, for the asymmetric logistic distribution, the histogram deviates significantly from the

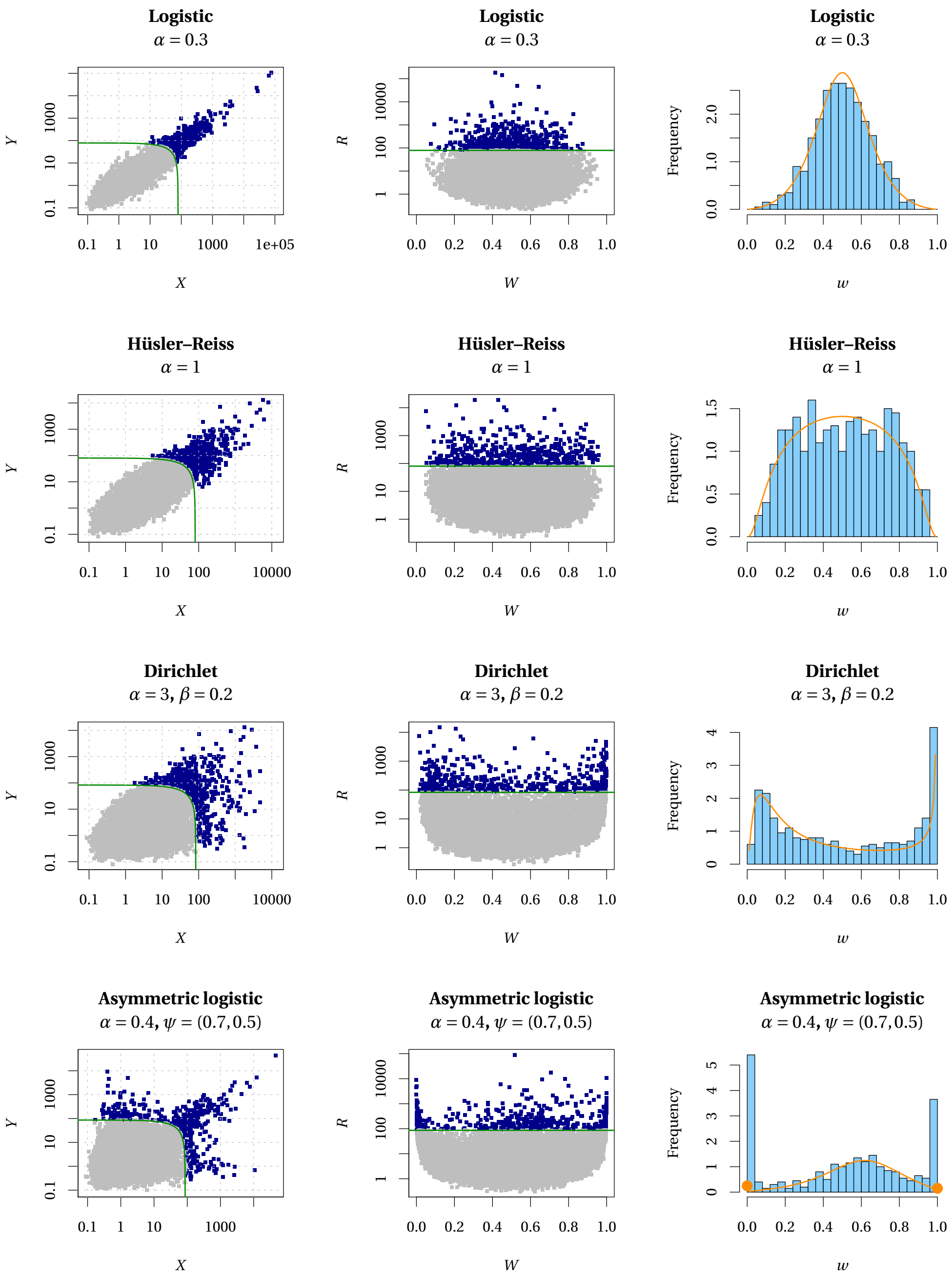

Figure 2.1 – *Scatter plots of Y versus X (left) and $R = X + Y$ versus $W = X/R$ (middle), and histograms of W given $R > z$ with z corresponding to the 97.5% empirical quantile of R (right) for pseudo-random samples of size 20'000 (500 threshold exceedances) from the logistic, Hüsler–Reiss, Dirichlet, and asymmetric logistic models with Fréchet margins. The solid green line in the scatter plots shows the threshold z, and observations with $R > z$ are shown in blue. The solid orange line in the histograms shows the spectral density function and orange dots show point masses of the spectral distribution.*

spectral density near the edges of the support. The most frequently observed pseudo-angles are those near the edges of the support, though the spectral density in these regions is close to zero. These discrepencies are due to the point masses of the spectral distribution at 0 and 1 and the fact the corresponding pseudo-angles can never be observed, since the magnitude of observations is both strictly positive and finite, so $0 < w_i < 1$ for $i = 1, \ldots, N$. Loosely speaking, the point masses of the spectral distribution spread onto the support of the distribution of observed pseudo-angles, mainly near the edges of the support. As a consequence, when the spectral distribution has point masses at 0 and/or 1, the distribution of observed pseudo-angles is inflated near the edge of the support, and it lacks the point masses. These discrepencies near the edges of the support may seem minor but can have a significant impact on inference, as illustrated below.

Maximum likelihood estimation of the spectral distribution using the natural approach is straightforward. Consider a parametric family of spectral distribution $\mathcal{H}_{\boldsymbol{\theta}}$ with parameter $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$ and PDF $h(\cdot;\boldsymbol{\theta})$. The maximum likelihood estimate is simply

$$\hat{\boldsymbol{\theta}}_{\mathrm{a}} = \operatorname*{arg\,max}_{\boldsymbol{\theta} \in \Theta} \sum_{j=1}^{n} \log h(w_j;\boldsymbol{\theta}), \tag{2.4}$$

where the subscript 'a' specifies that the estimator is obtained by fitting the asymptotic distribution of the pseudo angles.

Figure 2.2 shows maximum likelihood estimates of the spectral distribution of the logistic, Hüsler–Reiss, Dirichlet, and asymmetric logistic models based on observed pseudo-angles of the $n = 1'000$ largest observations out of $N = 40'000$ pseudo-random samples generated from the same models with Fréchet margins. The set-up for these fits is close to ideal for inference as the sample size is quite large, with 1'000 observations to estimate at most three parameters for univariate distributions having bounded support, the pseudo-angles correspond to observations with only the largest 2.5% pseudo-radiuses, and there is supposedly no model misspecification, as the logistic spectral distribution is fitted to data sampled from the logistic distribution, and so on. The estimates are quite close to the true parameter values for the first three distributions, but not for the asymmetric logistic distribution, where parameter estimates are quite far from the true values, resulting in a rather poor estimate of the spectral distribution. This poor estimate underestimates the occurrence of joint extreme events corresponding to the center of the spectral distribution, and could yield wrong conclusions in applications. The worst estimate in these examples occurs for the model having a spectral distribution with point masses at 0 and 1. The fact that one can't observe pseudo-angles equal to 0 and 1 corresponding to the point masses of the underlying spectral distribution can have a disastrous impact on inference, as estimation is biased towards a model of the parametric family having a PDF with high values near the edges of the support. Problems can also arise in the non-parametric setting and when the underlying distribution has no point masses, as illustrated below.

Figure 2.3 shows kernel density estimates of the spectral density h from the logistic, Hüsler–

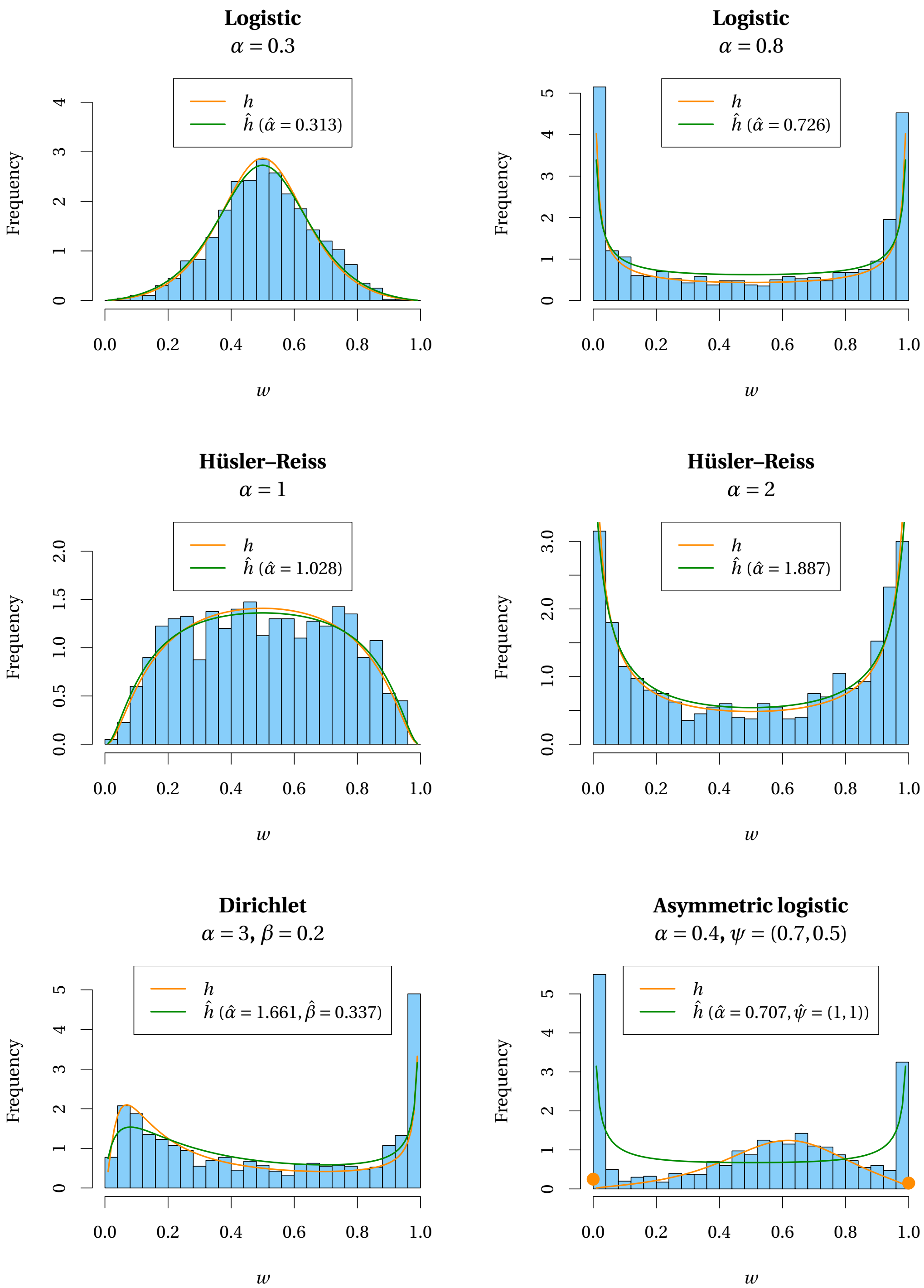

Figure 2.2 – *Histograms of the pseudo-angles W of the 1'000 largest observations from pseudo-random samples of size 40'000 from the logistic, Hüsler–Reiss, Dirichlet, and asymmetric logistic models with Fréchet margins. The solid orange lines show the spectral density function and orange dots show point masses of the spectral distribution. The solid green lines show the spectral density fitted by maximum likelihood.*

Reiss, Dirichlet, and asymmetric logistic distributions with Fréchet margins based on pseudo-random samples. Each kernel density estimator is based on $n = 1'000$ observed pseudo-angles of 'large' observations for several choices of high thresholds using a beta kernel with bandwidth of $n^{-1/2} \approx 0.0316$, which lead to kernel estimates with a good trade-off between smoothness and adaptability to the data. These estimators are unlikely to be valid spectral distributions, as there is no constraint in the estimation to ensure that they satisfy the mean constraint (2.1). The distribution with the most striking difference between the spectral distribution and the kernel density estimates is the asymmetric logistic, due to the point masses at 0 and 1. But despite the large number of observations used in the estimation, differences, though less blatant, are also present for the three other distributions, which don't have point masses. These differences are not exclusively due to sampling effects but are repeatedly present in simulations and simply illustrate the fact that the angular distribution of 'large' observations can be quite different from the spectral distribution, though the former converges to the latter as the threshold used to select 'large' observations tends to infinity. Obviously, the choice of the threshold level entails a bias-variance trade-off, but the bias might be much larger than intuition may suggest, even in ideal (and maybe unrealistic) situations where the sample size is very large, allowing one to choose a quite high threshold z. However, in applications where only hundreds or a few thousands of observations are available, one might be forced to use lower thresholds in order to have a reasonable number of data points for inference.

To summarize, the above examples show that the distribution of pseudo-angles of 'large' observations can differ significantly from the spectral distribution when the latter is asymmetric, has point masses at 0 or 1, or has weak extremal dependence, which can then yield significantly biased estimates.

2.3 Angular distribution of excesses

In this section, we derive a characterization of the angular distribution of excesses and illustrate it with some examples.

The angular distribution of excesses for $z > 0$ corresponds to the distribution of pseudo-angles of observations larger than z, that is the distribution of $W \mid R > z$. Let $\mathcal{H}_z$ denote the angular distribution of excesses associated to the spectral distribution $\mathcal{H}$ for some $z > 0$. Let H_z denote its CDF and h_z its generalized PDF. Recall that the spectral distribution $\mathcal{H}$ has support $[0,1]$, while the angular distribution of excesses $\mathcal{H}_z$ has support $(0,1)$, since $W = X/(X+Y)$ and $X, Y > 0$.

2.3.1 Characterisation of the angular distribution of excesses

The following result states that a bivariate extreme value distribution $\mathcal{G}$ has a PDF if and only if its associated spectral distribution $\mathcal{H}$ has no point mass in $(0,1)$.

Proposition 2.1. *A distribution $\mathcal{G} \in \mathbb{G}_2$ if and only if its associated spectral distribution $\mathcal{H} \in \mathbb{H}_2$.*

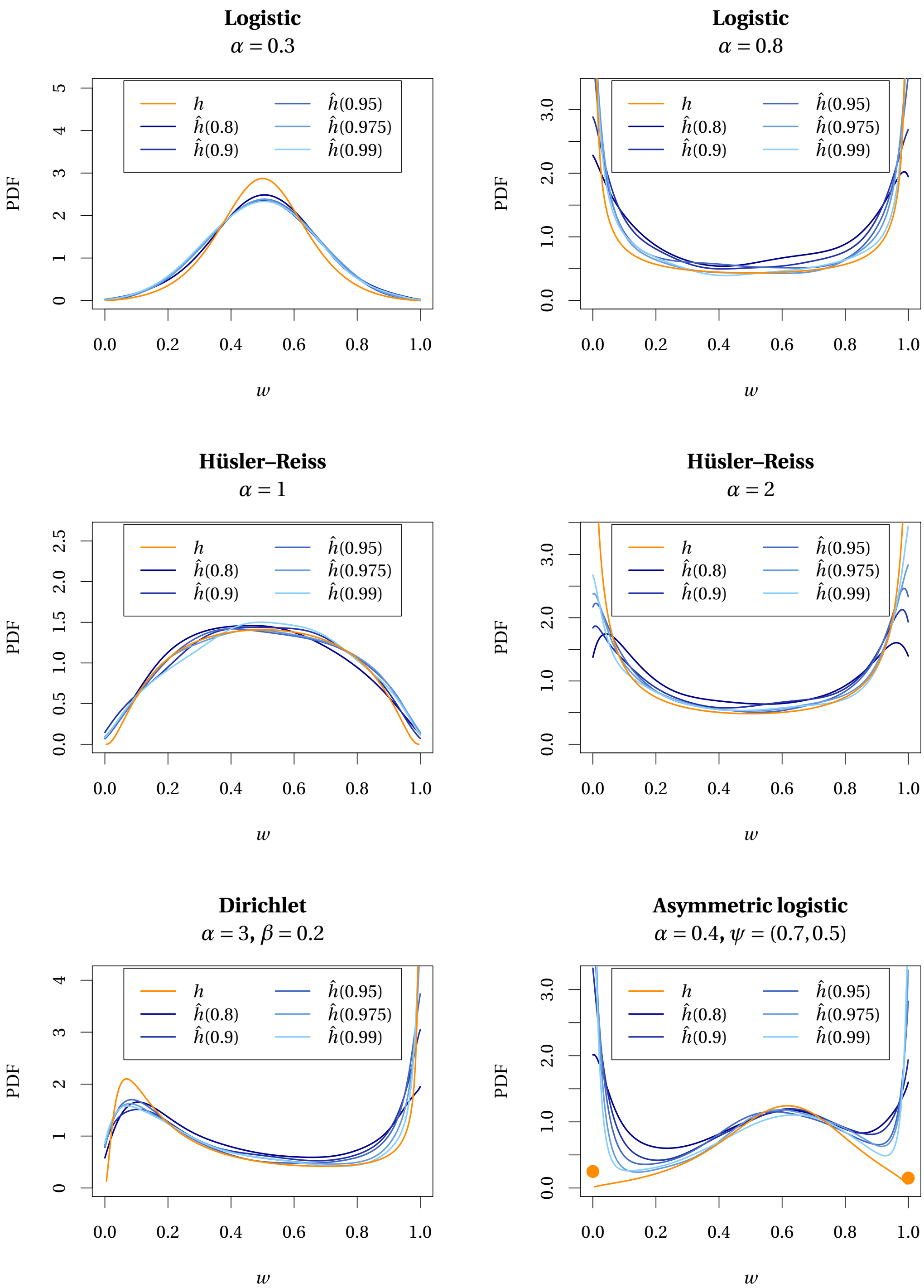

Figure 2.3 – *Kernel density estimates of the spectral density h from the logistic, Hüsler–Reiss, Dirichlet, and asymmetric logistic models with Fréchet margins based on the largest* $n = 1'000$ *observations from pseudo-random samples of size* $N = n/(1-p)$ *for* $p = 0.8, 0.9, 0.95, 0.975, 0.99$.

Proof. Suppose that $\mathcal{H} \in \mathbb{H}_2$. By definition $\mathcal{H}$ has no point mass in $(0,1)$. Denote potential point masses at 0 and 1 by $\delta_0 = H(\{0\})$ and $\delta_1 = H(\{1\})$ respectively. Then the exponent function is

$$\begin{aligned}
V(x,y) &= 2\int_{t\in[0,1]} \max\left(\frac{t}{x}, \frac{1-t}{y}\right) \mathrm{d}H(t) \\
&= 2\left(\frac{\delta_0}{y} + \frac{\delta_1}{x}\right) + 2\int_{t\in(0,1)} \max\left(\frac{t}{x}, \frac{1-t}{y}\right) \mathrm{d}H(t) \\
&= 2\left(\frac{\delta_0}{y} + \frac{\delta_1}{x}\right) + 2\int_0^1 \max\left(\frac{t}{x}, \frac{1-t}{y}\right) \tilde{h}(t)\,\mathrm{d}t \\
&= 2\left(\frac{\delta_0}{y} + \frac{\delta_1}{x}\right) + 2\int_0^{x/(x+y)} \frac{t}{x}\tilde{h}(t)\,\mathrm{d}t + 2\int_{x/(x+y)}^1 \frac{1-t}{y}\tilde{h}(t)\,\mathrm{d}t, \quad x,y>0.
\end{aligned}$$

Clearly, the exponent function is differentiable for all $x,y>0$, and so is the CDF $G(x,y) = \exp\{-V(x,y)\}$, $x,y>0$. Thus, $\mathcal{G} \in \mathbb{G}_2$.

Now, suppose that $\mathcal{H} \notin \mathbb{H}_2$, so $\mathcal{H}$ has at least one point mass in $(0,1)$ and potential point masses at 0 and 1. Let $a_1,\ldots,a_n$ be the locations of the point masses in $(0,1)$, and denote the generalized density function of $\mathcal{H}$ by

$$h(t) = H(\{0\})\delta(t) + H(\{1\})\delta(t-1) + \sum_{i=1}^n H(\{a_i\})\delta(t-a_i) + \tilde{b}(t), \quad 0 \le t \le 1,$$

where $\delta(\cdot)$ denotes the Dirac delta function, and $\tilde{b}(t) = \mathrm{d}/\mathrm{d}tH(t)$ on $(0,1)\setminus\{a_1,\ldots,a_n\}$ and is extended by continuity at $0, a_1,\ldots,a_n$, and 1. Then the exponent function is

$$\begin{aligned}
V(x,y) &= 2\int_{t\in(0,1)} \max\left(\frac{t}{x}, \frac{1-t}{y}\right) \mathrm{d}H(t) \\
&= 2\left(\frac{H(\{0\})}{y} + \frac{H(\{1\})}{x}\right) + 2\sum_{i=1}^n H(\{a_i\}) \max\left(\frac{a_i}{x}, \frac{1-a_i}{y}\right) \\
&\quad + 2\left\{\frac{1}{y}\int_0^{x/(x+y)} (1-t)h(t)\,dt + \frac{1}{x}\int_{x/(x+y)}^1 th(t)\,dt\right\}.
\end{aligned}$$

Clearly, the exponent function is not differentiable everywhere for $x,y>0$ due to the terms with maxima. In fact, the set where V is not differentiable is $\{(x,y): x/(x+y) \in \{a_1,\ldots,a_n\}\}$, since V is not differentiable when $a_i/x = (1-a_i)/y$ for $i = 1,\ldots,n$. So, the CDF $G(x,y) = \exp\{-V(x,y)\}$, $x,y>0$ is not differentiable (everywhere). Thus, $\mathcal{G} \notin \mathbb{G}_2$. □

In the following and unless stated otherwise, spectral distributions $\mathcal{H}$ are assumed to be 01-continuous. This set of spectral distributions is quite broad, since by Proposition 2.1 the set of 01-continuous spectral distributions corresponds to the set of all bivariate extreme value distributions $\mathcal{G}$ having a differentiable CDF.

The following result provides a characterization of the angular distribution of excesses $\mathcal{H}_z$, that is the marginal distribution of the pseudo-angle W conditioned on the event that the pseudo-radius R is larger than some finite threshold $z > 0$.

Theorem 2.1 (Angular distribution of excesses). *Let (X, Y) be a bivariate random vector having distribution $\mathcal{G} \in \mathbb{G}_2$, and let $\mathcal{H} \in \mathbb{H}_2$ be the spectral distribution associated with $\mathcal{G}$ having generalized density function*

$$h(t) = \delta_0 \delta(t) + \tilde{h}(t) + \delta_1 \delta(t-1), \quad 0 \le t \le 1,$$

as in equation (2.2).

Then the density function of the angular distribution $\mathcal{H}_z$ of excesses for a finite $z > 0$ is

$$h_z(w) = \frac{\tilde{f}_z(w)}{\int_0^1 \tilde{f}_z(t)\,\mathrm{d}t}, \quad 0 < w < 1,$$

where

$$\tilde{f}_z(w) = \frac{2\tilde{h}(w)}{B(w)} \left\{1 - e^{-B(w)/z}\right\} + \frac{4B_0(w)B_1(w)}{B(w)^2 w^2 (1-w)^2} \left\{1 - e^{-B(w)/z} - \frac{B(w)}{z} e^{-B(w)/z}\right\}, \tag{2.5}$$

and

$$B_0(w) = \delta_0 + \int_0^w (1-t)\tilde{h}(t)\,\mathrm{d}t, \tag{2.6}$$

$$B_1(w) = \delta_1 + \int_w^1 t\tilde{h}(t)\,\mathrm{d}t, \tag{2.7}$$

$$B(w) = 2\left\{\frac{B_0(w)}{1-w} + \frac{B_1(w)}{w}\right\}. \tag{2.8}$$

Proof. As usual, let $R = X + Y$ and $W = X/R$ denote the pseudo-radius and the pseudo-angle. First, we determine the joint density of (R, W), and then we compute h_z, the marginal density of W conditioned on the event $R > z$.

From the spectral representation theorem (Theorem 1.9), the joint distribution function of (X, Y) can be written as

$$G(x, y) = \exp\left\{-V(x, y)\right\}, \quad x, y > 0,$$

where

$$V(x, y) = 2\int_0^1 \max\left(\frac{t}{x}, \frac{1-t}{y}\right) \mathrm{d}H(t).$$

Here $t/x < (1-t)/y$ when $t < x/(x+y)$, so the exponent measure can be written as

$$
\begin{aligned}
V(x,y) &= 2\left\{\int_0^{x/(x+y)} \frac{1-t}{y}\,\mathrm{d}H(t) + \int_{x/(x+y)}^1 \frac{t}{x}\,\mathrm{d}H(t)\right\} \\
&= 2\left[\frac{1}{y}\left\{\delta_0 + \int_0^{x/(x+y)} (1-t)\tilde{h}(t)\,\mathrm{d}t\right\} + \frac{1}{x}\left\{\delta_1 + \int_{x/(x+y)}^1 t\tilde{h}(t)\,\mathrm{d}t\right\}\right].
\end{aligned}
$$

The CDF G is differentiable since $\mathcal{G} \in \mathbb{G}_2$. The joint PDF of (X,Y) is

$$
g(x,y) = \left\{V_x(x,y)V_y(x,y) - V_{xy}(x,y)\right\}\exp\left\{-V(x,y)\right\}, \quad x,y>0,
$$

where

$$
\begin{aligned}
V_x(x,y) &= \frac{\partial}{\partial x}V(x,y) = -\frac{2}{x^2}\left\{\delta_1 + \int_{x/(x+y)}^1 t\tilde{h}(t)\,\mathrm{d}t\right\}, \\
V_y(x,y) &= \frac{\partial}{\partial y}V(x,y) = -\frac{2}{y^2}\left\{\delta_0 + \int_0^{x/(x+y)} (1-t)\tilde{h}(t)\,\mathrm{d}t\right\}, \\
V_{xy}(x,y) &= \frac{\partial^2}{\partial x\partial y}V(x,y) = -\frac{2}{(x+y)^3}\tilde{h}\left(\frac{x}{x+y}\right).
\end{aligned}
$$

Derivations of expressions for $V_x(x,y)$, $V_y(x,y)$, and $V_{x,y}(x,y)$ are provided in Appendix A.1.

Consider the change of variables $r = u_1(x,y) = x+y$, and $w = u_2(x,y) = x/(x+y)$, which has Jacobian $J(x,y) = (x+y)^{-1}$. The corresponding inverse transformations are $x = v_1(r,w) = rw$, and $y = v_2(r,w) = r(1-w)$. Thus, the joint density function of (R,W) is

$$
\begin{aligned}
f_{R,W}(r,w) = {}& g_{X,Y}(x,y) \times \left|J(x,y)\right|^{-1}\Big|_{x=rw,\,y=r(1-w)} \\
= {}& \left[\frac{2}{x^2}\left\{\delta_1 + \int_{x/(x+y)}^1 t\tilde{h}(t)\,\mathrm{d}t\right\} \times \frac{2}{y^2}\left\{\delta_0 + \int_0^{x/(x+y)} (1-t)\tilde{h}(t)\,\mathrm{d}t\right\} + \frac{2}{(x+y)^3}\tilde{h}\left(\frac{x}{x+y}\right)\right] \\
& \times \exp\left\{-2\int_0^1 \max\left(\frac{t}{x},\frac{1-t}{y}\right)\mathrm{d}H(t)\right\} \times (x+y)\Big|_{x=rw,\,y=r(1-w)} \\
= {}& \left[\frac{4}{r^4w^2(1-w)^2}\left\{\delta_1 + \int_w^1 t\tilde{h}(t)\,\mathrm{d}t\right\} \times \left\{\delta_0 + \int_0^w (1-t)\tilde{h}(t)\,\mathrm{d}t\right\} + \frac{2\tilde{h}(w)}{r^3}\right] \\
& \times \exp\left\{-\frac{2}{r}\int_0^1 \max\left(\frac{t}{w},\frac{1-t}{1-w}\right)\mathrm{d}H(t)\right\} \times r \\
= {}& \left[\frac{4}{r^4w^2(1-w)^2}\left\{\delta_1 + \int_w^1 t\tilde{h}(t)\,\mathrm{d}t\right\} \times \left\{\delta_0 + \int_0^w (1-t)\tilde{h}(t)\,\mathrm{d}t\right\} + \frac{2\tilde{h}(w)}{r^3}\right] \\
& \times \exp\left[-\frac{2}{r}\left\{\frac{\delta_0}{1-w} + \int_0^w \frac{1-t}{1-w}\tilde{h}(t)\,\mathrm{d}t + \frac{\delta_1}{w} + \int_w^1 \frac{t}{w}\tilde{h}(t)\,\mathrm{d}t\right\}\right], \quad r>0,\, 0<w<1.
\end{aligned}
$$

Let $B_0(w) = \delta_0 + \int_0^w (1-t)\tilde{h}(t)\,\mathrm{d}t$, $B_1(w) = \delta_1 + \int_w^1 t\tilde{h}(t)\,\mathrm{d}t$, and $B(w) = 2\{B_0(w)/(1-w) +$

$B_1(w)/w\}$. Then,

$$\begin{aligned} f_{R,W}(r,w) &= \left\{\frac{2\tilde{h}(w)}{r^2} + \frac{4B_0(w)B_1(w)}{r^3 w^2 (1-w)^2}\right\} \exp\left[-\frac{2}{r}\left\{\frac{B_0(w)}{1-w} + \frac{B_1(w)}{w}\right\}\right] \\ &= \left\{\frac{2h(w)}{r^2} + \frac{4B_0(w)B_1(w)}{r^3 w^2 (1-w)^2}\right\} \exp\left\{-\frac{B(w)}{r}\right\}, \quad r>0, 0<w<1. \end{aligned}$$

For a finite value $z>0$, the density function for the distribution of excesses corresponding to the marginal density of W conditioned on the event $R>z$ is

$$h_z(w) = f_{W|R}(w \mid R>z) = \frac{\tilde{f}_z(w)}{\int_0^1 \tilde{f}_z(t)\,\mathrm{d}t}, \quad 0<w<1,$$

where $\tilde{f}_z(w) = \int_z^\infty f_{R,W}(r,w)\,\mathrm{d}r$.

Recall that the PDF and the CDF of the inverse gamma distribution with shape parameter $\alpha>0$ and rate parameter $\beta>0$ are respectively

$$f(x;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1}\exp\left(-\frac{\beta}{x}\right), \quad F(x;\alpha,\beta) = \frac{\Gamma\left(\alpha, \frac{\beta}{x}\right)}{\Gamma(\alpha)}, \quad x>0,$$

where $\Gamma(\cdot)$ and $\Gamma(\cdot,\cdot)$ respectively denote the gamma function and the upper incomplete gamma function.

The joint density function $f_{R,W}(r,w)$ can be expressed as the sum of two kernels of the inverse gamma density function. Hence,

$$\begin{aligned} \tilde{f}_z(w) &= \left[\frac{2\tilde{h}(w)}{B(w)}\Gamma\{1, B(w)/r\} + \frac{4B_0(w)B_1(w)}{B(w)^2 w^2(1-w)^2}\Gamma\{2, B(w)/r\}\right]_z^\infty \\ &= \frac{2\tilde{h}(w)}{B(w)}\gamma\{1, B(w)/z\} + \frac{4B_0(w)B_1(w)}{B(w)^2 w^2(1-w)^2}\gamma\{2, B(w)/z\} \\ &= \frac{2\tilde{h}(w)}{B(w)}\left\{1-e^{-B(w)/z}\right\} + \frac{4B_0(w)B_1(w)}{B(w)^2 w^2(1-w)^2}\left\{1-e^{-B(w)/z} - \frac{B(w)}{z}e^{-B(w)/z}\right\}, \end{aligned}$$

where $\gamma(\cdot,\cdot)$ denotes the lower incomplete gamma function. □

The PDF h_z is the theoretical PDF of pseudo-angles of observations having their pseudo-radius exceeding the threshold $z>0$ when the observations are i.i.d. and come from a bivariate extreme-value distribution; there is no asymptotic approximation.

Recall that the support of the spectral distribution $\mathcal{H}$ is $[0,1]$, whilst the support of the angular distribution of excesses $\mathcal{H}_z$ is $(0,1)$. The following result shows that the density function h_z converges pointwise and up to a factor to the spectral density h when $z\to\infty$.

Proposition 2.2. *Under the assumptions of Theorem 2.1,*

$$\lim_{z\to\infty} h_z(w) = \frac{\tilde{h}(w)}{1-\delta_0-\delta_1}, \quad 0 < w < 1.$$

Proof. A second-order expansion of $\tilde{f}_z(w)$ (see equation (2.5)) for $w \in (0,1)$ when $z > 0$ is large yields

$$\begin{aligned}
\tilde{f}_z(w) &= \frac{2\tilde{h}(w)}{B(w)}\left[1-\left\{1-\frac{B(w)}{z}+\frac{B(w)^2}{2z^2}+O\left(z^{-3}\right)\right\}\right]+\frac{4A(w)B(w)}{B(w)^2w^2(1-w)^2}\\
&\quad\times\left[1-\left\{1-\frac{B(w)}{z}+\frac{B(w)^2}{2z^2}+O\left(z^{-3}\right)\right\}-\frac{B(w)}{z}\left\{1-\frac{B(w)}{z}+O\left(z^{-2}\right)\right\}\right]\\
&= \frac{1}{z}2\tilde{h}(w)+\frac{1}{z^2}\left\{-B(w)\tilde{h}(w)+\frac{2A(w)B(w)}{w^2(1-w)^2}\right\}+O\left(z^{-3}\right),
\end{aligned}$$

and thus

$$\lim_{z\to\infty} h_z(w) = \lim_{z\to\infty}\frac{\tilde{h}(w)+\frac{1}{2z}\left\{-B(w)\tilde{h}(w)+\frac{2A(w)B(w)}{w^2(1-w)^2}\right\}+O\left(z^{-2}\right)}{\int_0^1\left[\tilde{h}(t)+\frac{1}{2z}\left\{-B(t)\tilde{h}(t)+\frac{2A(t)B(t)}{t^2(1-t)^2}\right\}+O\left(z^{-2}\right)\right]\mathrm{d}t}.$$

Noting that the integrand converges uniformly to $\tilde{h}(t)$ as $z\to\infty$, one can interchange the limit and the integral. Thus,

$$\lim_{z\to\infty} h_z(w) = \frac{\tilde{h}(w)}{\int_0^1\tilde{h}(t)\,\mathrm{d}t} = \frac{\tilde{h}(w)}{1-\delta_0-\delta_1}, \quad 0 < w < 1.$$

□

Proposition 2.2 shows that in the limit the 'shape' of the PDF h_z is the same as the 'shape' of the PDF h on the open interval $(0,1)$, and the two are equivalent if and only if the spectral distribution doesn't have point masses. But if the spectral distribution has point masses at 0 or 1, then the two PDFs are proportional up to a factor $1-\delta_0-\delta_1$. The consequence of this for inference is that there is no hope recovering the value of spectral distribution's point masses at 0 and 1 directly from the estimated spectral distribution on $(0,1)$ and without additional assumptions, since this would imply solving a system with two unknowns (δ_0 and δ_1) but only one equation (the mean constraint).

2.3.2 Properties of B_0 and B_1

The functions B_0 and B_1 are defined in terms of integrals involving the generalized density of the spectral distribution $\mathcal{H}$. In applications the computation of these functions might require numerical integration, which can be slow and unstable if the (generalized) spectral density explodes near 0 or 1, as is the case for example for the logistic distribution with dependence parameter $\theta > 0.5$. The following result provides expressions that link B_0, B_1, and B with the Pickands dependence function and its right-hand derivative, as well as the CDF of the spectral distribution. These expressions are particularly useful in numerical computations when the Pickands dependence function is available in an analytic form, which is the case for many common parametric models.

Lemma 2.1. *Let functions B_1, B_2, and B be as defined in equations* (2.6) *to* (2.8)*. Then*

$$\begin{aligned} B_0(w) &= \frac{1}{2}\left\{A(w) + (1-w)A'(w)\right\}, \quad 0 < w < 1, \\ B_1(w) &= \frac{1}{2}\left\{A(w) - wA'(w)\right\}, \quad 0 < w < 1, \\ B(w) &= \frac{A(w)}{w(1-w)}, \quad 0 < w < 1, \\ H([0,w]) &= \frac{1}{2} + B_0(w) - B_1(w), \quad 0 < w < 1, \end{aligned}$$

where A denotes the Pickands dependence function and A' is the right-hand derivative of A.

Proof. First, note that equation (1.14) gives

$$H((0,w)) = H([0,w]) - \delta_0 = \frac{1}{2}\left\{1 + A'(w)\right\} - \delta_0, \quad 0 \le w < 1. \tag{2.9}$$

By the definitions of B_0 (equation (2.6)) and H, we have

$$B_0(w) = \delta_0 + \int_0^w (1-t)\tilde{h}(t)\,\mathrm{d}t = H([0,w]) - \int_0^w t\tilde{h}(t)\,\mathrm{d}t, \quad 0 < w < 1.$$

Integrating by parts and using equation (2.9) yields

$$\begin{aligned} \int_0^w t\tilde{h}(t)\,\mathrm{d}t &= wH((0,w)) - \int_0^w H((0,t))\,\mathrm{d}t \\ &= \frac{1}{2}w\left\{1 + A'(w)\right\} - \delta_0 w - \int_0^w \left[\frac{1}{2}\left\{1 + A'(t)\right\} - \delta_0\right]\mathrm{d}t \\ &= \frac{1}{2}\left\{wA'(w) - A(w) + 1\right\}. \end{aligned}$$

Using equation (1.14) yields

$$\begin{aligned} B_0(w) &= \frac{1}{2}\left\{1 + A'(w)\right\} - \frac{1}{2}\left\{wA'(w) - A(w) + 1\right\} \\ &= \frac{1}{2}\left\{A(w) + (1-w)A'(w)\right\}, \quad 0 < w < 1. \end{aligned} \tag{2.10}$$

From the definition of B_1 (equation (2.7)) and integrating by parts, we have

$$
\begin{aligned}
B_1(w) &= \delta_1 + \int_w^1 t\tilde{h}(t)\,\mathrm{d}t \\
&= \delta_1 + H((0,1)) - wH((0,w)) - \int_w^1 H((0,t))\,\mathrm{d}t, \quad 0 < w < 1.
\end{aligned}
$$

Since $H((0,1)) = 1 - \delta_0 - \delta_1$, and using equation (2.9), we have

$$
\begin{aligned}
B_1(w) &= 1 - \delta_0 - \frac{1}{2}w\left\{1 + A'(w)\right\} + \delta_0 w - \int_w^1 \left[\frac{1}{2}\left\{1 + A'(t)\right\} - \delta_0\right]\mathrm{d}t \\
&= 1 - \frac{1}{2}w\left\{1 + A'(w)\right\} - \frac{1}{2}\{1 + A(1) - w - A(w)\} \\
&= \frac{1}{2}\left\{A(w) - wA'(w)\right\}, \quad 0 < w < 1. \qquad (2.11)
\end{aligned}
$$

By the definition of B (equation (2.8)), and using equations (2.10) and (2.11), we have

$$
\begin{aligned}
B(w) &= 2\left\{\frac{B_0(w)}{1-w} + \frac{B_1(w)}{w}\right\} = \frac{A(w) + (1-w)A'(w)}{1-w} + \frac{A(w) - wA'(w)}{w} \\
&= \frac{A(w)}{w(1-w)}, \quad 0 < w < 1.
\end{aligned}
$$

Consider the expression for H. Using equations (1.14), (2.10) and (2.11), we have

$$
\frac{1}{2} + B_0(w) - B_1(w) = \frac{1}{2}\left\{1 + A'(w)\right\} = H([0,w]), \quad 0 < w < 1.
$$

□

2.3.3 Examples

2.3.3.1 Independence

Recall that the spectral distribution associated with indepencence has point masses $\delta_0 = \delta_1 = 1/2$, and no mass on the interior of $(0,1)$, that is, $\tilde{h}(w) \equiv 0$. So, the generalized PDF of the spectral distribution is

$$
h(w) = \frac{1}{2}\{\delta(w) + \delta(w-1)\}, \quad 0 \le w \le 1.
$$

From Theorem 2.1, we have

$$
B_0(w) = B_1(w) = \frac{1}{2}, \quad B(w) = \frac{1}{w(1-w)}, \quad 0 < w < 1,
$$

and

$$\tilde{f}_z(w) = 1 - e^{-B(w)/z} - \frac{B(w)}{z} e^{-B(w)/z} = \gamma\left\{2, \frac{B(w)}{z}\right\}, \quad 0 < w < 1,$$

where $\gamma(\cdot,\cdot)$ denotes the lower incomplete Gamma function, so the PDF of the angular distribution of excesses is

$$h_z(w) = \frac{\gamma\left[2, \{zw(1-w)\}^{-1}\right]}{\int_0^1 \gamma\left[2, \{zt(1-t)\}^{-1}\right] \mathrm{d}t}, \quad 0 < w < 1. \tag{2.12}$$

Figure 2.4 shows the theoretical angular distribution of excesses in equation (2.12) for threshold levels corresponding to the approximate p-quantiles of R with $p = 0.5, 0.75, 0.9, 0.95$, and histograms of pseudo-angles W of the largest $n = 5'000$ observations from bivariate pseudo-random samples of independent unit Fréchet random variables of sizes $N = n/(1-p)$. These histograms and the theoretical PDFs show how the point masses of the spectral distribution spread mainly near the edges of the support for high thresholds z to the entire support as the threshold lowers.

2.3.3.2 Some parametric models

Theorem 2.1 allows us to derive the expression for the PDF of the angular distribution of excesses from the spectral distribution. For most parametric models the expression for h_z is intractable and involves an integral which needs to be computed numerically. Figure 2.5 shows the PDF of the spectral distribution h and the theoretical angular distribution of excesses h_z at threshold levels z corresponding to the approximate p-quantiles of the pseudo-radius R for $p = 0.8, 0.9, 0.95, 0.975, 0.99$ for the logistic, Hüsler–Reiss, Dirichlet, and asymmetric logistic models. The density of the angular distribution of excesses of the logistic and Hüsler–Reiss models appears to be quite close to the spectral density when the extremal dependence of the model is somewhat strong, that is when the PDF has a bell shape ($\alpha < 0.5$ for the logistic model, and, say, $\alpha < 1.1$ for the Hüsler–Reiss model). In this case, direct inference on the spectral distribution using pseudo-angles of 'large' observations instead of using the angular distribution of excesses is likely to yield nearly unbiased estimates even with low thresholds. In the other cases, the angular distribution of excesses increasingly differs from the spectral distribution as the threshold level lowers, which can yield strongly biased inference even with high thresholds, as already illustrated in Section 2.2.

2.4 Parametric inference using the angular distribution of excesses

In this section, we describe how to perform maximum likelihood inference on the spectral distribution using the angular distribution of excesses, and we compare this approach with two classical parametric approaches.

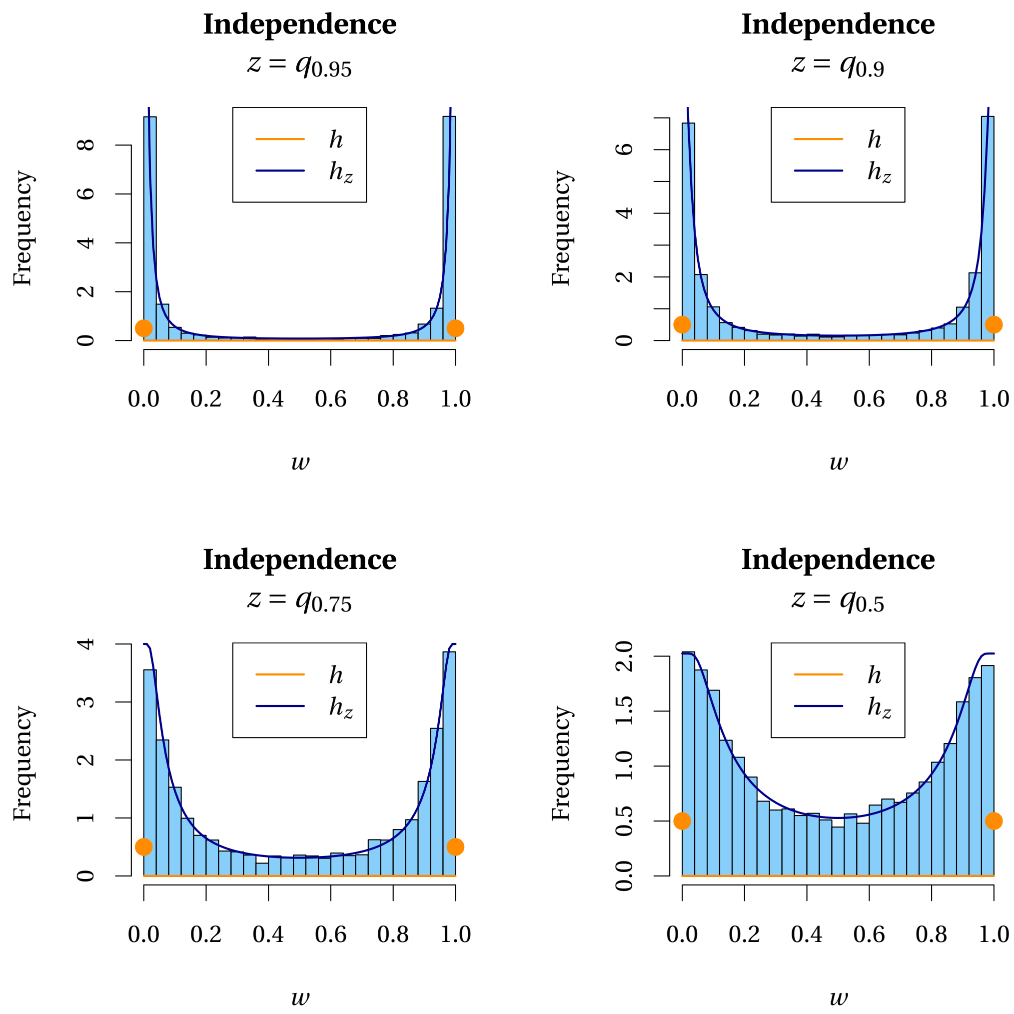

Figure 2.4 – *Histograms of the pseudo-angle W of the largest $n = 5'000$ observations from bivariate pseudo-random samples of independent unit Fréchet random variables of size $N = n/(1-p)$ for $p = 0.5, 0.75, 0.9, 0.95$. The solid blue lines show the PDF of the theoretical angular distribution of excesses for threshold levels z approximately equal to the p-quantiles of the pseudo-radius R. The solid orange lines show the generalized density function of the spectral distribution corresponding to independence with point masses shown as dots.*

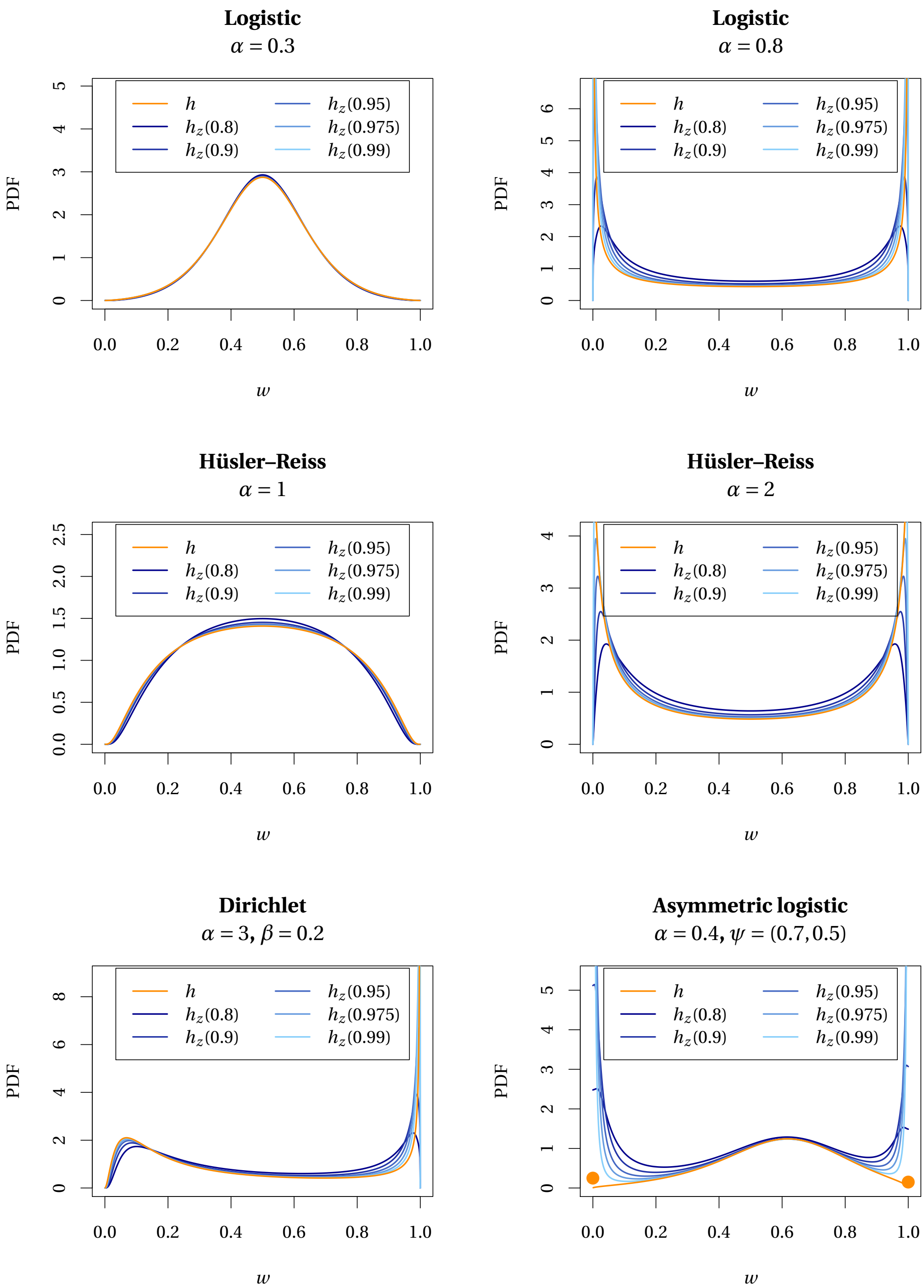

Figure 2.5 – *PDF of the spectral distribution h and the angular distribution of excesses h_z at threshold levels z corresponding to the approximate p-quantiles of the pseudo-radius R for $p = 0.8, 0.9, 0.95, 0.975, 0.99$ for the logistic, Hüsler–Reiss, Dirichlet, and asymmetric logistic models.*

As illustrated in Section 2.2, parametric inference on the spectral distribution can be strongly biased when done directly with the PDF of the spectral distribution using observed pseudo-angles of ‘large’ observations. This bias is caused by the fact that the angular distribution of excesses, which is the distribution corresponding to the observed pseudo-angles of ‘large’ observations, can differ significantly from the spectral distribution, especially for moderately low thresholds sometimes required in applications due to the scarcity of available data. So, in order to perform unbiased or at least bias-reduced inference, it seems natural to use the angular distribution of excesses associated with the spectral distribution.

Let $\mathcal{G}_{\boldsymbol{\theta}}$ be a parametric family of bivariate extreme value distribution with Fréchet margins indexed by a parameter $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$. Let $\mathcal{H}_{\boldsymbol{\theta}}$ be the spectral distribution associated with $\mathcal{G}_{\boldsymbol{\theta}}$. Let $H(\cdot;\boldsymbol{\theta})$ and $h(\cdot;\boldsymbol{\theta})$ respectively denote the CDF and the PDF of the spectral distribution, and let $h_z(\cdot;\boldsymbol{\theta})$ denote the PDF of the angular distribution of excesses associated with $\mathcal{H}_{\boldsymbol{\theta}}$ for a threshold $z > 0$ as in Theorem 2.1.

Let $(x_1, y_1), \ldots, (x_N, y_N)$ be a sample from $\mathcal{G}_{\boldsymbol{\theta}}$. Expressed in pseudo-polar coordinates, these observations are (r_i, w_i) where $r_i = x_i + y_i$ and $w_i = x_i / r_i$ for $i = 1, \ldots, N$. Let $z > 0$ be a large threshold, and let n be the number of ‘large’ observations such that $r_i > z$. Without loss of generality and to simplify notation, suppose that the observations are ordered in decreasing order of pseudo-radius, that is $r_1 \geq \cdots \geq r_N$. So, the set of pseudo-angles for ‘large’ observations is $\{w_1, \ldots, w_n\}$.

Maximum likelihood estimation of the spectral distribution using the angular distribution of excesses is straightforward, at least conceptually, since h_z is expressed in terms of the spectral density h (see Theorem 2.1). The likelihood function is based on the PDF of the angular distribution of excesses h_z, and the corresponding maximum likelihood estimate of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}}_{\mathrm{p}} = \underset{\boldsymbol{\theta} \in \Theta}{\arg\max} \sum_{j=1}^{n} \log h_z(w_j; \boldsymbol{\theta}).$$

In the numerical simulations presented below we compare the ‘asymptotic’, ‘penultimate’, and ‘censored’ approaches to maximum likelihood inference on the spectral distribution. The ‘asymptotic’ approach corresponds to the natural approach described in Section 2.2 where maximum likelihood inference is done with the PDF of the spectral distribution h using the pseudo-angles of the n observations with pseudo-radius larger than a given threshold z, that is $\{w_1, \ldots, w_n\}$. The corresponding maximum likelihood estimate of $\boldsymbol{\theta}$ is given in Equation (2.4). The ‘penultimate’ approach corresponds to maximum likelihood inference based on the angular distribution of excesses described above using $\{w_1, \ldots, w_n\}$. The ‘censored’ approach corresponds to the bivariate peaks over threshold method described in Section 1.2.2. The thresholds z_x and z_y have been chosen to be equal and such that there are n observations where at least one of the coordinates is not censored in the likelihood, that is $\sum_{j=1}^{N} I(x_j < z_x) I(y_j < z_y) = N - n$. In the ‘asymptotic’ and ‘penultimate’ approaches, the likelihood is based on only the pseudo-angles of the n largest observations, whilst in the 'censored' approach the likelihood is based on the N observations in Euclidean coordinates, of which n have at most one censored coordinate in the likelihood.

These approaches are compared for several parametric models using the mean integrated squared error

$$\mathrm{MISE}(\cdot; H, \boldsymbol{\theta}) = \mathrm{E}\left[\int_0^1 \{H(w;\cdot) - H(w;\boldsymbol{\theta})\}^2 \,\mathrm{d}w\right],$$

which is estimated empirically from a set of B estimates of $\boldsymbol{\theta}$, each obtained from a Monte Carlo pseudo-random sample.

Simulation results for data simulated from max-stable models and a non-max-stable model are respectively presented in Section 2.4.1 and Section 2.4.2. Computations were done using `R` (R Core Team, 2017). The data for the max-stable models were generated using the `rbvevd` function of the `evd` package (Stephenson, 2002), and the `rCopula` of the `copula` package (Hofert *et al.*, 2017) was used for non-max-stable model. The `fbvpot` function of the `evd` was used for fitting in the 'censored' approach.

The implementation of the 'penultimate' approach can be tricky and requires some care. The evaluation of the PDF h_z requires numerical integration of the functions B_0, B_1, and $\tilde{f}_z$, which can be slow and unstable, for example when the spectral distribution explodes to infinity near the edges of the support. The alternative expressions for B_0 and B_1 provided in Lemma 2.1 allow us to avoid some of the numerical integrations. In addition, the numerical optimization of the likelihood function can be slow to converge, depending on the algorithm used. Improving the efficiency of the likelihood optimization requires some testing and potential transformations of the optimization problem. We found that the Brent algorithm of the `optim` function worked best when the model has one parameter, and the Nelder–Mead algorithm worked best when the model had two parameters or more.

2.4.1 Data from max-stable models

This section presents the simulation results to compare the 'asymptotic', 'penultimate', and 'censored' approaches to maximum likelihood inference on the spectral distribution in the case where the data are simulated from max-stable models.

Max-stable models, or more specifically the PDF of their associated spectral distributions, might be symmetric or asymmetric, they might have point masses or no point masses, and they can have various levels of extremal dependence. In order to cover a large spectrum of the possible cases, we used the logistic model (Section 1.2.1.5), the Hüsler–Reiss model (Section 1.2.1.5), the Dirichlet model (Section 1.2.1.5), and the asymmetric logistic model (Section 1.2.1.5), and for each of them we selected three sets of parameters. Table 2.1 lists the selected sets of parameters for each model and provides their point masses, their level

Table 2.1 – *List of max-stable models and parameter sets considered in the simulation study comparing several parametric inference approaches on the spectral distribution. The characteristics or each model are the point masses at 0 (δ_0) and 1 (δ_1), the tail dependence index (χ), and the coefficient of extremal asymmetry (φ). The presence of point masses (PM) in the model is identified with a checkmark. The level of extremal dependence is marked as weak (W) if $\chi \leq 0.35$, strong (S) if $\chi > 0.65$, and moderate (M) otherwise. The level of extremal asymmetry is marked as absent (A) if $\varphi = 0$, strong (G) if $|\varphi| > 0.5$, and moderate (M) otherwise.*

						Extr. Dep.			Extr. Asym.		
Model	$\boldsymbol{\delta_0}$	$\boldsymbol{\delta_1}$	$\boldsymbol{\chi}$	$\boldsymbol{\varphi}$	**PM**	**W**	**M**	**S**	**A**	**M**	**S**
Logistic											
$\boldsymbol{\alpha = 0.3}$	0	0	0.77	0	–	–	–	✓	✓	–	–
$\boldsymbol{\alpha = 0.55}$	0	0	0.54	0	–	–	✓	–	✓	–	–
$\boldsymbol{\alpha = 0.8}$	0	0	0.26	0	–	✓	–	–	✓	–	–
Hüsler–Reiss											
$\boldsymbol{\alpha = 0.5}$	0	0	0.80	0	–	–	–	✓	✓	–	–
$\boldsymbol{\alpha = 1}$	0	0	0.62	0	–	–	✓	–	✓	–	–
$\boldsymbol{\alpha = 2}$	0	0	0.32	0	–	✓	–	–	✓	–	–
Dirichlet											
$\boldsymbol{\alpha = 3, \beta = 3}$	0	0	0.69	0	–	–	–	✓	✓	–	–
$\boldsymbol{\alpha = 3, \beta = 0.2}$	0	0	0.33	-0.29	–	✓	–	–	–	✓	–
$\boldsymbol{\alpha = 0.2, \beta = 0.2}$	0	0	0.20	0	–	✓	–	–	✓	–	–
Asymmetric logistic											
$\boldsymbol{\alpha = 0.2, \psi = (0.8, 0.8)}$	0.10	0.10	0.68	0	✓	–	–	✓	✓	–	–
$\boldsymbol{\alpha = 0.4, \psi = (0.7, 0.5)}$	0.25	0.15	0.39	0.31	✓	–	✓	–	–	✓	–
$\boldsymbol{\alpha = 0.6, \psi = (0.6, 0.2)}$	0.40	0.20	0.14	0.52	✓	✓	–	–	–	–	✓

of extremal dependence as measured by the tail dependence index χ (see Section 1.2.4), and their level of extremal asymmetry as measured by the coefficient of extremal asymmetry φ (see Section 3.2.1).

For each distribution and each parameter set we generated $B = 100$ pseudo-random samples of size $N = n/(1-p)$ for $n = 100, 1000$ and $p = 0.9, 0.95, 0.99$. Then, for each sample we fitted the parametric model corresponding to the model from which the data were generated using the 'asymptotic', 'penultimate', and 'censored' approaches, allowing us to compare their performance at estimating the true underlying parameter values. The mean integrated squared errors of these fits are shown in Tables 2.2 to 2.5. Boxplots of the integrated squared errors of each fit and the parameter estimates are shown in Appendix B.

The results for the logistic model are shown in Table 2.2 and Figures B.1 to B.6. The 'penultimate' approach outperforms the 'censored' approach in all cases and for all values of n and p. The 'penultimate' and 'asymptotic' approaches perform similarly in the first case ($\alpha = 0.3$)

Table 2.2 – *Mean integrated squared error (×1000) of $B = 100$ maximum likelihood estimates of the logistic model (standard deviation in parenthesis) for data simulated from the logistic model for three parameter choices using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n.*

			Logistic		
n	p	**Approach**	$\alpha = 0.3$	$\alpha = 0.55$	$\alpha = 0.8$
100	0.90	A	1.4 (1.7)	3.4 (3.9)	57.3 (14.1)
100	0.90	P	1.5 (1.9)	2.1 (2.6)	4.1 (6.8)
100	0.90	C	5.8 (18.8)	21.5 (80.0)	12.0 (18.8)
100	0.95	A	1.2 (1.8)	2.6 (3.5)	37.6 (11.2)
100	0.95	P	1.3 (1.8)	2.2 (3.2)	2.9 (4.3)
100	0.95	C	9.3 (27.0)	21.9 (76.0)	26.5 (105.6)
100	0.99	A	1.1 (1.6)	2.0 (3.2)	13.7 (5.9)
100	0.99	P	1.1 (1.6)	2.0 (3.0)	1.7 (2.2)
100	0.99	C	22.9 (42.1)	42.0 (67.0)	31.5 (67.0)
1000	0.90	A	0.2 (0.2)	2.9 (1.1)	57.7 (5.0)
1000	0.90	P	0.1 (0.2)	0.2 (0.2)	0.5 (0.7)
1000	0.90	C	0.3 (0.4)	0.7 (1.0)	0.6 (1.0)
1000	0.95	A	0.1 (0.2)	1.3 (0.8)	36.5 (3.5)
1000	0.95	P	0.1 (0.2)	0.2 (0.2)	0.3 (0.4)
1000	0.95	C	0.3 (0.5)	0.7 (0.9)	1.4 (5.4)
1000	0.99	A	0.1 (0.2)	0.3 (0.6)	13.1 (2.1)
1000	0.99	P	0.1 (0.2)	0.2 (0.5)	0.2 (0.3)
1000	0.99	C	10.0 (34.2)	19.1 (87.8)	13.5 (54.1)

where the extremal dependence is strong, whilst the 'penultimate' approach outperforms the 'asymptotic' approach in the second case ($\alpha = 0.55$, mild extremal dependence) when the threshold is low ($p = 0.9$ and $p = 0.95$), and in the third case ($\alpha = 0.8$, low extremal dependence) for all thresholds. The bias of the parameter estimators with the 'penultimate' approach is close to zero for all values of p and n, whilst the bias of the parameter estimators with the 'asymptotic' approach decreases as the threshold level increases. Increasing the sample size improves the performance of all approaches.

The results for the Hüsler–Reiss model are shown in Table 2.3 and Figures B.7 to B.12. Overall, the comparative performances of the three approaches are similar to those for the logistic model. The 'penultimate' approach outperforms the 'censored' approach in all cases and for all values of n and p. The 'penultimate' and 'asymptotic' approaches perform similarly in the first case ($\alpha = 0.5$) where the extremal dependence is strong, whilst the 'penultimate' approach outperforms the 'asymptotic' approach in the second case ($\alpha = 0.1$, mild extremal dependence) when the threshold is low ($p = 0.9$ and $p = 0.95$), and in the third case ($\alpha = 2$,

Table 2.3 – *Mean integrated squared error ($\times 1000$) of $B = 100$ maximum likelihood estimates of the Hüsler–Reiss model (standard deviation in parenthesis) for data simulated from the Hüsler–Reiss model for three parameter choices using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n.*

			Hüsler–Reiss		
n	p	Approach	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 2$
100	0.90	A	0.8 (1.0)	1.6 (2.1)	19.8 (8.6)
100	0.90	P	0.8 (1.0)	1.4 (2.0)	2.5 (4.7)
100	0.90	C	5.9 (17.6)	16.0 (58.8)	16.6 (32.5)
100	0.95	A	0.9 (1.3)	1.4 (1.9)	8.9 (6.0)
100	0.95	P	0.9 (1.3)	1.5 (2.0)	2.1 (2.8)
100	0.95	C	7.0 (19.4)	15.1 (53.3)	26.0 (87.9)
100	0.99	A	0.9 (1.2)	1.5 (2.0)	2.5 (3.4)
100	0.99	P	0.9 (1.2)	1.5 (2.0)	1.7 (2.6)
100	0.99	C	20.1 (33.6)	34.9 (82.5)	53.9 (108.3)
1000	0.90	A	0.1 (0.2)	0.6 (0.6)	18.4 (3.2)
1000	0.90	P	0.1 (0.1)	0.2 (0.3)	0.3 (0.5)
1000	0.90	C	1.1 (8.9)	0.5 (0.6)	0.9 (1.3)
1000	0.95	A	0.1 (0.1)	0.3 (0.4)	8.5 (1.9)
1000	0.95	P	0.1 (0.1)	0.2 (0.2)	0.2 (0.3)
1000	0.95	C	2.1 (12.5)	4.2 (36.6)	1.8 (7.9)
1000	0.99	A	0.1 (0.1)	0.1 (0.2)	1.0 (0.6)
1000	0.99	P	0.1 (0.1)	0.1 (0.2)	0.1 (0.2)
1000	0.99	C	8.4 (25.6)	15.5 (72.1)	8.5 (19.1)

low extremal dependence) for all thresholds. The bias of the parameter estimators with the 'penultimate' approach is close to zero for all values of p and n, whilst the bias of the parameter estimators with the 'asymptotic' approach decreases as the threshold increases (most visibly when n is large). Increasing the sample size improves the performance of all approaches.

The results for the Dirichlet model are shown in Table 2.4 and Figures B.13 to B.21. The 'penultimate' approach outperforms the 'censored' approach in all cases and for all values of n and p. The 'penultimate' and 'asymptotic' approaches perform similarly in the first case ($\alpha = 3$ and $\beta = 3$) where the extremal dependence is strong, whilst the 'penultimate' approach outperforms the 'asymptotic' approach in the second case ($\alpha = 3$ and $\beta = 0.2$, asymmetry) and in the third case ($\alpha = 0.2$ and $\beta = 0.2$, low extremal dependence) for all thresholds. The bias of the parameter estimators with the 'penultimate' approach is close to zero for all values of p and n, whilst the bias of the parameter estimators with the 'asymptotic' approach decreases as the threshold level increases. Increasing the sample size improves the performance of all approaches.

Table 2.4 – *Mean integrated squared error (×1000) of $B = 100$ maximum likelihood estimates of the Dirichlet model (standard deviation in parenthesis) for data simulated from the Dirichlet model for three parameter choices using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n.*

			Dirichlet		
n	p	**Approach**	$\alpha = 3$, $\beta = 3$	$\alpha = 3$, $\beta = 0.2$	$\alpha = 0.2$, $\beta = 0.2$
100	0.90	A	2.1 (2.0)	35.6 (10.9)	87.1 (15.0)
100	0.90	P	2.2 (2.2)	3.1 (4.1)	6.9 (16.1)
100	0.90	C	11.8 (31.6)	32.0 (118.2)	24.4 (27.3)
100	0.95	A	2.2 (2.2)	21.7 (8.5)	56.3 (11.4)
100	0.95	P	2.3 (2.2)	2.7 (3.8)	4.4 (4.3)
100	0.95	C	12.4 (32.0)	31.6 (111.3)	28.9 (34.6)
100	0.99	A	1.9 (1.9)	8.8 (5.2)	23.1 (6.9)
100	0.99	P	1.9 (1.9)	2.0 (2.6)	2.4 (2.5)
100	0.99	C	41.9 (56.2)	71.5 (117.7)	51.7 (66.2)
1000	0.90	A	0.4 (0.4)	34.1 (3.8)	85.2 (5.0)
1000	0.90	P	0.2 (0.2)	0.4 (0.5)	1.7 (10.2)
1000	0.90	C	1.0 (4.8)	1.3 (1.5)	1.6 (1.6)
1000	0.95	A	0.2 (0.3)	21.8 (2.8)	56.8 (4.0)
1000	0.95	P	0.2 (0.2)	0.3 (0.4)	0.5 (0.6)
1000	0.95	C	1.2 (4.6)	2.1 (7.8)	2.2 (4.2)
1000	0.99	A	0.2 (0.2)	8.0 (1.5)	22.2 (2.1)
1000	0.99	P	0.2 (0.2)	0.2 (0.3)	0.2 (0.2)
1000	0.99	C	21.2 (54.3)	28.7 (84.0)	10.2 (16.9)

The results for the asymmetric logistic model are shown in Table 2.5 and Figures B.22 to B.33. The 'penultimate' approach outperforms both the 'asymptotic' and 'censored' approaches in all cases and for all values of n and p. The bias of the parameter estimators with the 'penultimate' approach is close to zero for all values of p and n, except for the dependence parameter in the third case ($\alpha = 0.6$ and $\psi = (0.6, 0.2)$) when $n = 100$ and $p = 0.9, 0.95$ (see Figure B.31). The bias of the dependence parameter estimators with the 'asymptotic' approach increases as the threshold level p increases. The 'asymptotic' approach is unable to recover that the spectral distribution has point masses, as all estimates of the coefficients of asymmetry are (roughly) equal to 1. Increasing the sample size improves the performance of the 'penultimate' and 'censored' approaches, but not for the 'asymptotic' approach, which remains unchanged.

To summarize, we find that:

- the 'penultimate' approach performs better than the 'censored' approach in all cases of the four distributions;

Table 2.5 – *Mean integrated squared error (×1000) of $B = 100$ maximum likelihood estimates of the asymmetric logistic model (standard deviation in parenthesis) for data simulated from the asymmetric logistic model for three parameter choices using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n.*

			Asymmetric logistic		
			$\boldsymbol{\alpha = 0.2}$	$\boldsymbol{\alpha = 0.4}$	$\boldsymbol{\alpha = 0.6}$
$\boldsymbol{n}$	$\boldsymbol{p}$	**Approach**	$\boldsymbol{\psi = (0.8, 0.8)}$	$\boldsymbol{\psi = (0.7, 0.5)}$	$\boldsymbol{\psi = (0.6, 0.2)}$
100	0.90	A	83.7 (22.3)	44.9 (7.2)	144.9 (17.6)
100	0.90	P	6.4 (5.4)	9.1 (8.5)	9.1 (9.2)
100	0.90	C	18.8 (33.9)	26.8 (37.9)	26.7 (27.3)
100	0.95	A	103.4 (30.7)	38.6 (3.7)	107.2 (12.4)
100	0.95	P	6.3 (6.5)	6.1 (5.3)	6.0 (5.2)
100	0.95	C	17.9 (23.8)	35.2 (71.1)	19.4 (13.1)
100	0.99	A	163.0 (38.2)	42.4 (4.7)	56.6 (9.0)
100	0.99	P	5.0 (4.9)	6.2 (6.9)	3.4 (3.4)
100	0.99	C	31.8 (37.5)	55.1 (82.6)	26.1 (27.1)
1000	0.90	A	82.4 (6.8)	43.8 (2.0)	148.2 (6.1)
1000	0.90	P	0.5 (0.6)	0.8 (0.8)	0.8 (0.7)
1000	0.90	C	2.1 (4.8)	2.5 (3.4)	4.4 (4.7)
1000	0.95	A	103.7 (8.0)	37.3 (0.7)	107.5 (4.6)
1000	0.95	P	0.5 (0.4)	0.6 (0.5)	0.5 (0.7)
1000	0.95	C	1.7 (4.9)	2.1 (2.5)	2.9 (5.4)
1000	0.99	A	156.5 (13.5)	41.5 (1.8)	54.7 (2.7)
1000	0.99	P	0.5 (0.5)	0.6 (0.6)	0.3 (0.4)
1000	0.99	C	16.4 (35.7)	28.3 (91.4)	21.5 (87.4)

- the 'penultimate' approach performs better than the 'asymptotic' approach when the spectral distribution has point masses, is asymmetric, or when the extremal dependence is weak, which corresponds to situations where the angular distribution of excesses differs from the spectral distribution for high thresholds;
- the 'penultimate' approach doesn't improve much on the 'asymptotic' approach when the spectral distribution is symmetric with no point mass and extremal dependence is strong, which corresponds to the situation where the angular distribution of excesses doesn't differ much different from the spectral distribution for high thresholds;
- the bias of the parameter estimators with the 'penultimate' approach is close to zero for the three threshold levels and the two numbers of excesses;
- the parameter estimators with the 'asymptotic' approach are consistently biased, though the bias decreases as the threshold increases, except for the dependence parameter of

the asymmetric logistic distribution, whose bias increases as the threshold increases;

- increasing the threshold level (for a fixed number of excesses) improves the performance of both the 'penultimate' and the 'asymptotic' approaches by reducing the variance of the parameter estimates for both approaches and by reducing the bias of parameter estimators of the 'asymptotic' approach (the bias of the 'penultimate' approach is colse to zero), except for the asymmetric logistic model, as mentioned above;
- increasing the threshold level (for a fixed number of excesses) tends to decrease the performance of the 'censored' approach;
- increasing the number of observations (for a fixed threshold level) improves the performance of all approaches except the 'asymptotic' approach with the asymmetric logistic model, where the performance is unchanged.

These findings underscore that using the PDF of the spectral distribution to perform direct inference on the spectral distribution using pseudo-angles of 'large' observations yields poor results when the spectral distribution is asymmetric, when it has point masses, or when the extremal dependence is low. Using the 'penultimate' approach always yields better (or at least similar) results than the 'asymptotic' approach. Another advantage of the 'penultimate' approach is that, unlike the 'asymptotic' approach, the threshold choice doesn't involve a bias-variance trade-off when the data come from a max-stable model, as the angular distribution of excesses is the exact distribution of the pseudo-angles of observations above a threshold, whilst the spectral distribution is the asymptotic distribution of pseudo-angles when the pseudo-radius tends to infinity.

2.4.2 Data from non-max-stable models

In applications, it is unlikely to have data from exactly max-stable models, and the aim of inference is to estimate the max-stable attractor. This section presents simulation results to compare the 'asymptotic', 'penultimate', and 'censored' approaches to maximum likelihood inference on the spectral distribution when the data are simulated from non-max-stable models.

2.4.2.1 Non-max-stable models

The non-max-stable models considered for the simulation study are: (i) the 'Joe model' and (ii) the 'asymmetric Joe model', both having Fréchet margins and their structure of dependence determined by a copula.

Joe model The Joe model is based on the bivariate Joe/B5 copula (Joe, 2015, Sec. 4.7.1)

$$C_{\mathrm{J}}(u,v;\theta) = 1 - \left[1 - \left\{1-(1-u)^{\theta}\right\}\left\{1-(1-v)^{\theta}\right\}\right]^{1/\theta}, \quad 0 \le u, v \le 1,$$

with parameter $\theta \geq 1$, transformed to have Fréchet margins. The copula C_{J} is Archimedean respectively with generator and inverse generator

$$\psi_{\mathrm{J}}(s;\theta) = 1 - \left(1 - e^{-s}\right)^{1/\theta}, \quad s \geq 0, \quad \phi_{\mathrm{J}}(t;\theta) = -\log\left\{1 - (1-t)^{\theta}\right\}, \quad 0 < t \leq 1.$$

Genest and Rivest (1989) showed that the max-stable attractor of an Archimedean copula is also an Archimedean copula and that the Gumbel copula is the only max-stable copula that is Archimedean, so the only possible attractor of the bivariate Joe/B5 copula is the Gumbel copula, and its parameter value can be obtained from Theorem 1.8. We have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\frac{\phi_{\mathrm{J}}(t;\theta)}{\phi_{\mathrm{J}}'(t;\theta)} &= \frac{\mathrm{d}}{\mathrm{d}t}\frac{\{1-(1-t)^{\theta}\}\log\{1-(1-t)^{\theta}\}}{\theta(1-t)^{\theta}} \\ &= 1 + \log\left\{1-(1-t)^{\theta}\right\} + \frac{\theta-1}{\theta}\left\{1-(1-t)^{\theta}\right\}\frac{\log\{1-(1-t)^{\theta}\}}{(1-t)^{\theta}},\end{aligned}$$

and by l'Hôpital's rule

$$\lim_{t\uparrow 1}\frac{\log\{1-(1-t)^{\theta}\}}{(1-t)^{\theta}} = \lim_{t\uparrow 1}\frac{-1}{1-(1-t)^{\theta}} = -1,$$

yielding

$$\lim_{t\uparrow 1}\frac{\mathrm{d}}{\mathrm{d}t}\frac{\phi_{\mathrm{J}}(t;\theta)}{\phi_{\mathrm{J}}'(t;\theta)} = \frac{1}{\theta},$$

so by Theorem 1.8 the bivariate Joe/B5 copula with parameter $\theta \geq 1$ is attracted by the Gumbel copula with the same parameter θ. Therefore, the Joe model with parameter $\theta > 1$ is in the max-domain of attraction of the logistic model with parameter $\alpha = 1/\theta$, since the copula of the logistic model is the Gumbel copula.

Asymmetric Joe model The 'asymmetric Joe model' is based on an asymmetric version of the Joe copula with Fréchet margins. The asymmetric version of the Joe copula used here is obtained with the Khoudraji device (see Section 1.2.1.2) mixing the independence copula and the bivariate Joe/B5 copula. Consider the asymmetric Joe copula

$$C_{\mathrm{AJ}}(u,v;\psi_1,\psi_2,\theta) = u^{1-\psi_1}v^{1-\psi_2}C_{\mathrm{J}}(u^{\psi_1},v^{\psi_2};\theta), \quad 0 \leq u,v \leq 1,$$

where C_{J} is bivariate Joe/B5 copula with parameter $\theta \geq 1$, and mixing parameters $0 \leq \psi_1,\psi_2 \leq 1$. The Joe/B5 copula is the special case where $\psi_1 = \psi_2 = 1$. The max-stable attractor of the asymmetric Joe copula is the limiting copula of $C_{\mathrm{AJ}}\left(u^{1/n},v^{1/n}\right)^n$ as $n \to \infty$. We have

$$C_{\mathrm{AJ}}\left(u^{1/n},v^{1/n}\right)^n = u^{1-\psi_1}v^{1-\psi_2}C_{\mathrm{J}}\left(u^{\psi_1/n},v^{\psi_2/n};\theta\right)^n.$$

Since the Joe/B5 copula C_{J} is in the domain of attraction of the Gumbel copula C_{G} with the same parameter θ, we have

$$\begin{aligned}\lim_{n\to\infty} C_{\mathrm{AJ}}\left(u^{1/n}, v^{1/n}\right)^n &= u^{1-\psi_1} v^{1-\psi_2} C_{\mathrm{G}}\left(u^{\psi_1}, v^{\psi_2};\theta\right)\\ &= u^{1-\psi_1} v^{1-\psi_2} \exp\left[-\left\{(-\log u)^\theta + (-\log v)^\theta\right\}^{1/\theta}\right],\end{aligned}$$

which is the copula of the asymmetric logistic model, so the asymmetric Joe model is in the max-domain of attraction of the asymmetric logistic model.

2.4.2.2 Simulation study

For each non-max-stable model and each parameter set we generated $B = 100$ pseudo-random samples of size $N = n/(1-p)$ for $n = 100, 1000$ and $p = 0.9, 0.95, 0.99$. Then, for each sample we fitted the parametric max-stable attractor of the model from which the data were generated using the 'asymptotic', 'penultimate', and 'censored' approaches, allowing us to compare their performance at estimating the parameter values of the max-stable limiting distribution. The mean integrated squared errors of these fits are shown in Tables 2.6 and 2.7. Boxplots of the integrated squared errors of each fit and the parameter estimates are shown in Appendix B.

The results for the Joe model are shown in Table 2.6 and Figures B.34 to B.39. The comparative performance of the three approaches varies from one case to the other, but overall the performance of the 'asymptotic' and the 'penultimate' approaches improves as the threshold level p increases, unlike the 'censored' approach, for which it deteriorates. In the first case ($\theta = 1/0.3$), the 'asymptotic' approach outperforms the 'penultimate' approach for $p = 0.9$, but both approaches perform similarly for $p = 0.95, 0.99$, and they both outperform the 'censored' approach for all values of n and p. In the second case ($\theta = 1/0.55$), the 'asymptotic' approach performs best for all values of n and p, but the performance of the 'penultimate' approach is comparable to the 'asymptotic' for $p = 0.99$, whilst the 'censored' approach doesn't perform well for most values of p and n. In the third case ($\theta = 1/0.8$), the 'penultimate' approach performs best and the 'asymptotic' approach performs worst. In all cases, increasing the sample size improves the performance of the three approaches, and, as one would expect, the bias of the three approaches reduces as the threshold level p increases. The 'censored' approach has largest variance, whilst the other two approaches have similar variances.

The results for the asymmetric Joe model are shown in Table 2.7 and Figures B.40 to B.51. In all cases and for all values of n and p, the 'penultimate' approach performs best and the 'asymptotic' performs worst. The bias of the parameter estimates with the 'penultimate' approach is close to zero for all values of p and n, whilst the bias of the dependence parameter estimates with the 'asymptotic' approach increases as the threshold level p increases. The 'asymptotic' approach is unable to recover that the spectral distribution has point masses as all estimates of the coefficients of asymmetry are (roughly) equal to 1. Increasing the sample size improves the performance of the 'penultimate' estimator, whilst the performance of the

Table 2.6 – *Mean integrated squared error* ($\times 1000$) *of* $B = 100$ *maximum likelihood estimates of the logistic model (standard deviation in parenthesis) for data simulated from the Joe model for three parameter choices using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level* p *and number of excesses* n.

			Joe		
n	p	Approach	$\theta = 1/0.3$	$\theta = 1/0.55$	$\theta = 1/0.8$
100	0.90	A	1.4 (2.1)	1.1 (1.6)	44.1 (9.9)
100	0.90	P	1.8 (2.7)	6.2 (7.4)	9.0 (10.1)
100	0.90	C	4.5 (14.5)	11.7 (51.2)	11.9 (18.3)
100	0.95	A	1.8 (2.2)	1.7 (2.6)	27.2 (9.2)
100	0.95	P	1.9 (2.4)	4.0 (4.6)	6.8 (7.6)
100	0.95	C	7.1 (22.9)	13.7 (48.3)	20.0 (92.4)
100	0.99	A	1.1 (1.2)	2.0 (2.8)	10.2 (5.8)
100	0.99	P	1.1 (1.2)	2.6 (3.8)	3.1 (3.3)
100	0.99	C	25.2 (45.2)	37.4 (91.9)	40.6 (103.4)
1000	0.90	A	0.3 (0.3)	0.2 (0.2)	44.5 (3.9)
1000	0.90	P	0.6 (0.5)	3.8 (2.0)	5.6 (3.3)
1000	0.90	C	0.6 (0.7)	1.7 (1.7)	2.0 (2.4)
1000	0.95	A	0.2 (0.2)	0.2 (0.2)	27.4 (3.0)
1000	0.95	P	0.2 (0.3)	1.7 (1.3)	3.0 (2.0)
1000	0.95	C	0.4 (0.6)	0.9 (1.2)	1.3 (1.7)
1000	0.99	A	0.1 (0.2)	0.2 (0.3)	10.0 (2.1)
1000	0.99	P	0.1 (0.2)	0.4 (0.5)	0.9 (0.9)
1000	0.99	C	13.7 (40.1)	16.5 (73.4)	15.8 (77.0)

'asymptotic' approach improves or deteriorates depending on the case.

To summarize, we find that:

- the 'penultimate' approach performs better than the 'censored' approach in nearly all cases of the two non-max-stable models;
- the 'penultimate' approach performs better than the 'asymptotic' approach when the spectral distribution of the max-stable attractor has point masses and is asymmetric;
- the 'penultimate' approach doesn't improve much on the 'asymptotic' approach when the spectral distribution of the max-stable attractor is symmetric with no point mass, and extremal dependence is strong, which corresponds to the situation where the angular distribution of excesses doesn't differ much different from the spectral distribution of the attractor for high thresholds;

Table 2.7 – *Mean integrated squared error* ($\times 1000$) *of* $B = 100$ *maximum likelihood estimates of the asymmetric logistic model (standard deviation in parenthesis) for data simulated from the asymmetric Joe model for three parameter choices using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level* p *and number of excesses* n.

			Asymmetric Joe		
n	p	**Approach**	$\theta = 1/0.2$ $\psi = (0.8, 0.8)$	$\theta = 1/0.4$ $\psi = (0.7, 0.5)$	$\theta = 1/0.6$ $\psi = (0.6, 0.2)$
100	0.90	A	84.9 (22.3)	45.5 (9.0)	148.2 (20.0)
100	0.90	P	6.5 (5.3)	9.0 (9.0)	7.2 (6.7)
100	0.90	C	20.5 (31.1)	30.5 (70.2)	23.4 (16.1)
100	0.95	A	107.5 (28.2)	38.3 (2.7)	107.0 (15.1)
100	0.95	P	5.5 (4.5)	6.5 (5.0)	6.1 (5.9)
100	0.95	C	22.9 (44.8)	28.4 (47.3)	22.0 (17.6)
100	0.99	A	161.7 (42.4)	42.9 (4.7)	54.0 (7.8)
100	0.99	P	5.3 (5.3)	4.5 (4.1)	2.8 (2.4)
100	0.99	C	26.8 (34.8)	42.0 (56.7)	35.0 (59.9)
1000	0.90	A	84.7 (7.4)	43.0 (1.8)	146.0 (5.5)
1000	0.90	P	0.6 (0.6)	0.9 (0.8)	0.8 (0.6)
1000	0.90	C	3.3 (13.9)	7.2 (53.4)	7.8 (38.3)
1000	0.95	A	103.4 (9.1)	37.2 (0.7)	106.1 (4.3)
1000	0.95	P	0.6 (0.5)	0.8 (0.9)	0.6 (0.6)
1000	0.95	C	2.4 (6.6)	4.1 (14.8)	3.6 (5.0)
1000	0.99	A	156.3 (14.0)	41.6 (1.8)	54.0 (2.7)
1000	0.99	P	0.6 (0.7)	0.7 (0.8)	0.4 (0.4)
1000	0.99	C	25.8 (43.4)	25.3 (78.7)	17.7 (31.3)

- parameter estimators with the 'penultimate' approach have some bias for low thresholds but it is close to zero for large thresholds, unlike those of the 'asymptotic' and 'censored' approaches which might be biased for all considered thresholds, and in some cases their bias might even increase as the threshold level increases;

- increasing the threshold level (for a fixed number of excesses) improves the performance of the 'penultimate' approach, but counterintuitively, it might damage the performance of the 'asymptotic' and the 'censored' approaches;

- increasing the threshold level (for a fixed number of excesses) tends to decrease the performance of the 'censored' approach;

- increasing the number of observations (for a fixed threshold level) improves the performance of all approaches, though not necessarily a lot for the 'asymptotic' and 'censored' approaches.

These findings underscore that using the PDF of the spectral distribution to perform direct inference on the spectral distribution using pseudo-angles of 'large' observations yields poor results when the spectral distribution is asymmetric, when it has point masses, or when the extremal dependence is low. Using the 'penultimate' approach always yields better (or at least similar) results than the 'asymptotic' approach.

2.5 Summary

In this chapter, we provided a characterization of angular distribution of excesses corresponding to the distribution of pseudo-angles of 'large' observations, in order to improve direct inference on the spectral distribution in the bivariate setting.

In particular, we illustrated the pitfalls of inference on the spectral distribution using the standard 'asymptotic' approach based on the convergence of the angular distribution of 'large' observations to the spectral distribution as the radial threshold tends to infinity (Section 2.2). Then we provided a characterization of the angular distribution of excesses corresponding to the distribution of pseudo-angles of 'large' observations, allowing us to avoid model misspecification due to the use of finite observations when making inference on the spectral distribution (Section 2.3), and we showed that potential point masses of the spectral distribution cannot be recovered from direct inference using the 'asymptotic' approach. Finally, we investigated the performance of maximum likelihood inference on spectral distributions using the angular distribution of excesses, and compared it with two classical parametric approaches, namely the natural 'asymptotic' approach and a censored likelihood approach, through numerical simulations with datasets from both max-stable and non-max-stable models (Section 2.4). When the data come from a max-stable model, our simulation study showed that our 'penultimate' approach outperforms both the 'asymptotic' and the 'censored' approaches, except when the underlying spectral distribution is symmetric with no point mass, and extremal dependence is strong, in which case the 'penultimate' doesn't improve much on the 'asymptotic' approach. In addition, inference with the 'penultimate' approach is unbiased for all radial threshold levels, unlike the 'asymptotic' approach where counterintuitively the bias might even increase as the threshold level increases. We obtained similar results when data come from a non-max-stable model, which is most often the case in practice. However, in this case the estimator using our 'penultimate' approach is biased like the other approaches due to model misspecification, as one estimates the spectral distribution of the max-stable attractor of the distribution from which the data come.

To conclude, the angular distribution of excesses that we introduced improves maximum likelihood inference on the spectral distribution by removing model misspecification due to the use of finite observations.

3 Extremal asymmetry

Extremal dependence is at the heart of extreme value modelling and numerous measures to quantify it have been proposed in the literature, including the extremal coefficient θ_d, the tail dependence index η proposed by Ledford and Tawn (1996), and the pair of coefficients $(\chi, \overline{\chi})$ advocated by Coles *et al.* (1999) that are briefly discussed in Section 1.2.4.

In many applications, datasets seem to exhibit asymmetry in the dependence structure between the variables. Many parametric MEV models can accommodate asymmetry in the sense that the spectral density can be asymmetric, resulting in a non-exchangeable dependence structure. Such models include the bilogistic (Joe *et al.*, 1992), the Dirichlet (Coles and Tawn, 1991), the asymmetric logistic (Tawn, 1988a), and the skew-t and skew-normal (Padoan, 2011). However, there has been little if any attention on quantifying asymmetry at extreme levels, which can be useful for diagnosis and model checking. In this chapter, we propose a coefficient of extremal asymmetry that quantifies the asymmetry at extreme levels for pairs of variables.

3.1 The coefficients χ and $\overline{\chi}$ revisited

Coles *et al.* (1999) advocated measuring extremal dependence with coefficients χ and $\overline{\chi}$ defined as the limits of quantile-based functions (see Section 1.2.4). However, these functions lack the probabilistic interpretation of their respective limits, so we prefer to consider alternative functions that also have χ and $\overline{\chi}$ as their limits but have such interpretations.

3.1.1 Definitions and notation

Let (X, Y) be a bivariate random vector with joint CDF F, underlying copula C, and margins F_X and F_Y. For simplicity, suppose that F_X and F_Y are continuous, in which case the copula C is unique; see Theorem 1.7. Let $U = F_X(X)$ and $V = F_Y(Y)$ be the transformed variables with uniform margins.

The (upper) tail dependence index introduced by Joe (1993) is

$$\chi = \lim_{u\uparrow 1} \Pr\{F_Y(Y) > u \mid F_X(X) > u\} = \lim_{u\uparrow 1} \Pr(V > u \mid U > u),$$

provided the limit exists. The number $\chi \in [0,1]$ can be interpreted as the asymptotic tendency for one variable to be extreme given that the other variable is extreme. If $\chi = 0$, the variables are asymptotically independent, and are asymptotically dependent otherwise.

To quantify the upper tail dependence at any quantile level u, Coles *et al.* (1999) introduced the functions

$$\chi_l(u) = 2 - \frac{\log C(u,u)}{\log u}, \quad \overline{\chi_l}(u) = \frac{2\log(1-u)}{\log \overline{C}(u,u)} - 1, \quad 0 < u < 1,$$

and one has $\lim_{u\uparrow 1} \chi(u) = \chi$ and $\overline{\chi} = \lim_{u\uparrow 1} \overline{\chi_l}(u)$.

To have probabilistic interpretations of the quantile-based functions defining χ and $\overline{\chi}$, we use the function

$$\chi_p(u) = \Pr(V > u \mid U > u) = 2 - \frac{1 - C(u,u)}{1-u}, \quad 0 < u < 1,$$

which was used by Joe (1993) to define χ, and we introduce the function

$$\overline{\chi_p}(u) = \frac{2u}{1 - \overline{C}(u,u)} - 1, \quad 0 < u < 1.$$

Based on the definition of χ, we define an alternative extremal dependence function

$$\chi_p(u) = \Pr(V > u \mid U > u) = \frac{1 - 2u + C(u,u)}{1-u}, \quad 0 < u < 1, \tag{3.1}$$

and it follows that

$$\chi = \lim_{u\uparrow 1} \chi_p(u).$$

Definition (3.1) differs slightly from the function $\chi(u)$ introduced by Coles *et al.* (1999), denoted here by $\chi_l(u)$. The index ‘l’ in the above quantile-based function indicates that they are based on logarithms, whilst the index ‘p’ indicates that the functions have a probability interpretation.

3.1.2 Properties of χ_p

By the Fréchet bounds for copulas, any bivariate copula C has the bounds $\max(u_1 + u_2 - 1, 0) \le C(u_1, u_2) \le \min(u_1, u_2)$ for all $(u_1, u_2) \in [0,1]^2$, so the extremal dependence function is bounded from below and above by

$$\frac{1-2u+\max(2u-1,0)}{1-u} \le \chi_{\mathrm{P}}(u) \le 1, \quad 0<u<1.$$

In the case of independence $C(u,u)=u^2$, and the corresponding extremal dependence function is $\chi_{\mathrm{P}}(u)=1-u$. Therefore, variables are positively (resp. negatively) associated above quantile u if and only $\chi_{\mathrm{P}}(u)>1-u$ (resp. $\chi_{\mathrm{P}}(u)<1-u$).

3.1.3 Example: max-mixture model

Let's compute $\chi_{\mathrm{P}}(u)$ and χ when one of the variables is a max-mixture that depends on the second variable.

Let Z_1 and Z_2 be two independent unit Fréchet random variables. For some $0\le\alpha\le 1$, let $X=\max\{\alpha Z_1,(1-\alpha)Z_2\}$ and $Y=Z_1$. By construction, the random variable Y is unit Fréchet, and so is X, since

$$\Pr(X\le x)=\Pr\left\{Z_1\le\frac{x}{\alpha}, Z_2\le\frac{x}{(1-\alpha)}\right\}=e^{-\alpha/x}e^{-(1-\alpha)/x}=e^{-1/x}, \quad x>0.$$

Let F_Z denote the unit Fréchet distribution function, and define $U=F_Z(X)$, $V=F_Z(Y)$, and $z=F_Z^{-1}(u)$ for $u\in[0,1]$. Then,

$$\Pr(U>u,V>u)=\Pr(X>z,Y>z)=\Pr(Y>z)-\Pr(Y>z,X\le z).$$

Noting that $\Pr(Y>z)=\Pr(Z_1>z)=1-e^{-1/z}$, and

$$\begin{aligned}\Pr(Y>z,X\le z)&=\Pr[Z_1>z,\max\{\alpha Z_1,(1-\alpha)Z_2\}\le z]\\&=\Pr\left\{z<Z_1\le\frac{z}{\alpha}, Z_2\le\frac{z}{(1-\alpha)}\right\}\\&=\left(e^{-\alpha/z}-e^{-1/z}\right)e^{-(1-\alpha)/z}\\&=e^{-1/z}-e^{-(2-\alpha)/z},\end{aligned}$$

we have

$$\Pr(U>u,V>u)=1-2e^{-1/z}+e^{-(2-\alpha)/z}=1-2u+u^{2-\alpha}, \quad 0<u<1.$$

Thus,

$$\chi_{\mathrm{P}}(u)=\frac{1-2u+u^{2-\alpha}}{1-u}, \quad 0<u<1,$$

and by l'Hôpital's rule

$$\chi = \lim_{u\uparrow 1} 2-(2-\alpha)u^{1-\alpha} = \alpha,$$

so X and Y are asymptotically dependent if $\alpha > 0$, the case $\alpha = 0$ corresponding to independence.

3.1.4 Extreme-value distributions

In this section, we provide a characterization of the extremal dependence function $\chi_{\mathrm{p}}(u)$ based on the extremal coefficient θ_2.

Recall from the characterization of extreme value distributions (Theorem 1.9) that bivariate extreme value distributions with unit Fréchet margins have the form

$$G(x,y) = \exp\{-V(x,y)\}, \quad x,y>0,$$

where one can write

$$V(x,y) = \int_0^1 \max\left(\frac{w}{x}, \frac{1-w}{y}\right) \mathrm{d}H(w),$$

and H is a measure on $[0,1]$ satisfying the mean constraint

$$\int_0^1 w \,\mathrm{d}H(w) = 1.$$

The exponent function V is homogenous of order -1, so $V(x,x) = x^{-1}V(1,1)$, and the extremal coefficient is defined as $\theta_2 = V(1,1) = 2A(1/2)$, where A is the Pickands dependence function.

The following result gives the expression for the extremal dependence function $\chi_{\mathrm{p}}(\cdot)$ for extreme value distributions and recalls the link between the coefficient χ and the extremal coefficient θ_2, which is well known (Section 1.2.4).

Proposition 3.1 (Characterization of extremal dependence)**.** *Let (X,Y) be a bivariate random vector following an extreme value distribution G with extremal coefficient θ_2 and unit Fréchet margins. Then, the extremal dependence function is*

$$\chi_p(u) = \frac{1-2u+u^{\theta_2}}{1-u}, \quad 0<u<1, \tag{3.2}$$

and the extremal dependence coefficient is $\chi = 2-\theta_2$.

Proof. Let C be the copula underlying the CDF G, let F_Z be the unit Fréchet CDF $F_Z(z) = e^{-1/z}$, $z>0$, and let $u = F_Z^{-1}(z)$. Noting that, for $0<u<1$,

$$C(u,u) = G\left\{F_Z^{-1}(u), F_Z^{-1}(u)\right\} = \exp\left[-V\left\{-1/\log u, -1/\log u\right\}\right] = \exp\left\{\log(u)V(1,1)\right\} = u^{\theta_2},$$

it follows from equation (3.1) that the extremal dependence function is

$$\chi_{\mathrm{p}}(u) = \frac{1-2u+u^{\theta_2}}{1-u}, \quad 0 < u < 1,$$

and by l'Hopital's rule the extremal dependence coefficient is

$$\chi = \lim_{u \uparrow 1} \chi_{\mathrm{p}}(u) = 2 - \theta_2.$$

□

This result implies that for extreme value distributions with Fréchet margins the probability $\Pr(Y > z \mid X > z)$ for any $z > 0$ depends only on the extremal coefficient θ_2. In addition, rewriting equation (3.2) yields

$$\theta_2 = \frac{\log\{2u - 1 + (1-u)\chi_{\mathrm{p}}(u)\}}{\log u}, \quad 0 \le u \le 1,$$

which may be the basis of a new estimator of θ_2.

3.2 A coefficient of extremal asymmetry

In this section, we introduce a coefficient of extremal asymmetry which quantifies the level of asymmetry in the structure of dependence between two random variables at extreme levels.

3.2.1 Definitions and notations

Our characterization of extremal asymmetry is based on the relative tendency of one variable to be larger than the other, given that both are extreme.

Definition 3.1 (Partial extremal dependence coefficients)**.** Define the partial extremal dependence functions

$$\psi^{+}(u) = \Pr(V > U \mid U > u, V > u), \quad 0 \le u < 1,$$
$$\psi^{0}(u) = \Pr(V = U \mid U > u, V > u), \quad 0 \le u < 1,$$
$$\psi^{-}(u) = \Pr(V < U \mid U > u, V > u), \quad 0 \le u < 1,$$

provided $\Pr(U > u, V > u) > 0$, in which case $\psi^{+}(u) + \psi^{0}(u) + \psi^{-}(u) = 1$. If $\Pr(U > u, V > u) = 0$, set $\psi^{+}(u) = \psi^{0}(u) = \psi^{-}(u) = 0$. The partial extremal dependence coefficients are defined as

$$\psi^{+} = \lim_{u \uparrow 1} \psi^{+}(u), \quad \psi^{0} = \lim_{u \uparrow 1} \psi^{0}(u), \quad \psi^{-} = \lim_{u \uparrow 1} \psi^{-}(u),$$

provided the limits exist.

The number $\psi^+ \in [0,1]$ can be interpreted as the tendency for the second variable to be more extreme than the first, given that both variables are extreme. Similar interpretations hold for ψ^0 and ψ^-. The partial extremal dependence functions are linked to the extremal dependence function $\chi_{\mathrm{p}}(\cdot)$ by

$$\psi^+(u) = \frac{\Pr(V > U, U > u)}{\chi_{\mathrm{p}}(u)(1-u)}, \quad \psi^0(u) = \frac{\Pr(V = U, U > u)}{\chi_{\mathrm{p}}(u)(1-u)}, \quad \psi^-(u) = \frac{\Pr(V < U, V > u)}{\chi_{\mathrm{p}}(u)(1-u)},$$

for $0 < u < 1$.

Definition 3.2 (Coefficient of extremal asymmetry)**.** Define the extremal asymmetry function

$$\varphi(u) = \frac{\psi^+(u) - \psi^-(u)}{\psi^+(u) + \psi^-(u)}, \quad 0 < u < 1, \tag{3.3}$$

with the convention that $\varphi(u) = 0$ when $\psi^+(u) + \psi^-(u) = 0$, and define the coefficient of extremal asymmetry

$$\varphi = \lim_{u \uparrow 1} \varphi(u),$$

when the limit exists.

The number $\varphi \in [-1,1]$ reflects the tendency of one variable to be more extreme than the other, given that both variables are extreme. If $\varphi > 0$, then the first variable tends to be asymptotically larger than the second, and vice versa if $\varphi < 0$. If $\varphi = 0$, we say that variables X and Y are asymptotically symmetric, and asymptotically asymmetric otherwise. If $\psi^0(u) \equiv 0$, then the extremal asymmetry function simplifies to $\varphi(u) = 2\psi^+(u) - 1$.

It is obvious from the definitions that if (X, Y) are symmetric in the sense that their copula is such that $C(u, v) = C(v, u)$, $0 \leq u, v \leq 1$, then X and Y are asymptotically symmetric. Many bivariate copulas are asymptotically symmetric, including the boundary cases of comonotonicity (perfect dependence) and countermonotonicity, and independence, archimedian copulas, and copulas of elliptical distributions.

3.2.2 Example: max-mixture model

Let's compute the coefficient of extremal asymmetry for the max-mixture model defined in Section 3.1.3, using the same notation.

First, let's compute $\psi^+(u)$. We have

$$\Pr(V > U, U > u) = \Pr(Y > X, X > z) = \Pr(Y > X) - \Pr(Y > X, X \leq z). \tag{3.4}$$

The first term of the right-hand side of (3.4) may be written as

$$
\begin{aligned}
\Pr(Y > X) &= \Pr[Z_1 > \max\{\alpha Z_1, (1-\alpha) Z_2\}] \\
&= \Pr[Z_1 > (1-\alpha) Z_2] \\
&= \int_0^\infty \mathrm{d}z_1 \frac{1}{z_1^2} e^{-1/z_1} e^{-(1-\alpha)/z_1} \\
&= \frac{1}{2-\alpha} \left[e^{-(2-\alpha)/z_1}\right]_0^\infty \\
&= \frac{1}{2-\alpha}.
\end{aligned}
$$

The second term of the right-hand side of (3.4) may be written as

$$
\begin{aligned}
\Pr(Y > X, X \leq z) &= \Pr[Z_1 > \max\{\alpha Z_1, (1-\alpha) Z_2\}, \max\{\alpha Z_1, (1-\alpha) Z_2\} \leq z] \\
&= \Pr\left[(1-\alpha) Z_2 < Z_1 \leq \frac{z}{\alpha}, Z_2 \leq \frac{z}{(1-\alpha)}\right] \\
&= \int_0^{z/(1-\alpha)} \mathrm{d}z_2 \frac{1}{z_2^2} e^{-1/z_2} \left[e^{-\alpha/z} - e^{-1/\{z_2(1-\alpha)\}}\right] \\
&= e^{-\alpha/z} e^{-(1-\alpha)/z} - \int_0^{z/(1-\alpha)} \mathrm{d}z_2 \frac{1}{z_2^2} e^{-(2-\alpha)/\{z_2(1-\alpha)\}} \\
&= e^{-1/z} - \frac{1-\alpha}{2-\alpha} e^{-(2-\alpha)/z}.
\end{aligned}
$$

Substituting these two terms into (3.4) and using $u = e^{-1/z}$ yields

$$
\Pr(V > U, U > u) = \frac{1-(2-\alpha)u + (1-\alpha)u^{2-\alpha}}{2-\alpha}.
$$

Thus

$$
\psi^+(u) = \frac{1-(2-\alpha)u + (1-\alpha)u^{2-\alpha}}{(2-\alpha)\left(1-2u+u^{2-\alpha}\right)},
$$

and by l'Hôpital's rule

$$
\psi^+ = \lim_{u \uparrow 1} \frac{(1-\alpha)u^{1-\alpha} - 1}{(2-\alpha)u^{1-\alpha} - 2} = 1.
$$

Now, let's compute $\psi^-(u)$. We have

$$
\Pr(V < U, V > u) = \Pr(Y < X, Y > z) = \Pr(Y > z) - \Pr(Y \geq X, Y > z).
$$

Since $\Pr(Y > z) = \Pr(Z_1 > z) = 1 - e^{-1/z}$, and

$$\begin{aligned}\Pr(Y \geq X, Y > z) &= \Pr[Z_1 \geq \max\{\alpha Z_1, (1-\alpha)Z_2\}, Z_1 > z] \\ &= \Pr\left\{Z_1 > z, Z_2 \leq \frac{Z_1}{(1-\alpha)}\right\} \\ &= \int_z^\infty \mathrm{d}z_1 \frac{1}{z_1^2} e^{-1/z_1} e^{-(1-\alpha)/z_1} \\ &= \frac{1 - e^{-(2-\alpha)/z}}{2-\alpha},\end{aligned}$$

and using $u = e^{-1/z}$ we have

$$\Pr(V < U, V > u) = \frac{(2-\alpha)(1-u) - 1 + u^{2-\alpha}}{2-\alpha}.$$

Therefore

$$\psi^-(u) = \frac{(2-\alpha)(1-u) - 1 + u^{2-\alpha}}{(2-\alpha)\left(1 - 2u + u^{2-\alpha}\right)},$$

and by l'Hôpital's rule

$$\psi^- = \lim_{u\uparrow 1} \frac{u^{1-\alpha} - 1}{(2-\alpha)u^{1-\alpha} - 2} = 0.$$

It follows that $\psi^0(u) = 1 - \psi^+(u) - \psi^-(u) = 0$ for all $u \in [0,1]$, reflecting the fact that $\Pr(U = V) = 0$.

Finally, the extremal asymmetry function is

$$\varphi(u) = \frac{\alpha\left(1 - u^{2-\alpha}\right)}{(2-\alpha)\left(1 - 2u + u^{2-\alpha}\right)}, \quad 0 < u < 1.$$

If $\alpha = 0$, corresponding to the case of independence between Z_1 and Z_2, the extremal asymmetry function simplifies to $\varphi(u) = 0$, and thus $\varphi = 0$. If $0 < \alpha \leq 1$, then by l'Hôpital's rule the coefficient of extremal asymmetry is $\varphi = 1$, so X and Y are asymptotically asymmetric if $0 < \alpha \leq 1$.

3.2.3 Extreme-value distributions

The following result gives the extremal asymmetry function $\varphi(\cdot)$ and the coefficient of extremal asymmetry φ for bivariate extreme value distributions.

Proposition 3.2 (Characterization of extremal asymmetry). *Let (X, Y) be a bivariate random vector following an extreme value distribution G with Pickands dependence function A and unit Fréchet margins, and suppose that the exponent distribution associated with G is differentiable.*

Then, the extremal asymmetry function is

$$\varphi(u) = \frac{A'(1/2)\left(1-u^{2A(1/2)}\right)}{\theta\left(1-2u+u^{2A(1/2)}\right)}, \quad 0<u<1, \tag{3.5}$$

where A' is the derivative of A, and the coefficient of extremal asymmetry is

$$\varphi = \frac{A'(1/2)}{2-2A(1/2)}. \tag{3.6}$$

Proof. Let H be the spectral distribution associated with G, let h be its density function, and let V be the exponent function such that $G(x,y) = \exp\{-V(x,y)\}$, $x,y>0$. Let's start by deriving the expressions of the partial tail functions $\psi^+(u)$ and $\psi^-(u)$ and the extremal asymmetry function $\varphi(u)$.

The probability that both X and Y exceed a threshold $z>0$ is

$$\Pr(Y>X, X>z) = \int_z^\infty \int_x^\infty \frac{\partial^2}{\partial x \partial y} G(x,y)\,\mathrm{d}y\,\mathrm{d}x = -\int_z^\infty \left[G(x,y)V_x(x,y)\right]_{y=x}^\infty \mathrm{d}x.$$

We have $G(x,y) \to F_Z(x) = e^{-1/x}$ as $y\to\infty$, and $G(x,x) = \exp\{-V(x,x)\} = e^{-V(1,1)/x}$. Recall that

$$V(x,y) = 2\int_0^1 \max\left(\frac{w}{x}, \frac{1-w}{y}\right)\mathrm{d}H(w), \quad x,y>0,$$

and since $w/x < (1-w)/y$ when $w < x/(x+y)$, we have

$$V(x,y) = 2\int_0^{x/(x+y)} \frac{1-w}{y}\,\mathrm{d}H(w) + 2\int_{x/(x+y)}^1 \frac{w}{x}\,\mathrm{d}H(w).$$

Letting $V_x(x,y) = \partial V(x,y)/\partial x$, and recalling equation (A.1) we have

$$V_x(x,y) = -\frac{2}{x^2}\int_{x/(x+y)}^1 w h(w)\,\mathrm{d}w. \tag{3.7}$$

Thus,

$$V_x(x,x) = -\frac{2}{x^2}\int_{1/2}^1 w h(w)\,\mathrm{d}w = -\frac{2}{x^2}H^-,$$

where $H^- := \int_{1/2}^1 w\,\mathrm{d}H(w)$, and by the mean constraint,

$$\lim_{y\to\infty} V_x(x,y) = -\frac{2}{x^2}\int_0^1 w h(w)\,\mathrm{d}w = -\frac{1}{x^2},$$

so

$$\begin{aligned}
\Pr(Y > X, X > z) &= -\int_z^\infty \left\{ e^{-1/x} \frac{(-1)}{x^2} - e^{-V(1,1)/x} \frac{(-2)}{x^2} H^- \right\} \mathrm{d}x \\
&= \left[e^{-1/x} - \frac{2H^-}{V(1,1)} e^{-V(1,1)/x} \right]_z^\infty \\
&= 1 - e^{-1/z} - \frac{2H^-}{V(1,1)} \left\{ 1 - e^{-V(1,1)/z} \right\},
\end{aligned}$$

and thus

$$\Pr(V > U, U > u) = 1 - u - \frac{2H^-}{\theta} \left(1 - u^\theta \right).$$

Writing $H^+ = \int_0^{1/2} (1-w)\,\mathrm{d}H(w)$, we have

$$\theta = 2\int_0^1 \max(w, 1-w)\,\mathrm{d}H(w) = 2\int_0^{1/2} (1-w)\,\mathrm{d}H(w) + \int_{1/2}^1 w\,\mathrm{d}H(w) = 2(H^+ + H^-). \quad (3.8)$$

Hence,

$$\psi^+(u) = \frac{1 - u - 2H^-\theta^{-1}(1-u^\theta)}{1 - 2u + u^\theta},$$

and by l'Hopital's rule

$$\psi^+ = \lim_{u \uparrow 1} \frac{2H^- u^{\theta-1} - 1}{\theta u^{\theta-1} - 2} = \frac{1 - 2H^-}{2 - \theta} = \frac{1 - 2H^-}{1 - 2H^- + 1 - 2H^+} = \frac{1}{1 + R_H}, \quad (3.9)$$

where $R_H = (1 - 2H^+)/(1 - 2H^-)$.

We have $0 \le H^- \le 1/2$ since $h(w) > 0$, and by the mean constraint

$$H^- = \int_{1/2}^1 w\,\mathrm{d}H(w) \le \int_0^1 w\,\mathrm{d}H(w) = \frac{1}{2},$$

thus $0 \le 1 - 2H^- \le 1$. Note also that $0 \le H^+ \le 1/2$ since $(1-w)h(w) > 0$, and

$$H^+ = \int_0^{1/2} (1-w)\,\mathrm{d}H(w) \le \int_0^1 (1-w)\,\mathrm{d}H(w) = \int_0^1 \mathrm{d}H(w) - \int_0^1 w\,\mathrm{d}H(w) = 1 - \frac{1}{2} = \frac{1}{2},$$

implying that $0 \le 1 - 2H^+ \le 1$, and thus from equation (3.9) $0 \le \psi^+ \le 1$. If $H^+ = H^-$, then $\psi^+ = 1/2$, and if $H^- > H^+$, then $\psi^+ < 1/2$.

Similarly, the probability of a joint excess can also be written as

$$\Pr(Y < X, Y > z) = \int_z^\infty \int_y^\infty \frac{\partial^2}{\partial x \partial y} G(x,y)\,\mathrm{d}y\,\mathrm{d}x = -\int_z^\infty \left[G(x,y) V_y(x,y) \right]_{x=y}^\infty \mathrm{d}y.$$

We have $G(x,y) \to F_Z(x) = e^{-1/y}$ as $x \to \infty$, and $G(y,y) = \exp\{-V(y,y)\} = e^{-V(1,1)/y}$. Also,

$$\begin{aligned}
V_y(x,y) = \frac{\partial}{\partial y} V(x,y) &= \frac{2}{y}\left(1 - \frac{x}{x+y}\right) h\left(\frac{x}{x+y}\right) \frac{(-1)}{(x+y)^2} \\
&\quad + 2\int_0^{x/(x+y)} (1-w)\frac{(-1)}{y^2} h(w)\,\mathrm{d}w \\
&\quad - \frac{2}{x}\left(\frac{x}{x+y}\right) h\left(\frac{x}{x+y}\right) \frac{(-1)}{(x+y)^2} \\
&= -\frac{2}{y^2}\int_0^{x/(x+y)} (1-w)h(w)\,\mathrm{d}w.
\end{aligned}$$

Thus,

$$V_y(y,y) = -\frac{2}{y^2}\int_0^{1/2} (1-w)h(w)\,\mathrm{d}w = -\frac{2}{y^2}H^+,$$

and

$$\lim_{x\to\infty} V_y(x,y) = -\frac{2}{y^2}\int_0^{1} (1-w)h(w)\,\mathrm{d}w = -\frac{1}{y^2}$$

by the mean constraint, so

$$\begin{aligned}
\Pr(Y < X, Y > z) &= -\int_z^\infty \left\{ e^{-1/y}\frac{(-1)}{y^2} - e^{-V(1,1)/y}\frac{(-2)}{y^2}H^+ \right\} \mathrm{d}y \\
&= \left[e^{-1/y} - \frac{2H^+}{V(1,1)} e^{-V(1,1)/y} \right]_z^\infty \\
&= 1 - e^{-1/z} - \frac{2H^+}{V(1,1)}\left\{1 - e^{-V(1,1)/z}\right\},
\end{aligned}$$

and thus

$$\Pr(V < U, V > u) = 1 - u - \frac{2H^+}{\theta}\left(1 - u^\theta\right).$$

Hence,

$$\psi^-(u) = \frac{1 - u - 2H^+\theta^{-1}(1-u^\theta)}{1 - 2u + u^\theta},$$

and by l'Hopital's rule

$$\psi^- = \lim_{u\uparrow 1} \frac{2H^+u^{\theta-1} - 1}{\theta u^{\theta-1} - 2} = \frac{1-2H^+}{2-\theta} = \frac{1-2H^+}{1-2H^- + 1 - 2H^+} = \frac{1}{1+R_H^{-1}}.$$

Then, by equation (3.8) the extremal asymmetry function is

$$\varphi(u) = \frac{\theta^{-1}(1-u^\theta)(H^+ - H^-)}{1 - u - \theta^{-1}(1-u^\theta)(H^- + H^+)} = \frac{2\theta^{-1}(1-u^\theta)(H^+ - H^-)}{1 - 2u + u^\theta}. \tag{3.10}$$

Now, note that

$$H^+ - H^- = \int_0^{1/2} (1-w)\,\mathrm{d}H(w) - \int_{1/2}^{1} w\,\mathrm{d}H(w) = \int_0^{1/2} \mathrm{d}H(w) - \frac{1}{2} \tag{3.11}$$

by the mean constraint that applies to $H(w)$. From the definition of the Pickands dependence function, $A(t) = V\{(1-t)^{-1}, t^{-1}\}$ for $0 \le t \le 1$. Thus

$$\begin{aligned}
A(t) &= 2\int_0^1 \max\{w(1-t), (1-w)t\}\,\mathrm{d}H(w) \\
&= 2t\int_0^t (1-w)\,\mathrm{d}H(w) + 2(1-t)\int_t^1 w\,\mathrm{d}H(w) \\
&\quad + 2(1-t)\int_0^t w\,\mathrm{d}H(w) - 2(1-t)\int_0^t w\,\mathrm{d}H(w) \\
&= (1-t) + 2\int_0^t (t-w)\,\mathrm{d}H(w),
\end{aligned}$$

and

$$A'(t) = 2\int_0^t \mathrm{d}H(w) - 1. \tag{3.12}$$

Substituting equations (3.11) and (3.12) into (3.10) yields

$$\varphi(u) = \frac{A'(1/2)(1-u^\theta)}{\theta(1-2u+u^\theta)}, \tag{3.13}$$

and by l'Hopital's rule

$$\varphi(u) = \frac{A'(1/2)(1-u^\theta)}{\theta(1-2u+u^\theta)}, \quad u \in [0,1).$$

Since $\chi = 2-\theta$ and $\theta = 2A(1/2)$, we have the following equivalent expressions

$$\varphi = \frac{A'(1/2)}{2-\theta} = \frac{A'(1/2)}{\chi} = \frac{A'(1/2)}{2\{1-A(1/2)\}}.$$

□

By the convexity of A, the slope of the Pickand dependence function is bounded from below and above by

$$-2\{1-A(1/2)\} \le A'(1/2) \le 2\{1-A(1/2)\}.$$

The denominator in is just a normalization factor ensuring that $\varphi \in [-1,1]$.

3.2.3.1 Example: Asymmetric logistic distribution

The bivariate asymmetric logistic distribution function is

$$G(x,y) = \exp\left[-\frac{1-\psi_1}{x} - \frac{1-\psi_2}{y} - \left\{\left(\frac{\psi_1}{x}\right)^{1/\alpha} + \left(\frac{\psi_2}{y}\right)^{1/\alpha}\right\}^{\alpha}\right], \quad x_1, x_2 > 0,$$

where $0 \le \psi_1, \psi_2 \le 1$, and $0 < \alpha \le 1$. The corresponding exponent function is

$$V(x,y) = \frac{1-\psi_1}{x} + \frac{1-\psi_2}{y} + \left\{\left(\frac{\psi_1}{x}\right)^{1/\alpha} + \left(\frac{\psi_2}{y}\right)^{1/\alpha}\right\}^{\alpha},$$

and the stable extremal dependence function is

$$l(v_1, v_2) = (1-\psi_1)v_1 + (1-\psi_2)v_2 + \left\{(\psi_1 v_1)^{1/\alpha} + (\psi_2 v_2)^{1/\alpha}\right\}^{\alpha}.$$

The spectral density for $0 < w < 1$ is

$$\begin{aligned} h(w) &= -\frac{1}{2}\frac{\partial^2}{\partial x \partial y} V(x,y)\Big|_{x=w, y=1-w} \\ &= \frac{1}{2}(1/\alpha - 1)\psi_1^{1/\alpha}\psi_2^{1/\alpha}\{w(1-w)\}^{-1/\alpha-1}\left\{\left(\frac{\psi_1}{w}\right)^{1/\alpha} + \left(\frac{\psi_2}{1-w}\right)^{1/\alpha}\right\}^{\alpha-2} \\ &= \frac{1}{2}(1/\alpha - 1)(\psi_1\psi_2)^{1/\alpha}\{w(1-w)\}^{1/\alpha-2}\left[(\psi_2 w)^{1/\alpha} + \{\psi_1(1-w)\}^{1/\alpha}\right]^{\alpha-2}. \end{aligned}$$

Pickand's dependence function is

$$A(t) = l(1-t, t) = (1-\psi_1)(1-t) + (1-\psi_2)t + \left[\{\psi_1(1-t)\}^{1/\alpha} + (\psi_2 t)^{1/\alpha}\right]^{\alpha},$$

and its derivative is

$$A'(t) = \psi_1 - \psi_2 + \left[\psi_2(\psi_2 t)^{1/\alpha-1} - \psi_1\{\psi_1(1-t)\}^{1/\alpha-1}\right]\left[\{\psi_1(1-t)\}^{1/\alpha} + (\psi_2 t)^{1/\alpha}\right]^{\alpha-1}.$$

Thus,

$$A(1/2) = \frac{1}{2}\left\{2 - \psi_1 - \psi_2 + \left(\psi_1^{1/\alpha} + \psi_2^{1/\alpha}\right)^{\alpha}\right\},$$

and

$$A'(1/2) = \psi_1 - \psi_2 + \left(\psi_2^{1/\alpha} - \psi_1^{1/\alpha}\right)\left(\psi_1^{1/\alpha} + \psi_2^{1/\alpha}\right)^{\alpha-1}$$

From equation (3.2) with $\theta = 2A(1/2)$ we have

$$\chi(u) = \frac{1 - 2u + u^{2A(1/2)}}{1-u},$$

and

$$\chi = \psi_1 + \psi_2 - \left(\psi_1^{1/\alpha} + \psi_2^{1/\alpha}\right)^{\alpha}.$$

From equation (3.13) we have

$$\varphi(u) = \frac{A'(1/2)\left\{1 - u^{2A(1/2)}\right\}}{2A(1/2)\left\{1 - 2u + u^{2A(1/2)}\right\}},$$

and

$$\varphi = \frac{A'(1/2)}{2\{1 - A(1/2)\}}.$$

3.2.3.2 Extremal asymmetry function in terms of the copula

This section provides expressions for $\varphi(u)$ and φ in terms of the underlying copula.

Since $A(t) = -\log C\left(e^{-(1-t)}, e^{-t}\right)$, we have

$$A'(t) = \frac{e^{-t} C_y\left(e^{-(1-t)}, e^{-t}\right) - e^{-(1-t)} C_x\left(e^{-(1-t)}, e^{-t}\right)}{C\left(e^{-(1-t)}, e^{-t}\right)},$$

and thus

$$A'(1/2) = \frac{e^{-1/2}\left\{C_y\left(e^{-1/2}, e^{-1/2}\right) - C_x\left(e^{-1/2}, e^{-1/2}\right)\right\}}{C\left(e^{-1/2}, e^{-1/2}\right)}.$$

Substituting $u^\theta = C(u,u)$ into equation (3.13) yields

$$\varphi(u) = \frac{A'(1/2)\{1 - C(u,u)\}\log u}{\log C(u,u)\{1 - 2u + C(u,u)\}}.$$

3.3 Estimation

In this section, we describe two non-parametric approaches to estimating the coefficient of extremal asymmetry.

3.3.1 Empirical approach

Natural estimators of the functions of extremal dependence and extremal asymmetry are obtained using empirical estimates of the probabilities involved in the definition of these functions. Rewriting equations (3.1) and (3.3) in terms of the transformed variables $U = F_X(X)$ and $V = F_Y(Y)$ yields

$$\chi_{\mathrm{p}}(u) = \frac{\Pr(V > u, U > u)}{1 - u},$$

and

$$\varphi(u) = \frac{\Pr(V > U, U > u) - \Pr(V < U, V > u)}{\Pr(V > U, U > u) + \Pr(V < U, V > u)}.$$

On the basis of independent realizations $(x_1, y_1), \dots, (x_N, y_N)$ and transformed observations $u_i = \hat{F}_X(x_i)$ and $v_i = \hat{F}_Y(y_i)$, the estimators of these functions are

$$\hat{\chi}_{\mathrm{p}}(u) = \frac{N^{-1}\sum_{i=1}^{N} I(u_i > u, v_i > u)}{1-u},$$
$$\hat{\varphi}(u) = \frac{\sum_{i=1}^{N}\{I(v_i > u_i, u_i > u) - I(v_i < u_i, v_i > u)\}}{\sum_{i=1}^{N}\{I(v_i > u_i, u_i > u) + I(v_i < u_i, v_i > u)\}}.$$

Natural estimators of the marginal distributions F_X and F_Y are obtained with the modified empirical distribution function $\hat{F}_X(t) = (N+1)^{-1}\sum_{i=1}^{N} I(x_i < t)$ and $\hat{F}_Y(t) = (N+1)^{-1}\sum_{i=1}^{N} I(y_i < t)$.

Coefficients χ and φ can be estimated respectively by $\hat{\chi}_{\mathrm{p}}(u)$ and $\hat{\varphi}(u)$ for some appropriate threshold u close to one, since χ and φ are respectively the limits of $\chi_{\mathrm{p}}(u)$ and $\varphi(u)$ when $u \uparrow 1$ (Section 3.1.1 and Definition 3.2).

3.3.2 Empirical likelihood approach

An alternative non-parametric approach to estimate the coefficients χ and φ is to estimate the spectral distribution non-parametrically and to derive estimates of the coefficients from the relation between the coefficients and the spectral distribution. We illustrate this approach with the maximum empirical likelihood estimator introduced by Einmahl and Segers (2009).

Let $(x_1, y_1), \dots, (x_N, y_N)$ be a sample from a bivariate max-stable distribution with unit Fréchet margins. In practice, data can be transformed to have unit Fréchet margins (see Section 1.2.1.3). Expressed in pseudo-polar coordinates these observations are (r_i, w_i), where $r_i = x_i + y_i$ and $w_i = x_i / r_i$ for $i = 1, \dots, N$. Let $z > 0$ be a large threshold, and let n be the number of 'large' observations such that $r_i > z$. Without loss of generality and to simplify notation, suppose that the observations are ordered in decreasing order of pseudo-radius, that is $r_1 \geq \cdots \geq r_N$, so the set of pseudo-angles for 'large' observations is $\{w_1, \dots, w_n\}$.

The maximum empirical likelihood estimator places masses $p_1, \dots, p_n$ on $w_1, \dots, w_n$ in order to maximize the log empirical likelihood

$$l(p_1, \dots, p_n) = \sum_{i=1}^{n} \log p_i,$$

subject to the probability constraint $\sum_{i=1}^{n} p_i = 1$ and the mean constraint $\sum_{i=1}^{n} w_i p_i = 1/2$. By the method of Lagrange multipliers, the solution is

$$p_i = \frac{1}{n}\frac{1}{1+\lambda(w_i - 1/2)}, \quad i = 1, \dots, n,$$

where $\lambda \in \mathbb{R}$ is the Lagrange multiplier associated to the mean constraint, defined implicitly by

$$\sum_{i=1}^{n} \frac{w_i - 1/2}{1 + \lambda(w_i - 1/2)} = 0,$$

which has a unique solution in $(-1/\max(w_i - 1/2), -1/\min(w_i - 1/2))$, provided that $\min w_i < 1/2 < \max w_i$, which happens with probability one as $n \to \infty$.

Let $\hat{H}([0,t]) = \sum_{i=1}^{n} p_i I(w_i \le t)$ be the non-parametric estimator of the spectral distribution. From Equation (1.13) the estimator of the Pickands dependence function is

$$\hat{A}(t) = 1 - t + 2\int_0^t \hat{H}([0,w])\,\mathrm{d}w = 1 - t + 2\sum_{i=1}^{n} p_i (t - w_i) I(w_i \le t), \quad 0 \le t \le 1,$$

and from Equation (1.14) the estimator of the right-hand derivative of the Pickands dependence function is

$$\hat{A}'(t) = 2\hat{H}([0,t]) - 1 = 2\sum_{i=1}^{n} p_i I(w_i \le t) - 1, \quad 0 \le t < 1.$$

From Proposition 3.1 and recalling that $\theta_2 = 2A(1/2)$, the estimator of the coefficient of extremal dependence is

$$\hat{\chi} = 1 - 4\sum_{i=1}^{n} p_i (1/2 - w_i) I(w_i \le 1/2),$$

and from Proposition 3.2 the estimator of the coefficient of extremal asymmetry is

$$\hat{\varphi} = \frac{\sum_{i=1}^{n} p_i I(w_i \le 1/2) - 1/2}{1/2 - 2\sum_{i=1}^{n} p_i (1/2 - w_i) I(w_i \le 1/2)}.$$

3.4 Simulation study

This section presents the simulation results to compare the 'empirical' and 'empirical likelihood' approaches to estimate the coefficient of extremal asymmetry in the case where the data are simulated from max-stable models.

Max-stable models, or more specifically the PDF of their associated spectral distributions, might be symmetric or asymmetric, they might have point masses or no point masses, and they can have various levels of extremal dependence. In order to cover a large spectrum of the possible cases, we used the logistic model (Section 1.2.1.5), the Hüsler–Reiss model (Section 1.2.1.5), the Dirichlet model (Section 1.2.1.5), and the asymmetric logistic model (Section 1.2.1.5), and for each of them we selected three sets of parameters, as in Chapter 2. Table 2.1 lists the selected sets of parameters for each model and provides their point masses, their level of extremal dependence as measured by the tail dependence index χ (see Section 1.2.4), and their level of extremal asymmetry as measured by the coefficient of extremal

Table 3.1 – *Bias (×100) and RMSE (×100) of $B = 1000$ estimates of the coefficient of extremal asymmetry for data simulated from the logistic model for three parameter choices using the 'empirical' (E) and 'empirical likelihood' (EL) approaches for several choices of data sample size N and threshold level u.*

			Logistic		
N	***u***	**Approach**	$\boldsymbol{\alpha = 0.3}$	$\boldsymbol{\alpha = 0.55}$	$\boldsymbol{\alpha = 0.8}$
1000	0.9	E	-22.6 / 23.7	-43.4 / 44	-68.2 / 68.5
1000	0.9	EL	0.7 / 9.4	1.5 / 13.2	1.8 / 23.2
1000	0.95	E	-24.7 / 27	-46 / 47.2	-71.9 / 72.5
1000	0.95	EL	0.6 / 13.4	1 / 19.6	1 / 38.4
1000	0.99	E	-31.9 / 41.2	-53.7 / 59.2	-77.8 / 80.1
1000	0.99	EL	-0.2 / 34.9	-3.6 / 51.3	-28.8 / 82.3
5000	0.9	E	-21.8 / 22.1	-43 / 43.1	-67.6 / 67.7
5000	0.9	EL	0 / 4	0.2 / 5.6	0.5 / 10.9
5000	0.95	E	-22.8 / 23.3	-45 / 45.2	-70.9 / 71.1
5000	0.95	EL	0.1 / 5.7	-0.1 / 8.5	0.8 / 16.3
5000	0.99	E	-24.9 / 27	-47.8 / 49	-74.2 / 74.7
5000	0.99	EL	0.5 / 13.2	0 / 19	1.9 / 42

asymmetry φ.

For each distribution and each parameter set we generated $B = 1000$ pseudo-random samples of sizes $N = 1000, 5000$. Then, for each sample we estimated the coefficient of extremal asymmetry using the 'empirical' and 'empirical likelihood' approaches with the same number of observations n, which corresponds to the number of transformed observations (u_i, v_i) having both components larger than the parameter $u = 0.9, 0.95, 0.99$. The number n can vary from one sample to the other but both estimation approaches use the same number of observations on the same sample, allowing us to compare their performance at estimating the true underlying coefficient values. The biases and root mean squared errors (RMSE) for these estimates are shown in Tables 3.1 to 3.4.

The results are similar for the four models and the three sets of parameters. Unsurprisingly, increasing the sample size yields better estimators, and raising the threshold u yields worse estimators. The 'empirical likelihood' method performs better than the 'empirical' method in terms of both bias and RMSE for all parameter values, with a few exceptions when the threshold u is the highest and the extremal dependence of the model is weak, which yields a small sample size n. The 'empirical likelihood' estimators have low biases and most of their RMSEs are due to their variance. Conversly, most of the 'empirical' estimators' RMSE is due to bias.

The difference between these two approaches, which yield estimators with drastically different characteristics, lies in the subsets of observations that are used in estimation and the way

Table 3.2 – *Bias ($\times 100$) and RMSE ($\times 100$) of $B = 1000$ estimates of the coefficient of extremal asymmetry for data simulated from the Hüsler–Reiss model for three parameter choices using the 'empirical' (E) and 'empirical likelihood' (EL) approaches for several choices of data sample size N and threshold level u.*

			Hüsler–Reiss		
N	u	**Approach**	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 2$
1000	0.9	E	-19.5 / 20.8	-36.6 / 37.3	-63.5 / 63.8
1000	0.9	EL	0.1 / 8.4	0.1 / 10.8	0.1 / 19.1
1000	0.95	E	-21 / 23.4	-38.7 / 40	-66.4 / 67
1000	0.95	EL	0.1 / 11.8	-0.4 / 15.3	-0.5 / 28.8
1000	0.99	E	-30 / 39.1	-47.4 / 53	-73.6 / 76.3
1000	0.99	EL	-1.7 / 30.4	-3.4 / 43.1	-21.3 / 73.9
5000	0.9	E	-18.9 / 19.1	-35.8 / 35.9	-62.7 / 62.7
5000	0.9	EL	0 / 3.5	0.4 / 4.7	0.7 / 8.5
5000	0.95	E	-19.6 / 20	-37.2 / 37.5	-65.6 / 65.7
5000	0.95	EL	0.1 / 4.9	0.5 / 6.7	0.5 / 12.7
5000	0.99	E	-22.3 / 24.4	-39.7 / 40.9	-68.8 / 69.4
5000	0.99	EL	-0.4 / 11.7	0.4 / 15.9	1.1 / 32

they are used. The 'empirical' estimator is computed directly from the coefficient's definition but $\hat{\varphi}$ is approximated by $\hat{\varphi}(u)$, wheras the 'empirical likelihood' estimator is computed indirectly from a biased estimator of the spectral distribution but $\hat{\varphi}$ is not approximated by $\hat{\varphi}(u)$. It seems from the simulations that it is better to estimate φ indirectly from a biased estimator of the spectral distribution rather than through estimating the extremal asymmetry function at level $u < 1$.

3.5 Summary

In this chapter, we introduced a coefficient of extremal asymmetry for pairs of variables and which quantifies the relative tendency of one variable to be larger than the other, given that both are extreme. This can be used for diagnosis and model checking in the bivariate setting.

We provided a characterization of extremal asymmetry in the case where the variables follow a bivariate extreme value distribution. The coefficient of asymmetry has strong ties with the Pickands dependence function and its derivative.

We introduced two non-parametric estimators of the coefficient of extremal asymmetry and compared their performance through numerical simulations. The two estimators have diametrically opposed bias-variance trade-offs. Compared to the empirical estimator based on the definition of the coefficient, the estimator based on maximum empirical likelihood performed better, often much better, both in terms of root mean squared error and bias.

Table 3.3 – *Bias* ($\times 100$) *and RMSE* ($\times 100$) *of* $B = 1000$ *estimates of the coefficient of extremal asymmetry for data simulated from the Dirichlet model for three parameter choices using the 'empirical' (E) and 'empirical likelihood' (EL) approaches for several choices of data sample size* N *and threshold level* u.

			Dirichlet		
N	u	**Approach**	$\alpha = 3, \beta = 3$	$\alpha = 3, \beta = 0.2$	$\alpha = 0.2, \beta = 0.2$
1000	0.9	E	-30.1 / 31	-82.1 / 82.4	-73 / 73.3
1000	0.9	EL	-0.2 / 10.1	-4.9 / 19	0.3 / 27.3
1000	0.95	E	-31.8 / 33.5	-85.2 / 85.7	-77.3 / 77.7
1000	0.95	EL	-0.4 / 14.4	-2.5 / 28.1	-0.6 / 47.4
1000	0.99	E	-40.5 / 47.6	-94.6 / 97	-82.9 / 84.7
1000	0.99	EL	-2.7 / 37.7	-21.2 / 78.5	-42.2 / 88.3
5000	0.9	E	-29.4 / 29.6	-81.1 / 81.1	-72.6 / 72.7
5000	0.9	EL	0.1 / 4.2	-5.3 / 9.6	-0.2 / 12.9
5000	0.95	E	-30.5 / 30.8	-84.1 / 84.2	-76.5 / 76.6
5000	0.95	EL	0.2 / 6.1	-2.9 / 12.2	-0.6 / 20
5000	0.99	E	-33 / 34.6	-87.6 / 88.2	-79.9 / 80.3
5000	0.99	EL	0.1 / 14.7	0.5 / 30.2	-0.6 / 52.7

Table 3.4 – *Bias* ($\times 100$) *and RMSE* ($\times 100$) *of* $B = 1000$ *estimates of the coefficient of extremal asymmetry for data simulated from the asymmetric logistic model for three parameter choices using the 'empirical' (E) and 'empirical likelihood' (EL) approaches for several choices of data sample size* N *and threshold level* u.

			Asymmetric logistic		
N	u	**Approach**	$\alpha = 0.2$ $\psi = (0.8, 0.8)$	$\alpha = 0.4$ $\psi = (0.7, 0.5)$	$\alpha = 0.6$ $\psi = (0.6, 0.2)$
1000	0.9	E	-31.3 / 34.4	-36.3 / 37.3	-32.9 / 33.3
1000	0.9	EL	-0.2 / 12.8	4.7 / 18.7	13.7 / 36.4
1000	0.95	E	-33.7 / 38.8	-39.4 / 41.1	-36.9 / 37.5
1000	0.95	EL	-0.2 / 18.5	1.5 / 28.3	5.4 / 54.9
1000	0.99	E	-45 / 56.8	-42.5 / 48.2	-41 / 42.7
1000	0.99	EL	-0.4 / 48.4	-5.7 / 67.8	-28.9 / 59.4
5000	0.9	E	-30 / 30.8	-36.7 / 36.9	-32.9 / 33
5000	0.9	EL	-0.1 / 5.7	4.7 / 9.3	15.5 / 21.5
5000	0.95	E	-31.5 / 32.9	-39.4 / 39.8	-37.2 / 37.3
5000	0.95	EL	0 / 8.1	2.1 / 12.4	7.4 / 26.9
5000	0.99	E	-33.9 / 39	-41.7 / 43.2	-40.4 / 40.8
5000	0.99	EL	0.8 / 19.4	0 / 30.5	-1.4 / 58.7

4 Perspective

This chapter discusses aspects that were not investigated in this thesis and that are left for future work.

In Chapter 2, we introduced the angular distribution of excesses in the bivariate setting. It can be expected that the difference between the spectral distribution, which is often of interest in inference, and the angular distribution of excesses, from which the observations used in inference are drawn, also negatively impacts, possibly in a worse way, direct inference in dimensions $d > 2$. It seems clear that the characterization of the angular distribution of excesses could be extended to the multivariate setting and used to improve inference on the spectral distribution.

The angular distribution of excesses was used to improve parametric inference on the spectral distribution. It can be expected that the angular distribution of excesses can also be used to improve inference in a semi-parametric or nonparametric setting. We made an attempt to develop such a non-parametric estimator based on the idea of the empirical likelihood estimator introduced by Einmahl and Segers (2009). The expression of the empirical likelihood for the angular distribution of excesses is rather intricate, and we couldn't find an analytical simplification for the maximization problem using Lagrange multipliers like Einmahl and Segers (2009), so we tried to solve the high-dimensional non-linear optimization problem with non-linear constraints using numerical routines in `R`. The results were not conclusive. Moreover, this approach has two conceptual drawbacks. First, the fitted spectral distribution might have mass were there shouldn't be mass, as the locations of the atoms are given by the observed pseudo-angles. Second, this approach fails to put potential mass at the edges of the support, though it might be possible to address this by adapting the empirical likelihood approach. We also made an attempt to develop a semi-parametric estimator which approximates the spectral distribution by a mixture of beta distributions and a discrete distribution with point masses at the edges of the support. This approach seems appealing conceptually, as mixtures of beta distribution are dense in the space of probability density functions on the unit interval and point masses at the edges can easily be accommodated. However, we also encountered difficulties in numerically solving the optimization problem to obtain the fitted distribution.

In Chapter 3, we proposed a coefficient of extremal asymmetry and two non-parametric estimators. The asymptotic properties of these estimators are not formally investigated in this thesis, but it seems clear that both can be shown to be consistent and asymptotically normal under suitable conditions. The empirical estimator is purely moment-based, so it can be expected to be consistent and asymptotically normal under the usual conditions for extremal estimation, i.e., $u \to 1$, $N \to \infty$ and $N(1-u) \to \infty$. The properties of the empirical likelihood-based estimator will depend on those of the empirical likelihood estimator of the spectral distribution, whose good asymptotic properties were established by Einmahl and Segers (2009) using empirical process theory, so it too can be expected to have standard limiting behaviour.

A Computational details

A.1 Derivation of V_x, V_y, and $V_{x,y}$

This section contains the derivation of the partial and mixed partial derivatives of the bivariate exponent measure used in proofs on page 45, and on page 75.

Recall that

$$V(x,y) = \frac{2}{y}\left\{\delta_0 + \int_0^{x/(x+y)} (1-t)\tilde{h}(t)\,\mathrm{d}t\right\} + \frac{2}{x}\left\{\delta_1 + \int_{x/(x+y)}^1 t\tilde{h}(t)\,\mathrm{d}t\right\},$$

where the $\tilde{h}$ is the continuous part of the generalized density function of a spectral distribution $H \in \mathcal{H}_1$, and δ_0 and δ_1 respectively denote the point masses at 0 and 1.

The partial derivative of $V(x,y)$ with respect to x is

$$\begin{aligned}
V_x(x,y) &= \frac{\partial}{\partial x} V(x,y) \\
&= \frac{\partial}{\partial x}\frac{2}{y}\left\{\delta_0 + \int_0^{x/(x+y)} (1-t)\tilde{h}(t)\,\mathrm{d}t\right\} + \frac{\partial}{\partial x}\frac{2}{x}\left\{\delta_1 + \int_{x/(x+y)}^1 t\tilde{h}(t)\,\mathrm{d}t\right\} \\
&= \frac{2}{y}\left\{\left(1 - \frac{x}{x+y}\right)\tilde{h}\left(\frac{x}{x+y}\right)\frac{y}{(x+y)^2}\right\} \\
&\quad - \frac{2}{x^2}\left\{\delta_1 + \int_{x/(x+y)}^1 t\tilde{h}(t)\,\mathrm{d}t\right\} - \frac{2}{x}\frac{x}{x+y}\tilde{h}\left(\frac{x}{x+y}\right)\frac{y}{(x+y)^2} \\
&= -\frac{2}{x^2}\left\{\delta_1 + \int_{x/(x+y)}^1 t\tilde{h}(t)\,\mathrm{d}t\right\}.
\end{aligned} \tag{A.1}$$

The partial derivative of $V(x,y)$ with respect to y is

$$
\begin{aligned}
V_y(x,y) &= \frac{\partial}{\partial y} V(x,y) \\
&= \frac{\partial}{\partial y}\frac{2}{y}\left\{\delta_0 + \int_0^{x/(x+y)} (1-t)\tilde{h}(t)\,\mathrm{d}t\right\} + \frac{\partial}{\partial y}\frac{2}{x}\left\{\delta_1 + \int_{x/(x+y)}^{1} t\tilde{h}(t)\,\mathrm{d}t\right\} \\
&= -\frac{2}{y^2}\left\{\delta_0 + \int_0^{x/(x+y)} (1-t)\tilde{h}(t)\,\mathrm{d}t\right\} - \frac{2}{y}\left(1-\frac{x}{x+y}\right)\tilde{h}\left(\frac{x}{x+y}\right)\frac{x}{(x+y)^2} \\
&\quad + \frac{2}{x}\frac{x}{x+y}\tilde{h}\left(\frac{x}{x+y}\right)\frac{x}{(x+y)^2} \\
&= -\frac{2}{y^2}\left\{\delta_0 + \int_0^{x/(x+y)} (1-t)\tilde{h}(t)\,\mathrm{d}t\right\}.
\end{aligned}
$$

Finally, the mixed partial derivative of $V(x,y)$ is

$$
\begin{aligned}
V_{xy}(x,y) &= \frac{\partial^2}{\partial x \partial y} V(x,y) \\
&= -\frac{\partial}{\partial x}\frac{2}{y^2}\left\{\delta_0 + \int_0^{x/(x+y)} (1-t)\tilde{h}(t)\,\mathrm{d}t\right\} \\
&= -\frac{2}{y^2}\left(1-\frac{x}{x+y}\right)\tilde{h}\left(\frac{x}{x+y}\right)\frac{y}{(x+y)^2} \\
&= -\frac{2}{(x+y)^3}\tilde{h}\left(\frac{x}{x+y}\right).
\end{aligned}
$$

B Simulation results for parametric fits of spectral distribution

This appendix contains graphical results of numerical simulations performed to compare the performance of the 'asymptotic', 'penultimate', and 'censored' approaches to estimate the spectral distribution of the logistic, Hüsler–Reiss, Dirichlet, and asymmetric logistic models for three parameter choices each and several choices of threshold level p and number of threshold excesses n. The simulation procedure and the estimation approaches are described in Section 2.4.

B.1 Logistic model

B.1.1 Case $\alpha = 0.3$

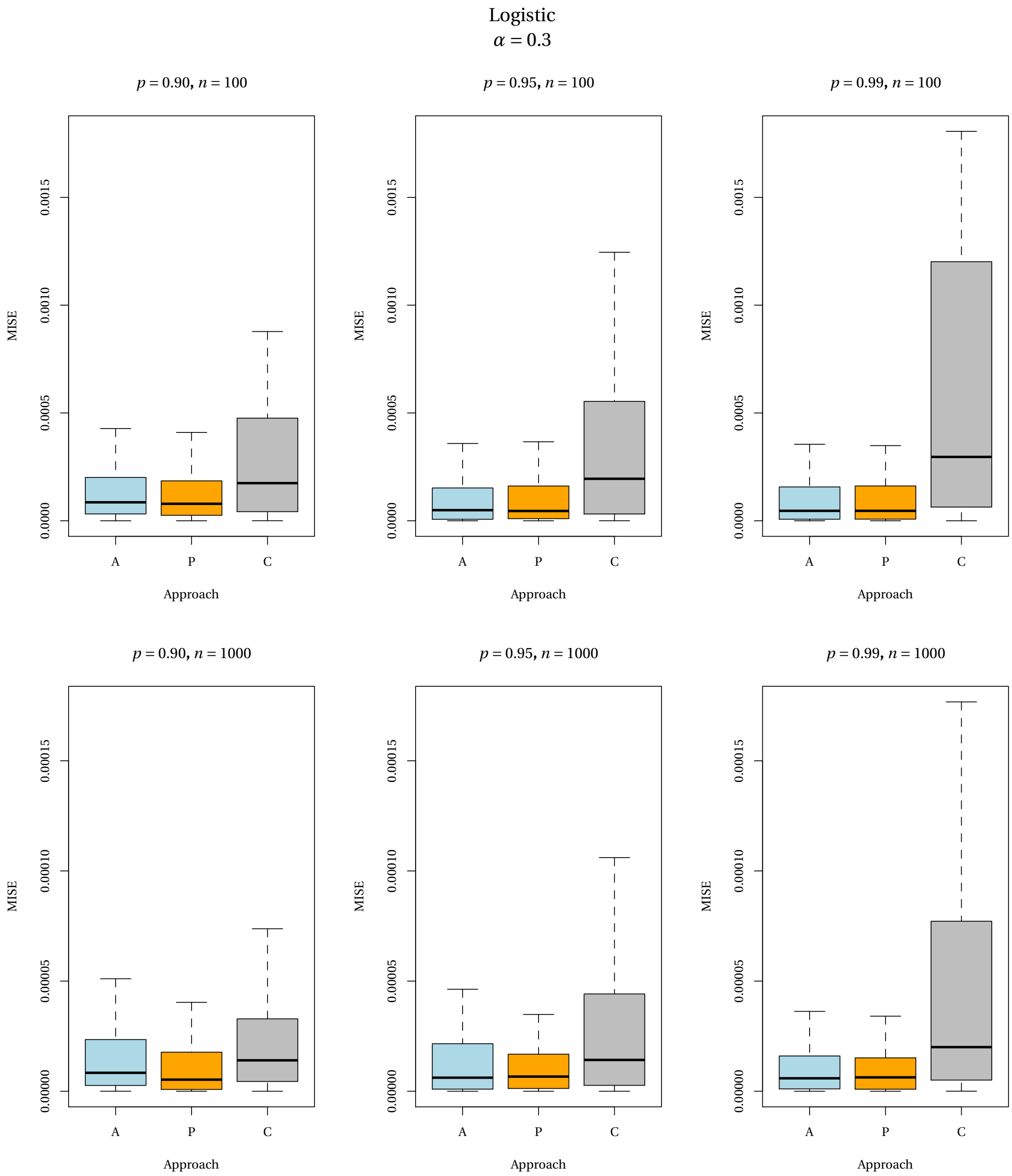

Figure B.1 – *Boxplots comparing the integrated squared error of maximum likelihood estimates of the logistic model based on data simulated from the logistic model with dependence parameter $\alpha = 0.3$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n.*

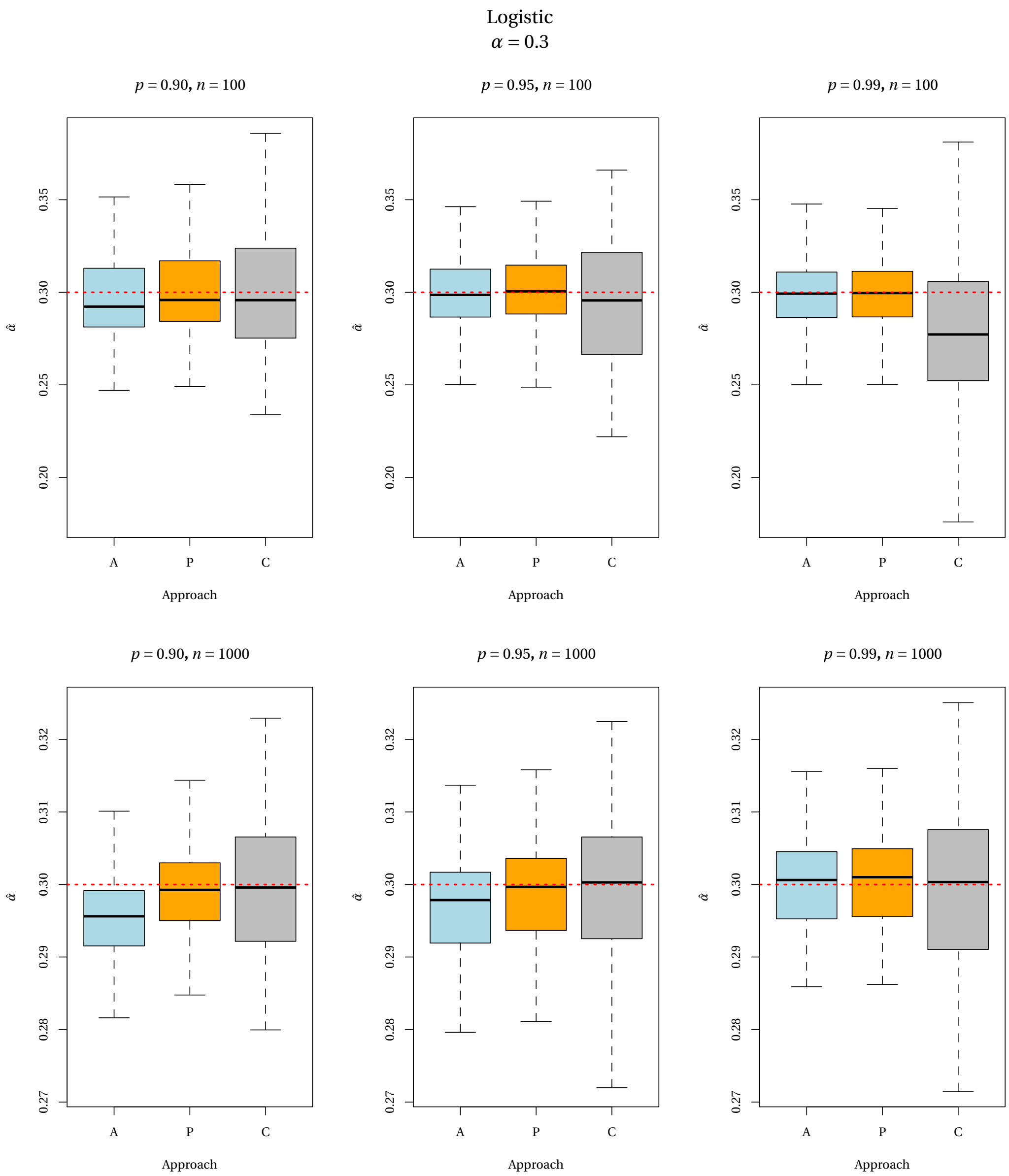

Figure B.2 – *Boxplots comparing dependence parameter estimates of the logistic model based on data simulated from the logistic model with dependence parameter* $\alpha = 0.3$ *using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level* p *and number of excesses* n*. The dashed red lines show the true parameter value.*

B.1.2 Case $\alpha = 0.55$

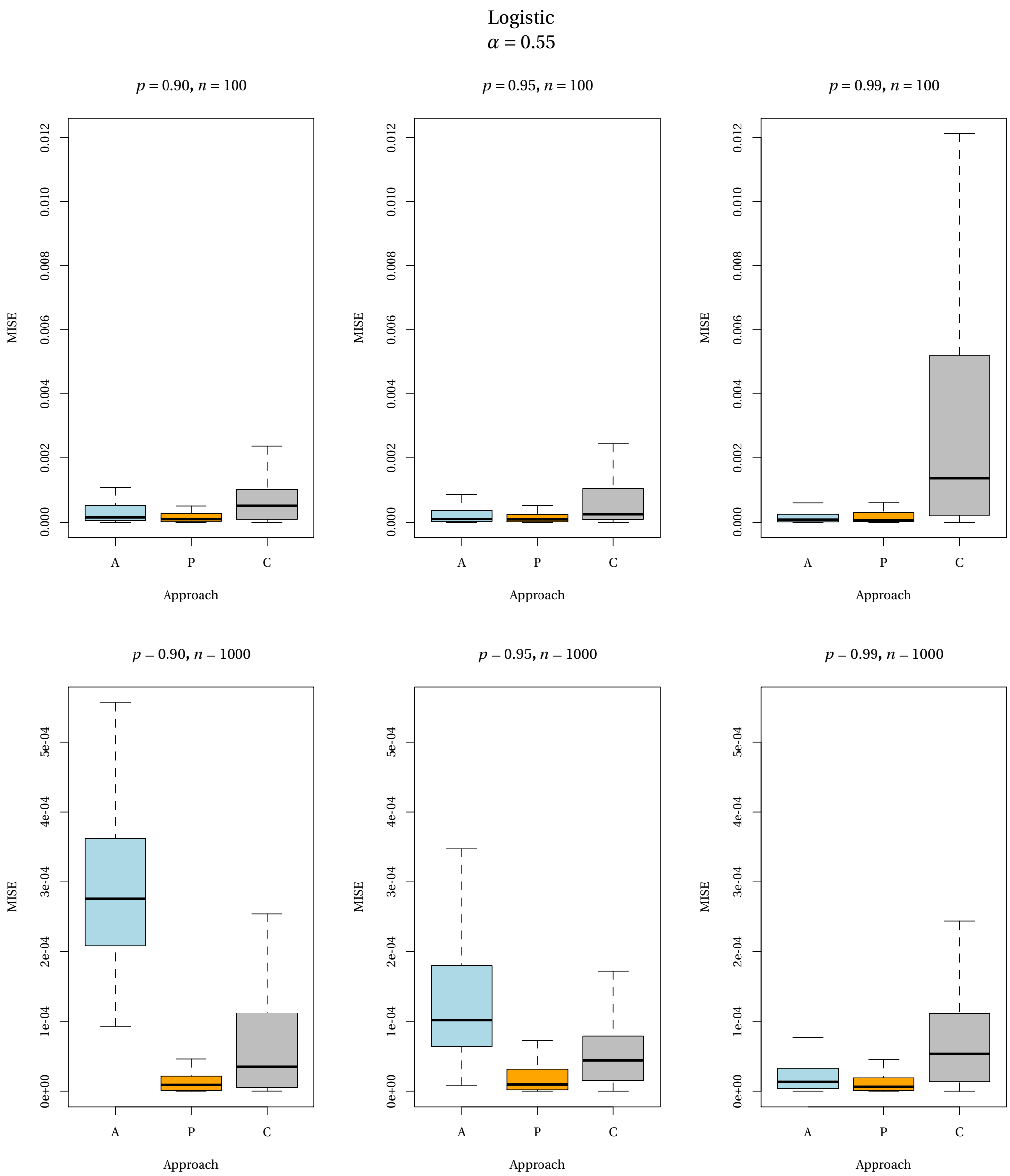

Figure B.3 – *Boxplots comparing the integrated squared error of maximum likelihood estimates of the logistic model based on data simulated from the logistic model with dependence parameter $\alpha = 0.55$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n.*

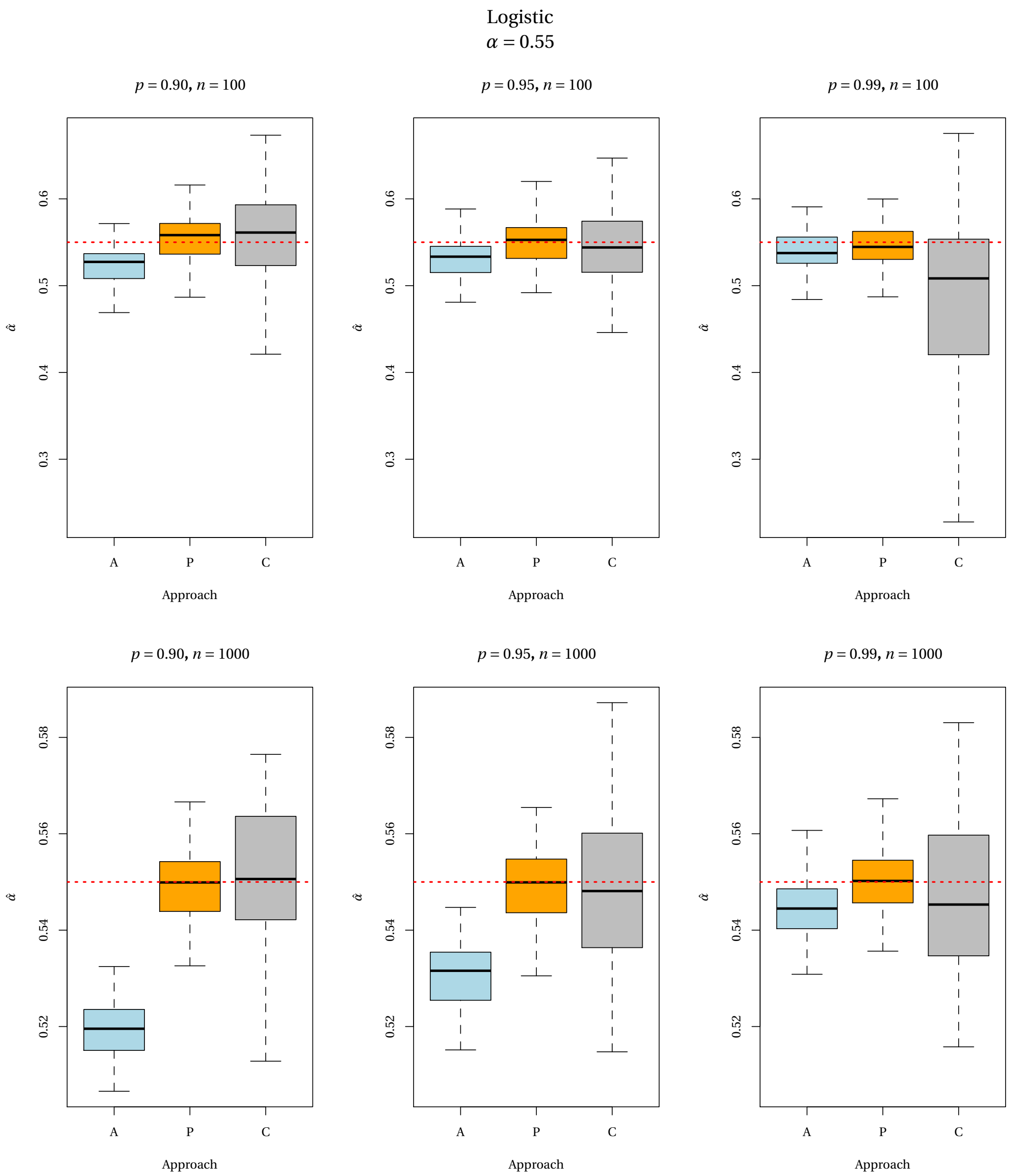

Figure B.4 – *Boxplots comparing dependence parameter estimates of the logistic model based on data simulated from the logistic model with dependence parameter* $\alpha = 0.55$ *using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level* p *and number of excesses* n*. The dashed red lines show the true parameter value.*

B.1.3 Case $\alpha = 0.8$

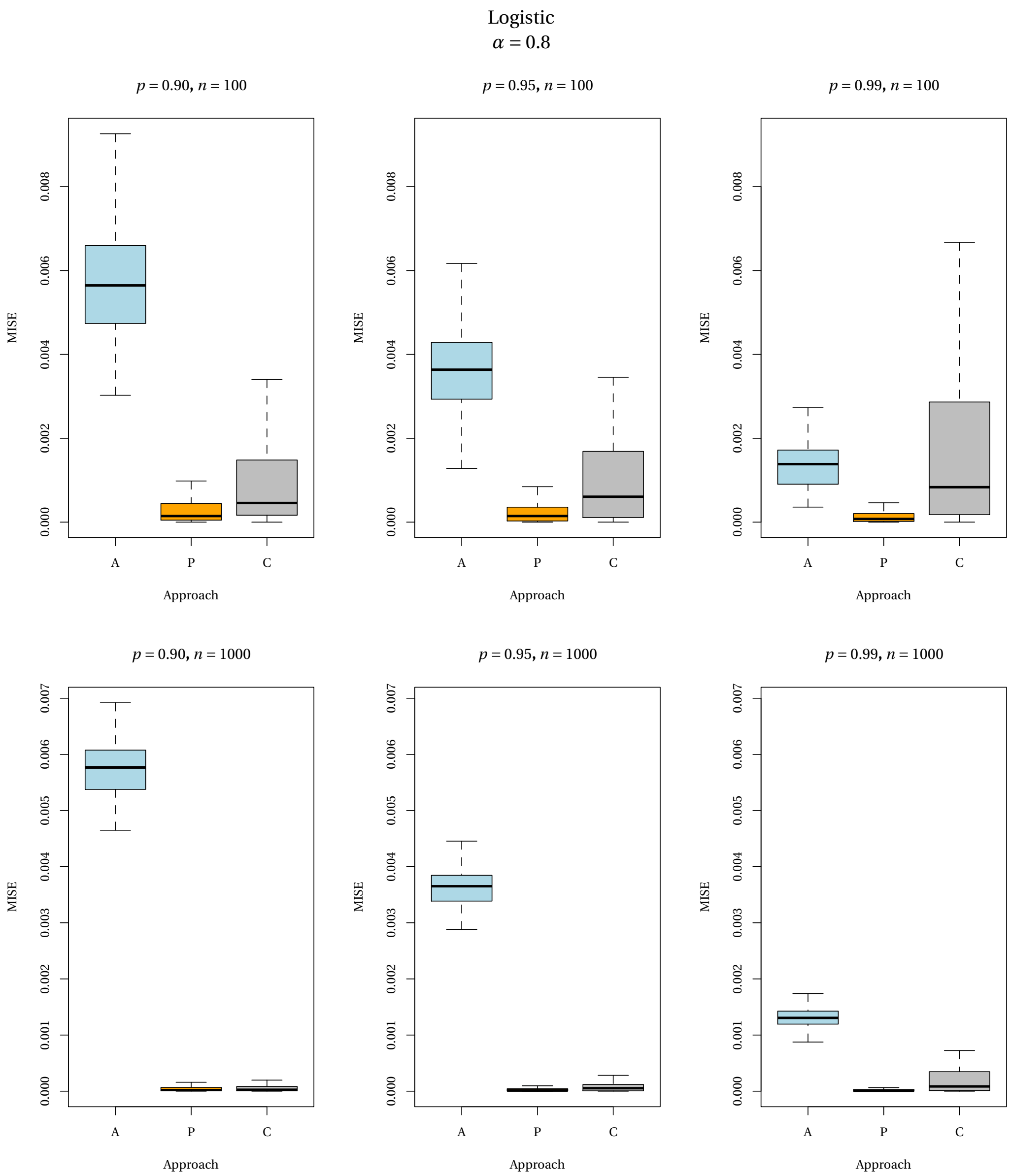

Figure B.5 – *Boxplots comparing the integrated squared error of maximum likelihood estimates of the logistic model based on data simulated from the logistic model with dependence parameter $\alpha = 0.8$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n.*

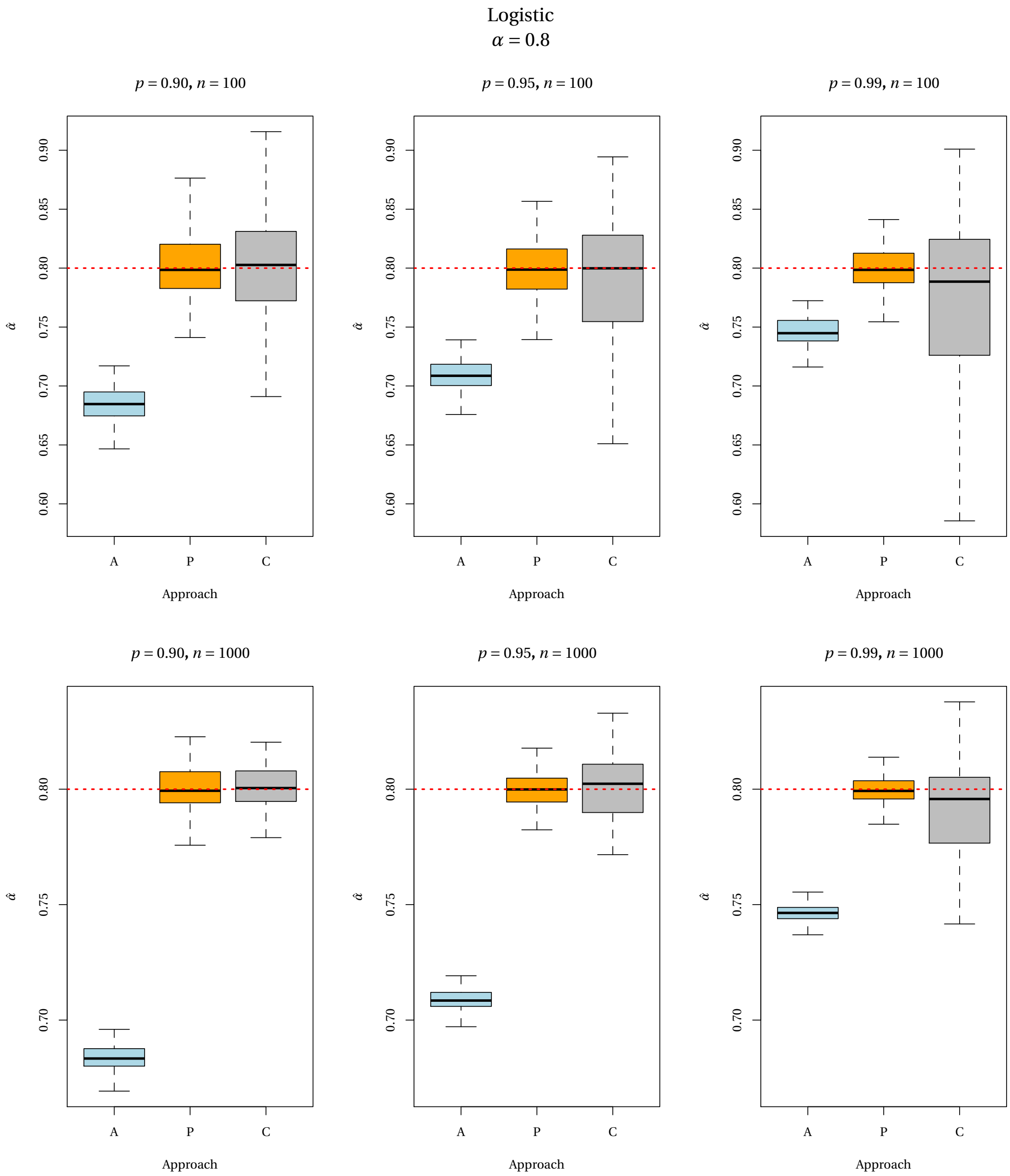

Figure B.6 – *Boxplots comparing dependence parameter estimates of the logistic model based on data simulated from the logistic model with dependence parameter* $\alpha = 0.8$ *using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level* p *and number of excesses* n*. The dashed red lines show the true parameter value.*

B.2 Hüsler–Reiss model

B.2.1 Case $\alpha = 0.5$

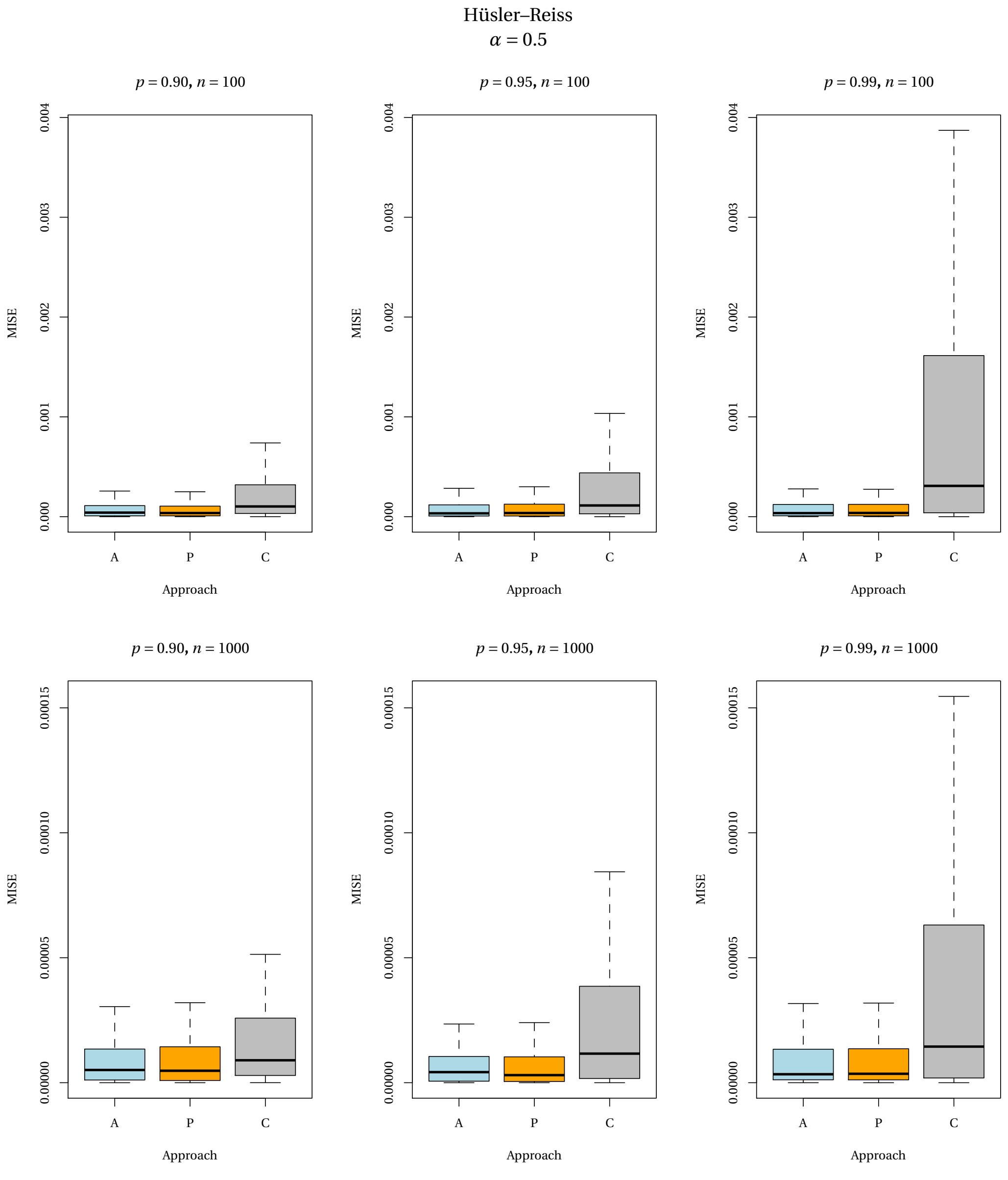

Figure B.7 – *Boxplots comparing the integrated squared error of maximum likelihood estimates of the Hüsler–Reiss model based on data simulated from the Hüsler–Reiss model with dependence parameter* $\alpha = 0.5$ *using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level* p *and number of excesses* n.

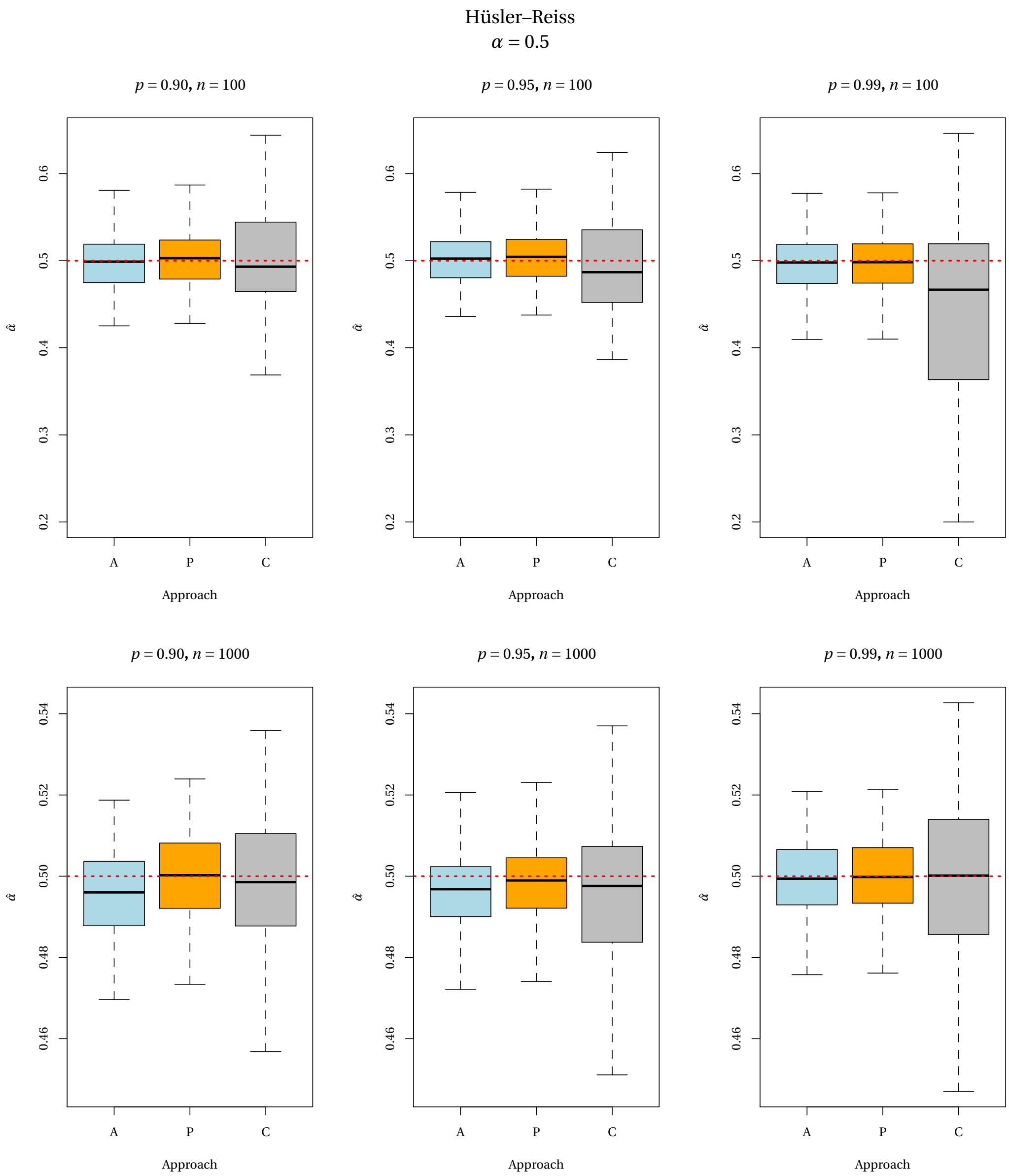

Figure B.8 – *Boxplots comparing dependence parameter estimates of the Hüsler–Reiss model based on data simulated from the Hüsler–Reiss model with dependence parameter $\alpha = 0.5$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n. The dashed red lines show the true parameter value.*

B.2.2 Case $\alpha = 1$

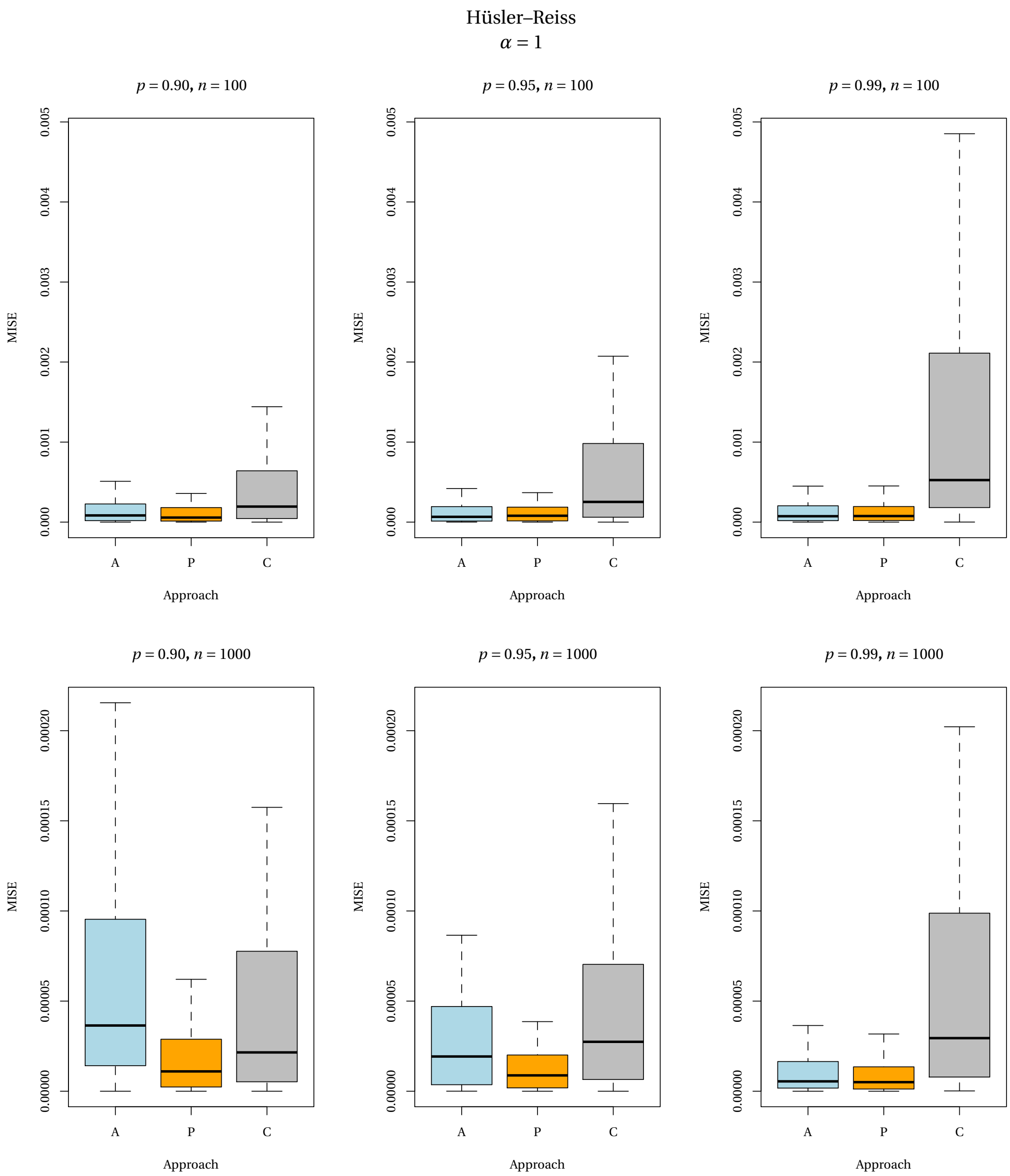

Figure B.9 – *Boxplots comparing the integrated squared error of maximum likelihood estimates of the Hüsler–Reiss model based on data simulated from the Hüsler–Reiss model with dependence parameter $\alpha = 1$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n.*

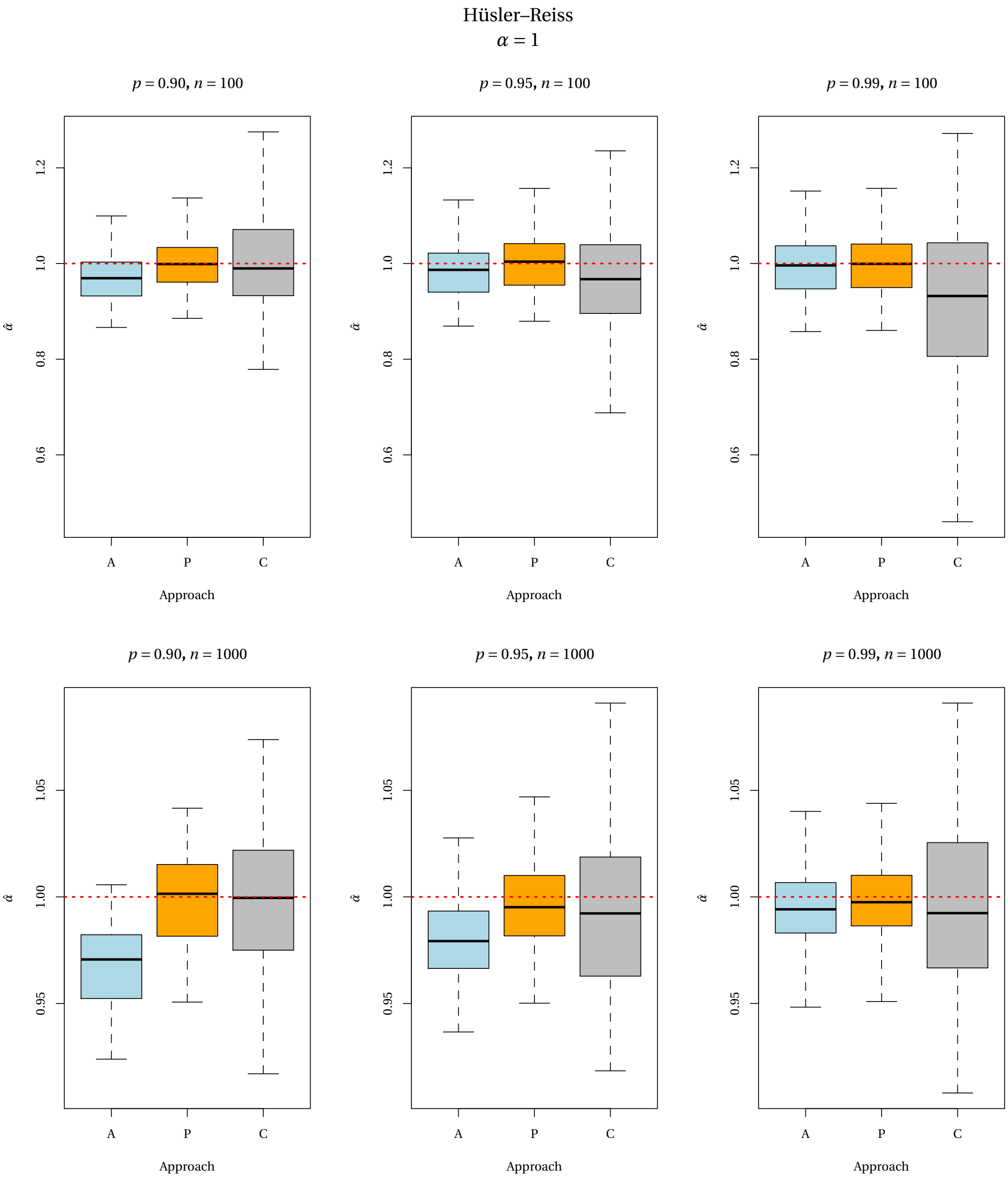

Figure B.10 – *Boxplots comparing dependence parameter estimates of the Hüsler–Reiss model based on data simulated from the Hüsler–Reiss model with dependence parameter $\alpha = 1$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n. The dashed red lines show the true parameter value.*

B.2.3 Case $\alpha = 2$

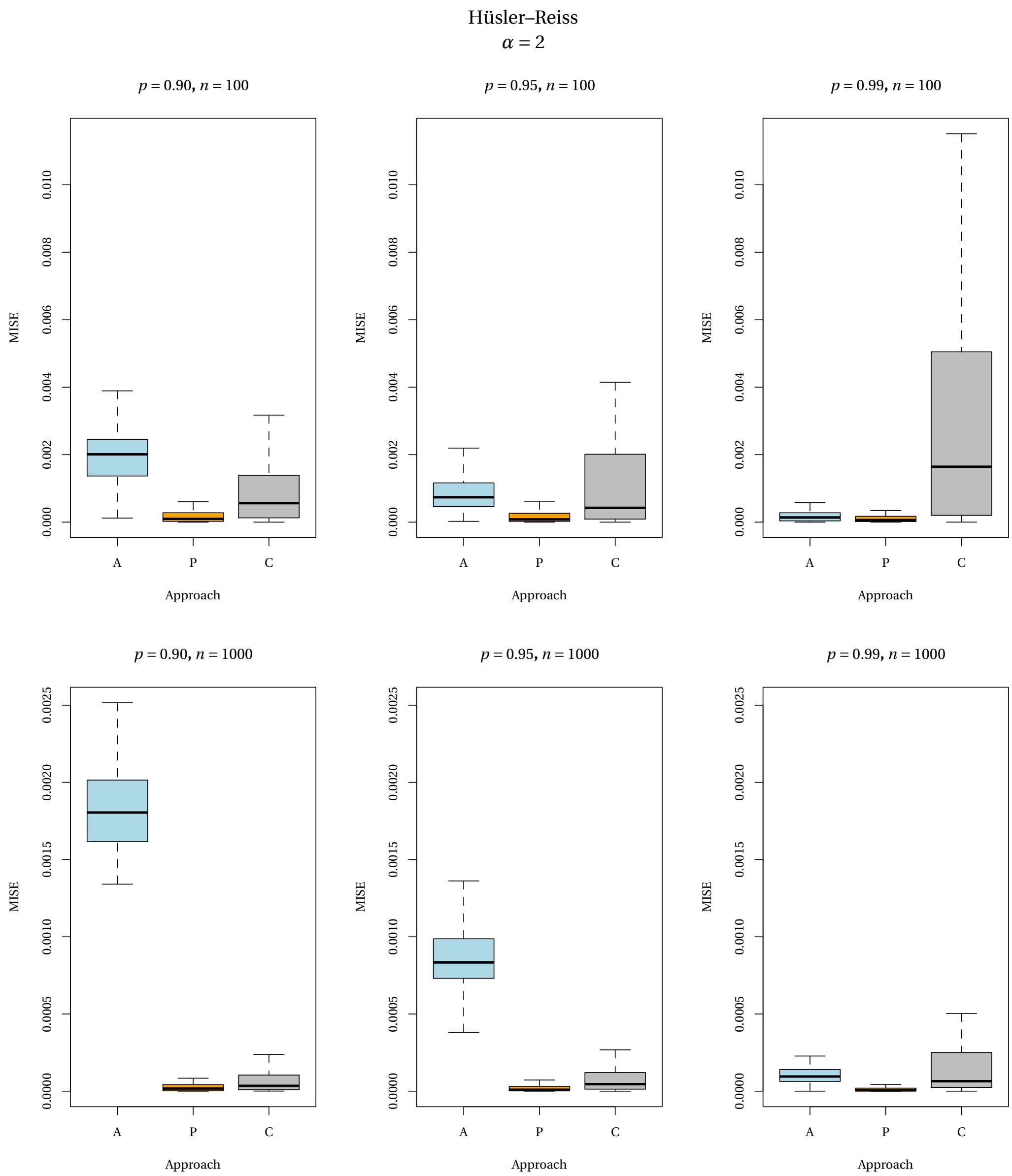

Figure B.11 – *Boxplots comparing the integrated squared error of maximum likelihood estimates of the Hüsler–Reiss model based on data simulated from the Hüsler–Reiss model with dependence parameter $\alpha = 2$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n.*

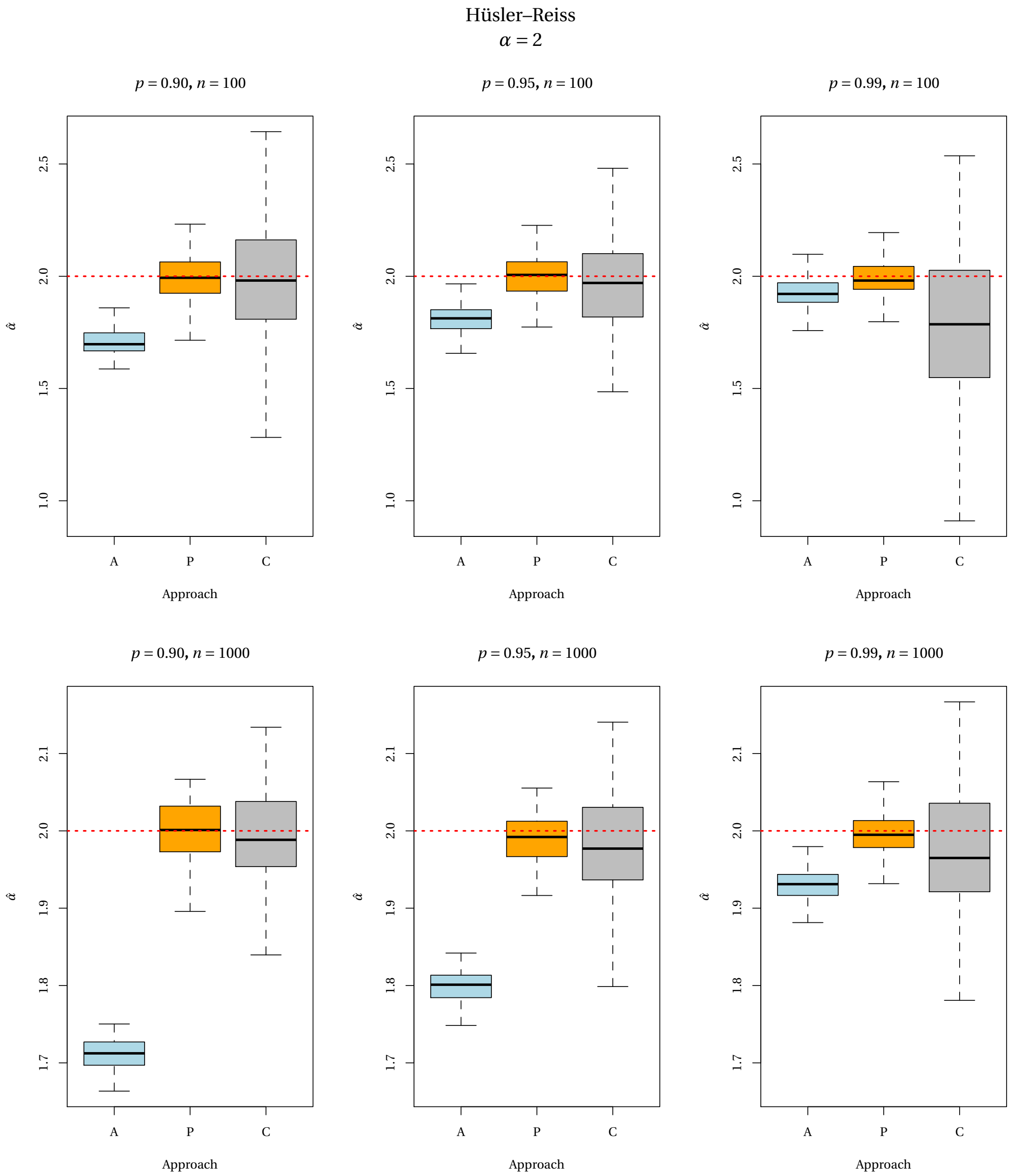

Figure B.12 – *Boxplots comparing dependence parameter estimates of the Hüsler–Reiss model based on data simulated from the Hüsler–Reiss model with dependence parameter $\alpha = 2$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n. The dashed red lines show the true parameter value.*

B.3 Dirichlet model

B.3.1 Case $\alpha = 0.2$ and $\beta = 0.2$

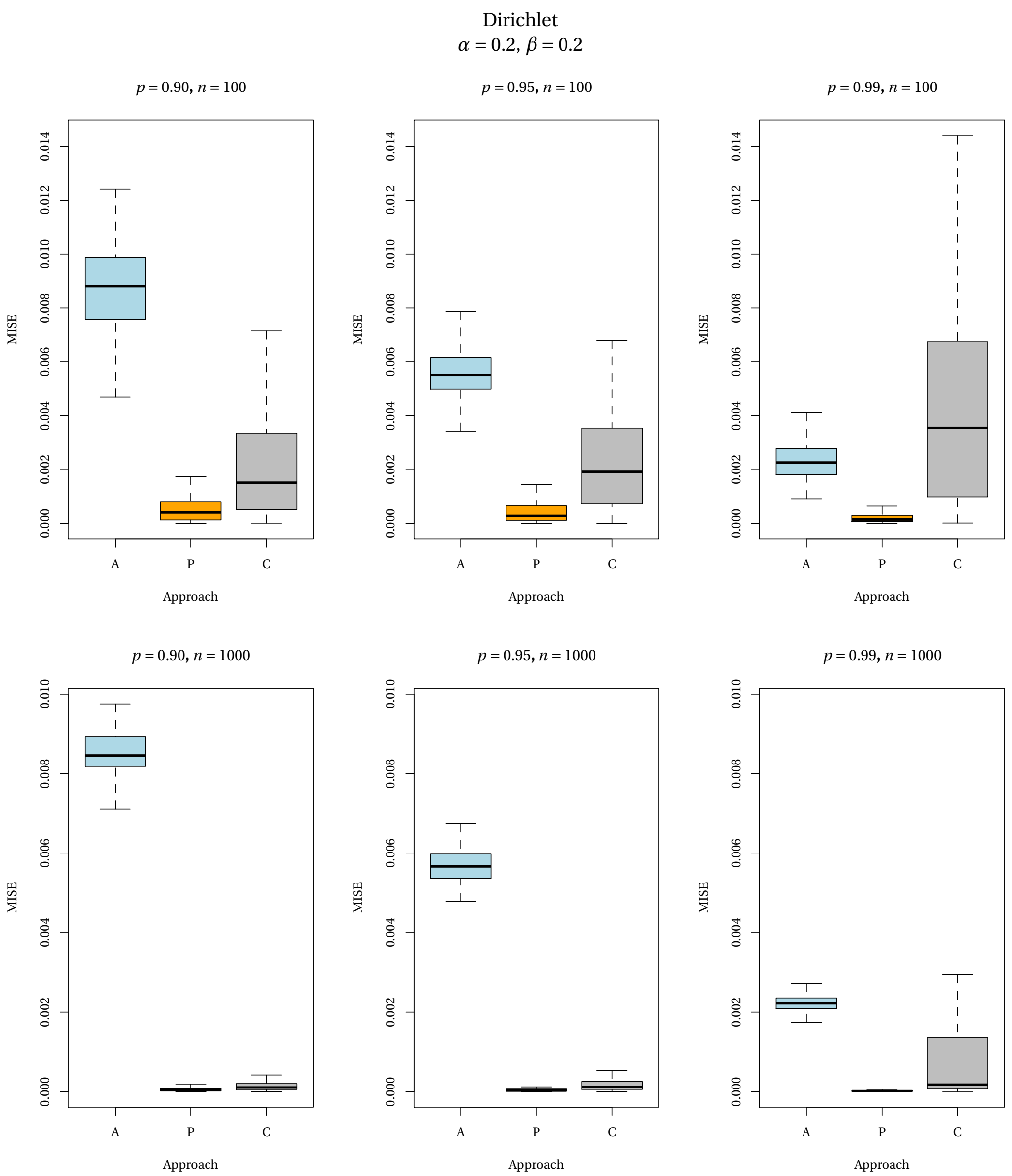

Figure B.13 – *Boxplots comparing the integrated squared error of maximum likelihood estimates of the Dirichlet model based on data simulated from the Dirichlet model with parameters $\alpha = 0.2$ and $\beta = 0.2$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n.*

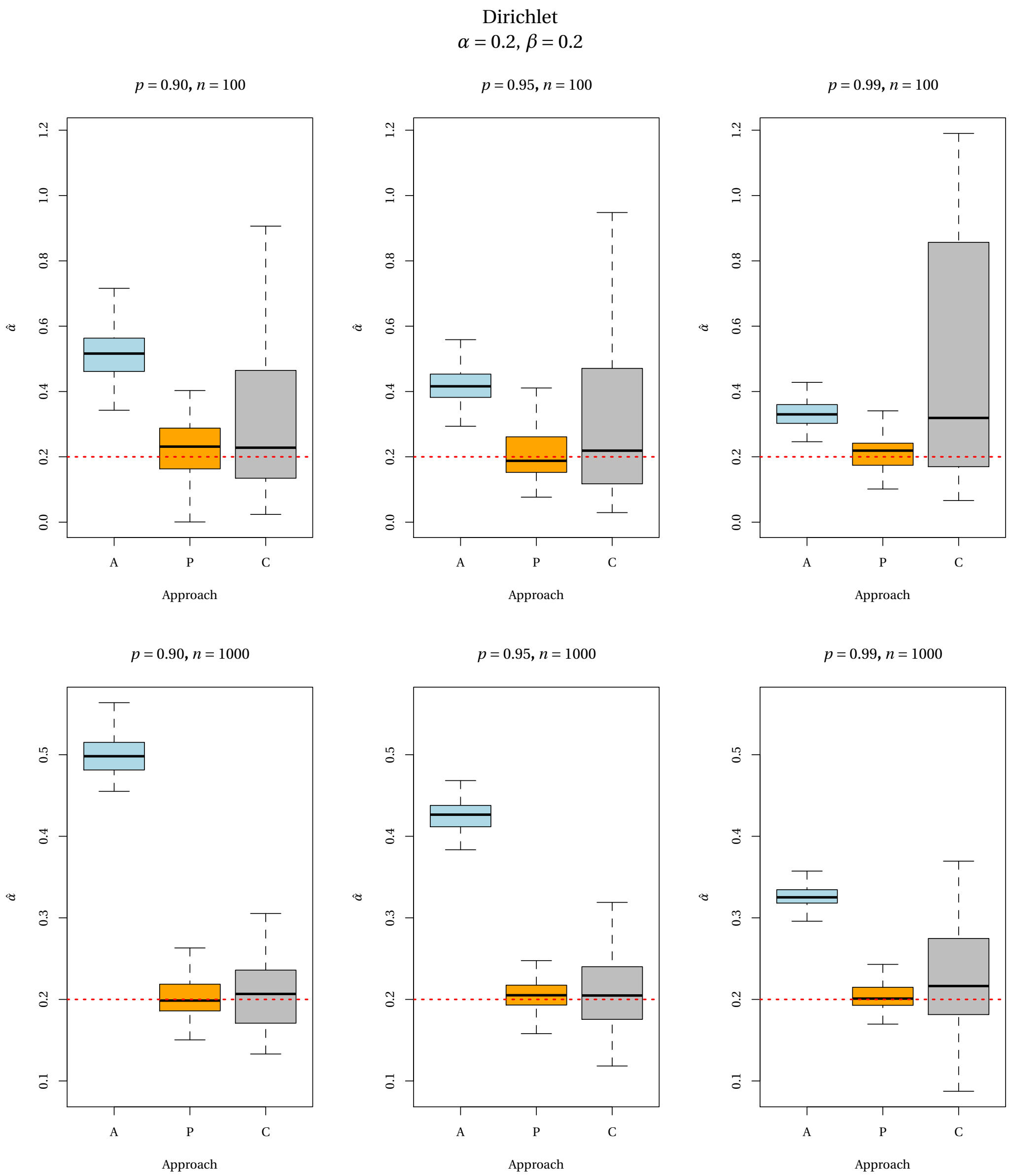

Figure B.14 – *Boxplots comparing first parameter estimates of the Dirichlet model based on data simulated from the Dirichlet model with parameters $\alpha = 0.2$ and $\beta = 0.2$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n. The dashed red lines show the true parameter value.*

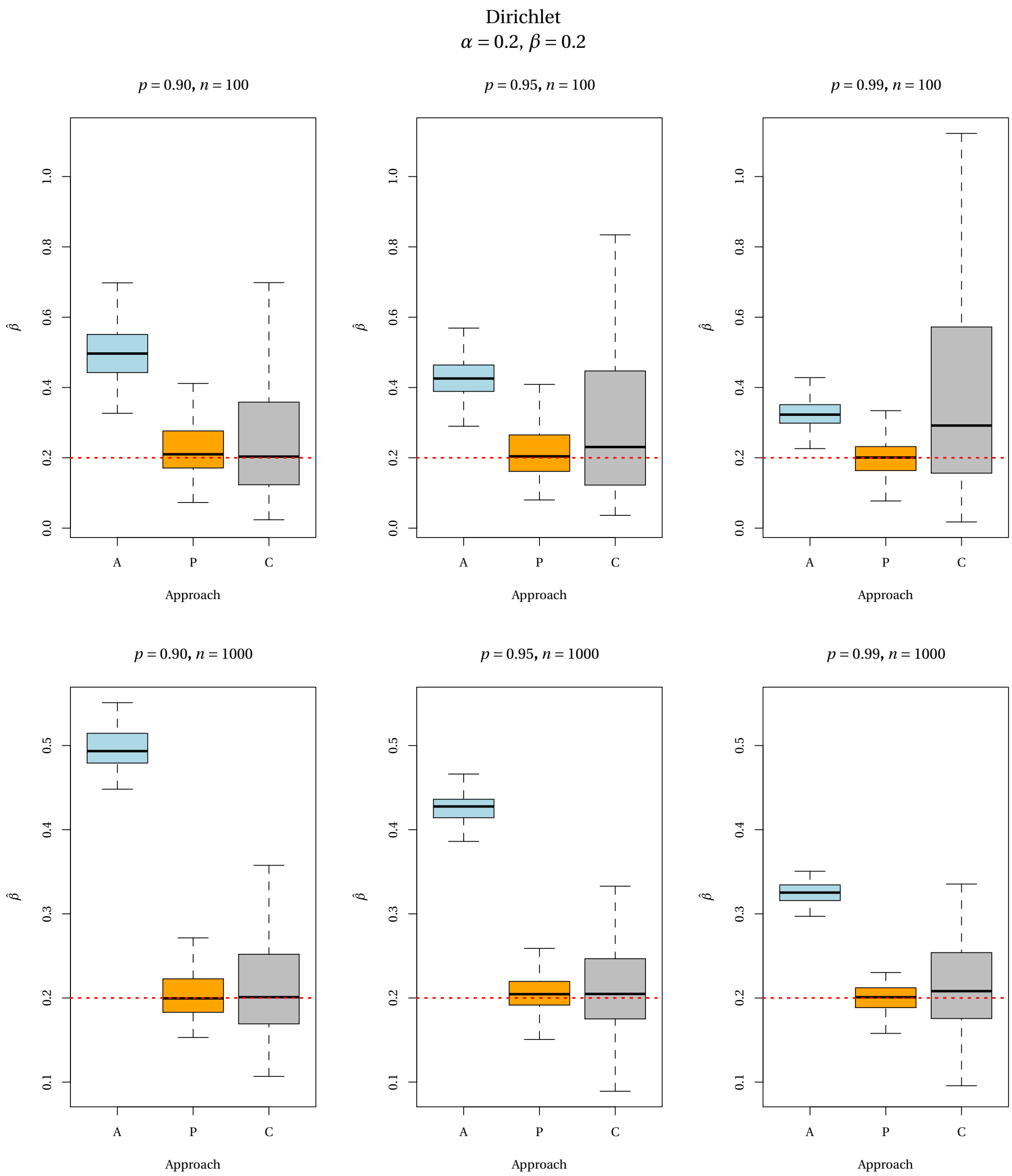

Figure B.15 – *Boxplots comparing second parameter estimates of the Dirichlet model based on data simulated from the Dirichlet model with parameters $\alpha = 0.2$ and $\beta = 0.2$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n. The dashed red lines show the true parameter value.*

B.3.2 Case $\alpha = 3$ and $\beta = 0.2$

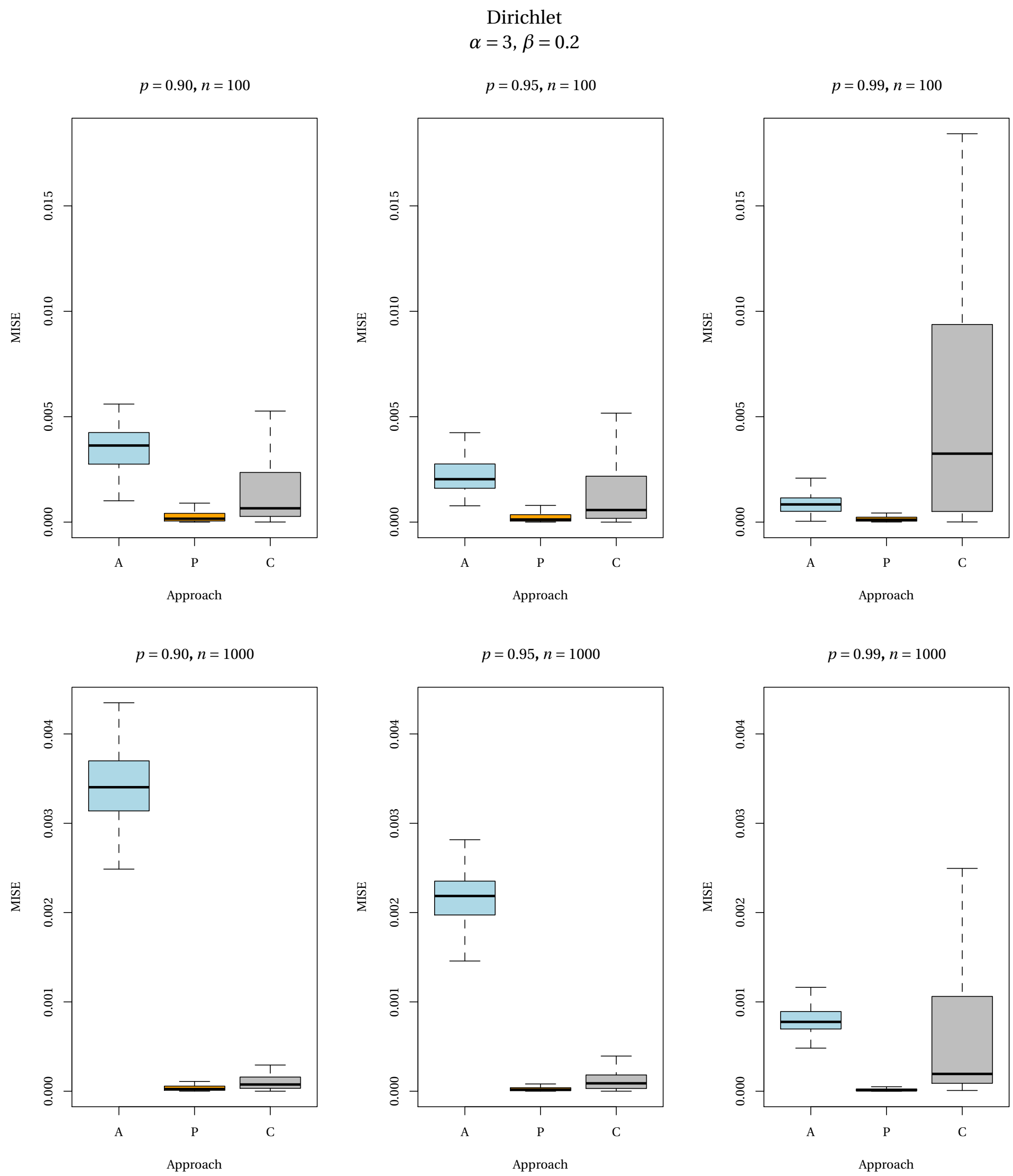

Figure B.16 – *Boxplots comparing the integrated squared error of maximum likelihood estimates of the Dirichlet model based on data simulated from the Dirichlet model with parameters $\alpha = 3$ and $\beta = 0.2$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n.*

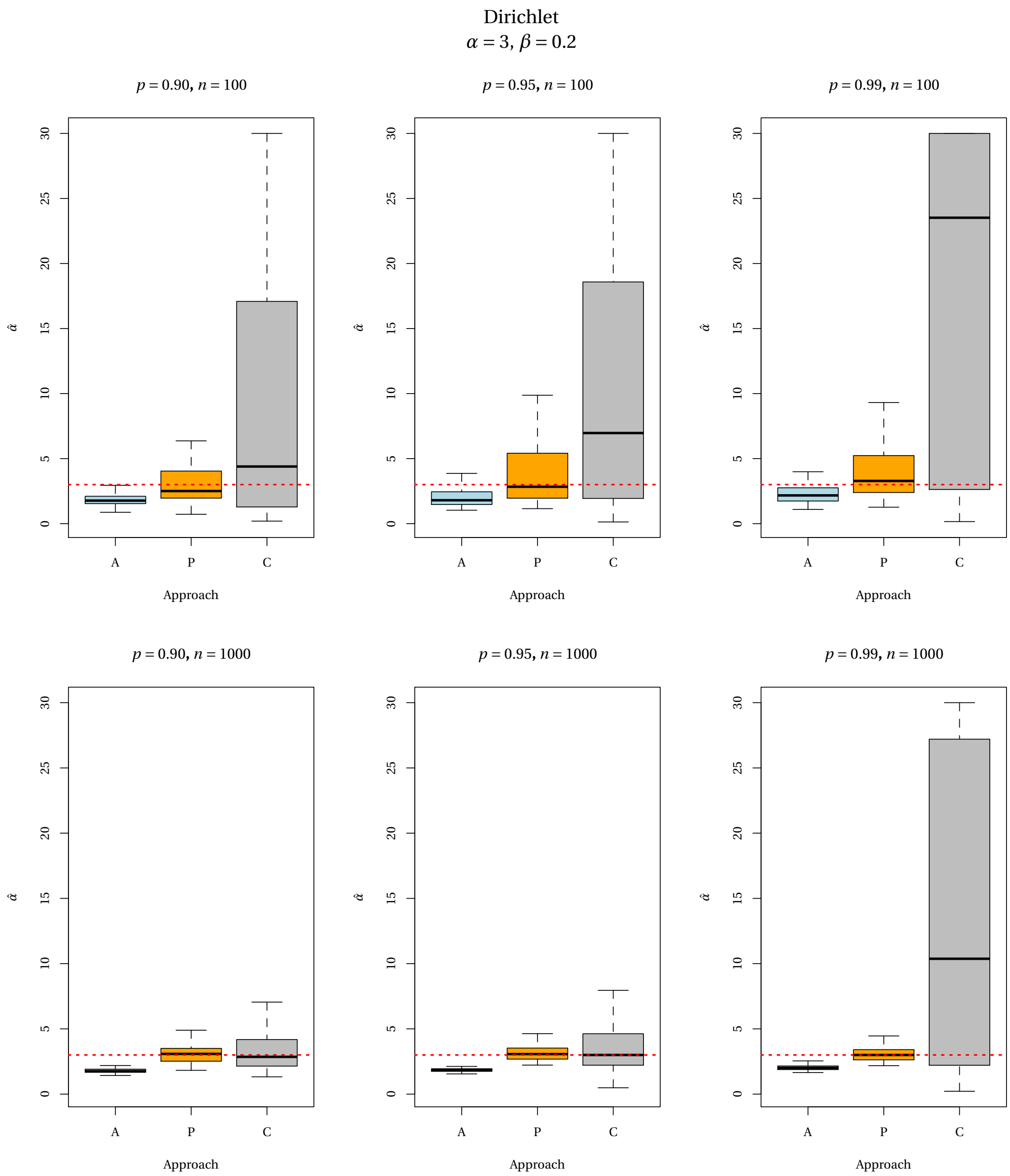

Figure B.17 – *Boxplots comparing first parameter estimates of the Dirichlet model based on data simulated from the Dirichlet model with parameters $\alpha = 3$ and $\beta = 0.2$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n. The dashed red lines show the true parameter value.*

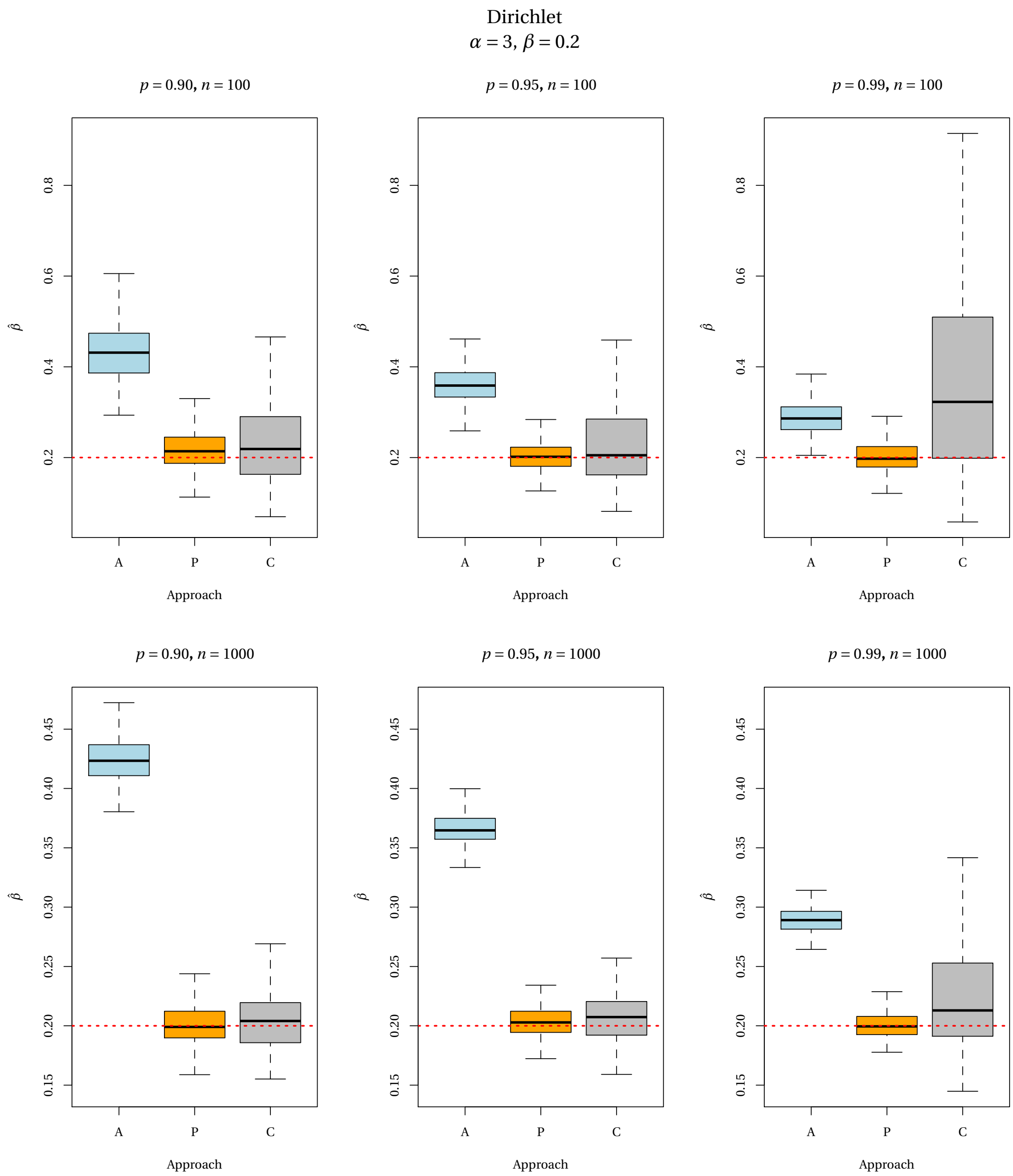

Figure B.18 – *Boxplots comparing second parameter estimates of the Dirichlet model based on data simulated from the Dirichlet model with parameters $\alpha = 3$ and $\beta = 0.2$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n. The dashed red lines show the true parameter value.*

B.3.3 Case $\alpha = 3$ and $\beta = 3$

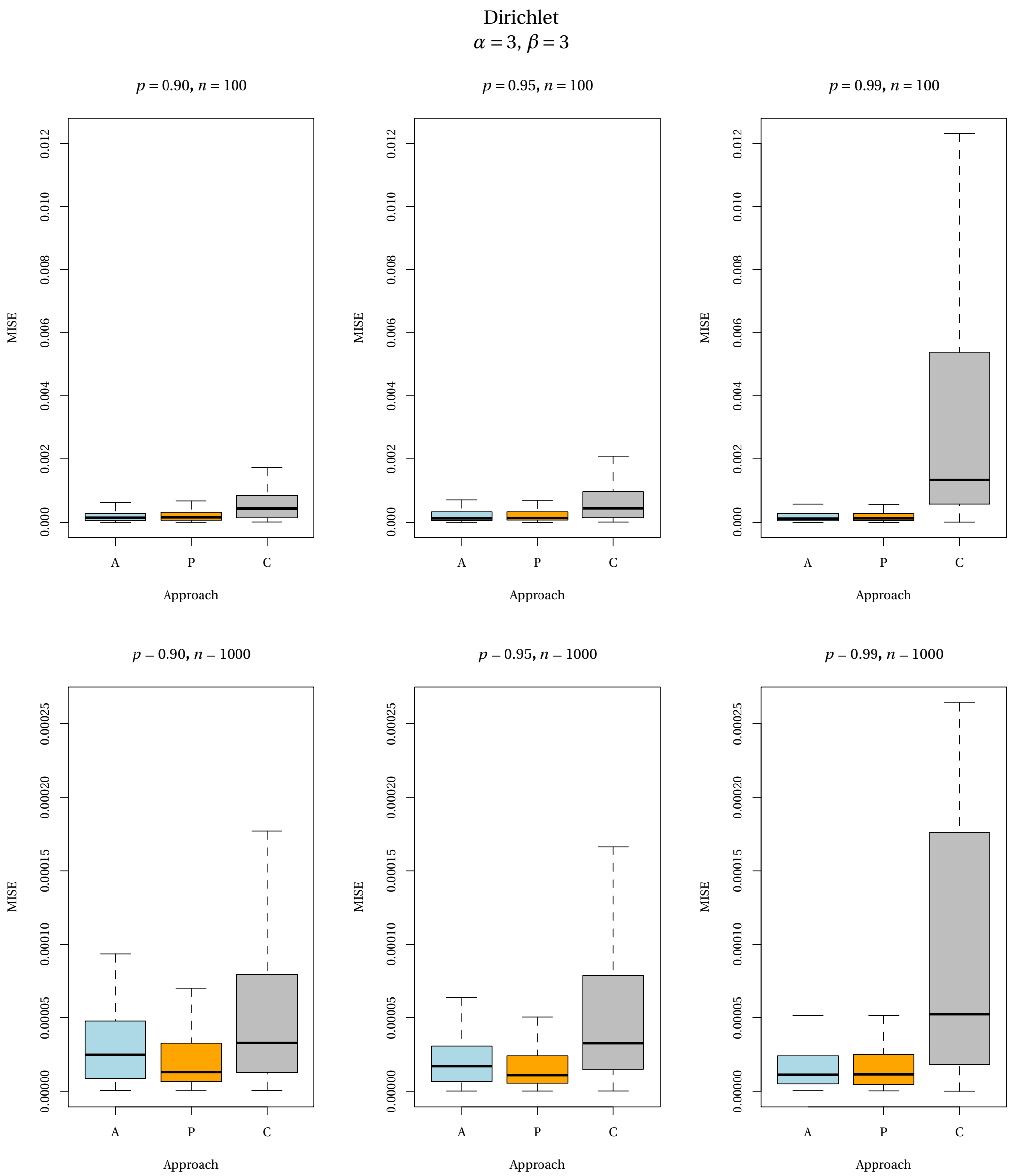

Figure B.19 – *Boxplots comparing the integrated squared error of maximum likelihood estimates of the Dirichlet model based on data simulated from the Dirichlet model with parameters $\alpha = 3$ and $\beta = 3$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n.*

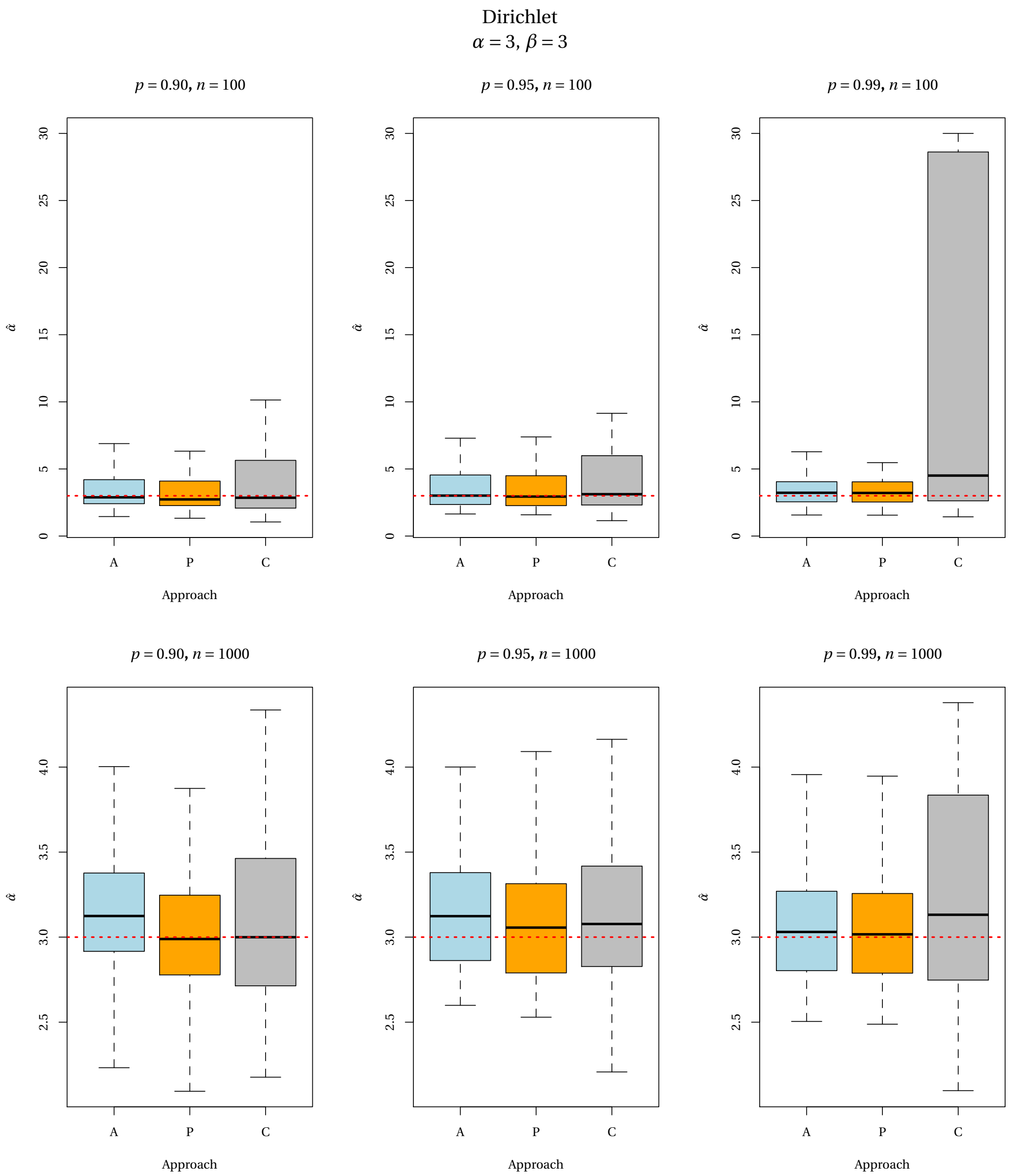

Figure B.20 – *Boxplots comparing first parameter estimates of the Dirichlet model based on data simulated from the Dirichlet model with parameters $\alpha = 3$ and $\beta = 3$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n. The dashed red lines show the true parameter value.*

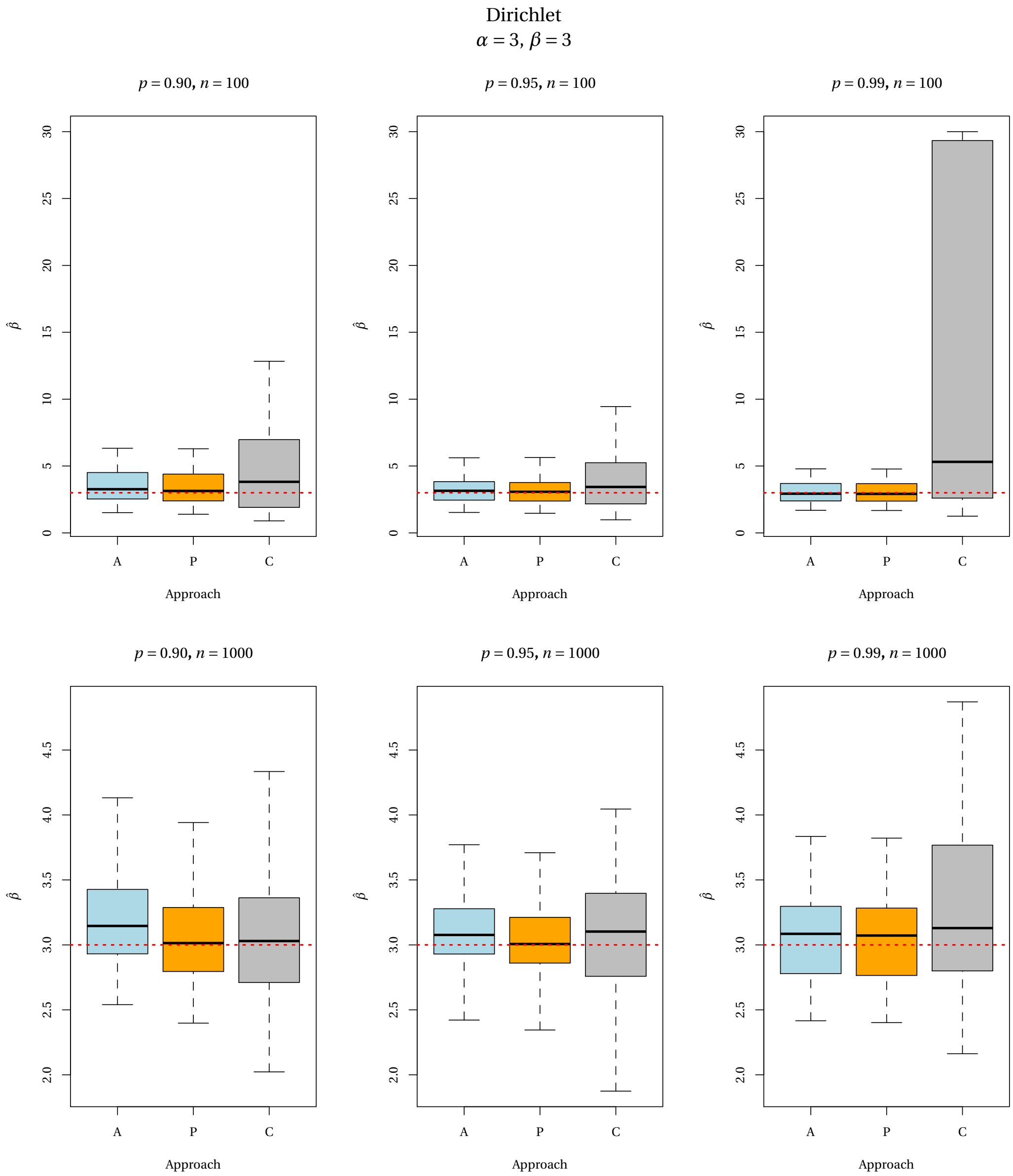

Figure B.21 – *Boxplots comparing second parameter estimates of the Dirichlet model based on data simulated from the Dirichlet model with parameters $\alpha = 3$ and $\beta = 3$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n. The dashed red lines show the true parameter value.*

B.4 Asymmetric logistic model

B.4.1 Case $\alpha = 0.2$ and $\psi = (0.8, 0.8)$

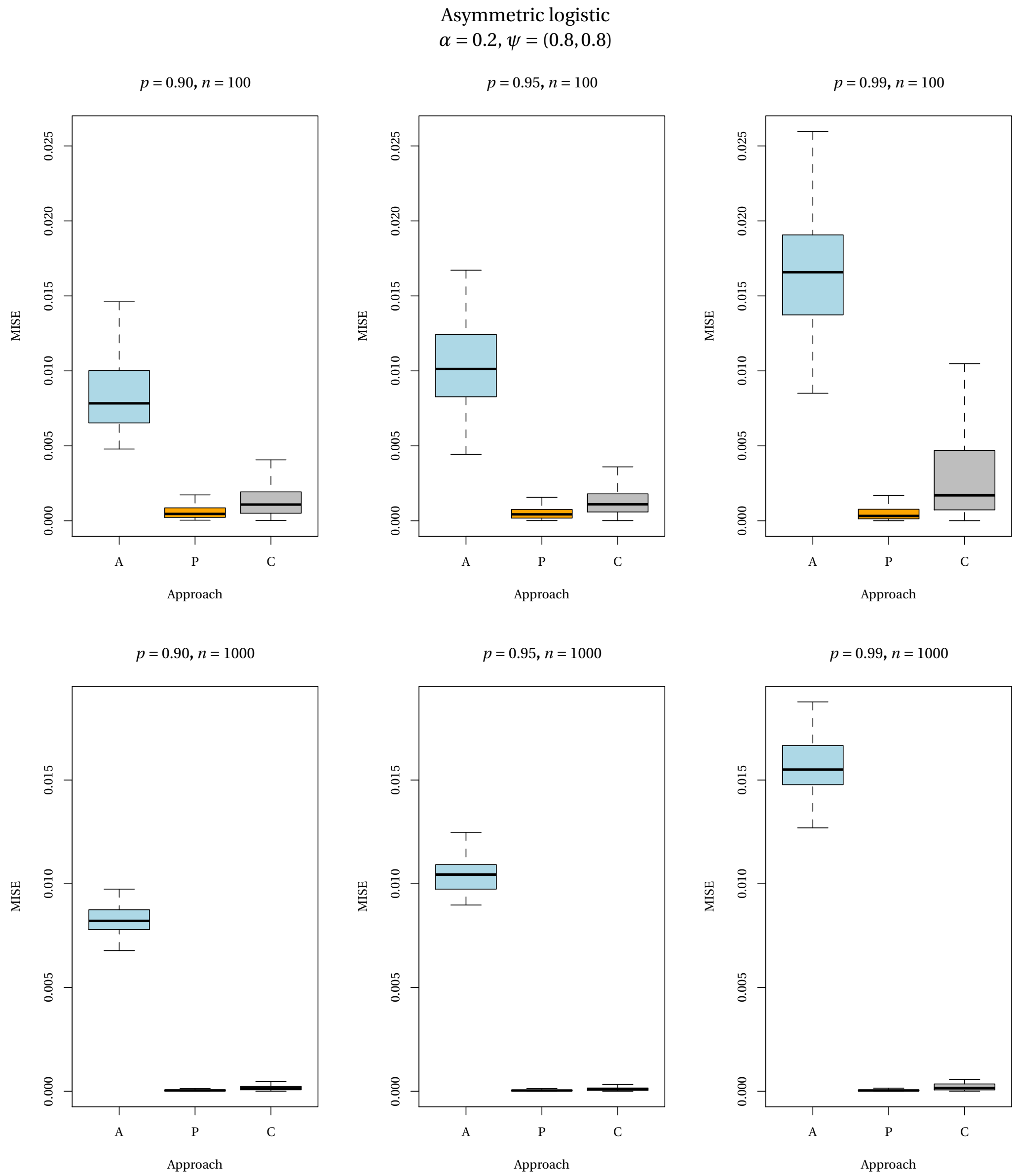

Figure B.22 – *Boxplots comparing the integrated squared error of maximum likelihood estimates of the asymmetric logistic model based on data simulated from the asymmetric logistic model with parameters $\alpha = 0.2$ and $\psi = (0.8, 0.8)$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n.*

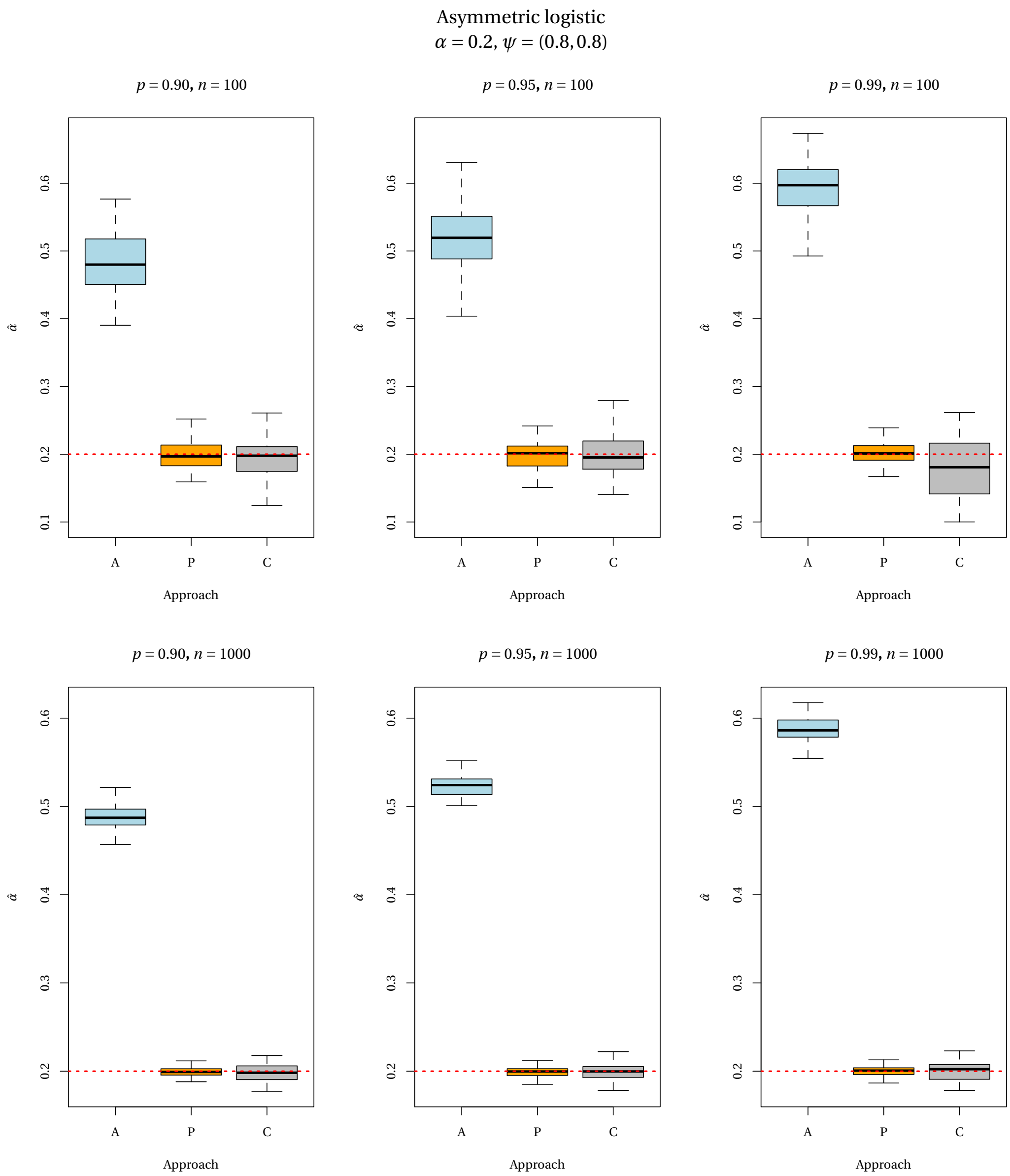

Figure B.23 – *Boxplots comparing dependence parameter estimates of the asymmetric logistic model based on data simulated from the asymmetric logistic model with parameters $\alpha = 0.2$ and $\psi = (0.8, 0.8)$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n. The dashed red lines show the true parameter value.*

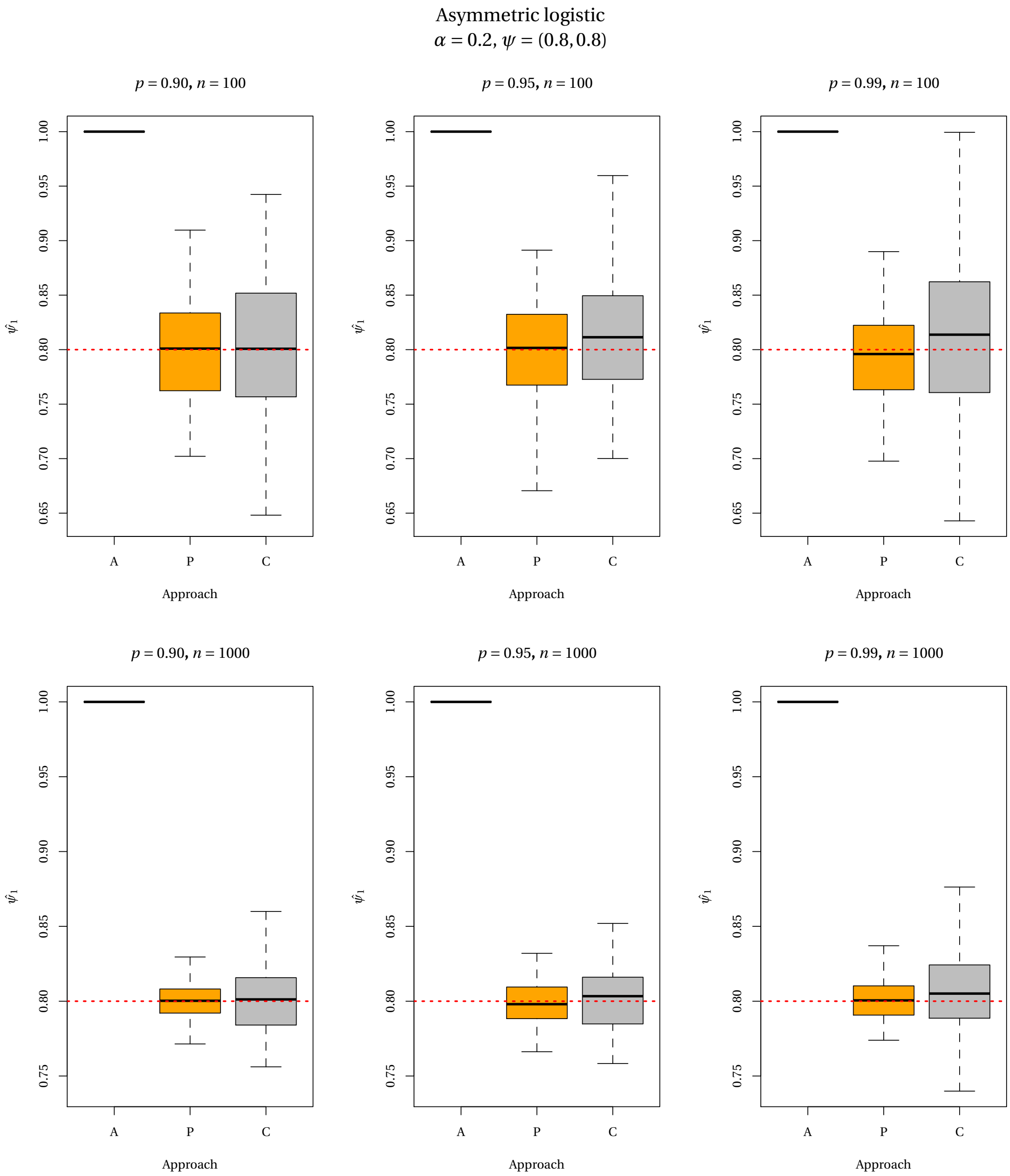

Figure B.24 – *Boxplots comparing first asymmetry parameter estimates of the asymmetric logistic model based on data simulated from the asymmetric logistic model with parameters* $\alpha = 0.2$ *and* $\psi = (0.8, 0.8)$ *using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n. The dashed red lines show the true parameter value.*

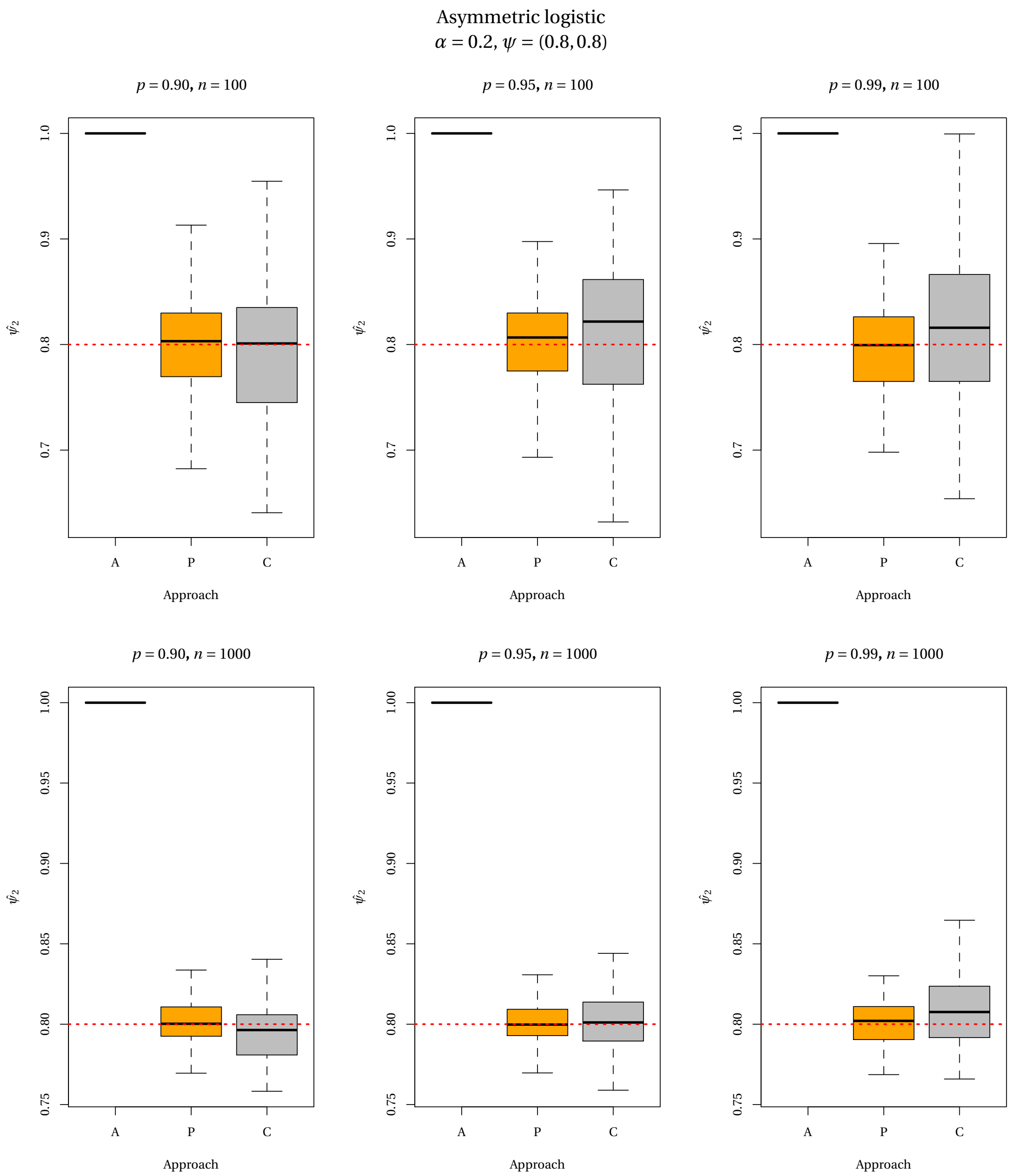

Figure B.25 – *Boxplots comparing second asymmetry parameter estimates of the asymmetric logistic model based on data simulated from the asymmetric logistic model with parameters* $\alpha = 0.2$ *and* $\psi = (0.8, 0.8)$ *using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level* p *and number of excesses* n*. The dashed red lines show the true parameter value.*

B.4.2 Case $\alpha = 0.4$ and $\psi = (0.7, 0.5)$

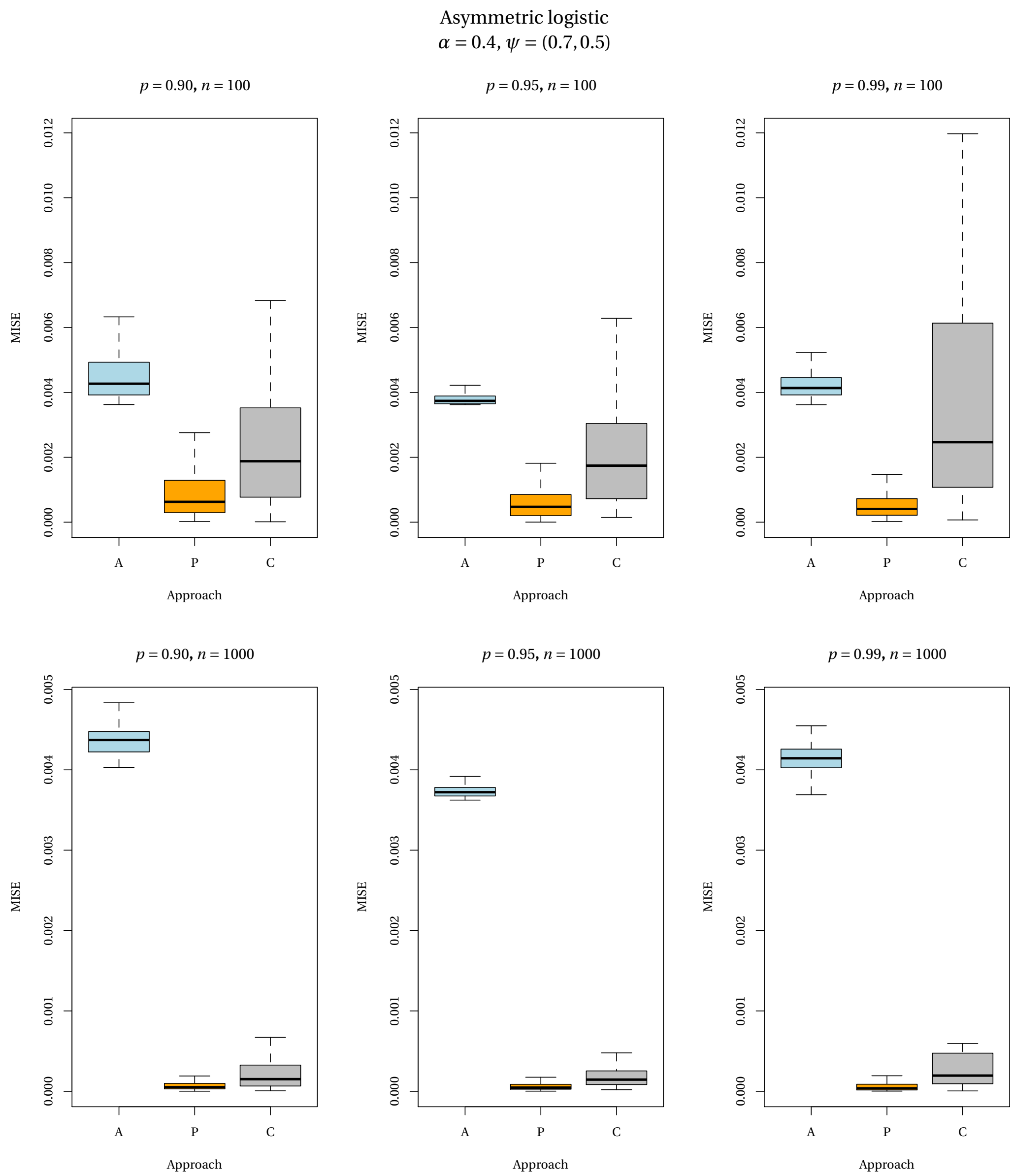

Figure B.26 – *Boxplots comparing the integrated squared error of maximum likelihood estimates of the asymmetric logistic model based on data simulated from the asymmetric logistic model with parameters $\alpha = 0.4$ and $\psi = (0.7, 0.5)$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n.*

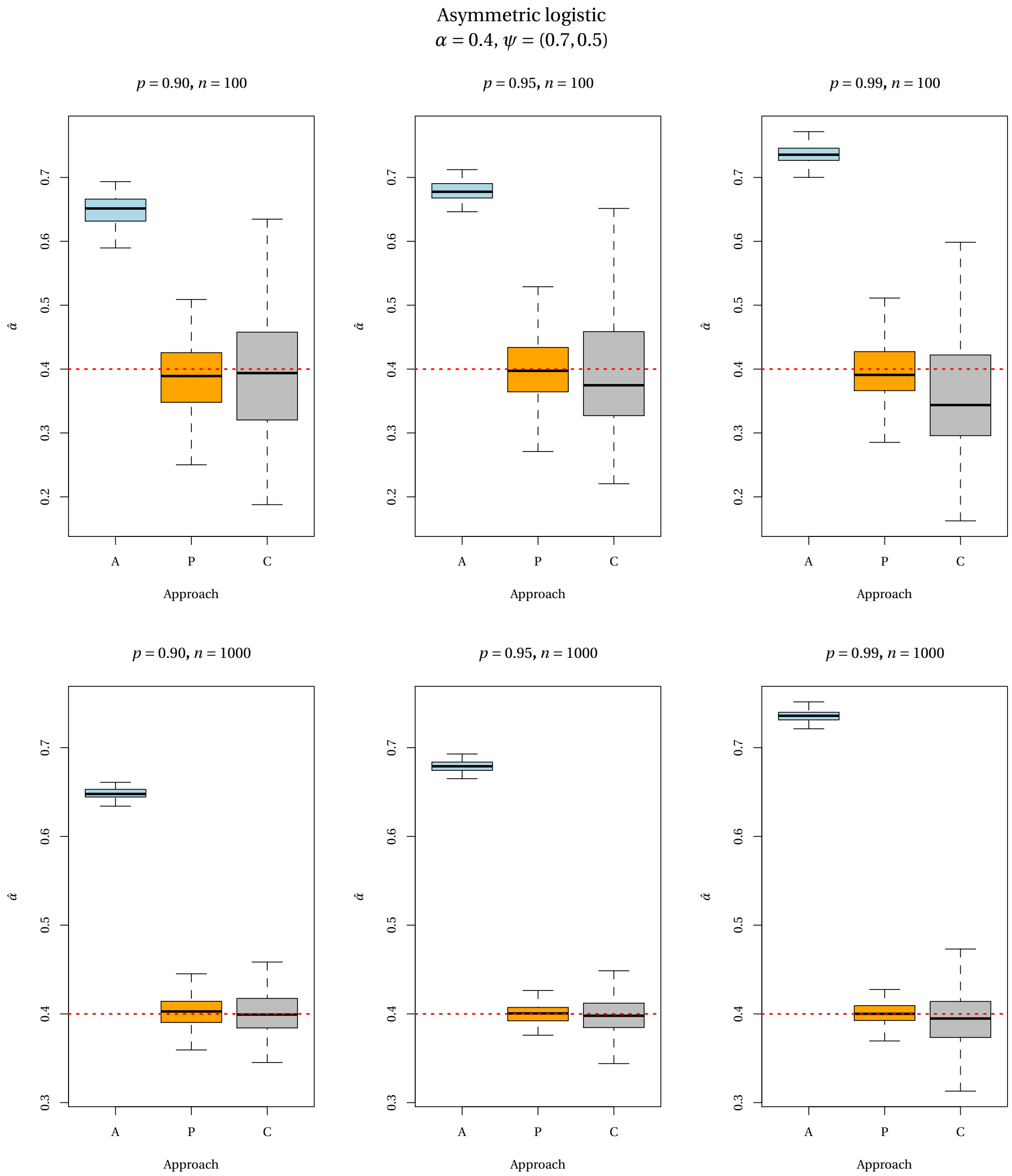

Figure B.27 – *Boxplots comparing dependence parameter estimates of the asymmetric logistic model based on data simulated from the asymmetric logistic model with parameters* $\alpha = 0.4$ *and* $\psi = (0.7, 0.5)$ *using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses* n*. The dashed red lines show the true parameter value.*

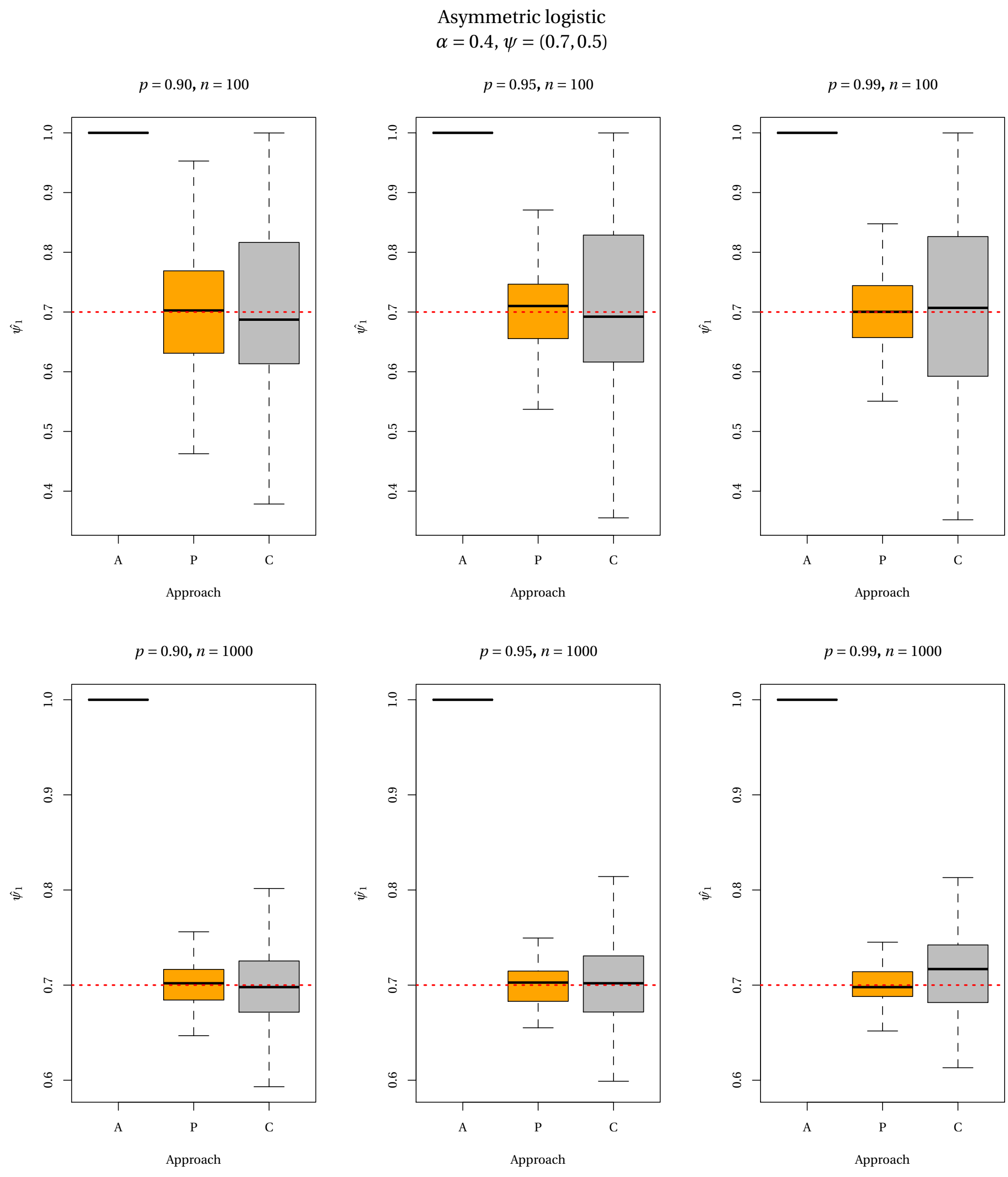

Figure B.28 – *Boxplots comparing first asymmetry parameter estimates of the asymmetric logistic model based on data simulated from the asymmetric logistic model with parameters* $\alpha = 0.4$ *and* $\psi = (0.7, 0.5)$ *using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n. The dashed red lines show the true parameter value.*

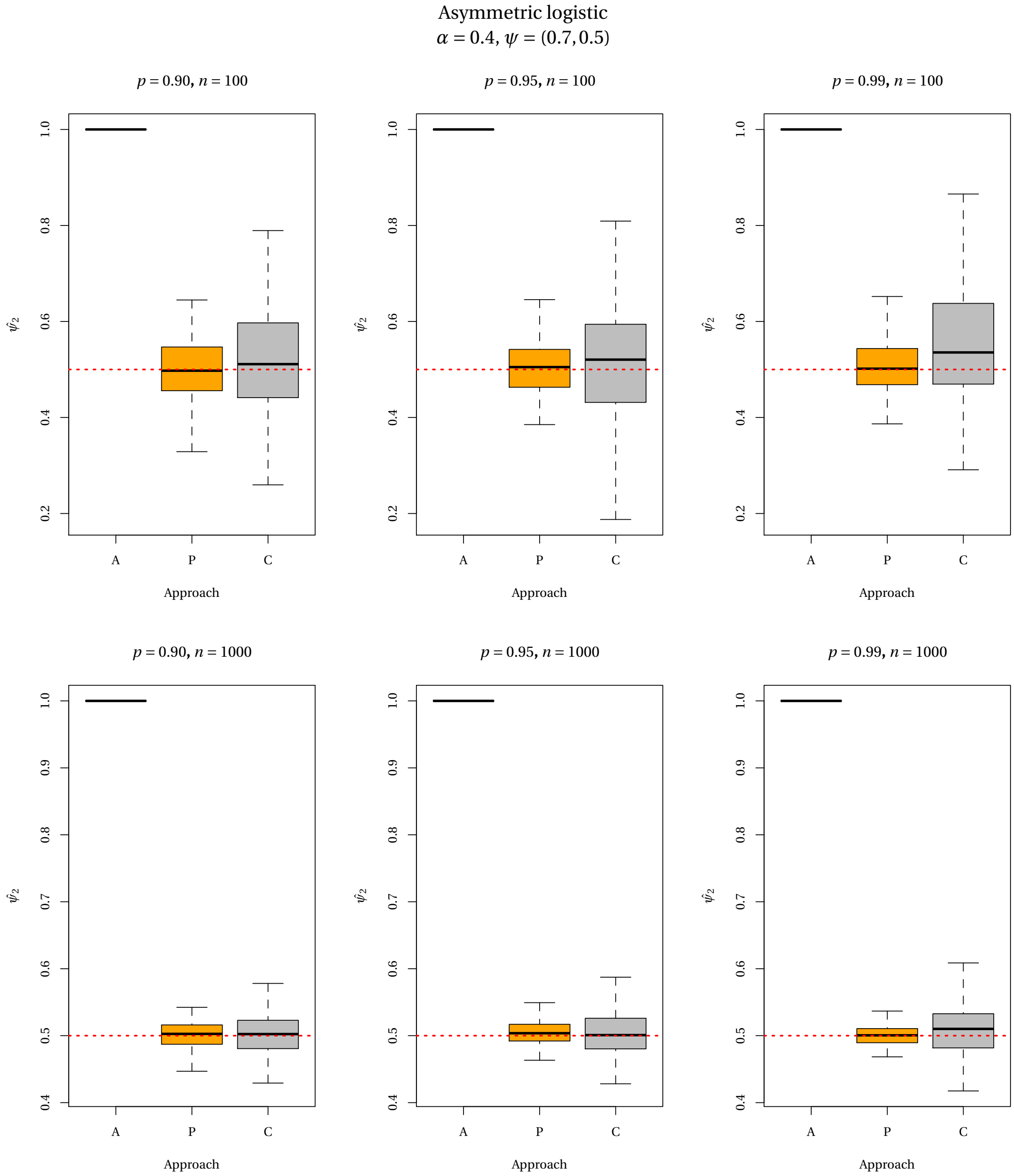

Figure B.29 – *Boxplots comparing second asymmetry parameter estimates of the asymmetric logistic model based on data simulated from the asymmetric logistic model with parameters $\alpha = 0.4$ and $\psi = (0.7, 0.5)$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n. The dashed red lines show the true parameter value.*

B.4.3 Case $\alpha = 0.6$ and $\psi = (0.6, 0.2)$

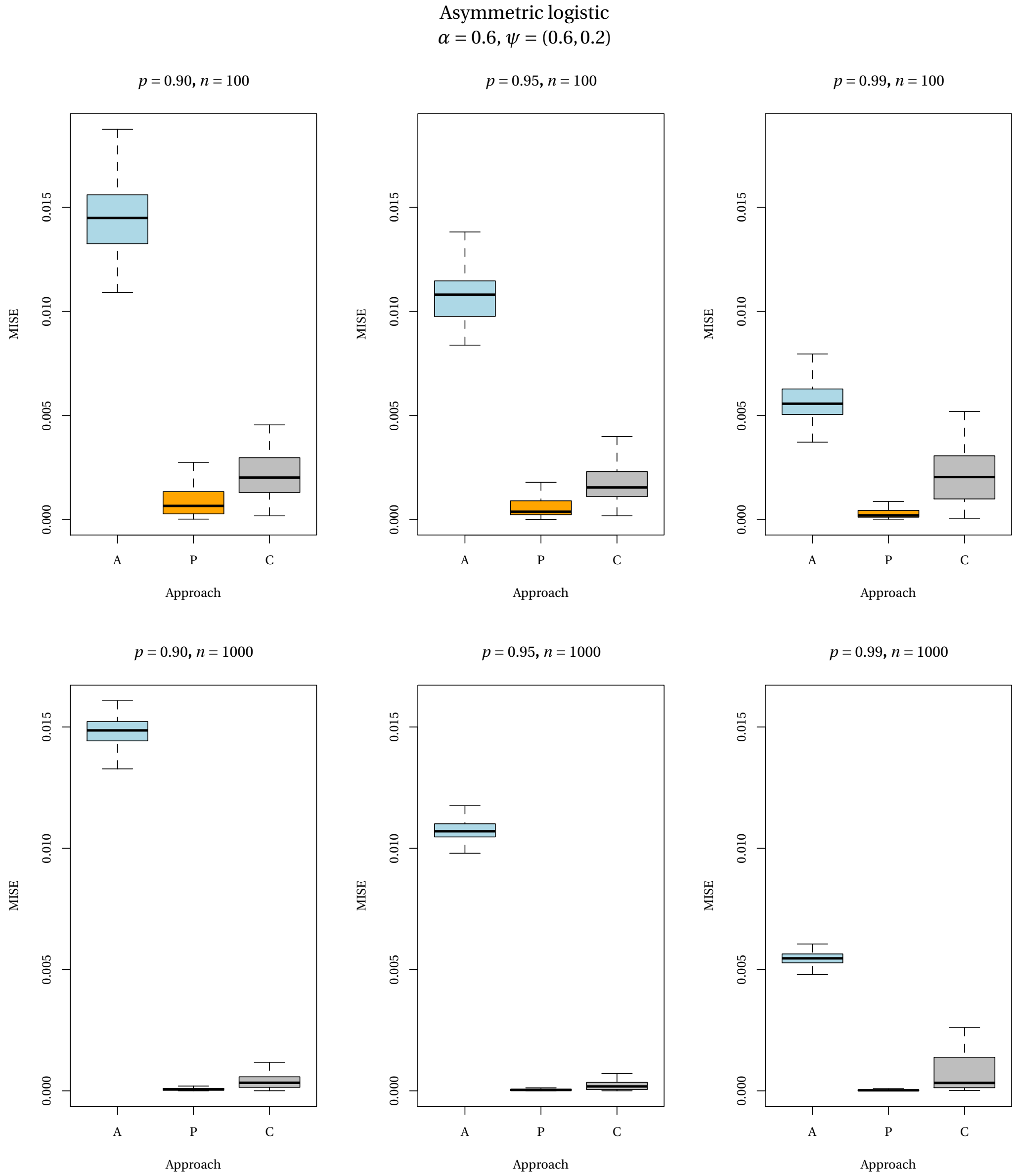

Figure B.30 – *Boxplots comparing the integrated squared error of maximum likelihood estimates of the asymmetric logistic model based on data simulated from the asymmetric logistic model with parameters* $\alpha = 0.6$ *and* $\psi = (0.6, 0.2)$ *using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n.*

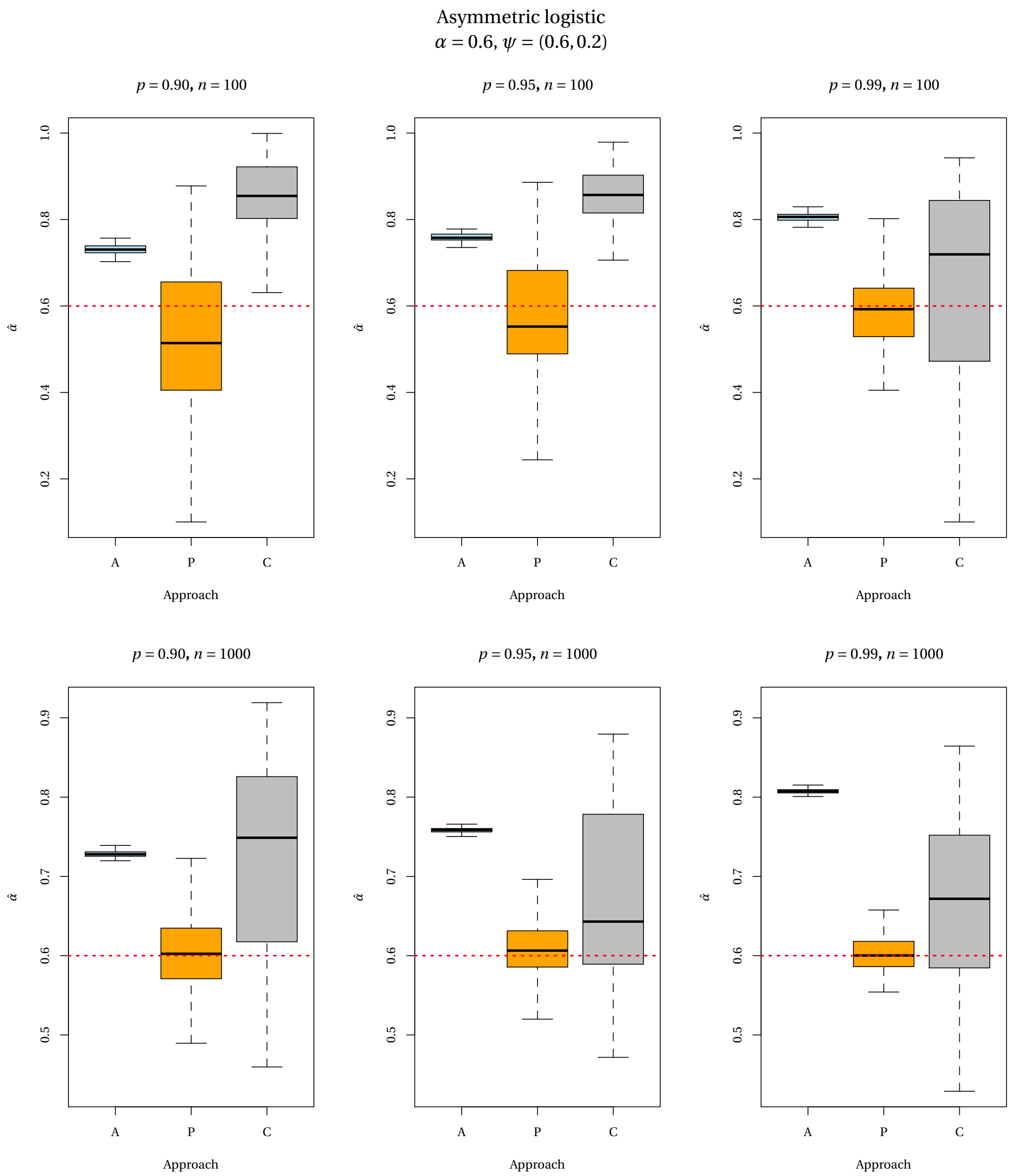

Figure B.31 – *Boxplots comparing dependence parameter estimates of the asymmetric logistic model based on data simulated from the asymmetric logistic model with parameters $\alpha = 0.6$ and $\psi = (0.6, 0.2)$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n. The dashed red lines show the true parameter value.*

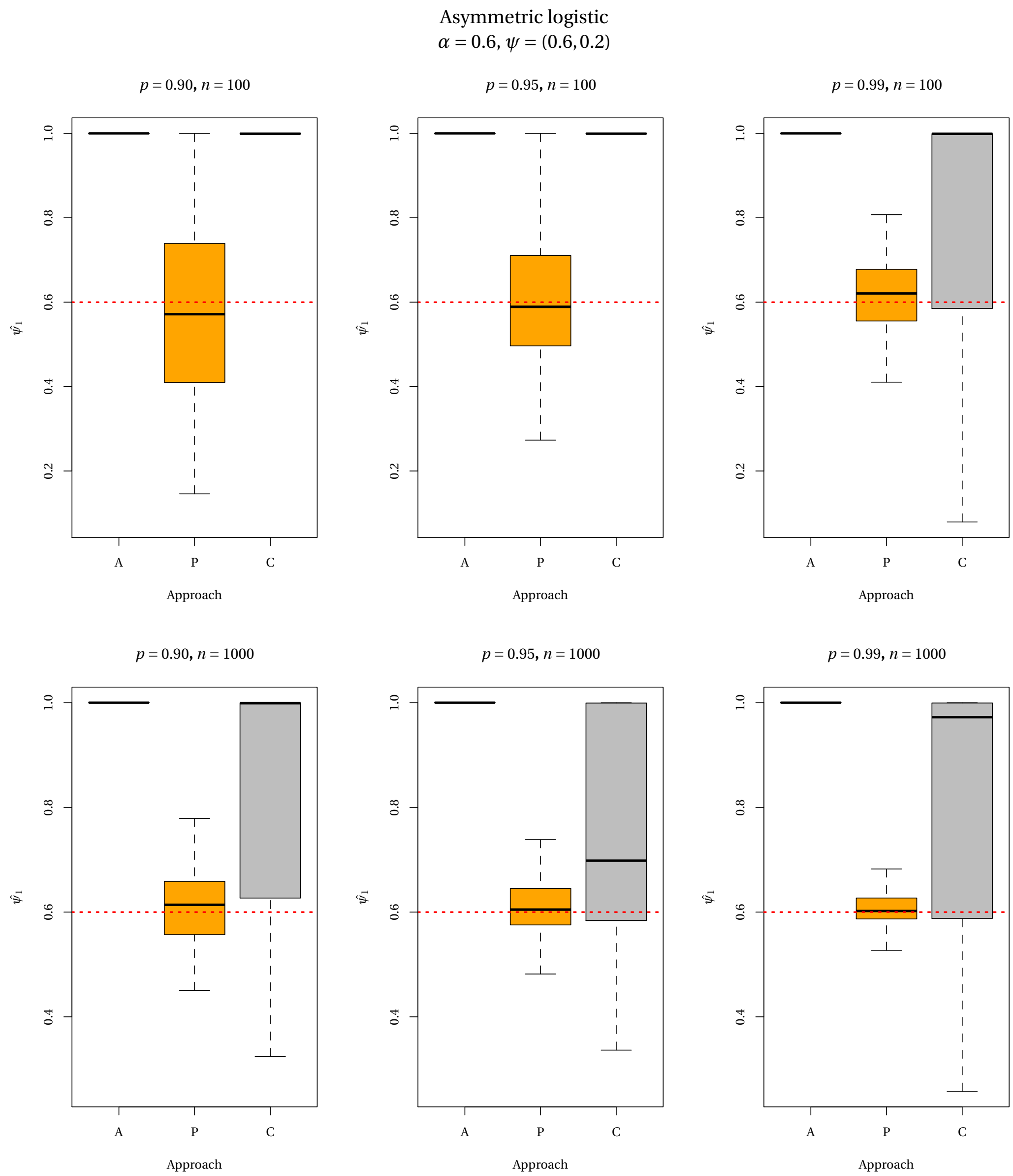

Figure B.32 – *Boxplots comparing first asymmetry parameter estimates of the asymmetric logistic model based on data simulated from the asymmetric logistic model with parameters $\alpha = 0.6$ and $\psi = (0.6, 0.2)$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n. The dashed red lines show the true parameter value.*

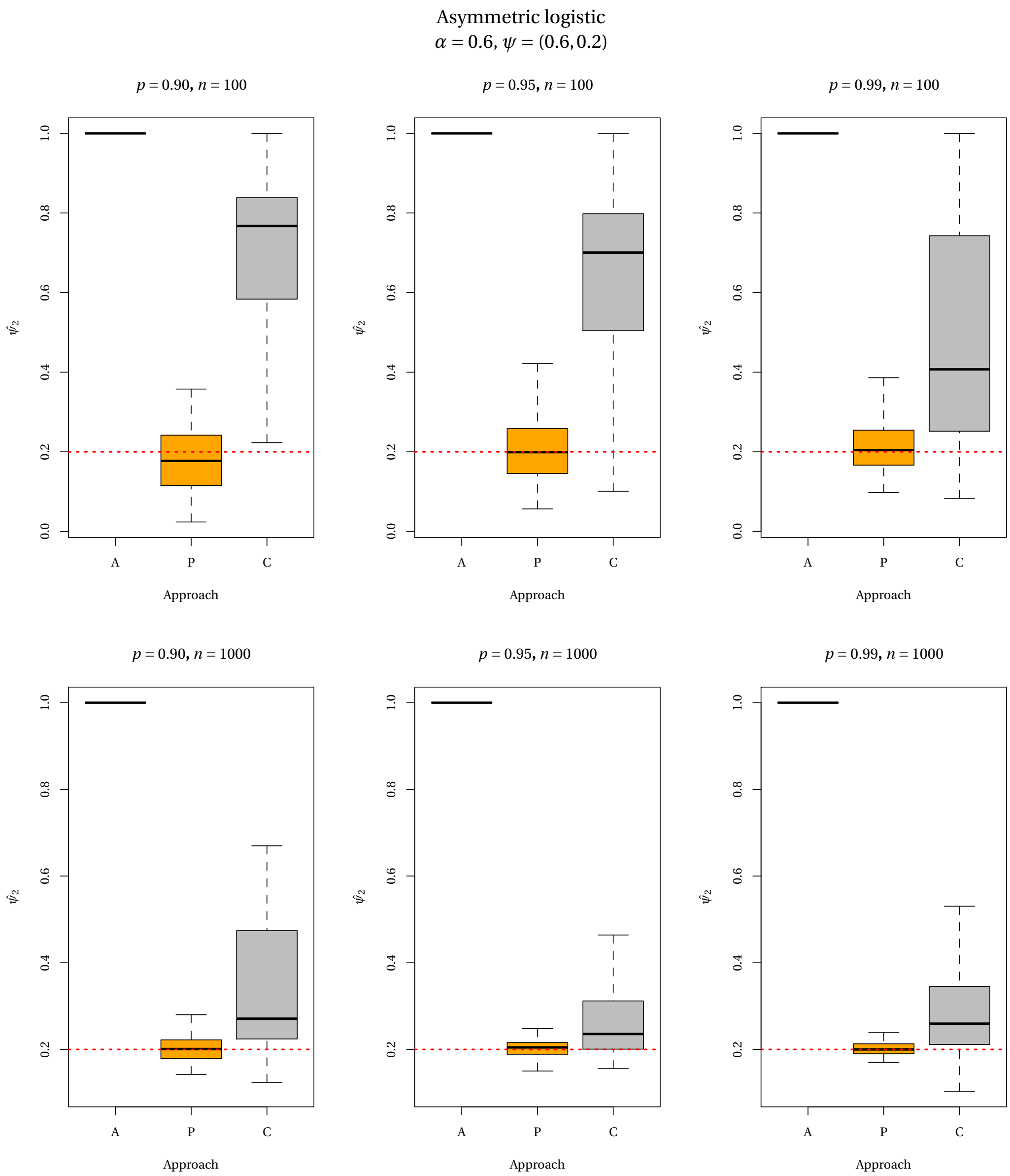

Figure B.33 – *Boxplots comparing second asymmetry parameter estimates of the asymmetric logistic model based on data simulated from the asymmetric logistic model with parameters $\alpha = 0.6$ and $\psi = (0.6, 0.2)$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n. The dashed red lines show the true parameter value.*

B.5 Joe model

B.5.1 Case $\theta = 1/0.3$

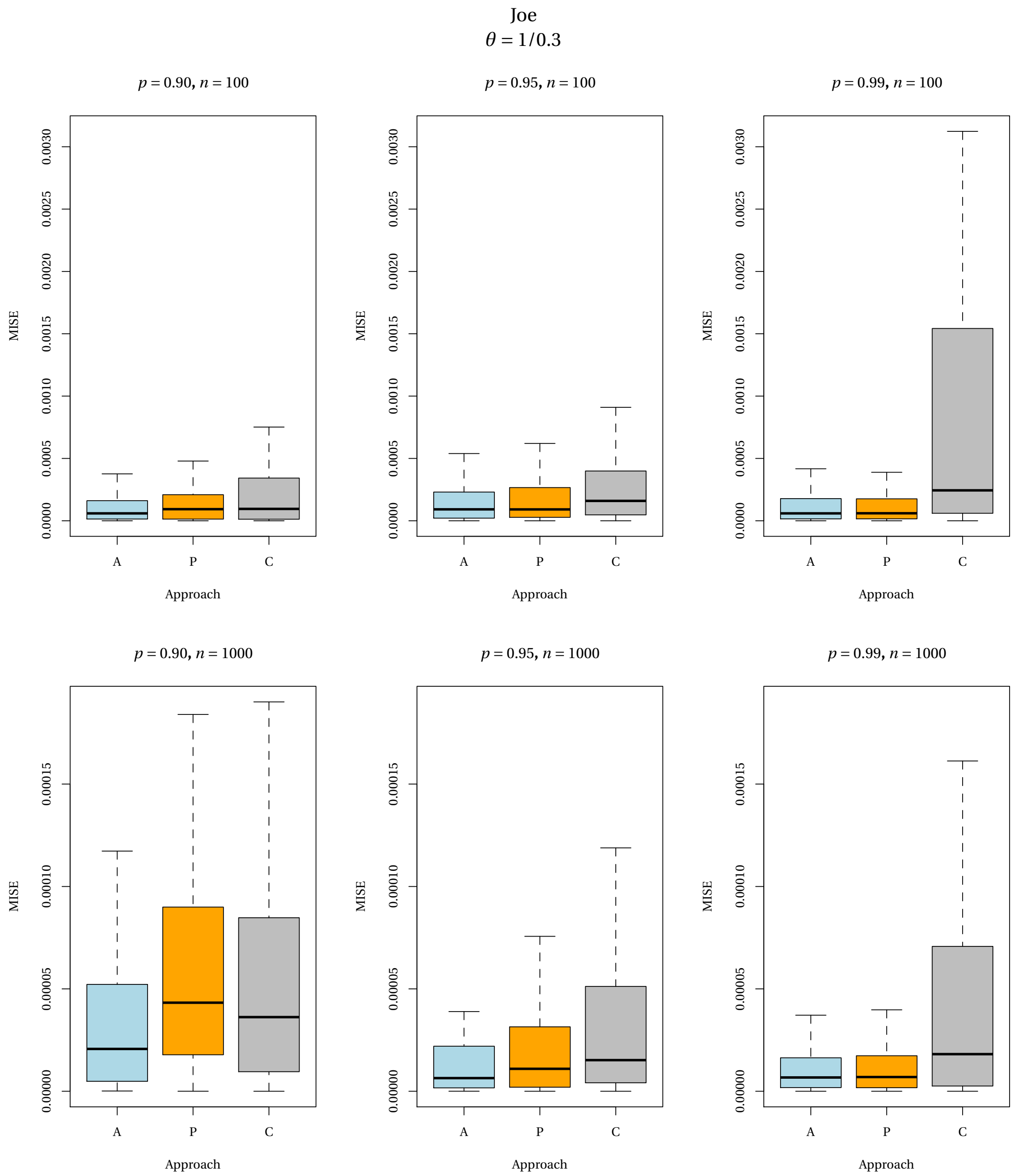

Figure B.34 – *Boxplots comparing the integrated squared error of maximum likelihood estimates of the logistic model based on data simulated from the Joe model with dependence parameter $\theta = 1/0.3$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n.*

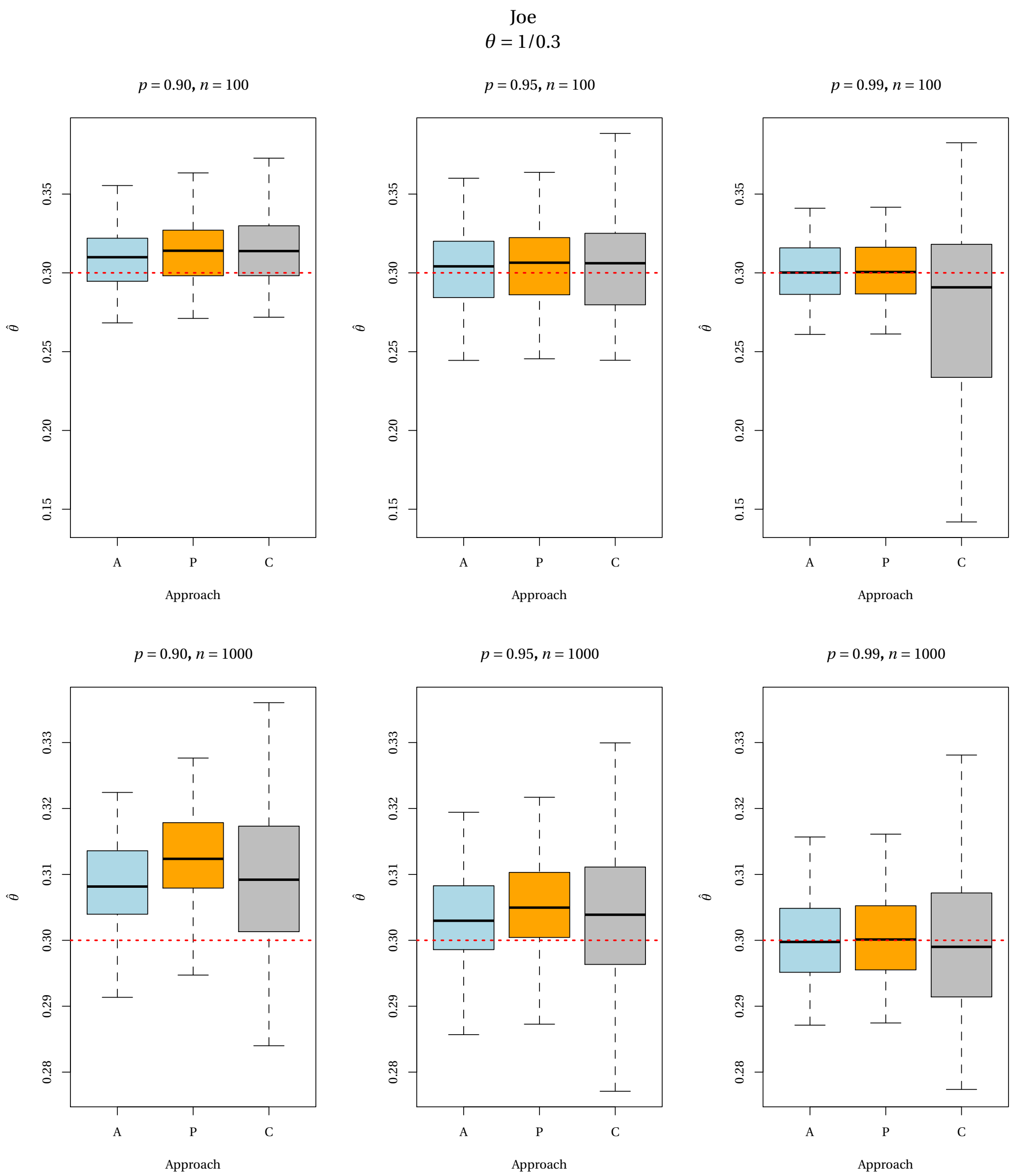

Figure B.35 – *Boxplots comparing dependence parameter estimates of the logistic model based on data simulated from the Joe model with dependence parameter* $\theta = 1/0.3$ *using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level* p *and number of excesses* n*. The dashed red lines show the true parameter value.*

B.5.2 Case $\theta = 1/0.55$

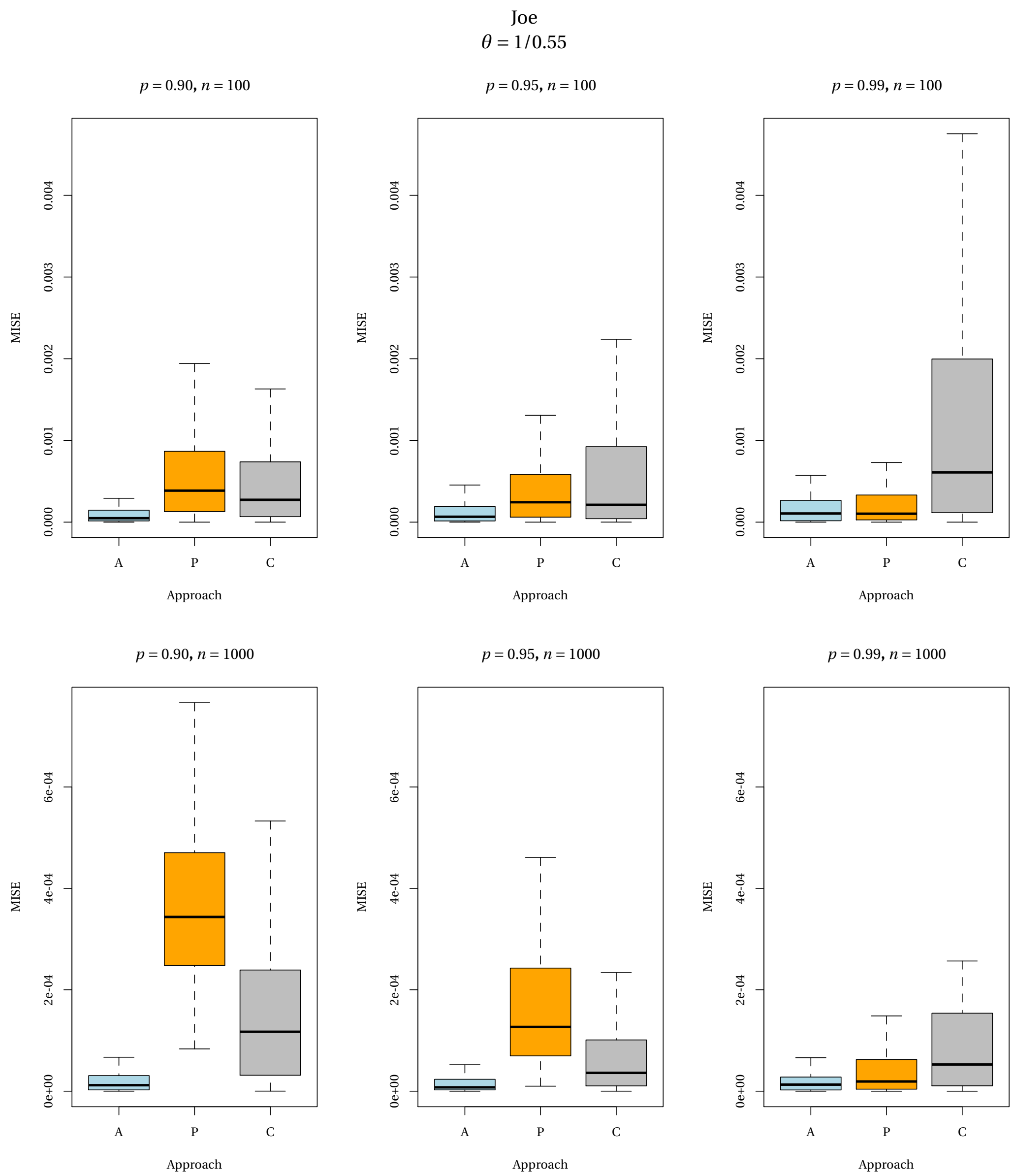

Figure B.36 – *Boxplots comparing the integrated squared error of maximum likelihood estimates of the logistic model based on data simulated from the Joe model with dependence parameter $\theta = 1/0.55$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n.*

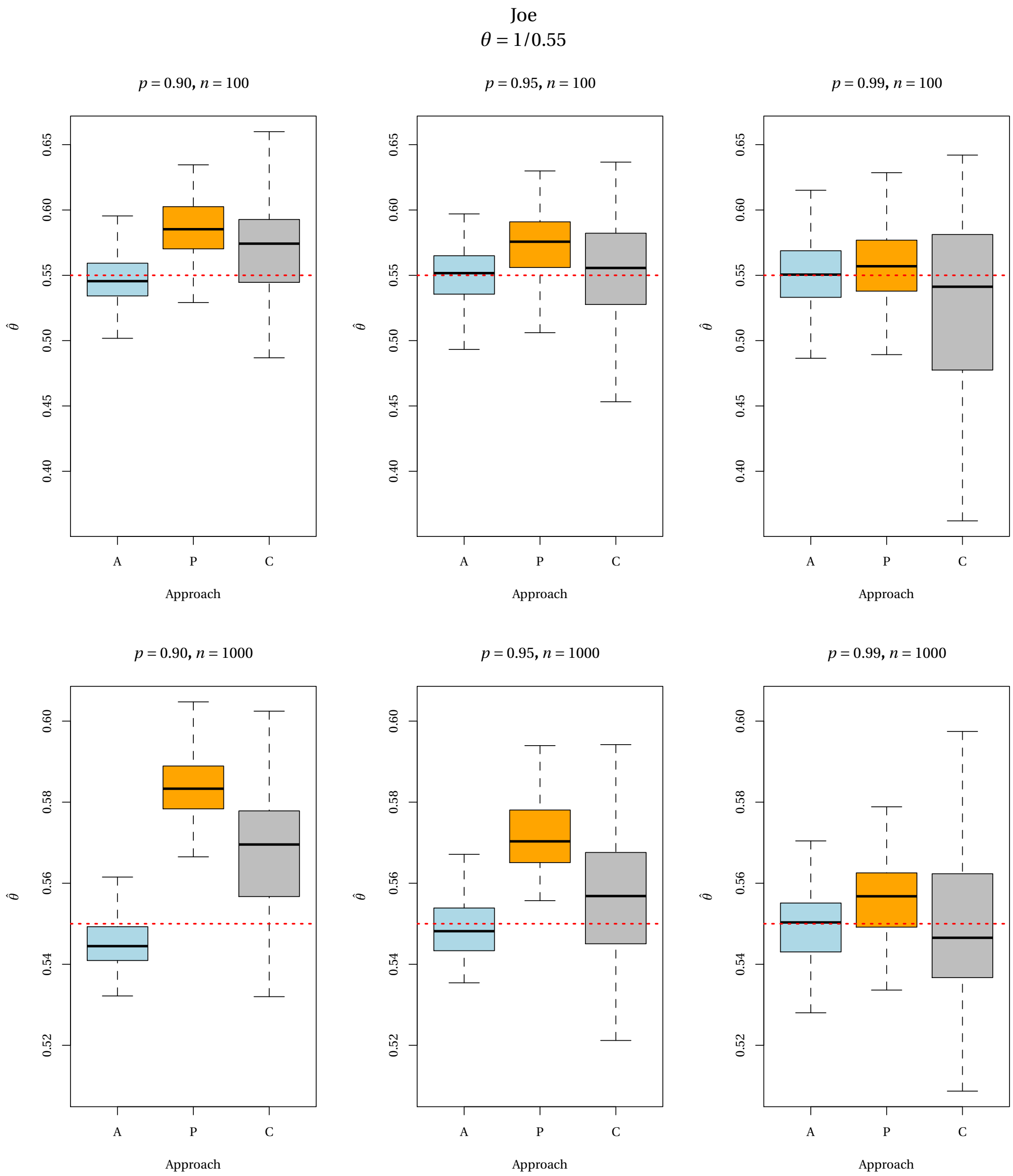

Figure B.37 – *Boxplots comparing dependence parameter estimates of the logistic model based on data simulated from the Joe model with dependence parameter $\theta = 1/0.55$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n. The dashed red lines show the true parameter value.*

B.5.3 Case $\theta = 1/0.8$

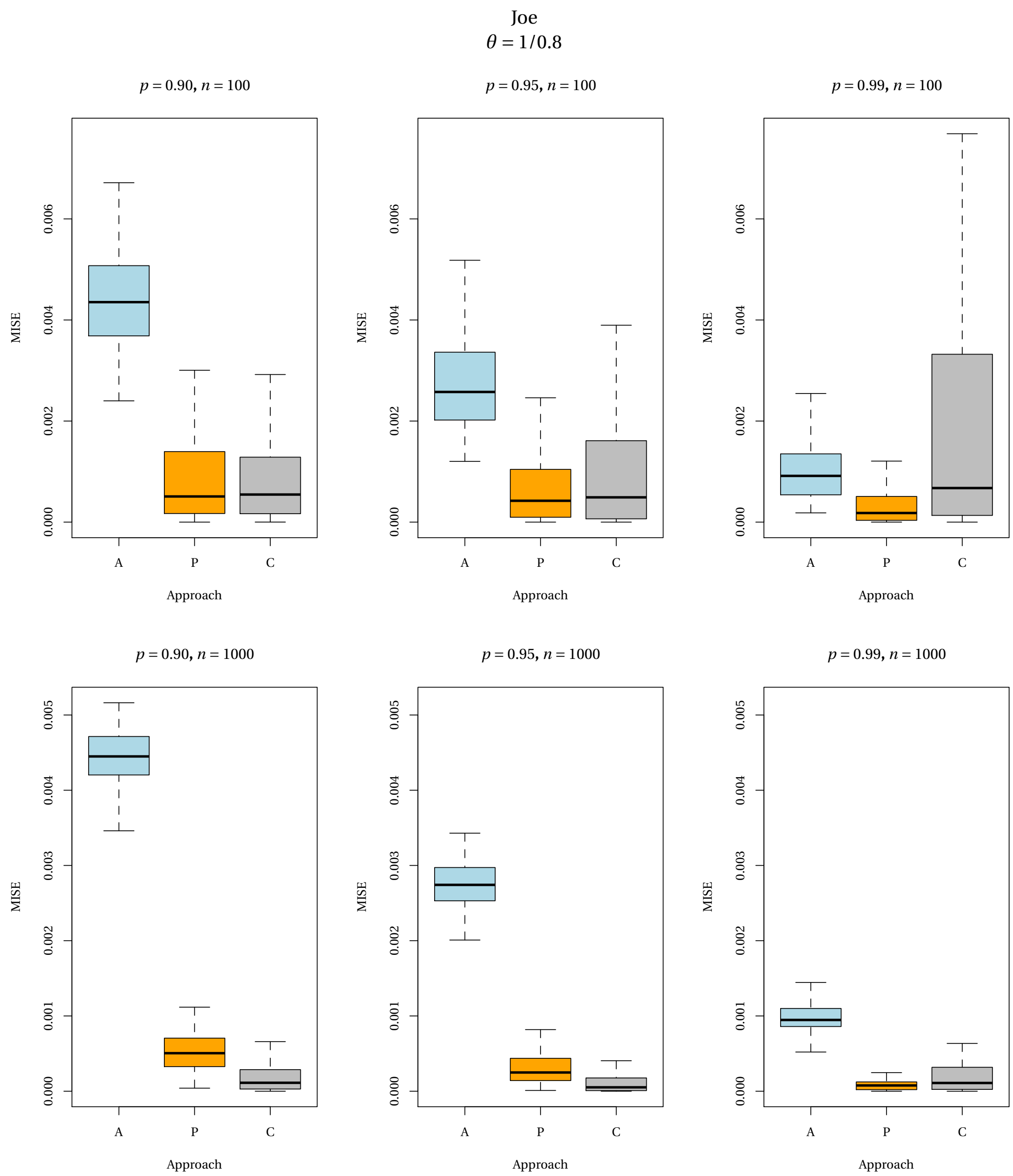

Figure B.38 – *Boxplots comparing the integrated squared error of maximum likelihood estimates of the logistic model based on data simulated from the Joe model with dependence parameter* $\theta = 1/0.8$ *using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses* n.

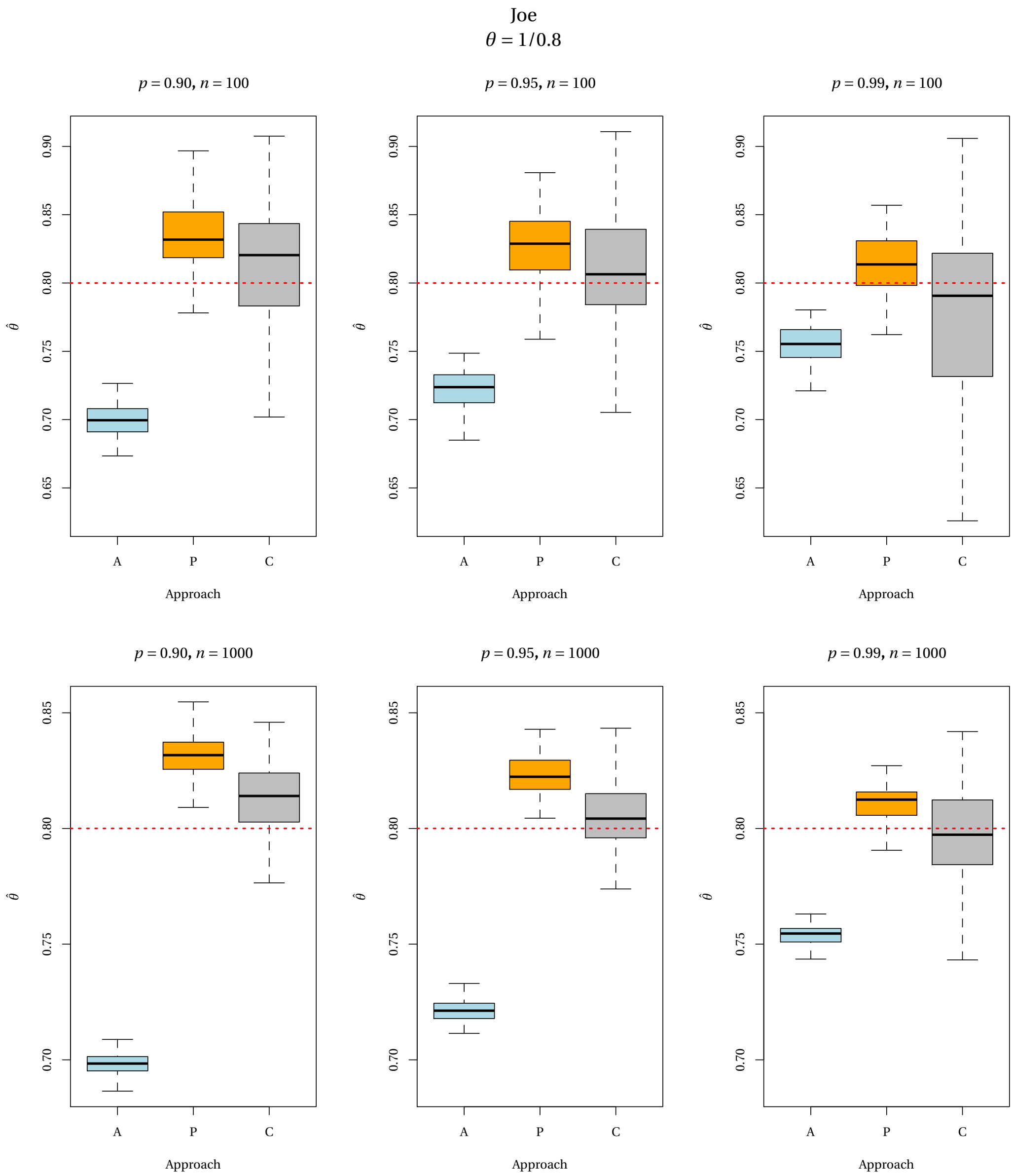

Figure B.39 – *Boxplots comparing dependence parameter estimates of the logistic model based on data simulated from the Joe model with dependence parameter* $\theta = 1/0.8$ *using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level* p *and number of excesses* n*. The dashed red lines show the true parameter value.*

B.6 Asymmetric Joe model

B.6.1 Case $\alpha = 1/0.2$ and $\psi = (0.8, 0.8)$

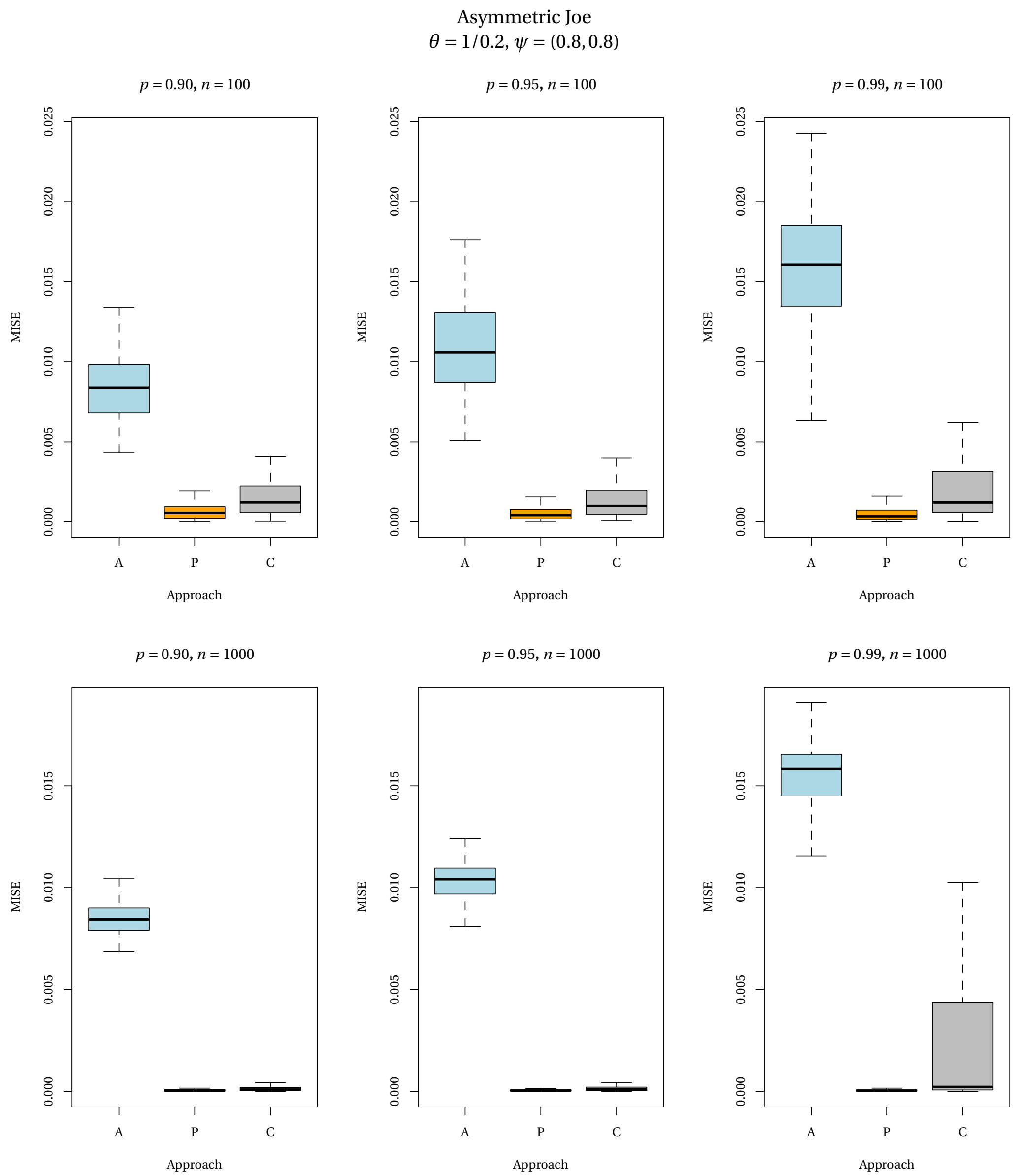

Figure B.40 – *Boxplots comparing the integrated squared error of maximum likelihood estimates of the asymmetric logistic model based on data simulated from the Asymmetric Joe model with parameters $\alpha = 1/0.2$ and $\psi = (0.8, 0.8)$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n.*

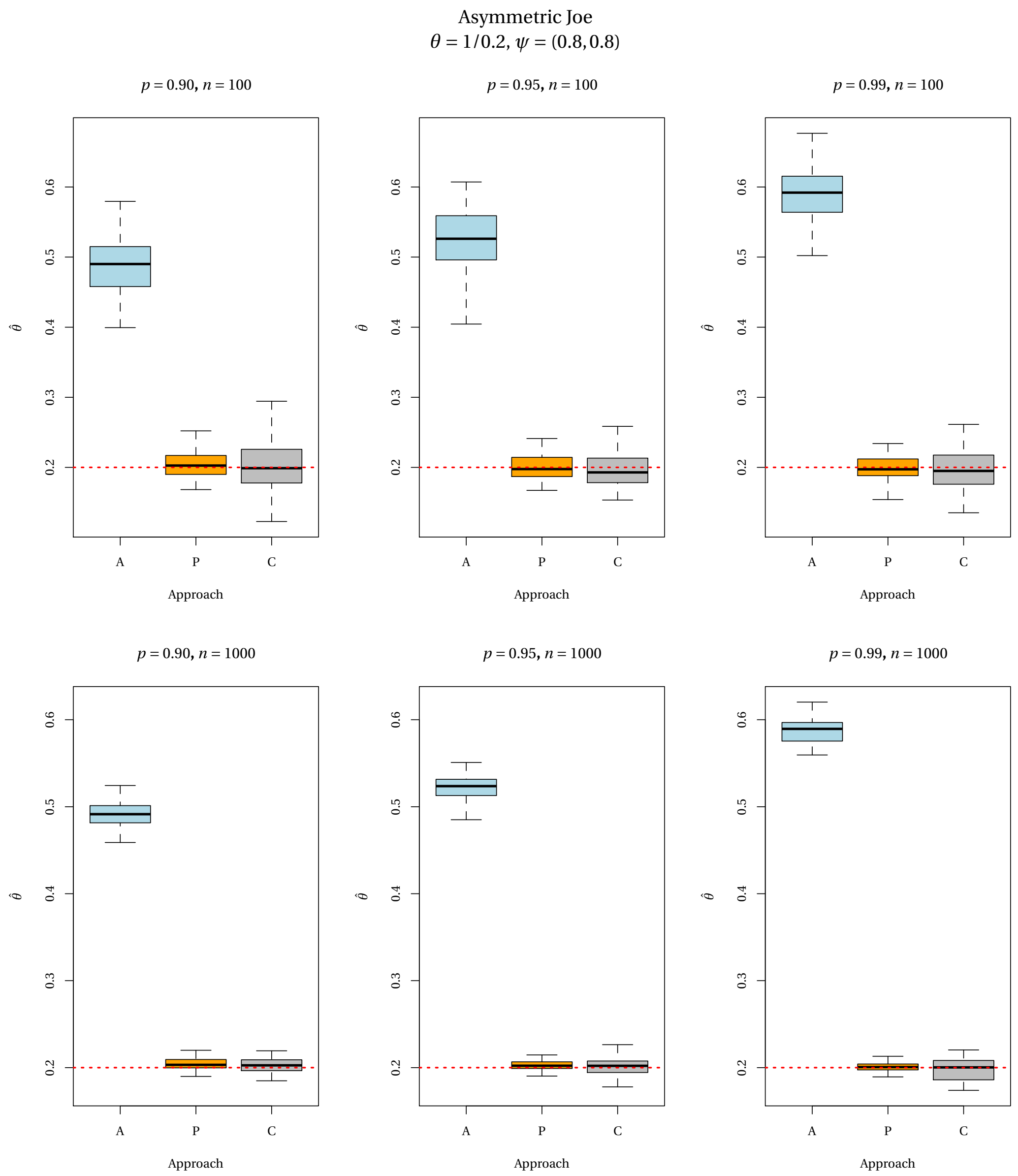

Figure B.41 – *Boxplots comparing dependence parameter estimates of the asymmetric logistic model based on data simulated from the Asymmetric Joe model with parameters $\alpha = 1/0.2$ and $\psi = (0.8, 0.8)$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n. The dashed red lines show the true parameter value.*

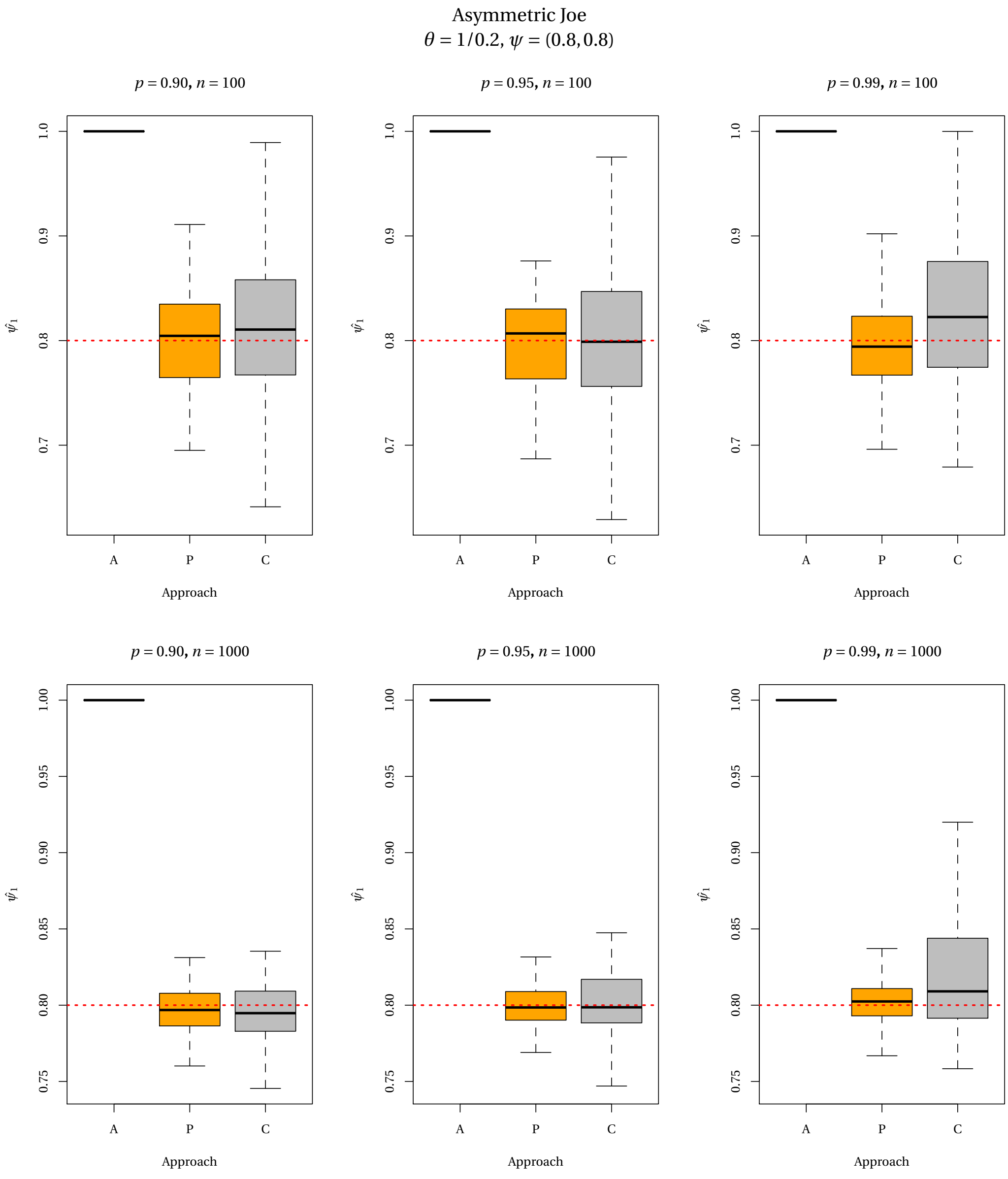

Figure B.42 – *Boxplots comparing first asymmetry parameter estimates of the asymmetric logistic model based on data simulated from the Asymmetric Joe model with parameters $\alpha = 1/0.2$ and $\psi = (0.8, 0.8)$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n. The dashed red lines show the true parameter value.*

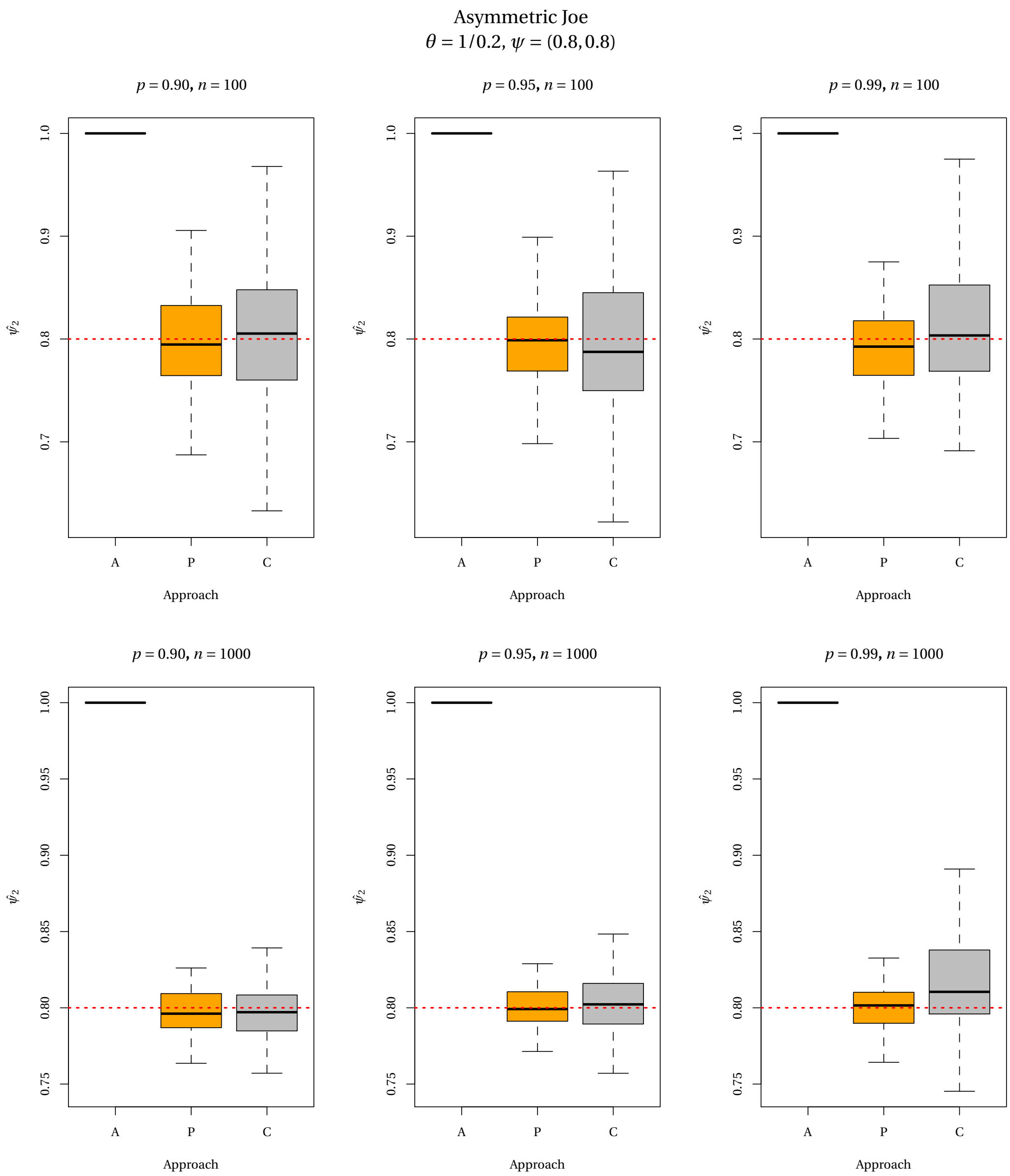

Figure B.43 – *Boxplots comparing second asymmetry parameter estimates of the asymmetric logistic model based on data simulated from the Asymmetric Joe model with parameters $\alpha = 1/0.2$ and $\psi = (0.8, 0.8)$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n. The dashed red lines show the true parameter value.*

B.6.2 Case $\alpha = 1/0.4$ and $\psi = (0.7, 0.5)$

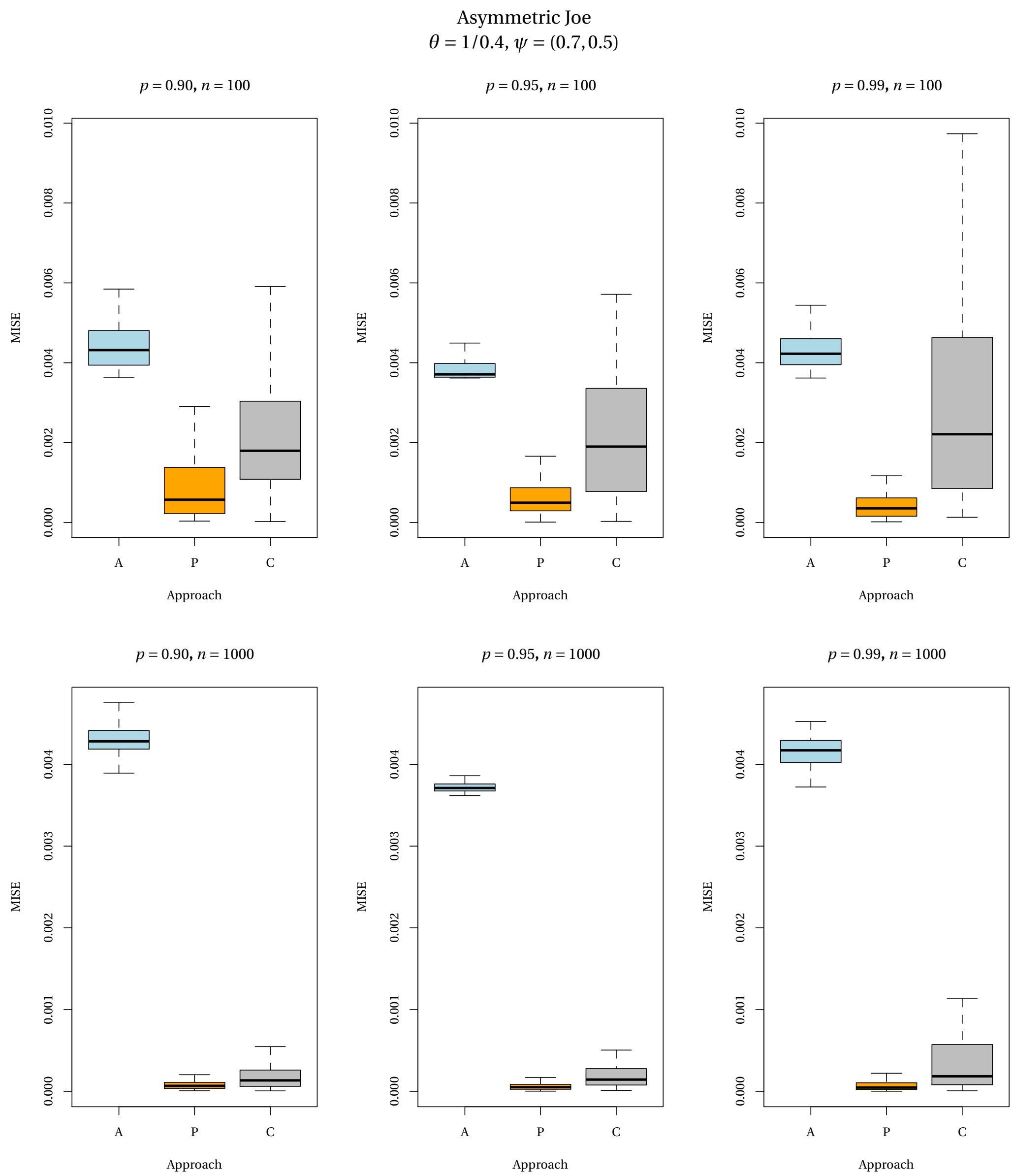

Figure B.44 – *Boxplots comparing the integrated squared error of maximum likelihood estimates of the asymmetric logistic model based on data simulated from the Asymmetric Joe model with parameters $\alpha = 1/0.4$ and $\psi = (0.7, 0.5)$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n.*

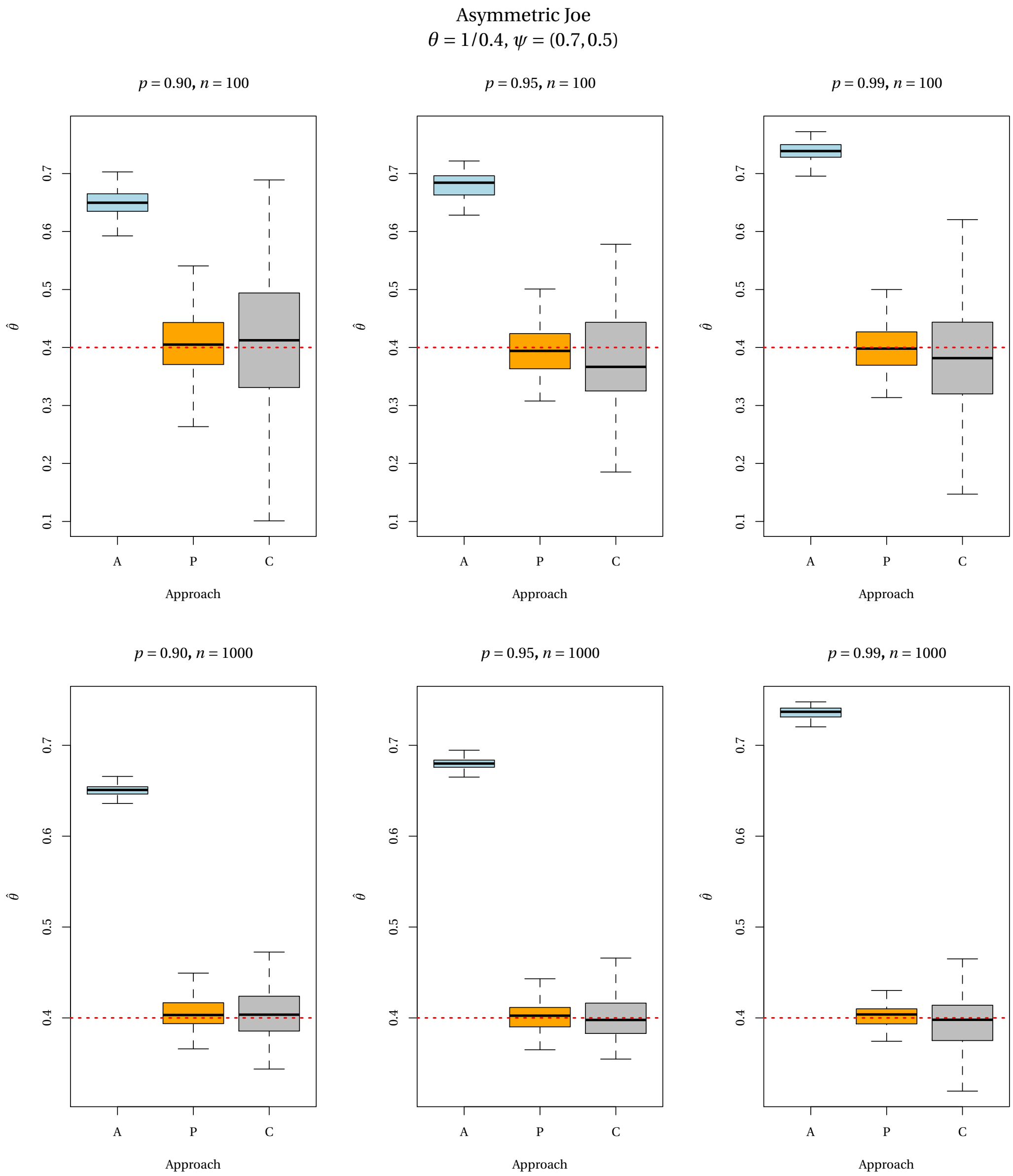

Figure B.45 – *Boxplots comparing dependence parameter estimates of the asymmetric logistic model based on data simulated from the Asymmetric Joe model with parameters $\alpha = 1/0.4$ and $\psi = (0.7, 0.5)$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n. The dashed red lines show the true parameter value.*

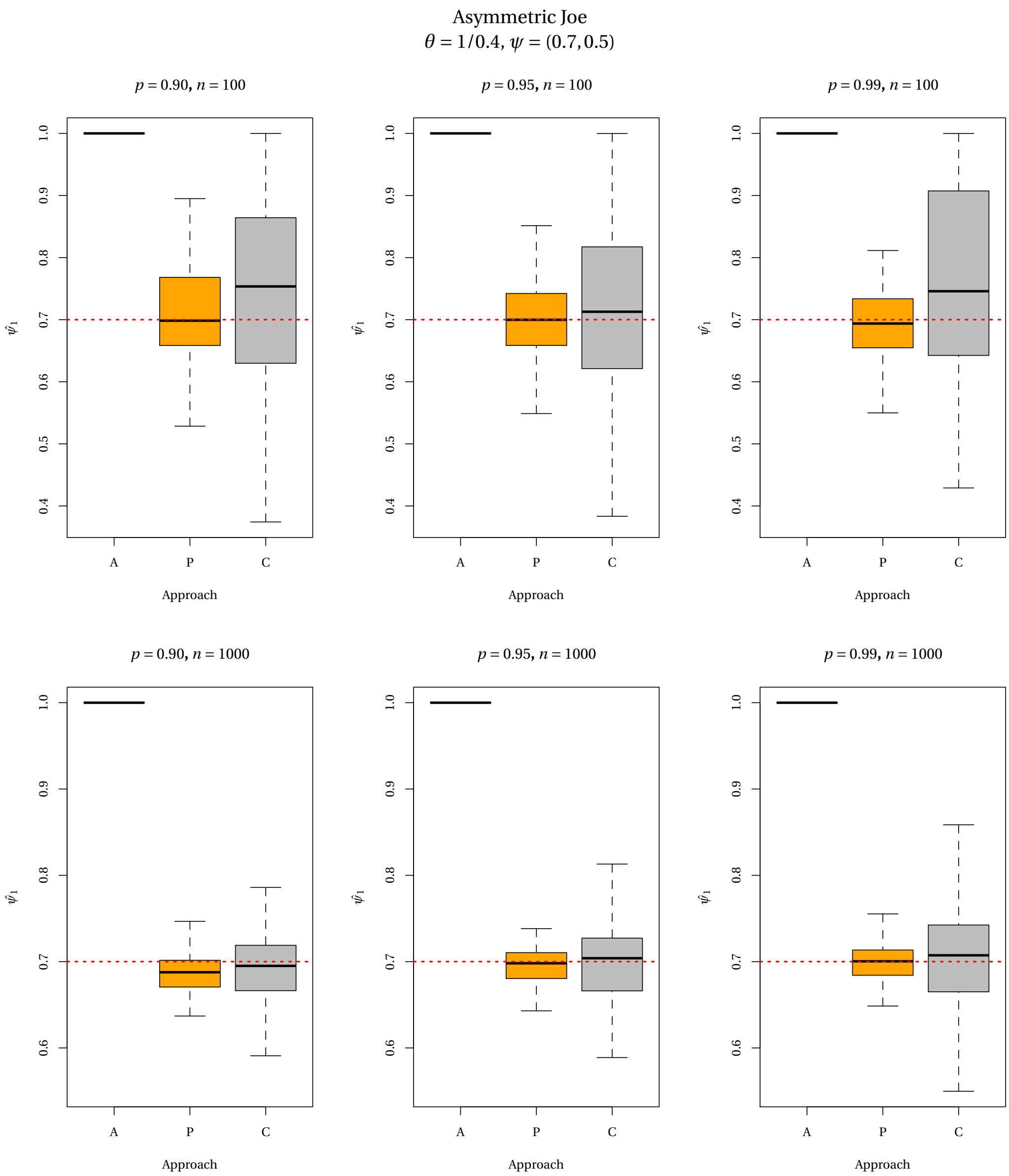

Figure B.46 – *Boxplots comparing first asymmetry parameter estimates of the asymmetric logistic model based on data simulated from the Asymmetric Joe model with parameters $\alpha = 1/0.4$ and $\psi = (0.7, 0.5)$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n. The dashed red lines show the true parameter value.*

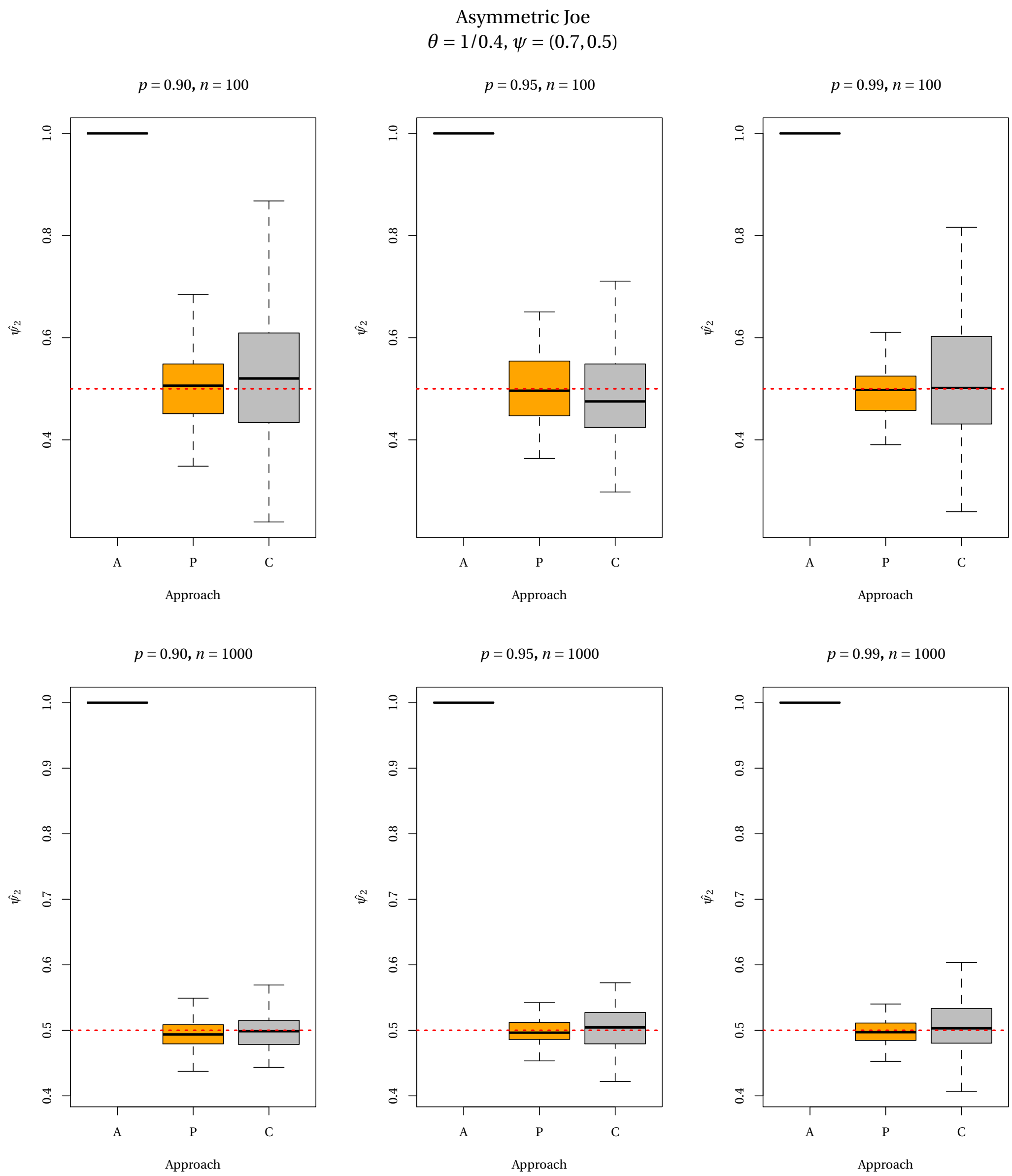

Figure B.47 – *Boxplots comparing second asymmetry parameter estimates of the asymmetric logistic model based on data simulated from the Asymmetric Joe model with parameters $\alpha = 1/0.4$ and $\psi = (0.7, 0.5)$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n. The dashed red lines show the true parameter value.*

B.6.3 Case $\alpha = 1/0.6$ and $\psi = (0.6, 0.2)$

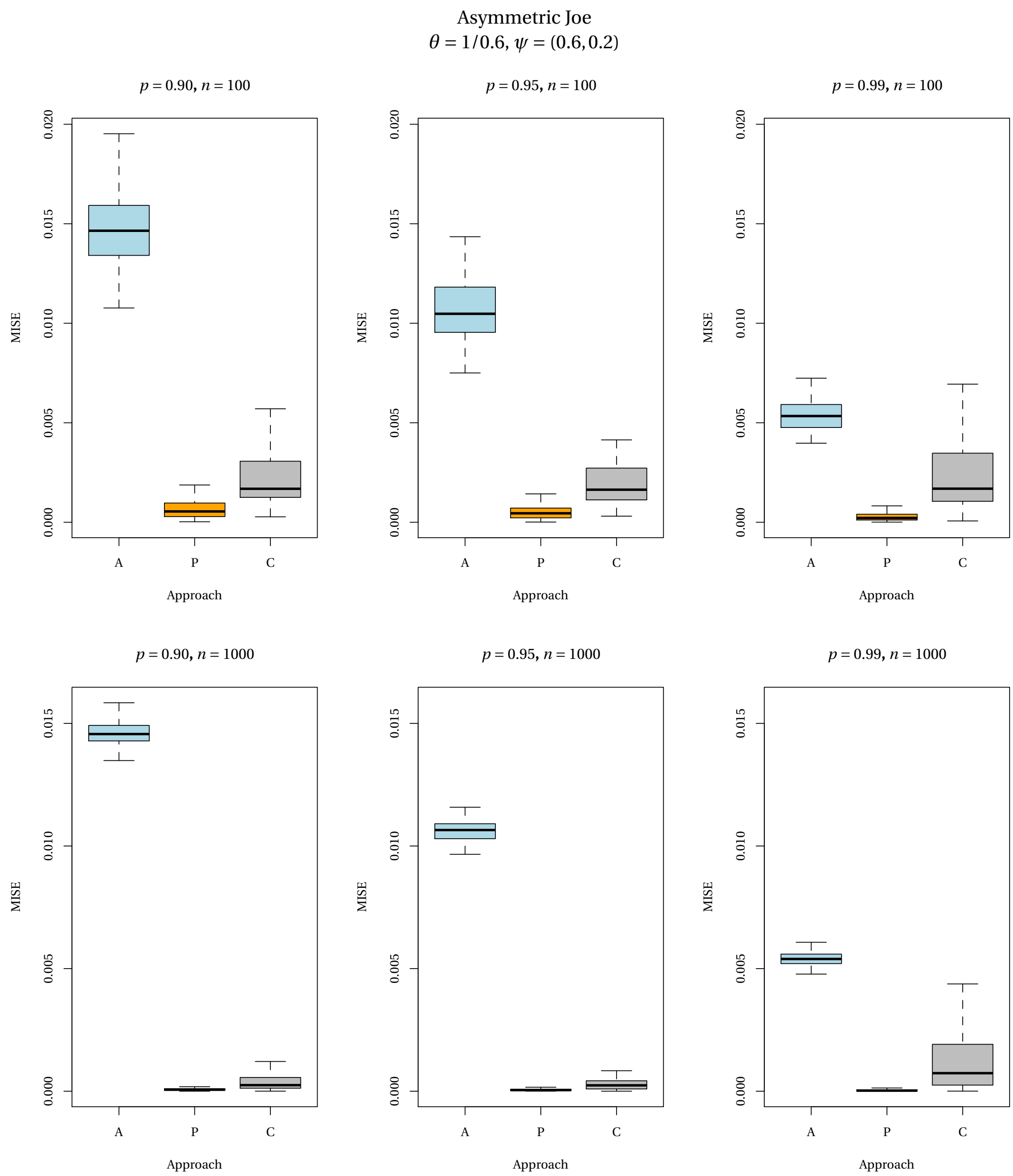

Figure B.48 – *Boxplots comparing the integrated squared error of maximum likelihood estimates of the asymmetric logistic model based on data simulated from the Asymmetric Joe model with parameters $\alpha = 1/0.6$ and $\psi = (0.6, 0.2)$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n.*

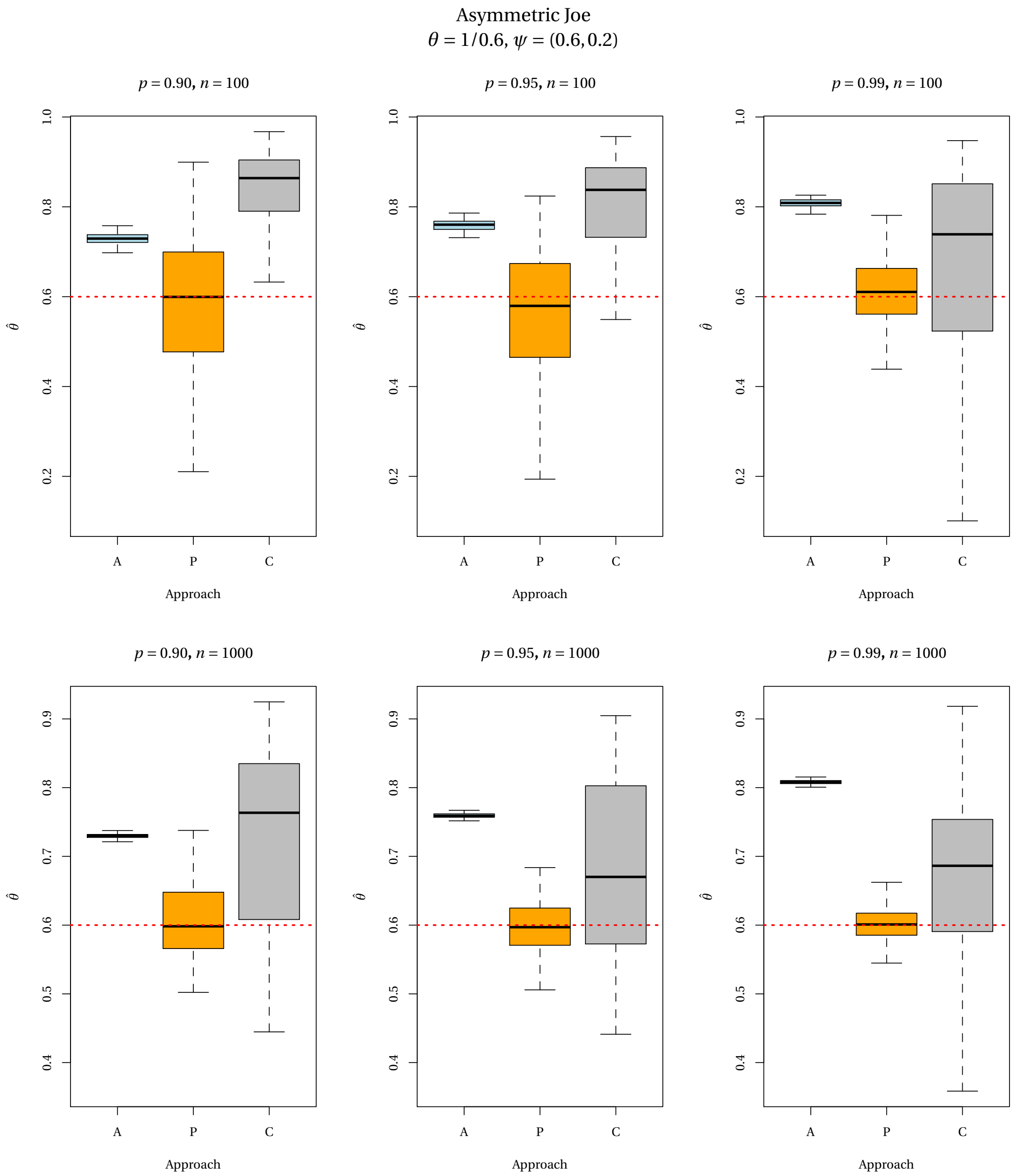

Figure B.49 – *Boxplots comparing dependence parameter estimates of the asymmetric logistic model based on data simulated from the Asymmetric Joe model with parameters $\alpha = 1/0.6$ and $\psi = (0.6, 0.2)$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n. The dashed red lines show the true parameter value.*

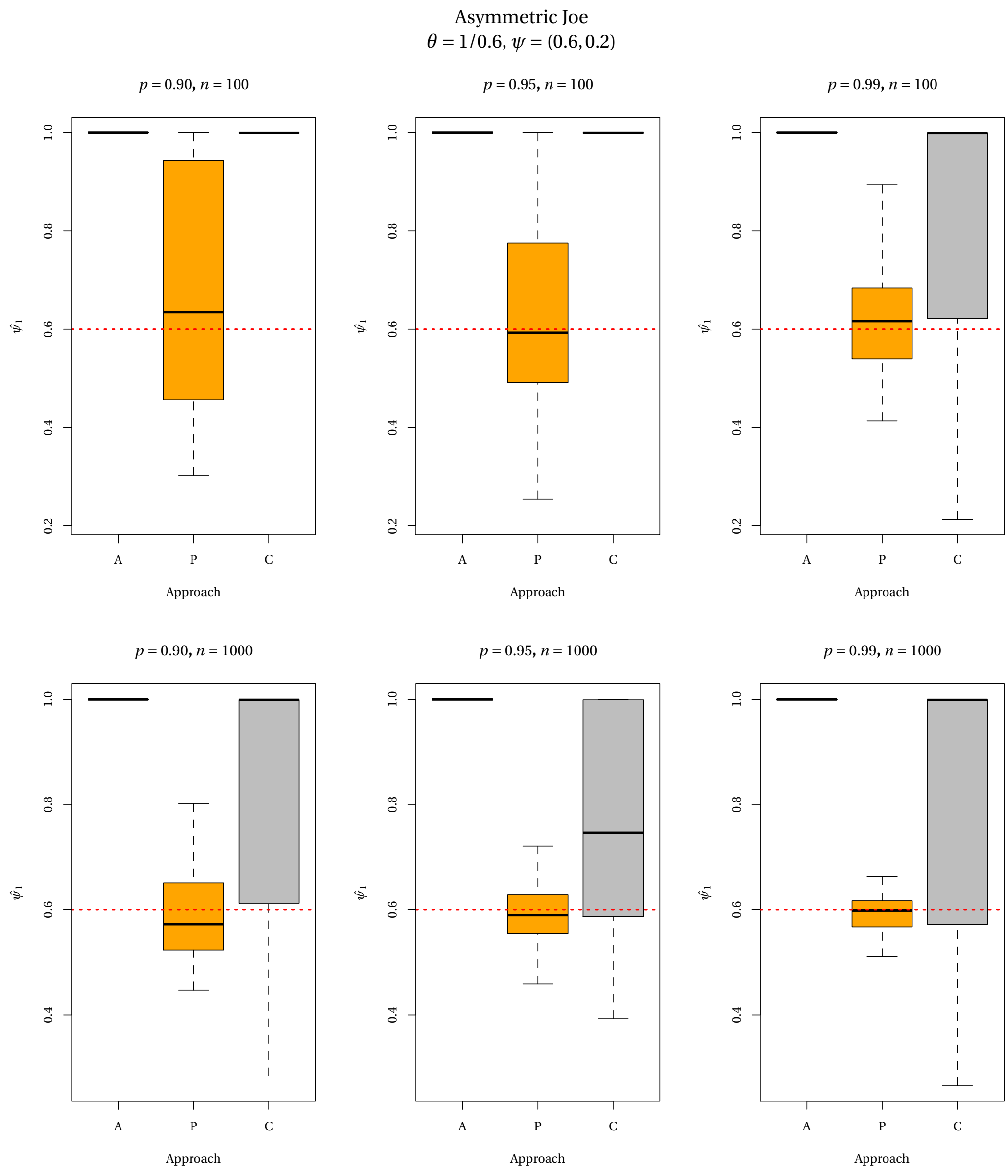

Figure B.50 – *Boxplots comparing first asymmetry parameter estimates of the asymmetric logistic model based on data simulated from the Asymmetric Joe model with parameters $\alpha = 1/0.6$ and $\psi = (0.6, 0.2)$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n. The dashed red lines show the true parameter value.*

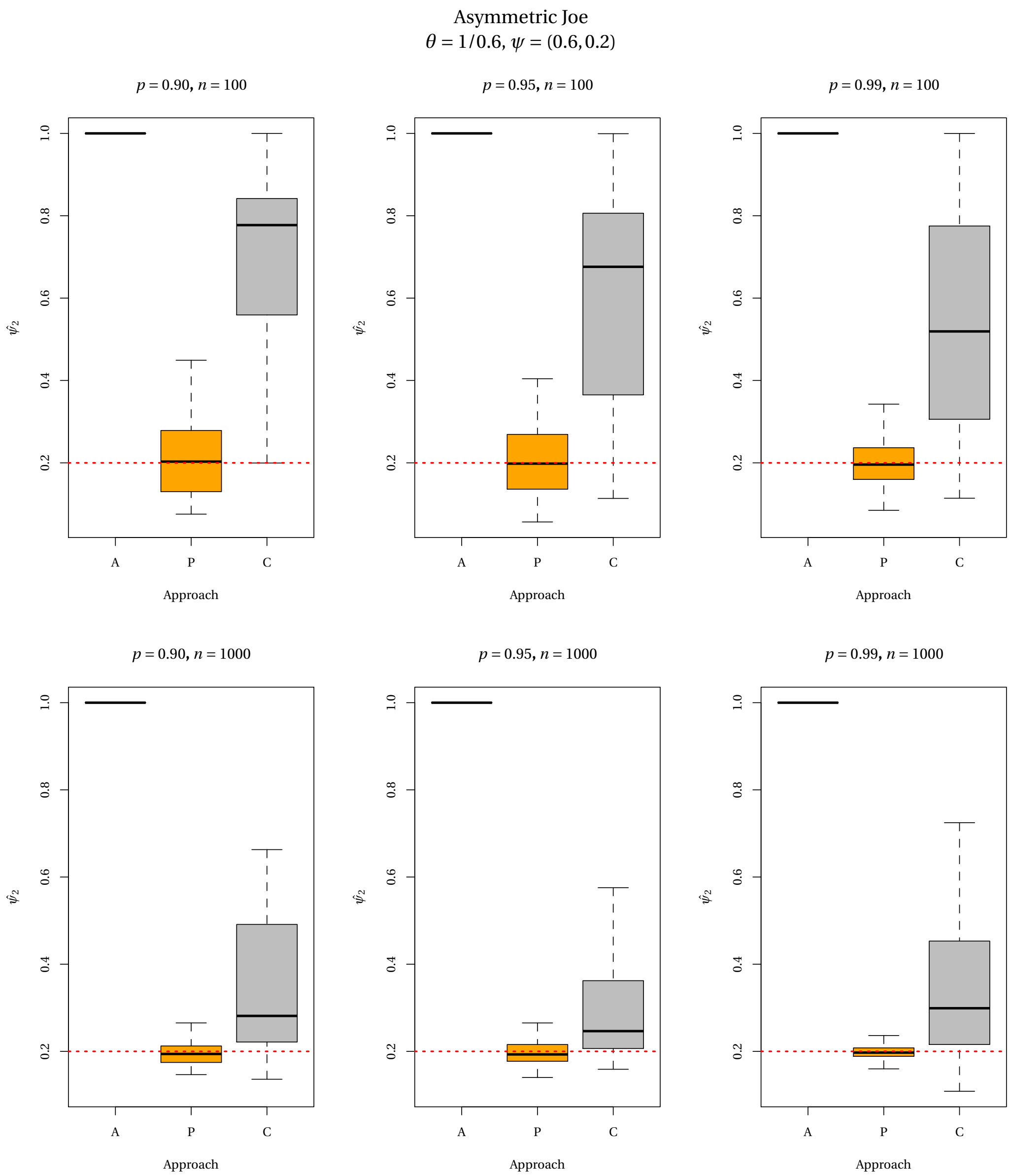

Figure B.51 – *Boxplots comparing second asymmetry parameter estimates of the asymmetric logistic model based on data simulated from the Asymmetric Joe model with parameters $\alpha = 1/0.6$ and $\psi = (0.6, 0.2)$ using the 'asymptotic' (A), 'penultimate' (P), and 'censored' (C) approaches for several choices of threshold level p and number of excesses n. The dashed red lines show the true parameter value.*

Bibliography

The number(s) in square brackets at the end of each item refer to the page(s) on which the respective work is cited.

Bacro, J.-N., Bel, L. and Lantuéjoul, C. (2010) Testing the independence of maxima: From bivariate vectors to spatial extreme fields. *Extremes* **13**(2), 155–175. [28]

Balkema, A. A. and de Haan, L. (1990) A convergence rate in extreme-value theory. *Journal of Applied Probability* **27**(03), 577–585. [5]

Balkema, G. and de Haan, L. (1974) Residual life time at great age. *Annals of Probability* **2**(5), 792–804. [9]

Ballani, F. and Schlather, M. (2011) A construction principle for multivariate extreme value distributions. *Biometrika* **98**(3), 633 –645. [19]

Beirlant, J., Goegebeur, Y., Teugels, J. and Segers, J. (2004) *Statistics of Extremes: Theory and Applications.* New York: Wiley. [1, 5, 7, 12, 14, 19, 31]

Berman, S. M. (1961) Convergence to bivariate limiting extreme value distributions. *Annals of the Institute of Statistical Mathematics* **13**(1), 217–223. [27]

Bortot, P. and Coles, S. (2003) Extremes of Markov chains with tail switching potential. *Journal of the Royal Statistical Society: Series B* **65**(4), 851–867. [14]

Caeiro, F. and Gomes, M. I. (2015) Threshold selection in extreme value analysis. In *Extreme Value Modeling and Risk Analysis: Methods and Applications,* eds D. Dey and J. Yan. Boca Raton, FL: CRC Press. [10]

Capéraà, P., Fougères, A.-L. and Genest, C. (1997) A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika* **84**(3), 567–577. [15, 26]

Capéraà, P., Fougères, A.-L. and Genest, C. (2000) Bivariate distributions with given extreme value attractor. *Journal of Multivariate Analysis* **72**(1), 30–49. [15]

de Carvalho, M. and Ramos, A. (2012) Bivariate extreme statistics, II. *Revstat* **10**(1), 83–107. [28, 30]

Castillo, E., Hadi, A., Balakrishnan, N. and Sarabia, J. (2005) *Extreme Value and Related Models with Applications in Engineering and Science.* Hoboken N.J.: Wiley. [1]

Castillo, E. and Hadi, A. S. (1997) Fitting the generalized Pareto distribution to data. *Journal of the American Statistical Association* **92**(440), 1609–1620. [10]

Chavez-Demoulin, V. and Davison, A. C. (2005) Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society: Series C* **54**(1), 207–222. [12]

Chavez-Demoulin, V. and Davison, A. C. (2012) Modelling time series extremes. *Revstat* **10**(1), 109–133. [12]

Coles, S. (2001) *An Introduction to Statistical Modeling of Extreme Values.* London: Springer. [1, 6, 7, 10, 11, 12]

Coles, S., Heffernan, J. and Tawn, J. (1999) Dependence measures for extreme value analyses. *Extremes* **2**(4), 339–365. [15, 28, 30, 31, 32, 67, 68]

Coles, S. and Tawn, J. (1991) Modelling extreme multivariate events. *Journal of the Royal Statistical Society: Series B* **53**(2), 377–392. [18, 21, 35, 67]

Coles, S. G. and Powell, E. A. (1996) Bayesian methods in extreme value modelling: A review and new developments. *International Statistical Review / Revue Internationale de Statistique* **64**(1), 119–136. [7]

Cooley, D., Cisewski, J., Erhardt, R. J., Jeon, S., Mannshardt, E., Omolo, B. O. and Sun, Y. (2012) A survey of spatial extremes: Measuring spatial dependence and modeling spatial effects. *Revstat* **10**(1), 135–165. [1]

Cooley, D., Davis, R. A. and Naveau, P. (2010) The pairwise beta distribution: A flexible parametric multivariate model for extremes. *Journal of Multivariate Analysis* **101**(9), 2103–2117. [19, 35]

Davison, A. C., Padoan, S. A. and Ribatet, M. (2012) Statistical modeling of spatial extremes (with discussion). *Statistical Science* **27**(2), 161–201. [1, 16]

Davison, A. C. and Smith, R. L. (1990) Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society: Series B* **52**(3), 393–442. [7, 10]

Deheuvels, P. (1991) On the limiting behavior of the Pickands estimator for bivariate extreme-value distributions. *Statistics & Probability Letters* **12**(5), 429–439. [26]

Dekkers, A. L. M., Einmahl, J. H. J. and de Haan, L. (1989) A moment estimator for the index of an extreme-value distribution. *Annals of Statistics* **17**(4), 1833–1855. [7]

Demarta, S. and McNeil, A. J. (2005) The t copula and related copulas. *International Statistical Review* **73**(1), 111–129. [15, 35]

Dey, D. K. and Yan, J. (eds) (2016) *Extreme Value Modeling and Risk Analysis: Methods and Applications.* Boca Raton, FL: CRC Press. [1]

Diebolt, J., Guillou, A., Naveau, P. and Ribereau, P. (2008) Improving probability-weighted moment methods for the generalized extreme value distribution. *Revstat* **6**(1), 33–50. [7]

Dupuis, D. J. and Tawn, J. A. (2001) Effects of mis-specification in bivariate extreme value problems. *Extremes* **4**(4), 315–330. [25]

Eastoe, E. F. and Tawn, J. A. (2009) Modelling non-stationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society: Series C* **58**(1), 25–45. [12]

Einmahl, J. and Segers, J. (2009) Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. *The Annals of Statistics* **37**(5B), 2953–2989. [81, 87, 88]

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events: For Insurance and Finance.* Berlin: Springer. [1, 3, 5, 6, 10, 11]

Falk, M. and Michel, R. (2006) Testing for tail independence in extreme value models. *Annals of the Institute of Statistical Mathematics* **58**(2), 261–290. [28]

Fawcett, L. and Walshaw, D. (2007) Improved estimation for temporally clustered extremes. *Environmetrics* **18**(2), 173–188. [14]

Fawcett, L. and Walshaw, D. (2012) Estimating return levels from serially dependent extremes. *Environmetrics* **23**(3), 272–283. [14]

Finkenstädt, B. and Rootzén, H. (2004) *Extreme Values in Finance, Telecommunications, and the Environment.* Chapman and Hall/CRC. [1]

Fisher, R. A. and Tippett, L. H. C. (1928) Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society* **24**, 180–190. [3]

Fougères, A.-L. (2003) Multivariate extremes. In *Extreme Values in Finance, Telecommunications, and the Environment.* New York: Chapman and Hall/CRC. [26]

Galambos, J. (1987) *The Asymptotic Theory of Extreme Order Statistics.* Second edition. Malabar, Fl: Krieger. [1]

Genest, C. and Rivest, L.-P. (1989) A characterization of Gumbel's family of extreme value distributions. *Statistics & Probability Letters* **8**(3), 207–211. [17, 61]

Gilleland, E. (2016) Computing Software. In *Extreme Value Modeling and Risk Analysis: Methods and Applications*, eds D. Dey and J. Yan. Boca Raton, FL: CRC Press. [1]

Gilleland, E., Ribatet, M. and Stephenson, A. G. (2012) A software review for extreme value analysis. *Extremes* **16**(1), 103–119. [1]

Gnedenko, B. (1943) Sur la distribution limite du terme maximum d'une série aléatoire. *The Annals of Mathematics* **44**(3), 423–453. [3]

Gudendorf, G. and Segers, J. (2010) Extreme-value copulas. In *Copula Theory and Its Applications*, eds P. Jaworski, F. Durante, W. K. Härdle and T. Rychlik, pp. 127–145. Berlin, Heidelberg: Springer. [16]

Gumbel, E. J. (1958) *Statistics of Extremes.* New York: Columbia University Press. [1]

Gumbel, E. J. (1960a) Distributions des valeurs extrêmes en plusieurs dimensions. *Publications de l'Institut de Statistique de l'Université de Paris* **9**, 171—173. [20]

Gumbel, E. J. (1960b) Bivariate exponential distributions. *Journal of the American Statistical Association* **55**, 697–707. [19]

Gumbel, E. J. (1961) Bivariate Logistic Distributions. *Journal of the American Statistical Association* **56**(294), 335–349. [35]

de Haan, L. and Ferreira, A. (2006) *Extreme Value Theory: An Introduction.* New York: Springer. [1, 5]

de Haan, L. and Resnick, S. I. (1977) Limit theory for multivariate sample extremes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **40**(4), 317–337. [26]

Hall, P. and Tajvidi, N. (2000) Distribution and dependence-function estimation for bivariate extreme-value distributions. *Bernoulli* **6**(5), 835–844. [26]

Heffernan, J. E. (2000) A directory of coefficients of tail dependence. *Extremes* **3**(3), 279–290. [15, 29, 31]

Heffernan, J. E. and Tawn, J. A. (2004) A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B* **66**(3), 497–546. [28]

Hill, B. M. (1975) A simple general approach to inference about the tail of a distribution. *The Annals of Statistics* **3**(5), 1163–1174. [7]

Hofert, M., Kojadinovic, I., Maechler, M. and Yan, J. (2017) *copula: Multivariate Dependence with Copulas.* R package version 0.999-18. [54]

Hosking, J. R. M. (1990) L-Moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society: Series B* **52**(1), 105–124. [7]

Hosking, J. R. M. and Wallis, J. R. (1987) Parameter and quantile estimation for the generalized pareto distribution. *Technometrics* **29**(3), 339–349. [10]

Hosking, J. R. M., Wallis, J. R. and Wood, E. F. (1985) Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics* **27**(3), 251. [6]

Hüsler, J. and Reiss, R.-D. (1989) Maxima of normal random vectors: Between independence and complete dependence. *Statistics & Probability Letters* **7**, 283–286. [20, 35]

Jenkinson, A. (1955) The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society* **81**(348), 158–171. [3]

Joe, H. (1993) Parametric families of multivariate distributions with given margins. *Journal of Multivariate Analysis* **46**(2), 262–282. [68]

Joe, H. (2015) *Dependence Modeling with Copulas.* Boca Raton, FL: CRC Press. [15, 60]

Joe, H., Smith, R. L. and Weissman, I. (1992) Bivariate threshold methods for extremes. *Journal of the Royal Statistical Society: Series B* **54**(1), 171–183. [67]

Katz, R. W., Parlange, M. B. and Naveau, P. (2002) Statistics of extremes in hydrology. *Advances in Water Resources* **25**(8), 1287–1304. [12]

Kotz, S. and Nadarajah, S. (2000) *Extreme Value Distributions: Theory and Applications.* London: Imperial College Press. [1, 19, 26]

Laurini, F. and Tawn, J. A. (2003) New estimators for the extremal index and other cluster characteristics. *Extremes* **6**(3), 189–211. [14]

Leadbetter, M., Lindgren, G. and Rootzén, H. (1983) *Extremes and Related Properties of Random Sequences and Processes.* New York: Springer. [1, 12]

Leadbetter, M. R. (1974) On extreme values in stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **28**(4), 289–303. [13]

Leadbetter, M. R. (1983) Extremes and local dependence in stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **65**(2), 291–306. [13]

Ledford, A. W. and Tawn, J. A. (1996) Statistics for near independence in multivariate extreme values. *Biometrika* **83**(1), 169–187. [28, 29, 30, 31, 67]

Ledford, A. W. and Tawn, J. A. (1997) Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society: Series B* **59**(2), 475–499. [28, 29]

Ledford, A. W. and Tawn, J. A. (1998) Concomitant tail behaviour for extremes. *Advances in Applied Probability* **30**(1), 197–215. [19, 28, 29]

Li, J., Peng, Z. and Nadarajah, S. (2008) A class of unbiased location invariant Hill-type estimators for heavy tailed distributions. *Electronic Journal of Statistics* **2**, 829–847. [7]

Longin, F. M. (ed.) (2017) *Extreme events in finance: a handbook of extreme value theory and its applications.* Wiley handbooks in financial engineering and econometrics. Hoboken, New Jersey: Wiley. [1]

Maraun, D., Rust, H. W. and Osborn, T. J. (2009) The annual cycle of heavy precipitation across the United Kingdom: A model based on extreme value statistics. *International Journal of Climatology* **29**(12), 1731–1744. [12]

Marcon, G., Padoan, S., Naveau, P., Muliere, P. and Segers, J. (2017) Multivariate nonparametric estimation of the Pickands dependence function using Bernstein polynomials. *Journal of Statistical Planning and Inference* **183**, 1–17. [26]

McNeil, A. J. and Frey, R. (2000) Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance* **7**(3-4), 271–300. [12]

Mikosch, T. (2006) Copulas: Tales and facts (with discussion). *Extremes* **9**, 3–20. [15]

von Mises, R. (1936) La distribution de la plus grande de n valeurs. *Revue Mathématique Union Interbalcanique* **1**, 141–160. Reprinted in Selected Papers of Richard von Mises, Volume II, American Mathematical Society, (1954), pp. 271–294. [3]

Naveau, P., Guillou, A., Cooley, D. and Diebolt, J. (2009) Modelling pairwise dependence of maxima in space. *Biometrika* **96**(1), 1–17. [30]

Nelsen, R. B. (2006) *An Introduction to Copula*. Second edition. New York: Springer. [15, 16, 31]

Omey, E. and Rachev, S. (1991) Rates of convergence in multivariate extreme value theory. *Journal of Multivariate Analysis* **38**(1), 36–50. [23]

Padoan, S. A. (2011) Multivariate extreme models based on underlying skew-t and skew-normal distributions. *Journal of Multivariate Analysis* **102**(5), 977–991. [19, 67]

Padoan, S. A. and Wand, M. (2008) Mixed model-based additive models for sample extremes. *Statistics & Probability Letters* **78**(17), 2850–2858. [12]

Pickands, J. (1971) The two-dimensional Poisson process and extremal processes. *Journal of Applied Probability* **8**(4), 745–756. [11]

Pickands, J. (1975) Statistical inference using extreme order statistics. *The Annals of Statistics* **3**(1), 119–131. [7, 8, 9]

Pickands, J. (1981) Multivariate extreme value distributions. In *Bulletin of the International Statistical Institute, Proceedings of the 43rd Session*, volume 2, pp. 859–878. The Hague: International Statistical Institute. [18, 21, 26]

Prescott, P. and Walden, A. T. (1980) Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika* **67**(3), 723–724. [6]

R Core Team (2017) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. [54]

Ramesh, N. and Davison, A. C. (2002) Local models for exploratory analysis of hydrological extremes. *Journal of Hydrology* **256**(1-2), 106–119. [12]

Ramos, A. and Ledford, A. (2005) Regular score tests of independence in multivariate extreme values. *Extremes* **8**(1), 5–26. [28]

Ramos, A. and Ledford, A. (2009) A new class of models for bivariate joint tails. *Journal of the Royal Statistical Society: Series B* **71**(1), 219–241. [28, 30]

Ramos, A. and Ledford, A. (2011) An alternative point process framework for modeling multivariate extreme values. *Communications in Statistics-Theory and Methods* **40**(12), 2205–2224. [28]

Reich, B. J. and Shaby, B. A. (2016) Time series of extremes. In *Extreme value modeling and risk analysis: Methods and applications*, eds D. Dey and J. Yan. Boca Raton, FL: CRC Press. [12]

Reiss, R. and Thomas, M. (2007) *Statistical Analysis of Extreme Values: With Applications to Insurance, Finance, Hydrology and Other Fields*. Third edition. Basel: Birkhäuser. [1]

Resnick, S. I. (1987) *Extreme Values, Regular Variation and Point Processes*. New York: Springer. [1, 26]

Resnick, S. I. (2007) *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. New York: Springer. [1]

Ribatet, M., Dombry, C. and Oesting, M. (2016) Spatial extremes and max-stable processes. In *Extreme value modeling and risk analysis: Methods and applications*, eds D. Dey and J. Yan. Boca Raton, FL: CRC Press. [1]

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

Schlather, M. (2001) Examples for the coefficient of tail dependence and the domain of attraction of a bivariate extreme value distribution. *Statistics & Probability Letters* **53**(3), 325–329. [29]

Schlather, M. and Tawn, J. A. (2003) A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika* **90**(1), 139–156. [30]

Segers, J. (2012) Max-stable models for multivariate extremes. *Revstat* **10**(1), 61–82. [19, 35]

Shumway, R. H. and Stoffer, D. S. (2011) *Time Series Analysis and Its Applications: With R examples*. Third edition. New York: Springer. [12]

Sibuya, M. (1960) Bivariate extreme statistics, I. *Annals of the Institute of Statistical Mathematics* **11**(2), 195–206. [27]

Sklar, A. (1959) Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris* **8**(1), 229–231. [16]

Smith, R. L. (1982) Uniform rates of convergence in extreme-value theory. *Advances in Applied Probability* **14**(3), 600–622. [5]

Smith, R. L. (1985) Maximum likelihood estimation in a class of nonregular cases. *Biometrika* **72**(1), 67–90. [6]

Smith, R. L. (1986) Extreme value theory based on the *r* largest annual events. *Journal of Hydrology* **86**(1–2), 27–43. [7]

Smith, R. L. (1989) Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone (with discussion). *Statistical Science* **4**(4), 367–393. [11, 12]

Smith, R. L. (1990) Max-stable processes and spatial extremes. Unpublished manuscript. [30]

Smith, R. L., Tawn, J. A. and Coles, S. G. (1997) Markov chain models for threshold exceedances. *Biometrika* **84**(2), 249 –268. [14]

Stephenson, A. G. (2002) evd: Extreme value distributions. *R News* **2**(2), 31–32. [54]

Stephenson, A. G. (2009) High-dimensional parametric modelling of multivariate extreme events. *Australian & New Zealand Journal of Statistics* **51**(1), 77–88. [19]

Stephenson, A. G. and Tawn, J. A. (2005) Exploiting occurrence times in likelihood inference for componentwise maxima. *Biometrika* **92**(1), 213–227. [25]

Süveges, M. (2007) Likelihood estimation of the extremal index. *Extremes* **10**, 41–55. [14]

Tawn, J. A. (1988a) Bivariate extreme value theory: Models and estimation. *Biometrika* **75**(3), 397–415. [20, 35, 67]

Tawn, J. A. (1988b) An extreme-value theory model for dependent observations. *Journal of Hydrology* **101**(1–4), 227–250. [7]

Varin, C. (2008) On composite marginal likelihoods. *AStA Advances in Statistical Analysis* **92**(1), 1. [25]

Varin, C., Reid, N. and Firth, D. (2011) An overview of composite lilelihood methods. *Statistica Sinica* **21**(1), 5–42. [25]

Varin, C. and Vidoni, P. (2005) A note on composite likelihood inference and model selection. *Biometrika* **92**(3), 519–528. [25]

Wadsworth, J. L. and Tawn, J. A. (2014) Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. *Biometrika* **101**(1), 1–15. [25]

Index

Claudio Semadeni, FRM, CAIA, PRM

Swiss & French citizen
45 years old
One child born 2009

Professional experience

Banque Cantonale Vaudoise (BCV)

Aug-18 – Present Assistant Vice President, **Quantitative Investment Manager** (Asset Management, Lausanne)

- Research and development of quantitative investment strategies (equities).
- Portfolio management of two commodities funds (AUM: USD 160mio. and USD 35mio.).

Edge Laboratories AG

Oct-14 – Jun-18 **Senior Quantitative Analyst** (Lausanne)

- Research on high dimensional statistical modelling of assets and risk factors.
- Development of asset and portfolio risk classification methods for suitability purposes.
- Constrained multi-asset portfolio optimization.

Ecole Polytechnique Fédérale de Lausanne (EPFL)

Sep-11 – Feb-18 **PhD Candidate and Teaching Assistant** (Institute of Mathematics – Chair of Statistics, Lausanne)

- Doctoral research on the modelling of extreme values with applications to financial data.
- Teaching assistant for mathematics courses at bachelor and master level.
- Supervision of student's research projects at master level (incl. two master theses).

Université Clermont Auvergne

2011 – 2017 **Lecturer** (Ecole de Management, Clermont-Ferrand)

Teaching a 2-day workshop on financial time series modelling and dynamic market risk quantification with R in the master program “Marchés Financiers”.

Haute Ecole de Gestion de Genève (HEG)

Feb-12 – Aug-15 **Lecturer** (Business Administration program, Geneva)

Teaching of bachelor level mathematics and statistics courses.

Unigestion SA

Apr-11 – Aug-11 Senior Vice President, **Quantitative Research Analyst** (Hedge Funds department, Geneva)

Jan-07 – Mar-11 Vice President, **Quantitative Research Analyst** (Hedge Funds department, Geneva)

In charge of the funds and portfolios modelling and risk monitoring.

- *Quantitative Analysis and Research:* Significantly contributed to the development of the portfolio/fund quantitative analysis and reporting platform.
- *Risk Control:* Performed on-going risk monitoring of portfolios and underlying funds. Determined expected behaviour and risk limits for portfolios and underlying funds, and performed periodical reviews of these limits.

Société de Gestion Financière et Commerciale SA (SGFC)

Jul-05 – Dec-06 **Head of Quantitative Analysis** (Fund of Hedge Funds team, Nyon area)

In charge of the team (3-5 junior analysts) managing a multi-strategy and multi-advisor fund of hedge funds within the family office of a major European family.

- *Investment management:* selection of investment recommendations from external advisors, and portfolio management according to the strategic allocation determined by the Investment Committee.
- *Quantitative analysis and risk management:* supervised in-house development of quantitative analysis and risk management reporting applications.
- *Operations:* developed and supervised the family office's fund of hedge funds portfolio administration activity.

The Finance Development Centre Ltd

Nov-03 – Jun-05 **Quantitative Analyst** (London)

Investment consulting and research projects in close collaboration with the company's Partners

- *Research and development:* improved and implemented proprietary portfolio optimisation and risk management methods based on Omega functions and associated metrics (generalisation of the Sharpe ratio for non-normal distributions) for hedge fund portfolios.
- *Portfolio optimisation:* quantitative selection of hedge funds, and optimisation of hedge funds portfolios.
- *Analysis:* prepared key sections of reports for clients/prospects and industry conference presentations.

Crédit Lyonnais Securities

Apr-03 – Oct-03 **Junior Quantitative Analyst** (Equity proprietary trading team, London)

- *Trading strategy development:* developed and backtested an equity market neutral trading strategy based on value, growth, momentum and earnings revisions investment style.
- *Performance and risk analysis:* developed an Excel application enabling in-depth performance and risk analysis for equity investment strategies.
- *Assistance to traders:* developed and improved an Excel application to generate buy/sell signals for an equity market neutral trading strategy.

Ecole Polytechnique Fédérale de Lausanne (EPFL)

May-02 – Mar-03 **Research Assistant** (Institute of Mathematics – Chair of Statistics, Lausanne)

- *Probability intervals in Bayesian wavelet estimation:* improved the method developed in my degree thesis enabling the determination of confidence intervals in Bayesian wavelet regression.
- *Detection of flammable liquids in fire debris:* calibrated and compared statistical inference methods for the detection of flammable liquids in fire debris; collaboration with the Institut de Police Scientifique et de Criminologie of the University of Lausanne.

Education

Sep-11 – present **Doctoral Studies in Mathematics**, Ecole Polytechnique Fédérale de Lausanne (**EPFL**)
PhD thesis: *Inference on the Angular Distribution of extremes*.

Oct-94 – Mar-02 **MSc in Mathematics**, Ecole Polytechnique Fédérale de Lausanne (**EPFL**)
Degree thesis in statistics completed at UC Santa Barbara, California (5 months): *Posterior Probability Intervals for Wavelet Models using Saddlepoint Methods.*

Professional certifications

CFA	Level 2 candidate of the *Chartered Financial Analyst* program
FRM	Earned *Financial Risk Manager* designation (Jan-17)
PRM	Earned *Professional Risk Manager* designation (Dec-10)
CAIA	Earned *Chartered Alternative Investment Analyst* designation (Sep-07)

Honours and awards

2012, 2014, 2016	*Dean's award* attributed by the School of Basic Sciences at EPFL for outstanding teaching.
1999	*Prix des Sports* awarded by the University of Lausanne and the EPFL to a university sportsman for outstanding commitment and achievement.

Publications

- Posterior probability intervals in Bayesian wavelet estimation, Biometrika (2004), with A. Davison & D. Hinkley.
- The evaluation of evidence in the forensic investigation of fire incidents (Part I), Forensic Science International (2005), with A. Biedermann, F. Taroni, O. Delemont & A. Davison.
- The evaluation of evidence in the forensic investigation of fire incidents (Part II), Forensic Science International (2005), with A. Biedermann, F. Taroni, O. Delemont & A. Davison.

Language skills

French:	Mother tongue.
English:	Fluent.
German:	Conversational (including Swiss German, needs refreshment).

Computer skills

Programming:	R, Matlab, C, SQL, VBA, LaTeX.
Software:	MS Office, Bloomberg, Pertrac.
Operating System:	Linux/Unix, Windows, Mac.

Past sport and extracurricular activities

- Top level volleyball career, 4 national titles and 45 selections with the Swiss national team (1991-2001)
- Board member and communication delegate of the *Association des Mathématiciens de l'EPFL* (2001-2003)
- Founder and President of the *Lausanne Université Club Beach Volleyball* (2001-2004)