FORECASTING OF NONLINEAR EXTREME QUANTILES USING COPULA MODELS

by

VINCENZO COIA

B.Sc. (Mathematics), Brock University, June 2011
B.Sc. (Biology), Brock University, October 2011
M.Sc., Brock University, October 2012

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Statistics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

February 2017

Abstract

Forecasts of extreme events are useful in order to prepare for disaster. Such forecasts are usefully communicated as an upper quantile function, and in the presence of predictors, can be estimated using quantile regression techniques. This dissertation proposes methodology that seeks to produce forecasts that (1) are consistent in the sense that the quantile functions are valid (non-decreasing); (2) are flexible enough to capture the dependence between the predictors and the response; and (3) can reliably extrapolate into the tail of the upper quantile function. To address these goals, a family of proper scoring rules is first established that measure the goodness of upper quantile function forecasts. To build a model of the conditional quantile function, a method that uses pair-copula Bayesian networks or vine copulas is proposed. This model is fit using a new class of estimators called the composite nonlinear quantile regression (CNQR) family of estimators, which optimize the scores from the previous scoring rules. In addition, a new parametric copula family is introduced that allows for a non-constant conditional extreme value index, and another parametric family is introduced that reduces a heavy-tailed response to a light tail upon conditioning. Taken altogether, this work is able to produce forecasts satisfying the three goals. This means that the resulting forecasts of extremes are more reliable than other methods, because they more adequately capture the insight that predictors hold on extreme outcomes. This work is applied to forecasting extreme flows of the Bow River at Banff, Alberta, for flood preparation, but can be used to forecast extremes of any continuous response when predictors are present.

Preface

This dissertation was completed under the supervision of Dr. Natalia Nolde and Dr. Harry Joe, both of whom provided several ideas put forth in this dissertation.

The motivating problem introduced in Section 2.1 came from the town of Canmore, Alberta, and the Alberta Environment and Parks. Dr. Nolde had the idea of addressing their need by forecasting high quantiles, making use of Extreme Value Theory. Dr. Nolde suggested the use of calibration and proper scoring rules to assess forecasts. I modified the forecasts to be the entire tail of a predictive distribution, and modified existing proper scoring rules to this type of forecast, in Section 3.2.

The idea of using vine copulas to model the conditional quantiles came from Dr. Joe. I worked out the details behind the non-identifiability issues that arose, as well as the fitting procedure, as discussed in Section 5.1.

Dr. Joe presented the issue of finding a copula family that allows for a non-constant conditional EVI. I developed the IG copula family with $k = 2$, with the IGL(2) copula[1] as a limit. Dr. Joe suggested generalizing the limit copula to attain higher dependence. I identified exactly how that can be done, by choosing a distribution function satisfying certain properties. This function, for the IGL(2) copula, was recognized by Dr. Joe as the Gamma $(2, 1)$ distribution function. I was then able to generalize this to the Gamma $(k, 1)$ distribution function, to obtain the IG and IGL families presented in Section 6.1.

Dr. Joe presented the problem of generalizing the estimation method of Bouyé and Salmon (2009). I developed an *ad hoc* generalization, which ended up being the CNQR estimator with identity transformation functions. Dr. Joe identified the addition of a transformation function, and I generalized that further to allow for different transformation functions for each quantile level. I later found that the family of CNQR estimators is rooted in the theory of proper scoring rules, so that these estimators were no longer *ad hoc*. Dr. Nolde and Dr. Joe were instrumental in identifying errors in my proofs of the

[1]The construction of the IGL(2) copula that I was working with was different than what is presented in Section 6.1. I was working with the $\alpha = 1$ version of the DJ copula family, presented in Proposition 6.2.2.

asymptotic properties of the CNQR estimators, and putting forth suggestions for fixing them. The results can be found in Section 5.2.

Chapter 7 makes use of two R (R Core Team, 2015) packages that I developed, and am still in the process of developing. They are the `copsupp` and `cnqr` packages. The `copsupp` package was originally developed with the intention of supplementing the `CopulaModel` (Joe and Krupskii, 2014) package with miscellaneous useful functions, but has evolved a purpose of providing a user-friendly interface for working with vine copula models. As such, its name will likely change to something more appropriate, like `rvine` or `regvine`, before being published. The `copsupp` package has some code written by Dr. Joe and Bo Chang. The `cnqr` package allows for the implementation of the proposed regression methodology, discussed in Chapter 5. For both packages, Dr. Joe was instrumental in helping me get some tricky aspects of the code to work. In addition, the code for the analysis in Chapter 7 benefited greatly by the careful eyes of Dr. Nolde and Dr. Joe.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

CCEVI Copula Conditional Extreme Value Index. See Definition 2.4.2.

CNQR Composite Nonlinear Quantile Regression. See Section 5.2.

DJ The Durante and Jaworski class of copulas. See Section 6.2.1 for a description.

ds Deseasonalized, referring to deseasonalized discharge of the Bow River at Banff. See Section F.1 for a definition.

$\mathrm{Exp}\,(\mu)$ The exponential distribution with mean $\mu > 0$, defined with distribution function

$$F_{\mathrm{Exp}}\,(x;\mu) = 1 - \exp\left(-\frac{x}{\mu}\right)$$

for $x > 0$.

extDJ The extended DJ copula family. See Section 6.2.2 for a description.

EVI Extreme Value Index. See Section 2.3 for a description.

$\mathrm{IG}\,(k,\theta)$ The Integrated Gamma copula with parameters $k > 1$ and $\theta > 0$. See Equation (6.1.12) for a definition.

$\mathrm{IGL}\,(k)$ The Integrated Gamma Limit copula with parameter $k > 1$. See Equation (6.1.9) for a definition.

iid Independent and identically distributed, referring to a set of random variables.

$\mathrm{Gamma}\,(k,\theta)$ The gamma distribution with shape parameter $k > 0$ and scale parameter $\theta > 0$, defined by

$$F_{\mathrm{Gamma}}\,(x;k,\theta) = 1 - \frac{\Gamma^*\left(k, x/\theta\right)}{\Gamma\,(k)}$$

for $x \geq 0$.

GPD Generalized Pareto Distribution.

$\mathcal{N}\left(\mu, \sigma^2\right)$ The Gaussian distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$. This is also used to indicate the multivariate Gaussian distribution.

$\mathrm{Par}\left(\sigma, \alpha\right)$ The Type I Pareto distribution, with scale parameter $\sigma > 0$ and shape parameter $\alpha > 0$, defined by

$$F_{\mathrm{Par}}\left(x; \sigma, \alpha\right) = 1 - \left(\frac{x}{\sigma}\right)^{-\alpha}$$

for $x \geq \sigma$.

PCBN Pair-Copula Bayesian Network. See Section 2.4.2 for a definition.

PIT Probability Integral Transform, of a continuous random variable. If W is a random variable, then $F_W\left(W\right)$ is the PIT score of W, and follows a $\mathrm{Unif}\left(0, 1\right)$ distribution.

resp. Respectively; used within braces to complement a result.

SWE Snow Water Equivalent, which is the depth of water obtained after having melted a snowpack.

$\mathrm{Unif}\left(a, b\right)$ The continuous uniform distribution from a to b, where $a < b$.

List of Symbols

This section defines the key symbols used throughout this dissertation. To aid in defining some symbols, let $\boldsymbol{W} = (W_1, \ldots, W_d)^\top$ be a random vector of continuous random variables for some integer $d \geq 2$, and let h be a real-valued function with domain in $\mathbb{R}$. Whenever h is defined such that $h(x_0)$ does not exist when x_0 is in the domain of h, then the convention is adopted where this quantity is defined as the appropriate limit of h at x_0.

$:=$	Used for defining symbols. The quantity on the left-hand side is defined as the quantity on the right-hand side. The symbol $=:$ means that the quantity on the right-hand side is defined as the quantity on the left-hand side.
$\to_d$	Convergence in distribution.
$\to_p$	Convergence in probability.
$a : b$	The set of integers from a to b inclusive, where $a \leq b$ are integers – that is, $[a, b] \cap \mathbb{Z}$. The convention is adopted that $b : a = \emptyset$.
$C_{\boldsymbol{W}}$	The copula of $\boldsymbol{W}$.
$c_{\boldsymbol{W}}$	The copula density of $C_{\boldsymbol{W}}$, defined as $\mathrm{D}_{1:d}\, C_{\boldsymbol{W}}$.
R	The reflection operator acting on a d-dimensional copula, defined as $\mathrm{R} = \mathrm{R}_{1:d}$.
$\mathrm{R}_{\mathcal{I}}$	For a set of integers $\mathcal{I} \subset 1 : d$ and d-dimensional copula C, $\mathrm{R}_{\mathcal{I}} C$ is the $\mathcal{I}$-reflection of copula C, defined in Appendix A.1.
C^+	The comonotonicity copula, $C^+ : (u, v) \mapsto \min(u, v)$.
C^-	The countermonotonicity copula, $C^- : (u, v) \mapsto \max(u + v - 1, 0)$.
$C^\perp$	The independence copula, $C^\perp : (u, v) \mapsto uv$.

$C_{a_1 a_2;\boldsymbol{a}_{3:d}}$ The copula distribution function joining the pair $(W_{a_1}, W_{a_2}) \mid \boldsymbol{W}_{\boldsymbol{a}_{3:d}}$, where $\boldsymbol{a} = (a_1, \ldots, a_d)^\top$ is some permutation of $1:d$. This notation is shorthand for the more proper notation that replaces each a_i with W_{a_i}, $i = 1, \ldots, d$.

$c_{a_1 a_2;\boldsymbol{a}_{3:d}}$ The copula density function joining the pair $(W_{a_1}, W_{a_2}) \mid \boldsymbol{W}_{\boldsymbol{a}_{3:d}}$, where $\boldsymbol{a} = (a_1, \ldots, a_d)^\top$ is some permutation of $1:d$. This notation is shorthand for the more proper notation that replaces each a_i with W_{a_i}, $i = 1, \ldots, d$.

$C_{a_1|a_2;\boldsymbol{a}_{3:d}}$ A conditional distribution of a bivariate copula, defined as $\mathrm{D}_1\, C_{a_1,a_2;\boldsymbol{a}_{3:d}}$, where $\boldsymbol{a} = (a_1, \ldots, a_d)^\top$ is some permutation of $1:d$. For example, when $d = 2$ and C is a bivariate copula, $C_{2|1} = \mathrm{D}_1\, C$ and $C_{1|2} = \mathrm{D}_2\, C$ are conditional distribution functions. This notation is shorthand for the more proper notation that replaces each a_i with W_{a_i}, $i = 1, \ldots, d$.

Cal Calibration, defined in Equation (3.1.1).

C_{DJ} The DJ copula family, defined in Equation (6.2.1).

C_{extDJ} The extDJ copula family, defined in Equation (6.2.6).

C_{IG} The IG copula family, defined in Equation (6.1.9).

C_{IGL} The IGL copula family, defined in Equation (6.1.12).

$\hat{c}_{\boldsymbol{W}}$ The reflection copula density, $\mathrm{D}_{12}\, \hat{C}_{\boldsymbol{W}}$.

$\hat{C}_{\boldsymbol{W}}$ The reflection copula $\mathrm{R}C_{\boldsymbol{W}}$, so that the "hat" can also be thought of as the reflection operator.

δ Mapping from a calendar date to the day of year. The range is $1:366$.

D The differential operator, defined as $\mathrm{D}h = h'$.

D_j The differential operator on the j'th argument. If $h_n : \mathbb{R}^n \to \mathbb{R}$ is a function differentiable in its j'th argument for some integer $n \geq 1$, then $\mathrm{D}_j\, h_n$ for $1 \leq j \leq n$ is the derivative with respect to the j'th argument, defined as

$$\mathrm{D}_j\, h_n\left(x_1, \ldots, x_n\right) = \frac{\partial}{\partial x_j}\, h_n\left(x_1, \ldots, x_n\right).$$

$\mathcal{D}\left(G_\xi\right)$ The (maximum) domain of attraction, loosely defined as the set of all distribution functions having sample maxima that tend to G_ξ in distribution (up

to location and scale). A distribution function F is said to have *extreme value index* $\xi \in \mathbb{R}$ if $F \in \mathcal{D}\left(G_\xi\right)$.

$\mathbb{E}$ Expectation. For W_1, the expectation is $\mathbb{E}\left(W_1\right) = \int w \, \mathrm{d}F_{W_1}\left(w\right)$, sometimes written $\mathbb{E}_{W_1}$ to emphasize expectation over the distribution of W_1.

$F_{\boldsymbol{W}}$ The distribution function of $\boldsymbol{W}$. When it is clear, the convention is adopted that $F_{a:b} = F_{\boldsymbol{W}_{a:b}}$ for integers $1 \leq a \leq b \leq d$.

$f_{\boldsymbol{W}}$ The density of $\boldsymbol{W}$, if it exists. When it is clear, the convention is adopted that $f_{a:b} = f_{\boldsymbol{W}_{a:b}}$ for integers $1 \leq a \leq b \leq d$.

$\mathcal{F}$ The set of all possible conditional distributions of Y.

$\hat{\mathcal{F}}$ The set of all upper distribution functions corresponding to the forecast space, defined as $\hat{\mathcal{F}} = \left\{Q^{\leftarrow} : Q \in \hat{\mathcal{Q}}\right\}$.

$\mathscr{F}$ A sigma field of sets in Ω.

g Transformation function, part of the proper scoring rule in Equation (3.2.8).

G_ξ The distribution function of the standard generalized extreme value distribution for $\xi \in \mathbb{R}$, defined as

$$G_\xi\left(x\right) = \begin{cases} \exp\left(-\left(1+\xi x\right)^{-1/\xi}\right), & \xi \neq 0 \\ \exp\left(-\exp\left(-x\right)\right), & \xi = 0 \end{cases}$$

for $x \in \left[-1/\xi, \infty\right)$ when $\xi > 0$, $x \in \mathbb{R}$ when $\xi = 0$, and $x \in \left(-\infty, -1/\xi\right]$ when $\xi < 0$.

$\mathscr{G}$ A sigma field within $\mathscr{F}$, interpreted as an "observation" or "experiment". Used for discussions of conditional probability.

H_k A generating function of the extDJ copula class that generates an IG copula; defined in Equation (6.2).

Γ The gamma function. For $k > 1$[2], the gamma function is

$$\Gamma\left(k\right) = \int_0^\infty x^{k-1} \exp\left(-x\right) \mathrm{d}x.$$

[2] Although the gamma function is also defined when k is complex, only $k > 1$ is needed.

Γ^* The (upper) incomplete gamma function. For $(k,t) \in (1,\infty) \times [0,\infty)$,

$$\Gamma^*(k,t) = \int_t^\infty x^{k-1} \exp(-x)\,\mathrm{d}x.$$

$h^\leftarrow$ The generalized inverse function of h. This is defined when h is non-decreasing by $h^\leftarrow(s) = \inf\left\{x \in \mathbb{R} : h(x) \geq s\right\}$ for all s in the range of h. When h is non-increasing, the generalized inverse is defined as $h^\leftarrow(s) = \bar{h}^\leftarrow(-s)$, where $\bar{h} = -h$, resulting in $h^\leftarrow(s) = \sup\left\{x \in \mathbb{R} : h(x) < s\right\}$. If a and b are two monotone real functions, then $a \circ b$ is monotone. If a and b have opposite monotonicity, then $a \circ b$ is decreasing. If a and b are both increasing, $(a \circ b)^\leftarrow = b^\leftarrow \circ a^\leftarrow$ is increasing.

h', h'' First and second derivatives of h, respectively.

$\mathbb{I}$ The indicator function. For a set $\mathcal{S}$,

$$\mathbb{I}_{\mathcal{S}}(x) = \begin{cases} 0, & x \in \mathcal{S}; \\ 1, & x \notin \mathcal{S}. \end{cases}$$

κ_ψ The kappa function for DJ generating function ψ, defined in Equation (E.1.1).

Ω The sample space that is the domain of Y.

p The number of predictors, $\boldsymbol{X}$.

$\mathcal{P}$ Probability measure on $\mathscr{F}$. Also used as the generic symbol for "probability".

ϕ The density of the standard Gaussian distribution.

Φ The distribution function of the standard Gaussian distribution.

Ψ_k A DJ generating function that generates an IGL copula; defined in Equation (6.1.7).

Q_{W_1} The quantile function of W_1, equal to $F_{W_1}^\leftarrow$.

$\mathcal{Q}$ The set of all conditional quantile functions of Y, defined as $\{F^\leftarrow : F \in \mathcal{F}\}$.

$\hat{\mathcal{Q}}$ The set of forecasts forming the forecast space.

RV_α The set of regularly varying functions (at infinity) with index of variation $\alpha \in \mathbb{R}$. A function $h \in \mathrm{RV}_\alpha$ if and only if

$$\lim_{x\to\infty} \frac{h(ux)}{h(x)} = u^\alpha$$

for any $u > 0$. Further, $h(x) = x^\alpha \ell(x)$ for x in the domain of h, where $\ell \in \mathrm{RV}_0$ is a slowly varying function. We say a function $h \in \mathrm{RV}_\alpha$ at 1^- if $x \mapsto h\left(1 - \frac{1}{x}\right) \in \mathrm{RV}_\alpha$ at infinity, so that $h(x) = (1-x)^{-\alpha}\,\ell\left((1-x)^{-1}\right)$ where $\ell \in \mathrm{RV}_0$ at 1^-.

ρ_τ The asymmetric absolute deviation loss function, defined in Equation (3.2.6).

τ_c Lower quantile level cutoff, defining the lower limit of the quantile function forecasts.

θ^+ For $\theta \in \mathbb{R}$, is the positive part of θ, defined as $\max(0, \theta)$.

θ^- For $\theta \in \mathbb{R}$, is the negative part of θ, defined as $\min(0, \theta)$.

$\boldsymbol{v}_{a:b}$ The subvector of $\boldsymbol{v} = \left(v_1, \ldots, v_p\right)^\top$, defined as $\boldsymbol{v}_{a:b} = (v_a, \ldots, v_b)^\top$ for integers $1 \le a \le b \le p$. The convention is adopted that $\boldsymbol{v}_{b:a}$ is a vector of length zero.

w Cross-quantile weight function, part of the proper scoring rule in Equation (3.2.8).

$\boldsymbol{X}$ A vector of p predictors.

$\mathscr{X}$ The predictor space; $\boldsymbol{X} \in \mathscr{X} \subset \mathbb{R}^p$.

Ξ_λ The Box-Cox power transformation, defined in Equation (2.5.2).

ξ_{W_1} The extreme value index of F_{W_1}, if it exists.

Y The response variable, to be forecast; a random variable with probability space $(\Omega, \mathscr{F}, \mathcal{P})$.

Y_t For times $t = 1, 2, \ldots$ are (not necessarily independent) replicates of Y, and is the response at time t.

$\mathscr{Y}$ The support of the distribution of Y, often taken to be $\mathbb{R}$.

Acknowledgements

It's incredible the impact a person can have on another's life. During my PhD program, I have had the fortune of being able to surround myself with truly inspiring people who have believed in my potential from the start. Without these people, not only would I have not accomplished this dissertation, but I would not have developed much as a person either.

First and foremost, my supervisors, Dr. Natalia Nolde and Dr. Harry Joe, are inspiring teachers and provided an effective learning environment. They have dedicated a generous amount of time and patience to help get me where I am today. Through regular meetings and feedback on my work, I have gained insight into their ideas, big-picture mindset, and pragmatic sense of what's important. Their patience has been a godsend, having dealt with a lot from me, from ignorance in research to the emotional ups and downs that proved inhibitive at times. In particular, Dr. Nolde's openness to learning new things is inspiring. Dr. Joe's computing practices are inspiring, and his extra help checking my code has been very helpful. Also, I appreciate Dr. Matías Salibián Barrera for his involvement with my supervisory committee – his input has been very valuable.

Of course, this work would not have been possible without my funders. I am grateful for the Natural Science and Engineering Research Council of Canada for a postgraduate scholarship, as well as the University of British Columbia for funding through the Four-Year Fellowship and Faculty of Science Graduate Award. Dr. Nolde and Dr. Joe have also graciously provided me with funding through research assistantships.

I owe it to the Department of Statistics for believing in my ability from the start, and allowing me to partake in this journey under its umbrella. The department offered a nurturing environment in several ways. Its incredibly inspiring members gave me valuable insight into professional effectiveness. The many seminars, conferences, visiting speakers, department teas, and grad trips offered valuable new perspectives and friendships. The volunteer positions that were made available, such as the grad rep position and membership on a search committee, have given me valuable leadership skills. This environment

allowed me to build valuable communication skills and a big-picture mindset of how statistics fits into the world.

Discussions with my colleagues have also proved invaluable. For instance, Bo Chang provided valuable input through discussions of software development, and David Lee provided insight through discussions on my research.

It was totally eye-opening to hear about the needs surrounding the Alberta 2013 flood, from speaking with BGC Engineering, the town of Canmore, the Alberta Environment and Parks, and Transalta. These conversations have provided me with a valuable new perspective of how government and industry respond and plan for natural disasters. I am grateful for the opportunity to be involved with addressing the issue of flooding.

Friends and family have played an integral role in the development of this dissertation. All of my friends have leant an ear and offered their support, and provided an outlet for stress relief and character building. Each person has helped me complete this dissertation in their own unique way. I'm grateful for my parents for always encouraging me to pursue education, and for fostering a love of learning. Colleen Lau, for helping free my mind and exposing me to new activities, through things such as our spontaneous days. Christina Bouchard, for helping me stay strong in the face of difficulties. Antonio Coia, for helping me connect with the outdoors. Derek Kief, for reminding me of the beauty of mathematics and science, and my research. Adam Higgs, for helping me gain confidence in myself. Al Cannon, for his empathy of my struggles, and his ideas. Tyler Kolpin, for enabling me to focus more of my energy on my research. Gareth Sirotnik, for bringing Zen meditation practice to UBC. Elena Shchurenkova and Andres Sanchez-Ordonez, for incredible hikes. Daniel Dinsdale and colleagues, for dinner parties.

This dissertation is important to me, as I am thrilled to be part of our civilization's pursuit of knowledge. There is still much that must be learned so that we can drive our civilization forward. I dream of being a driver of such progress, and this dissertation has given me more competence and a better idea of how this can be done. So, these people have not only greatly assisted with the completion of this dissertation, but by extension, have helped me step closer to my life vision. I am forever grateful for this. Thank you all.

To Zio Claudio, whose curiosity of nature is contagious.

Chapter 1

Introduction

On June 19, 2013, Alberta saw the Bow River and its tributaries swell beyond historical levels and devastate the river's nearby towns in a costly 3-day flood. The town of Canmore was one such town, whose unprepared inhabitants ended up stranded on an "island" surrounded by rushing water (Doll and Geddes, 2014).

How can one gain insight into such extreme events in advance? There will always be some amount of uncertainty in the outcome, but sometimes there are "precursors" that can provide insight to reducing such uncertainty. For the Alberta flood, it was heavy rain, a saturated ground, and a heavy snowpack together that made flooding almost certain. Such precursors are called *predictors* (or sometimes *covariates* or *input variables*), and when informative ones exist, the task becomes one of effectively garnering the insight that the predictors conceal about the outcome – extreme outcomes, in particular. Statistically, this "insight" is the conditional distribution of the response given the predictors.

A plethora of regression and machine learning methods exist that attempt to extract such insight on a *typical* outcome (i.e., the expectation of the conditional distribution) – linear regression being an example. But the expected outcome will be exceeded often, by its very nature of being a "middle" value. If large enough, this exceedance can be devastating, and unforeseen. Instead, one should garner insight into the chance of an *extreme* outcome. An extreme is usefully communicated as an extreme quantile – a predicted outcome that is exceeded with some small chance called an *exceedance probability*[1].

However, issuing a single or a few quantiles (point forecasts) is still quite limited in the amount of information conveyed, and the choice of exceedance probabilities is arbitrary. It has been recognized as early as Cooke (1906), and stressed by such authors as Krzysztofowicz (2001) and Gneiting and Raftery (2007), that it is important to communicate quantiles for *all* exceedance probabilities (a probabilistic forecast) – that is, convey

[1] One minus the exceedance probability is called a *quantile level*.

the entire conditional distribution. But quantiles having small exceedance probabilities are uninteresting in the context of extremes, and estimating these can compromise the quality of the estimates of extreme quantiles. Instead, it is more sensible to focus only on forecasting all extreme quantiles (i.e., the tail of the conditional distribution). The goal of this dissertation is to find a way to build a model that produces "good" forecasts of this "upper quantile function" type.

A model that produces forecasts is called a *forecaster*. To build a forecaster, a formula for obtaining a quantile from both the exceedance probability and the predictor is required. Three qualities are sought in a forecaster, as evaluated by forecasts on future events (as opposed to past events already "known" by the forecaster):

1. consistency of all forecasts produced, so that forecasts are valid upper quantile functions;

2. ability to extrapolate into the distributional tail reliably, so that extremes are properly represented; and

3. flexibility across the space of predictors, so that the unique effect of predictors (such as tail dependence) can be accommodated.

Existing methods to create such forecasts fall under the "quantile regression" umbrella. Many methodologies exist that can handle such regression, but have downfalls.

One commonly used method is linear quantile regression, introduced by Koenker and Bassett (1978). This method presumes that each quantile has a generic linear trend in the predictors, unrelated to other quantiles (non-parametric with respect to exceedance level). The estimation method enforces an upper bound to the distributional tails, violating the extrapolation requirement. Further, this method can result in non-monotonic predictive distributions, since no relation across exceedance probabilities is enforced. Modifications have been proposed to fix the non-monotonicity problem (cf. Bondell et al., 2010; Chernozhukov et al., 2010). Others add parametric assumptions across exceedance levels (cf. Zou and Yuan, 2008), but do not properly focus on the decay rate of the distribution's tail, and do not allow for flexibility across the predictor space.

There are also local fitting methods, such as a kernel method (Spokoiny et al., 2013), that do not impose any structure across the predictor space (i.e., non-parametric in the predictors), except that nearby predictors should convey similar predictions. However, these methods only use information from past scenarios where predictors were observed to be similar. When more predictors are considered, the chance of finding similar past situations dramatically decreases, so that less data are available for making a future prediction. And whatever data *are* available are likely "typical" values under that scenario,

so that minimal information about extremes is available. Ultimately, this "data sparsity" issue results in highly uncertain forecasts.

Newer methods are available that indicate a (parametric) relationship amongst both exceedance probabilities and predictors, using copula models. These have the benefit of producing consistent forecasts, and are able to incorporate Extreme Value Theory to justify a model for the distributional tail. Bouyé and Salmon (2009) use bivariate copulas in the one predictor setting for quantile regression. Kraus and Czado (2017) use vine copula models to accommodate multiple predictors, but focus on fitting the entire joint distribution of the predictors and response, and do not optimize the fit of the predictive distribution's tail.

This dissertation also uses vine copula models (extended to pair-copula Bayesian networks). Such models provide a flexible approach for modelling dependence amongst the predictors and response, and offer far more flexibility than the multivariate Gaussian distribution can provide. For example, perhaps the combination of heavy rain, a saturated ground, and heavy snowpack had more of a detrimental impact on the Alberta flood than the independent sum of those effects. This type of "tail dependence" can be captured with vine copulas. This modelling technique is adapted to fitting the conditional distribution's tail, and is estimated using the proposed *composite nonlinear quantile regression* (CNQR) family of estimators. This approach to quantile regression is one of the major new contributions of this dissertation. Also, the CNQR estimator has roots in the theory of proper scoring rules. Describing a proper scoring rule for upper quantile function forecasts is a major contribution of this dissertation.

The decay rate of the conditional distribution into the tail is determined by the conditional extreme value index (EVI), which might not be constant in the predictors. Trends in the conditional EVI must be captured by the underlying copula. A new copula family allowing for a non-constant conditional EVI is proposed, as well as one that turns a heavy-tailed marginal distribution into a light-tailed conditional distribution – another major new contribution of this dissertation.

The outline of the remainder of this dissertation is as follows. In Chapter 2, the example motivating the work of this dissertation is introduced in Section 2.1, followed by a formal description of the statistical objective in Section 2.2. The chapter continues by describing some requisite background in Extreme Value Theory (Section 2.3) and copulas (Section 2.4), the two of which are tied together with new concepts related to the conditional EVI (Section 2.5). Chapter 3 describes what it means for a forecaster to be "good". The new research here begins with Section 3.2.2, where a new class of proper scoring rules is proposed, and new research continues for the remainder of the chapter. A

more detailed description of existing methodology for building a forecaster is discussed in Chapter 4, which is followed by a discussion of the proposed methodology and proposed estimation in Chapter 5 (and contains all new research). To allow for greater flexibility in modelling, new copula families are introduced in Chapter 6 that allow for a non-constant or 0 conditional EVI (the entire chapter is new research). The methodologies discussed in this dissertation are compared through an application to predict extreme flows of the Bow River at Banff, Alberta, in Chapter 7, which is all new research. Chapter 8 concludes with a summary of the findings, and a discussion of some future research.

This dissertation also contains several appendices to supplement the material in the body. Appendix A gives formulas for vine and reflection copulas. Appendices B to E consist of new research, and contains proofs related to the conditional EVI of Section 2.5, the new proper scoring rules of Chapter 3, the CNQR estimator of Section 5.2, and the new copula families of Chapter 6 (respectively). Appendix F ends with details about the data analysis in Chapter 7, and is all new research.

Chapter 2

Background

The town of Canmore, Alberta experienced a devastating flood in the year 2013, motivating the need to forecast the "worst case scenario" river flow in the near future on a day-to-day basis. To do this, data on rainfall, snowpack, and river discharge are available at various stations since the 1980s. Section 2.1 elaborates on this motivation and the data.

A forecast is a communication of one's belief of a future outcome. This belief is represented as a probability (or density) for each possible outcome, and forms the predictive distribution. When there are observations on predictors, the predictive distribution should be as close as possible to the conditional distribution of the response given the predictors. An extreme forecast can be formed by communicating the tail of this distribution in the form of an upper quantile function. These concepts are formally stated in Section 2.2.

To build a forecaster of extremes, the methodology proposed in this dissertation makes use of concepts in Extreme Value Theory and copulas. The requisite background is provided in Sections 2.3–2.5.

2.1 Data and Motivating Application

The interest is in forecasting the *discharge* of the Bow River at Banff[1], one day ahead. The discharge of a river can be defined as the volume of water that flows past a cross-section of the river over a time period, measured in m^3/s. Discharge data beginning in the year 1980 are available, and are displayed in Figure 2.1. For time periods where data are available, see Figure 2.2.

[1]Discharge data are from the Water Survey of Canada of the Ministry, a branch of Environment and Climate Change Canada, https://www.ec.gc.ca/rhc-wsc/.

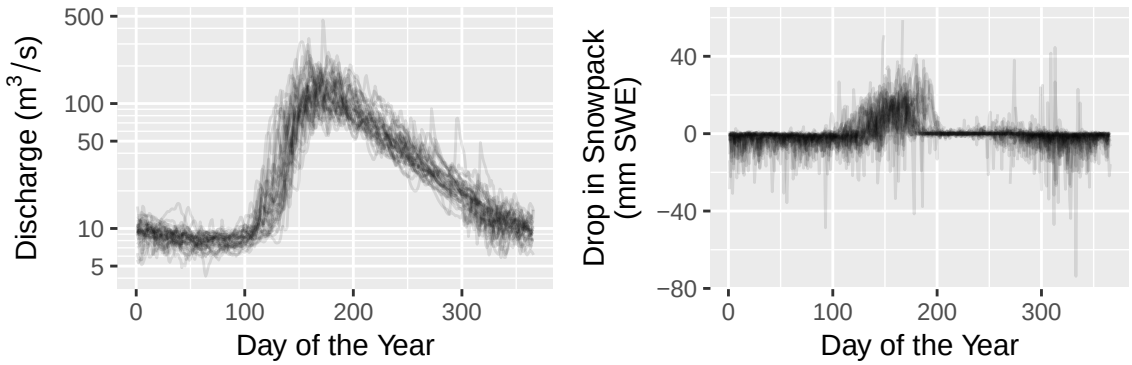

Figure 2.1: Time series plots of the discharge data (left) and drop in snowpack data (right), with each year overlaid. Note the log-scale on the discharge time series, and that "SWE" is snow water equivalent.

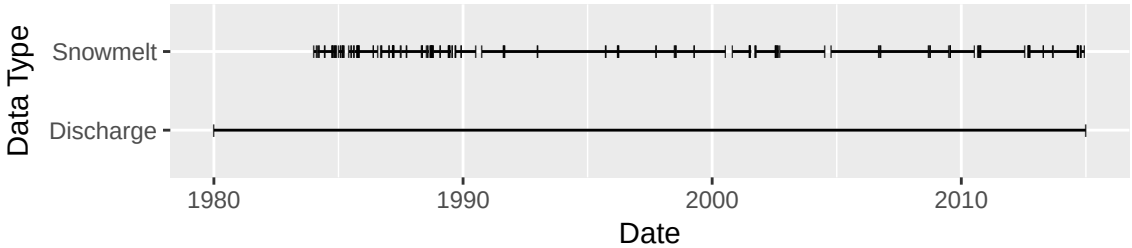

Figure 2.2: Intervals of time where data are available. Each interval represents an unbroken record, and is displayed as a solid horizontal line between two short vertical lines. A continuous record of discharge data is available.

To forecast discharge using information other than lagged ("recent") discharge and day of the year, variables that influence surface runoff are typically sought (cf. Beven, 2012). These include snowmelt, rainfall, soil infiltration, and evapotranspiration, and are typically computed from other measurements such as snowpack and temperature data. For simplicity of modelling, only snowmelt is used of these surface runoff variables – simple hydrologic models have been shown to compete well with complex hydrologic models anyway (Micovic, 2005).

Omitting rainfall as a predictor might seem surprising. Newspaper articles often report flash flooding of urban creeks due to heavy rainfall, such as in North Vancouver in British Columbia (Li, 2016). Perhaps rainfall has a heavy influence on urban creeks due to a higher abundance of impermeable surfaces, such as concrete, in the watershed. Further, the media may be more likely to report such floods by the very nature of them occurring in urban areas. As for rivers in the Rocky Mountains (including the Bow River at Banff), there is evidence that snowmelt (not rainfall) is the major driver of surface runoff, whereas rainfall becomes the predominate driver in the prairies (Environment and Climate Change Canada, 2009). Empirical evidence of available data confirm that

discharge is more dependent on snowmelt than rainfall.

The snowpack data[2] are available at various stations, shown in Table 2.1. Snow pillows are used to measure the "Snow Water Equivalent" (SWE) of the snowpack, which is the depth of water that would be obtained if the snowpack was melted (in millimeters). As a proxy for snowmelt, the drop in snowpack data are averaged across the stations for all data available on a given day. These data are displayed in Figure 2.1. More advanced methods exist to estimate quantities such as total basin snowmelt, but these are not considered in this dissertation.

Station Name	Station ID	Measurement Type
Bow River at Banff	05BB001	Discharge (m^3/s)
Three Isle Lake	05BF824	Snow Water Equivalent (mm)
Little Elbow Summit	05BJ805	
Mount Odlum	05BL812	
Skoki Lodge	05CA805	
Sunshine Village	05BB803	

Table 2.1: Information about the stations where data are collected in Alberta.

2.2 The Forecasting Task

Let $Y : \Omega \to \mathscr{Y} \subset \mathbb{R}$ be a random variable on the probability space $(\Omega, \mathscr{F}, \mathcal{P})$, where $\mathscr{F}$ is a sigma field of subsets of a sample space Ω, and $\mathcal{P}$ is a probability measure[3]. The marginal distribution function satisfies

$$F_Y(y) = \mathcal{P}\left(\left\{\omega \in \Omega : Y(\omega) \leq y\right\}\right)$$

for all $y \in \mathscr{Y}$. A *forecast* is an upper quantile function $\hat{Q} : (\tau_c, 1) \to \mathbb{R}$ for non-decreasing $\hat{Q}$, where $\tau_c \in (0, 1)$ is a quantile level cutoff, which is chosen by the analyst to suit the needs of the forecast users. Surely, without any information, the best forecast that can be issued is the true distribution $F_Y^{\leftarrow}$ over $(\tau_c, 1)$.

Sometimes, partial information about the response is available. This partial information takes the form of knowing whether the outcome $\omega \in G$ for all $G \in \mathscr{G}$, where $\mathscr{G}$ is a sigma field contained in $\mathscr{F}$ that can be interpreted as an "observation". In this case, the

[2]The snowpack data were provided to the authors by the Alberta Ministry of Environment and Parks, http://aep.alberta.ca/.

[3]Throughout this dissertation, $\mathcal{P}$ is used in general to represent "probability" of any event.

conditional distribution of Y given knowledge of $\mathscr{G}$ should be the forecast, because this distribution by setup is the most one can know under this information.

Of course, such conditional distribution is unknown, and requires estimation. Further, this estimation can only practically be done after selecting only the most useful information. Thus, it is supposed that this information takes the form of observations on p predictors $\boldsymbol{X} \in \mathscr{X} \subset \mathbb{R}^p$. Under the information that $\boldsymbol{X} = \boldsymbol{x}$ is observed, the true quantile function $Q_{Y|\boldsymbol{X}=\boldsymbol{x}}$ over $(\tau_c, 1)$ is the best forecast. The task of issuing a forecast therefore becomes one of estimating the tail of the conditional distribution of Y given the predictors.

Since forecasts depend on observed information, a forecast is stochastic. A *forecaster* is a stochastic draw from a set $\hat{\mathcal{Q}}$ of forecasts[4]. The corresponding set of upper distribution functions is $\hat{\mathcal{F}} = \left\{\hat{Q}^{\leftarrow} : \hat{Q} \in \hat{\mathcal{Q}}\right\}$. In this vein, the set of all (true) conditional distribution functions of Y given the results of any such observation $\mathscr{G} \subset \mathscr{F}$ is denoted $\mathcal{F}$, and its quantile functions are $\mathcal{Q} = \{F^{\leftarrow} : F \in \mathcal{F}\}$.

The response occurring at the (discrete) time t is Y_t. Gneiting et al. (2007) speak of a sequence of distributions from which nature generates the respective responses $Y_1, Y_2, \ldots$. However, the concept of a "generating distribution" depends highly on philosophy – for instance, it could be argued that nature only draws an outcome from a degenerate distribution. Instead, we think of the responses $Y_1, Y_2, \ldots$ as only being stochastic relative to an observer, with sequences of distributions belonging to $\mathcal{F}$, depending on what information is observed.

2.3 Extreme Value Index

This section briefly discusses the tail behaviour of a univariate distribution through a concept called the extreme value index (EVI), and why such behaviour is interesting to know in practice. For a more extensive review, see de Haan and Ferreira (2006).

We can begin our exploration of tail behaviour by investigating the asymptotic distribution of the sample maximum. Let $Y_1, \ldots, Y_n$ be an independent random sample drawn from the distribution function F_Y with support $\mathbb{R}$, and denote $Y_{n:n} = \max(Y_1, \ldots, Y_n)$ as the sample maximum. Of course, $Y_{n:n}$ approaches the right-endpoint of F_Y as $n \to \infty$ and its distribution thus becomes degenerate. But if there exist real sequences $\{a_n > 0\}$ and $\{b_n\}$ such that the distribution of $(Y_{n:n} - b_n)/a_n$ as $n \to \infty$ is *not* degenerate, what might that distribution look like? That is, if the following assumption holds, can something be

[4]This set is not necessarily the set of all upper quantile functions. We will see in Chapter 3.2 that information about such set is required for assessing goodness.

said about G?

Assumption 2.3.1. For $y \in \mathbb{R}$,

$$\lim_{n\to\infty} F_Y^n\left(a_n y + b_n\right) = G\left(y\right), \tag{2.3.1}$$

where G is not degenerate.

Proposition 2.3.1. *(Fisher and Tippett, 1928; Gnedenko, 1943) Suppose Assumption 2.3.1 holds for G non-degenerate. Then G belongs to the class of Generalized Extreme Value (GEV) distributions $G_\xi\left(\left(x-\mu\right)/\sigma\right)$ for location parameter $\mu \in \mathbb{R}$, scale parameter $\sigma > 0$, and shape parameter $\xi \in \mathbb{R}$, where*

$$G_\xi\left(z\right) = \exp\left(-\left(1+\xi z\right)^{-1/\xi}\right) \tag{2.3.2}$$

for $1 + \xi z > 0$.

See de Haan and Ferreira (2006, Theorem 1.1.6) for a proof. We will see later that larger values of ξ correspond to distributions F_Y with heavier tails.

In addition, other valid sequences of $\{a_n > 0\}$ and $\{b_n\}$ satisfying the convergence property in Equation (2.3.1) do not change the shape parameter ξ of the resulting GEV distribution – only the location and scale parameters might change. We therefore speak of convergence to a GEV distribution *of a certain* $\xi \in \mathbb{R}$. All such distribution functions F_Y satisfying the convergence property in Equation (2.3.1) leading to a GEV distribution with some common shape parameter $\xi \in \mathbb{R}$ are said to belong to the *(maximum) domain of attraction* of G_ξ, which is denoted $\mathcal{D}\left(G_\xi\right)$. Most continuous distributions that are dealt with in practice belong to a domain of attraction.

Proposition 2.3.1 is particularly important for the identification of the shape parameter defining the limiting GEV distribution of an underlying distribution F_Y. When thought of as a property of a distribution F_Y, the shape parameter of the GEV distribution is called the *extreme value index* (EVI) of F_Y, and is denoted ξ_Y (with the random variable in the subscript). The EVI is an important quantity because it describes the decay rate of the underlying distribution's tail. In particular, the tail can be heavy ($\xi_Y > 0$), decaying like a power function; light ($\xi_Y = 0$), decaying rapidly or exponentially; or short ($\xi_Y < 0$), having a finite right-endpoint.

Proposition 2.3.2 (Tail behaviour)**.** *Let $y^* = \sup\left\{y \in \mathbb{R} : F_Y\left(y\right) < 1\right\}$ be the right-endpoint of the support of a distribution function F_Y, which is possibly infinite. For some $\xi_Y \in \mathbb{R}$, $F_Y \in \mathcal{D}\left(G_{\xi_Y}\right)$ if and only if*

1. *for* $\xi_Y > 0$, $y^* = \infty$ *and*

$$\lim_{t\to\infty} \frac{1 - F_Y(ts)}{1 - F_Y(t)} = s^{-1/\xi_Y} \tag{2.3.3}$$

for all $s > 0$*;*

2. *for* $\xi_Y = 0$, $y^* \leq \infty$ *and there exists a positive function* g *such that*

$$\lim_{t\uparrow y^*} \frac{1 - F_Y\left(t + sg(t)\right)}{1 - F_Y(t)} = \exp(-s) \tag{2.3.4}$$

for all $s \in \mathbb{R}$*; and*

3. *for* $\xi_Y < 0$, $y^* < \infty$ *and*

$$\lim_{t\downarrow 0} \frac{1 - F_Y(y^* - st)}{1 - F_Y(y^* - t)} = s^{-1/\xi_Y} \tag{2.3.5}$$

for all $s > 0$.

See de Haan and Ferreira (2006, Theorem 1.2.1) for a proof.

Since the EVI determines the tail behaviour of the underlying distribution F_Y, the EVI should be an important consideration when modelling a distribution to evaluate the exceedance probability of extremes. This is especially true when F_Y is heavy-tailed: due to the slow decay of a heavy-tailed distribution, a realization that is much larger than "typical" realizations is not an unusual occurrence. The extent of such extreme values is more likely to be large if ξ_Y is large. Such "rogue values" are sometimes seen in practice, such as with natural disasters or insurance claims. If the EVI is not considered when modelling the tail behaviour of such phenomena, then the occurrence frequency of extreme events (and thus the risk of danger) could be seriously underestimated.

Proposition 2.3.2 tells us one way to find the EVI of a distribution, supposing one knows its sign. The von Mises condition is another useful way to find the EVI, although it is sufficient, and not necessary.

Proposition 2.3.3 (Tail behaviour)**.** *Let* $y^* = \sup\left\{y \in \mathbb{R} : F(y) < 1\right\}$ *be the right-endpoint of a distribution function* F*, which is possibly infinite. Suppose that, in some left neighbourhood of* y^**,* F'' *exists and* $F' > 0$*. If*

$$\lim_{y\uparrow y^*} \mathrm{D}\left(\frac{1 - F}{F'}\right)(y) = \xi \in \mathbb{R}, \tag{2.3.6}$$

or equivalently,

$$\lim_{y \uparrow y^*} \frac{\left(1 - F(y)\right) F''(y)}{\left[F'(y)\right]^2} = -\xi - 1 \in \mathbb{R}, \tag{2.3.7}$$

then $F \in \mathcal{D}\left(G_\xi\right)$.

See de Haan and Ferreira (2006, Theorem 1.1.8) for a proof.

We do not discuss estimation of a distribution's tail in this dissertation, or how to model the tail of a univariate distribution, but rather focus on modelling the conditional[5] EVI in a regression setting.

2.4 Copulas

The proposed methodology (along with some existing methods) makes use of *copulas* for modelling. An overview of modelling joint distributions with copulas is discussed here. For more details, see Joe (2014).

2.4.1 Copula Basics

Essentially, a function $C : [0,1]^d \to [0,1]$ for integer $d \geq 2$ is a *copula* if it is a distribution function with Uniform $(0,1)$ univariate marginals. The importance of copulas is made clear by Sklar's Theorem.

Theorem 2.4.1 (Sklar's theorem)**.** *A distribution function* $F_{1:d} : \mathbb{R}^d \to [0,1]$ *having univariate marginals* $F_1, \ldots, F_d$*,* $d \geq 2$*, can be written as*

$$F_{1:d}(x_1, \ldots, x_d) = C\left(F_1(x_1), \ldots, F_d(x_d)\right),$$

where C *is a copula. Further,* C *is unique if* $F_1, \ldots, F_d$ *are continuous. The converse is also true: given a copula* C *and univariate distribution functions* $F_1, \ldots, F_d$*,* $C\left(F_1(x_1), \ldots, F_d(x_d)\right)$ *for* $(x_1, \ldots, x_d) \in \mathbb{R}^d$ *is a distribution function having univariate marginals* $F_1, \ldots, F_d$*.*

Sklar's Theorem implies that the dependence structure within a multivariate (continuous) distribution is determined by the underlying copula. Consequently, multivariate distributions can be modelled using copulas together with the well-known methods of modelling univariate distributions. The theorem also describes a way to find the copula

[5] In this dissertation, "conditional EVI" means the EVI of the response given the predictors.

C underlying a joint distribution $F_{1:d}$:

$$C : (u_1, \ldots, u_d) \mapsto F_{1:d}\left(F_1^{\leftarrow}(u_1), \ldots, F_d^{\leftarrow}(u_d)\right) \tag{2.4.1}$$

for left-continuous inverse functions $F_1^{\leftarrow}, \ldots, F_d^{\leftarrow}$.

It is useful to consider different reflections of copulas when modelling, which we define here.

Definition 2.4.1 (Copula reflections)**.** *Suppose C is a copula, representing the distribution function of $\boldsymbol{U} = (U_1, \ldots, U_d)$ for integer $d \geq 2$. Suppose K is an integer, $1 \leq K \leq d$, and consider a subset $\{i_1, \ldots, i_K\} \subset \{1, \ldots, d\}$, $i_1 < \cdots < i_K$. We make the following definitions.*

- *The* $(i_1, \ldots, i_K)$-reflected copula *is the distribution function of $\boldsymbol{U}$ with the i_j'th entry replaced by $1 - U_{i_j}$ for each $j = 1, \ldots, K$. Such a copula is denoted by $R_{i_1 \cdots i_K} C$, where $R_{i_1 \cdots i_K}$ is a* reflection operator *defined in Equation (A.1.1).*

- *The $(1, \ldots, d)$-reflected copula is called the* reflected *or* reflection *copula.*

- *For any reflection operator R, the copula RC is in general referred to as* a reflection *copula.*

- *For $d = 2$, the 1-reflected copula is also called the* horizontally-reflected copula*, and the 2-reflected copula is also called the* vertically-reflected copula*.*

Formulas for computing a reflection copula are given in Appendix A.1.

We reserve the letter C to indicate a copula distribution function, and its lower case $c = \mathrm{D}_{12}\, C$ to indicate its density. For bivariate copulas, the conditional distributions are denoted $C_{2|1} = \mathrm{D}_1\, C$ and $C_{1|2} = \mathrm{D}_2\, C$. A "hat" on these quantities refers to the reflection copula – for example, $\hat{C}_{2|1} = \mathrm{D}_1\, \mathrm{R}_{12} C$.

There are many parametric copula families that have been identified in the literature (cf. Joe, 2014, Chapter 4), as well as many different interesting properties that are characteristic of these copulas. Some of these properties are asymmetry, quadrant dependence, and tail dependence. Tail dependence describes the dependence between variables, given that each is large. For a detailed overview of properties, see Joe (2014, Chapter 2).

Since a copula determines the dependence structure between multiple random variables, we would expect it to describe much of the behaviour of the conditional EVI. This notion is formalized in the CCEVI, which is the conditional EVI of a copula variable that is transformed to have a standard Type I Pareto margin. For random variable $U \sim \mathrm{Unif}\,(0, 1)$, $(1 - U)^{-1}$ is the required transformation.

Definition 2.4.2 (CCEVI)**.** *Suppose* $\boldsymbol{U} = (U_1, \ldots, U_d)^\top$ *is a d-dimensional random vector,* $d \geq 2$*, with joint distribution given by some copula* C*. Take* $i \in \{1, \ldots, d\}$*, and let* $\boldsymbol{U}_{-i}$ *be the vector* $\boldsymbol{U}$ *with the i'th entry removed. Then, if there exists a function* $\xi_i : [0,1]^{d-1} \to \mathbb{R}$ *such that*

$$F_{(1-U_i)^{-1}|\boldsymbol{U}_{-i}=\boldsymbol{u}_{-i}} \in \mathcal{D}\left(G_{\xi_i(\boldsymbol{u}_{-i})}\right)$$

for all $\boldsymbol{u}_{-i} \in [0,1]^{d-1}$*, we call* ξ_i *the* (upper) i'th Copula Conditional Extreme Value Index *(CCEVI). We refer to* the *CCEVI of a d-dimensional copula* C *as its upper d'th CCEVI, and denote this CCEVI by* ξ_C*.*

The lower i'th CCEVI can be defined analogously, but for the lower tail; that is, the EVI of $U_i^{-1} \mid \boldsymbol{U}_{-i} = \boldsymbol{u}_{-i}$. The i'th CCEVI could have been defined as the EVI of $U_i \mid \boldsymbol{U}_{-i} = \boldsymbol{u}_{-i}$, but since this random variable has a finite right-endpoint (at 1), its EVI is nonpositive and therefore uninteresting.

Examples of CCEVI's for some bivariate copula families are listed in Table 2.2.

	Parameter	CCEVI	CCEVI of Reflected
Gumbel	$\theta > 1$	$1/\theta$	1
Frank	$\theta \in \mathbb{R}$	1	1
Bivariate Gaussian	$-1 < \rho < 1$	$1 - \rho^2$	$1 - \rho^2$

Table 2.2: CCEVI's of some existing bivariate copula models, all of which are constant. See Appendix B.2 for proofs. Note that the bivariate Gaussian copula has the same CCEVI for ρ and $-\rho$, the latter being a 1-reflection.

The CCEVI is a real function with domain $[0,1]^{d-1}$. If this function is constant throughout its domain, we sometimes refer to the CCEVI as that constant with the implicit meaning of a constant function.

To find the CCEVI of a bivariate copula C, first fix a $u \in (0,1)$. One can then use the von Mises condition in Proposition 2.3.3 with $F = C_{2|1}\left(\cdot|u\right)$. Or, if one suspects the CCEVI is positive, the CCEVI can be obtained by the index of variation of the distribution function $C_{2|1}\left(1 - y^{-1}|u\right)$ for $y > 1$ through Equation (2.3.3). In particular,

$$\lim_{t\to\infty} \frac{1 - C_{2|1}\left(1 - (ts)^{-1}|u\right)}{1 - C_{2|1}\left(1 - t^{-1}|u\right)} = s^{-1/\xi_C(u)}, \tag{2.4.2}$$

or equivalently, using the copula density,

$$\lim_{t\to\infty} \frac{c\left(u, 1-(ts)^{-1}\right)}{c\left(u, 1-t^{-1}\right)} = s^{1-1/\xi_C(u)} \tag{2.4.3}$$

(which can be derived using L'hôpital's rule). In fact, this suggests that a copula having a finite and non-zero density at a boundary has a CCEVI of 1.

Proposition 2.4.2. *Suppose C is a bivariate copula with*

$$\lim_{v\uparrow 1} c\left(u, v\right) = c^* \in (0, \infty) \tag{2.4.4}$$

for all $u \in \mathscr{U} \subset (0, 1)$. Then the CCEVI of C over $\mathscr{U}$ is 1.

A straight-forward proof follows by Equation (2.4.3).

2.4.2 Pair-Copula Bayesian Networks

Modelling a bivariate distribution using copulas is straight-forward relative to modelling higher dimensional distributions. Once the univariate distributions are modelled, then one of the many parametric bivariate copula families listed in the literature (cf. Joe, 2014, Chapter 4) can be chosen to fit the dependence. Such dependence can be visualized in the data with a normal scores plot – a scatterplot of the data with marginals transformed to standard Gaussian. These plots are useful because it allows the dependence to be compared to a bivariate Gaussian distribution, and may reveal features such as tail dependence that one would be interested in capturing in a model.

However, modelling a multivariate distribution with more than two variables is more difficult. Although there are multivariate copula families identified in the literature that can be used, they are often too simple to capture the intricate dependence features demonstrated by real data. In addition, the dependence amongst the variables cannot be visualized in a straight-forward way like it can with a bivariate normal scores plot.

To deal with the issue of modelling the joint distribution of some p-dimensional random vector $\boldsymbol{W}$, we can consider introducing one variable at a time. As a preliminary, the univariate marginal distributions of each variable in $\boldsymbol{W}$ should be modelled in full. Here are the steps involved in building a p-variate copula model to describe the dependence in $\boldsymbol{W}$, and obtaining its joint distribution.

1. Choose two variables to start with – say W_1 and W_2 (letting the subscripts indicate the *introduction order*). Select a bivariate copula model to describe their dependence.

2. Introduce another variable, W_3, and "pair" it with each of the previous variables W_1 and W_2. However, copulas cannot be chosen freely for each pair, otherwise the joint distribution of (W_1, W_2, W_3) might not be valid. Instead, the pairs are chosen according to some chosen order (which we call the *pairing order* for W_3), and a copula model is chosen for each pair *conditional on* previously paired variables. In this case, we can choose copula models for (W_1, W_3) and $(W_2, W_3) \mid W_1$, or (W_2, W_3) and $(W_1, W_3) \mid W_2$. The copula models for each pair are assumed not to depend on the conditioning ("given") variables – an assumption known as the *simplifying assumption*.

3. Step 2 is repeated, each time introducing a new variable and pairing it with the previously introduced variables, until all variables are introduced. A model for the resulting copula underlying $\boldsymbol{W}$ can then be constructed, along with the joint distribution that is called a *pair-copula Bayesian network* (PCBN) (Bauer and Czado, 2015).

Here are some additional modifications that are useful.

- To prevent overfitting, a PCBN model can be simplified by choosing the independence copula to describe the dependence in pairs at some point in a pairing order. This is called *truncation*, and carries the interpretation that variability is already accounted for by the conditioning variables.

- To avoid an otherwise computationally expensive evaluation of the PCBN joint distribution (which can involve multi-dimensional integration), some restrictions on the pairing order can be applied. The details of these restrictions are not discussed here (see Joe, 2014, Chapter 3), but are only needed for variables introduced after the third (i.e. $W_4, \ldots, W_p$). Under the restrictions, the PCBN is called a *regular vine* and the resulting distribution of $\boldsymbol{W}$ is called a *vine distribution*, and its copula a *vine copula*.

For equations related to the PCBN and vine distributions, see Appendix A.2.

2.4.3 PCBN Representations

A PCBN is identified by its variable pairs and the copulas for each pair. Writing the variable pairs is quite verbose, and can be summarized by an array called the *PCBN*

array (or *vine array*). An example of such an array with five variables is

$$M = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ & 1 & 1 & 2 & 1 \\ & & 2 & & 4 \\ & & & & 3 \\ & & & & 2 \end{pmatrix}. \tag{2.4.5}$$

The introduction order of the variables is indicated by the first row, and each variables' pairing order is listed in order below it. Specifically, denoting M_{ij} as the entry in row i and column j, then for $i > 1$ and $j \geq i$, the pairs in the PCBN are $\left(W_{M_{ij}}, W_{M_{1j}}\right) \mid \left(W_{M_{2j}}, \ldots, W_{M_{(i-1)j}}\right)$ whenever M_{ij} is not blank (a blank entry indicates a truncation). This array has pairs (W_1, W_2), (W_1, W_3), $(W_2, W_3) \mid W_1$, (W_2, W_4), (W_1, W_5), $(W_4, W_5) \mid W_1$, $(W_3, W_5) \mid (W_1, W_4)$, $(W_2, W_5) \mid (W_1, W_3, W_4)$, and happens to be a vine. Different ways of writing vine arrays exist in the literature (with variables along the diagonal), but the advantage of this representation is that it makes truncation easy to express.

The pairs can also be represented by a directed acyclic graph with labelled edges[6]. Upon introducing the kth variable W_k, arrows are drawn from previous variables towards W_k with edges labelled according to their pairing order. Pairs that are (conditionally) independent can omit an arrow. See Figure 2.3 for a graphical representation of a PCBN and a vine. To obtain the pairs from a graphical representation, it is easiest to follow the graph "backwards" to get the reverse introduction order. Begin with a variable not at any arrow tail, and obtain its pairing order according to the labelled edges. Then remove that node and its edges, and repeat until no variables remain.

2.5 Conditional EVI

To extrapolate into the tail of a univariate distribution (i.e., extract estimates of higher and higher quantiles), Extreme Value Theory provides justification for a model that decays according to the EVI. If the EVI is estimated to be too small, then extremes will occur more often than anticipated (and vice-versa).

When using predictors to make forecasts, it is equally important to estimate the EVI of the conditional distribution of the response given the predictors. There are examples in nature where this *conditional EVI* is non-constant across the predictor space.

[6]Note that vines can also be represented using a different graphical representation due to their restrictions on the pairing order.

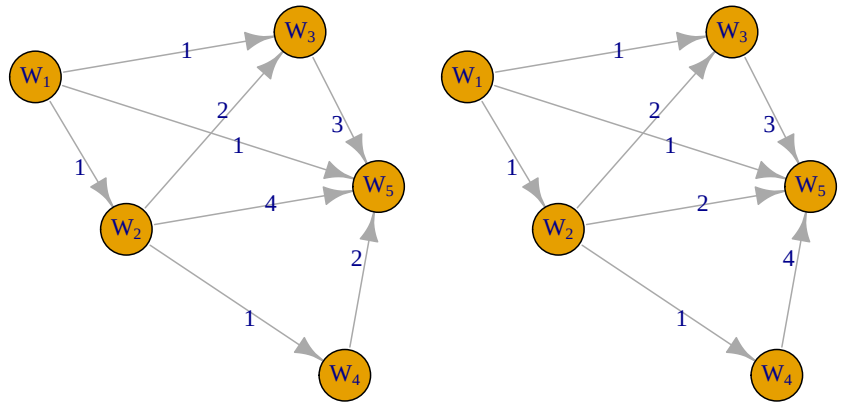

Figure 2.3: Graphical representation of a pair-copula Bayesian network (left) with array in (2.4.5), and a regular vine (right). The vine array on the right is the same as in (2.4.5), but with the last column replaced by $(5, 1, 2, 3, 4)$.

Section 2.5.1 demonstrates this with pollution data. Without capturing a non-constant conditional EVI in a model, extrapolation into a forecast's tail would be less reliable.

The link between conditional quantiles and the conditional EVI is explored briefly in Section 2.5.2, and shows that a conditional EVI that is non-constant results in quantile surfaces that are highly non-linear. Section 2.5.3 discusses how copulas influence the conditional EVI through a new concept called the *copula conditional EVI* (CCEVI), and this is useful for incorporating a non-constant (in the predictors) conditional EVI in a model.

2.5.1 Motivation

The dependence amongst atmospheric pollutants is an important consideration when making environmental impact assessments (Heffernan and Tawn, 2004). Examination of the dependence amongst the concentration of some pollutants reveals that a non-constant conditional EVI is tangible in nature.

As an example, data of daily maximum pollutant concentrations of five pollutants are available at ground level in Leeds city center, UK, between the years 1994 and 1998 (inclusive) during the months of April to July, and can be found in Heffernan and Tawn (2004). Two pollutants—ozone (O_3) and sulphur dioxide (SO_2)—demonstrate the need for a model that allows for a non-constant conditional EVI. These data are shown in Figure 2.4. The conditional EVI of SO_2 concentrations given O_3 concentrations can be estimated by using a moving-window approach. Estimates can be found in Figure 2.5, which includes a sensitivity analysis to ensure that the results are not an artifact of the estimation parameters. It appears that there is indeed a negative trend in the conditional

EVI.

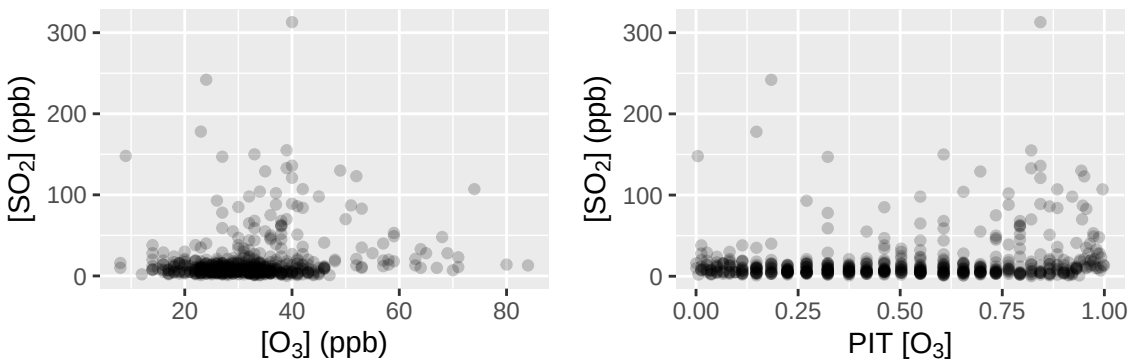

Figure 2.4: Relationship between atmospheric sulphur dioxide concentrations ([SO_2]) and atmospheric ozone concentrations ([O_3]) (left) or probability integral transformed (PIT) [O_3] (right). Measurements are in parts per billion (ppb), except for PIT scores, which are unit-less. Data are daily maximum concentrations collected at ground level in Leeds city center, UK, between the years 1994 and 1998 (inclusive) during the months of April to July.

Note that the direction of trend in the conditional EVI need not match the sign of the dependence, as these data seem to suggest. Although there is evidence that the conditional EVI is decreasing, there is evidence[7] that the overall dependence is decreasing. Indeed, with a decreasing conditional EVI, the upper quantile curves must be decreasing, as identified in Proposition 2.5.2; but this does not force the central quantile curves (or the conditional expectation curve) to be decreasing.

Other pairs of pollutants show evidence of a non-constant conditional EVI as well, but this example is perhaps the most convincing.

2.5.2 Conditional Quantiles

Suppose $\boldsymbol{X} \in \mathscr{X} \subset \mathbb{R}^p$ is a random vector of $p \geq 1$ predictors, and $Y \in \mathbb{R}$ is a response random variable. We make the following assumptions.

Assumption 2.5.1. The response Y has distribution function $F_Y \in \mathcal{D}\left(G_{\xi_Y}\right)$ for some $\xi_Y \in \mathbb{R}$, with right-endpoint $y^* = \sup\left\{y \in \mathbb{R} : F_Y(y) < 1\right\}$.

Assumption 2.5.2. There exists a measurable function $\xi_{Y|\boldsymbol{X}} : \mathscr{X} \to \mathbb{R}$ such that, for each $\boldsymbol{x} \in \mathscr{X}$, the conditional response $Y|\,\boldsymbol{X}{=}\boldsymbol{x}$ has distribution function $F_{Y|\boldsymbol{X}=\boldsymbol{x}} \in \mathcal{D}\left(G_{\xi_{Y|\boldsymbol{X}}(\boldsymbol{x})}\right)$ with right-endpoint

$$y^*_{Y|\boldsymbol{X}}(\boldsymbol{x}) = \sup\left\{y \in \mathbb{R} : F_{Y|\boldsymbol{X}}\left(y|\boldsymbol{x}\right) < 1\right\}. \tag{2.5.1}$$

[7]For the raw data, an estimate of Pearson's correlation is 0.16, and a test for significance of simple linear regression results in a p-value less than 0.001.

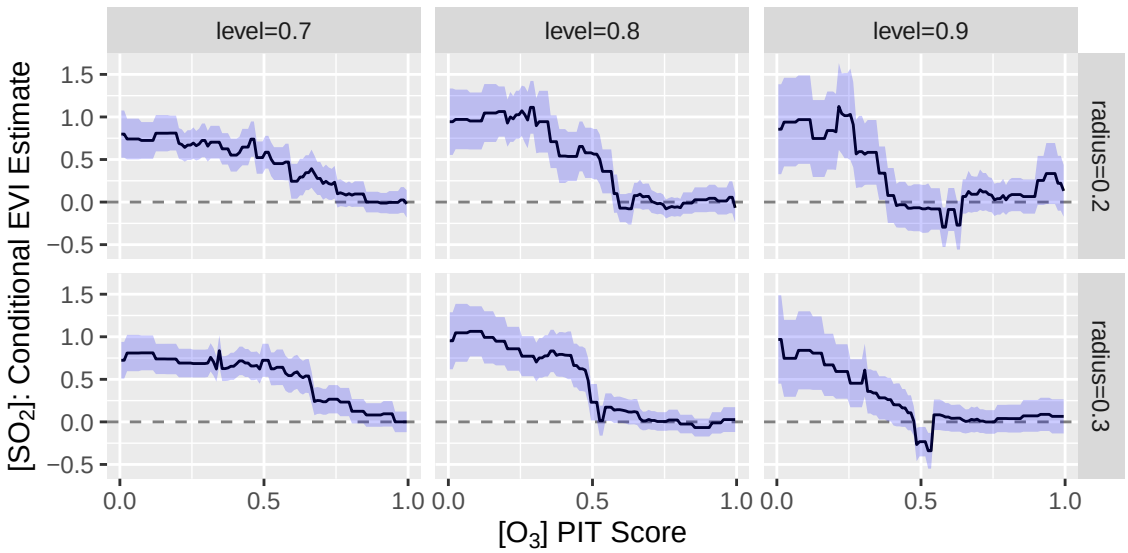

Figure 2.5: Local estimates of the conditional EVI of atmospheric sulphur dioxide concentrations ($[SO_2]$), conditional on PIT atmospheric ozone concentrations ($[O_3]$). To estimate the conditional EVI at a particular $[O_3]$ PIT score, a window with some radius (indicated in the panels on the right) is constructed, and the corresponding subsample of $[SO_2]$ data is extracted. A Generalized Pareto distribution (GPD) is fit to this univariate subsample using MLE, with the threshold parameter taken to be quantile estimates with levels indicated in the upper panels. Error bands represent one standard error of the EVI estimates. Despite different estimation parameters, there always appears to be a downward trend in the conditional EVI.

We begin by demonstrating that conditioning Y on $\boldsymbol{X}$ cannot result in a conditional EVI that is larger than the marginal EVI.

Proposition 2.5.1. *Under Assumptions 2.5.1 and 2.5.2 with $\xi_{Y|\boldsymbol{X}}$ continuous, if $\xi_Y \geq 0$, then $\xi_{Y|\boldsymbol{X}}(\boldsymbol{x}) \leq \xi_Y$ for almost all $\boldsymbol{x} \in \mathscr{X}$.*

See Appendix B.1 for a proof. This result indicates a "mixing" property that the marginal tail can become heavier than the conditional tails.

Note that, even if the conditional EVI is constant over the covariate space, it need not equal the marginal extreme value index (examples are shown in Table 2.2). This proposition tells us that knowledge of predictors cannot introduce "extra" tail heaviness into the response. The proposition also introduces the notion that some predictors might be more "informative" than others in terms of their ability to describe the extreme behaviour of the response: predictors $\boldsymbol{X}$ that induce a larger reduction of ξ_Y to $\xi_{Y|\boldsymbol{X}}$ are better in this light. This is because such predictors are able to describe some of the tail heaviness of the response distribution.

We turn our attention to the relationship between the conditional extreme value index and upper τ-*quantile surfaces* $Q_{Y|\boldsymbol{X}}(\tau|\cdot)$ (or quantile *curves* if the function is one-

dimensional), where $Q_{Y|\boldsymbol{X}}\left(\tau|\boldsymbol{x}\right) = F^{\leftarrow}_{Y|\boldsymbol{X}}\left(\tau|\boldsymbol{x}\right)$ for each $\tau \in (0,1)$ and $\boldsymbol{x} \in \mathscr{X}$. First, the conditional EVI over a path in the predictor space is non-decreasing if the quantile curves are non-decreasing over the path.

Proposition 2.5.2. *Suppose Assumption 2.5.2 holds with $\xi_{Y|\boldsymbol{X}} \geq 0$, and let $\tilde{\boldsymbol{x}} : (0,1) \to \mathscr{X}$ be a path in the predictor space from $\tilde{\boldsymbol{x}}(0)$ to $\tilde{\boldsymbol{x}}(1)$. If there exists an $\varepsilon > 0$ such that the τ-quantile curves $Q_{Y|\boldsymbol{X}}\left(\tau \mid \tilde{\boldsymbol{x}}(\cdot)\right)$ along the path are non-decreasing for each $\tau \in (1-\varepsilon, 1)$, then $\xi_{Y|\boldsymbol{X}}\left(\tilde{\boldsymbol{x}}(\cdot)\right)$ is non-decreasing.*

See Appendix B.1 for a proof.

One might think of the conditional EVI in relation to the marginal EVI as measuring "informativeness" of the predictors about the behaviour of extremes. When the conditional EVI and marginal EVI are close, then observing that predictor did not "lighten" the tail of the response much, and that observation of the predictors is therefore not very informative about the extremes of the response. Likewise, a conditional EVI that is much lower than the marginal EVI "lightens" the tail of the response, and is therefore informative.

Proposition 2.5.2 suggests that, when upper quantile curves are increasing, larger values of the predictors cannot become more informative regarding the extreme behaviour of the response. If a predictor in fact becomes less informative of the tail of Y, so that the conditional EVI increases as the predictor increases, it is important to capture this in a model so that the tail behaviour of the response is not underestimated when the predictor is observed to be large. Capturing such a trend requires a highly non-linear model of the quantile surfaces, as the following proposition suggests.

Proposition 2.5.3. *Suppose Assumption 2.5.2 holds with $\xi_{Y|\boldsymbol{X}} > 0$. If there exists an $\varepsilon > 0$ and a one-to-one mapping $\boldsymbol{T} : \mathscr{X} \to \mathbb{R}^p$ of the predictor space such that*

$$Q_{Y|\boldsymbol{X}}\left(\tau|\boldsymbol{T}^{\leftarrow}(\boldsymbol{w})\right) = \alpha + \boldsymbol{w}^{\top}\boldsymbol{\beta}(\tau)$$

for $\boldsymbol{w} \in \mathbb{R}^p$ and each $\tau \in (1-\varepsilon, 1)$, where $\boldsymbol{\beta} : (1-\varepsilon, 1) \to \mathbb{R}^p$, then $\xi_{Y|\boldsymbol{X}}$ is constant.

See Appendix B.1 for a proof.

This proposition suggests that, in order for a model to accommodate a varying conditional EVI, its quantile surfaces cannot be linear after any transformation of the predictor space. This proposition extends the result of Wang and Li (2013), which indicates that modelling linear quantile curves in one predictor (such as with Koenker and Basset's (1978) linear quantile regression method) is not appropriate if $\xi_{Y|\boldsymbol{X}}$ is non-constant. This makes intuitive sense, since the spacing between any two conditional quantiles is just a

multiple of the same conditional quantiles under a different observed predictor, and this cannot change the decay rate of the quantile function.

As an attempt to be able to use the linear model and still have the potential for a non-constant conditional EVI, Wang and Li (2013) propose a transformation method (in the single-predictor $\boldsymbol{X} = X_1$ setting). However, their method does not actually achieve their desired result. Let $\Xi_\lambda : [0, \infty) \to \mathbb{R}$ be the Box-Cox power transform given by

$$\Xi_\lambda : y \mapsto \begin{cases} \frac{y^\lambda - 1}{\lambda}, & \lambda \in \mathbb{R} \setminus \{0\} \\ \log y, & \lambda = 0. \end{cases} \tag{2.5.2}$$

They claim that there exists a $\lambda \in \mathbb{R}$ such that $\Xi_\lambda(Y) \mid X_1$ has linear τ-quantile curves for all $\tau \in [1 - \varepsilon, 1]$ for some $\varepsilon > 0$, and that the conditional EVI $\xi_{Y|X_1}$ of the distribution of $Y \mid X_1$ can be non-constant. However, $\xi_{Y|X_1}$ must in fact be constant in this scenario, unless $\lambda = 0$ or $F_{Y|X_1}$ is not "well behaved", as the following proposition indicates.

Proposition 2.5.4. *Suppose Assumption 2.5.1 holds, and F''_Y exists and $F'_Y > 0$ in some left neighbourhood of y^*. Further, assume $Y > 0$ almost surely. Then for $\lambda \in \mathbb{R}$, the distribution of $\Xi_\lambda(Y)$ has EVI*

$$\xi_{\Xi_\lambda(Y)} = \begin{cases} \lambda \xi_Y, & \xi_Y \geq 0; \\ \xi_Y, & \xi_Y < 0, \end{cases} \tag{2.5.3}$$

where Ξ_λ is the Box-Cox power transform defined in Equation (2.5.2).

This proposition is shown more generally in Wadsworth et al. (2010, Theorem 1), but Equation (2.5.3) is given in a less convenient form. See Appendix B.1 for a proof.

If there is to exist a $\lambda \in \mathbb{R}$ such that the upper quantile surfaces of $\Xi_\lambda(Y) \mid \boldsymbol{X}$ are linear, then the EVI of the distribution of $\Xi_\lambda(Y) \mid \boldsymbol{X}$ is constant, and this can only happen if any one of the following scenarios holds:

1. the original $\xi_{Y|\boldsymbol{X}}$ is constant;
2. $\lambda = 0$ (which results in a log-transform); or
3. there exists an $\boldsymbol{x} \in \mathscr{X}$ such that, near the upper endpoint $y^*_{Y|\boldsymbol{X}}(\boldsymbol{x})$ defined in Equation (2.5.1), $F'_{Y|\boldsymbol{X}=\boldsymbol{x}} = 0$ or $F''_{Y|\boldsymbol{X}=\boldsymbol{x}}$ does not exist.

Scenario 1 defeats the purpose of attempting a power transformation; Scenario 2 restricts the power transformation to a log-transform, so that there is no use looking for a $\lambda \neq 0$;

and Scenario 3 only holds for "misbehaved" conditional distributions, which are not useful to consider in practice. Further, Scenario 3 does not guarantee that the EVI of $Y \mid \boldsymbol{X} = \boldsymbol{x}$ is non-constant.

2.5.3 Modelling the Conditional EVI

Section 2.5.2 warns that a high amount of non-linearity in the upper quantile surfaces of $Y \mid \boldsymbol{X}$ is required for the conditional EVI to be non-constant. Such non-linearity should be accommodated when modelling if one wishes to allow for a varying conditional EVI.

Few methods exist to allow for a non-constant conditional EVI. Local estimation of high quantiles is a very flexible option, since one (indirectly) models the conditional EVI as a non-parametric function (cf. Koenker et al., 1994; Daouia et al., 2011; Spokoiny et al., 2013, for examples including one predictor). But local methods suffer from poor estimation quality of the conditional EVI: estimating any EVI is prone to error due to the data-scarcity nature of extreme values, and introducing predictors exacerbates this issue by thinning out the data even further.

Another option is to use an approach similar to a generalized linear model. For example, the conditional distribution's tail can be modelled as a Generalized Pareto distribution whose parameters (including the EVI) are modelled as a linear function of the predictors, up to some link function (cf. Coles, 2001, Chapter 6). Or, a neural network can be used to link the predictors to the parameters, as shown by Cannon (2010) for the generalized extreme value distribution. However, the choice of link function is not clear. In addition, the number of parameters involved in such a model can be quite large, especially if one wishes to account for interaction between predictors.

An alternative method is to use copulas with a non-constant CCEVI, as defined in Definition 2.4.2.

First, notice that a CCEVI cannot exceed 1.

Proposition 2.5.5. *Let ξ_i be the upper i'th CCEVI for a d-dimensional copula, $d \geq 2$. Then for almost all $\boldsymbol{u}_{-i} \in [0,1]^{d-1}$, $\xi_i(\boldsymbol{u}_{-i}) \leq 1$.*

See Appendix B.1 for a proof.

It is possible for the i'th CCEVI to be negative. In this case, the copula would not have support fully on the unit cube $[0,1]^d$. This is because the conditional distribution of the Pareto-transformed i'th variable has a finite right-endpoint, as opposed to infinity. The support does not extend all the way to 1 in variable i for the subset of the unit cube where the CCEVI is negative.

Given information about marginals, the conditional EVI of $Y \mid \boldsymbol{X}$ can be easily determined from the CCEVI.

Theorem 2.5.6. *Suppose Assumptions 2.5.1 and 2.5.2 hold, and that the dependence of* $\left(\boldsymbol{X}^\top, Y\right)$ *is described by a copula* C *with CCEVI* ξ_C*, which is defined in Definition 2.4.2. Suppose further that the right-endpoints* $\sup\left\{y \in \mathbb{R} : F_{Y|\boldsymbol{X}}\left(y|\boldsymbol{x}\right)\right\} = \infty$ *for all* $\boldsymbol{x} \in \mathscr{X}$ *(so that* $\xi_Y \geq 0$*). Suppose that the copula density* $c > 0$*,* $\mathrm{D}_2\, c$ *exists,* $F_Y' > 0$*, and* F_Y'' *exists in some left neighbourhood of* y^**, and that the von Mises limit in Equation (2.3.6) exists for both* $F = F_Y$ *and* $F = C_{2|1}\left(\cdot|u\right)$ *for each* $u \in (0,1)$*. Then*

$$\xi_{Y|\boldsymbol{X}}\left(\boldsymbol{x}\right) = \xi_Y \xi_C\left(\left(F_{X_1}\left(x_1\right), \ldots, F_{X_p}\left(x_p\right)\right)^\top\right) \tag{2.5.4}$$

for $\boldsymbol{x} \in \mathscr{X}$*, where* $F_{X_1}, \ldots, F_{X_p}$ *are the respective univariate marginal distribution functions of* $\boldsymbol{X}$*.*

See Appendix B.1 for a proof.

The CCEVI is a useful property to consider when using copulas for regression, particularly regression of extremes. Theorem 2.5.6 suggests that copulas with a CCEVI of 0 makes the response light-tailed upon conditioning on the predictor variable. We call such copulas *fully (upper) tail-lightening*. If this relationship is found in nature, then knowledge of this predictor would be very valuable, since it would entirely describe the extreme behaviour of the response. We describe such a copula family in Section 6.1. On the contrary, a copula having CCEVI of 1 suggests that conditioning on the predictor results in a conditional EVI that is unchanged from the marginal EVI. Loosely, this can be interpreted as the predictor not carrying any information about the extremes of a response. We call such copulas *tail-preserving*.

Chapter 3

Assessment of a Forecaster

Before discussing the construction of a forecaster, it is important to identify how it will be evaluated so that we know what to aim for. An important diagnostic assessment is the "calibration" of a forecaster, which assesses whether a forecaster is "on target", and is discussed in Section 3.1. To quantitatively evaluate forecasters, we use proper scoring rules, as discussed in Section 3.2. We end by emphasizing in Section 3.3 that forecasters should still be good under extreme circumstances.

3.1 Calibration

An important quality of a forecaster is its *calibration*[1]. A forecaster is *calibrated* if its forecasts of the τ-quantile are exceeded $(1-\tau)\,100\%$ of the time, for each $\tau \in (\tau_c, 1)$. The calibration of a forecaster can be defined over $(\tau_c, 1)$ as

$$\mathrm{Cal} : \tau \mapsto \mathbb{E}_{Y,\hat{Q}}\left(\mathcal{P}\left(Y \leq \hat{Q}\left(\tau\right)\right)\right) = \mathbb{E}_{\hat{Q}}\left(F_Y\left(\hat{Q}\left(\tau\right)\right)\right), \tag{3.1.1}$$

where the expectations are taken over the stochastic quantities listed in the subscripts. The forecaster is calibrated if $\mathrm{Cal}\left(\tau\right) = \tau$ for all $\tau \in (\tau_c, 1)$. Christoffersen (1998) introduced the concept of calibration in the context of interval forecasts, calling the concept "efficiency", and is discussed in terms of predictive distribution forecasts by Gneiting et al. (2007).

[1]Sometimes "calibration" is used to describe the procedure of selecting model parameters to obtain a forecaster, often referred to by statisticians as "estimation" and "model selection". This is not what we mean by calibration. The proposed estimation and selection procedures are discussed in Sections 5.2 and 5.3, respectively.

Calibration can be estimated by a sample average,

$$\widehat{\mathrm{Cal}}\left(\tau\right) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{[Y_i,\infty)}\left(\hat{Q}_i\left(\tau\right)\right), \tag{3.1.2}$$

which converges in probability to $\mathrm{Cal}\left(\tau\right)$ by the law of large numbers. Indeed, calibration is a measure of "correctness" of a forecaster, and so a check for calibration should be a "prerequisite" for a forecaster to be considered at all.

The calibration defined in Equation (3.1.1) is a sort-of "overall" calibration, where the expectation and probability is taken relative to the marginal distributions of Y and $\hat{Q}$. But calibration can also be defined under stricter circumstances – for instance, under the observation of a set of predictors $\boldsymbol{X} = \boldsymbol{x}$. In this case, a forecaster is *calibrated relative to the event* $\boldsymbol{X} = \boldsymbol{x}$ if its forecasts of the τ-quantile are exceeded $(1-\tau)\,100\%$ of the time whenever the partial information $\boldsymbol{X} = \boldsymbol{x}$ is obtained, for each $\tau \in (\tau_c, 1)$. The *relative calibration* to such event is

$$\mathrm{Cal}_{\boldsymbol{X}=\boldsymbol{x}} : \tau \mapsto \mathbb{E}_{\hat{Q}}\left(F_{Y|\boldsymbol{X}=\boldsymbol{x}}\left(\hat{Q}\left(\tau\right)\right) \mid \boldsymbol{X} = \boldsymbol{x}\right), \tag{3.1.3}$$

where the expectation is taken over the stochastic forecast $\hat{Q}$. Relative calibration across the predictor space is difficult to communicate and interpret, but a weighted version can be appropriate if one is interested in a particular region of the parameter space. This weighted analysis is discussed in Section 3.3.

Calibration can be visualized in a *calibration plot* that plots $\tau \in (\tau_c, 1)$ on the horizontal axis, and $1 - \widehat{\mathrm{Cal}}\left(\tau\right)$ on the vertical axis. Any deviation from the diagonal $1 - \widehat{\mathrm{Cal}}\left(\tau\right) = 1 - \tau$ signifies miscalibration.

Alternatively, one can consider calibration *histograms*. Notice that $\widehat{\mathrm{Cal}}\left(\tau\right)$ is simply a distribution function of $\hat{F}\left(Y\right)$, the probability integral transformed (PIT) response. This distribution function should be uniform over $\tau \in (\tau_c, 1)$. We can therefore construct a histogram of the transformed response to check the uniformity of these values. The PIT response can be interpreted as a measurement of "location" within the predictive distribution that the response realizes – larger values (closer to 1) indicate a realization that is further into the tail of the distribution $\hat{F}$. A histogram that is left-skewed (i.e., has more mass over larger values) means that the response falls into the tail more often than it should, suggesting that the forecasts need more mass in the tails. On the contrary, a right-skewed histogram suggests that the forecasts tend to have too much mass in the tail.

3.2 Proper Scoring Rules

Calibration plots do not allow for a quantitative comparison of forecasters. We therefore seek a scoring rule to carry out such comparison. According to the prequential principle of Dawid (1984), a scoring rule should assign a score using only the forecast and the observed response. A desirable type of scoring rule is the so-called *proper* scoring rule, which has the property that the expected score is optimal whenever the issued forecast equals "the distribution of Y". We clarify this concept and define such a class of scoring rules here. Throughout the discussion, we assume functions are measurable wherever appropriate.

3.2.1 Definition

We denote $S : \mathscr{Y} \times \hat{\mathcal{Q}} \to [0, \infty)$ as a scoring rule for quantile function forecasts, representing a non-negative penalty so that smaller values are better, and denote $y \in \mathscr{Y}$ as a realized value of Y.

The class $\mathcal{F}$ of possible distributions of Y is assumed to be such that the expected score given a forecast $\hat{Q} \in \hat{\mathcal{Q}}$ exists:

$$\mu_F\left(\hat{Q}\right) := \int S\left(y, \hat{Q}\right) \mathrm{d}F\left(y\right) \in \mathbb{R} \tag{3.2.1}$$

for all $F \in \mathcal{F}$. Also, the class $\hat{\mathcal{Q}}$ should only contain quantile functions that are non-decreasing, so that forecasts are sensible. The scoring rule S should have the property that its expectation under the forecast $\hat{Q}$ is smallest when $\hat{Q} = Q$, the "true" distribution.

Definition 3.2.1. *The scoring rule S is* proper *if, for all forecasts $\hat{Q} \in \hat{\mathcal{Q}}$,*

$$\int S\left(y, \hat{Q}\right) \mathrm{d}F\left(y\right) \geq \int S\left(y, Q\right) \mathrm{d}F\left(y\right) \tag{3.2.2}$$

for any $F \in \mathcal{F}$, where $Q = F^{\leftarrow}$. The scoring rule is called strictly proper *if the equality holds if and only if $\hat{Q} = Q$.*

The idea of proper scoring rules is formalized by Matheson and Winkler (1976), and advanced by Gneiting and Raftery (2007). Proper scoring rules are attractive because they reward the forecaster for issuing an honest forecast: if a forecaster truly believes in the distribution having quantile function $\hat{Q}$, then relative to this distribution, the forecaster can expect to optimize their score by issuing $\hat{Q}$ as the forecast – and $\hat{Q}$ exactly if S is strictly proper.

The goodness of a forecaster can be determined by its expectation:

$$\bar{\mu}_F := \mathbb{E}_{\hat{Q}}\left(\mu_F\left(\hat{Q}\right)\right), \tag{3.2.3}$$

where the expectation $\mathbb{E}_{\hat{Q}}$ is taken over the (stochastic) forecaster, and μ_F is defined in Equation (3.2.1), representing the mean score given a forecast relative to $F \in \mathcal{F}$. This expectation can be compared to the expectations of other forecasters to determine which is best – though, an interpretation of these expectations is lacking. In practice, the mean scores would be estimated using T forecasts $\hat{Q}_1, \ldots, \hat{Q}_T$ and realizations $Y_1, \ldots, Y_T$:

$$\hat{\bar{\mu}}_F = \frac{1}{T}\sum_{t=1}^{T} S\left(Y_t, \hat{Q}_t\right). \tag{3.2.4}$$

Though, other central estimates such as the median or trimmed mean might be considered. This is an estimate of the expectation relative to the marginal distribution $F = F_Y$. But expectations relative to other distributions $F \in \mathcal{F}$ can be used as well. For example, one might be interested in assessing the forecaster when a set of predictors $\boldsymbol{X} = \boldsymbol{x}$ is observed. In this case, the expectation is relative to the conditional distribution $F_{Y|\boldsymbol{X}=\boldsymbol{x}}$, and can be estimated by regression of the single-observation scores over the predictor space $\mathscr{X}$. This regression surface can be used to compare the performance of forecasters over different regions of the predictor space. This concept is extended further in Section 3.3 when considering weighted scores.

There are many proper scoring rules one can use, as is discussed in the next section. However, a common scoring rule should be used when comparing different forecasts so as not to introduce bias into the comparison.

3.2.2 A Class of Proper Scoring Rules

A class of proper scoring rules for a single quantile forecast has been described by Gneiting and Raftery (2007). For any quantile level $\tau \in (0,1)$, if $y \in \mathscr{Y}$ is the observed response and $\hat{Q}(\tau)$ is the forecast τ-quantile, then $s_\tau\left(y, \hat{Q}(\tau); \varphi\right) + h(y)$ is a proper scoring rule, where

$$s_\tau\left(y, \hat{Q}(\tau); \varphi\right) = \rho_\tau\left(\varphi(y;\tau) - \varphi\left(\hat{Q}(\tau); \tau\right)\right), \tag{3.2.5}$$

$h : \mathscr{Y} \to \mathbb{R}$ is arbitrary, $\varphi(\cdot;\tau) : \mathbb{R} \to \mathbb{R}$ is a non-decreasing function that is a transformation of the realization y and forecast $\hat{Q}(\tau)$, and $\rho_\tau : \mathbb{R} \to [0, \infty)$ is the asymmetric

absolute deviation function defined as

$$\rho_\tau(s) = \left(\tau - \mathbb{I}_{(-\infty,0)}(s)\right) s = \begin{cases} (1-\tau)|s|, & s < 0 \\ \tau |s|, & s \geq 0. \end{cases} \tag{3.2.6}$$

Gneiting and Raftery (2007, Equation 42) extend this class to the case of forecasting a finite set of K quantiles with levels $0 < \tau_1 < \cdots < \tau_K < 1$. Denoting $\hat{\boldsymbol{q}} = \left(\hat{Q}(\tau_1), \ldots, \hat{Q}(\tau_K)\right)$ as the forecast quantiles, and $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_K)$, a proper scoring rule is $S_c(y, \hat{\boldsymbol{q}}; \varphi, \boldsymbol{\tau}) + h(y)$, where

$$S_c(y, \hat{\boldsymbol{q}}; \varphi, \boldsymbol{\tau}) = \frac{1}{K} \sum_{k=1}^{K} \rho_{\tau_k}\left(\varphi(y; \tau_k) - \varphi\left(\hat{Q}(\tau_k); \tau_k\right)\right) \tag{3.2.7}$$

and each $\varphi(\cdot; \tau_k) : \mathbb{R} \to \mathbb{R}$ is nondecreasing. Note that the class of scoring rules defined in this way may not be exhaustive.

The composite scoring rule S_c above is simply an aggregation of the single-quantile scoring rules s_τ over some quantile levels τ. A natural extension to a class of proper scoring rules for the upper quantile function follows.

Theorem 3.2.1. *Let $h : \mathscr{Y} \to \mathbb{R}$ be an arbitrary function, and $\varphi(y; \tau) = g(y)\, w(\tau)$ where $g : \mathbb{R} \to \mathbb{R}$ is non-decreasing and $w : (\tau_c, 1) \to [0, \infty)$. Under the regularity conditions listed in Appendix C.1, the scoring rule $S\left(y, \hat{Q}; \varphi\right) + h(y)$ for $y \in \mathscr{Y}$ and $\hat{Q} \in \hat{\mathcal{Q}}$ exists and is proper relative to $\mathcal{F}$, where*

$$S\left(y, \hat{Q}; \varphi\right) = \int_{\tau_c}^{1} \rho_\tau\left(\varphi(y; \tau) - \varphi\left(\hat{Q}(\tau); \tau\right)\right) \mathrm{d}\tau. \tag{3.2.8}$$

See Appendix C.1 for a proof. Matheson and Winkler (1976) similarly discuss the aggregation of quantile scores, but in less detail and for $\tau_c = 0$.

Remark 3.2.1. The function g can be interpreted as a transformation of the response, and is therefore called the *transformation function* of S. The function w is called the *weight function* of S, since it assigns different weights across the quantile levels in $(\tau_c, 1)$.

Remark 3.2.2. In practice, the choice of h does not make a difference when comparing scores across forecasters, so it is taken to be the zero function. In addition, it is more computationally convenient to work with finite sums instead of an integral over τ. In this case, it is convenient to use the composite scoring rule S_c in place of S, with $\tau_k = \tau_c + (1 - \tau_c)(2k - 1)/(2K)$ for $k = 1, \ldots, K$.

The existence of the scoring rule S requires the concept of the *extreme value index* (EVI), discussed in Section 2.3. The EVI of a distribution can be thought of as the behaviour of the distribution's tail. Roughly, a positive EVI suggests that the survival function decays like a power function, and these distributions are said to be *heavy-tailed* (and more so when the EVI is larger); a negative EVI suggests that the distribution has a finite right-endpoint; and a zero EVI suggests that the survival function decays exponentially. It is important to note that distributions having EVI ≥ 1 do not have a finite mean. The set of distribution functions having some EVI $\xi \in \mathbb{R}$ is denoted by $\mathcal{D}\left(G_\xi\right)$, and is called the *domain of attraction* of G_ξ, where G_ξ is the so-called Extreme Value distribution, the details of which are not necessary for this dissertation. For more details of concepts in Extreme Value Theory, see de Haan and Ferreira (2006).

The conditions of Theorem 3.2.1 reveal that the scoring rule (through φ) and the forecast space $\hat{\mathcal{Q}}$ cannot be chosen independently. So as not to restrict the forecasters' beliefs, it is most sensible to gather the space of possible forecasts $\hat{\mathcal{Q}}$ before choosing a function φ. An analyst who wishes to "bypass" the choice of φ by selecting the identity transformation function and constant weight function so that $\varphi\left(y;\tau\right) = y$ inadvertently restricts the forecasters to issue forecast distributions that are not too heavy-tailed, as the following example illustrates.

Example 3.2.1. Suppose the scoring rule with $\varphi\left(y;\tau\right) = y$ is selected. Then by Condition 2 in Appendix C.1, the space of potential forecasts $\hat{\mathcal{Q}}$ can only contain distributions with extreme value indices less than 1. In addition, by Condition 3, it is further assumed that the response Y has a finite mean, and this may not be the case if, for example, Y has distribution with EVI at least 1.

If heavy-tailed distributions are allowed to be forecast, the following example illustrates a valid choice of the scoring rule.

Example 3.2.2. Suppose the space of allowed forecasts are in some domain of attraction, which covers most distributions one would use in practice. Suppose it is known by the forecasters that $\mathscr{Y} \subset (0, \infty)$. Then assuming $\mathbb{E}\left(\log Y\right)$ exists, a scoring rule with $g = \log$ is a valid choice when the weight function $w \in \mathrm{RV}_\alpha$ at 1^- for some $\alpha < 1$, and is continuous almost everywhere. Most of the conditions in Appendix C.1 follow as an immediate consequence of this definition, except for Condition 2, which follows after recognizing that $\log \circ \hat{Q}$ is the quantile function of a distribution with extreme value index in $(-\infty, 0]$.

There are some choices of φ that are always valid, as the following example illustrates.

Example 3.2.3. Take a transformation function g that is bounded above and below, and a weight function satisfying $w \in \mathrm{RV}_\alpha$ at 1^- for some $\alpha < 1$. Then the conditions in Appendix C.1 follow as a direct consequence, after recognizing that $g \circ \hat{Q}$ has a non-positive extreme value index (as long as the distribution function $\hat{Q}^{\leftarrow} \circ g^{\leftarrow}$ is in some domain of attraction, which might not happen if g and $\hat{Q}$ are too "misbehaved").

3.2.3 Choice of Quantile Weight Function

There are restrictions on the choice of φ that leads to a valid scoring rule, as discussed in Section 3.2.2. The present section investigates whether we can find scoring rules that are more desirable than others, but concludes that we cannot find any under certain criteria.

In the context of a single-quantile scoring rule s_τ, it is not well understood what transformation functions g are desirable. We instead turn our attention to finding scoring rules that have desirable properties *across* quantiles by choosing an appropriate weight function w.

One desirable feature is the notion that a single-quantile score s_τ (the integrand in Equation 3.2.8) should provide "equally good scores" to a forecast having "equally good quantiles". Such a scoring rule based on goodness alone would not assign different weights across quantiles.

Issuing the correct distribution $\hat{Q} = Q \in \mathcal{Q}$ is one example of a forecast that should sufficiently satisfy any notion of an "equally good forecast" across quantiles, so we will only consider this. The notion of an "equally good score" can be expressed through the *mean single-quantile score* at the true distribution, $\mathbb{E}\left(s_\tau\left(Y, Q(\tau); \varphi\right)\right)$. Lemma 3.2.2 identifies an expression for this expectation, and makes it clear that selecting a desirable scoring rule in this context amounts to selecting a weight function w.

Lemma 3.2.2. *Suppose the quantile function Q of Y is continuous, and $\mu = \mathbb{E}\left(g(Y)\right)$ exists. Then*

$$\mathbb{E}\left(s_\tau\left(Y, Q(\tau); \varphi\right)\right) = w(\tau)\,\tau\left(\mu - \mu_L(\tau)\right), \tag{3.2.9}$$

where $\mu_L(\tau) = \mathbb{E}\left(g(Y) \mid Y < Q(\tau)\right)$.

A proof can be found in Appendix C.2. Note that $\mu - \mu_L(\tau) > 0$ for $\tau \in (0, 1)$, by definition. Some examples of this expectation are shown in Table 3.1.

One way to define the notion of a scoring rule assigning "equally good scores" would be if the mean single-quantile score is constant (and non-zero) across $\tau \in (\tau_c, 1)$, for all distributions $Q \in \mathcal{Q}$. However, it is clear from Equation (3.2.9) that no such w exists

Distribution	$F_Y(y)$	$\mathbb{E}\left(s_\tau\left(Y, F_Y^{\leftarrow}(\tau); \varphi\right)\right)$
$Y \sim \mathcal{N}\left(\mu, \sigma^2\right)$ for $\mu \in \mathbb{R}$, $\sigma > 0$	$\Phi\left(\frac{y-\mu}{\sigma}\right)$, $y \in \mathbb{R}$	$\sigma\phi\left(\Phi^{\leftarrow}(\tau)\right) = \sigma(1-\tau)\ell(\tau)$ where $\ell \in \mathrm{RV}_0$ at 1^-.
$Y \sim \mathrm{Exp}(\mu)$ for $\mu > 0$	$1 - \exp\left(-\frac{y}{\mu}\right)$, $y > 0$	$\mu(1-\tau)\log\left(\frac{1}{1-\tau}\right)$
$Y \sim \mathrm{Par}\left(\sigma, \frac{1}{\xi}\right)$ for $\sigma > 0$, $\xi \in (0,1)$	$1 - \left(\frac{y}{\sigma}\right)^{-1/\xi}$, $y > \sigma$	$\frac{\sigma}{1-\xi}(1-\tau)\left[(1-\tau)^{-\xi} - 1\right] \sim \frac{\sigma}{1-\xi}(1-\tau)^{1-\xi}$ as $\tau \uparrow 1$.

Table 3.1: Examples of the expected single-quantile score with $w(\tau) = 1$ and $g(y) = y$, as given in Equation 3.2.9. The distributions are Gaussian (with ϕ and Φ denoting the density and distribution function of the standard Gaussian distribution, respectively), Exponential, and Type I Pareto. Proofs can be found in Appendix C.2.

due to the expectation's dependence on Q. Even if the definition was relaxed to hold for only one distribution (such as the marginal of Y), one must have *a priori* knowledge of Q (in the form of μ_L) to select such a w.

This raises the question as to whether an *a posteriori* selection of the weight function based on the fitting data is appropriate. For example, one may estimate $\tau \mapsto \tau\left(\mu - \mu_L(\tau)\right)$ using the fitting data, and choose its reciprocal as the weight function so that $\mathbb{E}\left(s_\tau\left(Y, Q(\tau); \varphi\right)\right)$ is estimated to be 1 relative to the marginal distribution of Y. We recommend against any *a posteriori* selection of the weight function. Although the resulting scoring rule would be proper when assessing new data, it may not be proper when assessing the fitting data, as is done in estimation, for example (see Section 5.2). This may result in an estimator that does not converge to the true distribution, because "dishonest" forecasts may be encouraged.

Lemma 3.2.2 demonstrates that, without the weight function (i.e. with $w(\tau) = 1$), the mean single-quantile score tends towards 0 as $\tau \uparrow 1$, since $\mu_L(\tau) \to \mu$. This means that higher quantiles have less of an influence on the overall score than others. Proposition 3.2.3 investigates the rate of decay of the mean single-quantile score.

Proposition 3.2.3. *Suppose $w \in RV_\alpha$ at 1^- for some $\alpha \in \mathbb{R}$, and $g : \mathbb{R} \to \mathbb{R}$ is a non-decreasing function such that the distribution function $F \circ g^{\leftarrow} \in \mathcal{D}\left(G_\xi\right)$ for some $\xi < 1$. Then*

$$\tau \mapsto \mathbb{E}\left(s_\tau\left(Y, Q(\tau); \varphi\right)\right) \in RV_{\alpha+\xi^+-1} \text{ as } \tau \uparrow 1,$$

where $\xi^+ = \max(0, \xi)$.

Perhaps instead of asking for the mean single-quantile score to be constant, we can ask for it to be slowly varying as $\tau \uparrow 1$, achieved through an appropriate choice of α. If this

is achieved, then the mean single-quantile score would display "constant-like behaviour" near $\tau = 1$, so that upper quantiles would be weighed similarly. To achieve this for a single $F \in \mathcal{F}$ with $F \circ g^{\leftarrow} \in \mathcal{D}\left(G_\xi\right)$ for $\xi < 1$, we would need to choose $\alpha = 1 - \xi^+$. Such a choice is valid when the correct forecasts are issued: according to Condition 2' in Appendix C.1, we only require $\alpha < 2 - \xi^+$ in order for the scoring rule in Equation (3.2.8) to exist.

It is realistic to select a g so that ξ^+ (and thus α) is assumed to be known *a priori*. It is often reasonable to assume *a priori* that each distribution function in $\mathcal{F}$ belongs to some domain of attraction. Taking $g = \log$ will then ensure that, for each $F \in \mathcal{F}$, $F \circ g^{\leftarrow} \in \mathcal{D}\left(G_\xi\right)$ where $\xi \leq 0$, so that $\xi^+ = 0$. One can then take $\alpha = 1$, using a weight function such as $w(\tau) = 1/(1-\tau)$.

However, in using Condition 2', we have assumed that all forecasts will never have their right-endpoint exceeded, and this may *not* be realistic in practice. Otherwise, to ensure the existence of the scoring rule in Equation (3.2.8), one must consult Condition 2 in Appendix C.1 as opposed to Condition 2'. The former condition states the requirement $\alpha < 1 - \xi^+$, meaning the selection of $\alpha = 1 - \xi^+$ is no longer an option. If one decides to choose $w \in \mathrm{RV}_\alpha$ at 1^- with e $\alpha = 1 - \xi^+$ anyway, forecasts would receive a score of infinity whenever the response exceeds the right-endpoint of the forecast distribution (see the existence proof of Theorem 3.2.1 in Appendix C.1 to see why).

In fact, if one must choose $\alpha < 1 - \xi^+$, then the mean single-quantile score will always decay to zero. Empirical evidence suggests that the closer α is to $1 - \xi^+$, the later this decay happens. Hence, one may wish to choose α close to $1 - \xi^+$.

3.3 Integrity of Extreme Forecasts

The calibration and proper scoring rule assessments discussed in Sections 3.1 and 3.2 are in general applicable to any forecasting method that issues a predictive distribution. When forecasting extremes, however, we ask that a forecaster's performance not be compromised in the tail of $\boldsymbol{X}$ – in other words, the forecaster should keep its integrity (or "goodness").

Integrity of forecasts in a distributional tail of $\boldsymbol{X}$ is important because the most extreme responses are typically more prone to occur here. Forecasters that keep their integrity in the tail of $\boldsymbol{X}$ should produce scores that are not compromised when $\boldsymbol{X}$ is observed in its tail. This can be evaluated by estimating the mean score surface $\boldsymbol{x} \mapsto \bar{\mu}_{F_{Y|\boldsymbol{X}=\boldsymbol{x}}}$ in Equation (3.2.3) over the tail, and comparing this to the marginal mean $\bar{\mu}_{F_Y}$.

To obtain a concrete score for the mean over the tail of $\boldsymbol{X}$, a weighted average over the predictor space can be used. The weight function $w_{\boldsymbol{X}} : \mathscr{X} \to [0, \infty)$ should be chosen to put more weight in the tail of $\boldsymbol{X}$. In the application in Chapter 7, a weight function based on the logistic function is used. A weighted estimator of the mean score in Equation (3.2.3) is

$$\hat{\bar{\mu}}_{F_Y}^{(w_{\boldsymbol{X}})} = \frac{\sum_t w_{\boldsymbol{X}}\left(\boldsymbol{X}_t\right) S\left(Y_t, \hat{Q}_t\right)}{\sum_t w_{\boldsymbol{X}}\left(\boldsymbol{X}_t\right)}.$$

For example, in the context of flooding, a forecaster that scores equally good when a heavy snowmelt is observed is desirable – it keeps its integrity.

Similarly, integrity in the tail of $\boldsymbol{X}$ can also be assessed with a calibration plot, where an estimate of the calibration is

$$\widehat{\mathrm{Cal}}_{w_{\boldsymbol{X}}}\left(\tau\right) = \frac{\sum_t w_{\boldsymbol{X}}\left(\boldsymbol{X}_t\right) \mathbb{I}_{\left(\hat{Q}_t(\tau), \infty\right)}\left(Y_t\right)}{\sum_t w_{\boldsymbol{X}}\left(\boldsymbol{X}_t\right)}.$$

Chapter 4

Existing Forecasting Methodology

Under the framework in Chapter 2, this chapter discusses how existing methodology might be used to build forecasters, along with their shortcomings. The methods discussed are local regression methods, linear regression, and fully parametric regression. In short, parametric assumptions are needed to extrapolate into the upper tail of the conditional distribution of $Y \mid \boldsymbol{X}$, as well as into the tails of the predictor space. Yet, these existing methods fall short in some way.

4.1 Local Regression

These methods model the quantile surfaces for any quantile level τ as a non-parametric function of the predictors. There are many such methods, and we discuss only a few here.

Smoothing methods for the mean can be extended to quantile regression. Local polynomials are discussed by Spokoiny et al. (2013), who concisely review other similar methods. Splines are discussed by Koenker et al. (1994) in the single-predictor setting, and is extended to the two-predictor setting in Koenker and Mizera (2004).

In the context of extreme quantiles, Daouia et al. (2011) consider fitting quantile functions at each point on the predictor space. For one such point $\boldsymbol{x}$, they propose estimating the conditional quantile function using the inverse empirical distribution function of the response data weighted proportionally to their predictors' vicinity to $\boldsymbol{x}$. This quantile function is used to obtain a Pickands-type estimate of the conditional extreme value index. Asymptotics of the quantile estimators are given by Daouia et al. (2013) when the conditional response belongs to the domain of attraction of some extreme value distribution.

Regardless of the specifics, these methods suffer from data sparsity when the number of predictors is not small, so that the fitted quantile functions will generalize poorly to

new data. This is true for any local fitting technique, but the problem is exacerbated in the context of extreme forecasting because even less data are found in either the $\boldsymbol{X}$ or $Y \mid \boldsymbol{X}$ tail. However, local methods are still useful to obtain a preliminary sense of the shape of the quantile surfaces.

4.2 Linear Regression

These methods model quantile surfaces as linear (or polynomial) functions, but typically do not impose a parametric form across quantile levels.

Possibly the most well-known method of this type is linear quantile regression, introduced by Koenker and Bassett (1978), where each quantile is modelled as a linear function of the predictors:

$$\hat{Q}_{Y|\boldsymbol{X}}\left(\tau|\boldsymbol{x};\alpha,\boldsymbol{\beta}\right) = \alpha\left(\tau\right) + \boldsymbol{x}^{\top}\boldsymbol{\beta}\left(\tau\right), \tag{4.2.1}$$

where $\boldsymbol{\beta} : (0,1) \to \mathbb{R}^p$ and $\alpha : (0,1) \to \mathbb{R}$. The parameters for a particular quantile level τ can be estimated by optimizing the score for a single quantile prediction,

$$\left(\hat{\alpha}\left(\tau\right), \hat{\boldsymbol{\beta}}\left(\tau\right)\right) = \underset{(a,\boldsymbol{b})}{\arg\min} \sum_{t=1}^{T} \rho_\tau\left(y_t - a - \boldsymbol{x}_t^{\top}\boldsymbol{b}\right), \tag{4.2.2}$$

where ρ_τ is the asymmetric absolute deviation function in Equation (3.2.6). See Koenker (2005) for a comprehensive review of quantile regression.

One problem with Koenker's method is that it may result in non-monotonic forecasts. This happens due to the "crossing" of different quantile surfaces, and is noted by Koenker (2005). This may be particularly problematic when predictors are observed to be extreme, as the application in Chapter 7 demonstrates.

There are several papers that attempt to address the crossing issue. Chernozhukov et al. (2010) "rearrange" the invalid quantile function to obtain a valid quantile function, both of which share a common stochastic process when generating data. Bondell et al. (2010) restrict estimation so that the quantile surfaces do not cross within the convex hull of the covariates, but this may not be appropriate if an extreme predictor is observed.

Another method to avoid inconsistent forecasts is to force the quantile surfaces to be parallel, as in the model used by Zou and Yuan (2008). Their model restricts Koenker's model so that $\boldsymbol{\beta}$ is constant. To estimate the common slope parameter, they propose a composite quantile regression estimator similar to Equation (4.2.2) that borrows strength

across K quantile levels $0 < \tau_1 < \cdots < \tau_K < 1$, again optimizing the predictor's score:

$$\left(\hat{\alpha}\left(\tau_1\right), \ldots, \hat{\alpha}\left(\tau_K\right), \hat{\boldsymbol{\beta}}\right) = \underset{(a_1,\ldots,a_K,\boldsymbol{b})}{\arg\min} \sum_{k=1}^{K} \sum_{t=1}^{T} \rho_{\tau_k}\left(y_t - a_k - \boldsymbol{x}_t^\top \boldsymbol{b}\right). \tag{4.2.3}$$

Even in light of these solutions to the inconsistency problem, models enforcing linear quantile surfaces are restrictive (Bernard and Czado, 2015). In particular, the fitted quantile functions over the tail of $\boldsymbol{X}$ can be highly biased, and therefore noncompliant with the $\boldsymbol{X}$ tail integrity requirement discussed in Section 3.3. Higher order terms can be added in an attempt to address this issue, but at the risk of losing the ability to generalize to new data, due to overfitting. In addition, the original problem of poor fit persists when we wish to extrapolate into the predictors' tail. This situation would happen if we wish to issue a forecast when record predictors are observed (such as record-high snowmelt).

To improve upon the fit in the tail of $\boldsymbol{X}$, Huang et al. (2015) propose weighting the quantile regression so that the tails carry more weight. This would improve the fit closer to the tail, but at the cost of compromising the fit over "central" values of the predictors, as seen in the unweighted score. In addition, the resulting forecaster would still be unable to extrapolate further into the tail of $\boldsymbol{X}$.

There is yet another problem: as long as the quantile functions are non-parametric in the quantile level, the quantile functions cannot effectively extrapolate into the distributional tail – the largest responses impose an unrealistic upper bound. A parametric assumption over the quantile level is needed to extrapolate into the tail beyond the bound imposed by the response variable.

Overall, these regression methods are generally useful when making predictions on "intermediate" quantiles after observing typical values of the predictors.

4.3 Fully Parametric Regression

A parametric modelling approach is to first fit a parametric distribution to $\left(\boldsymbol{X}^\top, Y\right)$, from which the conditional quantile function of Y given the predictors can be derived. If a multivariate Gaussian distribution is used to model the joint distribution, the conditional quantiles become linear in the predictors, with univariate Gaussian error. A more modern approach to modelling $F_{\boldsymbol{X},Y}$ is to use parametric copula families to describe the dependence. Vine copulas are a flexible approach to such modelling in the multi-predictor setting.

In the case of a single predictor, Bouyé and Salmon (2009) first suggest modelling

the conditional quantiles by specifying a copula model in a univariate time series. They propose to use nonlinear quantile regression over one quantile to estimate the copula parameters, and call the estimator a "Copula Quantile Regression" (CQR) estimator. Koenker (2005, Section 8.4) mentions as a research topic the use of copula models to build non-linear quantile models (and refers to a 2002 version of Bouyé and Salmon, 2009).

Chen et al. (2009) use a similar model for time series data as Bouyé and Salmon (2009), but suggest that the CQR estimator can be used separately to estimate different quantiles (so that the copula parameters need not be common across all quantile levels). Xie (2015) proposes estimating the copula parameters as being common across quantile levels by taking a weighted average of CQR estimates across different quantile levels. However, their recommended weights depend on the value of the predictor and it is therefore unclear exactly how the weights are computed (especially in comparison with the maximum likelihood estimator).

To generalize the copula model approach to p predictors, Kraus and Czado (2017) recommend modelling the joint distribution of $\left(\boldsymbol{X}^\top, Y\right)$ using a special type of vine called a D-Vine, and discuss how to compute the quantiles of Y given the predictors. Instead of using quantile regression to estimate the quantile function, they fully estimate the joint D-Vine distribution first.

However, fitting the joint distribution of $\left(\boldsymbol{X}^\top, Y\right)$ in its entirety will not necessarily lead to a good fit when the upper quantiles of Y given the predictors are extracted. The same can be said about the weighted estimator of Xie (2015). In addition, the joint distribution of $\left(\boldsymbol{X}^\top, Y\right)$ can be modelled more flexibly than the D-Vine of Kraus and Czado (2017). Improvements in these respects are addressed next in Chapter 5.

Chapter 5

Proposed Forecasting Methodology

In this chapter, we propose a method to obtain a forecaster. The new approach allows for flexible shapes of high quantile surfaces, and accounts for dependence amongst the predictors and response that is non-Gaussian. Essentially, the procedure involves specifying a joint distribution for $\left(\boldsymbol{X}^\top, Y\right)$, from which the quantile function of $Y \mid \boldsymbol{X}$ can be derived.

Making a forecaster requires the following steps.

1. Model the quantile function of $Y \mid \boldsymbol{X}$ by modelling the joint distribution of $\left(\boldsymbol{X}^\top, Y\right)$ with a vine copula or PCBN; see Section 5.1.

2. Fit the model to obtain a forecaster by optimizing the score on the training data; see Section 5.2.

3. Repeat Steps 1-2 with other candidate models, and select the optimal forecaster on the validation set; see Section 5.3.

5.1 Building a Model

We are interested in obtaining a model for the quantile function of $Y \mid \boldsymbol{X}$ that is parametric (or semi-parametric) across both predictors and quantile level. The main idea is to specify a PCBN model for the distribution of $\left(\boldsymbol{X}^\top, Y\right)$, and then derive the conditional quantile function.

Using copula models to describe the dependence amongst variables is advantageous because it allows for tail dependence to be modelled, resulting in nonlinear quantile surfaces. For example, it could be that a predictor does not influence the response much unless the predictor is large. Or, observing two large predictors could have a greater effect

on the response than they independently would have. Bernard and Czado (2015) discuss the importance of capturing tail dependence when modelling quantile curves (they only consider one predictor). They show the relationship between tail dependence and the asymptotic ($x \to \pm\infty$) quantile curves, and demonstrate that the assumption of linear quantile curves is a poor one.

Examples of quantile curves for some bivariate copulas can be found in Figure 5.1, under various marginals. The properties of these copulas are observable in the shapes of the quantile curves, and is is discussed more extensively by Bernard and Czado (2015). For example, the Frank copulas have no tail dependence, and the flattening of the quantile curves for larger predictor values is telling of this tail independence. The Gumbel copulas have quantile curves that approach the positive diagonal, indicating tail dependence. Overall, linearity of quantile curves (and by extension, quantile surfaces) is generally a poor assumption outside of the "central" predictor space. As such, the flexibility of the PCBN (or vine) approach allows for the forecaster to perform well, even when extreme predictors are observed.

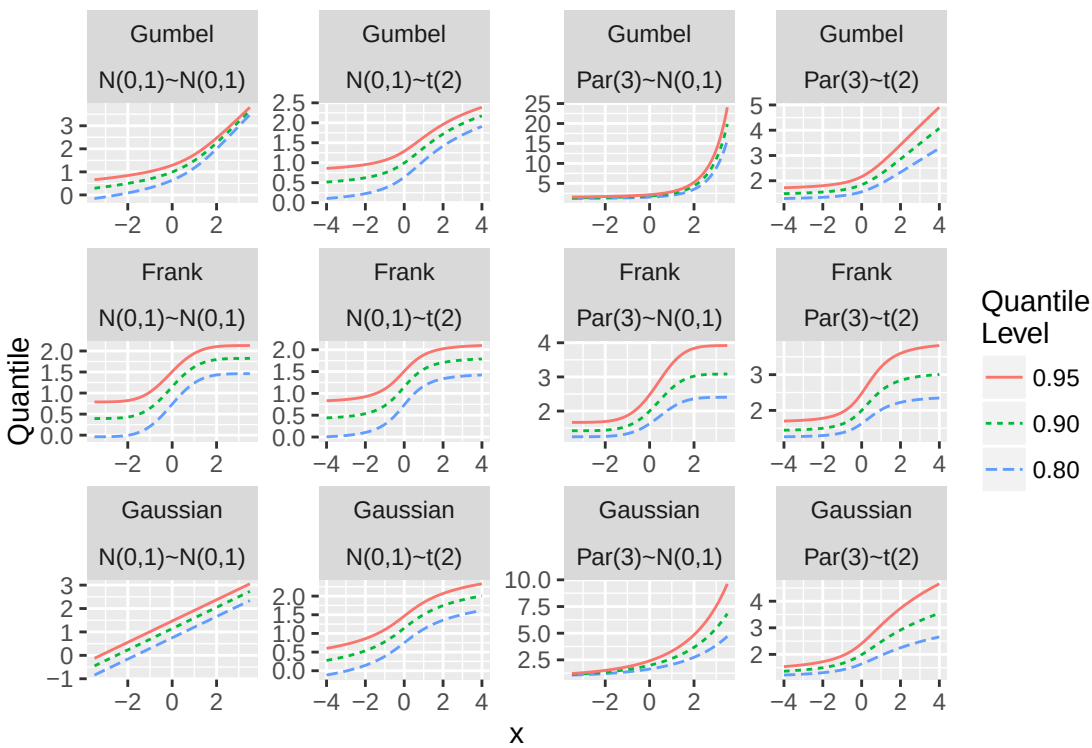

Figure 5.1: Examples of quantile curves of some bivariate distributions determined by copulas. Parameters for the copula families are chosen to have a Kendall's tau of 0.3. The marginal distributions are indicated in the form of "response distribution" ~ "predictor distribution".

After modelling the joint distribution of $\left(\boldsymbol{X}^\top, Y\right)$ by a parametric distribution, the

resulting parametric model for the quantile surfaces may be non-identifiable. This is because some of the parameters are strictly associated with the joint distribution of the predictors, as the following example demonstrates.

Example 5.1.1. Consider a Gaussian model for the joint distribution of three random variables:

$$(X_1, X_2, Y) \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} 1 & \rho_1 & \sigma\rho_3 \\ \rho_1 & 1 & \sigma\rho_2 \\ \sigma\rho_3 & \sigma\rho_2 & \sigma^2 \end{bmatrix}\right).$$

The conditional quantile function of $Y \mid \boldsymbol{X}$ is

$$\hat{Q}_{Y|(X_1,X_2)}\left(\tau \mid (x_1, x_2); \boldsymbol{\beta}\right) = \beta_0 \Phi^{\leftarrow}(\tau) + \beta_1 x_1 + \beta_2 x_2,$$

where $\Phi^{\leftarrow}$ is the probit function, and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^\top$ can be found from the parameters $\boldsymbol{\theta} = (\rho_1, \rho_2, \rho_3, \sigma)^\top$ by

$$\beta_0 = \sigma\sqrt{\frac{1 + 2\rho_1\rho_2\rho_3 - \rho_1^2 - \rho_2^2 - \rho_3^2}{1 - \rho_1^2}},$$

$$\beta_1 = \sigma\frac{\rho_3 - \rho_1\rho_2}{1 - \rho_1^2},$$

and

$$\beta_2 = \sigma\frac{\rho_2 - \rho_1\rho_3}{1 - \rho_1^2}.$$

The parameters $\boldsymbol{\theta}$ are non-identifiable for $\hat{Q}_{Y|(X_1,X_2)}$ because there exist more than one (in fact, infinitely many) members of the parameter space of $\boldsymbol{\theta}$ that result in the same value of $\boldsymbol{\beta}$ – that is, results in the same quantile surfaces. More generally, the regression slopes $\boldsymbol{\beta}$ is a matrix product of the inverse covariance matrix of the predictors, times the covariance vector between the response and predictors. If the former is left as a parameter, then this linear combination is non-identifiable.

A simple solution to the non-identifiability problem is to estimate those parameters associated with the distribution of $\boldsymbol{X}$ first, by maximum likelihood, for example.

Computing the quantile function of $Y \mid \boldsymbol{X}$ and the distribution of $\boldsymbol{X}$ from a PCBN is in general cumbersome and computationally expensive. The situation improves if the PCBN is modelled so that Y is introduced last. This is the proposed modelling method, and can be summarized in the following steps:

1. Fit a model to $F_{\boldsymbol{X}}$, the joint distribution of $\boldsymbol{X}$. For example, use a vine copula or PCBN.

2. Choose an order in which to link the predictors to the response. Without loss of generality, suppose this order is $1, \ldots, p$.

3. Choose a copula family to model $\left(X_j, Y\right) \mid \boldsymbol{X}_{1:(j-1)}$, for the selected order $j = 1, \ldots, p$.

4. Use Equation (A.2.3) in Appendix A.2 to calculate the quantile function of $Y \mid \boldsymbol{X}$.

Using this modelling method, the model for $\boldsymbol{X}$ is easier to fit, because it does not require the response to be integrated out of the full joint distribution. In addition, the equation for the quantile function of $Y \mid \boldsymbol{X}$ indicated in Step 4 is a closed-form equation.

This modelling method also carries an interpretation. A subsequent variable in the pairing order determined in Step 2 explains some of the "leftover" uncertainty in the response that previous variables cannot explain. Like in principal components analysis, the result is a division of the uncertainty of the response amongst the predictors. Consequently, we obtain a natural variable selection method, similar to that described in Kraus and Czado (2017): a predictor can be removed if it does not have much more to contribute in addition to the previously fitted predictors. However, since model error is compounded further down the linkage order of Step 2, the most descriptive predictors ought to be included early in the linkage order.

In the computation of the conditional quantile function in Step 4, the joint predictor distribution is only needed to transform the predictors X_j to an independent set of uniform predictors $U_j := F_{j|1:(j-1)}\left(X_j \mid \boldsymbol{X}_{1:(j-1)}\right)$ for $j = 1, \ldots, p$, forming the vector $\boldsymbol{U}$. This transformation sometimes coincides with the so-called *Rosenblatt transform* if the linkage order of Step 2 coincides with the introduction order of the vine/PCBN of the predictors. As such, estimation of the joint predictor distribution in Step 1 is not as critical as fitting the link between the predictors and response, whose fit will adapt appropriately to the resulting transformation of the predictors.

However, estimation of the joint predictor distribution *is* still important. If this estimation is done poorly, then there is more likely to be issues with multicollinearity, as the transformed predictors are no longer spread uniformly throughout the unit p-cube. But more importantly, the copulas modelled between the response and transformed predictors in Step 3 may not actually represent the copulas underlying the model. Unintended consequences may result, such as unintended tail dependence or CCEVI trend.

However, computing $\boldsymbol{U}$ can be computationally expensive if the PCBN on $\boldsymbol{X}_{1:j}$ for $j = 2, \ldots, p$ are not vines, and could each involve up to $(j-1)$-dimensional integrals. One way to simplify this is to model the distribution of $\boldsymbol{X}$ with a vine that has an introduction order equal to the pairing order of Y. Since vine models are extremely flexible, this should not negatively affect the flexibility of the modelling procedure. Then a closed-form expression exists for each transformed variable. In this case, the PCBN for $\left(\boldsymbol{X}^\top, Y\right)$ is also a vine.

When the PCBN for $\left(\boldsymbol{X}^\top, Y\right)$ is a D-Vine, then the model is similar to that of Kraus and Czado (2017) – the difference being that they fit the entire joint distribution of $\left(\boldsymbol{X}^\top, Y\right)$ using maximum likelihood so that the entire quantile function of $Y \mid \boldsymbol{X}$ can be found.

5.2 CNQR

With a parametric model of the conditional quantile function available, the next task becomes fitting the model. To do this, an estimation method called composite nonlinear quantile regression (CNQR) estimation is proposed in Section 5.2.1. The idea behind the estimation is to select the parameter that optimizes the forecaster's average score over the training data. Since there are a family of scores that can be optimized, there is a corresponding *family* of CNQR estimators.

Asymptotics of the CNQR estimators is established in Section 5.2.2. Under some regularity conditions, a CNQR estimator is consistent and asymptotically Gaussian, and this information can be used to make inference. Estimator choice is briefly investigated by comparing asymptotic variances, and it is found that about 10 quantile levels are appropriate for estimation, and that choosing a heavy-tailed transformation function might compromise the performance of the estimator.

Other existing estimation methods in the literature are special cases of CNQR estimation, as is discussed in Section 5.2.3.

5.2.1 Estimation

Suppose the upper quantile function of Y conditional on the predictors is modelled by the family $\hat{Q}_{Y|\boldsymbol{X}}\left(\tau|\boldsymbol{x};\boldsymbol{\theta}\right)$, indexed by (identifiable) parameters $\boldsymbol{\theta} \in \Theta$. We would like to fit a model from the family using observations $\left(\boldsymbol{X}_i^\top, Y_i\right)$ for $i = 1, \ldots, n$.

The *optimum score estimator* is a natural choice for estimation. It chooses the $\boldsymbol{\theta} \in \Theta$ so that the average score, as indicated in Equation (3.2.7), is optimized. Specifically, for

a quantile level cutoff $\tau_c \in (0,1)$, a selection of quantile levels $\tau_c \leq \tau_1 < \cdots < \tau_K < 1$, and non-decreasing transformation functions $g_1, \ldots, g_K$, the parameter $\hat{\boldsymbol{\theta}}_n \in \Theta$ is chosen, where

$$\hat{\boldsymbol{\theta}}_n = \underset{\boldsymbol{\theta}\in\Theta}{\arg\min} \frac{1}{nK} \sum_{i=1}^{n} \sum_{k=1}^{K} \rho_{\tau_k} \left(g_k\left(Y_i\right) - g_k\left(\hat{Q}_{Y|\boldsymbol{X}}\left(\tau_k|\boldsymbol{X}_i;\boldsymbol{\theta}\right)\right)\right), \tag{5.2.1}$$

and ρ_τ is indicated in Equation (3.2.6). The resulting family of estimators is referred to as the *composite nonlinear quantile regression* (CNQR) family. When

$$\tau_k = \frac{2k-1}{2K}\left(1-\tau_c\right) + \tau_c \tag{5.2.2}$$

for $k = 1, \ldots, K$, the CNQR estimators become a Riemann Sum approximation to the integral scoring rule in Equation (3.2.8).

The CNQR objective function can have a cusp at its minimum, as demonstrated in Figure D.1 of Appendix D. In addition, the function is non-differentiable at many points along its domain. This means minimization methods that rely on derivatives, such as gradient descent, may run into difficulties. However, the objective function approximates a smooth function for a large sample size, so that gradient descent methods may be sufficient. If not, Koenker and Park (1996) describe an algorithm for computing such a minimum. See Appendix D for a discussion on the smoothness of the objective function.

The cutoff τ_c is intended to be chosen "near" 1. In doing so, CNQR estimation selects a model that more accurately describes the distributional tail of Y given the predictors. However, choosing a larger τ_c comes at a cost of accepting a higher variance in the quantile estimates. This bias-variance tradeoff is similar to that of estimating the extreme value index of a univariate sample with the peaks over threshold method.

Note that the nonlinear quantile regression estimator of Koenker (2005) results as a special case of the estimator in Equation (5.2.1) when $K = 1$, g_1 is the identity function, and $\hat{Q}_{Y|\boldsymbol{X}}\left(\tau|\boldsymbol{x};\boldsymbol{\theta}\right) = \left(1, \boldsymbol{x}^\top\right)\boldsymbol{\theta}$. The CNQR estimators are similar to the weighted average estimator of Xie (2015), but the latter provides no guarantee that the scoring rule is optimized.

Instead of CNQR estimation, it is possible to find the likelihood from the quantile model, and use the corresponding maximum likelihood estimator (MLE). However, the lower endpoint of the distribution of $Y \mid \boldsymbol{X} = \boldsymbol{x}$ is $\lim_{\tau\downarrow\tau_c} \hat{Q}_{Y|\boldsymbol{X}}\left(\tau|\boldsymbol{x};\boldsymbol{\theta}\right)$, which depends on the model parameters. The regularity conditions are therefore not all satisfied, meaning that the MLE may have undesirable properties. In addition, there is no guarantee that the resulting forecaster would optimize the scoring rule.

5.2.2 Inference

To estimate the conditional quantile function of Y given the predictors, a parametric or semi-parametric family of such functions is chosen. To study asymptotic properties, the family is assumed to contain the true quantile function.

Assumption 5.2.1. The data $\left(\boldsymbol{X}_i^\top, Y_i\right)$ for $i = 1, \ldots, n$ are iid replicates of the random vector $\left(\boldsymbol{X}^\top, Y\right)$.

Assumption 5.2.2. Consider K quantile levels $\tau_c \leq \tau_1 < \cdots < \tau_K < 1$ for some $\tau_c \in (0, 1)$. Further, consider a family of quantile functions $(\tau, \boldsymbol{x}) \mapsto \hat{Q}_{Y|\boldsymbol{X}}\left(\tau|\boldsymbol{x}; \boldsymbol{\theta}\right)$ with domain $(\tau_c, 1) \times \mathscr{X}$, each strictly increasing in its first argument, indexed by parameters $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^{d_\Theta}$, $d_\Theta \geq 1$. There exists a $\boldsymbol{\theta}_0 \in \Theta$ such that $Q_{Y|\boldsymbol{X}}\left(\tau_k|\boldsymbol{x}\right) = \hat{Q}_{Y|\boldsymbol{X}}\left(\tau_k|\boldsymbol{x}; \boldsymbol{\theta}_0\right)$ for all $k = 1, \ldots, K$ and $\boldsymbol{x} \in \mathscr{X}$.

The composite non-linear quantile regression (CNQR) family of estimators of $\boldsymbol{\theta}_0$ are introduced in Equation (5.2.1) in Section 5.2. The estimators optimize the fit of K upper quantile surfaces with levels above a cutoff $\tau_c \in (0, 1)$, via a scoring rule indicated in Equation (3.2.7) in Section 3.2.2. Although it is indicated that the cutoff τ_c be non-zero, the CNQR estimators are also valid for $\tau_c = 0$. However, in this case, the family of models would describe the entire conditional distribution of Y given the predictors, so that the maximum likelihood estimator may be more appropriate. As discussed in Section 5.2, the motivation for choosing τ_c near 1 is to improve accuracy of high quantile estimation.

For simplicity of notation, for each $i = 1, \ldots, n$ and $k = 1, \ldots, K$, define

$$q_{ki} : \boldsymbol{\theta} \mapsto g_k\left(\hat{Q}_{Y|\boldsymbol{X}}\left(\tau_k|\boldsymbol{X}_i; \boldsymbol{\theta}\right)\right) \tag{5.2.3}$$

over the parameter space Θ, having gradient

$$\dot{\boldsymbol{q}}_{ki} : \boldsymbol{\theta} \mapsto \frac{\partial}{\partial \boldsymbol{\theta}} q_{ki}\left(\boldsymbol{\theta}\right) \tag{5.2.4}$$

and Hessian

$$\ddot{\boldsymbol{q}}_{ki} : \boldsymbol{\theta} \mapsto \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} q_{ki}\left(\boldsymbol{\theta}\right). \tag{5.2.5}$$

The i subscript is sometimes dropped to refer to the generic random vector $\left(\boldsymbol{X}^\top, Y\right)$.

Theorem 5.2.1 (Consistency)**.** *Suppose Assumptions 5.2.1 and 5.2.2 hold, and consider non-decreasing functions $g_k : \mathbb{R} \to \mathbb{R}$, $k = 1, \ldots, K$. Under the regularity conditions listed in Condition 1 in Appendix D, the CNQR estimator $\hat{\boldsymbol{\theta}}_n$ defined in Equation (5.2.1)*

converges in probability to $\boldsymbol{\theta}_0$ *as* $n \to \infty$*. That is, the CNQR family of estimators are consistent estimators.*

Theorem 5.2.2 (Asymptotic Normality)**.** *Suppose Assumptions 5.2.1 and 5.2.2 hold, and consider non-decreasing functions* $g_k : \mathbb{R} \to \mathbb{R}$*,* $k = 1, \ldots, K$*. Under the regularity conditions listed in Conditions 1 and 2, the CNQR estimator* $\hat{\boldsymbol{\theta}}_n$ *defined in Equation (5.2.1) satisfies* $\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \to_d \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_{CNQR}\right)$*, where*

$$\boldsymbol{\Sigma}_{CNQR} = \boldsymbol{D}_1^{-1}\left[\sum_{k,k'=1}^{K} \boldsymbol{D}_{0kk'} \min\left(\tau_k, \tau_{k'}\right)\left(1 - \max\left(\tau_k, \tau_{k'}\right)\right)\right]\boldsymbol{D}_1^{-1}, \tag{5.2.6}$$

$$\boldsymbol{D}_{0kk'} = \mathbb{E}\left(\left[\dot{\boldsymbol{q}}_k\left(\boldsymbol{\theta}_0\right)\right]\left[\dot{\boldsymbol{q}}_{k'}\left(\boldsymbol{\theta}_0\right)\right]^\top\right), \tag{5.2.7}$$

for each k, k'*,*

$$\boldsymbol{D}_1 = \sum_{k=1}^{K} \boldsymbol{D}_{1k}, \tag{5.2.8}$$

$$\boldsymbol{D}_{1k} = \mathbb{E}\left(f_{g_k(Y)|\boldsymbol{X}}\left(q_k\left(\boldsymbol{\theta}_0\right)|\boldsymbol{X}\right)\left[\dot{\boldsymbol{q}}_k\left(\boldsymbol{\theta}_0\right)\right]\left[\dot{\boldsymbol{q}}_k\left(\boldsymbol{\theta}_0\right)\right]^\top\right), \tag{5.2.9}$$

and q_k *and* $\dot{\boldsymbol{q}}_k$ *are respectively defined in Equations (5.2.3) and (5.2.4).*

Proofs of Theorems 5.2.1 and 5.2.2 are provided in Appendix D. These proofs consider the limit of the CNQR objective function defined in Equation (5.2.1), as opposed to the finite-sample objective function that is used in estimating equations theory. This is because the gradient may not exist at the minimum.

Theorem 5.2.2 is useful when making inference on the parameter $\boldsymbol{\theta}_0$, and more importantly, on the quantiles themselves (using the delta method). For a large enough sample size, one could approximate the sampling distribution of the CNQR estimator using the Gaussian distribution, and estimate the variance of this distribution using an estimate of Equation (5.2.6). However, the asymptotic distribution under intermediate and extreme order sequences might be a better finite-sample approximation than the Gaussian when τ_c is large – that is, by allowing $\tau_c \uparrow 1$ as $n \to \infty$. For an example of this in the linear quantile regression case, see Chernozhukov (2005).

The asymptotic covariance in Theorem 5.2.2 can provide insight into a desirable choice of a CNQR estimator. In particular, it is interesting to ask what transformation functions g_k result in a better estimator. For example, what effect does transforming the response to a light-tailed, heavy-tailed, or short-tailed distribution have on the estimator performance? In addition, when using the Riemann sum approximation with quantile levels computed using Equation (5.2.2), how is the variance related to the number of

quantile levels K, and at what point does the variance stabilize? How does the choice of τ_c affect the estimator variance?

These questions are investigated in Figure 5.2 using bivariate data generated from three models (a Frank copula, Gumbel copula, and reflected Gumbel copula). Estimation is done using the correct copula families and correct marginals. As expected, the precision worsens as larger values of τ_c are chosen, since less information is being drawn from the data. Also, it appears that when the response is transformed to have a heavy-tailed distribution, the estimator performance may be compromised. Perhaps this is due to the "spreading out" of the quantile curves when the tail is heavy. Another way to interpret this finding is that, if a response variable is suspected to have a heavy-tailed (Pareto-like) distribution, then it should be transformed to have a lighter tail (such as exponential or uniform). It seems that choosing $K = 10$ quantile levels is sufficient for convergence of the variance.

The asymptotic precision can be interpreted as a measure of goodness, indicating how well the estimator can identify the correct parameter from the parameter space. But this precision measure changes with parameterization, and it is less obvious how to extend precision (and variance) in the case of multiple parameters. For future research, it would be useful to investigate how well the estimator can identify the correct *model* from the *model space* (as opposed to *parameter* from the *parameter space*). For example, estimates of the quantiles themselves instead of the parameters would be more meaningful here. This would provide a "standardized" approach, allowing for a natural comparison in the case of different numbers of parameters and predictors, and would be independent of parameterization. It would allow for estimation on vines with more than two dimensions – a useful investigation indeed.

In extending the exploration to the case of multiple parameters and predictors, perhaps one would find the same result: that a transformation to a heavier-tailed response leads to less precise estimates. This is the case with the empirical univariate quantile estimator, which has variance proportional to the square of the slope of the true quantile function. This slope steepens in the tail with increasing tail heaviness. Some generalization of this is to be expected in the regression setting.

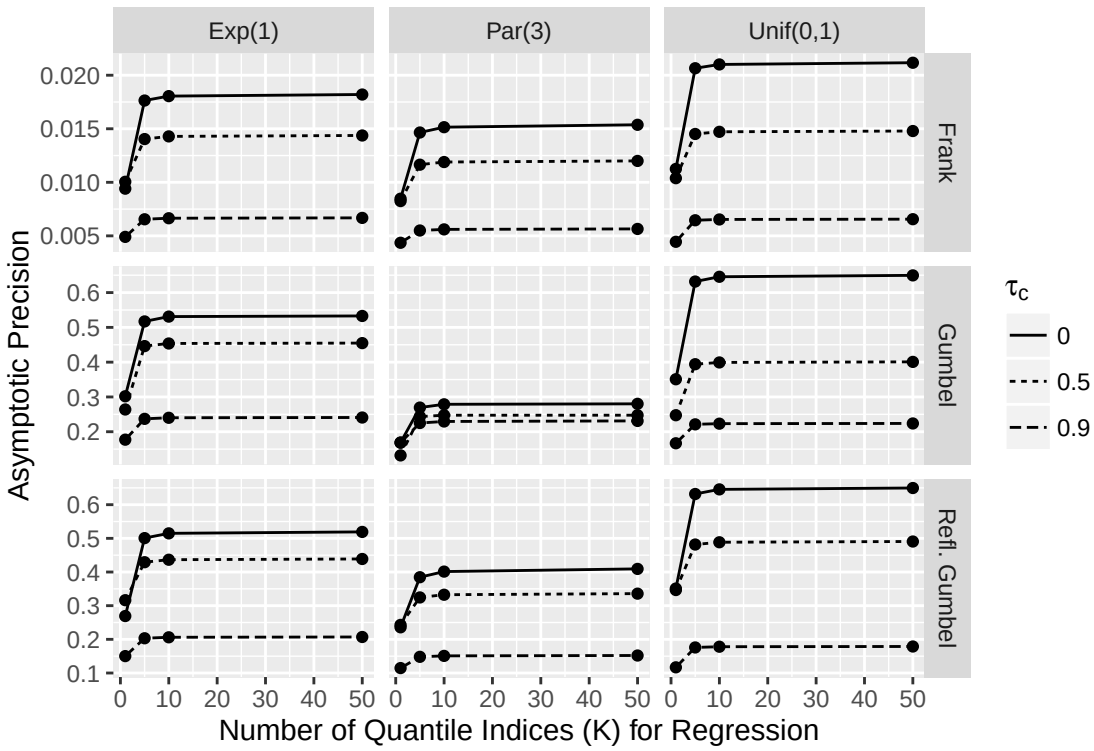

Figure 5.2: Asymptotic precision (inverse of the variance) of the family of CNQR estimators using K evenly spaced quantile levels calculated with Equation (5.2.2). The single predictor is linked to the response with the copula indicated in the side panels having a Kendall's tau of 0.3. The copulas were chosen to see whether tail dependence influences the asymptotic precision: the Frank copula has upper tail quadrant independence; the Gumbel copula has strong upper tail dependence; and the reflected Gumbel copula has intermediate tail dependence (similar to the Gaussian copula). A common transformation function $g_k = g$ is taken such that the transformed response $Y \mapsto g(Y)$ has marginal distributions indicated in the upper panels (the columns should therefore be interpreted as different CNQR estimators, not as different underlying distributions). Using approximately 10 quantile levels appears sufficient to achieve the converged ($K \to \infty$) precision. It appears that transforming the response to a heavy-tailed distribution may compromise the performance of the resulting CNQR estimator.

5.2.3 Special Cases

It is reassuring that the asymptotic covariance of a CNQR estimator in Equation (5.2.6) decomposes to the special case presented by the parallel linear case in Zou and Yuan (2008), and the single-quantile linear case in the seminal paper of Koenker and Bassett (1978), as the following remarks indicate.

5.2.3.1 Parallel Linear Quantile Surfaces

Suppose the quantile surfaces of Y conditional on the predictors are linear and parallel, so that

$$Q_{Y|\boldsymbol{X}}\left(\tau \mid \boldsymbol{x}\right) = \boldsymbol{x}^{\top}\boldsymbol{\theta}_0 + Q_{\varepsilon}\left(\tau\right) \tag{5.2.10}$$

for $\left(\tau, \boldsymbol{x}^\top\right) \in (0,1) \times \mathbb{R}^p$, where $\boldsymbol{\theta}_0 \in \mathbb{R}^p$, and $Q_\varepsilon : (0,1) \to \mathbb{R}$ is a strictly increasing function having $Q_\varepsilon(0.5) = 0$. The quantile function Q_ε describes an error distribution, determining the "spacing" of the parallel quantile surfaces.

Consider modelling these conditional quantiles with a family indexed by parameters $\boldsymbol{\theta} \in \mathbb{R}^p$ and functions $\hat{Q}_\varepsilon$ in the space of strictly increasing quantile functions with $\hat{Q}_\varepsilon(0.5) = 0$, by

$$\hat{Q}_{Y|\boldsymbol{X}}\left(\tau \mid \boldsymbol{x}; \hat{Q}_\varepsilon, \boldsymbol{\theta}\right) = \boldsymbol{x}^\top \boldsymbol{\theta} + \hat{Q}_\varepsilon(\tau)$$

for $\left(\tau, \boldsymbol{x}^\top\right) \in (0,1) \times \mathbb{R}^p$. Asymptotics for the CNQR estimator for $\boldsymbol{\theta}_0$ that uses K quantile levels $0 < \tau_1 < \cdots < \tau_K < 1$ and identity transformation functions g_k is considered by Zou and Yuan (2008, Theorem 2.1) (the estimator simultaneously estimates parameters $Q_\varepsilon(\tau_k)$ for $k = 1, \ldots, K$, but these are not considered). To compute the covariance in Equation (5.2.6), first notice that

$$f_{Y|\boldsymbol{X}}\left(y \mid \boldsymbol{x}\right) = f_\varepsilon\left(y - \boldsymbol{x}^\top \boldsymbol{\theta}_0\right)$$

for $\left(y, \boldsymbol{x}^\top\right) \in \mathbb{R}^{p+1}$, where f_ε is the density corresponding to the distribution with quantile function Q_ε. Now, for $k, k' = 1, \ldots, K$, we have $\dot{\boldsymbol{q}}_k(\boldsymbol{\theta}_0) = \boldsymbol{X}$, and using Equations (5.2.7) and (5.2.8),

$$\boldsymbol{D}_{0kk'} = \mathbb{E}\left(\boldsymbol{X}\boldsymbol{X}^\top\right) =: \boldsymbol{D}_0$$

and

$$\boldsymbol{D}_1 = \sum_{k=1}^{K} \mathbb{E}\left(f_\varepsilon\left(Q_\varepsilon(\tau_k)\right) \boldsymbol{X}\boldsymbol{X}^\top\right) = \boldsymbol{D}_0 \sum_{k=1}^{K} f_\varepsilon\left(Q_\varepsilon(\tau_k)\right).$$

The covariance becomes

$$\boldsymbol{\Sigma}_{\text{CNQR}} = \boldsymbol{D}_0^{-1} \frac{\sum_{k,k'=1}^{K} \min\left(\tau_k, \tau_{k'}\right)\left(1 - \max\left(\tau_k, \tau_{k'}\right)\right)}{\left[\sum_{k=1}^{K} f_\varepsilon\left(Q_\varepsilon(\tau_k)\right)\right]^2}, \tag{5.2.11}$$

which matches the result of Zou and Yuan (2008, Theorem 2.1).

5.2.3.2 Linear Quantile Surfaces

Suppose the quantile surfaces of Y conditional on the predictors are linear, so that

$$Q_{Y|\boldsymbol{X}}\left(\tau \mid \boldsymbol{x}\right) = \boldsymbol{x}^\top \boldsymbol{\theta}_0(\tau) \tag{5.2.12}$$

for $(\tau, \boldsymbol{x}) \in (0,1) \times \mathbb{R}^p$, where $\boldsymbol{\theta}_0 : (0,1) \to \mathbb{R}^p$ is a function such that $Q_{Y|\boldsymbol{X}}\left(\cdot \mid \boldsymbol{x}\right)$ is a valid quantile function for each $\boldsymbol{x}$. Fixing $\tau \in (0,1)$, a family of models for the τ-quantile

surface is

$$\hat{Q}_{Y|\boldsymbol{X}}\left(\tau \mid \boldsymbol{x}\right) = \boldsymbol{x}^\top \boldsymbol{\theta}$$

for $\boldsymbol{x} \in \mathbb{R}^p$, indexed by parameters $\boldsymbol{\theta} \in \mathbb{R}^p$. The CNQR estimator of $\boldsymbol{\theta}_0(\tau)$ using $K = 1$ quantile level $\tau_1 = \tau$ and identity transformation function g_1 reduces to the quantile regression estimator of Koenker and Bassett (1978). To compute the covariance in Equation (5.2.6), we have $\dot{\boldsymbol{q}}_1\left(\boldsymbol{\theta}_0(\tau)\right) = \boldsymbol{X}$, and using Equations (5.2.7) and (5.2.8),

$$\boldsymbol{D}_{011} = \mathbb{E}\left(\boldsymbol{X}\boldsymbol{X}^\top\right) =: \boldsymbol{D}_0$$

and

$$\boldsymbol{D}_1 = \mathbb{E}\left(f_{Y|\boldsymbol{X}}\left(\boldsymbol{X}^\top \boldsymbol{\theta}_0(\tau) \mid \boldsymbol{X}\right)\boldsymbol{X}\boldsymbol{X}^\top\right).$$

The covariance becomes

$$\boldsymbol{\Sigma}_{\text{CNQR}} = \tau(1-\tau)\boldsymbol{D}_1^{-1}\boldsymbol{D}_0\boldsymbol{D}_1^{-1}, \tag{5.2.13}$$

which matches the result of Koenker (2005, Theorem 4.1).

Although the estimation considered here is a special case of the estimation of the parallel linear model in Equation (5.2.10) with $K = 1$ quantile levels, the asymptotic covariances are not equal in general. That is, Equation (5.2.11) with $K = 1$ does not reduce to Equation (5.2.13) in general. This is because the quantity $\boldsymbol{D}_1$ depends on the true conditional quantile function in a neighbourhood of τ_1, which is more specific in the case of parallel quantile surfaces. In general, the $\boldsymbol{D}_1$ term can be reduced whenever $Q'_{Y|\boldsymbol{X}}\left(\tau \mid \boldsymbol{x}\right)$ does not depend on $\boldsymbol{x}$, because in general,

$$f_{Y|\boldsymbol{X}}\left(Q_{Y|\boldsymbol{X}}\left(\tau \mid \boldsymbol{x}\right) \mid \boldsymbol{x}\right) = \frac{1}{Q'_{Y|\boldsymbol{X}}\left(\tau \mid \boldsymbol{x}\right)}$$

for $(\tau, \boldsymbol{x}) \in (0,1) \times \mathbb{R}^p$.

5.3 Selecting a Model

An added complication to the proposed PCBN modelling method is that there are many candidate models that one could consider. Fitting m copula models to p predictors for all pairing orders results in $m^p p!$ possible forecasters. We are tasked with choosing the most appropriate forecaster – the one that generalizes best to new data.

One need not consider all $p!$ possible pairing orders. Instead, we recommend putting

the predictors that are most dependent with the response near the beginning of the pairing order (using any dependence measure). This is most desirable because error is propagated further down the pairing order.

With a pairing order selected, we propose a sequential method for selecting a forecaster. The method adds a predictor one at a time, each time choosing the optimal forecaster.

1. Select a model for Q_Y that optimizes the expected out-of-sample score. Set the predictor number $k = 1$.

2. Select candidate copula models for $C_{k,Y;1:(k-1)}$, fitting the corresponding model for $Q_{Y|1:k}$ for each – possibly re-estimating the parameters from previous steps to further optimize the score on the training set.

3. Choose the forecaster that has the best score on $Q_{Y|1:k}$ given in Equation (A.2.3). To avoid overfitting, consider estimating the scores using out-of-sample data, such as with cross validation or using a validation set; otherwise, one can use the training data to estimate the score.

4. Increase k by one and repeat Steps 2 and 3 until all predictors are considered.

As long as the independence copula is included as a candidate in Step 2—equivalent to leaving the k'th predictor out—then the estimated out-of-sample score will never worsen by adding a predictor. Of course, this does not guarantee that the true score does not worsen. This forecaster progression allows us to see the effect of adding a particular predictor (conditional on the already-included ones). Variable selection occurs here by removing predictors with the independence copula fitted.

Of course, candidate copula models should be motivated by an exploratory analysis. Normal scores plots of each predictor paired with the response will often give valuable insight as to what copula models to consider. In addition, the rough shape of the quantile curves in the normal scores plots can be investigated using the local fitting techniques discussed in Section 4.1, and compared to the quantile curves of copula families. For example, Bernard and Czado (2015) show that quantile curves that flatten out in the tails suggest tail independence. For a study on the quantile curves of copula families and their relationship to tail dependence, see Bernard and Czado (2015).

Chapter 6

New Bivariate Copula Families

The flexible PCBN/vine copula method of Chapter 5 uses a sequence of bivariate copula models having a variety of tail behaviour and CCEVI's. This variety is what gives the new approach its flexibility. Particularly, copulas having a non-constant CCEVI are important for capturing different tail behaviours of the upper quantile function forecasts across the predictor space.

However, an example of a copula having a non-constant CCEVI remains elusive – many (if not all) simple one- and two-parameter bivariate copulas in Joe (2014) have constant CCEVI's. This problem is compounded by the often non-trivial computations that are involved in finding the CCEVI of some copulas. In addition, fully tail-lightening copula families (i.e., those allowing a CCEVI of 0) remain elusive as well.

To solve this issue, a new copula family having a non-constant CCEVI is proposed in Section 6.1, and is called the *Integrated Gamma (IG)* copula family. An additional copula family, called the *Integrated Gamma Limit (IGL)* copula family, exists on a boundary of the IG copula family. The IGL family is an example of a family containing fully tail-lightening copulas, and is therefore described in some detail in this chapter as well.

The IG and IGL copula families have a natural non-parametric extension. The extension of the IG copula family is identified by Durante and Jaworski (2012). These extensions are discussed in Section 6.2.

For this chapter, it is especially pertinent to recall the convention that, if a function lacks a definition at a point in its domain, then it is defined as the limit at that point.

6.1 The IG and IGL Copula Families

This section introduces a copula family (the IG family) having non-constant CCEVI's, as well as a copula family (the IGL family) existing on a boundary of the IG family having

CCEVI's of zero. As discussed in Section 6.1.1, these families were discovered by finding the copula underlying a Type I Pareto conditional distribution having a linear shape parameter in the conditioning variable. This derivation can be safely skipped to obtain a formal definition of the families in Section 6.1.2. Some properties of the copula families are discussed in Section 6.1.3.

This section refrains from describing extensive properties of the IG and IGL copula families, because these families have a non-parametric generalization that is discussed in Section 6.2.

6.1.1 Beginnings

The approach taken to construct a bivariate copula family with a non-constant CCEVI is to specify a conditional distribution having an EVI that depends on the predictor. In particular, a response variable conditional on a uniform predictor is taken to have a Type I Pareto distribution with linear shape parameter. The corresponding copula can then be derived. Section 6.1.1.1 discusses the details.

The resulting copula family does not achieve high dependence, so a generalization of the family is sought in Section 6.1.1.2. A generalization is found after recognizing that the limit copula is related to the Gamma $(2, 1)$ distribution, and generalizing this to the Gamma $(k, 1)$ distribution.

6.1.1.1 Initial Derivation

Building a bivariate copula family with non-constant CCEVI can be done as follows.

1. Choose a distribution for a random variable Y conditional on a uniform predictor U, such that the conditional EVI is non-constant.

2. Obtain the joint distribution of (U, Y).

3. Obtain the marginal quantile function of Y.

4. Derive the underlying copula of (U, Y) using Sklar's theorem – specifically, Equation (2.4.1).

It is difficult to find a distribution in Step 1 so that the integral in Step 2 is tractable. If integration is tractable, the quantile function in Step 3 is unlikely to have a convenient closed form.

One workable example is to choose a Type I Pareto distribution in Step 1 having a shape parameter that is linear in the predictor:

$$F_{Y|U}\left(y \mid u\right) = 1 - y^{-\alpha(u)}$$

for $y \geq 1$ and $u \in (0,1)$, where $\alpha : (0,1) \to (0,\infty)$ is a positive linear function. The EVI of this distribution is $1/\alpha\left(u\right)$, which is non-constant in u.

To ensure the positivity of α, it is convenient to write

$$\alpha\left(u\right) = \nu\left(1 - \theta^{-} + \theta u\right)$$

for some $\theta \in \mathbb{R}$ and $\nu > 0$, where $\theta^{-} = \min\left(\theta, 0\right)$. With $\nu = 1$, α is the increasing line segment from the point $(0,1)$ to $(1, 1+\theta)$ when $\theta > 0$, and is the decreasing line segment from the point $\left(0, 1+|\theta|\right)$ to $(1,1)$ when $\theta < 0$. The positive ν parameter acts as an "amplifier" so that the space of positive linear functions with domain $(0,1)$ is spanned.

Next, the joint distribution of (U,Y) can be found rather easily by integration, after a change of variables to $t = \alpha\left(w\right)$:

$$\begin{aligned} F_{U,Y}\left(u,y\right) &= \int_0^u F_{Y|U}\left(y|w\right) \mathrm{d}w \\ &= u - \frac{1}{\nu\theta}\int_{\alpha(0)}^{\alpha(u)} y^{-t}\,\mathrm{d}t \\ &= u - y^{-\nu\left(1-\theta^{-}\right)}\frac{1 - y^{-\nu\theta u}}{\nu\theta\log y} \end{aligned}$$

for $(u,y) \in (0,1) \times [1,\infty)$. It will be more convenient to write this joint distribution as

$$F_{U,Y}\left(u,y\right) = u\left(1 - y^{\nu\theta^{-}} H_2\left(y^{\nu};\theta u\right)\right), \tag{6.1.1}$$

where

$$H_2\left(t;\eta\right) = \frac{1}{t}\Psi_2\left(\frac{1}{\eta\log t}\right) \tag{6.1.2}$$

for $(t,\eta) \in [1,\infty) \times (0,\infty)$, and

$$\Psi_2\left(t\right) = \frac{1 - \exp\left(-1/t\right)}{1/t} \tag{6.1.3}$$

for $t \geq 0$ (the motivation for using the subscript "2" will become clear in the following sections).

The marginal distribution function of Y can be found by

$$F_Y(y) = \lim_{u \uparrow 1} F_{U,Y}(u, y) = 1 - y^{\nu\theta^-} H_2(y^\nu; \theta)$$

for $y \geq 1$. But notice that

$$\begin{aligned} t^{\eta_-} H_2(t; \eta) &= t^{\eta_-} \frac{1}{t} \frac{1 - t^{-\eta}}{\eta \log t} \\ &= \frac{1}{t} \frac{t^{\eta_-} - t^{-\eta_+}}{\eta \log t} \\ &= \frac{1}{t} \frac{1 - t^{-|\eta|}}{|\eta| \log t} \\ &= H_2\left(t; |\eta|\right) \end{aligned} \tag{6.1.4}$$

for $(t, \eta) \in [1, \infty) \times (0, \infty)$, where $\eta_+ = \max(0, \eta)$, so that

$$F_Y(y) = 1 - H_2\left(y^\nu; |\theta|\right).$$

The quantile function is therefore

$$Q_Y(\tau) = \left[H_2^{\leftarrow}\left(1 - \tau; |\theta|\right)\right]^{1/\nu}$$

for $\tau \in (0, 1)$, where $H_2^{\leftarrow}(\cdot; \eta)$ is the unique inverse function of $H_2(\cdot; \eta)$ for each $\eta > 0$.

Lastly, the copula underlying (U, Y) can be found by Sklar's theorem, using Equation (2.4.1):

$$C_{U,Y}(u, v; \theta) = F_{U,Y}\left(u, Q_Y(v)\right) = u - u\left[H_2^{\leftarrow}\left(1 - v; |\theta|\right)\right]^{\theta^-} H_2\left(H_2^{\leftarrow}\left(1 - v; |\theta|\right); \theta u\right).$$

The parameter ν cancels out, leaving only the θ parameter.

Practically, we need only consider $\theta > 0$, because changing the sign of θ only results in a 1-reflection of the copula. This can be shown by proving $\mathrm{R}_1 C_{U,Y}(u, v; -\theta) = C_{U,Y}(u, v; \theta)$ for any $\theta \neq 0$. Using the formula of a 1-reflection in Equation (A.1.3), and letting $h = H_2^{\leftarrow}\left(1 - v; |\theta|\right)$, we have

$$\begin{aligned} \mathrm{R}_1 C_{U,Y}(u, v; -\theta) &= v - C_{U,Y}(1 - u, v; -\theta) \\ &= u - (1 - v) + (1 - u)\, h^{(-\theta)_-} H_2\left(h; -\theta(1 - u)\right). \end{aligned}$$

Now, notice that $(-\theta)^- = -\theta + \theta^-$, and use the identity in Equation (6.1.4) to obtain

$$1 - v = H_2\left(h; |\theta|\right) = h^{\theta^-} H_2\left(h; \theta\right).$$

Making these substitutions,

$$\begin{aligned}
\mathrm{R}_1 C_{U,Y}\left(u, v; -\theta\right) &= u - h^{\theta^-}\left[H_2\left(h; \theta\right) - \left(1-u\right) h^{-\theta} H_2\left(h; -\theta\left(1-u\right)\right)\right] \\
&= u - u h^{\theta^-} H_2\left(h; \theta u\right) \\
&= C_{U,Y}\left(u, v; \theta\right),
\end{aligned}$$

after some routine algebra to obtain the second line, by substituting the expression for H_2 found in Equation (6.1.2).

Taking $\theta > 0$, then, we obtain the family conveniently represented by the 2-reflection

$$\mathrm{R}_2 C_{U,Y}\left(u, v; \theta\right) = u - C_{U,Y}\left(u, 1-v; \theta\right) = u H_2\left(H_2^{\leftarrow}\left(v; \theta\right); \theta u\right) \tag{6.1.5}$$

for $(u, v) \in (0, 1)^2$.

6.1.1.2 Extension

One question to ask when confronted with a copula class is whether a full range of dependence is included (i.e., up to perfect positive and negative dependence). Empirical investigations of the copula family in Equation (6.1.5) have shown that dependence increases as θ increases, so the range of dependence can be investigated through the θ limits. It can be shown that the $\theta \downarrow 0$ limit of the copula family in Equation (6.1.5) is the independence copula, but that the $\theta \to \infty$ limit does not result in the comonotonicity copula (this is shown in Proposition 6.2.10 in a later section). Instead, we obtain

$$\lim_{\theta \to \infty} \mathrm{R}_2 C_{U,Y}\left(u, v; \theta\right) = u \Psi_2\left(u^{-1} \Psi_2^{\leftarrow}\left(v\right)\right) \tag{6.1.6}$$

for $(u, v) \in (0, 1)^2$, where Ψ_2 is defined in Equation (6.1.3). This copula has a low Kendall's tau of approximately 0.39, so it is of interest to extend this copula so that a larger dependence is reached.

The limit copula in Equation (6.1.6) belongs to a larger class of copulas described by Durante and Jaworski (2012) that generalizes the function Ψ_2. This class is described in a later section in Equation (6.2.1), but requires that $\Psi_2 : [0, \infty) \to [0, 1]$ be a concave distribution function (or convex survival function). Equivalently, Ψ_2' should be a

decreasing density function on $[0, \infty)$.

For Ψ_2 defined in Equation (6.1.3), the derivative is

$$\Psi_2'(t) = 1 - \left(\frac{1}{t} + 1\right) \exp\left(-\frac{1}{t}\right)$$

for $t \geq 0$, for which $\Psi_2'\left(1/t\right)$ is the Gamma $(2, 1)$ distribution function. Since the integral of a Gamma distribution function over $(0, \infty)$ exists, $\Psi_2'\left(1/t\right)$ can be replaced by the normalized Gamma $(k, 1)$ distribution function for $k > 1$ (the scale parameter of the Gamma distribution does not affect the results). Doing so results in a family of functions Ψ_k that can replace Ψ_2 in general, and defines a class of copulas called the Integrated Gamma Limit (IGL) family of copulas. The functions Ψ_k are defined in Equation (6.1.7).

It is shown in a later section that the IGL copula family has the comonotonicity copula occur as $k \to \infty$, so that the full range of dependence can now be reached.

With the limit ($\theta \to \infty$) copula in Equation (6.1.6) generalized, it is interesting to check whether an interior ($\theta \in (0, \infty)$) copula can be generalized by replacing H_2 in Equation (6.1.2) with "H_k", which replaces Ψ_2 with Ψ_k. This generalization is possible, and results in the Integrated Gamma (IG) family of copulas.

6.1.2 Definition

The IG family of bivariate two-parameter copulas, derived in Section 6.1.1, is defined here. Its boundary case – the IGL copula family – is also defined here.

The following families of functions for $k > 1$ are central to the IG and IGL copula families:

$$\Psi_k(y) = y\frac{\Gamma(k) - \Gamma^*\left(k, y^{-1}\right)}{\Gamma(k-1)} + \frac{\Gamma^*\left(k-1, y^{-1}\right)}{\Gamma(k-1)} \tag{6.1.7}$$

for $y \in [0, \infty)$, where Γ and Γ^* are (respectively) the gamma and (upper) incomplete gamma functions, and

$$H_k(y; \eta) = \frac{1}{y}\Psi_k\left(\frac{1}{\eta \log y}\right) \tag{6.1.8}$$

for $(y, \eta) \in [1, \infty) \times (0, \infty)$. Plots of some of these functions are shown in Figures 6.1 (for Ψ_k) and 6.2 (for H_k). It is important to recognize that each Ψ_k is a concave distribution function, and each $H_k(\cdot; \eta)$ is a survival function. Both are strictly monotone, so that the inverses (in the first argument) exist. Properties of these functions are examined more rigorously in Appendices E.3.1 and E.4.1.

Note that the IG and IGL copula families have a non-parametric generalization by generalizing the Ψ_k functions, as discussed in Section 6.2. As such, proofs related to

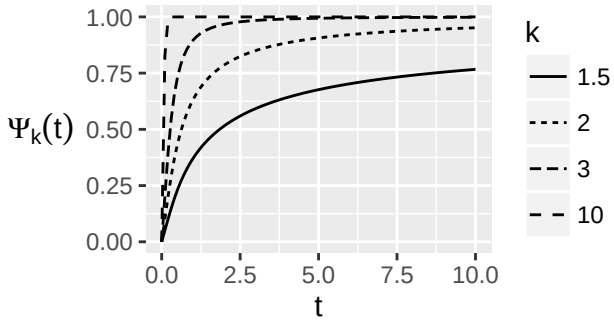

Figure 6.1: Some generating functions Ψ_k, described in Equation (6.1.7).

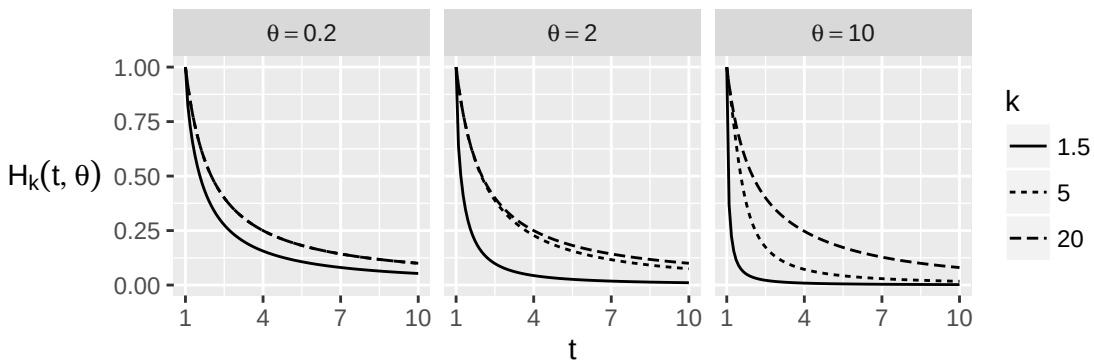

Figure 6.2: Some generating functions $H_k(\cdot;\theta)$, described in Equation (6.1.8).

the IG and IGL copula families, outlined in Appendices E.4 and E.3, rely on results regarding the non-parametric generalizations. These appendices therefore appear after the appendices for the non-parametric generalizations.

Theorem 6.1.1. *For each $(k,\theta) \in (1,\infty) \times (0,\infty)$, the function $H_k(\cdot;\theta)$ in Equation (6.1.8) is strictly decreasing and has unique inverse $H_k^{\leftarrow}(\cdot;\theta)$, and*

$$C_{IG}(u,v;k,\theta) = u + v - 1 + (1-u)\, H_k\left(H_k^{\leftarrow}(1-v;\theta);\theta(1-u)\right) \tag{6.1.9}$$

with $(u,v) \in (0,1)^2$ is a copula.

A proof follows from a further generalization introduced in Section 6.2 (through a stochastic representation given in Proposition 6.2.11).

Definition 6.1.1. *The copula family described by Equation (6.1.9) for parameters $(k,\theta) \in (1,\infty) \times (0,\infty)$ is called the* Integrated Gamma (IG) *copula family, having members $IG(k,\theta)$. The reflection copula is*

$$\hat{C}_{IG}(u,v;k,\theta) = uH_k\left(H_k^{\leftarrow}(v;\theta);\theta u\right).$$

The name "Integrated Gamma" comes from the fact that $\Psi_k'\left(1/y\right)$ is proportional to the Gamma$(k,1)$ distribution function for $y \geq 0$ (see Equation (E.3.1)), and the "IG" notation should not be confused with the "*incomplete* gamma" function. For formulas related to the IG copula family, see Appendix E.4.2.

The independence copula is achieved when $\theta \downarrow 0$, whereas the IGL copula family results when $\theta \to \infty$ (an investigation of concordance is left for future research).

Theorem 6.1.2. *For any $k > 1$, the IG copula family described in Equation (6.1.9) has*

$$\lim_{\theta \downarrow 0} C_{IG}\left(u, v; k, \theta\right) = C^{\perp}\left(u, v\right) = uv \tag{6.1.10}$$

and

$$\lim_{\theta \to \infty} C_{IG}\left(u, v; k, \theta\right) = C_{IGL}\left(u, v; k\right) \tag{6.1.11}$$

for $(u, v) \in (0,1)^2$, where

$$C_{IGL}\left(u, v; k\right) = u + v - 1 + \left(1 - u\right)\Psi_k\left(\left(1-u\right)^{-1}\Psi_k^{\leftarrow}\left(1 - v\right)\right), \tag{6.1.12}$$

and Ψ_k is defined in Equation (6.1.7). Furthermore, $(u, v) \mapsto C_{IGL}\left(u, v; k\right)$ is a copula.

This theorem is given more generally in Section 6.2, and is therefore not directly proven. The limit results are given in Proposition 6.2.10, and a stochastic representation of a generalized copula is given in Proposition 6.2.4.

Definition 6.1.2. *The copula family described by Equation (6.1.12) for parameter $k > 1$ is called the* Integrated Gamma Limit (IGL) *copula family, having members IGL (k). The reflection copula is*

$$\hat{C}_{IGL}\left(u, v; k\right) = u\Psi_k\left(u^{-1}\Psi_k\left(v\right)\right).$$

For formulas related to the IGL copula family, see Appendix E.3.2.

There may exist a natural multivariate extension to the IG and IGL copula families. One such extension might be achieved by adding to the definition of $H_k\left(t; \eta\right)$ in Equation (6.1.8) by multiplying by $p > 1$ copies of the Ψ_k function, to obtain

$$H_k^{(p)}\left(t; \eta_1, \ldots, \eta_d\right) = \frac{1}{t}\prod_{j=1}^{p}\Psi_k\left(\frac{1}{\eta_j \log t}\right)$$

for $t > 1$ and $\left(\eta_1, \ldots, \eta_p\right) \in (0, \infty)^p$. Taking $k = 2$, one can show that

$$\left(u_1, \ldots, u_p, v\right) \mapsto u_1 \cdots u_p\left(1 - H_2^{(p)}\left(H_2^{(p)\leftarrow}\left(1 - v; \theta_1, \ldots, \theta_p\right); \theta_1 u_1, \ldots, \theta_p u_p\right)\right)$$

with $\left(\theta_1, \ldots, \theta_p\right) \in (0, \infty)^p$ is the copula resulting from the following stochastic representation: $U_j \sim \text{Unif}\,(0,1)$ for $j = 1, \ldots, p$ are iid, and

$$Y \mid \left(U_1 = u_1, \ldots, U_p = u_p\right) \sim \text{Par}\left(1, 1 + \theta_1 u_1 + \cdots + \theta_p u_p\right).$$

However, it is questionable whether such copula holds any value in practice, since the predictor variables are independent.

Another option to generalize the IG and IGL copula families is to consider the following stochastic representation: $U \sim \text{Unif}\,(0,1)$, and iid variables $Y_j \mid U = u$ for $j = 1, \ldots, p$ for some $p > 1$ having standard Type I Pareto distributions with shape parameter linear in u. But the joint distribution of $\left(U, Y_1, \ldots, Y_p\right)$ involves a cumbersome binomial expansion. Yet another option is to specify a stochastic representation having a hierarchical form. For example, with three variables, one can take $U \sim \text{Unif}\,(0,1)$,

$$Y_1 \mid U = u \sim \text{Par}\left(1, \nu_1\left(1 + \theta_1 u\right)\right)$$

for $\nu_1, \theta_1 > 0$ and $u \in (0,1)$, and

$$Y_2 \mid \left(Y_1 = y_1, U = u\right) \sim \text{Par}\left(1, \nu_2\left(1 + \theta_2 F_{Y_1|U}\left(y_1 \mid u\right)\right)\right)$$

for $\nu_2, \theta_2 > 0$ and $y_1 > 1$. The joint distribution of (U, Y_1, Y_2) is that of (U, Y_1) (derived in Section 6.1.1.1), plus another term involving elliptic integrals. Overall, these ideas are mathematically cumbersome, and possibly intractable.

6.1.3 Properties

Some properties of the IG(k, θ) and IGL(k) copula families are discussed here. First, normal scores plots of simulated data from some of these copulas are explored in Figure 6.3. The following observations can be made *ab initio*.

- IG and IGL copulas are not permutation symmetric. This means that switching the horizontal and vertical variables results in a different copula, and can be seen by the asymmetry about the $y = x$ line.

- Overall, larger values of θ and k allow for stronger dependence, particularly when $\theta \gg k$ are both large. The comonotonicity copula results as $\theta \to \infty$ faster than $k \to \infty$. Independence results as $\theta \downarrow 0$. For a plot of the dependence in terms of Kendall's tau, see Figure 6.4.

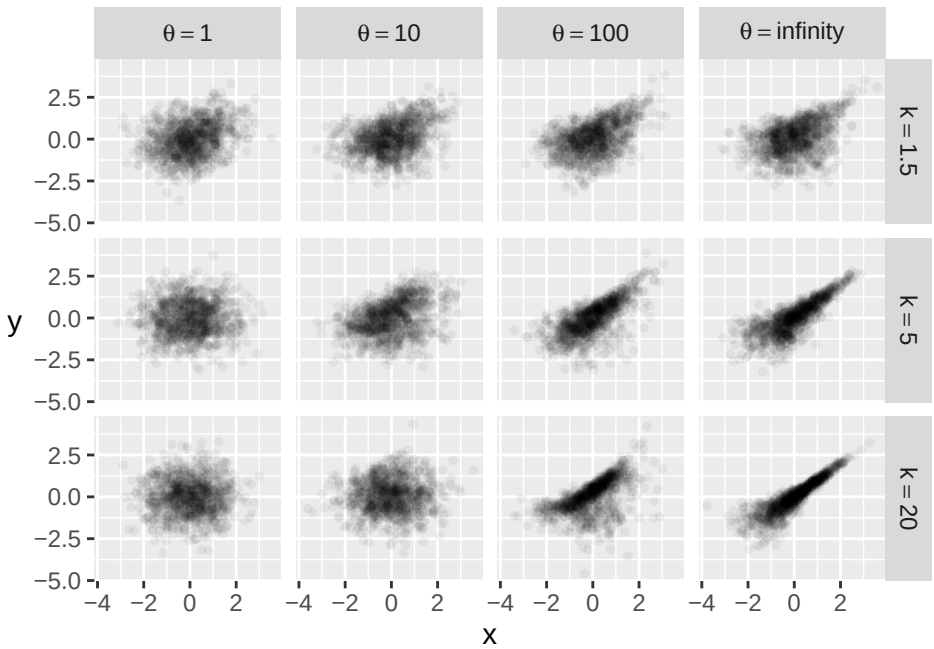

Figure 6.3: Normal scores plots of simulated data from members of the IG (θ, k) and IGL (k) copula families, defined respectively in Definitions 6.1.1 and 6.1.2. The marginal distributions are standard Gaussian, and the dependence is described by an IG or IGL copula. Each panel displays 1000 randomly generated observations.

- When $\theta < k$, the copulas appear to be close to independent
- There is no tail dependence, except in the upper tails of the IGL copulas, in which case a larger k results in a larger tail dependence coefficient.

Next, notice that as θ and k grow comparably, the resulting copula is *not* the comonotonicity copula. See the normal scores plots in Figure 6.5. The limiting copula as both θ and k increase appears to have a "hollowed out" upper-left corner with the formation of a "ridge".

Some of these observations will be proven. From these observations, it appears that the position of θ relative to k is a meaningful parameter. Additionally, the family may not show concordance in θ (an investigation is left for future research). It is left to future research to find an appropriate reparameterization – for example, possibly including the ratio θ/k.

The IG and IGL families indeed achieve the comonotonicity and independence copulas as bounds.

Proposition 6.1.3. *For* $(u, v) \in (0, 1)^2$,

$$C_{IGL}(u, v; k) \to C^+(u, v)$$

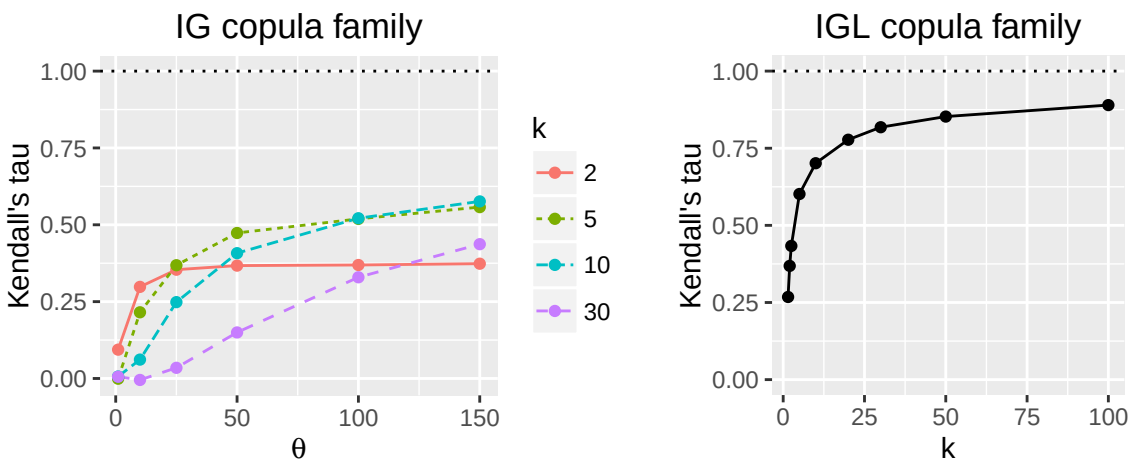

Figure 6.4: Kendall's tau dependence of the IG and IGL copula families, estimated by a random sample of 10,000 observations.

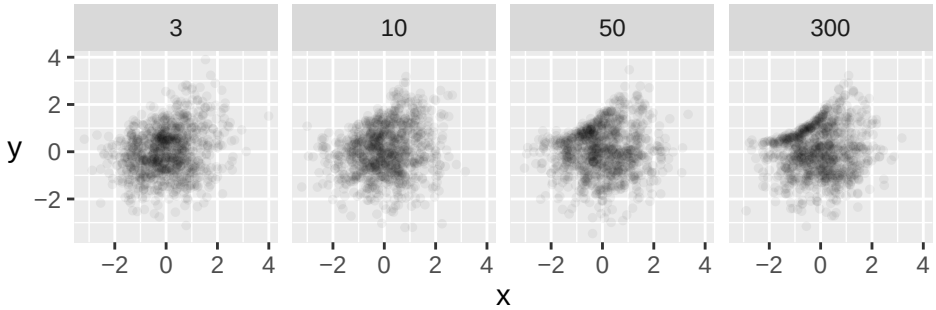

Figure 6.5: Normal scores plots of simulated data from the IG copula family (Definition 6.1.1) with θ and k parameters being equal, indicated in the panel labels. The marginal distributions are standard Gaussian, and the dependence is described by an IG copula.

as $k \to \infty$*, where* $C^{+}(u, v) = \min(u, v)$ *is the comonotonicity copula.*

A proof can be found in Appendix E.3.3. For future research, an extension of this result might hold for the IG copula family as both $\theta \to \infty$ and $k \to \infty$, with θ increasing faster so that $\theta/k \to \infty$.

The IG family has no tail dependence, but the IGL family does. An investigation of tail order is left for future research.

Proposition 6.1.4 (IG tail dependence)**.** *The IG copula family* C_{IG}*, defined in Equation (6.1.9), has zero lower tail dependence coefficient whenever* $k \geq 2$*, and zero upper tail dependence coefficient for all* $k > 1$*.*

A proof can be found in Appendix E.4.3. The case with $k \in (1, 2)$ is left for future research.

Proposition 6.1.5 (IGL tail dependence)**.** *The IGL copula family C_{IGL}, defined in Equation (6.1.12) has zero lower tail dependence coefficient and an upper tail dependence coefficient of*

$$\lambda_U^{(IGL)} = 1 - \frac{\left[(k-1)\,e^{-1}\right]^{k-1}}{\Gamma(k)}. \tag{6.1.13}$$

A proof can be found in Appendix E.3.3.

Of course, the primary interest for describing the IG and IGL families is their CCEVI's.

Proposition 6.1.6. *The IG copula with $\theta > 0$ and $k > 1$, defined in Definition 6.1.1, has CCEVI*

$$\xi_{IG} : u \mapsto \frac{1}{1 + \theta\,(1-u)}, \tag{6.1.14}$$

which is an increasing function.

A proof can be found in Appendix E.4.3.

Proposition 6.1.7. *Any IGL copula, defined in Equation (6.1.12), has a CCEVI of 0.*

A proof can be found in Appendix E.3.3.

The IG copula family has an increasing CCEVI, whose reciprocal is linear in the first copula variable. Modelling an increasing CCEVI is useful for capturing an increasing level of "risk" of an extreme response given larger values of a predictor. In particular, the CCEVI of an IG copula sharply increases towards 1 near $u = 1$, meaning that occurrences of extreme values of the predictor are mostly responsible for an extreme response.

The IGL copula family has a CCEVI of 0, so that conditioning a heavy-tailed response on a predictor would result in a light-tailed conditional distribution. A predictor displaying this phenomenon in practice is very useful, since the tail heaviness of a response variable is entirely described by the predictor. An IGL copula can model such phenomenon.

Some quantile curves of the IG and IGL copula families are shown in Figure 6.6. The zero tail dependence of the IG copula family is manifest in the flattening of its quantile curves for increasing values of the predictor variable. In contrast, the IGL copula has increasing quantile curves. The IGL copula also has upper quantile curves that are close together, and this is possibly a testament to its fully tail-lightening property "compressing" the marginal tail. Interestingly, near-linear quantile curves result from the IGL copula under exponential marginals.

As indicated throughout this section, the IG and IGL copula families have a non-parametric extension by generalizing the function Ψ_k. This extension is explored in the following section.

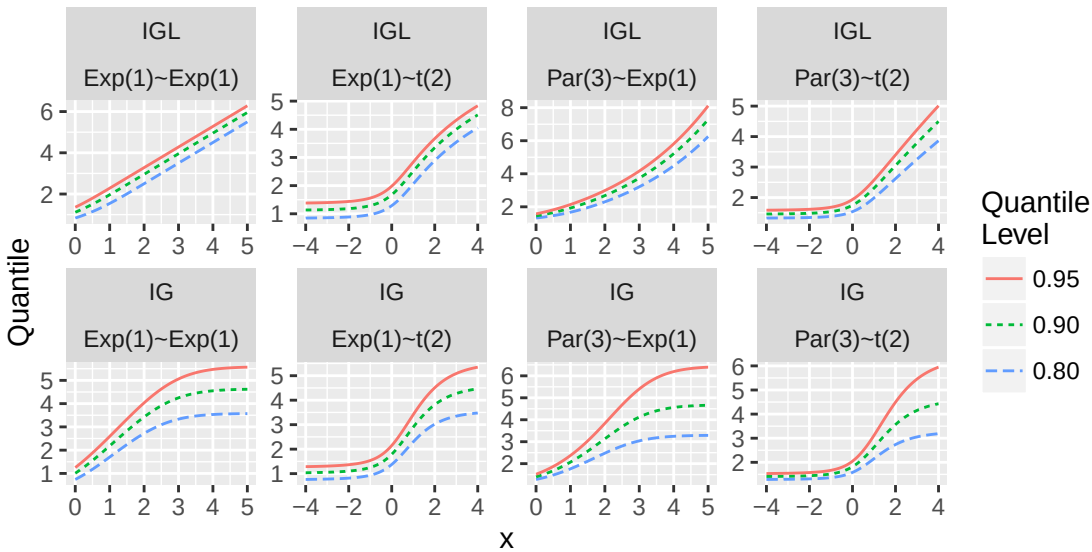

Figure 6.6: Examples of quantile curves of bivariate distributions described by the IGL(1.6) and IG(8.9, 3) copulas, defined respectively in Equations (6.1.12) and (6.1.9). Both copulas are chosen to have a Kendall's tau of approximately 0.3. The marginal distributions are indicated in the form "response distribution" ~ "predictor distribution".

6.2 Non-parametric Extensions

The IG copula family is defined in Section 6.1.2, and indexed by $(k, \theta) \in (1, \infty) \times (0, \infty)$. Upon deriving the IG family in Section 6.1.1, there is no knowledge of the k parameter – only $k = 2$ is inadvertently used. This narrower family suffers from a poor range of dependence. Generalizing to the IG copula family by allowing Ψ_2 to be any one of Ψ_k for $k > 1$, defined in Equation (6.1.7), adds the flexibility of a full range of dependence.

But the IG copula family is not the only generalization possible. In fact, replacing Ψ_2 in an IG copula to be any function ψ that describes a valid copula leads to a much larger class of copulas. The $\theta \to \infty$ version of this class, corresponding to the IGL copula family, is investigated by Durante and Jaworski (2012), and is referred to as the *DJ copula class* in this dissertation. The finite-θ version can be thought of as an extension to the DJ copula class, and is therefore referred to as the *extended DJ (extDJ) copula class*.

In this section, properties of these large copula classes are discussed to open up the possibility of describing other parametric families – especially those with non-constant CCEVI, like the IG copula family. Indeed, another such family is found, and is identified briefly.

Section 6.2.1 discusses the DJ copula class and its properties, and Section 6.2.2 discusses the extDJ copula class.

6.2.1 Durante and Jaworski Copula Class

The DJ copula class is indexed by functions referred to in this dissertation as *generating functions*. These generating functions are identified to have a type of "parity", and also form "clusters" due to an invariance property. This forms a discussion on defining the DJ copula class that is held in Section 6.2.1.1. Some properties of the DJ copula class are discussed in Section 6.2.1.2.

6.2.1.1 Definition

Durante and Jaworski (2012, Equation (2)) describe a class[1] of bivariate copulas that contains the IGL family as a special case. This *DJ copula class* can be defined by

$$C_{\mathrm{DJ}}(u, v; \psi) = u\psi\left(u^{-1}\psi^{\leftarrow}(v)\right) \tag{6.2.1}$$

for $(u, v) \in (0, 1)^2$, with *DJ generating function* or *DJ generator* $\psi : [0, \infty) \to (0, 1)$ that is concave and non-decreasing (or convex and non-increasing) with inverse $\psi^{\leftarrow}$. The DJ copula with generator ψ is said to be *generated* by ψ. Formulas related to the DJ copula class can be found in Appendix E.1.2. The DJ class is investigated here in additional detail to that described by Durante and Jaworski (2012).

Notice that the IGL copula family, defined in Equation (6.1.12), has reflection copulas that belong to the DJ copula class. The reflected IGL family results by restricting the space of DJ generating functions to Ψ_k for $k > 1$, defined in Equation (E.3.1).

There is a natural bipartition of the DJ generators into distribution functions (increasing) and survival functions (decreasing). In fact, these subclasses are naturally paired element-wise, connected by a vertical reflection of the generated copulas.

Proposition 6.2.1. *If ψ is a DJ generator that is a distribution function (resp. survival function), then $\tilde{\psi} = 1 - \psi$ is a survival function (resp. distribution function) and*

$$C_{DJ}\left(u, v; \tilde{\psi}\right) = R_2 C_{DJ}(u, v; \psi)$$

for $(u, v) \in (0, 1)^2$.

The proposition can be proved by a direct application of Equation (A.1.4), and is therefore omitted.

[1] Their motivation for describing this class is to characterize copulas that have the "invariance under univariate truncation" property. We omit details.

Not only does Proposition 6.2.1 identify that a valid distribution function and corresponding survival function generate copulas that are related by vertical reflection, but it also identifies closure of the DJ copula class under vertical reflection.

The definition of the DJ copula class in Equation (6.2.1) is not the only way the class can be defined.

Proposition 6.2.2. *Fix $\alpha \in \mathbb{R}\backslash\{0\}$. Let $\psi : [0,\infty) \to [0,1]$ be such that $(u,v) \mapsto C_{DJ}(u,v;\psi)$, defined in Equation (6.2.1), is a copula. Then*

$$C_{DJ}(u,v;\psi) = C_{DJ}^{(\alpha)}\left(u,v;\psi \circ g_{-1/\alpha}\right)$$

for all $(u,v) \in (0,1)^2$, where $g_{-1/\alpha}(t) = at^{-1/\alpha}$ for $t \in [0,\infty)$, $a > 0$, and

$$C_{DJ}^{(\alpha)}(u,v;\psi_\alpha) = u\psi_\alpha\left(u^\alpha \psi_\alpha^{\leftarrow}(v)\right) \tag{6.2.2}$$

for some $\psi_\alpha : [0,\infty) \to [0,1]$. The contrary is also true: let $\psi_\alpha : (0,\infty) \to [0,1)$ be such that $(u,v) \mapsto C_{DJ}^{(\alpha)}(u,v;\psi_\alpha)$ is a copula. Then

$$C_{DJ}^{(\alpha)}(u,v;\psi_\alpha) = C_{DJ}(u,v;\psi_\alpha \circ g_{-\alpha})$$

for all $(u,v) \in (0,1)^2$, where $g_{-\alpha}(t) = bt^{-\alpha}$ for $t \in [0,\infty)$, $b > 0$.

It follows by taking common values of α that a DJ generator does not generate a unique DJ copula, because the generators are invariant to a change in scale.

Corollary 6.2.3 (Invariance to scale changes)**.** *Suppose that $(u,v) \mapsto C_{DJ}^{(\alpha)}(u,v;\psi)$, defined in Equation (6.2.2), is a copula for some real $\alpha \neq 0$ and generating function $\psi : [0,\infty) \to [0,1]$. Then if $g(t) = \alpha t$ for any $\alpha > 0$, $C_{DJ}^{(\alpha)}(u,v;\psi) = C_{DJ}^{(\alpha)}(u,v;\psi \circ g)$.*

This class of copulas has a convenient stochastic representation.

Proposition 6.2.4 (Stochastic representation)**.** *Let $\psi : [0,\infty) \to [0,1]$ be a concave distribution function with density ψ'. Consider random variables Y and Z with $F_Y = \psi$, and for each $y \in [0,\infty)$,*

$$F_{Z|Y}\left(z \mid y\right) = 1 - \frac{\psi'(z)}{\psi'(y)} \tag{6.2.3}$$

for $z \geq y$. Letting $U = Y/Z$, the joint distribution function of (U,Y) is

$$F_{UY}(u,y) = u\psi\left(u^{-1}y\right) \tag{6.2.4}$$

for $(u, y) \in (0, 1) \times [0, \infty)$*, and the corresponding copula is the DJ copula in Equation (6.2.1) generated by* ψ*. The copula describing the dependence of* $\left(U, 1/Y\right)$ *is the DJ copula generated by* $1 - \psi$*, a convex survival function.*

Proofs of these results can be found in Appendix E.1.3.

6.2.1.2 Properties

A DJ copula generated by a distribution function has positive dependence, whereas as DJ copula generated by a survival function has negative dependence.

Proposition 6.2.5 (Quadrant dependence)**.** *The DJ copula generated by an increasing DJ generating function* ψ *is positive quadrant dependent; that is,*

$$C_{DJ}(u, v; \psi) \geq uv$$

for $(u, v) \in (0, 1)^2$*, where* C_{DJ} *is defined in Equation (6.2.1). The DJ copula generated by a decreasing generating function* $1 - \psi$ *is negative quadrant dependent; that is,*

$$C_{DJ}(u, v; 1 - \psi) \leq uv$$

for $(u, v) \in (0, 1)^2$*.*

This is proved by Durante et al. (2011, Proposition 2.1).

The comonotonicity and countermonotonicity copulas belong to the DJ copula class, and the independence copula arises as a limiting case.

Proposition 6.2.6 (Comonotonicity/Countermonotonicity)**.** *The DJ copula in Equation (6.2.1) with generating function* $\psi(y) = \min(y, 1)$ *for* $y \in [0, \infty)$ *and inverse* $\psi^{\leftarrow}(\tau) = \tau$ *for* $\tau \in [0, 1]$ *is the comonotonicity copula* C^{+}*. Likewise, the generating function* $\psi(y) = \max(1 - y, 0)$ *for* $y \in [0, \infty)$ *and inverse* $\psi^{\leftarrow}(\tau) = 1 - \tau$ *for* $\tau \in [0, 1]$ *yields the countermonotonicity copula* C^{-}*.*

This is proved by Jaworski (2015, Proposition 8).

Proposition 6.2.7 (Independence Limit)**.** *Consider the subclass of DJ copulas* $(u, v) \mapsto C_{DJ}(u, v; \psi_\theta)$ *described in Equation (6.2.1) with generating functions indexed by* $\theta \in (0, 1)$ *given by* $\psi_\theta(y) = \min\left(y^\theta, 1\right)$ *for* $y \in [0, \infty)$ *with inverses* $\psi_\theta^{\leftarrow}(\tau) = \tau^{1/\theta}$*. Then the independence copula arises as the* $\theta \downarrow 0$ *limit:*

$$\lim_{\theta \downarrow 0} C_{DJ}(u, v; \psi_\theta) = C^{\perp}(u, v)$$

for all $(u, v) \in (0, 1)^2$.

Some members of this class have non-constant CCEVI. We give an example where the DJ generating function is a Weibull distribution function.

Proposition 6.2.8. *Let* $F_\beta(y) = 1 - \exp\left(-y^{1/\beta}\right)$ *for* $\beta \geq 1$ *and* $y \geq 0$ *be a Weibull distribution function, with decreasing density. The copula family defined by* $C_{DJ}\left(u, v; 1 - F_\beta\right)$ *is referred to as the* DJ Weibull *copula family in this dissertation. These copulas have CCEVI of 1, and the reflected copulas have CCEVI* $u \mapsto u^{1/\beta}$.

Proofs of these results can be found in Appendix E.1.3. It is left for future research to discover other properties of the DJ copula class.

6.2.2 An Extended Class

It is natural to ask whether an extension of the DJ copula class be constructed, similar to how the IG copula family in Equation (6.1.9) can be thought of as an extension to the IGL copula family in Equation (6.1.12) by linking their generating functions via Equation (6.1.8).

Theorem 6.2.9. *Let* $\psi : [0, \infty) \to [0, 1]$ *be a concave distribution function, and* $\eta > 0$. *Define* $H_\psi(\cdot; \eta) : [1, \infty) \to [0, 1]$ *by generalizing Equation (6.1.8):*

$$H_\psi(t; \eta) = \frac{1}{t}\psi\left(\frac{1}{\eta \log t}\right). \tag{6.2.5}$$

Then, $H_\psi(\cdot; \eta)$ *is monotone decreasing, and*

$$C_{extDJ}(u, v; \psi, \theta) = uH_\psi\left(H_\psi^{\leftarrow}(v; \theta); \theta u\right) \tag{6.2.6}$$

for $(u, v) \in (0, 1)^2$ *and* $\theta > 0$ *is a copula defining an* extended DJ (extDJ) *copula class, where* $H_\psi^{\leftarrow}(\cdot; \theta)$ *is the inverse of* $H_\psi(\cdot; \theta)$.

A proof is given through a stochastic representation in Proposition 6.2.11. For formulas related to the extDJ copula class, see Appendix E.2.2.

The pair (ψ, θ) is called an *extDJ generator*, and an extDJ copula in Equation (6.2.6) is said to be *generated* by an extDJ generator. The corresponding bivariate function $H_\psi : [1, \infty) \times (0, \infty) \to [0, 1]$, defined in Equation (6.2.5), is called a *extDJ generating function*. The extDJ copula class is semi-parametric, with the space of generators being defined by concave distribution functions $\psi : [0, \infty) \to [0, 1]$ and a real parameter $\theta > 0$. See Appendix E.2.1 for properties of the extDJ generating functions.

Unlike the DJ generating functions, the extDJ generating functions are all survival functions, defined by the product of two survival functions. The space of extDJ generating functions are based only on those DJ generating functions that are increasing. This is so that the IG copula class, defined in Equation (6.1.9), follows as a special case of the extDJ copula class. In particular, restricting the space of extDJ generators to (Ψ_k, θ) for $k > 1$ and $\theta > 0$, where Ψ_k is defined in Equation (6.1.7), generate the (parametric) reflected IG copula family. To generalize the extDJ copula class further, one might consider defining H_ψ as a product of two generic survival functions or distribution functions, but this is not considered in this dissertation.

We saw in Proposition 6.2.7 that the independence copula arises as a limiting case of the DJ copula class, but otherwise, it seems that the independence copula is not a DJ copula. The extDJ copula class effectively provides a "bridge" between the DJ copulas and the independence copula.

Proposition 6.2.10 (Limit classes)**.** *Let $\psi : [0, \infty) \to [0, 1]$ be a concave distribution function. Then for $(u, v) \in (0, 1)^2$,*

$$\lim_{\theta \to \infty} C_{extDJ}(u, v; \psi, \theta) = C_{DJ}(u, v; \psi)$$

where C_{DJ} is defined in Equation (6.2.1), and

$$\lim_{\theta \downarrow 0} C_{extDJ}(u, v; \psi, \theta) = C^{\perp}(u, v).$$

Proposition 6.2.11 (Stochastic representation)**.** *Fix $\theta > 0$ and a concave distribution function $\psi : [0, \infty) \to [0, 1]$, which has density ψ'. Consider random variables Y and U described by the (valid) distribution functions*

$$F_Y(y) = 1 - H_\psi(y; \theta)$$

for $y \geq 1$, and

$$F_{U|Y}\big(u \mid y\big) = \frac{u \, \mathrm{D}_1 H_\psi(y; \theta u)}{\mathrm{D}_1 H_\psi(y; \theta)} \tag{6.2.7}$$

for $u \in (0, 1)$ and any $y \geq 1$, where H_ψ is defined in Equation (6.2.5). The random vector (U, Y) has joint distribution

$$F_{U,Y}(u, y) = u - uH_\psi(y; \theta u) \tag{6.2.8}$$

for $(u, y) \in (0, 1) \times [1, \infty)$, and the random vector $\big(U, 1/Y\big)$ has copula equal to the extDJ

copula in Equation (6.2.6) generated by (ψ, θ).

The extDJ generating functions can be defined in other ways besides Equation (6.2.5), due to their invariance under composition with invertible functions.

Proposition 6.2.12 (Invariance to function composition)**.** *Let* $\psi : [0, \infty) \to [0, 1]$ *be a concave distribution function,* $\eta > 0$*, and* g *be a univariate strictly monotone function having a domain such that the range is* $[1, \infty)$*. The extDJ copula with generating function* H_ψ*, defined in Equation (6.2.5), is equivalent to the extDJ copula with generating function*

$$\tilde{H}_\psi (t; \eta) = H_\psi \left(g (t) ; \eta \right)$$

for $(t; \eta) \in [1, \infty) \times (0, \infty)$*. That is, for* $(u, v) \in (0, 1)^2$ *and* $\theta > 0$*,*

$$uH_\psi \left(H_\psi^{\leftarrow} (v; \theta) ; \theta u \right) = u\tilde{H}_\psi \left(\tilde{H}_\psi^{\leftarrow} (v; \theta) ; \theta u \right).$$

Notice that composition of $H_\psi (\cdot; \eta)$ with a strictly decreasing function changes the direction of monotonicity. As such, the extDJ generating functions can be taken as distribution functions instead of survival functions.

Proposition 6.2.13 (Quadrant dependence)**.** *All extDJ copulas are positive quadrant dependent; that is, for any increasing DJ generator* ψ *and* $\theta > 0$*,*

$$C_{extDJ} (u, v; \psi, \theta) \geq uv$$

for $(u, v) \in (0, 1)^2$*, where* C_{extDJ} *is defined in Equation (6.2.6).*

The IG copula with $\theta > 0$ and $k > 1$ is the reflected extDJ copula generated by (Ψ_k, θ). It therefore follows that IG copulas are positive quadrant dependent.

Proofs of these results can be found in Appendix .E.2.3.

Chapter 7

Application to Flood Forecasting

This chapter focusses on applying the concepts in this dissertation to the application discussed in Section 2.1. Specifically, forecasters are built to forecast the extremes of *one-day-ahead discharge* of the Bow River at Banff using three methodologies: linear quantile regression, local regression, and the proposed CNQR method of Chapter 5 using vines. Each forecaster uses the same set of predictors, so as to assess each method's ability to leverage the information available in the predictors about the response. Forecasts are quantile functions above a lower quantile level of $\tau_c = 0.9$.

The data are first pre-processed in Section 7.1 by computing predictors and response variables, and are split into fitting and test sets. Building the three forecasters is discussed in Section 7.2.

It is found that the proposed method has the best score on the test set. However, the three methods perform comparably when considering uncertainty in the mean score, likely due to the near-linearity and/or lack of informativeness of the predictors. The evaluation and comparison of the forecasters is made in Section 7.3 using proper scoring rules, calibration plots and histograms, and weighted scores. The forecasts that would have been made during the Alberta flood are shown and discussed in Section 7.4.

Computational details of this application can be found in Appendix F.

7.1 Pre-processing the Data

The data presented in Section 2.1 need pre-processing before fitting forecasters. First, in Section 7.1.1, data are split into a fitting set, used for building the forecasters, and a test set, used for evaluating. The river discharge data are then deseasonalized in Section 7.1.2 to remove the within-year cycle, so that the deseasonalized discharge is approximately stationary across time. Then, effective predictors and a response can be

identified and computed from the raw data, discussed in Section 7.1.3. The response is chosen to be the *change in deseasonalized discharge*, and the four predictors are lags 0 and 1 change in deseasonalized discharge, and lags 0 and 1 drop in snowpack.

7.1.1 Separating the Data

In addition to *building* a forecaster, we wish to be able to *evaluate* the forecaster. To perform an adequate evaluation, one should use data (*test data*) that are different from those (*fitting data*) used in the model-building process. We reserve approximately 25% of the overall data for the test set.

Where model selection is required (such as with vine CNQR, for estimating the response marginal distribution), the fitting data are further split into a *training set* (50% of the data) used for model estimation, and a *validation set* (25% of the data) for model selection. To avoid "contamination" between the data sets due to autocorrelation, entire years are randomly selected for each set (except for 2013, the year of Alberta's big flood, reserved for the test set). The final selection of years can be seen in Table 7.1. In addition, a subset of data belonging to a "flood season" is selected, taken to be between the julian days 110 and 200.

Data Set		**Years Selected**
Test		1984, 1985, 1992, 2004, 2005, 2009, 2013, 2014
Fitting	Training	1980, 1981, 1982, 1983, 1988, 1990, 1991, 1993, 1994, 1996, 1998, 1999, 2000, 2001, 2002, 2003, 2007, 2010, 2012
	Validation	1986, 1987, 1989, 1995, 1997, 2006, 2008, 2011

Table 7.1: The years selected for each data set. The vine CNQR method splits the fitting data into training and validation, whereas the linear and local methods use the entire fitting data.

7.1.2 Deseasonalizing the Discharge

A glance at the overlaid time series plot in Figure 2.1 makes it clear that discharge is heavily dependent on the day of year. We consider this dependence separately from other predictors because of its cyclic trend, which cannot be adequately addressed using vine CNQR and linear quantile regression techniques (also because it is deterministic). We will therefore pre-process the data by removing this trend.

If Y_t is the discharge on date t, and δ is the mapping from date to the day of the year, then we assume

$$Z_t := \frac{\log Y_t - \mu_{\delta(t)}}{\sigma_{\delta(t)}} \tag{7.1.1}$$

is strictly stationary for some sequences $\{\mu_i\}$ and $\{\sigma_i > 0\}$, $i = 1, \ldots, 366$ (that is, Z_t has the same distribution across t). Such transformed data are called *deseasonalized* (ds). Estimates of $\{\mu_i\}$ and $\{\sigma_i\}$ are shown in Figure 7.1, and are obtained using a local method, described in detail in Appendix F.1. The resulting overlaid time series plots of the deseasonalized data Z_t can be found in Figure 7.2.

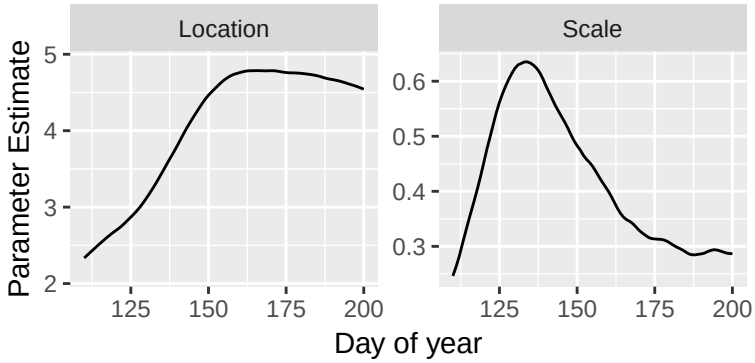

Figure 7.1: Estimated seasonal components $\{\mu_i\}$ and $\{\sigma_i\}$ (respectively, location and scale) of the discharge for the selected "flood season".

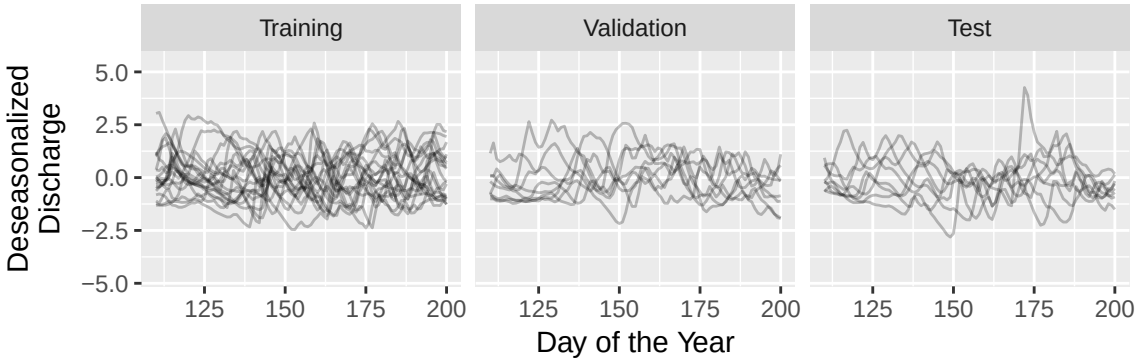

Figure 7.2: Overlaid time series plots of deseasonalized discharge during the "flood season", on each of the training, validation, and test sets.

7.1.3 Choosing Predictors and a Response

Although forecasts are to be made on the discharge of the Bow River at Banff, choosing discharge (even deseasonalized) as a response variable turns out to be a poor choice. Discharge is known to exhibit long-range dependence, meaning that there is a strong serial correlation in the time series. This is intuitive, because the discharge separated by two days ought to be rather similar. Such high serial correlation is visible in the autocor-

relation plot of ds discharge, found in Figure 7.3. Although there are 1965 observations on the fitting set, the "effective" sample size is much smaller.

On the contrary, *difference in ds discharge* displays very minimal serial correlation (also shown in Figure 7.3). This means that the effective sample size is close to the full sample size, and so more information is available on the day-to-day change in ds discharge than there is for the absolute ds discharge. Since the discharge can be recovered from the change in ds discharge, the *one-day ahead change in ds discharge* is chosen to be the response variable.

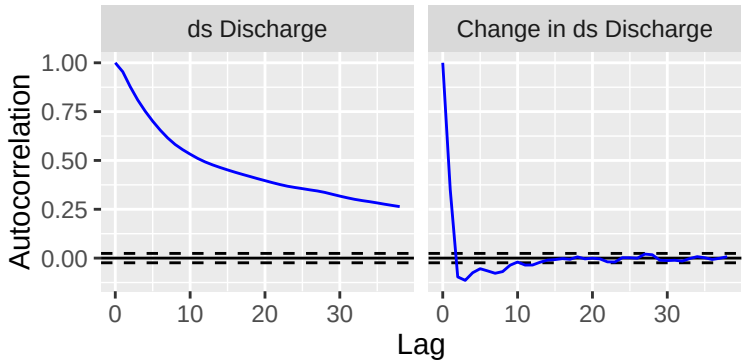

Figure 7.3: Autocorrelation plots of ds discharge and change in ds discharge. Error bands represent 95% confidence bands about zero. The ds discharge is highly serially correlated, whereas the change in ds discharge has minimal serial correlation.

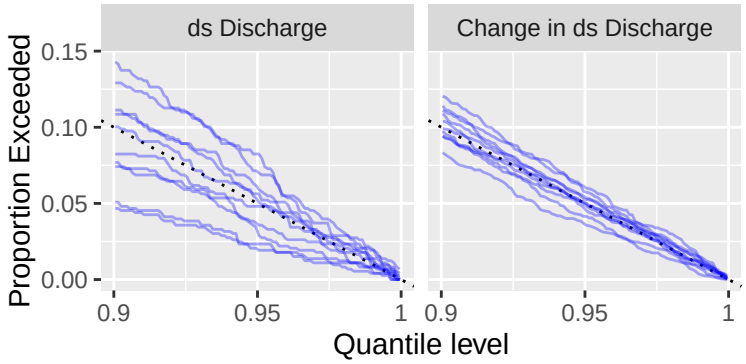

Figure 7.4: Calibration plots of the ds discharge marginal forecaster, and the change in ds discharge marginal forecaster. A marginal forecaster uses the marginal quantile function as a forecast. The ten curves in each panel represent the results under ten randomly chosen splits of the data into fitting and test sets.

To verify that change in ds discharge is in fact a better choice for the response than the ds discharge itself, one can compare the respective marginal forecasts, which do not use any predictors. Fit using the fitting set, the marginal forecaster of a variable is simply its empirical quantile function, which is issued as a forecast under any circumstance. Due

to high serial correlation, one would expect more disparity in ds discharge between the fitting and test sets, so that the forecaster based on ds discharge is more susceptible to give over- or under-estimates. Indeed, this is visible in the calibration plots of Figure 7.4.

Next, predictors are needed. Four predictors are chosen, based on the snowpack and discharge data shown in Figure 2.1 of Section 2.1: lags 0 and 1 change in deseasonalized discharge, and lags 0 and 1 drop in snowpack. Other predictors were considered, such as lagged ds discharge, but they were not as informative (as indicated by Kendall's tau).

Scatterplots of the response against the predictors can be seen in Figure 7.5, along with local estimates of upper quantiles. An assumption of linearity on these data so far does not appear to be a bad one, particularly because the dependence in each pair is not very strong. Certainly, a more rigorous choice of predictors would be an asset, such as a better estimate of snowmelt within the Bow River watershed, but these predictors are used regardless.

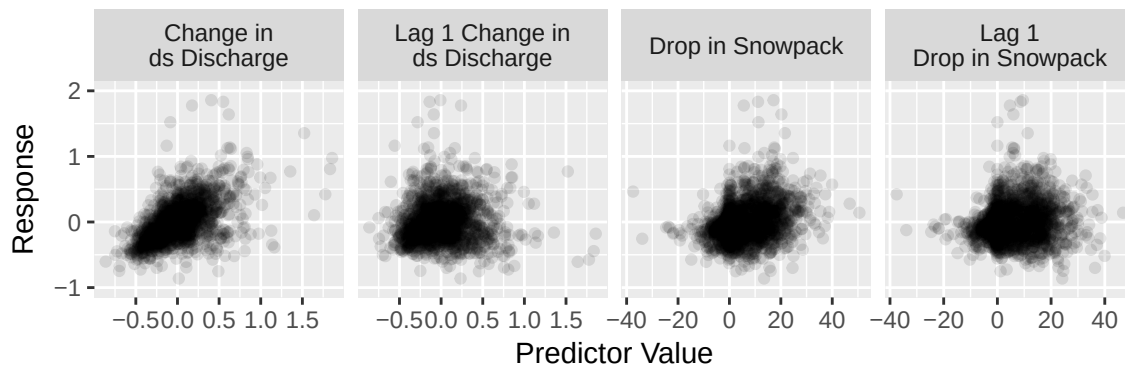

Figure 7.5: Scatterplots of predictors (listed in the upper panels) against the response, with local estimates of upper quantile curves. Variables are unit-less, aside from lag 0 and lag 1 drop in snowpack, which are in millimeters. The fitted curves are local estimates of $K = 10$ upper quantiles with levels above $\tau_c = 0.9$, taken to be $\tau_k = \tau_c + (1 - \tau_c)\,(2k - 1)\,/\,(2K)$ for $k = 1, \ldots, K$. The curves are displayed only to convey trends. The curves are obtained by fitting a GPD tail with MLE over a moving-window having a bandwidth of 0.5 standard deviations of the predictor.

7.2 Building Forecasters

With the data pre-processed, we are now ready to build the forecasters using the three competing techniques.

The application of linear quantile regression on the deseasonalized fitting data is routine – we use the `rq()` function in the R (R Core Team, 2015) package `quantreg` (Koenker, 2016). As for the local smoothing method, a Gaussian kernel with a bandwidth of 0.5 Mahalanobis units over the predictor space is used to obtain weights for the response.

The weights are used to obtain a weighted empirical quantile function, as discussed by Daouia et al. (2011).

For the vine CNQR method, the CNQR estimator is used with $K = 10$ quantile levels

$$\tau_k = \frac{2k-1}{2K}(1-\tau_c) + \tau_c$$

for $k = 1, \ldots, K$, and each g_k being the identity function.

The first stage of building a forecaster with vine CNQR is to model the univariate marginals of each variable. Although there are five variables between the four predictors and one response, these make up only two variables, up to lag: change in ds discharge, and drop in snowpack.

We consider two marginal distributions for change in ds discharge:

1. the empirical distribution, and
2. the empirical distribution with a Generalized Pareto distribution (GPD) tail,

both of which are estimated using the entire fitting data. Strictly using the empirical distribution forces a finite right-endpoint to the forecasts, whereas the GPD tail is justified by Extreme Value Theory and allows for a more reliable extrapolation into the predictive distribution's tail. See Appendix F.2 for details on fitting this marginal distribution.

From the overlaid time series plot in Figure 2.1, it appears that there is a cyclic trend in the distribution of snowmelt. This trend should be considered so that the appropriate marginal distribution can be used when computing the conditional predictors (computed with Equation (A.2.2)). However, it is not clear what components of the distribution change, as it is when deseasonalizing the discharge. We will therefore use a local empirical distribution method on the fitting data to estimate the marginal distribution, as discussed by Daouia et al. (2011). See Appendix F.3 for details on fitting the marginal distribution for snowmelt.

The second stage of building a forecaster with vine CNQR is to model the dependence between the variables. A vine copula is first fit on the predictors using MLE and selected using AIC. The vine copula is extended by introducing the response, where two pairing orders are considered (see Appendix F.4). Bivariate copula models are fit for each variable in the pairing order using the sequential method introduced in Section 5.3. Estimation is done on the entire fitting set – splitting the fitting data into training and validation here is not necessary, since the sample size of the fitting data (1965) is not that large anyway.

Lastly, we must select one of the resulting candidate vine CNQR forecasters. Scores and calibration plots of these forecasters on the fitting set are investigated to make a se-

lection (again, using a separate validation set for this is not necessary). See Appendix F.4 for details on modelling the dependence, and for model selection. Of the five candidates considered, "Candidate 5" was selected as the best.

To investigate the sensitivity of this method to the splitting of the data into fitting and test sets, this fitting procedure was repeated under a different random split. The vines fit to the predictors were almost identical. The fitted copulas linking the response and predictors were also similar, except for the last two vine CNQR forecasters, which had a BB8 copula instead of a t copula fit to the first predictor. Overall, this suggests that the vine CNQR method is not overly sensitive if fit on another set of data.

7.3 Evaluation

The three competing forecasters can now be evaluated using calibration plots (see Section 3.1) and a proper scoring rule (see Section 3.2.2). For comparison, the forecasters are also compared to the GPD marginal forecaster – that is, the forecaster that always issues the upper quantile function of the fitted GPD distribution of the response. This is to provide a sense of the value added by the predictors.

Calibration plots and calibration histograms of the forecasters can be found in Figure 7.6. They show that the forecasters are well-calibrated, and since the histograms all appear fairly uniform (or at least have no trend), it seems the tails of the forecasts decay adequately.

Next, mean scores are estimated both for forecasts made on the response (one-day-ahead change in ds discharge) and the original discharge scale. Regarding the choice of a scoring rule, we use Equation (3.2.8) with a constant cross-quantile weight of $w(\tau) = 1$, and transformation functions $g(x) = 1$ and $g(x) = \log(x)$ for the response and discharge, respectively. The computational simplifications mentioned in Remark 3.2.2 are also used.

The mean score estimates can be seen in Figure 7.7. Although the proposed method has the best score, at least the linear method appears to perform similarly[1], and the methods do not show much difference between scoring on the response or the discharge. Each competing forecaster shows some improvement from the marginal forecaster, though could be better if predictors that are more informative are used.

[1] A Diebold-Mariano test (Diebold and Mariano, 1995) for differences between the linear and vine CNQR mean scores produces p-values of 0.14 and 0.06 for the response and discharge scores, respectively. This suggests that any difference in the mean scores of the two forecasters is insignificant at the 0.05 level.

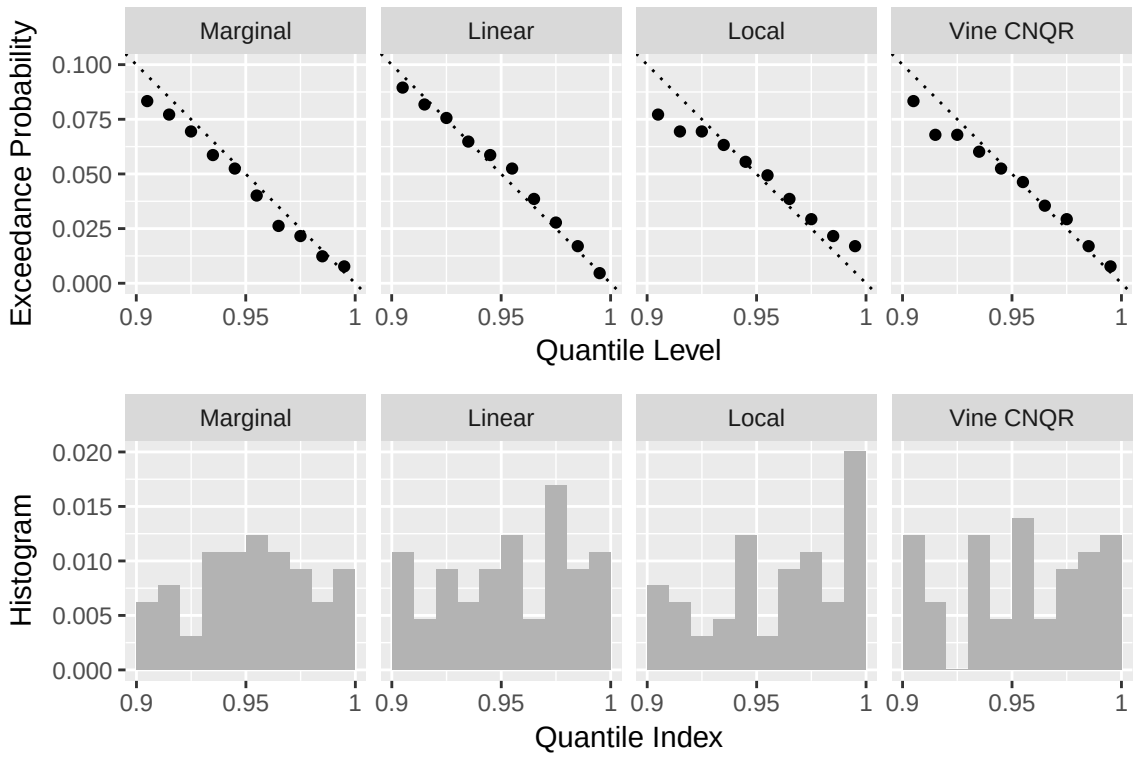

Figure 7.6: Calibration plots (above) and histograms (below) of the competing forecasters, on the test set. The dashed lines represent a perfectly calibrated forecaster. For the calibration plots, the region above the dashed line represents under-estimation (i.e., an exceedance occurs more often than suggested), and the region below represents over-estimation (i.e., an exceedance occurs less often than suggested).

Although both the linear and vine CNQR methods appear to be more or less equally good, the vine CNQR method still has additional properties that make it more desirable: its forecasts are consistent and is motivated by Extreme Value Theory in the Y marginal distribution (see Appendix F.2). The linear method gives an inconsistent forecast 11.3% of the time, and seems to be more likely to be inconsistent when the discharge is larger, as demonstrated in Figure 7.8. Note that nine other random splits of the data into fitting and test sets were tried, using the same copula models, and the results were very similar.

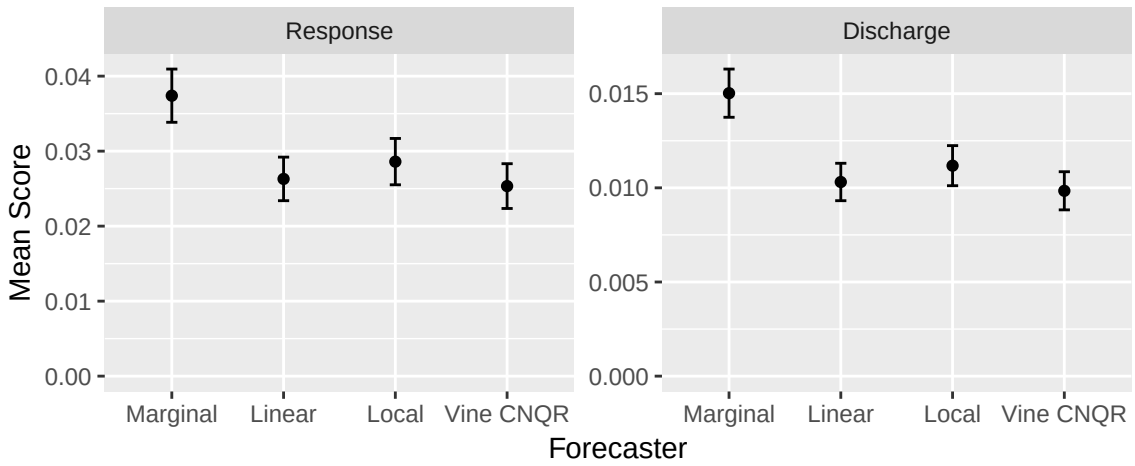

Figure 7.7: Estimates of the mean score for each of the three competing forecasters, computed on the test set. Error bars represent the standard error of the mean estimate, ignoring autocorrelation. Smaller scores are better.

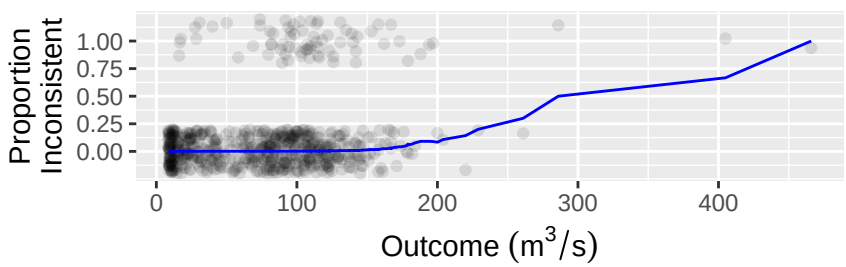

Figure 7.8: Of the forecasts issued by the linear forecaster when the observed outcome/discharge is at least as big as the horizontal axis, the solid blue line represents the (interpolated) proportion of forecasts that are inconsistent (that is, are non-monotonic quantile functions). Inconsistency was determined using 10 quantile levels. Approximately 11.3% of forecasts are inconsistent, and inconsistencies appear more likely when the outcome is large. In addition, the raw data are shown: each instance of an inconsistent forecast is plotted with jitter around a proportion of 1, and around 0 for each consistent forecast.

Lastly, we investigate how the performance of each forecaster changes when predictors are observed to be large, as well as how performance compares between forecasters when predictors are large. Weights are used under four scenarios: larger weights are taken when we observe

- a large discharge,
- a large snowmelt,
- either a large snowmelt or a large discharge, and
- both a large snowmelt and a large discharge.

Within each scenario, we consider different weight functions to ensure that the results are not an artifact of the weight function parameters. See Appendix F.5 for the specific choice of weights.

The weighted scores and calibration plots can be seen in Figures 7.9 and 7.10, respectively. The scores typically worsen when the weights are added, but less so when only discharge is weighted. This suggests that the forecasters perform slightly worse when predictors are observed to be large.

As for the weighted calibration plots, it appears that the calibration of each forecaster does not change much when predictors are large. That is, the curves in each panel are mostly clustered and not much different from the unweighted versions. However, when both snowmelt and discharge are weighted, it appears that the vine CNQR forecaster is able to retain its calibration more so than the linear and local methods, which have calibration plots that are more "scattered". This weighting scenario is different from the others, because it puts weight over the joint tail of the predictors. The relative performance of the vine CNQR forecaster suggests that it is able to capture the tail properties better than the local and linear methods can.

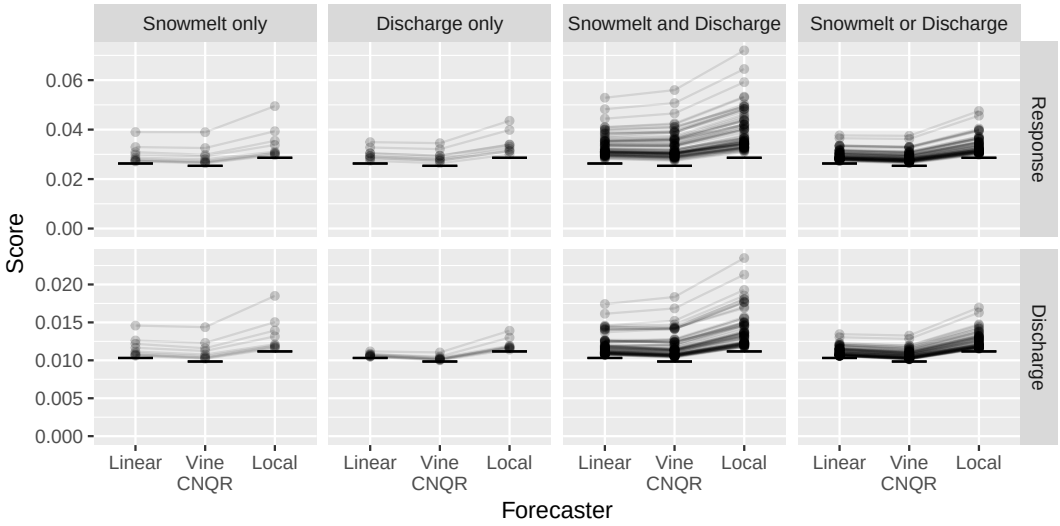

Figure 7.9: Comparison of weighted scores amongst the competing forecasters, estimated using the test set. The weighting scenario is indicated in the panel labels. Lines connect scores under common weighting schemes, and the solid horizontal lines represent the unweighted scores.

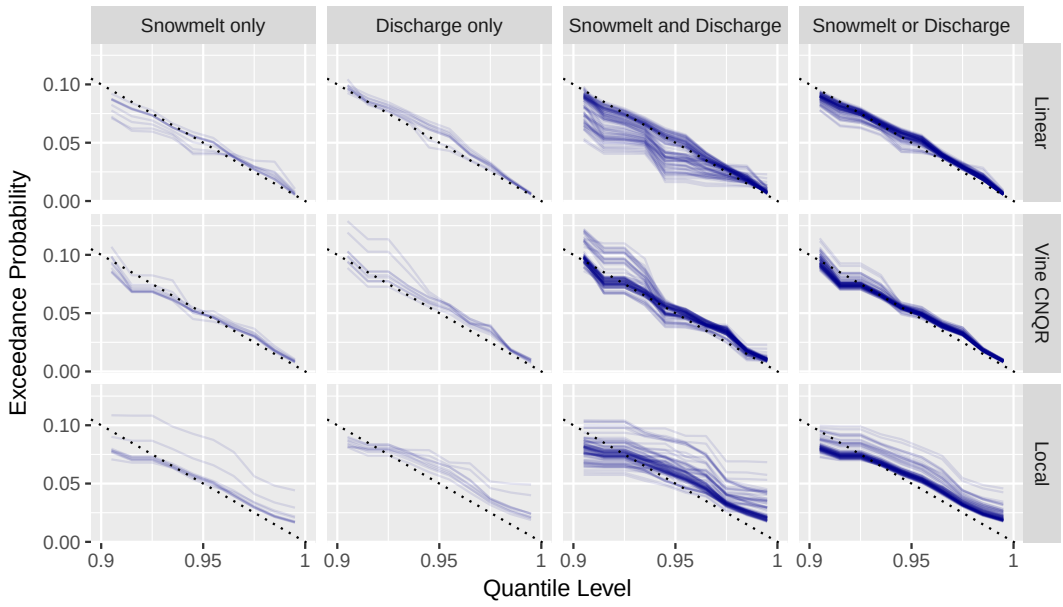

Figure 7.10: Calibration plots of the competing forecasters under different weighting schemes, estimated using the test set. The forecaster is indicated in the side panels; the weighting scenario is indicated in the upper panels. The dashed line represents a perfectly calibrated forecaster; the region above the dashed line represents under-estimation (i.e., an exceedance occurs more often than suggested), and the region below represents over-estimation (i.e., an exceedance occurs less often than suggested).

7.4 Forecasting the 2013 Flood

The days of the Alberta 2013 flood were deliberately included in the test set so that the forecasts for those days could be evaluated without being influenced by the fitting procedure. The forecasts can be seen in Figure 7.11.

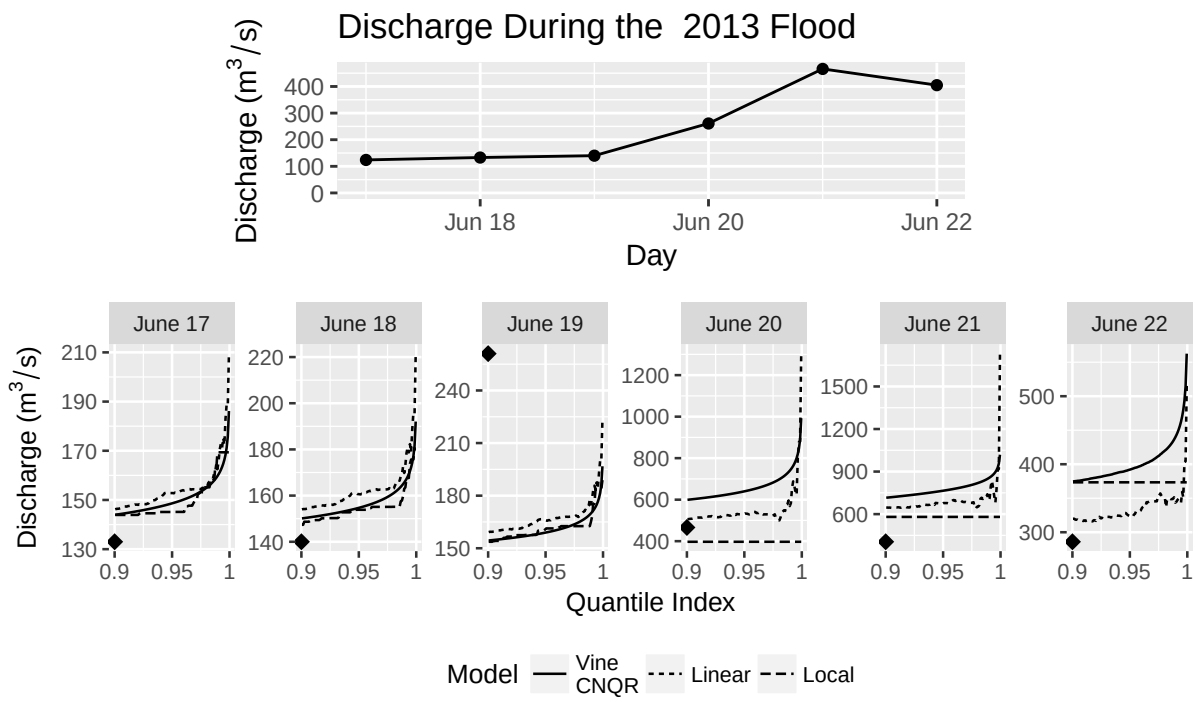

Figure 7.11: Above: discharge during the start of the 2013 flood, on the indicated days, observed at the Bow River at Banff. Below: Forecasts that would have been issued for the Bow River at Banff for the day *after* that which is indicated. The diamond-shaped bullet indicates the actual observed value on the next day.

Recall that, since the predictive distribution's tail is being forecast, we expect the forecasts to be well above the observed discharge. This is not the case when the discharge increased on June 20. This may be because the relatively unchanging discharge over the two previous days was not suggestive of an increase in discharge, leaving snowmelt as the sole indicator.

Notice that the linear method does not always produce consistent forecasts, as is especially true for June 21 and 22. In addition, the local method has a flat quantile function from June 20–22, corresponding to a large jump in the distribution function. This suggests that there are very few observations in the fitting set with predictors similar to what was observed on those days (because at least one observation has a weight > 0.1).

7.5 Simulation

The predictors used in the application are not very informative, and so trends (including nonlinear ones) are not very pronounced. The vine CNQR method's performance is investigated further here on two simulated data sets: one with linear quantile surfaces, and another with nonlinear quantile surfaces.

In both cases, 1000 observations on three variables are generated as training data, and 10,000 observations as test data. The vine CNQR method is compared against linear quantile regression. For vine CNQR, the marginals are estimated by the empirical distribution function, and when linking the response, the predictor with the highest Kendall's tau is chosen first in the pairing order.

The linear simulation has data generated from the quantile function

$$Q_{Y|X_1,X2}\left(\tau \mid x_1, x_2\right) = -\log\left(1-\tau\right) + \tau x_1 + \sqrt{(1-\tau)}x_2$$

for $x_1, x_2 > 0$ and $\tau \in (0,1)$, where the predictors have standard Exponential marginals linked by a $t\left(0.3, 3\right)$ copula. The non-linear simulation has data generated from a vine with the array

$$\begin{bmatrix} Y & X_1 & X_2 \\ & Y & X_1 \\ & & Y \end{bmatrix}$$

and corresponding copulas

$$\begin{bmatrix} \text{Gumbel}\left(3\right) & t\left(0.3, 3\right) \\ & \text{MTCJ}\left(2\right) \end{bmatrix}.$$

Figures 7.12 and 7.13 show the fitted quantile surfaces. For the linear data, the average scores on the test set were 0.14 for the linear method, and 0.15 for the nonlinear method. This means that, although the quantile surfaces are truly linear, the vine CNQR method still performs comparably to the linear method. The vine CNQR method produces surfaces that look approximately linear, but the linear method produces quantile surfaces that cross. As for the non-linear data, the average scores on the test set were 0.072 for the linear method, and 0.037 for the nonlinear method. The vine CNQR is a clear winner here, not only by its superior score, but by its visually excellent fit to the data. Also, note that the non-linearity of the quantile surfaces do not appear smooth – this is due to the empirical distributions being used as the marginals.

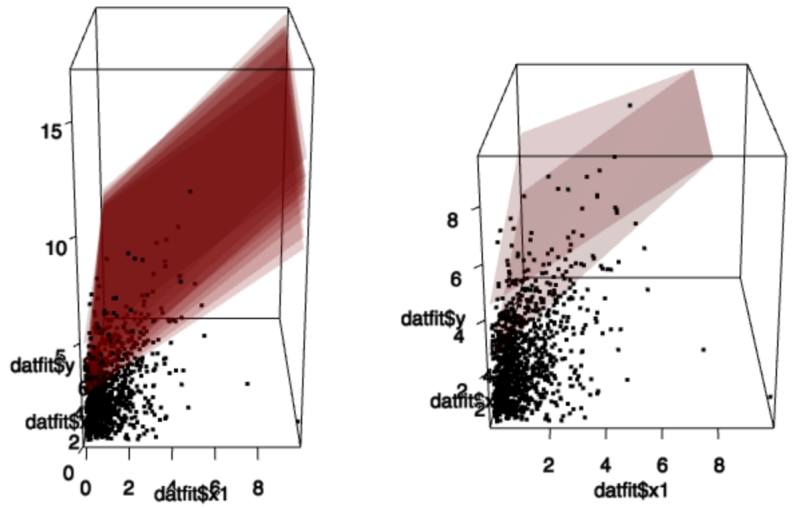

Figure 7.12: Linear simulation data with fitted linear (left) and non-linear (right) quantile surfaces. The linear surfaces were fit using linear quantile regression, and the non-linear surface with vine CNQR.

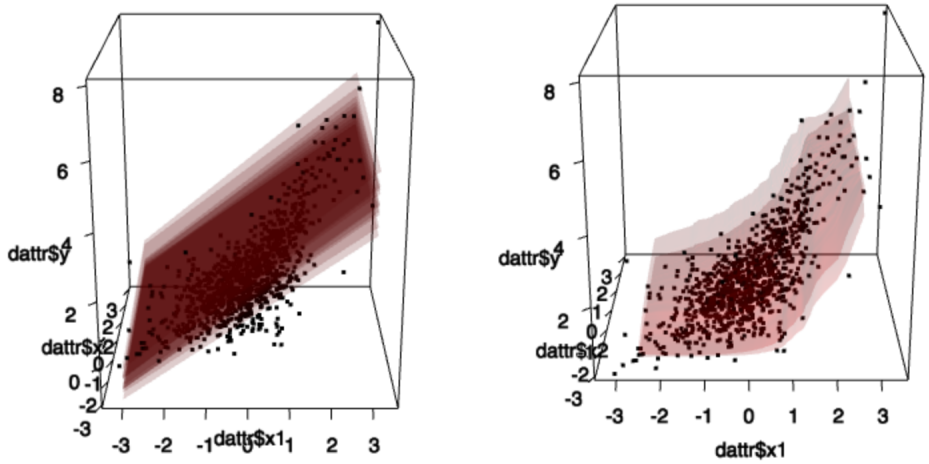

Figure 7.13: Non-linear simulation data with fitted linear (left) and non-linear (right) quantile surfaces. The linear surfaces were fit using linear quantile regression (with ten quantile levels above 0.9), and the non-linear surface with vine CNQR (with quantile levels 0.905 and 0.995).

Chapter 8

Conclusion

The main objective of this work is to issue extreme forecasts in terms of upper quantile functions, by garnering insight from predictors. This type of forecast is useful because it communicates the forecaster's entire belief about the chance of extremes.

Specifically, a method is sought in this dissertation that ensures quantile function forecasts are non-decreasing (consistent); that allows for extrapolation into the tail of the forecast distribution in a justified way; and that is flexible enough to depart from the assumption of linear quantile surfaces, so that phenomena such as tail dependence can be modelled. This work follows up on the call for further research in copula quantile regression by Koenker (2005).

A family of proper scoring rules is developed in order to assess the goodness of a forecaster – though it still remains unclear which scoring rules are more desirable than others. This family of scoring rules can be used to define a new family of estimators, called the CNQR estimators, which chooses the model that optimizes the score on the training data. Asymptotic properties of these estimators are established, and can be used for making inference. These results provide some evidence that data should be transformed to a random variable having a light- or short- tailed distribution to reduce the asymptotic variance.

The PCBN or vine copula approach to modelling quantile surfaces is very flexible, and allows for intricate aspects of quantile surfaces—such as tail dependence—to be captured. This is particularly important when forecasting extremes, because often the most extreme values occur when the predictors are large. In addition, this method always issues consistent forecasts. However, the PCBN or vine approach cannot handle a large number of predictors, because the model error compounds further down the linkage order. The method will become increasingly more flexible as more copula models are identified in the literature.

To allow for the most reliable extrapolation into the tail of a forecast, the conditional EVI should be carefully estimated. This can be done by first fitting an appropriate univariate marginal, such as a GPD, and then selecting a copula family having an appropriate CCEVI. The latter step is done indirectly through CNQR estimation. Since copulas having a non-constant CCEVI seem to be unavailable, two new parametric copula families are identified. The IG copula family is made up of copulas having a non-constant CCEVI, allowing for the conditional EVI to differ depending on the value of the predictor. The IGL copula family is made up of copulas having a zero CCEVI, allowing for a predictor to entirely describe the tail heaviness of a response variable. This finding is significant, because other parametric copula families identified in the literature do not seem to have these properties.

The work presented in this dissertation has many applications. The example of flood forecasting is given, but by extension, this work can be applied to forecast any extreme weather or natural disaster having some sort of "precursor". Other examples of application areas are finance, to forecast crashes in stock prices; insurance, to ensure an unusually high amount of claims can be paid out; and supply-chain management, to ensure that enough resources are available to handle unusually high demands from customers. In general, this work can be a big asset for decision makers who need to prepare for potential disaster.

There are many new research questions that arise from this work. The following are some interesting ones.

- Perhaps the sampling distribution of a CNQR estimator can be better approximated by considering intermediate and extreme order sequences of τ_c. Chernozhukov (2005) found this result with the linear quantile regression estimator, when regressing on high quantiles.

- What is the asymptotic distribution of the CNQR estimators when $K \to \infty$? Can this be used to gain insight as to what transformation functions g_k are appropriate to use for estimation?

- Koenker and Park (1996) introduce an algorithm that can be used to find the minimum of a CNQR objective function. Are there ways in which this algorithm can be improved specifically for CNQR?

- The proposed IG and IGL families still have many properties that need discovery, possibly aided by a new parameterization.

- Can the DJ Weibull copula, identified in Section 6.2.1.2, be generalized to attain higher dependence?
- Can a family of proper scoring rules for the EVI be obtained from the proper scoring rules identified for upper quantile functions, in Equation (3.2.8)? This would be useful to assess the goodness of fit of a model for the conditional EVI. Perhaps a limit as $\tau_c \uparrow 1$ while the integrand expands (for example, by increasing the weight function w) will result in such a scoring rule.
- If one knows the CCEVI's of each bivariate copula in a vine, how can the CCEVI of the entire vine copula be computed?
- When forecasting the mean in a time series, the "last value carried forward" technique is a naïve method used for comparison. This method forecasts the previously observed response. Is there an extension to quantiles that can be used?

Bibliography

Amemiya, T. (1985). *Advanced Econometrics*. Harvard University Press, Cambridge, Massachusetts.

Bauer, A. and Czado, C. (2015). Pair-copula Bayesian networks. *Journal of Computational and Graphical Statistics (to appear)*.

Bernard, C. and Czado, C. (2015). Conditional quantiles and tail dependence. *Journal of Multivariate Analysis*, 138:104–126.

Beven, K. (2012). *Rainfall Runoff Modelling*. Wiley-Blackwell.

Bondell, H. D., Reich, B. J., and Wang, H. (2010). Noncrossing quantile regression curve estimation. *Biometrika*, 97(4):825–838.

Bouyé, E. and Salmon, M. (2009). Dynamic copula quantile regressions and tail area dynamic dependence in Forex markets. *The European Journal of Finance*, 15(7-8):721–750.

Cannon, A. J. (2010). A flexible nonlinear modelling framework for nonstationary generalized extreme value analysis in hydroclimatology. *Hydrological Processes*, 24(6):673–685.

Chen, X., Koenker, R., and Xiao, Z. (2009). Copula-based nonlinear quantile autoregression. *Econometrics Journal*, 12(S1):S50–S67.

Chernozhukov, V. (2005). Extremal quantile regression. *The Annals of Statistics*, 33(2):806–839.

Chernozhukov, V., Fernández-Val, I., and Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica*, 78(3):1093–1125.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4):841–862.

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

Cooke, E. (1906). Forecasts and verifications in western Australia. *Monthly Weather Review*, 34(1):23–24.

Daouia, A., Gardes, L., and Girard, S. (2013). On kernel smoothing for extremal quantile regression. *Bernoulli*, 19(5B):2557–2589.

Daouia, A., Gardes, L., Girard, S., and Lekina, A. (2011). Kernel estimators of extreme level curves. *Test*, 20(2):311–333.

Dawid, A. P. (1984). Statistical theory: the prequential approach. *Journal of the Royal Statistical Society Series A*, 147(2):278–292.

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer.

Diebold, F. X. and Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, 13:253–265.

Doll, J. and Geddes, L. (2014). Flood reflections: reporter Jayme Doll returns to Canmore's Cougar Creek. *Global News*.

Durante, F. and Jaworski, P. (2012). Invariant dependence structure under univariate truncation. *Statistics*, 46(2):263–277.

Durante, F., Jaworski, P., and Mesiar, R. (2011). Invariant dependence structures and Archimedean copulas. *Statistics and Probability Letters*, 81(12):1995–2003.

Environment and Climate Change Canada (2009). *Flooding Events in Canada - Prairie Provinces*. http://www.ec.gc.ca/eau-water/default.asp?lang=En&n=E0399791-1 [Accessed: February 4, 2016].

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2):180–190.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics*, 44(3):423–453.

Gneiting, T., Balabdaoui, F., and Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society Series B (Methodological)*, 69(2):243–268.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378.

Heffernan, J. E. and Tawn, J. a. (2004). A conditional approach to modelling multivariate extreme values. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 66(3):497–546.

Hjort, N. L. and Pollard, D. (1993). Asymptotics for minimisers of convex processes. Technical report, University of Oslo.

Huang, M. L., Xu, X., and Tashnev, D. (2015). A weighted linear quantile regression. *Journal of Statistical Computation and Simulation*, 85(13):2596–2618.

Jaworski, P. (2015). Univariate conditioning of vine copulas. *Journal of Multivariate Analysis*, 138:89–103.

Jiang, X., Jiang, J., and Song, X. (2012). Oracle model selection for nonlinear models based on weighted composite quantile regression. *Statistica Sinica*, 22:1479–1506.

Joe, H. (2014). *Dependence Modeling with Copulas*. CRC Press.

Joe, H. and Krupskii, P. (2014). *CopulaModel: Dependence Modeling with Copulas*. R package version 0.6.

Knight, K. (1998). Limiting distributions for L_1 regression estimators under general conditions. *The Annals of Statistics*, 26(2):755–770.

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press.

Koenker, R. (2016). *quantreg: Quantile Regression*. R package version 5.26.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.

Koenker, R. and Mizera, I. (2004). Penalized triograms: total variation regularization for bivariate smoothing. *Journal of the Royal Statistical Society Series B*, 66(1):145–163.

Koenker, R., Ng, P., and Portnoy, S. (1994). Quantile smoothing splines. *Biometrika*, 81(4):673–680.

Koenker, R. and Park, B. (1996). An interior point algorithm for nonlinear quantile regression. *Journal of Econometrics*, 71:265–283.

Kraus, D. and Czado, C. (2017). D-vine copula based quantile regression. *Computational Statistics and Data Analysis*, 110:1–18.

Krzysztofowicz, R. (2001). The case for probabilistic forecasting in hydrology. *Journal of Hydrology*, 249(1-4):2–9.

Li, W. (2016). Flooding forces residents out of homes. *Vancouver Metro News*, page 4.

Matheson, J. E. and Winkler, R. L. (1976). Scoring rules for continuous probability distributions. *Management Science*, 22(10):1087–1096.

Micovic, Z. (2005). *Minimum Watershed Model Structure for Representation of Runoff Processes*. PhD thesis, The University of British Columbia.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In Engle, R. F. and McFadden, D. F., editors, *Handbook of Econometrics*, volume 4, chapter 36, pages 2111–2245.

Noh, H., El Ghouch, A., and Van Keilegom, I. (2015). Semiparametric conditional quantile estimation through copula-based multivariate models. *Journal of Business and Economic Statistics*, 33(2):167–178.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Schepsmeier, U., Stoeber, J., Brechmann, E. C., Graeler, B., Nagler, T., and Erhardt, T. (2016). *VineCopula: Statistical Inference of Vine Copulas*. R package version 2.0.1.

Spokoiny, V., Wang, W., and Härdle, W. K. (2013). Local quantile regression. *Journal of Statistical Planning and Inference*, 143(7):1109–1129.

Wadsworth, J. L., Tawn, J. A., and Jonathan, P. (2010). Accounting for choice of measurement scale in extreme value modeling. *Annals of Applied Statistics*, 4(3):1558–1578.

Wang, H. J. and Li, D. (2013). Estimation of extreme conditional quantiles through power transformation. *Journal of the American Statistical Association*, 108(503):1062–1074.

Xie, Q. (2015). Computation and application of copula-based weighted average quantile regression. *Journal of Computational and Applied Mathematics*, 281:182–195.

Zhao, Z. and Xiao, Z. (2014). Efficient regressions via optimally combining quantile information. *Econometric Theory*, 30(6):1272–1314.

Zou, H. and Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *The Annals of Statistics*, 36(3):1108–1126.

Appendix A

Formulas

Some formulas related to copulas are provided here. Section A.1 provides formulas for copula reflections, and Section A.2 provides formulas for PCBN/vine copula conditional distribution functions and conditional quantile functions.

A.1 Reflection Copulas

Let C be a d-dimensional copula, for $d \geq 2$. For some $m \in \{1, \ldots, d\}$, we derive the formula for the reflection copula $\mathrm{R}_{1\cdots m}C$, which defines a reflection operator.

Notice that $\mathrm{R}_{1\cdots m} = \mathrm{R}_m \mathrm{R}_{m-1} \cdots \mathrm{R}_1$ (or any permutation of R_j, $j = 1, \ldots, m$). The formula for $\mathrm{R}_1 C$ (which can be extended to any $\mathrm{R}_j C$) is simply

$$\begin{aligned} \mathrm{R}_1 C\left(u_1, \ldots, u_d\right) &= \mathcal{P}\left(U_1 \geq 1 - u_1, U_2 \leq u_2, \ldots, U_d \leq u_d\right) \\ &= C\left(1, u_2, \ldots, u_d\right) - C\left(1 - u_1, u_2, \ldots, u_d\right). \end{aligned}$$

Applying this formula recursively results in an inclusion-exclusion type formula that can be written compactly as

$$\mathrm{R}_{1\cdots m}C\left(u_1, \ldots, u_d\right) = \sum_{S \subset \{1, \ldots, m\}} (-1)^{|S|} C_{S \cup (m+1):d}\left(1 - u_i, i \in S, u_{m+1}, \ldots, u_d\right), \quad \text{(A.1.1)}$$

where for set $S \subset \{1, \ldots, m\}$, $|S|$ denotes the cardinality of S, and

$$C_{S \cup (m+1):d}\left(u_i, i \in S, u_{m+1}, \ldots, u_d\right)$$

is a $\big(d - m + |S|\big)$-dimensional margin of the copula C; it is C with the i'th argument being u_i for each $i \in S$, the j'th argument being 1 for each $j \in \{1, \ldots, m\} \setminus S$, and

arguments $m+1,\ldots,d$ being (respectively) $u_{m+1},\ldots,u_d$. Informally, but more useful in practice, we can write this formula as

$$\begin{aligned}\mathrm{R}_{1\cdots m}C\left(u_1,\ldots,u_d\right) &= C\left(1,\ldots,1,u_{m+1},\ldots,u_d\right)\\ &\quad-\sum_{i=1}^{m}C\left(1,\ldots,1-u_i\text{ (position }i),\ldots,1,u_{m+1},\ldots,u_d\right)\\ &\quad+\sum_{j<i\leq m}C\left(1,\ldots,1-u_j\text{ (position }j),\ldots,1-u_i\text{ (position }i),\ldots,1,u_{m+1},\ldots,u_d\right)\\ &\quad-+\cdots.\end{aligned} \tag{A.1.2}$$

The formulas for the bivariate case are

$$\mathrm{R}_1C\left(u,v\right)=v-C\left(1-u,v\right), \tag{A.1.3}$$

$$\mathrm{R}_2C\left(u,v\right)=u-C\left(u,1-v\right), \tag{A.1.4}$$

and

$$\hat{C}\left(u,v\right)=u+v-1+C\left(1-u,1-v\right). \tag{A.1.5}$$

A.2 PCBN Equations

For other formulas relating to vine distributions, see Joe (2014, Section 3.9.3).

Theorem A.2.1. *Suppose* $\boldsymbol{W}=\left(W_1,\ldots,W_p\right)^\top$ *is a vector of absolutely continuous random variables whose distribution function can be described by a PCBN with* $(p,1,\ldots,p-1)^\top$ *as the last column of its PCBN array (that is,* W_p *is introduced last and has a pairing order* $W_1,\ldots,W_{p-1}$*). Then if* $p'<p$*, the conditional distribution function of* $W_p\mid\left(W_1,\ldots,W_{p'}\right)$ *can be calculated using the recursive relation*

$$F_{p|1:p'}\left(w_p\mid\boldsymbol{w}_{1:p'}\right)=C_{p|p';1:(p'-1)}\left(F_{p|1:(p'-1)}\left(w_p\mid\boldsymbol{w}_{1:(p'-1)}\right)\mid u_{p'}\right), \tag{A.2.1}$$

where

$$u_{p'}=F_{p'|1:(p'-1)}\left(w_{p'}\mid\boldsymbol{w}_{1:(p'-1)}\right). \tag{A.2.2}$$

The corresponding quantile function is

$$Q_{p|1:p'}\left(\tau\mid\boldsymbol{w}_{1:p'}\right)=Q_{p|1:(p'-1)}\left(C^{\leftarrow}_{p|p';1:(p'-1)}\left(\tau\mid u_{p'}\right)\mid\boldsymbol{w}_{1:(p'-1)}\right). \tag{A.2.3}$$

Proof. Omitting function arguments where obvious, we have

$$\begin{aligned} F_{p|1:p'} &= \frac{D_1 F_{p',p|1:(p'-1)}}{f_{p'|1:(p'-1)}} \\ &= \frac{D_1 C_{p',p;1:(p'-1)}\left(F_{p'|1:(p'-1)}, F_{p|1:(p'-1)}\right) f_{p'|1:(p'-1)}}{f_{p'|1:(p'-1)}} \\ &= C_{p|p';1:(p'-1)}\left(F_{p|1:(p'-1)} \mid F_{p'|1:(p'-1)}\right), \end{aligned}$$

where D_1 is the differential operator on the first argument. The quantile function can be obtained by a simple function inversion exercise.

□

So, for example, if there are $p = 2$ predictors, the quantile function of $Y \mid (X_1, X_2)$ would be

$$Q_{Y|1:2}\left(\tau \mid (x_1, x_2)\right) = Q_{Y|1}\left(C^{\leftarrow}_{Y|2;1}\left(\tau \mid u_2\right)\right) = Q_Y\left(C^{\leftarrow}_{Y|1}\left(C^{\leftarrow}_{Y|2;1}\left(\tau \mid u_2\right) \mid u_1\right)\right), \tag{A.2.4}$$

where $u_1 = F_1\left(w_1\right)$ and $u_2 = F_{2|1}\left(w_2 \mid w_1\right)$.

Appendix B

Proofs related to the Conditional EVI

This appendix contains the proofs of the general results on the conditional EVI contained in Section 2.5. Then, the CCEVI's of some bivariate copula families are computed.

B.1 Proofs of Conditional EVI Results

Proof of Proposition 2.5.1. Proof by contradiction: suppose there exists a set $\mathscr{X}_0 \subset \mathscr{X}$ of non-zero measure such that $\xi_{Y|\boldsymbol{X}}(\boldsymbol{x}) > \xi_Y \geq 0$ for all $\boldsymbol{x} \in \mathscr{X}_0$, where $\xi_{Y|\boldsymbol{X}}(\boldsymbol{x})$ is the extreme value index of $Y \mid \boldsymbol{X} = \boldsymbol{x}$, assumed to be continuous in $\boldsymbol{x}$ for simplicity. We can further suppose that $\mathscr{X}_0$ is compact, so that $\xi_{Y|\boldsymbol{X}}(\boldsymbol{x}) \geq \xi_Y + \varepsilon > 0$ for some $\varepsilon > 0$.

Computing the marginal distribution from the conditional, we have for $y \in \mathbb{R}$,

$$\begin{aligned} 1 - F_Y(y) &\geq \int_{\mathscr{X}_0} \left(1 - F_{Y|\boldsymbol{X}}\left(y|\boldsymbol{x}\right)\right) \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) \\ &\geq \int_{\mathscr{X}_0} \ell_{\boldsymbol{x}}(y)\, y^{-1/(\xi_Y+\varepsilon)}\, \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) \\ &= \tilde{\ell}(y)\, y^{-1/(\xi_Y+\varepsilon)}, \end{aligned} \tag{B.1.1}$$

where $\ell_{\boldsymbol{x}} \in \mathrm{RV}_0$ for all $\boldsymbol{x} \in \mathscr{X}_0$, and

$$\tilde{\ell}(y) = \int_{\mathscr{X}_0} \ell_{\boldsymbol{x}}(y)\, \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x})$$

exists. Note that $\tilde{\ell}$ is non-zero since $\mathscr{X}_0$ has non-zero measure.

Since $\mathscr{X}_0$ is compact, $\tilde{\ell}$ is slowly varying because each $\ell_{\boldsymbol{x}}$ is: by the Uniform Convergence theorem,

$$\lim_{t\to\infty} \frac{\tilde{\ell}(ty)}{\tilde{\ell}(y)} = \lim_{t\to\infty} \frac{\int_{\mathscr{X}_0} \ell_{\boldsymbol{x}}(ty)\, \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x})}{\int_{\mathscr{X}_0} \ell_{\boldsymbol{x}}(t)\, \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x})} = 1$$

for $y \geq y_0 > 1$ such that $\tilde{\ell}$ is locally integrable over $[y_0, \infty)$.

The inequality in Equation (B.1.1) suggests that Y must be heavy-tailed, so that $\xi_Y = 0$ is not possible. But if $\xi_Y > 0$, we have $1 - F_Y(\cdot) \in \mathrm{RV}_{-1/\xi_Y}$, which decays more quickly than the lower bound in Equation (B.1.1) allows. We conclude that no set $\mathscr{X}_0 \subset \mathscr{X}$ with non-zero measure exists. □

Before proving Proposition 2.5.2, we present a lemma stating that a heavier-tailed distribution always "overtakes" a lighter-tailed distribution.

Lemma B.1.1. *Suppose $F_1 \in \mathcal{D}\left(G_{\xi_1}\right)$ and $F_2 \in \mathcal{D}\left(G_{\xi_2}\right)$ for some $\xi_1 > \xi_2 \geq 0$, so that F_1 is heavier tailed than F_2. Then there exists an $\varepsilon > 0$ such that $F_1^{\leftarrow}(\tau) > F_2^{\leftarrow}(\tau)$ for all $\tau \in (1-\varepsilon, 1)$.*

This result is intuitive, but a proof is shown anyway.

Proof. We have

$$F_1^{\leftarrow}(\tau) = \ell_1(\tau)(1-\tau)^{-\xi_1}$$

and

$$F_2^{\leftarrow}(\tau) = \ell_2(\tau)(1-\tau)^{-\xi_2}$$

for all $\tau \in (1-\varepsilon_1, 1)$ and some $\varepsilon_1 > 0$, where ℓ_1 and ℓ_2 are slowly varying at 1^-. Their ratio is therefore

$$\frac{F_1^{\leftarrow}(\tau)}{F_2^{\leftarrow}(\tau)} = \ell(\tau)(1-\tau)^{-(\xi_1-\xi_2)},$$

where ℓ is slowly varying at 1^-. Since $\xi_1 > \xi_2$, this ratio tends to infinity as $\tau \uparrow 1$, suggesting the existence of an $\varepsilon \in (0, \varepsilon_1]$ such that $F_1^{\leftarrow}(\tau) > F_2^{\leftarrow}(\tau)$ for all $\tau \in (1-\varepsilon, 1)$. □

Now we can prove the main proposition.

Proof of Proposition 2.5.2. Take $0 < t_1 < t_2 < 1$, and define

$$Q_1 = Q_{Y|\boldsymbol{X}}\left(\cdot \mid \tilde{\boldsymbol{x}}(t_1)\right)$$

and

$$Q_2 = Q_{Y|\boldsymbol{X}}\left(\cdot \mid \tilde{\boldsymbol{x}}(t_2)\right).$$

Since the τ-quantile curves are non-decreasing along the path $\tilde{\boldsymbol{x}}$ for all $\tau \in (1-\varepsilon, 1)$, we have $Q_1(\tau) \leq Q_2(\tau)$. By the contrapositive of Lemma B.1.1, $\xi_{Y|\boldsymbol{X}}\left(\tilde{\boldsymbol{x}}(t_1)\right) \leq \xi_{Y|\boldsymbol{X}}\left(\tilde{\boldsymbol{x}}(t_2)\right)$, so that the conditional EVI is non-decreasing from $\tilde{\boldsymbol{x}}(t_1)$ to $\tilde{\boldsymbol{x}}(t_2)$. Since t_1 and t_2 are arbitrary, this must hold for the entire path $\tilde{\boldsymbol{x}}(t)$ for $t \in (0, 1)$. □

Proof of Proposition 2.5.3. Define $\boldsymbol{W} = \boldsymbol{T}(\boldsymbol{X})$, and consider $\boldsymbol{x} \in \mathscr{X}$. Since $\boldsymbol{T}$ is one-to-one, the event $\boldsymbol{X} = \boldsymbol{x}$ is the same as the event $\boldsymbol{W} = \boldsymbol{T}(\boldsymbol{x})$, so that the conditional distributions of $Y \mid \boldsymbol{X} = \boldsymbol{x}$ and $Y \mid \boldsymbol{W} = \boldsymbol{T}(\boldsymbol{x})$—and their respective EVI's $\xi_{Y|\boldsymbol{X}}(\boldsymbol{x})$ and $\xi_{Y|\boldsymbol{W}}\left(\boldsymbol{T}(\boldsymbol{x})\right)$—are the same. This also means that

$$Q_{Y|\boldsymbol{W}}\left(\tau|\boldsymbol{w}\right) = Q_{Y|\boldsymbol{X}}\left(\tau|\boldsymbol{T}^{\leftarrow}(\boldsymbol{w})\right) = \alpha + \boldsymbol{w}^{\top}\boldsymbol{\beta}(\tau)$$

for each $\tau \in (1-\varepsilon, 1)$, so that $Y \mid \boldsymbol{W}$ has linear quantile surfaces. By Wang and Li (2013, Proposition 1), it follows that $\xi_{Y|\boldsymbol{W}} = \xi$ for some constant $\xi > 0$ (they show this for one predictor, but their argument extends to p predictors). Therefore,

$$\xi_{Y|\boldsymbol{X}}(\boldsymbol{x}) = \xi_{Y|\boldsymbol{W}}\left(\boldsymbol{T}(\boldsymbol{x})\right) = \xi,$$

so that $\xi_{Y|\boldsymbol{X}}$ is constant as well. □

Proof of Proposition 2.5.4. The proposition for $\xi_Y \leq 0$ is already proven in Wadsworth et al. (2010, Theorem 1). Consider, then, $\xi_Y > 0$. The right-endpoint is $y^* = \infty$, and since F_Y'' exists and $F_Y' > 0$ in a neighbourhood of y^*, by Wadsworth et al. (2010, Theorem 1),

$$\xi_{\Xi_\lambda(Y)} = \xi_Y + (\lambda - 1) \lim_{y\to\infty} \frac{\left(\frac{1-F_Y(y)}{F_Y'(y)}\right)}{y}. \tag{B.1.2}$$

Now, since $\xi_Y > 0$, we have $1 - F_Y \in \mathrm{RV}_{-1/\xi_Y}$ and $F_Y' \in \mathrm{RV}_{-1/\xi_Y - 1}$. This means

$$\frac{1 - F_Y(y)}{F_Y'(y)} \in \mathrm{RV}_1,$$

so that we can apply L'hôpital's rule to Equation (B.1.2) to obtain

$$\xi_{\Xi_\lambda(Y)} = \xi_Y + (\lambda - 1) \lim_{y\to\infty} \left(\frac{1 - F_Y}{F_Y'}\right)'(y). \tag{B.1.3}$$

Since $\xi_{\Xi_\lambda(Y)} \in \mathbb{R}$, the limit under consideration must exist. By the von Mises condition, the limit in Equation (B.1.3) must be the EVI of F_Y; that is,

$$\xi_{\Xi_\lambda(Y)} = \xi_Y + (\lambda - 1)\xi_Y = \lambda\xi_Y,$$

which completes the proof. □

Proof of Corollary 2.5.5. Since $U_i \sim \mathrm{Uniform}(0,1)$, $1 - U_i^{-1}$ follows a Type I Pareto

distribution with distribution function

$$F_{1-U_i^{-1}}(x) = 1 - x^{-1}$$

for $x > 1$. This Pareto distribution has an EVI of 1. Since the CCEVI ξ_i is defined as the EVI of $1 - U_i^{-1} \mid \boldsymbol{U}_{-i}$, by Proposition 2.5.1, $\xi_i \leq 1$. □

Proof of Theorem 2.5.6. Let $\boldsymbol{U} = \left(F_1(X_1), \ldots, F_p(X_p)\right)^\top$ and $\boldsymbol{u} = \left(F_1(x_1), \ldots, F_p(x_p)\right)^\top$. Since the event $\boldsymbol{X} = \boldsymbol{x}$ is equivalent to the event $\boldsymbol{U} = \boldsymbol{u}$, the EVI of the distribution of $Y \mid \boldsymbol{X} = \boldsymbol{x}$ is the same as the EVI of the distribution of $Y \mid \boldsymbol{U} = \boldsymbol{u}$. We will focus on finding the latter.

Let $V = F_Y(Y)$, and

$$W = \frac{1}{1 - V}, \tag{B.1.4}$$

so that $W \sim \text{Pareto}(1)$. By Sklar's Theorem (Theorem 2.4.1), $\left(\boldsymbol{U}^\top, V\right)$ has distribution function given by the copula C. The EVI of $W \mid \boldsymbol{U} = \boldsymbol{u}$ is just $\xi_C(\boldsymbol{u})$, by definition of the CCEVI. We find $\xi_{Y|\boldsymbol{X}}(\boldsymbol{x})$ by back-transforming W to Y by

$$Y = F_Y^{\leftarrow}\left(1 - \frac{1}{W}\right), \tag{B.1.5}$$

and using a corollary of the von Mises condition (cf. de Haan and Ferreira, 2006, Corollary 1.1.10). Note that $\xi_Y \geq 0$.

We first need the quantile function of $Y \mid \boldsymbol{U} = \boldsymbol{u}$, which, by Equation (B.1.5), is

$$F_{Y|\boldsymbol{U}=\boldsymbol{u}}^{\leftarrow}(\tau) = F_Y^{\leftarrow}\left(1 - \frac{1}{F_{W|\boldsymbol{U}=\boldsymbol{u}}^{\leftarrow}(1 - t^{-1})}\right)$$

for $\tau \in (0, 1)$, where $\tau = 1 - t^{-1}$. Written in terms of the $\left(1 - t^{-1}\right)$-quantile functions[1] for $t > 1$,

$$G_{Y|\boldsymbol{U}=\boldsymbol{u}}(t) = G_Y \circ G_{W|\boldsymbol{U}=\boldsymbol{u}}(t),$$

where $G(t) = F^{\leftarrow}\left(1 - t^{-1}\right)$ for distribution function F.

Next, we need the first and second derivatives of $G_{Y|\boldsymbol{U}=\boldsymbol{u}}$. These derivatives are

$$G'_{Y|\boldsymbol{U}=\boldsymbol{u}}(t) = \left[G'_Y \circ G_{W|\boldsymbol{U}=\boldsymbol{u}}(t)\right]\left[G'_{W|\boldsymbol{U}=\boldsymbol{u}}(t)\right]$$

[1] These functions are sometimes denoted by U (cf. de Haan and Ferreira, 2006), but we use G to avoid overloading $\boldsymbol{U}$.

and

$$G''_{Y|\boldsymbol{U}=\boldsymbol{u}}(t) = \left[G''_Y \circ G_{W|\boldsymbol{U}=\boldsymbol{u}}(t)\right]\left[G'_{W|\boldsymbol{U}=\boldsymbol{u}}(t)\right]^2 + \left[G'_Y \circ G_{W|\boldsymbol{U}=\boldsymbol{u}}(t)\right]\left[G''_{W|\boldsymbol{U}=\boldsymbol{u}}(t)\right].$$

By the corollary to the von Mises condition under consideration (cf. de Haan and Ferreira, 2006, Corollary 1.1.10), we have

$$\lim_{t\to\infty} t\frac{G''_{W|\boldsymbol{U}=\boldsymbol{u}}(t)}{G'_{W|\boldsymbol{U}=\boldsymbol{u}}(t)} = \xi_C(\boldsymbol{u}) - 1 \tag{B.1.6}$$

and

$$\lim_{t\to\infty} t\frac{G''_Y(t)}{G'_Y(t)} = \xi_Y - 1. \tag{B.1.7}$$

Now investigate a similar limit to find $\xi_{Y|\boldsymbol{X}}(\boldsymbol{x})$:

$$\begin{aligned}\lim_{t\to\infty} t\frac{G''_{Y|\boldsymbol{U}=\boldsymbol{u}}(t)}{G'_{Y|\boldsymbol{U}=\boldsymbol{u}}(t)} &= \lim_{t\to\infty} t\frac{\left[G''_Y \circ G_{W|\boldsymbol{U}=\boldsymbol{u}}(t)\right]\left[G'_{W|\boldsymbol{U}=\boldsymbol{u}}(t)\right]}{G'_Y \circ G_{W|\boldsymbol{U}=\boldsymbol{u}}(t)} + \lim_{t\to\infty} t\frac{G''_{W|\boldsymbol{U}=\boldsymbol{u}}(t)}{G'_{W|\boldsymbol{U}=\boldsymbol{u}}(t)} \\ &= \lim_{t\to\infty}\left(G_{W|\boldsymbol{U}=\boldsymbol{u}}(t)\frac{G''_Y \circ G_{W|\boldsymbol{U}=\boldsymbol{u}}(t)}{G'_Y \circ G_{W|\boldsymbol{U}=\boldsymbol{u}}(t)}\right)\left(t\frac{G'_{W|\boldsymbol{U}=\boldsymbol{u}}(t)}{G_{W|\boldsymbol{U}=\boldsymbol{u}}(t)}\right) + \xi_C(\boldsymbol{u}) - 1,\end{aligned} \tag{B.1.8}$$

by Equation (B.1.6). To continue, we consider two cases.

Consider the case where $\xi_C(\boldsymbol{u}) > 0$. Then $G_{W|\boldsymbol{U}=\boldsymbol{u}}(t) \to \infty$ as $t \to \infty$, since $Y \mid \boldsymbol{U} = \boldsymbol{u}$ has infinite right-endpoint. Then,

$$\begin{aligned}\lim_{t\to\infty} t\frac{G''_{Y|\boldsymbol{U}=\boldsymbol{u}}(t)}{G'_{Y|\boldsymbol{U}=\boldsymbol{u}}(t)} &= \left(\lim_{s\to\infty} s\frac{G''_Y(s)}{G'_Y(s)}\right)\left(\lim_{t\to\infty}\frac{tG'_{W|\boldsymbol{U}=\boldsymbol{u}}(t)}{G_{W|\boldsymbol{U}=\boldsymbol{u}}(t)}\right) + \xi_C(\boldsymbol{u}) - 1 \\ &= (\xi_Y - 1)\left(\lim_{t\to\infty}\frac{tG'_{W|\boldsymbol{U}=\boldsymbol{u}}(t)}{G_{W|\boldsymbol{U}=\boldsymbol{u}}(t)}\right) + \xi_C(\boldsymbol{u}) - 1,\end{aligned}$$

by Equation (B.1.7). Noting that $G_{W|\boldsymbol{U}=\boldsymbol{u}}(t) \to \infty$ and $tG'_{W|\boldsymbol{U}=\boldsymbol{u}}(t) \to$ as $t \to \infty$ so that L'hôpital's rule applies (both functions are in $\mathrm{RV}_{\xi_C(\boldsymbol{u})}$), we obtain

$$\begin{aligned}\lim_{t\to\infty} t\frac{G''_{Y|\boldsymbol{U}=\boldsymbol{u}}(t)}{G'_{Y|\boldsymbol{U}=\boldsymbol{u}}(t)} &= (\xi_Y - 1)\left(1 + \lim_{t\to\infty}\frac{tG''_{W|\boldsymbol{U}=\boldsymbol{u}}(t)}{G'_{W|\boldsymbol{U}=\boldsymbol{u}}(t)}\right) + \xi_C(\boldsymbol{u}) - 1 \\ &= (\xi_Y - 1)\left(1 + \xi_C(\boldsymbol{u}) - 1\right) + \xi_C(\boldsymbol{u}) - 1 \\ &= \xi_Y\xi_C(\boldsymbol{u}) - 1,\end{aligned}$$

by Equation (B.1.6). By the von Mises corollary, $\xi_{Y|\boldsymbol{X}}(\boldsymbol{x}) = \xi_Y \xi_C(\boldsymbol{u})$.

Next, consider the case where $\xi_C(\boldsymbol{u}) < 0$. Then the right-endpoint of $W \mid \boldsymbol{U} = \boldsymbol{u}$ is

$$w^* = \lim_{t\to\infty} G_{W|\boldsymbol{U}=\boldsymbol{u}}(t) \in (1, \infty)$$

(it must be greater than 1 due to the definition of W in Equation (B.1.4), but less than infinity, meaning that the copula does not have support on the entire hypercube). Continuing with Equation (B.1.8), the first limit is

$$\lim_{t\to\infty} G_{W|\boldsymbol{U}=\boldsymbol{u}}(t) \frac{G''_Y \circ G_{W|\boldsymbol{U}=\boldsymbol{u}}(t)}{G'_Y \circ G_{W|\boldsymbol{U}=\boldsymbol{u}}(t)} = w^* \frac{G''_Y(w^*)}{G'_Y(w^*)} \in \mathbb{R},$$

since $G'_Y > 0$ and G''_Y exists. To evaluate the second limit, notice that $G'_{W|\boldsymbol{U}=\boldsymbol{u}} \in \mathrm{RV}_{\xi_C(\boldsymbol{u})-1}$ and $G_{W|\boldsymbol{U}=\boldsymbol{u}} \in \mathrm{RV}_0$, so that the second limit is the limit of a function in $\mathrm{RV}_{\xi_C(\boldsymbol{u})}$:

$$\lim_{t\to\infty} t \frac{G'_{W|\boldsymbol{U}=\boldsymbol{u}}(t)}{G_{W|\boldsymbol{U}=\boldsymbol{u}}(t)} = 0.$$

Continuing from Equation (B.1.8), we obtain

$$\lim_{t\to\infty} t \frac{G''_{Y|\boldsymbol{U}=\boldsymbol{u}}(t)}{G'_{Y|\boldsymbol{U}=\boldsymbol{u}}(t)} = \xi_C(\boldsymbol{u}) - 1.$$

By the von Mises corollary, $\xi_{Y|\boldsymbol{X}}(\boldsymbol{x}) = \xi_C(\boldsymbol{u})$. □

Proof of Proposition 2.4.2. Take $u \in \mathscr{U}$, and use the limit in Equation (2.4.3) to obtain

$$s^{1-1/\xi_C(u)} = \frac{c^*}{c^*} = 1,$$

so that $\xi_C(u) = 1$. □

B.2 Derivation of CCEVI's

The CCEVI's of the Gaussian, Frank, and Gumbel copulas are provided here.

B.2.1 Gaussian Copula

The Gaussian copula with dependence parameter $\rho \in [-1, 1]$ is

$$C_{\text{Gauss}}(u, v; \rho) = \Phi_2\left(\Phi^{\leftarrow}(u), \Phi^{\leftarrow}(v); \rho\right),$$

where $\Phi^{\leftarrow}$ is the probit function, and Φ_2 is the cdf of the bivariate normal distribution,

$$\Phi_2(x, y; \rho) = \int_{-\infty}^{x} {}_{-\infty}^{y} \phi_2(s, t; \rho) \, \mathrm{d}t \, \mathrm{d}s$$

and

$$\phi_2(x, y; \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{y^2 - 2\rho xy + x^2}{2(1-\rho^2)}\right)$$

for $(x, y) \in \mathbb{R}^2$.

We find the CCEVI through the copula density. Letting $x = \Phi^{\leftarrow}(u)$ and $y = \Phi^{\leftarrow}(v)$, the copula density is

$$\begin{aligned} c_{\text{Gauss}}(u, v; \rho) &= \frac{\phi_2(x, y; \rho)}{\phi(x)\,\phi(y)} \\ &= \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right) \exp\left(\frac{x^2}{2}\right) \exp\left(\frac{y^2}{2}\right). \end{aligned}$$

Viewing x and ρ as constants, the density can be further decomposed:

$$\begin{aligned} c_{\text{Gauss}}(u, v; \rho) &\propto \exp\left(-\frac{-2\rho xy + y^2}{2(1-\rho^2)} + \frac{y^2}{2}\right) \\ &= \left[\exp\left(2\rho xy - y^2 + y^2\left(1-\rho^2\right)\right)\right]^{1/\left(2\left(1-\rho^2\right)\right)} \\ &= \left[\exp\left(-\left(\rho y - x\right)^2\right)\right]^{1/\left(2\left(1-\rho^2\right)\right)}. \end{aligned}$$

Next, consider the regular variation of the function

$$h : t \mapsto \exp\left\{\left(\rho\Phi^{\leftarrow}\left(1 - \frac{1}{t}\right) - x\right)^2\right\}$$

in some neighbourhood of infinity. This is easier to explore through its inverse

$$h^{\leftarrow} : t \mapsto \frac{1}{1 - \Phi\left(\frac{\sqrt{\log t} + x}{\rho}\right)},$$

also in a neighbourhood of infinity. For $s > 0$, we have

$$\begin{aligned}
\lim_{t\to\infty} \frac{h^{\leftarrow}(ts)}{h^{\leftarrow}(t)} &= \lim_{t\to\infty} \frac{1-\Phi\left(\frac{\sqrt{\log t}+x}{\rho}\right)}{1-\Phi\left(\frac{\sqrt{\log(st)}+x}{\rho}\right)} \\
&= \lim_{t\to\infty} \frac{\phi\left(\frac{\sqrt{\log t}+x}{\rho}\right)\frac{1}{\rho}\frac{1}{2}(\log t)^{-1/2}\frac{1}{t}}{\phi\left(\frac{\sqrt{\log(st)}+x}{\rho}\right)\frac{1}{\rho}\frac{1}{2}\left(\log(st)\right)^{-1/2}\frac{1}{st}s} \\
&= \lim_{t\to\infty} \frac{\phi\left(\frac{\sqrt{\log t}+x}{\rho}\right)}{\phi\left(\frac{\sqrt{\log(st)}+x}{\rho}\right)} \\
&= \left[\lim_{t\to\infty} \exp\left\{-\left(\sqrt{\log t}+x\right)^2+\left(\sqrt{\log(st)}+x\right)^2\right\}\right]^{1/\left(2\rho^2\right)} \\
&= \left[s\lim_{t\to\infty} \exp\left\{2x\sqrt{\log(st)}-2x\sqrt{\log t}\right\}\right]^{1/\left(2\rho^2\right)} \\
&= s^{1/\left(2\rho^2\right)},
\end{aligned}$$

so that $h^{\leftarrow} \in \mathrm{RV}_{1/\left(2\rho^2\right)}$ at infinity. This means that $h \in \mathrm{RV}_{2\rho^2}$.

Lastly, the CCEVI of the Gaussian copula can be found by investigating the following limit for $s > 0$:

$$\begin{aligned}
s^{-1}\lim_{t\to\infty} \frac{c_{\text{Gauss}}\left(u, 1-(st)^{-1};\rho\right)}{c_{\text{Gauss}}\left(u, 1-t^{-1};\rho\right)} &= s^{-1}\left[\lim_{t\to\infty}\frac{h(ts)}{h(t)}\right]^{-1/\left(2\left(1-\rho^2\right)\right)} \\
&= s^{-1}\left[s^{-2\rho^2}\right]^{-1/\left(2\left(1-\rho^2\right)\right)} \\
&= s^{-1/\left(1-\rho^2\right)}.
\end{aligned}$$

The CCEVI of the Gaussian copula family is therefore $1-\rho^2$. Since a Gaussian copula is reflection-symmetric, the reflection copula family shares the same CCEVI.

B.2.2 Frank Copula

The Frank copula family is defined by

$$C_{\text{Frk}}(u,v;\theta) = -\frac{1}{\theta}\log\left(1-\frac{\left(1-\exp(-\theta u)\right)\left(1-\exp(-\theta v)\right)}{1-\exp(-\theta)}\right)$$

for $\theta \in \mathbb{R}$ (the independence copula is obtained when $\theta = 0$). We find the CCEVI through the copula density.

Fixing θ and u, let $x = 1 - \exp(-\theta u)$ and $y = 1 - \exp(-\theta v)$ so that the copula density is

$$c_{\text{Frk}}(u, v; \theta) \propto \left(1 - \exp(-\theta) - xy\right)^{-2} \exp(-\theta v).$$

Evaluated at $v = 1$ (and $y = 1 - \exp(-\theta)$), we have

$$c_{\text{Frk}}(u, v; \theta) \propto \left(1 - \exp(-\theta)\right)^{-2} (1 - x)^{-2} \exp(-\theta),$$

which is non-zero (so long as $\theta \neq 0$, in which case we obtain the independence copula anyway). To find the CCEVI, consider the following limit for $s > 0$:

$$s^{-1} \lim_{t \to \infty} \frac{c_{\text{Frk}}\left(u, 1 - (st)^{-1}; \theta\right)}{c_{\text{Frk}}(u, 1 - t^{-1}; \theta)} = s^{-1} \lim_{t \to \infty} \frac{c_{\text{Frk}}(u, 1; \theta)}{c_{\text{Frk}}(u, 1; \theta)} = s^{-1}.$$

The CCEVI of the Frank copula family is therefore 1. This is also the CCEVI of the reflected Frank copula family due to the family's reflection symmetry.

B.2.3 Gumbel Copula

For the Gumbel copula family, we have for $(u, v) \in (0, 1)^2$ and $\theta > 1$,

$$C_{\text{Gum}}(u, v; \theta) = \exp\left\{-\psi_{\text{Gum}}(x, y; \theta)\right\},$$

where

$$\psi_{\text{Gum}}(x, y; \theta) = \left(x^\theta + y^\theta\right)^{1/\theta},$$

and $x = -\log u$, $y = -\log v$. The density is

$$\begin{aligned} c_{\text{Gum}}(u, v; \theta) &= \frac{C_{\text{Gum}}(u, v; \theta)}{uv} \left(\mathrm{D}_2\, \psi_{\text{Gum}}(x, y; \theta) - \mathrm{D}_{12}\, \psi_{\text{Gum}}(x, y; \theta)\right) \\ &= \frac{C_{\text{Gum}}(u, v; \theta)}{uv} y^{\theta-1} \left[\left(x^\theta + y^\theta\right)^{-1+1/\theta} + \left(x^\theta + y^\theta\right)^{-2+1/\theta} x^{\theta-1}(\theta - 1)\right] \\ &\sim y^{\theta-1}\left(x^{-\theta+1} + x^{-\theta}(\theta - 1)\right) \end{aligned}$$

as $v \uparrow 1$ (simply substitute $v = 1$ and $y = 0$ everywhere except the $y^{\theta-1}$ term).

To find the CCEVI, investigate the following limit for $s > 0$:

$$\begin{aligned} s^{-1} \lim_{t\to\infty} \frac{c_{\text{Gum}}\left(u, 1-(ts)^{-1};\theta\right)}{c_{\text{Gum}}\left(u, 1-(ts)^{-1};\theta\right)} &= s^{-1} \lim_{t\to\infty} \frac{(ts)^{-\theta+1}}{t^{-\theta+1}} \\ &= s^{-\theta}, \end{aligned}$$

since $-\log\left(1-(ts)^{-1}\right) \sim (ts)^{-1}$ as $t \to \infty$. The CCEVI is therefore $1/\theta$.

To find the CCEVI of the survival copula, first notice that

$$c_{\text{Gum}}(u, v; \theta) \sim \frac{C_{\text{Gum}}(u, v; \theta)}{uv}$$

as $v \downarrow 0$ (so $y \to \infty$). Now investigate the following limit for $s > 0$, this time taking $x = -\log(1-u)$:

$$\begin{aligned} s^{-1} \lim_{t\to\infty} \frac{\hat{c}_{\text{Gum}}\left(u, 1-(ts)^{-1};\theta\right)}{\hat{c}_{\text{Gum}}\left(u, 1-(ts)^{-1};\theta\right)} &= s^{-1} \lim_{t\to\infty} \frac{c_{\text{Gum}}\left(1-u, (ts)^{-1};\theta\right)}{c_{\text{Gum}}\left(1-u, (ts)^{-1};\theta\right)} \\ &= s^{-1} \lim_{t\to\infty} \frac{C_{\text{Gum}}\left(1-u, (ts)^{-1};\theta\right) ts}{C_{\text{Gum}}\left(1-u, t^{-1};\theta\right) t} \\ &= \exp\left\{\lim_{t\to\infty}\left[\left(x^{\theta} + (\log t)^{\theta}\right)^{1/\theta} - \left(x^{\theta} + \left(\log(ts)\right)^{\theta}\right)^{1/\theta}\right]\right\} \\ &= \exp\{-\log s\} \\ &= s^{-1}. \end{aligned}$$

The CCEVI of the survival Gumbel family is therefore 1.

Appendix C

Proofs related to Proper Scoring Rules

This appendix provides the proofs related to propriety and tail balance, discussed in Section 3.2.

C.1 Propriety

Here are the regularity conditions required for Theorem 3.2.1 in Section 3.2.2, followed by a proof of the theorem. Recall that the transformation function $g : \mathbb{R} \to \mathbb{R}$ is non-decreasing, the weight function $w : (\tau_c, 1) \to [0, \infty)$, and $\varphi(x; \tau) = g(x)\, w(\tau)$ for $(x, \tau) \in \mathbb{R} \times (\tau_c, 1)$ – all of which are fixed.

1. The function $\tau \mapsto \varphi\left(\hat{Q}(\tau); \tau\right)$ is continuous almost everywhere for all $\hat{Q} \in \hat{\mathcal{Q}}$.

2. $w \in \mathrm{RV}_\alpha$ at 1^-, satisfying $\alpha + \hat{\xi}^+ < 1$ for all $\hat{\xi} \in \hat{\Xi}_g$, where $\hat{\xi}^+ = \max\left(0, \hat{\xi}\right)$, and $\hat{\Xi}_g$ exists and is the set of extreme value indices of the distribution functions $\left\{\hat{F} \circ g^{\leftarrow} : \hat{F} \in \hat{\mathcal{F}}\right\}$.

3. The mean
$$\mu = \mathbb{E}\left(g(Y)\right) = \int g(y)\, \mathrm{d}F(y)$$
exists for all $F \in \mathcal{F}$.

Note that Condition 2 can be relaxed if the response never exceeds the support of the forecasts. If the right endpoints $\lim_{\tau \uparrow 1} \hat{Q}(\tau) \geq \lim_{\tau \uparrow 1} Q(\tau)$ (which is either finite or infinite) for all $\hat{Q} \in \hat{\mathcal{Q}}$ and $Q \in \mathcal{Q}$, then the condition can be relaxed to

2'. $w \in \mathrm{RV}_\alpha$ at 1^- , satisfying $\alpha + \hat{\xi}^+ < 2$ for all $\hat{\xi} \in \hat{\Xi}_g$, where $\hat{\Xi}_g$ exists and is the set of extreme value indices of the distribution functions $\left\{\hat{F} \circ g^{\leftarrow} : \hat{F} \in \hat{\mathcal{F}}\right\}$.

Proof of Theorem 3.2.1 – Existence. Let $\hat{Q} \in \hat{\mathcal{Q}}$ and $y \in \mathscr{Y}$. To show the existence of S in Equation (3.2.8), we must show that the integral

$$\mathcal{I}_{\varphi} := \int_{\tau_c}^{1} \left(\tau - \mathbb{I}_{\left(-\infty, \hat{Q}(\tau)\right)}(y)\right) \left(\varphi(y;\tau) - \varphi\left(\hat{Q}(\tau);\tau\right)\right) \mathrm{d}\tau \tag{C.1.1}$$

exists. As a preliminary, Condition 1 is needed to satisfy Lebesgue's integrability condition so that the integral is well-defined.

It is important to note that Condition 2 implies that $\alpha < 1$, since $\hat{\xi}^{+} \geq 0$ for all $\hat{\xi} \in \mathbb{R}$ whether or not $\hat{Q}$ has a finite right-endpoint. Similarly, Condition 2' implies that $\alpha < 2$.

We define $\hat{q}^{*} = \lim_{\tau \uparrow 1} \hat{Q}(\tau)$ as the right-endpoint of the forecast distribution, which is possibly infinity. The proof involves three cases.

- **Case 1:** $y \geq \hat{q}^{*}$.
- **Case 2:** $y < \hat{q}^{*}$, and $g \circ \hat{Q}$ is bounded above by some $\hat{g}^{*} \in \mathbb{R}$.
- **Case 3:** $y < \hat{q}^{*}$, and $g \circ \hat{Q}$ is not bounded above.

First, consider **Case 1**. The integral $\mathcal{I}_{\varphi}$ in Equation (C.1.1) becomes

$$\int_{\tau_c}^{1} w(\tau)\, \tau \left(g(y) - g \circ \hat{Q}(\tau)\right) \mathrm{d}\tau.$$

Since $g \circ \hat{Q}(\tau)$ is bounded below by $\hat{g}_{*} := \lim_{\tau \downarrow \tau_c} g \circ \hat{Q}(\tau)$ and above by $g(y)$, $\mathcal{I}_{\varphi}$ exists if

$$\int_{\tau_c}^{1} w(\tau) \left(g(y) - g \circ \hat{Q}(\tau_c)\right) \mathrm{d}\tau = \left(g(y) - \hat{g}_{*}\right) \int_{\tau_c}^{1} w(\tau)\, \mathrm{d}\tau$$

exists, and this holds by Condition 2. However, if $\lim_{\tau \uparrow 1} Q(\tau) \leq \hat{q}^{*}$ for all $Q \in \mathcal{Q}$, then observing $y \geq \hat{q}^{*}$ cannot happen. This would happen if all forecasts have an infinite right-endpoint.

Now, all the other cases have $y < \hat{q}^{*}$, meaning that there exists a $\tau_L \in (\tau_c, 1)$ such that $\hat{Q}(\tau) > y$ for all $\tau \in (\tau_L, 1)$. We now investigate the existence of $\mathcal{I}_{\varphi}$ after replacing the lower limit of integration with τ_L to obtain

$$\mathcal{I}_{2\varphi} := \int_{\tau_L}^{1} (1-\tau) \left(\varphi\left(\hat{Q}(\tau);\tau\right) - \varphi(y;\tau)\right) \mathrm{d}\tau = \int_{\tau_L}^{1} w(\tau)\,(1-\tau) \left(g \circ \hat{Q}(\tau) - g(y)\right) \mathrm{d}\tau, \tag{C.1.2}$$

whose existence implies the existence of the original integral $\mathcal{I}_{\varphi}$.

Next, consider **Case 2**. Since $g \circ \hat{Q}(\tau) \leq \hat{g}^*$, the integral $\mathcal{I}_{2\varphi}$ exists if

$$\left(\hat{g}^* - g(y)\right) \int_{\tau_L}^{1} w(\tau)(1-\tau)\,\mathrm{d}\tau$$

does, happening whenever $w \in \mathrm{RV}_\alpha$ at 1^- with $\alpha < 2$ – true by Condition 2'.

Finally, consider **Case 3**. Since $g \circ \hat{Q}(\tau) \to \infty$ as $\tau \uparrow 1$, the integral $\mathcal{I}_{2\varphi}$ exists if

$$\int_{\tau_L}^{1} w(\tau)(1-\tau)\, g \circ \hat{Q}(\tau)\,\mathrm{d}\tau$$

does. By Condition 2', $g \circ \hat{Q}$ is the quantile function of a distribution in some domain of attraction (because the existence of $\hat{\Xi}_g$ is assumed), and its extreme value index is $\hat{\xi} \geq 0$ since $g \circ \hat{Q}$ is not bounded above. It follows that $g \circ \hat{Q} \in \mathrm{RV}_{\hat{\xi}^+}$. Also by Condition 2',

$$\tau \mapsto w(\tau)(1-\tau)\, g \circ \hat{Q}(\tau) \in \mathrm{RV}_{\alpha-1+\hat{\xi}^+}$$

at 1^- where $\alpha - 1 + \hat{\xi}^+ < 1$, and this guarantees the existence of the integral.

□

Proof of Theorem 3.2.1 – Propriety. To show propriety (Definition 3.2.1), the goal is to show that for any fixed $\hat{Q} \in \hat{\mathcal{Q}}$ and $Q \in \mathcal{Q}$, the expected difference of scores satisfies

$$\int \left(S(y, Q; \varphi) - S\left(y, \hat{Q}; \varphi\right)\right) \mathrm{d}F(y) \leq 0, \tag{C.1.3}$$

where $F = Q^{\leftarrow}$.

We first must show that the expected score exists, so that the left-hand side of Equation (C.1.3) exists. We consider the existence of the expectation

$$\int S\left(y, \hat{Q}; \varphi\right) \mathrm{d}F(y) = \int_{\tau_c}^{1} \mathbb{E}_Y\left(s_\tau\left(Y, \hat{Q}(\tau); \varphi\right)\right) \mathrm{d}\tau, \tag{C.1.4}$$

and begin by examining the expected single-quantile score (note that $\hat{Q}$ is not random here, and the subscript Y is used for indicating this). Denoting

$$\hat{\tau}(\tau) = \mathcal{P}\left(Y < \hat{Q}(\tau)\right) \tag{C.1.5}$$

and

$$\mu_s\left(\tau; \hat{Q}(\tau)\right) = \mathbb{E}_Y\left(s_\tau\left(Y, \hat{Q}(\tau); \varphi\right)\right),$$

from the single-quantile score in Equation (3.2.5), we have

$$
\begin{aligned}
&\mu_s\left(\tau;\hat{Q}\left(\tau\right)\right)\\
&=\int w\left(\tau\right)\rho_\tau\left(g\left(y\right)-g\circ\hat{Q}\left(\tau\right)\right)\mathrm{d}F\left(y\right) && \text{(C.1.6)}\\
&=w\left(\tau\right)\int\left[\left(\tau-\mathbb{I}_{\left(-\infty,\hat{Q}(\tau)\right)}\left(y\right)\right)g\left(y\right)-\left(\tau-\mathbb{I}_{\left(-\infty,\hat{Q}(\tau)\right)}\left(y\right)\right)g\circ\hat{Q}\left(\tau\right)\right]\mathrm{d}F\left(y\right)\\
&=w\left(\tau\right)\left[\left(\tau\mu-\int_{-\infty}^{\hat{Q}(\tau)}g\left(y\right)\mathrm{d}F\left(y\right)\right)-g\circ\hat{Q}\left(\tau\right)\left(\tau-\hat{\tau}\left(\tau\right)\right)\right]\\
&=w\left(\tau\right)\tau\left(\mu-\hat{\mu}_L\left(\tau\right)\right)-\varphi\left(\hat{Q}\left(\tau\right);\tau\right)\left(\tau-\hat{\tau}\left(\tau\right)\right), && \text{(C.1.7)}
\end{aligned}
$$

where $\mu=\mathbb{E}\left(g\left(Y\right)\right)$ and

$$\hat{\mu}_L\left(\tau\right)=\frac{1}{\tau}\int_{-\infty}^{\hat{Q}(\tau)}g\left(y\right)\mathrm{d}F\left(y\right).$$

Denoting $\hat{q}^*=\lim_{\tau\uparrow 1}\hat{Q}\left(\tau\right)$ and $q^*=\lim_{\tau\uparrow 1}Q\left(\tau\right)$, notice that

$$\lim_{\tau\uparrow 1}\hat{\mu}_L\left(\tau\right)=\begin{cases}\mu, & \hat{q}^*\geq q^*;\\ \mu-\varepsilon, & \hat{q}^*<q^*\end{cases} \tag{C.1.8}$$

for some $\varepsilon>0$. Condition 3 then guarantees the existence of μ and $\hat{\mu}_L\left(\tau\right)$, which extends to the existence of $\mu_s\left(\tau;\hat{Q}\left(\tau\right)\right)$.

We need only show existence of the integral on the right-hand side of Equation (C.1.4) by using the expression in Equation (C.1.7).

First, investigate the existence of

$$\int_{\tau_c}^{1}w\left(\tau\right)\tau\left(\mu-\hat{\mu}_L\left(\tau\right)\right)\mathrm{d}\tau.$$

Since $\tau\left(\mu-\hat{\mu}_L\left(\tau\right)\right)\leq\mu$ for $\tau\in\left(\tau_c,1\right)$, we have

$$\int_{\tau_c}^{1}w\left(\tau\right)\tau\left(\mu-\hat{\mu}_L\left(\tau\right)\right)\mathrm{d}\tau\leq\mu\int_{\tau_c}^{1}w\left(\tau\right)\mathrm{d}\tau,$$

which exists since $w\in\mathrm{RV}_\alpha$ at 1^- where $\alpha<1$ by Condition 2.

Secondly, we show that

$$\int_{\tau_c}^{1} \varphi\left(\hat{Q}\left(\tau\right);\tau\right)\left(\tau-\hat{\tau}\left(\tau\right)\right)\mathrm{d}\tau$$

exists by showing that

$$\int_{\tau_c}^{1} w\left(\tau\right)\left|g\circ\hat{Q}\left(\tau\right)\right|\left|\tau-\hat{\tau}\left(\tau\right)\right|\mathrm{d}\tau \tag{C.1.9}$$

exists. Since $0 \leq \left|\tau-\hat{\tau}\left(\tau\right)\right| \leq 1$ for all $\tau \in (\tau_c, 1)$, and $\int_{\tau_c}^{1} w\left(\tau\right)\left|g\circ\hat{Q}\left(\tau\right)\right|\mathrm{d}\tau$ exists due to Condition 2, it follows that the desired integral in Equation (C.1.9) exists.

Now that we have shown existence of the expected score, we turn to showing propriety. We adopt a more rigorous notation to aid the discussion. For $k = 1, \ldots, K$, take $\tau_{k:K} = \tau_c + (1-\tau_c)(2k-1)/(2K)$, and $\boldsymbol{\tau}_K = (\tau_{1:K}, \ldots, \tau_{K:K})$. Denote $\boldsymbol{q}_K = \left(Q\left(\tau_{1:K}\right), \ldots, Q\left(\tau_{K:K}\right)\right)$ and $\hat{\boldsymbol{q}}_K = \left(\hat{Q}\left(\tau_{1:K}\right), \ldots, \hat{Q}\left(\tau_{K:K}\right)\right)$. We know from Gneiting and Raftery (2007, Corollary 1) that S_c—defined in Equation (3.2.7)—is a proper scoring rule, so that

$$\int \left(S_c\left(y, \boldsymbol{q}_K; \varphi, \boldsymbol{\tau}_K\right) - S_c\left(y, \hat{\boldsymbol{q}}_K; \varphi, \boldsymbol{\tau}_K\right)\right)\mathrm{d}F\left(y\right) \leq 0. \tag{C.1.10}$$

Also, we have convergence of the Riemann sum:

$$\lim_{K\to\infty}\left(S_c\left(y, \boldsymbol{q}_K; \varphi, \boldsymbol{\tau}_k\right) - S_c\left(y, \hat{\boldsymbol{q}}_K; \varphi, \boldsymbol{\tau}_k\right)\right) = S\left(y, Q; \varphi\right) - S\left(y, \hat{Q}; \varphi\right)$$

uniformly over $y \in \mathscr{Y}$. Therefore, by the Dominated Convergence Theorem, it follows that

$$\begin{aligned}&\int \left(S\left(y, Q; \varphi\right) - S\left(y, \hat{Q}; \varphi\right)\right)\mathrm{d}F\left(y\right)\\ &= \lim_{K\to\infty}\int \left(S_c\left(y, \boldsymbol{q}_K; \varphi, \boldsymbol{\tau}_k\right) - S_c\left(y, \hat{\boldsymbol{q}}_K; \varphi, \boldsymbol{\tau}_k\right)\right)\mathrm{d}F\left(y\right).\end{aligned} \tag{C.1.11}$$

The desired inequality in Equation (C.1.3) follows by the bound in Equation (C.1.10).

□

C.2 Tail Balance

Here is a proof of the expected score formula in Equation (3.2.9).

Proof of Lemma 3.2.2. We seek the expectation of

$$s_\tau\left(Y, Q\left(\tau\right); \varphi\right) = w\left(\tau\right)\left(\tau - \mathbb{I}_{\left(-\infty, Q(\tau)\right)}\left(Y\right)\right)\left(g\left(Y\right) - g\left(Q\left(\tau\right)\right)\right)$$

for almost all $\tau \in (\tau_c, 1)$ when Y has distribution function F, and $Q = F^{\leftarrow}$. Since

$$\mathbb{E}\left(\mathbb{I}_{\left(-\infty, Q(\tau)\right)}\left(Y\right)\right) = \tau$$

for almost all $\tau \in (\tau_c, 1)$,

$$\begin{aligned}\mathbb{E}\left(s_\tau\left(Y, Q\left(\tau\right); \varphi\right)\right) &= w\left(\tau\right)\mathbb{E}\left(\left(\tau - \mathbb{I}_{\left(-\infty, Q(\tau)\right)}\left(Y\right)\right) g\left(Y\right)\right) \\ &= w\left(\tau\right)\left[\tau\mu - \int_{-\infty}^{Q(\tau)} g\left(y\right) \mathrm{d}F\left(y\right)\right], \end{aligned} \tag{C.2.1}$$

where $\mu = \mathbb{E}\left(g\left(Y\right)\right)$. Now, notice that the distribution function of $Y \mid Y < Q\left(\tau\right)$ is just the original distribution function of Y adjusted by τ:

$$F_{Y|Y<Q(\tau)}\left(y\right) = \frac{\mathcal{P}\left(Y \leq y, Y < Q\left(\tau\right)\right)}{\mathcal{P}\left(Y < Q\left(\tau\right)\right)} = \frac{F\left(y\right)}{\tau},$$

$y < Q\left(\tau\right)$, for almost all $\tau \in (\tau_c, 1)$. The conditional expectation $\mu_L\left(\tau\right) = \mathbb{E}\left(g\left(Y\right) \mid Y < Q\left(\tau\right)\right)$ is thus

$$\mu_L\left(\tau\right) = \frac{1}{\tau}\int_{-\infty}^{Q(\tau)} g\left(y\right) \mathrm{d}F\left(y\right).$$

Factoring τ out of the square brackets in Equation (C.2.1) gives the desired result.

□

Here is a proof of the tail stability theorem.

Proof of Proposition 3.2.3. Let $u > 0$. To find the index of variation, we must compute,

using Equation (3.2.9),

$$\lim_{t\to\infty} \frac{\mathbb{E}\left(s_{1-(ut)^{-1}}\left(Y,Q;\varphi\right)\right)}{\mathbb{E}\left(s_{1-t^{-1}}\left(Y,Q;\varphi\right)\right)} = \lim_{t\to\infty} \frac{w\left(1-\frac{1}{ut}\right)\left(1-\frac{1}{ut}\right)\left(\mu-\mu_L\left(1-\frac{1}{ut}\right)\right)}{w\left(1-\frac{1}{t}\right)\left(1-\frac{1}{t}\right)\left(\mu-\mu_L\left(1-\frac{1}{t}\right)\right)} \tag{C.2.2}$$

$$= u^{\alpha} \lim_{t\to\infty} \frac{\mu-\mu_L\left(1-\frac{1}{ut}\right)}{\mu-\mu_L\left(1-\frac{1}{t}\right)} \tag{C.2.3}$$

$$= u^{\alpha-1} \lim_{t\to\infty} \frac{\mu_L'\left(1-\frac{1}{ut}\right)}{\mu_L'\left(1-\frac{1}{t}\right)}, \tag{C.2.4}$$

where μ_L' is the derivative of μ_L (the last equality holds due to L'Hôpital's rule). Note that the expectation exists because $\xi < 1$. We seek an asymptotic form for μ_L'.

Computed using the quantile function, the lower mean can be written

$$\mu_L\left(\tau\right) = \frac{1}{\tau}\int_0^{\tau} g\left(Q\left(p\right)\right)\mathrm{d}p. \tag{C.2.5}$$

Differentiating, we obtain

$$\mu_L'\left(\tau\right) = \frac{1}{\tau} g\circ Q\left(\tau\right) - \frac{1}{\tau^2}\int_0^{\tau} g\circ Q\left(p\right)\mathrm{d}p,$$

which exists due to the continuity of Q. But as $\tau \uparrow 1$,

$$\int_0^{\tau} g\circ Q\left(p\right)\mathrm{d}p \to \mu$$

and

$$g\circ Q\left(\tau\right) \to \begin{cases} \infty, & \xi \geq 0; \\ g^*, & \xi < 0, \end{cases}$$

for right-endpoint $g^* \in \mathbb{R}$. For the case where $\xi < 0$, we have $\mu_L'\left(\tau\right) \to g^* - \mu$ as $\tau \uparrow 1$, so that the limit in Equation (C.2.5) becomes $u^{\alpha-1}$. For the case where $\xi \geq 0$, we have

$$\mu_L'\left(\tau\right) \sim \frac{1}{\tau} g\circ Q\left(\tau\right)$$

as $\tau \uparrow 1$, so that the limit in Equation (C.2.2) becomes

$$u^{\alpha-1} \lim_{t\to\infty} \frac{\left(1-\frac{1}{t}\right) g\circ Q\left(1-\frac{1}{ut}\right)}{\left(1-\frac{1}{ut}\right) g\circ Q\left(1-\frac{1}{t}\right)} = u^{\alpha-1} \lim_{t\to\infty} \frac{g\left(Q\left(1-\frac{1}{ut}\right)\right)}{g\left(Q\left(1-\frac{1}{t}\right)\right)}.$$

The result follows after recognizing that $g \circ Q \in \mathrm{RV}_\xi$ at 1^-, since $F \circ g^{\leftarrow}$ has extreme value index $\xi \geq 0$.

□

Finally, we derive the equations of the expected single-quantile score found in Table 3.1 by computing μ_L.

Gaussian

Computing the lower conditional mean using the density, we have

$$\mu_L(\tau) = \frac{1}{\tau} \int_{-\infty}^{\mu+\sigma\Phi^{\leftarrow}(\tau)} y\phi\left(\frac{y-\mu}{\sigma}\right) \mathrm{d}y,$$

and through a change of variables to $z = \mu + \sigma y$, we obtain

$$\mu_L(\tau) = \frac{1}{\tau} \int_{-\infty}^{\Phi^{\leftarrow}(\tau)} (\sigma z + \mu)\, \phi(z)\, \mathrm{d}z.$$

The integration is made easier by noting that $-y\phi(y) = \phi'(y)$. Separating the integral into two terms and integrating, we obtain

$$\begin{aligned} \mu_L(\tau) &= \frac{1}{\tau}\left[\sigma\left(-\phi\left(\Phi^{\leftarrow}(\tau)\right)\right) + \mu\tau\right] \\ &= \mu - \sigma\frac{\phi\left(\Phi^{\leftarrow}(\tau)\right)}{\tau}, \end{aligned}$$

leading to the desired result.

Exponential

The quantile function under consideration is $Q_Y(\tau) = \mu \log\left(\frac{1}{1-\tau}\right)$. Now,

$$\mu_L(\tau) = \frac{1}{\tau} \int_0^{Q_Y(\tau)} y f_Y(y)\, \mathrm{d}y.$$

Using integration by parts (and integrating f_Y to its survival function $\bar{F}_Y(y) = 1 - F_Y(y) = \exp\left(-y/\mu\right)$ for $y \geq 0$), we obtain

$$\mu_L(\tau) = \frac{1}{\tau}\left[-Q_Y(\tau)(1-\tau) + \int_0^{Q_Y(\tau)} \bar{F}_Y(y)\, \mathrm{d}y\right].$$

The integration is made easier by noting that $-\frac{1}{\mu}\bar{F}_Y(y) = \bar{F}'_Y(y)$. Continuing,

$$\begin{aligned}\mu_L(\tau) &= \frac{1}{\tau}\left[-Q_Y(\tau)(1-\tau)+\mu\tau\right] \\ &= \mu - \mu\frac{1-\tau}{\tau}\log\left(\frac{1}{1-\tau}\right),\end{aligned}$$

leading to the desired result after recognizing that $\mathbb{E}(Y) = \mu$.

Type I Pareto

The quantile function under consideration is $Q_Y(\tau) = \sigma(1-\tau)^{-\xi}$, and the density is $f_Y(y) = \frac{1}{\xi}\sigma^{1/\xi}y^{-1/\xi-1}$ for $y > \sigma$ and $0 < \xi < 1$. Now,

$$\begin{aligned}\mu_L(\tau) &= \frac{1}{\tau}\int_{\sigma}^{Q_Y(\tau)} y f_Y(y)\,\mathrm{d}y. \\ &= \frac{1}{\tau}\frac{1}{\xi}\sigma^{1/\xi}\int_{\sigma}^{Q_Y(\tau)} y^{-1/\xi}\,\mathrm{d}y \\ &= \frac{1}{\tau}\frac{1}{\xi}\sigma^{1/\xi}\left(\frac{\xi}{\xi-1}\right)\left[\left(Q_Y(\tau)\right)^{1-1/\xi} - \sigma^{1-1/\xi}\right] \\ &= \frac{1}{\tau}\frac{1}{\xi-1}\sigma\left[(1-\tau)^{1-\xi}-1\right].\end{aligned}$$

The mean of a Pareto distribution exists when $0 < \xi < 1$ and is $\mu = \mathbb{E}(Y) = \sigma/(1-\xi)$, so then

$$\begin{aligned}\mu - \mu_L(\tau) &= \frac{\sigma}{1-\xi}\left[1+\frac{1}{\tau}\left((1-\tau)^{1-\xi}-1\right)\right] \\ &= \frac{\sigma}{1-\xi}\frac{1}{\tau}\left(\tau+(1-\tau)^{1-\xi}-1\right) \\ &= \frac{\sigma}{1-\xi}\frac{1-\tau}{\tau}\left((1-\tau)^{-\xi}-1\right).\end{aligned}$$

Appendix D

Proofs related to CNQR Asymptotics

This section contains proofs of consistency and asymptotic normality of the CNQR family of estimators. The estimator is defined in Equation (5.2.1), and the asymptotic results are outlined in Section 5.2.2.

D.1 Preliminaries

Standard estimating equations theory cannot be used to prove the asymptotic results. Such method requires the existence of the gradient at the objective function's minimum, but this is not always the case with CNQR estimation. Figure D.1 displays an example of an objective function whose minimum is at a cusp. Instead, the smoothness of the objective function's limit is used, as described in Newey and McFadden (1994, Chapter 7).

The cusp in the objective function is caused by the cusp in the asymmetric absolute deviation function, ρ_τ. Each term in the CNQR objective function (say the i'th observation and the k'th quantile) contributes a cusp at the parameter(s) whose τ_k-quantile surface passes through the i'th observation. This is because the argument of ρ_τ is zero – the location of which ρ_τ is non-differentiable. For there not to be a cusp at the minimum, none of the K quantile surfaces must intersect an observation. It is not obvious whether this is possible, but we at least know that sometimes there is a cusp at the minimum.

However, this cusp may not cause numerical issues when minimizing. The objective function seems to quickly become approximately smooth when the sample size is moderate, and when the number of regression quantiles K is more than a few. In fact, it seems that integrating over quantiles leads to a smooth objective function. A possible reason for this is, for any parameter value, there could exist a quantile surface that passes through a particular observation (unless that observation lies below the τ_c-quantile surface), in

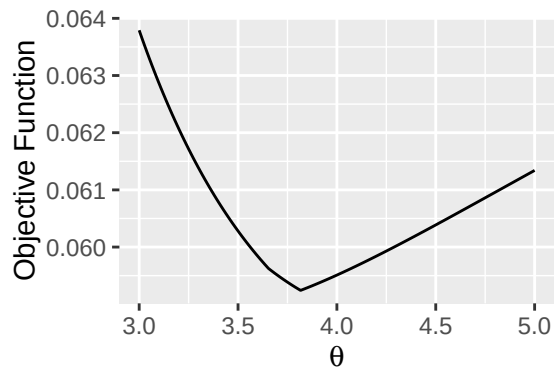

Figure D.1: An example of an objective function of the CNQR estimator in Equation (5.2.1) with two quantile levels $\tau_1 = 0.6$ and $\tau_2 = 0.8$, and transformation functions $g_1 = g_2$ equal to the identity function. Ten bivariate observations are generated from a Gumbel copula with parameter $\theta_0 = 5$, and given standard Exponential marginals. The objective function is created with these data having known marginals but unknown copula parameter. A region of the parameter space containing the minimum is displayed.

which case there is a cusp for that quantile surface and that observation. For most "nice" models, there is at most one such surface for one value of the parameter, and certainly not an uncountable number of such surfaces. This quantile surface has zero measure in the integral, so does not influence the resulting integral.

Before proving the asymptotic results, some useful identities are introduced in the following lemma, which extends a result of Knight (1998).

Lemma D.1.1. *For $x, y \in \mathbb{R}$ and asymmetric absolute deviation function ρ_τ defined in Equation (3.2.6) for $\tau \in (0,1)$,*

$$\rho_\tau(x-y) - \rho_\tau(x) = y\left(\mathbb{I}_{(-\infty,0)}(x) - \tau\right) + \mathcal{I}(x,y), \tag{D.1.1}$$

where $\mathcal{I}(x,y)$ can be written in the following ways:

$$\mathcal{I}(x,y) = \int_0^y \left[\mathbb{I}_{(-\infty,s)}(x) - \mathbb{I}_{(-\infty,0)}(x)\right] \mathrm{d}s, \tag{D.1.2}$$

$$\mathcal{I}(x,y) = \begin{cases} \int_0^y \mathbb{I}_{[0,s]}(x)\,\mathrm{d}s, & y \geq 0; \\ \int_y^0 \mathbb{I}_{[s,0]}(x)\,\mathrm{d}s, & y < 0, \end{cases} \tag{D.1.3}$$

and

$$\mathcal{I}(x,y) = \int_0^{|y|} \mathbb{I}_{[0,s]}\left(sign(y)\,x\right) \mathrm{d}s. \tag{D.1.4}$$

Note that the lower limit of integration need not be less than the upper limit.

Proof of Lemma D.1.1. The result with Equation (D.1.2) is shown first. Rewrite the

left-hand-side of Equation (D.1.1):

$$\begin{aligned}\rho_\tau(x-y)-\rho_\tau(x) &= \left(\tau-\mathbb{I}_{(-\infty,0)}(x-y)\right)(x-y)-\left(\tau-\mathbb{I}_{(-\infty,0)}(x)\right)x\\ &= -\tau y+(y-x)\,\mathbb{I}_{(0,\infty)}(y-x)+x\mathbb{I}_{(-\infty,0)}(x).\end{aligned}$$

Rewrite the right-hand side of Equation (D.1.1) using the form of $\mathcal{I}(x,y)$ in Equation (D.1.2):

$$\begin{aligned}&y\left(\mathbb{I}_{(-\infty,0)}(x)-\tau\right)+\int_0^y\left[\mathbb{I}_{(-\infty,s)}(x)-\mathbb{I}_{(-\infty,0)}(x)\right]\mathrm{d}s\\ =&-\tau y+\mathbb{I}_{(-\infty,0)}(x)\,y+\int_0^y\mathbb{I}_{(x,\infty)}(s)\,\mathrm{d}s-y\mathbb{I}_{(-\infty,0)}(x)\\ =&-\tau y+\int_0^y\mathbb{I}_{(x,\infty)}(s)\,\mathrm{d}s.\end{aligned}$$

The above integral can be further evaluated:

$$\begin{aligned}\int_0^y\mathbb{I}_{(x,\infty)}(s)\,\mathrm{d}s &= \int_0^x\mathbb{I}_{(x,\infty)}(s)\,\mathrm{d}s+\int_x^y\mathbb{I}_{(x,\infty)}(s)\,\mathrm{d}s\\ &= \int_0^x\mathbb{I}_{(x,\infty)}(s)\,\mathrm{d}s+\int_0^{y-x}\mathbb{I}_{(0,\infty)}(s)\,\mathrm{d}s\\ &= x\mathbb{I}_{(-\infty,0)}(x)+(y-x)\,\mathbb{I}_{(0,\infty)}(y-x),\end{aligned}$$

so that the identity holds with $\mathcal{I}(x,y)$ in Equation (D.1.2).

$\mathbb{I}_{(-\infty,s)}(x)$	$\mathbb{I}_{(-\infty,0)}(x)$	$\mathbb{I}_{(-\infty,s)}(x)-\mathbb{I}_{(-\infty,0)}(x)$	**Happens when...**	
			For $s>0$	**For** $s<0$
1	1	0	$x<0$	$x<s$
0	1	-1	impossible	$s\le x\le 0$
1	0	1	$0\le x\le s$	impossible
0	0	0	$x>s$	$x>0$

Table D.1: Indicator functions in the formulation of $Z_n(\boldsymbol{\theta})$

To show Equation (D.1.3), rewrite the integrand of Equation (D.1.2) by considering the cases outlined in Table D.1. For $s\in\mathbb{R}$,

$$\mathbb{I}_{(-\infty,s)}(x)-\mathbb{I}_{(-\infty,0)}(x)=\begin{cases}\mathbb{I}_{[0,s]}(x), & s\ge 0;\\ -\mathbb{I}_{[s,0]}(x), & s<0,\end{cases}$$

so that Equation (D.1.3) follows.

To show Equation (D.1.4), reduce the right-hand side of Equation (D.1.4) considering

the two cases $y < 0$ and $y \geq 0$. The $y \geq 0$ case is trivial, since it directly matches the $y \geq 0$ case in Equation (D.1.3). Consider, then, $y < 0$. Changing the variable of integration in the right-hand side of Equation (D.1.4) to $t = -s$, we have

$$\begin{aligned}\int_0^{|y|} \mathbb{I}_{[0,s]} \left(\operatorname{sign}(y)\, x\right) \mathrm{d}s &= -\int_0^{y} \mathbb{I}_{[0,-t]} (-x)\, \mathrm{d}t \\ &= \int_y^{0} \mathbb{I}_{[t,0]} (x)\, \mathrm{d}t,\end{aligned}$$

which matches the $y < 0$ case in Equation (D.1.3). Therefore, Equation (D.1.4) holds. □

The identity in Lemma D.1.1 using Equation (D.1.2) follows from Knight (1998), and is often used in the literature when proving asymptotic properties of quantile regression estimators (cf. Chernozhukov, 2005; Jiang et al., 2012; Noh et al., 2015; Zhao and Xiao, 2014; Huang et al., 2015; Xie, 2015). The other two identities are also useful, yet seem to be absent from the literature.

D.2 Proofs

To prove the asymptotic results of the CNQR family of estimators, it is useful to work with an alternative objective function that has a zero minimum. Recall by Assumption 5.2.2 that $\hat{Q}_{Y|\boldsymbol{X}}\left(\cdot|x;\boldsymbol{\theta}_0\right) = Q_{Y|\boldsymbol{X}}\left(\cdot|x\right)$; let

$$\varepsilon_{ki} = g_k\left(Y_i\right) - g_k\left(Q_{Y|\boldsymbol{X}}\left(\tau_k|\boldsymbol{X}_i\right)\right) = g_k\left(Y_i\right) - q_{ki}\left(\boldsymbol{\theta}_0\right), \tag{D.2.1}$$

and

$$\Delta q_{ki}\left(\boldsymbol{\theta}\right) = q_{ki}\left(\boldsymbol{\theta}\right) - q_{ki}\left(\boldsymbol{\theta}_0\right) \tag{D.2.2}$$

for each $i = 1, \ldots, n$ and $k = 1, \ldots, K$, where $\boldsymbol{\theta} \in \Theta$, and q_{ki} is defined in Equation (5.2.3). The i subscript is dropped to refer to the generic random vector $\left(\boldsymbol{X}^\top, Y\right)$. The conditional distribution function of ε_k given the predictors is then

$$F_{\varepsilon_k|\boldsymbol{X}}\left(t \mid \boldsymbol{x}\right) = F_{g_k(Y)|\boldsymbol{X}}\left(q_k\left(\boldsymbol{\theta}_0\right) + t \mid \boldsymbol{x}\right) \tag{D.2.3}$$

for $t \in \mathbb{R}$ and all $\boldsymbol{x} \in \mathscr{X}$, with density

$$f_{\varepsilon_k|\boldsymbol{X}}\left(t \mid \boldsymbol{x}\right) = f_{g_k(Y)|\boldsymbol{X}}\left(q_k\left(\boldsymbol{\theta}_0\right) + t \mid \boldsymbol{x}\right). \tag{D.2.4}$$

The new objective function that we seek is

$$Z_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{k=1}^{K}\sum_{i=1}^{n}\left[\rho_{\tau_k}\left(\varepsilon_{ki} - \Delta q_{ki}(\boldsymbol{\theta})\right) - \rho_{\tau_k}(\varepsilon_{ki})\right] \tag{D.2.5}$$

for $\boldsymbol{\theta} \in \Theta$. Notice that the corresponding CNQR estimator[1] is equivalent to $\hat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}} Z_n(\boldsymbol{\theta})$. This objective function is convenient because $Z_n(\boldsymbol{\theta}_0) = 0$, since $\Delta q_{ki}(\boldsymbol{\theta}_0) = 0$. Using the identity in Lemma D.1.1, this objective function can be written as

$$Z_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{k=1}^{K}\sum_{i=1}^{n}\left[\Delta q_{ki}(\boldsymbol{\theta})\left(\mathbb{I}_{(-\infty,0)}(\varepsilon_{ki}) - \tau_k\right) + \mathcal{I}\left(\varepsilon_{ki}, \Delta q_{ki}(\boldsymbol{\theta})\right)\right]. \tag{D.2.6}$$

To show consistency of the CNQR family of estimators, the following regularity conditions are needed.

Condition 1. Consider a CNQR estimator in Equation (5.2.1).

(a) The parameter space Θ is a compact set.

(b) There exist real numbers $\epsilon_1 > 0$ and $\epsilon_2 > 0$ such that, for at least one $k = 1, \ldots, K$, the set

$$\mathscr{X}_k = \left\{\boldsymbol{x} \in \mathscr{X} : f_{g_k(Y)|\boldsymbol{X}}\left(q_{ki}(\boldsymbol{\theta}_0) + t|\boldsymbol{x}\right) \geq \epsilon_1 \,\forall\, t \in (-\epsilon_2, \epsilon_2)\right\} \tag{D.2.7}$$

has $\mathcal{P}(\boldsymbol{X} \in \mathscr{X}_k) = p_{\mathscr{X}_k} > 0$.

(c) For all $\boldsymbol{\theta} \in \Theta \backslash \{\boldsymbol{\theta_0}\}$ and at least one $k = 1, \ldots, K$,

$$\mathcal{P}\left(\left|\Delta q_k(\boldsymbol{\theta})\right| > 0 \;\middle|\; \boldsymbol{X} \in \mathscr{X}_k\right) > 0.$$

Condition 1(a) can be relaxed in practice, as is discussed in Newey and McFadden (1994, Section 2.6). Condition 1(b) allows us to restrict focus of the predictor space $\mathscr{X}$ to regions $\mathscr{X}_k$ having non-zero measure, over which the response has a non-zero (conditional) probability of taking values near the τ_k-quantile. Condition 1(c) ensures that the parameter $\boldsymbol{\theta}_0$ is identifiable – that is, the quantile surface $\hat{Q}_{Y|\boldsymbol{X}}\left(\tau_k \mid \cdot; \boldsymbol{\theta}\right)$ is different from $Q_{Y|\boldsymbol{X}}\left(\tau_k \mid \cdot\right)$ over some subset of $\mathscr{X}_k$ having non-zero measure, unless $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

To prove consistency, we make use of the fact that a CNQR objective function has a unique minimum at the true parameter value in expectation, as stated in the following

[1] In practice, estimation cannot be carried out with this alternative objective function Z_n, which contains unknown quantities. But this objective function is useful for theoretical considerations.

lemma.

Lemma D.2.1. *Suppose Assumptions 5.2.1 and 5.2.2 hold, and consider non-decreasing functions $g_k : \mathbb{R} \to \mathbb{R}$, $k = 1, \ldots, K$. Under the regularity conditions listed in Condition 1, $\mathbb{E}\left(Z_n(\boldsymbol{\theta})\right)$ is uniquely minimized at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, where $Z_n(\boldsymbol{\theta})$ is defined in Equation (D.2.5). That is, $\mathbb{E}\left(Z_n(\boldsymbol{\theta})\right) > \mathbb{E}\left(Z_n(\boldsymbol{\theta}_0)\right)$ for all $\boldsymbol{\theta} \in \Theta \backslash \{\boldsymbol{\theta}_0\}$.*

Proof. The main idea of this proof is from Chen et al. (2009).

First, we will decompose $\mathbb{E}\left(Z_n(\boldsymbol{\theta})\right)$ for all $\boldsymbol{\theta} \in \Theta$. Using the formulation of $Z_n(\boldsymbol{\theta})$ in Equation (D.2.6) together with the law of total expectation,

$$\mathbb{E}\left(Z_n(\boldsymbol{\theta})\right) = \sum_{k=1}^{K} \mathbb{E}\left(\mathbb{E}\left(\left[\Delta q_k(\boldsymbol{\theta})\left(\mathbb{I}_{(-\infty,0)}(\varepsilon_k) - \tau_k\right) + \mathcal{I}\left(\varepsilon_k, \Delta q_k(\boldsymbol{\theta})\right)\right] \,\middle|\, \boldsymbol{X}\right)\right),$$

due to the iid assumption of the data, where ε_k and Δq_k are defined respectively in Equations (D.2.1) and (D.2.2). Since

$$\mathbb{E}\left(\mathbb{I}_{(-\infty,0)}(\varepsilon_k) - \tau_k \mid \boldsymbol{X}\right) = \mathcal{P}\left(Y < Q_{Y|\boldsymbol{X}}\left(\tau_k \mid \boldsymbol{X}\right) \mid \boldsymbol{X}\right) - \tau_k = 0,$$

we need only focus on the integral term $\mathcal{I}\left(\varepsilon_k, \Delta q_k(\boldsymbol{\theta})\right)$. Using the two-case form of the integral term in Equation (D.1.3), we have $\mathbb{E}\left(Z_n(\boldsymbol{\theta})\right) = \sum_k \mathbb{E}\left(\mathcal{T}_{1k} + \mathcal{T}_{2k}\right)$, where "terms 1 and 2" are

$$\mathcal{T}_{1k} = \mathbb{I}_{(0,\infty)}\left(\Delta q_k(\boldsymbol{\theta})\right) \mathbb{E}\left[\int_0^{\Delta q_k(\boldsymbol{\theta})} \mathbb{I}_{(0,s)}(\varepsilon_k)\,\mathrm{d}s \mid \boldsymbol{X}\right]$$

and

$$\mathcal{T}_{2k} = \mathbb{I}_{(-\infty,0)}\left(\Delta q_k(\boldsymbol{\theta})\right) \mathbb{E}\left[\int_{\Delta q_k(\boldsymbol{\theta})}^{0} \mathbb{I}_{(s,0)}(\varepsilon_k)\,\mathrm{d}s \mid \boldsymbol{X}\right].$$

It is now clear that $\mathbb{E}\left(Z_n(\boldsymbol{\theta})\right) \geq 0$ since $\mathcal{T}_{1k} \geq 0$ and $\mathcal{T}_{2k} \geq 0$ almost surely. Also, notice that $Z_n(\boldsymbol{\theta}_0) = 0$ since $\Delta q_k(\boldsymbol{\theta}_0) = 0$. Thus, to show that $\mathbb{E}\left(Z_n(\boldsymbol{\theta})\right)$ is uniquely minimized at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ is to show that $\mathbb{E}\left(Z_n(\boldsymbol{\theta})\right)$ is positive for all $\boldsymbol{\theta} \in \Theta \backslash \{\boldsymbol{\theta}_0\}$.

Now, take $\boldsymbol{\theta} \in \Theta \backslash \{\boldsymbol{\theta}_0\}$ and $k \in \{1, \ldots, K\}$. Restricting the predictor space to $\mathscr{X}_k$ defined in Equation (D.2.7), we have

$$\begin{aligned}
\mathcal{T}_{1k} &\geq \mathbb{I}_{\mathscr{X}_k}(\boldsymbol{X})\,\mathcal{T}_{1k} \\
&= \mathbb{I}_{\mathscr{X}_k}(\boldsymbol{X})\,\mathbb{I}_{(0,\infty)}\left(\Delta q_k(\boldsymbol{\theta})\right) \int_{-\infty}^{\infty} \left[\int_0^{\Delta q_k(\boldsymbol{\theta})} \mathbb{I}_{(0,s)}(\varepsilon)\,\mathrm{d}s\right] f_{g_k(Y)|\boldsymbol{X}}\left(q_k(\boldsymbol{\theta}_0) + \varepsilon \mid \boldsymbol{X}\right) \mathrm{d}\varepsilon \\
&= \mathbb{I}_{\mathscr{X}_k}(\boldsymbol{X})\,\mathbb{I}_{(0,\infty)}\left(\Delta q_k(\boldsymbol{\theta})\right) \int_0^{\Delta q_k(\boldsymbol{\theta})} \int_0^{s} f_{g_k(Y)|\boldsymbol{X}}\left(q_k(\boldsymbol{\theta}_0) + \varepsilon \mid \boldsymbol{X}\right) \mathrm{d}\varepsilon\,\mathrm{d}s,
\end{aligned}$$

using the density of ε_k conditional on the predictors in Equation (D.2.4). By Condition 1(b), confining $\boldsymbol{X} \in \mathscr{X}_k$ ensures that the random function $\varepsilon \mapsto f_{g_k(Y)|\boldsymbol{X}}\left(q_k\left(\boldsymbol{\theta}_0\right) + \varepsilon \mid \boldsymbol{X}\right)$ is almost surely larger than ϵ_1 over $(-\epsilon_2, \epsilon_2)$. So we integrate over the smaller function $\varepsilon \mapsto \epsilon_1 \mathbb{I}_{(-\epsilon_2,\epsilon_2)}(\varepsilon)$ instead, to obtain

$$\mathcal{T}_{1k} \geq \mathbb{I}_{\mathscr{X}_k}(\boldsymbol{X})\, \mathbb{I}_{(0,\infty)}\left(\Delta q_k\left(\boldsymbol{\theta}\right)\right) \frac{\epsilon_1}{2} \left(\begin{cases} \left[\Delta q_k\left(\boldsymbol{\theta}\right)\right]^2, & \Delta q_k\left(\boldsymbol{\theta}\right) < \epsilon_2; \\ \epsilon_2^2, & \Delta q_k\left(\boldsymbol{\theta}\right) \geq \epsilon_2 \end{cases} \right)$$

almost surely.

Similarly, we obtain

$$\mathcal{T}_{2k} \geq \mathbb{I}_{\mathscr{X}_k}(\boldsymbol{X})\, \mathbb{I}_{(-\infty,0)}\left(\Delta q_k\left(\boldsymbol{\theta}\right)\right) \frac{\epsilon_1}{2} \left(\begin{cases} \left[\Delta q_k\left(\boldsymbol{\theta}\right)\right]^2, & \Delta q_k\left(\boldsymbol{\theta}\right) > -\epsilon_2; \\ \epsilon_2^2, & \Delta q_k\left(\boldsymbol{\theta}\right) \leq -\epsilon_2 \end{cases} \right)$$

almost surely.

Putting $\mathcal{T}_{1k}$ and $\mathcal{T}_{2k}$ together, we obtain

$$\begin{aligned}
&\mathbb{E}\left(\mathcal{T}_{1k} + \mathcal{T}_{2k}\right) \\
\geq &\frac{\epsilon_1}{2}\mathbb{E}\left(\mathbb{I}_{\mathscr{X}_k}(\boldsymbol{X})\left[\left[\Delta q_k\left(\boldsymbol{\theta}\right)\right]^2 \mathbb{I}_{(-\epsilon_2,\epsilon_2)}\left(\Delta q_k\left(\boldsymbol{\theta}\right)\right) + \epsilon_2^2 \mathbb{I}_{[\epsilon_2,\infty)}\left(\left|\Delta q_k\left(\boldsymbol{\theta}\right)\right|\right)\right]\right) \\
= &\frac{\epsilon_1}{2}\mathbb{E}\left(\left[\Delta q_k\left(\boldsymbol{\theta}\right)\right]^2 \mathbb{I}_{(-\epsilon_2,\epsilon_2)}\left(\Delta q_k\left(\boldsymbol{\theta}\right)\right) + \epsilon_2^2 \mathbb{I}_{[\epsilon_2,\infty)}\left(\left|\Delta q_k\left(\boldsymbol{\theta}\right)\right|\right) \,\middle|\, \boldsymbol{X} \in \mathscr{X}_k\right) \mathcal{P}\left(\boldsymbol{X} \in \mathscr{X}_k\right) \\
= &\frac{\epsilon_1}{2} p_{\mathscr{X}_k} \mathbb{E}\left(\left[\Delta q_k\left(\boldsymbol{\theta}\right)\right]^2 \,\middle|\, \boldsymbol{X} \in \mathscr{X}_k, \left|\Delta q_k\left(\boldsymbol{\theta}\right)\right| < \epsilon_2\right) \mathcal{P}\left(\left|\Delta q_k\left(\boldsymbol{\theta}\right)\right| < \epsilon_2 \,\middle|\, \boldsymbol{X} \in \mathscr{X}_k\right) \\
&+ \frac{\epsilon_1 \epsilon_2^2}{2} p_{\mathscr{X}_k} \mathcal{P}\left(\left|\Delta q_k\left(\boldsymbol{\theta}\right)\right| \geq \epsilon_2 \,\middle|\, \boldsymbol{X} \in \mathscr{X}_k\right).
\end{aligned}$$

In the case that

$$\mathcal{P}\left(\left|\Delta q_k\left(\boldsymbol{\theta}\right)\right| \geq \epsilon_2 \,\middle|\, \boldsymbol{X} \in \mathscr{X}_k\right) > 0,$$

we have $\mathbb{E}\left(\mathcal{T}_{1k} + \mathcal{T}_{2k}\right) > 0$ and so $\mathbb{E}\left(Z_n\left(\boldsymbol{\theta}\right)\right) > 0$. In the case that

$$\mathcal{P}\left(\left|\Delta q_k\left(\boldsymbol{\theta}\right)\right| \geq \epsilon_2 \,\middle|\, \boldsymbol{X} \in \mathscr{X}_k\right) = 0,$$

so that most of the mass of $\Delta q_k(\boldsymbol{\theta})$ is near 0, we have

$$\mathbb{E}\left(\left[\Delta q_k(\boldsymbol{\theta})\right]^2 \,\middle|\, \boldsymbol{X} \in \mathscr{X}_k, \left|\Delta q_k(\boldsymbol{\theta})\right| < \epsilon_2\right) > 0$$

under Condition 1(c), and so $\mathbb{E}\left(Z_n(\boldsymbol{\theta})\right) > 0$. □

We are now equipped to prove consistency.

Proof of Theorem 5.2.1 (Consistency). Following Chen et al. (2009), Newey and McFadden (1994, Theorem 2.1) is used to prove consistency, which follows if the following four conditions hold.

Firstly, $\mathbb{E}\left(Z_n(\boldsymbol{\theta})\right)$ must be uniquely minimized at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. This holds by Lemma D.2.1.

Secondly, Θ must be compact, which follows by Condition 1(a).

Thirdly, $\mathbb{E}\left(Z_n(\boldsymbol{\theta})\right)$ must be continuous, which follows by the continuity of $\boldsymbol{\theta} \mapsto Q_{Y|\boldsymbol{X}}\left(\tau|\boldsymbol{x};\boldsymbol{\theta}\right)$ for any $\tau \in (0,1)$, and $\boldsymbol{x} \in \mathscr{X}$.

Lastly, we require $Z_n(\boldsymbol{\theta}) \to_p \mathbb{E}\left(Z_n(\boldsymbol{\theta})\right)$ uniformly over $\boldsymbol{\theta} \in \Theta$; that is,

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|Z_n(\boldsymbol{\theta}) - \mathbb{E}\left(Z_n(\boldsymbol{\theta})\right)\right| \to_p 0.$$

This follows by a uniform version of the weak law of large numbers (cf. Amemiya, 1985, Theorem 4.2.1), since the loss functions $\rho_{\tau_1}, \ldots, \rho_{\tau_K}$, defined in Equation (3.2.6), are "well behaved". In particular, since Θ is compact, the data are iid, Z_n is continuous (conditional on the observations), and the loss functions are finite over Θ, then uniform convergence in probability holds.

□

Next, we turn our attention to proving the asymptotic distribution of the CNQR estimator. The idea behind the proof comes from Koenker (2005); Zou and Yuan (2008). The proof of the asymptotic distribution of the CNQR estimator relies on the following regularity conditions.

Condition 2. Consider a CNQR estimator in Equation (5.2.1).

(a) For each $i = 1, 2, \ldots,$

  (i) the gradient $\dot{\boldsymbol{q}}_{ki}$, defined in Equation (5.2.4), exists and is continuous in a neighbourhood of $\boldsymbol{\theta}_0$ and is non-zero at $\boldsymbol{\theta}_0$; and

  (ii) the Hessian $\ddot{\boldsymbol{q}}_{ki}$, defined in Equation (5.2.5), exists and is continuous in a neighbourhood of $\boldsymbol{\theta}_0$.

(b) For each $k = 1, \ldots, K$ and $k' = 1, \ldots, K$, each entry of $\boldsymbol{D}_{0kk'}$ defined in Equation (5.2.7) is finite.

(c) For each $k = 1, \ldots, K$, each entry of $\ddot{\boldsymbol{q}}_k(\boldsymbol{\theta}_0)$ defined in Equation (5.2.5) is almost surely finite.

(d) For each $k = 1, \ldots, K$, each entry of $\boldsymbol{D}_{1k}$ defined in Equation (5.2.9) is finite.

Condition 2(a) allows for the use of Taylor's theorem with a second degree Taylor polynomial, and implies that $\boldsymbol{\theta}_0$ must be in the interior of Θ. Conditions 2(b)–(d) prevent divergence of the objective function as the sample size increases.

Proof of Theorem 5.2.2 (Asymptotic Normality). We begin by identifying an objective function whose arg min is $\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right)$. Since $\hat{\boldsymbol{\theta}}_n$ minimizes $Z_n(\boldsymbol{\theta})$, defined in Equation (D.2.5), it follows that $\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right)$ minimizes $Z_n^* : \boldsymbol{\delta} \mapsto nZ_n\left(\boldsymbol{\theta}_0 + \boldsymbol{\delta}/\sqrt{n}\right)$, which evaluates to

$$Z_n^*(\boldsymbol{\delta}) = \sum_{k=1}^{K}\sum_{i=1}^{n}\left[\Delta q_{ki}\left(\boldsymbol{\theta}_0 + \frac{\boldsymbol{\delta}}{\sqrt{n}}\right)\left(\mathbb{I}_{(-\infty,0)}(\varepsilon_{ki}) - \tau_k\right) + \mathcal{I}\left(\varepsilon_{ki}, \Delta q_{ki}\left(\boldsymbol{\theta}_0 + \frac{\boldsymbol{\delta}}{\sqrt{n}}\right)\right)\right] \tag{D.2.8}$$

for $\boldsymbol{\delta}$ in some neighbourhood of $\mathbf{0}$, where the function $\mathcal{I}$ takes one of the forms described in Lemma D.1.1. The strategy is to find the limiting distribution of $\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right)$ through the limiting objective function, $Z_\infty^* : \boldsymbol{\delta} \mapsto \lim_{n\to\infty} Z_n^*(\boldsymbol{\delta})$.

Consider the second order Taylor polynomial of $\boldsymbol{\delta} \mapsto \Delta q_{ki}\left(\boldsymbol{\theta}_0 + \boldsymbol{\delta}/\sqrt{n}\right)$ about $\mathbf{0}$ for each i and k. By Taylor's theorem,

$$\begin{aligned}\Delta q_{ki}\left(\boldsymbol{\theta}_0 + \boldsymbol{\delta}/\sqrt{n}\right) &= \frac{1}{\sqrt{n}}\left[\dot{\boldsymbol{q}}_{ki}(\boldsymbol{\theta}_0)\right]^\top \boldsymbol{\delta} + \frac{1}{2n}\boldsymbol{\delta}^\top\left[\ddot{\boldsymbol{q}}_{ki}(\boldsymbol{\theta}_0)\right]\boldsymbol{\delta} + o_p\left(n^{-1}\right) \\ &=: \Delta q_{nki}^*(\boldsymbol{\delta})\end{aligned}$$

as $n \to \infty$ for $\boldsymbol{\delta}$ in some neighbourhood of $\mathbf{0}$, where $\dot{\boldsymbol{q}}_{ki}$ and $\ddot{\boldsymbol{q}}_{ki}$ are respectively defined in Equations (5.2.4) and (5.2.5). Note that $\Delta q_{nki}^*(\boldsymbol{\delta}) = O_p\left(n^{-1/2}\right)$ as $n \to \infty$. This expansion exists due to Condition 2(a). Substituting this expansion into $Z_n^*(\boldsymbol{\delta})$ in Equation (D.2.8), we obtain

$$Z_n^*(\boldsymbol{\delta}) = \sqrt{n}\boldsymbol{\delta}^\top \bar{\boldsymbol{Z}}_{1n} + \frac{1}{2}\bar{Z}_{2n}(\boldsymbol{\delta}) + \sum_{k=1}^{K}\bar{Z}_{3nk}(\boldsymbol{\delta}) + o_p(1)$$

as $n \to \infty$, where

$$\bar{\boldsymbol{Z}}_{1n} = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} \left(\mathbb{I}_{(-\infty,0)} \left(\varepsilon_{ki} \right) - \tau_k \right) \dot{\boldsymbol{q}}_{ki} \left(\boldsymbol{\theta}_0 \right),$$

$$\begin{aligned} \bar{Z}_{2n} \left(\boldsymbol{\delta} \right) &= \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} \left(\mathbb{I}_{(-\infty,0)} \left(\varepsilon_{ki} \right) - \tau_k \right) \boldsymbol{\delta}^\top \ddot{\boldsymbol{q}}_{ki} \left(\boldsymbol{\theta}_0 \right) \boldsymbol{\delta} \\ &=: \frac{1}{n} \sum_{i=1}^{n} Z_{2ni} \left(\boldsymbol{\delta} \right), \end{aligned}$$

and

$$\bar{Z}_{3nk} \left(\boldsymbol{\delta} \right) = \frac{1}{n} \sum_{i=1}^{n} Z_{3nki} \left(\boldsymbol{\delta} \right),$$

where

$$Z_{3nki} \left(\boldsymbol{\delta} \right) = n \mathcal{I} \left(\varepsilon_{ki}, \Delta q^*_{nki} \left(\boldsymbol{\delta} \right) \right).$$

The asymptotic behaviour of each $\sqrt{n} \bar{\boldsymbol{Z}}_{1n}$, $\bar{Z}_{2n} \left(\boldsymbol{\delta} \right)$, and $\bar{Z}_{3nk} \left(\boldsymbol{\delta} \right)$ are now sought.

Firstly, it is shown that $\sqrt{n} \bar{\boldsymbol{Z}}_{1n}$ converges to a normal distribution as $n \to \infty$. For its expectation, since

$$\begin{aligned} \mathbb{E} \left[\mathbb{I}_{(-\infty,0)} \left(\varepsilon_{ki} \right) \right] &= \mathbb{E} \left[\mathbb{E} \left(\mathbb{I}_{(-\infty,0)} \left(\varepsilon_{ki} \right) \mid \boldsymbol{X} \right) \right] \\ &= \mathbb{E} \left[\mathcal{P} \left(Y < Q_{Y|\boldsymbol{X}} \left(\tau_k \mid \boldsymbol{X} \right) \mid \boldsymbol{X} \right) \right] \\ &= \tau_k, \end{aligned}$$

we have $\mathbb{E} \left(\bar{\boldsymbol{Z}}_{1n} \right) = \boldsymbol{0}$. For the covariance matrix, notice that

$$\mathbb{I}_{(-\infty,0)} \left(\varepsilon_{ki} \right) \sim \text{Bernoulli} \left(\tau_k \right)$$

and

$$\mathbb{I}_{(-\infty,0)} \left(\varepsilon_{ki} \right) \mathbb{I}_{(-\infty,0)} \left(\varepsilon_{k'i} \right) \sim \text{Bernoulli} \left(\min \left(\tau_k, \tau_{k'} \right) \right)$$

for each k,k'. Now,

$$\begin{aligned}
\operatorname{Cov}\left(\sqrt{n}\bar{\boldsymbol{Z}}_{1n}\right) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{k,k'}\mathbb{E}\left(\operatorname{Cov}\left(\mathbb{I}_{(-\infty,0)}\left(\varepsilon_{ki}\right)\dot{\boldsymbol{q}}_{ki}\left(\boldsymbol{\theta}_0\right),\mathbb{I}_{(-\infty,0)}\left(\varepsilon_{k'i}\right)\dot{\boldsymbol{q}}_{k'i}\left(\boldsymbol{\theta}_0\right)\mid\boldsymbol{X}_i\right)\right)\\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{k,k'}\mathbb{E}\left(\operatorname{Cov}\left(\mathbb{I}_{(-\infty,0)}\left(\varepsilon_{ki}\right),\mathbb{I}_{(-\infty,0)}\left(\varepsilon_{k'i}\right)\mid\boldsymbol{X}_i\right)\left[\dot{\boldsymbol{q}}_{ki}\left(\boldsymbol{\theta}_0\right)\right]\left[\dot{\boldsymbol{q}}_{k'i}\left(\boldsymbol{\theta}_0\right)\right]^\top\right)\\
&= \sum_{k,k'}\min\left(\tau_k,\tau_{k'}\right)\left(1-\max\left(\tau_k,\tau_{k'}\right)\right)\boldsymbol{D}_{0kk'}\\
&=: \boldsymbol{\Sigma}_1
\end{aligned}$$

where $\boldsymbol{D}_{0kk'}$ is defined in Equation (5.2.7), because $\tau_k\tau_{k'} = \min\left(\tau_k,\tau_{k'}\right)\max\left(\tau_k,\tau_{k'}\right)$. By Condition 2(a)(i), $\boldsymbol{\Sigma}_1$ is nonzero because each $\dot{\boldsymbol{q}}_{ki}\left(\boldsymbol{\theta}_0\right)$ is nonzero, and by Condition 2(b), $\boldsymbol{\Sigma}_1$ has finite entries. Thus $\sqrt{n}\bar{\boldsymbol{Z}}_{1n} \to_d \boldsymbol{Z}_{1\infty}$ by the Central Limit theorem, where $\boldsymbol{Z}_{1\infty} \sim \mathcal{N}\left(\boldsymbol{0},\boldsymbol{\Sigma}_1\right)$.

Secondly, it is shown that $\bar{Z}_{2n}\left(\boldsymbol{\delta}\right) \to_p 0$ as $n \to \infty$. We have $\mathbb{E}\left(\bar{Z}_{2n}\left(\boldsymbol{\delta}\right)\right) = 0$ for the same reason that $\mathbb{E}\left(\bar{\boldsymbol{Z}}_{1n}\right) = \boldsymbol{0}$. By the law of total covariance, and since the data are iid, we have

$$\begin{aligned}
&\operatorname{Var}\left(\bar{Z}_{2n}\left(\boldsymbol{\delta}\right)\right)\\
=&\frac{1}{n^2}\sum_{i=1}^{n}\sum_{k,k'}\mathbb{E}\left(\operatorname{Cov}\left(\mathbb{I}_{(-\infty,0)}\left(\varepsilon_{ki}\right),\mathbb{I}_{(-\infty,0)}\left(\varepsilon_{ki}\right)\mid\boldsymbol{X}_i\right)\boldsymbol{\delta}^\top\ddot{\boldsymbol{q}}_{ki}\left(\boldsymbol{\theta}_0\right)\boldsymbol{\delta}\boldsymbol{\delta}^\top\ddot{\boldsymbol{q}}_{k'i}\left(\boldsymbol{\theta}_0\right)\boldsymbol{\delta}\right)\\
=&\frac{1}{n}\sum_{k,k'}\min\left(\tau_k,\tau_{k'}\right)\left(1-\max\left(\tau_k,\tau_{k'}\right)\right)\mathbb{E}\left(\boldsymbol{\delta}^\top\ddot{\boldsymbol{q}}_k\left(\boldsymbol{\theta}_0\right)\boldsymbol{\delta}\boldsymbol{\delta}^\top\ddot{\boldsymbol{q}}_{k'}\left(\boldsymbol{\theta}_0\right)\boldsymbol{\delta}\right)\\
\to&0
\end{aligned}$$

as $n \to \infty$, by Condition 2(c). Thus $\bar{Z}_{2n}\left(\boldsymbol{\delta}\right) \to_p 0$ by the weak law of large numbers.

Lastly, it is shown that $\bar{Z}_{3nk}\left(\boldsymbol{\delta}\right)$ converges in probability to the limit of its expectation as $n \to \infty$ for each $k = 1,\ldots,K$. To find the expectation of $Z_{3nki}\left(\boldsymbol{\delta}\right)$, use the form of the integral $\mathcal{I}$ found in Equation (D.1.2), and consider a change of variables to $t = \sqrt{n}s$, to obtain

$$Z_{3nki}\left(\boldsymbol{\delta}\right) = \int_0^{\left[\dot{\boldsymbol{q}}_{ki}\left(\boldsymbol{\theta}_0\right)\right]^\top\boldsymbol{\delta}+o_p(1)} t\left[\mathbb{I}_{\left(-\infty,t/\sqrt{n}\right)}\left(\varepsilon_{ki}\right)-\mathbb{I}_{(-\infty,0)}\left(\varepsilon_{ki}\right)\right]/\left(\frac{t}{\sqrt{n}}\right)\mathrm{d}t.$$

By the law of total expectation, and since the data are iid,

$$
\begin{aligned}
&\mathbb{E}\left(Z_{3nki}\left(\boldsymbol{\delta}\right)\right)\\
=&\mathbb{E}\left(\int_0^{\boldsymbol{\delta}^\top \dot{\boldsymbol{q}}_k(\boldsymbol{\theta}_0)+o_p(1)} t\frac{\left[F_{g_k(Y)|\boldsymbol{X}}\left(q_k\left(\boldsymbol{\theta}_0\right)+t/\sqrt{n}\mid \boldsymbol{X}\right)-F_{g_k(Y)|\boldsymbol{X}}\left(q_k\left(\boldsymbol{\theta}_0\right)\mid \boldsymbol{X}\right)\right]}{t/\sqrt{n}}\,\mathrm{d}t\right)\\
=&\mathbb{E}\left(\int_0^{\boldsymbol{\delta}^\top \dot{\boldsymbol{q}}_k(\boldsymbol{\theta}_0)+o_p(1)} t\left[f_{g_k(Y)|\boldsymbol{X}}\left(q_k\left(\boldsymbol{\theta}_0\right)\mid \boldsymbol{X}\right)+o_p\left(1\right)\right]\mathrm{d}t\right)\\
=&\mathbb{E}\left(\frac{1}{2}\left[\boldsymbol{\delta}^\top\dot{\boldsymbol{q}}_k\left(\boldsymbol{\theta}_0\right)+o_p\left(1\right)\right]^2\left[f_{g_k(Y)|\boldsymbol{X}}\left(q_k\left(\boldsymbol{\theta}_0\right)\mid \boldsymbol{X}\right)+o_p\left(1\right)\right]\right)\\
=&\frac{1}{2}\boldsymbol{\delta}^\top\boldsymbol{D}_{1k}\boldsymbol{\delta}+R_{3nk}
\end{aligned}
$$

as $n\to\infty$, where the remainder $R_{3nk}=o\left(1\right)$, and $\boldsymbol{D}_{1k}$ is defined in Equation (5.2.9). To find the variance of $Z_{3nki}\left(\boldsymbol{\delta}\right)$, use the form of the integral $\mathcal{I}$ found in Equation (D.1.4), so that

$$
\begin{aligned}
Z_{3nki}\left(\boldsymbol{\delta}\right)&=n\int_0^{\left|\Delta q^*_{nki}(\boldsymbol{\delta})\right|}\mathbb{I}_{[0,s]}\left(\operatorname{sign}\left(\Delta q^*_{nki}\left(\boldsymbol{\delta}\right)\right)\varepsilon_{ki}\right)\mathrm{d}s\\
&=n\left(\left|\Delta q^*_{nki}\left(\boldsymbol{\delta}\right)\right|-\operatorname{sign}\left(\Delta q^*_{nki}\left(\boldsymbol{\delta}\right)\right)\varepsilon_{ki}\right)\mathbb{I}_{\left[0,\left|\Delta q^*_{nki}(\boldsymbol{\delta})\right|\right]}\left(\operatorname{sign}\left(\Delta q^*_{nki}\left(\boldsymbol{\delta}\right)\right)\varepsilon_{ki}\right)\\
&=n\left(\Delta q^*_{nki}\left(\boldsymbol{\delta}\right)-\varepsilon_{ki}\right)W_{nki},
\end{aligned}
$$

as $n\to\infty$, where

$$
\begin{aligned}
W_{nki}&=\operatorname{sign}\left(\Delta q^*_{nki}\left(\boldsymbol{\delta}\right)\right)\mathbb{I}_{\left[\min\left(0,\Delta q^*_{nki}(\boldsymbol{\delta})\right),\max\left(0,\Delta q^*_{nki}(\boldsymbol{\delta})\right)\right]}\left(\varepsilon_{ki}\right)\\
&=\begin{cases}1, & 0\le\varepsilon_{ki}<\Delta q^*_{nki}\left(\boldsymbol{\delta}\right);\\ 0, & \varepsilon_{ki}<\min\left(0,\Delta q^*_{nki}\left(\boldsymbol{\delta}\right)\right)\ \text{ or }\ \varepsilon_{ki}>\max\left(0,\Delta q^*_{nki}\left(\boldsymbol{\delta}\right)\right);\\ -1, & \Delta q^*_{nki}\left(\boldsymbol{\delta}\right)<\varepsilon_{ki}\le 0,\end{cases}
\end{aligned}
$$

which can only take two values with non-zero probability, depending on the sign of $\Delta q^*_{nki}\left(\boldsymbol{\delta}\right)$. The probabilities are defined by $1-p_{nk}:=\mathcal{P}\left(W_{nk}=0\mid\boldsymbol{X}\right)$ (recall that a

dropped i subscript refers to the generic random vector $\left(\boldsymbol{X}^\top, Y\right)$), which evaluate to

$$\begin{aligned}
p_{nk} =& \mathcal{P}\left(\min\left(0, \Delta q^*_{nk}\left(\boldsymbol{\delta}\right)\right) < \varepsilon_k \leq \max\left(0, \Delta q^*_{nk}\left(\boldsymbol{\delta}\right)\right) \mid \boldsymbol{X}\right) \\
=& F_{g_k(Y)|\boldsymbol{X}}\left(\max\left(0, \Delta q^*_{nk}\left(\boldsymbol{\delta}\right)\right) + q_k\left(\boldsymbol{\theta}_0\right) \mid \boldsymbol{X}\right) \\
& - F_{g_k(Y)|\boldsymbol{X}}\left(\min\left(0, \Delta q^*_{nk}\left(\boldsymbol{\delta}\right)\right) + q_k\left(\boldsymbol{\theta}_0\right) \mid \boldsymbol{X}\right) \\
=& \operatorname{sign}\left(\Delta q^*_{nk}\left(\boldsymbol{\delta}\right)\right) \frac{F_{g_k(Y)|\boldsymbol{X}}\left(\Delta q^*_{nk}\left(\boldsymbol{\delta}\right) + q_k\left(\boldsymbol{\theta}_0\right) \mid \boldsymbol{X}\right) - F_{g_k(Y)|\boldsymbol{X}}\left(q_k\left(\boldsymbol{\theta}_0\right) \mid \boldsymbol{X}\right)}{\Delta q^*_{nk}\left(\boldsymbol{\delta}\right)} \Delta q^*_{nk}\left(\boldsymbol{\delta}\right) \\
=& \left[f_{g_k(Y)|\boldsymbol{X}}\left(q_k\left(\boldsymbol{\theta}_0\right) \mid \boldsymbol{X}\right) + o_p\left(1\right)\right] \left|\Delta q^*_{nk}\left(\boldsymbol{\delta}\right)\right|
\end{aligned}$$

as $n \to \infty$, since $\Delta q^*_{nk}\left(\boldsymbol{\delta}\right) \to_p 0$. Note that $p_{nk} = O_p\left(n^{-1/2}\right)$ as $n \to \infty$. Now, let us find the variance of $Z_{3nki}\left(\boldsymbol{\delta}\right)$ in a few stages. First,

$$\begin{aligned}
\operatorname{Var}\left(Z_{3nk}\left(\boldsymbol{\delta}\right)\right) &= n^2 \operatorname{Var}\left(\left(\Delta q^*_{nk}\left(\boldsymbol{\delta}\right) - \varepsilon_k\right) W_{nk}\right) \\
&\leq n^2 \mathbb{E}\left(\left(\Delta q^*_{nk}\left(\boldsymbol{\delta}\right) - \varepsilon_k\right)^2 W_{nk}^2\right) \\
&= n^2 \mathbb{E}\left(\mathbb{E}\left(\left(\Delta q^*_{nk}\left(\boldsymbol{\delta}\right) - \varepsilon_k\right)^2 W_{nk}^2 \mid \boldsymbol{X}\right)\right),
\end{aligned}$$

based on the definition of variance in terms of expectation, and based on the law of total expectation (double expectation). Next, the inner expectation can be decomposed by the law of total expectation on the partition $W_{nk} = 0$ and $W_{nk} \neq 0$:

$$\begin{aligned}
\operatorname{Var}\left(Z_{3nk}\left(\boldsymbol{\delta}\right)\right) &\leq n^2 \mathbb{E}\left(0\left(1 - p_{nk}\right) + \mathbb{E}\left(\left(\Delta q^*_{nk}\left(\boldsymbol{\delta}\right) - \varepsilon_k\right)^2 \mid \boldsymbol{X}, W_{nk} \neq 0\right) p_{nk}\right) \\
&\leq n^2 \mathbb{E}\left(\mathbb{E}\left(\Delta q^*_{nk}\left(\boldsymbol{\delta}\right)^2 \mid \boldsymbol{X}, W_{nk} \neq 0\right) p_{nk}\right),
\end{aligned}$$

where the second line holds because $\left(\Delta q^*_{nk}\left(\boldsymbol{\delta}\right) - \varepsilon_k\right)^2 \leq \Delta q^*_{nk}\left(\boldsymbol{\delta}\right)^2$ almost surely when $W_{nk} \neq 0$. But now the inner expectation is the expectation of a constant, because $\Delta q^*_{nk}\left(\boldsymbol{\delta}\right)$ is fixed when $\boldsymbol{X}$ is fixed. This means that

$$\operatorname{Var}\left(Z_{3nk}\left(\boldsymbol{\delta}\right)\right) \leq n^2 \mathbb{E}\left(\Delta q^*_{nk}\left(\boldsymbol{\delta}\right)^2 p_{nk}\right) = n^2 O\left(n^{-3/2}\right) = O\left(n^{1/2}\right)$$

as $n \to \infty$. Now it can be shown that the variance of the average vanishes as $n \to \infty$:

$$\operatorname{Var}\left(\bar{Z}_{3nk}\left(\boldsymbol{\delta}\right)\right) = \frac{1}{n} \operatorname{Var}\left(Z_{3nk}\left(\boldsymbol{\delta}\right)\right) = O\left(n^{-1/2}\right).$$

Now, by Chebyshev's inequality, for any $\epsilon > 0$,

$$\mathcal{P}\left(\left|\bar{Z}_{3nk}\left(\boldsymbol{\delta}\right) - \frac{1}{2}\boldsymbol{\delta}^\top \boldsymbol{D}_{1k}\boldsymbol{\delta} - R_{3nk}\right| > \epsilon\right) \leq \frac{1}{\epsilon^2}\mathrm{Var}\left(\bar{Z}_{3nk}\left(\boldsymbol{\delta}\right)\right) \to 0$$

as $n \to \infty$, so that $\bar{Z}_{3nk}\left(\boldsymbol{\delta}\right) \to_p \boldsymbol{\delta}^\top \boldsymbol{D}_{1k}\boldsymbol{\delta}/2$ (because the remainder $R_{3nk} \to 0$), and $\sum_k \bar{Z}_{3nk}\left(\boldsymbol{\delta}\right) \to_p \boldsymbol{\delta}^\top \boldsymbol{D}_1\boldsymbol{\delta}/2$.

Combining the asymptotic behaviour of $\sqrt{n}\bar{\boldsymbol{Z}}_{1n}$, $\bar{Z}_{2n}\left(\boldsymbol{\delta}\right)$, and $\bar{Z}_{3nk}\left(\boldsymbol{\delta}\right)$, we obtain

$$Z_n^*\left(\boldsymbol{\delta}\right) \to_d Z_\infty^*\left(\boldsymbol{\delta}\right) = \boldsymbol{\delta}^\top \boldsymbol{Z}_{1\infty} + \frac{1}{2}\boldsymbol{\delta}^\top \boldsymbol{D}_1\boldsymbol{\delta}.$$

It follows that the arg min of these objective functions also converge in distribution (cf. Hjort and Pollard, 1993). Condition 2(a)(i) ensures that $\boldsymbol{D}_1$ is non-zero and is therefore positive definite, and this ensures that $Z_\infty^*\left(\boldsymbol{\delta}\right)$ is uniquely minimized at $-\boldsymbol{D}_1^{-1}\boldsymbol{Z}_{1\infty}$, which exists because of Condition 2(d). Finally,

$$\begin{aligned}\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) &= \arg\min_{\boldsymbol{\delta}} Z_n^*\left(\boldsymbol{\delta}\right) \\ &\to_d \arg\min_{\boldsymbol{\delta}} Z_\infty^*\left(\boldsymbol{\delta}\right) \\ &= -\boldsymbol{D}_1^{-1}\boldsymbol{Z}_{1\infty}.\end{aligned}$$

□

Appendix E

Proofs related to the New Copula Families

This appendix first provides details and proofs of the DJ and extDJ copula families, introduced in Section 6.2. Then, details and proofs of the IGL and IG copula families – special cases of the DJ and extDJ classes, respectively – are provided, introduced in Section 6.1. Note that the DJ and extDJ families are discussed here first, because these results follow through to the IGL and IG copula families. This order is reversed in Chapter 6 because it makes more sense to introduce the various families by generalizing the IGL and IG families to the DJ and extDJ families.

Note that Durante and Jaworski (2012), the founders of the DJ copula class, do not discuss much of the properties of the DJ class, nor do they discuss much about parametric cases of this copula class.

Each section in this appendix is laid out in the same way. First, properties of the generating function of the copula class are outlined. Second, some key formulas related to the copula class are provided. Third, proofs of the properties outlined in Chapter 6 are given.

E.1 DJ Copula Class

E.1.1 Generating Function Properties

Not much can be said about the DJ generating functions outside of their definition. However, the following family of functions plays an important role in the IG and other copula families.

Definition E.1.1. *Let $\psi : [0, \infty) \to [0, 1]$ be a concave distribution function or a convex*

survival function with derivative ψ'. The kappa function of ψ *is defined as*

$$\kappa_\psi(t) = \psi(t) - t\psi'(t) \tag{E.1.1}$$

for $t > 0$.

Proposition E.1.1. *Let $\psi : [0, \infty) \to [0, 1]$ be a concave distribution function (resp. convex survival function) with derivative ψ'. Then κ_ψ, defined in Equation (E.1.1), is a distribution function (resp. survival function). If ψ'' exists, then*

$$\kappa'_\psi(t) = -t\psi''(t)$$

for $t > 0$.

Proof. When ψ'' exists,

$$\kappa'_\psi(t) = \psi'(t) - \psi'(t) - t\psi''(t) = -t\psi''(t)$$

for $t > 0$. The monotonicity of κ_ψ follows directly from the concavity of ψ, because ψ'' does not change sign.

It remains to evaluate κ_ψ at the endpoints of the support. First, notice that the existence of the upper integral of ψ' to infinity implies $\psi'(t) = o\left(1/t\right)$ as $t \to \infty$, so that $t\psi'(t) \to 0$. Now,

$$\lim_{t\to\infty} \kappa_\psi(t) = \lim_{t\to\infty} \psi(t) - \lim_{t\to\infty} t\psi'(t) = \lim_{t\to\infty} \psi(t).$$

Similarly, the existence of the lower integral of ψ' from zero implies that $\psi'(t) = o\left(1/t\right)$ as $t \downarrow 0$, so that $t\psi'(t) \to 0$. Now,

$$\lim_{t\downarrow 0} \kappa_\psi(t) = \lim_{t\downarrow 0} \psi(t) - \lim_{t\downarrow 0} t\psi'(t) = \lim_{t\downarrow 0} \psi(t).$$

Since κ_ψ is monotone and evaluates to the same values as ψ at the endpoints of the support (either 0 or 1), then κ_ψ is a distribution function when ψ is a distribution function, and κ_ψ is a survival function when ψ is a survival function. □

E.1.2 Formulas

The DJ copula generated by ψ has the following distributional quantities, for $(u, v) \in (0, 1)^2$.

- Whenever ψ' exists, the conditional distributions are

$$C_{\mathrm{DJ},2|1}\left(v|u;\psi\right) = \kappa_\psi\left(u^{-1}\psi^{\leftarrow}\left(v\right)\right), \tag{E.1.2}$$

where κ_ψ is defined in Equation (E.1.1), and

$$C_{\mathrm{DJ},1|2}\left(u|v;\psi\right) = \frac{\psi'\left(u^{-1}\psi^{\leftarrow}\left(v\right)\right)}{\psi'\left(\psi^{\leftarrow}\left(v\right)\right)}. \tag{E.1.3}$$

- Whenever ψ'' exists, the copula density is

$$c_{\mathrm{DJ}}\left(u,v;\psi\right) = -\frac{\psi^{\leftarrow}\left(v\right)\psi''\left(u^{-1}\psi^{\leftarrow}\left(v\right)\right)}{u^2\psi'\left(\psi^{\leftarrow}\left(v\right)\right)}. \tag{E.1.4}$$

E.1.3 Proofs of Properties

Proof of Proposition 6.2.2 (equivalence class). Consider ψ so that $(u,v) \mapsto C_{\mathrm{DJ}}\left(u,v;\psi\right)$ is a copula. Then

$$\left(\psi \circ g_{-1/\alpha}\right)^{\leftarrow} = g_{-1/\alpha}^{\leftarrow} \circ \psi^{\leftarrow}$$

(the same direction of monotonicity of ψ and $g_{-1/\alpha}$ is not required), and for $(u,v) \in (0,1)^2$,

$$\begin{aligned}
C_{\mathrm{DJ}}^{(\alpha)}\left(u,v;\psi\circ g_{-1/\alpha}\right) &= u\psi\left(g_{-1/\alpha}\left(u^{\alpha}g_{-1/\alpha}^{\leftarrow}\left(\psi^{\leftarrow}\left(v\right)\right)\right)\right)\\
&= u\psi\left(u^{-1}g_{-1/\alpha}\left(g_{-1/\alpha}^{\leftarrow}\left(\psi^{\leftarrow}\left(v\right)\right)\right)\right)\\
&= u\psi\left(u^{-1}\psi^{\leftarrow}\left(v\right)\right)\\
&= C_{\mathrm{DJ}}\left(u,v;\psi\right).
\end{aligned}$$

It follows that the class of copulas defined in Equation (6.2.2) is not empty, since the original DJ class of copulas is non-empty.

For the converse, suppose ψ_α is such that $(u,v) \mapsto C_{\mathrm{DJ}}^{(\alpha)}\left(u,v;\psi_\alpha\right)$ is a copula. Then

for $(u,v) \in (0,1)^2$,

$$\begin{aligned} C_{\mathrm{DJ}}\left(u,v;\psi_\alpha \circ g_{-\alpha}\right) &= u\psi_\alpha\left(g_{-\alpha}\left(u^{-1}g_{-\alpha}^{\leftarrow}\left(\psi_\alpha^{\leftarrow}(v)\right)\right)\right) \\ &= u\psi_\alpha\left(u^{\alpha}g_{-\alpha}\left(g_{-\alpha}^{\leftarrow}\left(\psi_\alpha^{\leftarrow}(v)\right)\right)\right) \\ &= u\psi_\alpha\left(u^{\alpha}\psi_\alpha^{\leftarrow}(v)\right) \\ &= C_{\mathrm{DJ}}^{(\alpha)}\left(u,v;\psi_\alpha\right). \end{aligned}$$

□

Proof of Corollary 6.2.3 (scale invariance). Consider the function $g_{-\alpha}(t) = t^{-\alpha}$ for $t \in (0,\infty)$, and take $(u,v) \in (0,1)^2$. By Proposition 6.2.2,

$$C_{\mathrm{DJ}}^{(\alpha)}\left(u,v;\psi \circ g\right) = C_{\mathrm{DJ}}\left(u,v;\psi \circ g \circ g_{-\alpha}\right).$$

Now take

$$g_{-1/\alpha} = \left(g \circ g_{-\alpha}\right)^{\leftarrow} = g_{-\alpha}^{\leftarrow} \circ g^{\leftarrow}$$

(the direction of monotonicity of g and $g_{-\alpha}$ is not required), defined by

$$g_{-1/\alpha}(t) = \left(\frac{t}{a}\right)^{-1/\alpha} = \tilde{a}t^{-1/\alpha}$$

for $t \in (0,\infty)$, where $\tilde{a} = a^{1/\alpha} > 0$. By Proposition 6.2.2,

$$\begin{aligned} C_{\mathrm{DJ}}\left(u,v;\psi \circ g \circ g_{-\alpha}\right) &= C_{\mathrm{DJ}}^{(\alpha)}\left(u,v;\psi \circ g \circ g_{-\alpha} \circ g_{-1/\alpha}\right) \\ &= C_{\mathrm{DJ}}^{(\alpha)}\left(u,v;\psi \circ g \circ g_{-\alpha} \circ g_{-\alpha}^{\leftarrow} \circ g^{\leftarrow}\right) \\ &= C_{\mathrm{DJ}}^{(\alpha)}\left(u,v;\psi\right), \end{aligned}$$

so that $C_{\mathrm{DJ}}^{(\alpha)}\left(u,v;\psi \circ g\right) = C_{\mathrm{DJ}}^{(\alpha)}\left(u,v;\psi\right)$. □

Proof of Proposition 6.2.4 (stochastic representation). First, notice that the distribution function of Z conditional on Y, described in Equation (6.2.4), is in fact a distribution function, since ψ' is decreasing due to the concavity of ψ. Now, the joint distribution

function of (U, Y) in Equation (6.2.4) can be shown using Equation (6.2.3):

$$\begin{aligned}
\mathcal{P}(U \leq u, Y \leq y) &= \int_0^y \mathcal{P}\left(\frac{Y}{Z} \leq u \,\middle|\, Y = w\right) \psi'(w)\,\mathrm{d}w \\
&= \int_0^y \mathcal{P}\left(Z \geq u^{-1}w \mid Y = w\right) \psi'(w)\,\mathrm{d}w \\
&= \int_0^y \psi'\left(u^{-1}w\right)\mathrm{d}w \\
&= u\psi\left(u^{-1}y\right)
\end{aligned}$$

for $(u, y) \in (0, 1) \times [0, \infty)$. It follows that U has a $\mathrm{Unif}(0, 1)$ distribution:

$$\mathcal{P}(U \leq u) = \lim_{y\to\infty} \mathcal{P}(U \leq u, Y \leq y) = u \lim_{y\to\infty} \psi\left(u^{-1}y\right) = u$$

for $u \in (0, 1)$. The copula linking (U, Y) is therefore

$$C_{UY}(u, v) = \mathcal{P}\left(U \leq u, Y \leq \psi^{\leftarrow}(v)\right) = u\psi\left(u^{-1}\psi^{\leftarrow}(v)\right),$$

which is the DJ copula with generating function ψ.

Now, using the distribution function of $\tilde{Y} = 1/Y$ given by

$$F_{\tilde{Y}}(y) = 1 - \psi\left(\frac{1}{y}\right),$$

the copula linking $\left(U, \tilde{Y}\right)$ is

$$C_{U\tilde{Y}}(u, v) = \mathcal{P}\left(U \leq u, F_{\tilde{Y}}\left(\tilde{Y}\right) \leq v\right) = \mathcal{P}\left(U \leq u, 1 - \psi(Y) \leq v\right) = \mathrm{R}_2 C_{UY}(u, v),$$

by definition of the R_2 reflection in Definition 2.4.1. Since C_{UY} is the DJ copula with generating function ψ, by Proposition 6.2.1, $C_{U\tilde{Y}}$ is the DJ copula with the convex survival function $1 - \psi$. □

Proof of Proposition 6.2.7 (independence copula). Fix $(u, v) \in (0, 1)^2$. Using Equation (6.2.1), the copula family evaluates to

$$C_{\mathrm{DJ}}(u, v; \psi_\theta) = u \min\left(u^{-\theta}v, 1\right)$$

for $\theta \in (0,1)$. Since $u^{-\theta}v < 1$ for θ in some neighbourhood of 0,

$$\lim_{\theta \downarrow 0} C_{\mathrm{DJ}}\left(u, v; \psi_\theta\right) = \lim_{\theta \downarrow 0} u^{1-\theta} v = uv = C^{\perp}\left(u, v\right).$$

□

Proof of Proposition 6.2.8 (DJ Weibull CCEVI). The DJ Weibull copula family is described for $\beta \geq 1$ by

$$C_{\mathrm{DJW}}\left(u, v; \beta\right) = uv^{u^{-1/\beta}},$$

and results when taking the DJ generating function $\psi_\beta : t \mapsto \exp\left(-t^{1/\beta}\right)$ in Equation (6.2.1), which is strictly decreasing with inverse function $\psi_\beta^{\leftarrow}\left(w\right) = \left(-\log w\right)^\beta$ for $w \in (0,1)$ and derivative

$$\psi_\beta'\left(t\right) = -\frac{1}{\beta} t^{1/\beta - 1} \psi_\beta\left(t\right).$$

To find the CCEVI, fix $u \in (0,1)$ and let $x = u^{-1/\beta} > 1$. We use the copula density, which is

$$c_{\mathrm{DJW}}\left(u, v; \beta\right) = \frac{x v^{x-1}\left(\beta - 1 - x \log v\right)}{\beta},$$

where $v \in (0,1)$. Since

$$\lim_{v \uparrow 1} c_{\mathrm{DJW}}\left(u, v; \beta\right) = x \frac{(\beta - 1)}{\beta} \in (0, \infty),$$

the CCEVI is 1, by Proposition 2.4.2.

To find the CCEVI of the reflection copula, we use the conditional distribution of the DJ family found in Equation (E.1.2),

$$\hat{C}_{\mathrm{DJW},2|1}\left(v|u; \beta\right) = 1 - \kappa_\beta\left(\left(1-u\right)^{-1}\left(-\log\left(1-v\right)\right)^\beta\right),$$

where

$$\begin{aligned}
\kappa_\beta\left(t\right) &= \psi_\beta\left(t\right) - t\psi_\beta'\left(t\right) \\
&= \exp\left(-t^{1/\beta}\right)\left(1 + \frac{1}{\beta} t^{1/\beta}\right) \\
&\sim \exp\left(-t^{1/\beta}\right)
\end{aligned}$$

as $t \to \infty$. Now, consider the following limit for any $s > 0$:

$$\begin{aligned}
\lim_{t\to\infty} \frac{1 - \hat{C}_{\mathrm{DJW},2|1}\left(1 - (ts)^{-1}\,|u;\beta\right)}{1 - \hat{C}_{\mathrm{DJW},2|1}\left(1 - t^{-1}|u;\beta\right)} &= \lim_{t\to\infty} \frac{\kappa_\beta\left((1-u)^{-1}\left(\log(ts)\right)^{\beta}\right)}{\kappa_\beta\left((1-u)^{-1}\left(\log t\right)^{\beta}\right)} \\
&= \left[\lim_{t\to\infty} \frac{\exp\left(-\log(ts)\right)}{\exp\left(-\log t\right)}\right]^{(1-u)^{-1/\beta}} \\
&= s^{-1/(1-u)^{1/\beta}},
\end{aligned}$$

so that the CCEVI of the reflection copula is $u \mapsto (1-u)^{1/\beta}$, which is decreasing. $\square$

E.2 extDJ Copula Class

E.2.1 Generating Function Properties

This section examines properties of the extDJ generating functions.

Proposition E.2.1. *Let $\psi : [0,\infty) \to [0,1]$ be a distribution function. The function $H_\psi : [1,\infty) \times (0,\infty) \to [0,1]$, defined in Equation (6.2.5), has the following properties:*

1. *If ψ is differentiable, then H_ψ has derivatives*

$$\mathrm{D}_1 H_\psi(t;\eta) = -\frac{1}{t^2}\left[\psi\left(\frac{1}{\eta\log t}\right) + \frac{1}{\eta(\log t)^2}\psi'\left(\frac{1}{\eta\log t}\right)\right] \tag{E.2.1}$$

and

$$\mathrm{D}_2 H_\psi(t;\eta) = -\frac{1}{t\eta^2\log t}\psi'\left(\frac{1}{\eta\log t}\right). \tag{E.2.2}$$

If ψ is twice-differentiable, then

$$\mathrm{D}_{12} H_\psi(t;\eta) = \frac{1}{(t\eta\log t)^2}\left[(1+\log t)\,\psi'\left(\frac{1}{\eta\log t}\right) + \frac{1}{\eta\log t}\psi''\left(\frac{1}{\eta\log t}\right)\right] \tag{E.2.3}$$

for $(t,\eta) \in (1,\infty) \times (0,\infty)$.

2. *For any $\eta > 0$, $H_\psi(\cdot;\eta)$ is strictly decreasing, so that $H_\psi^{\leftarrow}(\cdot;\eta)$ is the unique inverse function.*

3. *For any $\eta > 0$, $H_\psi(\cdot;\eta)$ is a survival function over $[1,\infty)$.*

4. *For $t > 1$, we have*

$$H_\psi(t;0) := \lim_{\eta\downarrow 0} H_\psi(t;\eta) = \frac{1}{t} \tag{E.2.4}$$

and

$$\lim_{\eta\to\infty} H_\psi(t;\eta) = 0. \tag{E.2.5}$$

5. *For any $t > 1$, $H_\psi(t;\cdot)$ is non-increasing, and is strictly decreasing if ψ is strictly increasing. It follows that the functions $t \mapsto H_\psi(t;\eta)$ are bounded above by $t \mapsto 1/t$ for each $\eta > 0$.*

6. *If ψ is differentiable, then*

$$H_\psi(t;\eta) + \eta\, \mathrm{D}_2 H_\psi(t;\eta) = \frac{1}{t}\kappa_\psi\left(\frac{1}{\eta \log t}\right) \tag{E.2.6}$$

$$= H_{\kappa_\psi}(t;\eta) \tag{E.2.7}$$

for $(t,\eta) \in (1,\infty) \times (0,\infty)$, where κ_ψ is defined in Equation (E.1.1).

7. *If ψ is strictly increasing, then*

$$\frac{1}{\eta \log H_\psi^{\leftarrow}(v;\eta)} = \psi^{\leftarrow}\left(v H_\psi^{\leftarrow}(v;\eta)\right) \tag{E.2.8}$$

for $\eta > 0$ and $v \in (0,1)$.

Proof. Property 1 follows by direct differentiation of H_ψ.

Property 2: take $1 < t_1 < t_2$. Now,

$$H_\psi(t_1;\eta) = \frac{1}{t_1}\psi\left(\frac{1}{\eta \log t_1}\right) > \frac{1}{t_2}\psi\left(\frac{1}{\eta \log t_1}\right) \geq \frac{1}{t_2}\psi\left(\frac{1}{\eta \log t_2}\right) = H_\psi(t_2;\eta),$$

where the second inequality holds because ψ is increasing.

Property 3: we have

$$\lim_{t\downarrow 1} H_\psi(t;\eta) = \lim_{t\downarrow 1}\psi\left(\frac{1}{\eta \log t}\right) = \lim_{x\to\infty}\psi(x) = 1,$$

and

$$\lim_{t\to\infty} H_\psi(t;\eta) = \left(\lim_{t\to\infty}\frac{1}{t}\right)\left(\lim_{x\downarrow 0}\psi(x)\right) = 0,$$

since ψ is a distribution function on $[0,\infty)$. Since $H_\psi(\cdot;\eta)$ is decreasing, as identified in Property 2, $H_\psi(\cdot;\eta)$ is a survival function.

Property 4: for t >1,

$$\lim_{\eta\downarrow 0} H_\psi(t;\eta) = \frac{1}{t}\lim_{\eta\downarrow 0}\psi\left(\frac{1}{\eta \log t}\right) = \frac{1}{t}\lim_{x\to\infty}\psi(x) = \frac{1}{t}$$

and

$$\lim_{\eta\to\infty} H_\psi(t;\eta) = \frac{1}{t}\lim_{\eta\to\infty}\psi\left(\frac{1}{\eta \log t}\right) = \frac{1}{t}\lim_{x\downarrow 0}\psi(x) = 0,$$

since ψ is a distribution function on $[0,\infty)$.

Property 5: take $0 < \eta_1 < \eta_2$. Now,

$$H_\psi(t;\eta_1) = \frac{1}{t}\psi\left(\frac{1}{\eta_1 \log t}\right) \geq \frac{1}{t}\psi\left(\frac{1}{\eta_2 \log t}\right) = H_\psi(t;\eta_2),$$

where the inequality arises since ψ is non-decreasing. If ψ is strictly increasing, then the inequality also becomes strict.

Property 6: the result follows by simple algebra, using the form of $\mathrm{D}_2 H_\psi$ in Equation (E.2.2).

Property 7: by Property 2, $H_\psi(\cdot;\eta)$ is strictly decreasing, so that $H_\psi^{\leftarrow}(\cdot;\eta)$ is its unique inverse. From the definition of $H_\psi(\cdot;\eta)$ in Equation (6.2.5), the inverse function $H_\psi^{\leftarrow}(\cdot;\eta)$ satisfies

$$v = \frac{1}{H_\psi^{\leftarrow}(v;\eta)}\psi\left(\frac{1}{\eta \log H_\psi^{\leftarrow}(v;\eta)}\right)$$

for $v \in (0,1)$. Multiplying both sides by $H_\psi^{\leftarrow}(v;\eta)$ and applying the inverse function $\psi^{\leftarrow}$ arrives at the desired identity. □

E.2.2 Formulas

Let (ψ,θ) be an extDJ generator, with corresponding extDJ generating function H_ψ defined in Equation (6.2.5). The corresponding extDJ copula, defined in Equation (6.2.6), has the following distributional quantities for $(u,v) \in (0,1)^2$, using the shorthand notation $t = H_\psi^{\leftarrow}(v;\theta)$ when appropriate.

- Whenever ψ' exists, the 2|1 conditional distribution function is

$$C_{\mathrm{extDJ},2|1}\left(v \mid u;\psi;\theta\right) = \frac{1}{t}\kappa_\psi\left(\frac{1}{\theta u \log t}\right) = H_{\kappa_\psi}\left(H_\psi^{\leftarrow}(v;\theta);\theta u\right), \tag{E.2.9}$$

  where κ_ψ is defined in Equation (E.1.1). This can be found using Proposition E.2.1(6). The corresponding quantile function is

$$C_{\mathrm{extDJ},2|1}^{\leftarrow}\left(\tau \mid u;\psi,\theta\right) = H_\psi\left(H_{\kappa_\psi}^{\leftarrow}(\tau;\theta u);\theta\right) \tag{E.2.10}$$

  for $\tau \in (0,1)$.

- Whenever ψ' exists, the 1|2 conditional distribution function is

$$C_{\text{extDJ},1|2}\left(u \mid v;\psi,\theta\right) = u\frac{\mathrm{D}_1 H_\psi\left(H_\psi^{\leftarrow}(v;\theta);\theta u\right)}{\mathrm{D}_1 H_\psi\left(H_\psi^{\leftarrow}(v;\theta);\theta\right)} \tag{E.2.11}$$
$$= u\frac{\psi\left(\frac{1}{\theta u\log t}\right)+\frac{1}{\theta u(\log t)^2}\psi'\left(\frac{1}{\theta u\log t}\right)}{\psi\left(\frac{1}{\theta\log t}\right)+\frac{1}{\theta(\log t)^2}\psi'\left(\frac{1}{\theta\log t}\right)}.$$

This result can be determined using the formula for $\mathrm{D}_1 H_\psi$, found in Equation (E.2.1).

- Whenever ψ is twice-differentiable, the copula density is

$$c_{\text{extDJ}}\left(u,v;\psi,\theta\right) = \frac{\mathrm{D}_1 H_{\kappa_\psi}\left(H_\psi^{\leftarrow}(v;\theta);\theta u\right)}{\mathrm{D}_1 H_\psi\left(H_\psi^{\leftarrow}(v;\theta);\theta\right)} \tag{E.2.12}$$
$$= \frac{\kappa_\psi\left(\frac{1}{\theta u\log t}\right)+\frac{1}{\theta u(\log t)^2}\kappa_\psi'\left(\frac{1}{\theta u\log t}\right)}{\psi\left(\frac{1}{\theta\log t}\right)+\frac{1}{\theta(\log t)^2}\psi'\left(\frac{1}{\theta\log t}\right)},$$

where κ_ψ is defined in Equation (E.1.1). Note that $\mathrm{D}_1 H_{\kappa_\psi}$ and $\mathrm{D}_1 H_\psi$ are both negative, due to Proposition E.2.1(2), so that this density is non-negative.

E.2.3 Proofs of Properties

Proof of Proposition 6.2.10 (limit copulas). Take $(u,v)\in(0,1)^2$. To show the $\theta\to\infty$ limit, rewrite the extDJ copula by substituting the definition of H_ψ from Equation (6.2.5), then apply the identity in Proposition E.2.17, to obtain

$$C_{\text{extDJ}}\left(u,v;\psi,\theta\right) = u\frac{1}{H_\psi^{\leftarrow}(v;\theta)}\psi\left(u^{-1}\frac{1}{\theta\log H_\psi^{\leftarrow}(v;\theta)}\right)$$
$$= \frac{u\psi\left(u^{-1}\psi^{\leftarrow}\left(vH_\psi^{\leftarrow}(v;\theta)\right)\right)}{H_\psi^{\leftarrow}(v;\theta)}.$$

Since $H_k^{\leftarrow}(v;\theta)\to 1$ as $\theta\to\infty$ for any $v\in(0,1)$, we obtain the DJ copula as a limit:

$$\lim_{\theta\to\infty} C_{\text{extDJ}}\left(u,v;\psi,\theta\right) = \lim_{h\to 1}\frac{u\psi\left(u^{-1}\psi^{\leftarrow}(vh)\right)}{h} = u\psi\left(u^{-1}\psi^{\leftarrow}(v)\right) = C_{\text{DJ}}\left(u,v;\psi\right).$$

The $\theta\downarrow 0$ limit follows directly due to the limiting curve $H_\psi(t;\theta)\to 1/t$ for $t\in(1,\infty)$

identified in Equation (E.2.4):

$$\lim_{\theta\downarrow 0} C_{\mathrm{extDJ}}\left(u,v;\psi,\theta\right) = u\frac{1}{1/v} = C^{\perp}\left(u,v\right).$$

□

Proof of Proposition 6.2.11 (stochastic representation). First, the validity of the distribution functions F_Y and $F_{U|Y}$ are verified. The validity of F_Y follows directly by Proposition E.2.1(3). Next, for any $y > 1$, after interchanging the derivative and limit, we have

$$\lim_{u\downarrow 0} F_{U|Y}\left(u \mid y\right) = \left(\lim_{u\downarrow 0} u\right)\frac{1}{\mathrm{D}_1 H_\psi\left(y;\theta\right)}\frac{\mathrm{d}}{\mathrm{d}y}\frac{1}{y} = 0$$

due to Proposition E.2.1(4), and

$$\lim_{u\uparrow 1} F_{U|Y}\left(u \mid y\right) = \frac{\mathrm{D}_1 H_\psi\left(y;\theta\right)}{\mathrm{D}_1 H_\psi\left(y;\theta\right)} = 1.$$

The density is non-negative:

$$f_{U|Y}\left(u \mid y\right) = \frac{\mathrm{D}_1 H_\psi\left(y;\theta u\right) + \theta u\,\mathrm{D}_{12} H_\psi\left(y;\theta u\right)}{\mathrm{D}_1 H_\psi\left(y;\theta\right)} = \frac{\mathrm{D}_1 H_{\kappa_\psi}\left(y;\theta u\right)}{\mathrm{D}_1 H_\psi\left(y;\theta\right)},$$

by Proposition E.2.1(6), where κ_ψ is defined in Equation (E.1.1). Proposition E.1.1 states that κ_ψ is a distribution function, so that by Proposition E.2.1(2), both $\mathrm{D}_1 H_{\kappa_\psi}$ and $\mathrm{D}_1 H_\psi$ are negative, and therefore $f_{U|Y}$ is positive.

The joint distribution of (U,Y) is

$$\begin{aligned} F_{U,Y}\left(u,y\right) &= \int_1^y F_{U|Y}\left(u \mid t\right) f_Y\left(t\right)\mathrm{d}t \\ &= -\int_1^y \frac{u\,\mathrm{D}_1 H_\psi\left(t;\theta u\right)}{\mathrm{D}_1 H_\psi\left(t;\theta\right)}\mathrm{D}_1 H_\psi\left(t;\theta\right)\mathrm{d}t \\ &= u - uH_\psi\left(y;\theta u\right) \end{aligned}$$

for $(u,y) \in (0,1)\times(1,\infty)$, since $H_\psi\left(1;\theta u\right) = 1$ by Proposition E.2.1(3).

To find the copula, we need

$$F_U\left(u\right) = \lim_{y\to\infty} F_{U,Y}\left(u,y\right) = u - u\lim_{y\to\infty} H_\psi\left(y;\theta u\right) = u,$$

since $H_\psi(\cdot;\theta u)$ is a survival function, as identified in Proposition E.2.1(3). We also need

$$Q_Y(\tau) = H_\psi^{\leftarrow}(1-\tau;\theta)$$

for $\tau \in (0,1)$.

Now, relevant distributional quantities related to $\left(U, 1/Y\right)$ are

$$F_{U,1/Y}(u,t) = F_U(u) - F_{U,Y}\left(u, \frac{1}{t}\right) = u - F_{U,Y}\left(u, \frac{1}{t}\right)$$

for $(u,t) \in (0,1) \times (1,\infty)$, and

$$Q_{1/Y}(\tau) = \frac{1}{Q_Y(1-\tau)}$$

for $\tau \in (0,1)$. Sklar's theorem gives us the copula

$$\begin{aligned} C_{U,1/Y}(u,v) &= F_{U,1/Y}\left(u, Q_{1/Y}(v)\right) \\ &= u - F_{U,Y}\left(u, Q_Y(1-v)\right) \\ &= uH_\psi\left(H_\psi^{\leftarrow}(v;\theta);\theta u\right) \end{aligned}$$

for $(u,v) \in (0,1)^2$, which is the extDJ copula generated by (ψ,θ). □

Proof of Proposition 6.2.12. Evaluating the extDJ copula using $\tilde{H}_\psi$, we obtain

$$\begin{aligned} u\tilde{H}_\psi\left(\tilde{H}_\psi^{\leftarrow}(v;\theta);\theta u\right) &= uH_\psi\left(g\left(g^{\leftarrow}\left(H_\psi^{\leftarrow}(v;\theta)\right)\right);\theta u\right) \\ &= uH_\psi\left(H_\psi^{\leftarrow}(v;\theta);\theta u\right) \end{aligned}$$

for all $(u,v) \in (0,1)^2$. Note that the direction of monotonicity of g does not matter. □

Proof of Proposition 6.2.13 (quadrant dependence). Take ψ to be a concave distribution function, $\theta > 0$, and $(u,v) \in (0,1)^2$. The argument is made clearer by letting $y = H_\psi^{\leftarrow}(v;\theta)$. Since $H_\psi(y;\cdot)$ is non-increasing, as indicated in Proposition E.2.1(5), we have $H_\psi(y;\theta u) \geq H_\psi(y;\theta) = v$. Multiplying both sides by u gives the desired result,

$$uH_\psi\left(H_\psi^{\leftarrow}(v;\theta);\theta u\right) \geq uv.$$

□

E.3 IGL Copula Class

E.3.1 Generating Function Properties

The mathematics underlying the IGL family of copulas relies heavily on the functions Ψ_k for $k > 1$, defined in Equation (6.1.7). These functions are also DJ generating functions. Properties of Ψ_k are discussed here.

Proposition E.3.1. *For $k > 1$, Ψ_k defined in Equation (6.1.7) has the following properties.*

1. *Derivatives: for $t > 0$,*
$$\Psi_k'(t) = \frac{\Gamma(k) - \Gamma^*\left(k, 1/t\right)}{\Gamma(k-1)} \tag{E.3.1}$$
and
$$\Psi_k''(t) = -\frac{t^{-k-1}\exp\left(-t^{-1}\right)}{\Gamma(k-1)}. \tag{E.3.2}$$

2. $\lim_{t\to\infty}\Psi_k(t) = 1$.

3. $\Psi_k' \in RV_{-k}$ *at infinity.*

4. $\Psi_k \in RV_{-1}$ *at* 0^+.

5. $\lim_{t\downarrow 0}\Psi_k(t) = 0$.

6. $\lim_{t\downarrow 0}\Psi_k'(t) = k-1$.

7. $\lim_{t\downarrow 0}\Psi_k''(t) = 0$.

8. *The kappa function of Ψ_k, defined in Equation (E.1.1), is*
$$\kappa_k(t) := \kappa_{\Psi_k}(t) = \frac{\Gamma^*\left(k-1, 1/t\right)}{\Gamma(k-1)} \tag{E.3.3}$$
for $t > 0$.

Proof. Property 1 follows by direct differentiation.

Property 2:

$$\begin{aligned}\lim_{t\to\infty}\Psi_k(t) &= \lim_{x\downarrow 0}\left(\frac{\Gamma(k) - \Gamma^*(k,x)}{x\Gamma(k-1)} + \frac{\Gamma^*(k-1,x)}{\Gamma(k-1)}\right)\\ &= \lim_{x\downarrow 0}\frac{-x^{k-1}\exp(-x)}{\Gamma(k-1)} + 1\\ &= 1,\end{aligned}$$

since $k > 1$.

Property 3: for $s > 0$,

$$\begin{aligned}
\lim_{t\to\infty} \frac{\Psi_k'(ts)}{\Psi_k'(ts)} &= \lim_{t\to\infty} \frac{\Gamma(k) - \Gamma^*\left(k, (ts)^{-1}\right)}{\Gamma(k) - \Gamma^*(k, t^{-1})} \\
&= \lim_{t\to\infty} \frac{(ts)^{-(k-1)} \exp\left(-(ts)^{-1}\right) s^{-1}\left(-t^{-2}\right)}{t^{-(k-1)} \exp\left(-t^{-1}\right)\left(-t^{-2}\right)} \\
&= s^{-k} \lim_{t\to\infty} \exp\left(t^{-1}\left(1 - s^{-1}\right)\right) \\
&= s^{-k}.
\end{aligned}$$

Property 4: for $s > 0$,

$$\begin{aligned}
\lim_{t\to\infty} \frac{\Psi_k\left((ts)^{-1}\right)}{\Psi_k(t^{-1})} &= s^{-1} \lim_{t\to\infty} \frac{\Psi_k'\left((ts)^{-1}\right)}{\Psi_k'(t^{-1})} \\
&= s^{-1} \lim_{t\to\infty} \frac{\Gamma(k) - \Gamma^*(k, ts)}{\Gamma(k) - \Gamma^*(k, t)} \\
&= s^{-1} \frac{\Gamma(k)}{\Gamma(k)} \\
&= s^{-1}.
\end{aligned}$$

Property 5:

$$\lim_{t\downarrow 0} \Psi_k(t) = \lim_{y\to\infty} \frac{\Gamma(k) - \Gamma^*(k, y)}{y\Gamma(k-1)} + \lim_{y\to\infty} \frac{\Gamma^*(k-1, y)}{\Gamma(k-1)} = 0$$

Property 6:

$$\lim_{t\downarrow 0} \Psi_k'(t) = \lim_{y\to\infty} \frac{\Gamma(k) - \Gamma^*(k, y)}{\Gamma(k-1)} = \frac{\Gamma(k)}{\Gamma(k-1)} = k - 1.$$

Property 7:

$$\lim_{t\downarrow 0} \Psi_k''(t) = -\lim_{y\to\infty} \frac{y^{k+1} \exp(-y)}{\Gamma(k-1)} = 0.$$

Property 8 follows by a direct substitution of Ψ_k and Ψ_k', found respectively in Equations (6.1.7) and (E.3.1). □

E.3.2 Formulas

Distributional quantities related to the IGL copula family are presented here. For neatness, results are bundled as a proposition.

Proposition E.3.2. *Let $k > 1$. The IGL copula $(u, v) \mapsto C_{IGL}(u, v; k)$ has*

$$C_{IGL,2|1}\left(v \mid u; k\right) = 1 - \frac{\Gamma^*\left(k-1, (1-u)/\Psi_k^{\leftarrow}(1-v)\right)}{\Gamma(k-1)}, \tag{E.3.4}$$

$$C_{IGL,1|2}\left(u \mid v; k\right) = 1 - \frac{\Gamma(k) - \Gamma^*\left(k, (1-u)/\Psi_k^{\leftarrow}(1-v)\right)}{\Gamma(k) - \Gamma^*\left(k, 1/\Psi_k^{\leftarrow}(1-v)\right)}, \tag{E.3.5}$$

and

$$c_{IGL}(u, v; k) = \frac{(1-u)^{k-1}\left[\Psi_k^{\leftarrow}(1-v)\right]^{-k}\exp\left\{-(1-u)/\Psi_k^{\leftarrow}(1-v)\right\}}{\Gamma(k) - \Gamma\left(k, 1/\Psi_k^{\leftarrow}(1-v)\right)} \tag{E.3.6}$$

for $(u, v) \in (0, 1)^2$ and $k > 1$.

Note that $C_{\mathrm{IGL},2|1}\left(\cdot \mid u; k\right)$ is the Gamma$(k-1, 1)$ distribution function composed with

$$v \mapsto \frac{1-u}{\Psi_k^{\leftarrow}(1-v)}.$$

Proof. We work with the reflection copula, which has a simpler form. Fix $k > 1$ and consider $(u, v) \in (0, 1)^2$.

For the $2|1$ distribution function, use the DJ copula $2|1$ distribution in Equation (E.2.9) to obtain

$$\hat{C}_{\mathrm{IGL},2|1}\left(v \mid u; k\right) = \kappa_k\left(u^{-1}\psi^{\leftarrow}(v)\right) = \frac{\Gamma^*\left(k-1, u/\Psi_k^{\leftarrow}(v)\right)}{\Gamma(k-1)},$$

where κ_k is defined in Equation (E.3.3). The result follows since $C_{\mathrm{IGL},2|1}\left(v \mid u; k\right) = 1 - \hat{C}_{\mathrm{IGL},2|1}\left(1-v \mid 1-u; k\right)$.

For the $1|2$ distribution function, use the DJ copula $1|2$ distribution in Equation (E.2.11) to obtain

$$\begin{aligned}\hat{C}_{\mathrm{IGL},1|2}\left(u \mid v; k\right) &= \frac{\Psi_k'\left(u^{-1}\Psi_k^{\leftarrow}(v)\right)}{\Psi_k'\left(\Psi_k^{\leftarrow}(v)\right)} \\ &= \frac{\Gamma(k) - \Gamma^*\left(k, u/\Psi_k^{\leftarrow}(v)\right)}{\Gamma(k) - \Gamma^*\left(k, 1/\Psi_k^{\leftarrow}(v)\right)}.\end{aligned}$$

The result follows since $C_{\mathrm{IGL},1|2}\left(u \mid v; k\right) = 1 - \hat{C}_{\mathrm{IGL},2|1}\left(1-u \mid 1-v; k\right)$.

For the density, use the DJ copula density in Equation (E.1.4), and substitute the form of Ψ_k' and Ψ_k'' found respectively in Equations (E.3.1) and (E.3.2):

$$\begin{aligned}\hat{c}_{\mathrm{IGL}}(u,v;k) &= -\frac{\Psi_k^{\leftarrow}(v)\,\Psi_k''\left(u^{-1}\Psi_k^{\leftarrow}(v)\right)}{u^2\Psi_k'\left(\Psi_k^{\leftarrow}(v)\right)} \\ &= -\frac{\Psi_k^{\leftarrow}(v)}{u^2}\frac{\left[u/\Psi_k^{\leftarrow}(v)\right]^{k+1}\exp\left(-u/\Psi_k^{\leftarrow}(v)\right)}{\Gamma(k)-\Gamma^*\left(k,1/\Psi_k^{\leftarrow}(v)\right)} \\ &= -\frac{u^{k-1}\left[\Psi_k^{\leftarrow}(v)\right]^{-k}\exp\left(-u/\Psi_k^{\leftarrow}(v)\right)}{\Gamma(k)-\Gamma^*\left(k,1/\Psi_k^{\leftarrow}(v)\right)}.\end{aligned}$$

The result follows since $c_{\mathrm{IGL}}(u,v;k) = \hat{c}_{\mathrm{IGL}}(1-u,1-v;k)$. □

E.3.3 Proofs of Properties

Proof of Proposition 6.1.3 (comonotonicity). The strategy is to describe the IGL family using a different subclass of DJ generating functions whose $k \to \infty$ limit is a comonotonicity-generating construction function.

For any $k > 1$, the DJ copula with generating function

$$\Psi_{2k} : y \mapsto \Psi_k\left(\frac{y}{k}\right)$$

for Ψ_k defined in Equation (6.1.7) is the IGL (k) copula, due to the property of invariance under scale change described in Corollary 6.2.3. By Proposition 6.2.6, it only needs to be shown that $\Psi_{2k}(y) \to \min(y,1)$ as $k \to \infty$ for almost all $y > 0$. It is sufficient to show this by restricting k to the integers.

Take $k > 1$. Out generating function for $y > 0$ evaluates to

$$\Psi_{2k}(y) = \frac{k-1}{k}yF_k\left(\frac{k}{y}\right) + 1 - F_{k-1}\left(\frac{k}{y}\right),$$

where F_k is the Gamma $(k,1)$ distribution function. Letting $E_1,\ldots,E_k \sim \mathrm{Exp}(1)$ be an iid sample, F_k is the distribution function of $S_k = \sum_i^k E_i$. As $k \to \infty$, the Central Limit theorem states that

$$\frac{S_k - k}{\sqrt{k}} \to_d \mathcal{N}(0,1),$$

so that

$$F_k\left(\frac{k}{y}\right) = \mathcal{P}\left(S_k \le \frac{k}{y}\right) = \mathcal{P}\left(\frac{S_k-k}{\sqrt{k}} \le \sqrt{k}\left(\frac{1}{y}-1\right)\right) = \Phi\left(\sqrt{k}\left(\frac{1}{y}-1\right)\right) + o(1).$$

As such,

$$\lim_{k\to\infty} F_k\left(\frac{k}{y}\right) = \begin{cases} 0, & y > 1; \\ 1, & y < 1. \end{cases}$$

The generating function, therefore, has the comonotonicity-generating function as a limit:

$$\lim_{k\to\infty} \Psi_{2k}(y) = y\mathbb{I}_{(0,1)}(y) + \mathbb{I}_{(1,\infty)}(y) = \min(y, 1)$$

for almost all $y > 0$. □

Proof of Proposition 6.1.5 (tail dependence). Fix $k > 1$ and $\theta > 0$.

The upper tail dependence $\lambda_U^{(\mathrm{IGL})}$ of the IGL copula family is the lower tail dependence of the reflection copula family:

$$\begin{aligned}
\lambda_U^{(\mathrm{IGL})} &= \lim_{u\downarrow 0} \frac{\hat{C}_{\mathrm{IGL}}(u, u; k)}{u} \\
&= \lim_{u\downarrow 0} \Psi_k\left(\frac{\Psi_k^{\leftarrow}(u)}{u}\right) \\
&= \lim_{t\downarrow 0} \Psi_k\left(\frac{t}{\Psi_k(t)}\right) \\
&= \Psi_k\left(\lim_{t\downarrow 0} \frac{1}{\Psi_k'(t)}\right) \\
&= \Psi_k\left(\frac{1}{k-1}\right),
\end{aligned}$$

by Property 6 of Proposition E.3.1. Using the incomplete gamma function identity

$$\Gamma^*(k, x) = (k-1)\,\Gamma^*(k-1, x) + x^{k-1}\exp(-x)$$

for any $x \geq 0$, the tail dependence can be further evaluated:

$$\begin{aligned}
\Psi_k\left(\frac{1}{k-1}\right) &= \frac{\Gamma(k) - \Gamma^*(k, k-1)}{(k-1)\,\Gamma(k-1)} + \frac{\Gamma^*(k-1, k-1)}{\Gamma(k-1)} \\
&= \frac{\Gamma(k) - \Gamma^*(k, k-1)}{\Gamma(k)} + \frac{(k-1)\,\Gamma^*(k-1, k-1)}{\Gamma(k)} \\
&= \frac{\Gamma(k) - (k-1)^{k-1}\exp\big(-(k-1)\big)}{\Gamma(k)} \\
&= 1 - \frac{\left[(k-1)\,e^{-1}\right]^{k-1}}{\Gamma(k)}.
\end{aligned}$$

The lower tail dependence $\lambda_L^{(\mathrm{IGL})}$ of the IGL copula family is

$$\begin{aligned}
\lambda_L^{(\mathrm{IGL})} &= \lim_{u\downarrow 0} \frac{C_{\mathrm{IGL}}\left(u,u;k\right)}{u} \\
&= \lim_{u\downarrow 0} \frac{1-2\left(1-u\right)+\left(1-u\right)\Psi_k\left(\left[\Psi_k^{\leftarrow}\left(1-u\right)\right]/\left(1-u\right)\right)}{1-\left(1-u\right)} \\
&= \lim_{t\to\infty} \frac{1-2\Psi_k\left(t\right)+\Psi_k\left(t\right)\Psi_k\left(t/\Psi_k\left(t\right)\right)}{1-\Psi_k\left(t\right)} \\
&= 1-\lim_{t\to\infty} \frac{1-\Psi_k\left(t/\Psi_k\left(t\right)\right)}{1-\Psi_k\left(t\right)},
\end{aligned}$$

since $\Psi_k\left(t\right)\to 1$ as $t\to\infty$. To continue, we make use of L'hôpital's rule twice. We will need the form of Ψ_k' found in Equation (E.3.1), and the knowledge that $\frac{\mathrm{d}}{\mathrm{d}t}t/\Psi_k\left(t\right)\to 1$ as $t\to\infty$ because $t/\Psi_k\left(t\right)\sim t$. The desired limit is

$$\begin{aligned}
\lim_{t\to\infty} \frac{1-\Psi_k\left(t/\Psi_k\left(t\right)\right)}{1-\Psi_k\left(t\right)} &= \lim_{t\to\infty} \frac{\Psi_k'\left(t/\Psi_k\left(t\right)\right)}{\Psi_k'\left(t\right)} \\
&= \lim_{t\to\infty} \frac{\Gamma\left(k\right)-\Gamma^*\left(k,\Psi_k\left(t\right)/t\right)}{\Gamma\left(k\right)-\Gamma^*\left(k,1/t\right)} \\
&= \lim_{t\to\infty} \frac{\left[\Psi_k\left(t\right)\right]^{k-1}t^{-(k-1)}\exp\left(-\Psi_k\left(t\right)/t\right)}{t^{-(k-1)}\exp\left(-1/t\right)} \\
&= \exp\left(\lim_{t\to\infty}\frac{1-\Psi_k\left(t\right)}{t}\right) \\
&= 1,
\end{aligned}$$

so that $\lambda_L^{(\mathrm{IGL})}=0$. □

Proof of Proposition 6.1.7 (CCEVI).. Fix $u\in(0,1)$ and $k>1$. Using the conditional distribution found in Equation (E.3.4), let

$$\begin{aligned}
F\left(y\right) &= 1-C_{\mathrm{IGL},2|1}\left(1-\frac{1}{y}\mid u;k\right) \\
&= 1-\frac{\Gamma^*\left(k-1,R_1\left(y\right)\right)}{\Gamma\left(k-1\right)} \\
&= F_{k-1}\left(R_1\left(y\right)\right)
\end{aligned}$$

for $y > 0$, where F_{k-1} is the Gamma $(k-1, 1)$ distribution function, and

$$R_1(y) = \frac{1-u}{\Psi_k(y^{-1})}.$$

Note that R_1 has a unique inverse function since Ψ_k does, and by Property (4) of Proposition (E.3.1), $R_1 \in \mathrm{RV}_1$.

The CCEVI of the IGL copula family is just the EVI of F, which we will find using Theorem 1.2.6 of de Haan and Ferreira (2006). By the theorem, since $F_{k-1} \in \mathcal{D}(G_0)$, we have

$$1 - F_{k-1}(t) = b_1(t) \exp\left\{-\int_{t_0}^{t} \frac{\mathrm{d}s}{b_2(s)}\right\}, \tag{E.3.7}$$

for $t \in (t_0, \infty)$, $t_0 \in (0, \infty)$, where b_1 (continuous) and b_2 are positive real functions satisfying

$$\lim_{t\to\infty} b_1(t) = b_1^* \in (0, \infty), \tag{E.3.8}$$

$$\lim_{t\to\infty} b_2'(t) = 0, \tag{E.3.9}$$

and

$$\lim_{t\to\infty} b_2(t) = 0. \tag{E.3.10}$$

Now investigate the form of $1 - F$, using Equation (E.3.7) and changing the variable of integration to $r = R_1^{\leftarrow}(s)$:

$$\begin{aligned} 1 - F(y) &= b_1\big(R_1(y)\big) \exp\left\{-\int_{t_0}^{R_1(y)} \frac{\mathrm{d}s}{b_2(s)}\right\} \\ &= \tilde{b}_1(y) \exp\left\{-\int_{y_0}^{y} \frac{\mathrm{d}r}{\tilde{b}_2(r)}\right\}, \end{aligned}$$

where $y_0 = R_1^{\leftarrow}(t_0)$,

$$\tilde{b}_1(y) = b_1\big(R_1(y)\big), \tag{E.3.11}$$

and

$$\tilde{b}_2(y) = \frac{b_2\big(R_1(y)\big)}{R_1'(y)} \tag{E.3.12}$$

for $y \in (y_0, \infty)$. The DOA condition of F depends on $\tilde{b}_1$ and $\tilde{b}_2$. First, both functions are positive: since $b_1 > 0$, we have $\tilde{b}_1 > 0$, and since $b_2 > 0$ and $R_1 \in \mathrm{RV}_1$, it follows that $\tilde{b}_2 > 0$ since $R_1' > 0$ in some neighbourhood of infinity (which we can take to be (y_0, ∞), without loss of generality). In addition, $\tilde{b}_2$ is continuous because b_2 is, and because R_1 is

differentiable everywhere on (y_0, ∞). Second, by Equations (E.3.8) and (E.3.11),

$$\lim_{y\to\infty} \tilde{b}_1(y) = \lim_{y\to\infty} b_1(y) = 0.$$

Third, we need the derivative

$$\begin{aligned}\tilde{b}_2'(y) &= \frac{\left[R_1'(y)\right]^2 b_2'\left(R_1(y)\right) - R_1''(y)\, b_2\left(R_1(y)\right)}{\left[R_1'(y)\right]^2} \\ &= b_2'\left(R_1(y)\right) - b_2\left(R_1(y)\right)\frac{R_1''(y)}{\left[R_1'(y)\right]^2},\end{aligned}$$

for $y \in (y_0, \infty)$. By Equations (E.3.9) and (E.3.10), and since $R_1''/\left(R_1'\right)^2 \in \mathrm{RV}_{-1}$ so that

$$\lim_{y\to\infty} \frac{R_1''(y)}{\left[R_1'(y)\right]^2} = 0,$$

we have

$$\lim_{y\to\infty} \tilde{b}_2'(y) = \lim_{y\to\infty} b_2'(y) - \left(\lim_{y\to\infty} b_2(y)\right)\left(\lim_{y\to\infty} \frac{R_1''(y)}{\left[R_1'(y)\right]^2}\right) = 0.$$

By Theorem 1.2.6 of de Haan and Ferreira (2006), $F \in \mathcal{D}(G_0)$, so that the CCEVI of the IGL copula family is 0. □

E.4 IG Copula Class

E.4.1 Generating Function Properties

$$\mathrm{D}_1 H_k(t;\eta) = -\frac{(\log t+1)\left(\Gamma(k)-\Gamma^*(k,\eta\log t)\right)+\eta(\log t)^2\Gamma^*(k-1,\eta\log t)}{\eta(t\log t)^2\Gamma(k-1)}$$

$$\mathrm{D}_2 H_k(t;\eta) = -\frac{1}{t\eta^2\log t}\frac{\Gamma(k)-\Gamma^*(k,\eta\log t)}{\Gamma(k)}$$

$$\mathrm{D}_{12} H_k(t;\eta) = \frac{\left(\Gamma(k)-\Gamma^*(k,\eta\log t)\right)(\log t+1)-t^{-\eta}(\eta\log t)^k}{(t\eta\log t)^2\Gamma(k-1)}$$

Proposition E.4.1. *For $k>1$, the function H_k defined in Equation (6.1.8) has the following properties.*

1. *Initial slope:*

$$\lim_{t\downarrow 1}\mathrm{D}_1 H_k(t;\theta) = \begin{cases} -\infty, & 1<k<2; \\ -\left(1+\frac{\theta}{2}\right), & k=2; \\ -1 & k>2. \end{cases} \tag{E.4.1}$$

2. *For $t>0$,*

$$H_k(t;\theta)+\theta\,\mathrm{D}_2 H_k(t;\theta) = \frac{\Gamma^*(k-1,\theta\log t)}{t\Gamma(k-1)}. \tag{E.4.2}$$

3. *For $\theta>0$, $H_k^{\leftarrow}(\cdot;\theta)\in RV_{-1}$ at 0^+.*

Proof. Property 1: by Equation (E.2.1),

$$\begin{aligned}\lim_{t\downarrow 1}\mathrm{D}_1 H_k(t;\theta) &= -\lim_{t\downarrow 1}\left[\Psi_k\left(\frac{1}{\theta\log t}\right)+\frac{1}{\theta(\log t)^2}\Psi_k'\left(\frac{1}{\theta\log t}\right)\right] \\ &= -\left[1+\theta\lim_{x\to\infty}x^2\Psi_k'(x)\right].\end{aligned}$$

From Proposition E.3.1(3), $x\mapsto x^2\Psi_k'(x)\in \mathrm{RV}_{2-k}$ at infinity, so that the limit is clear

whenever $k \neq 2$. Since $\Psi_2'(x) = 1 - \left(x^{-1}+1\right)\exp\left(-x^{-1}\right)$,

$$\begin{aligned}\lim_{x\to\infty} x^2\Psi_2'(x) &= \lim_{t\downarrow 0}\frac{1-(t+1)\exp(-t)}{t^2}\\ &= \lim_{t\downarrow 0}\frac{-\exp(-t)+\exp(-t)(1+t)}{2t}\\ &= \frac{1}{2}.\end{aligned}$$

Property 2: using the identity in Equation (E.2.6),

$$H_k(t;\theta) + \theta\,\mathrm{D}_2 H_k(t;\theta) = \frac{1}{t}\kappa_k\left(\frac{1}{\theta\log t}\right),$$

where κ_k can be found in Equation (E.3.3).

Property 3: First, find the index of variation of $H_k(\cdot;\theta)$ at infinity: for $s>0$,

$$\lim_{t\to\infty}\frac{H_k(ts;\theta)}{H_k(t;\theta)} = s^{-1}\lim_{t\to\infty}\frac{\Psi_k\left(\left(\theta\log(ts)\right)^{-1}\right)}{\Psi_k\left(\left(\theta\log(t)\right)^{-1}\right)}.$$

By Property 4 of Proposition E.3.1, $\Psi_k \in \mathrm{RV}_{-1}$ at 0^+. Since $(\theta\log t)^{-1} \in \mathrm{RV}_0$ as $t\to\infty$, it follows that

$$\Psi_k\left(\left(\theta\log(t)\right)^{-1}\right) \in \mathrm{RV}_{(-1)(0)} = \mathrm{RV}_0$$

as $t\to\infty$. Finally,

$$\lim_{t\to\infty}\frac{H_k(ts;\theta)}{H_k(t;\theta)} = s^{-1}.$$

□

E.4.2 Formulas

Distributional quantities related to the IG copula family are presented here. For neatness, results are bundled as a proposition.

Proposition E.4.2. *Let $\theta>0$ and $k>1$. The IG copula $(u,v)\mapsto C_{IG}(u,v;k,\theta)$ has*

$$C_{IG,2|1}\left(v \mid u;k,\theta\right) = 1 - \frac{\Gamma^*\left(k-1,\theta(1-u)\log\left(H_k^{\leftarrow}(1-v;\theta)\right)\right)}{H_k^{\leftarrow}(1-v;\theta)\,\Gamma(k-1)}, \tag{E.4.3}$$

$$C_{IG,1|2}\left(v \mid u; k, \theta\right) = 1 - (1-u)\,\frac{\mathrm{D}_1\, H_\psi\left(H_\psi^{\leftarrow}\left(1-v;\theta\right);\theta\left(1-u\right)\right)}{\mathrm{D}_1\, H_\psi\left(H_\psi^{\leftarrow}\left(1-v;\theta\right);\theta\right)}$$

(which is presented this way because the long-hand form is very cumbersome), and

$$c_{IG}\left(u,v;k,\theta\right) = -\frac{\theta^{k-1}\left(1-u\right)^{k-1}\left(\log t_v\right)^{k-2} t_v^{-\theta(1-u)} + \Gamma^*\left(k-1,\theta\left(1-u\right)\log t_v\right)}{\Gamma\left(k-1\right) t_v^2\, \mathrm{D}_1\, H_k\left(t_v;\theta\right)}, \tag{E.4.4}$$

where $t_v = H_k^{\leftarrow}\left(1-v;\theta\right)$.

Proof. We work with the reflection copula, which has a simpler form. Fix $\theta > 0$ and $k > 1$, and consider $(u,v) \in (0,1)^2$.

For the 2|1 distribution function,

$$\begin{aligned}\hat{C}_{\mathrm{IG},2|1}\left(v|u;k,\theta\right) &= \tfrac{\mathrm{d}}{\mathrm{d}u} u H_k\left(H_k^{\leftarrow}\left(v;\theta\right);\theta u\right)\\ &= H_k\left(H_k^{\leftarrow}\left(v;\theta\right);\theta u\right) + \theta u\, \mathrm{D}_2\, H_k\left(H_k^{\leftarrow}\left(v;\theta\right);\theta u\right)\\ &= \frac{\Gamma^*\left(k-1,\theta u \log\left(H_k^{\leftarrow}\left(v;\theta\right)\right)\right)}{H_k^{\leftarrow}\left(v;\theta\right)\Gamma\left(k-1\right)},\end{aligned} \tag{E.4.5}$$

by Property 2 of Proposition E.2.1. The result follows since

$$C_{\mathrm{IG},2|1}\left(v|u;k\right) = 1 - \hat{C}_{\mathrm{IG},2|1}\left(1-v|1-u;k\right).$$

For the 1|2 distribution function,

$$\begin{aligned}\hat{C}_{\mathrm{IG},1|2}\left(u|v;k,\theta\right) &= \tfrac{\mathrm{d}}{\mathrm{d}v} u H_k\left(H_k^{\leftarrow}\left(v;\theta\right);\theta u\right)\\ &= \frac{u\, \mathrm{D}_1\, H_k\left(H_k^{\leftarrow}\left(v;\theta\right);\theta u\right)}{\mathrm{D}_1\, H_k\left(H_k^{\leftarrow}\left(v;\theta\right);\theta\right)}.\end{aligned}$$

For the density function, differentiate Equation (E.4.5) with respect to v, letting $t_v = H_k^{\leftarrow}\left(1-v;\theta\right)$:

$$\begin{aligned}\hat{c}_{\mathrm{IG}}\left(u,v;k,\theta\right) &= \frac{-\left(\theta u \log t_{1-v}\right)^{k-2}\exp\left(-\theta u \log t_{1-v}\right)\theta - \Gamma^*\left(k-1,\theta u \log t_{1-v}\right)}{\Gamma\left(k-1\right) t_{1-v}^2}\,\tfrac{\mathrm{d}t_{1-v}}{\mathrm{d}v}\\ &= -\frac{\theta^{k-1}u^{k-1}\left(\log t_{1-v}\right)^{k-2} t_{1-v}^{-\theta u} + \Gamma^*\left(k-1,\theta u \log t_{1-v}\right)}{\Gamma\left(k-1\right) t_{1-v}^2\, \mathrm{D}_1\, H_k\left(t_{1-v};\theta\right)}.\end{aligned}$$

□

E.4.3 Proofs of Properties

Proof of Proposition 6.1.4. Fix $k > 1$ and $\theta > 0$.

The upper tail dependence $\lambda_U^{(\mathrm{IG})}$ of the IG copula family is the lower tail dependence of the reflection IG copula family. So,

$$\begin{aligned}
\lambda_U^{(\mathrm{IG})} &= \lim_{u \downarrow 0} \frac{\hat{C}_{\mathrm{IG}}(u, u; k, \theta)}{u} \\
&= \lim_{u \downarrow 0} H_k\left(H_k^{\leftarrow}(u; \theta); \theta u\right) \\
&\leq \lim_{u \downarrow 0} \frac{1}{H_k^{\leftarrow}(u; \theta)} \\
&= 0,
\end{aligned}$$

due to the upper bound of H_k (see Property 5 in Proposition E.2.1). Since H_k is non-negative, it follows that $\lambda_U^{(\mathrm{IG})} = 0$.

For the lower tail dependence $\lambda_L^{(\mathrm{IG})}$ of the IG copula family,

$$\begin{aligned}
\lambda_L^{(\mathrm{IG})} &= \lim_{u \downarrow 0} \frac{C_{\mathrm{IG}}(u, u; k, \theta)}{u} \\
&= \lim_{u \uparrow 1} \frac{1 - 2u + uH_k\left(H_k^{\leftarrow}(u; \theta); \theta u\right)}{1 - u} \\
&= 1 - \lim_{u \uparrow 1} \frac{1 - H_k\left(H_k^{\leftarrow}(u; \theta); \theta u\right)}{1 - u} \\
&= 1 - \lim_{u \uparrow 1} \left[\frac{\mathrm{D}_1 H_k\left(H_k^{\leftarrow}(u; \theta); \theta u\right)}{\mathrm{D}_1 H_k\left(H_k^{\leftarrow}(u; \theta); \theta\right)} + \theta\, \mathrm{D}_2 H_k\left(H_k^{\leftarrow}(u; \theta); \theta u\right)\right].
\end{aligned}$$

For the first limit, since $\mathrm{D}_1 H_k(1; \theta)$ is finite and non-zero whenever $k \geq 2$ (Equation (E.4.1)), we obtain

$$\lim_{u \uparrow 1} \frac{\mathrm{D}_1 H_k\left(H_k^{\leftarrow}(u; \theta); \theta u\right)}{\mathrm{D}_1 H_k\left(H_k^{\leftarrow}(u; \theta); \theta\right)} = \frac{\mathrm{D}_1 H_k(1; \theta)}{\mathrm{D}_1 H_k(1; \theta)} = 1.$$

For the second limit, set $t_u = H_k^{\leftarrow}(u; \theta) \uparrow 1$ as $u \uparrow 1$ and use Equation (E.2.2) to obtain

$$\begin{aligned}
\lim_{u \uparrow 1} \mathrm{D}_2 H_k(t_u; \theta u) &= -\frac{1}{\theta} \lim_{u \uparrow 1} \frac{1}{\theta u \log t_u} \Psi_k'\left(\frac{1}{\theta u \log t_u}\right) \\
&= -\frac{1}{\theta} \lim_{x \to \infty} x \Psi_k'(x) \\
&= 0,
\end{aligned}$$

since $x\Psi_k'(x) \in \mathrm{RV}_{-(k-1)}$ by Property 3 of Proposition E.3.1. It follows that $\lambda_L^{(\mathrm{IG})} = 0$ when $k \geq 2$. □

Proof of Proposition 6.1.6 (CCEVI).. Take $k > 1$, $\theta > 0$, and $u \in (0,1)$, and let $\bar{u} = 1 - u$. Define $v(t) = H_k^{\leftarrow}\left(t^{-1};\theta\right)$ for $t > 1$, and note that by Property 3 of Proposition E.2.1, $v \in \mathrm{RV}_1$ at infinity.

To find the CCEVI, investigate the following limit for $s > 0$, using Equation (E.4.3):

$$
\begin{aligned}
&\lim_{t\to\infty} \frac{1 - C_{\mathrm{IG},2|1}\left(1-(ts)^{-1} \mid u;k,\theta\right)}{1 - C_{\mathrm{IG},2|1}\left(1-t^{-1} \mid u;k,\theta\right)} \\
=&\left(\lim_{t\to\infty} \frac{\Gamma^*\left(k-1,\theta\bar{u}\log\left(v(ts)\right)\right)}{\Gamma^*\left(k-1,\theta\bar{u}\log\left(v(t)\right)\right)}\right)\left(\lim_{t\to\infty}\frac{v(ts)}{v(t)}\right)^{-1} \\
=&s^{-1}\lim_{t\to\infty}\frac{\left[\log\left(v(ts)\right)\right]^{k-2}\left[v(ts)\right]^{-1-\theta\bar{u}}v'(ts)\,s}{\left[\log\left(v(t)\right)\right]^{k-2}\left[v(t)\right]^{-1-\theta\bar{u}}v'(t)} \\
=&\left(\lim_{t\to\infty}\frac{\log\left(v(ts)\right)}{\log\left(v(t)\right)}\right)^{k-2}\left(\lim_{t\to\infty}\frac{v(ts)}{v(t)}\right)^{-1-\theta\bar{u}}\left(\lim_{t\to\infty}\frac{v'(ts)}{v'(t)}\right) \\
=&s^{-1-\theta\bar{u}},
\end{aligned}
$$

since $\log\circ v \in \mathrm{RV}_0$ at infinity and $v' \in \mathrm{RV}_0$ at infinity. It follows that the CCEVI is

$$
u \mapsto \frac{1}{1+\theta(1-u)}.
$$

□

Appendix F

Application Supplement

This appendix provides the details of the analysis discussed in Chapter 7. First, deseasonalization of the river discharge variable is discussed in Section F.1. The marginals of the response and predictors are modelled in Sections F.2 and F.3, respectively. Then, the dependence structure amongst the variables is modelled using vine copulas in Section F.4. The final section, Section F.5, discusses the weights chosen to assess the competing forecasters.

F.1 Deseasonalization of Discharge on a Log Scale

This section discusses the details of estimating the sequences $\{\mu_i\}$ and $\{\sigma_i\}$ in Equation (7.1.1) using discharge data Y_t on the fitting set.

To estimate the location parameter μ_i, we use a kernel-weighted sample average of $\log Y$:

$$\hat{\mu}_i = \frac{\sum_{t=1}^{T} K_b\left(\delta\left(t\right) - i\right)\left(\log Y_t\right)}{\sum_{t=1}^{T} K_b\left(\delta\left(t\right) - i\right)},$$

where $K_b : \mathbb{R} \to [0, \infty)$ is a kernel density function symmetric about 0, which we take to be a tri-cubic kernel for some bandwidth $b > 0$ defined as

$$K_b\left(x\right) = \begin{cases} \left(1 - \left|\frac{x}{b}\right|^3\right)^3, & -b \leq x \leq b \\ 0, & \text{otherwise.} \end{cases} \tag{F.1.1}$$

Here, $b-1$ represents the furthest number of days away from i to take in the computation of the weighted average. We choose $b = 4$.

To estimate the scale parameter σ_i, we similarly use a kernel-weighted sample standard

deviation:

$$\hat{\sigma}_i = \sqrt{\frac{\sum_{t=1}^{T} K_b\left(\delta\left(t\right) - i\right)\left(\log Y_t - \hat{\mu}_i\right)^2}{\sum_{t=1}^{T} K_b\left(\delta\left(t\right) - i\right)}},$$

where $K_b : \mathbb{R} \to [0, \infty)$ is again the tri-cubic kernel for which we use $b = 8$.

F.2 Marginal of the Response

This section discusses fitting a marginal distribution to the change in deseasonalized discharge, $\Delta Z_t := Z_t - Z_{t-1}$, with a Generalized Pareto tail.

Let $\mathcal{T}$ be the set of all dates where discharge data are to be used for analysis. Further, denote $\mathcal{T}_{\text{tr}}$, $\mathcal{T}_{\text{val}}$, $\mathcal{T}_{\text{fit}}$, and $\mathcal{T}_{\text{test}}$ as the sets of dates corresponding to the training, validation, fitting, and test sets, respectively, so that $\mathcal{T}_{\text{fit}} = \mathcal{T}_{\text{tr}} \cup \mathcal{T}_{\text{val}}$ and $\mathcal{T} = \mathcal{T}_{\text{fit}} \cup \mathcal{T}_{\text{test}}$.

The GPD-tail model using some subset $\mathcal{T}_* \subset \mathcal{T}$ and parameters $\ell > 0$ (threshold), $\sigma > 0$ (scale), and $\xi \in \mathbb{R}$ (shape, or extreme value index) is given by

$$\hat{F}_{\Delta Z_t}\left(z; \mathcal{T}_*, \ell, \sigma, \xi\right) = \begin{cases} \hat{F}_{\text{Emp}}\left(z; \mathcal{T}_*\right), & z \leq \ell; \\ p_\ell + \left(1 - p_\ell\right) F_{\text{GPD}}\left(\frac{z-\ell}{\sigma}; \xi\right), & z > \ell, \end{cases}$$

where $\hat{F}_{\text{Emp}}\left(\cdot; \mathcal{T}_*\right)$ is the empirical distribution function of the data $\{\Delta Z_t : t \in \mathcal{T}_*\}$, $p_\ell = \hat{F}_{\text{Emp}}\left(\ell; \mathcal{T}_*\right)$, and F_{GPD} is the standardized Generalized Pareto distribution function

$$F_{\text{GPD}}\left(x; \xi\right) = 1 - \left(1 + \xi x\right)^{-1/\xi}$$

for $x \geq 0$ (and $x \leq -1/\xi$ if $\xi < 0$). Our objective is to choose $\mathcal{T}_*$, ℓ, σ, and ξ so that $\hat{F}_{\Delta Z_t}\left(z; \mathcal{T}_*, \ell, \sigma, \xi\right)$ scores well on an out-of-sample data set.

We first fit the parameters ℓ, σ, and ξ using the following steps:

1. Build the set of distributions

$$\left\{\hat{F}_{\Delta Z_t}\left(z; \mathcal{T}_{\text{tr}}, \ell, \hat{\sigma}_{\text{MLE}}, \hat{\xi}_{\text{MLE}}\right) : \ell \in \mathbb{R}\right\}, \tag{F.2.1}$$

   where $\hat{\sigma}_{\text{MLE}}$ and $\hat{\xi}_{\text{MLE}}$ are the maximum likelihood estimators of σ and ξ on the training set.

2. Choose the threshold parameter $\hat{\ell}$ whose distribution in Equation (F.2.1)—as a forecaster—scores optimally on the validation set.

The final model of the marginal distribution is $\hat{F}_{\Delta Z_t}\left(\cdot; \mathcal{T}_{\text{fit}}, \hat{\ell}, \hat{\sigma}_{\text{MLE}}, \hat{\xi}_{\text{MLE}}\right)$.

Note that the entire fitting data should not be used in Step 1, because out-of-sample data are needed for scoring in Step 2. Scoring in Step 2 should not occur on the same data set as is used in Step 1, since the entire empirical distribution would be favoured over any GPD tail due to the empirical distribution's optimal in-sample score.

The final model has a threshold of 0.2 (corresponding to the empirical 0.81-quantile), with scale and shape parameter estimates being 0.22 and 0.09, respectively, so that the discharge distribution is modelled to have a heavier tail than exponential. See Figure F.1 for QQ and PP plots of the model fit. Finally, the distribution estimates are transformed so that they estimate the distribution of Z_t.

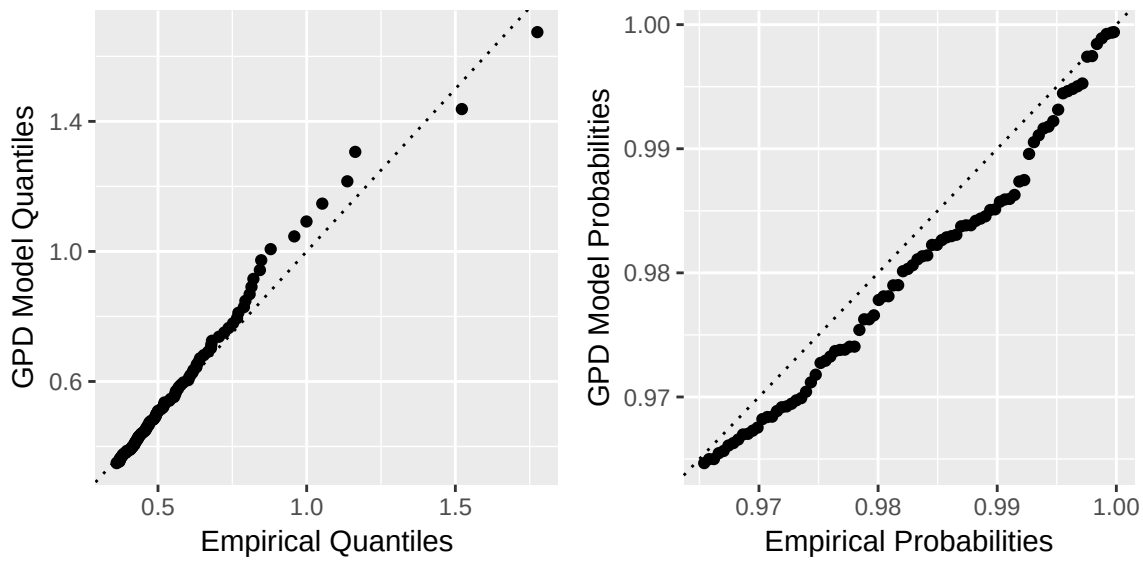

Figure F.1: QQ and PP plots of the marginal distribution estimate of deseasonalized discharge, against the empirical distribution for the fitting data. Only the upper corners of the plots are shown, since both distributions are equal in the lower corners.

F.3 Marginal of Snowmelt

The marginal distribution of snowmelt is estimated in a similar way to the deseasonalization of discharge.

Let X_t be the snowmelt observed on date t. The cdf of X_t is assumed to only depend on the day of year $\delta(t)$, and is modelled with a weighted empirical distribution on the fitting set:

$$\hat{F}_{X_t}(x) = \frac{\sum_{r \in \mathcal{T}_{\text{fit}}} K_b\left(\delta(r) - \delta(t)\right) \mathbb{I}_{[X_r, \infty)}(x)}{\sum_{r \in \mathcal{T}_{\text{fit}}} K_b\left(\delta(r) - \delta(t)\right)},$$

where K_b is the tri-cubic kernel given in Equation (F.1.1). A bandwidth of $b = 4$ is used. See Figure F.2 for diagnostic plots of the resulting model.

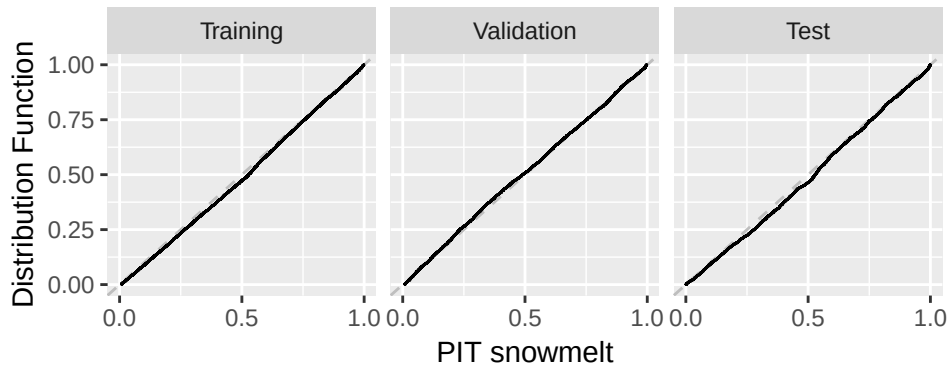

Figure F.2: Empirical distribution functions of the probability integral transformed (PIT) snowmelt, using the estimated cdf of snowmelt. The PIT samples should be from a Uniform(0,1) distribution if the cdf estimate is correct, occurring when the empirical distribution falls on the diagonal.

F.4 Dependence

This section discusses fitting vine copula models to the response and four predictors.

First, a 2-truncated vine is fit to the four predictors after applying a probability integral transform using the marginal models. Since we consider two marginal models for the deseasonalized discharge, we obtain two such vines. The vine array is chosen using sequential minimum spanning trees (cf. Joe, 2014, Section 6.17). Then, the following bivariate copula models are fit using MLE, and are selected using AIC with the function `RVineCopSelect()` from the R package `VineCopula` (Schepsmeier et al., 2016): Gaussian, t, MTCJ, Gumbel, Frank, Joe, BB1, BB7, and BB8. Coding the variables according to

1. one-day-ahead change in deseasonalized discharge (response),
2. change in deseasonalized discharge: lag 0,
3. change in deseasonalized discharge: lag 1,
4. drop in snowpack: lag 0, and
5. drop in snowpack: lag 1,

we obtain the vine array

$$\begin{bmatrix} 5 & 4 & 3 & 2 \\ 0 & 5 & 5 & 3 \\ 0 & 0 & 4 & 5 \end{bmatrix} \tag{F.4.1}$$

in both cases. The resulting copula models corresponding to the edges in the vine array are similar in both vines, and are as follows:

Empirical Marginal		
bb8(6, 0.768)	bb8(2.43, 0.828)	bvtcop(0.612, 3.11)
	mtcjv(0.0847)	frk(1.14)

GPD-tail Marginal		
bb8(6, 0.768)	bb8(2.42, 0.831)	bvtcop(0.613, 2.98)
	mtcjv(0.0876)	frk(1.15)

The copula indicated in row i and column j corresponds to the edge in row $i+1$ and column $j+1$ of the vine array. The abbreviations of the copula models can be found in Table F.1.

Abbreviation	Copula Family
bvncop	Gaussian
bvtcop	t
mtcj	MTCJ
gum	Gumbel
frk	Frank
joe	Joe
bb1	BB1
bb7	BB7
bb8	BB8
indepcop	Independence Copula

Table F.1: Abbreviations of the copula families considered. A 'v' appended after an abbreviation represents the *vertically reflected* copula family, an appended 'u' represents the *horizontally reflected* copula family, and an appended 'r' represents the *reflected* copula family.

Next, pairing orders for the response was chosen so that the resulting PCBN is a vine (for computational simplicity):

- Order one: 2, 3, 5, 4
- Order two: 4, 5, 3, 2

The copula models considered for each edge are the independence copula, Gaussian, t, MTCJ, Gumbel, Frank, Joe, BB1, and BB8, and these are fit with CNQR using the sequential method. Also, the IG and IGL copulas were fit for edges that were not close to independence (to avoid computational difficulty), but these did not lead to an optimal score. If there was evidence that a non-constant CCEVI is present, then one might consider choosing the IG copula, so long as the resulting score is "acceptable". The fitted copulas for each scenario can be seen in Table F.2. These copula families exhibit a range of tail dependence – see Table F.3 for a summary of symmetry and tail dependence of these copula models.

Empirical, Order 1	GPD tail, Order 1	Empirical, Order 2	GPD tail, Order 2
bb1r(0.24, 1.6)	bvncop(0.61)	bvtcop(0.4, 16)	bvtcop(0.38, 14)
bb1v(0.23, 1.2)	bvtcop(-0.38, 9.2)	bvtcop(-0.25, 23)	bvncop(-0.25)
bb1r(0.015, 1)	frk(0.088)	bvncop(0.07)	gumr(1.1)
bvtcop(0.29, 95)	bvncop(0.29)	bvncop(0.54)	frk(4.5)

Table F.2: Copulas linking the response and the predictors using CNQR. The marginal distribution, followed by the linkage order of the predictors, are indicated in the top row. The fitted copulas are displayed corresponding to the linkage order, from top to bottom.

Family	**Symmetry**	**Tail Dependence**	
		Lower	**Upper**
Gaussian	symmetric	intermediate	intermediate
t	symmetric	strong	strong
MTCJ	permutation	yes	tail quadrant independence
Gumbel	permutation	intermediate	strong
Frank	symmetric	tail quadrant independence	tail quadrant independence
Joe	permutation	tail quadrant independence	yes
BB1	permutation	yes	yes
BB7	permutation	yes	yes
BB8	permutation	tail quadrant independence	tail quadrant independence

Table F.3: Properties of some bivariate copula families. All families have members that are at least permutation symmetric (symmetry after swapping the two variables); those that are listed as "symmetric" also have reflection symmetry (a copula C is reflection symmetric if $\mathrm{R}C = C$, where R is the reflection operator defined in Definition 2.4.1). Dependence in either tail of each copula are also indicated, sometimes indicating the strength of tail dependence (a concept related to *tail order* – see Joe, 2014, Section 2.16). These copulas cover a range of tail behaviours in comparison to the Gaussian copula.

Next, it was identified that variables 2, 3 and variables 1, 2 (both pairs are change in ds discharge separated by a lag of 1) have dependence that is not permutation symmetric (as seen in a normal scores plot). To accommodate this, a skewed BB1 copula was fit to these edges (for a GPD marginal). The BB1 copula family is described by parameters

$\theta > 0$ and $\delta > 1$ (details can be found in Joe, 2014, Section 4.17.1), and the upper-skewed version has an additional parameter $\alpha \in (0, 1)$ (larger values imply more skewness). The relationship between the copulas is

$$C_{\text{BB1sk}}(u, v; \theta, \delta, \alpha) = C_{\text{BB1}}\left(u, v^{1-\alpha}; \theta, \delta\right) v^{\alpha}$$

for $(u, v) \in (0, 1)^2$, where C_{BB1sk} and C_{BB1} are skewed BB1 and BB1 copulas, respectively. Because the skewed BB1 family has three parameters, this family was not considered for all edges, for computational simplicity. The skewed BB1 copula linking variables 2, 3 (in the predictor vine) has parameter estimates $\hat{\theta} = 1.21$, $\hat{\delta} = 1.44$, and $\hat{\alpha} = 0.26$. This resulted in an improvement on the predictor vine of approximately 49 AIC. The resulting copulas linking the predictors and the response, fit by CNQR, are shown in Table F.4.

Candidate 5: with skew
bb1sk(0.41, 1.6, 0.17)
bvtcop(-0.42, 7.8)
frk(0.12)
bvncop(0.27)

Table F.4: Copulas linking the response and the predictors using CNQR, for Order 1 with a GPD marginal, using a skewed BB1 copula linking change in ds discharge separated by a lag of 1. The fitted copulas are displayed corresponding to the linkage order, from top to bottom. "bb1sk" is short for "skewed BB1".

Of the five vine CNQR models, we must select an optimal one. We consider the mean scores (Figure F.3), calibration plots (Figure F.4), and calibration histograms (Figure F.5) on the validation set. It appears that all five are generally acceptable. Candidate 5 seems to have the best score, and produces forecasts with a GPD tail (as opposed to forecasts with a right-endpoint, as in Candidates 1 and 3), so it is selected.

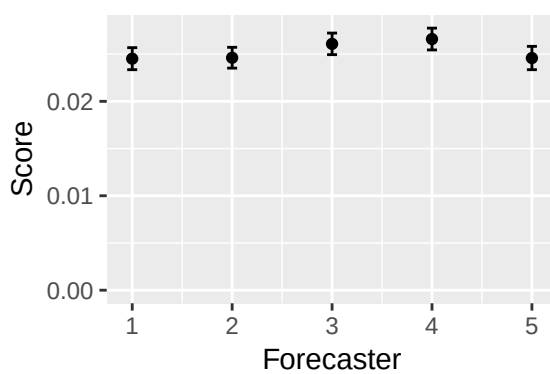

Figure F.3: Mean score estimates of the candidate vine CNQR forecasters, estimated using the training set. Error bars represent standard error of the mean estimate (which ignore autocorrelation).

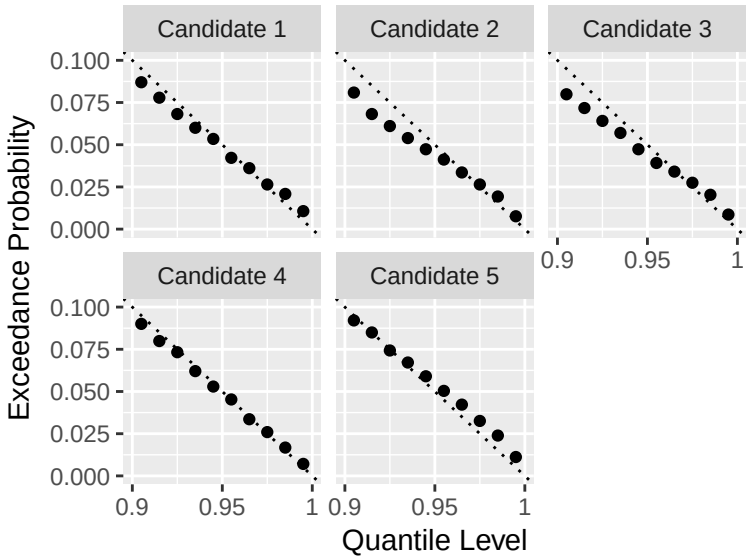

Figure F.4: Calibration plots on the candidate vine CNQR forecasters, using the training set. Each appear to be sufficiently calibrated.

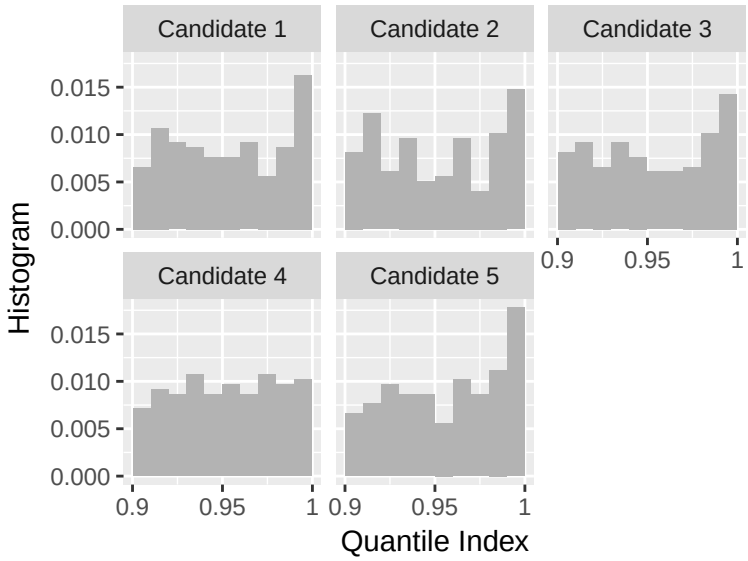

Figure F.5: Calibration histograms on the candidate vine CNQR forecasters, using the training set.

F.5 Weights

This section discusses the choice of weight functions to be used for evaluating forecasters when predictors are large. We seek weight functions over the space of the original predictors (that is, not deseasonalized). Practically, the weight functions can be thought of as the "amount of concern" that one has upon observing the predictors.

First, consider a weight function on one predictor. We would like weights to transition from 0 at $-\infty$ to a weight of 1 at $+\infty$, and therefore choose the logistic curve, defined for location parameter $x_0 \in \mathbb{R}$ and rate parameter $r > 0$ as

$$w_0\left(x; r, x_0\right) = \frac{1}{1 + \exp\left(-r\left(x - x_0\right)\right)} \tag{F.5.1}$$

for $x \in \mathbb{R}$. We will use this function to construct the weight functions for each scenario.

F.5.1 Scenario 1: Large Discharge

For this scenario, we suppose that observing a large discharge on either today or yesterday (lags 0 or 1) is enough for concern (i.e., one would not be "doubly concerned" upon seeing a large discharge two days in a row). An appropriate weight function motivated by inclusion-exclusion is

$$w_1\left(x_2, x_3; r_D, x_{0D}\right) = w_0\left(x_2; r_D, x_{0D}\right) + w_0\left(x_3; r_D, x_{0D}\right) - w_0\left(x_2; r_D, x_{0D}\right) w_0\left(x_3; r_D, x_{0D}\right), \tag{F.5.2}$$

where $(x_2, x_3) \in \mathbb{R}^2$ are observed discharge for both lags (labels are chosen for consistency with variable labels in Appendix F.4). We choose parameters $r_D \in \{0.02, 0.06, 0.1\}$ and $x_{0D} \in \{40, 90, 140\}$. The single-variable logistic function w_0 under these parameters can be seen contrasted against a histogram of discharge in Figure F.6.

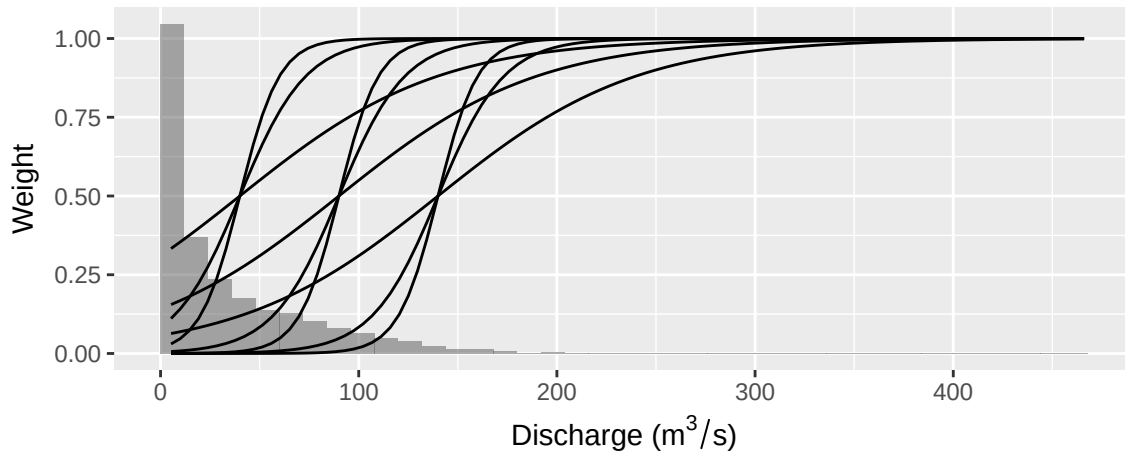

Figure F.6: Single-variable logistic weight functions resulting from the weight parameter choices for discharge. The histogram is that of discharge data on the test set (and is without scale).

F.5.2 Scenario 2: Large Snowmelt

For this scenario, we suppose that observing a large snowmelt on both today and yesterday (lags 0 and 1) is "doubly concerning". An appropriate weight function is

$$w_2\left(x_4, x_5; r_S, x_{0S}\right) = w_0\left(x_4; r_S, x_{0S}\right) + w_0\left(x_5; r_S, x_{0S}\right), \tag{F.5.3}$$

where $(x_4, x_5) \in \mathbb{R}^2$ are observed snowmelt for both lags (labels are chosen for consistency with variable labels in Appendix F.4). We choose parameters $r_S \in \{0.05,\ 0.2,\ 0.7\}$ and $x_{0S} \in \{0,\ 7.5,\ 15\}$. The single-variable logistic function w_0 under these parameters can be seen juxtaposed against a histogram of snowmelt in Figure F.7.

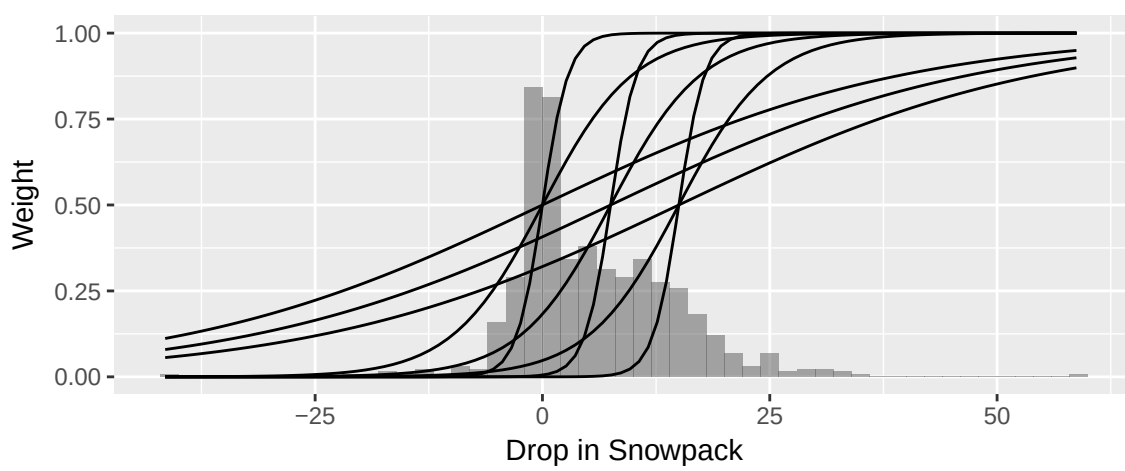

Figure F.7: Single-variable logistic weight functions resulting from the weight parameter choices for snowmelt. The histogram is that of snowmelt data on the test set (and is without scale).

F.5.3 Scenarios 3 and 4: Large Snowmelt and/or Large Discharge

Scenario 3 assigns a high weight if either a large snowmelt or a large discharge is observed. We suppose that observing both is "doubly concerning", and take a weight function of

$$w_3\left(x_2,\ldots,x_5;r_D,r_S,x_{0D},x_{0S}\right)=w_1\left(x_2,x_3;r_D,x_{0D}\right)+w_2\left(x_4,x_5;r_S,x_{0S}\right), \tag{F.5.4}$$

where $(x_2,\ldots,x_5)\in(0,\infty)^2\times\mathbb{R}^2$ are the observed lags 0 and 1 discharge and lags 0 and 1 snowmelt, respectively. The same parameters for r_D, r_S, x_{0D}, and x_{0S} as used in Scenarios 1 and 2 are used here.

Scenario 4 assigns a high weight if both a large discharge or large snowmelt is observed. An appropriate weight function is thus

$$w_4\left(x_2,\ldots,x_5;r_D,r_S,x_{0D},x_{0S}\right)=w_1\left(x_2,x_3;r_D,x_{0D}\right)w_2\left(x_4,x_5;r_S,x_{0S}\right). \tag{F.5.5}$$