Modeling Evolving Dependence between Bivariate Extremes through Multivariate Distortion Functions

Nahid Sadr

A Thesis

in

The Department

of

Mathematics and Statistics

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Science (Mathematics) at

Concordia University

Montreal, Quebec, Canada

August 2021

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Nahid Sadr

Entitled: Modeling Evolving Dependence between Bivariate Extremes through Multivariate Distortion Functions

and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Mathematics)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

_____ Thesis Supervisor

Dr. M. Mailhot

_____ Examiner

Dr. K. Herrmann

_____ Examiner

Dr. Y. Lu

Approved by _____

Dr. G. Dafni (Graduate Program Director)

_____

Dr. P. Sicotte (Dean of Faculty)

_____

Date

Abstract

Modeling Evolving Dependence between Bivariate Extremes through Multivariate Distortion Functions

Nahid Sadr

Probability distortion has been a means of pricing in insurance and finance for a long time. It is often utilized to transform the loss probability distribution to another distribution that assigns more weight to the outstanding potential losses. Parametric models for multivariate distributions can be proposed based on the extension of distortion transformations to the multivariate framework, which allows for generating new families of copulas from an existing one. These parametric representations are used in order to relate the distribution of bivariate climate extreme realizations to the distribution of projected extremes in the long term. The focus of this thesis is on modeling the bivariate distribution of temperature and precipitation annual maxima in Montreal by Extreme Value Theory, and propose a distortion of this model to explain the future projections based on three emissions scenarios representing different atmospheric concentrations of greenhouse gases (RCP 2.6, RCP 4.5 and RCP 8.5). Lastly, Akaike information criterion (AIC) and Bayesian information criterion (BIC) are employed to compare the performance of different distortions.

Acknowledgments

I would like to express my deepest gratitude to my supervisor, Dr. Mélina Mailhot for her invaluable guidance and continuous support during my studies at Concordia University. She not only made me a much better student but also helped me be a better person today than I was the day before.

I especially thank my friend and colleague, Emmanuel Osei Mireku, who never hesitates to share and provided a great amount of assistance in every step of this journey.

I would also like to extend my sincere thanks to my family for the moral support they have always given me.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The need to better understand extreme events is applied in a variety of fields such as hydrology (modeling floods), finance (modeling extreme losses), insurance (modeling extreme claims), engineering (designing structures to withstand extreme events), etc. Extreme weather events can have catastrophic physical and economic consequences in both urban and rural areas. Some of these impacts are studied by Kunkel et al. (1999), Easterling et al. (2000) and McMichael et al. (2006). Extreme Value Theory which was first introduced by Fisher and Tippett (1928) and then generalized by Von Mises (1954) and Jenkinson (1955), is the branch of probability that attempts to model the events that occur infrequently, but carry a high cost to society. Return level or Value at Risk (VaR) analysis is a univariate extreme value measure that relies on properly estimating the upper tail of a distribution.
The dependence between extreme values is also of special importance and can cause long-term and disastrous damages. In the bivariate case, tail dependence has been studied using the tool of copulas. A copula is the joint distribution of the percentiles of the random variables. By transforming the marginal distributions to uniform distributions on the unit cube, the tail dependence structure is separated from the margins (Sklar (1959), Joe (1997), Nelsen (2007), and Gudendorf and Segers (2010)). Copula-based methods are widely used in financial risk management and actuarial science for statistical modeling. However, the choice

of copula is crucial since selecting the wrong copula can lead to incorrect conclusions and flawed decisions in terms of the severity of these extreme events. Dupuis and Tawn (2001) explore inference problems caused by copula model misspecification. According to Watts (2016), the global financial crisis during 2007-8 is an example of the effect of wrong copula selection and neglecting extremal dependence. There have been numerous analyses on the dependence between climatological variables and extreme events. AghaKouchak et al. (2014) show that the traditional univariate risk assessment methods based on precipitation condition may substantially underestimate the risk of extreme events such as the 2014 California drought because of ignoring the effects of temperature. They argue that a bivariate model is necessary for assessing the risk of extreme events, especially in a warming climate. Cong and Brady (2012) and Wazneh et al. (2020) also investigate the interdependence between temperature and precipitation extremes. They found that ignoring the joint distribution and confounding impacts of precipitation and temperature lead to the underestimation of occurrence of probabilities for these two concurrent extreme events.

Distortion functions have been used in pricing of insurance contracts for a long time (Pflug (2006)). Distortion of copulas is a natural extension of distortion within the univariate framework. The study of multivariate distortions is of general interest because they can be used for generating, in a flexible way, new families of copulas. Valdez and Xiao (2011) introduced three approaches to distort a copula. We focus on the third approach which is proved by Morillas (2005) as a method to obtain new copulas from a given one. Such a transformation has been considered several times in the literature, under different names like distortion or transformation of a copula. It has originated from the study of distorted probability distribution functions and has been considered by several authors such as Frees and Valdez (1998), Durrleman et al. (2000), Genest and Rivest (2001), Durante and Sempi (2005), Klement et al. (2005), Wang et al. (2005), Charpentier (2008), Alvoni et al. (2009), Durante et al. (2010), and Lin et al. (2018).

In the long term, the extremes of the climate variables such as temperature and precipita-

tion are expected to be greater than the historical realizations due to global climate change. We aim to apply the multivariate distortion functions to the joint distribution of historical temperature and precipitation extremes as a model for future extremes. This method relates the distribution of historical observations to the distribution of future projections and allows for an evolving dependence structure.

This thesis is structured as follows. In Chapter 2, extreme events and univariate extreme value modeling by block maxima and threshold excess method, are explained. In Chapter 3, the concept of dependence and copula models are discussed in order to model bivariate extremes. Univariate and multivariate distortions are introduced in Chapter 4. Lastly, we apply our method to Montreal data in Chapter 5 to present a model for future temperature and precipitation annual maxima, by distorting the historical joint distribution.

Chapter 2

Univariate Extremes

2.1 Extreme Events

Catastrophic events have had a profound influence on our environment. Not only the causes of such occurrences should be investigated, but also plans should be made to deal with the massive financial loss that will result. The study of extreme events could help to understand the underlying risk of catastrophic events. Extreme events are cases that are unexpected and rare but have been observed through time. The probability of occurrence of each extreme event is small. Hydrologists use the term t-year return level (x_t) to define the frequency at which a particular event will occur

$$x_t = \inf\left\{x \in \mathbb{R} : F(x) \geq \left(1 - \frac{1}{t}\right)\right\}. \tag{2.1}$$

Assume that the probability of a specific occurrence for each year is p, with p being a small probability. Given that each year's events are independent of one another, the number of years in which a case occurs follows a geometric distribution with parameter p and expected value of $\frac{1}{p}$. To put it another way, a 100-year return level means that the chance of an extreme event exceeding the 100-year event is approximately 0.01. Statisticians call the t-year event the upper $\left(1 - \frac{1}{t}\right)$-percent quantile of the distribution of a random variable.

Statistical models are used to make inferences from historical data in the analysis of extreme values. The issue that arises by utilizing the classical statistical models is that extreme values may be treated as any other data points, and even sometimes are identified as "outliers" and excluded from the study. The point of our interest is the distribution of the tails of those commonly used statistical models. In this case, we use the Extreme Value Theory (EVT) to analyze the behavior of the tail of those distributions.

2.2 Extreme Value Theory

Let $X_1, X_2, \ldots, X_n$ be a set of independent random variables with the same distribution function F. The focus of EVT is about the distribution of the upper tail, i.e., $M_n = \max\{X_1, X_2, \ldots, X_n\}$, where X_i represents values of a process recorded on a regular period such as a year, month, or week. For example, in the case where $X_1, X_2, \ldots, X_n$ are the observations throughout a year, M_n is called the annual maximum. According to David and Nagaraja (2004), Galambos (1978), and Gumbel (1958), the distribution of maximum or the largest order statistic of independent and identically distributed random variables is obtained as follows as a result of probability theory:

$$\begin{aligned} \mathbb{P}(M_n \leq x) &= \mathbb{P}(X_1 \leq x, X_2 \leq x, \ldots, X_n \leq x) \\ &= \mathbb{P}(X_1 \leq x)\,\mathbb{P}(X_2 \leq x), \ldots, \mathbb{P}(X_n \leq x) \\ &= [F(x)]^n. \end{aligned} \tag{2.2}$$

Thus, the distribution of the maximum depends on the distribution of random variables. Suppose that G is the distribution of the maximum, z is the smallest possible value for which $F(z) = 1$,

$$z = \inf\{x \in \mathbb{R} \mid F(x) = 1\},$$

then,

$$F_X(x) = \begin{cases} 0 \leqslant F(x) < 1 & , x < z \\ 1 & , x \geqslant z \end{cases}, \tag{2.3}$$

hence,

$$G(x) = \lim_{n\to\infty} [F_X(x)]^n = \begin{cases} 0 & , x < z \\ 1 & , x \geqslant z \end{cases}. \tag{2.4}$$

Therefore, for large n, G is a degenerate distribution with a jump from zero to 1 at z. In order to avoid this problem, the M_n variable needs to be normalized by sequences of constants $\{a_n\}$ and $\{b_n\}$. As n increases, appropriate choices of $\{a_n\} > 0$ and $\{b_n\}$ can stabilize the location and scale:

$$\frac{M_n - b_n}{a_n}.$$

Theorem 2.1. *Extreme Value Distribution for Maxima [Fisher and Tippett (1928)] Suppose $X_1, X_2, \ldots, X_n$ are independent and identically distributed (iid) with distribution (df) F. If there exist constants $\{a_n > 0\}$ and $\{b_n\} \in \mathbb{R}$ such that:*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{M_n - b_n}{a_n} \leqslant z\right) \quad \to \quad G(z),$$

where $M_n = max(X_1, \ldots, X_n)$, G is non-degenerate distribution function. Then G is of one the following types:

$$\textit{I: } G(z) = \exp\left\{-\exp\left[-\left(\tfrac{z-b}{a}\right)\right]\right\}, \quad -\infty < z < \infty,$$

$$\textit{II: } G(z) = \begin{cases} 0, & z \leq b \\ \exp\left\{-\left(\tfrac{z-b}{a}\right)^{-\alpha}\right\}, & z > b \end{cases},$$

$$\textit{III: } G(z) = \begin{cases} \exp\left\{-\left[-\left(\tfrac{z-b}{a}\right)\right]^{\alpha}\right\}, & z < b \\ 1, & z \geq b \end{cases},$$

for parameters $a > 0, b$ and, in the case of families II and III, $\alpha > 0$.

According to Theorem 2.1, as the sample size gets large, the distribution of the renormalized maximum converges to a non-degenerate distribution within one of the types I, II, and III. The notable feature of this theorem is that it covers all possibilities for the limiting distribution of maxima, regardless of the population distribution F.
The three limits appeared in Theorem 2.1 I, II, and III are respectively named Gumbel, Frechet, and Weibull families, corresponding to different kinds of behaviors in the tail.

2.3 Generalized Extreme Value Distribution (GEV)

The Generalized Extreme Value (GEV) distribution was later introduced in 1954, to reformulate the three types of limiting distributions of maxima into one family of distributions.

Theorem 2.2. *[Independently proposed by Von Mises (1954) and Jenkinson (1955)] If there exists sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that:*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{M_n - b_n}{a_n} \leqslant z\right) \quad \to \quad G(z),$$

for a non-degenerate distribution function G, then G is a member of the Generalized Extreme

Value (GEV) family:

$$G(z) = \exp\left\{-\left(1+\xi\left(\frac{z-\mu}{\sigma}\right)\right)^{-1/\xi}\right\}, \tag{2.5}$$

for $\{z : 1+\xi\frac{(z-\mu)}{\sigma} > 0\}$*, where* μ*,* $\xi \in \mathbb{R}$ *and* $\sigma \in \mathbb{R}^+$*, and* $G(z) = 0$ *otherwise.*

A location parameter μ, a scale parameter $\sigma > 0$, and a shape parameter $\xi \in \mathbb{R}$ are defined in the GEV model. The cases $\xi > 0$ and $\xi < 0$ refer to the generalized Frechet and Weibull distributions, respectively. The Gumbel family of distribution is also derived by taking the limit of the GEV distribution as $\xi \to 0$:

$$G(z) = \exp\left\{-\exp\left\{-\frac{(z-\mu)}{\sigma}\right\}\right\}. \tag{2.6}$$

The expected value, variance, and mode of the GEV distribution are as follows:

$$\begin{aligned}
\mathrm{E}(X) &= \mu + \tfrac{\sigma}{\xi}\left(g_1 - 1\right), \\
\mathrm{Var}(X) &= \tfrac{\sigma^2}{\xi^2}\left(g_2 - g_1^2\right), \\
\mathrm{Mode}(X) &= \mu + \tfrac{\sigma}{\xi}\left[(1+\xi)^{-\xi} - 1\right],
\end{aligned}$$

where $g_k = \Gamma(1-k\xi)$ for $k = 1, 2$, and Γ is the gamma function $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$.
Note that the expected value of the GEV distribution is finite when $\xi < 1$ and the variance exists when $\xi < 0.5$.
As noted in the previous section, The support and tail behavior of the three families differ. In case $\xi = 0$, which is the Gumbel family, the upper tail decays exponentially. Whereas for the Frechet family ($\xi > 0$), the upper tail is heavy. The Weibull family of distributions ($\xi < 0$) limits the right tail to a finite upper bound. Figure 2.1 shows three different probability density functions in Gumbel, Frechet and Weibull families. The support of the distribution that is in the Frechet family is limited to a lower bound but does not have an upper bound. Since the tail is heavy in the Frechet case, the 0.99 quantile of this distribution can be very

large (dashed line). However, the large quantile of the distribution in the Weibull family is limited to the upper bound, which can be considered an advantage when it comes to hydrological data analysis. In hydrological data such as precipitation or wave height, one would prefer the support of the model to be limited to an upper bound, in order to estimate the height of a hydraulic structure. The support of the GEV distribution is shown in Table 2.1.

Family	Shape	Support of GEV
Gumbel	$\xi = 0$	$x \in (-\infty\,,\,+\infty)$
Frechet	$\xi > 0$	$x \in \left[\mu - \frac{\sigma}{\xi}\,,\,+\infty\right)$
Weibull	$\xi < 0$	$x \in \left(-\infty\,,\,\mu - \frac{\sigma}{\xi}\right]$

Table 2.1: Support of GEV distribution for each family of distributions in Theorem 2.1.

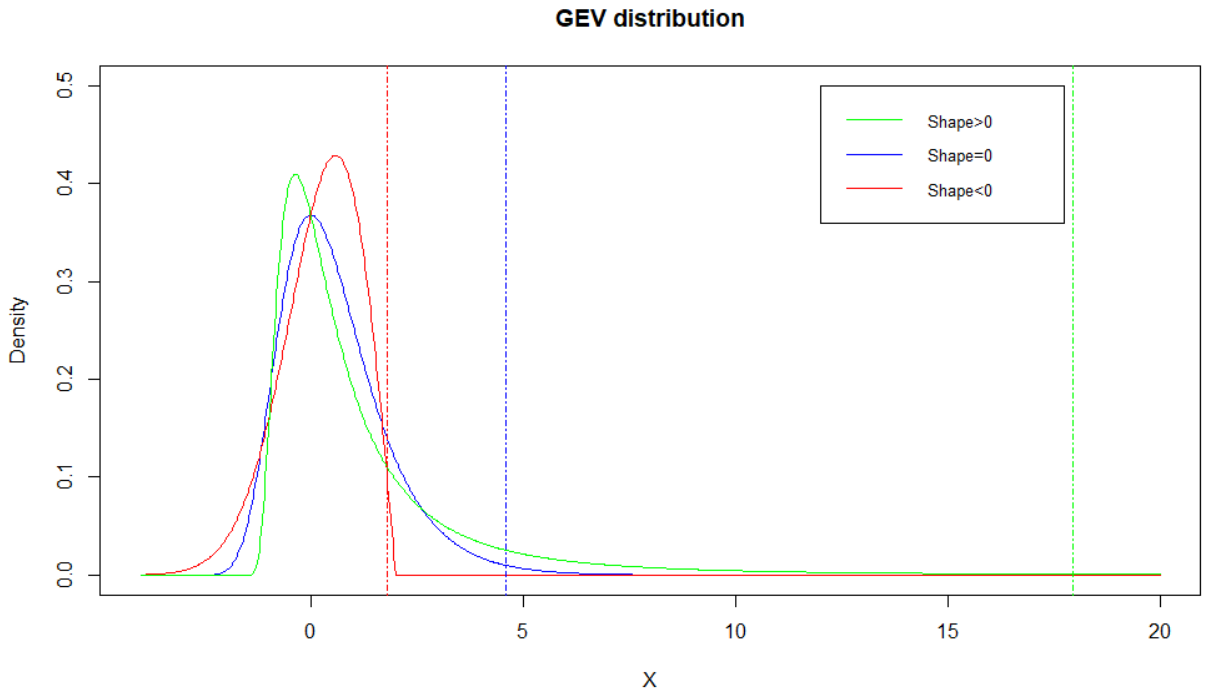

Figure 2.1: Probability density function of GEV distribution for different values of the shape parameter ($\xi \in \{-0.5, 0, 0.5\}$) and fixed values of the location and scale parameters ($\mu = 0, \sigma = 1$). The 0.99 percentile of each distribution is shown in dashed line.

Example 2.3.1. *If $X_1, X_2, \ldots$ is a sequence of independent standard Exponential variables, letting $a_n = 1$ and $b_n = log(n)$, the limit distribution of the maxima shifted and scaled by the normalizing constants, for large n is:*

$$\begin{aligned}
\mathbb{P}\left(\left(M_n - b_n\right)/a_n \leq z\right) &= F^n(z + \log n) \\
&= \left[1 - e^{-(z+\log n)}\right]^n \\
&= \left[1 - n^{-1}e^{-z}\right]^n \\
&\to \exp\left(-e^{-z}\right),
\end{aligned}$$

which is the Gumbel distribution in the GEV family of distributions ($\xi = 0$).

Example 2.3.2. *If $X_1, X_2, \ldots$ is a sequence of iid Uniform(0,1) variables. Let $a_n = 1/n$ and $b_n = 1$. The limit distribution of the renormalized maxima, for large n is:*

$$\begin{aligned}
\mathbb{P}\left(\left(M_n - b_n\right)/a_n \leq z\right) &= F^n\left(\frac{z}{n} + 1\right) \\
&= \left[1 + \frac{z}{n}\right]^n \\
&\to \exp\left(z\right),
\end{aligned}$$

which is in the GEV family of distribution ($\xi = -1$).

In practice, the normalizing sequences $\{a_n > 0\}$ and $\{b_n\}$ will be unknown since the parent distribution F is unknown. Fortunately, this issue can be resolved by the following approximation. Assuming that n is large enough for Theorem 2.2 to hold approximately:

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leqslant z\right) \quad \approx \quad G(z),$$

then it follows that:

$$\begin{aligned}
\mathbb{P}(M_n \leqslant z) &\approx G\left\{\left(z - b_n\right)/a_n\right\} \\
&= G^*(z),
\end{aligned}$$

where G^* is also a member of the GEV family of distributions. Thus, the GEV distribution

could be fitted directly to the sample maximum M_n without any normalization, and statistical inference for the tail behavior of the parent distribution F will be obtained by the estimation of G^*.

2.4 Generalized Pareto Distribution (GPD)

Another approach to the estimation of the extreme tail of the unknown parent distribution is provided by the Generalized Pareto distribution.

Theorem 2.3. *[Davison and Smith (1990) and Coles et al. (2001)] Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed random variables with cumulative distribution function F. Let $M_n = max\{X_1, \ldots, X_n\}$ be the sample maximum. Suppose F satisfies theorem 2.2 so that there exists a generalized extreme value distribution G as the limiting distribution of M_n. Then for a large enough threshold u, the conditional distribution function of the threshold excess, $X - u$ on $X - u > 0$, is approximated by:*

$$P(X - u \leqslant y | X > u) = H(y) = 1 - \left(1 + \frac{\xi y}{\sigma^*}\right)^{-1/\xi}, \tag{2.7}$$

for $\{y : y > 0$ and $(1+\xi y/\sigma^*) > 0\}$, with $\sigma^* = \sigma + \xi(u - \mu)$, where ξ, μ and σ are parameters of the corresponding GEV distribution G, where

$$G(x) = \exp\left\{-\left[1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right]^{-1/\xi}\right\}, \tag{2.8}$$

H is said to be a member of the Generalized Pareto (GP) distribution family.

Proof. If the limiting distribution of M_n is in the GEV family, for large enough n:

$$F^n(z) \approx \exp\left\{-\left[1 + \xi\frac{z - \mu}{\sigma}\right]^{-1/\xi}\right\},$$

$$n \log(F(z)) \approx -\left[1 + \xi \frac{z-\mu}{\sigma}\right]^{-1/\xi}.$$

For a large threshold u, a first-order Taylor Expansion implies that:

$$\log(F(u) \approx F(u) - 1 = -[-F(u) + 1],$$

therefore,

$$[1 - F(u)] \approx \frac{1}{n}\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]^{\frac{-1}{\xi}}.$$

Similarly, for $y > 0$:

$$[1 - F(u+y)] \approx \frac{1}{n}\left[1 + \xi\left(\frac{u+y-\mu}{\sigma}\right)\right]^{\frac{-1}{\xi}}.$$

Thus,

$$\begin{aligned}
\mathbb{P}(X > u + y \mid X > u) &= \frac{\mathbb{P}(X > u+y)}{\mathbb{P}(X > u)} = \frac{1 - F(u+y)}{1 - F(u)} \\
&\approx \frac{\frac{1}{n}\left[1 + \xi\left(\frac{u+y-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}}{\frac{1}{n}\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}} \\
&= \left[\frac{1 + \xi\left(\frac{u-\mu}{\sigma}\right)}{1 + \xi\left(\frac{u-\mu}{\sigma}\right)} + \frac{\xi\left(\frac{y}{\sigma}\right)}{1 + \xi\left(\frac{u-\mu}{\sigma}\right)}\right]^{-\frac{1}{\xi}} \\
&= \left[1 + \xi\frac{y}{\sigma^*}\right]^{-1/\xi}, \\
\Rightarrow \mathbb{P}(X - u \leqslant y \mid X > u) &= 1 - \left[1 + \xi\frac{y}{\sigma^*}\right]^{-1/\xi}.
\end{aligned}$$

■

By Theorem 2.3, if there exists a limiting distribution G for the sample maxima of a random sequence, then the sample threshold excesses have a corresponding conditional limiting distribution H, which belongs to the GP family. Furthermore, the shape parameter ξ of the GP distribution is equal to that of the corresponding GEV distribution, and the other parameter scale σ^* is determined by the threshold u and all three parameters of the

corresponding GEV distribution. By taking the limit of GP distribution as ξ approaches 0, the GP distribution for $\xi = 0$ takes the form of:

$$H(y) = 1 - \exp\left(-\frac{y}{\sigma}\right), \tag{2.9}$$

for $y > 0$, and $\sigma^* = \sigma$, where σ is the scale parameter of the corresponding Gumbel-type GEV distribution G,

$$G(x) = \exp\left[-\exp\left(-\frac{x-\mu}{\sigma}\right)\right]. \tag{2.10}$$

Therefore, the GP distribution is an Exponential distribution when $\xi = 0$.

The expected value and variance of the GP(σ, ξ) distribution are as follows:

$$\begin{aligned} \mathrm{E}(X) &= \tfrac{\sigma}{1-\xi}, \\ \mathrm{Var}(X) &= \tfrac{\sigma^2}{(1-\xi)^2(1-2\xi)}. \end{aligned}$$

Similar to the GEV distribution, the expected value of the GP distribution exists when $\xi < 1$ and the variance is finite for $\xi < 0.5$.

Example 2.4.1. *Let $X_1, X_2, \ldots$ be a sequence of independent Exponential(1) random variables and u be a high enough threshold. The limit distribution of the excess above u is:*

$$\begin{aligned} \mathbb{P}\{X - u > y \mid X > u\} &= \frac{\mathbb{P}(X > u + y)}{\mathbb{P}(X > u)} \\ &= \frac{e^{-(u+y)}}{e^{-u}} = e^{-y}. \end{aligned}$$

Because of the memory-less property of Exponential, the limit distribution of excess is again the Exponential distribution and is in the GP family of distributions with $\xi = 0$.

Example 2.4.2. *Let $X_1, X_2, \ldots$ be a sequence of iid Uniform(0,1) random variables and u*

be a high enough threshold. The limit distribution of the excess above u is:

$$\begin{aligned}\mathbb{P}\{X - u > y \mid X > u\} &= \frac{\mathbb{P}(X > u + y)}{\mathbb{P}(X > u)} \\ &= \frac{1 - (u + y)}{1 - u} = 1 - \frac{u}{1 - u}\,,\end{aligned}$$

which is in the GP family of distributions with $\xi = -1$.

Note that the shape parameters in Examples 2.3.2 and 2.4.2 are equal to -1 and imply the same behavior in the tail of Uniform distribution in both methods of modeling the tail of a parent distribution. The same result holds for Examples 2.3.1 and 2.4.1.

Similar to the GEV distribution, for the GP distribution, ξ is the shape parameter that determines the tail behavior of the conditional distribution of threshold excess. The case $\xi = 0$, is exactly the Exponential(σ) distribution. Whereas when $\xi > 0$, the upper tail is heavy. The support of GP distribution when $\xi < 0$, is limited to a finite upper bound. Figure 2.2 shows three different probability density functions in the GP family of distributions. The 0.99 quantile of this distribution is very large (dashed line) when ξ is positive since the tail is the heaviest in this case. The support of the GP distribution is shown in Table 2.2. Note that there always exists a lower bound of zero, because the excesses cannot be negative.

Shape	Support of GP
$\xi = 0$	$x \in [0\,,\, +\infty)$
$\xi > 0$	$x \in [0\,,\, +\infty)$
$\xi < 0$	$x \in [0\,,\, -\frac{\sigma^*}{\xi}]$

Table 2.2: Support of GEV distribution for each family of distributions in Theorem 2.1.

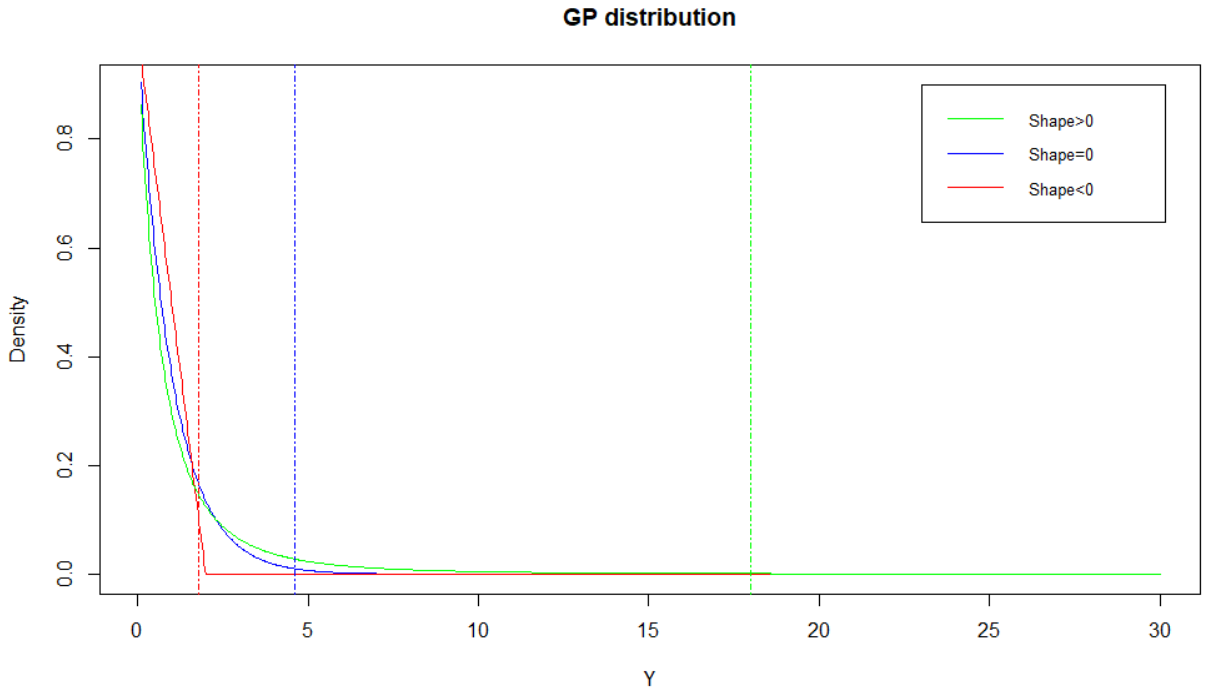

Figure 2.2: Probability density function of GP distribution for different values of the shape parameter $\xi \in \{-0.5, 0, 0.5\}$ and a fixed scale parameter $\sigma^* = 1$. The 0.99 percentile of each distribution is shown in dashed line.

2.5 Modeling Univariate Extremes

2.5.1 Block Maxima Method

The Block Maxima method is an approach to estimate the distribution of the sample maxima by the GEV distribution. Consider $X_1, X_2, \ldots$ is a series of independent observations recorded on a regular time-scale (e.g. daily precipitations, weekly temperatures). If $X_1, X_2, \ldots$ are blocked into sequences of observations of length n, $M_n = max\{X_1, X_2, \ldots, X_n\}$ would be the maximum of a block of size n. In the case where n is the number of observations in a year, $M_{n,1}, M_{n,2}, \ldots, M_{n,m}$ values are annual maxima throughout m years. The block-maxima model will divide the sample into blocks of equal size n, and then fit the GEV(μ, σ, ξ) distribution to the set of corresponding block maxima, for some μ, σ, and ξ parameters.

A critical issue for the block maxima method is the choice of the block size, which leads to a bias-variance trade-off. If the selected block size is too small, it violates the asymptotic properties of Theorem 2.2 and the fitted GEV can be poor with biased estimation since some of the block maxima may not even belong to extreme values. On the contrary, if the block size is too large, there will be only a few block maxima, leading to large variances in the estimation.

In terms of parameter estimations for the GEV distribution by the block-maxima model, various methods have been proposed, the maximum likelihood method is the most widely used. However, there are possible violations of the regularity conditions when fitting the GEV distribution by the Maximum likelihood method (Smith (1985) and Bücher and Segers (2017)):

- when $\xi > -0.5$ there exist regular maximum likelihood estimators, which have regular asymptotic properties;
- when $-1 < \xi < -0.5$ the maximum likelihood estimators may exist, but do not possess regular asymptotic properties;
- when $\xi < -1$ the maximum likelihood estimators do not exist.

When $\xi < -0.5$, the GEV distribution has a very short bounded upper-tail which is rare in extreme value data analysis.

2.5.2 Threshold-based Method

Another method to model the extreme upper tail of an unknown parent distribution is the threshold-based method which is done by fitting the GP distribution to observations whose value exceeds some high threshold (Hosking and Wallis (1987)). Consider $X_1, X_2, \ldots$ is a series of independent observations recorded on a regular time-scale (e.g. daily precipitations, weekly temperatures). Then for a large enough threshold u, the extreme events are defined

as $Y_j = X_{(j)} - u$, where $X_{(j)} \in \{X_i : X_i > u, i = 1, \ldots, n\}$, and $j = 1, \ldots, k$. Then $Y_1, \ldots, Y_k$ can be modeled by a GP Distribution.

$$Y_1, \ldots, Y_k \quad \sim \text{GP}(\sigma^*, \xi),$$

where $\sigma^* = \sigma + \xi(u - \mu)$ for some μ, σ, and ξ parameters.

Threshold-based models are more popular in extreme value analysis since more data points are taken into account. Also, in the Block Maxima method some information may be ignored by only considering maximum of each block. Figure 2.3 compares the two approaches for Danish insurance claim data (previously studied by other authors such as McNeil (1997), Resnick (1997), Emberchts et al. (1997), Sastry and Sinha (2010)), where the extreme claims are shown in green. It can be seen that the number of extreme values is larger in a Threshold-based method, with a threshold of size 10 which has been considered appropriate in the literature. On the other hand, in the Block Maxima method for instance, the maximum of the forth block is considered an extreme value while it is smaller than the second largest order statistic of some other blocks. Therefore, some extremes may be ignored in the Block Maxima method.

Like the choice of the block size in the block-maxima method for fitting a GEV distribution, choosing a large enough threshold u is a trade-off between bias and variance. In practice, it is expected to choose a threshold as low as possible so that adequate data could be obtained to fit the GP distribution well. However, a threshold set too low may lead to biased estimations of the GP distribution because extreme value assumptions for the model could be violated.

There have been numerous methods of threshold selection given a data set. Mean Residual Life (MRL) plots are the most basic method to choose a threshold but it has not been proved useful in practice (Zoglat et al. (2013)). Parameter Stability plots are widely used in threshold selection, however one has to make a careful selection since the estimations are

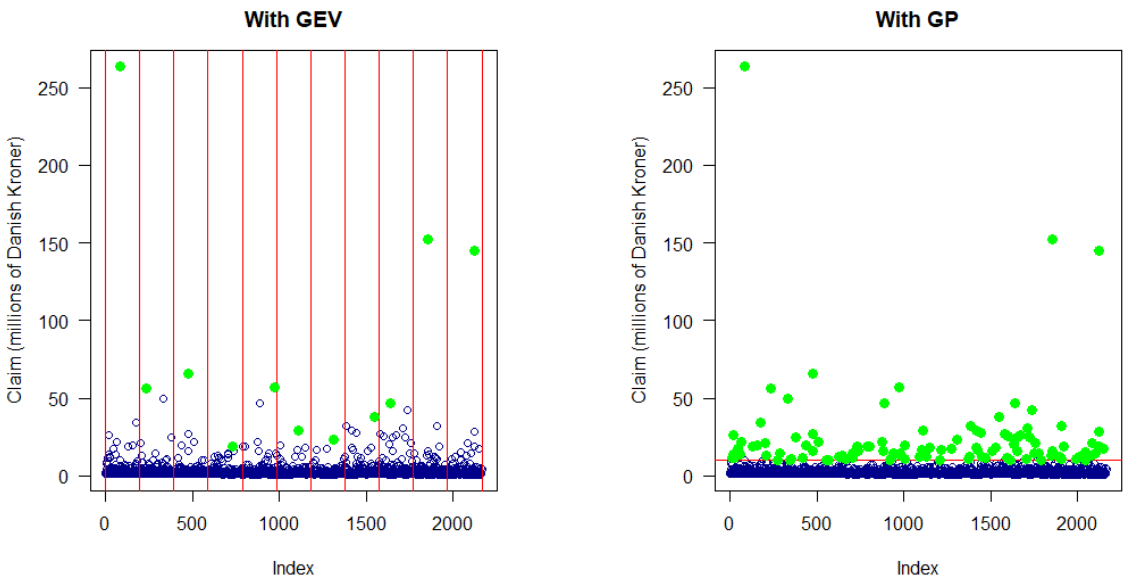

Figure 2.3: Block Maxima vs. Threshold-based method for Danish insurance claims.

pretty sensitive to the choice of threshold. Recent test-based approaches have been introduced to improve threshold diagnostic plots or threshold selection rules such as Northrop and Coleman (2014) and Wadsworth (2016). Some of these methods are explained in brief:

- **Method 1.** Mean Residual Life Plots:

  This approach is based on the expected value of the excess over a suitable choice of threshold. Assume $Y_1, \ldots, Y_k$ are excess over a high enough threshold u_0 and have the limit distribution $GP(\sigma_{u_0}, \xi)$ with $\sigma_{u_0} = \sigma + \xi(u_0 - \mu)$. If the GP is valid for the excess over the threshold u_0, then it should be valid for the excess over any larger threshold $u > u_0$. Hence, for $u > u_0$:

$$\mathrm{E}[X - u \mid X > u] = \frac{\sigma_u}{1 - \xi},$$

  where $\sigma_u = \sigma + \xi(u - \mu)$. Thus, σ_u can be written in terms of σ_{u_0}:

$$\sigma_u = \sigma_{u_0} - \xi(u_0 - \mu) + \xi(u - \mu) = \sigma_{u_0} - \xi(u - u_0),$$

which follows that:

$$\mathrm{E}[X-u \mid X>u]=\frac{\sigma_u}{1-\xi}=\frac{\sigma_{u_0}+\xi\,(u-u_0)}{1-\xi}=\frac{\sigma_{u_0}-\xi u_0}{1-\xi}+\frac{\xi}{1-\xi}u.$$

Since the expected value of excess of a threshold higher than the suitable threshold is a linear function of u, by plotting the mean of excess against the corresponding choices of thresholds, the threshold is selected at a point after which the plot looks linear.

- **Method 2.** Parameter Stability Plots:

  The main idea of a Parameter Stability plot is to fit GP over a range of thresholds and assess the stability of parameter estimates. If GP is a valid distribution of the excess above a high enough threshold u_0:

  - The estimate of the shape parameter (ξ) should not change for any $u > u_0$;
  - Since $\sigma_u = \sigma_{u_0} - \xi(u-u_0)$, the modified scale parameter $\sigma^* = \sigma_u - \xi u = \sigma_{u_0} - \xi u_0$ should also be stable for any $u > u_0$.

  Hence, by plotting the estimates of ξ and σ^* against different choices of the threshold, the threshold is selected at a point after which the two plots of $\hat{\xi}$ and $\hat{\sigma}^*$ are constant.

- **Method 3.** A Test-based Approach:

  One of the test-based approaches to improve threshold selection was introduced by Wadsworth (2016). Consider $u_1 < \ldots < u_k$ are different choices of threshold. The shape parameters of the GP models associated with each threshold are $\hat{\xi}_1, \ldots, \hat{\xi}_k$. Then the increments scaled by their asymptotic variance are defined as:

$$\begin{pmatrix} \xi_1^* \\ \xi_2^* \\ \vdots \\ \xi_{k-1}^* \end{pmatrix} := m^{1/2} \begin{pmatrix} \frac{(\hat{\xi}_1-\hat{\xi}_2)}{\left\{\left(I_2^{-1}-I_1^{-1}\right)_{\xi,\xi}\right\}^{1/2}} \\ \frac{(\hat{\xi}_2-\hat{\xi}_3)}{\left\{\left(I_3^{-1}-I_2^{-1}\right)_{\xi,\xi}\right\}^{1/2}} \\ \vdots \\ \frac{(\hat{\xi}_{k-1}-\hat{\xi}_k)}{\left\{\left(I_k^{-1}-I_{k-1}^{-1}\right)_{\xi,\xi}\right\}^{1/2}} \end{pmatrix},$$

where I_i is the Fisher information matrix on the interval (u_i, ∞), for $i = 1, \ldots, k$.

The idea of this method is that as the number of points (m) gets large, the scaled increments $\hat{\xi}^*$ should behave like a white noise process:

$$m \to \infty : \quad \xi_1^*, \ldots, \xi_{k-1}^* \sim N_{k-1}\left(0, 1_{k-1}\right).$$

A likelihood ratio test for the distribution of the scaled increments is proposed and the threshold is selected at a point after which the standard normal distribution cannot be rejected and the shape parameter is stable. This method can only be used for large number of data points. Although the threshold is automatically selected by this method, one has to choose k, the number of thresholds based on the number of data points.

Once a threshold is selected, the GP distribution can be fitted to the excess of the selected threshold using the Maximum Likelihood method of estimation. Results from Smith (1985) on the regularity of MLE as a function of the ξ also apply in fitting the GP distribution.

2.6 Diagnostics

Diagnostic plots such as PP plots, QQ plots, return level plots and density plots are used to assess the performance of a fitted model in extreme value analysis.

2.6.1 Diagnostic Plots for GEV

- **PP Plot:**

  A PP plot is a comparison of the empirical and fitted distribution functions. With ordered block maxima $z_{(1)} \leq z_{(2)} \leq \ldots \leq z_{(m)}$, the empirical distribution function evaluated at $z_{(i)}$ is given by

$$G\left(z_{(i)}\right) = i/(m+1).$$

By substitution of parameter estimates into equation 2.5, the corresponding model-based estimates, when $\hat{\xi} \neq 0$, are:

$$\hat{G}\left(z_{(i)}\right) = \exp\left\{-\left[1+\hat{\xi}\left(\frac{z_{(i)}-\hat{\mu}}{\hat{\sigma}}\right)\right]^{-1/\hat{\xi}}\right\}.$$

If the GEV model is performing well,

$$G\left(z_{(i)}\right) \approx \hat{G}\left(z_{(i)}\right),$$

for each i, so a probability plot, consisting of the points

$$\left\{\left(G\left(z_{(i)}\right), \hat{G}\left(z_{(i)}\right)\right), i = 1, \ldots, m\right\},$$

should lie close to the unit diagonal. Any substantial departures from linearity are indicative of some failing in the fit.

- **QQ Plot:**

  Similarly, a QQ plot consists:

$$\left\{\left(z_i\,,\, \hat{G}^{-1}(i/(m+1))\right), i = 1, \ldots, m\right\},$$

  and it should lie close to the unit diagonal. Like the PP plot, any substantial departures from linearity are indicative of some failing in the GEV model.

- **Return Level Plot:**

  Estimates of extreme quantiles of the annual maximum distribution are obtained by inverting equations 2.5 and 2.6 as follows:

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi}\left[1 - \{-\log(1-p)\}^{-\xi}\right], & \text{for } \xi \neq 0 \\ \mu - \sigma\log\{-\log(1-p)\}, & \text{for } \xi = 0 \end{cases},$$

where $G(z_p) = 1 - p$ and z_p is called a $\frac{1}{p}$ years-return level, the value that is expected to be exceeded once every $\frac{1}{p}$ observations. According to Coles et al. (2001), if we let $y_p = -\log(1-p)$, then a return level plot is obtained by plotting z_p against y_p on a logarithmic scale:

$$\{(\log y_p, \hat{z}_p), 0 < p < 1\}.$$

Figure 2.4 shows the following properties of a return level plot for three different shape parameters $\xi \in \{-0.2, 0, 0.2\}$.

- For $\xi < 0$, the return level plot is concave;
- For $\xi = 0$, the return level plot is linear;
- For $\xi > 0$, the return level plot is convex.

Confidence intervals are also added to the plot to increase its informativeness.

- **Density Plot:**

  For completeness, an equivalent diagnostic based on the density function is a comparison of the probability density function of the fitted GEV model with a histogram of the data. This is generally less informative than the previous plots, since the form of a histogram can vary substantially with the choice of grouping intervals.

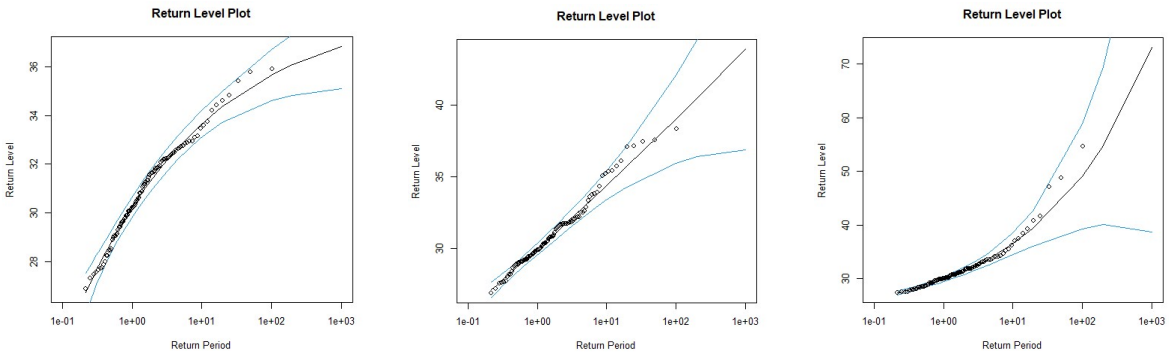

Figure 2.4: Return level plots for Weibull (left), Gumbel (middle), and Frechet (right) family of distributions.

2.6.2 Diagnostic Plots for GPD

After obtaining the proper threshold of the fitted GP distribution, similar to the GEV, the quality of the fit needs to be assessed by the following diagnostic plots.

- **PP Plot:**

  With the threshold u, threshold excesses $y_{(1)} \leq \ldots \leq y_{(k)}$ and an estimated model $\hat{H}$, the PP plot is given by:

$$\left\{\left(i/(k+1), \hat{H}\left(y_{(i)}\right)\right); i = 1, \ldots, k\right\},$$

  where $\hat{H}$ is the estimated distribution function of GPD.

  For $\hat{\xi} \neq 0$, $\hat{H}$ is of the form:

$$\hat{H}(y) = 1 - \left(1 + \frac{\hat{\xi} y}{\hat{\sigma}}\right)^{-1/\hat{\xi}}.$$

- **QQ Plot:**

  The QQ plot is given in a similar way:

$$\left\{\left(\hat{H}^{-1}(i/(k+1)), y_{(i)}\right), i = 1, \ldots, k\right\},$$

  where $\hat{H}^{-1}$ is of the following form when $\hat{\xi} \neq 0$:

$$\hat{H}^{-1}(y) = u + \frac{\hat{\sigma}}{\hat{\xi}}\left[y^{-\hat{\xi}} - 1\right].$$

  According to the theory, if excesses of u fit into the GP model, both the PP and QQ plots should consist of approximately linear points.

- **Return Level Plot:**

  The return level plot is given by the set of points $\{(m, \hat{x}_m)\}$ for large values of m,

where $\hat{x}_m$ is:

$$\hat{x}_m = u + \frac{\hat{\sigma}}{\hat{\xi}}\left[\left(m\hat{\zeta}_u\right)^{\hat{\xi}} - 1\right],$$

and

$$\hat{\zeta}_u = \Pr\{X > u\}.$$

The same properties of a return level plot of a GEV distribution hold for a return level plot of the GP model:

- For $\xi < 0$, the return level plot is concave;
- For $\xi = 0$, the return level plot is linear;
- For $\xi > 0$, the return level plot is convex.

- **Density Plot:**

  Similar to the GEV fit, the density function of the fitted GP model can be compared to a histogram of the threshold excess.

Chapter 3

Bivariate Modeling

3.1 Copulas

Consider the random vector $\underline{X}$ which contains n random variables $X_1, \ldots, X_n$ linked together through a joint density function f and joint distribution function F which is defined by:

$$F(x_1, x_2, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n),$$

which in bivariate setting is simplified to the joint distribution:

$$F(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2).$$

Copulas have become a popular tool in multivariate analysis. Copula functions are often used in financial risk management and actuarial science for modeling dependence structures in a sample of data. The basic idea is to transform all the marginal variables so that each of them follows the uniform distribution on $[0, 1]$. For any continuous random variable, the distribution function (d.f.) has a uniform distribution. Similarly, for a 2-dimensional random vector $X = (X_1, X_2)$ with continuous marginal distributions F_1 and F_2, the transformed vector $(F_1(X_1), F_2(X_2))$ has uniformly distributed margins and the copula of $\underline{X}$ is the joint

distribution of the transformed vector.

3.1.1 Definition and Properties

The following definition reviews the required properties for the function C to be a copula and holds for any bivariate copula by letting $n = 2$.

Definition 3.1.1. *(Morillas (2005) and Nelsen (2007)) An d-copula is the restriction to the unit n-cube* $[0,1]^{\mathrm{d}}$ *of a multivariate cumulative distribution function whose marginals are uniform on* $[0,1]$*, more precisely, an* d*-copula is a function* $C : [0,1]^{\mathrm{d}} \to [0,1]$ *that satisfies:*

- $C(x_1, \ldots, x_d) = 0$ *if* $x_i = 0$ *for any* $i = 1, \ldots, d$*;*
- $C(1, \ldots, 1, x_i, 1, \ldots, 1) = x_i$ *for each* $i = 1, \ldots, d$ *and all* $x_i \in [0,1]$*;*
- C *is d-increasing, i.e., the* C*-volume of every d-box is non-negative.*

The first and second conditions are known as boundary conditions whereas the third condition is known as monotonicity. The third condition means that for all d-box

$$B = [a_1,\ b_1] \times \ldots \times [a_d, b_d] \subseteq [0,1]^d,$$

with $a_i < b_i, i = 1, \ldots, d$, we have:

$$V_c(B) = \sum_{i_1=1}^{2} \ldots \sum_{i_d=1}^{2} (-1)^{i_1 + \ldots + i_d} C\left(x_{1i_1}, \ldots, x_{di_n}\right) \geq 0.$$

For example in the bivariate case, 2-increasing means that for every a_1, a_2, b_1, b_2 in $[0,1]$ such that $a_1 \leq a_2$ and $b_1 \leq b_2$, the rectangle inequality holds:

$$C(a_2, b_2) - C(a_2, b_1) - C(a_1, b_2) + C(a_1, b_1) \geq 0.$$

Figure 3.1 illustrates the rectangle inequality in the bivariate case.

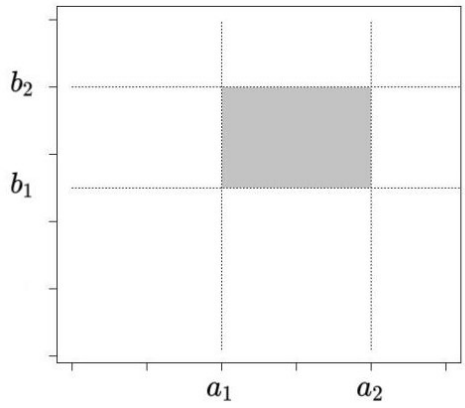

Figure 3.1: Illustration of the rectangle inequality for a bivariate distribution.

Sklar's Theorem is the foundation of the theory of copulas. It reveals the relationship between multivariate distribution functions and their univariate margins.

Theorem 3.1. *(Sklar (1959)) Let F be a d-dimensional joint distribution function with margins $F_1, \ldots, F_d$. Then there exists a copula C such that for all $x \in \mathbb{R}^d$,*

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)). \tag{3.1}$$

If $F_1, \ldots, F_d$ are all continuous, then the corresponding copula C is unique and can be written as $C(u_1, \ldots, u_d) = F\left(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\right), (u_1, \ldots, u_d) \in [0,1]^d$. Thus, for multivariate distributions with continuous margins, the univariate margins and the multivariate dependence structure (as described by their copulas) can be separated. For a random vector X with distribution function F and continuous margins $F_1, \ldots, F_d$, the copula of F (or X) is the distribution function C of $(F_1(X_1), \ldots, F_d(X_d))$.

Therefore, the joint distribution function for a bivariate random vector based on Sklar's Theorem ($d = 2$), is:

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

and the joint density function of the bivariate random vector joint by copula C is also derived

as follows:

$$\begin{aligned} f(x_1, x_2) &= \frac{\partial^2 F(x_1, x_2)}{\partial x_1 \partial x_2} \\ &= \frac{\partial^2 C(F_1(x_1), F_2(x_2))}{\partial x_1 \partial x_2} \\ &= \frac{\partial^2 C(F_1(x_1), F_2(x_2))}{\partial F_1(x_1)\, \partial F_2(x_2)} \frac{\partial F_1(x_1)}{\partial x_1} \frac{\partial F_2(x_2)}{\partial x_2} \\ &= c(F_1(x_1), \ldots, F_2(x_2))\, f_1(x_1)\, f_2(x_2), \end{aligned}$$

where c is called the density copula:

$$c(u_1, u_2) = \tfrac{\partial^2}{\partial u_1 \partial u_2} C(u_1, u_2).$$

Fréchet bounds refer to another important property of copulas that shows any copula is bounded to a lower bound and an upper bound.

Theorem 3.2. *(Fréchet-Hoeffding Bounds) For every copula $C(u_1, \ldots, u_d)$,*

$$\max\left\{\sum_{i=1}^{d} u_i + 1 - d, 0\right\} \leq C(u) \leq \min\{u_1, \ldots, u_d\}.$$

The lower and upper bounds are denoted by $W(u_1, \ldots, u_d)$ and $M(u_1, \ldots, u_d)$ respectively.

Proof. If C is the distribution function of $(U_1, \ldots, U_d)$, then

$$C(u_1, \ldots, u_d) = P(U_1 \leq u_1, \ldots, U_d \leq u_d).$$

Given that

$$P(U_1 \leq u_1, \ldots, U_d \leq u_d) \leq P(U_i \leq u_i), \quad \text{for } i = 1, \ldots, d,$$

then

$$\begin{aligned} P(U_1 \leq u_1, \ldots, U_d \leq u_d) &\leq \min(P(U_1 \leq u_1), \ldots, P(U_d \leq u_d)) \\ &= \min(u_1, \ldots, u_d) = M(u_1, \ldots, u_d). \end{aligned}$$

Also,

$$\begin{aligned} P\left(U_1 \leq u_1, \ldots, U_d \leq u_d\right) &= P\left(\bigcap_{1\leq i\leq d} \{U_i \leq u_i\}\right) \\ &= 1 - P\left(\bigcup_{1\leq i\leq d} \{U_i > u_i\}\right) \\ &\geq 1 - \sum_{i=1}^{d} P\left(U_i > u_i\right) \\ &\geq 1 - \sum_{i=1}^{d} (1 - u_i) \\ &\geq 1 - d + \sum_{i=1}^{d} u_i = W(u_1, \ldots, u_d). \end{aligned}$$

■

Note that $M(u_1, \ldots, u_d)$ is a copula for any value of d, however, $W(u_1, \ldots, u_d)$ is only a copula for $d = 2$. M and W are referred to as the comonotonic and counter-comonotonic copulas, respectively. There exist different families of copulas that are introduced in the next sections.

3.1.2 Perfect Dependence and Independence Copulas

The first family of copulas consists of the cases where the random variables are perfectly dependent or independent:

- **Independence Copula:** If X_1, X_2 are independent, then $C(\boldsymbol{u}) = \prod_{i=1}^{2} u_i = u_1 u_2$, which is called the independence copula. It is sometimes denoted by $\Pi(u)$.

- **Comonotonic Copula:** If X_1, X_2 are perfectly positively dependent, i.e., $X_2 = T\left(X_1\right)$ where T is an almost surely strictly increasing function, then $C(\boldsymbol{u}) = M\left(u_1, u_2\right) = \min\left\{u_1, u_2\right\}$, which is called the comonotonicity copula.

- **Counter-comonotonic Copula:** If X_1, X_2 are perfectly negatively dependent, i.e., $X_2 = T\left(X_1\right)$ where T is an almost surely strictly decreasing function, then $C(\boldsymbol{u}) =$

$W(u_1, u_2) = \max\{u_1 + u_2 - 1, 0\}$, which is called the counter-comonotonicity copula.

3.1.3 Elliptical Copulas

Elliptical bivariate distributions have contour shapes of ellipses. These copulas can induce negative and positive associations and are appropriate when radial symmetry (i.e. equivalent upper and lower tail dependence) is evident between the two random variables of interest. Two well-known examples of elliptical distributions are the bivariate Normal and Student t.

- **Normal Copula:** The form of the bivariate Normal copula is given by:

$$C_{\rho}^{\text{Gauss}}(u_1, u_2) = \Phi_{\rho}\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2)\right),$$

  where Φ is the CDF of a standard univariate normal random vector and Φ_{ρ} is the CDF of a bivariate normal random variable with mean $\underline{\mu} = (0, 0)$, marginal variances 1, and correlation $\rho \in [-1, 1]$. This copula assigns equal degrees of positive and negative dependence to the joint distribution whilst assuming no tail dependence. As the ρ approaches -1 and 1, the Normal copula attains the Fréchet lower and upper bound, respectively.

- **Student t Copula:** The Student t copula has two parameters, ρ and ν (degrees of freedom), where the latter controls the heaviness of the tails. Its function is given by:

$$C_{\nu,\rho}^{t}(u_1, u_2) = t_{\nu,\rho}\left(t_{\nu}^{-1}(u_1), t_{\nu}^{-1}(u_2)\right),$$

  where t_{ν} is the cumulative distribution function (CDF) of a standard univariate t-distribution with ν degrees of freedom and $t_{\nu,\rho}$ is the CDF of a bivariate t-distribution with mean $\underline{\mu} = (0, 0)$, correlation $\rho \in [-1, 1]$ and ν degrees of freedom. This copula has more probability mass in the tails than the Normal copula. As the value of ν increases, it approximates a Gaussian distribution.

3.1.4 Archimedean Copulas

A widely used class of copulas are the Archimedean copulas, which can be expressed in terms of their generator $\psi : [0, \infty) \to [0, 1]$ as

$$C(u_1, u_2) = \psi\left\{\psi^{-1}(u_1) + \psi^{-1}(u_2)\right\}, \tag{3.2}$$

for the bivariate case. The generator function of a bivariate Archimedean copula has the following properties:

(1) Strictly decreasing convex function;

(2) $\psi(0) = \infty$ and $\psi(1) = 0$;

and ψ^{-1} represents the inverse of the generator function. Clayton, Frank and Gumbel copula are the most commonly used examples of Archimedean copulas that will be defined in this section.

- **Clayton Copula:** Clayton copula is known to have a lower tail dependence. The bivariate Clayton copula can be written as the following:

$$C_\theta^{\mathrm{Cl}}(u_1, u_2) = [u_1^{-\theta} + u_2^{-\theta} - 1]^{-1/\theta}, \quad \text{for } \theta \in [-1, \infty)\backslash\{0\}, \tag{3.3}$$

  which is an Archimedean copula with generator function $\psi(t) = \frac{1}{\theta}(t^{-\theta} - 1)$. Since Clayton copula exhibits no upper tail dependency, it is often used to model lower tail dependency. The limiting case $\theta \to 0$ implies independence, while the case $\theta \to \infty$ allows for perfect dependence. The Counter-comonotonic copula is also attained for $\theta = -1$, in the bivariate setting.

- **Frank Copula:** The bivariate Frank copula is of the form:

$$C_\theta^{Fr}(u_1, u_2) = -\frac{1}{\theta}\ln\left[1 + \frac{\left(e^{-\theta u_1} - 1\right)\left(e^{-\theta u_2} - 1\right)}{e^{-\theta} - 1}\right], \quad \text{for } \theta \in \mathrm{R}\backslash\{0\}, \tag{3.4}$$

where θ is the parameter of copula that controls the amount of dependence. When $\theta \to 0$, the copula approaches the independence case, while for $\theta \to \infty$ the Comonotonic copula is obtained. In the bivariate case, the Counter-comonotonic copula is attained when $\theta \to -\infty$. The generator function associated with the Frank copula is $\psi(t) = -\ln\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right)$.

- **Gumbel Copula:** In the context of insurance claim payments, the Gumbel copula is riskier in the upper tail for an insurer than the Clayton copula. Let the generator function be $\psi(t) = (-\ln t)^{\theta}$, then the Gumbel copula is given by:

$$C_{\theta}^{Gu}(u_1, u_2) = \exp\left\{-\left[(-\ln u_1)^{\theta} + (-\ln u_2)^{\theta}\right]^{1/\theta}\right\}, \quad \theta \in [1, \infty). \tag{3.5}$$

  Although the Gumbel copula is widely used to model the upper tail dependence, it does not allow for negative dependence. In the case $\theta = 1$, the Gumbel copula is equal to the independence copula and as θ gets large, the amount of positive dependence increases between the variables.

Figure 3.2 shows the contour plot of the bivariate density function of the same marginal distributions $N(0, 1)$, jointed by different copulas. It visualizes the fact that Clayton copula and Gumbel copula are appropriate for lower tail dependence and upper tail dependence modeling, respectively. It can also be seen that the tails of the Student t copula are heavier than the Normal copula.

3.1.5 Extreme Value Copulas

In cases where the interest is in joint extreme events, extreme value copulas might be a good choice for modeling the dependence structure between rare events. Let (X_i, Y_i), for $i = 1, \ldots, n$ be a random vector, then $X = max(X_1, \ldots, X_n)$ and $Y = max(Y_1, \ldots, Y_n)$ are the component-wise maxima. If by Theorem 2.2, the limiting distribution of renormalized X and the limiting distribution of renormalized Y converge to a non-degenerate distribution in

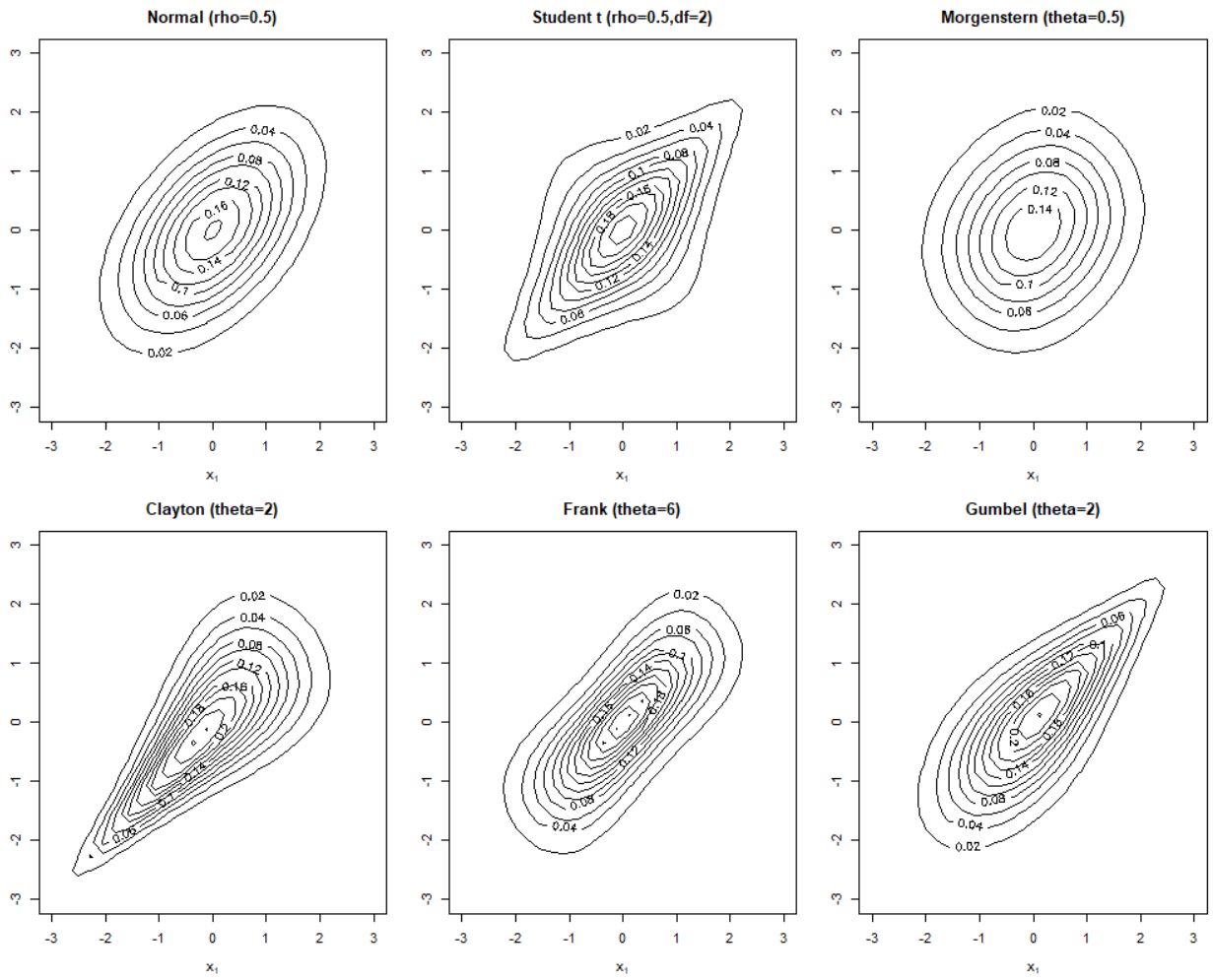

Figure 3.2: Contour plots of the joint density function for standard normal marginals connected by different copulas.

the GEV family of distributions, the joint distribution of (X, Y) is also non-degenerate and is of the following form:

$$C(u, v) = \exp\{-V(-1/\log u, -1/\log v)\}, \tag{3.6}$$

where

$$V(x, y) = \int_0^1 \max\left(\frac{w}{x}, \frac{1-w}{y}\right) 2dH(w), \tag{3.7}$$

and H the spectral measure, is a distribution function on $[0,1]$ with mean $\frac{1}{2}$. If H is differentiable with density h, then $V(x,y)$ is equal to:

$$V(x,y) = 2\int_0^1 \max\left(\frac{w}{x}, \frac{1-w}{y}\right) h(w)dw. \tag{3.8}$$

Since there could be many possibilities for the spectral measure H, there is no general parametric form to bivariate extreme value family. The following two non-differentiable cases for H can be considered as the independence and perfect dependence scenarios:

- When H is a distribution function with $P(w=0) = 0.5$ and $P(w=1) = 0.5$, the mean of the distribution H would be $\frac{1}{2}$ and

  $$V(x,y) = 2\left[\frac{1}{2}\left(\frac{1}{y}\right) + \frac{1}{2}\left(\frac{1}{x}\right)\right] = \frac{1}{y} + \frac{1}{x},$$

  by the substitution of V in equation 3.6, we have:

  $$C(u,v) = \exp\{-[-\log(u) - \log(v)]\} = uv,$$

  which is the independence copula.

- When H is a distribution function with $P(w=0.5) = 1$, the mean is obviously $\frac{1}{2}$ and

  $$V(x,y) = 2\ \max\left\{\frac{1}{2x}, \frac{1}{2y}\right\} = \max\left\{\frac{1}{x}, \frac{1}{y}\right\}.$$

  By the substitution of V in equation 3.6, we have:

  $$C(u,v) = \exp\{-\max(-\log(u), -\log(v))\} = \min(u,v),$$

  which is the Comonotonic copula.

Unfortunately, there is no finite parametric form of V that covers all possible extreme

value copulas (Tawn (1988)). In this section, some certain parametric forms of extreme value copulas are reviewed.

- **Gumbel Copula:** As already noted in Section 3.1.4, Gumbel copula allows for extremal (upper tail) dependence behavior and is the only Archimedean copula (Genest and Rivest (1989)). Due to the fact that it has only one parameter, Gumbel copula is the simplest and most popular extreme value copula. The form of Gumbel copula (equation 3.5) can be derived by the spectral measure H, for which:

$$h(w) = \frac{1}{2}(\delta - 1)\{w(1-w)\}^{-1-\delta}\left\{w^{-\delta} + (1-w)^{-\delta}\right\}^{\frac{1}{\delta}-2}.$$

  The issue that arises when modeling bivariate extremes using this copula is that it does not allow for negative dependence and is symmetric in the extreme levels of dependence.

- **Bilogistic Copula:** Bilogistic copula which is a generalized form of the Gumbel copula and has two parameters α and β, is derived by the spectral measure that has the density:

$$h(w) = \frac{1}{2}(1-\alpha)(1-w)^{-1}w^{-2}(1-u)u^{1-\alpha}\{\alpha(1-u) + \beta u\}^{-1},$$

  where $0 < w < 1, 0 < \alpha < 1, 0 < \beta < 1$ and u is the solution of

$$(1-\alpha)(1-w)(1-u)^{\beta} - (1-\beta)wu^{\alpha} = 0.$$

  Unlike the Gumbel copula, the Bilogistic copula allows for asymmetry in the dependence between random variables. The larger the difference between α and β is, the more asymmetric the dependence structure is, and the case where $\alpha = \beta$, the exact Gumbel copula is attained which is symmetric.

- **Coles-Tawn Copula:** Another extreme value copula that can explain the asymmetry

in the dependence was proposed by Coles and Tawn (1991), which is sometimes called Dirichlet copula. The density of the spectral measure to build this copula is of the form:

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

for $0 < w < 1$, where the parameters α and β are positive. The case $\alpha = \beta$ is the only case that this copula is symmetric.

3.2 Dependence Measures

The choice of the copula in bivariate extreme value analysis is crucial. A real-life example of choosing the inappropriate copula model is discussed by Watts (2016). After the 2007-2008 financial crisis, many contributing causes were given. The financial model that came under particularly strong criticism was the Normal copula. Normal copula was described as mathematically "beautiful", but in some cases a flawed model that is widely used for its simplicity and tractability. The choice of Normal copula in 2007-08 caused trillions of dollars of losses only because it does not allow for the upper tail dependence and in some cases could underestimate the extreme losses.

Although there exist different methods to assess the performance of different copula fits, dependence measures can be used as a means that leads us to an appropriate family of copulas. Some important measures of overall dependence (e.g. Kendall's tau, Spearman's rho, etc.) and tail dependence are discussed in this section.

3.2.1 Pearson's rho

Let X_1, X_2 be two random variables. The Pearson's ρ of X_1 and X_2 is defined as:

$$\rho\left(X_1, X_2\right) = \frac{\operatorname{Cov}\left(X_1, X_2\right)}{\sqrt{\operatorname{Var}\left(X_1\right)\operatorname{Var}\left(X_2\right)}},$$

with range $[-1, 1]$. Despite its popularity, Pearson's correlation is not a particularly appropriate measure of the dependence between X_1 and X_2. Pearson's rho is usually an appropriate measure of dependence for Normal multivariate distributions and it may not always exist. Pearson's rho is defined when the variances of X_1 and X_2 are finite and nonzero. These are drawbacks for analyzing dependence structures of heavy tailed distributions. In many cases, the data follow a joint heavy tailed distribution and the variances are not finite. Another disadvantage of Pearson's correlation is not necessarily invariant to increasing transformations.

3.2.2 Spearman's rho

Spearman's rho is a measure of rank correlation. Unlike linear correlation (Pearson's rho), rank correlation depends only on the copula of a bivariate distribution and not on the margins. Rank correlation coefficients are invariant under strictly increasing transformations. Rank correlation coefficients of a pair of random variables measure the extent to which one variable follows the other in an increasing or decreasing behavior. The dependence that they measure is known as concordance. Let the random pair (X_1, X_2) have the joint distribution F, and F_i for $i = 1, 2$ be the univariate margins. Then the Spearman's rho (or Spearman's rank correlation) is defined by:

$$\rho_S = \rho\left(F_1\left(X_1\right), F_2\left(X_2\right)\right) = \frac{\mathbb{E}\left[\left(F_1\left(X_1\right) - \mu_{F_1}\right)\left(F_2\left(X_2\right) - \mu_{F_2}\right)\right]}{\sigma_{F_1}\sigma_{F_2}}, \tag{3.9}$$

where $F_1\left(X_1\right)$ and $F_2\left(X_2\right)$ are $U(0,1)$ so $\mathrm{E}\left[F_i\left(X_i\right)\right] = 1/2$ and $\mathrm{Var}\left[F_i\left(X_i\right)\right] = 1/12, i = 1, 2$ and thus

$$\rho_S = 12 \iint F_1\left(x_1\right) F_2\left(x_2\right) dF\left(x_1, x_2\right) - 3 = 12 \iint u_1 u_2 dC\left(u_1, u_2\right) - 3. \tag{3.10}$$

3.2.3 Kendall's Tau

Kendall's tau is another popular measure of rank correlation. Thus, it is invariant under strictly increasing transformations and can be written in terms of the copula function. It measures the difference between the probability of two random concordant pairs and the probability of two random discordant pairs. Let the random pair (X_1, X_2) have the joint distribution F, and F_i for $i = 1, 2$ be the univariate margins. If (X_1^*, X_2^*) is a pair independent of (X_1, X_2), Kendall's tau is

$$\begin{aligned}\tau &= \mathrm{P}\left(\left(X_1 - X_1^*\right)\left(X_2 - X_2^*\right) > 0\right) - \mathrm{P}\left(\left(X_1 - X_1^*\right)\left(X_2 - X_2^*\right) < 0\right) \\ &= 2\mathrm{P}\left(\left(X_1 - X_1^*\right)\left(X_2 - X_2^*\right) > 0\right) - 1 \\ &= 4\int F dF - 1 = 4\int C\left(u_1, u_2\right) dC\left(u_1, u_2\right) - 1,\end{aligned}$$

with $1 - 0 = 1$ as one limit, and $0 - 1 = -1$ as the other limit.

3.2.4 Lower and Upper Tail Dependence Parameters

In bivariate extreme value analysis, the amount of dependence in the upper quadrant tail or lower quadrant tail, is of interest. The upper and lower tail dependence parameters are defined in terms of limiting conditional probabilities. For upper tail dependence, one examines the probability that X_1 exceeds its α-level quantile given that X_2 exceeds its α-level quantile, by letting α approach 1 and vice versa. The upper tail dependence parameter is defined as:

$$\begin{aligned}\lambda_{upper}(X_1, X_2) &= \lim_{u \to 1} P\left[X_1 > \mathrm{VaR}_u\left(X_1\right) \mid X_2 > \mathrm{VaR}_u\left(X_2\right)\right] \\ &= \lim_{u \to 1} \frac{P\left[X_1 > \mathrm{VaR}_u\left(X_1\right), X_2 > \mathrm{VaR}_u\left(X_2\right)\right]}{P\left[X_2 > \mathrm{VaR}_u\left(X_2\right)\right]} \\ &= \lim_{u \to 1} \frac{P(V > u, U > u)}{P(U > u)} \\ &= \lim_{u \to 1} \frac{1 - 2u + C(u, u)}{1 - u}.\end{aligned}$$

Variables are asymptotically independent in the upper tail if they have a copula for which $\lambda_{upper} = 0$. If $\lambda_{upper} > 0$, the variables are asymptotically dependent, with higher values indicating stronger dependence.

The lower tail dependence parameter is defined in a similar way:

$$\begin{aligned}
\lambda_{lower}(X_1, X_2) &= \lim_{u \to 0} P\left[X_1 \leqslant \mathrm{VaR}_u\left(X_1\right) \mid X_2 \leqslant \mathrm{VaR}_u\left(X_2\right)\right] \\
&= \lim_{u \to 0} \frac{P\left[X_1 \leqslant \mathrm{VaR}_u\left(X_1\right), X_2 \leqslant \mathrm{VaR}_u\left(X_2\right)\right]}{P\left[X_2 \leqslant \mathrm{VaR}_u\left(X_2\right)\right]} \\
&= \lim_{u \to 0} \frac{P(V \leqslant u, U \leqslant u)}{P(U \leqslant u)} \\
&= \lim_{u \to 0} \frac{C(u, u)}{u},
\end{aligned}$$

provided the limits exist.

As noted in section 3.2, choosing the wrong copula could result in massive losses due to the extremal dependence structure of the data. Figure 3.3 shows that for the same amount of overall dependence (Kendall's tau = 0.4), the behavior in the upper tail of the Normal and Gumbel copulas are very different. Hence, using the Normal copula to model data with a non-zero λ_{upper} results in underestimation of the upper tail. Since the upper tail dependence parameter for the Gaussian copula is exactly zero, it is asymptotically independent regardless of the correlation parameter ρ. Whereas for the Gumbel copula the upper tail dependence parameter depends on the parameter of the copula ($\lambda_{upper}(X_1, X_2) = 2 - 2^{1/\theta}$). Therefore, the Gumbel copula is asymptotically dependent in the upper tail and asymptotic independence is obtained only if the variables are exactly independent ($\theta = 1 \ \rightarrow \ \lambda_{upper} = 0$).

3.2.5 Coefficient of Tail Dependence

When the random variables are asymptotically independent, one would be interested to investigate the rate at which asymptotic independence is approached. For instance, for the

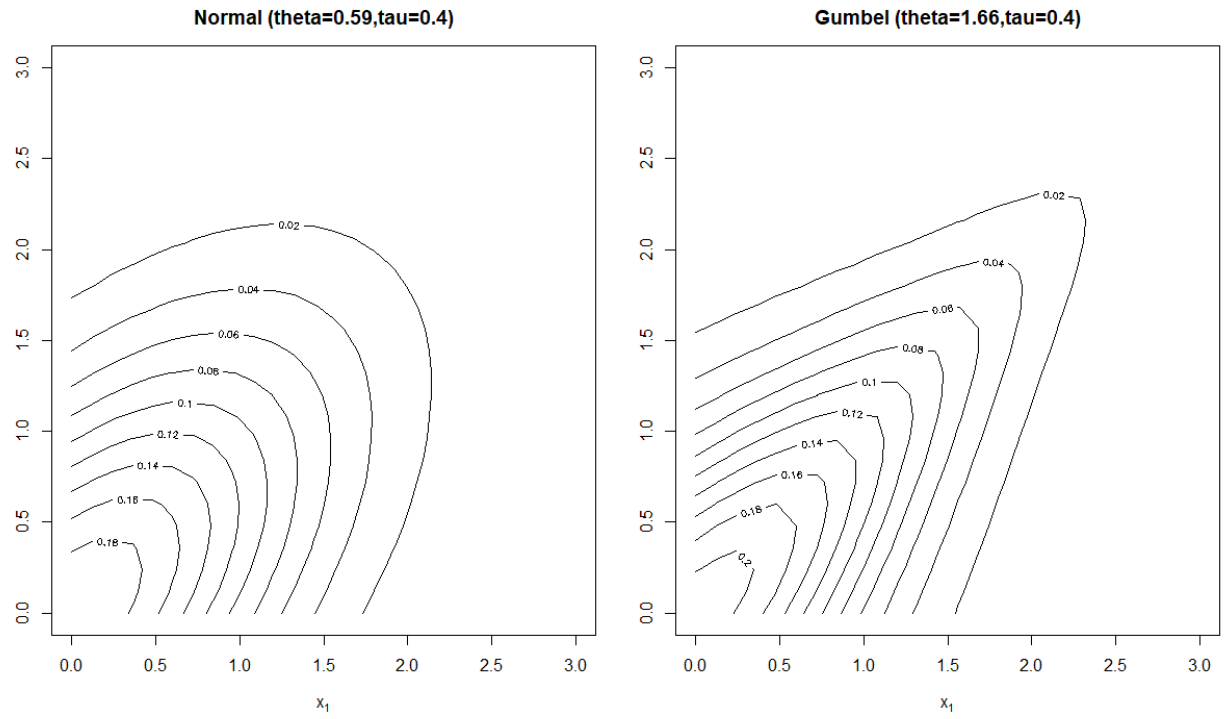

Figure 3.3: The upper tail contours of the joint density function for standard normal marginals with $\tau = 0.4$, connected by Normal copula (right) and the Gumbel copula (left).

Normal copula, when the random variables are negatively correlated ($\rho < 0$) the convergence $\lambda_{upper} \to 0$ is faster than the case where the variables are positively correlated ($\rho > 0$). The Coefficient of Tail Dependence is a measure that shows the strength of dependence in high levels, before asymptotic independence is approached. Ledford and Tawn (1996) suggested that the following approximation for the survival probability holds for large u:

$$P\left(Z_1 > u, Z_2 > u\right) \sim L\left(\frac{1}{1-u}\right)\left(\frac{1}{1-u}\right)^{-1/\eta}, \tag{3.11}$$

where $0 < \eta \leq 1$ is a constant, and L is a slowly varying function, that is $L(r)$ satisfies $L(tr)/L(r) \to 1$ as $r \to \infty$ for all fixed $t > 0$. Then by taking the log of both sides:

$$\log(1 - 2u + C(u,u)) \sim \frac{1}{\eta}\log(1-u).$$

Thus, η can be estimated as:

$$\hat{\eta} = \lim_{u \to 1} \frac{\log(1-u)}{\log(1-2u+C(u,u))} .$$

The cases $\eta < 1$ refer to asymptotic independence. Values of $\eta > 1/2$ and $\eta < 1/2$ indicate positive and negative extremal association (before asymptotic independence is approached), respectively. While $\eta = 1/2$ corresponds to near independence. The case $\eta = 1$ is of special importance as this is implied by the variables being asymptotically dependent.

3.3 Classification for Asymptotic Independence

Ledford and Tawn (1996), also classified the type of extremal behavior of asymptotic independent variables by the rate at which 0 is approached in $\lim_{u \to 1} \frac{1-2u+C(u,u)}{1-u}$. The bounding cases for perfect dependence and exact independence are:

$$P(U > u, V > u) \sim \begin{cases} (1-u) & \text{perfect dependence} \\ (1-u)^2 & \text{exact independence} \end{cases} .$$

They sought a model which smoothly links these two bounding cases presented the model in equation 3.11. The classification for the asymptotically independent case is as follows:

- Class 1: Asymptotic dependence, $\eta = 1$ and $L(r) \not\to 0$.

  This kind of extremal behavior can be modeled by the extreme value family of copulas. As an example for Gumbel copula, $L\left(\frac{1}{1-u}\right) = 2 - 2^{1/\theta}$ is not zero for $\theta > 1$.

- Class 2: Positive association, $\frac{1}{2} < \eta < 1$.

  A possible choice of copula, in this case, could be the bivariate Normal with correlation $0 < \rho < 1$. We have $\eta = (1+\rho)/2$ and $L\left(\frac{1}{1-u}\right) = C_\rho \left(\log\left(\frac{1}{1-u}\right)\right)^{-\rho/(1-\rho)}$ where $C_\rho = (1+\rho)^{3/2}(1-\rho)^{-1/2}(4\pi)^{-\rho/(1+\rho)}$. This result is obtained by approximating $\mathrm{P}(U > u, V > u)$ for a Normal copula.

- Class 3: Near independence, $\eta = \frac{1}{2}$ and $L\left(\frac{1}{1-u}\right) \geq 1$.

  When no extremal dependence exists ($\lambda_{upper} = 0\,,\ \eta = \frac{1}{2}$), the dependence is modeled by Clayton or Morgenstern copula. For Clayton copula which is of the form in equation 3.3, $L\left(\frac{1}{1-u}\right) = 1 + \theta$, for $\theta \geq 0$.

  The Morgenstern (FGM) copula is another option with no extremal dependence behavior for which $\eta = \frac{1}{2}$, and is defined by:

$$C^{FGM}(u, v) = uv[1 + \delta(1 - u)(1 - v)], \quad -1 \leq \delta \leq 1\,,$$

  for which $L\left(\frac{1}{1-u}\right) = 1 + \delta \geq 1$ when $0 \leq \delta \leq 1$.

Chapter 4

Distortions

Distortion functions have been used in the pricing of insurance contracts for a long time. It is used to transform the probabilities of the loss distribution to another probability distribution by re-weighing the original one. The essential idea of distorted insurance pricing is to over-value outstanding potential losses. This approach provides a risk-adjusted premium, which always exceeds the net premium and has a safety margin. In this chapter, the distortion of univariate and multivariate distribution functions are reviewed.

4.1 Probability Distortion

Let X be a random variable describing the non-negative loss distribution of an insurance contract, and F be the pertaining distribution function. Then, $F_X(x) := P(X \leq x)$ is the CDF, and

$$F_X^{-1}(u) := \inf \{x : F_X(x) \geq u\},$$

denotes the generalized inverse or the quantile function. The random variable X can be given by employing the probability integral transform as:

$$X = F_X^{-1}(U),$$

where U is a uniformly distributed random variable on $[0,1]$ on the same probability space as X.

Definition 4.1.1. *(Distortion function) Assume that $g : [0,1] \to [0,1]$. We say g is a distortion function if it satisfies the following properties:*

- *$g(0) = 0$ and $g(1) = 1$;*
- *g is continuous and non-decreasing.*

In the case where g is convex, then we have what we call a convex distortion function. On the other hand, if g is concave, we have what we call concave distortion function.

For instance, the following are the best-known distortion functions that have been proposed in the literature:

- **Proportional Hazard:**

  The convex Proportional Hazard distortion function is (Wang (1995)):

  $$g(u) = u^{1/\gamma}, \quad 0 < \gamma \leq 1.$$

  For the values $\gamma > 1$, g is a concave distortion function.

- **Dual Power:**

  The convex Dual Power distortion function is given by:

  $$g(u) = 1 - (1-u)^{\gamma}, \quad \gamma \leq 1,$$

  where g is concave if $\gamma > 1$.

- **Wang's Transformation:**

  The Wang distortion introduced by Wang (2002) which is widely used in insurance

pricing, is of the form:

$$g(u) = \Phi\left[\Phi^{-1}(u) + \gamma\right], \quad \gamma \leq 0,$$

where Φ is the standard normal distribution function. This transformation is concave for positive values of γ.

For more examples of distortion functions, see Morillas (2005) and Valdez and Xiao (2011). Figure 4.1 illustrates the three examples of convex transformations noted above.

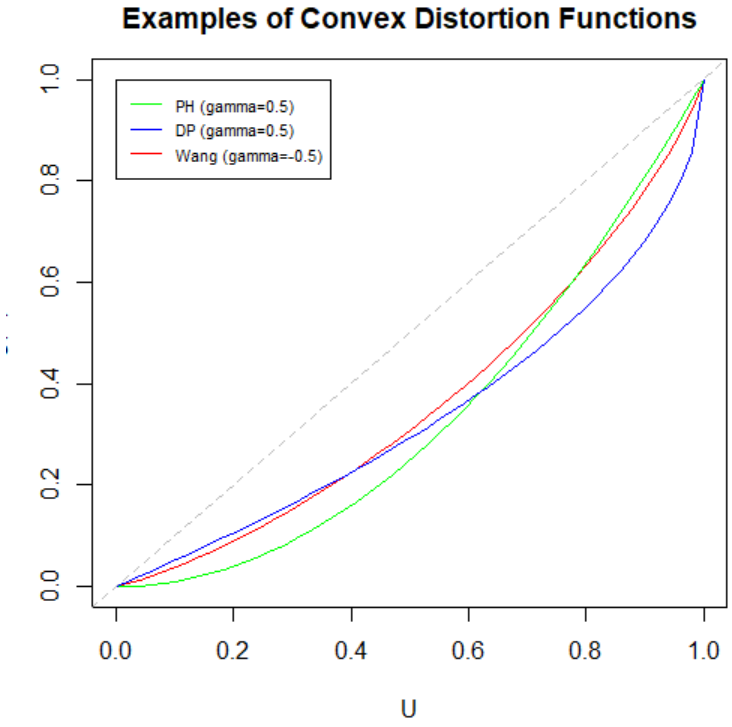

Figure 4.1: The Convex Proportional Hazard (green), Dual Power (blue), and Wang (red) distortion functions.

Definition 4.1.2. *(Distorted probability) Let g be a distortion function and X a random variable with distribution function F_X. Then the transformation of the distribution function:*

$$F_{X^*}(x) = g\left[F_X(x)\right] = g \circ F_X(x),$$

is the distribution function of X^* that leads to a probability distortion of X to X^*.

In insurance pricing and financial risk management, the transformation of the distribution function typically represents a change in the probability measure. Note that for insurance premium purposes, this distorted expectation must be at least equal to the expectation under the original probability measure. Such is the case only when g is convex. A distorted premium principle was introduced by Wang (1996), motivated by Yaari (1987). If g is indeed convex, then direct application of Jensen's inequality leads us to:

$$F_{X^*}(x) \ \leq F_X(x),$$

from which it follows that

$$\mathbb{E}(X^*) = \int_0^\infty \left[1 - g\left[F_X(x)\right]\right] dx \geq \int_0^\infty \left[1 - F_X(x)\right] dx = \mathbb{E}(X)\,.$$

Thus,

$$\mathbb{E}(X^*) - \mathbb{E}(X) \geq 0\,,$$

and this difference is often referred to as the risk premium. Figure 4.2 shows the univariate standard normal density and distribution function, distorted by the three examples of convex distortion functions. It can be seen in the three top plots (CDF's), that for a fixed amount of probability function, the distorted random variable (X^*) always exceeds the random variable itself (X). Note that Wang's transformation preserves the Normal and Log-normal distributions:

- $X \sim \text{Normal}\,(\mu, \sigma^2)$ implies $X^* \sim \text{Normal}\,(\mu - \gamma\sigma, \sigma^2)$
- $X \sim \text{Lognormal}\,(\mu, \sigma^2)$ implies $X^* \sim \text{Lognormal}\,(\mu - \gamma\sigma, \sigma^2)$

In the actuarial literature, it is a more common practice to distort the survival function, $S_X(x) = 1 - F_X(x)$, instead of the distribution function. If we distort F_X with a distortion

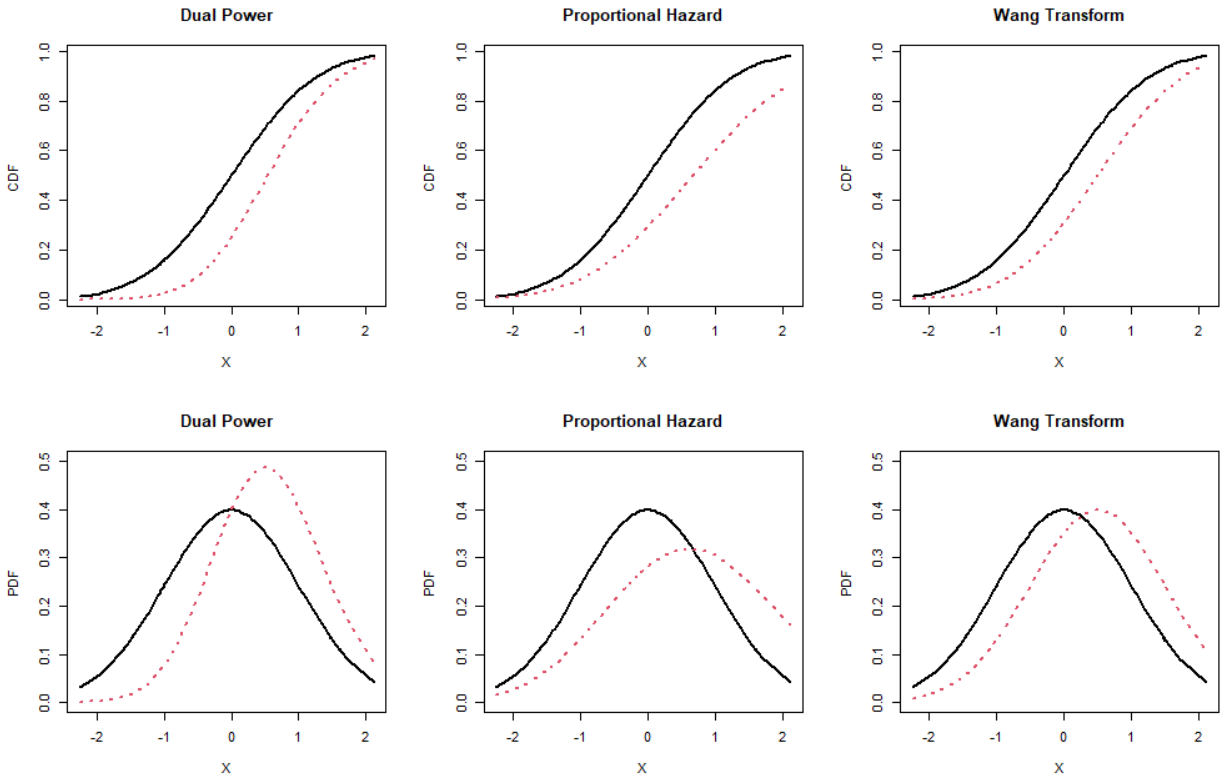

Figure 4.2: The distortion of standard Normal distribution by three different convex distortion functions; the distorted distribution is shown in dashed red line.

function g, this implies that

$$S_{X^*}(x) = 1 - g\left[1 - S_X(x)\right].$$

By defining the function

$$\tilde{g}(t) = 1 - g(1-t),$$

for which it may be called the conjugate of the distortion function g, this demonstrates the equivalence of the distortion between the distribution and survival functions.

4.2 Distorted Copula

The study of multivariate distortions is of general interest because they can be used for generating, in a flexible way, new families of multivariate distributions from a given one. The extension of multivariate distortion can be practically implemented in risk management where one would be interested in portfolios of correlated risks. The first concept that is defined is the externally distorted multivariate distribution, but before that, there are regularity conditions that need to be satisfied.

Definition 4.2.1. *(Absolute monotony) A function $g(t)$ is said to be absolutely monotonic, of order n, on an interval I if the following conditions are satisfied:*

- *g is continuous on I; and*
- *g has non-negative derivatives of orders up to, and including, n, i.e. $g^{(k)}(t) \geq 0$ for all t on the interior of I and for $k = 0, 1, \ldots, n$.*

A sufficient condition for a function to be absolutely monotonic is if g is absolutely monotonic of order n on I, then it must be non-negative, non-decreasing, convex, and continuous everywhere on I (Valdez and Xiao (2011)).

Definition 4.2.2. *(Externally distorted distribution) Let $F \in \mathcal{F}$ be a d-dimensional distribution function. If T is a distortion function that is absolutely monotonic of order d on the interval $[0, 1]$, we define the externally distorted multivariate distribution function by $\widetilde{F}^{ext}(x) = T \circ F(x)$ with $x \in \mathbb{R}_+^d$.*

The external distortion of the upper β-level curve of the joint distribution is the boundary of the set:

$$\widetilde{L}^{ext}(\beta) = \{x \in \mathbb{R}_+^d,\ \widetilde{F}^{ext}(x) \geq \beta\},$$

which follows that:

$$\widetilde{L}^{ext}(\beta) = \{x \in \mathbb{R}_+^d,\ T \circ F(x) \geq \beta\} = \{x \in \mathbb{R}_+^d,\ F(x) \geq T^{-1}(\beta)\}.$$

Note that if T is absolutely monotonic, T^{-1} is a concave function. Then in particular $\alpha := T^{-1}(\beta) > \beta$, for $\beta \in (0,1)$. This means the distortion function T transforms the upper β-level set of the multivariate distribution F into the upper α-level set, with higher risk level α.

The dependence structure of the externally distorted distribution does not necessarily remain the same. Let X and Y be two independent real random variables (i.e., $F(x,y) = F_X(x)F_Y(y)$) and T be a distortion function. Let $(\widetilde{X}, \widetilde{Y})$ be a vector with distribution function $\widetilde{F}^{\text{ext}}(x,y) = T \circ F(x,y)$. Since:

$$T \circ (F_X(x)F_Y(y)) \neq (T \circ F_X(x)) \cdot (T \circ F_Y(y)),$$

$\widetilde{X}$ and $\widetilde{Y}$ are not necessarily independent.

Now, let $X = (X_1, \ldots, X_n)$ be an n-dimensional random vector connected by a copula denoted by $C_X(u_1, \ldots, u_n)$ with marginal probabilities $u_i = F_i(x_i)$, for $i = 1, \ldots, n$. Valdez and Xiao (2011) consider the extension to three different kinds of multivariate distortion:

- **Distortion of the first kind:**

  Let $g_1, \ldots, g_n$ be n distortion functions. Then the transformation of the copula associated with X defined by:

  $$C_X(u_1^*, \ldots, u_n^*) = C_X(g_1(u_1), \ldots, g_n(u_n)), \tag{4.1}$$

  which is a multivariate probability distortion of X to X^*. This type of distortion leads to a simple distortion of the margins that preserves the copula structure. Since the copula is the same structure, this type of distortion does not provide a new copula but creates a new multivariate distribution function. A multivariate extension of the Wang's transformation is constructed by the distortion of the first kind (see Kijima (2006)).

- **Distortion of the second kind:**

  Similar to the distortion of the first kind, this case does not provide a new copula, but rather a specification of a new copula. the distortion of the second kind is derived by:

$$\hat{C}(u_1^*, \ldots, u_n^*) = \hat{C}(g_1(u_1), \ldots, g_n(u_n)), \tag{4.2}$$

  where $\hat{C}$ is another specific copula function. Hence, the dependence structure could change and if $\hat{C} = C_X$, then the distortion of the first kind is obtained.

- **Distortion of the third kind:**

  Let g be an external distortion function with an inverse function g^{-1} that is absolutely monotonic of order n on the interval $[0, 1]$ and $g_1, \ldots, g_n$ be n distortion functions that are called internal distortion functions. Then the transformation of the copula associated with X is defined by:

$$C_g\left(u_1, \ldots, u_n\right) = g^{-1}\left(C_{\mathbf{X}}\left(g_1\left(u_1\right), \ldots, g_n\left(u_n\right)\right)\right), \tag{4.3}$$

  which provides a multivariate probability distortion of X to $\widetilde{X}$. The function C_g that is obtained by this distortion indeed satisfies the properties of a copula function if $g^{-1} = g_1 = \cdots = g_n$ and is then the copula that is associated with the distorted random vector $\widetilde{X}$.

  Unlike the first two kinds, this transformation leads us to a new method of constructing new copulas from a given one, when the internal and external distortion functions are equal. Furthermore, since C_g is a copula, the marginal distributions associated with

the distorted copula are preserved:

$$\begin{aligned} C_g\left(1,\ldots,1,u_k,1,\ldots,1\right) &= g^{-1}\left(C_X\left(g_1(1),\ldots,g_{k-1}(1),g_k\left(u_k\right),g_{k+1}(1),\ldots,g_n(1)\right)\right. \\ &= g^{-1}\left(C_X\left(1,\ldots,1,g_k\left(u_k\right),1,\ldots,1\right)\right. \\ &= g^{-1}\left(g_k\left(u_k\right)\right) = u_k\,. \end{aligned}$$

Hence, if the marginal distributions of the joint distribution $\tilde{F}$ are denoted by $\tilde{F}_1,\ldots\tilde{F}_n$:

$$\tilde{F}_i = g^{-1}\circ g_i^{-1}\circ F_i,\ \text{for}\quad i=1,\ldots,n\,, \tag{4.4}$$

and the bivariate α-level curve defined by Cossette et al. (2013), associated with the distorted distribution in the bivariate setting would be the set of points:

$$\underline{VaR}^*_\alpha(X) = \left\{\left(\hat{F}_2^{-1}(\alpha),x_2\right),x_2\geq F_2^{-1}\left(g_2^{-1}(g(\alpha))\right)\right\},$$

or equivalently,

$$\underline{VaR}^*_\alpha(X) = \left\{\left(x_1,\bar{F}_1^{-1}(\alpha)\right),x_1\geq F_1^{-1}\left(g_1^{-1}(g(\alpha))\right)\right\}.$$

For the special case $g=g_1=g_2$, we have:

$$\underline{VaR}^*_\alpha(X) = \left\{\left(\hat{F}_2^{-1}(\alpha),x_2\right),x_2\geq F_2^{-1}\left(g^{-1}(g(\alpha))\right)\right\} = \left\{\left(\hat{F}_2^{-1}(\alpha),x_2\right),x_2\geq F_2^{-1}\left(\alpha\right)\right\},$$

which has the same limits as the α-level curve of the non-distorted bivariate distribution. In this case, distortion of the third kind preserves the margins.

Example 4.2.1. *Distortion of the comonotonic copula. [Valdez and Xiao (2011)] Let $g=g_1=\cdots=g_n$ be the internal distortion functions and g^{-1} is the external distortion function. Then by the distortion of the third kind, the distorted comonotonic copula is the comonotonic copula itself.*

As noted in section 3.1.2, the comonotonic copula is defined by:

$$C_U\left(u_1,\ldots,u_n\right)=\min\left(u_1,\ldots,u_n\right).$$

This is straightforward to show by noting that because g is increasing, if $u_i=\min\left(u_1,\ldots,u_n\right)$ for some $i=1,\ldots,n$, then $g\left(u_i\right)=\min\left(g\left(u_1\right),\ldots,,\ g\left(u_n\right)\right)$. Therefore, it follows that

$$C_g\left(u_1,\ldots,u_n\right)=g^{-1}\left(C_U\left(g\left(u_1\right),\ldots,g\left(u_n\right)\right)\right)=\min\left(u_1,\ldots,u_n\right).$$

This result is also an immediate consequence of the fact that distortion of the third kind preserves the margins, in the case $g=g_1=\cdots=g_n$.

Example 4.2.2. *A bivariate case. [Di Bernardino and Rullière (2012)] Consider a distortion function T_f of the form:*

$$T_f(u)=\begin{cases}0 & \text{if } u=0\\ \text{logit}^{-1}(f(\text{logit}(u))) & \text{if } 0<u<1\\ 1 & \text{if } u=1\,,\end{cases}$$

where f is the hyperbolic conversion function H:

$$H_{m,h,\rho_1,\rho_2,\eta}(x)=m-h+\left(e^{\rho_1}+e^{\rho_2}\right)\frac{x-m-h}{2}-\left(e^{\rho_1}-e^{\rho_2}\right)\sqrt{\left(\frac{x-m-h}{2}\right)^2+e^{\eta-\frac{\rho_1+\rho_2}{2}}},$$

with $m,h,\rho_1,\rho_2\in\mathbb{R}$, and one smoothing parameter $\eta\in\mathbb{R}$. The inverse of H is also given by:

$$H^{-1}_{m,h,\rho_1,\rho_2,\eta}(x)=H_{m,-h,-\rho_1,-\rho_2,\eta}(x).$$

As internal distortion the hyperbolic conversion functions are considered with $m=0.536, h=0, \rho_1=-\rho_2=0.321, \eta=5$. Let $C(u,v)=uv$ and $F_1(x)=F_2(x)=1-\mathrm{e}^{-x}$. Then $F(x,y)=\left(1-\mathrm{e}^{-x}\right)\left(1-\mathrm{e}^{-y}\right)$. The bivariate distorted distribution function is considered

of the following form:

$$\tilde{F}(x,y) = T \circ C\left(T^{-1}F_1(x), T^{-1}F_2(y)\right),$$

where $T = T_f$, *with* $f = H_{m,h,\rho_1,\rho_2,\eta}$. *Figure 4.3 illustrates the distorted level curves. Note that since the internal distortion functions are equal and the same as the inverse of the external distribution, the margins are preserved.*

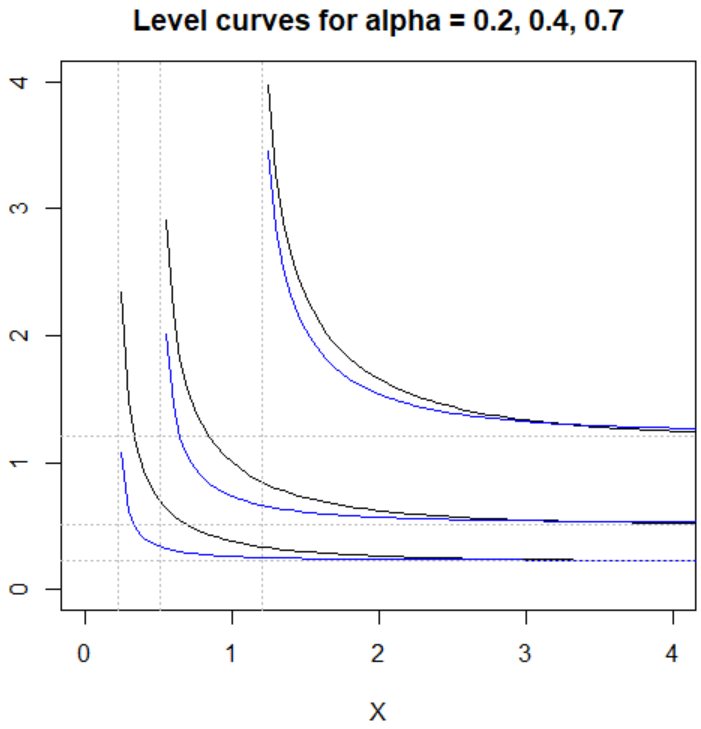

Figure 4.3: Level curves for $\alpha \in \{0.2, 0.4, 0.7\}$ of the distorted distribution (blue) and the non-distorted independent distribution (black).

Chapter 5

Application for Montreal Data

Temperature and precipitation are important climatic variables, especially in the context of climate change. They are among major climatic variables that are used to characterize extreme weather events, which can have profound impacts on ecosystems and society. It has been shown that ignoring the joint distribution of precipitation and temperature lead to the underestimation of occurrence of probabilities for these two concurrent extreme modes. This underestimation can also lead to incorrect conclusions and flawed decisions in terms of the severity of these extreme events. However, accurate analysis of the joint distribution of rainfall and temperature could be challenging due to possible interdependence between them. In this chapter, we aim to model the joint distribution of the historical annual temperature and precipitation maxima using bivariate copulas and introduce a distortion of the historical model for future projections.

5.1 Modeling Univariate Maxima

Considering the historical daily climate data from 1942 to 2012 at Montréal International Airport (YUL), the annual maxima (block maxima) of temperature ($^{\circ}C$) and precipitation (mm) are obtained with 71 blocks of size 365. According to Theorem 2.2, as the block size

gets large, the limiting distribution of the block maxima is in the GEV family of distributions. The block size (365) is considered large in this case and we have enough data points (71) to fit a GEV distribution to the annual maximum temperatures and precipitations. Firstly, the Ljung–Box test could be used to ensure that there is no auto-correlation in each of the time series. The Ljung-Box test is defined as:

$$\begin{cases} H_0 : \text{The data are independently distributed} \\ H_1 : \text{The data are not independently distributed} \end{cases},$$

with the test statistic:

$$Q = n(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{n-k},$$

where n is the sample size, $\hat{\rho}_k$ is the sample autocorrelation at lag k, and h is the number of lags being tested. The test statistic Q is compared with $(1-\alpha)$-quantile of the chi-squared distribution with h degrees of freedom. The independence in a time series is rejected if $Q > \chi^2_{(1-\alpha)}$. The p-values of the Ljung-Box test for temperature and precipitation annual maxima over 71 years are obtained in Table 5.1. P-values being greater than $\alpha = 0.05$ imply that there is not enough statistical evidence to reject the independence among the series, at 0.05 level of significance. Figure 5.1 shows the autocorrelation function (ACF) plots for temperature and precipitation annual maxima. The blue dashed lines in the ACF plots give the values beyond which the auto-correlations are significantly different from zero. It is evident from Figure 5.1 that the sample autocorrelations at different lags are not significant.

Annual Maxima	P-value
Temperature	0.1846
Precipitation	0.2286

Table 5.1: P-values of the Ljung-Box test for the temperature and precipitation annual maxima over 71 years at YUL.

By the maximum likelihood method of estimation, the asymptotic univariate distributions

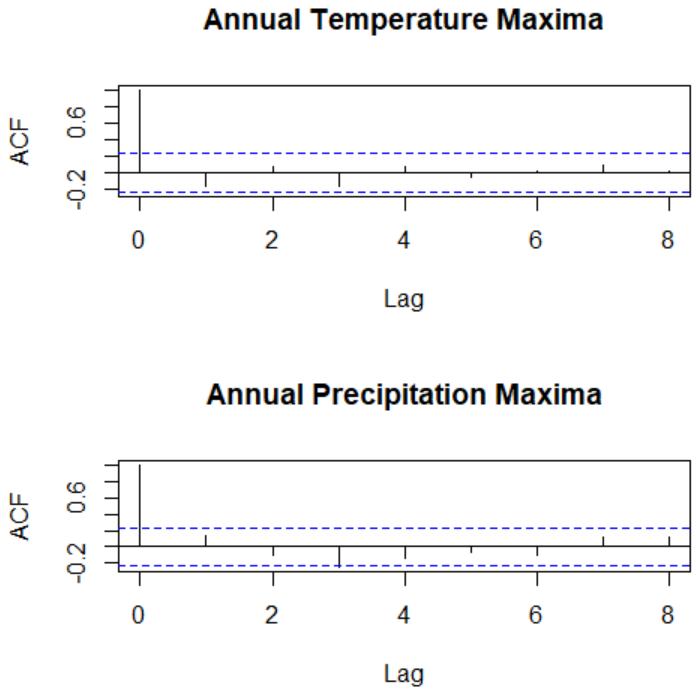

Figure 5.1: Auto-correlation plots for series of maxima.

of annual maxima are obtained in Table 5.2.

The negative shape parameter of the GEV fit associated with the temperature maxima

Annual Maxima	Location $\hat{\mu}$	Scale $\hat{\sigma}$	Shape $\hat{\xi}$	MSE	$\hat{z}_{0.01}$
Temperature	32.24 (0.19)	1.42 (0.13)	-0.14 (0.08)	0.00026	37.01 (0.67)
Precipitation	42.21 (1.41)	10.46 (1.04)	0.02 (0.09)	0.00031	93.02 (10.72)

Table 5.2: Fitted GEV distribution to the annual temperature and precipitation maxima and the corresponding standard error of estimation shown in parentheses.

indicates that it is in the Weibull family of distributions and the maximum temperature does not get higher than a specific upper bound of $\hat{\mu} - \hat{\sigma}/\hat{\xi} = 42.11$. The 0.99 quantile or the estimated 100-year return level ($\hat{z}_{0.01}$) is the value that is expected to be exceeded with probability of approximately 0.01. The Mean Square Error (MSE) of the fits are also calculated in Table 5.2. Small amounts of MSE for both fitted models indicate appropriate fits. As noted in section 2.6, diagnostic plots of the fitted GEV distribution are also useful means to access the performance of the fitted model. Since the shape parameter is negative,

the return level plot in Figure 5.2 is concave. It can be seen in the return level plot that the largest observed temperature value is underestimated by the model, yet the model has a good overall fit and even captures the second largest observation.

Figure 5.3 shows the diagnostic plots of the GEV distribution fitted to the historical

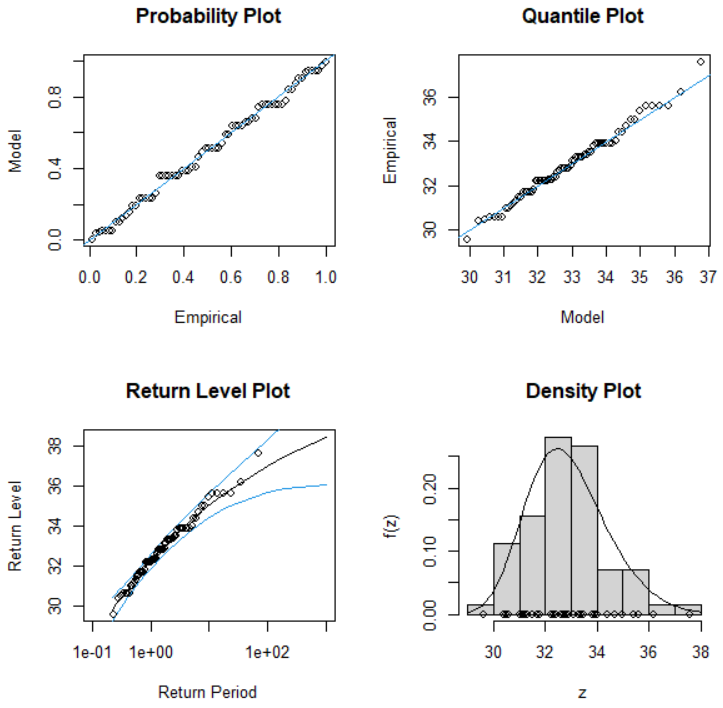

Figure 5.2: Diagnostic plots for the GEV distribution fitted to the temperature maxima.

annual precipitation maxima. The return level plot, in this case, is linear since the shape parameter is pretty close to zero (Gumbel family). The blue range which represents the normal confidence interval of the return level is wider than the one in Figure 5.2 due to the larger variance and error of estimation (see Table 5.2). Overall, both diagnostics for temperature and precipitation GEV fits are acceptable.

Other than the point estimates, confidence interval of the estimated return level is very important in univariate extreme value analysis. Since the return level is a function of the

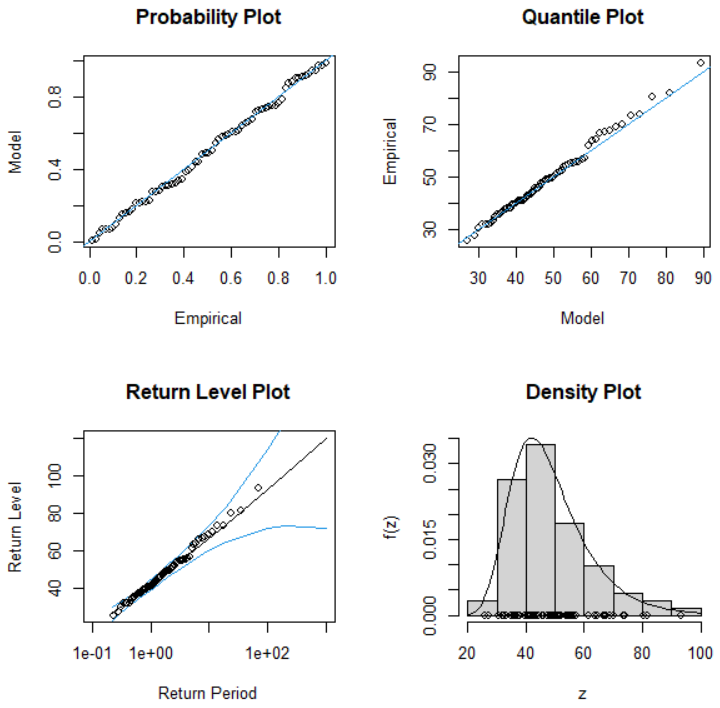

Figure 5.3: Diagnostic plots for the GEV distribution fitted to the precipitation maxima.

parameters (μ, σ, ξ):

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi}\left[1 - \{-\log(1-p)\}^{-\xi}\right] & \text{for } \xi \neq 0 \\ \mu - \sigma \log\{-\log(1-p)\} & \text{for } \xi = 0 \end{cases},$$

for $G(z_p) = 1-p$, the 0.95 level confidence interval is found based on the delta method. Given a function ϕ, the maximum likelihood estimator (MLE) for $\phi(\mu, \sigma, \xi)$ is given by $\hat{\phi}_0(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ and follows the approximate Normal distribution:

$$\hat{\phi}_0 \sim \mathrm{N}\left(\phi_0,\ \mathrm{V}_\phi\right), \quad \text{where} \quad \mathrm{V}_\phi = \nabla_\phi^T\ \mathrm{V}_{(\mu,\sigma,\xi)} \nabla_\phi\,,$$

with $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ being the MLE's of (μ, σ, ξ), ∇_ϕ the gradient of z_p, and $\mathrm{V}_{(\mu,\sigma,\xi)}$ the inverse of information matrix. Hence, the CI based on asymptotic normality is of the form:

$$\phi_0 \in \hat{\phi}_0 \pm 1.96\sqrt{\mathrm{V}_\phi}\,.$$

Another method of estimating confidence intervals is the profile likelihood method which is preferable to the intervals based on the delta method (asymptotic normality). The profile log-likelihood function is the log-likelihood function that is maximized with respect to all other parameters. The profile log-likelihood function associated with the return level can be derived as follows:

$$\ell_p(z_p) = \max_{\sigma,\xi} \ell(z_p, \sigma, \xi),$$

given that μ can be written in terms of z_p:

$$\mu = \begin{cases} z_p + \frac{\sigma}{\xi}\left[1 - \{-\log(1-p)\}^{-\xi}\right] & \text{for } \xi \neq 0 \\ z_p + \sigma \log\{-\log(1-p)\} & \text{for } \xi = 0 \end{cases}.$$

It follows that for large n, under suitable regularity conditions, an approximate $(1-\alpha)100\%$ confidence interval for z_p is:

$$C_\alpha = \left\{ z_p : D_p(z_p) \leq F^{-1}_{\chi^2_1}(\alpha) \right\},$$

where $D_p(z_p)$ is the deviance statistic:

$$D_p(z_p) = 2\{\log \mathrm{ML} - \ell_p(z_p)\} \sim \chi^2_1.$$

The estimated 95% intervals for the 100-year return level of annual temperature and precipitation maxima at YUL are summarized in Table 5.3. The estimated intervals based on

Annual Maxima	Support of GEV	$\hat{z}_{0.01}$	Normality CI	Profile CI
Temperature	$(-\infty, 42.11]$	37.01	(35.70, 38.32)	(36.13, 39.26)
Precipitation	$(-\infty, +\infty)$	93.02	(71.99, 114.04)	(79.29, 129.59)

Table 5.3: Confidence intervals for the 100-year return level of annual temperature and precipitation maxima at YUL based on asymptotic normality and profile likelihood.

the profile likelihood method has proven to be more accurate than the asymptotic normal intervals, because the intervals based on normality are asymptotic and used when the num-

ber of data points is large enough. Also, the normal intervals are symmetrical and may underestimate the upper bound of the range in which $\hat{z}_p$ lies.

5.2 Dependence Analysis

It is generally believed that there exists a significant correlation between the annual precipitation and temperature. However, in reality, it is difficult to model this dependence structure since the correlation between them might change between months. If we consider the historical daily climate data from 1942 to 2012 introduced in section 5.1 and take the monthly maxima throughout 71 years, the monthly maximum temperature and monthly maximum precipitation are shown in Figure 5.4. The overall dependence between maximum temperature and precipitation for each month of the year is measured by Kendall's tau in Table 5.4. It is evident that for some months such as March, May, and July, the two random variables are negatively correlated. Whereas for some other months like January and February the Kendall's tau indicates positive association.

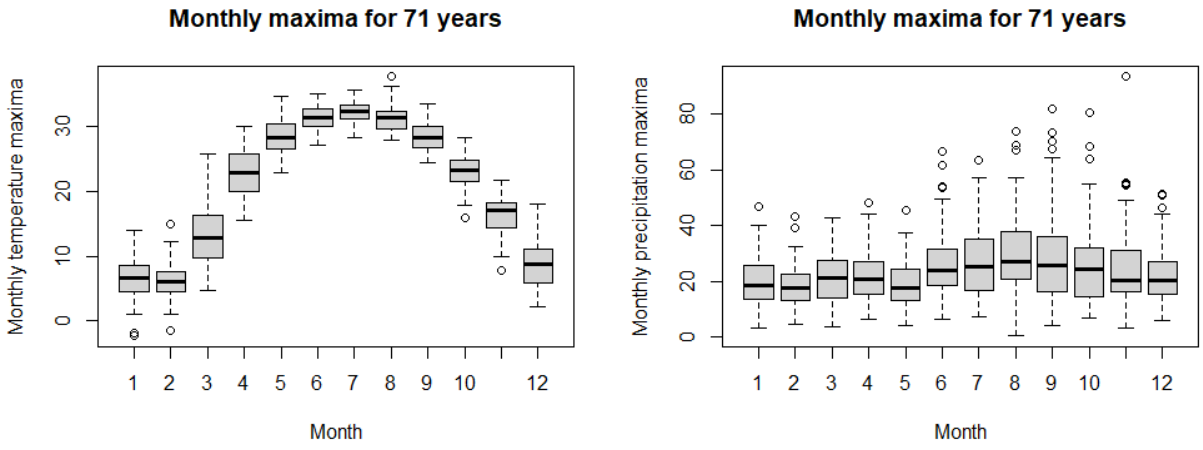

Figure 5.4: Box-plots of monthly maxima over 71 years for temperature (left) and precipitation (right).

The interest of this work is to investigate the dependence behavior between the annual temperature and precipitation maxima by the measures that were introduced in section 3.2.

Monthly Maxima	Jan.	Feb.	Mar.	Apr.	May	Jun.
Kendall's tau	0.11	0.17	-0.19	0.09	-0.20	0.09
Monthly Maxima	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
Kendall's tau	-0.12	0.00	0.01	-0.05	0.08	0.08

Table 5.4: Kendall's tau between maximum temperature and precipitation for each month of the year over 71 years.

Because of the inter-dependence between the maximum temperature and maximum precipitation at YUL that was explained earlier, the amount of dependence between the annual variables is pretty small. In this case, the rank correlation measures Kendall's tau and Spearman's rho are 0.013 and 0.017, respectively. Note that these two measure the overall amount of dependence in a bivariate data set.

To be able to give the choice of copula careful consideration, one should also look into the dependence in the tail. The upper tail dependence parameter (λ_{upper}) which is sometimes referred to as χ in extreme value analysis, represents the amount of asymptotic dependence. The upper plot in Figure 5.5 shows the limit function $lim_{u\to 1}P(V > u \mid U > u)$ for annual maximum temperature and annual maximum precipitation from 1942 to 2012. Since the χ-plot is pretty stable around zero, the variables could be considered asymptotically independent.

Other than the coefficient of tail dependence η, there exists another tail dependence measure $\bar{\chi}$ that represents the strength of dependence among asymptotic independence (Coles et al. (1999)):

$$\bar{\chi} = \lim_{u\to 1} \frac{2\log \mathrm{P}(U > u) - \log \mathrm{P}(U > u, V > u)}{\log \mathrm{P}(U > u, V > u)}.$$

Given that the coefficient of tail dependence η is defined as (see section 3.2.5):

$$\eta = \lim_{u\to 1} \frac{\log(1-u)}{\log(1-2u+C(u,u))},$$

$\bar{\chi}$ can be written in terms of η,

$$\begin{aligned}\bar{\chi}(u) &= \frac{2\log \mathrm{P}(U > u) - \log \mathrm{P}(U > u, V > u)}{\log \mathrm{P}(U > u, V > u)} \\ &= 2\frac{\log \mathrm{P}(U > u)}{\log \mathrm{P}(U > u, V > u)} - 1 \\ &= 2\,\eta(u) - 1.\end{aligned}$$

The lower plot in Figure 5.5 shows the estimated $\bar{\chi}$ for annual temperature and annual precipitation maxima from 1942 to 2012. It is clear that the estimated $\bar{\chi}$ is again stable around zero which implies that the coefficient of tail dependence η should be stabilized around 0.5, since:

$$\bar{\chi} \to 0 \quad \Rightarrow \quad \eta = \frac{\bar{\chi} + 1}{2} \to 0.5,$$

which indicates that there is no evidence of extremal dependence between annual temperature and precipitation maxima.

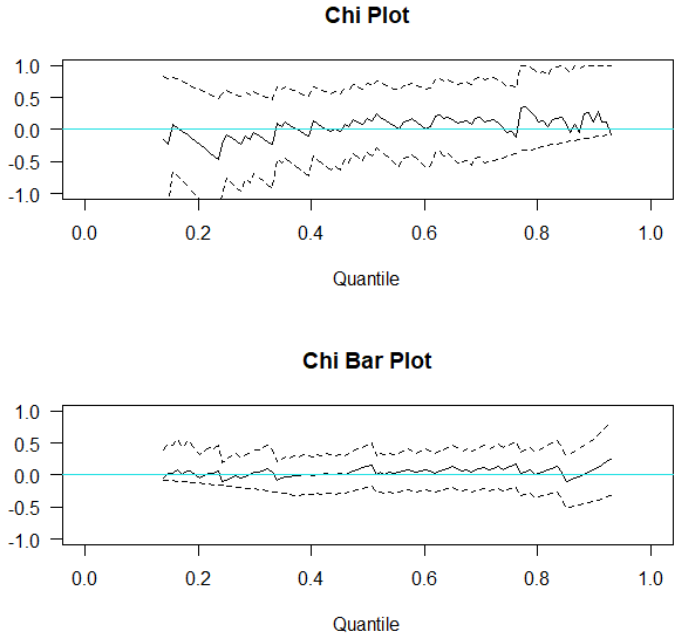

Figure 5.5: The estimated upper tail dependence parameter χ (top) and $\bar{\chi}$ (bottom) between annual temperature and precipitation maxima at YUL, for different levels.

To double-check the extremal independence in our data set, a series of tail independence tests introduced by Falk and Michel (2006) are employed. Let (X, Y) be a random vector that follows in its upper tail a bivariate extreme value distribution. The conditional distribution function of $X + Y$, given that $X + Y > c$, converges to $F(t) = t^2$, $t \in [0, 1]$, if $c \to 0^-$ if X and Y are asymptotically independent:

$$F_c(t) := P(X + Y > tc \mid X + Y > c) \approx t^2, \quad 0 \leq t \leq 1.$$

Otherwise, the limit is $F(t) = t$. This is utilized to test for the tail independence of X, Y via various tests such as Fisher, Kolmogorov-Smirnov, and Chi-Square. The p-values associated with these tests for annual temperature and precipitation maxima are summarized in Table 5.5. P-value being significant (> 0.05) implies that the independence in the tail cannot be rejected.

Tail Independence Test	P-value
Fisher	0.25
Kolmogorov-Smirnov	0.51
Chi-Square	0.39

Table 5.5: Three tests of tail independence for a bivariate extreme value data set.

The information on the overall and extremal dependence behavior in the bivariate data set can lead us to select the appropriate family of copula. When there is no dependence in the upper tail, any member of the extreme value family of copulas would be a flawed choice. Moreover, $\hat{\eta} = 0.5$ implies that the random variables are nearly independent before asymptotic independence is approached. Thus, Clayton copula or Morgenstern copula might be our potential options (see section 3.3). Using the maximum likelihood estimation three different asymptotically independent copulas are fitted to annual temperature and precipitation maxima at YUL. After calculating the parameters of each copula, it is necessary to decide which family is the best representation of the dependence structure between the variables of interest. One of the techniques to select the best copula is based on approaches related to

information criteria, such as Akaike and Bayesian Information Criteria. The Akaike information criterion (AIC) and Bayesian information criterion (BIC) are measures of the relative goodness of fit of a statistical model. the definition of AIC and BIC are as follows:

$$\mathrm{AIC} = 2k - 2\ln(L),$$

$$\mathrm{BIC} = -2\ln(L) + k\ln(N),$$

where k is the number of parameters in the copula, L is the maximized value of the likelihood function for the copula and N is the sample size. Table 5.6 summarizes the estimation of the three different asymptotically independent copula functions by the maximum likelihood method.

Although the dependence behaviour between the precipitation and temperature maxima in our case is pretty close to independence, we prefer not to ignore the very small amount of dependence and consider the Clayton copula with parameter 0.02. The difference between the values of AIC and BIC associated with the three copula fits in Table 5.6 are not very pronounced, but Clayton copula has the smallest AIC and BIC among the three cases.

	Morgenstern	Clayton	Normal
Copula parameter $\hat{\theta}$	0.031	0.020	0.013
Max. log-likelihood	0.004	0.015	0.006
AIC	1.991	1.969	1.987
BIC	4.254	4.232	4.250

Table 5.6: Results of different copula models for annual temperature and precipitation maxima.

5.3 Distorted Model

In this section, a distortion of the historical distribution is proposed as a model for future annual temperature and precipitation maxima. We consider a bivariate projected data set from an ensemble of 24 climate models provided by climate experts for three emissions

scenarios representing different atmospheric concentrations of greenhouse gases (RCP 2.6, RCP 4.5 and RCP 8.5). The data set consists of the two random variables,

$\hat{X}_1$: the annual maximum of daily maximum temperature values, and

$\hat{X}_2$: the annual maximum of daily total precipitations,

that are simulated from the model fitted to a grid (approximately 10 km $\times$ 10 km) that includes the Montreal International Airport (YUL). Since we are interested in the upper tail, the 0.90 quantiles simulated from the projected model are going to be used as projections. Figure 5.6 and Figure 5.7 provide a visualization of the historical and projected data for three different emission scenarios. It is apparent from Figure 5.6 that the annual maximum temperature is projected to get higher based on the emission rate. However, this is not the case for annual maximum precipitation in Figure 5.7. Although the highest annual maximum precipitation projections (113 and 121 millimeters) belong to the pessimistic emission case, different projected scenarios overlap.

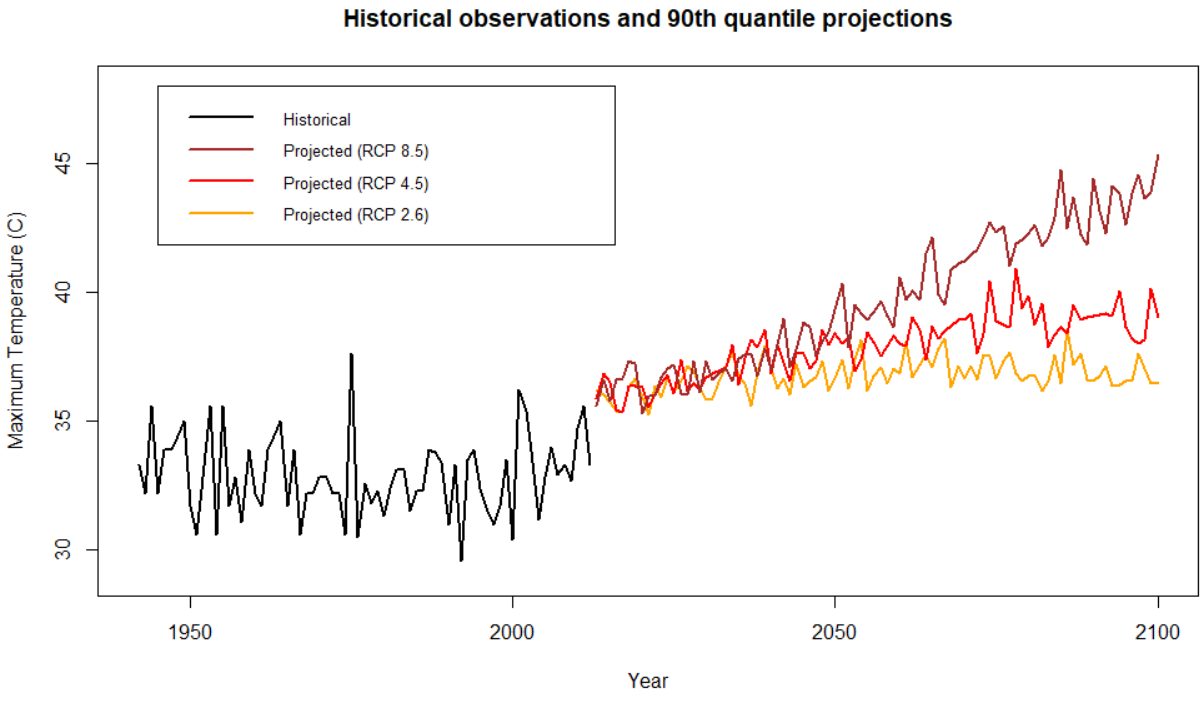

Figure 5.6: Plot of temperature data at Montreal International Airport (YUL).

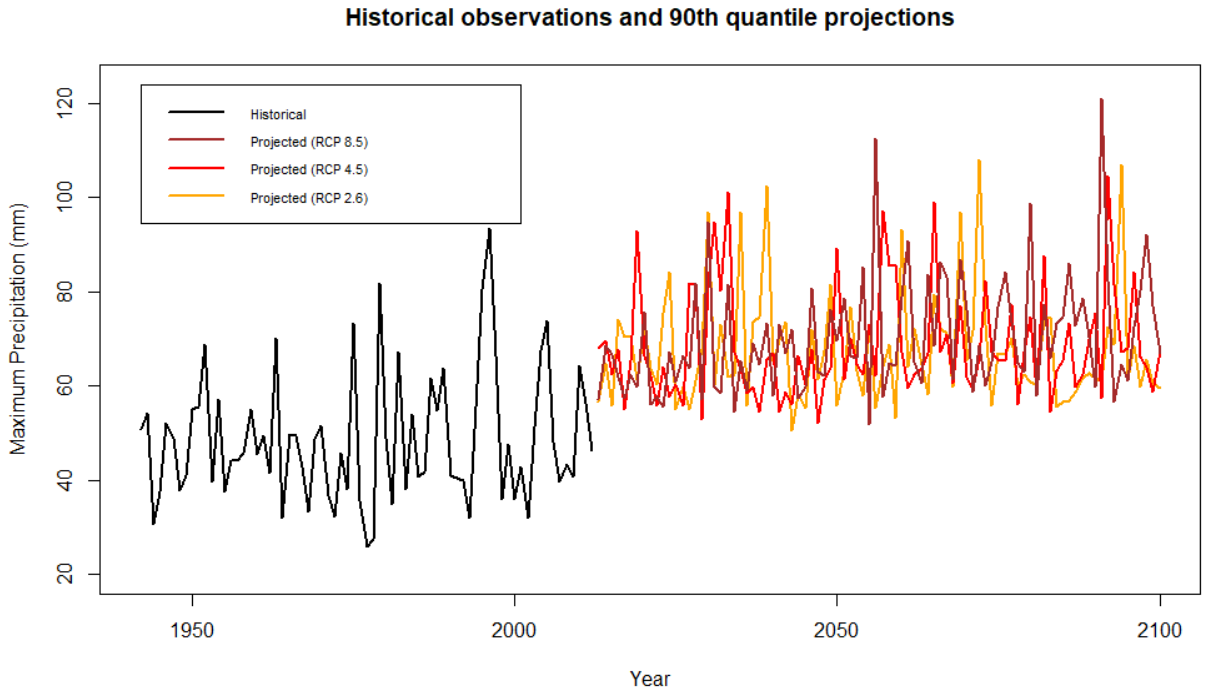

Figure 5.7: Plot of precipitation data at Montreal International Airport (YUL).

In section 5.2, the analysis led us to a bivariate Clayton copula with GEV marginals as the joint distribution of the historical annual temperature (X_1) and precipitation (X_2) maxima:

$$F(x_1, x_2) = [(F_1(x_1))^{-\theta} + F_2(x_2))^{-\theta} - 1]^{-1/\theta}, \quad \theta = 0.02,$$

where $F_1 \sim GEV(\hat{\mu}_1 = 32.24, \hat{\sigma}_1 = 1.42, \hat{\xi}_1 = -0.14)$ and $F_2 \sim GEV(\hat{\mu}_2 = 42.21, \hat{\sigma}_2 = 10.46, \hat{\xi}_2 = 0.02)$ are the asymptotic marginal distributions of maximum temperature and precipitation, respectively.

In order to estimate the bivariate distortion of the historical Clayton distribution, we estimate the internal distortions first. Given the annual maximum temperature projections $\hat{X}_{1,1}, \hat{X}_{1,2}, \ldots, \hat{X}_{1,n}$ and the internal distortion function g_1 with parameter γ_1, the univariate distorted distribution function is,

$$F_1^*(x_1) = g_1 \circ F_1(x_1),$$

with the distorted density of the form,

$$f_1^*(x_1) = \frac{\partial}{\partial x_1} F_1^*(x_1) = g_1'(F_1(x_1)) f_1(x_1).$$

Thus, to estimate the first internal distortion parameter γ_1, the log-likelihood function to be maximized would be:

$$\ell(\gamma_1; \hat{\mu}_1, \hat{\sigma}_1, \hat{\xi}_1, \hat{X}_1) = \sum_{i=1}^{n} \ln\{g_1'(F_1(\hat{X}_{1,i})) f_1(\hat{X}_{1,i})\}, \tag{5.1}$$

where F_1 is the asymptotic distribution of the historical annual maximum temperature and $\hat{X}_{1,i},\ i = 1, \ldots, n$ are the projected values for maximum temperatures.

The same procedure holds for annual maximum precipitation to estimate the other internal distortion parameter by maximizing:

$$\ell(\gamma_2; \hat{\mu}_2, \hat{\sigma}_2, \hat{\xi}_2, \hat{X}_2) = \sum_{i=1}^{n} \ln\{g_2'(F_2(\hat{X}_{2,i})) f_2(\hat{X}_{2,i})\}, \tag{5.2}$$

with F_2 being the asymptotic GEV of the historical annual maximum precipitation and $\hat{X}_{2,i},\ i = 1, \ldots, n$ the projected values for precipitation maxima.

The external distortion parameter γ_3 can also be estimated by maximum likelihood method. By 4.3, the distortion of the historical bivariate distribution would be of the form:

$$F^*_{X_1,X_2}(x_1, x_2) = g_3([(g_1 \circ F_1(x_1))^{-\theta} + g_2 \circ F_2(x_2))^{-\theta} - 1]^{-1/\theta}), \quad \theta = 0.02.$$

Hence, the following log-likelihood function is maximized at $\hat{\gamma}_3$:

$$\ell(\gamma_3; \hat{\mu}_1, \hat{\sigma}_1, \hat{\xi}_1, \hat{\mu}_2, \hat{\sigma}_2, \hat{\xi}_2, \hat{X}_1, \hat{X}_2) = \sum_{i=1}^{n} \ln\{f^*_{X_1,X_2}(x_1, x_2)\}, \tag{5.3}$$

where the corresponding density function is:

$$f^*_{X_1,X_2}(x_1,x_2) = \frac{\partial^2}{\partial F_1(x_1)\partial F_2(x_2)} F^*_{X_1,X_2}(x_1,x_2) f_1(x_1) f_2(x_2).$$

Consider that the distortions g_1, g_2, and g_3 are of the form of Wang distortion function with different parameters γ_1, γ_2, and γ_3, respectively. Table 5.7 summarizes the MLE's for each emission scenario. The marginal GEV parameters are obtained by fitting GEV to the historical data (1942 to 2012) similar to the section 5.1. Given the estimates of GEV parameters and the projected data, the distortion parameters are estimated by maximizing the log-likelihood functions introduced in equations 5.1, 5.2, and 5.3. Each row in Table 5.7 refers to a projected scenario based on RCP, where the distortion parameters are estimated by fitting the distorted copula to the prjected values (2013 to 2100). The MSE, AIC, and BIC associated with the distorted model in each case are calculated to assess the performance of the three different distortions in this case. Tables 5.8 and 5.9 show the same estimations for Dual Power and Proportional Hazard distortion functions.

Figures 5.8, 5.9, and 5.10 show the level curves of the bivariate Clayton copula fitted to historical data and the distorted level curves fitted to the projections for the three different emission scenarios. Dashed lines in these plots represent the univariate VaR of the distorted margins introduced in equation 4.4.

	Temp. GEV Parameters			Prec. GEV Parameters			Distortion Parameters					
	$\hat{\mu}_1$	$\hat{\sigma}_1$	$\hat{\xi}_1$	$\hat{\mu}_2$	$\hat{\sigma}_2$	$\hat{\xi}_2$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	MSE	AIC	BIC
$(1942-2012)-$(RCP 2.6)	32.24	1.42	-0.14	42.21	10.46	0.02	-2.20	-1.28	0.31	0.0218	-477.2	-452.4
$(1942-2012)-$(RCP 4.5)							-2.83	-1.33	0.35	0.0127	-764.3	-739.5
$(1942-2012)-$(RCP 8.5)							-3.33	-1.41	0.58	0.0378	-740.9	-716.1

Table 5.7: Estimated parameters of the bivariate distorted distribution by Wang Distortion based on projected values for the three emission scenarios.

	Temp. GEV Parameters			Prec. GEV Parameters			Distortion Parameters					
	$\hat{\mu}_1$	$\hat{\sigma}_1$	$\hat{\xi}_1$	$\hat{\mu}_2$	$\hat{\sigma}_2$	$\hat{\xi}_2$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	MSE	AIC	BIC
(1942 – 2012) – (RCP 2.6)	32.24	1.42	−0.14	42.21	10.46	0.02	0.23	0.41	0.921	0.0314	−403.6	−378.8
(1942 – 2012) – (RCP 4.5)							0.16	0.40	0.922	0.0221	−689.5	−664.8
(1942 – 2012) – (RCP 8.5)							0.14	0.38	0.881	0.0141	−766.7	−741.9

Table 5.8: Estimated parameters of the bivariate distorted distribution by Dual Power Distortion based on projected values for the three emission scenarios.

	Temp. GEV Parameters			Prec. GEV Parameters			Distortion Parameters					
	$\hat{\mu}_1$	$\hat{\sigma}_1$	$\hat{\xi}_1$	$\hat{\mu}_2$	$\hat{\sigma}_2$	$\hat{\xi}_2$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	MSE	AIC	BIC
(1942 – 2012) – (RCP 2.6)	32.24	1.42	−0.14	42.21	10.46	0.02	0.018	0.14	1.00	0.0049	−707.4	−682.7
(1942 – 2012) – (RCP 4.5)							0.008	0.13	0.99	0.0048	−854.4	−829.6
(1942 – 2012) – (RCP 8.5)							0.001	0.11	3.51	0.0852	−784.3	−759.5

Table 5.9: Estimated parameters of the bivariate distorted distribution by Proportional Hazard Distortion based on projected values for the three emission scenarios.

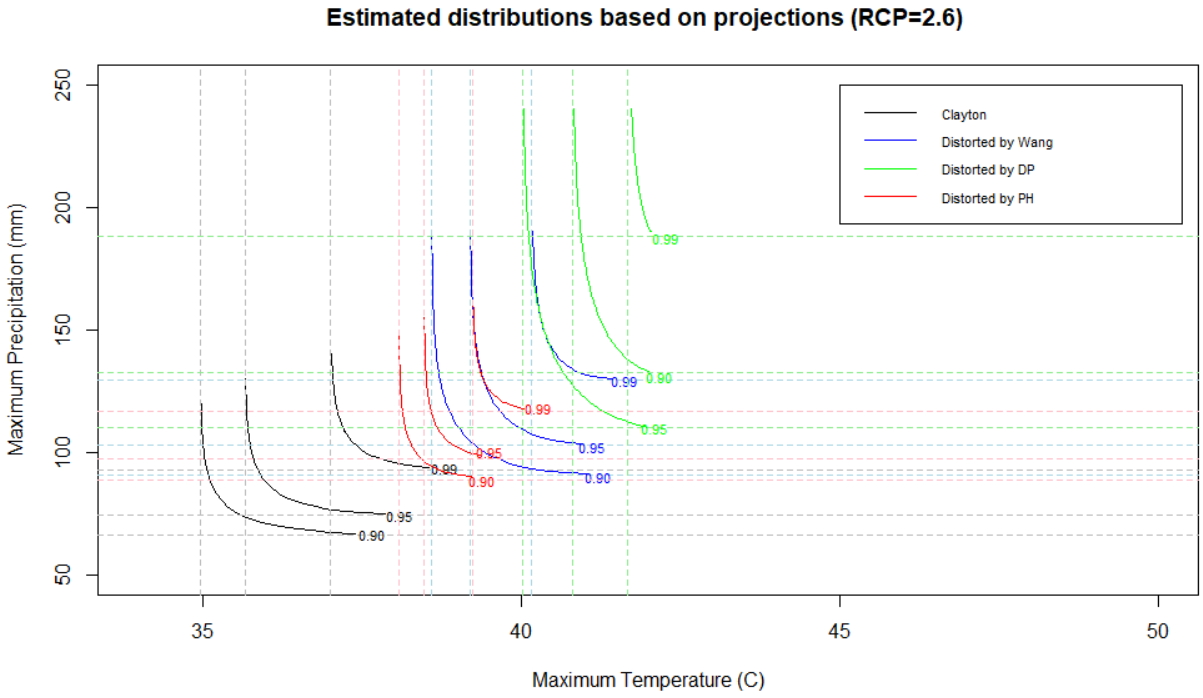

Figure 5.8: Distortion of the level curves based on the optimistic emission scenario (RCP=2.6).

5.4 Discussion

In the previous section, the historical joint distribution of the annual temperature and precipitation maxima at YUL was distorted to propose a model for future projections. For an absolutely monotonic external distortion function, the overall dependence between the

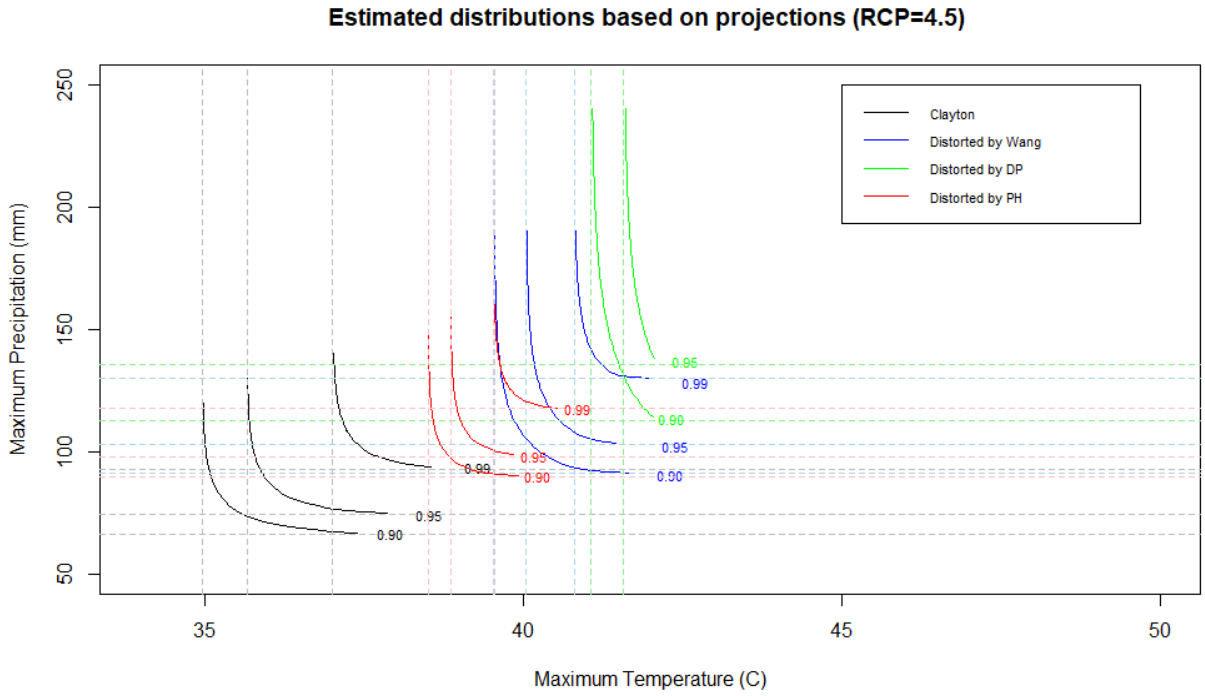

Figure 5.9: Distortion of the level curves based on the realistic emission scenario (RCP=4.5).

random variable is expected to increase because of the relationship between Kendall's tau and the copula function,

$$\tau = \int C(u_1, u_2) dC(u_1, u_2) - 1.$$

If T is an absolutely monotonic distortion function,

$$\tilde{C}^{ext}(u_1, u_2) = T \circ C(u_1, u_2) = \alpha,$$

$$C(u_1, u_2) = T^{-1}(\alpha) > \alpha,$$

which implies a greater probability of u_1 and u_2 level quantiles occurring together and a larger Kendall's tau. The fact that the overall dependence between the projected data set is slightly stronger ($\tau = 0.07$) than the historical one ($\tau = 0.01$), can be explained by absolute monotonocity of the external distortion function. The estimations for three different distortion functions Wang's distortion, Dual Power distortion, and Proportional Hazard dis-

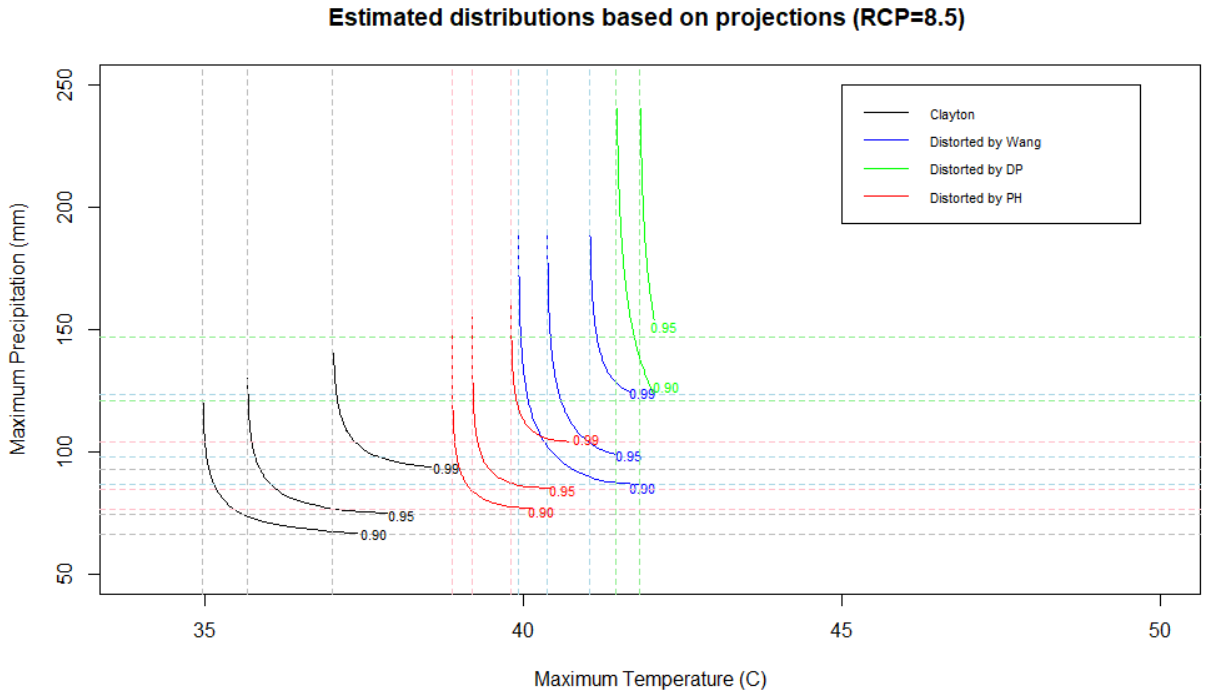

Figure 5.10: Distortion of the level curves based on the pessimistic emission scenario (RCP=8.5).

tortion show that the internal distortion functions are always convex for the three projected scenarios. As the rate of emission (RCP) in projections gets larger, the internal distortion parameters get smaller, meaning that the internal distortion functions become more convex. However, the estimated external distortion functions are not necessarily convex. For example, for Wang's distortion, the external distortion function is maximized at a point for which the function is concave, while for the Dual Power distortion the estimated external distortions are convex.

The Proportional Hazard distortion function has the best overall performance in terms of MSE among the three distortions, except for the RCP= 8.5 scenario. For the pessimistic projected scenario (RCP= 8.5), the bivariate distorted distribution by Dual Power has the smallest MSE among the three cases. However, the MSE of the fit is still significant.

Information criteria are adopted here because they can describe the trade-off between bias and variance in model construction. By comparing the calculated AIC and BIC for each

model in Tables 5.7, 5.8, and 5.9, it evident that the Proportional Hazard distortion has the smallest AIC and BIC for all projected scenarios, even though the MSE is significant for the pessimistic scenario.

In Table 5.3, the estimations show that the support of the asymptotic distribution of annual maximum temperature is limited to 42° Celsius and we can say (with 95% confidence) that the value which is expected to be exceeded with probability 0.01 is somewhere between 36.13 and 39.26. On the other hand, Figure 5.6 shows that the maximum temperature is projected to get as high as 45° Celsius for the pessimistic scenario, which is in contrast with what the Extreme Value Theory suggests based on the historical data. Also, in Figures 5.9 and 5.10, the distorted level curves especially by the Dual Power distortion are limited to an upper bound, due to the fact that the marginal distribution of temperature is GEV with a negative shape parameter ($\xi = -0.14$). Therefore, even the distortion of the bivariate historical model, does not allow for temperatures higher than 42° Celsius.

Chapter 6

Conclusion

In this thesis, we sought to present a bivariate model for future temperature and precipitation maxima by applying distortion to bivariate distributions, which is a method of constructing new multivariate distributions from a given one. We estimated the dependence structure and asymptotic marginal distributions for historical annual temperature and precipitation in Montreal and presented the distortions of the joint distribution in order to relate the distribution of historical observations to the distribution of the projections for the long-term future. Utilizing multivariate distortions allows for a transformed dependence structure which could be practical to model evolving dependence between the two random variables. This method not only provides a bivariate model but also fits a univariate marginal distorted distribution to the projections of each random variable.

However, the fact that the GEV is an asymptotic result and might not be applicable for small sample sizes, is a drawback to the taken approach. The error of estimation associated with the GEV fitted to small samples, say, less than 40 years of data, is significantly larger and the small sample size violates the asymptotic properties of the GEV model. Another potential problem arises when the asymptotic marginal distribution of historical data is in the Weibull family ($\xi < 0$), where the support of GEV is limited to an upper bound. Because of the limited support of marginal GEV in some cases, the bivariate model and its

distortion do not allow for values higher than the upper bound. In our case (see figures 5.9 and 5.10), the limited support of GEV could be either a disadvantage of our model or a proof of an overestimation of the projected annual temperature maxima (out of the support of the historical GEV). One more issue is that the estimated external distortion function might not always be convex, which is in contrast with absolute monotony in Definition 4.2.2. For instance, the external Wang's distortion that is estimated in Table 5.7 is concave and not absolutely monotonic. Whereas the estimated external distortion in Table 5.8 is convex for all projected scenarios and fits the absolute monotony property of the external distortion function.

The method proposed in this thesis could further be utilized for other variables that are projected to increase in the long-term future due to the warming climate such as British Columbia's increasing wildfire claims. In future pursuits, we are interested in implementing other types of distortion functions including hyperbolic composite distortions. Di Bernardino and Rullière (2012) present a level sets-based approach that consists of two estimation algorithm to estimate the external and internal hyperbolic composite distortion functions of a multivariate distorted distribution. These composite distortion functions might make the model more complicated but represent a distorted model without the need for projected values.

Bibliography

AghaKouchak, A., Cheng, L., Mazdiyasni, O., and Farahmand, A. (2014). Global warming and changes in risk of concurrent climate extremes: Insights from the 2014 california drought. *Geophysical Research Letters*, 41(24):8847–8852.

Alvoni, E., Papini, P. L., and Spizzichino, F. (2009). On a class of transformations of copulas and quasi-copulas. *Fuzzy Sets and Systems*, 160(3):334–343.

Bücher, A. and Segers, J. (2017). On the maximum likelihood estimator for the generalized extreme-value distribution. *Extremes*, 20(4):839–872.

Charpentier, A. (2008). Dynamic dependence ordering for archimedean copulas and distorted copulas. *Kybernetika*, 44(6):777–794.

Coles, S., Bawa, J., Trenner, L., and Dorazio, P. (2001). *An introduction to statistical modeling of extreme values*, volume 208. Springer.

Coles, S., Heffernan, J., and Tawn, J. (1999). Dependence measures for extreme value analyses. *Extremes*, 2(4):339–365.

Coles, S. G. and Tawn, J. A. (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(2):377–392.

Cong, R.-G. and Brady, M. (2012). The interdependence between rainfall and temperature: copula analyses. *The Scientific World Journal*, 2012.

Cossette, H., Mailhot, M., Marceau, É., and Mesfioui, M. (2013). Bivariate lower and upper orthant value-at-risk. *European actuarial journal*, 3(2):321–357.

David, H. A. and Nagaraja, H. N. (2004). *Order statistics*. John Wiley & Sons.

Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(3):393–425.

Di Bernardino, E. and Rullière, D. (2012). Distortions of multivariate risk measures: a level-sets based approach.

Dupuis, D. J. and Tawn, J. A. (2001). Effects of mis-specification in bivariate extreme value problems. *Extremes*, 4(4):315–330.

Durante, F., Foschi, R., and Sarkoci, P. (2010). Distorted copulas: constructions and tail dependence. *Communications in Statistics—Theory and Methods*, 39(12):2288–2301.

Durante, F. and Sempi, C. (2005). Copula and semicopula transforms. *International Journal of Mathematics and Mathematical Sciences*, 2005(4):645–655.

Durrleman, V., Nikeghbali, A., and Roncalli, T. (2000). A simple transformation of copulas. *Available at SSRN 1032543*.

Easterling, D. R., Meehl, G. A., Parmesan, C., Changnon, S. A., Karl, T. R., and Mearns, L. O. (2000). Climate extremes: observations, modeling, and impacts. *science*, 289(5487):2068–2074.

Emberchts, P., Klüppelberg, C., and Mikosch, T. (1997). Risk theory. In *Modelling extremal events*, pages 21–57. Springer.

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.

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical proceedings of the Cambridge philosophical society*, volume 24, pages 180–190. Cambridge University Press.

Frees, E. W. and Valdez, E. A. (1998). Understanding relationships using copulas. *North American actuarial journal*, 2(1):1–25.

Galambos, J. (1978). The asymptotic theory of extreme order statistics. Technical report.

Genest, C. and Rivest, L.-P. (1989). A characterization of gumbel's family of extreme value distributions. *Statistics & Probability Letters*, 8(3):207–211.

Genest, C. and Rivest, L.-P. (2001). On the multivariate probability integral transformation. *Statistics & probability letters*, 53(4):391–399.

Gudendorf, G. and Segers, J. (2010). Extreme-value copulas. In *Copula theory and its applications*, pages 127–145. Springer.

Gumbel, E. J. (1958). *Statistics of extremes*. Columbia university press.

Hosking, J. R. and Wallis, J. R. (1987). Parameter and quantile estimation for the generalized pareto distribution. *Technometrics*, 29(3):339–349.

Jenkinson, A. F. (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.

Joe, H. (1997). *Multivariate models and multivariate dependence concepts*. CRC Press.

Kijima, M. (2006). A multivariate extension of equilibrium pricing transforms: The multivariate esscher and wang transforms for pricing financial and insurance risks. *ASTIN Bulletin: The Journal of the IAA*, 36(1):269–283.

Klement, E. P., Mesiar, R., and Pap, E. (2005). Transformations of copulas. *Kybernetika*, 41(4):425–434.

Kunkel, K. E., Andsager, K., and Easterling, D. R. (1999). Long-term trends in extreme precipitation events over the conterminous united states and canada. *Journal of climate*, 12(8):2515–2527.

Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187.

Lin, F., Peng, L., Xie, J., and Yang, J. (2018). Stochastic distortion and its transformed copula. *Insurance: Mathematics and Economics*, 79:148–166.

McMichael, A. J., Woodruff, R. E., and Hales, S. (2006). Climate change and human health: present and future risks. *The Lancet*, 367(9513):859–869.

McNeil, A. J. (1997). Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin: The Journal of the IAA*, 27(1):117–137.

Morillas, P. M. (2005). A method to obtain new copulas from a given one. *Metrika*, 61(2):169–184.

Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media.

Northrop, P. J. and Coleman, C. L. (2014). Improved threshold diagnostic plots for extreme value analyses. *Extremes*, 17(2):289–303.

Pflug, G. C. (2006). On distortion functionals. *Statistics & Risk Modeling*, 24(1):45–60.

Resnick, S. I. (1997). Discussion of the danish data on large fire insurance losses. *ASTIN Bulletin: The Journal of the IAA*, 27(1):139–151.

Sastry, D. and Sinha, R. (2010). A revisit to danish fire loss data. In *Conference Proceedings of the 13th Global Conference of Actuaries (GCA) held in Mumbai in February*.

Sklar, M. (1959). Distribution functions in dimensions and their margins. *Publ. inst. statist. univ. Paris*, 8:229–231.

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1):67–90.

Tawn, J. A. (1988). Bivariate extreme value theory: models and estimation. *Biometrika*, 75(3):397–415.

Valdez, E. A. and Xiao, Y. (2011). On the distortion of a copula and its margins. *Scandinavian Actuarial Journal*, 2011(4):292–317.

Von Mises, R. (1954). La distribution de la plus grande de n valeurs. in (ed.), selected papers (vol. ii, pp. 271-294). providence, ri. *American Mathematical Society*.

Wadsworth, J. L. (2016). Exploiting structure of maximum likelihood estimators for extreme value threshold selection. *Technometrics*, 58(1):116–126.

Wang, S. (1995). Insurance pricing and increased limits ratemaking by proportional hazards transforms. *Insurance: Mathematics and Economics*, 17(1):43–54.

Wang, S. (1996). Premium calculation by transforming the layer premium density. *ASTIN Bulletin: The Journal of the IAA*, 26(1):71–92.

Wang, S., Nelsen, R., and Valdez, E. (2005). Distortion of multivariate distributions: adjustment for uncertainty in aggregating risks. Technical report, Mimeo.

Wang, S. S. (2002). A universal framework for pricing financial and insurance risks. *ASTIN Bulletin: The Journal of the IAA*, 32(2):213–234.

Watts, S. (2016). The gaussian copula and the financial crisis: A recipe for disaster or cooking the books?

Wazneh, H., Arain, M. A., Coulibaly, P., and Gachon, P. (2020). Evaluating the dependence between temperature and precipitation to better estimate the risks of concurrent extreme weather events. *Advances in Meteorology*, 2020.

Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica: Journal of the Econometric Society*, pages 95–115.

Zoglat, A., El Adlouni, S., Ezzahid, E., Amar, A., Okou, C., and Badaoui, F. (2013). Statistical methods to expect extreme values: Application of pot approach to cac40 return index.