Novel methodology for the estimation of extremal bivariate return curves

Callum John Rowlandson Murphy-Barltrop,

B.Sc.(Hons.), M.Res

Submitted for the degree of Doctor of Philosophy at Lancaster University.

April 2023

STOR-i
excellence with impact

Abstract

The aim of this thesis is to develop novel methodology for estimating an extreme risk measure, known as a return curve, for bivariate random vectors. In doing this, we also aim to develop novel techniques for estimating the extremal dependence structure of bivariate random vectors, and then to compare these techniques to existing methodology.

In many practical applications, understanding joint extreme risks from pairs of random variables is crucial for ensuring robust risk analyses and informed decision making. Return curves provide a means of both quantifying and visualising such risks. However, estimation of these curves has not been well studied, particularly in the case when data exhibits asymptotic independence. Furthermore, under the influence of climate change, the joint extremal behaviour for pairs of environmental variables is likely to change; techniques are required to ensure such trends are captured when estimating return curves.

We first propose a range of novel estimation methods for return curves; unlike several existing techniques, our estimates are based on bivariate extreme value models that can capture both key forms of extremal dependence. We devise tools for validating return curve estimates, as well as representing their uncertainty, and compare a selection of curve estimation techniques through simulation studies. Curve estimates are obtained for two metocean data sets, with diagnostics indicating generally good performance.

In the context of extremes, few methods have been proposed for modelling trends

in extremal dependence, even though capturing this feature is important for quantifying joint tail behaviour. Motivated by observed dependence trends in data from the UK Climate Projections, we propose a novel semi-parametric modelling framework for non-stationary, bivariate extremal dependence structures. This framework allows us to capture a wide variety of dependence trends for datasets exhibiting asymptotic independence. We compare our model to an existing technique through a simulation study, obtaining competitive results over a range of extremal dependence structures. When applied to a climate projection dataset, our model is able to capture observed dependence trends and, in combination with models for marginal non-stationarity, can be used to produce estimates of return curves in future climates.

Whilst asymptotic independence is frequently observed in practice, the majority of approaches for bivariate extremes are based on the framework of regular variation. In practice, this is problematic since this framework is is unable to accurately extrapolate into the joint tail for data sets exhibiting this class of extremal dependence. Motivated by this shortcoming, we introduce a range of novel estimators for the so-called 'angular dependence function', a quantity which summarises the dependence structure for asymptotically independent variables. We compare the proposed estimators to existing techniques through a systematic simulation study, obtaining competitive results in many cases. The proposed methodology is also applied to river flow data from the north of England, UK, and used to obtain return curve estimates for different pairs of gauge sites.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisors Jennifer Wadsworth and Emma Eastoe. Thank you both for your never-ending patience, kindness, understanding and guidance over the last four years. There were many moments throughout this process when I strongly doubted myself and my own abilities; you both helped me to challenge these doubts, ultimately giving me the confidence and reassurance to continue in academia. This also made the PhD a lot less daunting and intimidating, and I thoroughly enjoyed the process because of you.

Thank you also to my director of studies Jonathan Tawn; you have been such a fantastic mentor and guide throughout the PhD, and I am constantly in awe of both your research ethic and integrity.

Of course, no PhD is complete without some difficulties along the way. Mine happened to coincide with some periods of challenging personal circumstances where, perhaps unsurprisingly, I struggled with my mental health. It is because of these periods, however, that I feel I have become the more resilient and optimistic person that I am today. I would therefore like to acknowledge everyone in my life who provided invaluable support during these more difficult moments - especially the mental health professionals I have worked with over the last four years.

From my year 6 teacher who predicted that I would one day make a 'great mathematician' to all of the Lancaster University academics who encouraged me to consider a PhD in the first place: many thanks to all of the teachers and lecturers who have

inspired me over the years. I am forever grateful for all of your patience and enthusiasm, and I promise I will try my best to 'pass on the favour' to others down the line.

I am very grateful to the Office for Nuclear Regulation for sponsoring my PhD and providing the initial motivation for the work. They have been an incredibly supportive and encouraging industrial partner, and I have thoroughly enjoyed doing research with such an important application. In particular, I would like to thank External Hazards Inspector Dr Aiden Parkes; it has been an absolute pleasure working with you over the last four years, and I hope we can continue to collaborate for many years to come.

I was fortunate enough to complete my PhD at the EPSRC-funded STOR-i Centre for Doctoral Training; this provided an inclusive and vibrant research community full of many wonderful individuals. Thank you to all of the STOR-i staff members for your hard work in managing the day-to-day running of the centre. I would also like to acknowledge all of the STOR-i students I have had the pleasure to work alongside; the PhD would not have been half as fun without the pub trips, football sessions, lunchtime chats and coffee breaks. A special thanks goes out to everyone involved in the extremes research group; I count myself 'extremely' lucky to be part of such a wonderful and supportive group of individuals, and I look forward to continuing our chats and collaborations for many years to come.

Many thanks to all of my friends, both in and outside of STOR-i; there are too many of you to name individually, but know that I appreciate each and every one of you. Thank you for stopping me from taking life too seriously and for providing such a fantastic support network during my time at Lancaster. I cherish all of the memories that we have made together, and look forward to making new ones in this next chapter of my life.

Last but not least, I would like to thank my family: Mum, Dad, Anya and Cara. I would never have made it to this point in my life without your constant encouragement, support and love. You helped a little boy to achieve one of his dreams and for that, I

am forever grateful. This PhD is as much yours as it is mine.

Declaration

I declare that the work in this thesis has been done by myself and has not been submitted elsewhere for the award of any other degree.

Chapter 3 has been published as Murphy-Barltrop, C. J. R., Wadsworth, J. L., & Eastoe, E. F. (2023). New estimation methods for extremal bivariate return curves. Environmetrics, e2797. https://doi.org/10.1002/env.2797

Chapter 4 has been submitted for publication as Murphy-Barltrop, C. J. R. & Wadsworth, J. L. (2023). Modelling non-stationarity in asymptotically independent extremes.

Chapter 5 has been submitted for publication as Murphy-Barltrop, C. J. R., Wadsworth, J. L., & Eastoe, E. F. (2023). Improving estimation for asymptotically independent bivariate extremes via global estimators for the angular dependence function.

The word count for this thesis is approximately 31716.

Callum John Rowlandson Murphy-Barltrop

Contents

List of Figures

List of Tables

List of Abbreviations

AD	Asymptotic Dependence
ADF	Angular Dependence Function
AI	Asymptotic Independence
BEV	Bivariate Extreme Value
GAM	Generalised Additive Model
GEV	Generalised Extreme Value
GP	Generalised Pareto
GPD	Generalised Pareto Distribution
IB	Integrated Bias
IID	Independent and Identically Distributed
ISE	Integrated Squared Error
IV	Integrated Variance
MISE	Mean Integrated Squared Error
PIT	Probability Integral Transform
PDF	Pickands' Dependence Function
ONR	Office for Nuclear Regulation
RMISE	Root Mean Integrated Squared Error
UKCP18	2018 UK Climate Projections

Chapter 1

Introduction

1.1 Motivation

The modelling of extreme values is important for a wide variety of sectors, including actuarial science, insurance, environmental science, and structural engineering. One area of structural engineering in which this modelling is particularly crucial is the nuclear industry; extreme natural hazards, such as floods, solar storms and earthquakes, can pose high risks to the safety of nuclear sites. Due to the potential safety breach, nuclear facilities are often built to be able to withstand very extreme natural events, and such events must be estimated during the engineering design process.

By definition, historical datasets will contain few extreme observations with which to guide estimation; extreme value theory provides a mathematical framework for modelling such values, allowing one to extrapolate beyond the observed range of data. This allows for a more robust inference from historical extreme events and, for many practical applications, fitted models are often used to obtain estimates of extreme risk measures. Such measures act as a guide for decision-making and risk assessment.

In many cases, it does not suffice to consider the extremes of a single hazard. For example, during the 1999 flood of the Blayais nuclear power plant in France, a com-

bination of extreme winds and sea levels resulted in the plant's flood defences being overwhelmed, causing damage to the plant's power supply and cooling facilities (Mattei et al., 2001). Cooling is a crucial process for keeping nuclear reactors at safe temperatures. This event was ranked as a level 2 on the International Nuclear Event Scale, and while no major accident occurred, it exposed key weaknesses within the safety measures and procedures in place at the time for many nuclear sites (Institute for Nuclear Protection and Safety, 2000). This led to the development of new methods for evaluating flood risk, resulting in a significant cost to upgrade many existing sites.

In contrast to the Blayais incident, the direct impact of the 2011 Fukushima nuclear disaster in Japan, ranked as a level 7 on the International Nuclear Event Scale, was huge. Damage to the plant's cooling facilities led to the meltdown of three reactors, resulting in $187 billion in damages and 154,000 people being evacuated. As in Blayais, this event was also triggered by the combination of two extreme events: in this case, an earthquake and a tsunami (International Atomic Energy Agency, 2015). Prior to the accident, regulatory and safety procedures for the Fukushima-Daiichi plant had only considered the separate risks from these two hazards. Owing to this shortcoming, a government report released following the accident stated that it was a "profoundly man-made disaster – that could and should have been foreseen and prevented" (National Diet of Japan, 2012). The Fukushima disaster also led to wide-spread changes in nuclear safety assessments and procedures, both in Japan and internationally. In the most drastic cases, certain countries, such as Germany and Belgium, opted to phase-out all nuclear energy operations in response to the events of Fukushima.

As a consequence of these two incidents, international nuclear regulatory practices now require that the impact from combined extreme events be taken into consideration when designing nuclear facilities. A recent report from the Office for Nuclear Regulation (ONR), the independent safety regulator for the UK's nuclear industry, states that licensees "should take into account combinations of external hazards that could

reasonably be expected to occur at a given site" (Office for Nuclear Regulation, 2018). Moreover, the ONR requires sites to be built such that they can withstand external hazards at a 10^{-4} annual exceedance probability (Office for Nuclear Regulation, 2014). This corresponds to events with a 10,000 year return period (see Chapter 2.1.3). Similar standards are in place for other nuclear regulatory bodies, such as the US Nuclear Regulatory Commission (NRC) and the French Nuclear Safety Authority (ASN).

These regulatory requirements raise two important questions. Firstly, how does one define an extreme event for a combination of two variables? Unlike in the univariate setting, bivariate vectors have no natural ordering, and the most extreme values of two variables need not occur simultaneously. Moreover, in many practical settings there exist wide ranges of bivariate events which can be high impact, implying the definition of a joint extreme event may not be unique; see, for instance, Ross et al. (2020).

Assuming one has a means of defining a joint extreme event, an important question remains: how does one estimate the risk from such events? This typically requires an assessment of how the extremes of one variable depend on the extremes of another. Formally, this is known as the extremal dependence structure, and many frameworks exist for its estimation. However, this assessment is further complicated by the fact that many variables considered by nuclear regulators, such as temperature, relative humidity and wind speed (Office for Nuclear Regulation, 2021), exhibit complex distributional trends over time. Such trends occur due to various factors, including anthropogenic climate change. As such, both the definition of a joint extreme event and the underlying extremal dependence structure may change over time.

In this thesis, we attempt to answer both of these questions. We introduce a risk measure, known as a return curve, which provides a summary of joint extreme events for two variables. These curves are based on the concept of the joint survivor region, and are used to evaluate extreme responses in a wide range of sectors; see Chapters 2.4 and 3 for further details. We provide novel techniques for estimating this risk measure

for a wide range of data structures, including structures exhibiting complex dependence trends. Furthermore, we provide uncertainty quantification and diagnostic tools for the return curve, allowing for a more robust and detailed practical analysis.

Alongside return curves, we also introduce novel methods for estimating the extremal dependence structures of bivariate random vectors. These techniques build upon the current theory, allowing for a more accurate assessment of joint tail properties. The methods we propose can also be used to model trends in the extremal dependence structure, and hence our approach can be used to evaluate joint extremes in future climates.

1.2 Overview of thesis

As stated, our objective is to answer both of the following questions: for a bivariate random vector, what constitutes a joint extreme event, and how does one make inference on such events? There exists a rich literature on both topics, and our techniques build upon this work while simultaneously addressing some of the limitations of existing approaches.

To begin, Chapter 2 gives a brief overview of extreme value modelling techniques. We introduce approaches for modelling univariate and bivariate extremes, including the cases when the underlying random variables or vectors exhibit non-stationarity. We also introduce the notion of a return curve, and discuss its utility and interpretation.

In Chapter 3, we introduce novel return curve estimation techniques. These techniques are shown to outperform existing approaches and provide accurate curve estimates for both simulated and observed datasets. We also introduce a means for quantifying uncertainty for this risk measure, which is important in a practical setting. Furthermore, we provide a diagnostic tool for curve estimates, allowing one to evaluate goodness of fit for observed datasets.

In Chapter 4, we introduce a novel modelling framework for capturing trends in the extremal dependence structure under asymptotic independence. The majority of existing approaches for modelling non-stationary extremal dependence assume asymptotic dependence, even though asymptotic independence is often shown to exist in practice. In practice, assuming the incorrect form of extremal dependence can lead to an inaccurate extrapolation in the joint upper tail. Our work overcomes these limitations, while simultaneously outperforming an existing approach. Furthermore, our framework can be used to estimate return curves in a non-stationary setting, allowing one to assess the impact of climate change (for example) on joint extreme events.

In Chapter 5, we introduce techniques for estimating the extremal dependence structure of bivariate random vectors. In particular, we consider the model of Wadsworth and Tawn (2013) and provide novel methods for estimating a key quantity known as the 'angular dependence function'. Until recently, this quantity has only been estimated using a pointwise estimator, resulting in rough functional estimates. Our proposed techniques, which give smooth functional estimates, outperform the pointwise estimator, resulting in better estimates of extremal dependence structures. In a practical setting, this helps improve the accuracy of return curve estimates.

Finally, in Chapter 6, we briefly summarise the contributions of this thesis, and discuss avenues for further work. In particular, we consider how the proposed techniques could be extended beyond the bivariate setting, and how the underlying inferential procedures could be improved. We also highlight the impact of this work within the context of nuclear regulation, and discuss how the proposed methodology could be utilised to improve regulatory practices.

Chapter 2

Literature review

2.1 Univariate extreme value theory

In the univariate setting, there are two widely used methodologies for modelling extreme values: the block maxima approach and the peaks over threshold approach. We briefly introduce both approaches; a detailed discussion of these modelling frameworks can be found in Coles (2001). We restrict attention to the upper tail, noting that both approaches can be easily adapted for the lower tail.

2.1.1 Block maxima approach

Let $X_1, X_2 \ldots, X_n$ denote a sequence of $n \in \mathbb{N}$ independent and identically distributed (IID) random variables with distribution function F. We consider the random variable

$$M_n := \max\{X_1, \ldots, X_n\}.$$

Interest lies in understanding how M_n is distributed, since this gives some insight into the extremal behaviour of F. In theory, it is straightforward to derive the distribution

of M_n since

$$\begin{aligned}\Pr(M_n \leq x) &= \Pr(X_1 \leq x, \ldots, X_n \leq x) \\ &= \Pr(X_1 \leq x) \times \ldots \times \Pr(X_n \leq x) \\ &= F^n(x).\end{aligned}$$

However, for any $x < x_u$, where x_u denotes the upper end-point of F, we have that $F(x) < 1$ and $F^n(x) \to 0$ as $n \to \infty$, implying the theoretical distribution of M_n is degenerate.

This issue can be overcome by renormalising the variable M_n. The extremal types theorem of Leadbetter et al. (1983) states that if there exist series $a_n > 0$ and b_n such that

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

as $n \to \infty$, where G is non-degenerate, then G is a generalised extreme value (GEV) distribution taking the form

$$G(z) = \exp\left\{-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}, \tag{2.1.1}$$

with $x_+ := \max\{x, 0\}$ and $(\mu, \sigma, \xi) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}$. In this case, we say that F lies in the domain of attraction of G. Given a GEV distributed variable Z, we write $Z \sim \text{GEV}(\mu, \sigma, \xi)$.

We refer to μ, σ, and ξ as the location, scale and shape parameters, respectively. Moreover, the GEV distribution has three distinct subclasses: Fréchet, Gumbel, and negative Weibull, corresponding to $\xi > 0$, $\xi = 0$, and $\xi < 0$, respectively. These subclasses exhibit different and distinct forms of tail behaviour. Note that the Gumbel subclass is interpreted as the limit of equation (2.1.1) as $\xi \to 0$.

In practice, the normalising series a_n and b_n, which depend on F, are unknown.

However, we can assume equation (2.1.1) holds approximately for large n, implying

$$\Pr(M_n \leq z) \approx G\left(\frac{z - b_n}{a_n}\right) = G^*(z),$$

where G^* denotes a GEV(μ^*, σ^*, ξ) distribution, for which the location and scale have been shifted in comparison to the limiting values from equation (2.1.1). This avoids the need to estimate the normalising series, resulting in the block maxima statistical model. Data are first split into large blocks of equal length and the maxima of each block is recorded. A GEV distribution is then fitted to these block maxima using standard statistical techniques, and the fitted model can be used to approximate the upper tail of F. This approach is illustrated in Figure 2.1.1.

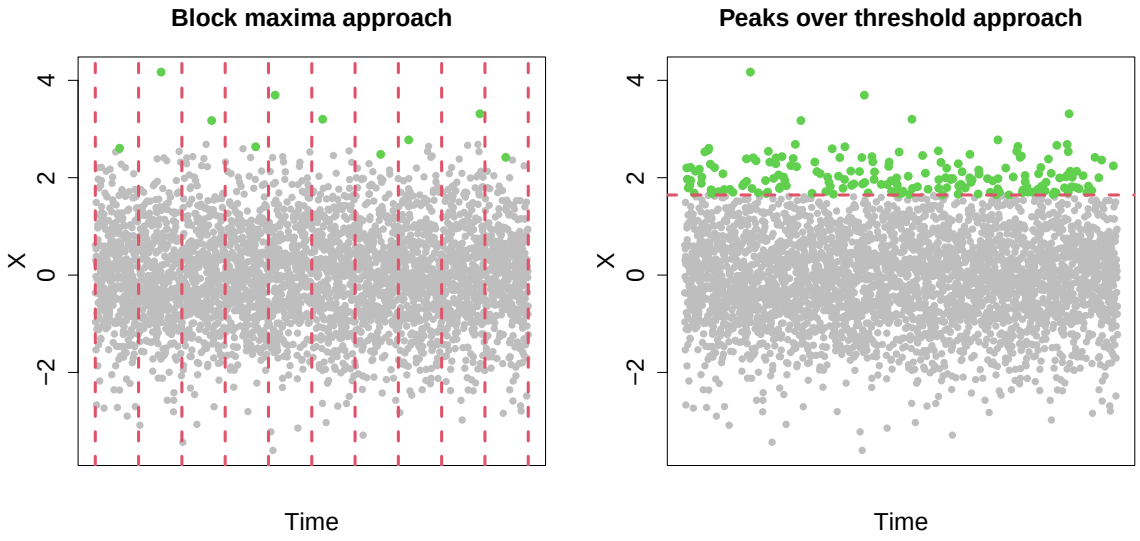

Figure 2.1.1: Block maxima (left) and peaks over threshold (right) approaches. The green points represent the observations used in each modelling approach, while the red dotted lines represent the block boundaries and threshold for the block maxima and peaks over threshold approaches, respectively.

2.1.2 Peaks over threshold approach

One drawback of the block maxima approach is that only the maxima of each block are used in inference. This can result in some extreme observations being disregarded, leading to an incomplete picture of the upper tail.

A more widely used alternative is the peaks over threshold approach. Given $X \sim F$, with F in the domain of attraction of some GEV distribution G, the Pickands-Balkema-de Haan theorem (Balkema and de Haan, 1974; Pickands, 1975) implies that under certain regularity conditions, there exists a normalising function $c(u) > 0$ such that for all $z > 0$,

$$\Pr\left(\frac{X-u}{c(u)} \leq z \;\middle|\; X > u\right) \to H(z) := 1 - \left[1 + \frac{\xi z}{\tilde{\sigma}}\right]_+^{-1/\xi}, \tag{2.1.2}$$

as $u \to x_u$, with $(\tilde{\sigma}, \xi) \in \mathbb{R}^+ \times \mathbb{R}$. Here, H is the cumulative distribution function of a generalised Pareto (GP) distribution, with scale and shape parameters $\tilde{\sigma}$ and ξ, respectively. Note that ξ is equal to the shape parameter of G.

Like the normalising series of the block maxima approach, the normalising function $c(u)$ depends on F and hence is unknown in practice. However, assuming equation (2.1.2) approximately holds for some large threshold u,

$$\Pr\left(X - u \leq z \;\middle|\; X > u\right) \approx H\left(\frac{z}{c(u)}\right) = H^*(z),$$

where H^* denotes a GP distribution with a shifted scale parameter $\tilde{\sigma}^*$. This implies we can model the exceedances of u using a GP distribution - hence the term peaks over threshold. In this case, we write $X - u \mid X > u \sim \text{GP}(\tilde{\sigma}^*, \xi)$. This approach is illustrated in Figure 2.1.1, alongside the block maxima approach. One can clearly observe the difference in the observations incorporated by either approach.

To apply either of the univariate modelling approaches, one must first select either the block size or the threshold. In both cases, this selection represents a bias-variance trade off; selecting a block size (threshold) too large will result in high variability, while

too small a block size (threshold) will give biased results. A more detailed discussion of this trade off can be found in Coles (2001). In the case of the peaks over threshold approach, a wide range of techniques have been proposed for threshold selection; see Wadsworth (2016), Northrop et al. (2017) and Varty et al. (2021), for instance.

2.1.3 Return levels

With a GEV or GP distribution fitted to extreme observations, interest lies in understanding the potential risk from extreme events. Return levels are the most widely used metric for quantifying this risk. Given a variable X and a small probability p, the return level x_p is defined as the $(1-p)$-th quantile of X, i.e., $\Pr(X > x_p) = p$. Estimates of x_p can be obtained by inverting either equation (2.1.1) or (2.1.2) depending on whether X represents block maxima or threshold exceedances. We would expect X to exceed the corresponding return level x_p once, on average, every $1/p$ observations, with the quantity $1/p$ termed the return period.

For example, suppose X denotes annual maxima of some time series (e.g., temperature) and $p = 0.1$. In this case, the x_p return level is the temperature we would expect X to exceed once every 10 years on average. In many applications, return levels estimates are obtained for return periods that correspond to some number of years; for example, nuclear regulators often specify a $10,000$ year time period for design standards.

For both univariate approaches, uncertainty for return levels can be easily quantified using standard statistical methods. Combined, these techniques allow us to quantify extremal risks at different return periods, along with the corresponding uncertainties, allowing for a robust risk analysis.

2.2 Bivariate extreme value theory

2.2.1 Defining extreme events

Unlike in the univariate setting, there is no standard practice for defining extreme events when considering two variables. This is in part due to the lack of a natural ordering for bivariate vectors, i.e., given $(x_1, y_1), (x_2, y_2) \in \mathbb{R}^2$, the statement $(x_1, y_1) < (x_2, y_2)$ lacks meaning. Barnett (1976) suggests several possible methods for defining bivariate extremes; however, the suitability of each method is likely to be context dependent. The earliest approaches directly extended the block maxima approach to the bivariate setting, resulting in componentwise maxima. However, as shown in Chapter 2.2.3, componentwise maxima need not correspond to actual observations, and hence such approaches have limited applicability in practice. Alternative approaches include constructing a one-dimensional structure variable from the bivariate vector (Coles and Tawn, 1994), constructing a convex hull from the data (Chatterjee and Chatterjee, 1990), and conditioning on at least one variable being extreme (Rootzén and Tajvidi, 2006), with the latter approach resulting in a bivariate extension to the GP distribution.

2.2.2 Copula modelling

Unlike in the univariate setting, bivariate modelling requires us to consider the dependence between variables. In the context of extremes, particular interest lies in understanding how the extremes of one variable depend on the extremes of another. However, this dependence can be obscured for variables with different marginal behaviours. Bivariate extreme value modelling therefore makes use of copulas, which allow for the separation of marginal and dependence modelling.

Given a continuous random vector (X, Y) with joint distribution function F and $X \sim F_X$, $Y \sim F_Y$, Sklar's theorem (Sklar, 1959) implies that there exists a unique

distribution function $C : [0,1]^2 \to [0,1]$ such that

$$F(x,y) = C(F_X(x), F_Y(Y)).$$

One can observe that C is simply the joint distribution function on standard uniform margins since, by the probability integral transform (PIT), we have that the variables $F_X(X)$ and $F_Y(Y)$ are Uniform$[0,1]$ distributed.

A consequence of Sklar's theorem is that the joint distribution function of a random vector (X,Y) is invariant under marginal transformation, implying the choice of margins is somewhat arbitrary for bivariate modelling procedures. This result is often applied in bivariate extremes, where certain choices of margins are favourable for quantifying the dependence between extremes. Two common choices of margins are standard Fréchet and standard exponential; the corresponding joint distribution functions for these marginal choices are

$$F(x,y) = C_F\left\{-\frac{1}{\log F_X(x)}, -\frac{1}{\log F_Y(y)}\right\},$$
$$F(x,y) = C_E\left\{-\log(1-F_X(x)), -\log(1-F_Y(y))\right\},$$

where C_F and C_E denote the joint distribution functions on standard Fréchet and exponential margins, respectively. Figure 2.2.1 illustrates data simulated from a bivariate Gaussian copula with $\rho = 0.5$ on standard Gaussian, uniform, Fréchet, and exponential margins. One can observe how each marginal distribution exaggerates different aspects of the underlying dependence structure.

2.2.3 Componentwise maxima

Let $(X_1,Y_1),(X_2,Y_2),\ldots,(X_n,Y_n)$ denote $n \in \mathbb{N}$ IID copies of a random vector (X,Y) with joint distribution function F. In analogy to the block maxima approach, we

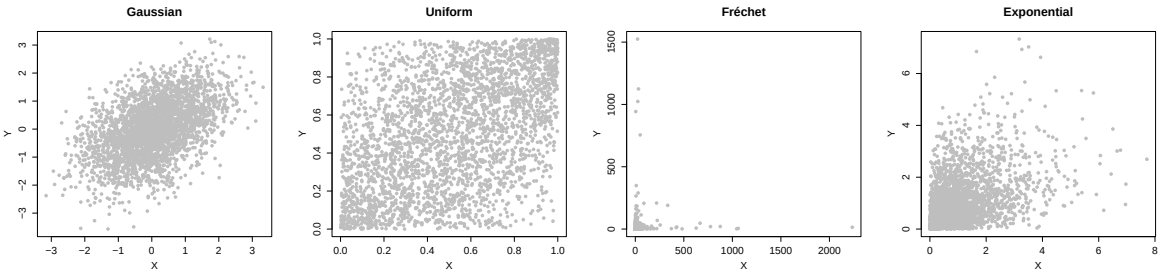

Figure 2.2.1: Data from a bivariate Gaussian copula, with $\rho = 0.5$ and $n = 3,000$, on four different margins. From left to right: standard Gaussian, uniform, exponential and Fréchet marginal distributions.

consider the distribution of the componentwise maxima

$$\mathbf{M}_n := (\max_{i \leq n} X_i, \max_{i \leq n} Y_i).$$

This structure variable is illustrated in Figure 2.2.2. Understanding how $\mathbf{M}_n$ is distributed provides insight into the joint extremal behaviour of (X, Y). Suppose there exist series of vectors $\mathbf{a}_n = (a_{n,1}, a_{n,2}) \in \mathbb{R}^2_+$ and $\mathbf{b}_n = (b_{n,1}, b_{n,2}) \in \mathbb{R}^2$ such that

$$\Pr(\mathbf{a}_n^{-1}(\mathbf{M}_n - \mathbf{b}_n) \leq \mathbf{z}) \to G(\mathbf{z}), \tag{2.2.1}$$

as $n \to \infty$, where G is a non-degenerate distribution function and vector algebra is interpreted as being componentwise. Then G is a bivariate extreme value (BEV) distribution, and we say that F is in the domain of attraction of G. In contrast to the univariate setting, there is no finite dimensional closed parametric form for G.

The required normalising series $\mathbf{a}_n$ and $\mathbf{b}_n$ depends on the marginal distributions of (X, Y), and it is standard practice to assume common margins for both X and Y. In particular, letting X and Y be standard Fréchet, it suffices to set $\mathbf{a}_n = (n, n)$ and $\mathbf{b}_n = (0, 0)$ to achieve the convergence of equation (2.2.1). In this case, G is given by

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

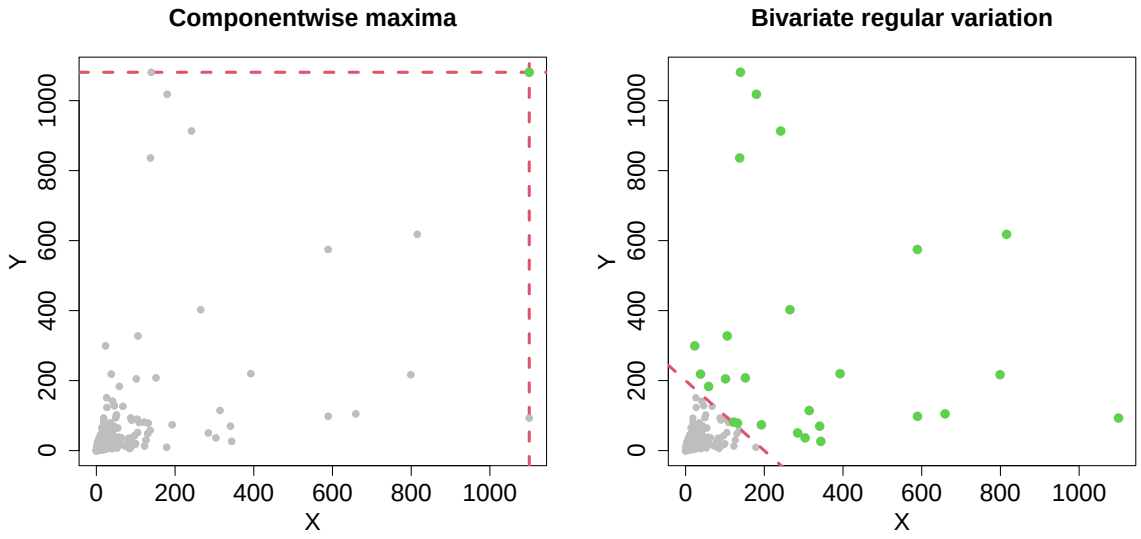

Figure 2.2.2: Componentwise maxima (left) and regular variation (right) modelling approaches for bivariate extremes. In either case, the green points represent the extreme values, while the red dotted lines represent marginal maxima and a high radial threshold for the left and right plots, respectively.

where V is known as the exponent measure (de Haan and Resnick, 1977). This measure takes the form

$$V(x,y) = 2\int_0^1 \max\left(\frac{w}{x}, \frac{1-w}{y}\right) \mathrm{d}H(w), \tag{2.2.2}$$

where H is a distribution function on $[0,1]$, known as the spectral measure, satisfying the mean constraint

$$\int_0^1 w \mathrm{d}H(w) = 1/2. \tag{2.2.3}$$

This implies that the class of BEV distributions are in one-to-one correspondence with the set of distributions functions H satisfying equation (2.2.3). Moreover, H (equivalently, V) captures the dependence structure between the extreme values, otherwise known as the extremal dependence structure. Special cases include independence, for which $H(\{0\}) = H(\{1\}) = 0.5$, equivalently $V(x,y) = x^{-1} + y^{-1}$, and complete dependence, for which $H(\{0.5\}) = 1$, equivalently $V(x,y) = \max\{x^{-1}, y^{-1}\}$.

From equation (2.2.2), it is straightforward to see that the exponent measure V

is homogeneous of order -1, i.e., $V(sx, sy) = s^{-1}V(x, y)$ for any $s > 0$. In practice, this implies that a fitted BEV distribution can be used to evaluate probability regions outside the observed data range, allowing for extrapolation of joint extreme events.

Several parametric forms for V have been proposed, including the logistic distribution (Gumbel, 1960) for which $V(x, y) = \left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha}$ with $\alpha \in (0, 1]$. Here, $\alpha = 1$ corresponds to independence and the degree of positive dependence increases as $\alpha \to 0$. Other popular parametric models include the asymmetric logistic (Tawn, 1988), Dirichlet (Coles and Tawn, 1991), and Hüsler-Reiss (Hüsler and Reiss, 1989) distributions.

Finally, we note that the extremal dependence structure can be equivalently characterised using Pickands' dependence function (PDF) (Pickands, 1981). Given a valid exponent measure $V(x, y)$, the PDF is given by

$$A(w) := V\left(\frac{1}{w}, \frac{1}{1-w}\right),$$

where $w := x/(x+y)$. This function is convex, and satisfies $\max(w, 1-w) \leq A(w) \leq 1$ for all $w \in [0, 1]$. The convexity of A implies that it is not possible to model negative dependence using the componentwise maxima approach, illustrating a drawback of this framework. In the cases of independence and complete dependence, we have that $A(w) = 1$ and $A(w) = \max(w, 1-w)$, respectively, for all $w \in [0, 1]$.

2.2.4 Bivariate regular variation

Again, let (X, Y) be a random vector on standard Fréchet margins and consider $R := X + Y$ and $W := X/R$; we refer to these variables as the radial and angular components of (X, Y), respectively. We say that (X, Y) is bivariate regularly varying if, for any

measurable $B \subset [0,1]$,

$$\lim_{r\to\infty} \Pr(W \in B, R > sr \mid R > r) = H(B)s^{-1}, \tag{2.2.4}$$

for any $s \geq 1$, where H is the spectral measure of equation (2.2.2) and $H(\partial B) = 0$, with ∂B denoting the boundary of B (Resnick, 1987). Limit (2.2.4) implies that for the largest radial values, the radial and angular components are approximately independent.

Evaluation of the distribution of $W \mid R > u$ for some large threshold u, can provide insight into the joint extremal behaviour of (X, Y). In this sense, this modelling framework can be viewed as a bivariate extension to the peaks over threshold approach in which R represents the magnitude of the joint extreme events, while W represents the relative contribution of each component of (X, Y). This approach is also illustrated in Figure 2.2.2. In analogy with the univariate framework, one can again observe the difference in the observations incorporated by this modelling technique and the componentwise maxima approach.

For both the componentwise maxima and bivariate regular variation frameworks, the extremal dependence is characterised by the spectral measure H or, equivalently, the PDF A. Consequently, a broad range of modelling approaches have been proposed for estimating these quantities; an overview of such techniques can be found in Chapter 9 of Beirlant et al. (2004). For examples of recent estimation approaches, see Einmahl and Segers (2009), Eastoe et al. (2014), Guillotte and Perron (2016) and Marcon et al. (2016).

2.2.5 Dependence measures

When modelling bivariate extremes, it is clear that capturing the extremal dependence structure is of paramount importance. A fundamental classification of this dependence is whether the most extreme events can occur jointly or not. Given any random vector

(X, Y), with $X \sim F_X, Y \sim F_Y$, consider the coefficient χ (Joe, 1997), defined as

$$\chi := \lim_{u \to 1} \Pr(F_Y(Y) > u \mid F_X(X) > u) \in [0, 1],$$

where the limit exists. If $\chi > 0$, the most extreme values of either variable can occur simultaneously, and we say that X and Y are asymptotically dependent. The strength of dependence increases as $\chi \to 1$, with perfect dependence arising at $\chi = 1$. Conversely, if $\chi = 0$, we say that X and Y are asymptotically independent since the most extreme values of either variable occur separately.

In practice, using the spectral measure (or PDF) to model asymptotically independent data leads to unsatisfactory extrapolation in the joint tail (Ledford and Tawn, 1996, 1997; Heffernan and Tawn, 2004). This occurs due to the fact for any random vector (X, Y) exhibiting asymptotic independence, the spectral measure H places all mass on the points $\{0\}$ and $\{1\}$; equivalently, $A(w) = 1$ for all $w \in [0, 1]$. Asymptotic independence is therefore a degenerate case for both modelling frameworks introduced in Chapters 2.2.3 and 2.2.4 since neither framework can distinguish between it and exact independence. This limits the use of these modelling frameworks to data sets exhibiting asymptotic dependence, and has consequently led to the development of flexible modelling approaches that are able to capture both classes of extremal dependence.

2.2.6 Modelling asymptotic independence

The first framework to capture asymptotic independence was proposed by Ledford and Tawn (1996). Given a random vector (X, Y) with standard exponential margins, the joint upper tail is assumed to take the form

$$\Pr(X > u, Y > u) = \Pr(\min(X, Y) > u) = L(e^u)e^{-u/\eta}, \tag{2.2.5}$$

as $u \to \infty$, where L is a slowly varying function at infinity, i.e., $\lim_{u\to\infty} L(cu)/L(u) = 1$ for $c > 0$, and $\eta \in (0, 1]$. The parameter η is termed the coefficient of tail dependence, with $\eta = 1$ and $\lim_{u\to\infty} L(u) > 0$ corresponding to asymptotic dependence and $\eta < 1$, or $\eta = 1$ and $\lim_{u\to\infty} L(u) = 0$, corresponding to asymptotic independence.

Similarly to χ, η provides a classification of extremal dependence, with the strength of dependence increasing as η approaches 1. The coefficient of tail dependence also indicates the form of asymptotic independence, with $\eta \in (1/2, 1)$, $\eta = 1/2$ and $\eta < 1/2$ corresponding to cases of positive association, (near) extremal independence and negative association between variables. In practical settings, both χ and η are unknown, and a wide range of methods exist for their estimation; see, for instance, Coles et al. (1999), Peng (1999), Frahm et al. (2005) and Aghakouchak et al. (2010).

Resnick (2002) formalised the theory of the limit expression denoted in equation (2.2.5), resulting in the concept of bivariate hidden regular variation. Given a random vector (X, Y) on standard Pareto margins, we say (X, Y) exhibits hidden regular variation if there exists a function $k : \mathbb{R}_+ \to \mathbb{R}_+$ such that, for all measurable sets $A \subset \mathbb{R}^2_+$,

$$u \Pr((X, Y)/k(u) \in A) \to \nu(A), \tag{2.2.6}$$

as $u \to \infty$, where k is regular varying with index η, i.e., $\lim_{u\to\infty} k(cu)/k(u) = c^\eta$, $c > 0$, and ν is a non-degenerate measure on $\mathbb{R}^2_+$ satisfying $\nu(tA) = t^{-1/\eta}\nu(A)$, $t > 0$. In the case of $\eta = 1$ (i.e., asymptotic dependence), equation (2.2.6) is equivalent to bivariate regular variation.

Extensions of the framework in equation (2.2.5) include Ledford and Tawn (1997), Draisma et al. (2004) or Ramos and Ledford (2009). However, like the original, all of these extensions are applicable only in regions where both variables are large. As such, statistical techniques based on equation (2.2.5) offer limited practical applicability. To address this central limitation, Heffernan and Tawn (2004) proposed a conditional modelling approach for bivariate extremes: given a random vector (X, Y) on standard

Laplace margins, it is assumed there exist normalising functions $a : \mathbb{R} \to \mathbb{R}$ and $b : \mathbb{R} \to \mathbb{R}_+$ such that

$$\lim_{u \to \infty} \Pr\left[(Y - a(X))/b(X) \leq z, X - u > x \mid X > u\right] = D(z)e^{-x}, \; x > 0, \tag{2.2.7}$$

for a non-degenerate distribution function D. This framework is able to capture both classes of extremal dependence, with asymptotic dependence arising when $a(x) = x$ and $b(x) = 1$, and asymptotic independence otherwise.

This method is a flexible approach for modelling bivariate extremes and, unlike equation (2.2.5), is not restricted to regions where both variables are large. Note that one can equivalently condition on the event $Y > u$ in equation (2.2.7) and assume the existence of normalising functions for X. In practice, the functions a and b are typically estimated parametrically under a misspecified model for D, while the distribution function D is subsequently estimated non-parametrically. In particular, setting $a(x) = \alpha x$ and $b(x) = x^\beta$, with $\alpha \in [-1, 1]$ and $\beta \in (-\infty, 1]$, captures the limiting joint conditional distributions for a large variety of bivariate distributions (Heffernan and Tawn, 2004; Keef et al., 2013a).

Wadsworth and Tawn (2013) subsequently introduced an alternative representation for bivariate tail probabilities via a general extension of the Ledford and Tawn (1996) framework. Allowing for different marginal growth rates of the variables, as in the conditional extremes approach of Heffernan and Tawn (2004), their model also captures both asymptotic dependence and asymptotic independence. Given (X, Y) on standard exponential margins, it is assumed that for each ray $w \in [0, 1]$,

$$\Pr(\min\{X/w, Y/(1-w)\} > u) = L(e^u; w)e^{-\lambda(w)u}, \tag{2.2.8}$$

as $u \to \infty$, where $L(\cdot\,; w)$ is slowly varying and $\lambda(w) \geq \max(w, 1 - w)$, termed the angular dependence function (ADF), generalises the coefficient of tail dependence, with

$\eta = 1/\{2\lambda(0.5)\}$. Asymptotic dependence implies the lower bound for the ADF, i.e., $\lambda(w) = \max(w, 1-w)$ for all $w \in [0,1]$. Evaluation of equation (2.2.8) for rays w close to 0 and 1 corresponds to regions where only one variable is extreme. Moreover, for the boundary cases $w = 0$ and $w = 1$, equation (2.2.8) reduces to the marginal exponential survivor functions, i.e., $\Pr(X > u) = \Pr(Y > u) = e^{-u}$. The quantity λ captures the extremal dependence structure for a wide variety of copula models; see Wadsworth and Tawn (2013) for further details. In practice, $\lambda(w)$ is estimated separately for each $w \in [0,1]$ via the Hill estimator (Hill, 1975).

Alongside these flexible approaches, there exist several copula-based modelling approaches that can capture both dependence regimes; see, for instance, Coles and Pauli (2002), Wadsworth et al. (2017) or Huser and Wadsworth (2019). For several of these proposed copula models, asymptotic dependence does not represent a boundary case. However, while this feature is practically advantageous, these copula models require specific parametric forms to be assumed for the joint distribution function. These strong assumptions significantly reduces the flexibility of these copula-based approaches, limiting their use in practice. In this sense, we prefer to adopt only those bivariate modelling approaches for which inference is semi- or non-parametric.

2.3 Non-stationary extremes

Thus far, all of the theory and subsequent statistical models have used the assumption that the underlying random variables are IID. However, this assumption is unlikely to hold in many practical settings. Many environmental hazards, such as temperatures, sea levels and river flows, exhibit complex distributional trends, thus invalidating the assumption of being 'identically distributed'. These trends can arise for a variety of reasons, such as seasonal behaviour and long-term shifts in climate conditions.

A random variable or vector for which the underlying distribution function(s) changes

over time is said to be non-stationary. In such cases, the standard extreme value modelling procedures described in Chapters 2.1 and 2.2, which assume stationarity, are no longer applicable, and additional statistical techniques are required to capture the underlying tail behaviours.

To accurately assess extremal risks from non-stationary random variables or vectors, the underlying trends should be incorporated in the statistical procedures. Moreover, such trends typically arise due to the influence of external covariates. For example, the observed increase in extreme global temperatures is driven, in large part, by the rise in atmospheric carbon dioxide emissions (Baker et al., 2018).

2.3.1 Univariate non-stationary extremes

Let $\{X_t\}$ denote a non-stationary univariate process with g-dimensional covariates $\mathbf{Z}_t, t \in \{1, 2, \ldots, n\}$, with $g \in \mathbb{N}$. Interest lies in understanding the tail behaviour of $X_t \mid \mathbf{Z}_t = \mathbf{z}_t$ for some covariate realisation $\mathbf{z}_t$.

The earliest approaches for modelling univariate non-stationary extremes involved allowing the parameters of the GEV or GP distributions to be covariate dependent; see Smith (1989) and Davison and Smith (1990). We restrict attention to the GP approach since it is more widely used in practice. In this case, the non-stationary model is given by

$$X_t - u \mid (X_t > u,\ \mathbf{Z}_t = \mathbf{z}_t) \sim \mathrm{GP}(\tilde{\sigma}(\mathbf{z}_t), \xi(\mathbf{z}_t)), \tag{2.3.1}$$

for a sufficiently large threshold u. The covariate functions, $\tilde{\sigma}$ and ξ, must be specified prior to applying this model; this is seldom a trivial task, and often requires a degree of subjective judgement.

Many extensions to equation (2.3.1) have been proposed. Several allow for the threshold u to also be covariate dependent: Kyselý et al. (2010) and Northrop and Jonathan (2011) use quantile regression to estimate a varying threshold with a constant exceedance probability, whereas Sigauke and Bere (2017) use a cubic smoothing spline.

A more recent class of models use random effects to capture non-stationary trends in the GP distribution parameters. Originally developed in the spatial context (Casson and Coles, 1999), these models have been used to represent a range of environmental variables, such as ozone levels (Gilleland et al., 2006), river flow data (Eastoe and Tawn, 2010) and precipitation data (Cooley et al., 2007).

Alongside these techniques, several flexible approaches have been proposed that use generalised additive models (GAMs) to capture non-stationary behaviour in model parameters for both the GEV and GP distributions (Chavez-Demoulin and Davison, 2005; Yee and Stephenson, 2007; Youngman, 2019). GAMs use smooth functions to capture covariate trends and are less rigid than linear regression models; see Wood (2017) for an overview.

Finally, Nogaj et al. (2007), Eastoe and Tawn (2009) and Mentaschi et al. (2016) propose pre-processing the data set prior to fitting the GP distribution. The pre-processing typically takes the form

$$\tilde{X}_t = \frac{X_t - m(\mathbf{z}_t)}{s(\mathbf{z}_t)},$$

for each $t \leq n$, with $m : \mathbb{R}^g \to \mathbb{R}$ and $s : \mathbb{R}^g \to \mathbb{R}_+$ denoting covariate functions relating to the mean and variance of $\{X_t\}$, respectively. Assuming the variable $\tilde{X}_t$ is stationary for all $t \leq n$, standard extreme value techniques can then be used to model tail behaviour. As with equation (2.3.1), the covariate functions must be specified prior to applying the pre-processing step.

Given the end goal is to quantify extremal risks, we observe that the traditional concept of a return level lacks meaning in the non-stationary setting, motivating an alternative definition. This is not a trivial task, and a variety of extended definitions have been proposed (Rootzén and Katz, 2013; Serinaldi, 2015). We restrict attention to one in particular: given any $t \in \{1, \ldots, n\}$ and a small probability p, the conditional return level $x_{p,t}$ is defined as the value satisfying $\Pr(X_t > x_{p,t} \mid \mathbf{Z}_t = \mathbf{z}_t) = p$ (Eastoe,

2019). Unlike its stationary counterpart, $x_{p,t}$ cannot be interpreted in terms of its return period, i.e., the process $\{X_t\}$ will not exceed $x_{p,t}$ once, on average, every $1/p$ observations, since the distribution varies over t. However, by fixing p and considering how the conditional return level varies over all t, one can evaluate the change in extremal risk across different covariate realisations. For example, this could allow one to assess changes in extreme temperatures over time as a result of increases in CO_2 emissions.

2.3.2 Bivariate non-stationary extremes

Let $\{X_t, Y_t\}$ denote a non-stationary bivariate process with covariates $\mathbf{Z}_t$ defined as before. In this setting, non-stationarity can exist in both the marginal and joint distribution functions. In the spirit of copula modelling, we opt for a separate treatment of marginal and dependence modelling and assume that the margins of $\{X_t, Y_t\}$ are stationary. In practice, if this is not the case, stationary marginals can be obtained by first fitting non-stationary marginal distributions, such as those described in Chapter 2.3.1, and then transforming to a stationary scale using the PIT.

Interest now lies in understanding how the extremal dependence structure of $\{X_t, Y_t\}$ varies under different realisations of the covariate process $\mathbf{Z}_t$. This problem has not been well studied in the literature, and consequently only a handful of modelling approaches exist. Of these approaches, the majority assume that the underlying process is bivariate regularly varying, and their use is thereby restricted to data sets exhibiting asymptotic dependence. de Carvalho and Davison (2014), Castro-Camilo et al. (2018) and Mhalla et al. (2019a) propose flexible modelling techniques for capturing non-stationary trends in the spectral measure under covariate influence, while Mhalla et al. (2017), Escobar-Bach et al. (2018) and Mhalla et al. (2019b) propose similar techniques for capturing non-stationarity in the PDF. Alongside these modelling frameworks, Bodik et al. (2022) propose a technique for detecting causal covariates influencing the PDF, allowing for a simplified analysis of dependence trends.

Far less consideration has been given to the case where the data is asymptotically independent. Jonathan et al. (2014b) and Guerrero et al. (2023) propose a non-stationary extension of the modelling framework described in equation (2.2.7), whereby the dependence parameters (α, β) are allowed to vary over time. In the former, the authors propose smooth covariate functions for a and b, while for the latter, the parameters (α, β) are allowed to vary smoothly over time for blocks of observations via a penalised log-likelihood. On the other hand, Mhalla et al. (2019b) propose non-stationary extension of equation (2.2.8), along with a flexible GAM-based technique for estimating the ADF under covariate influence.

2.4 Bivariate risk measures

As noted in Chapter 1, it is often important to quantify joint extreme risks from bivariate random vectors. However, this is not a trivial task, in part due to the lack of any unique definition for a 'bivariate extreme value'. As a result, a wide range of risk measures have been proposed within the literature.

Serinaldi (2015) introduces seven possible bivariate risk measures for extremes, with each measure defined by a different probability related to the joint distribution. Several additional risk measures have been proposed under the term 'environmental contour'; see Mackay and Haselsteiner (2021) and Haselsteiner et al. (2021) for an overview of such techniques. Whilst a wide range of practical applications exist, the suitability of each proposed measure is likely to be highly context dependent; see Lindt and Niedzwecki (2000), Salvadori and Michele (2004), Gouldby et al. (2017) and Velarde et al. (2019).

In this thesis, we restrict attention to one particular measure known as a return curve. Return curves are the natural bivariate extension of univariate return levels, and are also interpreted in terms of a return period. In addition, they are one of the measures that is of most interest in nuclear regulation (Office for Nuclear Regulation,

2017). Given a random vector (X, Y) and probability p, the p-probability return curve is given by the set

$$\mathrm{RC}(p) := \left\{(x, y) \in \mathbb{R}^2 : \Pr(X > x, Y > y) = p\right\}.$$

Clearly values of p close to zero correspond to rare joint exceedance events. Within the literature, this set has a variety of labels including level curves (Salvadori and Michele, 2004), quantile curves (Chebana and Ouarda, 2011), joint probability curves (Gouldby et al., 2017) and isolines (Cooley et al., 2019). In analogy to return levels, we also define the return period to be $1/p$; given any point $(x, y) \in \mathrm{RC}(p)$, we would expect to observe the joint exceedance event $\{X > x, Y > y\}$ once, on average, each return period. Equivalently, in a sample of size n from (X, Y), we expect to observe np points in the region $(x, \infty) \times (y, \infty)$.

The concept of a return curve is illustrated in Figure 2.4.1. Unlike in the univariate setting where each p-probability return level is unique, return curves are instead sets containing an infinite number of values. We also see that return curves include regions where only one variable is extreme as well as those in which both are extreme. Considering the left panel of Figure 2.4.1, one can observe how the behaviour of the return curves changes as the probability p (return period $1/p$) decreases (increases).

In many practical applications, estimates of $\mathrm{RC}(p)$ are required for return periods that exceed the number of observations. This extrapolation is only possible given an accurate description of the joint tail behaviour. As in univariate risk estimation, standard statistical techniques are unsuitable for this purpose, motivating the use of bivariate extreme value theory.

One must also quantify uncertainty when estimating return curves. Compared with univariate risk measures, this is an even more complex problem since uncertainty arises over two-dimensional space. It is also essential to evaluate goodness of fit of a given return curve estimate. In particular, one must ensure the curve estimate accurately

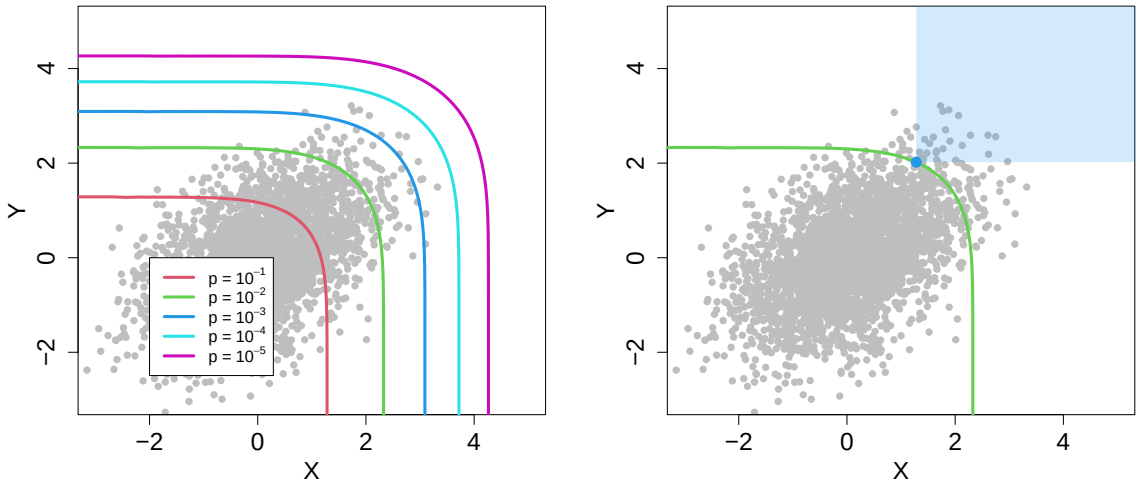

Figure 2.4.1: Illustrations of return curves for bivariate Gaussian data on standard normal margins, with $\rho = 0.5$ and $n = 3,000$. Left: return curves for five probabilities $p \in \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$. Right: the joint survivor set $(x, \infty) \times (y, \infty)$ (shaded blue region) for a point on the $p = 10^{-2}$ return curve. The probability of observing data in joint survivor sets of form $(x, \infty) \times (y, \infty)$ is constant and equal to p for any point on the return curve.

represents the probability p.

Within the literature for extremes, little attention has been given to the estimation of return curves. Moreover, the limitations of the few existing approaches clearly motivate the development of novel estimation techniques. These points are addressed in Chapter 3, where we introduce a range of novel estimation techniques for the return curve, as well as developing tools for evaluating uncertainty and goodness of fit.

Finally, like the univariate return level, the standard definition of a return curve lacks meaning in the non-stationary setting. We therefore extend the definition of this risk measure to be covariate-dependent. Given a process $\{X_t, Y_t\}$ with covariates $\mathbf{Z}_t,\ t \in \{1, 2, \ldots, n\}$, we define the p-probability return curve at the covariate realisation $\mathbf{z}_t$ to be

$$\mathrm{RC}_{\mathbf{z}_t}(p) := \{(x, y) \in \mathbb{R}^2 \mid \Pr(X_t > x, Y_t > y \mid \mathbf{Z}_t = \mathbf{z}_t) = p\}.$$

Thus far, no methods have been proposed for estimating return curves under covariate

influence within the literature, even though non-stationarity is commonly observed in many practical settings where quantifying joint risks is important. This gap in the literature is addressed in Chapter 4.

Chapter 3

New estimation methods for extremal bivariate return curves

3.1 Introduction

3.1.1 Univariate extremal risk measures

Statistical analysis of extreme values is important in a wide range of environmental contexts, from the modelling of ocean waves (Shooter et al., 2021) and wind speeds (Baran et al., 2021) to climatological assessments (Yu et al., 2022; Worms and Naveau, 2022). The two most common approaches to modelling the extreme behaviour (or tail) of a single variable are block maxima and peaks over threshold (Coles, 2001). For the latter, which is more popular in practice, the generalised Pareto distribution (GPD) is used to model exceedances of some high threshold. This is justified through the Pickands-Balkema-de Haan theorem (Balkema and de Haan, 1974; Pickands, 1975); given a random variable $X \sim F_X$ satisfying certain conditions, there exists a normalising function $c(u)$ such that

$$\Pr\left(\frac{X-u}{c(u)} \leq x \;\middle|\; X > u\right) \to G(x) := 1 - \left\{1 + \frac{\xi x}{\sigma}\right\}_+^{-1/\xi}, \;\; x > 0, \; (\sigma, \xi) \in \mathbb{R}^+ \times \mathbb{R},$$

as $u \to x^F := \sup\{x : F_X(x) < 1\}$. Here, G is the cumulative distribution function of a GPD, with scale and shape parameters, σ and ξ, respectively, and $z_+ = \max(0, z)$. The shape parameter ξ determines the behaviour of the tail: the cases $\xi < 0$, $\xi = 0$ and $\xi > 0$ correspond to bounded, exponential and heavy tails, respectively. Given a sufficiently high threshold u, we assume the statistical model $X - u \mid X > u \sim \text{GPD}(\sigma, \xi)$. A detailed discussion of peaks over threshold modelling can be found in Coles (2001).

Often, univariate extreme value models are used to estimate risk measures for events associated with small probabilities; these summary statistics provide a quantification of extremal behaviour that can be used to help mitigate against rare environmental events, such as floods, storms, or wildfires. One such measure is known as a return level. Given a real variable X, representing a measurement taken at regular time intervals, and probability p, the p-probability return level is the value x_p that satisfies the equation $\Pr(X > x_p) = p$. For small p, x_p represents a high quantile which can be estimated using the GPD. We restrict attention to the case when X is stationary since the interpretation of return levels is more straightforward in this setting. We define the return period of x_p to be the value $1/p$; one would expect the variable X to exceed x_p once, on average, during each return period. The relationship between return levels and periods can be illustrated using a return level plot; examples of three such plots with varying shape parameters are given in the left panel of Figure 3.1.1.

Return levels are widely used and provide a simple way to understand risk. However, many potentially impactful events arise due to the effect of more than one variable. For example, Mattei et al. (2001) describe how the combination of high sea levels and wind led to large-scale flooding at the Blayais nuclear power plant in 1999, causing significant damage. For this reason, it is desirable to have similar risk measures in the multivariate case, but thus far relatively little consideration has been given to this problem. This is in part due to the lack of natural ordering for multivariate vectors, which means there is no longer a single definition of an extreme event. A variety of measures have been

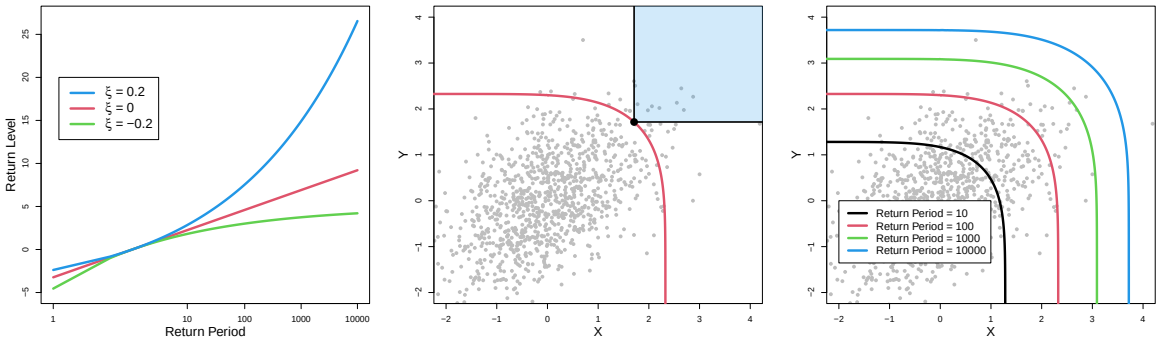

Figure 3.1.1: Left: Return level plots under varying shape parameters. Centre: Return curve RC(p) (red) of standard bivariate normal data with $\rho = 0.5$ and $n = 1000$ for return period of 100 ($p = 1/100$). In a sample of size n, we expect to observe np points in the blue shaded region; this is true for all equivalent shaded regions at any point on the curve. Right: Return curves of the same bivariate normal data set for return periods in the set $\{10, 100, 1000, 10000\}$.

proposed (Serinaldi, 2015), with each suited to a different analytical need. In this text, we focus on the bivariate case and one particular measure known as a return curve, which directly extends the return level concept to the bivariate setting.

3.1.2 Return curves

Consider the joint survival function of the continuous random vector (X, Y) at a given probability p, i.e. $\Pr(X > x, Y > y) = p$. The combinations $(x, y) \in \mathbb{R}^2$ satisfying this equation define a curve in the plane; we therefore define the p-probability return curve to be the set

$$\mathrm{RC}(p) := \left\{(x, y) \in \mathbb{R}^2 : \Pr(X > x, Y > y) = p\right\}.$$

We consider values of p close to zero, corresponding to rare joint exceedance events. Within the literature, this set has a variety of labels, including isolines (Cooley et al., 2019), hazard curves (Simpson and Wadsworth, 2017) and joint probability curves (Gouldby et al., 2017). In an analogue to return levels, we define the return period to be $1/p$, since given any point $(x, y) \in \mathrm{RC}(p)$, we would expect to observe the event

$\{X > x, Y > y\}$ once, on average, each return period. Equivalently, in a sample of size n from (X, Y), we expect to observe np points in the region $(x, \infty) \times (y, \infty)$.

Since return curves define a line in $\mathbb{R}^2$ rather than a single value, the two dimensional return level plot does not naturally extend to this setting. Instead, we can consider different return periods and plot the corresponding curves individually or simultaneously; examples of both are given in the centre and right panels of Figure 3.1.1 for a standard bivariate normal data set with correlation coefficient $\rho = 0.5$.

Return curves are arguably the most intuitive bivariate extension to return levels since they are also defined in terms of the survivor function. They have been used in practice to derive extremal environmental conditions for the design and analysis of ocean structures, such as oil rigs (Jonathan et al., 2014a), freight ships (Vanem et al., 2020) and wind turbines (Manuel et al., 2018; Velarde et al., 2019), and coastal structures, such as railway lines (Environmental Agency, 2005; Gouldby et al., 2017) and wave energy converters (Eckert-Gallup and Martin, 2016).

As motivating examples, we consider two environmental data sets of practical importance, both of which are illustrated in Figure 3.1.2. Our objective is to use return curve estimates to derive joint extremal conditions for each data set. Both data sets are comprised of metocean variables, which have previously been used in a comparison exercise for a risk measure known as an environmental contour (Haselsteiner et al., 2021); such measures also aim to summarise joint extremal behaviour. However, unlike return curves, they do not offer an intuitive interpretation in terms of return periods.

The first data set contains measured significant wave height (m) and zero up crossing period (s) between 1996-2005 obtained from a buoy on the east coast of Florida, USA. The second data set contains 25 years of wind speed (m/s) and significant wave height (m) observations obtained from the hindcast model coastDat -2 (Groll and Weisse, 2017) for a location in the North Sea near the east coast of the UK. These combinations of variables are of particular relevance for the structural reliability of offshore and coastal

structures, and bivariate risk measures are commonly used to inform the design basis for such structures (Jonathan et al., 2014a; Haselsteiner et al., 2019; Mackay and Jonathan, 2020). They therefore provide realistic examples with which to illustrate the utility of return curve estimates.

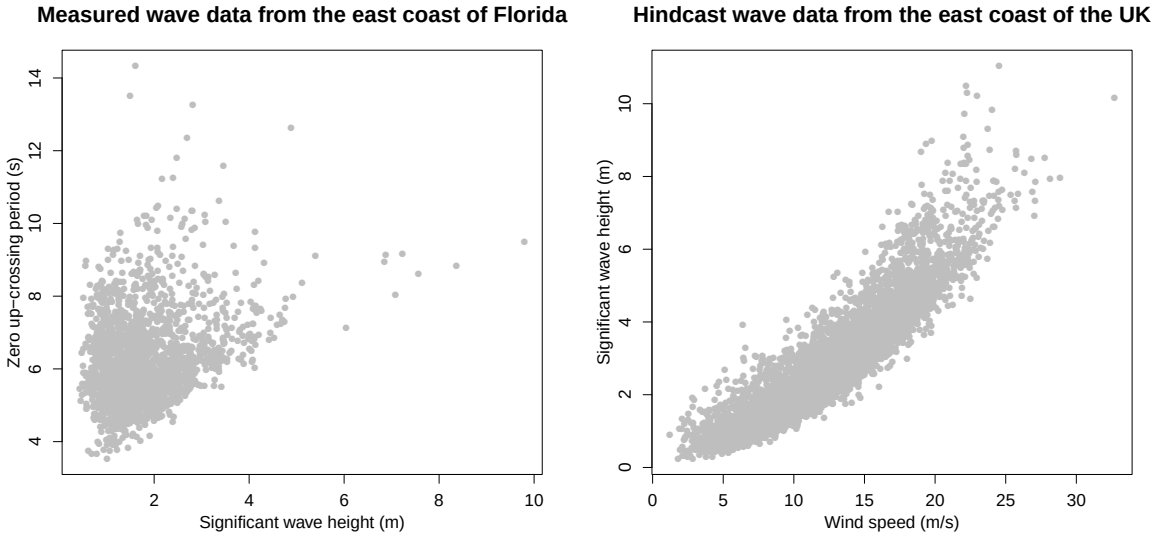

Figure 3.1.2: Left: Measured daily maxima metocean data for a buoy located on the east coast of Florida, USA. Right: Hindcast daily maxima metocean data for a location near the east coast of the UK.

Both of the original data sets are comprised of hourly observations, resulting in strong temporal dependence between observations. To reduce the temporal dependence while retaining a relevant data set, we restrict attention to daily maxima values. Furthermore, we account for seasonal non-stationarity by restricting attention to the months September-March; our analysis suggests the largest wave heights are observed in this interval. This results in $n = 2048$ and $n = 5306$ observations, respectively, for the measured and hindcast data sets.

3.1.3 Practical considerations for return curve estimation

To estimate return curves in practice, we must first estimate the joint survival function of an observed bivariate process at a fixed probability p. Since we restrict attention to

probabilities close to zero, we require an estimation method that can accurately capture the joint tail behaviour of the process and provide realistic extrapolation to estimate $\mathrm{RC}(p)$ for p outside of the observation period. For this reason, we restrict attention to models from multivariate extreme value theory. We must also consider methods for evaluating the uncertainty associated with return curve estimates. This is a more complex problem than assessing the uncertainty of return level estimates, since there is more than one dimension in which the risk measure can vary. Furthermore, given a data set for which a curve has been estimated at probability p, it is also essential to evaluate goodness of fit via a diagnostic tool. Little attention has been given to either of these problems within the extremes literature, motivating the development of novel approaches.

This article is structured as follows. Section 3.2 provides a brief background on bivariate extreme value theory. In Section 3.3, we introduce various properties of return curves, review existing approaches for curve estimation within the extremes literature and present two novel approaches. Section 3.4 introduces new tools for analysing uncertainty and goodness of fit in return curve estimates. Section 3.5 presents a simulation study to evaluate the performance of these tools and compare curve estimates from several models. In Section 3.6, we apply our methodology to the two aforementioned environmental data sets. We conclude in Section 3.7 with a discussion and outlook on future work.

3.2 Bivariate extreme value theory

When assessing the extremal behaviour of a continuous bivariate random vector (X, Y), one must consider tail behaviour within both marginal processes, along with the dependence between the largest observations of either variable, which we refer to as the extremal dependence. A fundamental classification of extremal dependence is whether

the most extreme events can occur jointly or not. This is quantified by the coefficient $\chi := \lim_{u \to 1} \chi(u) \in [0, 1]$, where

$$\chi(u) = \Pr(F_Y(Y) > u \mid F_X(X) > u),$$

for $X \sim F_X$, $Y \sim F_Y$. The cases $\chi = 0$ and $\chi > 0$ correspond to asymptotic independence (AI) and asymptotic dependence (AD), respectively. A larger χ corresponds to stronger positive dependence in the joint tail, with $\chi = 1$ occurring for perfectly dependent variables. In practice, we cannot estimate χ in the limit as $u \to 1$, meaning we must use $\chi(u)$, $u < 1$, to determine AD vs AI.

Early modelling techniques proposed for multivariate extreme values focused only on the AD case (de Haan and Resnick, 1977; Coles and Tawn, 1991; de Haan and de Ronde, 1998). Such methods are based on the framework of multivariate regular variation: given a random vector (X, Y) with standard Fréchet margins, we define radial and angular components $R := X + Y$ and $W := X/R$, respectively. We say (X, Y) is multivariate regularly varying if, for measurable $B \subset [0, 1]$, we have

$$\lim_{r \to \infty} \Pr(W \in B, R > sr \mid R > r) = H(B)s^{-1}, \; s \geq 1, \tag{3.2.1}$$

with $H(\partial B) = 0$, where ∂B is the boundary of B (Resnick, 1987). Limit (3.2.1) implies that for large radial values, R and W are approximately independent. The quantity H, which is known as the spectral measure, captures the extremal dependence structure of (X, Y) and must satisfy the moment constraint $\int_0^1 w \mathrm{d}H(w) = 1/2$. For all AI vectors, the spectral measure places mass on the atoms $\{0\}$ and $\{1\}$; consequently, under AI, this framework is degenerate and is unable to accurately extrapolate into the joint tail, except in the special case of independence (Ledford and Tawn, 1996, 1997).

In recent years, it has been shown that the AI case is at least as important as the AD case, and that assuming the incorrect form of extremal dependence leads to unsuitable

extrapolation in the joint upper tail (Ledford and Tawn, 1996, 1997; Heffernan and Tawn, 2004). Therefore, unless there is strong prior knowledge in favour of either AD or AI, it is desirable to use models which have sufficient flexibility to allow the data to directly inform the class of extremal dependence structure.

The first such approach was proposed in Ledford and Tawn (1996). Given (X, Y) with standard exponential margins, this model assumes the joint tail representation

$$\Pr(X > t, Y > t) = \Pr(\min(X, Y) > t) = L(e^t)e^{-t/\eta} \text{ as } t \to \infty, \tag{3.2.2}$$

where L is a slowly varying function at infinity, i.e., $\lim_{t\to\infty} L(ct)/L(t) = 1$ for $c > 0$, and $\eta \in (0, 1]$. The parameter η is termed the coefficient of tail dependence, with $\eta = 1$ and $\lim_{t\to\infty} L(t) > 0$ corresponding to AD and $\eta < 1$, or $\eta = 1$ and $\lim_{t\to\infty} L(t) = 0$, corresponding to AI. Estimation of η can be performed in practice using the Hill estimator (Hill, 1975). Several extensions to this approach exist (Ledford and Tawn, 1997; Ramos and Ledford, 2009): however, all techniques derived under this framework are applicable only within regions where both variables are large. Consequently, these methods are not appropriate for the estimation of return curves, since this measure is defined also in regions where only one variable is large; see Figure 3.1.1.

Wadsworth and Tawn (2013) provide an alternative representation for bivariate tail probabilities using a more general extension of the model described in equation (3.2.2) that allows for joint tail estimation in regions where only one variable is large. Given (X, Y) with standard exponential margins, they assume for each $w \in [0, 1]$

$$\Pr(\min\{X/w, Y/(1-w)\} > t) = L(e^t \mid w)e^{-\lambda(w)t},\ \lambda(w) \geq \max(w, 1-w), \tag{3.2.3}$$

as $t \to \infty$, where $L(\cdot \mid w)$ is slowly varying for each ray $w \in [0, 1]$. The function λ, which is termed the angular dependence function, is the key quantity in determining joint tail behaviour, and both AD and AI can be captured under this assumption,

with AD implying the lower bound $\lambda(w) = \max(w, 1-w)$. This quantity generalises the coefficient η, with $\eta = 1/(2\lambda(0.5))$, and can be estimated pointwise for any ray $w \in [0,1]$ using the Hill estimator. This approach can be used to estimate joint survivor probabilities where only one variable is large by taking values of w close to 0 or 1.

Heffernan and Tawn (2004) proposed a very general modelling tool for conditional probabilities. We consider the extension given in Keef et al. (2013a) since the formulation given in the original approach cannot easily accommodate structures exhibiting negative dependence. Given a random vector (X, Y) with standard Laplace margins, it is assumed that there exist normalising functions $a : \mathbb{R} \to \mathbb{R}$ and $b : \mathbb{R} \to \mathbb{R}_+$ such that

$$\lim_{t\to\infty} \Pr\left[(Y - a(X))/b(X) \leq z, X - t > x \mid X > t\right] = D(z)e^{-x}, \tag{3.2.4}$$

for a non-degenerate distribution function D. Similarly to Wadsworth and Tawn (2013), this framework is able to capture both AD and AI, with AD arising when $a(x) = x$ and $b(x) = 1$. This method is a flexible approach for modelling multivariate extremes and is also not restricted only to regions where both variables are large. Note that one could also condition on the event $Y > t$ and assume the existence of normalising functions for the variable X: in combination, these assumptions allow consideration of the region where either variable is large. The functions a and b are typically estimated parametrically under a misspecified model for D, while the distribution function D is subsequently estimated non-parametrically.

Alongside these approaches, we note that there exist a range of copula-based models that can capture both dependence regimes (Coles and Pauli, 2002; Wadsworth et al., 2017; Huser and Wadsworth, 2019). Such techniques aim to create a unified modelling framework for AD and AI. Moreover, the case of AD does not represent a boundary case for the latter two approaches, which could be practically advantageous. However, they all require stronger assumptions about the form of parametric family for the bivariate distribution, reducing their flexibility and limiting their use in practice. As a result,

we prefer instead to consider the more flexible modelling assumptions described in equations (3.2.3) and (3.2.4).

3.3 Bivariate return curve estimation

We now consider techniques for practical estimation of return curves. We begin by introducing theoretical results for return curves in Section 3.3.1. Naive implementation of statistical estimation methods would generally produce curves that fail to respect these results, but imposing them will typically improve estimation. Section 3.3.2 details how to perform marginal transformations so that the models in equations (3.2.3) and (3.2.4) can be applied in practice. In Section 3.3.3, we present existing approaches for return curve estimation before introducing novel estimation techniques in Section 3.3.4.

3.3.1 Return curve properties

A careful consideration of the theory surrounding the joint survival function allows us to deduce several properties about the shape and magnitude of $\mathrm{RC}(p)$ for a given $p \in (0,1)$. We begin by noting that the joint distribution function, $F_{X,Y}$, can be expressed in terms of the marginal distribution functions of X and Y, F_X, F_Y, respectively, and a copula C via $F_{X,Y}(x,y) = C(F_X(x), F_Y(y))$. The return curve is linked to the joint distribution function $F_{X,Y}$ by the equation $\Pr(X > x, Y > y) = 1 - F_X(x) - F_Y(y) + F_{X,Y}(x,y)$. Throughout this section, we make the assumption that the random vector (X,Y) has strictly continuous marginal distribution functions.

Property 3.3.1. *Let $x_p := F_X^{-1}(1-p)$ and $y_p := F_Y^{-1}(1-p)$ be the $(1-p)$-th quantiles of X and Y, respectively. Then for $(x,y) \in \mathrm{RC}(p)$, $x \leq x_p$ and $y \leq y_p$.*

Proof. We have $\Pr(X > x) \geq \Pr(X > x, Y > y) = p = \Pr(X > x_p)$ and hence $x \leq x_p$.

□

This result bounds the coordinate values that can be observed on the return curve. Next, by considering the limit of the joint survivor function as one variable converges to the lower limit of the marginal support, we obtain the following result.

Property 3.3.2. *Let* $supp(F)$ *denote the support of* F *and* $x_{inf} := \inf\{supp(F_X)\}$, $y_{inf} := \inf\{supp(F_Y)\}$. *We have that*

$$\Pr(X > x, Y > y) = \begin{cases} \Pr(Y > y) & \text{if } x \leq x_{inf}, \\ \Pr(X > x) & \text{if } y \leq y_{inf}. \end{cases}$$

Combining this statement with Property 3.3.1, Property 3.3.3 follows.

Property 3.3.3. *Let* $(x, y) \in \mathrm{RC}(p)$. *If* $x \leq x_{inf}$ $(y \leq y_{inf})$, *then* $y = y_p$ $(x = x_p)$.

These results allow us to easily compute the curve coordinates on the regions $(-\infty, x_{inf}) \times (y_p, \infty)$ and $(x_p, \infty) \times (-\infty, y_{inf})$, assuming we can accurately estimate the marginal quantiles (x_p, y_p) and the infima of marginal supports (x_{inf}, y_{inf}). Finally, by considering coordinates at different points on a return curve, we obtain the following result.

Property 3.3.4. *Suppose the copula,* C, *of* (X, Y) *on uniform margins has joint support on the whole of* $[0, 1]^2$ *and joint density function, denoted* c. *Given* $(x_1, y_1), (x_2, y_2) \in \mathrm{RC}(p)$ *with* $0 < F_X(x_1), F_X(x_2) < 1$ *and* $0 < F_Y(y_1), F_Y(y_2) < 1$, *we have that* $x_1 < x_2 \Leftrightarrow y_1 > y_2$.

Proof. Suppose $x_1 < x_2$ and $y_1 \leq y_2$. This implies that

$$\begin{aligned}
p &= \Pr(X > x_1, Y > y_1) = \Pr(F_X(X) > F_X(x_1), F_Y(Y) > F_Y(y_1)) \\
&= \int_{F_X(x_1)}^{1} \int_{F_Y(y_1)}^{1} c(u,v)\mathrm{d}v\mathrm{d}u \\
&= \int_{F_X(x_1)}^{F_X(x_2)} \int_{F_Y(y_1)}^{1} c(u,v)\mathrm{d}v\mathrm{d}u + \int_{F_X(x_2)}^{1} \int_{F_Y(y_1)}^{1} c(u,v)\mathrm{d}v\mathrm{d}u \\
&> \int_{F_X(x_2)}^{1} \int_{F_Y(y_1)}^{1} c(u,v)\mathrm{d}v\mathrm{d}u \quad \text{(since we have support on the whole of } [0,1]^2) \\
&\geq \int_{F_X(x_2)}^{1} \int_{F_Y(y_2)}^{1} c(u,v)\mathrm{d}v\mathrm{d}u = \Pr(X > x_2, Y > y_2) = p
\end{aligned}$$

implying $p > p$, a contradiction. Hence, $y_1 > y_2$. □

This result governs the shape of the contour defined by the return curve set. We note there is an alternative proof given in Cooley et al. (2019) under the assumption of monotonicity of the joint survivor function.

3.3.2 Marginal transformations

From Section 3.2, it is clear that in order to apply multivariate extreme value models, we need to standardise the marginal distributions of a random vector to achieve the form assumed by the model. Typically, inference involves two steps: forward transformation to get the data onto desired margins and back transformation to move any computed statistics, such as a return curve, back onto the original margins. For both steps, we use the semi-parametric approach given in Coles and Tawn (1991). Given an identically distributed sample $\{(x_i, y_i) : i = 1, \ldots, n\}$ from a random vector (X, Y) with unknown margins, we estimate the marginal distribution $\hat{F}_X$ (similarly $\hat{F}_Y$) by

$$\hat{F}_X(x) = \begin{cases} 1 - \{1 - \tilde{F}_X(u_X)\}\{1 + \xi_X(x - u_X)/\sigma_X\}_+^{-1/\xi_X}, & \text{for } x > u_X, \\ \tilde{F}_X(x), & \text{for } x \leq u_X, \end{cases} \tag{3.3.1}$$

where the first line represents the GPD above a high threshold u_X and $\tilde{F}_X$ is an empirical rank transform given by $\tilde{F}_X(x) = \sum_{i=1}^{n} \mathbb{1}(x_i \leq x)/(n+1)$. This approach ensures the marginal tail behaviour is captured within the transformation. Moreover, equation (3.3.1) can be easily inverted to perform the back transformation step.

3.3.3 Existing methodology

The literature on extremal return curve estimation is sparse, owing to the fact that little consideration has been given to this problem in practice. Of the existing approaches, each can be designated into one of three categories: approaches for AD data only, approaches for AI data only, and approaches applicable to data exhibiting either regime. The majority of the available literature falls within the first of these categories (Stephenson, 2002; Salvadori and Michele, 2004; Marcon et al., 2017a). In all cases, the authors assume the bivariate copula is in a family of distributions termed bivariate extreme value copulas. These copulas, which are directly related to the spectral measure from limit (3.2.1), imply AD and provide the basis for the majority of multivariate extreme value techniques. The necessity for making such a strict assumption is a well known drawback of this kind of model, since the form of extremal dependence is seldom known prior to analysis and AI is frequently observed in practice (Heffernan and Tawn, 2004; Huser and Wadsworth, 2019). We therefore choose not to consider such approaches further.

In Cooley et al. (2019), the authors propose separate techniques for the first and second categories. In both cases, extremal return curve estimates are obtained by 'projecting' empirical curves estimated for less extreme probabilities. For the case of AD, given a random vector (X, Y) with standard Fréchet margins, multivariate regular variation is exploited to obtain curve estimates. Consider two small probabilities p and p^* with $p^* > p$; multivariate regular variation implies that $\mathrm{RC}(p) \approx s^{-1}\mathrm{RC}(p^*)$, where $s := p^*/p > 1$. In practice, $\mathrm{RC}(p^*)$ is estimated empirically via a smooth

Gaussian-kernel estimate of the joint survivor function, and scaled by the coefficient s^{-1} to produce an estimate for RC(p).

In the case of AI, a similar estimation procedure is proposed based on the framework of hidden regular variation (Resnick, 2002), an elaboration of the assumption outlined in equation (3.2.2). However, as mentioned in Section 3.2, this approach only works in regions where both variables are large. To account for this, Cooley et al. (2019) proposed an ad-hoc procedure to link this region to the marginal axes, requiring additional steps and parameter estimation. In contrast, the assumption in equation (3.2.3) provides a theoretically sound link between regions where variables are of different magnitudes. Estimates of return curves from the Cooley et al. (2019) approach are illustrated in Figure 3.5.1. In Section 3.5, we compare the resulting curve estimates from this approach to the methods introduced in this paper and show that the techniques we present outperform this method in a wide range of scenarios.

For the third category, few approaches exist within the literature; this is in part because the bivariate extreme value methodologies that allow for dual estimation are relatively modern. All proposed techniques use a semi-parametric implementation of the conditional extremes model described in equation (3.2.4) (Jonathan et al., 2014a; Gouldby et al., 2017; Simpson and Wadsworth, 2017; Tilloy et al., 2020). However, like the other techniques introduced in this section, little to no consideration is given to the theory behind return curves, leading to curve estimates with undesirable properties. Moreover, as will be discussed in Section 3.3.4.1, utilising the Heffernan and Tawn (2004) modelling framework for return curve estimation is not straightforward, requiring delicate treatment and several steps; this has not been fully acknowledged in these existing approaches.

To the best of our knowledge, the modelling techniques discussed here cover all of the proposed methods for estimating return curves at extremal probabilities. Furthermore, we know of no attempt to compare curve estimates from these different methods.

No formal quantifications of return curve uncertainty or bias have been proposed previously, making it difficult to evaluate performance over different dependence structures. Some approaches (Simpson and Wadsworth, 2017; Cooley et al., 2019) instead provide bootstrap curve estimates which, while representing the uncertainty in curve estimates, do not provide interpretable confidence regions.

Alongside this issue, there is only one diagnostic tool in the literature for evaluating the accuracy of return curve estimates (Cooley et al., 2019). This tool utilised the result that, if data are independent and identically distributed, the number of points in each survival region on the return curve should theoretically be Binomial(n, p) distributed, where n denotes the size of the data set. This property can be used to construct a theoretical confidence region for the number of points in the joint survival set $(x, \infty) \times (y, \infty)$ for any point $(x, y) \in \text{RC}(p)$. While the authors show that these confidence regions capture the number of observations within estimated joint survival sets for the majority of considered examples, we argue that the resulting diagnostic is relatively uninformative since it lacks an intuitive interpretation in terms of the survival probability p. Moreover, this diagnostic strongly relies on the assumption of independent observations, which is seldom the case in practice. We present an alternative diagnostic tool in Section 3.4.2, where confidence intervals are instead obtained using sets of empirical probability estimates obtained through bootstrapping and compared to the true probability p. Temporal dependence can be incorporated through block bootstrapping, meaning this tool can be applied to a wider range of data sets; see Sections 3.4.2 and 3.6 for further details.

3.3.4 Novel methods for return curve estimation

We outline two methods for estimation of $\text{RC}(p)$ based on the modelling assumptions given in equations (3.2.3) and (3.2.4). Consider a random vector (X, Y) with standard exponential margins, for which the marginal support is given by the set $\mathbb{R}_+$; this implies

$x_{inf} = y_{inf} = 0$. We can immediately deduce from Property 3.3.3 that the coordinates of the return curve intersecting the margins are given by $(0, y_p)$ and $(x_p, 0)$, with $y_p = x_p = -\log(p)$, the $(1-p)$-th quantile. These coordinates give us 'start' and 'end' points for curve construction. Moreover, given a curve estimate $\widehat{\mathrm{RC}}(p)$ for this vector constructed with these boundary points, Properties 3.3.1 and 3.3.4 can be imposed to ensure the resulting curve is theoretically possible. For the former, if any $(x, y) \in \widehat{\mathrm{RC}}(p)$ satisfy $x > x_p$ (similarly, $y > y_p$), we set $x = x_p$ ($y = y_p$), thereby bounding values on the curve. For the latter, we treat the bounded curve estimate as a function of x and apply an iterative algorithm starting at the point $(0, y_p)$ to obtain a monotonic function. Incorporating additional theoretical knowledge into return curve estimation should lead to more accurate and robust estimates. For both of the methods introduced in this section, we impose the properties above retrospectively once curve estimates have been obtained.

3.3.4.1 Method based on the approach given in Heffernan and Tawn (2004)

In this section, we propose an implementation of the Heffernan and Tawn (2004) model, which builds on the existing methods that have applied this framework for return curve construction. Unlike these techniques, we incorporate the properties introduced in Section 3.3.1 into return curve estimates and provide an intuitive algorithm for combining the point estimates obtained from conditioning on both variables. Let (X_L, Y_L) denote the vector (X, Y) on standard Laplace margins and consider a small probability p for which we wish to obtain a return curve estimate. To achieve this, we fit the Heffernan and Tawn (2004) model twice, conditioning on both X_L and Y_L separately, thus allowing us to estimate the curve in different regions. In particular, we consider the regions defined by $R_{Y_L > X_L} := \{(x_L, y_L) \in \mathbb{R}^2 \mid y_L > x_L\}$ and $R_{Y_L \leq X_L} := \{(x_L, y_L) \in \mathbb{R}^2 \mid y_L \leq x_L\}$.

For $R_{Y_L > X_L}$, we first select a high quantile u_{Y_L} from the marginal distribution of Y_L

such that $\Pr(Y_L > u_{Y_L}) > p$. In particular, we select the 0.95 empirical quantile of this distribution, implying the return curve probability, p, must be smaller than 0.05. We assume the normalising functions are given by $a(y) = \alpha y$ and $b(y) = y^\beta$ for constants $\alpha \in [-1, 1]$ and $\beta \in (-\infty, 1)$: as noted in Keef et al. (2013a), these functions capture the limiting dependence structures for a wide range of distributions. The parameters α and β can be estimated under the working assumption that the distribution function D, which captures the stochastic behaviour of the variable $(X_L - \alpha Y_L)/Y_L^\beta \mid Y_L > u_{Y_L}$, is that of a Gaussian distribution. We denote the fitted values by $\hat{\alpha}$ and $\hat{\beta}$. These values can be used to simulate from the variable $X_L \mid Y_L > u_{Y_L}$; for example, see Jonathan et al. (2014a).

We then consider a decreasing sequence of high quantiles from Y_L that exist in the interval $(u_{Y_L}, F_{Y_L}^{-1}(1-p))$. The upper end point of this interval is the limit that Y_L values can attain on this curve and the lower end point represents the minimal quantile for which the fitted model is valid. We denote this set by $\mathcal{Y}$ and iteratively consider each $y_* \in \mathcal{Y}$ in turn, with $q := \Pr(Y_L > y_*)$. Using the fitted parameter values, we use the model to simulate from the conditional distribution $X_L \mid Y_L > y_*$. Letting x_* denote the (estimated) $(1 - p/q)$-th quantile from this distribution, we have that the resulting coordinate (x_*, y_*) is a member of the set $\widehat{\mathrm{RC}}(p)$ (defined for (X_L, Y_L)) since $\Pr(X_L > x_*, Y_L > y_*) = \Pr(X_L > x_* \mid Y_L > y_*) \Pr(Y_L > y_*) = \frac{p}{q} \times q = p$. We continue in this manner until we obtain a value x_{**} with $y_* \leq x_{**}$, or we have exhausted all values in the set $\mathcal{Y}$. The resulting coordinate set then gives an estimate of the curve in $R_{Y_L > X_L}$.

A near identical procedure is used to obtain the curve estimate in $R_{Y_L \leq X_L}$, this time selecting a high quantile u_{X_L} from the distribution of X_L and fitting the conditional model above this quantile. We then consider a set of quantiles in the interval $(x'_{**}, F_{X_L}^{-1}(1-p))$, where $x'_{**} = x_{**}$ if x_{**} exists and u_{X_L} otherwise, where u_{X_L} denotes the empirical 0.95 quantile from X_L. We label this set $\mathcal{X}$, ordered such that the quan-

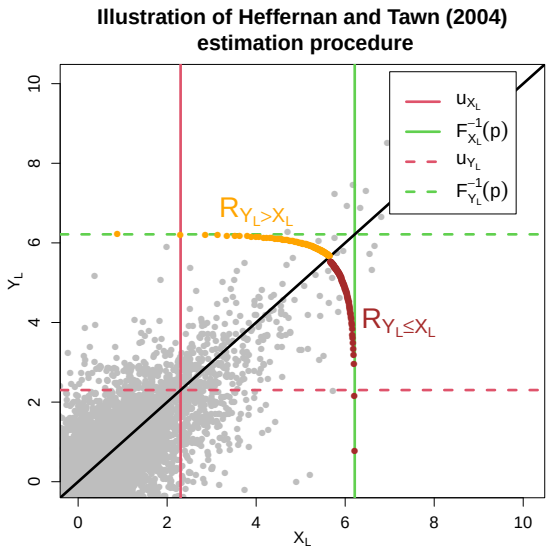

Figure 3.3.1: Illustration of return curve estimation procedure using Heffernan and Tawn (2004) modelling framework. Red and green lines give 0.95 and $(1-p)$-th quantiles for both margins, respectively. The brown and orange points denote the point estimates obtained for the regions $R_{Y_L>X_L}$ and $R_{Y_L\leq X_L}$, respectively.

tiles are increasing, and use the fitted model to obtain quantiles from the conditional distribution $Y_L \mid X_L > x_*$ for each $x_* \in \mathcal{X}$. The resulting coordinate sets from both regions are combined to give an estimate of the return curve over the entire joint support of (X_L, Y_L). An illustration of this procedure is given in Figure 3.3.1. As can be observed, two sets of points estimates are obtained by conditioning on either variable; these sets are then combined to give an estimate of the entire return curve. As a final step, we apply the probability integral transform to transform the curve estimate to standard exponential margins.

We note that the implementation of this model to estimate return curves is more complex than the methods proposed in Section 3.3.4.2 and Cooley et al. (2019). This is due to the fact the model requires a variable to condition on, meaning we have to fit the model twice and provide a technique for joining point estimates from regions $R_{Y_L>X_L}$ and $R_{Y_L\leq X_L}$.

3.3.4.2 Method based on the approach given in Wadsworth and Tawn (2013)

In this section, we propose a novel implementation of the model described in equation (3.2.3) to generate non-parametric return curve estimates. Consider a random vector (X, Y) with standard exponential margins and define a set $\mathcal{W}$ containing equally spaced rays in the interval $[0, 1]$, ordered from lowest to highest. Assuming $|\mathcal{W}|$ is sufficiently large, we are able to evaluate the joint extremal behaviour across the entire region for which at least one variable is extreme. For each $w \in \mathcal{W}$, we use the 95% empirical threshold of the variable $T_w := \min\{X/w, Y/(1-w)\}$ to obtain an estimate of the angular dependence function via the Hill estimator, which we denote $\hat{\lambda}(w)$. For large u, equation (3.2.3) implies that for any $w \in (0, 1)$ and $t > 0$,

$$\Pr\left(T_w > t + u \middle| T_w > u\right) \approx \exp\{-t\hat{\lambda}(w)\}.$$

Estimates of t and u, combined with the rays w, provide estimates of points in $\mathrm{RC}(p)$. We firstly select a small probability $p^* > p$ and estimate u as the $(1-p^*)$-th quantile of T_w, implying $\Pr(T_w > u) = p^*$. One can then estimate the value of $t > 0$ such that $\Pr(T_w > t + u) = p$ since

$$p = \Pr(T_w > t + u) = \Pr(T_w > u) \times \Pr(T_w > t + u \mid T_w > u) = p^* \exp\{-t\hat{\lambda}(w)\},$$

giving the estimate $t = -\frac{1}{\hat{\lambda}(w)} \log(p/p^*)$. Setting $(x, y) := (w(t+u), (1-w)(t+u))$, we have $(x, y) \in \widehat{\mathrm{RC}}(p)$, resulting in a return curve point estimate for each ray $w \in \mathcal{W}$.

The special case of asymptotic dependence, i.e., $\lambda(w) = \max(w, 1-w)$, is discussed in the appendix. In particular, we show that whilst asymptotic dependence corresponds to a boundary case for the Wadsworth and Tawn (2013) modelling framework, this framework can still be used to derive return curve estimates for this class of extremal dependence.

3.4 Uncertainty estimation and diagnostic tool

3.4.1 Capturing uncertainty in return curve estimates

While previous methods for return curve estimation have considered sampling uncertainty, none provide a means to construct interpretable confidence regions and/or 'average' estimates for return curves (Simpson and Wadsworth, 2017; Cooley et al., 2019). Here, we propose a new method for representing uncertainty in return curve estimates that addresses limitations in the existing methods and provides a formal framework for comparing curve estimates from different models where the truth is known.

Our goal is to represent sampling uncertainty in return curve estimates via some type of confidence region at a given significance level $\alpha \in (0, 1)$. Since these curves vary in two dimensions, careful consideration is needed to ensure the resulting region represents α in a straightforward and interpretable manner.

Figure 3.4.1 displays $n = 10000$ datapoints from inverted logistic (Ledford and Tawn, 1997) and asymmetric logistic (Tawn, 1988) copulas on standard exponential margins. The true return curves for $p = 1/10000$ are given in red while the curves estimated using the Wadsworth and Tawn (2013) model are given in green. A representation of sampling uncertainty will help to determine the quality of these estimates. To achieve this, we propose an adaptation of a tool given in Haselsteiner et al. (2019) for representing uncertainty in environmental contour estimates. The novelty in our approach comes from the fact the original tool has not been directly applied for return curve estimation, even though return curves and environmental contours bear many similarities (Haselsteiner et al., 2021). Moreover, no consideration is given to the theoretical justification of the resulting uncertainty representation in the original approach in terms of coverage properties.

On standard exponential margins, the joint support is given by the set $\mathbb{R}^2_+$; we consider a set of angles in the interval $(0, \pi/2)$ and use these angles to divide the $\mathbb{R}^2_+$

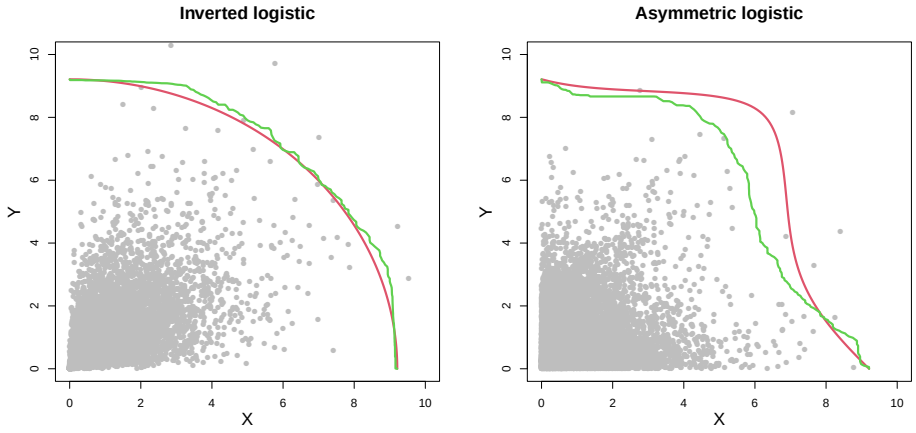

Figure 3.4.1: Estimated (green) vs true (red) curves for data sets simulated from inverted logistic (left) and asymmetric logistic (right) copulas.

plane. Specifically, given some large positive integer m, we define $\boldsymbol{\Theta} := \{\pi(m+1-j)/2(m+1) \mid 1 \leq j \leq m\}$, i.e., a sequence of decreasing angles starting near $\pi/2$ and approaching 0. For each $\theta \in \boldsymbol{\Theta}$, let $L_\theta := \{(x,y) \in \mathbb{R}^2_+ \mid \tan(\theta) = y/x\}$ denote the line segment intersecting the origin with gradient $\tan(\theta) > 0$. For any return curve estimate $\widehat{\mathrm{RC}}(p)$ satisfying properties 3.3.1 - 3.3.4, we have that L_θ intersects $\widehat{\mathrm{RC}}(p)$ exactly once for every $\theta \in \boldsymbol{\Theta}$, implying there is a one-to-one correspondence between angles and points on the estimated curve. An illustrative figure of this correspondence can be found in the supplementary material.

Letting $\{(\hat{x}_\theta, \hat{y}_\theta)\} := \widehat{\mathrm{RC}}(p) \cap L_\theta$, we let $\hat{d}_\theta$ denote the l_2-norm of this point estimate, i.e., $\hat{d}_\theta := (\hat{x}_\theta^2 + \hat{y}_\theta^2)^{1/2}$. Since the angle θ is fixed, this metric represents the aspect of $(\hat{x}_\theta, \hat{y}_\theta)$ that will vary across different curve estimates. Uncertainty in return curve estimates can consequently be quantified using the distribution of $\hat{d}_\theta$ at each angle $\theta \in \boldsymbol{\Theta}$. We propose the following bootstrap procedure: for $k = 1, \ldots, K$,

1. Bootstrap the original data sample to produce a new sample of the same size.

2. For each $\theta \in \boldsymbol{\Theta}$, obtain the l_2-norm for the corresponding point estimate obtained using a given model. Denote this value by $\hat{d}_{\theta,k}$.

If temporal dependence is shown to exist in the data, block bootstrapping (Kunsch,

1989) can be used for this procedure. This sampling scheme retains temporal dependence in the resampled datasets, which ensures the additional uncertainty that arises due to the lower effective sample size is accounted for (Politis and Romano, 1994). It is widely used for quantifying uncertainty in environmental applications of extreme value theory; see, for instance, Eastoe and Tawn (2009), Mhalla et al. (2019b) and Simpson and Wadsworth (2021).

Given $\theta \in \boldsymbol{\Theta}$, we construct empirical estimates of the mean, median, and $100(1-\alpha)\%$ confidence intervals for the l_2-norm values using the sample $\{\hat{d}_{\theta,k} \mid 1 \leq k \leq K\}$. Taking $\alpha = 0.05$, we estimate the 2.5% and 97.5% quantiles using this sample, which we denote $\hat{d}_\theta^{0.025}$ and $\hat{d}_\theta^{0.975}$ respectively. Assuming unbiased estimation, $\Pr(\hat{d}_\theta^{0.025} \leq d_\theta \leq \hat{d}_\theta^{0.975}) \approx 0.95$, where $d_\theta = (x_\theta^2 + y_\theta^2)^{1/2}$ is the l_2-norm of $(x_\theta, y_\theta) \in \mathrm{RC}(p) \cap L_\theta$. Hence, one can show that

$$\Pr\left[(x_\theta, y_\theta) \in \left\{(x,y) \in L_\theta \middle| d_\theta \in [\hat{d}_\theta^{0.025}, \hat{d}_\theta^{0.975}]\right\}\right] \approx 0.95,$$

implying the set $\left\{(x,y) \in L_\theta \middle| d_\theta \in [\hat{d}_\theta^{0.025}, \hat{d}_\theta^{0.975}]\right\}$ defines a confidence region for curve points along the line $y = \tan(\theta)x$. Taking the maximum and minimum x and y coordinates in this set, we obtain a pointwise confidence region for points along the line segment L_θ at each angle $\theta \in \boldsymbol{\Theta}$. These pointwise confidence regions, along with the x and y coordinates corresponding to the mean and median l_2-norm values, can be joined together in order of angle to construct estimates that represent mean, median, and 95% confidence interval estimates for the return curve.

Our procedure is illustrated in Figure 3.4.2 with $m = 150$. The confidence interval width appears to vary over angles in both cases - this is partly explained by implementation of Properties 3.3.1 - 3.3.4 in each bootstrap curve estimate. For the inverted logistic copula, the true curve is captured by the estimated confidence region at all angles. For the asymmetric logistic copula, the estimated confidence region only captures the true curve in certain regions of the $\mathbb{R}_+^2$ plane. This observation indicates some bias may exist for curve estimates from this model. This bias is likely a result of the modelling

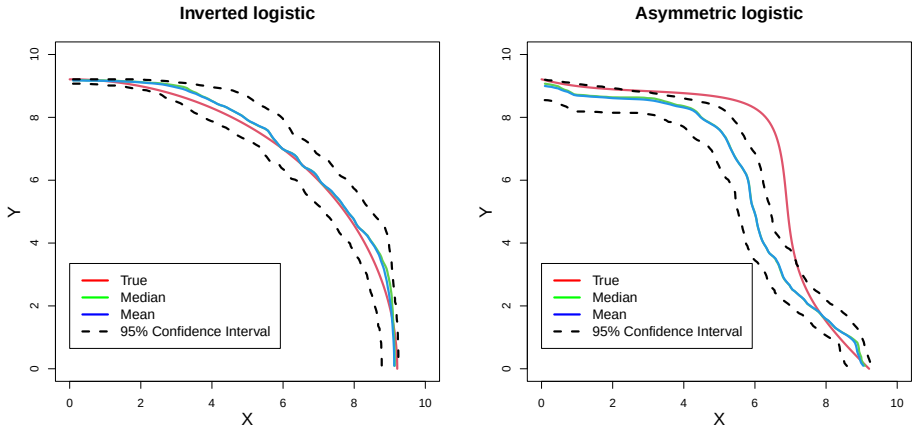

Figure 3.4.2: Comparison of median, median, and 95% confidence interval return curve estimates (green, blue and black dotted lines respectively) to the true return curves (red) for inverted logistic (left) and asymmetric logistic (right) copulas with $K = 1000$ bootstrap samples.

framework being unable to account for the complex asymmetric structure of this copula at finite levels due to a poor rate of convergence to the limiting angular dependence function introduced in equation (3.2.3). Many bivariate extreme value models fail to account for asymmetric dependence structures; see, for instance, Tendijck et al. (2021).

3.4.2 Return curve diagnostic tool

Since the true return curve is unknown in practice, we require a means of evaluating the goodness of fit for a curve estimate, $\widehat{\mathrm{RC}}(p)$, obtained from a particular sample. We propose such a technique and illustrate the method using a single data set simulated from a logistic copula (Gumbel, 1960) on standard exponential margins. This tool provides a means to assess the accuracy of a given curve estimate for a data set with no knowledge of marginal or copula distributions.

Consider the shaded survival regions defined in the left panel Figure 3.4.3 for an estimated return curve $\widehat{\mathrm{RC}}(p)$, where p is small but $\widehat{\mathrm{RC}}(p)$ is in the range of the data. Regions of the form $(x, \infty) \times (y, \infty)$ are illustrated at three different points on the curve. The probability of lying within each such region should, by definition, equal p.

To assess this, we consider fixed survival regions for a chosen subset of points on the estimated curve. For convenience, this subset is chosen such that points correspond to the set of angles $\boldsymbol{\Theta}$; again we take $m = 150$. This results in the set of points sufficing as a representation of the entire estimated curve, as demonstrated in the right panel of Figure 3.4.3. If this estimated curve accurately reflects the true return curve, the empirical probability of observing data within each survival region should be close to p.

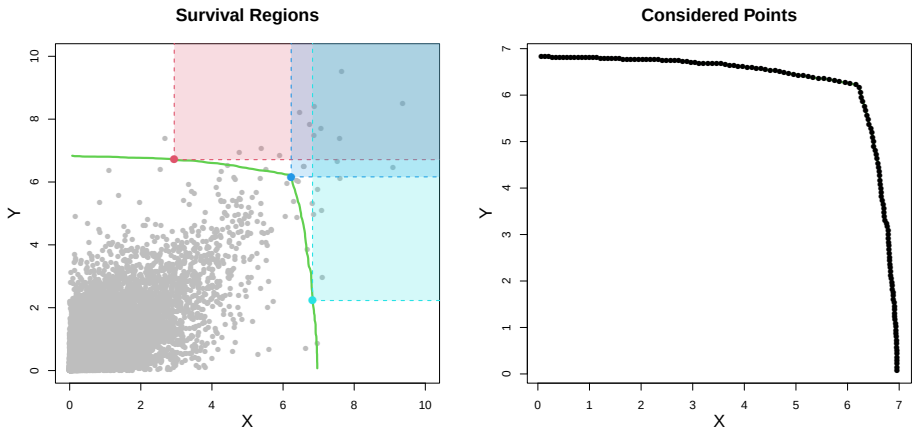

Figure 3.4.3: Left: Survival regions for three points on the estimated return curve. Right: Subset of points on the estimated curve considered in diagnostic.

Let $j \in [1, 150]$ denote the index of an angle $\theta_j \in \boldsymbol{\Theta}$, and let $(\hat{x}_{\theta_j}, \hat{y}_{\theta_j})$ denote the corresponding point on the estimated curve. Furthermore, let $(\mathbf{x}, \mathbf{y}) = \{(x_i, y_i) : 1 \leq i \leq n\}$ denote the observed sample that has been used to estimate the curve. The empirical estimate, which we denote $\hat{p}_j$, is given by the proportion of points lying in the region $(\hat{x}_{\theta_j}, \infty) \times (\hat{y}_{\theta_j}, \infty)$. We then apply the bootstrap to resample the original data set and this estimation procedure is repeated to obtain a range of empirical estimates. As in Section 3.4.1, block bootstrapping should be applied if temporal dependence is shown to exist in the data set. For each j, we let $\hat{\mathcal{P}}_j$ denote the set of empirical probability estimates obtained using bootstrapping. Finally, we estimate the median and 95% pointwise confidence intervals for the probabilities at index j by taking empirical 2.5%,

50% and 97.5% quantiles of the set $\hat{\mathcal{P}}_j$. These estimates provide a pointwise diagnostic at each angle, and can be combined over angles to represent the diagnostic procedure over the whole curve.

This procedure is illustrated in Figure 3.4.4 using the example given in Figure 3.4.3. The black line and shaded regions in the figure represent the empirical estimates of the median and 95% pointwise confidence intervals, respectively, for each index, with the red line denoting the true probability. As can be observed, for all indices, the confidence bounds contain the true value p, suggesting this estimated curve accurately represents this value. However, the median empirical estimates are greater than p at the majority of indices, suggesting a slight overestimation bias for this particular curve estimate.

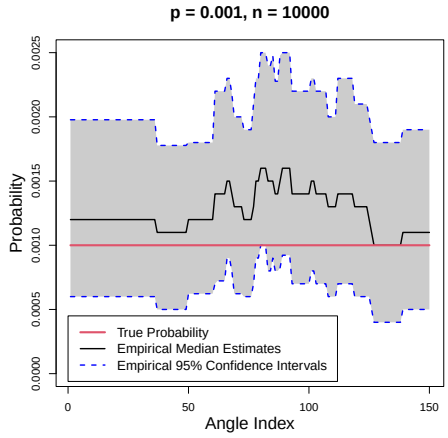

Figure 3.4.4: Illustration of diagnostic tool. Solid red and black lines denotes true and mean empirical estimates, respectively, and grey shaded region between dotted blues lines describe empirical 95% CI estimates.

We note that the confidence intervals produced through bootstrapping for the return curves in Section 3.4.1, and survival probabilities in Section 3.4.2, are all pointwise and dependent across angles. Although they cannot be interpreted across the whole range, they still provide a useful assessment of the utility of various curve estimation techniques. Both tools are adapted in Section 3.6 to account for the original margins of the environmental data sets, allowing us to analyse the quality of return curve estimates for these examples.

3.5 Simulation study

We compare the return curve estimates from the models discussed in Section 3.3.4 to those estimated using the methodology of Cooley et al. (2019). For this, we consider several simulated data sets on standard exponential margins, representing a range of different extremal dependence structures. Specifically, we consider the following copula families: logistic and asymmetric logistic copulas from the bivariate extreme value (BEV) family, the bivariate normal copula with correlation coefficient ρ, logistic and asymmetric logistic copulas from the inverted BEV family, the bivariate t copula with correlation coefficient ρ and degrees of freedom ν and the Frank copula with dependence parameter ζ.

For the methods introduced in Cooley et al. (2019), we transform data to standard Fréchet margins, use these methods to obtain return curve estimates, transform back to standard exponential margins and apply Properties 3.3.1, 3.3.3 and 3.3.4. For each example, the chosen estimation procedure is determined by the extremal dependence exhibited by the underlying copula; this must be specified prior to inference, illustrating a drawback of this approach. We use our knowledge of the true dependence structure to implement the correct procedure, but in practice such knowledge would not be available to us. The code for implementing this approach can be found at `https://www.stat.colostate.edu/~cooleyd/Isolines/`.

Examples of both estimated and true return curves for each copula, with $n = 10000$ and $p = 10^{-3}$, are illustrated in Figure 3.5.1. We note that for the Frank copula, there is a distinct 'linear' segment of the curve estimate from the Heffernan and Tawn (2004) model: this lack of fit arises because the model formulation described in equation (3.2.4) requires us to condition on either X_L or Y_L. This implies we can only evaluate joint tail behaviour in the region where at least one variable is large, i.e., $\{(x, y) \in \mathbb{R}^2 \mid x > u_{X_L} \text{ OR } y > u_{Y_L}\}$. As observed for the Frank copula, part of the true return curve can be defined outside of this region for negatively dependent

data sets; consequently, without additional methodology for approximating the joint survivor function in $\{(x, y) \in \mathbb{R}^2 \mid x \leq u_{X_L} \text{ AND } y \leq u_{Y_L}\}$, this curve region cannot be estimated using equation (3.2.4). This explains the linear segment, since the point estimates for the regions $R_{Y_L > X_L}$ and $R_{Y_L \leq X_L}$ are connected to obtain $\widehat{\text{RC}}(p)$.

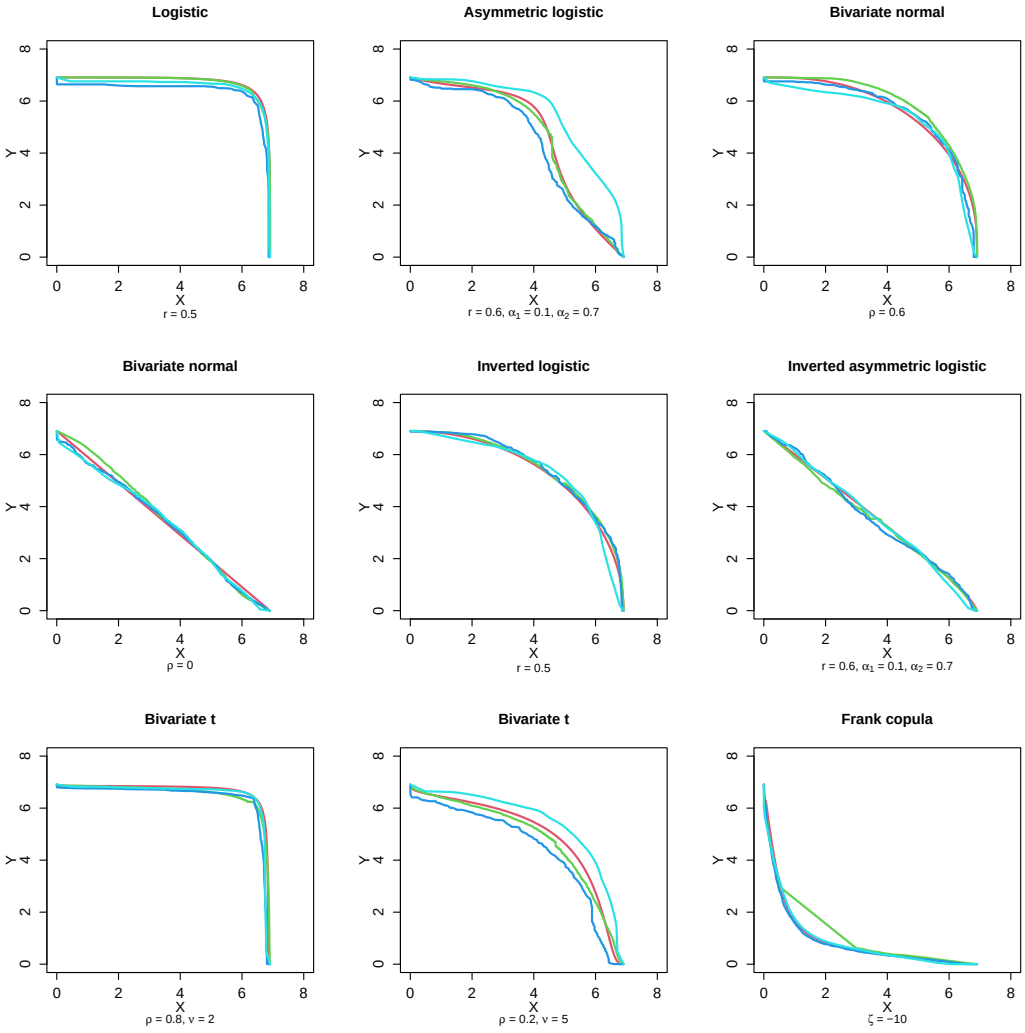

Figure 3.5.1: Examples curve estimates for each copula family. True curves are given in red, while the estimated curves from the Heffernan and Tawn (2004), Wadsworth and Tawn (2013) and Cooley et al. (2019) models are given in green, dark blue and light blue, respectively.

To compare curve estimates, we evaluated bias, computational cost and sampling uncertainty for each of the three curve estimation procedures. To assess bias, 1000 samples of size $n = 100000$ were simulated from each copula and the probabilities

$p = 10^{-3}$ and $p = 10^{-4}$ were considered. Adapting the procedure detailed in Section 3.4.1, median curve estimates for each copula were obtained over the 1000 samples and compared to the corresponding true curves. While a median curve cannot be computed in practice, this measure should provide a summary of the bias that arises from each procedure.

To summarise results, we consider a plot of the l_2-norm values at angles $\theta_j \in \boldsymbol{\Theta}$ against the corresponding indices $j \in [1, m]$ for the true and estimated median curves, with $m = 150$. Numerical methods can then be used to compute the absolute area between the resulting norm curves, with smaller area values corresponding to median curves with less bias. An illustration of this procedure can be found in the supplementary material.

The summary statistics of integrated absolute difference for each copula-model pairing are given in Table 3.5.1. The bias from each procedure appears to vary significantly over the different copula structures, suggesting that the bias in return curve estimates exhibited by a particular model varies with the form of extremal dependence. It is clear that the bias from the Cooley et al. (2019) curve estimates are significantly larger for all but three of the copulas considered; however, we note than in all cases, this method has an unrealistic advantage, namely that the extremal dependence classes have been correctly specified. On the other hand, the models proposed in Heffernan and Tawn (2004) and Wadsworth and Tawn (2013) appear to have similar amounts of bias across the majority of copula structures considered, and neither consistently outperforms the other. Since models for multivariate extremes are typically based on asymptotic arguments which sometimes hold better for one data set than another, this conclusion is most likely a reflection of the different asymptotic arguments for these models.

To evaluate the computational cost of each estimation technique, fifty samples of size $n = 10000$ were simulated from a logistic copula with dependence parameter 0.5. With $p = 10^{-3}$, a Windows machine with a 1.60 GHz Intel(R) Core(TM) i5-8250U

Table 3.5.1: Summary statistics for each model under different copula structures. In each case, 1000 samples of $n = 100000$ datapoints were simulated and the median curves were computed for $p = 10^{-3}$ and 10^{-4}. 'HT', 'WT' and 'CO' correspond to the median curve estimates from the Heffernan and Tawn (2004), Wadsworth and Tawn (2013) and Cooley et al. (2019) models, respectively.

Copula		$p = 10^{-3}$			$p = 10^{-4}$		
	Model	HT	WT	CO	HT	WT	CO
BEV Logistic		1.80	10.83	0.21	2.56	14.07	0.22
BEV Asymmetric Logistic		23.61	28.27	85.50	47.66	78.16	126.25
Bivariate Normal 1		2.49	3.01	25.31	4.07	7.69	37.21
Bivariate Normal 2		0.18	0.14	8.83	0.37	0.22	13.14
Inverted BEV Logistic		2.75	0.39	26.23	4.06	0.56	34.12
Inverted BEV Asymmetric Logistic		0.80	0.19	11.47	1.22	0.30	11.35
Bivariate T 1		8.26	7.17	2.36	11.24	10.44	3.87
Bivariate T 2		14.62	26.37	65.05	39.77	74.95	90.72
Frank		46.39	9.65	5.65	30.33	46.56	29.70

processor and 16GB of RAM was used to compute return curve estimates for each of the fifty samples, and the total computation times were recorded. For the Heffernan and Tawn (2004), Wadsworth and Tawn (2013) and Cooley et al. (2019) techniques, these times were 269.9s, 6.1s, and 2618.2s, respectively.

Application of the Wadsworth and Tawn (2013) modelling framework was significantly quicker than the other two approaches; this is likely due to the fact this technique does not involve any simulation and/or smoothing. Of the remaining estimation frameworks, application of the Heffernan and Tawn (2004) model was still significantly quicker than the method given in Cooley et al. (2019). This conclusion appears to be a result of the Gaussian-kernel density smoothing techniques that are applied when obtaining the empirical curve estimates for the latter approach. Combined with the fact the bias appears significantly lower for the other two models, we choose not consider the approach of Cooley et al. (2019) further.

To assess the sampling uncertainty from the remaining procedures, we computed the coverage for estimated confidence regions at fixed angles. For this, 500 simulated samples of size $n = 10000$ from each copula were considered. Using bootstrapping with $K = 200$ iterations, confidence regions were obtained following the procedure outlined

in Section 3.4.1 for probabilities of $p = 10^{-3}$ and 10^{-4}, and we assessed the coverage of these at five fixed angles $\theta \in \{\pi(m+1-j)/2(m+1) \mid j = 1, 38, 75, 112, 150\}$, allowing assessment of coverage for a variety of regions. Two of the angles are only considered for the BEV asymmetric logistic and inverted BEV asymmetric logistic copulas, since these are the only distributions not to exhibit symmetry. We consider 95% confidence regions for both probabilities. The results for $p = 10^{-3}$ are given in Table 3.5.2; the results for $p = 10^{-4}$ can be found in the supplementary material, along with an visual illustration of the coverage procedure.

Table 3.5.2: Coverage values of 95% confidence regions for $p = 10^{-3}$. 'HT' and 'WT' correspond to the Heffernan and Tawn (2004) and Wadsworth and Tawn (2013) models, respectively.

Copula	Probability	$p = 10^{-3}$									
	Model	HT					WT				
BEV Logistic	Angle	1	2	3	4	5	1	2	3	4	5
	Coverage	0.014	0.970	0.800	*	*	0.856	0.606	0.896	*	*
BEV Asymmetric Logistic	Angle	1	2	3	4	5	1	2	3	4	5
	Coverage	0.756	0.938	0.868	0.860	0.896	0.908	0.054	0.362	0.902	0.936
Bivariate Normal 1	Angle	1	2	3	4	5	1	2	3	4	5
	Coverage	0.358	0.942	0.956	*	*	0.872	0.930	0.912	*	*
Bivariate Normal 2	Angle	1	2	3	4	5	1	2	3	4	5
	Coverage	0.948	0.930	0.955	*	*	0.930	0.934	0.934	*	*
Inverted BEV Logistic	Angle	1	2	3	4	5	1	2	3	4	5
	Coverage	0.780	0.932	0.932	*	*	0.864	0.944	0.942	*	*
Inverted BEV Asymmetric Logistic	Angle	1	2	3	4	5	1	2	3	4	5
	Coverage	0.960	0.929	0.942	0.951	0.922	0.924	0.934	0.958	0.942	0.904
Bivariate T 1	Angle	1	2	3	4	5	1	2	3	4	5
	Coverage	0.504	0.928	0.512	*	*	0.920	0.758	0.884	*	*
Bivariate T 2	Angle	1	2	3	4	5	1	2	3	4	5
	Coverage	0.784	0.896	0.930	*	*	0.938	0.806	0.632	*	*
Frank	Angle	1	2	3	4	5	1	2	3	4	5
	Coverage	0.896	0	0	*	*	0.922	0.692	0.714	*	*

These coverage results provide an insight into differences in the Heffernan and Tawn (2004) and Wadsworth and Tawn (2013) models. Firstly, for angles close to 0 and $\pi/2$, the coverage from the Wadsworth and Tawn (2013) model tends to be closer to the nominal level then that from the Heffernan and Tawn (2004) model. This is especially apparent when examining the scores at both probabilities for the logistic and first bivariate normal copula examples. We note that imposing Property 3.3.1 will affect the coverage near the margins, since we do not allow return curve coordinate estimates that exceed the marginal $(1-p)$-th quantiles, resulting in constrained confidence intervals. We also note that the coverage values for the Frank copula from the Heffernan and Tawn (2004) framework are noticeably small; this relates to aforementioned shortcoming of this approach for data sets with negative dependence. On the other hand, certain coverage values obtained using the Wadsworth and Tawn (2013) approach are noticeably smaller than the corresponding values from the Heffernan and Tawn (2004) approach; for example, for the BEV asymmetric logistic copula at $p = 10^{-4}$ and the second bivariate t copula. On the whole, neither procedure consistently outperforms the other over the copulas and angles considered and encouragingly, the resulting coverage scores were, in many cases, close to the nominal level.

From these results, we suggest that the curve estimation technique derived using the Wadsworth and Tawn (2013) model is preferable in a practical setting; it is straightforward to implement and significantly outperforms the other considered techniques in terms of computation time. Combined with bias and coverage results, alongside the ability of the Wadsworth and Tawn (2013) model to capture negative dependence structures, this curve estimation technique offers clear advantages over the alternative methods, making it the best suited for practical applications.

3.6 Case study

We now apply the techniques proposed in Section 3.3.4 to the two metocean data sets introduced in Section 3.1. We first transform both data sets to standard exponential margins. Assuming each margin is identically distributed over time, we estimate the marginal distributions using equation (3.3.1) and apply the probability integral transform. We then use the techniques proposed in Section 3.3.4 to obtain curve estimates for the probability $p = 10^{-3}$, corresponding to a return period of approximately 4.7 years. The resulting curve estimates are illustrated in Figure 3.6.1 on the original margins (following back transformation). The corresponding parameter estimates for the Heffernan and Tawn (2004) and Wadsworth and Tawn (2013) modelling frameworks are given in the supplementary material.

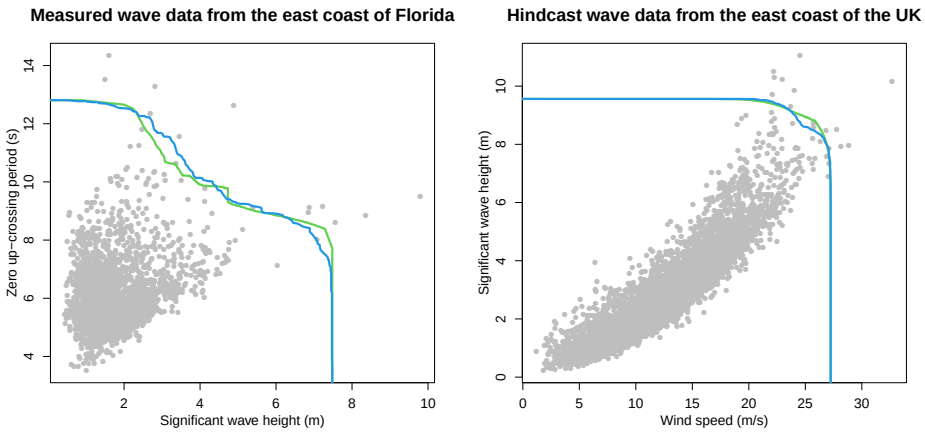

Figure 3.6.1: Return curves estimated for measured (left) and hindcast (right) data sets, with $p = 10^{-3}$. Green and blue lines represent the estimates from the Heffernan and Tawn (2004) and Wadsworth and Tawn (2013) models, respectively.

Next, we adapt the diagnostic tool introduced in Section 3.4 to assess the goodness of fit of return curve estimates for both data sets. To account for the additional uncertainty that arises during estimation of the marginal distributions, we apply the diagnostic on the original margins of the data. This is done as follows: letting $\{(x_i, y_i)\}_{i=1,\ldots,n}$ denote either data set, we define $(x_0, y_0) = (\min_i x_i, \min_i y_i)$. We use these coordinates as a

reference point from which we can evaluate return curve estimates. Given $\boldsymbol{\Theta}$ defined as before, we define the line segment $L'_\theta := \{(x, y) \in \mathbb{R}^2 \mid y = (x - x_0)\tan(\theta) + y_0\}$ for each $\theta \in \boldsymbol{\Theta}$ and, for any curve estimate $\widehat{\mathrm{RC}}(p)$, consider the intersection of the sets $\widehat{\mathrm{RC}}(p) \cap L'_\theta$. Illustrations of reference points, line segments and intersection points are given in the supplementary material for both data sets. Similarly to before, these intersection points are used to define the joint survival regions; the data is then resampled and sets of empirical probability estimates are obtained for each region.

We apply block bootstrapping for resampling, because even with pre-processing, both data sets still appear to exhibit some marginal temporal dependence. Block sizes of 5 and 10 were selected for the measured and hindcast data sets, respectively, by considering plots of the autocorrelation function and selecting values beyond which the dependence appeared insignificant for both variables. These block sizes were then used to bootstrap the original data sets. The resulting diagnostic plots are given in Figures 3.6.2 and 3.6.3.

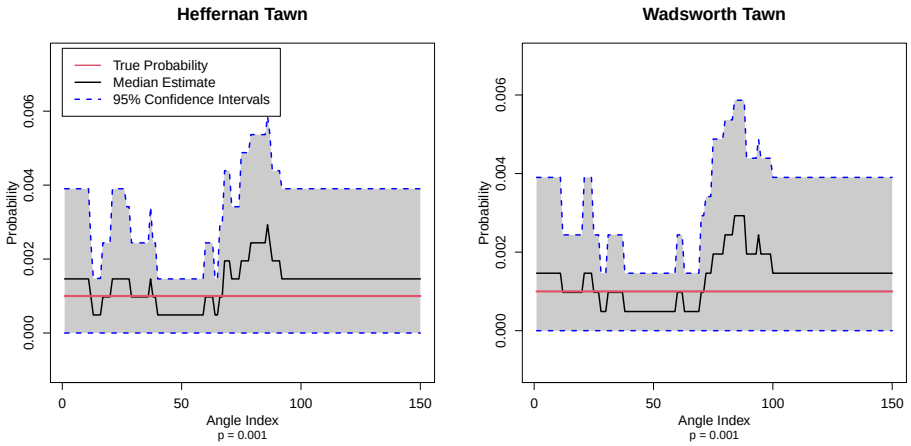

Figure 3.6.2: Measured data diagnostic plots with $K = 1000$ block bootstraps from Heffernan and Tawn (2004) (left) and Wadsworth and Tawn (2013) (right) models, respectively.

For the observed data, Figure 3.6.2 suggests both models perform similarly and provide accurate curve estimates, while for the hindcast data, Figure 3.6.3 suggests the Wadsworth and Tawn (2013) curve estimate outperforms the Heffernan and Tawn

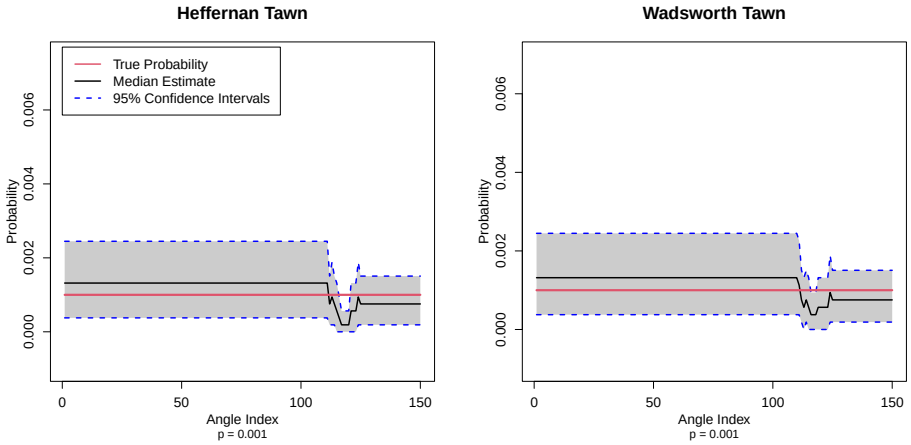

Figure 3.6.3: Hindcast data diagnostic plots with $K = 1000$ block bootstraps from Heffernan and Tawn (2004) (left) and Wadsworth and Tawn (2013) (right) models, respectively.

(2004) estimate at a subset of angles. This is reflected by the difference in curve estimates in the the joint upper tail of the data; see Figure 3.6.1. In both cases, the variability in curve estimates appears to vary with the angle. The estimated confidence intervals capture the true probability at the majority of considered angles, suggesting both estimation techniques can accurately capture joint tail behaviour for these data sets.

Finally, we apply an adaptation of the technique introduced in Section 3.4.1 to assess uncertainty in return curve estimates. Letting $(\hat{x}_\theta, \hat{y}_\theta) := \widehat{\mathrm{RC}}(p) \cap L'_\theta$, with L'_θ defined as before, uncertainty is captured by considering the distribution of l_2-norm values from the reference point, i.e., $\hat{d}_\theta = |(\hat{x}_\theta - x_0)^2 + (\hat{y}_\theta - y_0)^2|^{1/2}$. Using block bootstrapping, with the marginal distributions re-estimated for each bootstrapped sample, we obtain median and mean curve estimates, along with pointwise 95% confidence intervals across angles. For the reasons outlined in Section 3.4.2, we use the Wadsworth and Tawn (2013) approach to obtain these estimates. The resulting curve estimates are illustrated in Figure 3.6.4; we observe that taking mean and median curves appears to have a smoothing effect on the resulting estimates, relative to the original curve estimates given in Figure 3.6.1.

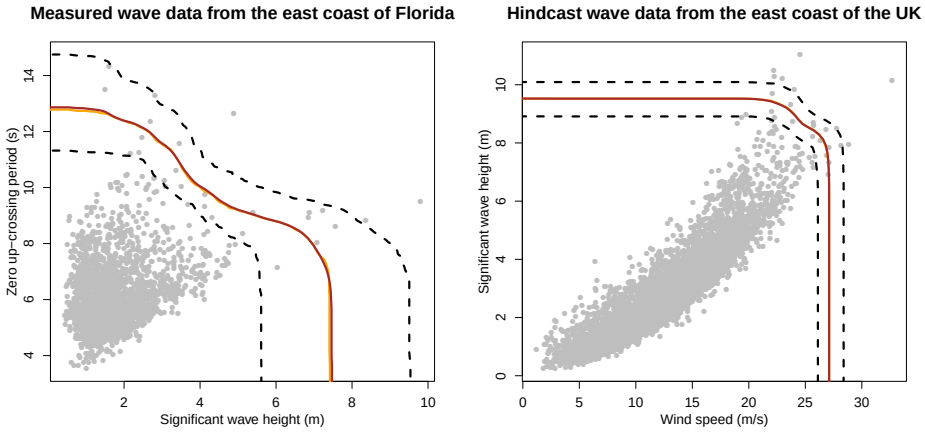

Figure 3.6.4: Median (orange) and mean (brown) curve estimate, along with 95% (black dotted) confidence regions obtained using block bootstrapping with $K = 250$ for the measured (left) and hindcast (right) data sets. The Wadsworth and Tawn (2013) model was used to obtain all estimates.

3.7 Discussion

We have considered the concept of a return curve as a bivariate extension to a return level and introduced novel estimation techniques, illustrating that these methods perform better than an existing approach. Furthermore, unlike Cooley et al. (2019), our methods do not require the form of extremal dependence to be pre-specified: this is an obvious advantage, since determining the extremal dependence structure is seldom straightforward. We have also proposed novel uncertainty representation and diagnostic tools for the risk measure.

For the diagnostic tool proposed in Section 3.4.2, we note that extreme survival region probabilities are estimated empirically, meaning the accuracy of such estimates will be directly related to the sample size n and the probability p. This represents a broader problem within the extremes literature, since, by definition, we will have observed very few extremes values that can be used to verify and justify a given approach. This issue is illustrated further in the supplementary material, where we consider decreasing probabilities for a fixed sample size. However, as with similar analyses, if the tool appears to illustrate a good fit to the data at less extreme probabilities, we can be

more confident in extrapolating to more extreme values.

While we have focused on bivariate random vectors, both concepts and methodology can be extended to the general multivariate setting. However, higher dimensional curves are difficult to visualise and capturing the dependence structures in higher dimensions becomes increasingly complex, since different two dimensional marginals can exhibit different forms of extremal dependence. Nevertheless, in the bivariate setting, we believe return curves are a useful tool for researchers to explore joint extremal behaviour and develop a better understanding of potential risks. Indeed, return curves are already utilised to analyse risks for ocean and coastal structures. It is also important to note this risk measure only denotes the rarity of events, not impact: therefore, in practice, researchers must carefully consider which regions of the multivariate space are impactful prior to inference.

As is common in many environmental contexts, the data sets considered in Section 3.6 both appear to exhibit non-negligible temporal dependence. We account for this feature by using block bootstrapping to quantify uncertainty, but this creates the additional challenge of block size selection. For this, we use an ad-hoc technique based on examinations of ACF plots. An in-depth investigation could improve on this approach through a more robust, theoretically justified resampling scheme.

For the hindcast data set discussed in Section 3.6, Figure 3.6.3 illustrates a downside of applying the diagnostic tools on the original marginal distributions. In particular, the majority of angles selected to represent the curve estimates correspond to the 'marginal limits' of the curves, i.e., the straight line segments connecting the curve estimates to the margins in Figure 3.6.4. Due to the strong positive dependence between the hindcast variables, empirical survival probabilities will be unchanging along these line segments, explaining the largely constant diagnostic probabilities in Figure 3.6.3. This is further demonstrated by the illustration of line segments given in the supplementary material. Future research could explore techniques for selecting angles such that the

corresponding return curve representation is exclusively in-between the aforementioned marginal limits.

Finally, we note that all techniques discussed in this paper are only applicable to data sets exhibiting stationarity; accounting for non-stationarity in the context of return curves presents many challenges, since return curves are defined in the stationary setting only and all models introduced in Section 3.3 assume stationarity. While a range of approaches exist for capturing non-stationary in the univariate setting (e.g. Eastoe, 2019), relatively few approaches exist in the multivariate setting. This topic has recently been explored in Murphy-Barltrop and Wadsworth (2022), whereby the authors extend the definition of return curves to the non-stationary setting and provide techniques for their estimation.

Chapter 4

Modelling non-stationarity in asymptotically independent extremes

4.1 Introduction

Modelling joint tail behaviour of multivariate datasets is important in a wide variety of applications, including nuclear regulation (Office for Nuclear Regulation, 2018), neuroscience (Guerrero et al., 2023) and flood risk analysis (Gouldby et al., 2017). When analysing multivariate extremes, it is important to capture the dependence structure at extreme levels appropriately. In certain applications, one would expect the extremes to occur simultaneously – a situation termed asymptotic dependence – whilst in others, joint occurrence of the very largest events cannot happen – a situation termed asymptotic independence. Section 4.2 explains these concepts in detail. The study of extremal dependence structures is well established, and a wide range of statistical modelling techniques have been proposed (Coles and Tawn, 1991; Ledford and Tawn, 1997; Heffernan and Tawn, 2004).

Extremal dependence between two variables may be summarised by bivariate risk measures. A variety of risk measures have been proposed in the literature (Serinaldi, 2015), and are selected according to the needs of an analysis. For this paper, we restrict attention to one particular measure known as the return curve due to its utilisation in a variety of practical applications (Murphy-Barltrop et al., 2023). Given a small probability p, the p-probability return curve is given by $\text{RC}(p) := \{(x, y) \in \mathbb{R}^2 \mid \Pr(X > x, Y > y) = p\}$, with corresponding return period $1/p$. This curve directly extends the concept of a return level from the univariate framework (Coles, 2001) to the bivariate setting. These curves, which provide a summary of joint tail behaviour, are widely used in practice to derive extremal conditions during the design analysis of many ocean and coastal structures, including oil rigs (Jonathan et al., 2014a), railway lines (Gouldby et al., 2017) and wind turbines (Manuel et al., 2018).

However, in many real world scenarios, datasets exhibit non-stationarity; this feature can result in extremal dependence structures that are not fixed due to covariate influences on the underlying processes. In this setting, there is no longer a meaningful or fixed definition of a return curve. We therefore expand the definition of this risk measure to be covariate-dependent, resulting in a non-stationary counterpart; see Rootzén and Katz (2013) and Serinaldi (2015) for related discussion. Given some covariates $\mathbf{Z}_t,\ t \in \{1, 2, \ldots, n\}$, where t denotes time, let $\{X_t, Y_t\}$ denote a conditionally stationary process, i.e., the distribution of $(X_t, Y_t) \mid \mathbf{Z}_t$ does not depend on t (Caires and Ferreira, 2005). In this setting, we define the p-probability return curve at a covariate realisation $\mathbf{z}_t$ to be $\text{RC}_{\mathbf{z}_t}(p) := \{(x, y) \in \mathbb{R}^2 \mid \Pr(X_t > x, Y_t > y \mid \mathbf{Z}_t = \mathbf{z}_t) = p\}$. Evaluation of $\text{RC}_{\mathbf{z}_t}(p)$ over different values of t allows one to explore joint extremal behaviour over time, and thus may be useful when designing ocean and coastal structures for future climates.

In a practical setting, we wish to derive estimates of non-stationary return curves for environmental datasets to evaluate the changing risk with covariates. Our methodology

is motivated particularly by non-stationarity observed in data obtained from the UK Climate Projections (UKCP18) under emissions scenario RCP 8.5. This corresponds to the 'worst-case' scenario, whereby greenhouse gas emissions continue to rise throughout the 21st century (Met Office Hadley Centre, 2018). As such, data from these projections can be used as a risk management tool to help mitigate against the impacts of climate change in a conservative manner. Specifically, we focus in this work on relative humidity and temperature projections over the summer months (June, July and August) at a grid cell containing the UK's Heysham nuclear power station. Denoting relative humidity by $\mathrm{RH}_t \in [0, 100]$ for $t \in \{1, 2, \ldots, n\}$, we define a 'dryness' variable as $\mathrm{Dr}_t := 100 - \mathrm{RH}_t \in [0, 100]$.

Data are only considered for summer months since extremal dependence structures vary significantly across meteorological seasons and worst extremes tend to occur in summer; see the Supplementary Material for further details. In the context of nuclear safety, both high temperature and high dryness (low humidity) values are independently identified as primary hazards by the UK's Office for Nuclear Regulation (ONR) (Office for Nuclear Regulation, 2018). As a result, pre-set 'design values' of either variable, corresponding to a '1 in 10,000-year' event, are used to inform the design bases of UK-based nuclear sites. Moreover, the combination of high temperature and high dryness values has been identified as a relevant safety consideration (Knochenhauer and Louko, 2004; Office for Nuclear Regulation, 2021), since this combination is characteristic of drought conditions. Such conditions have the potential for catastrophic consequences, including loss of foundation support to facilities and loss of water supply. Therefore, evaluating the joint extremal behaviour for this particular combination of hazards can provide useful information about joint risk over the observation period.

The dataset of temperature and dryness at the start and end of the time period, along with the temperature time series, are plotted in the left and centre panels of Figure 4.1.1, respectively. Clear non-stationary trends can be observed within both

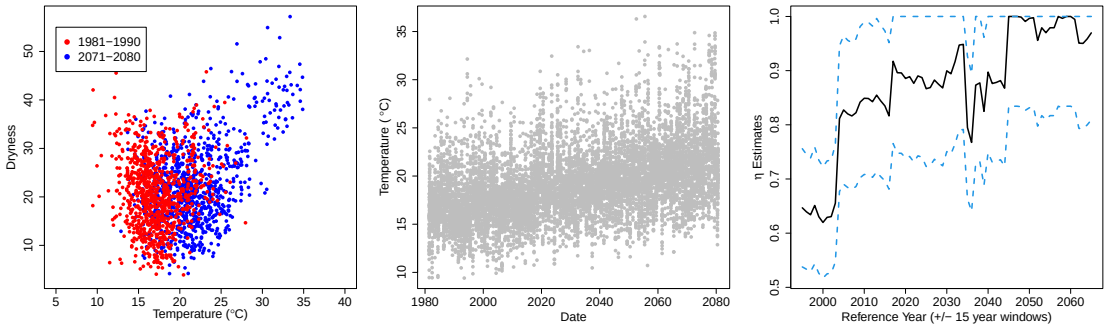

Figure 4.1.1: Left: Plot of first and last 10 years of combined projections, given in red and blue, respectively. Centre: Plot of Heysham temperature time series. Right: Plot of η estimates over rolling windows (solid black lines), alongside 95% pointwise confidence intervals (dotted blue lines).

marginal data sets; these trends are likely a result of seasonal behaviour combined with long term trends due to climate change.

To assess trends in dependence, we can calculate suitable coefficients using rolling windows of data, assuming local stationarity within each window. Rolling windows are defined by taking a reference year $y \in \{1995, 1996, \ldots, 2065\}$ and considering all data for the months June, July and August within the interval $[y - 15, y + 15]$: this results in $2,790$ observations for each window. The right panel of Figure 4.1.1 demonstrates a clear trend in an extremal dependence coefficient labelled η (Ledford and Tawn, 1996); this measure summarises the dependence between the most extreme observations, with larger values corresponding to a higher degree of positive dependence. Further discussion can be found in Section 4.2.2. The illustrated trend suggests the probability of extreme observations occurring simultaneously is increasing over time, motivating the need for modelling techniques that can capture trends of this nature. We return to a detailed analysis of this dataset in Section 4.5.

The majority of existing techniques for modelling multivariate extremes assume stationarity in the joint tail structure. Furthermore, of the approaches that can accommodate non-stationarity, most are suitable only for datasets exhibiting asymptotic dependence, as we discuss in Section 4.2.3. This is restrictive since in practice, asymp-

totic independence is often observed (Ledford and Tawn, 1996, 1997); this is further evidenced by estimated η values for the UKCP18 dataset, which indicate the presence of asymptotic independence, at least throughout most of the observation period.

We propose a new method for capturing non-stationary extremal dependence structures when asymptotic independence is present, based on a non-stationary extension to the Wadsworth and Tawn (2013) modelling framework. In doing so, we are able to evaluate and visualise trends across the entire extremal dependence structure. This is in contrast to other approaches, where implementation may be limited to trends in one-dimensional summary measures, such as the coefficient of tail dependence (Ledford and Tawn, 1996) or the extremal coefficient (Frahm, 2006).

This paper is structured as follows: Section 4.2 recalls existing methodology for capturing tail behaviour in the stationary and non-stationary settings for both univariate and multivariate random vectors. Section 4.3 introduces our modelling approach, which relies on quantile regression techniques to derive two estimators of a quantity describing extremal dependence in a non-stationary setting. We also propose methodology for non-stationary return curve estimation using these estimators. Section 4.4 details a simulation study, showing our estimators to be approximately unbiased over a range of simulated examples. In Section 4.5, we apply our model to the UKCP18 dataset. Our approach is able to reveal clear trends in the extremal dependence of this process, and estimates of return curves are obtained. We conclude in Section 4.6 with a discussion and outlook on future work.

4.2 Background

4.2.1 Univariate extreme value theory

In the univariate setting, one of the most popular techniques for capturing tail behaviour is known as the peaks-over-threshold approach, whereby a generalised Pareto distribu-

tion (GPD) is fitted to all exceedances of some high threshold. This is justified by the Pickands-Balkema-de Haan theorem (Balkema and de Haan, 1974; Pickands, 1975), which states that for a random variable X satisfying certain regularity conditions, there exists a normalising function $c(u)$ such that

$$\Pr\left(\frac{X-u}{c(u)} \leq x \;\middle|\; X > u\right) \to G(x) := 1 - \left\{1 + \frac{\xi x}{\tau}\right\}_+^{-1/\xi}, \quad x > 0, \; (\tau, \xi) \in \mathbb{R}^+ \times \mathbb{R}, \tag{4.2.1}$$

as $u \to x^F := \sup\{x : F(x) < 1\}$; see also Coles (2001). Here, $G(x)$ is the cumulative distribution function of a GPD, with scale and shape parameters, τ and ξ, respectively, and $z_+ = \max(0, z)$. The shape parameter dictates the behaviour of the tail, with $\xi < 0$, $\xi = 0$ and $\xi > 0$ corresponding to bounded, exponential and heavy tails, respectively. In practice, for an observed random variable with a finite sample size, a high threshold u is selected and a GPD is fitted to the positive exceedances: we write $X - u \mid X > u \sim \text{GPD}(\tau, \xi)$.

In many contexts, such as financial and environmental modelling, datasets exhibit non-stationarity, whereby the underlying distribution changes with time or other covariates. In most such cases, we can no longer expect a stationary GPD model to capture the tail adequately. This feature can be present in a range of different forms, as exhibited by the seasonal and long term trends present in the UKCP18 dataset introduced in Section 4.1. Davison and Smith (1990) addressed this issue by using covariates to capture trends in the GPD parameters. Given a non-stationary process $\{Y_t\}$, with covariates $\mathbf{Z}_t$, a non-stationary GPD model is given by

$$(Y_t - u \mid Y_t > u, \; \mathbf{Z}_t = \mathbf{z}_t) \sim \text{GP}(\tau(\mathbf{z}_t), \xi(\mathbf{z}_t)), \tag{4.2.2}$$

for a sufficiently large threshold u. More recent extensions to this model also allow the threshold u to be covariate dependent. For example, Kyselý et al. (2010) and Northrop and Jonathan (2011) use quantile regression to estimate a threshold with a

constant exceedance probability, whereas Sigauke and Bere (2017) use a cubic smoothing spline. More flexible approaches have been proposed using generalised additive models (GAMs) to capture non-stationary behaviour in univariate extremes (Chavez-Demoulin and Davison, 2005; Youngman, 2019). GAMs use smooth functions to capture trends due to covariates, and are less rigid than standard regression models. A wide range of statistical software is available for fitting these models (Wood, 2021).

All the approaches discussed thus far can only be used to model non-stationarity in the extremes. For many statistics corresponding to joint tail behaviour, such as return curves, one must also be able to capture non-stationary within the body of the data simultaneously. This is because extremes of one variable may occur with average values of another variable; see, for instance, the combined projections data for the 1981-1990 period in Figure 4.1.1. To address such challenges, a range of pre-processing techniques have been proposed that allow marginal non-stationarity to be captured in the body and tail of a dataset simultaneously (Nogaj et al., 2007; Eastoe and Tawn, 2009; Mentaschi et al., 2016). For these approaches, covariate functions are used to capture and effectively 'remove' non-stationarity from the body of the data. Once removed, any remaining trends in the tail can be captured using any of the methods introduced above. For a non-stationary process $\{Y_t\}$, with covariates $\mathbf{Z}_t$, the general set-up of these models is to assume

$$(Y_t \mid \mathbf{Z}_t = \mathbf{z}_t) = \mu(\mathbf{z}_t) + \sigma(\mathbf{z}_t)R_t, \tag{4.2.3}$$

with μ and $\log(\sigma)$ as linear functions of covariates. Here, the residual process $\{R_t\}$ is assumed to be approximately stationary, and assigning a distribution to this yields a likelihood for all parameters; Eastoe and Tawn (2009), for example, adopt a standard normal distribution, with the option to also include a shape transformation. Covariate functions are selected through an analysis of non-stationary trends within the body.

An alternative is given by Krock et al. (2021), who propose a single distribution for

capturing non-stationary behaviour in the body and tail simultaneously. This distribution, which extends the model for stationary data proposed in Stein (2021), accounts for both seasonal and long term trends. The idea behind this approach is to provide a 'one size fits all' model, hence the same trends are assumed to be present within the extremes as in the rest of the data. This is unlikely to be the case in all practical scenarios; as a result, we prefer to adopt pre-processing techniques.

4.2.2 Bivariate extreme value theory

We briefly recall approaches to modelling extremes in the bivariate setting. To begin, consider a random vector (X, Y) with respective marginal distribution functions F_X, F_Y. Consider the conditional probability $\chi(u) = \Pr(F_Y(Y) > u \mid F_X(X) > u)$ and define the coefficient $\chi := \lim_{u\to 1} \chi(u) \in [0, 1]$. The cases $\chi = 0$ and $\chi > 0$ correspond to the aforementioned asymptotic independence and asymptotic dependence schemes, respectively. This distinction is important since many models are suitable for data exhibiting one scheme only.

For mathematical simplicity in the description of extremal dependence, it is common to consider random vectors with standardised marginal distributions. This is achieved in practice through marginal estimation and application of the probability integral transform.

Classical modelling approaches are based on the framework of multivariate regular variation, and are applicable only to asymptotically dependent data. Given a random vector (X, Y) with standard Fréchet margins, we define the radial and angular components to be $V := X + Y$ and $W := X/V$, respectively. We say that (X, Y) is multivariate regularly varying if, for all Borel subsets $B \in [0, 1]$, we have

$$\lim_{v\to\infty} \Pr(W \in B, V > sv \mid V > v) = H(B)s^{-1},$$

for any $s > 1$, where H is termed the spectral measure (Resnick, 1987). This assumption implies that, for large radial values, V and W are independent. The spectral measure captures the extremal dependence structure of (X, Y). It must satisfy the moment constraint $\int_0^1 wH(\mathrm{d}w) = 1/2$, but has no closed parametric form. All asymptotically independent distributions have a spectral measure placing mass at the endpoints $\{0\}$ and $\{1\}$ of the unit interval, which is why this modelling framework is unable to capture tail properties under this scheme (Coles et al., 1999). Moreover, it has been shown that assuming the incorrect form of extremal dependence will lead to unsatisfactory extrapolation in the joint tail (Ledford and Tawn, 1997; Heffernan and Tawn, 2004). This has consequently led to the development of flexible modelling approaches that are able to theoretically capture both extremal dependence regimes.

The first such idea was proposed in Ledford and Tawn (1996, 1997). It is assumed that the joint tail of a random vector (X, Y) with standard exponential margins is given by

$$\Pr(X > u, Y > u) = \Pr(\min(X, Y) > u) = L(e^u)e^{-u/\eta} \text{ as } u \to \infty, \tag{4.2.4}$$

where L is a slowly varying function at infinity, i.e., $\lim_{x\to\infty} L(cx)/L(x) = 1$ for $c > 0$, and $\eta \in (0, 1]$. The parameter η is termed the coefficient of tail dependence, with $\eta = 1$ and $\lim_{u\to\infty} L(e^u) > 0$ corresponding to asymptotic dependence and $\eta < 1$, or $\eta = 1$ and $\lim_{u\to\infty} L(e^u) = 0$, corresponding to asymptotic independence. In Figure 4.1.1, our estimates of η suggest asymptotic independence is exhibited by the UKCP18 data throughout most of the observation period. In practice, this framework is limited by the fact it only characterises the joint tail where both variables are large, and hence is not applicable in regions where only one variable is extreme.

Alternative characterisations of the joint tail have been proposed to circumvent this issue. Heffernan and Tawn (2004) introduce a general, regression-based modelling tool for conditional probabilities. Given a random vector (X, Y) with standard Laplace margins (Keef et al., 2013a), it is assumed that normalising functions $a : \mathbb{R} \to \mathbb{R}$ and

$b : \mathbb{R} \to (0, \infty)$ exist such that the following convergence holds:

$$\lim_{u \to \infty} \Pr\left[(Y - a(X))/b(X) \leq z, X - u > x \mid X > u\right] = D(z)e^{-x}, \;\; x > 0,$$

for a non-degenerate distribution function D. Both regimes can be captured via the functions a and b, with asymptotic dependence arising when $a(x) = x$ and $b(x) = 1$. Note that one could instead condition on the event $Y > u$. The functions a and b are typically estimated parametrically, while the distribution function D is estimated non-parametrically. This model has been widely used in practice, with applications ranging from ocean engineering (Ross et al., 2020) to coastal flood mitigation (Gouldby et al., 2017).

Wadsworth and Tawn (2013) provide an alternative representation for the joint tail using a general extension of the framework described in equation (4.2.4). Given (X, Y) with standard exponential margins, they assume that for each $w \in [0, 1]$,

$$\Pr(\min\{X/w, Y/(1-w)\} > u) = L(e^u; w)e^{-\lambda(w)u}, \;\; \lambda(w) \geq \max(w, 1-w), \quad (4.2.5)$$

as $u \to \infty$, where $L(\cdot\,; w)$ is slowly varying for each ray $w \in [0, 1]$ and λ is the termed the angular dependence function (ADF). This function, which describes the dependence structure of the joint tail along the ray w, generalises the coefficient η, with $\eta = 1/\{2\lambda(0.5)\}$. Both extremal dependence regimes can be captured by λ, with asymptotic dependence implying the lower bound $\lambda(w) = \max(w, 1-w)$ for all $w \in [0, 1]$. Pointwise estimates of the ADF can be obtained in practice via the Hill estimator (Hill, 1975). Moreover, λ captures the joint tail behaviour of a wide range of data structures (Wadsworth and Tawn, 2013), and thus equation (4.2.5) provides a flexible modelling framework for bivariate extremes.

Alongside these approaches, we note that there exist several copula-based models that can theoretically capture both extremal dependence regimes, such as those given

in Coles and Pauli (2002), Wadsworth et al. (2017) and Huser and Wadsworth (2019). However, due to the stronger assumptions about the form of parametric family for the bivariate distribution, we prefer instead to use more flexible modelling techniques.

4.2.3 Non-stationary extremal dependence

Although many extreme value analyses seek to capture marginal non-stationarity, common practice is to assume stationarity in dependence, often without even assessing this feature. Relatively little consideration has been given to this problem in the literature, and most of the approaches that do exist rely on the multivariate regular variation framework, thereby being restricted to asymptotically dependent data. For example, Mhalla et al. (2017) and Mhalla et al. (2019b) propose semi-parametric models to capture trends in parameters of quantities related to the spectral measure, while de Carvalho and Davison (2014), Castro-Camilo et al. (2018) and Mhalla et al. (2019a) propose flexible modelling techniques for capturing non-stationary trends in the spectral measure under covariate influence.

Mhalla et al. (2019b) also propose a technique for data exhibiting asymptotic independence, using GAMs to capture trends in the non-stationary extension to the ADF defined in equation (4.3.1). Given a non-stationary process $\{X_t, Y_t\}$ with standard exponential margins and an external g-dimensional covariate $\mathbf{Z}_t$, the extended ADF $\lambda(\cdot : \mathbf{Z}_t = \mathbf{z}_t)$ is assumed to take the semi-parametric form

$$\lambda(w : \mathbf{Z}_t = \mathbf{z}_t, \boldsymbol{\phi}) = h_w^{-1}\left\{(\mathbf{z}_t^1)'\boldsymbol{\gamma} + \sum_{a=1}^{A} s^a(\mathbf{z}_{t,a}^2)\right\}, \tag{4.2.6}$$

where h_w is a link function, and $\mathbf{z}_t^1, \mathbf{z}_{t,1}^2, \ldots, \mathbf{z}_{t,A}^2$ are subvectors of $\mathbf{z}_t$, or products of covariates if interactions are considered. The vector $\boldsymbol{\gamma}$ gathers linear coefficients whereas s^a denote smooth functions for each $a \leq A$; the parameter vector $\boldsymbol{\phi}$ represents all parameters to be estimated, i.e., $\boldsymbol{\gamma}$ and the coefficients from each basis function s^a.

Estimation of $\boldsymbol{\phi}$ is via a penalised log-likelihood approach, where roughness penalties are imposed to ensure a smooth model fit. The link function $h_w(x) = \log[\{x - \max(w, 1 - w)\}/(1-x)]$ is used, resulting in fitted values contained in the interval $[\max(w, 1-w), 1]$. However, this range is restrictive since $\lambda(w) \leq 1$ implies positive extremal association. In practice, Mhalla et al. (2019b) only applied model (4.2.6) along the ray $w = 1/2$, corresponding to modelling non-stationarity in η only.

Non-stationary extensions to the Heffernan and Tawn (2004) model also exist: Jonathan et al. (2014b) propose smooth covariate functions for a and b, while Guerrero et al. (2023) allow these parameters to vary smoothly over time for blocks of observations via a penalised log-likelihood. However, we note that conditional extremes techniques have been shown to create additional complexities during implementation, requiring more steps compared to alternative approaches because of the need to condition on each variable being extreme separately; see Murphy-Barltrop et al. (2023). Our proposed method is simpler to implement in practice compared to the those derived under this framework.

4.3 Non-stationary angular dependence function

4.3.1 Introduction

We describe a non-stationary extension to the ADF λ of Wadsworth and Tawn (2013), which is the key building block for estimating non-stationary return curves. We assume stationary marginal distributions throughout this section, allowing us to separate out the two forms of trends; further discussion on the separate treatment of these trends can be found in Section 4.5.

Let $\{X_t, Y_t\}$ denote a non-stationary process with stationary margins and standard exponential marginal distributions. If this is not the case in practice, standard exponential margins can be obtained by first fitting non-stationary marginal distributions, such

as those described in Section 4.2.1, and then applying the probability integral transform. Given an external g-dimensional covariate $\mathbf{Z}_t$, we assume that for all $w \in [0,1]$ and $t \in \{1, 2, \ldots, n\}$,

$$\Pr\left(\min\left\{\frac{X_t}{w}, \frac{Y_t}{1-w}\right\} > u \mid \mathbf{Z}_t = \mathbf{z}_t\right) = L(e^u \mid w, \mathbf{Z}_t = \mathbf{z}_t)e^{-\lambda(w|\mathbf{Z}_t=\mathbf{z}_t)u} \text{ as } u \to \infty, \tag{4.3.1}$$

where L denotes a slowly varying function and $\lambda(\cdot \mid \mathbf{Z}_t = \mathbf{z}_t)$ denotes the non-stationary counterpart of the ADF at time t. This amounts to assuming that the joint tail of $(X_t, Y_t) \mid \mathbf{Z}_t$ can be captured by equation (4.2.5) for all $t \in \{1, 2, \ldots, n\}$: this seems reasonable, given the flexibility of the framework outlined in Wadsworth and Tawn (2013).

Define $K_{w,t} := \min\{X_t/w, Y_t/(1-w)\}$: we refer to this variable as the min-projection. Equation (4.3.1) implies that, for each $w \in [0,1]$, and $t \leq n$,

$$\Pr\left(K_{w,t} > v + u \middle| K_{w,t} > u, \mathbf{Z}_t = \mathbf{z}_t\right) \to \exp\{-v\lambda(w \mid \mathbf{Z}_t = \mathbf{z}_t)\} \text{ as } u \to \infty, \quad v > 0.$$

However, unlike its stationary counterpart, the non-stationary ADF cannot be estimated via the Hill estimator; this is because we typically do not have repeated observations for a covariate realisation. Moreover, even with repeated observations, the resulting sample sizes would typically be too small for reliable estimation. As such, a new estimation procedure is required for this function. Given $w \in [0,1]$ and two quantiles q_1, q_2 close to one with $q_1 < q_2 < 1$, we assume the existence of positive sequences $\{u_{w,t}\}_{t\leq n}$ and $\{v_{w,t}\}_{t\leq n}$ such that

$$\Pr\left(K_{w,t} \leq u_{w,t} \mid \mathbf{Z}_t = \mathbf{z}_t\right) = q_1, \quad \Pr\left(K_{w,t} \leq u_{w,t} + v_{w,t} \mid \mathbf{Z}_t = \mathbf{z}_t\right) = q_2, \tag{4.3.2}$$

for all $t \leq n$. Assuming strict monotonicity of the cumulative distribution function for $K_{w,t} \mid (\mathbf{Z}_t = \mathbf{z}_t)$, we deduce that $v_{w,t} > 0$ for all $t \leq n$. Furthermore, the quantile

q_1 being close to one implies values of the sequence $\{u_{w,t}\}_{t\leq n}$ are large in magnitude. Under the model assumptions, we can therefore deduce that

$$\frac{1-q_2}{1-q_1} = \Pr\left(K_{w,t} > v_{w,t} + u_{w,t} \middle| K_{w,t} > u_{w,t},\ \mathbf{Z}_t = \mathbf{z}_t\right) \approx \exp\{-v_{w,t}\lambda(w \mid \mathbf{Z}_t = \mathbf{z}_t)\},$$

which is rearranged to give

$$\lambda(w \mid \mathbf{Z}_t = \mathbf{z}_t) \approx -\frac{1}{v_{w,t}} \log\left(\frac{1-q_2}{1-q_1}\right), \tag{4.3.3}$$

for all $t \leq n$. Hence, estimates of the sequence $\{v_{w,t}\}_{t\leq n}$ lead to a point-wise estimator for the non-stationary ADF at a given angle $w \in [0,1]$. We denote this estimator by $\hat{\lambda}(\cdot \mid \mathbf{z}_t)$, and describe improvements to its stability in Section 4.3.3.

4.3.2 Estimating quantiles of the min-projection

The sequence $\{v_{w,t}\}_{t\leq n}$ corresponds to differences in covariate-varying quantiles of the univariate min-projection $K_{w,t} \mid (\mathbf{Z}_t = \mathbf{z}_t)$ for each $w \in [0,1]$. Quantile regression methods therefore provide a natural solution to the problem of its estimation. Such techniques have successfully been applied in a variety of contexts, ranging from ecology (Cade and Noon, 2003) to growth charts (Wei et al., 2006). Here, we describe the most commonly-used approach in terms of the min-projection variable. Given a value $q \in (0,1)$, the q-th quantile of $K_{w,t} \mid (\mathbf{Z}_t = \mathbf{z}_t) \sim F_{K_{w,t}|\mathbf{z}_t}$ is

$$Q_{K_{w,t}|(\mathbf{Z}_t=\mathbf{z}_t)}(q) = \inf\{x : F_{K_{w,t}|\mathbf{z}_t}(x \mid \mathbf{z}_t) \geq q\}.$$

We assume that the conditional quantile function is linear in $\mathbf{z}_t$, implying $Q_{K_{w,t}|(\mathbf{Z}_t=\mathbf{z}_t)}(q) = \mathbf{z}_t'\boldsymbol{\pi}$, where $\boldsymbol{\pi} \in \mathbb{R}^g$ denotes a vector of coefficients. The vector $\boldsymbol{\pi}$ is estimated through a minimisation of a suitable loss function; see Koenker et al. (2017) for further details.

There is also a range of literature available on the topic of extremal quantile re-

gression (Youngman, 2019; Velthoen et al., 2021), applicable when q is very close to one. However, many of the derived extreme value laws involve unknown constants which cannot be estimated without strong modelling assumptions (Chernozhukov et al., 2017). Furthermore, applying the simple 'rule-of-thumb' proposed in Chernozhukov and Fernández-Val (2011) for deciding between extreme value and normal quantile approximations, the quantile levels considered in Sections 4.4 and 4.5 would not be deemed extreme enough to warrant the former. Moreover, we have found standard quantile regression techniques are able to produce accurate estimates of non-stationary ADFs across a range of simulated examples.

4.3.3 Averaging over quantiles

Prior to applying the proposed model, one must first select q_1 and q_2 for estimating quantiles of $K_{w,t}$. This selection represents a bias-variance trade off, as is often observed in applications of extreme value theory: quantiles that are not sufficiently extreme (close to one) will induce bias in results, while quantiles that are too large will result in highly variable estimates. Moreover, considering only a single pair of quantiles will lead to higher variability in ADF estimates. To address these issues, we consider a range of quantile pairs simultaneously and compute an average estimator over these values. Specifically, let $\{(q_{1,j}, q_{2,j}) \mid 1 \leq j \leq m\}$ be quantiles near one, with $q_{1,j} < q_{2,j} < 1$ for $j = 1, \ldots, m$. For each j, the pair $(q_{1,j}, q_{2,j})$ is used to derive an estimator $\hat{\lambda}_j$, as in equation (4.3.3). Our final estimator is derived to be the average of these:

$$\bar{\lambda}_{QR}(w \mid \mathbf{z}_t) := \frac{1}{m} \sum_{j=1}^{m} \hat{\lambda}_j(w \mid \mathbf{z}_t),$$

for all $w \in [0, 1]$ and $t \leq n$. In unreported simulations, we have found this aggregated estimator to outperform estimators obtained from any individual pair of quantiles considered. Our choices for m and $\{(q_{1,j}, q_{2,j})\}$ are detailed in Section 4.4.3.

4.3.4 Bernstein-Bézier polynomial smooth estimator

One drawback of the average estimator $\bar{\lambda}_{QR}$ proposed in Section 4.3.3 is that it is pointwise for each ray $w \in [0,1]$. This typically leads to non-smooth estimates of the ADF that one would not expect to observe in practice. We therefore extend this estimator to give smooth functional estimates using a parametric family derived from the set of Bernstein-Bézier polynomials. These polynomials have been applied in many approaches to estimate Pickands' dependence function (Guillotte and Perron, 2016; Marcon et al., 2016, 2017b), a quantity related to the spectral measure which bears many similarities to the ADF (Wadsworth and Tawn, 2013). For the interval $[0,1]$, the family of Bernstein-Bézier polynomials of degree $k \in \mathbb{N}$ is defined to be

$$\mathcal{B}_k = \left\{ \sum_{i=0}^{k} \alpha_i \binom{k}{i} w^i (1-w)^{k-i} \ : \ \boldsymbol{\alpha} \in [0,1]^{k+1}, \ w \in [0,1] \right\}.$$

We use this family of functions to approximate $\lambda(\cdot \mid \mathbf{Z}_t = \mathbf{z}_t)$. However, for any $f \in \mathcal{B}_k$, we have $f(w) \leq 1$ for all $w \in [0,1]$. As such, this family of polynomials can only approximate ADFs representing non-negative dependence in the extremes. Furthermore, we wish to allow for covariate influence in the dependence structure; this corresponds to covariate influence in the coefficient vector $\boldsymbol{\alpha}$. We therefore propose extending this family of polynomials to the following set:

$$\mathcal{B}_k^* \mid (\mathbf{Z}_t = \mathbf{z}_t) = \left\{ \sum_{i=0}^{k} \beta_i(\mathbf{z}_t) \binom{k}{i} w^i (1-w)^{k-i} \ : \ \boldsymbol{\beta}(\mathbf{z}_t) \in [0,\infty)^{k+1}, \ w \in [0,1] \right\},$$

where $\beta_i : \mathbb{R}^g \to [0,\infty)$ denote functions of the covariates. For any $t \leq n$, let $\lambda_{BP}(\cdot \mid \mathbf{z}_t) \in \mathcal{B}_k^* \mid (\mathbf{Z}_t = \mathbf{z}_t)$ represent a form of the non-stationary ADF given by this family of functions. Our objective is to find an estimator $\bar{\lambda}_{BP}$ that minimises the equation

$$|\lambda(w \mid \mathbf{Z}_t = \mathbf{z}_t) - \lambda_{BP}(w \mid \mathbf{z}_t)| \tag{4.3.4}$$

over all rays $w \in [0,1]$ and $\mathbf{z}_t$ for $t \leq n$; this is achieved through estimation of the coefficient functions β_i. Since λ is unobserved in practice, we consider the objective function

$$S(\boldsymbol{\theta}) := \frac{1}{|\mathcal{W}|n} \sum_{w \in \mathcal{W}} \sum_{t=1}^{n} \left| \bar{\lambda}_{QR}(w \mid \mathbf{z}_t) - \lambda_{BP}(w \mid \mathbf{z}_t, \boldsymbol{\theta}) \right|,$$

with $\mathcal{W} := \{0, 0.01, 0.02, \ldots, 0.99, 1\}$ defining a finite set spanning the interval $[0,1]$ and $\boldsymbol{\theta}$ denoting the parameter vector corresponding to the coefficient functions $\beta_0, \beta_1, \ldots, \beta_k$. The intuition here is that $S(\boldsymbol{\theta})$ gives an approximation of the absolute value in (4.3.4) integrated over w and t: it is therefore desirable to find a value of $\boldsymbol{\theta}$ which minimises S.

To estimate $\boldsymbol{\theta}$, we must specify the form of coefficient functions. To start, we impose that $\beta_0(\mathbf{z}_t) = \beta_k(\mathbf{z}_t) = 1$ for all $t \leq n$; any function $f \in \mathcal{B}_k^* \mid (\mathbf{Z}_t = \mathbf{z}_t)$ satisfying these conditions has the property that $f(0) = f(1) = 1$, corresponding to the theoretical end-points of the ADF: $\lambda(0) = \lambda(1) = 1$. For $i \in \{1, 2, \ldots, k-1\}$, we assume that $\beta_i(\mathbf{z}_t) = h(\mathbf{z}_t' \boldsymbol{\theta}_i)$, where $h : \mathbb{R} \to [0, \infty)$ denotes a link-function and $\boldsymbol{\theta}_i \in \mathbb{R}^g$ denotes a vector of coefficients for each i. The entire parameter vector is therefore $\boldsymbol{\theta} := \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_{k-1}\}$, with an estimator defined as

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \mathbb{R}^{g(k-2)}}{\arg\min}\, S(\boldsymbol{\theta}).$$

Finally, a smooth estimator of $\lambda(\cdot \mid \mathbf{Z}_t = \mathbf{z}_t)$ is given by $\bar{\lambda}_{BP}(\cdot \mid \mathbf{z}_t) := \lambda_{BP}(\cdot \mid \mathbf{z}_t, \boldsymbol{\theta} = \hat{\boldsymbol{\theta}})$.

4.3.5 Incorporating theoretical properties

Wadsworth and Tawn (2013) show that the ADF is bounded from below, i.e, $\lambda(w) \geq \max(w, 1-w)$ for all $w \in [0,1]$. This bound corresponds to asymptotic dependence, with larger values corresponding to a form of asymptotic independence. Neither of our estimators $\bar{\lambda}_{QR}$ or $\bar{\lambda}_{BP}$ are required to satisfy this constraint, and so we impose this

bound retrospectively. In each case, we define $\bar{\lambda}^*_{QR}$ or $\bar{\lambda}^*_{BP}$ as

$$\bar{\lambda}^*_{QR}(w \mid \mathbf{z}_t) = \max\left\{\bar{\lambda}_{QR}(w \mid \mathbf{z}_t), \max(w, 1-w)\right\},$$

for all $w \in [0,1]$, and similarly for $\bar{\lambda}^*_{BP}$. In unreported results, we found that incorporating this lower bound improves estimation quality within the resulting ADF estimates.

For $\bar{\lambda}^*_{QR}$, we also impose end-point conditions retrospectively: $\bar{\lambda}^*_{QR}(0 \mid \mathbf{Z}_t = \mathbf{z}_t) = \bar{\lambda}^*_{QR}(1 \mid \mathbf{Z}_t = \mathbf{z}_t) = 1$ for all $t \leq n$. Recall that this requirement is already satisfied by the smooth estimator by setting $\beta_0(\mathbf{z}_t) = \beta_k(\mathbf{z}_t) = 1$ for all $t \leq n$.

4.3.6 Estimating non-stationary return curves

We now consider the problem of estimating $\mathrm{RC}_{\mathbf{z}_t}(p)$ at some fixed $t \leq n$ using an estimator of the non-stationary ADF. Let λ^* denote either the estimator $\bar{\lambda}^*_{QR}$ or $\bar{\lambda}^*_{BP}$. Given the set of rays $\mathcal{W}$ defined in Section 4.3.4 and a quantile q_1 close to one, we let $\{u_{w,t}\}_{w\in\mathcal{W}}$ be defined as in equation (4.3.2). Then, for all $w \in \mathcal{W}$, define $\{r_{w,t}\}_{w\in\mathcal{W}}$ as

$$r_{w,t} := -\frac{1}{\lambda^*(w \mid \mathbf{z}_t)} \log\left(\frac{p}{1-q_1}\right),$$

implying $p/(1-q_1) = \exp\{-r_{w,t}\lambda^*(w \mid \mathbf{z}_t)\} \approx \exp\{-r_{w,t}\lambda(w \mid \mathbf{Z}_t = \mathbf{z}_t)\}$. Define $(x_{w,t}, y_{w,t}) := (w(r_{w,t}+u_{w,t}), (1-w)(r_{w,t}+u_{w,t}))$. We have

$$\begin{aligned}
\Pr(X_t > x_{w,t}, Y_t > y_{w,t} \mid \mathbf{Z}_t = \mathbf{z}_t) &= \Pr(K_{w,t} > r_{w,t} + u_{w,t} \mid \mathbf{Z}_t = \mathbf{z}_t) \\
&= \Pr(K_{w,t} > r_{w,t} + u_{w,t} \mid K_{w,t} > u_{w,t}, \mathbf{Z}_t = \mathbf{z}_t) \\
&\quad \times \Pr(K_{w,t} > u_{w,t} \mid \mathbf{Z}_t = \mathbf{z}_t) \\
&\approx \exp\{-r_{w,t}\lambda^*(w \mid \mathbf{z}_t)\} \Pr(K_{w,t} > u_{w,t} \mid \mathbf{Z}_t = \mathbf{z}_t) \\
&= \frac{p}{1-q_1} \times 1 - q_1 = p,
\end{aligned}$$

meaning that the set $\{(x_{w,t}, y_{w,t})\}_{w \in \mathcal{W}}$ provides an approximation of $\text{RC}_{\mathbf{z}_t}(p)$. Similarly to the estimation of $\bar{\lambda}^*_{QR}$ and $\bar{\lambda}^*_{BP}$, as described in Sections 4.3.3 and 4.3.4, we denote $\widehat{\text{RC}}^j_{\mathbf{z}_t}(p) = \{(x^j_{w,t}, y^j_{w,t})\}_{w \in \mathcal{W}}$ for each quantile $q_{1,j}$ close to one, and take our final estimator of the return curve to be $\overline{\text{RC}}_{\mathbf{z}_t}(p) = \{(\sum_{j=1}^m x^j_{w,t}/m, \sum_{j=1}^m y^j_{w,t}/m)\}_{w \in \mathcal{W}}$.

4.4 Simulation study

4.4.1 Overview

We use simulation to evaluate the properties of the estimators $\bar{\lambda}^*_{QR}$ and $\bar{\lambda}^*_{BP}$ derived in Section 4.3. Section 4.4.2 introduces a range of examples exhibiting non-stationary extremal dependence. The variation in dependence structures allows us to assess the relative strengths and weaknesses of each estimator. In Section 4.4.3, we consider parametric forms for the covariate and link functions introduced in Section 4.3. In Section 4.4.4, we evaluate the bias and variability that arises from $\bar{\lambda}^*_{QR}$ and $\bar{\lambda}^*_{BP}$, which we find to outperform the corresponding estimator from Mhalla et al. (2019b). Finally, in Section 4.4.5, we illustrate that our proposed estimators can be used to derive accurate estimates of non-stationary return curves across each of the considered examples.

4.4.2 Simulated examples of non-stationary dependence structures

We introduce several examples exhibiting non-stationary extremal dependence under asymptotic independence. In each case, the non-stationarity is over the time covariate $t \in \{1, 2, \ldots, n\}$, with $n = 10,000$ denoting the sample size. The first two examples are obtained using the bivariate normal copula, for which the dependence is characterised by the coefficient $\rho \in [-1, 1]$. For the first example, we take $\rho(t) = (t-1)/(n-1)$, so that $\rho(1) = 0$ and $\rho(n) = 1$, i.e., moving from independence to perfect positive dependence.

For the second example, we take $\rho(t) = -0.9 + 0.9(t-1)/(n-1)$, giving $\rho(1) = -0.9$ and $\rho(n) = 0$, i.e., moving from a strong negative correlation to independence.

For the third, fourth and fifth examples, we use the inverted extreme value copula (Ledford and Tawn, 1997) with logistic, asymmetric logistic and Hüsler-Reiss families, respectively. For the logistic and asymmetric logistic, the dependence is characterised by the parameter $r \in (0, 1)$, with the degree of positive dependence increasing at r approaches 0. We take $r(t) = 0.01 + 0.98(t-1)/(n-1)$, hence moving from strong positive dependence at $t = 1$ to close to independence at $t = n$. The asymmetric logistic distribution also requires two asymmetry parameters $(\kappa_1, \kappa_2) \in [0, 1]^2$ (Tawn, 1988): we fix $(\kappa_1, \kappa_2) = (0.3, 0.7)$, noting this does not change the overall trend in dependence. The Hüsler-Reiss family is characterised by the dependence parameter $s > 0$, with independence and complete dependence obtained as s approaches 0 and ∞, respectively. We take $s(t) = 0.01 + 9.99(t-1)/(n-1)$, resulting in increasing dependence over time.

For the final example, we start with a specified ADF and use a method given in Nolde and Wadsworth (2022) to construct a copula with this ADF. Given the dependence parameter $c \in (0, 1)$, we take $\lambda(w) = \max\{(2w-1)/c, (1-2w)/c, 1/(2-c)\}$, which is the ADF of the density proportional to

$$\exp(-\max\{(x-y)/c, (y-x)/c, (x+y)/(2-c)\}), \tag{4.4.1}$$

and simulate from this density using MCMC. Such a distribution does not have exactly exponential margins, however in this case the transformation to exponential margins, via the probability integral transform, yields a density with the same ADF. We refer to this example as the copula of model (4.4.1) henceforth and set $c(t) = 0.1 + 0.8(t-1)/(n-1)$; this results in a similar dependence trend to the inverted logistic example. Illustrations of the resulting ADFs over time for each example are given in Figure 4.4.1.

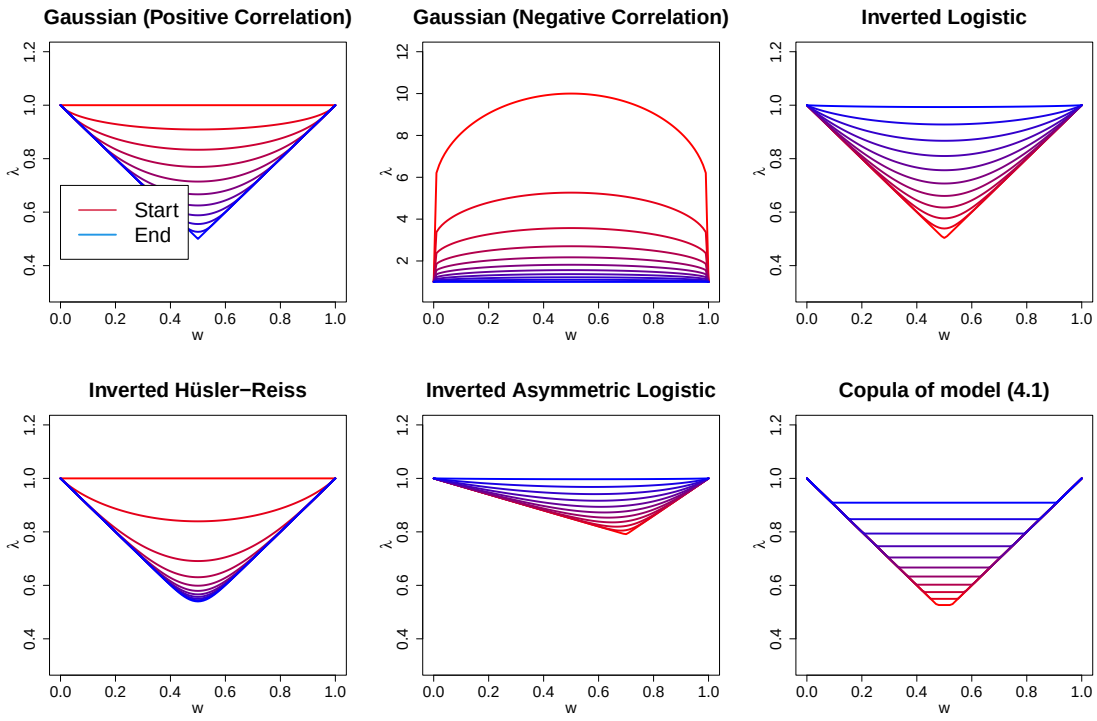

Figure 4.4.1: Illustration of true ADFs over time for each copula example. Colour change is used to illustrate trends in extremal dependence structure over the time frame, with red and blue corresponding to the start and end of time frame, respectively.

4.4.3 Covariate and link functions

Both estimators derived in Section 4.3 require estimates of sequences obtained via a quantile regression procedure. For this, one must first specify the functional form for the relationship between quantiles and covariates $\mathbf{z}_t$. Since the data we have simulated has a dependence structure directly related to the covariate t, we propose the covariate set $\mathbf{z}_t := \{1, t, t^2, t^3\}$. For any given quantile q and ray $w \in [0, 1]$, it is assumed that the quantile function is given by $Q_{K_{w,t}|\mathbf{Z}_t=\mathbf{z}_t}(q) = \mathbf{z}_t'\boldsymbol{\pi}$, where $\boldsymbol{\pi} \in \mathbb{R}^4$. Additional polynomial terms were considered, but we found that a cubic expression was flexible enough to accurately capture quantiles trends for each of the studied examples. Applying the methodology described in Koenker et al. (2017), estimates of the sequences described in equation (4.3.2) can be obtained for any pair of quantiles q_1, q_2.

To select m and $\{(q_{1,j}, q_{2,j})\}$, a range of quantile sets were compared for the examples

discussed in Section 4.4, with the resulting ADF estimates showing very little difference in variability or accuracy. Consequently, the exact choice of quantile sets does not appear to significantly alter the resulting ADF estimates, so long as a range of different quantile pairs are considered. Therefore, for $\{q_{1,j}\}_{j=1}^{m}$, we take $m = 30$ equally spaced points in the interval $[0.9, 0.95]$, and set $q_{2,j} = q_{1,j} + 0.04$ for $j = 1, \ldots, m$.

For $\bar{\lambda}^*_{QR}$, estimates of non-stationary ADFs can be derived directly using estimated sequences, while specification of coefficients functions is also required for $\bar{\lambda}^*_{BP}$. Defining $\mathbf{z}_t := \{1, t\}$, we set $\log(\beta_i(\mathbf{z}_t)) = \mathbf{z}_t'\boldsymbol{\psi}_i$, with $\boldsymbol{\psi}_i \in \mathbb{R}^2$ for each $i \in \{1, 2, \ldots, k-1\}$, thereby ensuring positive coefficient functions. We found this form to be flexible enough to capture the range of dependence trends considered in the study, though additional polynomial terms and/or coefficient functions could be added for more complex data structures.

4.4.4 Simulation study

We evaluate the properties of estimators proposed in Section 4.3 and compare these to the estimator derived in Mhalla et al. (2019b), which we denote $\bar{\lambda}^*_{GAM}$. For the latter approach, the thresholds used to define exceedances of the min-projection are set to be constant over the covariate space: this assumption is not required for $\bar{\lambda}^*_{QR}$ or $\bar{\lambda}^*_{BP}$.

To obtain ADF estimates using $\bar{\lambda}^*_{GAM}$, the same covariate set is used as for the quantile regression procedure, i.e., $\mathbf{z}_t = \{1, t, t^2, t^3\}$. This link function proposed in Mhalla et al. (2019b) imposes the lower bound of $\lambda(w) = \max(w, 1-w)$, but also imposes an upper bound of $\lambda(w) = 1$, meaning negative dependence cannot be captured. To relax this constraint, we instead consider the link function $h_w(x) = \log(x - \max(w, 1-w))$, $x \in [\max(w, 1-w), \infty)$, $w \in [0, 1]$ when estimating the ADFs for the negatively correlated Gaussian example. This ensures the lower bound is still satisfied for each example, while allowing all forms of extremal dependence to be captured. We also impose the boundary conditions discussed in Section 4.3.5 for $w = 0$ and $w = 1$. Furthermore,

to ensure comparability, the final estimator is obtained as an average over 10 equally spaced quantiles in the interval $[0.9, 0.95]$. Such quantiles are used to define threshold exceedances of the min-projection variable prior to GAM fitting. We observe that imposing the upper bound $\lambda(w) = 1$ for $\bar{\lambda}^*_{GAM}$ in the majority of considered examples results in this estimator having a practical advantage over $\bar{\lambda}^*_{QR}$ and $\bar{\lambda}^*_{BP}$, since the latter estimators can exceed this upper bound.

To evaluate the performance of the estimators, the mean integrated squared error (MISE), alongside the integrated squared error (ISE) of a median estimator calculated over 250 simulated examples, was computed for each simulated example at three fixed time points; $t = 1$, $t = n/2$ and $t = n$. These results can be found in the Supplementary Material. These metrics indicate a similar performance between our proposed estimators and $\bar{\lambda}^*_{GAM}$; however, this is not the case if we consider the time frame as a whole. For this, we fix rays $w \in \{0.1, 0.3, 0.5\}$ and compute estimates of the ADF across 250 simulated examples. The median of the ADF estimates, alongside 0.025 and 0.975 quantiles, are calculated for each ray and time point $t \in \{1, \ldots, n\}$ and these estimates are then plotted over time for fixed rays. The resulting plots for the inverted logistic copula are illustrated in Figure 4.4.2. As can be observed, the median estimates appear very close to the true ADF values for both our estimators proposed in Section 4.3. The uncertainty is lower for $\bar{\lambda}^*_{BP}$ compared to $\bar{\lambda}^*_{QR}$, owing to the former's semi-parametric form. Furthermore, both of our estimators appear to show lower bias on average compared to $\bar{\lambda}^*_{GAM}$, particularly for $w = 0.3$ and $w = 0.5$. Similar results are obtained for each of the copula examples, with the exception of the negatively correlated Gaussian copula: in this case, significant bias arises nearer the start of the time interval for each estimator, owing to the infinite upper bound for λ as $\rho \to -1$. The resulting plots can be found in the Supplementary Material. These results suggest higher bias for $\bar{\lambda}^*_{GAM}$ compared to the estimators proposed in Section 4.3.

Finally, we again fix the time points $t = 1, t = n/2$ and $t = n$ and evaluate the vari-

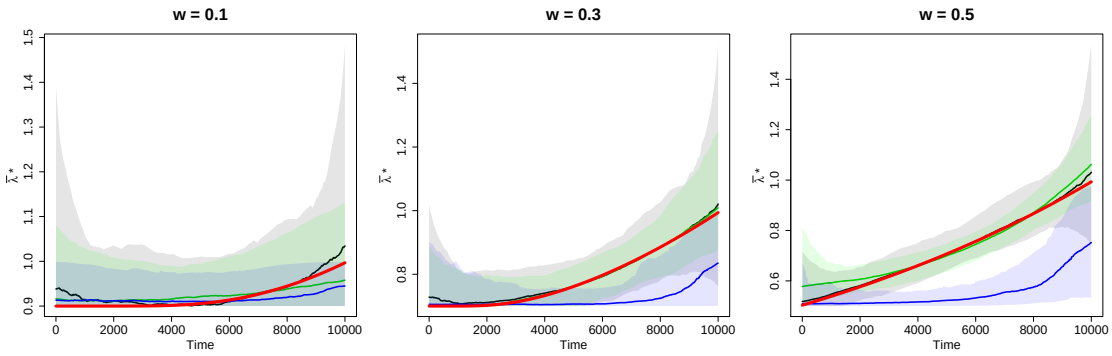

Figure 4.4.2: Plots of median and 95% confidence interval estimates over time at rays $w \in \{0.1, 0.3, 0.5\}$ for the inverted logistic copula. Red lines correspond to true ADF values, while the black, green and blue lines correspond to the median estimates for $\bar{\lambda}^*_{QR}$, $\bar{\lambda}^*_{BP}$ and $\bar{\lambda}^*_{GAM}$, respectively, with the coloured regions representing the estimated confidence intervals.

ability in ADF estimates. Median curve estimates, alongside pointwise 95% confidence intervals, are obtained for each of the copula examples. As with the analysis of MISE and ISE estimates, these results indicate similar performances across the three estimators. Plots of the estimated median curves and confidence regions for each estimator can be found in the Supplementary Material.

From these results combined, we conclude that the proposed estimators from Section 4.3 outperform the GAM-based estimator $\bar{\lambda}^*_{GAM}$ in a wide range of scenarios. This can be partly explained by the fact these estimators do not make the assumption of a constant threshold over time for the min-projection variable, leading to more realistic estimates. Moreover, we note that these results hold even though $\bar{\lambda}^*_{GAM}$ has the additional advantage of a link function which bounds the resulting ADF estimates for examples exhibiting positive extremal dependence. We therefore choose not to consider this GAM-based estimator further.

4.4.5 Return curve estimates

We now consider our ultimate goal of estimating return curves, $\mathrm{RC}_{\mathbf{z}_t}(p)$, for extreme survival probabilities. Curve estimates $\overline{\mathrm{RC}}_{\mathbf{z}_t}(p)$ at $p = 1/n$ under $\bar{\lambda}^*_{QR}$ and $\bar{\lambda}^*_{BP}$ were

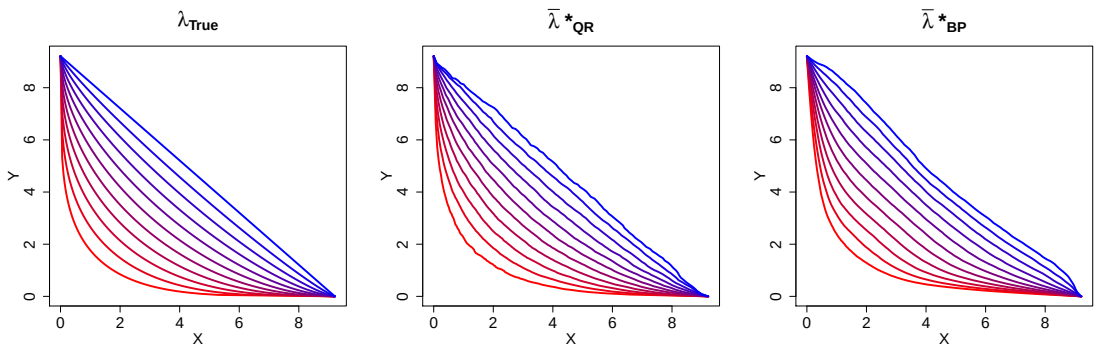

Figure 4.4.3: Plots of median return curve estimates over time with $p = 1/n$ for the Gaussian copula with negative correlation. Colour change illustrates extremal dependence trends over time, with red and blue corresponding to start and end of time frame, respectively. True curves given in left panel, with estimated curves from the median estimators of $\bar{\lambda}^*_{QR}$ and $\bar{\lambda}^*_{BP}$ given in centre and right panels, respectively.

obtained for 250 simulated examples from each copula. To give an overall impression of the bias from each estimator, we fix a time point t and plot the median of the 250 estimates for $\overline{\mathrm{RC}}_{\mathbf{z}_t}(p)$. Specifically, since each coordinate of $\overline{\mathrm{RC}}_{\mathbf{z}_t}(p)$ is associated with a ray $w \in [0, 1]$, we take the median along each ray. Median curve estimates for the negatively correlated Gaussian example are given alongside the true return curves in Figure 4.4.3. As can be observed, the estimated curves in each case closely resemble the true return curves. This is even true at the start of the observation period, for which significant bias in ADF estimators was observed. Return curve estimates for each of the other simulated examples were similar in accuracy; these plots can be found in the Supplementary Material.

4.5 UKCP18 temperature and dryness data

4.5.1 Properties of data

We denote the dataset introduced in Section 4.1 as $\{X_t, Y_t\}$ for $t \in \{1, \ldots, n\}$, with X_t and Y_t corresponding to the temperature and dryness variables, respectively. In this case, we have $n = 9000$, corresponding to 100 years of summer projections from

June 1981 to August 2080. We treat the time index t as a covariate for this data; while this does not correspond to any physical process, it can be used to capture the non-stationarity present in the data, which has been more fully explained by physical inputs to the climate model. For the marginal time series, empirical evidence indicates the presence of seasonal and long term trends within the main bodies of both variables. Further exploratory analysis suggests the presence of non-stationary behaviour within dryness extremes and that non-stationarity is present within the extremal dependence structure, as evidenced by the trend in η in Figure 4.1.1. In this section, we attempt to account for all three forms of non-stationary trends and produce return curve estimates up to the end of the observation period.

4.5.2 Capturing marginal non-stationarity

To capture marginal non-stationarity, we extend the pre-processing technique described in equation (4.2.3), with the goal of removing any marginal trends from the data. Rather than specifying linear parametric forms for the covariate functions, as in Eastoe and Tawn (2009), we instead assume the residual process R_t is a sequence of standard normal variables and allow μ and σ to be general smooth functions of covariate vectors. These functions can be estimated using a GAM framework, allowing for flexible modelling of the marginal trends.

The time series in both cases appear to exhibit long term trends in location and scale, along with seasonal behaviour within the former. Therefore, for the location function μ, and scale function σ, we take $\mathbf{z}_t = \{1, t, d_t\}$ and $\mathbf{z}_t = \{1, t\}$, respectively, where $d_t \in \{1, 2, \ldots, 90\}$ denotes the day index of the process at time t. For example, d_t equals 1, 45, and 90 for June 1st, July 15th and August 30th, respectively. A thin plate regression spline is used for the covariate t, while for d_t, a cyclic cubic regression spline of dimension 90 (corresponding to the number of data points in each year) is used.

The fitting of the location and scale covariate functions is carried out using the `R` package `mgcv` (Wood, 2021), with a `gaulss` family. Model optimisation is achieved via restricted maximum likelihood estimation, with smoothness penalties and knot quantities selected automatically using generalised cross validation. Further details for these modelling procedures can be found in Wood (2017). The resulting fitted trends are in good agreement with empirical estimates from the marginal time series: see the Supplementary Material for the corresponding plots. Moreover, the transformed series appear stationary, with no obvious long term or seasonal trends in either the location or scale.

With non-stationary trends in both marginal bodies accounted for, residual processes can be obtained through the transformations $R_{X_t} = [X_t - \mu_X(\mathbf{z}_t)]/\sigma_X(\mathbf{z}_t)$ and $R_{Y_t} = [Y_t - \mu_Y(\mathbf{z}_t)]/\sigma_Y(\mathbf{z}_t)$, where (μ_X, σ_X) and (μ_Y, σ_Y) denote the estimated covariate functions for X_t and Y_t, respectively. Assuming an accurate model fit, these processes should be approximately stationary within the body of the data. However, non-stationary trends are likely to remain in the tails since GAM fitting is driven by the body. Following Eastoe and Tawn (2009), we fit the non-stationary GPD described in equation (4.2.2) to capture any remaining trends. Significant linear and harmonic trends are shown to exist within the scale parameter of the residual process for dryness, while the shape parameter is assumed to be fixed over time. This assumption is common within the analysis of non-stationary univariate extremes (e.g., Eastoe and Tawn, 2009; Chavez-Demoulin and Davison, 2012), since the shape parameter is often seen seen as an intrinsic property of a physical process. No significant trends were found for the scale parameter corresponding to the temperature variable. Let $\{(r_{X_t}, r_{Y_t}) : t = 1, \ldots, n\}$ denote a data sample corresponding to the residuals vector (R_{X_t}, R_{Y_t}). An estimate of the marginal distribution function $F_{R_{Y_t}}$ is given by

$$\hat{F}_{R_{Y_t}}(r \mid \mathbf{Z}_t = \mathbf{z}_t) = \begin{cases} 1 - (1 - q_Y)\{1 + \hat{\xi}_Y(r - u_Y)/\hat{\tau}_Y(\mathbf{z}_t)\}^{-1/\hat{\xi}_Y} & \text{for } r > u_Y, \\ \sum_{t=1}^{n} \mathbb{1}(r_{Y_t} \leq r)/(n+1) & \text{for } r \leq u_Y, \end{cases}$$

where $\mathbb{1}$ denotes an indicator function, u_Y is the empirical q_Y quantile of R_{Y_t}, and $(\hat{\tau}_Y(\mathbf{z}_t), \hat{\xi}_Y)$ are the MLEs of the GPD scale and shape parameters. The stationary GPD, denoted in equation (4.2.1), is used to estimate the upper tail of $F_{R_{X_t}}$.

Finally, the data are transformed to standard exponential margins via the probability integral transform. To assess the outcome of the pre-processing procedure, exponential rate parameters were estimated over time for both marginal processes. The resulting estimates remain approximately constant at one (the target value) throughout the observation period; an illustrative plot can be found in the Supplementary Material.

4.5.3 Model fitting

With the data transformed to standard exponential margins, we apply the methodology proposed in Section 4.3 for return curve estimation. The extremal dependence trend observed in Figure 4.1.1 appears to suggest the extremes of the process are becoming more positively dependent over the time frame; therefore, one may expect the ADF estimates to tend towards the lower bound as $t \to n$.

We set $\mathbf{z}_t := \{1, t\}$ for both the quantile regression procedure and estimation of Bernstein polynomial coefficients. This reduced covariate space was flexible enough to capture the observed extremal dependence trend within the data. Moreover, the same set of quantile pairs $\{q_{1,j}, q_{2,j}\}_{j=1}^{m}$ was considered as defined in Section 4.4.3. Two sets of ADF estimates are obtained over the observation period corresponding to $\bar{\lambda}^*_{QR}$ and $\bar{\lambda}^*_{BP}$; these are plotted in the left and centre panels of Figure 4.5.1, respectively. The selected values of t for the plotted curves correspond to July 15th for an increasing subset of equally spaced years between 1981-2080.

We observe that for both estimators, we obtain ADFs which exceed one at several rays; this is unexpected, since exploratory analysis of the data appears to suggest the presence of positive dependence over the entire time frame. This could indicate the presence of some underlying issues within the proposed modelling framework; see

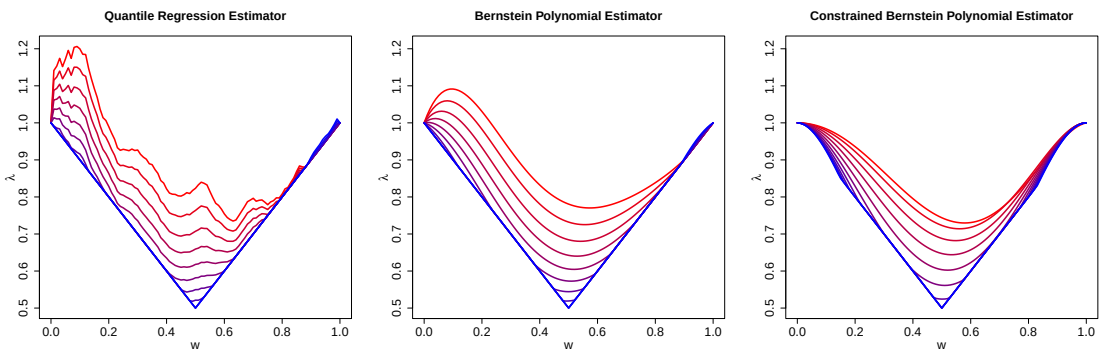

Figure 4.5.1: Left, centre and right: ADF estimates over time for $\bar{\lambda}^*_{QR}$, $\bar{\lambda}^*_{BP}$ and constrained $\bar{\lambda}^*_{BP}$, respectively. Curves change from red to blue over the observation period.

Section 4.6 for further discussion. While $\lambda(w) < 1$ or $\lambda(w) > 1$ for all $w \in (0, 1)$ indicate positive and negative extremal dependence, respectively, intermediate scenarios where $\lambda(w) > 1$ for some rays and $\lambda(w) < 1$ for others are theoretically possible. However, having investigated this aspect further, the evidence that $\lambda(w) > 1$ for small w is not strong, and so we seek to implement an estimator which is bounded above by one. We therefore re-apply the Bernstein polynomial estimator, this time constraining the coefficient functions to be in the interval $[0, 1]$ via a logit link function, i.e., $\text{logit}(\beta_i(\mathbf{z}_t)) = \mathbf{z}_t'\boldsymbol{\psi}_i$, $\boldsymbol{\psi}_i \in \mathbb{R}^2$, $i \in \{1, 2, \ldots, k-1\}$. The resulting ADF estimates are illustrated in the right panel of Figure 4.5.1. The estimated trends are again in good agreement with what is observed in the data, and the estimates approach the lower bound $\max(w, 1-w), w \in [0, 1]$, as $t \to n$. This convergence to asymptotic dependence is in agreement with the η estimates from Figure 4.1.1.

To evaluate uncertainty in estimates, we propose a block bootstrapping procedure. First, the data on exponential margins is split into segments of size 450, corresponding to five years of observations. The extremal dependence structure within such segments is then assumed to be approximately stationary. Within each segment, data are then resampled in blocks of size 15 with replacement to account for temporal dependence. These blocks are combined to form a resampled segment, then each of the segments are merged in order to obtain a new dataset. This process is repeated 250 times to generate

sets of ADF estimates while accounting for complex structures in the data. The median of the estimated ADFs for $t = n/2$, alongside 95% pointwise confidence intervals, are illustrated in the left panel of Figure 4.5.2 for the Bernstein polynomial estimator with constrained coefficients. Note that imposing the ADF lower bound, as described in Section 4.3.5, constrains the range of the confidence intervals, thus explaining why the median estimate appears close to the lower bound.

Finally, to assess the quality of the ADF estimates, we have compared the model estimates of η over time, obtained via the constrained version of $\bar{\lambda}^*_{BP}$, to the empirical estimates introduced in Section 4.1. For each rolling window, we have taken the average η estimate from the fitted model. As illustrated in the centre panel of Figure 4.5.2, the model estimates appear similar over time, suggesting we have accurately captured the extremal dependence trend at this ray. Similar results were also observed for model estimates at other rays.

4.5.4 Return curve estimates

We use our estimated ADFs to estimate return curves $\overline{\mathrm{RC}}_{\mathbf{z}_t}(p)$ up to the year 2080. In the stationary setting, we define '1 in 10,000-year' bivariate events to be those with joint survival probability of $p = 1/(10000 \times n_y)$, where n_y denotes the average number of observations per year (Brunner et al., 2016). We therefore obtain return curve estimates at this probability level for the vector (X_t, Y_t) using the constrained version of $\bar{\lambda}^*_{BP}$. This is done in two steps: first, we apply the method introduced in Section 4.3.6 to obtain return curve estimates on standard exponential margins. These curves are then transformed back to the original scale by applying the the inverse of the semi-empirical distribution given in equation (4.2.2), followed by the inverses of the transformations used to obtain the variables R_{X_t} and R_{Y_t}. The resulting curve estimates are illustrated in the right panel of Figure 4.5.2, with the selected values of t for the plotted curves again corresponding to July 15th for an increasing subset of years between 1981-2080. All of

the theoretical properties for return curves introduced in Murphy-Barltrop et al. (2023) have been imposed to ensure the resulting estimates are both theoretically possible and realistic.

From these curve estimates, two conclusions of relevance to nuclear regulators are evident. Firstly, clear marginal trends can be observed for both time series. For example, the temperature values at the point where the curves intersect the x-axis, which equate to $(1-p)$-th non-stationary quantile estimates (i.e., '1 in 10,000-year' univariate events), increase by over 15°C through the observation period. This implies that the regulatory design values will increase significantly over the time frame. Moreover, a dependence trend is evidenced by the changing shape in return curve estimates over time, with the curves becoming increasingly 'square' shaped from 1981-2080. This suggests that joint extremes of temperature and dryness are becoming more likely to occur over the observation period, implying drought-like conditions could be more common at the end of the time frame.

To better illustrate the shift in joint extremal behaviour over the observation period, consider the point labelled on the 1981 curve in green; this corresponds to the ray $w = 0.5$ when translated to standard exponential margins. These coordinates considered in the year 2080 would equate to a one in 1.25-year joint survival event. Moreover, the marginal return periods in 2080 would be 0.11 and 0.97 years for temperature and dryness, respectively. These values are orders of magnitude different from the 10,000-year regulatory standard. At the other end of the scale, values on the 2080 return curve lie above the fitted upper endpoints of the 1981 marginal distributions.

The non-stationarity in these return curves should be taken into account when considering the design basis for future nuclear installations - in particular, conservative principles suggest designing to values occurring at the end of the time frame. Moreover, due to the observed change in extremal dependence, future designs must be able to cope with the most extreme values of temperature and dryness occurring simultaneously,

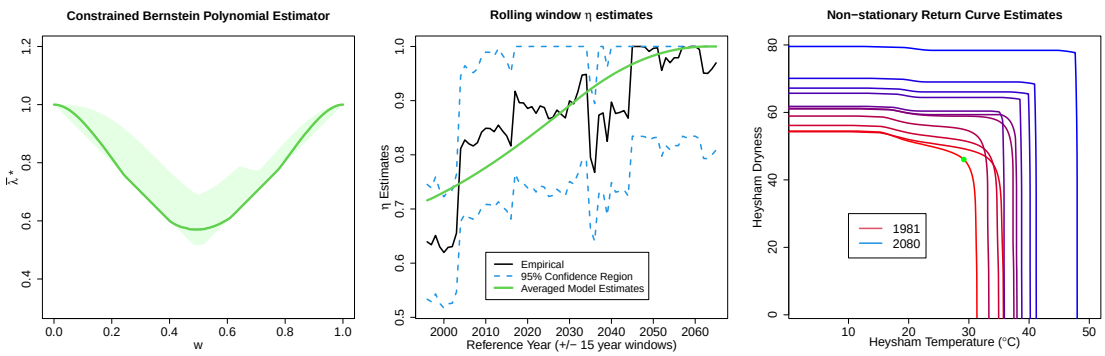

Figure 4.5.2: Left: median ADF estimate for $t = n/2$ obtained through bootstrapping procedure, with coloured region representing uncertainty bounds. Centre: Comparison of averaged model η estimates for rolling windows to empirical estimates, with black, green and dotted blue lines corresponding to empirical, model, and 95% confidence interval estimates respectively. Right: Return curve estimates on original margins with $p = 1/n$ at July 15th over the observation period. Time is illustrated using a colour transition, with the curves for the start and end of the time frame labelled. The green point corresponds to the ray $w = 0.5$ when the 1981 curve is transformed to standard exponential margins.

again resulting in a more complex design specification compared to the start of the observation period.

We note that there is a large degree of uncertainty in these curve estimates which cannot be quantified in a simple manner - see Section 4.6 for further discussion. Uncertainty arises in every step of the estimation procedure; see, for example, the estimated uncertainty for the fitted GAM functions illustrated in the Supplementary Material. Marginal trends play a key role in the changing values of the return curves, and differences in these point estimates would naturally impact the estimated curves.

4.6 Discussion

We have proposed a novel method for capturing non-stationary extremal dependence structures under asymptotic independence. Our method has been successfully applied to heavily non-stationary data from the UKCP18, allowing us to obtain return curve estimates up to the year 2080 and thereby illustrating how this framework could help

improve practical risk management under future climate scenarios.

We have investigated the properties of our estimators via simulation and have observed them generally perform well in terms of bias. While it would be desirable also to have theoretical results on bias of the estimators, this is very challenging in practice and is likely to require assumptions that are too strict to make the results worthwhile. Potentially a more promising line of future work would be the development of diagnostic plots for non-stationary ADF and return curve estimates. Our diagnostic procedures were limited to comparison of rolling window η estimates against those derived from the ADF. In general, diagnostic plots for return curves, such as those introduced in Murphy-Barltrop et al. (2023), require stationarity assumptions.

For the sake of simplicity, we have restricted attention to the bivariate setting. However, it is worth noting that all methods could in principle be extended to the general multivariate setting. This scenario results in additional complexities, since different extremal dependence regimes can exist within subvectors of a multivariate random vector, and more sophisticated model formulations may be required to capture such dependence structures. Moreover, applications of multivariate extremal risk measures are limited, owing in part to the fact visualisation and interpretation becomes more challenging in this setting.

As noted in Section 4.3.5, imposing the lower bound on the estimates of the ADF was shown to improve the estimation procedure. This was an ad-hoc post-processing step, with many of the obtained ADF estimates below the bound. Future work could explore how this lower bound could be incorporated into the theoretical quantile regression framework, and whether this would further improve the quality of ADF estimates. We note that when asymptotic dependence is present, $\lambda(w \mid \mathbf{z}_t) = \max(w, 1-w)$ will cease to depend on $\mathbf{z}_t$. However, return curve estimates $\overline{\mathrm{RC}}_{\mathbf{z}_t}(p)$ are also affected by the sequences $\{u_{w,t}\}$ as described in Section 4.3.6, so will still capture non-stationarity. Nonetheless, if asymptotic dependence is clearly present at all time points, other tech-

niques for estimating non-stationarity may be preferable.

When estimating non-stationary ADFs for the case study data in Section 4.5.3, we initially obtained ADFs which exceeded one for small w. While it is difficult to say for certain why this unexpected result occurred, we suspect this could be related to a marginal standardisation issue, i.e., the transformed margins are not exactly standard exponential. This issue would alter the tail properties of the min-projection variable, in-turn affecting the quality of the resulting ADFs estimates. Accurately accounting for marginal non-stationarity in the analysis of non-stationary multivariate extremes represents an important line of research.

We recognise that the method proposed for evaluating and representing uncertainty in Section 4.5 relies on strong assumptions and is an approximation of the true uncertainty. Theoretical derivation of uncertainty intervals for either of the estimators proposed in Section 4.3 is not possible, meaning any evaluation of uncertainty must be non-parametric. Uncertainty quantification is a general problem when modelling non-stationary processes, since the underlying datasets cannot be resampled using a straightforward bootstrapping procedure.

Finally, we note that the data within the climate projections exhibits non-negligible temporal dependence. This feature decreases the amount of information available, and we found that the ADF estimation procedures detailed in Section 4.3 performed worse when this feature was present. This feature may also be part of the explanation for the unexpected results discussed in Section 4.5.3. However, for ease of implementation, we have assumed independence in both marginal distributions. Techniques for incorporating temporal dependence with quantile regression (Koenker et al., 2017) could be incorporated into our methodology in future work.

Chapter 5

Improving estimation for asymptotically independent bivariate extremes via global estimators for the angular dependence function

5.1 Introduction

Bivariate extreme value theory is a branch of statistics that deals with the modelling of dependence between the extremes of two variables. This type of analysis is useful in a variety of fields, including finance (Castro-Camilo et al., 2018), engineering (Ross et al., 2020), and environmental science (Brunner et al., 2016), where understanding and predicting the behaviour of rare, high-impact events is important.

In certain applications, interest lies in understanding the risk of observing simultaneous extreme events at multiple locations; for example, in the context of flood risk

modelling, numerous simultaneous floods can result in damaging consequences to properties, businesses, infrastructure, communications and the economy (Lamb et al., 2010; Keef et al., 2013b). To support resilience planning, it it imperative to identify locations at high risk of joint extremes.

Classical theory for bivariate extremes is based on the framework of regular variation. Given a random vector (X, Y) with standard exponential margins, we say that (X, Y) is bivariate regularly varying if, for any measurable $B \subset [0, 1]$,

$$\lim_{r\to\infty} \Pr(V \in B, R > sr \mid R > r) = H(B)s^{-1}, \; s \geq 1, \tag{5.1.1}$$

with $R := e^X + e^Y$, $V := e^X/R$ and $H(\partial B) = 0$, where ∂B is the boundary of B (Resnick, 1987). Note that bivariate regular variation is most naturally expressed on standard Pareto margins, and the mapping $(X, Y) \mapsto (e^X, e^Y)$ performs this transformation. We refer to R and V as radial and angular components, respectively. Limit (5.1.1) implies that for the largest radial values, the radial and angular components are independent. Furthermore, the quantity H, which is known as the spectral measure, must satisfy the moment constraint $\int_0^1 v \mathrm{d}H(v) = 1/2$.

The spectral measure summarises the extremal dependence of (X, Y), and a wide range of approaches exist for its estimation (e.g., Einmahl and Segers, 2009; de Carvalho and Davison, 2014; Eastoe et al., 2014). Equivalently, one can consider Pickands' dependence function (PDF; Pickands, 1981), which has a direct relationship to H via

$$A(t) = \int_0^1 \max\{vt, (1-v)(1-t)\} 2\mathrm{d}H(v), \quad t \in [0, 1].$$

This function again captures the extremal dependence of (X, Y), and many approaches also exist for its estimation (e.g., Guillotte and Perron, 2016; Marcon et al., 2016; Vettori et al., 2018). Moreover, estimation approaches for the spectral measure and PDF encompass a wide range of statistical methodologies, with parametric, semi-parametric,

and non-parametric modelling techniques proposed in both Bayesian and frequentist settings.

However, methods based on bivariate regular variation are limited in the forms of extremal dependence they can capture. This dependence can be classified through the coefficient χ (Joe, 1997), defined as

$$\chi := \lim_{u \to \infty} \Pr(Y > u \mid X > u) \in [0, 1],$$

where this limit exists. If $\chi > 0$, then X and Y are asymptotically dependent, and the most extreme values of either variable can occur simultaneously. If $\chi = 0$, X and Y are asymptotically independent, and the most extreme values of either variable occur separately.

Under asymptotic independence, the spectral measure H places all mass on the points $\{0\}$ and $\{1\}$; equivalently, $A(t) = 1$ for all $t \in [0, 1]$. Consequently, for this form of dependence, the framework described in limit (5.1.1) is degenerate and is unable to accurately extrapolate into the joint tail (Ledford and Tawn, 1996, 1997). Practically, an incorrect assumption of asymptotic dependence between two variables is likely to result in an overly conservative estimate of joint risk.

To overcome this limitation, several models have been proposed that can capture both classes of extremal dependence. The first was given by Ledford and Tawn (1996), in which they assume that as $u \to \infty$, the joint tail can be represented as

$$\Pr(X > u, Y > u) = \Pr(\min(X, Y) > u) = L(e^u)e^{-u/\eta}, \tag{5.1.2}$$

where L is a slowly varying function at infinity, i.e., $\lim_{u\to\infty} L(cu)/L(u) = 1$ for $c > 0$, and $\eta \in (0, 1]$. The quantity η is termed the coefficient of tail dependence, with $\eta = 1$ and $\lim_{u\to\infty} L(u) > 0$ corresponding to asymptotic dependence and either $\eta < 1$ or $\eta = 1$ and $\lim_{u\to\infty} L(u) = 0$ corresponding to asymptotic independence. Many extensions to

this approach exist (e.g., Ledford and Tawn, 1997; Resnick, 2002; Ramos and Ledford, 2009); however, all such approaches are only applicable in regions where both variables are large, limiting their use in many practical settings. Since many extremal bivariate risk measures, such as environmental contours (Haselsteiner et al., 2021) and return curves (Murphy-Barltrop et al., 2023), are defined both in regions where both variables are extreme and in regions where only one variable is extreme, methods based on equation (5.1.2) are inadequate for their estimation.

Several copula-based models have been proposed that can capture both classes of extremal dependence, such as those given in Coles and Pauli (2002), Wadsworth et al. (2017) and Huser and Wadsworth (2019). Unlike equation (5.1.2), these can be used to evaluate joint tail behaviour in all regions where at least one variable is extreme. However, these techniques typically require strong assumptions about the parametric form of the bivariate distribution, thereby offering reduced flexibility.

Heffernan and Tawn (2004) proposed a modelling approach, known as the conditional extremes model, which also overcomes the limitations of the framework described in equation (5.1.2). This approach assumes the existence of normalising functions $a : \mathbb{R}_+ \to \mathbb{R}$ and $b : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$\lim_{u\to\infty} \Pr\left[(Y - a(X))/b(X) \leq z,\ X - u > x \mid X > u\right] = D(z)e^{-x},\ x > 0, \tag{5.1.3}$$

for a non-degenerate distribution function D. Note that the choice of conditioning on $X > u$ is arbitrary, and an equivalent formulation exists for normalised X given $Y > u$. This framework can capture both asymptotic dependence and asymptotic independence, with the former arising when $a(x) = x$ and $b(x) = 1$, and can also be used to describe extremal behaviour in regions where only one variable is large.

Finally, Wadsworth and Tawn (2013) proposed a general extension of equation

(5.1.2). As $u \to \infty$, they assume that for each ray $w \in [0, 1]$,

$$\Pr(\min\{X/w, Y/(1-w)\} > u) = L(e^u; w)e^{-\lambda(w)u}, \quad \lambda(w) \geq \max(w, 1-w), \tag{5.1.4}$$

where $L(\cdot\,; w)$ is slowly varying for each $w \in [0, 1]$ and λ is the angular dependence function (ADF), which generalises the coefficient $\eta = 1/\{2\lambda(0.5)\}$. This extension captures both extremal dependence regimes, with asymptotic dependence implying the lower bound, i.e., $\lambda(w) = \max(w, 1-w)$ for all $w \in [0, 1]$. Evaluation of the ADF for rays w close to 0 and 1 corresponds to regions where one variable is larger than the other.

The ADF can be viewed as the counterpart of the PDF for asymptotically independent variables, and shares many of its theoretical properties (Wadsworth and Tawn, 2013). It can be used to differentiate between different forms of asymptotic independence, with both positive and negative associations captured, alongside complete independence, which implies $\lambda(w) = 1$ for all $w \in [0, 1]$. Figure 5.1.1 illustrates the ADFs for three copulas. We observe a variety in shapes, corresponding to differing degrees of positive extremal dependence in the underlying copulas. The weakest dependence is observed for the inverted logistic copula, while the ADF for the asymptotically dependent logistic copula is equal to the lower bound.

Despite these modelling advances, the majority of approaches for quantifying the risk of bivariate extreme events still require bivariate regular variation. Many of the approaches that do allow for asymptotic independence use the conditional extremes model of equation (5.1.3) despite some well known limitations of this approach (Liu and Tawn, 2014).

One particular application of the model described in equation (5.1.4) is the estimation of so-called bivariate return curves, $\mathrm{RC}(p) := \{(x, y) \in \mathbb{R}^2 : \Pr(X > x, Y > y) = p\}$, which requires knowledge of extremal dependence in regions where either variable is large; see Section 5.5.4. Murphy-Barltrop et al. (2023) obtain estimates of return

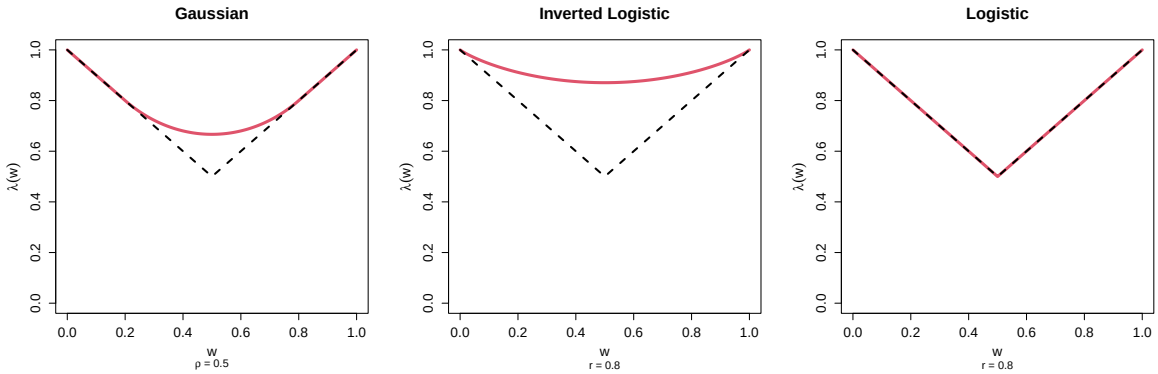

Figure 5.1.1: The true ADFs (given in red) for three example copulas. Left: bivariate Gaussian copula with coefficient $\rho = 0.5$. Centre: inverted logistic copula with dependence parameter $r = 0.8$. Right: logistic copula with dependence parameter $r = 0.8$. The lower bound for the ADF is denoted by the black dotted line.

curves, finding that estimates derived using equation (5.1.4) were preferable to those from the conditional extremes model. Mhalla et al. (2019b) and Murphy-Barltrop and Wadsworth (2022) also provide non-stationary extensions and inference methods for the ADF.

In this paper, we propose a global methodology for ADF estimation in order to improve extrapolation into the joint upper tail for bivariate random vectors exhibiting asymptotic independence. Until recently, the ADF has been estimated only in a pointwise manner using the Hill estimator (Hill, 1975) on the tail of $\min\{X/w, Y/(1-w)\}$, resulting in unrealistic rough functional estimates and, as we demonstrate in Section 5.4, high degrees of variability. Further, Murphy-Barltrop et al. (2023) showed that pointwise ADF estimates result in non-smooth return curve estimates, which are again unrealistic.

The first smooth ADF estimator was proposed in Simpson and Tawn (2022) based on a theoretical link between a limit set derived from the shape of appropriately scaled sample clouds and the ADF (Nolde and Wadsworth, 2022). The authors introduce global estimation techniques for the limit set, from which smooth ADF estimates follow; see Section 5.2 for further details.

We introduce several novel smooth ADF estimators, and compare their performance with the pointwise Hill estimator, as well as the estimator given in Simpson and Tawn (2022). In Section 5.2, we review the literature on ADF estimation. In Section 5.3, we introduce a range of novel estimators, and select tuning parameters for each proposed estimation technique. In Section 5.4, we compare each of the available estimators through a systematic simulation study, finding certain estimators to be favourable over others. A subset of estimators are then applied to river flow data sets in Section 5.5 and used to obtain estimates of return curves for different combinations of river gauges. We conclude in Section 5.6 with a discussion.

5.2 Existing techniques for ADF estimation

In this section, we introduce existing estimators for the ADF, with (X, Y) denoting a random vector with standard exponential margins throughout. To begin, for any ray $w \in [0, 1]$, define the min-projection at w as $T_w := \min\{X/w, Y/(1-w)\}$. Equation (5.1.4) implies that for any $w \in [0, 1]$ and $t > 0$,

$$\Pr(T_w > u + t \mid T_w > u) = \frac{L(e^{u+t}; w)}{L(e^u; w)} e^{-\lambda(w)t} \to e^{-\lambda(w)t} = t_*^{-\lambda(w)}, \tag{5.2.1}$$

as $u \to \infty$, with $t_* := e^t$. Since the expression in equation (5.2.1) has a univariate regularly varying tail with positive index, Wadsworth and Tawn (2013) propose using the Hill estimator (Hill, 1975) to obtain a pointwise estimator of the ADF; we denote this 'base' estimator $\hat{\lambda}_H$. A major drawback of this technique is that the estimator is pointwise, leading to rough and often unrealistic estimates of the ADF. Furthermore, this estimator need not satisfy the theoretical constraints on the ADF identified in Wadsworth and Tawn (2013), such as the endpoint conditions $\lambda(0) = \lambda(1) = 1$. Moreover, no information is shared across different rays, increasing the variability in the resulting estimates.

Simpson and Tawn (2022) recently proposed a novel estimator for the ADF using a theoretical link with the limiting shape of scaled sample clouds. Let $C_n := \{(X_i, Y_i)/\log n;\ i = 1, \ldots, n\}$ denote n scaled, independent copies of (X, Y). Nolde and Wadsworth (2022) explain how, as $n \to \infty$, the asymptotic shape of C_n provides information on the underlying extremal dependence structure. In many situations, C_n converges onto the compact limit set $\bar{G} = \{(x, y) : g(x, y) \leq 1\} \subseteq [0, 1]^2$, where g is the gauge function of $\bar{G}$. A sufficient condition for this convergence to occur is that the joint density, f, of (X, Y) exists, and that

$$-\log f(tx, ty) \sim tg(x, y), \quad x, y \geq 0,\ t \to \infty. \tag{5.2.2}$$

Following Nolde (2014), we also define the unit-level set $G = \{(x, y) : g(x, y) = 1\} \subset [0, 1]^2$. Given fixed margins, the shapes of $\bar{G}$ and G are completely determined by the extremal dependence structure of (X, Y). Furthermore, Nolde and Wadsworth (2022) show that the boundary set G is also directly linked to the modelling frameworks described in equations (5.1.2), (5.1.3) and (5.1.4), as well as the approach of Simpson et al. (2020). In particular, letting $R_w := (w/\max(w, 1-w), \infty] \times ((1-w)/\max(w, 1-w), \infty]$ for all $w \in [0, 1]$, we have that

$$\lambda(w) = \max(w, 1 - w) \times r_w^{-1}, \tag{5.2.3}$$

where

$$r_w = \min\{r \in [0, 1] : rR_w \bigcap G = \emptyset\}.$$

The boundary sets G for each of the copulas in Figure 5.1.1 are given in Figure 5.2.1, alongside the coordinates $(r_w w/\max(w, 1 - w), r_w(1 - w)/\max(w, 1 - w))$ for all $w \in [0, 1]$; these coordinates represent the relationship between G and the ADF via equation (5.2.3). One can again observe the variety in shapes. For the asymptotically dependent

logistic copula, we have that $(1, 1) \in G$; this is true for all asymptotically dependent bivariate random vectors.

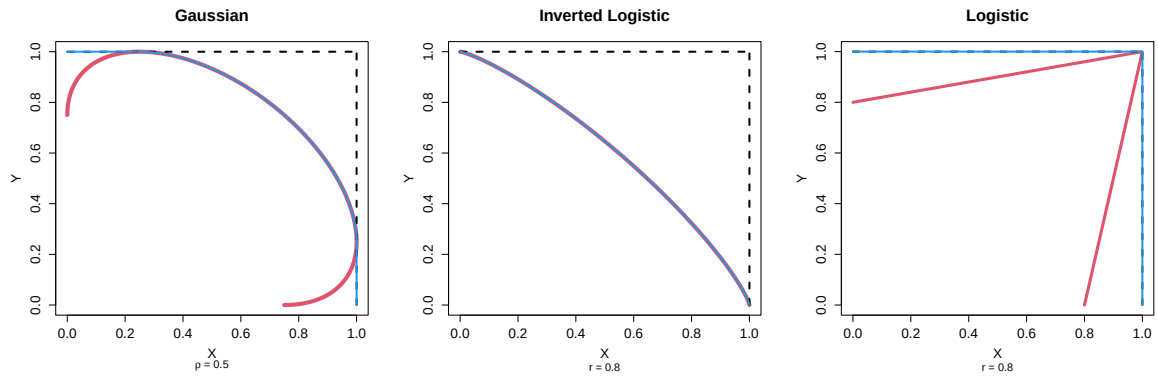

Figure 5.2.1: The boundary set G (given in red) for three example copulas, with coordinate limits denoted by the black dotted lines and the blue lines representing the coordinates $(r_w w/\max(w, 1-w), r_w(1-w)/\max(w, 1-w))$ for all $w \in [0, 1]$. Left: bivariate Gaussian copula with coefficient $\rho = 0.5$. Centre: inverted logistic copula with dependence parameter $r = 0.8$. Right: logistic copula with dependence parameter $r = 0.8$.

In practice, both the limit set, $\bar{G}$, and its boundary, G, are unknown. Simpson and Tawn (2022) propose an estimator for G, which is then used to derive an estimator $\hat{\lambda}_{ST}$ for the ADF via equation (5.2.3). The resulting estimator $\hat{\lambda}_{ST}$ was shown to outperform $\hat{\lambda}_H$ in a wide range of scenarios (Simpson and Tawn, 2022).

Estimation of G uses an alternative radial-angular decomposition of (X, Y), with $R^* := X + Y$ and $V^* := X/(X + Y)$. Simpson and Tawn (2022) assume the tail of $R^* \mid V^* = v^*$, $v^* \in [0, 1]$, follows a generalised Pareto distribution (GPD) (Davison and Smith, 1990) and then use generalised additive models to capture trends over angles in both the threshold and GPD scale parameter (Youngman, 2019). Next, high quantile estimates from the conditional distributions $R^* \mid V^* = v^*$, $v^* \in [0, 1]$ are computed using the fitted GPDs. They are then transformed back to the original scale using $X = R^*V^*$ and $Y = R^*(1 - V^*)$ and finally scaled onto the set $[0, 1]^2$ to give an estimate of G; see Simpson and Tawn (2022) for further details.

Wadsworth and Campbell (2022) also provide methodology for estimation of G,

though their focus is on estimation of tail probabilities more generally, including in dimensions greater than two. Furthermore, their approach requires prior selection of a parametric form for g. We therefore restrict our attention to the work of Simpson and Tawn (2022) as their main focus is non-parametric estimation for G in two dimensions.

When applying the estimators $\hat{\lambda}_H$ and $\hat{\lambda}_{ST}$ in Section 5.4 and 5.5, we use the tuning parameters suggested in the original approaches. In the case of $\hat{\lambda}_H$, we set u to be the empirical 90% quantile of T_w. The default tuning parameters for $\hat{\lambda}_{ST}$ can be found in Simpson and Tawn (2022), and example estimates of the set G obtained using the suggested parameters are given in the Supplementary Material.

5.3 Novel estimators for the ADF

Motivated by the goal of global estimation, we propose a range of novel estimators for the ADF. Since the ADF and PDF bear many theoretical similarities, we begin by reviewing estimation of the PDF. A smooth functional estimate for the ADF is desirable, so we restrict attention to approaches for the PDF which achieve this: spline-based techniques (Hall and Tajvidi, 2000; Cormier et al., 2014) and techniques that utilise the family of Bernstein-Bézier polynomials (Guillotte and Perron, 2016; Marcon et al., 2016, 2017b). In this paper, we focus on to the latter category, since spline-based techniques typically result in more complex formulations and a larger number of tuning parameters. Moreover, approaches based on Bernstein-Bézier polynomials have been shown to improve estimator performance across a wide range of copula examples (Vettori et al., 2018). For functions on the interval $[0, 1]$, the family of Bernstein-Bézier polynomials of degree $k \in \mathbb{N}$ is given by

$$\mathcal{B}_k = \left\{ \sum_{i=0}^{k} \beta_i \binom{k}{i} w^i (1-w)^{k-i} \;\middle|\; \boldsymbol{\beta} \in [0,1]^{k+1},\ w \in [0,1] \right\}.$$

Many approaches assume that the PDF $A \in \mathcal{B}_k$ and propose techniques for estimating the coefficient vector $\boldsymbol{\beta}$, resulting in an estimator $\hat{\boldsymbol{\beta}}$. This automatically ensures $A(t) \leq 1$ for all $t \in [0,1]$, thereby satisfying the theoretical upper bound of the PDF.

We make a similar assumption about the ADF, and use this to propose novel estimators. However, unlike the PDF, the ADF is unbounded from above, meaning functions in $\mathcal{B}_k$ cannot represent all forms of extremal dependence captured by equation (5.1.4). Moreover, the endpoint conditions $\lambda(0) = \lambda(1) = 1$ are not necessarily satisfied by functions in $\mathcal{B}_k$. We therefore propose an alternative family of polynomials: given $k \in \mathbb{N}$, let

$$\mathcal{B}_k^* = \left\{ (1-w)^k + \sum_{i=1}^{k-1} \beta_i \binom{k}{i} w^i (1-w)^{k-i} + w^k =: f(w) \;\middle|\; w \in [0,1], \right. \\ \left. \boldsymbol{\beta} \in [0,\infty)^{k-1} \text{ such that } f(w) \geq \max(w, 1-w) \right\}. \tag{5.3.1}$$

Functions in this family are unbounded from above, and $f(0) = f(1) = 1$ for all $f \in \mathcal{B}_k^*$. Furthermore, the parameter vector $\boldsymbol{\beta}$ is constrained such that each $f \in \mathcal{B}_k^*$ satisfies the lower bound of the ADF.

For the remainder of this section, let $\lambda(\cdot;\ \boldsymbol{\beta}) \in \mathcal{B}_k^*$ represent a form of the ADF from $\mathcal{B}_k^*$. Interest now lies in estimating the coefficient vector $\boldsymbol{\beta}$, which requires choice of the degree $k \in \mathbb{N}$. This is a trade-off between flexibility and computational complexity; polynomials with small values of k may not be flexible enough to capture all extremal dependence structures, resulting in bias, while high values of k will increase computational burden and parameter variance.

5.3.1 Composite likelihood approach

One consequence of equation (5.2.1) is that, for all $w \in [0,1]$, the conditional variable $T_w^* := (T_w - u_w \mid T_w > u_w) \sim \text{Exp}(\lambda(w))$, approximately, for large u_w. The density of this variable is $f_{T_w^*}(t_w^*) \approx \lambda(w) e^{-\lambda(w) t_w^*}, t_w^* > 0$, resulting in a likelihood function

for min-projection exceedances of u_w. Let $(\mathbf{x}, \mathbf{y}) := \{(x_i, y_i) : \ i = 1, \ldots, n\}$ denote n independent observations from the joint distribution of (X, Y). For each $w \in \mathcal{W}$, where $\mathcal{W}$ denotes some finite subset spanning the interval $[0, 1]$, let $\mathbf{t}_w := \{\min(x_i/w, y_i/(1-w)) : \ i = 1, \ldots, n\}$ and take u_w to be the empirical $100q\%$ quantile of $\mathbf{t}_w$, with q close to 1 and fixed across w. Letting $\mathbf{t}_w^* := \{t_w - u_w \mid t_w \in \mathbf{t}_w, t_w > u_w\}$, we have a set of realisations from the conditional variable T_w^*.

One approach to obtain an estimate of $\lambda(w)$ while considering all $w \in \mathcal{W}$ simultaneously is to use a composite likelihood, in which multiple likelihood components are treated as independent whether or not they are independent. Provided each component is a valid density function, the resulting likelihood function provides unbiased parameter estimates; see Varin et al. (2011) for further details. For this model, the likelihood function is given by

$$\mathcal{L}_C(\boldsymbol{\beta}) = \prod_{w \in \mathcal{W}} \prod_{t_w^* \in \mathbf{t}_w^*} \lambda(w;\, \boldsymbol{\beta}) e^{-\lambda(w;\, \boldsymbol{\beta}) t_w^*} = \left[\prod_{w \in \mathcal{W}} \lambda(w;\, \boldsymbol{\beta})^{|\mathbf{t}_w^*|} \right] \times e^{-\sum_{w \in \mathcal{W}} \sum_{t_w^* \in \mathbf{t}_w^*} \lambda(w;\, \boldsymbol{\beta}) t_w},$$

where $|\mathbf{t}_w^*|$ denotes the cardinality of the set $\mathbf{t}_w^*$. This composite likelihood function has equal weights across all $w \in \mathcal{W}$ (the 'components'). An estimator of the ADF, $\hat{\lambda}_{CL}$, is given by $\lambda(\cdot;\, \hat{\boldsymbol{\beta}}_{CL})$, where $\hat{\boldsymbol{\beta}}_{CL}$ is the maximum likelihood estimator of $\boldsymbol{\beta}$.

To apply this method in practice, one must first select a set $\mathcal{W}$ and a value for the probability q; see Section 5.3.4 for further details. The former is akin to selecting the degree of smoothing, while the latter is analogous to selecting a threshold for the GPD defined in Section 5.2 in the univariate setting.

5.3.2 Probability ratio approach

With $\mathcal{W}$ and $\mathbf{t}_w$ defined as in Section 5.3.1, consider two probabilities $q < p < 1$, both close to one. Given any $w \in \mathcal{W}$, let u_w and v_w denote the $100q\%$ and $100p\%$ empirical quantiles of $\mathbf{t}_w$, respectively. Assuming the distribution function of T_w is

strictly monotonic, equation (5.2.1) implies that

$$\frac{1-p}{1-q} = \Pr(T_w > v_w \mid T_w > u_w) \approx e^{-\lambda(w)(v_w - u_w)} \Rightarrow \left| \frac{1-p}{1-q} - e^{-\lambda(w)(v_w-u_w)} \right| \approx 0, \tag{5.3.2}$$

Similarly to Murphy-Barltrop and Wadsworth (2022), we exploit equation (5.3.2) to obtain an estimator for the ADF. Firstly, we observe that this representation holds for all $w \in \mathcal{W}$, hence

$$\sum_{w \in \mathcal{W}} \left| \frac{1-p}{1-q} - e^{-\lambda(w)(v_w-u_w)} \right| \approx 0.$$

To ensure equation (5.3.2) holds requires careful selection of q and p. This selection also represents a bias-variance trade off: probabilities too small (big) will induce bias (high variability). Moreover, owing to the different rates of convergence to the limiting ADF (Wadsworth and Tawn, 2013), a single pair (q, p) is unlikely to be appropriate across all data structures. We instead consider a range of probability pairs simultaneously. Specifically, letting $\{(q_j, p_j) \mid q_j < p_j < 1, 1 \leq j \leq m\}$, $m \in \mathbb{N}$, be pairs of probabilities near one, consider the expression

$$S(\boldsymbol{\beta}) := \sum_{w \in \mathcal{W}} \sum_{j=1}^{m} \left| \left[\frac{1-p_j}{1-q_j} \right] - e^{-\lambda(w;\boldsymbol{\beta})(v_{w,j}-u_{w,j})} \right|,$$

in which $u_{w,j}$ and $v_{w,j}$ denote $100q_j\%$ and $100p_j\%$ empirical quantiles of $\mathbf{t}_w$, respectively, for each $j = 1, \ldots, m$. Since it is desirable for $S(\boldsymbol{\beta}) = 0$, we set $\hat{\boldsymbol{\beta}}_{PR} = \arg\min_{\boldsymbol{\beta} \in [0,\infty)^{k-1}} S(\boldsymbol{\beta})$ and denote by $\hat{\lambda}_{PR}$ the estimator $\lambda(\cdot;\, \hat{\boldsymbol{\beta}}_{PR})$. Similarly to $\hat{\lambda}_{CL}$, one must select the sets $\mathcal{W}$ and $\{(q_j, p_j) \mid q_j < p_j < 1, 1 \leq j \leq m\}$, $m \in \mathbb{N}$ prior to applying this estimator; see Section 5.3.4.

5.3.3 Incorporating knowledge of conditional extremes parameters

As noted in Section 5.2, the set G links different representations of bivariate extremes. Recall the conditional extremes modelling framework described in equation (5.1.3). Let $a_{y|x}$ and $a_{x|y}$ be the normalising functions for conditioning on the events $X > t$ and $Y > t$ respectively, and let $\alpha_{y|x} := \lim_{t\to\infty} a_{y|x}(t)/t$ and $\alpha_{x|y} := \lim_{t\to\infty} a_{x|y}(t)/t$, with $\alpha_{y|x}, \alpha_{x|y} \in [0, 1]$. From Nolde and Wadsworth (2022), we have that $g(1, \alpha_{y|x}) = 1$ and $g(\alpha_{x|y}, 1) = 1$, with g defined as in equation (5.2.2). Assuming that the values of $\alpha_{y|x}$ and $\alpha_{x|y}$ are known, it follows from equation (5.2.3) that since the set $R_{1/(1+\alpha)}$ will always intersect the point $(1, \alpha_{y|x}) \in G$ for any $\alpha \in [0, \alpha_{y|x}]$,

$$\lambda\left(\frac{1}{1+\alpha}\right) = \max\left(\frac{1}{1+\alpha}, \frac{\alpha}{1+\alpha}\right) \times r^{-1}_{1/(1+\alpha)} = \frac{1}{1+\alpha} \times 1^{-1} = \frac{1}{1+\alpha}, \tag{5.3.3}$$

for all $\alpha \in [0, \alpha_{y|x}]$. Similarly, $\lambda(\alpha/(1+\alpha)) = 1/(1+\alpha)$ for all $\alpha \in [0, \alpha_{x|y}]$. Consequently, for all $w \in [0, \alpha^*_{x|y}] \bigcup [\alpha^*_{y|x}, 1]$, with $\alpha^*_{x|y} := \alpha_{x|y}/(1+\alpha_{x|y})$ and $\alpha^*_{y|x} := 1/(1+\alpha_{y|x})$, $\lambda(w) = \max(w, 1-w)$.

This result is illustrated in Figure 5.3.1 for a Gaussian copula with $\rho = 0.5$. Here, $\alpha_{x|y} = \alpha_{y|x} = 0.25$, implying $\lambda(w, 1-w) = \max(w, 1-w)$ for all $w \in [0, 0.2] \bigcup [0.8, 1]$; these rays correspond to the blue lines in the figure. One can observe that for any region R_w defined along either of the blue lines (such as the shaded regions illustrated for $w = 0.1$ and $w = 0.9$), we have that $r_w = 1$, since these regions will intersect G at either the coordinates $(0.25, 1)$ or $(1, 0.25)$.

In practice, $\alpha^*_{x|y}$ and $\alpha^*_{y|x}$ are unknown; however, estimates $\hat{\alpha}_{y|x}$ and $\hat{\alpha}_{x|y}$ are commonly obtained using a likelihood function based on a misspecified model for the distribution D in equation (5.1.3) (e.g. Jonathan et al., 2014a). The resulting estimates, denoted $\hat{\alpha}^*_{x|y}$, $\hat{\alpha}^*_{y|x}$, can be used to approximate the ADF for $w \in [0, \hat{\alpha}^*_{x|y}) \bigcup (\hat{\alpha}^*_{y|x}, 1]$. What now remains is to combine this with an estimator for $\lambda(w)$ on $[\hat{\alpha}^*_{x|y}, \hat{\alpha}^*_{y|x}]$.

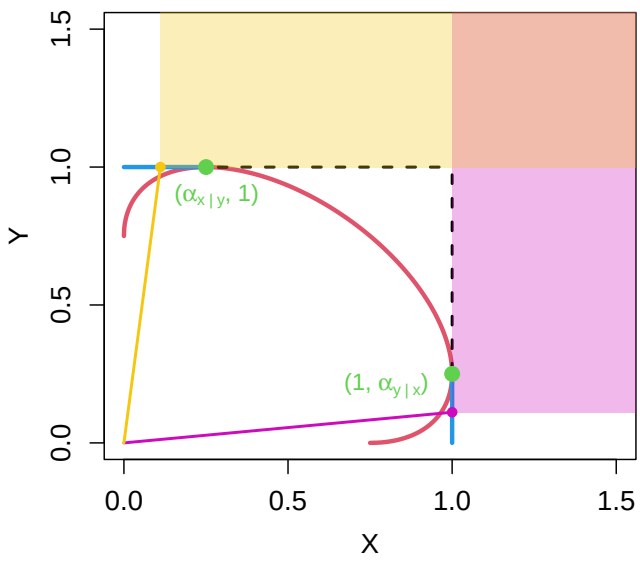

Figure 5.3.1: Pictoral illustration of the result described in equation (5.3.3). The boundary set G, given in red, is from the bivariate Gaussian copula with $\rho = 0.5$, with the points $(1, \alpha_{y|x})$ and $(\alpha_{x|y}, 1)$ given in green. The blue lines represent the rays $w \in [0, \alpha^*_{x|y}] \bigcup [\alpha^*_{y|x}, 1]$, while the yellow and pink shaded regions represent the set R_w for $w = 0.1$ and $w = 0.9$, respectively.

A crude way to obtain an estimator via this framework would be to set $\lambda(w) = \max(w, 1-w)$ for $w \in [0, \hat{\alpha}^*_{x|y}) \bigcup (\hat{\alpha}^*_{y|x}, 1]$ and $\lambda(w) = \hat{\lambda}_H(w)$, $\hat{\lambda}_{CL}(w)$ or $\hat{\lambda}_{PR}(w)$ for $w \in [\hat{\alpha}^*_{x|y}, \hat{\alpha}^*_{y|x}]$. However, this results in discontinuities at $\hat{\alpha}^*_{x|y}$ and $\hat{\alpha}^*_{y|x}$. Instead, for the smooth estimators, we rescale $\mathcal{B}^*_k$ such that the resulting ADF estimate is continuous for all $w \in [0, 1]$. Consider the set of polynomials given by

$$\mathcal{B}'_k = \left\{ \hat{\alpha}^*_{x|y} \left(1 - \gamma^*(v)\right)^k + \sum_{i=1}^{k-1} \beta_i \binom{k}{i} \gamma^*(v)^i \left(1 - \gamma^*(v)\right)^{k-i} + \hat{\alpha}^*_{y|x} \gamma^*(v)^k =: f(v) \right. \\ \left. \Big| v \in [\hat{\alpha}^*_{x|y}, \hat{\alpha}^*_{y|x}], \boldsymbol{\beta} \in [0, \infty)^{k-1} \text{ such that } f(v) \geq \max(v, 1-v) \right\}, \tag{5.3.4}$$

with $\gamma^*(v) := (v - \hat{\alpha}^*_{x|y})/(\hat{\alpha}^*_{y|x} - \hat{\alpha}^*_{x|y})$. For all $f \in \mathcal{B}'_k$, we have that $f(\hat{\alpha}^*_{x|y}) = \hat{\alpha}^*_{x|y}$ and $f(\hat{\alpha}^*_{y|x}) = \hat{\alpha}^*_{y|x}$, and each f is only defined on the interval $[\hat{\alpha}^*_{x|y}, \hat{\alpha}^*_{y|x}]$. Letting

$\lambda'(\cdot\,;\boldsymbol{\beta}) \in \mathcal{B}_k'$ represent a form of the ADF for $w \in [\hat{\alpha}^*_{x|y}, \hat{\alpha}^*_{y|x}]$, the techniques introduced in Sections 5.3.1 and 5.3.2 can be used to obtain estimates of the coefficient vectors, which we denote $\hat{\boldsymbol{\beta}}_{CL2}$ and $\hat{\boldsymbol{\beta}}_{PR2}$, respectively. The resulting estimators for λ are given by

$$\hat{\lambda}_{CL2}(w) = \begin{cases} \lambda'(w; \hat{\boldsymbol{\beta}}_{CL2}) & \text{for } w \in [\hat{\alpha}^*_{x|y}, \hat{\alpha}^*_{y|x}], \\ \max(w, 1-w) & \text{for } w \in [0, \hat{\alpha}^*_{x|y}) \bigcup (\hat{\alpha}^*_{y|x}, 1], \end{cases}$$

with $\hat{\lambda}_{PR2}$ defined analogously. We lastly define the discontinuous estimator $\hat{\lambda}_{H2}$ as

$$\hat{\lambda}_{H2} := \begin{cases} \hat{\lambda}_H(w) & \text{for } w \in [\hat{\alpha}^*_{x|y}, \hat{\alpha}^*_{y|x}], \\ \max(w, 1-w) & \text{for } w \in [0, \hat{\alpha}^*_{x|y}) \bigcup (\hat{\alpha}^*_{y|x}, 1]. \end{cases}$$

This is obtained by combining the pointwise Hill estimator with the information provided by the estimates $\hat{\alpha}^*_{x|y}, \hat{\alpha}^*_{y|x}$. Illustrations of all the estimators discussed in this section, as well as in Section 5.2, can be found in the Supplementary Material.

5.3.4 Tuning parameter selection

Prior to using any of the ADF estimators introduced in this section, we are required to select at least one tuning parameter. For the probability values required by the estimators introduced in Sections 5.3.1 and 5.3.2, we set $q = 0.90$, $\{q_j\}_{j=1}^m := \{0.87 + (j-1) \times 0.002\}_{j=1}^m$ and $p_j = q_j + 0.05$ for $j = 1, \ldots, m$, with $m = 31$. These values were chosen to evaluate whether the resulting estimators improve upon the base estimator $\hat{\lambda}_H$ using (approximately) the same amount of tail information in all cases. We tested a range of probabilities for both estimators and found that the overall quality of ADF estimates was similar across different extremal dependence structures. For example, for $\hat{\lambda}_{CL}$, a lower q resulted in mild improvements for asymptotically independent copulas, while simultaneously worsening the quality of ADF estimates for asymptotically dependent examples, while a higher q led to higher variance.

We also set $\mathcal{W} := \{0, 0.001, 0.002, \ldots, 0.999, 1\}$, i.e., a finite set of equally spaced rays spanning the interval $[0, 1]$. This set was sufficient to ensure a high degree of smoothness in the resulting ADF estimates without too high a computational burden.

For each of the novel estimators (except $\hat{\lambda}_{H2}$), we must also specify the degree $k \in \mathbb{N}$ for the polynomial families described by equations (5.3.1) and (5.3.4). In the case of the PDF, studies have found that higher values of k are preferable for very strong positive dependence, while the opposite is true for weak dependence (Marcon et al., 2017b; Vettori et al., 2018). We prefer to select a single value of k that performs well across a range of dependence structures, while minimising the computational burden; this avoids the need to select this parameter when obtaining ADF estimates in practice.

To achieve this objective, we estimated the root mean integrated squared error (RMISE), as defined in Section 5.4.1, of the estimators $\hat{\lambda}_{CL}$ and $\hat{\lambda}_{PR}$ with $k = 2, 3, \ldots, 11$ using 200 samples from two Gaussian copula examples, corresponding to strong ($\rho = 0.9$) and weak ($\rho = 0.1$) positive dependence. Assessment of how the RMISE estimates vary over k for both estimators suggests that $k = 7$ is sufficient to accurately capture both dependence structures. The full results can be found in the Supplementary Material.

For each of the 'combined' estimators in Section 5.3.3, we take the same tuning parameters as for the 'non-combined' counterpart, since the combined estimators have near identical formulations only defined on a subset of $[0, 1]$. For example, the empirical 90% threshold of the min-projection is used for both $\hat{\lambda}_H$ and $\hat{\lambda}_{H2}$. Finally, when estimating the conditional extremes parameters, empirical 90% conditioning thresholds are used.

5.4 Simulation Study

5.4.1 Overview

In this section, we use simulation to compare the estimators proposed in Section 5.3 to the existing techniques described in Section 5.2. For the comparison, we introduce nine copula examples, representing a wide variety of extremal dependence structures, and encompassing both extremal dependence regimes.

The first three examples are from the bivariate Gaussian distribution, for which the dependence is characterised by the correlation coefficient $\rho \in [-1, 1]$. We consider $\rho \in \{-0.6, 0.1, 0.6\}$, resulting in data structures exhibiting medium negative, weak positive, and medium positive dependence, respectively. Note that in the case of $\rho = -0.6$, the choice of exponential margins will hide the dependence structure (Keef et al., 2013a; Nolde and Wadsworth, 2022).

For the next two examples, we consider the bivariate extreme value copula with logistic and asymmetric logistic families (Gumbel, 1960; Tawn, 1988). In both cases, the dependence is characterised by the parameter $r \in (0, 1]$; we set $r = 0.8$, corresponding to weak positive extremal dependence. For the asymmetric logistic family, we also require two asymmetry parameters $(k_1, k_2) \in [0, 1]^2$, which we fix to be $(k_1, k_2) = (0.3, 0.7)$, resulting in a mixture structure.

We next consider the inverted bivariate extreme value copula (Ledford and Tawn, 1997) for the logistic and asymmetric logistic families, with the dependence again characterised by the parameters r and (r, k_1, k_2), respectively. We set $r = 0.4$, corresponding to moderate positive dependence, and again fix $(k_1, k_2) = (0.3, 0.7)$. Note that for this copula, the model described in equation (5.1.4) is exact: see Wadsworth and Tawn (2013).

Lastly, we consider the bivariate student t copula, for which dependence is characterised by the correlation coefficient $\rho \in [-1, 1]$ and the degrees of freedom $\nu > 0$. We

consider $\rho = 0.8$, $\nu = 2$ and $\rho = 0.2$, $\nu = 5$, corresponding to strong and weak positive dependence.

Illustrations of the true ADFs for each copula are given in Figure 5.4.1, showing a range of extremal dependence structures. For examples where the ADF equals the lower bound, the copula exhibits asymptotic dependence. While the fifth copula exhibits asymmetric dependence, the limiting ADF is symmetric; the same is not true for its inverted counterpart.

To evaluate estimator performance, we use the RMISE

$$\text{RMISE}\left(\hat{\lambda}_{-}\right) = \left(\mathbb{E}\left[\int_0^1 \left\{\hat{\lambda}_{-}(w) - \lambda(w)\right\}^2 \mathrm{d}w\right]\right)^{1/2},$$

where $\hat{\lambda}_{-}$ denotes an arbitrary estimator. Simple rearrangement shows that this metric is equal to the square root of the sum of integrated squared bias (ISB) and integrated variance (IV) (Gentle, 2002), i.e.,

$$\text{RMISE}\left(\hat{\lambda}_{-}\right) = \left[\underbrace{\int_0^1 \left\{\mathbb{E}\left[\hat{\lambda}_{-}(w)\right] - \lambda(w)\right\}^2 \mathrm{d}w}_{\text{ISB}} + \underbrace{\int_0^1 \mathbb{E}\left(\left\{\mathbb{E}\left[\hat{\lambda}_{-}(w)\right] - \hat{\lambda}_{-}(w)\right\}^2\right) \mathrm{d}w}_{\text{IV}}\right]^{1/2}.$$

Therefore, the RMISE summarises the quality of an estimator in terms of both bias and variance, and can be used as a means to compare different estimators.

5.4.2 Results

For the copulas described in Section 5.4.1, data from each copula example was simulated on standard exponential margins with a sample size of n =10,000, and the integrated squared error (ISE) of each estimator was approximated for $1,000$ samples using the trapezoidal rule; see the Supplementary Material for further information. The square root of the mean of these estimates was then computed, resulting in a Monte-Carlo estimate of the RMISE. For each of the estimators $\hat{\lambda}_{CL}, \hat{\lambda}_{PR}, \hat{\lambda}_{CL2}$ and $\hat{\lambda}_{PR2}$, the opti-

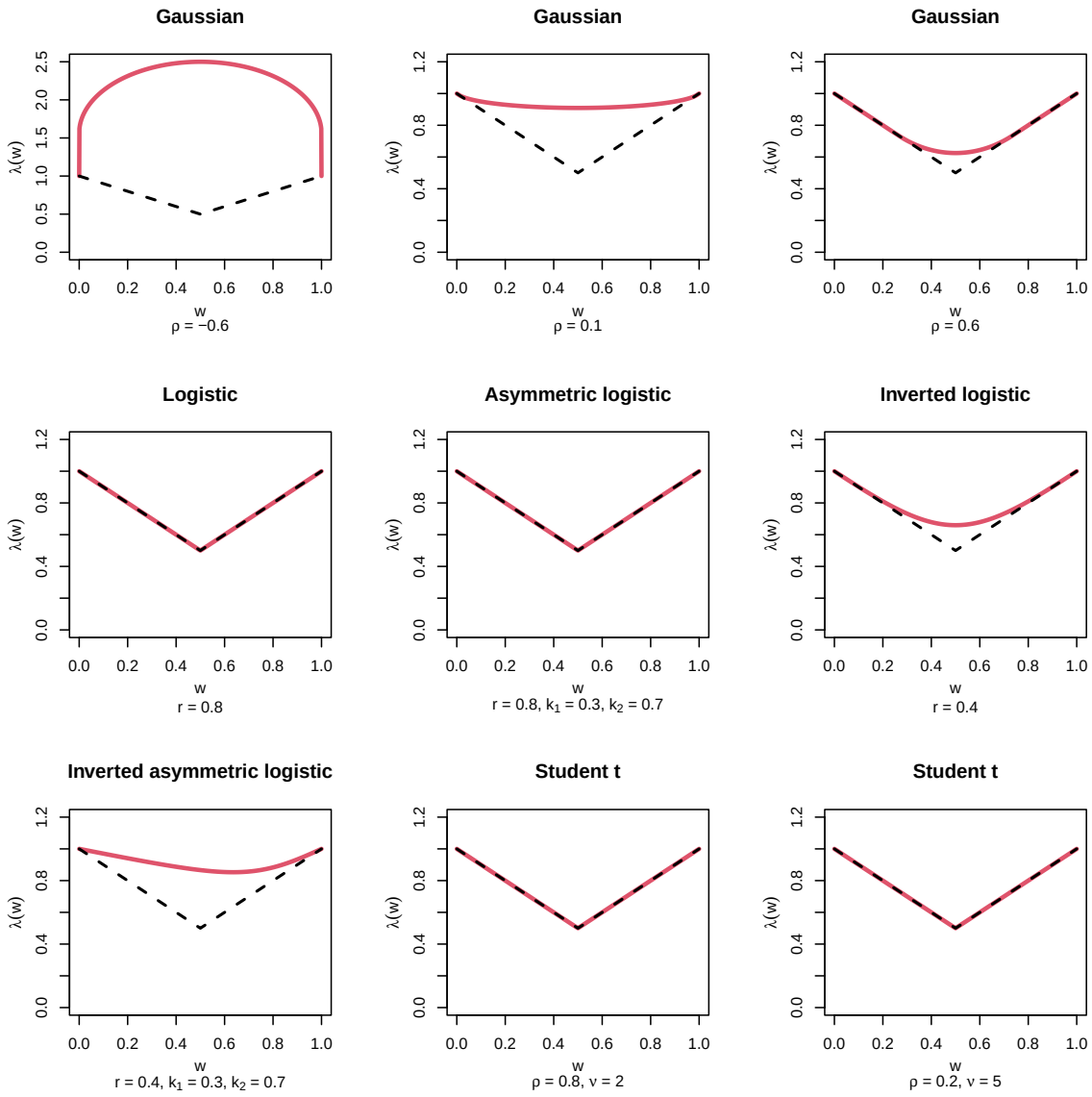

Figure 5.4.1: True ADFs (in red) for each copula introduced in Section 5.4.1, along with the theoretical lower bound (black dotted line).

misation functions were constrained to ensure the resulting ADF estimates satisfy the theoretical lower bound. The RMISE estimates for each estimator and copula combination are shown in Table 5.4.1. Tables for ISB and IV can be found in the Supplementary Material. For each estimator, the bias varies significantly across the different copulas. However, in the majority of cases, the bias/variance are similar across most of the estimators.

Table 5.4.1: RMISE values (multiplied by 100) for each estimator and copula combination. Smallest RMISE values in each row are highlighted in bold, with values reported to 3 significant figures.

Copula	$\hat{\lambda}_H$	$\hat{\lambda}_{CL}$	$\hat{\lambda}_{PR}$	$\hat{\lambda}_{H2}$	$\hat{\lambda}_{CL2}$	$\hat{\lambda}_{PR2}$	$\hat{\lambda}_{ST}$
Copula 1	**61.1**	61.3	66.2	61.4	61.9	66.7	63.7
Copula 2	3.55	3.33	3.64	3.51	3.33	3.63	**2.95**
Copula 3	3.78	3.48	3.84	3.27	3.22	3.57	**1.09**
Copula 4	4.9	4.79	6.92	4.28	4.25	6.17	**2.77**
Copula 5	14.1	14.1	17.1	14.1	14.1	17	**12.1**
Copula 6	2.51	1.97	2.15	2	**1.74**	1.9	2.12
Copula 7	2.93	**2.64**	2.88	2.87	2.66	2.89	3.96
Copula 8	2.49	2.72	2.95	0.66	**0.6**	0.789	1.87
Copula 9	12.1	12	14.9	12	12	14.9	**11.1**

While no estimator consistently outperforms the others, $\hat{\lambda}_{CL2}$ and $\hat{\lambda}_{ST}$ tend to have lower RMISE, ISB and IV values, on average. This is especially the case when comparing to the base estimator $\hat{\lambda}_H$. Furthermore, the 'combined' estimators outperform their non-combined counterparts in many cases, suggesting that incorporating parameter estimates from the conditional extremes model can reduce bias and variance. The Gaussian copula with $\rho = -0.6$ has significantly higher RMISE values, indicating that none of the estimators capture negative dependence well, though this is in part due to the choice of exponential margins.

Overall, these results indicate that no one estimator is preferable across all extremal dependence structures. However, we suggest using the estimators $\hat{\lambda}_{CL2}$ and $\hat{\lambda}_{ST}$ since, on average, these appeared to result in less bias and variance. The form of extremal dependence appears to affect the performance of both of these estimators; since this is often difficult to quantify a priori, we suggest using both estimators and evaluating relative performance via diagnostics, as we do in Section 5.5.

5.5 Case Study

5.5.1 Overview

Understanding the probability of observing extreme river flow events (i.e., floods) at multiple sites simultaneously is important in a variety of sectors, including insurance (Keef et al., 2013b; Quinn et al., 2019; Rohrbeck and Cooley, 2021) and environmental management (Lamb et al., 2010; Gouldby et al., 2017). Valid risk assessments therefore require accurate evaluation of the extremal dependence between different sites.

In this section, we estimate the ADF of river flow data sets obtained from gauges in the north of England, UK, which can be subsequently used to construct bivariate return curves. Daily average flow values (m^3/s) at six river gauge locations on different rivers were considered. The gauge sites are illustrated in Figure 5.5.1. For each location, data is available between May 1993 and September 2020; however, we only consider dates where a measurement is available for every location. To avoid seasonality, we consider the interval October-March only; from our analysis, it appears that the highest daily flow values are observed in this period. This results in $n = 4,659$ data points for each site. Plots of the daily flow time series can be found in the Supplementary Material; these plots suggest that an assumption of stationarity is reasonable for the extremes of each data set.

We fix the site on the river Lune to be our reference site and consider the extremal dependence between this and all other gauges. We first estimate the extremal dependence measure χ and the coefficient η using the upper 10% of the corresponding joint tails. Both χ and η are limiting values; however, in practice, we are unable to evaluate such limits without a closed form for the joint distribution. We therefore calculate these values empirically. Taking χ, for example, an estimate is given by $\hat{\chi}_q = \hat{\Pr}(X > \hat{x}_q, Y > \hat{y}_q)/\hat{\Pr}(X > \hat{x}_q)$, where $\hat{\Pr}(\cdot)$ denotes an empirical probability estimate and $\hat{x}_q$ and $\hat{y}_q$ denote empirical $100q\%$ quantiles estimates for the variables X

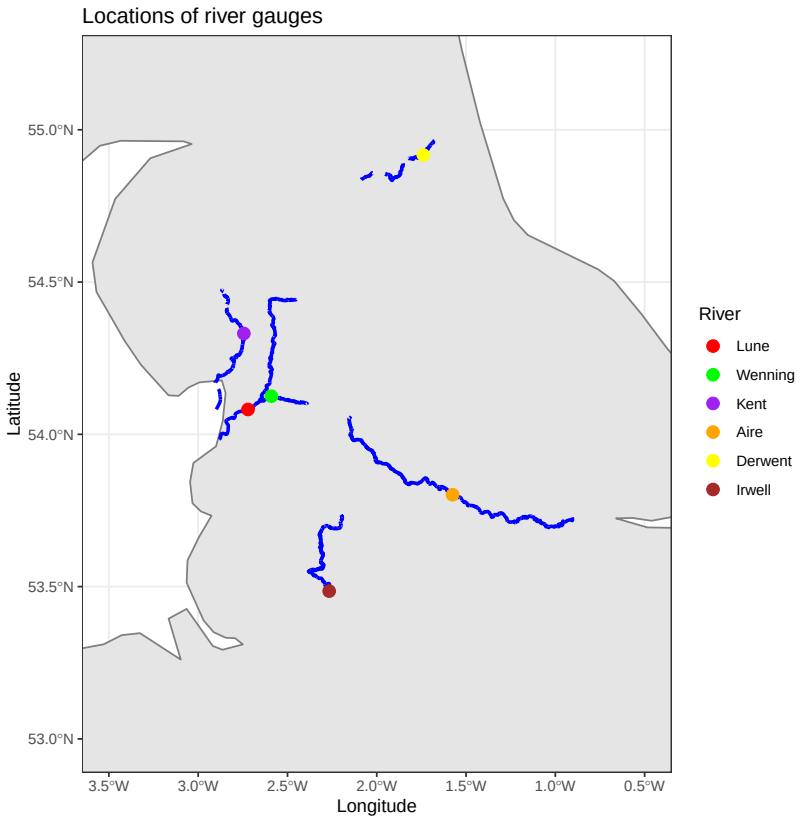

Figure 5.5.1: Locations of river gauges in the north of England, UK. Individual rivers illustrated in blue alongside the corresponding gauge locations.

and Y, respectively, and q is some value close to 1. Specifically, we take $q = 0.9$. In practice, we are unlikely to observe $\chi = \hat{\chi}_q$, even at the most extreme quantile levels, i.e., as $q \to 1$. This can be problematic when trying to quantify the form of extremal dependence, since $\hat{\chi}_q > 0$ may arise for asymptotically independent data sets (for example). Therefore, the estimated coefficients should act only as a rough guide for this quantification.

The dependence measure estimates and 95% confidence intervals are shown in Figure 5.5.2 as a function of distance from the reference site. Here and throughout, all confidence intervals are obtained via block bootstrapping with block size $b = 40$; this value appears appropriate to account for the varying degrees of temporal dependence observed across the six gauge sites. These estimates suggest that asymptotic independence may exist for at least three of the site pairings; therefore, modelling techniques

based on bivariate regular variation would likely fail to capture the observed extremal dependencies in this scenario.

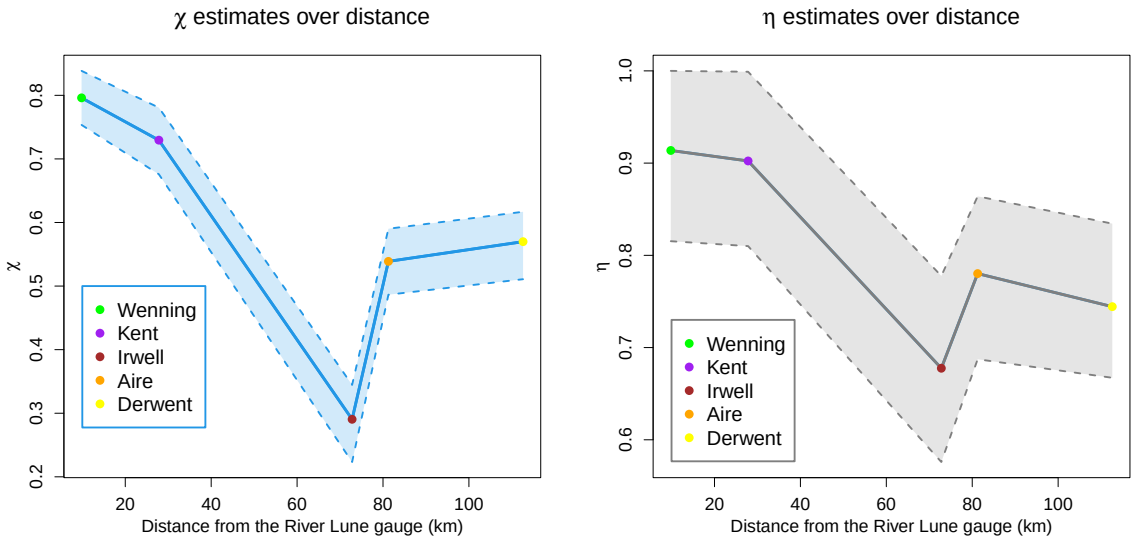

Figure 5.5.2: Estimated dependence coefficients as a function of distance from the Lune gauge, with 95% pointwise confidence intervals given by the shaded regions. Left: Estimates of χ (blue). Right: Estimates of η (grey).

5.5.2 ADF Estimation

We transform each marginal data set to exponential margins using the semi-parametric approach of Coles and Tawn (1991), whereby a GPD is fitted to the upper tail and the body is modelled empirically. The GPD thresholds are selected using the technique proposed in Varty et al. (2021). Diagnostic plots found in the Supplementary Material indicate good model fits. Since our results from Section 5.4 suggest that the estimators $\hat{\lambda}_{CL2}$ and $\hat{\lambda}_{ST}$ perform best overall, we used these, alongside the base estimator $\hat{\lambda}_H$, to estimate the ADF for each combination of the reference gauge and the other five gauges. The resulting ADF estimates can be found in Figure 5.5.3.

For the most part, there is decent agreement across the estimators. One can observe contrasting shapes across the different pairs of gauges, illustrating the variety in the observed extremal dependence structures. We note that the estimates obtained using $\hat{\lambda}_H$

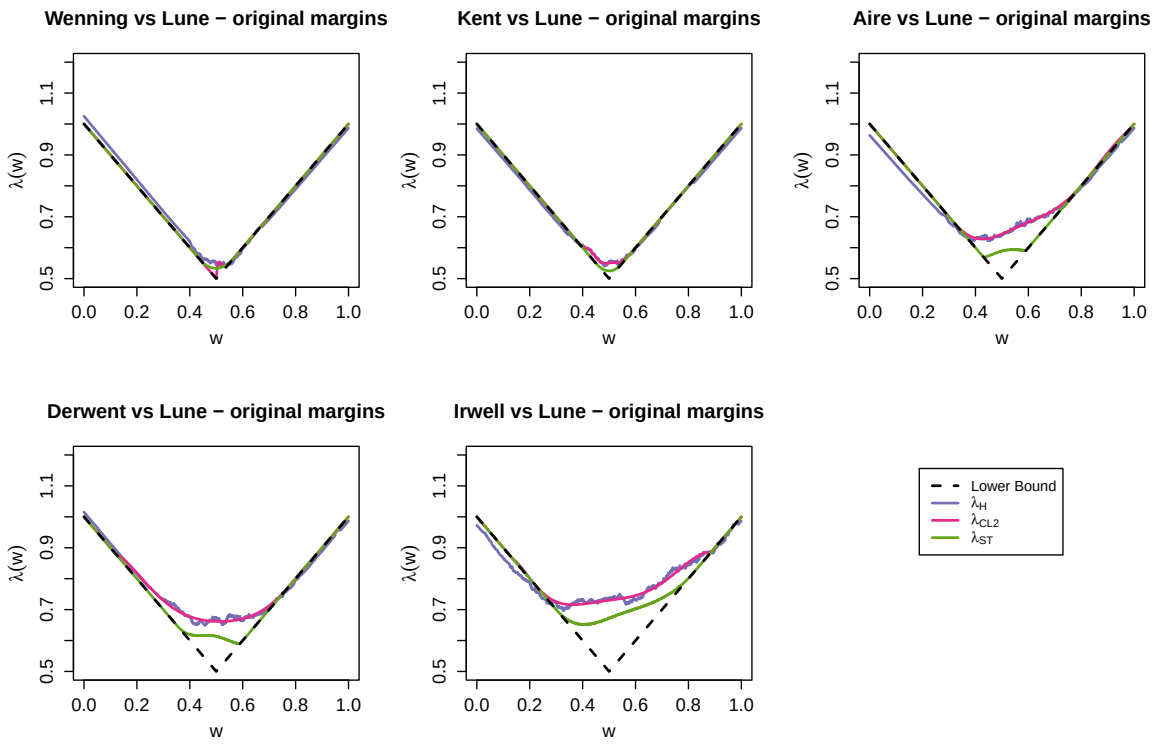

Figure 5.5.3: ADF estimates for each pair of gauge sites. The purple, pink and green lines represent the estimates from $\hat{\lambda}_H$, $\hat{\lambda}_{CL2}$ and $\hat{\lambda}_{ST}$, respectively, with the theoretical lower bound denoted by the black dotted lines.

demonstrate some of the drawbacks mentioned in Section 5.2, such as non-smoothness and not respecting theoretical results for the ADF. Moreover, these results illustrate that on the whole, the estimator $\hat{\lambda}_{CL2}$ is very much a smoothed version of $\hat{\lambda}_H$, owing to the form of likelihood function used.

5.5.3 Assessing goodness of fit for ADF estimates

Recall that, from equation (5.2.1), we have $T_w^* \sim \text{Exp}(\lambda(w))$ as $u_w \to \infty$ for all $w \in [0,1]$. We exploit this result to assess the goodness of fit for ADF estimates.

Let $\hat{\lambda}(w), w \in [0,1]$, denote an estimated ADF obtained using the sample $\{(x_i, y_i) : i = 1, \ldots, n\}$. Given $w \in [0,1]$, let u_w denote some high empirical quantile from the sample $\mathbf{t}_w$, and consider the sample $\mathbf{t}_w^*$, with $\mathbf{t}_w$ and $\mathbf{t}_w^*$ defined as in Section 5.3.2. If $\mathbf{t}_w^*$ is indeed a sample from an $\text{Exp}(\hat{\lambda}(w))$ distribution, we would expect good agreement

between the empirical and model quantiles. Letting $n_w = |\mathbf{t}_w^*|$ and $\mathbf{t}_{w(j)}^*$ denote the j-th order statistic of $\mathbf{t}_w^*$, we consider the set of pairs

$$\left\{ \left(\frac{-\log(1 - j/(n_w + 1))}{\hat{\lambda}(w)}, \mathbf{t}_{w(j)}^* \right) : j = 1, \ldots, n_w \right\},$$

for different rays $w \in [0, 1]$, corresponding to a range of joint survival regions. With u_w fixed to be the 90% empirical quantile of $\mathbf{t}_w$, quantile-quantile (QQ) plots for five individual rays, $w \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$, are given in Figure 5.5.4 for the first pair of gauges and the $\hat{\lambda}_{CL2}$ estimator. Uncertainty intervals are obtained via block bootstrapping. On the whole, the estimated exponential quantiles appear in good agreement with the observed quantiles, indicating the underlying ADF estimate accurately captures the tail behaviour for each min-projection variable. Analogous plots for $\hat{\lambda}_H$ and $\hat{\lambda}_{ST}$ are given in the Supplementary Material. Similar plots were obtained when the other pairs of gauges were considered.

5.5.4 Estimating return curves

To quantify the risk of joint flooding events across sites, we follow Murphy-Barltrop et al. (2023) and use the ADF to estimate a bivariate risk measure known as a return curve, as defined in Section 5.1. This measure is the direct bivariate extension of a return level, which is commonly used to quantify risk in the univariate setting (Coles, 2001). Taking probability values close to zero gives a summary of the joint extremal dependence, thus allowing for comparison across different data structures. In the context of extreme river flows, return curves can be used to evaluate at which sites joint extremes (floods) are more/less likely to occur. For illustration, we fix p to correspond to a 5 year return period, i.e., $p = 1/(5n_y)$, where n_y is the average number of points observed in a given year (Brunner et al., 2016). Excluding missing observations, we have 28 years of data, hence the resulting curve should be within the range of data

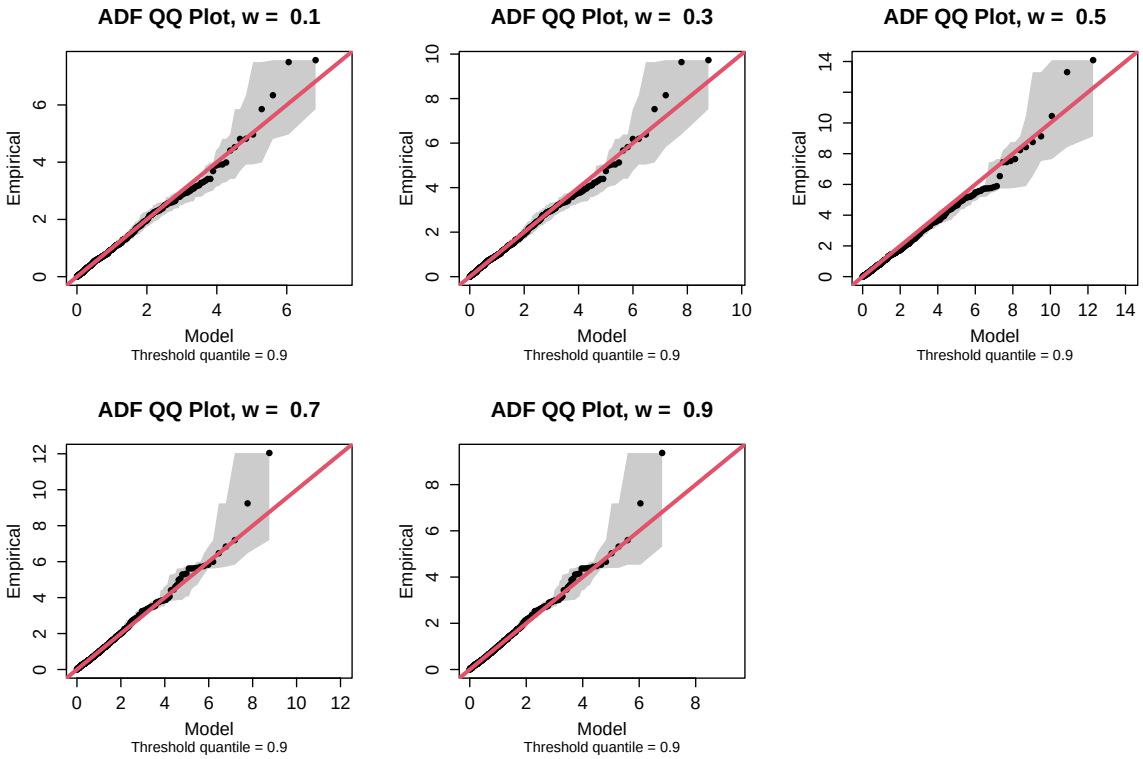

Figure 5.5.4: Individual ADF QQ plots for $w \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ for first pair of gauges, using the ADF estimate obtained via $\hat{\lambda}_{CL2}$. Estimates given in black, with 95% pointwise confidence intervals represented by the grey shaded regions. Red lines correspond to the $y = x$ line.

whilst simultaneously representing the joint tail. The resulting return curve estimates for each ADF estimator and pair of gauge sites can be found in Figure 5.5.5.

There is generally good agreement among the estimated curves. The almost-square shapes of the estimates for the first two pairs of gauges indicate higher likelihoods of observing simultaneous flood events at the corresponding gauge sites; this is as expected given the close spatial proximity of these sites. In all cases, the curves derived via $\hat{\lambda}_H$ are quite rough and unrealistic, and are subsequently ignored. To assess the goodness of fit of the remaining return curve estimates, we consider the first and fifth examples and apply the diagnostic introduced in Murphy-Barltrop et al. (2023). Our results suggest good quality model fits for both of the estimates obtained using $\hat{\lambda}_{CL2}$ and $\hat{\lambda}_{ST}$. Furthermore, we also obtain 95% return curve confidence intervals for these examples.

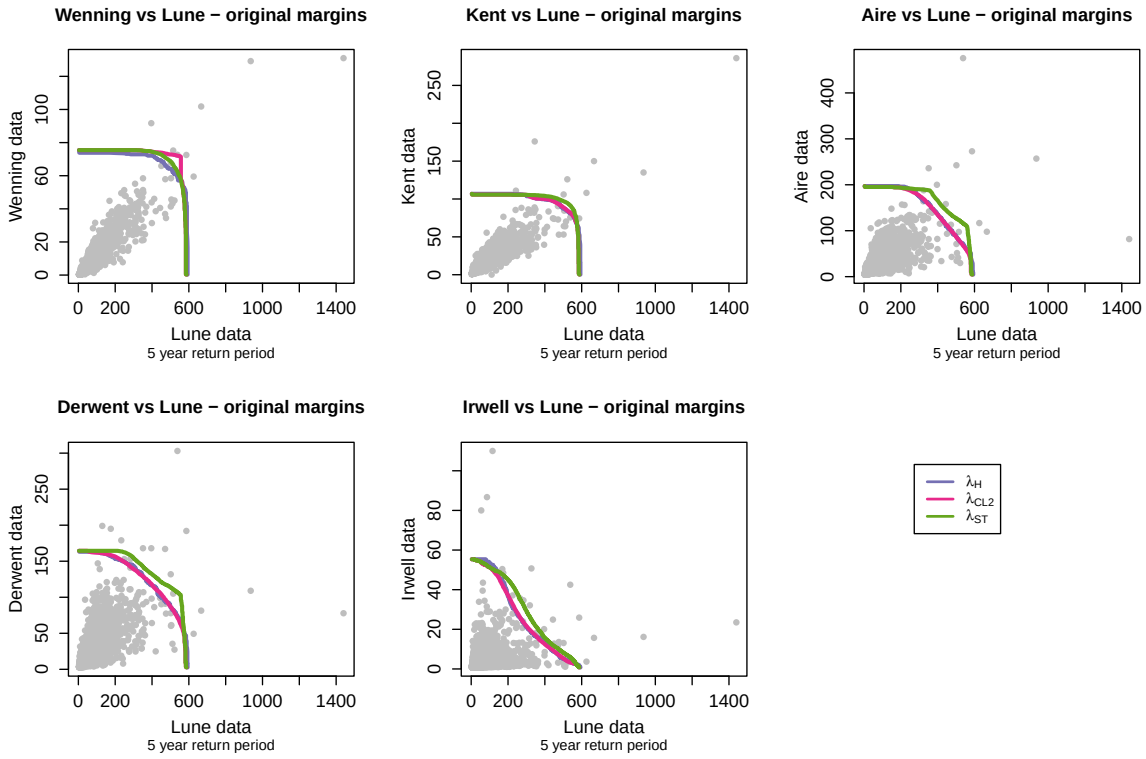

Figure 5.5.5: Estimated 5-year return curves (on original margins) for each pair of gauges. The purple, pink and green lines represent the curve estimates from $\hat{\lambda}_H$, $\hat{\lambda}_{CL2}$ and $\hat{\lambda}_{ST}$, respectively.

The resulting plots illustrating the diagnostics and confidence intervals, along with a brief explanation of the diagnostic tool, are given in the Supplementary Material.

5.6 Discussion

We have introduced a range of novel global estimators for the ADF, as detailed in Section 5.3. We compared these estimators to existing techniques through a systematic simulation study and found our novel estimators to be competitive in many cases. In particular, the estimators derived via the composite likelihood approach of Section 5.3.1, alongside the estimator of Simpson and Tawn (2022), appear to have lower bias and variance, on average, compared to alternative estimation techniques. We also applied ADF estimation techniques to a range of river flow data sets, and obtained estimates

of return curves for each data set. The results suggest that our estimation procedures are able to accurately capture the range of extremal dependence structures exhibited in the data, allowing for a more robust risk analysis of joint flooding events.

From Section 5.4, one can observe that the 'combined' estimators proposed in Section 5.3.3 outperform their 'uncombined' counterparts in the majority of instances. This indicates that incorporating the knowledge obtained from the conditional extremes parameters leads to improvements in ADF estimates. Furthermore, in most cases, ADF estimates obtained via approximations of the set G appeared to have lower bias compared to alternative estimation techniques. More generally, these results suggest that inferential techniques that exploit the results of Nolde and Wadsworth (2022) are superior to techniques which do not. Estimation of G, and its impact on estimation of other extremal dependence properties, represents an important line of research.

As noted in Section 5.1, few applications of the modelling framework described in equation (5.1.4) exist, even though this model offers advantages over the widely used approach of Heffernan and Tawn (2004) when evaluating joint tail probabilities. Inference via the ADF ensures consistency in extremal dependence properties, and one can obtain accurate estimates of certain risk measures, such as return curves.

For each of the existing and novel estimators introduced in Sections 5.2 and 5.3, we were required to select quantile levels, which is equivalent to selecting thresholds of the min-projection. With the exception of $\hat{\lambda}_{ST}$, similar quantile levels were considered for each estimator so as to provide some degree of comparability. However, due to variation in estimation procedures, we acknowledge that the selected quantile levels are not readily comparable since the quantity of joint tail data used for estimation varies across different estimators. Moreover, as noted in Section 5.3.4, trying to select 'optimal' quantile levels appears a fruitless exercise since the performance of each estimator does not appear to alter much across different quantile levels.

Finally, we acknowledge the lack of theoretical treatment for the proposed ADF

estimators which is an important consideration for understanding properties of the methodology. However, theoretical results of this form typically require in-depth analyses and strict assumptions, which themselves may be hard to verify, whilst in practice one can only ever look at diagnostics obtained from the data. We have therefore opted for a more practical treatment of the proposed estimators.

Chapter 6

Discussion

The aim of this thesis was to address two questions: how does one define extreme risk in a bivariate setting, and how does one estimate such risks? We believe these questions have been comprehensively addressed through the development of a number of novel estimation techniques for extremal dependence structures and return curves. Several of the modelling problems considered had previously been given little or no attention within the literature, and where methodology existed, the proposed methods were shown to outperform existing techniques in many cases.

This project was part-funded by the ONR, who provided the initial motivation for investigating extreme risks in a bivariate setting. As discussed in Chapter 1.1, accounting for joint risks from two (or more) variables is crucial for ensuring safe nuclear operations. In this context, the methods introduced in this thesis could allow for a more realistic and robust evaluation of joint extremes in a changing climate. For example, return curve estimates could be used to inform the design basis for future nuclear installations and thereby improve nuclear regulatory practices. An executive summary of the thesis will be published on the ONR's website upon completion, alongside a software package for obtaining return curve estimates. We expect these outputs to increase both the impact and exposure of the research.

There are several ways in which our work could be extended. Perhaps most obviously, we have restricted attention to the bivariate setting throughout. This choice was made for practical reasons, not least because quantifying the extremal dependence structures of multivariate random vectors becomes increasingly complicated as the dimension d increases, since different m-dimensional sub-vectors, where $m < d$, can exhibit different forms of extremal dependence. Furthermore, in the context of nuclear regulation, the quantification of joint extreme risks is a relatively recent area of consideration that has been largely motivated by the Blayais and Fukushima incidents. Applying the principle of parsimony, it is therefore best to first thoroughly consider the problem of joint risk quantification in the bivariate setting.

Given careful consideration, we believe all of the proposed methodology could be extended to the general multivariate setting. Take the trivariate setting for example: given a random vector (X, Y, Z) and small probability p, a p-probability return curve is given by the set

$$\mathrm{RC}(p) := \left\{(x, y, z) \in \mathbb{R}^3 : \Pr(X > x, Y > y, Z > z) = p\right\}.$$

Return curve definitions for higher dimensions ($d \geq 4$) are analogous. However, we note that both visualisation and interpretation become increasingly complex, especially for dimension $d \geq 4$. In such cases, one cannot produce meaningful return curve plots; such plots are essential for illustrating extreme risks in practical settings. Furthermore, as the dimension d increases, larger sample sizes are typically required to identify joint extreme events due to more complex estimation procedures. This is not always possible in practical settings, where sample sizes are often limited by observation periods.

These remarks raise several important questions: when is it beneficial to obtain extremal risk measures for three (or more) variables, and how can one estimate, utilise and interpret such measures? Of course, the answers to these questions are likely to be highly context dependent, and also require some degree of expert judgement.

Figure 6.0.1 illustrates a three-dimensional return curve for $n = 3,000$ data points simulated from a trivariate Gaussian copula on standard exponential margins with covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0.2 & 0.8 \\ 0.2 & 1 & 0.5 \\ 0.8 & 0.5 & 1 \end{pmatrix},$$

mean vector $\boldsymbol{\mu} = (0, 0, 0)$, and $p = 1/1000$. In this case, the curve is given by a surface in $\mathbb{R}^3$, and the probability of observing points in the joint survivor set $(x, \infty) \times (y, \infty) \times (z, \infty)$ is equal to p for any point on the curve.

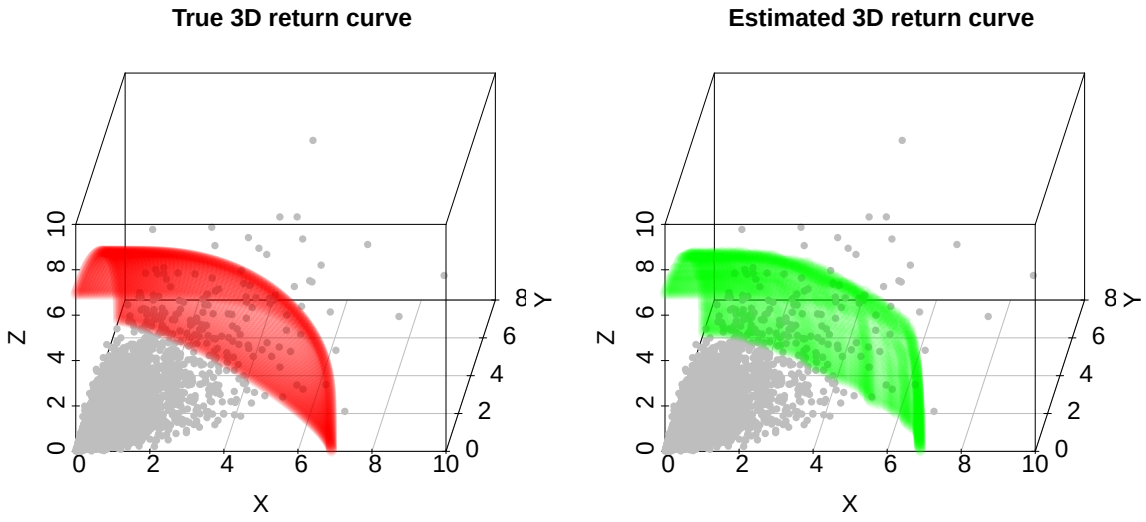

Figure 6.0.1: True (left) and estimated (right) return curves for a trivariate Gaussian sample on standard exponential margins with $p = 1/1000$.

By extending the modelling frameworks we have considered to the general multivariate setting, estimates of higher dimensional return curves can be obtained. In particular, the model of Wadsworth and Tawn (2013), which has been central to the methodology developed in this thesis, can be extended to an arbitrary number of dimensions under similar assumptions as in the bivariate case. In the trivariate case, for

any random vector (X, Y, Z) on standard exponential margins and $(w_1, w_2) \in [0, 1]^2$ satisfying $w_1 + w_2 \leq 1$, the model is defined as

$$\Pr(\min\{X/w_1, Y/w_2, Z/(1 - w_1 - w_2)\} > t) = L(e^t \mid w_1, w_2)e^{-\lambda(w_1,w_2)t},$$

as $t \to \infty$, with $\lambda(w_1, w_2) \geq \max(w_1, w_2, 1 - w_1 - w_2)$. Here, λ denotes the two-dimensional ADF describing the extremal dependence structure. Similar to the bivariate setting, the Hill estimator can be used to estimate λ and hence approximate the return curve. An example curve estimate is illustrated for the trivariate Gaussian copula in Figure 6.0.1; the curves appear similar in both shape and magnitude to the true curve.

These observations lead us to believe that most of the statistical procedures we have introduced, both for return curves and extremal dependence structures in both stationary and non-stationary settings, can be extended to higher dimensions. Of course, there are many inferential challenges that are likely to arise due to the increased number of parameters that must be estimated and the increased complexity of the extremal dependence structures. Such challenges offer scope for both novel and invaluable statistical techniques, thus presenting natural avenues for further work.

Alongside extensions to higher dimensions, there is also scope for further work on the estimation of the non-stationary ADF. The techniques introduced in Chapter 4 account for the fact that, unlike in the stationary setting, we are unlikely to have repeated observations of min-projection variables for a fixed covariate realisation. The proposed framework uses quantile regression to obtain pointwise estimates, which are then smoothed using Bernstein-Bézier polynomials and finally corrected in an ad-hoc manner to ensure theoretical properties of the ADF still hold.

We believe that some of the techniques introduced in Chapter 5 may be useful for improving estimation of the non-stationary ADF. In particular, in the stationary setting, we found that ADF estimators derived via a composite likelihood approach were

favourable in many cases. Since this approach involves multiplying different likelihood components together, regardless of whether they are independent, it could potentially be adapted for non-stationary case by multiplying likelihood functions over both rays and covariates. Moreover, we believe there is scope to formalise the techniques introduced in Chapter 4 for imposing the necessary theoretical constrains on the non-stationary ADF.

Throughout this thesis, we have used the Wadsworth and Tawn (2013) framework extensively for modelling the joint tail of bivariate data sets and estimating return curves. In every such instance, we were required to select quantiles of the min-projection variable prior to estimating the angular dependence function. This is analogous to selecting a threshold for the peaks-over-threshold technique, since selecting quantiles too low (high) will result in higher bias (variability). Statistical techniques for selecting these quantiles remains an unexplored area of research, and several of the proposed univariate threshold selection methods (e.g. Wadsworth, 2016; Varty et al., 2021) could likely be adapted for this setting.

Chapter 5 provided the additional interesting and useful insight that exploitation of the limit set theory developed by Nolde and Wadsworth (2022) leads to improved estimates of the extremal dependence structure in a wide variety of settings. In particular, the estimator of the ADF linked to G, where G denotes the unit level set defined in chapter 5.2, was preferable when compared to many alternative estimators. Thus far, only two approaches (Simpson and Tawn, 2022; Wadsworth and Campbell, 2022) exist for estimating G. We hope our results will provide motivation for future research on this topic, as well as further consideration of how the results of Nolde and Wadsworth (2022) can be applied in practical settings.

Finally, throughout this thesis, we chose to exclusively consider the bivariate risk measure known as a return curve. This was mainly because return curves provide a natural extension to return levels, which are the most widely used risk measure in the

univariate setting. However, as noted in Chapter 2.4, many alternative risk measures have been proposed, such as those described in Serinaldi (2015) and Haselsteiner et al. (2021). Similar to return curves, literature on the estimation of many of these risk measures is sparse, and similar theory from bivariate extreme value theory is likely to be applicable. Many avenues for future research therefore exist - though we recommend that developments are made with caution to ensure the resulting estimates offer utility within practical applications.

Appendix A

Appendix

A.1 Return curve estimation via the Wadsworth and Tawn (2013) model under asymptotic dependence

As noted in Section 3.3.4.2, whilst asymptotic dependence represents a boundary case for the Wadsworth and Tawn (2013) model, this framework can still be used to estimate return curves for data sets exhibiting this form of dependence. To see this, first assume that we have estimated $\hat{\lambda}(w) = \max(w, 1-w)$ for all $w \in [0,1]$ and let p, p^* and $\mathcal{W}$ be defined as in Section 3.3.4.2. Given any $w \in \mathcal{W}$, we set $t = -\log(p/p^*)/\max(w, 1-w)$ and compute a return curve point estimate as $(w(t+u), (1-w)(t+u))$. Whilst the values of t will, in theory, be the same for all asymptotically dependent data structures, the same is not true for the values of u required for the return curve point estimates.

We refer to u as the min-projection threshold since u is the value satisfying $\Pr(T_w > u) = p^*$, with $T_w = \min\{X/w, Y/(1-w)\}$. Consequently, u will vary depending on the distribution of T_w, which in turn will vary depending on the underlying copula distribution. In particular, unlike the angular dependence function, the distribution of

T_w will not be constant for all asymptotically dependent data structures.

As a demonstration, let (X, Y) be a random vector from the logistic distribution (Gumbel, 1960) on standard exponential margins with dependence parameter $\alpha \in (0, 1)$; see Section 2.2.3. For all $\alpha \in (0, 1)$, this random vector exhibits asymptotic dependence. Furthermore, it is straightforward to show that

$$\begin{aligned}\Pr(T_w > u) = e^{-wu} + e^{-(1-w)u} - 1 \\ + \exp\left\{-\left[\left(-\log(1-e^{-wu})\right)^{1/\alpha} + \left(-\log(1-e^{-(1-w)u})\right)^{1/\alpha}\right]^{\alpha}\right\}.\end{aligned}$$

Consequently, the min-projection quantiles will vary depending on the dependence parameter α. As such, return curve point estimates obtained using the Wadsworth and Tawn (2013) modelling framework will not be constant for asymptotically dependent data structures, even though the angular dependence function does not change in such cases.

We note that this boundary case is also relevant for the curve estimates obtained in Chapters 4 and 5. For the former chapter, where we consider non-stationarity in the extremal dependence structure, min-projection thresholds are obtained using quantile regression. In practice, this allows us to capture non-stationarity in return curves when asymptotic dependence in present, i.e., when the non-stationary angular dependence function is fixed at the lower bound.

On a final note, we remark that in cases where one suspects asymptotic dependence may be present, it may be more appropriate to use return curve estimation techniques based on the framework of regular variation. This is because we are unlikely to estimate $\hat{\lambda}(w) = \max(w, 1-w)$ exactly for all $w \in [0, 1]$ due to this being a boundary case for the angular dependence function. This could result in the asymptotic behaviour of the joint tail being misspecified, leading to poor quality return curve estimates. As noted in Section 3.3.3, a range of techniques for return curve estimation using

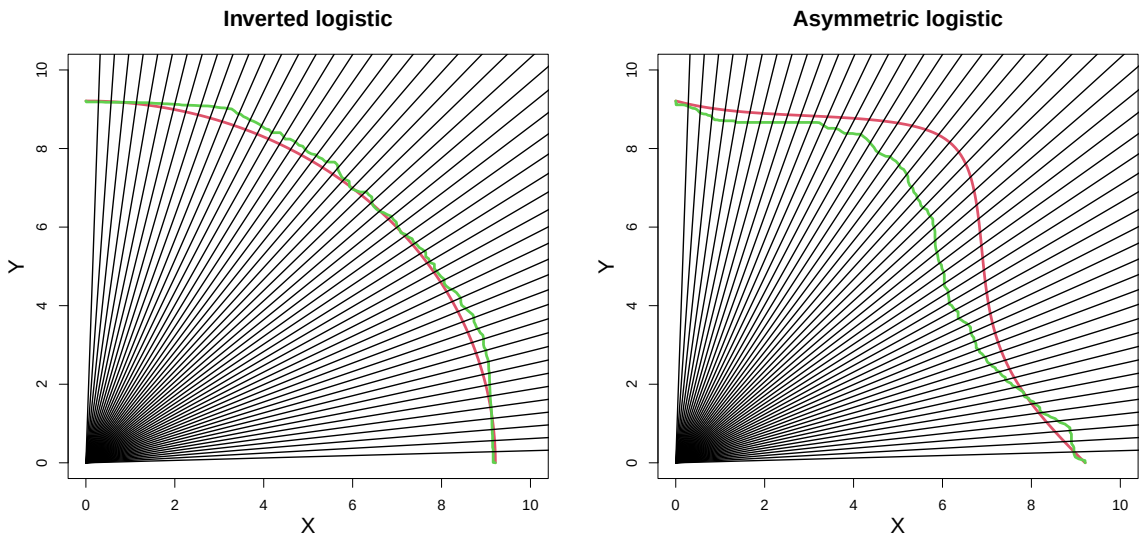

Figure A.2.1: Resulting line segments from a set of angles $\boldsymbol{\Theta}$ with $m = 25$ for both copula examples. True and estimated curves given in red and green, respectively.

the regular variation framework have been proposed; see, for instance, Salvadori and Michele (2004), Marcon et al. (2017a) and Cooley et al. (2019).

A.2 Supplementary material for Chapter 3

A.2.1 Angles and lines

Examples of angles $\theta \in \boldsymbol{\Theta}$ and corresponding line segments L_θ are illustrated in Figure A.2.1 for both copula examples discussed in Section 3.4.1. From Figure A.2.1, one can observe that each line segment intersects both the estimated and true return curves exactly once; this follows from the definition of a return curve. In this manner, each angle and corresponding line segment represents a common feature of both curves.

A.2.2 Diagnostic example

In Figure A.2.2, we illustrate the diagnostic tool at different probabilities. The sample size is fixed at $n = 10^4$ and four probabilities are considered ($p \in \{10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$).

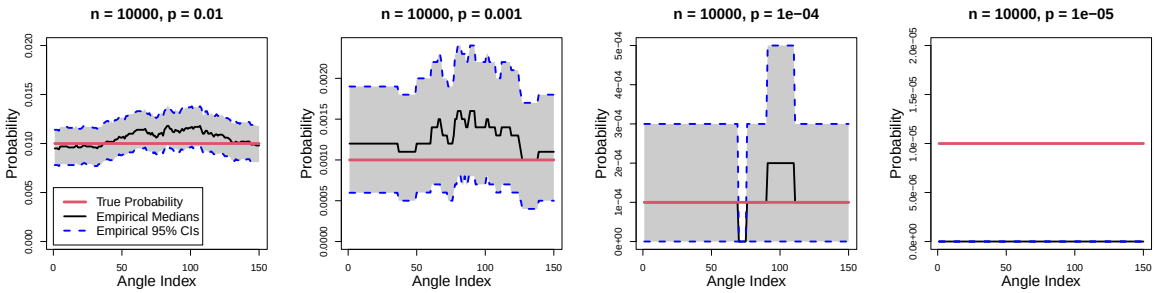

Figure A.2.2: Diagnostic tool for four p values with logistic copula sample and $n = 10000$. Red, black, and dotted blue lines represent true values, estimated medians, and estimated 95% confidence intervals, respectively, for each probability and index j.

A.2.3 Summary statistic of integrated absolute difference

An example of the plot corresponding to the summary statistic used to evaluate bias in return curve estimates is illustrated in the left panel of Figure A.2.3 for a standard bivariate normal copula with $\rho = 0.6$, along with the median return curve estimates from each model in the right panel. Letting d_{θ_j} and $\hat{d}_{\theta_j}$, $j = 1, \ldots, 150$, denote the l_2-norm values of true and estimated median curves for angles $\theta_j \in \boldsymbol{\Theta}$, the summary statistic is given by

$$A(d, \hat{d}) = \sum_{j=1}^{150} |d_{\theta_j} - \hat{d}_{\theta_j}|.$$

The closer this quantity is to zero, the closer the norm values are to the truth and hence the nearer the median curve estimates are to the true curve. Moreover, an area of zero corresponds to an unbiased curve estimate since this implies there is no difference between the estimated and true curves (at the points corresponding to angles in $\boldsymbol{\Theta}$). One can observe the median curve estimate obtained using the Cooley et al. (2019) framework appears to perform poorly for this particular example.

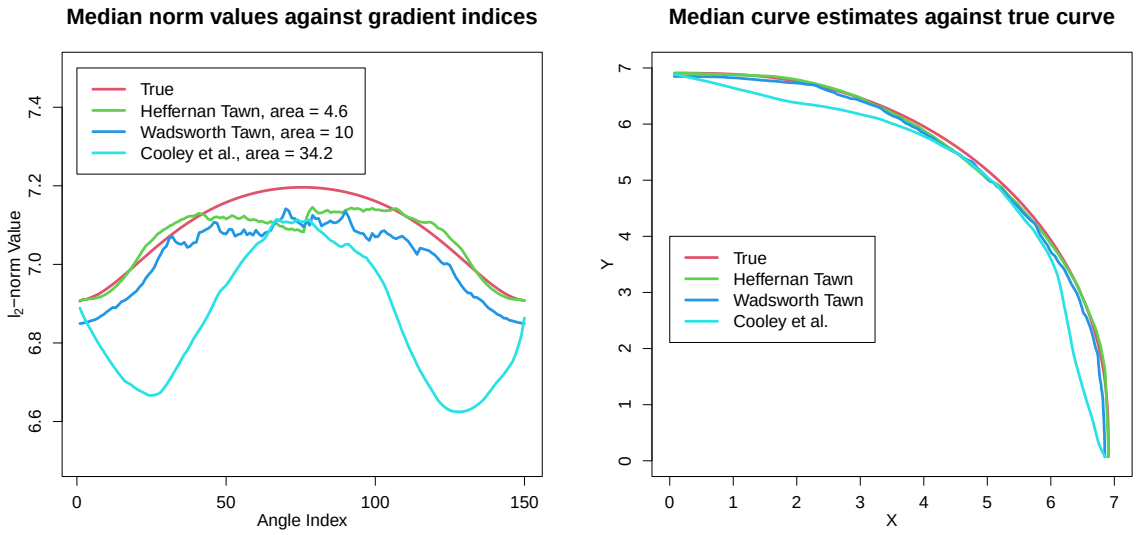

Figure A.2.3: Left: median d_θ estimates obtained using 50 simulated samples from a standard bivariate normal copula with $\rho = 0.6$ and true norm values against angle indices. Summary statistics given in plot legend. Right: median curve estimates from each model against true curve for the same example. True values are given in red while the estimated values from the Heffernan and Tawn (2004), Wadsworth and Tawn (2013) and Cooley et al. (2019) models are given in green, dark blue and light blue, respectively.

A.2.4 Illustration of procedure for estimating coverage

The left panel Figure A.2.4 illustrates the five angles used to evaluate coverage. We label these angles 1-5, clockwise from the y-axis. Angle 3 is close in value to $\pi/4$, resulting in a line similar in appearance to $y = x$. Angles 4 and 5 are obtained by applying the function $f(\theta) = \pi/2 - \theta$ to angles 1 and 2, respectively, and hence they can, in a sense, be considered symmetric.

For each simulated sample and angle θ, a confidence region is obtained for estimated coordinates on the line segment L_θ. One can record whether the true point at this angle lies within the estimated region; for an unbiased curve estimate, this would be expected $100(1-\alpha)\%$ of the time. An example confidence region is illustrated in the right panel of Figure A.2.4; we note that the true point (pink) lies within the region. Repeating this procedure over the 500 simulated samples, the proportion of times the true points lie

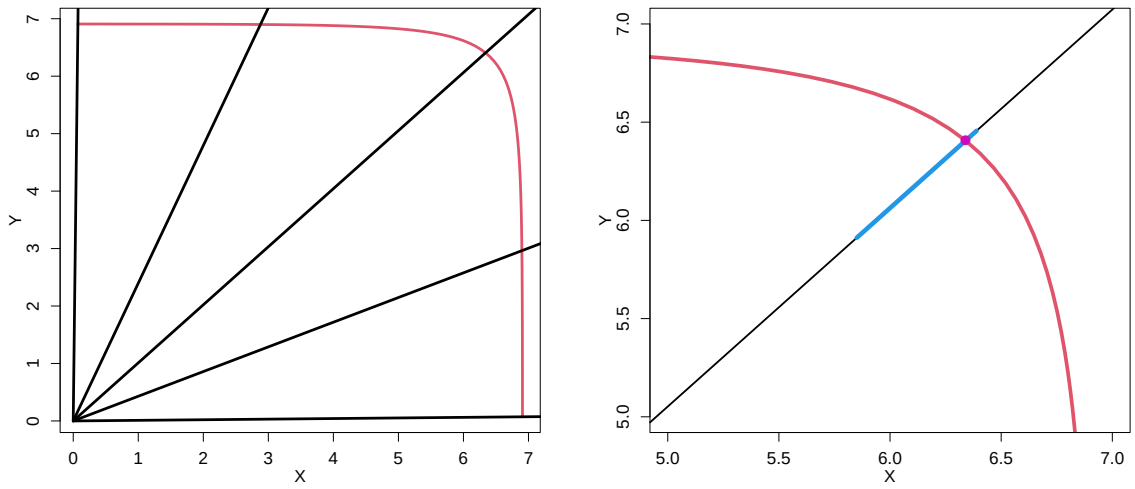

Figure A.2.4: Left: Line segments, L_θ, corresponding to the angles considered for evaluating coverage. Right: Confidence region computed for one sample at the third angle. The blue line represents the estimated 95% confidence region for l_2-norm values along the corresponding line segment. True return curve (red) in both plots obtained from the logistic copula with dependence parameter 0.5 and $p = 10^{-3}$.

within the estimated confidence regions can be computed, giving an estimated measure of coverage at each of the angles 1-5.

A.2.5 Additional coverage results

Table A.2.1 gives the additional coverage results for each copula considered in Section 3.5, with $p = 10^{-4}$.

A.2.6 Estimated dependence parameters for metocean data sets

Figure A.2.5 illustrates block bootstrapped α and β parameter estimates for both of the data sets considered in the case study of the article. Here, the parameters α and β correspond to the formulation of the conditional extremes model given in Keef et al. (2013a), for which we set the normalising functions $a(x) = \alpha x, \alpha \in [-1, 1]$, and

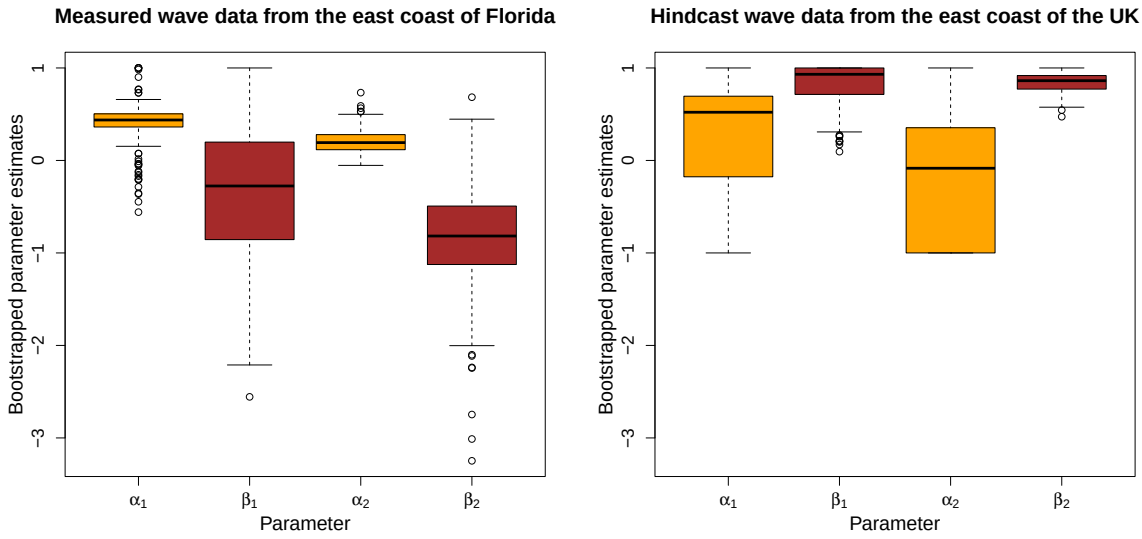

Figure A.2.5: Boxplots of block bootstrapped α and β parameter estimates for both metocean data sets. In both cases, the subscripts 1 and 2 correspond to the variable selected for conditioning: specifically, 1 (2) indicates that the variable plotted on the x-axis (y-axis) of Figure 3.1.2 was used for conditioning. The orange and brown shadings correspond to the α and β parameters, respectively.

$b(x) = x^{\beta}, \beta \in (-\infty, 1]$. For both data sets, 95% empirical quantiles of the conditioning variables were used for model fitting. As can be observed, the variability in parameter estimates makes it difficult to identify the form of extremal dependence present within each data set. In particular, it is not clear whether either data set exhibits asymptotic dependence, i.e., $\alpha = 1$ and $\beta = 0$, from these parameter estimates.

Block bootstrapping was also used to obtain median and 95% confidence interval estimates of the angular dependence function described in Equation (3.2.3); these estimates are illustrated in Figure A.2.6 for both metocean data sets. As noted in Section 3.3.4.2, estimation of the angular dependence function is carried out pointwise using the Hill estimator. Similarly to the conditional extremes parameters, it is difficult to identify the exact forms of extremal dependence exhibited by each data set. One can conclude, however, that a higher degree of positive dependence is present for the hindcast data set, as evidenced by the estimates angular dependence functions being closer to the lower bound. This is what one might expect given the shapes of the data clouds

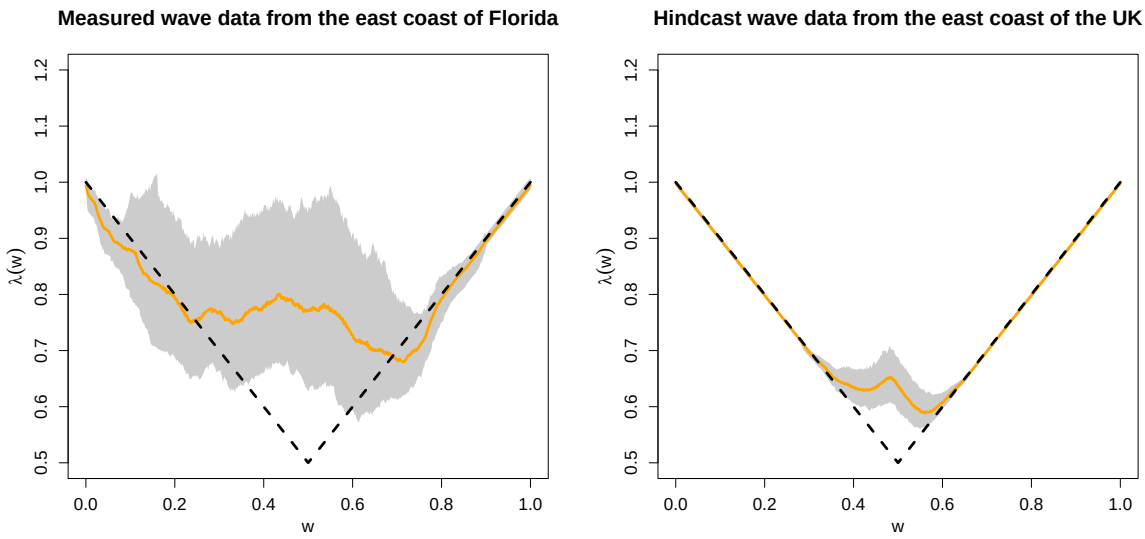

Figure A.2.6: Median and 95% confidence interval estimates of the angular dependence function obtained using block bootstrapping for both metocean data sets. In both panels, the orange thick and black dotted lines represent the median estimate and lower bound $\max(w, 1-w), w \in [0,1]$, respectively, while the grey regions illustrate the 95% confidence intervals.

in Figure 3.1.2. Furthermore, one can observe the contrast in variability between the two data sets that arises due to differences in sample sizes.

A.2.7 Illustration of reference points and corresponding line segments

Figure A.2.7 illustrations the reference points (x_0, y_0) and corresponding line segments for both data sets considered in the case study of the article. The points at which the line segments intersect return curve estimates are used to quantify uncertainty and define joint survival regions for the diagnostic tool. One can observe that in both cases, defining line segments from the origin $(0,0)$ would result in procedures whereby return curve estimates are evaluated outside of the region where data has been observed: such an evaluation is not meaningful in practice.

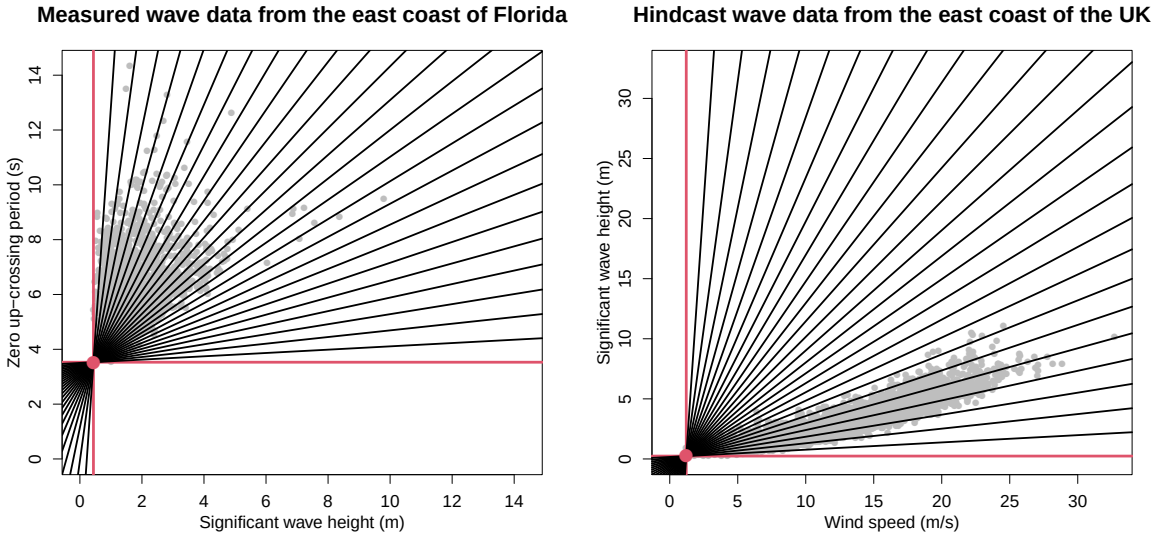

Figure A.2.7: Reference points (x_0, y_0) (red) for measured (left) and hindcast (right) data sets, alongside corresponding line segments (black) L'_θ for $\theta \in \mathbf{\Theta}$, with $m = 25$.

Table A.2.1: Coverage values of 95% confidence regions for $p = 10^{-4}$. 'HT' and 'WT' correspond to the Heffernan and Tawn (2004) and Wadsworth and Tawn (2013) models, respectively.

Copula	Probability	$p = 10^{-4}$									
	Model	HT					WT				
BEV Logistic	Angle	1	2	3	4	5	1	2	3	4	5
	Coverage	0.010	0.970	0.688	*	*	0.844	0.554	0.846	*	*
BEV Asymmetric Logistic	Angle	1	2	3	4	5	1	2	3	4	5
	Coverage	0.592	0.936	0.496	0.846	0.954	0.902	0	0.008	0.718	0.948
Bivariate Normal 1	Angle	1	2	3	4	5	1	2	3	4	5
	Coverage	0.474	0.946	0.964	*	*	0.868	0.936	0.906	*	*
Bivariate Normal 2	Angle	1	2	3	4	5	1	2	3	4	5
	Coverage	0.942	0.928	0.936	*	*	0.936	0.938	0.934	*	*
Inverted BEV Logistic	Angle	1	2	3	4	5	1	2	3	4	5
	Coverage	0.850	0.928	0.944	*	*	0.860	0.946	0.946	*	*
Inverted BEV Asymmetric Logistic	Angle	1	2	3	4	5	1	2	3	4	5
	Coverage	0.938	0.928	0.952	0.954	0.902	0.936	0.936	0.954	0.946	0.916
Bivariate T 1	Angle	1	2	3	4	5	1	2	3	4	5
	Coverage	0.481	0.933	0.419	*	*	0.920	0.672	0.834	*	*
Bivariate T 2	Angle	1	2	3	4	5	1	2	3	4	5
	Coverage	0.765	0.756	0.854	*	*	0.938	0.554	0.176	*	*
Frank	Angle	1	2	3	4	5	1	2	3	4	5
	Coverage	0.902	0.816	0	*	*	0.920	0.024	0.01	*	*

A.3 Supplementary material for Chapter 4

A.3.1 Varying extremal dependence structures across seasons for UKCP18 dataset

Trends in the extremal dependence structure for the UKCP18 projections were considered across each of the meteorological seasons independently. For a given season, the corresponding subset of data was transformed to exponential margins using the same techniques as described in the case study of the main text (Section 5). The coefficient η was then estimated across ± 15 year rolling windows over the observation period; the resulting plots for Autumn, Winter and Spring are given in the left, centre and right panels of Figure A.3.1. These plots illustrate significantly different behaviour across these seasons, justifying our choice to just consider summer data within the case study. Moreover, summer is likely to correspond to the highest temperature and dryness values, hence it makes most sense to consider joint extremal behaviour for this season.

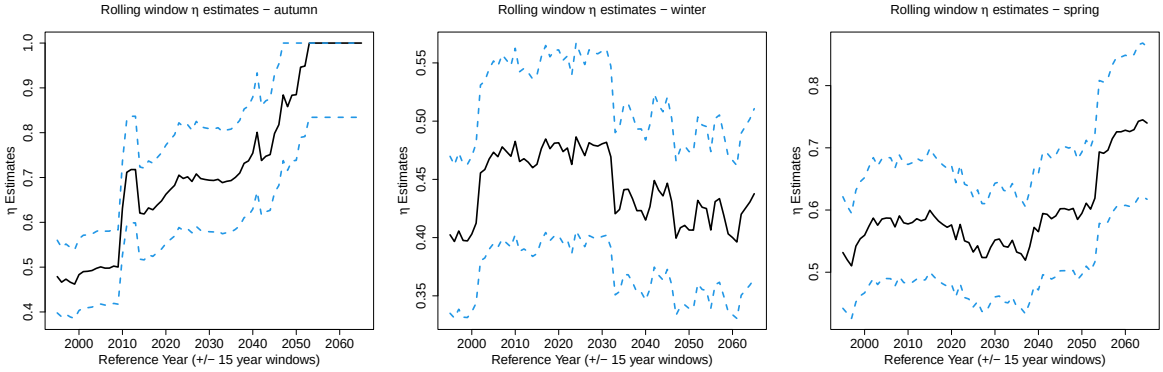

Figure A.3.1: Trends in η parameter estimates (solid black lines) over ± 15 year rolling windows for autumn, winter and spring, alongside 95% pointwise confidence intervals (dotted blue lines).

A.3.2 Additional simulation study results

A.3.2.1 MISE values of estimators

To evaluate the performance of the estimators, estimates of the mean integrated squared error (MISE) were obtained using 250 samples from each copula. Given an estimator $\bar{\lambda}^*$, the MISE at time t is given by

$$\text{MISE}(\bar{\lambda}^*(\cdot \mid \mathbf{z}_t)) = \mathbb{E}\left(\int_0^1 \left[\bar{\lambda}^*(w \mid \mathbf{z}_t) - \lambda(w \mid \mathbf{Z}_t = \mathbf{z}_t)\right]^2 \mathrm{d}w\right),$$

with smaller MISE values corresponding to estimators with lower bias and variance. Three different time points, $t = 1, t = n/2$ and $t = n$, were considered, corresponding to the start, middle and end of the simulated time frame, respectively.

Table A.3.1 gives MISE values for each estimator and copula example at each time point. One can observe the lowest MISE values are always for $\bar{\lambda}^*_{BP}$ or $\bar{\lambda}^*_{GAM}$; this is likely due to the reduced variance of these estimators compared to $\bar{\lambda}^*_{QR}$, owing to their semi-parametric forms.

A.3.2.2 ISE values of median estimators

Using 250 samples from each copula, median ADF estimators were computed pointwise over the set $\mathcal{W} = \{0, 0.01, 0.02, \ldots, 0.99, 1\}$ for three different time points: $t = 1, t = n/2$ and $t = n$, corresponding to the start, middle and end of the simulated time frame, respectively. The integrated squared error (ISE) of the median estimators is used to compare performance across $\bar{\lambda}^*_{QR}$, $\bar{\lambda}^*_{BP}$ and $\bar{\lambda}^*_{GAM}$. Although these median estimators are not computable in a given application, understanding their properties gives an insight into the bias of the estimators. Letting $\text{med}_t\bar{\lambda}^*$ denote a median estimator at time t, the ISE is given by

$$\text{ISE}(\text{med}_t\bar{\lambda}^*(\cdot \mid \mathbf{z}_t)) = \int_0^1 \left[\text{med}_t\bar{\lambda}^*(w \mid \mathbf{z}_t) - \lambda(w \mid \mathbf{Z}_t = \mathbf{z}_t)\right]^2 \mathrm{d}w,$$

Table A.3.1: MISE values for each estimator at start, middle and end of simulated time frame. Smallest MISE values in each row are highlighted in bold.

Copula	Times	MISE - $\bar{\lambda}^*_{QR}$	MISE - $\bar{\lambda}^*_{BP}$	MISE - $\bar{\lambda}^*_{GAM}$
Gaussian (Positive Correlation)	Start	0.03375	0.01012	**0.00497**
Gaussian (Positive Correlation)	Middle	0.00356	**0.0021**	0.00229
Gaussian (Positive Correlation)	End	0.01978	0.00407	**0.00143**
Gaussian (Negative Correlation)	Start	32.24126	35.76231	**30.10062**
Gaussian (Negative Correlation)	Middle	0.08607	0.08646	**0.06309**
Gaussian (Negative Correlation)	End	0.02546	0.00849	**0.00611**
Inverted Logistic	Start	0.02075	0.00541	**0.00315**
Inverted Logistic	Middle	0.00209	**0.00107**	0.00627
Inverted Logistic	End	0.03539	**0.00893**	0.02744
Inverted Hüsler-Reiss	Start	0.05748	**0.02017**	0.03849
Inverted Hüsler-Reiss	Middle	0.00177	**0.00065**	0.00089
Inverted Hüsler-Reiss	End	0.01741	0.0057	**0.00367**
Inverted Asymmetric Logistic	Start	0.02771	**0.00602**	0.01863
Inverted Asymmetric Logistic	Middle	0.00279	**0.00142**	0.01995
Inverted Asymmetric Logistic	End	0.02956	**0.00641**	0.01406
Copula of model (4.4.1)	Start	0.0156	**0.00621**	0.00683
Copula of model (4.4.1)	Middle	0.00208	**0.00088**	0.01775
Copula of model (4.4.1)	End	0.04873	0.01486	**0.00392**

with smaller ISE values corresponding to an estimator with lower bias. Table A.3.2 gives the ISE values for each median estimator and copula example at each of the three time points. As can be observed, the ISE values are very similar for both $\bar{\lambda}^*_{QR}$ and $\bar{\lambda}^*_{BP}$, indicating similar bias for these estimators. Moreover, $\bar{\lambda}^*_{QR}$ often results in the lowest ISE values compared to the other two estimators.

We observe that the MISE and ISE values for the negatively correlated Gaussian copula at the start of the interval ($t = 1$) are significantly larger than other values. This is due to the fact that for strongly negatively dependent data structures, the value of the true ADF tends towards infinity as ρ approaches -1 for any ray $w \in (0, 1)$, which is difficult to capture in practice. However, we note that while significant bias appears to exist in ADF estimates, we are still able to obtain accurate return curve estimates.

Table A.3.2: ISE values for median estimators at start, middle and end of simulated time frame. Smallest ISE values in each row are highlighted in bold.

Copula	Times	ISE - $\bar{\lambda}^*_{QR}$	ISE - $\bar{\lambda}^*_{BP}$	ISE - $\bar{\lambda}^*_{GAM}$
Gaussian (Positive Correlation)	Start	**0.00084**	0.00138	0.00167
Gaussian (Positive Correlation)	Middle	0.00115	0.00110	**0.00103**
Gaussian (Positive Correlation)	End	**0.00010**	0.00035	0.00016
Gaussian (Negative Correlation)	Start	33.12202	36.09158	**31.43350**
Gaussian (Negative Correlation)	Middle	0.08161	0.08196	**0.05865**
Gaussian (Negative Correlation)	End	0.00040	0.00116	**0.00025**
Inverted Logistic	Start	0.00086	0.00041	**0.00007**
Inverted Logistic	Middle	**0.00001**	0.00009	0.00654
Inverted Logistic	End	**0.00144**	0.00145	0.01910
Inverted Hüsler-Reiss	Start	0.00582	**0.00142**	0.03454
Inverted Hüsler-Reiss	Middle	**0.00001**	0.00006	0.00037
Inverted Hüsler-Reiss	End	0.00026	0.00097	**0.00019**
Inverted Asymmetric Logistic	Start	0.00123	**0.00033**	0.01837
Inverted Asymmetric Logistic	Middle	**0.00002**	**0.00002**	0.01942
Inverted Asymmetric Logistic	End	0.00073	**0.00038**	0.00965
Copula of model (4.4.1)	Start	0.00084	**0.00069**	0.00405
Copula of model (4.4.1)	Middle	**0.00004**	0.00013	0.01519
Copula of model (4.4.1)	End	0.00888	0.00596	**0.00381**

A.3.2.3 Non-stationary ADF estimates over time

Figures A.3.2 - A.3.6 illustrate median estimates, alongside 0.025 and 0.975 quantile estimates, of the ADF at fixed rays over time for each copula example, excluding the inverted logistic copula which has already been considered in the main text. In each case, the rays $w = 0.1$, $w = 0.3$ and $w = 0.5$ have been considered and the black, green and blue lines correspond to the median estimates for $\bar{\lambda}^*_{QR}$, $\bar{\lambda}^*_{BP}$ and $\bar{\lambda}^*_{GAM}$, respectively, with the coloured regions representing the area between the pointwise 0.025 and 0.975 quantiles. For the inverted asymmetric logistic copula, two additional rays ($w = 0.7$ and $w = 0.9$) have been considered to account for the asymmetry within this example.

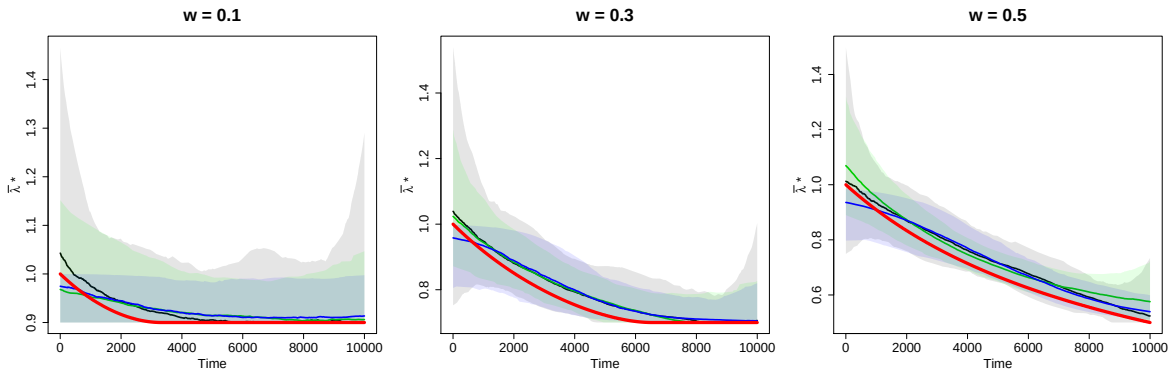

Figure A.3.2: Non-stationary ADF estimates over time for the Gaussian copula with positive correlation.

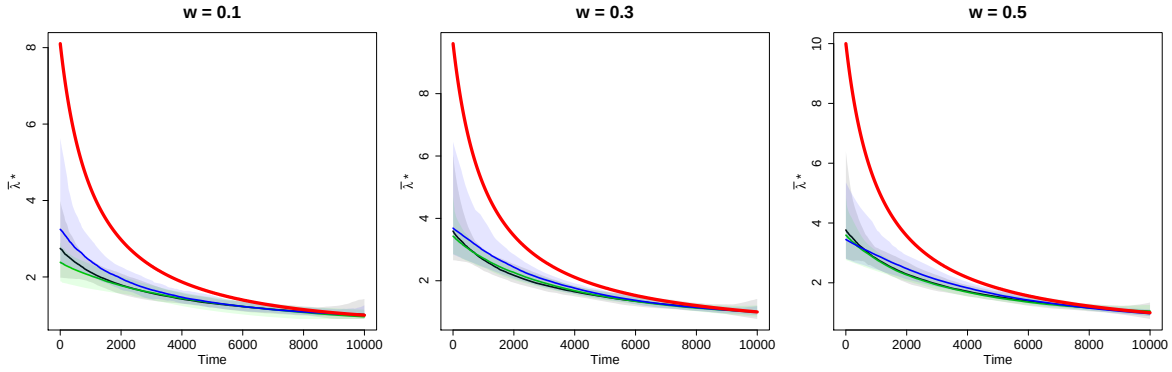

Figure A.3.3: Non-stationary ADF estimates over time for the Gaussian copula with negative correlation.

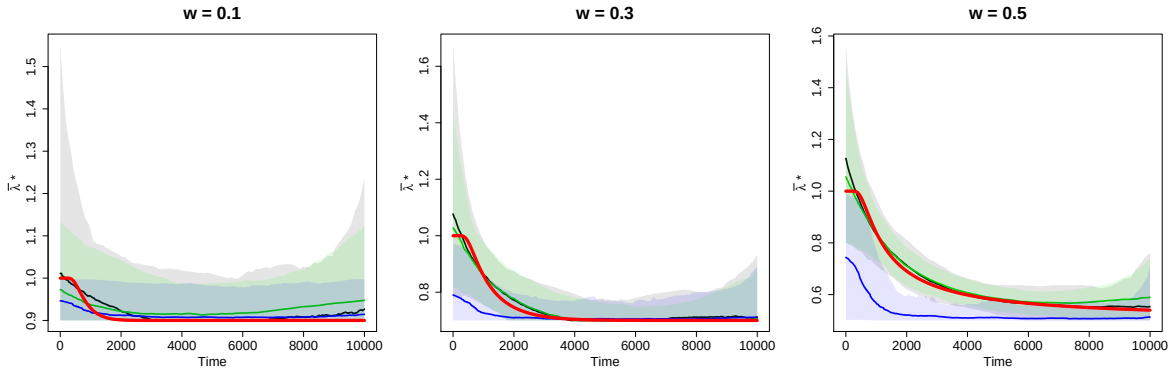

Figure A.3.4: Non-stationary ADF estimates over time for the inverted Hüsler-Reiss copula.

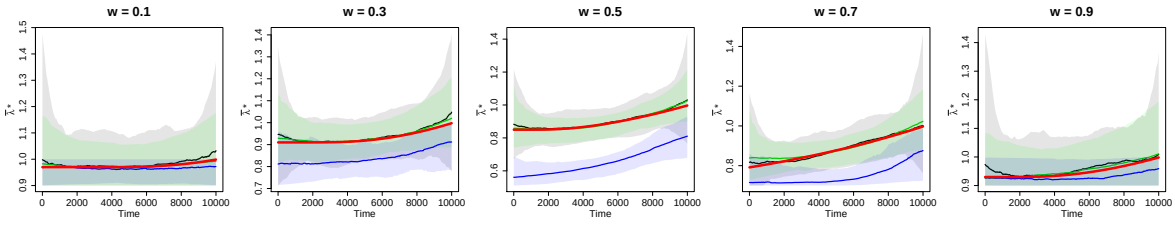

Figure A.3.5: Non-stationary ADF estimates over time for the inverted asymmetric logistic copula.

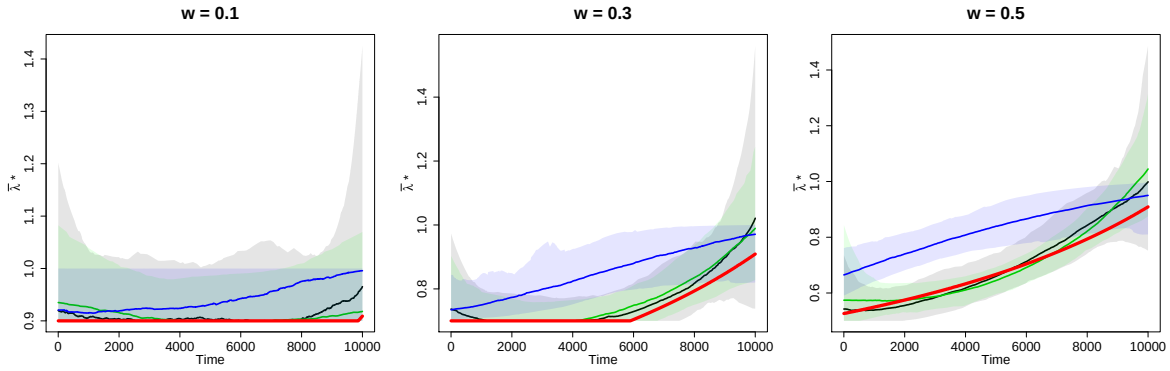

Figure A.3.6: Non-stationary ADF estimates over time for the copula of model (4.4.1)

A.3.2.4 Non-stationary ADF estimates at fixed time points

Figures A.3.7 - A.3.12 illustrate median estimates, alongside 0.025 and 0.975 quantile estimates, of the ADF at three fixed time points ($t = 1$, $t = n/2$ and $t = n$) over all rays $w \in [0, 1]$ for each copula example. In each case, the red lines represent the true ADF functions, while the black, green and blue lines correspond to the median estimates for $\bar{\lambda}^*_{QR}$, $\bar{\lambda}^*_{BP}$ and $\bar{\lambda}^*_{GAM}$, respectively, with the coloured regions representing the area between the pointwise 0.025 and 0.975 quantiles.

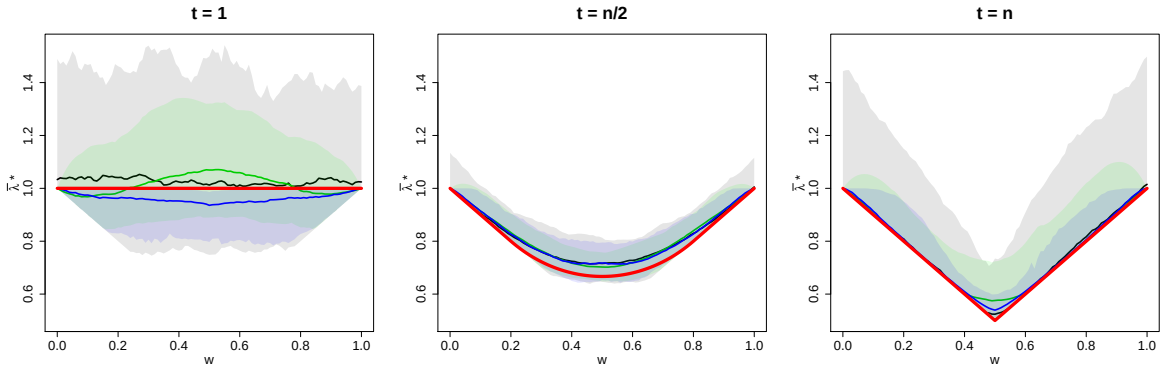

Figure A.3.7: Non-stationary ADF estimates at three fixed time points for the Gaussian copula with positive correlation.

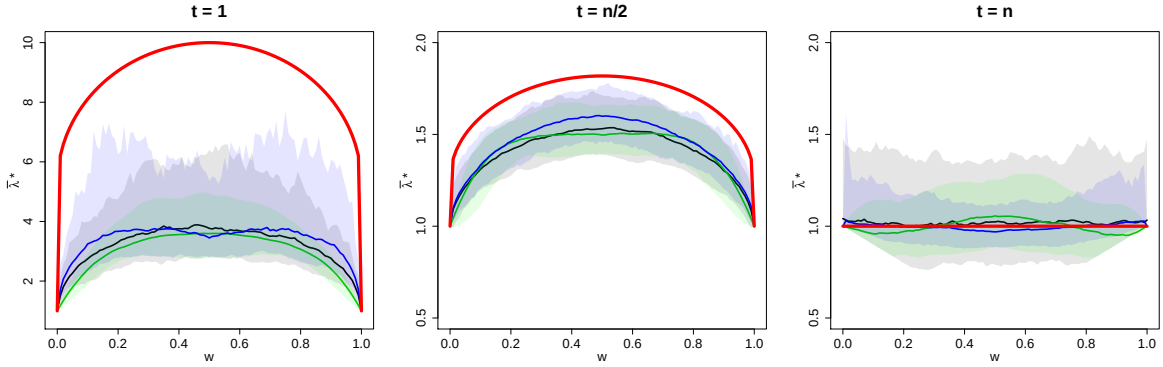

Figure A.3.8: Non-stationary ADF estimates at three fixed time points for the Gaussian copula with negative correlation.

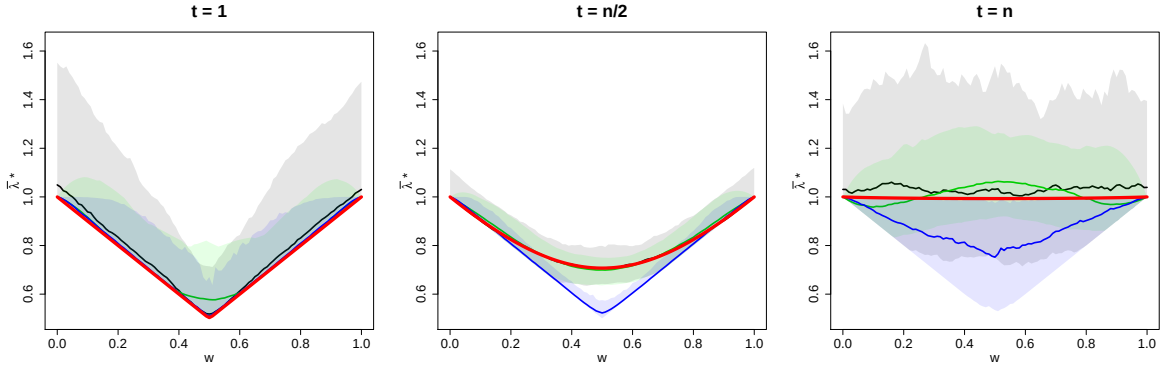

Figure A.3.9: Non-stationary ADF estimates at three fixed time points for the inverted logistic copula.

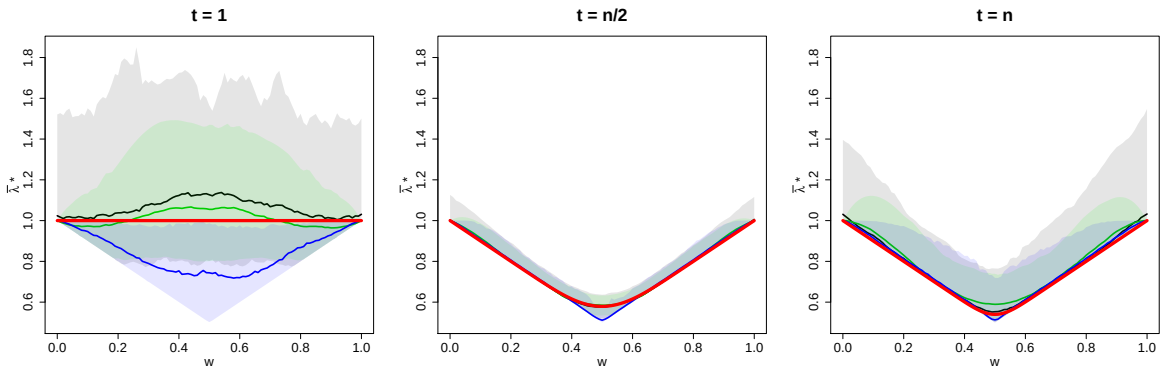

Figure A.3.10: Non-stationary ADF estimates at three fixed time points for the inverted Hüsler-Reiss copula.

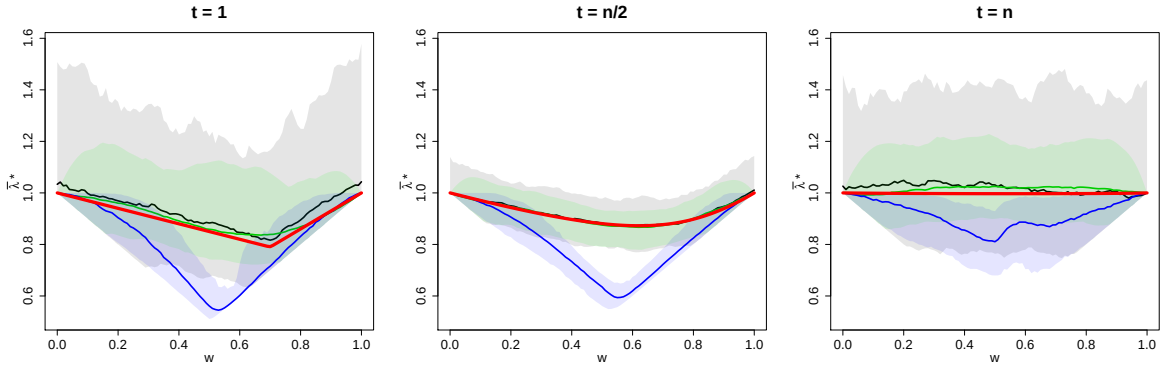

Figure A.3.11: Non-stationary ADF estimates at three fixed time points for the inverted asymmetric logistic copula.

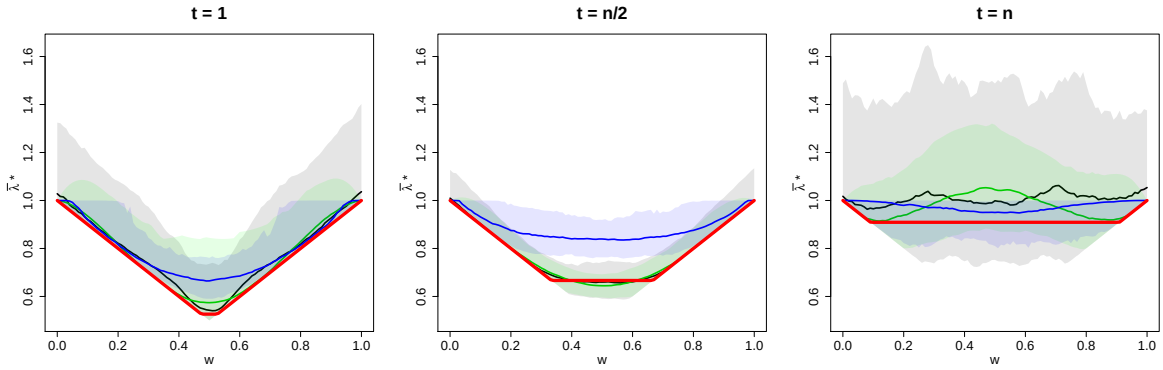

Figure A.3.12: Non-stationary ADF estimates at three fixed time points for the copula of model (4.4.1).

A.3.2.5 Non-stationary return curve estimates

Figures A.3.13 - A.3.17 illustrate median estimates of $\mathrm{RC}_{\mathbf{z}_t}(p)$ at $p = 1/n$ obtained under $\bar{\lambda}^*_{QR}$ and $\bar{\lambda}^*_{BP}$ for each of the copula examples, excluding the Gaussian copula with negative correlation which has already been considered in the main text. In each case, the true return curves are given in the left panels, while the curves corresponding to $\bar{\lambda}^*_{QR}$ and $\bar{\lambda}^*_{BP}$ are given in the centre and right panels, respectively. Time is illustrated using a smooth curve colour change, with red and blue curves corresponding to the start and end of the time frame, respectively. We observe that the quality of the median curve estimates appear to be worse near both the start and end of the time frame. This is likely due to the larger degree of variability that arises at such time points, as exhibited in Figures A.3.2 - A.3.6.

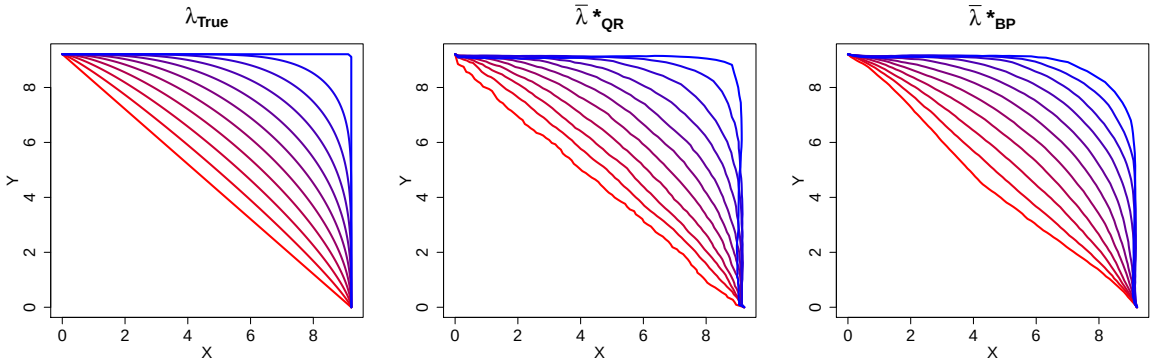

Figure A.3.13: Non-stationary return curve estimates over time for the Gaussian copula with positive correlation.

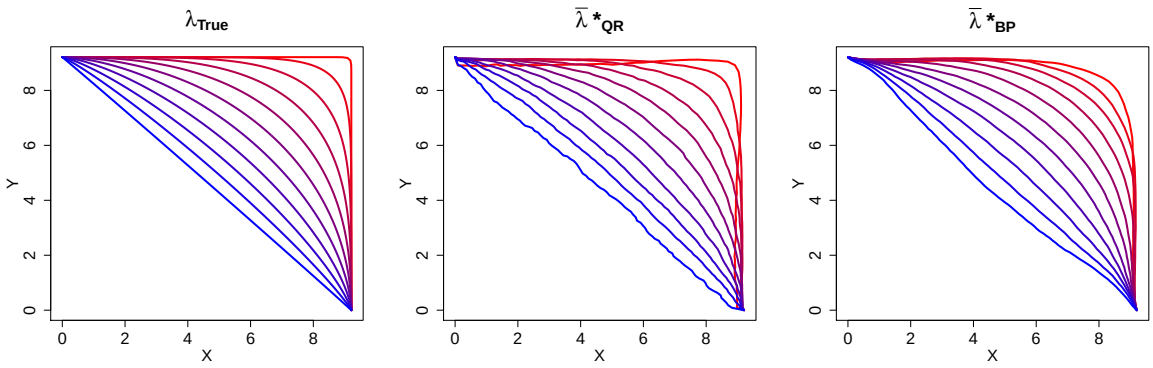

Figure A.3.14: Non-stationary return curve estimates over time for the inverted logistic copula.

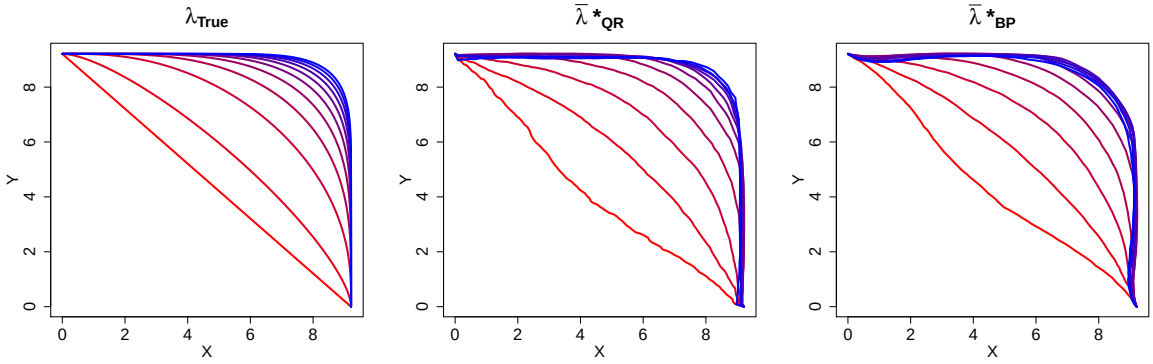

Figure A.3.15: Non-stationary return curve estimates over time for the inverted Hüsler-Reiss copula.

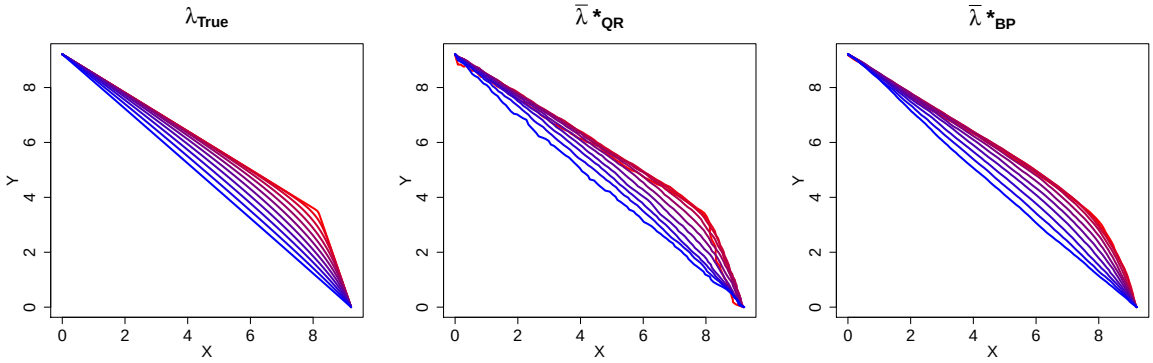

Figure A.3.16: Non-stationary return curve estimates over time for the inverted asymmetric logistic copula.

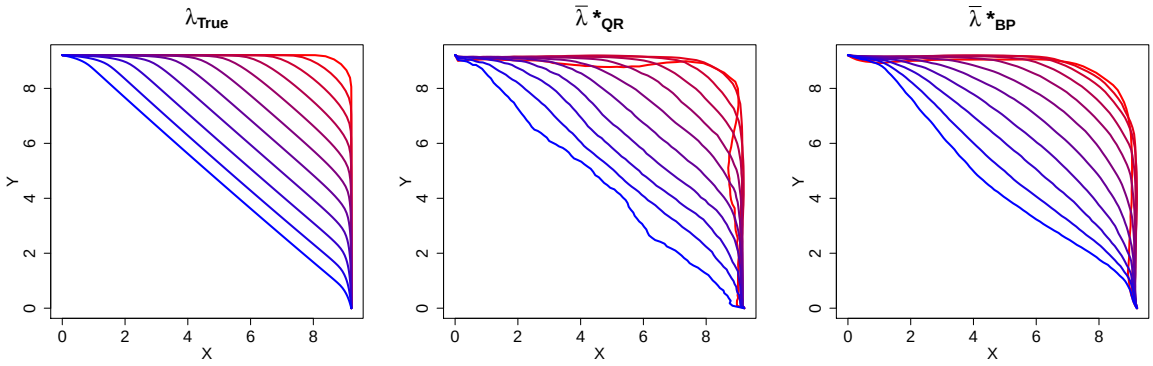

Figure A.3.17: Non-stationary return curve estimates over time for the copula of model (4.4.1).

A.3.3 Additional case study figures

A.3.3.1 Fitted pre-processing trend functions

Figure A.3.18 compares empirical estimates of the mean and standard deviations for both sets of projections against fitted location and scale functions, alongside estimated 95% confidence regions. Empirical estimates are obtained using the data for fixed years over the observation period. The fitted location and scale functions are then averaged over each year and compared to the empirical estimates. One can observe very similar trends for both variables, indicating the pre-processing technique is accurately capturing the marginal non-stationary trends within the body of data. Note that the estimated confidence regions are likely to be an under-representation of the true uncertainty due to the fact we have assumed independence between observations when fitting the GAMs.

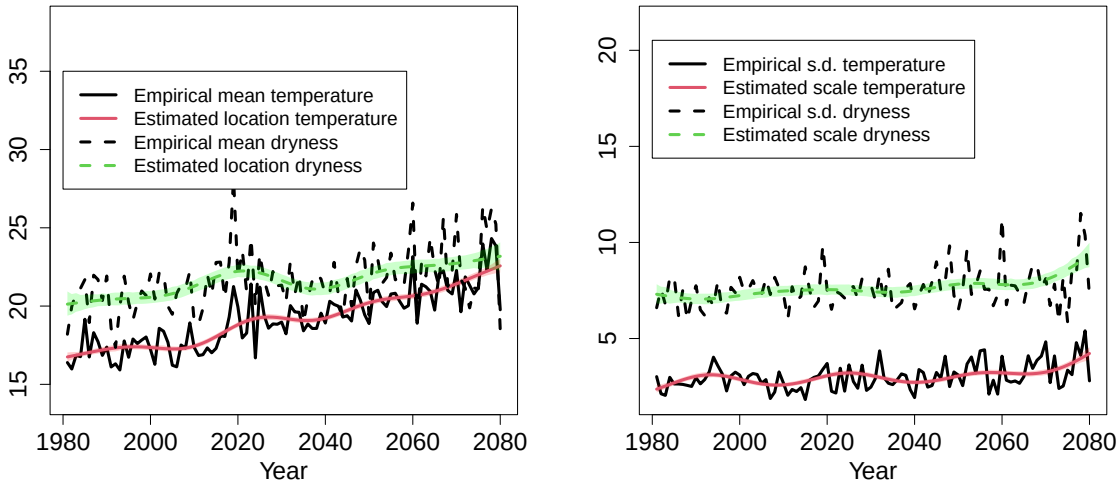

Figure A.3.18: Comparison of estimated location and scale function values (red and green for temperature and dryness, respectively) with 95% confidence intervals (shaded regions) against empirical mean and standard deviation estimates (black). For the fitted functions, the average value for a given year has been taken to ensure correspondence with the empirical values.

A.3.3.2 Estimated rate parameters

Figure A.3.19 illustrates exponential rate parameter estimates for the pre-processed data, alongside 95% pointwise confidence intervals, for ±15 year rolling windows over the observation period. As can be observed, the rate parameter estimates remain approximately constant at one throughout the entire observation period, suggesting a successful transformation to exponential margins.

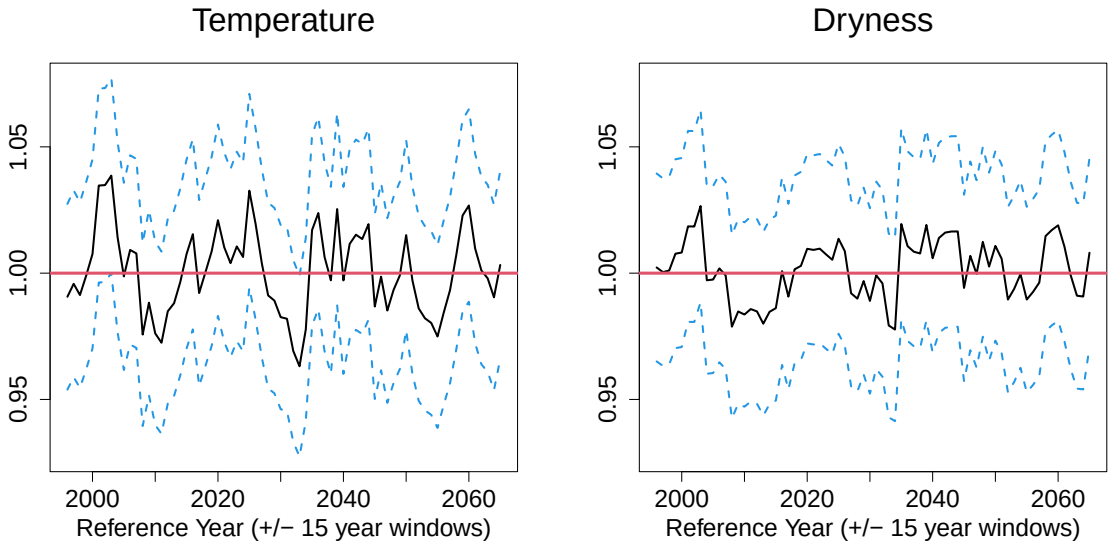

Figure A.3.19: Estimated exponential rate parameters (black) with 95% pointwise confidence intervals (dotted blue) over the time period. The target rate parameter is given in red.

A.4 Supplementary material for Chapter 5

A.4.1 Example ADF estimates

Examples of ADF estimates obtained using each of the estimators discussed in the main article are given in Figure A.4.1 for a bivariate Gaussian copula with $\rho = 0.5$. The different estimates are in good agreement, and one can observe the roughness in estimates obtained via the Hill estimator.

A.4.2 Example boundary set estimates

Figure A.4.2 illustrates estimates of the boundary set G obtained using the technique proposed in Simpson and Tawn (2022) for three copula examples.

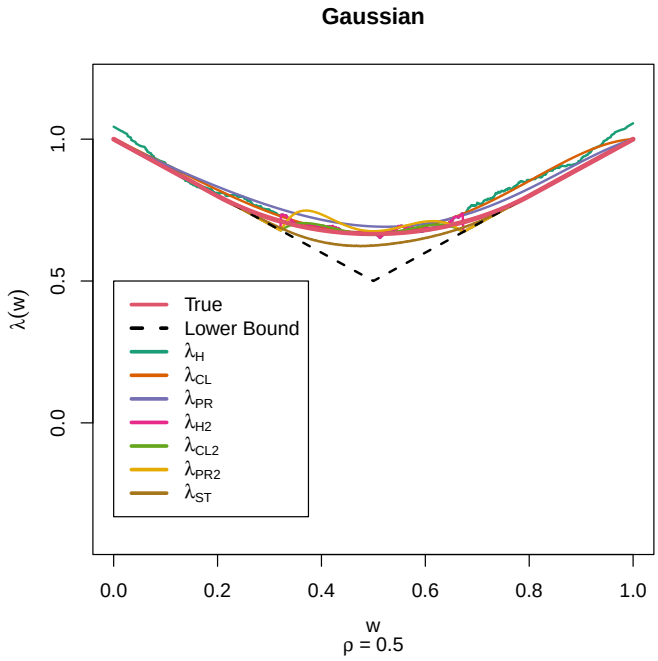

Figure A.4.1: ADF estimates from each of the estimators discussed in the main article. Red represents the true ADF, with the theoretical lower bound given by the dotted black lines.

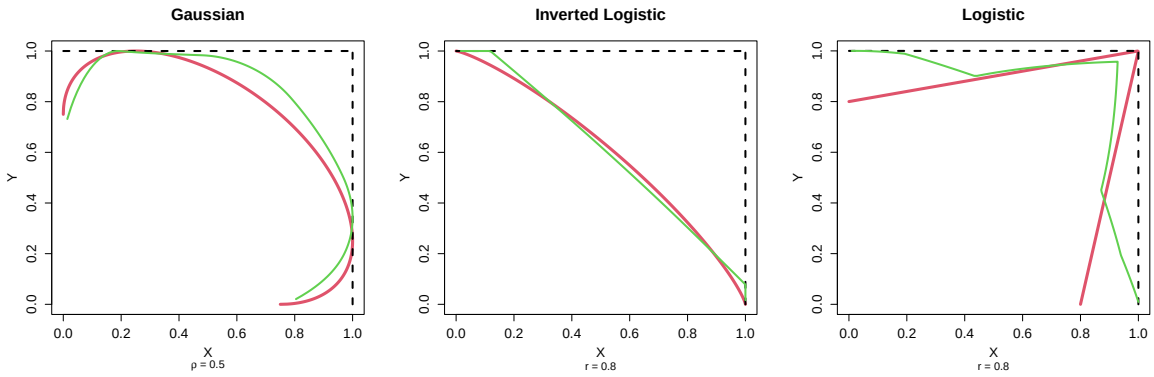

Figure A.4.2: The boundary set G (given in red) for three copula examples, with estimates from Simpson and Tawn (2022) given in green. Left: bivariate Gaussian copula with correlation coefficient $\rho = 0.5$. Centre: inverted logistic copula with dependence parameter $r = 0.8$. Right: logistic copula with dependence parameter $r = 0.8$. In each plot, the coordinate limits of G are denoted by the black dotted lines.

A.4.3 Tuning parameter selection

Figures A.4.3 and A.4.4 illustrate plots of scaled RMISE estimates against the polynomial degree k for the estimators $\hat{\lambda}_{CL}$ and $\hat{\lambda}_{PR}$, respectively. RMISE was estimated

using Monte-Carlo techniques: first, for each $j = 1, 2, \ldots, N$, where $N = 200$ denotes the number of samples, the ISE was estimated via the trapezium rule, i.e.,

$$\widehat{\text{ISE}}\left(\hat{\lambda}_j\right) = \frac{0.001}{2}\left((\hat{\lambda}_j(0) - \lambda(0))^2 + \sum_{w \in \mathcal{W}\setminus\{0,1\}} 2(\hat{\lambda}_j(w) - \lambda(w))^2 + (\hat{\lambda}_j(1) - \lambda(1))^2\right),$$

where $\hat{\lambda}_j$ denotes the ADF estimate for sample j and $\mathcal{W}$ denotes the set of rays spanning the interval $[0, 1]$, as defined in Section 3.4 the main article. An estimate of the RMISE is then given by

$$\widehat{\text{RMISE}} = \sqrt{\frac{1}{N}\sum_{j=1}^{N} \widehat{\text{ISE}}\left(\hat{\lambda}_j\right)}.$$

For the polynomial degree, we considered $k \in \{2, 3, \ldots, 11\}$; higher values of k were not considered due to computational complexity. The left and right panels of Figures A.4.3 and A.4.4 correspond to Gaussian copulas exhibiting strong ($\rho = 0.9$) and weak ($\rho = 0.1$) positive dependence, respectively.

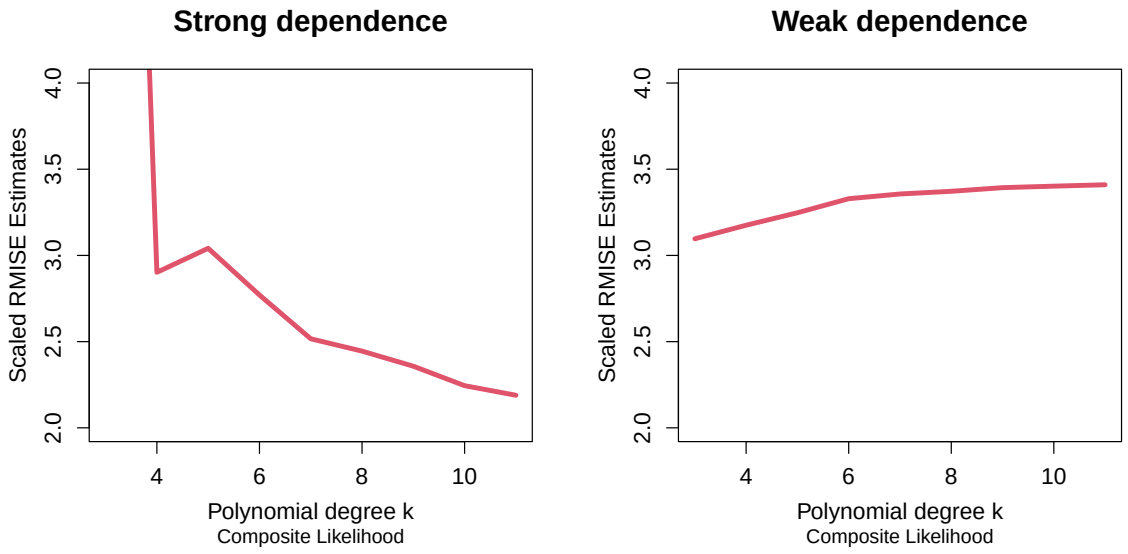

Figure A.4.3: RMISE estimates (multiplied by 100) over k obtained for $\hat{\lambda}_{CL}$ using $N = 200$ from Gaussian copulas with strong (left, $\rho = 0.9$) and weak (right, $\rho = 0.1$) positive dependence.

One can observe that, for the strongly dependent example, the RMISE estimates

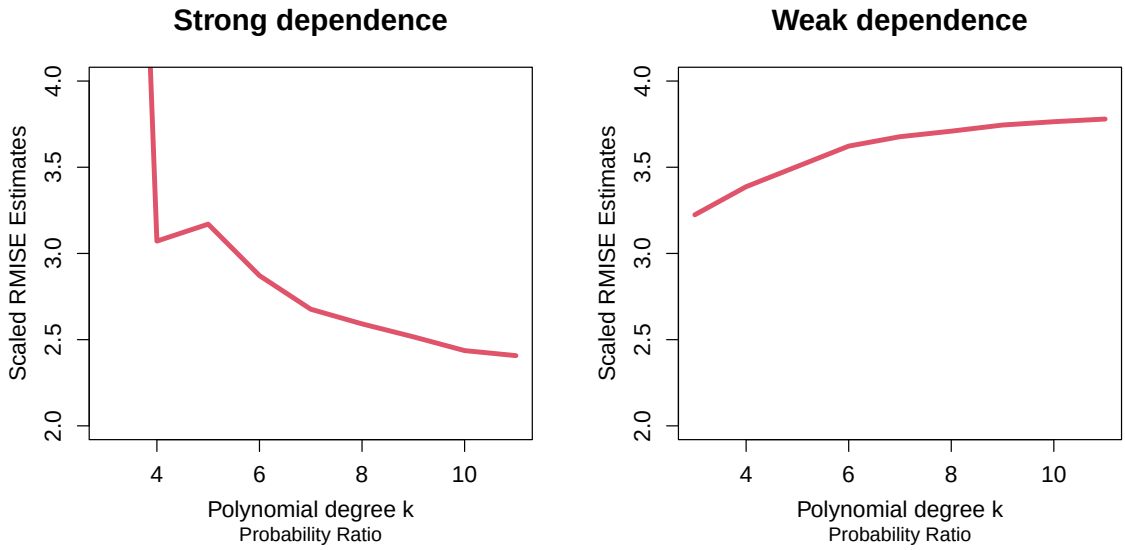

Figure A.4.4: RMISE estimates (multiplied by 100) over k obtained for $\hat{\lambda}_{PR}$ using $N = 200$ from Gaussian copulas with strong (left, $\rho = 0.9$) and weak (right, $\rho = 0.1$) positive dependence.

tend to decrease as the value of k increases. This is in agreement with findings for the PDF (e.g., Marcon et al., 2017b; Vettori et al., 2018); strongly dependent copulas require a higher degree of flexibility to capture the triangle-like shapes of the resulting dependence functions.

On the other hand, for the weakly dependent Gaussian copula, the value of k made little difference to the resulting RMISE estimates. There is even a slight increase in RMISE estimates for higher values of k; this suggests that having a higher polynomial degrees for such data structures may lead to over-fitting.

In practice, we set $k = 7$; this value is sufficient for obtaining low RMISE estimates under both copula examples. Furthermore, for the strongly dependent Gaussian copula, the reduction in RMISE values above this value of k are very marginal. This polynomial degree therefore appears to offer sufficient flexibility without high computational burden and/or parameter variability.

A.4.4 Additional simulation study results

The ISB and IV estimates for each estimator are given in Tables A.4.1 and A.4.2. One can observe that, while the $\hat{\lambda}_{ST}$ appears to perform best in terms of ISB, the estimators derived using the composite likelihood function ($\hat{\lambda}_{CL}$ and $\hat{\lambda}_{CL2}$) exhibit the least IV for eight out of the nine copula examples. For the most part, one can observe similar ISB and IV values across the different estimators.

Table A.4.1: ISB values (multiplied by 1,000) for each estimator and copula combination. Smallest ISB values in each row are highlighted in bold, with values reported to 3 significant figures.

Copula	$\hat{\lambda}_H$	$\hat{\lambda}_{CL}$	$\hat{\lambda}_{PR}$	$\hat{\lambda}_{H2}$	$\hat{\lambda}_{CL2}$	$\hat{\lambda}_{PR2}$	$\hat{\lambda}_{ST}$
Copula 1	**371**	374	436	374	380	442	402
Copula 2	0.361	0.378	0.468	0.34	0.338	0.416	**0.00815**
Copula 3	0.779	0.883	1.07	0.762	0.771	0.946	**0.0181**
Copula 4	1.74	1.95	4.25	1.5	1.49	3.22	**0.299**
Copula 5	19.2	19.3	28.4	19.1	19.1	28.1	**13.6**
Copula 6	0.0012	0.0526	0.0683	**0.000954**	0.00324	0.00531	0.0045
Copula 7	**0.000917**	0.00709	0.0131	0.00161	0.00276	0.00467	0.55
Copula 8	0.047	0.622	0.664	**0.00728**	0.00903	0.0186	0.0755
Copula 9	13.8	13.8	21.3	13.7	13.7	21.1	**10.9**

A.4.5 Additional case study figures

This section contains additional figures for the case study detailed in Section 4 of the main article. Figure A.4.5 illustrates daily river flow time series for each of the six gauges in the north of England, UK. These series suggest a stationarity assumption is reasonable for the extremes of each data set.

Figure A.4.6 illustrates the QQ plots from the fitted GPDs at each of the six gauges. One can observe that, in each case, the majority of points lie close to the $y = x$ line,

Table A.4.2: IV values (multiplied by 1,000) for each estimator and copula combination. Smallest IV values in each row are highlighted in bold, with values reported to 3 significant figures.

Copula	$\hat{\lambda}_H$	$\hat{\lambda}_{CL}$	$\hat{\lambda}_{PR}$	$\hat{\lambda}_{H2}$	$\hat{\lambda}_{CL2}$	$\hat{\lambda}_{PR2}$	$\hat{\lambda}_{ST}$
Copula 1	2.6	**2.02**	2.11	3.51	3.29	3.18	4.7
Copula 2	0.9	**0.732**	0.859	0.889	0.769	0.904	0.863
Copula 3	0.652	0.328	0.401	0.311	0.267	0.33	**0.1**
Copula 4	0.657	0.352	0.536	0.324	**0.315**	0.588	0.468
Copula 5	0.808	**0.651**	0.841	0.793	0.707	0.933	1.07
Copula 6	0.628	0.337	0.395	0.4	**0.299**	0.356	0.445
Copula 7	0.855	**0.692**	0.817	0.821	0.702	0.831	1.02
Copula 8	0.571	0.119	0.206	0.0363	**0.0269**	0.0437	0.273
Copula 9	0.799	**0.696**	0.895	0.839	0.761	0.98	1.28

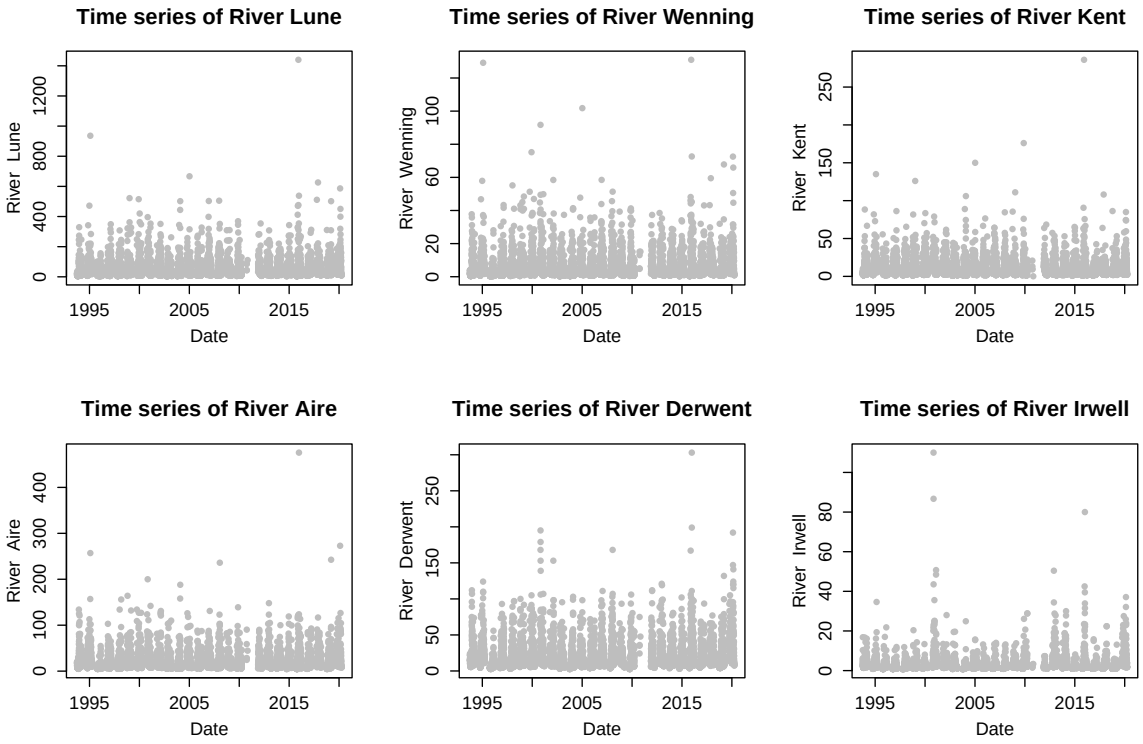

Figure A.4.5: Daily river flow time series for the six gauges in the north of England, UK.

indicating the fitted models capture the upper tails well.

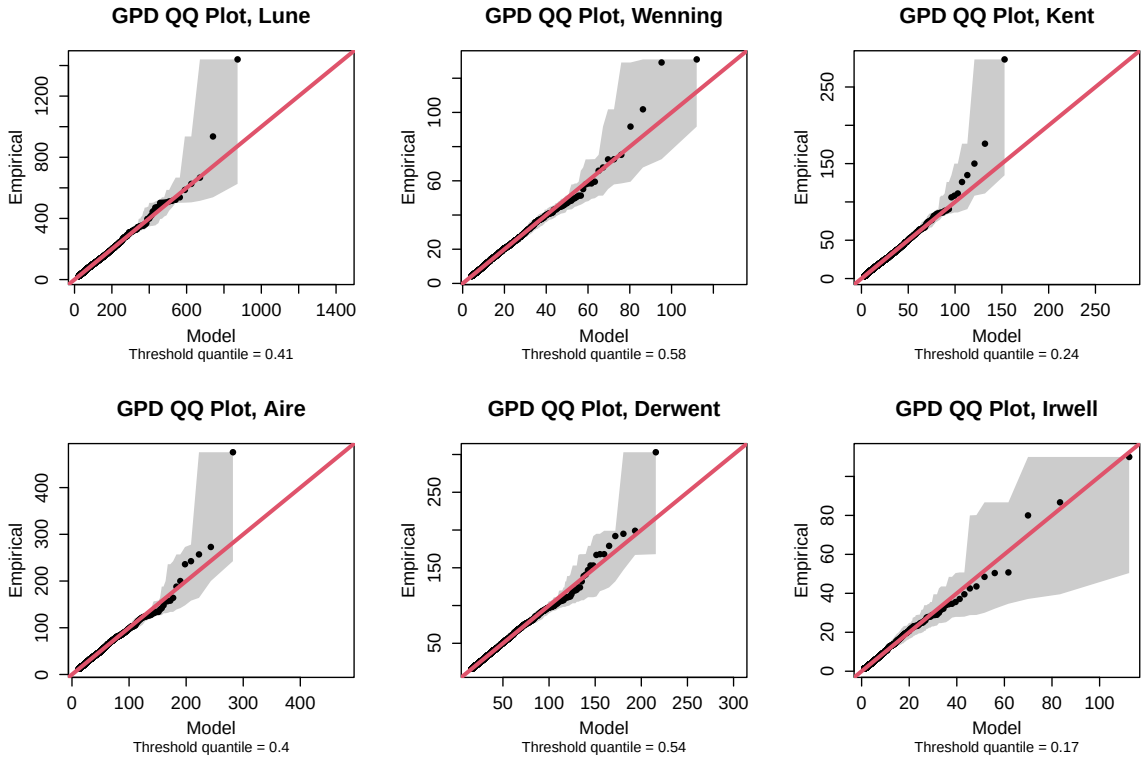

Figure A.4.6: QQ plots for each of the fitted GPDs at each of the six gauges. Estimates are given in black, with 95% pointwise confidence intervals represented by the grey shaded regions. The red line corresponds to the $y = x$ line. The corresponding threshold quantile levels are given in the subtitle of each plot.

Figures A.4.7 and A.4.8 illustrate the ADF QQ plots for the first pair of gauges using the estimates obtained via $\hat{\lambda}_{ST}$ and $\hat{\lambda}_{H}$, respectively. The estimated and observed quantiles appear in good agreement at each of the considered rays.

Figures A.4.9 and A.4.10 illustrate the return curve diagnostic of Murphy-Barltrop et al. (2023) for the estimators $\hat{\lambda}_{ST}$ and $\hat{\lambda}_{CL2}$, respectively. For this diagnostic, a subset of points are selected on a return curve estimate; these points correspond to a set of $m = 150$ equally spaced angles θ in the interval $[0, \pi/2]$, i.e., given $(x, y) \in \text{RC}(p)$, we have $\theta = \tan^{-1}(y/x)$. Empirical estimates of the joint survival function are computed for each point and bootstrapping is used to evaluate uncertainty. Finally, the median empirical estimates, alongside 95% pointwise confidence intervals, are plotted against

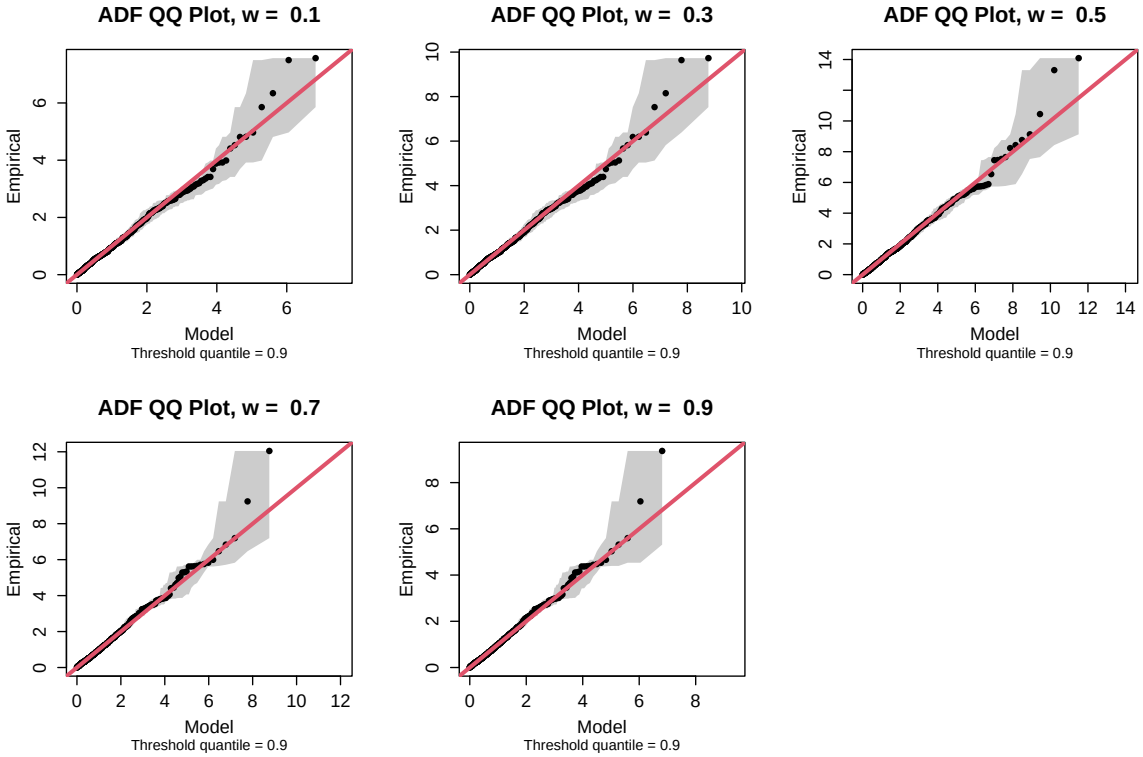

Figure A.4.7: Individual ADF QQ plots for $w \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ for first pair of gauges, using the ADF estimate obtained via $\hat{\lambda}_{ST}$. Estimates are given in black, with 95% pointwise confidence intervals represented by the grey shaded regions. The red line corresponds to the $y = x$ line.

the angle index and compared to the true probability; see Murphy-Barltrop et al. (2023) for further details.

Both estimators appear to give a similar level of accuracy, though for the fifth gauge site pairing, both $\hat{\lambda}_{ST}$ and $\hat{\lambda}_{CL2}$ fail to capture the true probability at all angles.

Finally, Figures A.4.11 and A.4.12 illustrate estimated return curve uncertainty intervals obtained using the $\hat{\lambda}_{ST}$ and $\hat{\lambda}_{CL2}$ estimators, respectively, for the first and fifth gauge site pairings. In both figures, one can observe the contrast in shapes of the uncertainty regions between the two site pairings.

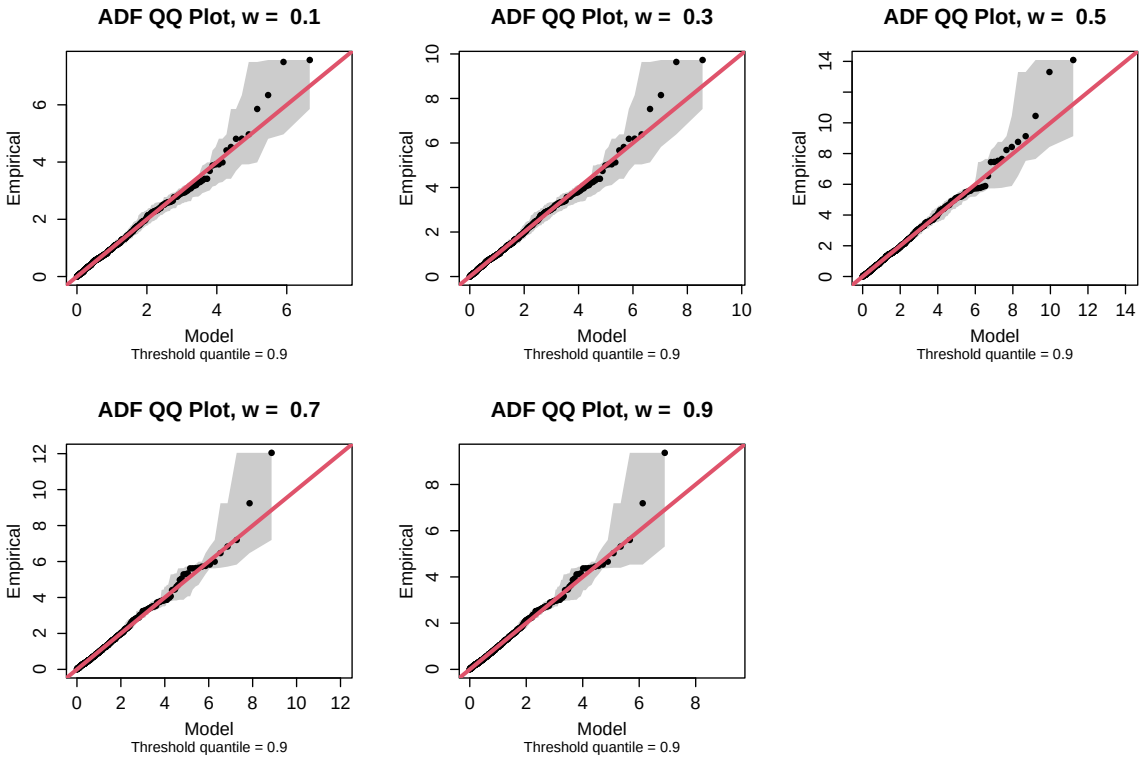

Figure A.4.8: Individual ADF QQ plots for $w \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ for first pair of gauges, using the ADF estimate obtained via $\hat{\lambda}_H$. Estimates are given in black, with 95% pointwise confidence intervals represented by the grey shaded regions. The red line corresponds to the $y = x$ line.

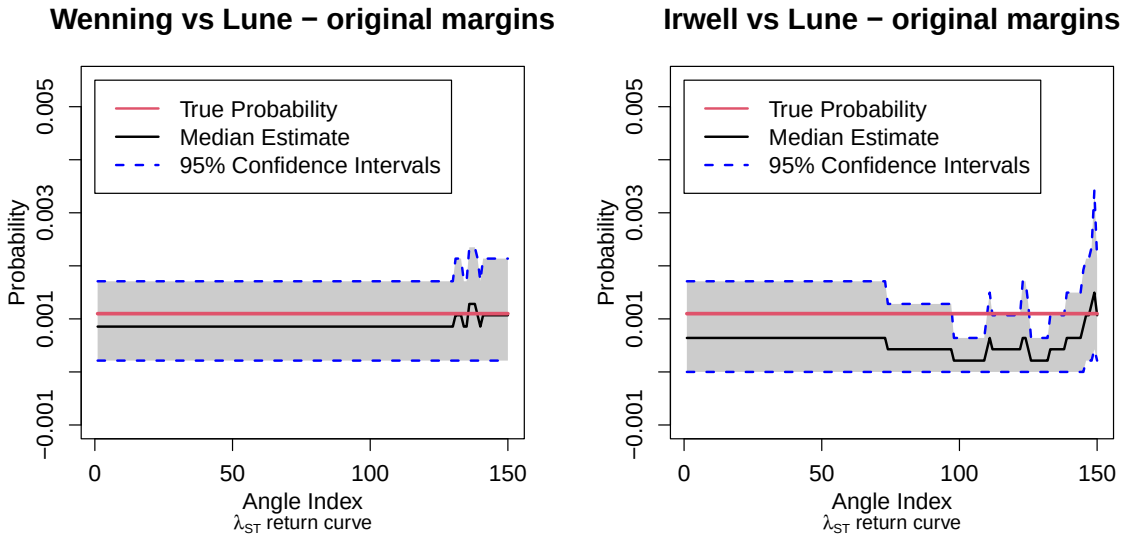

Figure A.4.9: Diagnostic plots of the return curve estimates from the $\hat{\lambda}_{ST}$ estimator for the first and fifth gauge site pairings. The black and red lines indicate the empirical median and true survival probabilities, respectively, with 95% bootstrapped confidence intervals denoted by the shaded regions.

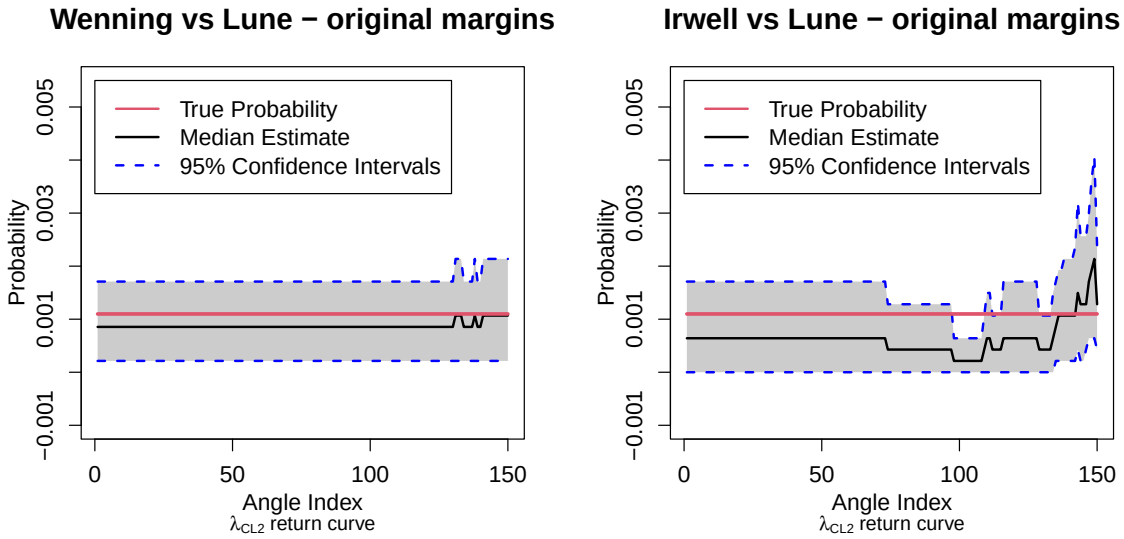

Figure A.4.10: Diagnostic plots of the return curve estimates from the $\hat{\lambda}_{CL2}$ estimator for the first and fifth examples. The black and red lines indicate the empirical median and true survival probabilities, respectively, with 95% bootstrapped confidence intervals denoted by the shaded regions

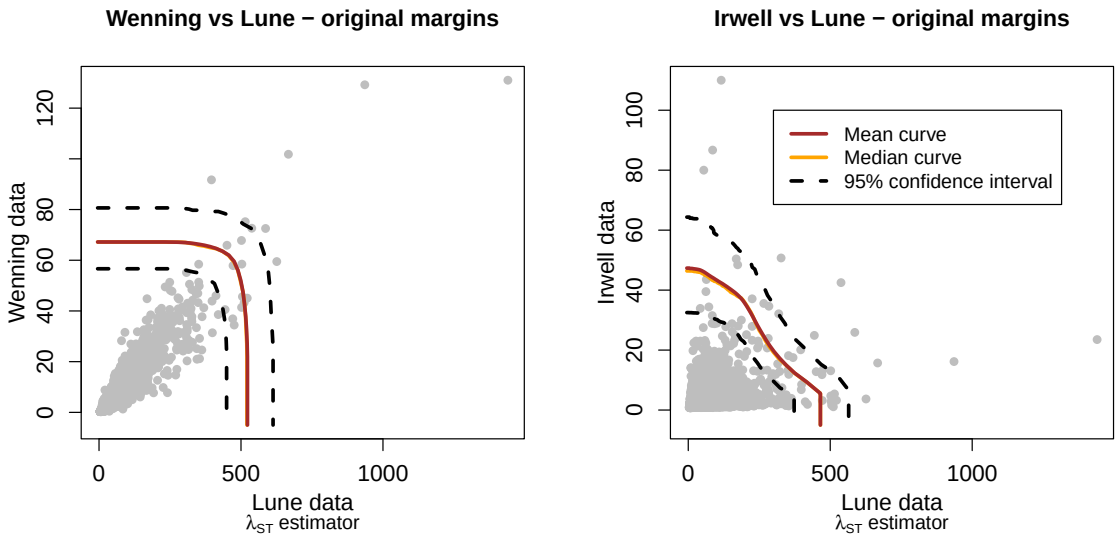

Figure A.4.11: Median and mean return curve estimates in orange and brown, respectively, obtained using the $\hat{\lambda}_{ST}$ estimator for the first and fifth examples. The black dotted lines indicate 95% confidence intervals.

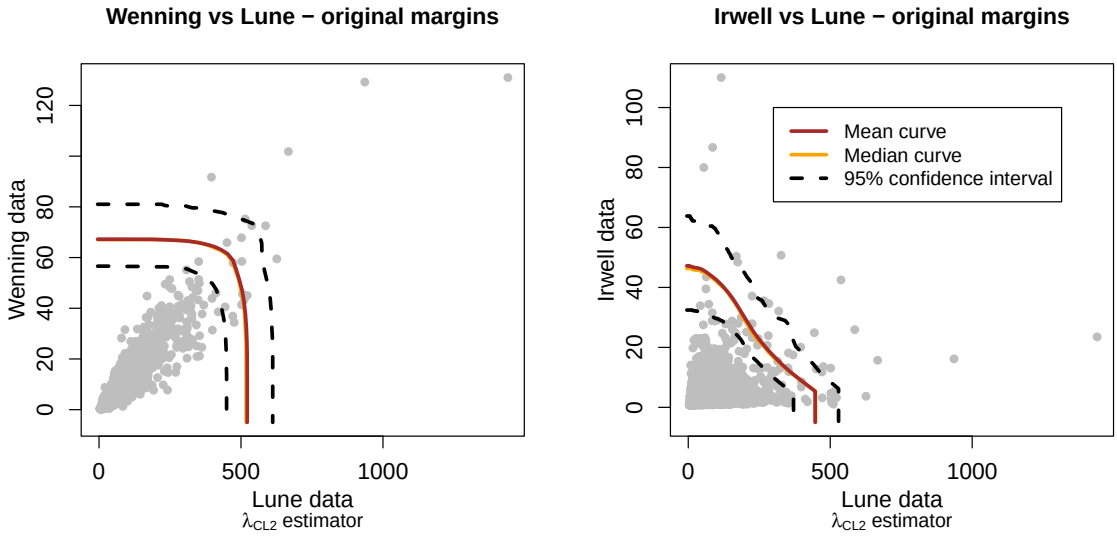

Figure A.4.12: Median and mean return curve estimates in orange and brown, respectively, obtained using the $\hat{\lambda}_{CL2}$ estimator for the first and fifth examples. The black dotted lines indicate 95% confidence intervals.

Bibliography

Aghakouchak, A., Ciach, G., and Habib, E. (2010). Estimation of tail dependence coefficient in rainfall accumulation fields. *Advances in Water Resources*, 33:1142–1149.

Baker, H. S., Millar, R. J., Karoly, D. J., Beyerle, U., Guillod, B. P., Mitchell, D., Shiogama, H., Sparrow, S., Woollings, T., and Allen, M. R. (2018). Higher CO2 concentrations increase extreme event risk in a 1.5 °c world. *Nature Climate Change*, 8:604–608.

Balkema, A. A. and de Haan, L. (1974). Residual Life Time at Great Age. *The Annals of Probability*, 2:792–804.

Baran, S., Szokol, P., and Szabó, M. (2021). Truncated generalized extreme value distribution-based ensemble model output statistics model for calibration of wind speed ensemble forecasts. *Environmetrics*, 32:1–24.

Barnett, V. (1976). The Ordering of Multivariate Data. *Journal of the Royal Statistical Society. Series A (General)*, 139:318–344.

Beirlant, J., Goegebeur, Y., Teugels, J., and Segers, J. (2004). *Statistics of Extremes*. Wiley.

Bodik, J., Mhalla, L., and Chavez-Demoulin, V. (2022). Detecting causal covariates for extreme dependence structures. *arXiv*, 2212.09831.

Brunner, M. I., Seibert, J., and Favre, A. C. (2016). Bivariate return periods and their importance for flood peak and volume estimation. *Wiley Interdisciplinary Reviews: Water*, 3:819–833.

Cade, B. S. and Noon, B. R. (2003). A Gentle Introduction to Quantile Regression for Ecologists. *Frontiers in Ecology and the Environment*, 1:412–420.

Caires, S. and Ferreira, J. A. (2005). On the Non-parametric Prediction of Conditionally Stationary Sequences. *Statistical Inference for Stochastic Processes*, 8:151–184.

Casson, E. and Coles, S. (1999). Spatial Regression Models for Extremes. *Extremes*, 1:449–468.

Castro-Camilo, D., de Carvalho, M., and Wadsworth, J. (2018). Time-varying extreme value dependence with application to leading European stock markets. *Annals of Applied Statistics*, 12:283–309.

Chatterjee, S. and Chatterjee, S. (1990). A note on finding extreme points in multivariate space. *Computational Statistics and Data Analysis*, 10:87–92.

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

Chavez-Demoulin, V. and Davison, A. C. (2012). Modelling time series extremes. *REVSTAT-Statistical Journal*, 10:109–133.

Chebana, F. and Ouarda, T. B. (2011). Multivariate quantiles in hydrological frequency analysis. *Environmetrics*, 22:63–78.

Chernozhukov, V. and Fernández-Val, I. (2011). Inference for extremal conditional quantile models, with an application to market and birthweight risks. *Review of Economic Studies*, 78:559–589.

Chernozhukov, V., Fernández-Val, I., and Kaji, T. (2017). *Handbook of Quantile Regression Chapter 18 - Extremal Quantile Regression*. Chapman and Hall/CRC.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer London.

Coles, S., Heffernan, J., and Tawn, J. (1999). Dependence measures for multivariate extremes. *Extremes*, 2:339–365.

Coles, S. and Pauli, F. (2002). Models and inference for uncertainty in extremal dependence. *Biometrika*, 89:183–196.

Coles, S. G. and Tawn, J. A. (1991). Modelling Extreme Multivariate Events. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 53:377–392.

Coles, S. G. and Tawn, J. A. (1994). Statistical Methods for Multivariate Extremes: An Application to Structural Design. *Applied Statistics*, 43:1–48.

Cooley, D., Nychka, D., and Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, 102:824–840.

Cooley, D., Thibaud, E., Castillo, F., and Wehner, M. F. (2019). A nonparametric method for producing isolines of bivariate exceedance probabilities. *Extremes*, 22:373–390.

Cormier, E., Genest, C., and Nešlehová, J. G. (2014). Using B-splines for nonparametric inference on bivariate extreme-value copulas. *Extremes*, 17:633–659.

Davison, A. C. and Smith, R. L. (1990). Models for Exceedances Over High Thresholds. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 52:393–425.

de Carvalho, M. and Davison, A. C. (2014). Spectral density ratio models for multivariate extremes. *Journal of the American Statistical Association*, 109:764–776.

de Haan, L. and de Ronde, J. (1998). Sea and Wind: Multivariate Extremes at Work. *Extremes*, 1:7–45.

de Haan, L. and Resnick, S. I. (1977). Limit theory for multivariate sample extremes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 40:317–337.

Draisma, G., Drees, H., Ferreira, A., and Haan, L. D. (2004). Bivariate tail estimation: dependence in asymptotic independence. *Bernoulli*, 10:251–280.

Eastoe, E. F. (2019). Nonstationarity in peaks-over-threshold river flows: A regional random effects model. *Environmetrics*, 30:1–18.

Eastoe, E. F., Heffernan, J. E., and Tawn, J. A. (2014). Nonparametric estimation of the spectral measure, and associated dependence measures, for multivariate extreme values using a limiting conditional representation. *Extremes*, 17:25–43.

Eastoe, E. F. and Tawn, J. A. (2009). Modelling non-stationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 58:25–45.

Eastoe, E. F. and Tawn, J. A. (2010). Statistical models for overdispersion in the frequency of peaks over threshold data for a flow series. *Water Resources Research*, 46:1–12.

Eckert-Gallup, A. and Martin, N. (2016). Kernel density estimation (KDE) with adaptive bandwidth selection for environmental contours of extreme sea states. *In: OCEANS 2016 MTS/IEEE. IEEE, Monterey, CA, USA*, pages 1–5.

Einmahl, J. H. and Segers, J. (2009). Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. *Annals of Statistics*, 37:2953–2989.

Environmental Agency (2005). CSG 15 Final Project Report. randd.defra.gov.uk/Document.aspx?Document=AE1143_3886_FRP.pdf.

Escobar-Bach, M., Goegebeur, Y., and Guillou, A. (2018). Local robust estimation of the Pickands dependence function. *The Annals of Statistics*, 46:2806–2843.

Frahm, G. (2006). On the extremal dependence coefficient of multivariate distributions. *Statistics and Probability Letters*, 76:1470–1481.

Frahm, G., Junker, M., and Schmidt, R. (2005). Estimating the tail-dependence coefficient: Properties and pitfalls. *Insurance: Mathematics and Economics*, 37:80–100.

Gentle, J. E. (2002). *Elements of Computational Statistics*. Springer-Verlag.

Gilleland, E., Nychka, D., and Schneider, U. (2006). Spatial models for the distribution of extremes. *Hierarchical Modelling for the Environmental Sciences: Statistical Methods and Applications*, 216:170–183.

Gouldby, B., Wyncoll, D., Panzeri, M., Franklin, M., Hunt, T., Hames, D., Tozer, N., Hawkes, P., Dornbusch, U., and Pullen, T. (2017). Multivariate extreme value modelling of sea conditions around the coast of England. *Proceedings of the Institution of Civil Engineers: Maritime Engineering*, 170:3–20.

Groll, N. and Weisse, R. (2017). A multi-decadal wind-wave hindcast for the North Sea 1949-2014: CoastDat2. *Earth System Science Data*, 9:955–968.

Guerrero, M. B., Huser, R., and Ombao, H. (2023). Conex–Connect: Learning patterns in extremal brain connectivity from MultiChannel EEG data. *The Annals of Applied Statistics*, 17:178–198.

Guillotte, S. and Perron, F. (2016). Polynomial Pickands functions. *Bernoulli*, 22:213–241.

Gumbel, E. J. (1960). Bivariate Exponential Distributions. *Journal of the American Statistical Association*, 55:698–707.

Hall, P. and Tajvidi, N. (2000). Distribution and dependence-function estimation for bivariate extreme-value distributions. *Bernoulli*, 6:835–844.

Haselsteiner, A. F., Coe, R. G., Manuel, L., Chai, W., Leira, B., Clarindo, G., Soares, C. G., Ásta Hannesdóttir, Dimitrov, N., Sander, A., Ohlendorf, J. H., Thoben, K. D., de Hauteclocque, G., Mackay, E., Jonathan, P., Qiao, C., Myers, A., Rode, A., Hildebrandt, A., Schmidt, B., Vanem, E., and Huseby, A. B. (2021). A benchmarking exercise for environmental contours. *Ocean Engineering*, 236:1–29.

Haselsteiner, A. F., Nguyen, P. T., Coe, R. G., Martin, N., Manuel, L., and Eckert-Gallup, A. (2019). A benchmarking exercise on estimating extreme environmental conditions: Methodology and Baseline results. *Proceedings of the International Conference on Offshore Mechanics and Arctic Engineering - OMAE*, 3:1–10.

Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 66:497–546.

Hill, B. M. (1975). A Simple General Approach to Inference About the Tail of a Distribution. *The Annals of Statistics*, 3:1163–1174.

Huser, R. and Wadsworth, J. L. (2019). Modeling Spatial Processes with Unknown Extremal Dependence Class. *Journal of the American Statistical Association*, 114:434–444.

Hüsler, J. and Reiss, R. D. (1989). Maxima of normal random vectors: Between independence and complete dependence. *Statistics & Probability Letters*, 7:283–286.

Institute for Nuclear Protection and Safety (2000). Report on the flooding of the Blayais site on 27 December 1999. https://www.irsn.fr/FR/expertise/rapports_expertise/Documents/surete/rapport_sur_l_inondation_du_site_du_blayais.pdf.

International Atomic Energy Agency (2015). The Fukushima Daiichi Accident: Report by the Director General. https://www-pub.iaea.org/MTCD/Publications/PDF/Pub1710-ReportByTheDG-Web.pdf.

Joe, H. (1997). *Multivariate Models and Multivariate Dependence Concepts*. Chapman and Hall/CRC.

Jonathan, P., Ewans, K., and Flynn, J. (2014a). On the estimation of ocean engineering design contours. *Journal of Offshore Mechanics and Arctic Engineering*, 136:1–8.

Jonathan, P., Ewans, K., and Randell, D. (2014b). Non-stationary conditional extremes of northern North Sea storm characteristics. *Environmetrics*, 25:172–188.

Keef, C., Papastathopoulos, I., and Tawn, J. A. (2013a). Estimation of the conditional distribution of a multivariate variable given that one of its components is large: Additional constraints for the Heffernan and Tawn model. *Journal of Multivariate Analysis*, 115:396–404.

Keef, C., Tawn, J. A., and Lamb, R. (2013b). Estimating the probability of widespread flood events. *Environmetrics*, 24:13–21.

Knochenhauer, M. and Louko, P. (2004). Guidance for External Events Analysis. *Probabilistic Safety Assessment and Management*, pages 1498–1503.

Koenker, R., Chernozhukov, V., He, X., and Peng, L. (2017). *Handbook of Quantile Regression*. Chapman and Hall/CRC.

Krock, M., Bessac, J., Stein, M. L., and Monahan, A. H. (2021). Nonstationary seasonal model for daily mean temperature distribution bridging bulk and tails. *arXiv*, 2110.10046:1–19.

Kunsch, H. R. (1989). The Jackknife and the Bootstrap for General Stationary Observations. *The Annals of Statistics*, 17:1217–1241.

Kyselý, J., Picek, J., and Beranová, R. (2010). Estimating extremes in climate change simulations using the peaks-over-threshold method with a non-stationary threshold. *Global and Planetary Change*, 72:55–68.

Lamb, R., Keef, C., Tawn, J., Laeger, S., Meadowcroft, I., Surendran, S., Dunning, P., and Batstone, C. (2010). A new method to assess the risk of local and widespread flooding on rivers and coasts. *Journal of Flood Risk Management*, 3:323–336.

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer New York.

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

Ledford, A. W. and Tawn, J. A. (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 59:475–499.

Lindt, J. W. V. D. and Niedzwecki, J. M. (2000). Environmental contour analysis in earthquake engineering. *Engineering Structures*, 22:1661–1676.

Liu, Y. and Tawn, J. A. (2014). Self-consistent estimation of conditional multivariate extreme value distributions. *Journal of Multivariate Analysis*, 127:19–35.

Mackay, E. and Haselsteiner, A. F. (2021). Marginal and total exceedance probabilities of environmental contours. *Marine Structures*, 75:1–24.

Mackay, E. and Jonathan, P. (2020). Assessment of return value estimates from stationary and non-stationary extreme value models. *Ocean Engineering*, 207:107406.

Manuel, L., Nguyen, P. T., Canning, J., Coe, R. G., Eckert-Gallup, A. C., and Martin, N. (2018). Alternative approaches to develop environmental contours from metocean data. *Journal of Ocean Engineering and Marine Energy*, 4:293–310.

Marcon, G., Naveau, P., and Padoan, S. (2017a). A semi-parametric stochastic generator for bivariate extreme events. *Stat*, 6:184–201.

Marcon, G., Padoan, S. A., and Antoniano-Villalobos, I. (2016). Bayesian inference for the extremal dependence. *Electronic Journal of Statistics*, 10:3310–3337.

Marcon, G., Padoan, S. A., Naveau, P., Muliere, P., and Segers, J. (2017b). Multivariate nonparametric estimation of the Pickands dependence function using Bernstein polynomials. *Journal of Statistical Planning and Inference*, 183:1–17.

Mattei, J., Vial, E., Rebour, V., Liemersdorf, H., and Turschmann, M. (2001). Generic results and conclusions of re-evaluating the flooding protection in French and German nuclear power plants. https://inis.iaea.org/collection/NCLCollectionStore/_Public/33/021/33021372.pdf.

Mentaschi, L., Vousdoukas, M., Voukouvalas, E., Sartini, L., Feyen, L., Besio, G., and Alfieri, L. (2016). The transformed-stationary approach: A generic and simplified methodology for non-stationary extreme value analysis. *Hydrology and Earth System Sciences*, 20:3527–3547.

Met Office Hadley Centre (2018). UKCP18 Regional Climate Model Projections for the UK. catalogue.ceda.ac.uk/uuid/b4d24b3df3754b9d9028447eb3cd89c6.

Mhalla, L., Chavez-Demoulin, V., and Naveau, P. (2017). Non-linear models for extremal dependence. *Journal of Multivariate Analysis*, 159:49–66.

Mhalla, L., de Carvalho, M., and Chavez-Demoulin, V. (2019a). Regression-type models for extremal dependence. *Scandinavian Journal of Statistics*, 46:1141–1167.

Mhalla, L., Opitz, T., and Chavez-Demoulin, V. (2019b). Exceedance-based nonlinear regression of tail dependence. *Extremes*, 22:523–552.

Murphy-Barltrop, C. J. R. and Wadsworth, J. L. (2022). Modelling non-stationarity in asymptotically independent extremes. *arXiv*, 2203.05860.

Murphy-Barltrop, C. J. R., Wadsworth, J. L., and Eastoe, E. F. (2023). New estimation methods for extremal bivariate return curves. *Environmetrics*, e2797:1–22.

National Diet of Japan (2012). The official report of The Fukushima Nuclear Accident Independent Investigation Commission - Executive summary. https://warp.da.ndl.go.jp/info:ndljp/pid/3856371/naiic.go.jp/en/report/.

Nogaj, M., Parey, S., and Dacunha-Castelle, D. (2007). Non-stationary extreme models and a climatic application. *Nonlinear Processes in Geophysics*, 14:305–316.

Nolde, N. (2014). Geometric interpretation of the residual dependence coefficient. *Journal of Multivariate Analysis*, 123:85–95.

Nolde, N. and Wadsworth, J. L. (2022). Linking representations for multivariate extremes via a limit set. *Advances in Applied Probability*, 54:688–717.

Northrop, P. J., Attalides, N., and Jonathan, P. (2017). Cross-validatory extreme value threshold selection and uncertainty with application to ocean storm severity. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 66:93–120.

Northrop, P. J. and Jonathan, P. (2011). Threshold modelling of spatially dependent non-stationary extremes with application to hurricane-induced wave heights. *Environmetrics*, 22:799–809.

Office for Nuclear Regulation (2014). Safety Assessment Principles - 2014 edition (Revision 1, January 2020). https://www.onr.org.uk/saps/saps2014.pdf.

Office for Nuclear Regulation (2017). Introduction to Extreme Value Theory and Constructing Hazard Curves. https://www.onr.org.uk/documents/2017/onr-rrr-054.pdf.

Office for Nuclear Regulation (2018). NS-TAST-GD-013. https://www.onr.org.uk/operational/tech_asst_guides/ns-tast-gd-013.htm.

Office for Nuclear Regulation (2021). NS-TAST-GD-013 – Annex 2 - Meteorological Hazards. https://www.onr.org.uk/consultations/2021/external-hazards/ns-tast-gd-013-annex-2.pdf.

Peng, L. (1999). Estimation of the coefficient of tail dependence in bivariate extremes. *Statistics and Probability Letters*, 43:399–409.

Pickands, J. (1975). Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics*, 3:119–131.

Pickands, J. (1981). Multivariate extreme value distribution. *Proceedings 43th, Session of International Statistical Institution, 1981*.

Politis, D. N. and Romano, J. P. (1994). The Stationary Bootstrap. *Journal of the American Statistical Association*, 89:1303–1313.

Quinn, N., Bates, P. D., Neal, J., Smith, A., Wing, O., Sampson, C., Smith, J., and Heffernan, J. (2019). The Spatial Dependence of Flood Hazard and Risk in the United States. *Water Resources Research*, 55:1890–1911.

Ramos, A. and Ledford, A. (2009). A new class of models for bivariate joint tails. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 71:219–241.

Resnick, S. (2002). Hidden Regular Variation, Second Order Regular Variation and Asymptotic Independence. *Extremes*, 5:303–336.

Resnick, S. I. (1987). *Extreme Values, Regular Variation and Point Processes*. Springer New York.

Rohrbeck, C. and Cooley, D. (2021). Simulating flood event sets using extremal principal components. *arXiv*, 2106.00630.

Rootzén, H. and Katz, R. W. (2013). Design Life Level: Quantifying risk in a changing climate. *Water Resources Research*, 49:5964–5972.

Rootzén, H. and Tajvidi, N. (2006). Multivariate generalized Pareto distributions. *Bernoulli*, 12:917–930.

Ross, E., Astrup, O. C., Bitner-Gregersen, E., Bunn, N., Feld, G., Gouldby, B., Huseby, A., Liu, Y., Randell, D., Vanem, E., and Jonathan, P. (2020). On environmental contours for marine and coastal design. *Ocean Engineering*, 195:106194.

Salvadori, G. and Michele, C. D. (2004). Frequency analysis via copulas: Theoretical aspects and applications to hydrological events. *Water Resources Research*, 40:1–17.

Serinaldi, F. (2015). Dismissing return periods! *Stochastic Environmental Research and Risk Assessment*, 29:1179–1189.

Shooter, R., Ross, E., Ribal, A., Young, I. R., and Jonathan, P. (2021). Spatial dependence of extreme seas in the North East Atlantic from satellite altimeter measurements. *Environmetrics*, 32:1–15.

Sigauke, C. and Bere, A. (2017). Modelling non-stationary time series using a peaks over threshold distribution with time varying covariates and threshold: An application to peak electricity demand. *Energy*, 119:152–166.

Simpson, E. S. and Tawn, J. A. (2022). Estimating the limiting shape of bivariate scaled sample clouds for self-consistent inference of extremal dependence properties. *arXiv*, 2207.02626.

Simpson, E. S. and Wadsworth, J. L. (2017). Introduction to Extreme Value Theory and Constructing Hazard Curves. https://www.onr.org.uk/documents/2017/onr-rrr-054.pdf.

Simpson, E. S. and Wadsworth, J. L. (2021). Conditional modelling of spatio-temporal extremes for Red Sea surface temperatures. *Spatial Statistics*, 41:1–17.

Simpson, E. S., Wadsworth, J. L., and Tawn, J. A. (2020). Determining the dependence structure of multivariate extremes. *Biometrika*, 107:513–532.

Sklar, A. (1959). Fonctions de repartition a n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231.

Smith, R. L. (1989). Extreme Value Analysis of Environmental Time Series: An Application to Trend Detection in Ground-Level Ozone. *Statistical Science*, 4:367–377.

Stein, M. L. (2021). A parametric model for distributions with flexible behavior in both tails. *Environmetrics*, 32:e2658.

Stephenson, A. (2002). evd: Extreme Value Distributions. *R News*, 2:31–32.

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

Tendijck, S., Eastoe, E., Tawn, J., Randell, D., and Jonathan, P. (2021). Modeling the Extremes of Bivariate Mixture Distributions With Application to Oceanographic Data. *Journal of the American Statistical Association*, 0:1–12.

Tilloy, A., Malamud, B. D., Winter, H., and Joly-Laugel, A. (2020). Evaluating the efficacy of bivariate extreme modelling approaches for multi-hazard scenarios. *Natural Hazards and Earth System Sciences*, 20:2091–2117.

Vanem, E., Guo, B., Ross, E., and Jonathan, P. (2020). Comparing different contour methods with response-based methods for extreme ship response analysis. *Marine Structures*, 69:1–19.

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21:5–42.

Varty, Z., Tawn, J. A., Atkinson, P. M., and Bierman, S. (2021). Inference for extreme earthquake magnitudes accounting for a time-varying measurement process. *arXiv*, 2102.00884.

Velarde, J., Vanem, E., Kramhøft, C., and Sørensen, J. D. (2019). Probabilistic analysis of offshore wind turbines under extreme resonant response: Application of environmental contour method. *Applied Ocean Research*, 93:1–16.

Velthoen, J., Dombry, C., Cai, J.-J., and Engelke, S. (2021). Gradient boosting for extreme quantile regression. *arXiv*, 2103.00808.

Vettori, S., Huser, R., and Genton, M. G. (2018). A comparison of dependence function estimators in multivariate extremes. *Statistics and Computing*, 28:525–538.

Wadsworth, J. and Campbell, R. (2022). Statistical inference for multivariate extremes via a geometric approach. *arXiv*, 2208.14951.

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

Wadsworth, J. L. and Tawn, J. A. (2013). A new representation for multivariate tail probabilities. *Bernoulli*, 19:2689–2714.

Wadsworth, J. L., Tawn, J. A., Davison, A. C., and Elton, D. M. (2017). Modelling across extremal dependence classes. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 79:149–175.

Wei, Y., Pere, A., Koenker, R., and He, X. (2006). Quantile regression methods for reference growth charts. *Statistics in Medicine*, 25:1369–1382.

Wood, S. (2021). Mixed GAM Computation Vehicle with Automatic Smoothness Estimation. *R Package*.

Wood, S. N. (2017). *Generalized Additive Models*. Chapman and Hall/CRC.

Worms, J. and Naveau, P. (2022). Record events attribution in climate studies. *Environmetrics*, 33:1–16.

Yee, T. W. and Stephenson, A. G. (2007). Vector generalized linear and additive extreme value models. *Extremes*, 10:1–19.

Youngman, B. D. (2019). Generalized Additive Models for Exceedances of High Thresholds With an Application to Return Level Estimation for U.S. Wind Gusts. *Journal of the American Statistical Association*, 114:1865–1879.

Yu, C., Blaha, O., Kane, M., Wei, W., Esserman, D., and Zelterman, D. (2022). Regression methods for the appearances of extremes in climate data. *Environmetrics*, 33:1–14.