

UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Scienze Statistiche  
Corso di Dottorato di Ricerca in Scienze Statistiche  
Ciclo XXXV

# On the modeling of discrete extreme values

**Coordinatore del Corso:** Prof. Nicola Sartori

**Supervisore:** Prof. Carlo Gaetan

**Co-supervisore:** Dr. Philippe Naveau

**Dottorando:** Touqeer Ahmad

02, May, 2023



# Abstract

The statistical modeling of integer-valued extremes has received less attention than its continuous counterparts in the extreme value theory (EVT) literature. In this dissertation, we mainly focus on two problems: one, how to introduce and deal with different kinds of dependence (either its simple or temporal) behavior over the tail when one is working with discrete threshold exceedances, and second, how to model the entire range (i.e., low, moderate and extremes) of discrete extreme data.

Firstly, to describe simple or temporal dependence in discrete exceedances above a threshold. The modeling framework is executed in two steps. In the first step, discrete exceedances are modeled through a discrete generalized Pareto distribution (DGPD), which can be obtained by mixing a Geometric variable with a Gamma distribution. In the second step, a model for discrete extreme values is built by injecting Gamma random variables or latent Gamma process via hierarchical framework, which confirms that the marginal distribution is a DGPD, as expected from classical discrete EVT. In that construction, we obtained a bivariate distribution with DGPD marginals through the Laplace transform of multivariate Gamma distribution with Gamma marginals. In addition, we further developed a bivariate geometric distribution through Farlie-Gumbel-Morgenstern Copula, mixed it into bivariate Gamma distribution, and found a bivariate distribution with DGPD marginals. In this scenario, we have two dependence parameters: one is the copula dependence parameter, and the other is linked with the layer induced through Gamma random variables associated with the hierarchical setting.

Further, we employ four distinct underlying stationary Gamma processes, each producing a different temporal dependency structure, either asymptotic independence or asymptotic dependence. Through the use of pairwise likelihoods, the proposed model

is applied to real discrete time series. Observations of both series over a finite threshold have shown asymptotic independent behavior. One can use a new model for the discrete-time series, which has asymptotic-dependent behavior over the tail. In both scenarios, the proposed model is more flexible.

Secondly, selecting the optimal threshold to define exceedances remains challenging when working with discrete extreme data. Moreover, within a regression framework, the treatment of the many data points (those below the chosen threshold) is either ignored or decoupled from extremes. One possibility is to model the bulk part (observation below the threshold) and tail part (observation above the threshold) by separate models with a mixture setting. Again, optimal threshold is needed, and this framework is computationally burdensome. Based on these considerations, we propose to enforce EVT compliance by using smooth transitions between the two tails (lower and upper). By extending Generalized Additive Models (GAM) to discrete extremes responses, we are able to incorporate covariates. A GAM model quantifies the parameters of the model as functions of covariates. We also develop models with an additional parameter representing the proportion of zero values in the data in the case of zero inflation. The maximum likelihood estimation procedure is implemented for estimation purposes. With the advantage of bypassing the threshold selection step, our findings indicate that the proposed models are more flexible and robust than competing models (i.e., DGPD, Poisson distribution, and negative binomial distribution).



# Sommario

La modellazione statistica dei valori estremi interi ha ricevuto meno attenzione rispetto a quelli continui nella letteratura sulla teoria dei valori estremi (EVT). In questa tesi, ci concentriamo principalmente su due problemi. Il primo si concentra su come introdurre e trattare diversi tipi di dipendenza (semplice o temporale) nella coda della distribuzione quando si tratta il superamento di soglie discrete. In secondo luogo, affrontiamo il problema di modellare l'intera gamma di dati estremi discreti.

Innanzitutto, poniamo l'attenzione sul primo obiettivo, vale a dire quello di descrivere la dipendenza semplice o temporale nel caso di superamenti di una soglia discreta. La modellazione statistica viene eseguita in due fasi. Nella prima, i superamenti sono modellati attraverso una distribuzione di Pareto generalizzata discreta (DGPD), che può essere ottenuta combinando una variabile geometrica con una distribuzione Gamma. Nella seconda fase, si costruisce un modello per i valori estremi discreti introducendo variabili casuali Gamma o processi latenti Gamma attraverso una struttura gerarchica, che conduce ad una distribuzione marginale DGPD, coerentemente alla teoria classica dei valori estremi discreti. In questa costruzione, abbiamo ottenuto una distribuzione bivariata con marginali DGPD attraverso la trasformata di Laplace della distribuzione multivariata Gamma con marginali Gamma. Inoltre, abbiamo sviluppato una distribuzione geometrica bivariata attraverso la copula Farlie-Gumbel-Morgenstern, l'abbiamo combinata alla distribuzione Gamma bivariata e abbiamo ottenuto una distribuzione bivariata con marginali DGPD. In questo scenario, abbiamo due parametri di dipendenza: uno è il parametro di dipendenza della copula, mentre l'altro è legato allo strato indotto dalle variabili casuali Gamma associate all'impostazione gerarchica.

Inoltre, impieghiamo quattro distinti processi Gamma stazionari sottostanti, ognuno dei quali produce una diversa struttura di dipendenza temporale, sia di indipendenza

che dipendenza asintotica. Il modello proposto viene applicato a serie temporali discrete reali utilizzando un approccio di verosimiglianza a coppie. Le osservazioni di entrambe le serie su una soglia finita hanno mostrato un comportamento asintotico indipendente. È possibile utilizzare un nuovo modello per le serie temporali discrete, che presenta un comportamento asintoticamente-dipendente sulla coda. In entrambi gli scenari, il modello proposto è più flessibile.

Tuttavia, la selezione della soglia ottimale per definire i superamenti rimane una sfida quando si lavora con dati estremi discreti. Inoltre, in un quadro di regressione, il trattamento dei molti punti di dati (quelli al di sotto della soglia scelta) viene ignorato o disaccoppiato dagli estremi. Una possibilità è quella di modellare la parte di massa (osservazioni al di sotto della soglia) e la parte di coda (osservazioni al di sopra della soglia) con modelli separati con un modello mistura. Anche in questo caso è necessaria una soglia ottimale e questo schema è computazionalmente oneroso. Sulla base di queste considerazioni, proponiamo di far rispettare l'EVT utilizzando transizioni morbide tra le due code (inferiore e superiore). Estendendo i modelli additivi generalizzati (GAM) a variabili risposta discrete, siamo in grado di incorporare covariate. Un modello GAM quantifica i parametri come funzioni delle covariate. Sviluppiamo anche modelli con un parametro aggiuntivo nel caso di inflazione di zeri. La procedura di massima verosimiglianza è stata implementata ai fini di stimare i parametri del modello proposto. Sfruttando il vantaggio di poter evitare la fase di selezione della soglia, i nostri risultati indicano che i modelli proposti sono più flessibili e robusti rispetto ai modelli concorrenti (ad esempio, DGPD, distribuzione di Poisson e distribuzione binomiale negativa).



*To my beloved family,  
especially my parents*



# Acknowledgements

First and foremost, I thank my supervisor Prof. Carlo Gaetan for his unwavering support, insightful feedback, and constant motivation throughout my research. Their exceptional expertise, patience, and generosity of time and spirit have been invaluable in introducing me to discrete extreme modeling. I would also like to thank my co-supervisor, Dr. Philippe Naveau (Senior Research Scientist at the Laboratoire des Sciences du Climat et de l'Environnement, IPSL-CNRS, France), for suggesting zero-inflated models for the entire domain of discrete extreme modeling. In addition, I am grateful to Prof. Julien Worms for the valuable discussions and hospitality and for providing me with a space in his office at the University of Versailles Saint-Quentin-en-Yvelines (University of Paris-Saclay) during my research stay with Philippe in France. I would also like to thank Dr. Linda Mhalla (Scientist at the Institute of Mathematics, EPFL, Switzerland) and Dr. Thomas Opitz (Research Associate at National Research Institute for Agriculture, Food and Environment, France) for the valuable discussion on the hierarchical modeling of discrete extremes in the context of time series. I would also like to thank Prof. Sebastian Engelke for allowing me to visit and work at the University of Geneva, Switzerland. During my visit to the University of Geneva, I was fortunate to meet many people who made me feel welcome and at home from the first day.

I wanted to take a moment to thank Prof. Manuel Scotto from the University of Lisbon, Portugal, and Prof. Valérie Chavez-Demoulin from HEC Lausanne, University of Lausanne, Switzerland, for your time and effort in evaluating my this dissertation. Your feedback was incredibly helpful and insightful, and I truly appreciate the attention you put into your comments.

I would like to gratefully acknowledge the financial support from Fondazione Cassa di Risparmio di Padova e Rovigo (CARIPARO), which has enabled me to pursue my research goals without financial constraints. I am also thankful to the faculty and non-teaching staff of the Department of Statistical Sciences, University of Padova, for their unconditional support and love.

My sincere thanks go to my colleagues and friends from the Department of Statistical Sciences, who have provided me with a stimulating intellectual environment, helpful feedback, and endless encouragement. In particular, I would like to thank Nicolas Bianco, Pietro Belloni, Christian Castiglione, Marco Girardi, Caizhu Huang, Erika

Banzato, Fariborz Setoudeh, and Giuseppe Alfonzetti for the valuable discussions, support, and friendship.

I also thank the Pakistani's students community living in Padova for their support, love, and respect. I also would like to thank my teachers (especially Dr. Irshad Ahmad Arshad and Dr. Ishfaq Ahmad), colleagues, and friends living in Pakistan for their continuous support.

My deepest gratitude goes to my parents, Fatah Muhammad and Irshad Begum, and siblings, Naveed Akhtar and Farooq Ahmad, and the rest of my family for their unwavering love, encouragement, and support throughout my academic journey. Their love, patience, and understanding have been invaluable to me, and I could not have accomplished this without them.

*Touqeer Ahmad*





# Contents

<b>List of Figures</b>	<b>xiv</b>
<b>List of Tables</b>	<b>xix</b>
<b>Introduction</b>	<b>1</b>
Overview . . . . .	1
Main contributions of the thesis . . . . .	2
<b>1 Modeling of bivariate discrete dependent extremes</b>	<b>5</b>
1.1 Introduction . . . . .	6
1.1.1 Bivariate extremes . . . . .	10
1.2 Discrete hierarchical models . . . . .	13
1.2.1 DGPD and its hierarchical representation . . . . .	13
1.2.2 Bivariate models . . . . .	14
1.2.3 Copula based models . . . . .	15
1.2.4 Dependence induction . . . . .	18
1.3 Tail dependence coefficient . . . . .	20
1.3.1 Gaver model (GM) . . . . .	22
1.3.2 Copula-based Gaver and Lewis model (CGM) . . . . .	23
1.3.3 Kibble model (KM) . . . . .	23
1.3.4 Copula-based Kibble model (CKM) . . . . .	24
1.3.5 Thinned Gamma model (TGM) . . . . .	24
1.3.6 Copula-based Thinned Gamma model (CTGM) . . . . .	25
1.3.7 Complete dependence model (CDM) . . . . .	25
1.3.8 Copula-based complete dependence model (CCDM) . . . . .	26
1.4 Inference . . . . .	26
1.4.1 Censored likelihood . . . . .	26
1.5 Simulation study . . . . .	29
1.5.1 Bivariate and copula-based bivariate models . . . . .	29
1.6 Real Applications . . . . .	30
1.7 Final remarks . . . . .	35
<b>2 A time series model for discrete extreme values</b>	<b>37</b>
2.1 Introduction . . . . .	38
2.2 Discrete time series hierarchical models . . . . .	39

2.2.1	Induction of temporal dependence layer . . . . .	40
2.3	Extremogram for tail behavior . . . . .	42
2.4	Pairwise likelihood . . . . .	44
2.5	Simulation study . . . . .	46
2.6	Real data applications . . . . .	49
2.7	Final remarks . . . . .	54
<b>3</b>	<b>Models for the entire range of count data with extreme observations</b>	<b>55</b>
3.1	Introduction . . . . .	56
3.2	Discrete extremes modeling . . . . .	59
3.2.1	The discrete Gamma spliced threshold DGPD model . . . . .	59
3.2.2	Discrete extended generalized Pareto distribution . . . . .	61
3.2.3	Zero-inflated discrete extended generalized Pareto distribution . . . . .	63
3.3	Generalized additive modelling . . . . .	66
3.4	Bivariate modeling of entire range discrete extremes data . . . . .	68
3.5	Simulation study . . . . .	70
3.5.1	Discrete extended Generalized Pareto distribution . . . . .	70
3.5.2	Zero-inflated discrete extended Generalized Pareto distribution . . . . .	73
3.5.3	GAM form modeling . . . . .	73
3.6	Bivariate discrete extended generalized Pareto distribution . . . . .	76
3.7	Real data applications . . . . .	79
3.7.1	Discrete extended generalized Pareto distribution . . . . .	79
3.7.2	GAM forms applications of DEGPD and ZIDEGPD to avalanches data . . . . .	81
3.8	Bivariate discrete extended generalized Pareto distribution . . . . .	88
3.9	Final remarks . . . . .	89
	<b>Conclusions</b>	<b>91</b>
	<b>Appendix Appendix A</b>	<b>95</b>
A.1	Proofs of propositions . . . . .	95
A.2	Properties of conditional tail dependence . . . . .	98
A.2.1	Gaver model . . . . .	98
A.2.2	Copula-based Gaver model . . . . .	100
A.2.3	Kibble model . . . . .	102
A.2.4	Copula based Kibble model . . . . .	104
A.2.5	Thinned Gamma model . . . . .	106
A.2.6	Copula-based Thinned Gamma model . . . . .	107
A.2.7	Complete dependence model . . . . .	109
A.2.8	Copula-based complete dependence model . . . . .	110
	<b>Appendix Appendix B</b>	<b>113</b>
B.1	Maximum likelihood procedure of DEGPD . . . . .	113
B.2	Maximum likelihood procedure of ZIDEGPD . . . . .	116

**Bibliography**



# List of Figures

1.1	Densities for GEV distribution with $\mu = 0$ , $\sigma = 1$ and varying $\xi \in (-0.5, 0, 0.5)$ . . . . .	7
1.2	Densities of the GPD for different values of the shape parameter $\xi$ and scale parameter is equal to 1. For the Fréchet family $\xi = 0.5$ , and for the Weibull family $\xi = -0.5$ . . . . .	9
1.3	<b>(a)</b> Upper tail values (upper right quadrant above red lines) of $Y_1$ and $Y_2$ generated from GM are locally correlated; <b>(b)</b> Upper tail values (upper right quadrant above red lines) of $Y_1$ and $Y_2$ generated from KM seem to be locally independent . . . . .	22
1.4	Censored likelihood working . . . . .	28
1.5	The 23 SAFRAN massifs of the French Alps. A number between brackets indicates the mentioned snow avalanches paths in each massif (Evin <i>et al.</i> , 2021). . . . .	32
1.6	Number of avalanche events in Haute-Maurienne and Maurienne massifs of the French Alps. . . . .	33
1.7	Empirical and model-based estimate of $\chi(k) = P(Y_1 > k Y_2 > k)$ for the counts of avalanches at Haute-Maurienne and Maurienne massifs of French Alps. The dashed lines give 0.95 bootstrap confidence bands . . . .	34
2.1	The dependence measure $\chi(u)$ and $\bar{\chi}(u)$ for simulated data from each of <b>(a)</b> Gaver and Lewis process; <b>(b)</b> Warren process; <b>(c)</b> Thinned gamma process, and <b>(d)</b> Markov change-point process. Lag $\Delta = 3$ is used here. . . . .	48
2.2	Monthly based number of police reports on narcotics trafficking in Sydney, Australia, along with sample extremogram. . . . .	49
2.3	Number of tick changes by a minute of EUR/GBP sterling exchange rate series with sample extremogram. . . . .	51
2.4	Empirical and fitted models plots of $\chi = P(Y_{t+1} > k^* Y_t > k^*)$ versus $k^*$ at lag 4 of two real-time series <b>(a)</b> monthly number of police reports on narcotics trafficking in Sydney, Australia, and <b>(b)</b> number of tick changes by a minute of EUR/GBP from 09:00 am to 09:00 pm on December 12th, 2019. The dashed lines in both figures give 0.95 bootstrap confidence bands. . . . .	52
2.5	Bootstrap boxplots of the estimated Kendall's tau coefficient centered at observed Kendall's tau coefficient using <b>(a)</b> police reports on narcotics trafficking data, <b>(b)</b> number of tick changes by minutes of the exchange rate of euro to British pound (EUR/GBP) data. . . . .	53

3.1	Probability mass function and the cumulative distribution function of discrete gamma GP spliced at threshold model corresponding to $\theta_B = (\alpha = 5, \beta = 0.25)$ , $u=25$ , $\theta_T = (\sigma = 15, \xi = 0, 0.1, 0.3, 0.5)$ . . . . .	60
3.2	(a) Probability mass function corresponding to model in (3.9) type (i) for $\sigma = 1$ , $\xi = 0.5$ and shape parameter for lower tail $\kappa = 1, 2, 5, 10$ ; (b) Probability mass function corresponding to model (3.9) combined with $G(u; \psi) = 1 - D_\delta(1 - u)^\delta$ for $\sigma = 1$ , $\xi = 0.5$ and $\delta = \infty, 5, 3, 1$ ; (c) Probability mass function corresponding to model (3.9) combined with $G(u; \psi) = [1 - D_\delta(1 - u)^\delta]^{\kappa/2}$ for $\sigma = 1$ , $\xi = 0.5$ , $\delta = 1, 2$ and $\kappa = 1, 2, 5$ , and (d) Probability mass function corresponding to model in (3.9) combined with $G(u; \psi) = pu^{\kappa_1} - (1 - p)u^{\kappa_2}$ for $\sigma = 1$ , $\xi = 0.2$ , $\kappa_1 = 1, 2$ and $\kappa_2 = 1, 2, 5, 10$ . . . . .	62
3.3	(a) Probability mass function corresponding to model in (3.12) with $G(u; \psi) = u^\kappa$ having $\pi = 0.2, \sigma = 1, \xi = 0.2$ and shape parameter for lower tail $\kappa = 1, 5, 10, 20$ ; (b) Probability mass function corresponding to model in (3.12) combined with $G(u; \psi) = 1 - D_\delta(1 - u)^\delta$ for $\pi = 0.2, \sigma = 1, \xi = 0.2$ and $\delta = 1, 3, 5, \infty$ ; (c) Probability mass function corresponding to model in (3.12) combined with $G(u; \psi) = [1 - D_\delta(1 - u)^\delta]^{\kappa/2}$ for $\pi = 0.2, \sigma = 1, \xi = 0.2$ , $\delta = 1, 2, 5$ and $\kappa = 1, 5, 10, 20$ , and (d) Probability mass function corresponding to model (3.12) combined with $G(u; \psi) = pu^{\kappa_1} - (1 - p)u^{\kappa_2}$ for $\pi = 0.2, p = 0.5, \sigma = 1, \xi = 0.2$ , $\kappa_1 = 1, 2$ and $\kappa_2 = 1, 2, 5, 10$ . The orange lines represent the zero-inflated Poisson probability mass function with different settings of their parameter. . . . .	65
3.4	Boxplots of maximum likelihood estimates of parameters of each model from $n = 1000$ with $10^4$ replication at different parameters settings: <b>(a)</b> corresponds to model type (i), <b>(b)</b> corresponds to model type (ii), <b>(c)</b> corresponds to model type (iii), and <b>(d)</b> corresponds to model type (iv). . . . .	71
3.5	Boxplots of maximum likelihood estimates of parameters of each ZIDEGPD model from $n = 1000$ with $10^4$ replication at different parameters settings: <b>(a)</b> corresponds to model type (i), <b>(b)</b> corresponds to model type (ii), <b>(c)</b> corresponds to model type (iii), and <b>(d)</b> corresponds to model type (iv). . . . .	74
3.6	<i>Left:</i> Fitted non-parametric GAM form models with true parameters generated through orthogonal polynomial of order 2, <i>Right:</i> Fitted non-parametric GAM form models with true parameters generated from the orthogonal polynomial of order 5. The black solid line in each curve represents the varying true parameters, and the red line represents the estimated one with 95% confidence intervals. The boxplots centered at true values represent the estimates of the fixed parameters in each model. . . . .	75
3.7	Frequency distribution of upheld complaints of automobile insurance companies in New York City (2009-2020). . . . .	79
3.8	Quantile-quantile plots of the fitted models with 95% bootstrap-based confidence intervals . . . . .	81
3.9	Correlation among covariates for avalanches data. . . . .	82
3.10	Estimated non-parametric effects of covariates in the $\sigma$ component of DEGPD model type (i). . . . .	84

---

3.11	Diagnostic plots of residual quantiles of the proposed DEGPD type (i) to type (iv) and competitor (DGPD and Poisson) distributions. . . . .	88
3.12	Diagnostic plots of residual quantiles for the proposed ZIDEGPD type (i) to ZIDEGPD type (iii) models. . . . .	89









# List of Tables

1.1	Bias and RMSE for parameter $\alpha$ and $\rho$ acquired from simulations of the models based on CDM, GM, KM and TGM, with $n= 20,000$ , and threshold $u$ equal to the 0.90 quantile and under different parameter configurations. . . . .	31
1.2	Bias and RMSE for parameter $\alpha, \phi$ and $\rho$ acquired from simulations of the models based on CCDM, CGM, CKM, and CTGM, with $n= 20,000$ , and threshold $u$ equal to the 0.90 quantile and under different parameter configurations. . . . .	31
1.3	Estimates of model parameters with their standard errors (in parenthesis) of CDM GM, KM, TGM, CCDM, CGM, CKM, and CTGM for Avalanches data of french Alps. . . . .	33
2.1	Root mean squared error and bias for $\alpha$ and $\rho$ found from simulations of the models based on GLP, WP, TGP, and MCP, with $n=30,000, (u = 0.90)$ quantile and under different parameter settings. . . . .	47
2.2	Estimates of the model parameter with their (standard errors) under GLP, WP, TGP, and MCP for Police reports on narcotics trafficking using PL with $\Delta = 3$ . . . . .	50
2.3	Estimates of the model parameter under GLP, WP, TGP and MCP for a number of tick changes by a minute of EUR/GBP sterling exchange rate series ( $u = 29$ ) at $\Delta = 4$ . . . . .	51
3.1	Root mean square errors of parameter estimates of DEGPD found from $10^4$ independent data sets of size $n = 1000$ . . . . .	72
3.2	Root mean square errors of the parameter estimates of BDEGPD found from $10^4$ independent data sets of size $n = 500$ . The cells contain RMSEs of parameters $\kappa/\sigma/\xi$ for BDEGPD associated with family $G(u; \psi) = u^\kappa$ , $\delta/\sigma/\xi$ for BDEGPD associated with family $G(u; \psi) = 1 - D_\delta\{(1 - u)^\delta\}$ , $\delta/\kappa/\sigma/\xi$ for BDEGPD associated with $G(u; \psi) = [1 - D_\delta\{(1 - u)^\delta\}]^{\kappa/2}$ , and $p/\kappa_1/\kappa_2/\sigma/\xi$ for BDEGPD linked with $G(u; \psi) = pu^{\kappa_1} + (1 - p)u^{\kappa_2}$ , respectively. . . . .	78
3.3	Estimated parameters for extended versions of discrete Pareto distribution with all four parametric families fitted to insurance complaints data of New York City. Standard errors are reported between parenthesis. The bootstrap confidence intervals at level 95% are reported between square brackets. . . . .	80

3.4	AIC and BIC associated with the fitted DEGPD and ZIDEGPD along with Chi-square goodness-of-fit test. P-values are reported between parenthesis. . . . .	82
3.5	Detailed information of covariates . . . . .	83
3.6	Estimated coefficients and smooth terms for GAM form DEGPD models fitted to avalanches data. . . . .	86
3.6	<b>(Cont...)</b> Estimated coefficients and smooth terms for GAM form DEGPD models fitted to avalanches data. . . . .	87
3.7	Kolmogorov–Smirnov (KS) test statistics (p-values between parentheses) of the proposed DEGPD type (i) to type (iv) and competitor (DGPD and Poisson) distributions. . . . .	87
3.8	Estimated parameters for bivariate extended discrete generalized Pareto distribution with all four parametric families fitted to Haute-Maurienne and Maurienne massifs of French Alps. Standard errors are reported between parenthesis. The bootstrap confidence intervals at level 95% are reported between square brackets. . . . .	90
S1	Root mean square errors of parameter estimates ZIDEGPD found from $10^4$ independent data sets of size $n = 1000$ . . . . .	118
S2	Estimated coefficients and smooth terms for GAM form ZIDEGPD models fitted to Avalanches data of Haute-Maurienne massif of French Alps .	119
S2	<b>(Cont...)</b> Estimated coefficients and smooth terms for GAM form ZIDEGPD models fitted to Avalanches data of Haute-Maurienne massif of French Alps . . . . .	120





# Introduction

## Overview

In many fields, such as climatology, hydrology, ecology, and oceanography, dependencies in continuous extreme-value data have been a topic of growing importance for the last few decades. EVT can be used to describe atypical situations that can significantly impact various applications, where data about how the tail of an actual distribution behaves is becoming increasingly important. Under suitable conditions, a generalized Pareto distribution (GPD) can approximate the distribution of exceedances over a high threshold, which is considered a common approach to modeling continuous extreme data. A major concern is how to take into account the simple or temporal dependence of exceedances, which can either be asymptotically dependent when observation of process or variables over the threshold lies in clusters or asymptotically independent when data points over the threshold are independent and identically distributed. In a continuous extreme perspective, literature has suggested different procedures see, for instance, (Smith *et al.*, 1997; Bortot and Tawn, 1998; Bortot and Gaetan, 2014; Noven *et al.*, 2018; Bacro *et al.*, 2020, and references therein).

One approach to moving from continuous to discrete extremes is to model threshold exceedances of integer random variables by the discrete version of the generalized Pareto distribution. Therefore in this manuscript, we develop a hierarchical framework to model discrete threshold exceedances and to observe different kinds of extremal behaviors. The hierarchical construction ensures that the marginal distribution of discrete exceedances converges to the discrete GPD (DGPD). The marginal DGPD is obtained by mixing Geometric and Gamma random variables. The valuable feature of this hierarchical mixture is the bivariate distribution with discrete GPD marginals is tractable via the Laplace transform (LT) of bivariate Gamma distribution. Further, to incorporate different kinds of extremal dependence, we focus on Markov chains with hierarchical

construction and have a Gamma marginal. Indeed, different choices of the underlying Gamma process can lead to different degrees of temporal dependence of discrete extremes, including asymptotic dependence or asymptotic independence.

On the other hand, the optimal threshold selection that defines exceedances remains a problematic issue. This thesis also extends the idea of using a smooth transition between the two tails (lower and upper) to force large and small discrete extreme values to comply with EVT. The regression-based framework called the Generalized additive modeling (GAM) framework is developed to model discrete extreme responses when one deals with the entire range of discrete data. In the case of zero inflation, we extended the GAM framework with an additional parameter representing the zero inflation problem. Due to the ability to bypass the threshold selection step, the proposed models appear more flexible and robust than their rivals.

## Main contributions of the thesis

### Modeling of bivariate discrete dependent extremes

Chapter 1 of the thesis provides a new modeling framework for bivariate discrete extremes in perspective to observe different kinds of extremal dependence over the tail. According to Hitz *et al.* (2017), the DGPD is considered an appropriate distribution for modeling discrete exceedances. By keeping this in mind, we developed a bivariate distribution with discrete DGPD marginal by mixing independent Geometric random variables with bivariate Gamma random variables. Bivariate LTs of multivariate Gamma distribution with Gamma marginal supported us in developing the probability mass function of the bivariate distribution.

A bivariate model constructed by hierarchical mixture representation allows dependence induction through Gamma random variables or both Gamma and Geometric random variables. We use hierarchical models with Gamma marginals to incorporate dependence through Gamma random variables. To introduce dependence through Geometric random variables, we first develop a bivariate Geometric distribution using Farlie-Gumbel-Morgenstern Copula (FGMC) (Morgenstern, 1956). Again, we get a bivariate distribution with DGPD margins by mixing bivariate Geometric and bivariate



Gamma distributions. This new bivariate distribution will have two dependence parameters: one is related to the Geometric variables dependence layer, and the other is related to the Gamma dependence layer.

To observe the extremal behavior of different hierarchical models when working with proposed models, chapter 1 also developed the tail dependence measure by following the idea of Coles *et al.* (1999). More formally, hierarchical models with Gamma marginal used here have the same correlation  $\rho$ . Furthermore, we observe through a tail dependence measure that the hierarchical model shows different extremal behavior when working with bivariate discrete exceedances.

We show through simulations how different choices of hierarchical models come with different extremal behavior. Later on, we applied the proposed models to the avalanche data of two different massifs of the French Alps.

## A time series model for discrete extreme values

Chapter 2 presents a new modeling paradigm for describing temporal dependence in discrete exceedances above a threshold. The modeling framework is executed in two steps. In the first step, discrete exceedances are modeled through DGPD, which can be obtained by mixing a Geometric variable with a Gamma distribution (see chapter 1). In the second step, a model for discrete extreme values is built by injecting a latent Gamma process via hierarchical framework, which confirms that the marginal distribution is DGPD, as expected from classical discrete extreme value theory. To introduce temporal dependence (either asymptotic independence or asymptotic dependence) in the model, we use four stationary Gamma processes, each producing a different temporal dependence. The proposed model is tested through a simulation study and applied to time series data of a monthly number of police reports on narcotics trafficking in Sydney, Australia, and the number of tick changes by minutes of the exchange rate of euro to British pound (EUR/GBP) on December 12th, 2019 by using a pairwise likelihood approach. Observations of both series over a finite threshold have shown asymptotic independent behavior. One can use a new model for the discrete-time series, which has asymptotic-dependent behavior over the tail. In both scenarios, our proposed model is more flexible.

## Models for the entire range of count data with extreme observations

In chapter 1 and chapter 2 of the thesis, we model threshold exceedances of integer-valued random variables by DGPD. Still, the optimal threshold selection that defines exceedances remains a problematic issue. Moreover, within a regression framework, the treatment of the many data points (those below the chosen threshold) is either ignored or decoupled from extremes. Considering these issues, Chapter 3 extends the idea of using a smooth transition between the two tails (lower and upper) to force large and small discrete extreme values to comply with EVT. For lower tail representation, we use the cumulative distribution function of different parametric distributions (say,  $G$ ) as discussed in Naveau *et al.* (2016). In order to implement the modeling framework, a discrete extended GPD distribution (DEGPD) is developed. In addition, discrete nature extreme events may contain many zero values. For instance, the number of insurance complaints or the number of avalanches may include many zero values. This type of data is generally referred to as zero-inflated (ZI), requiring specialized statistical methods for analysis. Therefore, we have introduced a zero-inflated version of discrete extended generalized Pareto distribution (ZIDEGPD). ZIDEGPD models have an additional parameter in the case of zero inflation.

To incorporate covariates, we extend the GAM framework to discrete extreme responses. In the GAM forms, the parameters of DEGPD and ZIDGPD models are quantified as a function of covariates. We applied our proposed model to avalanches data with environmental covariates. The maximum likelihood estimation procedure is implemented for estimation purposes. With the advantage of bypassing the threshold selection step, our findings indicate that the proposed models are more flexible and robust than competing models (i.e., discrete generalized Pareto distribution and Poisson distribution). Models proposed in chapter 3 can further apply to environmental variables count data by using their temporal or spatio-temporal characteristics.

In addition, chapter 3 proposes bivariate versions of DEGPD (BDEGPD) corresponding to different  $G$  families. To develop the probability mass function of BDEGPD, we use the LT of bivariate Gamma distribution with  $Gamma(1, 1)$  marginals. Again, the hierarchical model with Gamma marginals introduced in chapter 1 can use to induce dependence.

# Chapter 1

## Modeling of bivariate discrete dependent extremes

### Overview: Chapter 1

↪ The main aim of chapter 1 is to develop a modeling framework to observe a dependence in bivariate discrete extremes. The marginal DGPD is obtained through hierarchical construction by mixing the Geometric random variable with the Gamma variable. By keeping hierarchical construction in mind, we propose a bivariate discrete distribution with DGPD marginals by mixing independent Geometric random variables with Gamma random variables. This mixture ensures that the marginal distribution is DGPD as well. In order to induce different kinds of dependence in bivariate exceedances, hierarchical models with Gamma marginals are used. Furthermore, this chapter also focuses on constructing bivariate Geometric distribution through Farlie-Gumbel-Morgenstern Copula (FGMC) (Barbiero, 2019). It uses to derive the bivariate distribution by considering Geometric Gamma mixture distributions. Bivariate discrete distribution with DGPD margins is also obtained in copula settings by mixing the bivariate dependent Geometric random variables with Gamma random variables. Further, we discuss the tail dependence measure for the different hierarchical models with their extremal properties to observe tail dependence.

Section 1.1 briefly introduces bivariate extreme and discusses some relevant existing literature. Section 1.2 provides the detailed construction of discrete hierarchical models with both simple and copula-based settings. Section 1.3 discusses the basic concept of conditional tail dependence and proves their asymptotic properties. Inference information is provided in section 1.4. Simulation study results

are given in section 1.5. Section 1.6 of the chapter discusses the real applications. Final remarks are pinned in section 1.7

## 1.1 Introduction

Classical EVT is crucial in developing stochastic models for solving real-life problems associated with rare events. Let  $y_j, j = 1 \dots n$  be independent and identically distributed (i.i.d) realization of a random variable  $Y$ . Practitioners are interested in estimating the probability  $P(Y > x)$  for a large  $x \in \mathbb{R}$ . The probability can be obtained through the empirical estimate  $\hat{p} = |\{j : y_j > x\}|/n$ . In the case of large  $x$ , few points fall above  $x$  resulting in a noisy estimator. In addition, if  $x$  exceeds the maximal observation, then  $\hat{p} = 0$  which can underestimate important risks if  $P(Y > x) > 0$ . This demonstrates why extreme quantile estimation should sometimes not be performed naively but needs special techniques. Moreover, EVT motivates a parametric family distribution that can approximate the tails of a broad class of distributions.

Let  $Y$  be a continuous or discrete random variable taking values in  $[0, y_F)$  with cumulative distribution function (CDF)  $F$  where  $y_F \in \mathbb{R}$ . One of the most famous results in probability theory is the central limit theorem. It states that for any i.i.d. copies  $Y_j, j = 1 \dots n$  of a random variable  $Y$  whose variance is finite, there exist sequences  $c_n$  and  $d_n$  such that

$$d_n^{-1} \left( \sum_{j=1}^n Y_j - c_n \right) \xrightarrow{d} X \quad (1.1)$$

as  $n \rightarrow \infty$ , where  $X$  follows to a normal distribution; one can choose  $c_n = E(Y)$  and  $d_n = \sqrt{n}$ , see for instance, (Billingsley, 1995). Consider the maximum  $M_n = \max_{j=1, \dots, n}(Y_j)$  instead of sum. The counterpart of the central limit theorem, known as the Fisher—Tippett—Gnedenko theorem, describe that if there exist sequences  $c_n$  and  $d_n$  such that

$$d_n^{-1} (M_n - c_n) \xrightarrow{d} Z \quad (1.2)$$

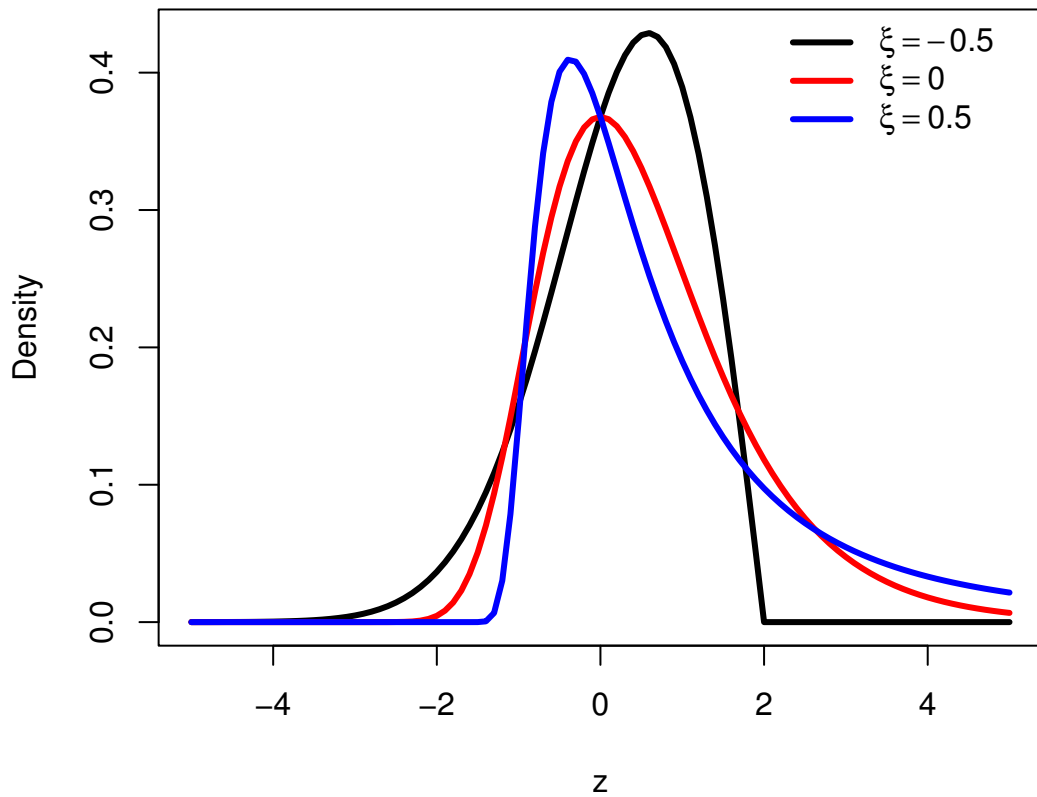


FIGURE 1.1: Densities for GEV distribution with  $\mu = 0$ ,  $\sigma = 1$  and varying  $\xi \in (-0.5, 0, 0.5)$ .

and  $Z$  is non-degenerate, then  $Z$  follows a generalized extreme value (GEV) distribution and defined by its CDF

$$G(z|\mu, \sigma, \xi) = \begin{cases} \exp \left[ - \left\{ 1 + \xi \left( \frac{z-\mu}{\sigma} \right) \right\}^{\frac{-1}{\xi}} \right] & \xi \neq 0 \\ \exp \left[ - \exp \left\{ - \left( \frac{z-\mu}{\sigma} \right) \right\} \right] & \xi = 0 \end{cases} \quad (1.3)$$

If (1.2) holds, one says that  $Y$  or its distribution belongs to the maximum domain of attraction (MDA) of an extreme value distribution that is  $Y \in \text{MDA}_\xi$ . Many common continuous distributions belong to the maximum domain of attraction of extreme value distributions, including Normal, Log-normal, Student, Exponential, Weibull, Fréchet, Beta, Gamma, and Uniform. Figure 1.1 explains how the shape parameter of the GEV distribution affects the tail behavior by keeping the remaining parameters as constant. When  $\xi = 0$ , the GEV distribution leads to the Gumbel family with a distribution function given in (1.3). When  $\xi < 0$ , the GEV distribution tends to a negative Weibull distribution with a finite upper endpoint. Last, when  $\xi > 0$  belongs to the Fréchet distribution with a heavier tail. Combining the original three extreme value distributions

into one simplifies statistical implementation. For more details about GEV distribution, see, for instance, Coles (2001).

Interestingly, when  $Y \in MDA_\xi$ , the behavior of  $Y|Y \geq u$  as  $u \rightarrow y_F$  is characterized as follows:  $Y \in MDA_\xi$  is equivalent to saying that there exists a strictly positive sequence  $\sigma_u$  such that

$$\sigma_u^{-1}(Y - u) | Y \geq u \xrightarrow{d} Z, \quad (1.4)$$

as  $u \rightarrow y_F$ , for some  $Z$  following a non-degenerate probability distribution on  $[0, \infty)$ . Under (1.4),  $Z$  is approximated to the generalized Pareto distribution (GPD) (Pickands, 1975), defined by its CDF

$$F(y; \sigma, \xi) = \begin{cases} 1 - (1 + \xi y/\sigma)_+^{-1/\xi} & \xi \neq 0 \\ 1 - \exp(-y/\sigma) & \xi = 0 \end{cases} \quad (1.5)$$

where  $(a)_+ = \max(a, 0)$ ,  $\sigma > 0$  and  $-\infty < \xi < +\infty$  represent the scale and shape parameters of the distribution, respectively.

The shape parameter  $\xi$  defines the tail behavior of the GPD. If  $\xi < 0$ , the upper tail is bounded. If  $\xi = 0$ , this tends to be the case of an exponential distribution, where all moments are finite. If  $\xi > 0$ , the upper tail is unbounded, but higher moments ultimately become infinite. Three defined cases are labeled “short-tailed”, “light-tailed”, and “heavy-tailed”, respectively. These types of tail behavior make GPD more flexible to model excesses.

Figure 1.2 explains how GPD behaves when the shape parameter varies with different values. This inspires the GPD technique, approximating the distribution of  $Y$  exceedances above a significant threshold  $u$  (Davison and Smith, 1990).

In the discrete case, approximating the tail of a discrete distribution using the GPD poses different challenges. First, an essential condition is that a discrete distribution  $F$  belongs to a maximum domain of attraction when the distribution is long-tailed, that is  $\bar{F}(u+1)/\bar{F}(u) \rightarrow 1$  (Anderson, 1970, 1980; Shimura, 2012; Hitz *et al.*, 2017). Many well-known distributions, including Geometric, Poisson, and Negative binomial distributions, are applied to model discrete data. But, these distributions are not heavy-tailed. Second, ties are not permitted, which presents a problem when attempting to approximate the tail of a discrete distribution by a GPD.

To overcome these problems, Hitz *et al.* (2017) prove that random variables  $X \in D\text{-}MDA_\xi$  and  $Y \in MDA_\xi$  for  $\xi > 0$  satisfy  $X \stackrel{d}{=} \lfloor Y \rfloor$ . Further, they develop two approaches to model the tails of discrete observations, each relying on a specific assumption of the underlying distribution. The first one is to approximate  $P(Y - u = k | Y \geq u)$  for large

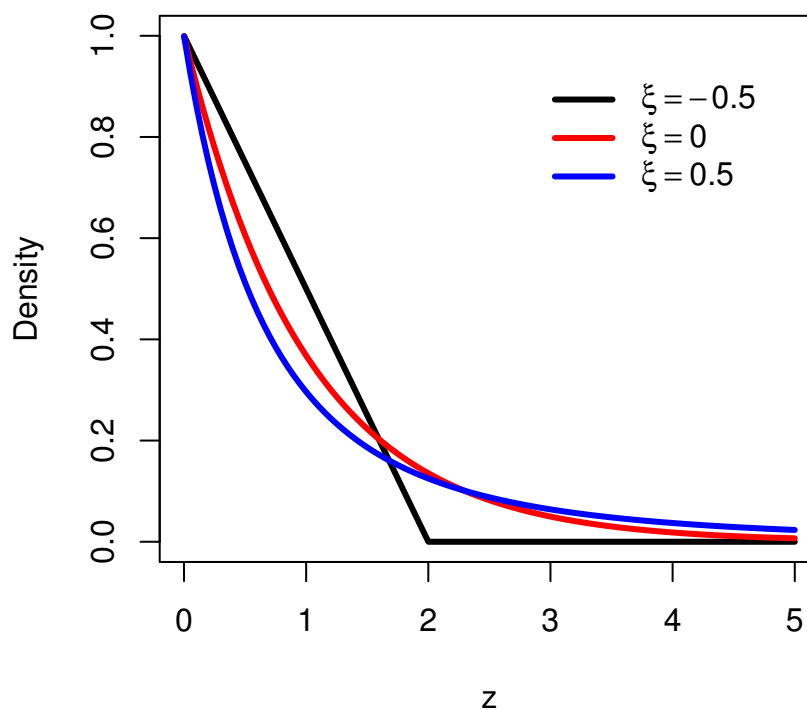


FIGURE 1.2: Densities of the GPD for different values of the shape parameter  $\xi$  and scale parameter is equal to 1. For the Fréchet family  $\xi = 0.5$ , and for the Weibull family  $\xi = -0.5$ .

integer threshold  $u$  by

$$P(Y = k) = F(k + 1; \sigma, \xi) - F(k; \sigma, \xi), \quad k \in \mathbb{N}_0 \quad (1.6)$$

where  $F(\cdot; \sigma, \xi)$  is the CDF of GPD. The distribution defined in (1.6) will be called DGPD. In the existing literature, the DGPD has been used by Prieto *et al.* (2014) to model road accidents data, and Ranjbar *et al.* (2022) applied in regression context to model extremes of seasonal viruses and hospital congestion in Switzerland. Except for extreme value applications, numerous features of discrete Pareto-type distributions were studied in Krishna and Pundir (2009); Buddana and Kozubowski (2014); Kozubowski *et al.* (2015). The second one approximate  $P(Y - u = k | Y \geq u)$  by

$$P(Y = k) = \frac{f(k; \sigma, \xi)}{\sum_j^\infty f(j; \sigma, \xi)}, \quad k \in \mathbb{N}_0 \quad (1.7)$$

The distribution defined in (1.7) is called generalized Zipf distribution (GZD). Zipf law distribution is sometimes presented as the counterpart of the Pareto distribution. In the case  $\xi > 0$ , the GZD tends to a Zipf–Mandelbrot distribution. When  $\xi < 0$ , the GZD has a finite endpoint. Finally, when  $\xi = 0$ , the GZD (as well as the DGPD) is simply

a Geometric distribution. For more details, we refer (Hitz *et al.*, 2017, and references therein).

In the current scenario, we are interested in modeling discrete exceedances by ensuring that the marginal distribution of exceedances is approximated to DGPD. We develop the DGPD marginal distribution through hierarchical construction by representing the DGPD as a mixture of Geometric and Gamma random variables. Like continuous GPD, DGPD retains the important stability property with respect to the exceedances (Budana and Kozubowski, 2014). In addition, the conditional distribution of threshold excesses by DGPD random variable is DGPD distributed too. The stability property of excesses may be damaged in the case of light-tail Geometric distribution (which is a special case of DGPD with  $\xi = 0$ ) (Anderson, 1970; Leadbetter *et al.*, 1983). Further, some conditions required to meet when modeling block maxima of discrete random variables are highlighted in Dkengne *et al.* (2016).

### 1.1.1 Bivariate extremes

A further focus is placed on bivariate discrete extreme distributions and we observe that the joint tail of the distribution exhibits different dependent behaviors. The multivariate extremes, especially bivariate extremes, are relatively new and attention-gaining subject in EVT. Extending extreme value statistics from the univariate to the bivariate cases is challenging. In the univariate context, the theory is based on manipulating a few concepts that were useful in analyzing extreme values. However, in the bivariate context, these notions lose their evident definition (Dutfoy *et al.*, 2014). We have to reconsider them in order to set an adapted signification.

Several settings, for instance, Gumbel's logistic and mixed models, see Gumbel (1960); Gumbel (1961); Gumbel and Mustafi (1967), seem to be quite well established. Nevertheless, these models have only asymptotic justifications for the univariate case. Asymptotic extreme value distributions were later derived based on Gumbel's logistic and mixed models. The general result is presented in Pickands (1981); Marshall and Olkin (1988).

Further, many studies have been conducted in continuous extremes to model and estimate the function that expresses the dependence structure between extreme points. Substantial efforts are made to develop dependence measures by, for instance, De Haan and Resnick (1977); Coles and Tawn (1991, 1994); Coles *et al.* (1999); Schlather and Tawn (2003), and Eastoe *et al.* (2013). Despite this, we cannot determine the joint distribution of the bivariate extremes from a given set of bivariate data using precise



estimation tools. Of course, the estimates of marginal extreme values can be derived from marginal data sets as in Zachary *et al.* (1998); De Haan and De Ronde (1998), but the joint distribution is still an arguable task.

Adopting a copula model to represent the joint distribution (Nelsen, 2007) is common when dealing with bivariate extremes. The marginal distribution is derived as a combination of the copula and asymptotic extreme value distributions, typically of the GEV type (Coles, 2001). For this specific purpose, several copula models have been developed; see, for instance, Tawn (1988); De Waal and Van Gelder (2005); Gudendorf and Segers (2010), and Ribatet and Sedki (2013). The major disadvantage of this scheme is that it is relatively ad hoc. Choosing one copula over another does not seem to be justified theoretically.

Later on, the conditional extremes models proposed by Heffernan and Tawn (2004), subsequently developed, for instance, by Jonathan *et al.* (2010); Gardes and Girard (2010); Das and Resnick (2011); Ewans and Jonathan (2014), and Simpson and Wadsworth (2021), provides an innovative paradigm to model marginals and dependence structures of multivariate extremes. Firstly, the marginal distribution of the variable is transformed into standard Gumbel distribution. Thus, the threshold exceedances of a marginal variable are modeled independently using the GPD. However, the dependence structure is modeled for pairs of transformed variables. In that case, the model assumptions may lead to the collection of distinct semi-parametric conditional models being fitted through ad hoc methods. The complicated nature of the methods may restrict model applicability.

In addition, the hierarchical construction of extreme value models has recently gained much attention in both theoretical and practical work. To observe the dependence behavior of the exceedances over the tail, Bortot and Gaetan (2014) developed the hierarchical framework and obtained GPD as marginal distribution by mixing Exponential and Gamma random variables. Further, they get bivariate distribution with GPD marginals through bivariate LT of multivariate Gamma distribution with Gamma marginals. In the same study, they use two distinct hierarchical constructions, representing different dependence structures over the tail. Subsequently, the hierarchical models are discussed by Casson and Coles (1999); Gaetan and Grigoletto (2004); Sang and Gelfand (2009); Bortot and Gaetan (2014); Reich *et al.* (2014); Bopp and Shaby (2017); Bacro *et al.* (2020); Yadav *et al.* (2020); Courceau and Veraart (2022); Bortot and Gaetan (2022), and Bacro *et al.* (2023).

As discussed above, Bortot and Gaetan (2014) uses a decomposition of the GPD to develop a hierarchical modeling framework for exceedances that preserves GPD as

marginal distribution. In contrast, if the values are discrete, one may wish to preserve and utilize the discreteness of the values. We are mainly concerned with asymptotic independence and dependence, which are the main limits for using discrete extreme value distributions to model the tail of the discrete multivariate distributions.

Therefore, the classical extremal techniques are appreciated, and a wide range of dependence can be considered at extreme levels. We can accomplish this by developing a bivariate distribution with DGPD marginals by mixing independent Geometric and Gamma random variables in hierarchical construction. The probability mass function (PMF) of bivariate discrete distribution can be expressed in terms of multivariate Gamma LTs with Gamma margins.

Considering dependent Geometric random variables instead of independent Geometric random variables while developing a bivariate discrete distribution. First, we need a bivariate Geometric distribution and use it to derive the bivariate distribution with DGPD marginal by considering the Geometric Gamma distribution mixture. The bivariate Geometric distribution is developed through the FGMC. The FGMC was introduced by Morgenstern (1956); later, many applications of this copula exist in literature. The copula theory, whose roots date back to the 20th century, has gained great attention in the last two decades, especially due to vital applications in hydrology, finance, climatology, and actuarial sciences. The copula paradigm permits us to use different marginal distributions from the study of the dependence models and then combine more marginal models with several possible dependence structures.

A simpler method to derive valid bivariate Geometric distribution is by pairing a copula distribution with Geometric marginals. The adoption of copula distribution is based on the well-known Sklar's theorem for continuous distributions. For instance, any continuous distribution can be putrefied into copula and marginal distributions, and conversely, a recipe of copula and marginal distributions provides a valid continuous joint distribution Sklar (1959). The copulas completely favor dependency structure for continuous distributions, thus separating marginal distributions from dependencies. (Inouye *et al.*, 2017). Generally, copula distributions with continuous marginals enjoy a wide variety of applications (see,, for instance, Cherubini *et al.* (2004) in finance and Peres *et al.* (2018)). However, copula models with discrete margins, such as Geometric or Poisson, are more challenging from both theoretical and computational points of view (Genest and Nešlehová, 2007; Nikoloulopoulos and Karlis, 2009; Nikoloulopoulos, 2013a,b). Furthermore, numerous explanations and recent developments have tried to tackle discussed limitations (see, for example, Inouye *et al.* (2017) and references therein). The purpose behind the development of bivariate Geometric distribution is

to take into account dependence among Geometric marginal via FGMC dependence parameter. The copula-based model has two dependence parameters: the copula dependence parameter and the other one is dependence parameter of the layer introduced through the hierarchy of Gamma random variables.

## 1.2 Discrete hierarchical models

### 1.2.1 DGPD and its hierarchical representation

Let  $Y$  be a discrete random variable. We are interested in modeling the data points above the fixed integer value high threshold  $u$ , which basically represents the tail behavior. Discrete observations above the high threshold  $u$  follow the DGPD, which has the PMF is

$$P(Y = k) = \left(1 + \frac{\xi k}{\sigma}\right)^{-\frac{1}{\xi}} - \left(1 + \frac{\xi(k+1)}{\sigma}\right)^{-\frac{1}{\xi}}, \quad k \in \mathbb{N}_0 \quad (1.8)$$

where  $(1 + \xi k/\sigma)^{-1/\xi}$  is the survival function (SF) of GPD,  $\sigma$  and  $\xi$  are the scale and shape parameters of DGPD, respectively.

For the purpose of subsequent constructions, we use  $\text{DGPD}(\beta, \alpha)$  with re-parametrization of scale and shape parameters as  $\beta = \sigma/\xi$  and  $\alpha = 1/\xi$ . After this re-parametrization, the PMF of DGPD is

$$P(Y = k) = \left(1 + \frac{k}{\beta}\right)^{-\alpha} - \left(1 + \frac{k+1}{\beta}\right)^{-\alpha}, \quad k \in \mathbb{N}_0 \quad (1.9)$$

We use the definition discussed by Buddana and Kozubowski (2014) to write  $\text{DGPD}(\beta, \alpha)$  as mixture of Geometric and Gamma distributions, that is

$$P(Y = k) = \int_0^\infty P(Y = k|\Lambda = \lambda)f(\lambda)d\lambda \quad (1.10)$$

where  $P(Y = k|\Lambda = \lambda)$  is Geometric distribution,  $\text{Geo}(q)$ , with  $q = 1 - e^{-\Lambda/\beta}$ , and  $\Lambda$  follows Gamma distribution with probability density function

$$f(\lambda, \alpha, \beta) = [\beta^\alpha/\Gamma(\alpha)]\lambda^{\alpha-1}e^{-\beta\lambda}. \quad (1.11)$$

Using the standard conditioning argument, the SF of DGPD can be derived by mixing the components of Geometric and Gamma distributions, that is

$$P(Y > k) = S(k) = \int_0^\infty e^{-\frac{\lambda k}{\beta}} f(\lambda) d\lambda = L^{(1)}(s)|_{(s=\frac{k}{\beta})}, \quad (1.12)$$

where  $e^{-\lambda k/\beta}$  is SF of Geo( $q$ ) distribution, and  $f(\lambda)$  is a probability density function of the standard Gamma distribution, i.e., Gamma( $\alpha, 1$ ). Here  $L^{(1)}(s)|_{(s=\frac{k}{\beta})}$  is the LT of Gamma( $\alpha, 1$ ) distribution. For  $\alpha > 0$ , the given consideration falls in D-MDA $_\alpha$  (Hitz *et al.*, 2017). Furthermore, similar to continuous GPD, the important stability property with respect to the exceedances is also retained for DGPD (Buddana and Kozubowski, 2014). The expression (1.12) has been used in other studies to extend the modeling framework in different scenarios. For instance, Bhati and Bakouch (2019) proposed a new geometric discrete Pareto distribution marginal by using a Geometric random variable with parameter  $q = 1 - e^{-(\omega+\Lambda)/\beta}$  instead of  $q = 1 - e^{-\Lambda/\beta}$  in (1.12). By using (1.12), Constantinescu *et al.* (2019) developed zero modified discrete Pareto distribution (generally called zero-inflated) by using zero modified Geometric random variable in place of Geometric random variables (i.e., ZMG( $\pi, q$ ), where  $\pi$  represent the proportion of zero values and  $q = 1 - e^{-\Lambda/\beta}$ , with  $\beta > 0$ ).

## 1.2.2 Bivariate models

An interesting aspect of our representation is the tractability of the bivariate distribution with DGPD margins. Moreover, the following propositions will clearly define our proposal regarding bivariate distributions. The proposition 1.1 is based on the Buddana and Kozubowski (2014) construction.

**Proposition 1.1.** *Let  $\Lambda$  have a standard Gamma distribution. Suppose, given  $\Lambda$ ,  $Y_i, i = 1, 2$  are independent Geometric random variables with parameter  $q_i = 1 - e^{-\lambda/\beta}, i = 1, 2$ , where  $\beta > 0$ . Then  $\mathbf{Y} = (Y_1, Y_2)$  follows a bivariate distribution with DGPD marginals and is defined through univariate LT of Gamma distribution with the support of both variables. That is, the joint survival function is written as*

$$S(k_1, k_2) = \int_0^\infty e^{-\frac{k_1 \lambda}{\beta} - \frac{k_2 \lambda}{\beta}} f(\lambda) d\lambda = L^{(1)}(s)|_{(s=\frac{k_1+k_2}{\beta})} \quad (1.13)$$

where  $e^{-k_1 \lambda/\beta}$  and  $e^{-k_2 \lambda/\beta}$  are the SFs of independent Geometric random variables  $Y_1$  and  $Y_2$ . The expression  $L^{(1)}(s)$  is the LT of  $\Lambda$ . Notice that expression (1.13) can easily be extended to the multivariate case.

We extend the idea using bivariate Gamma distributions instead of standard Gamma distributions as in proposition 1.1. Once again, our idea is based on representing bivariate distribution as Gamma mixtures of the Geometric distribution, enabling us to keep easily tractable marginal distributions that remain coherent with univariate EVT. When comparing with Buddana and Kozubowski (2014) construction, the major advantage of our proposal is that it allows us to introduce a dependence layer through bivariate Gamma distribution. Further, proposition 1.2 explains how we proceed with our proposal.

**Proposition 1.2.** *Let  $\Lambda = (\Lambda_1, \Lambda_2)$  have a bivariate distribution with Gamma margins. Suppose, given  $\Lambda = (\Lambda_1, \Lambda_2)$ ,  $Y_i, i = 1, 2$  are independent Geometric random variable with parameter  $q_i = 1 - e^{-\lambda_i/\beta}, i = 1, 2$ , where  $\beta > 0$ . Then  $\mathbf{Y} = (Y_1, Y_2)$  follows a bivariate distribution with DGPD marginals and is defined through bivariate LTs of bivariate Gamma distributions. That is, the joint survival function is written as*

$$S(k_1, k_1) = \int_0^\infty \int_0^\infty e^{-\frac{k_1 \lambda_1}{\beta} - \frac{k_2 \lambda_2}{\beta}} f(\lambda_1, \lambda_2) d\lambda_1 d\lambda_2 = L^{(2)}(s_1, s_2)|_{(s_1 = \frac{k_1}{\beta}, s_2 = \frac{k_2}{\beta})} \quad (1.14)$$

where  $e^{-k_1 \lambda_1 / \beta}$  and  $e^{-k_2 \lambda_2 / \beta}$  are the SFs of independent Geometric random variables  $Y_1$  and  $Y_2$ . The expression  $L^{(2)}(s_1, s_2)$  is the bivariate LT of  $\Lambda = (\Lambda_1, \Lambda_2)$ .

The proofs of the proposition 1.1 and 1.2 are given in Appendix A.1. Continuing our proposal, we incorporate the above propositions into a hierarchical framework. More formally, as shown in (1.10), a two-stage specification is easy to respect this constraint. In the second stage, we incorporate the dependence layer (either simple or temporal) through Gamma random variables or dependent Geometric random variables. In the case of dependent Geometric random variables, we first need to construct a bivariate Geometric distribution. This construction is done through a copula; the subsequent subsection contains a copula-based modeling framework.

### 1.2.3 Copula based models

In the previous section, we considered independent Geometric random variables and mixed them with Gamma random variables to get a joint distribution. Another possibility is to produce dependent Geometric random variables first and then represent them as a mixture with Gamma distribution (either univariate or bivariate). The attention is essential to construct a bivariate Geometric distribution and use it to derive the bivariate distribution with DGPD marginals by considering Geometric Gamma mixture representation. This construction can be done through the copula function. Before starting construction, the definition of copula is recalled.

A  $d$ -dimensional copula is a joint CDF in  $[0, 1]^d$  with standard uniform CDF's  $U_i, i = 1, 2, \dots, d$ , such that:

$$C(u_1, u_2, \dots, u_d) := P(U_1 \leq u_1, U_2 \leq u_2, \dots, U_d \leq u_d) \quad (1.15)$$

The key advantage of the copula in the study of multivariate distribution functions is summarized by Sklar's theorem (Sklar, 1959), which here we precisely recap.

Let  $F$  be a  $d$ -dimensional distribution function with marginals  $F_i, i = 1, 2, \dots, d$ . Then, there exist a  $d$ -dimensional copula i.e.,  $C_d : [0, 1]^d \rightarrow [0, 1]$ , such that  $\forall y_1, y_2, \dots, y_d$  in domain  $d$ . That is,

$$F(y_1, y_2, \dots, y_d) = C_d(F_1(y_1), F_2(y_2), \dots, F_d(y_d)). \quad (1.16)$$

If  $F_i(\cdot), i = 1, 2, \dots, d$  are continuous, then  $C_d$  is unique; otherwise;  $C_d$  is uniquely determined on  $Ran(F_1) \times \dots \times Ran(F_d)$ , where  $Ran(F_i)$  denote the range of  $F_i$ . Alternatively, if  $C_d$  is a copula and  $F_i(\cdot), i = 1, 2, \dots, d$  are univariate CDF's, then the function  $F$  given in (1.16) is joint CDF with  $F_i(\cdot), i = 1, 2, \dots, d$  marginals. If the marginals are continuous, then the unique copula  $C_d$  is defined as

$$C(u_1, u_2, \dots, u_d) = F(F_1^{-1}(u_1), F_2^{-1}(u_2), \dots, F_d^{-1}(u_d)) \quad (1.17)$$

where  $F_i^{-1}$  indicates the generalized inverse of the marginal CDF, according to Sklar's Theorem, it is natural to define the notion of copula of a CDF  $F$  with continuous marginal CDF's  $F_i(\cdot), i = 1, 2, \dots, d$  as the CDF  $C_d$  of  $(F_1(Y_1), F_2(Y_2), \dots, F_d(Y_d))$  (Biller and Corlu, 2012). Now, we will proceed to construct a bivariate Geometric distribution using FGMC. The FGMC was introduced by Morgenstern (1956). In general, the copula is a function that connects the cumulative distribution function to its marginals, and a distribution function characterizes the dependence structure of the model.

Let  $S(k_i) = e^{-k_i \lambda_i / \beta}, i = 1, 2$  are the univariate SFs of univariate Geometric random variables  $Y_i, i = 1, 2$ . Following Morgenstern (1956); Gumbel (1960); Farlie (1960), the bivariate FGMC is given by

$$C(u_1, u_2) = u_1 u_2 [1 + \phi(1 - u_1)(1 - u_2)], \quad u_1, u_2 \in [0, 1] \quad (1.18)$$

where  $\phi \in [-1, 1]$  is a measure of the dependence between  $u_1$  and  $u_2$ .

The joint SF of Geometric random variables  $Y_1$  and  $Y_2$  by using FMGC with  $u_i =$

$S(k_i), i = 1, 2$  is written as

$$S(k_1^*, k_2^* | \lambda_1, \lambda_2) = e^{-\left(\frac{\lambda_1 k_1}{\beta} + \frac{\lambda_2 k_2}{\beta}\right)} + \phi \left( e^{-\left(\frac{\lambda_1 k_1}{\beta} + \frac{\lambda_2 k_2}{\beta}\right)} - e^{-\left(\frac{2\lambda_1 k_1}{\beta} + \frac{\lambda_2 k_2}{\beta}\right)} - e^{-\left(\frac{\lambda_1 k_1}{\beta} + \frac{2\lambda_2 k_2}{\beta}\right)} + e^{-\left(\frac{2\lambda_1 k_1}{\beta} + \frac{2\lambda_2 k_2}{\beta}\right)} \right) \quad (1.19)$$

Further,  $\Lambda_1$  and  $\Lambda_2$  are Gamma random variables (see subsection 1.2.1). Notice that, when  $\phi = 0$  the joint SF given in equation (1.19) is reduces to  $S(k_1, k_2) = e^{-(\lambda_1 k_1/\beta + \lambda_2 k_2/\beta)}$  and  $Y_1$  and  $Y_2$  are independent. In addition, the PMF of bivariate Geometric distribution can be derived easily by using the relationship of bivariate PMF and SF as suggested by Barbiero (2019). Since we will use only the SF to move forward, we are not interested in the bivariate PMF of Geometric distribution.

The main objective is to propose a copula-based bivariate distribution that allows for dependence between variables while still having the DGPD margins. Subsequent propositions will support a new proposal regarding bivariate distribution.

**Proposition 1.3.** *Let  $\Lambda$  follow a standard distribution. Suppose, given  $\Lambda = \lambda$ ,  $S(k_1^*, k_2^* | \lambda)$  is the survival function of copula-based bivariate Geometric distribution with Geometric margins with parameter  $q_i = 1 - e^{-\Lambda/\beta}, i = 1, 2$ , where  $\beta > 0$ . Then  $\mathbf{Y} = (Y_1, Y_2)$  follows a copula-based bivariate distribution with DGPD marginals and is defined through LT of standard Gamma distribution. That is, the joint SF via (1.13) is written as*

$$S(k_1, k_2) = \int_0^\infty \left[ e^{-\left(\frac{\lambda k_1}{\beta} + \frac{\lambda k_2}{\beta}\right)} + \phi \left( e^{-\left(\frac{\lambda k_1}{\beta} + \frac{\lambda k_2}{\beta}\right)} - e^{-\left(\frac{2\lambda k_1}{\beta} + \frac{\lambda k_2}{\beta}\right)} - e^{-\left(\frac{\lambda k_1}{\beta} + \frac{2\lambda k_2}{\beta}\right)} + e^{-\left(\frac{2\lambda k_1}{\beta} + \frac{2\lambda k_2}{\beta}\right)} \right) \right] f(\lambda) d\lambda \quad (1.20)$$

$$S(k_1, k_2) = L^{(1)}(s) \Big|_{(s=\frac{k_1+k_2}{\beta})} + \phi \left( L^{(1)}(s) \Big|_{(s=\frac{k_1+k_2}{\beta})} - L^{(1)}(s) \Big|_{(s=\frac{2k_1+k_2}{\beta})} - L^{(1)}(s) \Big|_{(s=\frac{k_1+2k_2}{\beta})} + L^{(1)}(s) \Big|_{(s=\frac{2k_1+2k_2}{\beta})} \right) \quad (1.21)$$

where  $\phi \in (-1, 1)$  is the copula dependence parameter and  $L^{(1)}(\cdot)$  is the LT of random variable  $\Lambda$ .  $S(k_1, k_2)$  is the joint SF of copula-based bivariate distribution with DGPD margins.

When  $\phi = 0$ , the expression (1.21) reduces to Buddana and Kozubowski (2014) proposal, which is shown in proposition 1.1.

**Proposition 1.4.** *Let  $\mathbf{\Lambda} = (\Lambda_1, \Lambda_2)$  having a bivariate distribution with Gamma margins. Suppose, given  $\mathbf{\Lambda} = (\Lambda_1, \Lambda_2)$ ,  $S(k_1^*, k_2^* | \lambda_1, \lambda_2)$  is the SF of copula-based bivariate Geometric distribution with Geometric margins with parameter  $q_i = 1 - e^{-\Lambda_i/\beta}$ ,  $i = 1, 2$ , where  $\beta > 0$ . Then  $\mathbf{Y} = (Y_1, Y_2)$  follows a copula-based bivariate distribution with DGPD marginals and is defined through bivariate LTs of bivariate Gamma distributions. That is, the joint SF is written as*

$$S(k_1, k_2) = \int_0^\infty \int_0^\infty \left[ e^{-\left(\frac{\lambda_1 k_1}{\beta} + \frac{\lambda_2 k_2}{\beta}\right)} + \phi \left( e^{-\left(\frac{\lambda_1 k_1}{\beta} + \frac{\lambda_2 k_2}{\beta}\right)} - e^{-\left(\frac{2\lambda_1 k_1}{\beta} + \frac{\lambda_2 k_2}{\beta}\right)} - e^{-\left(\frac{\lambda_1 k_1}{\beta} + \frac{2\lambda_2 k_2}{\beta}\right)} + e^{-\left(\frac{2\lambda_1 k_1}{\beta} + \frac{2\lambda_2 k_2}{\beta}\right)} \right) \right] f(\lambda_1, \lambda_2) d\lambda_1 d\lambda_2 \quad (1.22)$$

$$S(k_1, k_2) = L^{(2)}(s_1, s_2) \Big|_{(s_1=\frac{k_1}{\beta}, s_2=\frac{k_2}{\beta})} + \phi \left( L^{(2)}(s_1, s_2) \Big|_{(s_1=\frac{k_1}{\beta}, s_2=\frac{k_2}{\beta})} - L^{(2)}(s_1, s_2) \Big|_{(s_1=\frac{2k_1}{\beta}, s_2=\frac{k_2}{\beta})} - L^{(2)}(s_1, s_2) \Big|_{(s_1=\frac{k_1}{\beta}, s_2=\frac{2k_2}{\beta})} + L^{(2)}(s_1, s_2) \Big|_{(s_1=\frac{2k_1}{\beta}, s_2=\frac{2k_2}{\beta})} \right) \quad (1.23)$$

where  $L^{(2)}(., .)$  is the bivariate LT of  $\mathbf{\Lambda} = (\Lambda_1, \Lambda_2)$ .  $S(k_1, k_2)$  is a joint SF of copula-based bivariate discrete distribution with DGPD marginals.

When  $\phi = 0$ , the expression (1.23) reduces to the construction proposed in proposition 1.2.

The proofs of the propositions 1.3 and 1.4 are also provided in Appendix A.1.

## 1.2.4 Dependence induction

This section deals with how dependence can be induced in bivariate models as well as in copula-based models. Interestingly, we can introduce the dependence layer in our proposal through Gamma random variables. From a modeling perspective, we require an analytical expression of bivariate LT of multivariate Gamma distribution. We use hierarchical models with Gamma margins and have different analytical expressions of their Laplace transforms, which can produce different kinds of dependence structures at the upper tail. Furthermore, the complete dependence model, which leads to the result of a proposition 1.1, is discussed later in the simulation study and real data examples. We are considering the following hierarchical models.



**1. Gaver model (GM):** This class of hierarchical model is firstly introduced by Gaver and Lewis (1980); Walker (2000). Let

$$\begin{aligned}
\Lambda_1 &\sim \text{Gamma}(\alpha, \beta) \\
P &\sim \text{Gamma}(\alpha, 1) \\
X|P &\sim \text{Poisson}\left(P\frac{(1-\rho)}{\rho}\right) \\
V|X &\sim \text{Gamma}\left(X, \frac{\beta}{\rho}\right) \\
\Lambda_2 &= \rho\Lambda_1 + V
\end{aligned} \tag{1.24}$$

with  $\Lambda_1$  independent of  $P$ ,  $X|P$  and  $V|X$ .  $\rho$  is the dependence parameter which represents the correlation between  $\Lambda_1$  and  $\Lambda_2$ , it ranges  $0 \leq \rho < 1$ . Given  $\mathbf{\Lambda} = (\Lambda_1, \Lambda_2)$  follow the marginal distributions  $\text{Gamma}(\alpha, \beta)$ .

**2. Kibble model (KM):** Among various bivariate Gamma models, we focus on the bivariate Gamma distribution introduced by Kibble (1941). The major advantage of using this model is that the analytical form of their LT exists. Further, to simulate dependent Gamma random variables, we use a hierarchical model introduced by Warren (1992) and is defined as

$$\begin{aligned}
\Lambda_1 &\sim \text{Gamma}(\alpha, \beta) \\
X|\Lambda_1 &\sim \text{Poisson}\left(\frac{\rho\Lambda_1\beta}{1-\rho}\right) \\
\Lambda_2|X &\sim \text{Gamma}\left(X + \alpha, \frac{1-\rho}{\beta}\right)
\end{aligned} \tag{1.25}$$

The resulting  $\mathbf{\Lambda} = (\Lambda_1, \Lambda_2)$  again follow  $\text{Gamma}(\alpha, \beta)$  margins with dependence parameter  $\rho$ , it ranges  $0 \leq \rho < 1$ .

**3. Thinned Gamma model (TGM):** This class of model was introduced by (Wolpert, 2021) by using the thinning layer generated from Beta distribution. Let

$$\begin{aligned}
\Lambda_1 &\sim \text{Gamma}(\alpha, \beta) \\
B &\sim \text{Beta}(\alpha\rho, \alpha(1-\rho)) \\
V &\sim \text{Gamma}(\alpha(1-\rho), \beta) \\
\Lambda_2 &= B\Lambda_1 + V
\end{aligned} \tag{1.26}$$

$\mathbf{\Lambda} = (\Lambda_1, \Lambda_2)$  again follow  $\text{Gamma}(\alpha, \beta)$  margins with dependence parameter  $\rho$ , it ranges  $0 < \rho < 1$ .

The interesting feature of the above hierarchical models is that they have the same correlation structure, that is,  $\text{corr}(\Lambda_1, \Lambda_2) = \rho$ . In addition, we will work with LTs of  $\Lambda_1$  and  $\Lambda_2$  because of Gamma( $\alpha, \beta$ ) margins, the univariate Laplace for the above-defined models is

$$L^{(1)}(s)|_{(s=\frac{k}{\beta})} = E(e^{-\frac{\Lambda k}{\beta}}) = \left(\frac{\beta}{\beta + k}\right)^\alpha \quad (1.27)$$

In addition, we will have to work with bivariate LTs of  $\Lambda_1$  and  $\Lambda_2$  as

$$L^{(2)}(s_1, s_2)|_{(s_1=\frac{k_1}{\beta}, s_2=\frac{k_2}{\beta})} = E\left(e^{-\frac{\Lambda_1 k_1}{\beta} - \frac{\Lambda_2 k_2}{\beta}}\right), \quad (1.28)$$

which is given by

$$L^{(2)}(s_1, s_2)|_{(s_1=\frac{k_1}{\beta}, s_2=\frac{k_2}{\beta})} = \left[\frac{(\beta + \rho k_2)\beta}{(\beta + k_2)(\beta + k_1 + \rho k_2)}\right]^\alpha, \quad 0 < \rho < 1, \quad (1.29)$$

for GM and by

$$L^{(2)}(s_1, s_2)|_{(s_1=\frac{k_1}{\beta}, s_2=\frac{k_2}{\beta})} = \left[\frac{\beta^2}{(k_1 + \beta)(k_2 + \beta) - \rho k_1 k_2}\right]^\alpha, \quad 0 < \rho < 1 \quad (1.30)$$

for KM (Lai and Balakrishnan, 2009), respectively. Further, the

$$L^{(2)}(s_1, s_2)|_{(s_1=\frac{k_1}{\beta}, s_2=\frac{k_2}{\beta})} = \left[\frac{\beta^{\alpha(2-\rho)}}{(\beta + k_1)^{\alpha(1-\rho)}(\beta + k_1 + k_2)^{\alpha\rho}(\beta + k_2)^{\alpha(1-\rho)}}\right] \quad (1.31)$$

with  $0 < \rho < 1$  is bivariate LT for TGM.

### 1.3 Tail dependence coefficient

Extreme value theory is regularly applied to model the data of different extreme events, for which the paradigm of dependence is commonly inherent. Dependence arises, for example, when the stochastic behavior of different processes under investigation is related to each other. Majority of dependence measures, like Pearson's product-moment correlation coefficient, Spearman's rank order correlation coefficient, and Kendall correlation coefficient, are intended to determine the dependence of random variables on their distributions. Most regularly used measures fail to represent the dependence of the distribution at upper and lower tails (Kotz and Nadarajah, 2000). Conversely, the upper part of the distribution may behave differently in terms of dependency than the central and/or lower parts of the distribution (Embrechts *et al.*, 2002).

In the continuous extreme value framework, several tail dependence measures were

developed to describe the behavior of two or more variables at tails (upper or lower) (Ledford and Tawn, 1997) and (Coles *et al.*, 1999). In this context, our main focus is on the tail dependence coefficient, which was introduced early by Sibuya (1960). For instance,  $Y_1$  and  $Y_2$  are two random variables with identical marginal distributions, one natural measure based on conditional probabilities, later defined by Coles *et al.* (1999)

$$\chi = \lim_{k \rightarrow k^*} P(Y_1 > k | Y_2 > k) = \lim_{k \rightarrow k^*} \frac{L^{(2)}(s_1, s_2)|_{(s_1=\frac{k}{\beta}, s_2=\frac{k}{\beta})}}{L^{(1)}(s)|_{(s=\frac{k}{\beta})}}, \quad (1.32)$$

where  $k^* \rightarrow \infty$  is the upper endpoint of the common marginal distribution. The variables are said to be asymptotically independent when  $\chi = 0$ ; otherwise asymptotically dependent. Several statistical methods for the general class of extreme value distributions have been proposed having  $\chi = 0$ , see, e.g., (Ledford and Tawn, 1996, 1997; Bortot and Tawn, 1998). Later, Coles *et al.* (1999) and Coles (2001, Ch 8) pointed out that within the class of asymptotically dependent distributions, it is evident that  $\chi$  gives the simple extremal dependence measure. However, it shows different behavior for asymptotically independent distributions at finite levels. That's why Coles *et al.* (1999) introduced another extremal dependence measure, so-called  $\bar{\chi}$ , to overcome this deficiency. We define  $\bar{\chi}$  as

$$\bar{\chi} = \lim_{k \rightarrow k^*} \frac{2 \log P(Y_1 > k)}{\log P(Y_1 > k, Y_2 > k)} - 1 = \lim_{k \rightarrow k^*} \frac{2 \log L^{(1)}(s)|_{(s=\frac{k}{\beta})}}{\log L^{(2)}(s_1, s_2)|_{(s_1=\frac{k}{\beta}, s_2=\frac{k}{\beta})}} - 1 \quad (1.33)$$

In the present scenario, we are working in the domain of the discrete extreme. The tail dependence concept is the same as a continuous framework. Furthermore, we explain the concept of tail dependence more precisely in discrete extreme cases by giving an example. Figure 1.3 describes the concept of discrete extreme models tail dependencies. Figure 1.3 shows two generated discrete random variables from different hierarchical settings with parameters ( $\beta = 200, \alpha = 30$ , and  $\rho = 0.95$ ). Figure 1.3 *left* is simulated using the GM, while Figure 1.3 *right* is generated using the KM. The integer variables ( $Y_1$  and  $Y_2$  with realization  $k_1$  and  $k_2$ ) are positive in nature in both figures. Notice that the upper right quadrant shaded with red color (above the red lines) is dissimilar in the left and right panels of Figure 1.3. The red lines are drawn corresponding to high threshold values. Hence, the points lying in the upper right quadrant of Figure 1.3 *left* are locally dependent, while the points lie in the upper right quadrant of Figure 1.3 *right* seem to be locally independent. Also, Bortot and Gaetan (2014) proved in a continuous framework

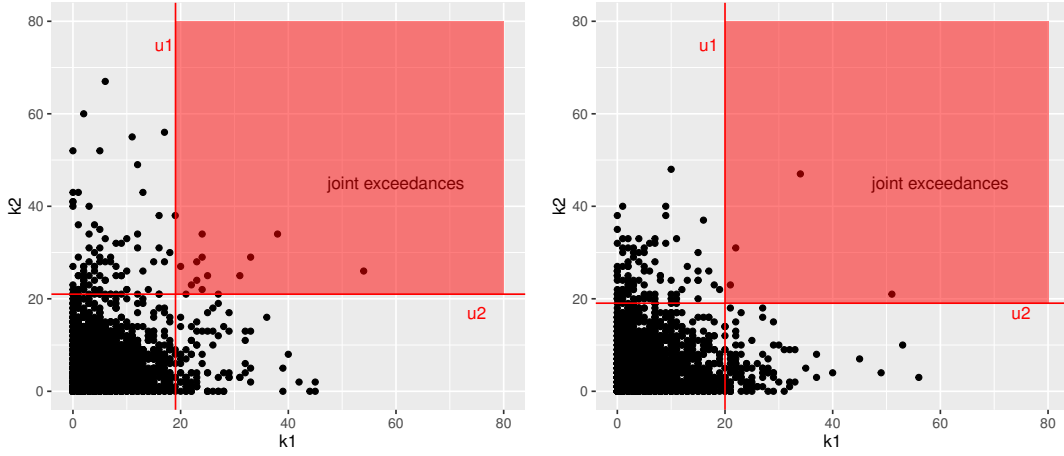


FIGURE 1.3: **(a)** Upper tail values (upper right quadrant above red lines) of  $Y_1$  and  $Y_2$  generated from GM are locally correlated; **(b)** Upper tail values (upper right quadrant above red lines) of  $Y_1$  and  $Y_2$  generated from KM seem to be locally independent

that the GM has asymptotic dependence while the KM has asymptotically independent behavior over the tail. Similarly, the model based on Buddana and Kozubowski (2014) construction has asymptotic dependent behavior while TGM has asymptotic independent behavior over the tail. For this reason, we must derive the asymptotic properties of conditional tail dependence measure for the hierarchical constructions proposed in section 1.2.

### 1.3.1 Gaver model (GM)

For a broader understanding of the extremal dependence in GM, we would like to use the simplified form of (1.32) and to prove the limiting measure ( $\chi$ ) to evaluate the dependence between the joint exceedances of discrete random variables  $Y_1$  and  $Y_2$ .

However, the definition of  $\chi$  measure given in (1.32) is based on the assumption that  $Y_1$  and  $Y_2$  are positive discrete-valued random variables with a marginal SF given in (1.12). Further, the joint SF corresponding GM is provided in the form of LT in (1.29). To get the limiting measure  $\chi$  for GM, we solve (1.32) by incorporating univariate and bivariate LTs of Gamma distribution linked with the GM, that is,

$$\chi = \lim_{k \rightarrow \infty} \frac{\left[ \frac{\{\beta + \rho k\} \beta}{(\beta + k) \{\beta + k + \rho k\}} \right]^\alpha}{\left[ \frac{\beta}{\beta + k} \right]^\alpha} = \left[ \frac{\rho}{1 + \rho} \right]^\alpha \quad (1.34)$$

The positive limit of  $\chi = [\rho / (1 + \rho)]^\alpha$  indicates that the discrete random variables ( $Y_1$  and  $Y_2$ ) are asymptotically dependent.

To attain a better characterization of the joint tail decay rate under asymptotic independence, quicker than the marginal tail the decay rate, we use the definition (1.33) through the limit relation as if  $\chi > 0$  then  $\bar{\chi} = 1$ . The derivation of the limiting form of  $\chi$  and  $\bar{\chi}$  corresponding to the GM are provided in Appendix A.2.1.

### 1.3.2 Copula-based Gaver and Lewis model (CGM)

To obtain the limiting measure  $\chi$  corresponding CGM, we use equation (1.32) with joint SF given in (1.23) Thus, we have

$$\chi = \lim_{k \rightarrow \infty} \frac{\left( \left[ \frac{\{\beta+\rho k\}\beta}{(\beta+k)\{\beta+k+\rho k\}} \right]^\alpha + \phi \left[ \frac{\{\beta+\rho k\}\beta}{(\beta+k)\{\beta+k+\rho k\}} \right]^\alpha - \left[ \frac{\{\beta+\rho k\}\beta}{(\beta+k)\{\beta+2k+\rho k\}} \right]^\alpha - \left[ \frac{\{\beta+2\rho k\}\beta}{\{\beta+2k\}\{\beta+k+2\rho k\}} \right]^\alpha + \left[ \frac{\{\beta+2\rho k\}\beta}{\{\beta+2k\}\{\beta+2k+2\rho k\}} \right]^\alpha \right)}{\left[ \frac{\beta}{\beta+k} \right]^\alpha}$$

$$\chi = \left( \frac{\rho}{1+\rho} \right)^\alpha + \phi \left[ \left( \frac{\rho}{1+\rho} \right)^\alpha - \left( \frac{\rho}{2+\rho} \right)^\alpha - \left( \frac{\rho}{1+2\rho} \right)^\alpha + \left( \frac{\rho}{2+2\rho} \right)^\alpha \right] \quad (1.35)$$

Notice that the  $\chi > 0$  indicates that CGM also has asymptotic-dependent behavior. Additionally,  $\phi$  involve in (1.35), which is a copula dependence parameter. Positive value  $\phi$  show more asymptotic dependence, while a negative value of  $\phi$  show weaker dependence as compared to GM. When  $\phi = 0$  CGM has similar asymptotic dependence behavior as GM. The detailed proof is given in Appendix A.2.2.

### 1.3.3 Kibble model (KM)

To observe the joint tail dependence behavior of KM via  $\chi$  and  $\bar{\chi}$ , we use definitions given in (1.32) and (1.33) with bivariate LT linked with KM, which is provided in 1.30. Thus, the

$$\chi = \lim_{k \rightarrow \infty} \frac{\left[ \frac{\beta^2}{(k+\beta)(k+\beta)-\rho k^2} \right]^\alpha}{\left[ \frac{\beta}{\beta+k} \right]^\alpha} = 0 \quad (1.36)$$

and

$$\bar{\chi} = \lim_{k \rightarrow \infty} \frac{2 \log \left[ \frac{\beta}{\beta+k} \right]^\alpha}{\log \left[ \frac{\beta^2}{(k+\beta)(k+\beta)-\rho k^2} \right]^\alpha} - 1 = 0 \quad (1.37)$$

In the case of KM, the measures  $\chi = \bar{\chi} = 0$  show no local clustering of extremes in  $Y_1$  and  $Y_2$  variables. This means that KM has asymptotic independence behavior over the tail. The detailed derivation of limiting measure  $\chi$  and  $\bar{\chi}$  corresponding to the KM are given in Appendix A.2.3.

### 1.3.4 Copula-based Kibble model (CKM)

To get the limiting form of  $\chi$  and  $\bar{\chi}$  corresponding CKM, we use the definitions given (1.32) and (1.33) with joint SF given in (1.23) based on bivariate LT associated with KM provided in (1.30). That is, the  $\chi$  is

$$\chi = \lim_{k \rightarrow \infty} \frac{\left( \begin{aligned} & \left[ \frac{\beta^2}{(k+\beta)^2 - \rho k^2} \right]^\alpha + \phi \left[ \frac{\beta^2}{(k+\beta)^2 - \rho k^2} \right]^\alpha \\ & - \left[ \frac{\beta^2}{(2k+\beta)(k+\beta) - 2\rho k^2} \right]^\alpha - \left[ \frac{\beta^2}{(k+\beta)(2k+\beta) - 2\rho k^2} \right]^\alpha \\ & + \left[ \frac{\beta^2}{(2k+\beta)^2 - 4\rho k^2} \right]^\alpha \end{aligned} \right)}{\left[ \frac{\beta}{\beta+k} \right]^\alpha} = 0$$

Again, the limiting value of  $\chi$  and  $\bar{\chi}$  based CKM tend to zero, indicating that there is no local clustering at an extreme level of  $Y_1$  and  $Y_2$  variables. It means that CKM also has asymptotic independence behavior over the tail. The detailed derivation of the properties  $\chi$  and  $\bar{\chi}$  are provided in Appendix A.2.4.

### 1.3.5 Thinned Gamma model (TGM)

Similar to GM and KM, we would like to derive the asymptotic properties of joint tail dependence measure  $\chi$  and  $\bar{\chi}$  corresponding to TGM by using the usual definitions with bivariate LT given in (1.31). The limiting measure  $\chi$  is

$$\chi = \lim_{k \rightarrow \infty} \frac{\left[ \frac{\beta^{\alpha(2-\rho)}}{(\beta+k)^{2\alpha(1-\rho)}(\beta+2k)^{\alpha\rho}} \right]}{\left[ \frac{\beta}{\beta+k} \right]^\alpha} = 0 \quad (1.38)$$

and, the  $\bar{\chi}$  is

$$\bar{\chi} = \lim_{k \rightarrow \infty} \frac{2 \log \left[ \frac{\beta}{\beta+k} \right]^\alpha}{\log \left[ \frac{\beta^{\alpha(2-\rho)}}{(\beta+k)^{2\alpha(1-\rho)}(\beta+2k)^{\alpha\rho}} \right]} - 1 = \frac{\rho}{2-\rho} \quad (1.39)$$

The measure  $\chi = 0$  and  $\bar{\chi} = \rho/(2-\rho)$  show that the  $Y_1$  and  $Y_2$  variables are asymptotically independent when generated from TGM. A larger value of  $\rho$  in  $\bar{\chi}$  may lead to dependence (Bacro *et al.*, 2020). The detailed derivation of asymptotic properties of  $\chi$  and  $\bar{\chi}$  is given in Appendix A.2.5.

### 1.3.6 Copula-based Thinned Gamma model (CTGM)

In order to obtain the limiting value of  $\chi$  and  $\bar{\chi}$  corresponding TGM, we use the definitions given (1.32) and (1.33) with joint SF given in (1.23) based on bivariate LT linked with TGM provided in (1.31). That is, the  $\chi$  is

$$\chi = \lim_{k \rightarrow \infty} \frac{\left( \left[ \frac{\beta^{\alpha(2-\rho)}}{(\beta+k)^{2\alpha(1-\rho)}(\beta+2k)^{\alpha\rho}} \right] + \phi \left[ \frac{\beta^{\alpha(2-\rho)}}{(\beta+k)^{2\alpha(1-\rho)}(\beta+2k)^{\alpha\rho}} \right] - \left[ \frac{\beta^{2\alpha-\alpha\rho}}{(\beta+2k)^{\alpha-\alpha\rho}(\beta+k)^{\alpha-\alpha\rho}(\beta+3k)^{\alpha\rho}} \right] - \left[ \frac{\beta^{2\alpha-\alpha\rho}}{(\beta+2k)^{\alpha-\alpha\rho}(\beta+k)^{\alpha-\alpha\rho}(\beta+3k)^{\alpha\rho}} \right] + \left[ \frac{\beta^{2\alpha-\alpha\rho}}{(\beta+2k)^{2\alpha-2\alpha\rho}(\beta+4k)^{\alpha\rho}} \right] \right)}{\left[ \frac{\beta}{\beta+k} \right]^\alpha} \quad (1.40)$$

and  $\bar{\chi}$  is  $\rho/(2-\rho)$ .

Again, measure  $\chi = 0$  and  $\bar{\chi} = \rho/(2-\rho)$  show that the  $Y_1$  and  $Y_2$  variables are asymptotically independent when generated from CTGM. In addition, the copula dependence parameter  $\phi$  has a neutral role in both  $\chi$  and  $\bar{\chi}$ . The derivation is provided in Appendix A.2.6.

### 1.3.7 Complete dependence model (CDM)

In order to get the limiting value of conditional tail dependence measure for CDM discussed in proposition (1.1), we use the usual definition given in (1.32) with LT of

Gamma distribution given in (1.27) with the involvement of  $Y_1$  and  $Y_2$  random variables.

$$\chi = \lim_{k \rightarrow \infty} \frac{\left[ \frac{\beta}{2k+\beta} \right]^\alpha}{\left[ \frac{\beta}{\beta+k} \right]^\alpha} = \left[ \frac{1}{2} \right]^\alpha \quad (1.41)$$

The positive limit of  $\chi = [1/2]^\alpha$  indicates that the discrete random variables ( $Y_1$  and  $Y_2$ ) are said to be asymptotically dependent. The related proofs are provided in Appendix A.2.7.

### 1.3.8 Copula-based complete dependence model (CCDM)

To obtain the limiting measure  $\chi$  corresponding CCDM, we use again expression (1.32) with joint SF. Thus, we have

$$\begin{aligned} \chi &= \lim_{k \rightarrow \infty} \frac{\left[ \frac{\beta}{2k+\beta} \right]^\alpha + \phi \left[ \left[ \frac{\beta}{2k+\beta} \right]^\alpha - \left[ \frac{\beta}{3k+\beta} \right]^\alpha - \left[ \frac{\beta}{3k+\beta} \right]^\alpha + \left[ \frac{\beta}{4k+\beta} \right]^\alpha \right]}{\left[ \frac{\beta}{\beta+k} \right]^\alpha} \\ \chi &= \left( \frac{1}{2} \right)^\alpha + \phi \left[ \left( \frac{1}{2} \right)^\alpha - \left( \frac{1}{3} \right)^\alpha - \left( \frac{1}{3} \right)^\alpha + \left( \frac{1}{4} \right)^\alpha \right] \end{aligned} \quad (1.42)$$

The limiting measure  $\chi > 0$  clearly indicates that the CCDM has asymptotic-dependent behavior. For CCDM, the parameter  $\phi$  plays a similar role as for CGM. The proofs are provided in Appendix A.2.8.

## 1.4 Inference

The section 1.2 provides the details of the main theoretical results of the proposed models. In section 1.4, the inference procedure of the models is presented since they will be exploited in a simulation study and real data analysis.

### 1.4.1 Censored likelihood

The censored likelihood is developed to estimate the bivariate and copula-based bivariate hierarchical models. Let  $(k_{11}, k_{21}), \dots, (k_{1n}, k_{2n})$  are the realizations of integer-valued random variables  $(Y_1, Y_2)$  with joint SF defined in expression (1.14) and (1.23), respectively. For suitable thresholds  $u_1$  and  $u_2$ , the marginal distribution of  $(Y_1, Y_2)$  follows DGPD with SF provided in equation (1.12). Moreover, the joint PMF can be constructed by using the joint LTs of  $\Lambda_1$  and  $\Lambda_2$ . Thus, the joint PMF, when comparing



the pair with suitable high thresholds  $u_1$  and  $u_2$ , whether the pair  $(k_1, k_2)$  lies above or below the thresholds. That is,

- for  $k_1 > u_1, k_2 > u_2$ .

$$P(k_1, k_2; \theta) = S(k_1, k_2) - S(k_1 + 1, k_2) - S(k_1, k_2 + 1) + S(k_1 + 1, k_2 + 1)$$

$$P(k_1, k_2; \theta) = L^{(2)}(s_1, s_2)|_{(s_1=\frac{k_1}{\beta}, s_2=\frac{k_2}{\beta})} - L^{(2)}(s_1, s_2)|_{(s_1=\frac{k_1+1}{\beta}, s_2=\frac{k_2}{\beta})} \\ - L^{(2)}(s_1, s_2)|_{(s_1=\frac{k_1}{\beta}, s_2=\frac{k_2+1}{\beta})} + L^{(2)}(s_1, s_2)|_{(s_1=\frac{k_1+1}{\beta}, s_2=\frac{k_2+1}{\beta})}$$

- for  $k_1 > u_1, k_2 \leq u_2$ .

$$P(k_1, u_2; \theta) = S(k_1) - S(k_1 + 1) - S(k_1, u_2 + 1) + S(k_1 + 1, u_2 + 1)$$

$$P(k_1, u_2; \theta) = L^{(1)}(s)|_{(s=\frac{k_1}{\beta})} - L^{(1)}(s)|_{(s=\frac{k_1+1}{\beta})} - L^{(2)}(s_1, s_2)|_{(s_1=\frac{k_1}{\beta}, s_2=\frac{u_2+1}{\beta})} \\ + L^{(2)}(s_1, s_2)|_{(s_1=\frac{k_1+1}{\beta}, s_2=\frac{u_2+1}{\beta})}$$

- for  $k_1 \leq u_1, k_2 > u_2$ .

$$P(u_1, k_2; \theta) = S(k_2) - S(k_2 + 1) - S(u_1 + 1, k_2) + S(u_1 + 1, k_2 + 1)$$

$$P(u_1, k_2; \theta) = L^{(1)}(s)|_{(s=\frac{k_2}{\beta})} - L^{(1)}(s)|_{(s=\frac{k_2+1}{\beta})} - L^{(2)}(s_1, s_2)|_{(s_1=\frac{u_1+1}{\beta}, s_2=\frac{k_2}{\beta})} \\ + L^{(2)}(s_1, s_2)|_{(s_1=\frac{u_1+1}{\beta}, s_2=\frac{k_2+1}{\beta})}$$

- for  $k_1 \leq u_1, k_2 \leq u_2$

$$P(u_1, u_2; \theta) = 1 - S(u_1 + 1) - S(u_2 + 1) + S(u_1 + 1, u_2 + 1)$$

$$P(u_1, u_2; \theta) = 1 - L^{(1)}(s)|_{(s=\frac{u_1+1}{\beta})} - L^{(1)}(s)|_{(s=\frac{u_2+1}{\beta})} + L^{(2)}(s_1, s_2)|_{(s_1=\frac{u_1+1}{\beta}, s_2=\frac{u_2+1}{\beta})}$$

In each case  $L^{(2)}(\cdot, \cdot)$  is the LT of  $\Lambda_1$  and  $\Lambda_2$ ,  $L^{(1)}(\cdot)$  is the univariate LT of  $\Lambda$ . Inference for both bivariate and copula-based bivariate models is complicated because a bivariate pair may lie in any of one regions defined above. For instance, a point  $(k_1, k_2)$  lie in

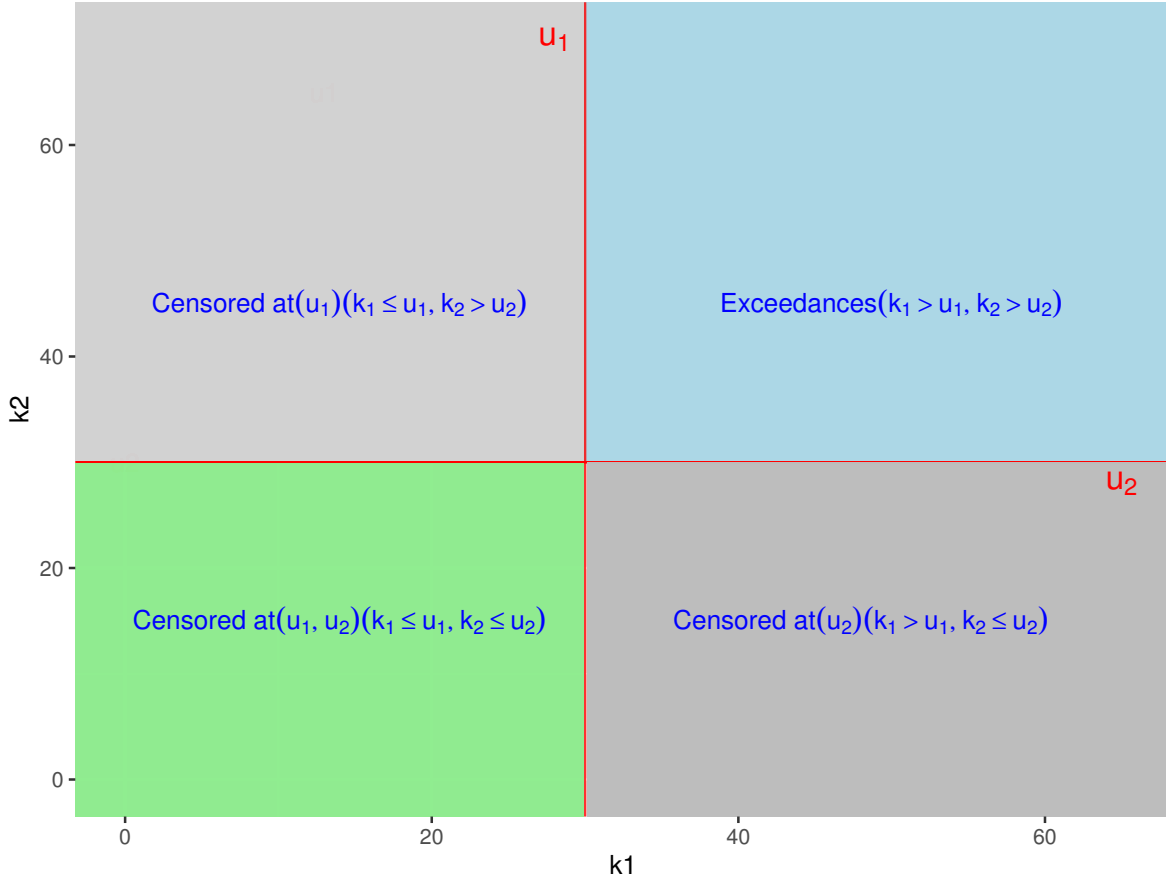


FIGURE 1.4: Censored likelihood working

region three (i.e.,  $k_1 \leq u_1, k_2 > u_2$ ). It means that the specific point of  $k_1$  is below the threshold  $u_1$ , and the point  $k_2$  exceeds the threshold  $u_2$ . For components that lie above the  $u_1$  and  $u_2$  thresholds, the joint PMF (i.e.,  $P(Y_1 = k_1, Y_2 = k_2)$ ) constitutes the suitable likelihood component. On the other hand, when  $(k_1, k_2) \in (k_1 \leq u_1, k_2 > u_2)$ , there is an information on observed data censoring the marginal  $k_1$  points, but not the  $k_2$  components. Figure 1.4 clearly explains how we censor the data points when working with censored likelihood. Thus the general form of censored likelihood is defined as

$$L(\theta; (k_{11}, k_{21}) \dots (k_{1n}, k_{2n})) = \prod_{i=1}^n P(\theta; (k_{1i}, k_{2i})) \quad (1.43)$$

where  $\theta = (\beta, \alpha)$  is the parameters vector for the CDM model,  $\theta = (\beta, \alpha, \rho)$  is the parameters vector for GM, KM, and TGM, respectively;  $\theta = (\beta, \alpha, \phi)$  is the parameters vector for CCDM and  $\theta = (\beta, \alpha, \phi, \rho)$  is the parameter vector for CGM and CKM and CTGM, respectively.

## 1.5 Simulation study

A simulation study is designed in order to assess the empirical properties and/or performance of our proposals given in section 1.2. Under the proposed models of section 1.2, we will examine the quality of the estimated parameters obtained using the censored likelihood approach.

### 1.5.1 Bivariate and copula-based bivariate models

A simulation study evaluates the censored likelihood estimator for bivariate and copula-based bivariate proposals. More precisely, we generated  $n = 20,000$  sample data point for variable  $Y_1$  and  $Y_2$  from CDM (as described in proposition 1.1), GM, KM and TGM using  $(\alpha = 25, \beta = 200, \rho = 0.40)$ ,  $(\alpha = 25, \beta = 200, \rho = 0.90)$ ,  $(\alpha = 30, \beta = 200, \rho = 0.40)$  and  $(\alpha = 30, \beta = 200, \rho = 0.95)$  with threshold  $u$  equal to the 0.90 order quantile. In the case of copula-based models, the Gamma random variables  $\Lambda_1$  and  $\Lambda_2$  are simulated from the hierarchical setting of GM and KM, and TGM. By using the following parameter scheme  $(\alpha = 25, \beta = 200, \phi = 0.4, \rho = 0.40)$ ,  $(\alpha = 25, \beta = 200, \phi = 0.8, \rho = 0.90)$ ,  $(\alpha = 30, \beta = 200, \phi = 0.4, \rho = 0.40)$  and  $(\alpha = 30, \beta = 200, \phi = 0.8, \rho = 0.95)$ , the variable  $Y_1$  and  $Y_2$  with incorporation of  $\Lambda_1$  and  $\Lambda_2$  are simulated by following below given steps.

**Step 1:** Simulate random pair  $(v_1, v_2)$  from independent uniform random variables, that is  $V_1 \sim U(0, 1)$  and  $V_2 \sim U(0, 1)$ .

**Step 2:** Set  $w_1 = v_1$  and  $Y_1|\Lambda_1 = F_{Y_1|\Lambda_1}^{-1}(w_1)$ , where  $F_{Y_1|\Lambda_1}^{-1}$  denotes the quantile function of Geometric distribution with parameter  $q_1 = 1 - e^{-\Lambda_1/\beta}$ , i.e.,  $Y_1|\Lambda_1 = \lceil \frac{\ln(1-w_1)}{\ln(1-q_1)} - 1 \rceil$ , with  $\lceil \cdot \rceil$  indicating the ceiling function.

**Step 3:** Set  $w_2 = \frac{2v_2}{(a+b)}$ , where  $a = 1 + \phi(1 - p_{k_1|\lambda_1}(k_1|\lambda_1) + 2F_{k_1|\lambda_1}(k_1|\lambda_1 - 1))$  and  $b = [a^2 - 4(a-1)v_2]$ . Then,  $Y_2|\Lambda_2 = \lceil \frac{\ln(1-w_2)}{\ln(1-q_2)} - 1 \rceil$ , where  $q_2 = 1 - e^{-\Lambda_2/\beta}$  and  $\phi$  is the copula dependence parameter.

The CDM, GM, KM, TGM, CCDM, CGM, CKM, and CTGM are fitted to the exceedances using the censored likelihood approach explained in section 1.4. Root mean square errors (RMSEs) and bias of the estimates are obtained by repeating the fitting process  $10^3$  times. Table 1.1 and Table 1.2 show the RMSEs and bias of parameter estimates for both bivariate and copula-based bivariate models, respectively.

During the simulation study, we observe that the scale parameter  $\beta$  is estimated correctly for every model in each parameter scheme. Therefore, Tables do not report the RMSEs and bias of the  $\beta$  parameter. Conversely, the RMSEs and bias are smaller for smaller values of  $\alpha$  with a combination of smaller  $\rho$  values. RMSEs and bias for  $\alpha$

slightly increase when parameter combination  $\alpha$  is high with smaller  $\rho$ . An improvement is observed in all models when both  $\alpha$  and  $\rho$  increase. Furthermore, the RMSEs and bias of  $\alpha$  and  $\phi$  in both the CDM and CCDM are quite reasonable. Since both CDM and CCDM use completely dependent Gamma variables with  $\rho = 1$ ; therefore, the  $\rho$  parameter results are not shown for both cases. Overall, all models exhibit reasonable statistical and computational efficiency across the parameter setting.

## 1.6 Real Applications

This section discusses the real-data application of the proposals made in section 1.2. Bivariate and copula-based bivariate models are applied to daily avalanches counts at two different massifs of the French Alps.

The Enquête Permanente sur les Avalanches (EPA) collected avalanches data from the French Alps, which has monitored about 3900 paths since the early 20th century (see, for instance, Mougin (1922); Evin *et al.* (2021)). Quantitative (run-out elevations, deposit volumes, etc.) and qualitative (flow regime, snow quality, etc.) information is collected for each event. It varies in quality from time to time, depending on the local observers (mostly forestry rangers). Natural avalanche activity is also uncertain because records tend to record paths visible from valleys, so that high-elevation activity may be underestimated.

This application uses the data of avalanches activity at two massifs (namely Haute-Maurienne and Maurienne) by the daily number of avalanche events recorded from 1958 to 2021 in the EPA. The daily number of avalanche events at Haute-Maurienne and Maurienne massifs are extracted from the whole data set, which was used by Evin *et al.* (2021) and Dkengne *et al.* (2016) as shown in Figure 1.5.

Some days after the release of the avalanche, an avalanche event is recorded. By this reason, this may lead to an approximate estimation of the exact day of the event estimated by the observers. It is too restrictive to select the events for which the date is known (too many events are lost) Dkengne *et al.* (2016). However, suppose the avalanche happened many days before the observation. In that case, the insertion of that day could lead to a bias in analysis (because of the wrong number of avalanches for that day and the difficulty concerning this event to meteorological and snow conditions). As a compromise, Dkengne *et al.* (2016) and Evin *et al.* (2021) only consider avalanche events that occurred within the previous three days of the observation. Therefore, this application uses aggregates of daily avalanche events in Haute-Maurienne and Maurienne regions with the same dates at massif scale for the 64 winter seasons around 1958 to

TABLE 1.1: Bias and RMSE for parameter  $\alpha$  and  $\rho$  acquired from simulations of the models based on CDM, GM, KM and TGM, with  $n= 20,000$ , and threshold  $u$  equal to the 0.90 quantile and under different parameter configurations.

True	$\alpha$				True	$\rho$			
	CDM	GM	KM	TGM		CDM	GM	KM	TGM
25	3.92 (0.46)	5.51 (0.97)	5.26 (0.91)	5.28 (0.85)	0.40	-	0.26 (-0.00)	0.23 (0.18)	0.25 (0.01)
25	-	4.31 (0.11)	4.86 (0.55)	4.29 (0.16)	0.90	-	0.23 (-0.08)	0.20 (-0.06)	0.20 (-0.06)
30	(5.74) 0.65	7.89 (1.18)	7.68 (1.40)	8.56 (1.54)	0.40	-	0.20 (0.012)	0.30 (0.025)	0.30 (0.02)
30	-	5.00 (0.08)	6.31 (0.21)	4.86 (0.05)	0.95	-	0.21 (-0.09)	0.23 (-0.10)	0.20 (-0.08)

TABLE 1.2: Bias and RMSE for parameter  $\alpha, \phi$  and  $\rho$  acquired from simulations of the models based on CCDM, CGM, CKM, and CTGM, with  $n= 20,000$ , and threshold  $u$  equal to the 0.90 quantile and under different parameter configurations.

True	$\alpha$				True	$\phi$				True	$\rho$			
	CCDM	CGM	CKM	CTGM		CCDM	CGM	CKM	CTGM		CCDM	CGM	CKM	CTGM
25	4.07 (0.25)	4.43 (0.23)	4.24 (0.37)	4.46 (0.46)	0.40	0.10 0.00	0.15 (-0.01)	0.15 (-0.01)	0.16 (-0.01)	0.40	-	0.36 (0.04)	0.36 (0.03)	0.37 (0.04)
25	3.70 (0.35)	3.84 (0.18)	4.12 (0.34)	3.64 (0.01)	0.80	0.10 (0.00)	0.13 (0.04)	0.12 (0.03)	0.13 (0.04)	0.90	-	0.28 (-0.12)	0.25 (-0.09)	0.27 (-0.12)
30	6.86 (0.77)	6.30 (0.94)	6.16 (0.43)	5.97 (0.57)	0.40	0.09 (-0.01)	0.15 (-0.00)	0.15 (-0.01)	0.15 (-0.02)	0.40	-	0.39 (0.05)	0.39 (0.06)	0.39 (0.06)
30	6.25 (0.52)	6.50 (0.29)	5.71 (0.50)	6.13 (0.60)	0.80	0.10 (-0.01)	0.13 (0.05)	0.12 (0.05)	0.13 (0.05)	0.95	-	0.29 (-0.15)	0.28 (-0.14)	0.30 (-0.15)

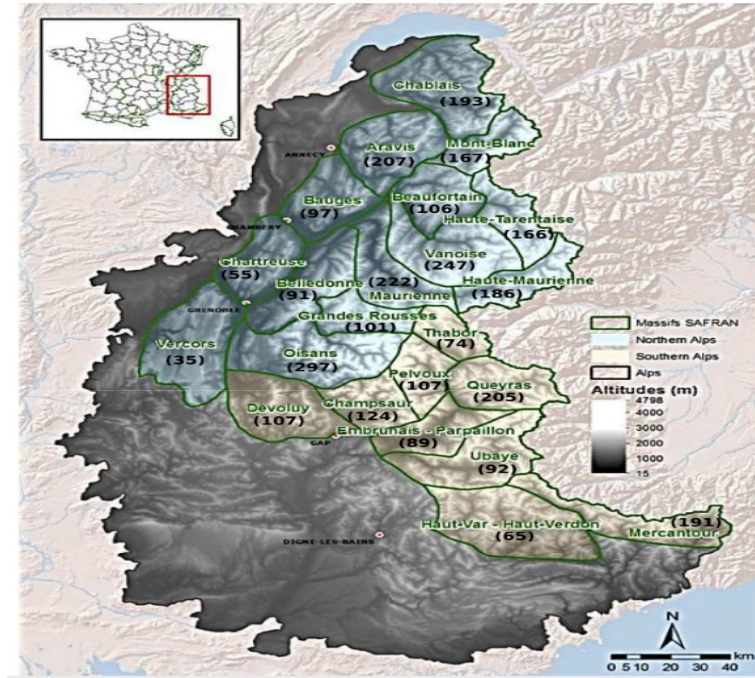


FIGURE 1.5: The 23 SAFRAN massifs of the French Alps. A number between brackets indicates the mentioned snow avalanches paths in each massif (Evin *et al.*, 2021).

2021. Further, 186 and 222 paths are covered by both massifs, as shown in Figure 1.5. In addition, the length of observations at both massifs is 427, with the same collection dates. Figure 1.6 shows the scatter plot of the Haute-Maurienne and Maurienne massifs avalanches counts.

The CDM, GM, KM, TGM, CCDM, CGM, CKM, and CTGM are applied to the bivariate integer-valued data with the threshold  $u_1 = 5$  and  $u_2 = 6$  corresponding to 0.80 quantiles of Haute-Maurienne and Maurienne data, respectively. Results of the fitted models with their AIC and BIC are shown in Table 1.3. Estimates of parameter  $\beta$  and  $\alpha$  are similar across the simple and copula-based models. The estimated  $\rho$  parameter deviates in both hierarchical and copula-based hierarchical models. In both cases, KM and TGM lead to a larger estimate. But, the copula dependence parameter  $\phi$  is over-estimated in copula-based hierarchical models as compared to CCDM. Also, the estimated  $\rho$  reduces in CGM, CKM, and CTGM as compared to GM, KM and TGM. This may happen by the reason that because we are introducing dependence in these models from bivariate Geometric distribution and Gamma random variables as well (theoretical details are provided subsection in 1.2.4 of the chapter). In proposed models, the  $\phi$  and  $\rho$  are basically related to dependence structure. In the fitting at finite thresholds, the dependence induced by GM and CGM is stronger as compared to

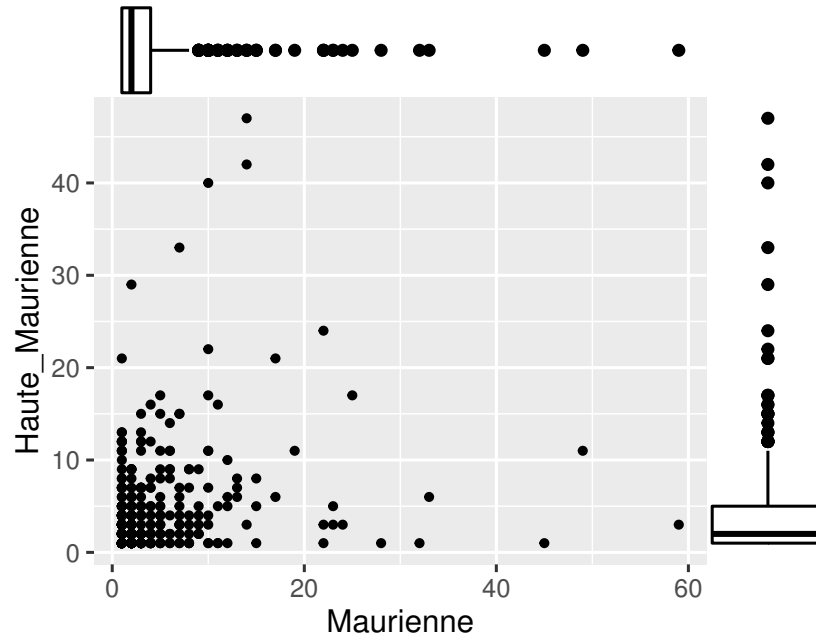


FIGURE 1.6: Number of avalanche events in Haute-Maurienne and Maurienne massifs of the French Alps.

other models. The larger value of  $\rho$  in both simple and copula-based models compensates for that tendency. The AIC and BIC indicate a clear performance in both simple and copula-based models to GM and CGM, respectively. This may tend to extremal dependence in Haute-Maurienne and Maurienne massifs avalanche at the extreme level.

TABLE 1.3: Estimates of model parameters with their standard errors (in parenthesis) of CDM GM, KM, TGM, CCDM, CGM, CKM, and CTGM for Avalanches data of french Alps.

Model	$\beta$	$\alpha$	$\phi$	$\rho$	LL	AIC	BIC
CDM	9.19 (2.35)	3.16 (0.59)	-	-	868.50	1741.01	1749.12
GM	8.89 (2.36)	3.09 (0.59)	-	0.90 (0.23)	868.41	1742.83	1755.00
KM	9.18 (2.38)	3.16 (0.60)	-	0.95 (0.13)	868.50	1743.01	1755.18
TGM	9.19 (2.39)	3.16 (0.60)	-	0.97 (0.16)	868.50	1743.00	1755.18
CCDM	10.34 (3.00)	3.42 (0.74)	0.39 (0.38)	-	867.98	1741.97	1760.19
CGM	9.83 (2.73)	3.31 (0.67)	0.91 (0.67)	0.53 (0.27)	865.92	1739.85	1756.08
CKM	10.55 (2.92)	3.48 (0.71)	0.92 (0.81)	0.68 (0.35)	866.89	1741.78	1758.00
CTGM	10.57 (2.96)	3.49 (0.72)	0.92 (0.86)	0.65 (0.39)	866.99	1741.99	1758.22

For the diagnostics of the fitted models, we inspected the behavior of the summary statistics that depend only on the characteristics of variable  $Y_1$  and  $Y_2$ : the conditional probabilities  $P(Y_1 > k | Y_2 > k)$  and the average cluster size of the discrete exceedances

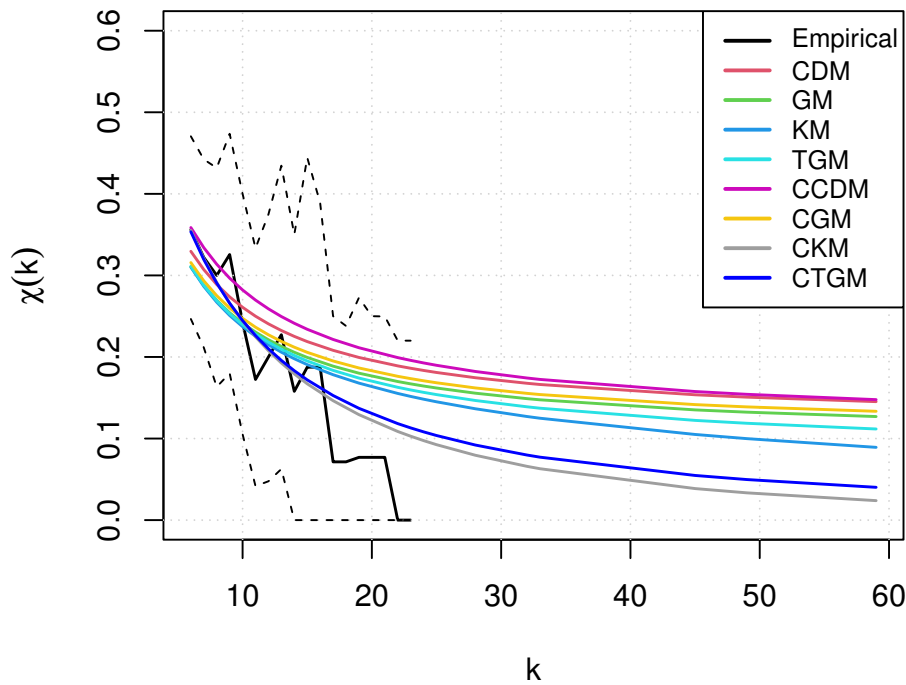


FIGURE 1.7: Empirical and model-based estimate of  $\chi(k) = P(Y_1 > k | Y_2 > k)$  for the counts of avalanches at Haute-Maurienne and Maurienne massifs of French Alps. The dashed lines give 0.95 bootstrap confidence bands

$k$ , for  $k > u$ . We adopted conditional probabilities to measure the strength of local extremal dependence between Haute-Maurienne and Maurienne massifs avalanches. Further, the conditional probabilities permit evaluation of the model for the short-term prediction abilities or inadequacies. In practical terms, the average cluster size summarizes the tendency for extreme occurrences. To evaluate the quality of the extrapolation over a fixed threshold  $u$ , the summary measures were analyzed as a function of exceedances  $k$ , with  $k > u$ . Asymptotic properties of the measure based on conditional probabilities are discussed in section 1.3 that joint variables are asymptotically independent when the conditional probabilities converge to zero, as  $k \rightarrow \infty$ . Figure 1.7 compares model-based and empirical estimates of the tail dependence measure (developed in section 1.3 of the chapter) based on conditional probabilities for the estimated models, namely CDM, GM, KM, TCM, CCDM, CGM, CKM, and CTGM. A decreasing degree of dependence can be seen in empirical values, and this corresponds to convergence to independence as  $k \rightarrow \infty$ . Values obtained from the CCDM, CGM, CKM, and CTGM match closely with the empirical patterns of the joint tail dependence, while CDM, GM, KM, and TGM slightly overestimate the dependence. In the proposed formulation, the copula-based model is preferable over simple models for the avalanches data of the French Alps; however, all models produce more stable estimates than the



empirical counterparts as  $k$  increases, resulting in the dependence being underestimated (overestimated) for low (high) values of  $k$ .

## 1.7 Final remarks

This thesis chapter proposes a model that allows us to assess the changes in the extremal dependence structure. Hierarchical settings and copula are considered the main tools to introduce different kinds of dependence in the proposed model. This newly developed framework is the first step to modeling bivariate discrete extremes having asymptotic dependence or asymptotic independence at the extreme. Moreover, the theoretical developments of section 1.3 have shown that both asymptotic dependence and asymptotic Independence are attainable in simple and copula-based models with an appropriate choice of Gamma random variables in hierarchical settings. Generally speaking, both scenarios can be tackled with the proposal presented in this chapter.

Furthermore, we perform inference by using the censored likelihood as used by Coles (2001) for continuous bivariate extremes. In general, the implementation of our proposal is straightforward, and inference is computationally convenient. In addition, we tested our models through a simulation study and later applied them to the avalanche count data of two massifs of the French Alps.

The proposed methodology in this chapter could also be applied to integer-valued environmental variables and the count data of seasonal viruses. In addition, regression-type modeling can be implemented for DGPD marginals. Regression-type methodology work by letting the parameters of marginal DGPD vary with covariates.



# Chapter 2

## A time series model for discrete extreme values

### Overview: Chapter 2

↪ This chapter presents a new model for describing temporal dependence in discrete exceedances above a threshold. The modeling framework is executed in two steps. In the first step, discrete exceedances are modeled through DGPD, which can be obtained by mixing a Geometric variable with a Gamma distribution (the main results associated with its hierarchical representation are derived in chapter 1). In the second step, a model for discrete extreme values is built by injecting a latent Gamma process via hierarchical framework, which confirms that the marginal distribution is DGPD, as expected from classical discrete extreme value theory. We employ four distinct underlying stationary Gamma processes, each producing a different temporal dependency structure, either asymptotic independence or asymptotic dependence. The proposed model is applied to two real discrete time series. Observations of both series over a finite threshold have shown asymptotic independent behavior. One can use a novel framework for discrete-time series, which has an asymptotic-dependent behavior over the tail. In both scenarios, our proposed model is more flexible.

Chapter 2 is organized as follows. Section 2.1 provides an introduction to temporal extremes containing a detailed literature review and study gap. Section 2.2 presents the latent process model with hierarchical construction for discrete time-dependent extremes. Asymptotic properties of tail dependence measure are discussed in Section 2.3. Section 2.4 deals with the inferential activities, especially the pairwise likelihood approach. The simulation study is given in Section 2.5 to

assess the model parameters and dependence behavior over the tail. Section 2.6 applies the novel model to two distinct examples of discrete time series having extreme observations (monthly number of police reports on narcotics trafficking and number of tick changes by a minute of EUR/GBP series). Section 2.7 concludes and provides some further recommendations.

## 2.1 Introduction

Classical EVT based on asymptotic extreme value models for block maxima or exceedances over high thresholds plays a crucial role in modeling extremes. Whereas the GEV distribution ascends as the only possible limit model for block maxima, the GPD is its counterpart for exceedances over a high threshold. In the univariate and multivariate framework, block maxima approaches are usually criticized for excluding many relevant observations. Moreover, they evaluate the joint distribution function of extreme values occurring inside a block whose coarse might not be appropriate for analysis. In contrast, exceedances over threshold methods engage all data points for which at least one coordinate exceeds a corresponding high threshold, employing the censored maximum likelihood technique where the likelihood of observation relies on which components surpass the corresponding thresholds coordinates (Dutfoy *et al.*, 2014).

In the context of stochastic modeling, threshold methods have been extensively developed in recent years. In this sense, the main question that appears in mind is how to capture temporal dependence when observations occur in clusters over the tail. For instance, the exceedances over a high threshold can be asymptotically dependent when they occur in clusters; otherwise, they may be asymptotically independent. For a review of modeling time series of extreme values, see Chavez-Demoulin and Davison (2012). Literature suggests two main approaches. Firstly, a Markov chain model is estimated using a likelihood function as part of the threshold approach, which assumes that the extreme model is only fitted to observations that exceed the threshold while the others are censored (Smith *et al.*, 1997; Bortot and Tawn, 1998). The transition kernel of the Markov chain can be specified using standard bivariate extremal models. Near-independence models can also be specified (Bortot and Tawn, 1998; Ramos and Ledford, 2009; De Carvalho and Ramos, 2012). A second approach considers a hierarchical model. The parameters of a standard extreme-value distribution are driven by a latent stochastic process (Casson and Coles, 1999; Gaetan and Grigoletto, 2004; Huerta and Sansó,

2007; Bortot and Gaetan, 2014; Bopp and Shaby, 2017). According to Bortot and Gaetan (2014), the marginal distributions generally no longer exhibit extreme values, and the induced dependence is not strong enough to allow asymptotically high clustering of extremes.

As discussed in chapter 1 of the thesis, a class of models has been proposed using Gamma latent variables in a continuous setting. Hierarchical models are flexible in their characteristics of serial dependence and result in a process with GPD margins. Similar to Bortot and Gaetan (2014), we want to model serial dependence above the threshold when working with discrete extremes. However, discrete distribution (e.g., DGPD) fulfills the extremal properties, and it remains the favorite for modeling discrete exceedances (Anderson, 1970; Hitz *et al.*, 2017).

The major contribution of this chapter is to propose a new discrete extreme value model that accommodates a variety of extremal dependence characteristics. This modeling framework is linked with a hierarchical approach, ensuring that the marginal distribution of exceedances is DGPD. Therefore, the classical extremal techniques are appreciated, and a wide range of dependence can be considered at the extreme level. We can accomplish this by representing the DGPD as a mixture of Geometric and Gamma random variables (Buddana and Kozubowski, 2014). DGPD is considered a natural candidate for modeling dependence in exceedances; for more details, see chapter 1. Further, we use four distinct stationary Gamma processes to generate a temporal layer with different kinds of dependence behavior. Furthermore, the hierarchical construction-based function was evaluated using the pairwise likelihood method when performing inference. In the case of nonstationary DGPD margins, the effect of the covariates can be taken into account by developing a GAM form modeling paradigm and dealing parameters of the model as a function of covariates.

## 2.2 Discrete time series hierarchical models

This section of the chapter presents a hierarchical model for discrete threshold exceedance time series, where a first-order latent Markov chain with stationary Gamma margins is used to model the extremal dependence. In this hierarchical construction, we are able to account for serial dependence and maintain DGPD margins for exceedances. The model is motivated by the following representation for DGPD: for shape parameter,  $\xi > 0$ , the DGPD given in (1.8) can be constructed as a mixture of Geometric and

Gamma random variables (see, for instance, section 1.2). This show that if

$$\begin{aligned} Y|\Lambda &\sim \text{Geo}(1 - e^{-\Lambda/\beta}) \\ \Lambda|\alpha, \beta &\sim \text{Gamma}(\alpha, \beta) \end{aligned} \tag{2.1}$$

then, by equation (1.10),  $Y \sim \text{DGPD}(\alpha, \beta)$  with PMF given in (1.9). Let  $\{Y_t\}$  be a stationary random sequence of discrete random variables. We are interested in modeling the observations above the fixed high threshold  $u$ , which basically represents the tail behavior. Moreover, we work with EVT and defend the DGPD as marginal distribution of  $\{Y_t, t \geq 1\}$ . From (2.1), firstly, we assume conditionally on  $\{\Lambda_t\}$ , the  $Y_t|\Lambda_t \sim \text{Geo}(1 - \exp[-\Lambda_t/\beta])$ . Secondly, we introduce the temporal dependence by using different stationary Gamma processes for  $\Lambda_t$  having  $\text{Gamma}(\alpha, \beta)$  margins. By compliance with the results of section 1.2, the marginal distribution of  $Y_t$  follows  $\text{DGPD}(\alpha, \beta)$ , concerning marginal  $\Lambda_t$ . In addition, the Gamma distribution parameters are positive; this representation also puts a constraint on the shape parameter as  $\alpha > 0$ , which leads to a heavy tail case of DGPD (Hitz *et al.*, 2017). Therefore, the interesting feature of the above representation is that the bivariate distributions with discrete Pareto-type margins are easily tractable. At this stage, it is much easier to incorporate the temporal dependence layer via hierarchical settings of stochastic processes with  $\text{Gamma}(\alpha, \beta)$  margins for  $\{\Lambda_t\}$ . The following section describes the considered stochastic processes along their hierarchical structure for introducing the temporal dependence layer in the proposed representation.

### 2.2.1 Induction of temporal dependence layer

This section deals with how temporal dependence can be induced in our proposed representation. Interestingly, we introduce temporal dependence in the proposed representation through stationary Gamma process  $\{\Lambda_t\}$ . Thus, we focus on different stationary Gamma processes, which are more flexible with Markov chains and have recursive forms which may lead to different extremal dependence structures. After inducing the temporal layer via  $\{\Lambda_t\}$ , the original variable  $\{Y_t\}$  may have non-Markovian nature model (Bortot and Gaetan, 2014). We use the following stationary Gamma processes as

**1. *Gaver and Lewis process (GLP)*:** This class of process is discussed by Gaver

and Lewis (1980); Walker (2000). The GLP is defined as

$$\begin{aligned}
\Lambda_{t-1} &\sim \text{Gamma}(\alpha, \beta) \\
P_t &\sim \text{Gamma}(\alpha, 1) \\
X_t|P_t &\sim \text{Poisson}\left(P_t \frac{(1-\rho)}{\rho}\right) \\
V_t|X_t &\sim \text{Gamma}\left(X_t, \frac{\beta}{\rho}\right) \\
\Lambda_t &= \rho\Lambda_{t-1} + V_t
\end{aligned} \tag{2.2}$$

with  $\Lambda_{t-1}$  independent of  $P_t$ ,  $X_t|P_t$  and  $V_t|X_t$ .  $\rho$  is the dependence parameter which is the correlation between  $\Lambda_{t-1}$  and  $\Lambda_t$ , it ranges  $0 \leq \rho < 1$ . Given  $\{\Lambda_t\}$  is stationary Gamma process whose marginal distribution is  $\text{Gamma}(\alpha, \beta)$ .

**2. Warren process (WP):** The WP was introduced by Warren (1992), and is defined as follows. Let

$$\begin{aligned}
\Lambda_{t-1} &\sim \text{Gamma}(\alpha, \beta) \\
X_t|\Lambda_{t-1} &\sim \text{Poisson}\left(\frac{\rho\Lambda_{t-1}\beta}{1-\rho}\right) \\
\Lambda_t|X_t &\sim \text{Gamma}\left(X_t + \alpha, \frac{1-\rho}{\beta}\right)
\end{aligned} \tag{2.3}$$

The resulting  $\{\Lambda_t\}$  again stationary Markov Gamma process with recursive form and follow  $\text{Gamma}(\alpha, \beta)$  margins with dependence parameter  $\rho$ , it ranges  $0 \leq \rho < 1$ .

**3. Thinned Gamma process (TGP):** This class of process was introduced in Wolpert (2021) by using the thinning layer generated from Beta distribution. The general form of TGP is defined as

$$\begin{aligned}
\Lambda_{t-1} &\sim \text{Gamma}(\alpha, \beta) \\
B_t &\sim \text{Beta}(\alpha\rho, \alpha(1-\rho)) \\
V_t &\sim \text{Gamma}(\alpha(1-\rho), \beta) \\
\Lambda_t &= B_t\Lambda_{t-1} + V_t
\end{aligned} \tag{2.4}$$

where  $\{\Lambda_{t-1}\}$  is independent of  $B_t$  and  $V_t$ . The  $\{\Lambda_t\}$  is a Markov process with gamma univariate marginal distribution  $\text{Gamma}(\alpha, \beta)$  with auto-correlation  $\rho$ , it ranges  $0 \leq \rho < 1$ . The process of passing from  $\{\Lambda_t\}$  to  $B_t\Lambda_{t-1}$  is called thinning (Wolpert, 2021). Hence, we will refer  $\{\Lambda_t\}$  as thinned gamma process.

**4. The Markov change-point process (MCPP):** Let  $\{\xi_n : n \in \mathbb{Z}\} \stackrel{\text{iid}}{\sim} \text{Gamma}(\alpha, \beta)$  be independent and identically distributed Gamma random variables and let  $P_t$  be a

standard Poisson process index by  $t \in \mathbb{R}$  ( so  $P_0 = 0$  and  $(P_t - P_{s=t-1}), \forall -\infty < s < t < \infty$ , independent increments), and let

$$\Lambda_t = \xi_n, \quad n = P_{ut}, \quad (2.5)$$

then each  $\Lambda_t \sim \text{Gamma}(\alpha, \beta)$  and, for  $t-1, t \in \mathbb{R}$ ,  $\{\Lambda_{t-1}\}$  and  $\{\Lambda_t\}$  are either identical or independent reminiscent of Metropolis Markov Chain Monte Carlo (for more details we refer to Wolpert (2021)). Once again  $\{\Lambda_t\}$  have  $\text{Gamma}(\alpha, \beta)$  marginal distribution. The dependence among exceedances is controlled through  $\rho$ , with the larger value of  $\rho$  tending to have a stronger level of dependence. It is worth mentioning that all considered processes have the same correlation function, that is

$$\text{Corr}(\Lambda_t, \Lambda_{t+j}) = \rho^{|j|}$$

We will work with LTs of  $\Lambda_t$  and  $\Lambda_{t+j}$  and defined as

$$L_j^{(2)}(s_1, s_2) \Big|_{(s_1 = \frac{k_1}{\beta}, s_2 = \frac{k_2}{\beta})} = E \left( e^{-\frac{\Lambda_t k_1}{\beta} - \frac{\Lambda_{t+j} k_2}{\beta}} \right), \quad (2.6)$$

The LTs corresponding to GLP, WP and TGP are already defined in (1.29),(1.30) and (1.31), respectively. We need to replace  $\rho$  as  $\rho^j$  when working with time series. In addition, the LT associated with MCPP is given by

$$L_j^{(2)}(s_1, s_2) \Big|_{(s_1 = \frac{k_1}{\beta}, s_2 = \frac{k_2}{\beta})} = \left[ \frac{\rho^j \beta^\alpha}{(\beta + k_1 + k_2)^\alpha} + \frac{(1 - \rho^j) \beta^{2\alpha}}{(\beta + k_1)^\alpha (\beta + k_2)^\alpha} \right], \quad (2.7)$$

$0 < \rho < 1$ , for MCPP, respectively.

## 2.3 Extremogram for tail behavior

There is a common measure of dependence in extremes called extremogram, which describes the conditional probability that one random variable will be extreme when the other is extreme (Davis and Mikosch, 2009; Chavez-Demoulin and Davison, 2012). For a strictly stationary  $\mathbb{N}_0^d$  integer-valued time series  $(Y_t)$ , by following Davis *et al.* (2012) the extremogram is defined for two sets  $A$  and  $B$  bounded away from zero by following as

$$\rho_{A,B}(h) = \lim_{k \rightarrow \infty} P(k^{-1}Y_h \in B | k^{-1}Y_0 \in A), \quad h = 0, 1, 2, \dots \quad (2.8)$$



with the given limit exists. Since  $A$  and  $B$  are bounded away from zero, the events  $k^{-1}Y_0 \in A$  and  $k^{-1}Y_h \in B$  are becoming extreme in a sense the probabilities of these events are converging to zero with  $k \rightarrow \infty$  (Davis *et al.*, 2012). For the special choice in the  $d = 1$  case of  $A = B = (1, \infty)$ , the extremogram reduces to the tail dependence coefficient ( $\chi$ ) between  $Y_0$  and  $Y_h$  that is often used in EVT and quantitative risk management (McNeil *et al.*, 2015). The  $\chi$  measure is discussed in detail in section 1.3 of chapter 1. The variables  $Y_0$  and  $Y_h$  are said to be asymptotically independent when extremal measure  $\chi = 0$  and dependent otherwise.

Another standard measure called the extremal index  $\theta \in (0, 1)$  is developed in the literature to deal with temporal dependence of a stationary process at asymptotically high levels that quantify the tendency of extreme values to cluster Leadbetter *et al.* (1983). Under certain conditions, the extremal index is the reciprocal of the limiting mean size of clusters of exceedances as the threshold goes to the upper endpoint of the univariate marginal distribution (Bortot and Gaetan, 2014). For  $\theta = 1$ , the tail behavior of the series  $(Y_t)$  likes to be asymptotically independent and identically distributed (no clusters of exceedances). Conversely, when  $\theta < 1$ , the series reveals temporal dependence, even at asymptotically high levels (clusters of exceedances). Further, In the case of discrete data, Scotto *et al.* (2018) analyzed the properties of the extremal index by working with the maximum First-order integer-valued autoregressive (max-INAR(1)) process, and they got closed-form results for the extremal index. They prove that the max-INAR(1) process in case of heavy tail holds the conditions defined by Leadbetter *et al.* (1983). In the current scenario, we will work with the tail dependence measure.

The model considered in section 2.2 for  $\{\Lambda_t\}$  leads to different dependence structures among  $Y_t$ . In response, GLP and MCPP induce asymptotic dependence for all  $0 < \rho \leq 1$ . The theoretical results of  $\chi$  corresponding to GLP and MCPP are given  $[\rho^j/(1+\rho^j)]^\alpha$  and  $[\rho^j/2^\alpha]$ , respectively. The non-zero  $\chi$  indicates that the GLP and MCPP models have asymptotic-dependent behavior. On the other hand, WP and TGP induce asymptotic Independence in  $Y_t$ . The theoretical result of the  $\chi$  measure corresponding to WP and TGP converges to zero and clearly shows asymptotic independent behavior. Moreover, WP and GLP for  $\Lambda_t$  lead to different extremal dependence characteristics, including asymptotic dependence Furthermore, the TGM and MCPP models also have different extremal dependence characteristics. When one uses the extremal index  $\theta$ , the model use from choosing  $\{\Delta_t\}$  as in (2.2) and (2.5) has extremal index  $\theta < 1$ , so that, the exceedances occur in clusters, whereas the model obtained by choosing (2.3) and (2.4) has  $\theta = 1$  and extremes are asymptotically independent. Related proofs can be found in Bortot and Gaetan (2014). The simulation-based exploratory analysis associated with

$\chi$  supports the above examples provided in section 2.5.

## 2.4 Pairwise likelihood

Let  $k_t, t = 1, \dots, n$  is the observed sequence of discrete nature, the likelihood inference for hierarchical models requires approximating the n-fold integral, that is

$$L_n(\theta) = \int \left[ \prod_{t=1}^n \{(1 - \exp(-\lambda_t/\beta)) \exp(-\lambda_t k_t/\beta)\} f(\lambda_1, \dots, \lambda_n; \alpha, \beta, \rho) \right] d\lambda_1, \dots, d\lambda_n \quad (2.9)$$

where  $\theta = (\alpha, \beta, \rho)$  and  $f(\lambda_1, \dots, \lambda_n; \alpha, \beta, \rho)$  is the joint density function of  $\Lambda_1, \dots, \Lambda_n$ . The formula (2.9) may proceed further via the filtering algorithm. In addition to its drawbacks, the filtering algorithm propagates numerical errors through nested integrals (Pedeli and Varin, 2020). In light of this, evaluating  $L_n(\theta)$  is not feasible because of the complex integral involved in (2.9). The following pairwise log-likelihood (PL), which is an example of composite likelihood (Lindsay, 1988) replaces the full likelihood in this situation:

$$pl_n(\theta) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{\min(i+\Delta, n)} \log P(k_i, k_j; \theta) \quad (2.10)$$

where  $P(k_i, k_j)$  is the joint PMF of  $(Y_i, Y_j)$  and  $1 \leq \Delta \leq n - 1$  is the constant which defines the maximum lag. We will compute the pairwise likelihood for the  $\Delta$  order using all pairs of observations with lag distances. When compared to the ordinary likelihood, the pairwise likelihood offers a significant reduction in computational cost. Moreover, when the pairs  $(k_i, k_j)$  are treated as independent, the PL viewed an example of composite likelihood (Lindsay, 1988).

The useful feature of our proposed models is that the PL evaluation is much easier. To this end, each PMF in (2.10) can be written in LTs of  $\{\Lambda_t\}$  process. It is easy to construct the joint PMF when comparing the pair with suitable high threshold  $u$ , weathering the pair  $(k_i, k_j)$  lie above or below the  $u$ . That is,

- for  $k_i > u, k_j > u$ .

$$\begin{aligned} P(k_i, k_j; \theta) &= L_{j-i}^{(2)}(s_1, s_2) \Big|_{(s_1=\frac{k_1}{\beta}, s_2=\frac{k_2}{\beta})} - L_{j-i}^{(2)}(s_1, s_2) \Big|_{(s_1=\frac{k_1+1}{\beta}, s_2=\frac{k_2}{\beta})} \\ &\quad - L_{j-i}^{(2)}(s_1, s_2) \Big|_{(s_1=\frac{k_1}{\beta}, s_2=\frac{k_2+1}{\beta})} + L_{j-i}^{(2)}(s_1, s_2) \Big|_{(s_1=\frac{k_1+1}{\beta}, s_2=\frac{k_2+1}{\beta})} \end{aligned}$$

- for  $k_i > u, k_j \leq u$ .

$$P(k_i, k_j; \theta) = L^{(1)}(s)|_{(s=\frac{k_1}{\beta})} - L^{(1)}(s)|_{(s=\frac{k_1+1}{\beta})} - L_{j-i}^{(2)}(s_1, s_2)|_{(s_1=\frac{k_1}{\beta}, s_2=\frac{u+1}{\beta})} \\ + L_{j-i}^{(2)}(s_1, s_2)|_{(s_1=\frac{k_1+1}{\beta}, s_2=\frac{u+1}{\beta})}$$

- for  $k_i \leq u, k_j > u$ .

$$P(k_i, k_j; \theta) = L^{(1)}(s)|_{(s=\frac{k_2}{\beta})} - L^{(1)}(s)|_{(s=\frac{k_2+1}{\beta})} - L_{j-i}^{(2)}(s_1, s_2)|_{(s_1=\frac{u+1}{\beta}, s_2=\frac{k_2}{\beta})} \\ + L_{j-i}^{(2)}(s_1, s_2)|_{(s_1=\frac{u+1}{\beta}, s_2=\frac{k_2+1}{\beta})}$$

- for  $k_i \leq u, k_j \leq u$ .

$$P(k_i, k_j; \theta) = 1 - L^{(1)}(s)|_{(s=\frac{u+1}{\beta})} - L^{(1)}(s)|_{(s=\frac{u+1}{\beta})} + L_{j-i}^{(2)}(s_1, s_2)|_{(s_1=\frac{u+1}{\beta}, s_2=\frac{u+1}{\beta})}$$

In each case  $L_{j-i}^{(2)}(\cdot, \cdot)$  is Laplace transform of  $\Lambda_i$  and  $\Lambda_j$ ,  $L^{(1)}(\cdot)$  is the univariate Laplace transform of  $\Lambda$ .

Under suitable regularity conditions, the  $\Delta$  order maximum PL estimator  $\hat{\theta}_{mpl}$  is consistent and asymptotically normal distributed with asymptotic mean and variance, that is

$$\hat{\theta}_n \sim N(\theta, G_n(\theta)^{-1} = H_n(\theta)^{-1} J_n(\theta) H_n(\theta)^{-1}) \quad (2.11)$$

where

$$H_n(\theta) = E[-\nabla^2 pl_n(\theta)] = E \left\{ - \sum_{i=1}^{n-1} \sum_{j=i+1}^{\min(i+\Delta, n)} \nabla^2 \log P(k_i, k_j; \theta) \right\}$$

and

$$J_n(\theta) = Var[\nabla pl_n(\theta)] = Var \left\{ \sum_{i=1}^{n-1} \sum_{j=i+1}^{\min(i+\Delta, n)} \nabla \log P(k_i, k_j; \theta) \right\}$$

The matrices  $H_n$  and  $J_n$  must be estimated consistently to evaluate standard errors. One can recover by using estimates  $\hat{H}_n = H_n(\hat{\theta}_n)$  and  $\hat{J}_n = J_n(\hat{\theta}_n)$ . As a result, the direct calculation of  $\hat{J}_n$  is difficult as  $O(n^4)$  number of calculations required. As an alternative, we estimate  $H_n(\theta)$  by  $H_n(\hat{\theta}) = -\nabla^2 pl_n(\hat{\theta})$  and used subsampling approach (Carlstein, 1986) to estimate  $J_n$ . The subsampling procedure contains  $J_n$  estimates over  $S$  overlapping windows  $W_j \subset \{1, \dots, n\}, j = 1, \dots, S$ , of size  $w_j$  through the following expression

$$J_n = \frac{d_n}{S} \sum_{j=1}^S \frac{\nabla pl_{W_j}(\hat{\theta}) \nabla pl_{W_j}(\hat{\theta})'}{w_j},$$

where  $d_n = \Delta(n - 1)$  and  $\nabla pl_{W_j}$  represent the pairwise score which is evaluated over the window  $W_j$ . The asymptotic variance can be estimated as  $V_n = H_n^{-1} J_n H_n^{-1}$ .

In our setup, selection among four processes is essential because all four lead to different tail behaviors and extrapolations. For this task, we followed Varin and Vidoni (2005) and adopted pairwise likelihood information criterion (PLIC) for model selection. By using PLIC, we determine a model which minimizes

$$PLIC(\hat{\theta}) = -\log pl_n(\hat{\theta}) + tr(J_n H_n^{-1}) \quad (2.12)$$

where  $J_n$  and  $H_n$  are estimated in accordance of the above discussion.

## 2.5 Simulation study

To evaluate the performance of PL estimator for the proposed models, we carried out a simulation study. A series  $n = 30,000$  was simulated from each of GLP, WP, TGP, and MCPP models using the following parameter settings  $(\alpha = 5, \beta = 50, \rho = 0.40)$ ,  $(\alpha = 5, \beta = 50, \rho = 0.90)$ ,  $(\alpha = 10, \beta = 50, \rho = 0.40)$  and  $(\alpha = 10, \beta = 50, \rho = 0.95)$  with threshold  $u$  equal to the 0.90 quantile. The choice of lag  $\Delta$  is necessary for PL. An increase in lag  $\Delta$  may increase the computational burden, but findings in the existing literature indicate that an improvement in estimation precision will not necessarily be gained (see, for instance, Bortot and Gaetan (2014)). To this end, the GLP, WP, TGP, and MCPP models were fitted with support of PL using lag  $\Delta \in (1, 2, 3, 5, 7)$ . To estimate the root mean square errors (RMSE) and bias of PL estimates for each proposed model, the fitting process is repeated  $10^3$  times.

For every value of  $\Delta$ , we observe that the parameter  $\beta$  is always estimated correctly. On the other hand, RMSEs for  $\alpha$  and  $\rho$  are smaller when true value of  $\alpha$  is comparatively small. It is astonishing to announce that the estimation precision for  $\alpha$  and  $\rho$  parameters is reasonable for a smaller value of  $\alpha$ , irrespective of choice  $\Delta$ . Simulation-based RMSEs and bias of the parameters  $\alpha$  and  $\rho$  are reported in Table 2.1. A minor improvement is noted when moving from  $\Delta = 1$  to  $\Delta = 3$ , while a further increase in  $\Delta$  appears to provide no benefit.

On the other side, it is difficult to generalize a particular pattern when  $\alpha$  is higher. RMSEs for all models seem higher for larger  $\alpha$  and smaller  $\rho$ . Moreover, it can be observed from Table 2.1 that the higher lag is advantageous only in some cases. Therefore,  $\Delta = 3$  is considered a more reasonable choice, which provides statistical and computational efficiency for the proposed models across each configuration of the parameters.

TABLE 2.1: Root mean squared error and bias for  $\alpha$  and  $\rho$  found from simulations of the models based on GLP, WP, TGP, and MCPP, with  $n=30,000, (u = 0.90)$  quantile and under different parameter settings.

lag	True value	$\alpha$				True value	$\rho$			
		GLP	WP	TGP	MCPP		GLP	WP	TGP	MCPP
1	5	0.27	0.30	0.28	0.28	0.40	0.05	0.06	0.06	0.05
		(0.002)	(0.02)	(0.02)	(0.02)		(0.00)	(0.00)	(-0.003)	(0.00)
	5	0.32	0.28	0.30	0.38	0.90	0.06	0.04	0.04	0.06
		(0.02)	(0.01)	(0.02)	(0.03)		(-0.002)	(-0.001)	(0.00)	(-0.002)
3	10	0.97	0.99	0.97	0.99	0.40	0.10	0.10	0.10	0.10
		(0.09)	(0.08)	(0.11)	(0.12)		(0.007)	(-0.001)	(-0.001)	(0.003)
	10	1.01	0.97	0.98	1.09	0.95	0.09	0.07	0.07	0.09
		(0.04)	(0.10)	(0.02)	(0.02)		(-0.03)	(-0.02)	(-0.02)	(-0.03)
7	5	0.27	0.25	0.28	0.28	0.40	0.04	0.05	0.05	0.04
		(0.006)	(0.02)	(0.02)	(0.02)		(-0.000)	(-0.001)	(-0.005)	(-0.002)
	5	0.32	0.28	0.30	0.38	0.90	0.03	0.02	0.02	0.03
		(0.02)	(0.001)	(0.02)	(0.03)		(-0.002)	(-0.002)	(-0.001)	(-0.001)
10	10	0.97	0.99	0.97	0.99	0.40	0.07	0.08	0.08	0.08
		(0.09)	(0.08)	(0.11)	(0.12)		(0.001)	(-0.007)	(-0.001)	(-0.002)
	10	1.04	0.99	1.02	1.12	0.95	0.04	0.03	0.03	0.04
		(0.10)	(0.15)	(0.08)	(0.08)		(-0.005)	(-0.002)	(-0.002)	(-0.005)
10	5	0.27	0.25	0.26	0.28	0.40	0.04	0.05	0.04	0.04
		(0.01)	(0.007)	(0.01)	(0.01)		(-0.001)	(-0.002)	(-0.005)	(0.00)
	5	0.32	0.28	0.30	0.38	0.90	0.01	0.02	0.01	0.01
		(0.02)	(-0.003)	(0.01)	(0.03)		(-0.001)	(-0.002)	(-0.002)	(-0.001)
10	0.97	0.99	0.97	1.00	0.40	0.07	0.08	0.08	0.07	
	(0.09)	(0.09)	(0.11)	(0.12)		(0.000)	(-0.008)	(-0.001)	(-0.003)	
10	1.02	1.00	1.03	1.14	0.95	0.02	0.02	0.02	0.02	
	(0.11)	(0.16)	(0.10)	(0.11)		(-0.001)	(-0.001)	(-0.001)	(-0.001)	

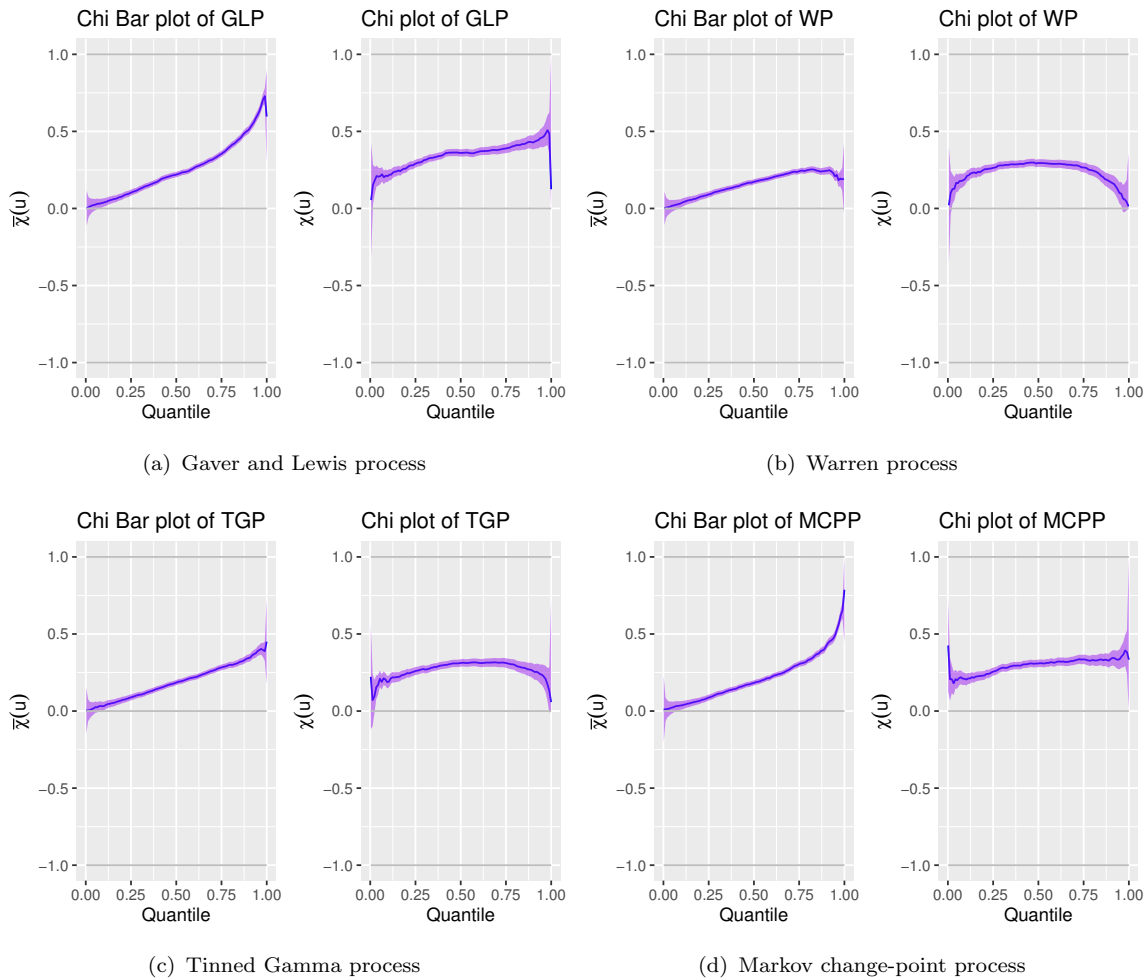


FIGURE 2.1: The dependence measure  $\chi(u)$  and  $\bar{\chi}(u)$  for simulated data from each of **(a)** Gaver and Lewis process; **(b)** Warren process; **(c)** Thinned gamma process, and **(d)** Markov change-point process. Lag  $\Delta = 3$  is used here.

To analyze the behavior of extremal dependence of our proposed models given in section 2.2. Empirical estimates of  $\chi(u)$  and  $\bar{\chi}(u)$  are calculated for the data, which is simulated from each of GLP, WP, TGP, and MCPP with parameters configuration as  $(\alpha = 1, \beta = 50, \rho = 0.90)$ . Estimates can be plotted to determine the limiting behavior as a function of marginal quantiles. The marginal quantile is linked with the corresponding threshold value from the data. Moreover, a graphical representation of dependence measures could facilitate discerning between asymptotically dependent and asymptotically independent relationships. For an illustration, the empirical estimates plots of  $\bar{\chi}(u)$  and  $\chi(u)$  with 95% confidence intervals for the simulated data from our proposed models are displayed in Figure 2.1. It can be observed that the  $\chi(u)$  converges to 0.49 and 0.32 for GLP and MCPP, corresponding to a higher threshold selection quantile. In fact, these values indicate that the observations generated from GLP and MCPP occur in clusters above sufficiently high thresholds. On the other hand,  $\chi(u)$

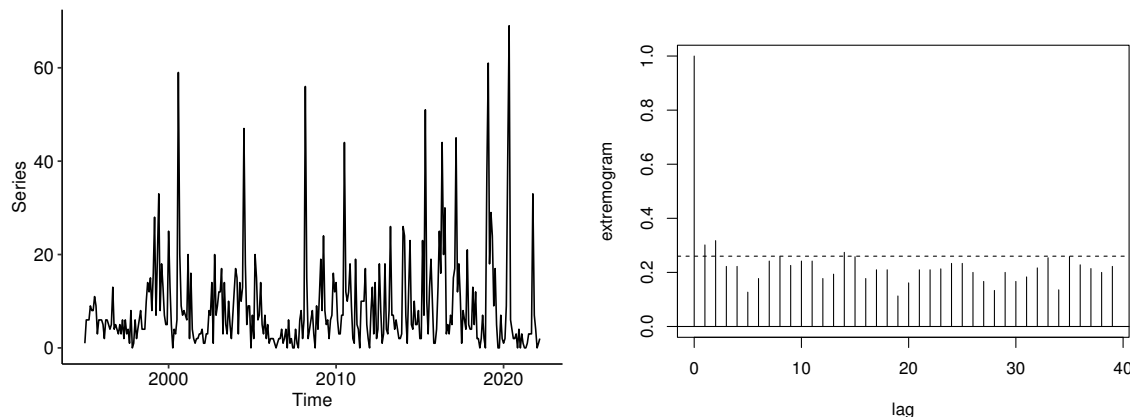


FIGURE 2.2: Monthly based number of police reports on narcotics trafficking in Sydney, Australia, along with sample extremogram.

converges to zero value for WP and TGP when using the exceedances above a sufficiently high threshold. The measure  $\chi(u)$  of TGP converges to zero slowly compared to WP. It is evident that WP and TGP have asymptotic independent behavior at the extreme. Thus, the simulation study supports our theoretical results of section 2.3.

## 2.6 Real data applications

This section deals with the real data applications of our proposed model. We apply our model to two real discrete time series as described subsequently.

In the first application, we apply GLP, WP, TGP, and MCPP models on the data of monthly police reports on narcotics trafficking in Sydney, Australia. The length of the series is 327 months, and it was recorded from January 1995 to March 2022. This data is part of the police reports of the New South Wales data set. The data is freely available at <http://www.bocsar.nsw.gov.au/>. Furthermore, Gorgi (2020) used this data recorded from January 1995 to December 2016 in terms of empirical applications of beta-negative binomial auto-regression modeling.

Figure 2.2 depicts the series time series and extremogram plots. It can be seen that the series have extreme observations. Especially high numbers of narcotics trafficking reports were recorded in August 2000, March 2008, July 2010, May 2015, May 2016, February 2019, and May 2020. The bird's-eye view of Figure 2.2 expresses that the reports series  $\{Y_t\}$  is stationary. For clarity, we test the stationary assumption concerning a monotonic trend in time using the procedure explained by Naghettini (2017, chp 7). The estimated test statistics is  $\hat{T} = -0.7458$ . At the significance level  $\alpha = 0.05$ , the critical value of the test statistic is  $t_{0.975,325} = 0.2279$ . Therefore, the decision is not to reject the null hypothesis that observed sample data are stationary. To this end, our

proposed model seems appropriate for describing the dependence structure of the series over the tail.

To fit GLP, WP, TGP, and MCPP models, we set threshold  $u = 13$ , corresponding to 0.80 order quantile. Each model was fitted to the time series  $\{Y_t\}$  (i.e., the monthly number of police reports on narcotics trafficking in Sydney) using PL with lag 3. Estimated parameters and standard errors and PLIC are reported in Table 2.2.

TABLE 2.2: Estimates of the model parameter with their (standard errors) under GLP, WP, TGP, and MCP for Police reports on narcotics trafficking using PL with  $\Delta = 3$ .

Model	$\beta$	$\alpha$	$\rho$	PLIC
GLP	40.99 (9.36)	5.64 (1.06)	0.80 (0.17)	-1267.20
WP	40.36 (9.11)	5.57 (1.03)	0.84 (0.13)	-1301.72
TGP	40.43 (9.15)	5.58 (1.04)	0.84 (0.13)	-1296.95
MCPP	40.59 (9.33)	5.59 (1.06)	0.81 (0.16)	-1276.23

In the second application, we also apply all four models to a number of tick changes by minutes of the exchange rate of euro to British pound (EUR/GBP) on December 12<sup>th</sup>, 2019. There were general elections in the UK on this day. Accordingly, we analyze the ticks between 9.00 a.m. to 9.00 p.m. (Greenwich mean time). Prices with high frequency from foreign exchange markets are used to construct the series. According to Gorgi (2020), the closing price of EUR/GBP is taken into account at every minute, and the number of tick changes is calculated by dividing the absolute price variation by the tick size  $10^{-5}$ . The series has 720 observations, corresponding to the number of minutes in 12 hours sample. The data is freely available and can be downloaded from the website <http://www.histdata.com/>. These time series were also used by Gorgi (2020) for empirical applications of beta-negative binomial auto-regression models.

Figure 2.3 display the time series plot and empirical extremogram of the series. Several extreme observations appear in the series and significant autocorrelation, as indicated by the autocorrelation functions. In light of these features, the proposed models are well suited to model dependence at the extreme level using the heavy tail bivariate distribution with DGPD marginals. Again using PL with  $\Delta = 4$ , all four models are fitted to exceedances above the threshold  $u = 29$ , which is fixed corresponding to the 0.90 order quantile of the series. Table 2.3 reports the results of the second application of our proposed models to the real data.

Overall, parameter estimates based on the numbers of narcotics trafficking reports series are similar across all four models. All four models show similarities in parameters



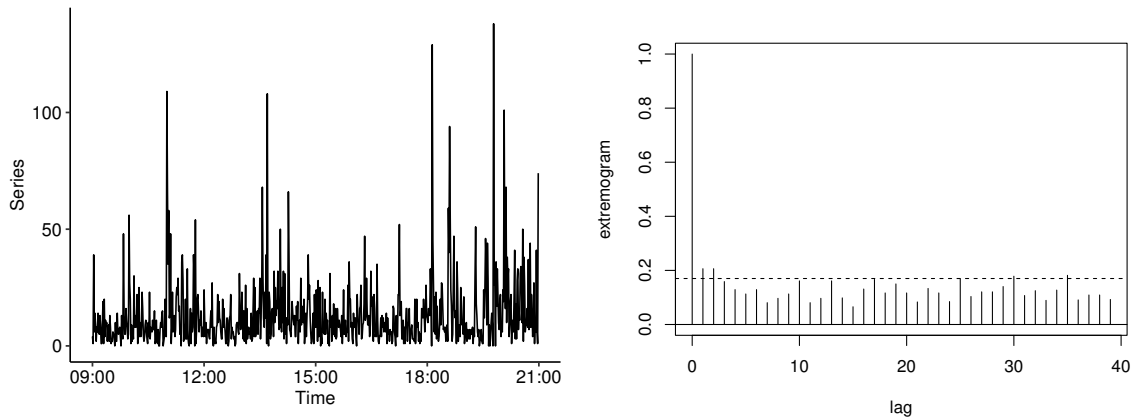


FIGURE 2.3: Number of tick changes by a minute of EUR/GBP sterling exchange rate series with sample extremogram.

in the second application as well. Notice that parameter  $\rho$  somehow deviates for WP and TGP leading to larger estimates. In applying our proposal,  $\rho$  is only the specific parameter linked entirely to the dependence structure. Compared to GLP and MCPP, WP and TGP induce a weaker extremal dependence. Therefore, the WP and TGP compensate for the tendency by increasing the estimate  $\rho$  when fitting at finite thresholds. WP and TGP have smaller PLIC, which tend to have asymptotic independence at the extreme in both applications. Employing this, both the monthly number of police reports on narcotics trafficking and the number of tick changes by minutes, the exchange rate of EUR/GBP may reveal temporal independence at the extreme level.

TABLE 2.3: Estimates of the model parameter under GLP, WP, TGP and MCP for a number of tick changes by a minute of EUR/GBP sterling exchange rate series ( $u = 29$ ) at  $\Delta = 4$ .

Model	$\beta$	$\alpha$	$\rho$	PLIC
GLP	45.25 (7.10)	4.79 (0.57)	0.85 (0.09)	1821.03
WP	44.86 (6.95)	4.76 (0.56)	0.90 (0.06)	1696.53
TGP	44.89 (6.99)	4.77 (0.56)	0.89 (0.07)	1715.64
MCPP	44.81 (7.09)	4.76 (0.57)	0.85 (0.09)	1804.58

In addition, both observed series are compared with all four respective fitted processes by examining their sub-asymptotic dependence properties. Thus, we employ tail dependence measures developed in section 2.3, which relies on  $P(Y_{t+1} > k^* | Y_t > k^*)$ . Using conditional probability, we summarize the relationship between successive exceedances of  $k^*$  (Coles (2001); Bortot and Gaetan (2014)), where  $k^*$  represents the varying level over the fixed threshold  $u$ . The variables  $Y_t$  and  $Y_{t+1}$  are said to be asymptotically

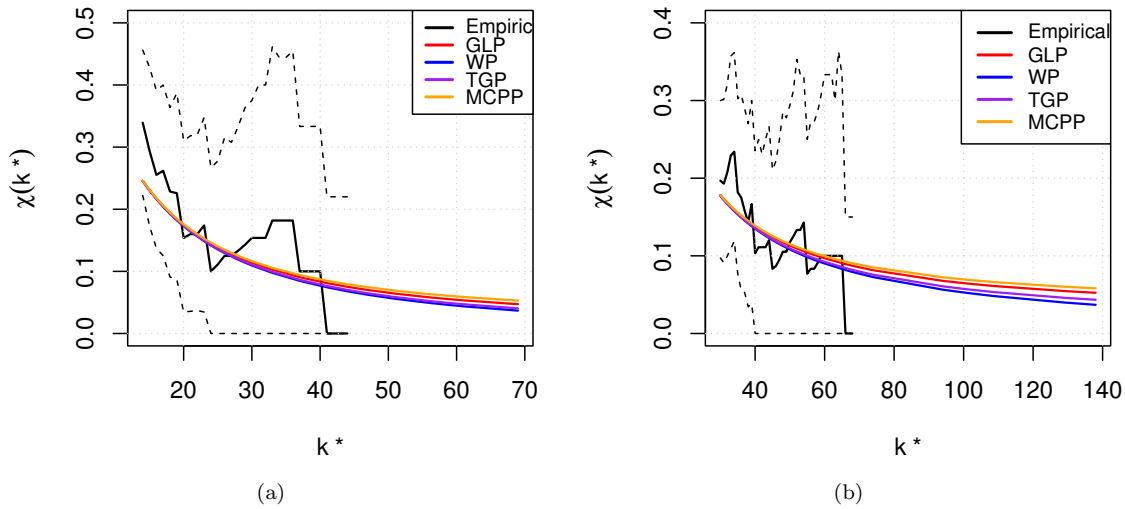


FIGURE 2.4: Empirical and fitted models plots of  $\chi = P(Y_{t+1} > k^* | Y_t > k^*)$  versus  $k^*$  at lag 4 of two real-time series (a) monthly number of police reports on narcotics trafficking in Sydney, Australia, and (b) number of tick changes by a minute of EUR/GBP from 09:00 am to 09:00 pm on December 12th, 2019. The dashed lines in both figures give 0.95 bootstrap confidence bands.

independent when  $\chi = P(Y_{t+1} > k^* | Y_t > k^*) = 0$  as  $k^* \rightarrow \infty$ ; otherwise, they are asymptotically dependent.

Figure 2.4 depicts empirical and fitted estimates of  $\chi = Pr(Y_{t+1} > k^* | Y_t > k^*)$  as function of  $k^* > u$  corresponding to both observed series. The estimates corresponding to fitted GLP, WP, TGP, and MCPP are obtained through the definition of  $\chi$  in terms of LTs using the estimated parameter values of the models. It can be seen from Figure 2.4(a, b) that the empirical estimates in both cases converge to 0, indicating that the consecutive observations over of finite threshold are asymptotically independent. Moreover, empirical and model-based curves mix well within an observed range of both applications. As we have seen in section 2.3 theoretically, the  $\chi$  measure based on conditional probability converges to zero for WP and TGP and to a positive constant for GLP and MCPP. Hence, the model-based estimates of WP and TGP converge to zero when  $k^* \rightarrow \infty$ . On the other hand, model-based estimates of GLP and MCPP tend to have positive constants.

To assess the model adequacy, we use Kendall's tau rank correlation coefficient ( $\tau$ ), a non-parametric measure of association between two variables, which measures the similarity of the rankings of the data points in two observed samples. The  $\tau$  coefficient is defined as

$$\tau = \frac{2(n_c - n_d)}{n(n-1)} \quad (2.13)$$

where  $n$  represents the sample size,  $n_c$  is the number of concordant pairs and  $n_d$  is the

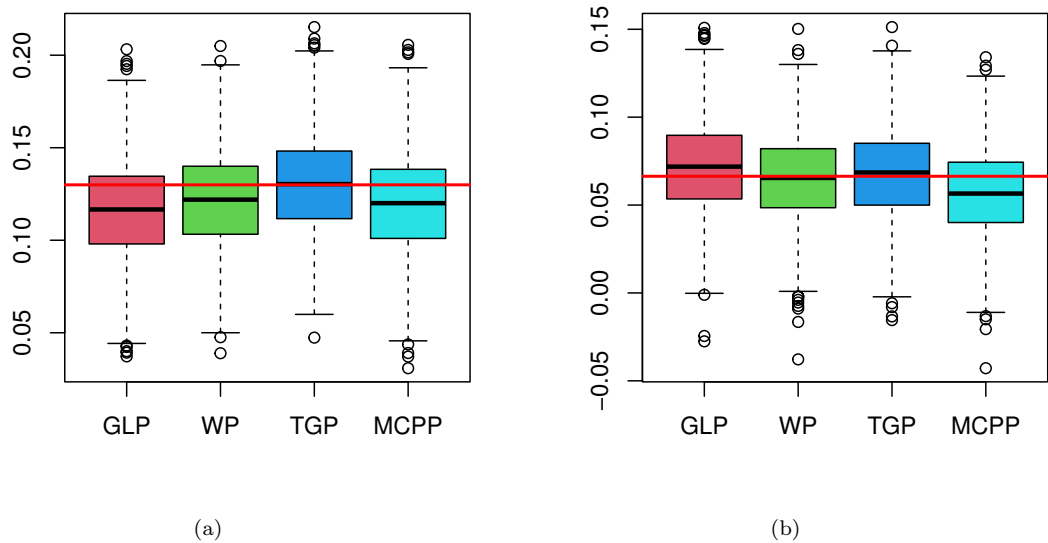


FIGURE 2.5: Bootstrap boxplots of the estimated Kendall's tau coefficient centered at observed Kendall's tau coefficient using (a) police reports on narcotics trafficking data, (b) number of tick changes by minutes of the exchange rate of euro to British pound (EUR/GBP) data.

number of disconcordant pairs. In the case of ties observations in observed samples, the formula given in (2.13) reflects as

$$\tau = \frac{2(n_c - n_d)}{n(n-1)} - T_x \quad \text{or} \quad \tau = \frac{2(n_c - n_d)}{n(n-1)} - T_y \quad (2.14)$$

where  $T_x = \sum t(t-1)$ ,  $t$  is the number of tied observations in each group of the ties in the first quantity (say  $X$  variable), while  $T_y = \sum t(t-1)$ ,  $t$  is the number of tied observations in each group of the ties in the second quantity (say  $Y$  variable)

Figure 2.5 shows boxplots of the bootstrap estimates of the estimated  $\hat{\tau}_{est}$  coefficient, the  $\hat{\tau}_{est}$  is obtained from the simulated data by using the estimated parameters of each model from Table 2.2 and Table 2.3. The length of the simulation was similar to the observed data examples. The red lines in Figure 2.5 (a, b) represent the observed  $\tau_{obs}$  coefficient, which is obtained from the observed data of considered examples. By looking at the boxplots in Figure 2.5(a), corresponding to the police reports on narcotics trafficking, we see that boxplot associated with TGP is close to  $\tau_{obs}$  suggesting an unbiased estimation of the model, while the boxplot of WP is closer to  $\tau_{obs}$  suggesting a smaller bias compared to GLP and MCPP. By looking at the boxplots in Figure 2.5(b), corresponding to the number of tick changes by minutes of the exchange rate of euro to British pound (EUR/GBP), the boxplots associated with WP and TGP are close to  $\tau_{obs}$  indicating the unbiased estimation of the models. In contrast, the boxplots deviating

from  $\tau_{obs}$  show some biasedness in model estimation. As a result, the WP and TGP models are more adequate and reliable to use for considered examples.

## 2.7 Final remarks

Chapter 2 is designed to propose a class of models for discrete time series extremes. We developed a discrete extreme value model that uses the latent process framework to obtain a flexible dependence structure. Our construction has the benefit of preserving the DGPD with a positive shape parameter for the marginals, which is reliable for modeling exceedances in discrete extreme value theory. Dependence was induced through latent Markov chains with a hierarchical setting discussed by Bortot and Gaetan (2014); Wolpert (2021). A remarkable feature of the developed models is that as you move further into the right tail of the marginal distribution, you see a substantial variation in the temporal dependence structure. To this end, the theoretical constructions of section 2.3 demonstrated that both asymptotic dependence and asymptotic independence of discrete extremes are possibly achievable with an appropriate choice of the latent process  $\Lambda_t$ .

Interestingly, the models proposed for discrete extremes have proven to be quite adaptable in apprehending different kinds of extremal dependence in both the simulation studies and the real applications. However, both the monthly number of police reports on narcotics trafficking and the number of tick changes by a minute of the EUR/GBP series have shown asymptotic independent behavior at the extreme. In the case of non-stationary marginals, it is no longer possible for a stationary GPD or DGPD model to adequately capture the tail behavior. To capture non-stationary behavior in univariate extremes, the GAM has been proposed as a more flexible approach by Chavez-Demoulin and Davison (2005); Ranjbar *et al.* (2022). The GAM uses smooth functions in order to dependence due to covariates and is less rigid than the standard regression model. The proposed framework can be modified by replacing the latent Markov chain with trawl processes in the present scenario (Noven *et al.*, 2018). Since extreme values are, by definition, rare, utilizing information from surrounding nearby geographical locations could lead to a substantial increase in power for capturing dependence. Although this would present a computational difficulty, it is an interesting subject for further research.

# Chapter 3

## Models for the entire range of count data with extreme observations

### Overview: Chapter 3

~> The statistical modeling of integer-valued extremes has received less attention than their continuous counterparts in the EVT literature. One approach to moving from continuous to discrete extremes is to model threshold exceedances of integer random variables by the discrete version of the generalized Pareto distribution. Still, the optimal threshold selection that defines exceedances remains a problematic issue. Moreover, within a regression framework, the treatment of the many data points (those below the chosen threshold) is either ignored or decoupled from extremes. Considering these issues, we extend the idea of using a smooth transition between the two tails (lower and upper) to force large and small discrete extreme values to comply with EVT. In the case of zero inflation, we also develop models with an additional parameter representing the proportion of zero values in the data. To incorporate covariates, we extend the Generalized Additive Models (GAM) framework to discrete extreme responses. In the GAM forms, the parameters of our proposed models are quantified as a function of covariates. The maximum likelihood estimation procedure is implemented for estimation purposes. With the advantage of bypassing the threshold selection step, our findings indicate that the proposed models are more flexible and robust than competing models (i.e., discrete generalized Pareto distribution and Poisson distribution). The chapter is organized as follows. Section 3.1 discusses the study background, spreads light on existing literature, and elaborates the gap of the study. Section 3.2 presented the discrete extended versions of GPD (DEGPD) and zero-inflated

versions of DEGPD (ZIDEGPD) along with a simple sampling scheme. The GAM forms procedure related to DEGPD and ZIDEGPD is given in Section 3.3. Section 3.4 deals with the construction of bivariate DEGPD. To assess the performance of the proposed models, Section 3.5 provides the results of the conducted extensive simulation study. In addition, simulation study results associated with bivariate DEGPD are provided in 3.6. Section 3.7 discusses applications of DEGPD and ZIDEGPD to the number of upheld complaints data of insurance companies. Real application of GAM form models to avalanches data with environmental covariates is also given in the same section. Bivariate DEGPD applied also applied to avalanches data, and findings are reported in 3.8 Finally, Section 3.9 concludes with final remarks and a discussion.

### 3.1 Introduction

EVT originating from the innovative work of Fisher and Tippett (1928) offers a facility of stochastic modeling related to very high and very low-frequency events (e.g., extreme temperature, heavy rainfall intensities, heavy floods, and extreme winds, etc.). For example, from the last three decades, Coles (2001), Beirlant *et al.* (2004) and de Hann and Ferreira (2006) discussed regularly adapted extreme value models to measure uncertainty for continuous extremes events. More precisely, the distribution of exceedances (i.e., the amount of data that appears over a given high threshold) is often approximated by the so-called GPD defined by its CDF in (1.5).

Optimal threshold selection in GPD application remains an arguable and elusive task (see, e.g. Dupuis, 1999; Scarrott and MacDonald, 2012). Numerous studies (for instance, Embrechts *et al.*, 1999; Choulakian and Stephens, 2001; Davison and Smith, 1990; Katz *et al.*, 2002; Boutsikas and Koutras, 2002) have established how the GPD can be fitted to continuous extreme events. This is vindicated by Pickands' theorem (Pickands, 1975), which states that, for most random variables, the distribution of the exceedances converges to a GPD as the threshold increases to the right endpoint. One of the major disadvantages of GPD is that it only models those observations which occur over a certain high threshold. This imposes an artificial dichotomy in the data (i.e., observations are either below or above the threshold) and the question of finding the optimal threshold remains complex for practitioners. In the continuous extreme value setting, many authors have attempted to model an entire range of data without

threshold selection. For example, Frigessi *et al.* (2002) proposed a dynamically weighted mixture model by combining light-tailed density and heavy-tailed density (i.e., GPD) through weight function. The dynamically weighted mixture approach can be valuable in unsupervised tail estimation, especially in heavy-tailed situations and for small percentiles. Frigessi's model has many advantages, but it has a drawback. For instance, the model has six parameters, and inference is not a straightforward task ( see Frigessi *et al.*, 2002, for more details). Carreau and Bengio (2009) proposed a semiparametric model called the "hybrid Pareto" model that stitches a Gaussian distribution with a heavy-tailed GPD. According to Carreau and Bengio (2009), the hybrid Pareto model offers efficient estimates of the tail of the distributions and converges faster in terms of log-likelihood than existing GPD. They used hybrid Pareto models in a regression context for statistical modeling of rainfall-runoff. MacDonald *et al.* (2011) combined a non-parametric kernel density estimator for the bulk of the distribution with a heavy-tailed GPD. One of the drawbacks of these approaches is that it still needs to select a suitable threshold.

To keep a low number of parameters, avoid mixture modeling, and simplify inference, Naveau *et al.* (2016) proposed a general procedure to extend the GPD class. This construction is based on the integral transform idea to simulate GPD random draws, that is  $F_{\sigma,\xi}^{-1}(U)$ , where  $U \sim \mathcal{U}(0, 1)$  represents an uniformly distributed random variable on  $(0, 1)$  and  $F_{\sigma,\xi}^{-1}$  denotes the inverse of the CDF (1.5). This leads to the class of random variables stochastically defined as

$$F_{\sigma,\xi}^{-1}\{G^{-1}(U)\}, \quad (3.1)$$

where  $G$  is a CDF on  $[0, 1]$  and  $U \sim \mathcal{U}(0, 1)$ . The key problem is to find a class for  $G$  that preserves the upper tail behavior with shape parameter  $\xi$  and also controls the lower tail behavior. Naveau *et al.* (2016) defined restrictions for validity of  $G$  families. For instance, the tail of  $G$  denoted by  $\bar{G} = 1 - G$  has to satisfy

$$\begin{aligned} \lim_{u \rightarrow 0} \frac{\bar{G}(1-u)}{u} &= a, \text{ for some finite } a > 0 \text{ (upper tail behavior),} \\ \lim_{u \rightarrow 0} \frac{G(u)}{u^\kappa} &= c, \text{ for some finite } c > 0 \text{ (lower tail behavior).} \end{aligned} \quad (3.2)$$

Four examples for parametric family  $G$  were studied in Naveau *et al.* (2016). By construction, this approach bypasses the elusive choice of a fixed and optimal threshold. Inference can be performed with classical methods such as maximum likelihood and probability weighted moments (see, e.g. Le Gall *et al.*, 2022; Furrer and Naveau, 2007).

Semi-parametric modeling based on this class has been studied by Tencaliec *et al.* (2019) and extensions to handle covariates have been proposed by Carrer and Gaetan (2022) and de Carvalho *et al.* (2022). Furthermore, Gamet and Jalbert (2022) modeled specific of tail estimation based on the same construction. Still, model (3.1) has to be yet tailored to handle discrete-valued random variables.

The subsequent development deals with the discrete extreme models. A PMF of discrete distribution is obtained in (1.6) by discretizing the CDF defined by (1.5). The distribution defined in (1.6) is called DGPD. Like continuous GPD, the DGPD is well approximated to discrete excesses over a high threshold (Hitz *et al.*, 2017). Again, an appropriate threshold selection procedure is required that can offer an optimal threshold for fitting DGPD. Also, a few questions arise the DGPD models the data above the threshold, but how to model the observations below the threshold or how to model the entire range of count data having extreme observations.

One possibility is using threshold spliced mixture representation to model the discrete observations. Again, the optimal threshold is needed for fitting the threshold spliced mixture model. The detail is provided in section 3.2.1. By keeping these arguments in mind, we want to develop a modeling framework that can be used to model the entire range of discrete extreme data without fixing the threshold. We take advantage of the constructions given in Naveau *et al.* (2016) to introduce such a modeling framework. Thus, DEGPD is proposed here by discretizing the CDF of continuous extended generalized Pareto distributions via equation (1.6). Discrete nature extreme events may contain a lot of zero values. For instance, the insurance complaints data or avalanches data may include many zero values. This data type is generally called zero-inflated (ZI), requiring specialized statistical methods for analysis. Therefore, we have introduced a zero-inflated version of DEGPD (ZIDEGPD). To model excess zeros, Lambert (1992) introduced a two-component mixture model, where one component is a point mass at zero and the other component is an assumed parametric count distribution. Lambert's specification is an example of a distributional regression model where "distributional" should emphasize that the conditional distribution of the count data is modeled in terms of covariates rather than the the mean.

So far, the main focus of the literature has been on the relationship between the mean, the variance with the covariates (Rigby and Stasinopoulos, 2005), less attention has been paid to the tail of the count distribution. Finally, we aim to correctly model excess zeros and the heavy tail of the distribution using DEGPD and ZIDEGPD GAM forms models.



## 3.2 Discrete extremes modeling

### 3.2.1 The discrete Gamma spliced threshold DGPD model

As we have seen in existing literature, the discrete and continuous GPD is well approximated to the data existing over a specified threshold. In the continuous extreme value framework, many authors have tried to model the bulk part with some specific distribution (for example, Gamma, Normal, Log-normal, Weibull, Beta distribution, etc.). The arrangement, where data above and below an unknown threshold is drawn from the “bulk” and “tail” distributions, respectively, meets into a family of models called spliced threshold models. The distinctive motivation for the model is the privilege that two underlying processes originally generate the data above and below the threshold. For a thorough review of the general spliced threshold model in the continuous domain, see, for example, Dey and Yan (2016).

Let  $F_m(y|\boldsymbol{\theta}_B)$  be the CDF of gamma distribution which corresponds to the bulk model, and  $F_m(y|\boldsymbol{\theta}_T)$  is CDF of GPD corresponds to tail model, where  $\boldsymbol{\theta}_B$  and  $\boldsymbol{\theta}_T$  indicates the parameter vectors of bulk and tail models, respectively. The CDF of the bulk and tail model spliced at the threshold is given by

$$H_m(y) = \begin{cases} (1 - \phi) \frac{F_m(y|\boldsymbol{\theta}_B)}{F_m(u|\boldsymbol{\theta}_B)} & \text{for } y \leq u \\ 1 - \phi + \phi F_m(y|\boldsymbol{\theta}_T, u) & \text{for } y > u \end{cases} \quad (3.3)$$

with corresponding probability density function

$$h_m(x) = \begin{cases} (1 - \phi) \frac{f_m(x|\boldsymbol{\theta}_B)}{F_m(u|\boldsymbol{\theta}_B)} & \text{for } x \leq u \\ 1 - \phi + \phi f_m(x|\boldsymbol{\theta}_T, u) & \text{for } x > u \end{cases} \quad (3.4)$$

When we have discrete extreme data that characteristically exhibits several ties on the lower tail. That's why we need to amend the model developed before for continuous paradigm to discrete form in order to account for the censored data. Let  $Y \sim \Gamma(\alpha, \beta)$ , we write the PMF of discrete gamma distribution (Chakraborty and Chakravarty, 2012) as

$$P_g(Y = k) = \frac{1}{\Gamma(\alpha)} [\gamma(\alpha, \beta(k+1)) - \gamma(\alpha, \beta k)] \quad (3.5)$$

where  $\alpha > 0, \beta > 0, Y \in \mathbb{N}_0$  and  $\gamma(\alpha, \beta k)$  is the lower incomplete gamma function

$$\gamma(\alpha, \beta k) = \int_0^{\beta k} t^{\alpha-1} e^{-t} dt \quad (3.6)$$

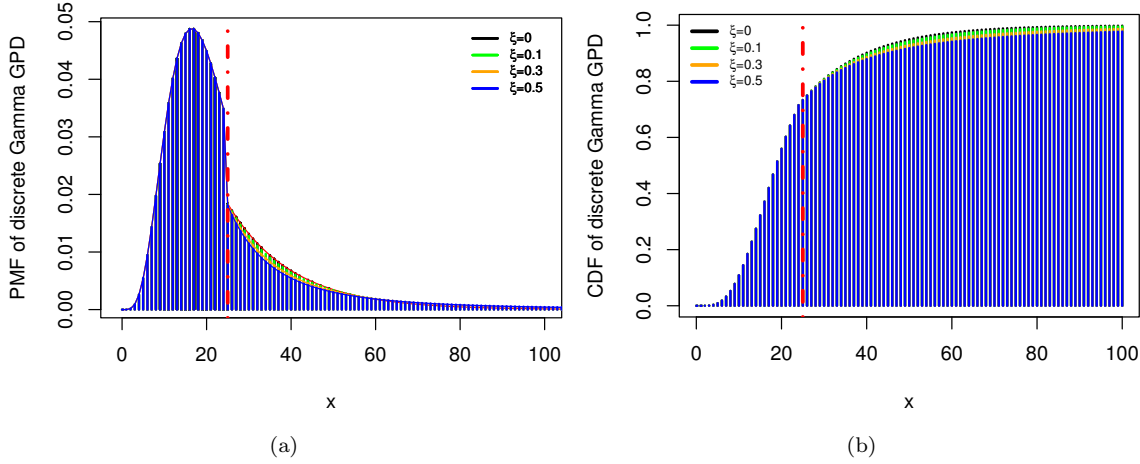


FIGURE 3.1: Probability mass function and the cumulative distribution function of discrete gamma GP spliced at threshold model corresponding to  $\theta_{\mathbf{B}} = (\alpha = 5, \beta = 0.25)$ ,  $u=25$ ,  $\theta_{\mathbf{T}} = (\sigma = 15, \xi = 0, 0.1, 0.3, 0.5)$

On the other hand, If  $Y \sim GPD(u, \sigma, \xi)$ , the PMF of DGPD is written by using (1.6) as

$$P_{dp}(Y = k) = \left(1 + \frac{\xi(k - u)}{\sigma}\right)^{-1/\xi} - \left(1 + \frac{\xi(k + 1 - u)}{\sigma}\right)^{-1/\xi} \quad (3.7)$$

for  $(-\infty < u, \xi < \infty)$ ,  $\sigma \in (0, \infty)$  and  $k \in \mathbb{N}_0$ . The DGPD support  $k \geq u$  when  $\xi \geq 0$  and  $u \leq k \leq u - \frac{\sigma}{\xi}$  when  $\xi < 0$ . Here, we observe  $\xi > 0$ . Hence, the discrete gamma generalized Pareto distribution (DGPPD) spliced at the threshold model is given by

$$P_{dggp}(Y = k) = \begin{cases} (1 - \phi) \frac{P_g(Y=k|\theta_{\mathbf{B}})}{F(u-d|\theta_{\mathbf{B}})} & \text{for } k \leq u - d \\ \phi P_{dp}(Y = k|\theta_{\mathbf{T}}, u) & \text{for } k \geq u \end{cases} \quad (3.8)$$

where  $P_g(X = k|\theta_{\mathbf{B}}) \sim \text{discrete Gamma}(\alpha, \beta)$ ,  $P_{dp}(X = k|\theta_{\mathbf{T}}) \sim \text{discrete GPD}$ ,  $d = \min(k)$ , which indicates that we model integer data and  $\phi$  is the proportion of observations over threshold  $u$ . For more details about mixture models and their implications, see, for instance, Hu and Scarrott (2018).

Figure 3.1 shows the bulk and tail parts behavior in terms of PMF and CDF of DG-GPD spliced at the threshold. Also, for fitting this mixture model, a suitable threshold is needed. However, a new framework to model a whole range of data has been developed to overcome this deficiency by avoiding the threshold selection issue.

### 3.2.2 Discrete extended generalized Pareto distribution

We start by considering the CDF  $G\{F(\cdot; \sigma, \xi)\}$  of EGPD where  $G$  meets the conditions (3.2). To model non-negative integer data, we discretize the CDF by

$$P(Y = k) = G\{F(k+1; \sigma, \xi)\} - G\{F(k; \sigma, \xi)\}, \quad k \in \mathbb{N}_0 \quad (3.9)$$

The distribution defined by (3.9) will be called discrete extended generalized Pareto distribution (DEGPD). The explicit formula of the CDF of DEGPD is developed as

$$P(Y \leq k) = G\{F(k+1; \sigma, \xi)\} \quad (3.10)$$

and the quantile function is derived as

$$q_p = \begin{cases} \lceil \frac{\sigma}{\xi} \left[ \{1 - G^{-1}(p)\}^{-\xi} - 1 \right] \rceil - 1, & \text{if } \xi > 0 \\ \lceil -\sigma \log \{1 - G^{-1}(p)\} \rceil - 1, & \text{if } \xi = 0 \end{cases} \quad (3.11)$$

with  $0 < p < 1$ .

For  $G$ , we use four parametric expressions  $G(\cdot, \psi)$ , already proposed in Naveau *et al.* (2016), namely

- i.  $G(u; \psi) = u^\kappa$ ,  $\psi = \kappa > 0$ ;
- ii.  $G(u; \psi) = 1 - D_\delta\{(1-u)^\delta\}$ ,  $\psi = \delta > 0$  where  $D_\delta$  is the CDF of a Beta random variable with parameters  $1/\delta$  and 2, that is:

$$D_\delta(u) = \frac{1+\delta}{\delta} u^{1/\delta} \left(1 - \frac{u}{1+\delta}\right)$$

- iii.  $G(u; \psi) = [1 - D_\delta\{(1-u)^\delta\}]^{\kappa/2}$ ,  $\psi = (\delta, \kappa)$  with  $\delta > 0$  and  $\kappa > 0$ ;
- iv.  $G(u; \psi) = pu^{\kappa_1} + (1-p)u^{\kappa_2}$ ,  $\psi = (p, \kappa_1, \kappa_2)$  with  $\kappa_2 \geq \kappa_1 > 0$  and  $p \in (0, 1)$ .

The parametric family (i) leads to PMF of DEGPD with three parameters ( $\kappa, \sigma$  and  $\xi$ ):  $\kappa$  deals the shape of the lower tail,  $\sigma$  is a scale parameter, and  $\xi$  controls the rate of upper tail decay. Thus, Figure 3.2(a) shows the behavior of PMF of DEGPD with fixed scale and upper tail shape parameter (i.e.,  $\sigma = 1$  and  $\xi = 0.5$ ) and with different values of lower tail behaviors ( $\kappa=1, 2, 5, 10$ ). Similar to the EGPD framework of Naveau *et al.* (2016), the DGPD is recovered when  $\kappa = 1$ , and additional flexibility for low values is attained by varying  $\kappa$ . For instance, more flexibility on the lower tail can be observed without losing upper tail behavior in Figure 3.2(a) when putting the value of  $\kappa = 10$ .

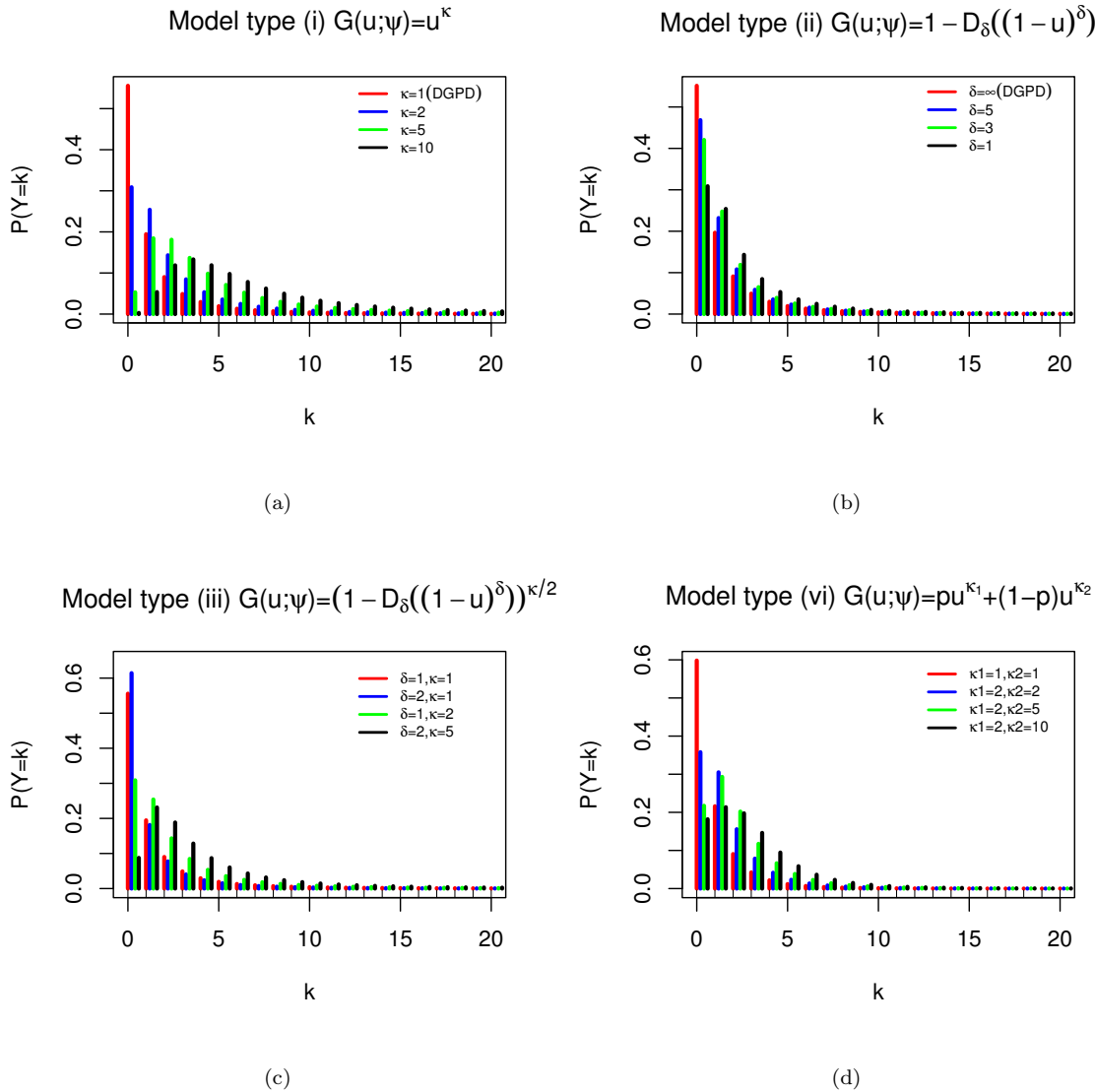


FIGURE 3.2: (a) Probability mass function corresponding to model in (3.9) type (i) for  $\sigma = 1$ ,  $\xi = 0.5$  and shape parameter for lower tail  $\kappa = 1, 2, 5, 10$ ; (b) Probability mass function corresponding to model (3.9) combined with  $G(u; \psi) = 1 - D_\delta(1 - u)^\delta$  for  $\sigma = 1$ ,  $\xi = 0.5$  and  $\delta = \infty, 5, 3, 1$ ; (c) Probability mass function corresponding to model (3.9) combined with  $G(u; \psi) = [1 - D_\delta(1 - u)^\delta]^{\kappa/2}$  for  $\sigma = 1$ ,  $\xi = 0.5$ ,  $\delta = 1, 2$  and  $\kappa = 1, 2, 5$ , and (d) Probability mass function corresponding to model in (3.9) combined with  $G(u; \psi) = pu^{\kappa_1} - (1 - p)u^{\kappa_2}$  for  $\sigma = 1$ ,  $\xi = 0.2$ ,  $\kappa_1 = 1, 2$  and  $\kappa_2 = 1, 2, 5, 10$ .

The parametric family (iv) is the mixture of power laws:  $\kappa_1$  identifies the shape of the lower tail, while  $\kappa_2$  modifies the shape of the central part of the distribution and  $\sigma$  and  $\xi$  are scale and upper tail parameters, respectively. It can be observed from Figure 3.2(d) that the DEGPD related to  $G(u; \psi) = pu^{\kappa_1} + (1 - p)u^{\kappa_2}$  is also showing flexibility with  $p = 0.5$ ,  $\sigma = 1$ ,  $\xi = 0.5$ ,  $\kappa_1 = 1, 2$ , different values of  $\kappa_2$ .

The parametric family (ii) is another interesting choice for constructing DEGPD.

This choice is fairly more complex than the previous two. Figure 3.2(b), illustrates the behavior of PMF with different values of  $\delta$ . the EGPD connected with this G family converges to GPD when  $\delta$  increases to infinity. Moreover, conditions in (3.2) are satisfied with  $\delta = 2$  (see Naveau *et al.*, 2016, for more details). In discrete settings, the DEGPD corresponding  $G(u; \psi) = 1 - D_\delta\{(1-u)^\delta\}$  also becomes very closer to the DGPD density when  $\delta$  increases to infinity.

In general, the parameter  $\delta$  describes the central part of the distribution. Thus, this parameter relatively improves the modeling flexibility for the central part of the distribution. The parameter  $\delta$  is sometimes interpreted as a "threshold tuning parameter". One of the drawbacks of DEGPD (ii) is that it models only the central and upper part of the distribution. On the other hand, the lower tail behavior could not be estimated directly. This drawback is addressed by implementing another parametric family.

The parametric family (iii) supports the lower tail of the distribution with the  $\kappa > 0$  parameter. Interestingly, this family also tends to the DEGPD with parameters  $(\kappa, \delta, \sigma$  and  $\xi)$ . The  $(\kappa, \delta$  and  $\xi)$  represents the lower, central, and upper parts of the distribution, respectively, and  $\sigma$  is a scale parameter as usual. In particular, Figure 3.2(c) showing the behavior of PMF of DEGPD linked with  $G(u; \psi) = [1 - D_\delta\{(1-u)^\delta\}]^{\kappa/2}$  at different settings of the parameters.

Overall, all types of DEGPD discussed above with a combination of different parametric families are more flexible for modeling discrete extremes except the DEGPD (ii). In fact, the DEGPD (ii) corresponding to the parametric family (ii) has limited flexibility on the lower tail.

In addition, a large number of zeros can be found in various practical application data sets. In that case, the usual statistical models with a flexible lower tail cannot be adjusted for the excessive zeros, which complicates a precise statistical analysis. An investigation of the origins of these zeros is essential. The subsequent section will explain the zero inflation modeling framework.

### 3.2.3 Zero-inflated discrete extended generalized Pareto distribution

We follow Lambert (1992) and we suppose that  $Z$  is observed with an excessive number of zeros relative to those observed under the DEGPD, the zero-inflated distribution (ZIDEGPD) is defined in a straightforward way as:

$$P(Z = m) = \begin{cases} \pi + (1 - \pi)G\{F(1, \sigma, \xi)\} & m = 0 \\ (1 - \pi)[G\{F(m+1, \sigma, \xi)\} - G\{F(m, \sigma, \xi)\}] & m = 1, 2, \dots \end{cases} \quad (3.12)$$

where  $0 \leq \pi \leq 1$  and the remaining parameters would be the same as DEGPD. It turns out that the CDF of ZIDEGPD is

$$P(Z \leq m) = \pi + (1 - \pi)G\{F(m + 1, \sigma, \xi)\}, \quad m \in \mathbb{N}_0, \quad (3.13)$$

and the quantile function is

$$q_{p^*} = \begin{cases} \lceil \frac{\sigma}{\xi} \left[ \{1 - G^{-1}(p^*)\}^{-\xi} - 1 \right] \rceil - 1, & \xi > 0 \\ \lceil -\sigma \log \{1 - G^{-1}(p^*)\} \rceil - 1, & \xi = 0 \end{cases} \quad (3.14)$$

with  $0 < p^* = (p - \pi)/(1 - \pi) < 1$ . Again, the above expressions are simple and straightforward with the existing four G families. The flexibility of the proposed extended versions is observed here through PMF with different lower tail parameters.

The parametric family (i) also leads to PMF of ZIDEGPD given in (3.12) with four parameters ( $\pi, \kappa, \sigma$  and  $\xi$ ):  $\pi$  is the proportion of zero observations in the sample that is inflating the data distribution and the interpretation of other parameters is same as DEGPD. Thus, Figure 3.3(a) shows the behavior of PMF of ZIDEGPD with fixed zero-inflation, scale, and upper tail shape parameters (i.e.,  $\pi = 0.20, \sigma = 1$  and  $\xi = 0.2$ ) and with different values of lower tail behaviors ( $\kappa = 1, 2, 5, 10$ ). The zero-inflated DGPD is recovered when  $\kappa = 1$  with more proportion of zero values. Additional flexibility for low values with the proportion of zero inflation is attained by varying  $\kappa$ . For instance, more flexibility on the lower tail with zero inflation can be observed without losing upper tail behavior in Figure 3.3(a) when putting the value of  $\kappa = 10$  or  $20$ . ZI Poisson and ZIDEGPD densities behave similarly for small and moderate values but differ at the upper tail.

The ZIDEGPD corresponding to parametric family (iv) has six parameters ( $\pi, p, \kappa_1, \kappa_2, \sigma$  and  $\xi$ ) by following the restriction ( $\kappa_1 \leq \kappa_2$ ). It can be noticed from Figure 3.3(d) that the ZIDEGPD based on  $G(u; \psi) = pu^{\kappa_1} + (1 - p)u^{\kappa_2}$  is also a flexible and produce zero inflation when using  $\pi = 0.2, p = 0.5, \sigma = 1, \xi = 0.2, \kappa_1 = 1, 2$ , different values of  $\kappa_2$ . Again, the density of ZI Poisson and ZIDEGPD show similar behavior for small and moderate values; it may change at the upper tail due to the heavy tail of ZIDEGPD. The ZIDEGPD proposed by using  $G(u; \psi) = 1 - D_\delta\{(1 - u)^\delta\}$  have four parameters ( $\pi, \delta, \sigma$  and  $\xi$ ). Figure 3.3(b) describes the behaviour of PMF with fixed parameters (i.e.,  $\pi = 0.2, \sigma = 1$  and  $\xi = 0.2$ ) and with different values of  $\delta$ . The ZIDEGPD follows ZIDGPD when  $\delta$  increases to infinity. It can be observed that the number of zeros increases when  $\delta$  increases. This behaves like ZIP at the lower tail and a central part when the mean of ZIP is small. The disadvantage of this kind of ZIDEGPD is that

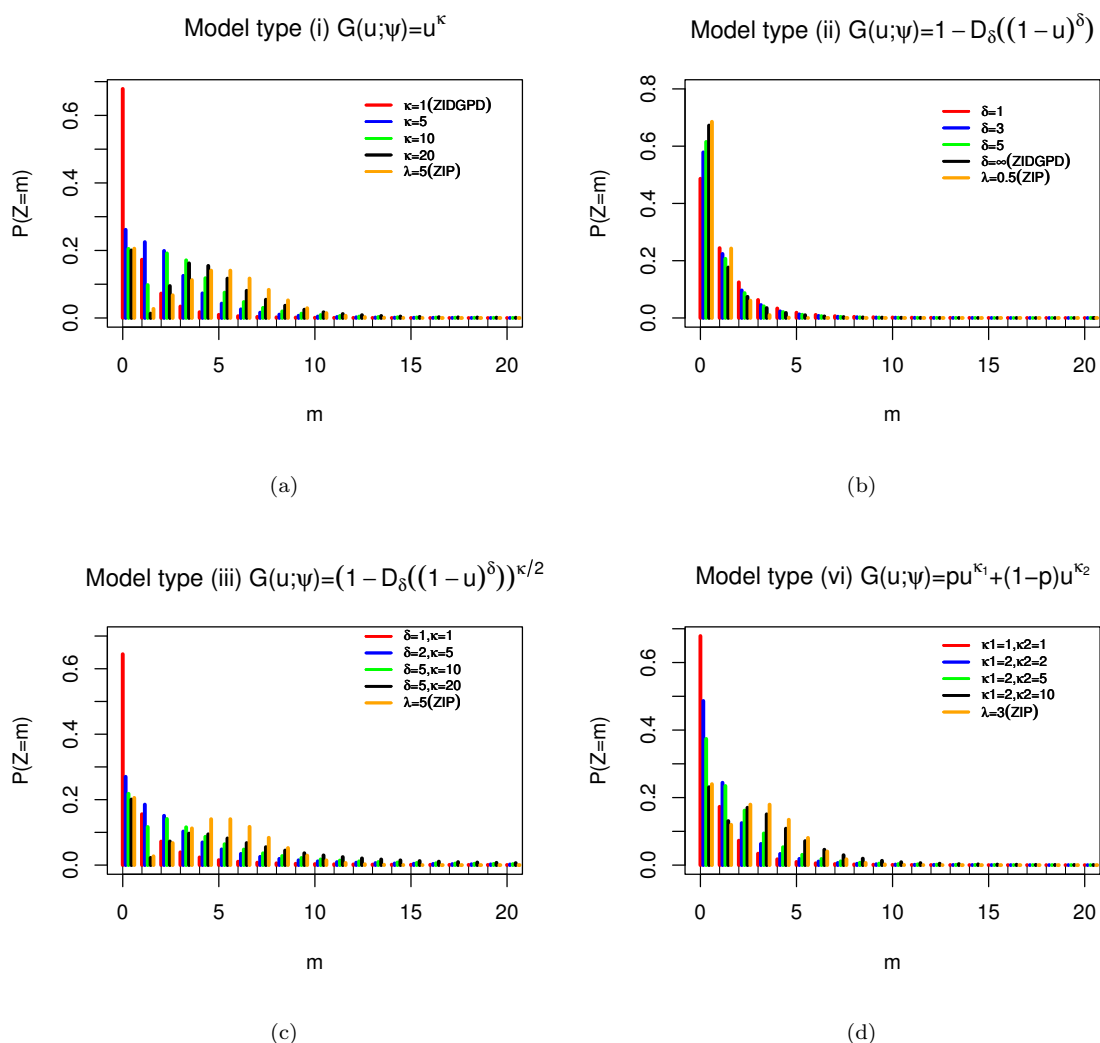


FIGURE 3.3: (a) Probability mass function corresponding to model in (3.12) with  $G(u; \psi) = u^\kappa$  having  $\pi = 0.2, \sigma = 1, \xi = 0.2$  and shape parameter for lower tail  $\kappa = 1, 5, 10, 20$ ; (b) Probability mass function corresponding to model in (3.12) combined with  $G(u; \psi) = 1 - D_\delta(1 - u)^\delta$  for  $\pi = 0.2, \sigma = 1, \xi = 0.2$  and  $\delta = 1, 3, 5, \infty$ ; (c) Probability mass function corresponding to model in (3.12) combined with  $G(u; \psi) = [1 - D_\delta(1 - u)^\delta]^{\kappa/2}$  for  $\pi = 0.2, \sigma = 1, \xi = 0.2, \delta = 1, 2, 5$  and  $\kappa = 1, 5, 10, 20$ , and (d) Probability mass function corresponding to model (3.12) combined with  $G(u; \psi) = pu^{\kappa_1} - (1 - p)u^{\kappa_2}$  for  $\pi = 0.2, p = 0.5, \sigma = 1, \xi = 0.2, \kappa_1 = 1, 2$  and  $\kappa_2 = 1, 2, 5, 10$ . The orange lines represent the zero-inflated Poisson probability mass function with different settings of their parameter.

it models only the central and upper parts of the distribution; the lower tail behavior could not be estimated directly even though we have zero inflation.

The proposed ZIDEGPD based on  $G(u; \psi) = [1 - D_\delta\{(1 - u)^\delta\}]^{\kappa/2}$  have five parameters:  $\pi$  supporting the proportion of zero values, the  $(\kappa, \delta$  and  $\xi)$  represents the lower, central and the upper parts of the distribution, respectively, and  $\sigma$  is a scale parameter as usual. Figure 3.3(c) showing the behavior of PMF of ZIDEGPD linked with

$G(u; \psi) = [1 - D_\delta\{(1 - u)^\delta\}]^{\kappa/2}$  at different settings of the parameters. Zero inflation also occurs when  $\kappa$  increases. Again, the density shape is similar to ZIP at the lower and central part of the distribution.

For modeling zero-inflated discrete extremes, all types of ZIDEGPD discussed above are flexible for both tails. The ZIDEGPD (ii) corresponding to parametric family (ii) has limited flexibility on the lower tail, but the  $\pi$  parameter explains zero proportion more correctly.

### 3.3 Generalized additive modelling

This section proposes regression-based discrete extreme models by letting the parameters of discrete extreme models vary with covariates. In a continuous framework, modeling continuous variables via extreme value model approximations, employing techniques that allow for the incorporation of flexible forms of dependence on covariates is very appealing. Davison and Smith (1990) used such models to model the size and occurrence of excesses over a high threshold through GPD. Pauli and Coles (2001) proposed smooth models for extreme value distribution parameters based on penalized likelihood. Later on, Chavez-Demoulin and Davison (2005) used Generalized Additive Model (GAM) that was originally proposed by Hastie and Tibshirani (1990) to estimate flexible GPD parameters with an orthogonal reparametrization. Yee and Stephenson (2007) developed vector generalized additive models to model generalized extreme value distribution parameters as linear or smooth functions of covariates. Vector generalized additive models can easily be implemented in an R package called **VGAM**. More recently, Youngman (2019) models threshold exceedances with GPD parameters of GAM forms. Generally, the GAM form models characteristically reflect additive smooths representations with splines.

In the sequel we denote the vector of parameters  $(\xi, \sigma, \psi^T)^T$  or  $(\xi, \sigma, \psi^T, \pi)^T$  with  $\theta = (\theta_1, \dots, \theta_d)^T$ . In practice, the parameters of the distribution of  $Y$  or  $Z$  may depend on some covariates  $\mathbf{x}$ , i.e.  $\theta(\mathbf{x}) = (\theta_1(\mathbf{x}), \dots, \theta_d(\mathbf{x}))^T$ . The specification is an instance of a distributional regression model (Stasinopoulos *et al.*, 2018).

For relating the distributional parameters  $(\theta_1(\mathbf{x}), \dots, \theta_d(\mathbf{x}))$  to the covariates, we consider additive predictors of the form

$$\eta_i(\mathbf{x}) = s_{i1}(\mathbf{x}) + \dots + s_{iJ_i}(\mathbf{x}) \quad (3.15)$$

where  $s_{i1}(\cdot), \dots, s_{iJ_i}(\cdot)$  are smooth functions of the covariates  $\mathbf{x}$ . The predictors are



linked to the distributional parameters via known monotonic and twice differentiable link functions  $h_i(\cdot)$ .

$$\theta_i(\mathbf{x}) = h_i(\eta_i(\mathbf{x})), \quad i = 1, \dots, d \quad (3.16)$$

In the case of model with  $G(u; \psi) = u^\kappa$ , common link functions are

$$\xi(\mathbf{x}) = \exp(\eta_\xi(\mathbf{x})), \quad \sigma(\mathbf{x}) = \exp(\eta_\sigma(\mathbf{x})), \quad \kappa(\mathbf{x}) = \exp(\eta_\kappa(\mathbf{x})), \quad \pi(\mathbf{x}) = \exp\left(\frac{\eta_p(\mathbf{x})}{1 + \eta_p(\mathbf{x})}\right)$$

The functions  $s_{ij}$  in (3.15) are approximated in terms of basis function expansions

$$s_{ij}(\mathbf{x}) = \sum_{k=1}^{K_{ij}} \beta_{ij,k} B_k(\mathbf{x}), \quad (3.17)$$

where  $B_k(\mathbf{x})$  are the basis functions and  $\beta_{ij,k}$  denote the corresponding basis coefficients. These basis can be of different types (see Wood, 2017, for instance). The basis function expansions can be written as  $s_{ij}(\mathbf{x}) = \mathbf{t}_{ij}(\mathbf{x})^T \boldsymbol{\beta}_{ij}$  where  $\mathbf{t}_{ij}(\mathbf{x})$  is still a vector of transformed covariates that depends on the basis functions and  $\boldsymbol{\beta}_{ij} = (\beta_{ij,1}, \dots, \beta_{ij,K_{ij}})^T$  is a parameter vector to be estimated.

The MLE method is practiced to estimate the parameters of the proposed models. Let  $y_1, \dots, y_n$  be  $n$  independent observations from (3.9) and  $\mathbf{x}_1, \dots, \mathbf{x}_n$  the related covariates. The log-likelihood function is given by,

$$l(\boldsymbol{\beta}) = \sum_{i=1}^n \log [G(F(y_i + 1; \sigma(\mathbf{x}_i), \xi(\mathbf{x}_i)); \psi(\mathbf{x}_i)) - G(F(y_i; \sigma(\mathbf{x}_i), \xi(\mathbf{x}_i)); \psi(\mathbf{x}_i))]. \quad (3.18)$$

where  $\boldsymbol{\beta}$  collects all unknown coefficient  $\beta_{ij,k}$  of the basis expansions.

Instead if we consider  $n$  independent observations  $z_1, \dots, z_n$  from (3.12) we get

$$\begin{aligned} l(\boldsymbol{\beta}) &= \sum_{i=1}^n I_0(z_i) \log [\pi(\mathbf{x}_i) + (1 - \pi(\mathbf{x}_i))G(F(1; \sigma(\mathbf{x}_i), \xi(\mathbf{x}_i)); \psi(\mathbf{x}_i))] + \\ &+ \sum_{i=1}^n (1 - I_0(z_i)) \log(1 - \pi(\mathbf{x}_i)) \times \\ &[G(F(z_i + 1; \sigma(\mathbf{x}_i), \xi(\mathbf{x}_i)); \psi(\mathbf{x}_i)) - G(F(z_i; \sigma(\mathbf{x}_i), \xi(\mathbf{x}_i)); \psi(\mathbf{x}_i))]. \end{aligned} \quad (3.19)$$

Derivatives with respect to unknown parameters of DEPGD and ZIDEGPD can be solved by standard numerical techniques to obtain the maximum likelihood estimators for unknown parameters.

To ensure regularization of the functions  $s_{ij}(\mathbf{x})$  so-called penalty terms are added to the objective log-likelihood function. Usually, the penalty for each function  $s_{ij}(\mathbf{x})$  are

quadratic penalty  $\lambda \boldsymbol{\beta}_{ij}^T \mathbf{G}_{ij}(\boldsymbol{\lambda}_{ij}) \boldsymbol{\beta}_{ij}$  where  $\mathbf{G}_{ij}(\boldsymbol{\lambda}_{ij})$  is a known semi-definite matrix and the vector  $\boldsymbol{\lambda}_{ij}$  regulates the amount of smoothing needed for the fit. A special case is when  $\mathbf{G}_{ij}(\boldsymbol{\lambda}_{ij}) = \lambda_{ij} \mathbf{G}_{ij}$ . Therefore the type and properties of the smoothing functions are controlled by the vectors  $\boldsymbol{x}_{ij}(\boldsymbol{x})$  and the matrices  $\mathbf{G}_{ij}(\boldsymbol{\lambda}_{ij})$ .

The penalized log-likelihood function for the latter models reads:

$$\begin{aligned} l_p(\boldsymbol{\beta}) &= l(\boldsymbol{\beta}) - \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^{J_i} \boldsymbol{\beta}_{ij}^T \mathbf{G}_{ij}(\boldsymbol{\lambda}_{ij}) \boldsymbol{\beta}_{ij} \\ &= l(\boldsymbol{\beta}) - \frac{1}{2} \boldsymbol{\lambda}_k \boldsymbol{\beta}^T \mathbf{G}^k \boldsymbol{\beta} \end{aligned} \quad (3.20)$$

where  $l(\boldsymbol{\beta})$  is the log-likelihood function (3.18) or (3.19). Wood (2011) proposed a restricted maximum likelihood by coupling with GAM forms to estimate  $\boldsymbol{\beta}$  and  $\boldsymbol{\lambda}$ , Wood *et al.* (2016) extends beyond the exponential family. Let  $\hat{\boldsymbol{\beta}}$  be the maximizer of  $l_p$  given  $\boldsymbol{\lambda}$  and let  $\mathcal{H} = -\nabla^2 l_p(\hat{\boldsymbol{\beta}})$ . The result in a restricted likelihood

$$l_R(\boldsymbol{\lambda}) = l(\hat{\boldsymbol{\beta}}) + \frac{1}{2} \log |\mathbf{G}^\lambda|_+ - \frac{1}{2} \log |\mathcal{H}| + \text{cst} \quad (3.21)$$

where  $\mathbf{G}^\lambda = \lambda_k \mathbf{G}^k$  and  $|\mathbf{G}^\lambda|_+$  the product of the positive eigen-values of  $\mathbf{G}^\lambda$ . Comprehensive details for implementing restricted maximum likelihood estimation using (3.21) is provided in Wood *et al.* (2016); Youngman (2019).

To fit DEGPD and ZIDEGPD with GAM forms, we have written an R code that implements the distributions as “new families” for `evgam` R package (Youngman, 2020). An example of R code with the name “Fit\_degpd\_zidegpd.R” and the complete source code of the function is provided on the GitHub <https://github.com/touqeerahmadunipd/degpd-and-zidegpd>.

### 3.4 Bivariate modeling of entire range discrete extremes data

In this section, we discuss bivariate versions of DEGPD (BDEGPD). Let  $Y \in \mathbb{N}_0$  follow the DEGPD with PMF and CDF defined in (3.9) and (3.10). To obtain BDEGPD, we start with the CDF of EGPD defined in Naveau *et al.* (2016). It turns out that  $Y \sim H(k)$  iff

$$Y = \sigma H_\xi^{-1}\{G^{-1}(U)\} \quad (3.22)$$

or

$$Y = \sigma H_\xi^{-1}[G^{-1}\{H_1(X)\}] \quad (3.23)$$

where  $G$  is a CDF on  $[0, 1]$ ,  $U \sim \mathcal{U}(0, 1)$  and  $H_1(X)$  is the CDF of GPD with shape  $\xi = 1$ . Let suppose  $(Y_1, Y_2) \in \mathbb{N}_0 \times \mathbb{N}_0$  are independent random variables, and the joint PMF is defined as

$$P(Y_1 = k_1, Y_2 = k_2) = P(k_1 < Y_1 \leq k_1 + 1, k_2 < Y_2 \leq k_2 + 1) \quad (3.24)$$

In general, the joint PMF can be written in terms of joint CDFs as

$$\begin{aligned} P(Y_1 = k_1, Y_2 = k_2) &= P(Y_1 \leq k_1 + 1, Y_2 \leq k_2 + 1) - P(Y_1 \leq k_1 + 1, Y_2 \leq k_2) \\ &\quad - P(Y_1 \leq k_1, Y_2 \leq k_2 + 1) + P(Y_1 \leq k_1, Y_2 \leq k_2) \end{aligned} \quad (3.25)$$

To simplify the above expression, we solve each part of the right side separately. Thus, we start with

$$\begin{aligned} P(Y_1 \leq k_1, Y_2 \leq k_2) &= P(\sigma H_\xi^{-1}[G^{-1}\{H_1(X_1)\}] \leq k_1, \sigma H_\xi^{-1}[G^{-1}\{H_1(X_2)\}] \leq k_2) \\ &= P\left(X_1 \leq H_1^{-1}\left[G\left\{H_\xi\left(\frac{k_1}{\sigma}\right)\right\}\right], X_2 \leq H_1^{-1}\left[G\left\{H_\xi\left(\frac{k_2}{\sigma}\right)\right\}\right]\right) \\ &= P(X_1 \leq x_1, X_2 \leq x_2) \end{aligned} \quad (3.26)$$

The term  $P(X_1 \leq x_1, X_2 \leq x_2)$  can be written as Laplace transform of Gamma distribution by means of  $X_i$  following GPD with shape parameter  $\xi = 1$ . By using this trick, the Laplace transform of Gamma distribution with  $\Lambda \sim \text{Gamma}(1, 1)$  is written as

$$\begin{aligned} P(X_1 > x_1, X_1 > x_2) &= \mathbb{E}_\Lambda (e^{-\Lambda x_1} e^{-\Lambda x_2}) \\ &= \frac{1}{(1 + x_1 + x_2)} \\ &= L^{(2)}(s)|_{(s=x_1+x_2)} \end{aligned} \quad (3.27)$$

Now, the joint CDF  $P(Y_1 \leq k_1, Y_1 \leq k_2) = P(X_1 \leq x_1, X_1 \leq x_2)$  can be derived easily with the support of Gamma Laplace transforms, that is

$$\begin{aligned} P(Y_1 \leq k_1, Y_2 \leq k_2) &= 1 - P(Y_1 > k_1) - P(Y_2 > k_2) + P(Y_1 > k_1, Y_1 > k_2) \quad (3.28) \\ &= 1 - L^{(1)}(s)|_{(s=k_1)} - L^{(1)}(s)|_{(s=k_2)} + L^{(2)}(s)|_{(s=k_1+k_2)} \end{aligned}$$

where  $L^{(1)}(\cdot)$  is the univariate Laplace transform of  $\text{Gamma}(1, 1)$ . Similarly, the expressions  $P(Y_1 \leq k_1 + 1, Y_2 \leq k_2 + 1)$ ,  $P(Y_1 \leq k_1 + 1, Y_2 \leq k_2)$  and  $P(Y_1 \leq k_1, Y_2 \leq k_2 + 1)$

can be derived by following the same procedure. By using (3.25), we get the PMF of BDEGPD in terms of LTs of bivariate Gamma distribution as

$$P(Y_1 = k_1, Y_2 = k_2) = L^{(2)}(s)|_{(s=k_1+1+k_2+1)} - L^{(2)}(s)|_{(s=k_1+1+k_2)} - L^{(2)}(s)|_{(s=k_1+k_2+1)} + L^{(2)}(s)|_{(s=k_1+k_2)} \quad (3.29)$$

where  $L^{(2)}(., .)$  is Laplace transform of *Gamma*(1, 1) distribution. In addition, the hierarchical models discussed in Chapter 1 can also be used in (3.27). Moreover, the LTs corresponding to GLM, WM, and TGM defined in (1.29), (1.30) and (1.31) tend to (3.27), when dependence parameter  $\rho = 1$  and marginal *Gamma*(1, 1) distribution. The above model is called the complete dependence model.

## 3.5 Simulation study

### 3.5.1 Discrete extended Generalized Pareto distribution

In this section, we discuss a simulation study intended to assess the accuracy of a maximum likelihood estimate (MLE). Different settings of parameters are tried to test each model. Moreover, the scale and upper tail shape parameters are permanently set to  $(\sigma = 1, \xi = 0.2)$  for all four models. The sample size  $n = 1000$  with  $10^4$  replications are used to calculate root mean square errors (RMSEs) corresponding to each model. The remaining parameters of the proposed models are set as follows:

- (i)  $G(u; \psi) = u^\kappa$  with lower tail parameter  $\kappa = 1, 2, 3, 10$ .
- (ii)  $G(u; \psi) = 1 - D_\delta\{(1 - u)^\delta\}$ , with  $\delta = 0.5, 1, 2, 5$ .
- (iii)  $G(u; \psi) = [1 - D_\delta\{(1 - u)^\delta\}]^{\kappa/2}$  with  $\delta = 0.5, 1, 2, 5$  and  $\kappa = 1, 2, 3, 10$ .
- (iv)  $G(u; \psi) = pu^{\kappa_1} + (1 - p)u^{\kappa_2}$  with  $p = 0.5$ ,  $\kappa_1 = 1, 2, 3, 10$  and  $\kappa_2 = 2, 3, 10, 20$ .

The boxplots of the MLEs are constructed to assess the models performance for the above representative cases. Figure 3.4 shows the boxplots of estimated parameters conforming to all four models with different simulation settings. Figure 3.4(a) show MLEs of DEGPD based on parametric family  $G(u; \psi) = u^\kappa$  with true parameter settings (i.e.,  $\kappa = 1, 2, 5, 10$ ,  $\sigma = 1$ , and  $\xi = 0.2$ ). In addition, the horizontal red line in each boxplot represents the true parameters. Figure 3.4(a) indicates that the MLEs of DEGPD (i) are quite reasonable with less variability.

Figure 3.4(b) reports much variability in parameter  $\delta$  estimates when we increase the true value of  $\delta$ . This may happen due to the skewness parameter  $\delta$ . Skewness

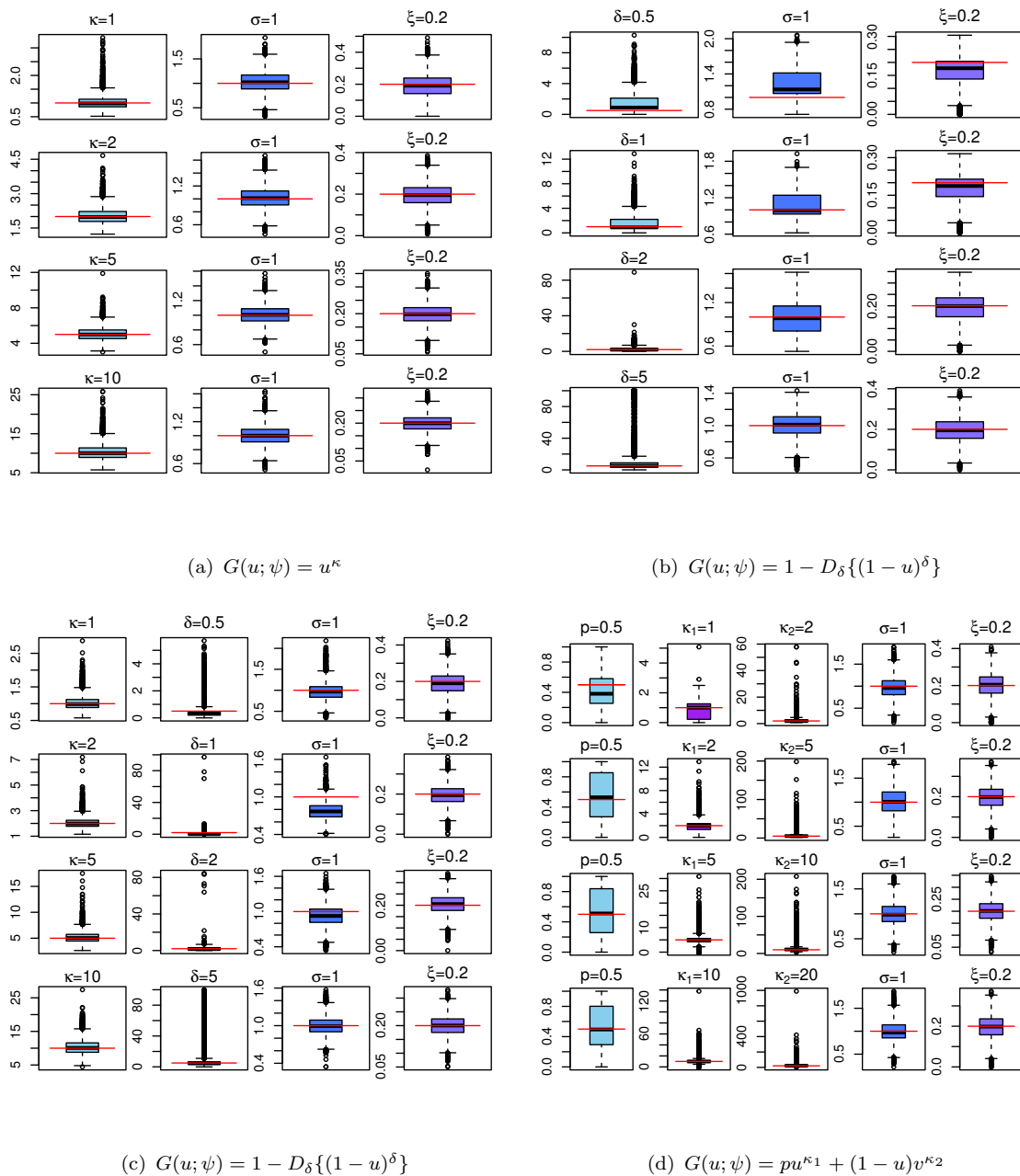


FIGURE 3.4: Boxplots of maximum likelihood estimates of parameters of each model from  $n = 1000$  with  $10^4$  replication at different parameters settings: **(a)** corresponds to model type (i), **(b)** corresponds to model type (ii), **(c)** corresponds to model type (iii), and **(d)** corresponds to model type (iv).

parameters like  $\delta$  and  $\kappa_2$  are hard to estimate by the MLE method; this was observed by Naveau *et al.* (2016) for extended generalized Pareto distributions, Sartori (2006) for the skew-normal and Ribereau *et al.* (2016) for skew generalized extreme value case.

Similarly, Figure 3.4(c) shows that the estimate of parameter  $\delta$  is again outperformed when the true value increases. Figure 3.4(d) shows the MLEs for all parameters are reasonable, with variability seen in  $\kappa_2$  and sometimes in  $\kappa_1$ . Again, this may be due to

the appearance of  $\kappa_2$  as skewness parameter (Naveau *et al.*, 2016).

TABLE 3.1: Root mean square errors of parameter estimates of DEGPD found from  $10^4$  independent data sets of size  $n = 1000$ .

$G(u; \psi) = u^\kappa$									
$\kappa$	RMSE	$\sigma$	RMSE	$\xi$	RMSE				
1	0.24	1	0.21	0.20	0.07				
2	0.35	1	0.16	0.20	0.05				
5	0.76	1	0.12	0.20	0.04				
10	2.05	1	0.13	0.20	0.03				
$G(u; \psi) = 1 - D_\delta\{(1 - u)^\delta\}$									
$\delta$	RMSE	$\sigma$	RMSE	$\xi$	RMSE				
0.5	1.52	1	0.33	0.20	0.06				
1	1.36	1	0.22	0.20	0.06				
2	2.13	1	0.20	0.20	0.06				
5	19.64	1	0.16	0.20	0.06				
$G(u; \psi) = [1 - D_\delta\{(1 - u)^\delta\}]^{\kappa/2}$									
$\delta$	RMSE	$\kappa$	RMSE	$\sigma$	RMSE	$\xi$	RMSE		
0.5	0.42	1	0.20	1	0.20	0.20	0.06		
1	0.94	2	0.39	1	0.20	0.20	0.05		
2	1.94	5	1.01	1	0.19	0.20	0.04		
5	30.15	10	2.33	1	0.13	0.20	0.04		
$G(u; \psi) = pu^{\kappa_1} + (1 - p)u^{\kappa_2}$									
$p$	RMSE	$\kappa_1$	RMSE	$\kappa_2$	RMSE	$\sigma$	RMSE	$\xi$	RMSE
0.5	0.28	1	0.58	2	1.83	1	0.23	0.20	0.06
0.5	0.32	2	0.98	5	6.43	1	0.26	0.20	0.06
0.5	0.32	5	2.13	10	12.19	1	0.23	0.20	0.04
0.5	0.30	10	4.62	20	25.47	1	0.23	0.20	0.04

For further investigation, the RMSEs of model parameters for each configuration are given in Table 3.1. Overall, the findings of the table show that the maximum likelihood estimator performed well for model type (i) when the lower shape parameter  $\kappa$  increases. Model type (ii) highlights that the MLEs are sensible when threshold tuning parameter  $\delta < 5$ . The RMSEs with respect to model type (iii) show that the MLEs are poor when  $\delta > 1$ . Finally, the case of model type (iv) intensifies that the parameters  $\kappa_1$  and  $\kappa_2$  entail much variability, especially when  $\kappa_1 > 2$  and  $\kappa_2 > 5$ .

### 3.5.2 Zero-inflated discrete extended Generalized Pareto distribution

To evaluate the maximum likelihood estimator for ZIDEGPD models, the simulation study has been conducted with different configurations of parameters. The scale and upper tail shape parameters are fixed to  $\sigma = 1$  and  $\xi = 0.2$  for all four models. Like DEGPD, the sample size  $n = 1000$  with  $10^4$  replications are used to calculate RMSEs for each model. The other parameters of the proposed ZIDEGPD are chosen as

- (i)  $G(u; \psi) = u^\kappa$  with lower tail parameter  $\kappa = 5, 10$
- (ii)  $G(u; \psi) = 1 - D_\delta\{(1 - u)^\delta\}$ , with  $\delta = 1, 5$ .
- (iii)  $G(u; \psi) = [1 - D_\delta\{(1 - u)^\delta\}]^{\kappa/2}$  with  $\delta = 1, 5$  and  $\kappa = 5, 10$ .
- (iv)  $G(u; \psi) = pu^{\kappa_1} + (1 - p)u^{\kappa_2}$  with  $p = 0.5$ ,  $\kappa_1 = 1, 5$  and  $\kappa_2 = 5, 10$ .

The zero inflation parameter (i.e., the proportion of zeros)  $\pi$  is considered 0.2 and 0.5 for all models, respectively.

Figure 3.5 clearly shows that the parameter of ZIDEGPD are estimated correctly for  $G(u; \psi) = u^\kappa$  even though when the proportion of zeros is higher. Similar to DEGPD, the estimates of  $\delta$  of ZIDEGPD model (ii),  $\kappa$  and  $\delta$  of ZIDEGPD model (iii) and  $\kappa_1$  and  $\kappa_2$  of ZIDEGPD model (iv) showing more variability. This is already noted for DEGPD cases.

Similar to DEGPD, we check the performance of the maximum likelihood estimator for ZIDEGPD by observing RMSEs of the parameters. We found that the  $\delta$  parameter involved in model type (ii) and model type (iii) and  $\kappa_2$  parameter of model type (iv) entail much variability when estimated by MLE. Similar characteristics have already been noted in DEGPD models. In addition, we found the ZIDEGPD based on  $G(u; \psi) = u^\kappa$  is more reliable than other models. Information regarding RMSEs of ZIDEGPD parameters is reported in Table S1 of "Appendix B".

### 3.5.3 GAM form modeling

To assess the parameter of our proposal in finite samples, we run a simulation study with GAM non-parametric forms. We consider the DEGPD type (i) model with three different settings to explore the complex nonlinear function. The following models are

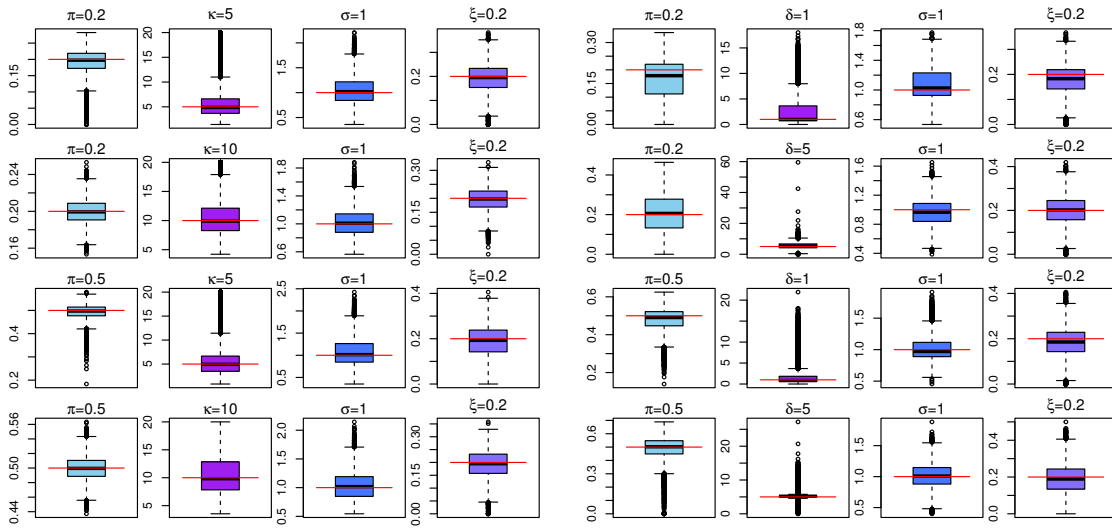
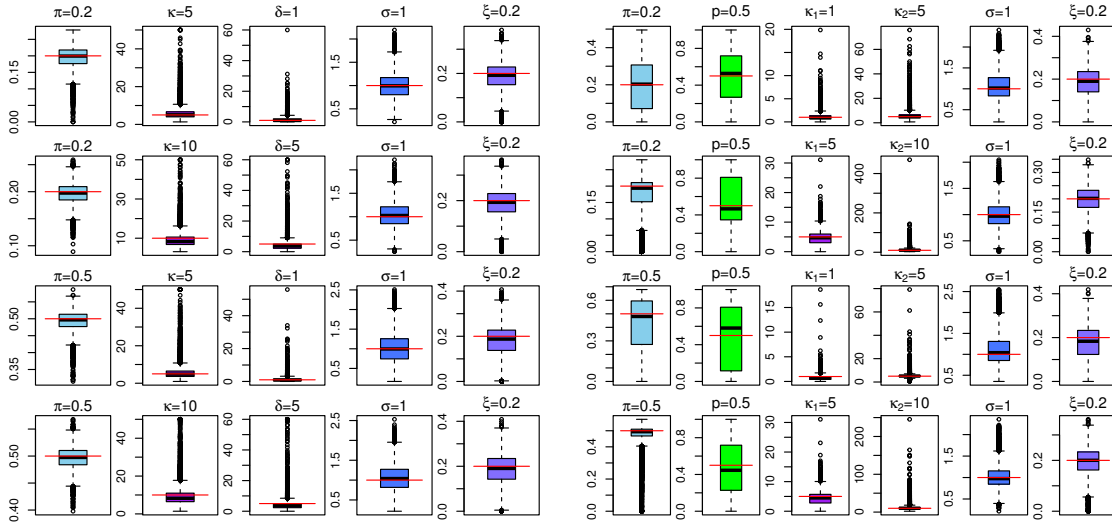
(a)  $G(u; \psi) = u^\kappa$ (b)  $G(u; \psi) = 1 - D_\delta\{(1 - u)^\delta\}$ (c)  $G(u; \psi) = 1 - D_\delta\{(1 - u)^\delta\}$ (d)  $G(u; \psi) = pu^{\kappa_1} + (1 - u)v^{\kappa_2}$ 

FIGURE 3.5: Boxplots of maximum likelihood estimates of parameters of each ZIDEGPD model from  $n = 1000$  with  $10^4$  replication at different parameters settings: **(a)** corresponds to model type (i), **(b)** corresponds to model type (ii), **(c)** corresponds to model type (iii), and **(d)** corresponds to model type (iv).

considered in a simulation study

$$M_1 : \sigma = s(x); \quad \kappa = cst; \quad \xi = cst$$

$$M_2 : \sigma = cst; \quad \kappa = s(x); \quad \xi = cst$$

$$M_3 : \sigma = cst; \quad \kappa = cst; \quad \xi = s(x)$$

where  $s(\cdot)$  indicates the smoothed predictor and covariate  $x$  is generated from  $\mathcal{N}(0.2, 1)$



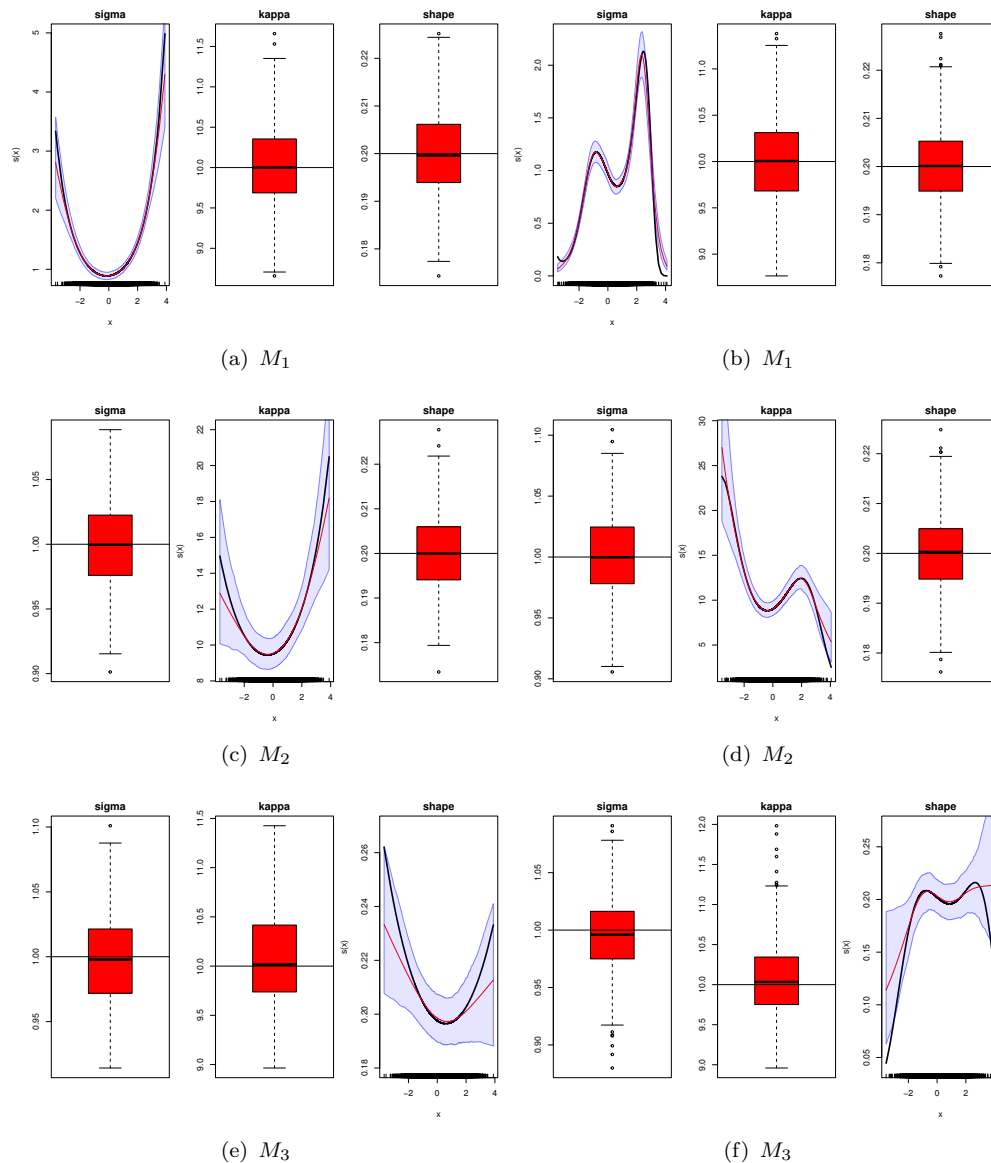


FIGURE 3.6: *Left:* Fitted non-parametric GAM form models with true parameters generated through orthogonal polynomial of order 2, *Right:* Fitted non-parametric GAM form models with true parameters generated from the orthogonal polynomial of order 5. The black solid line in each curve represents the varying true parameters, and the red line represents the estimated one with 95% confidence intervals. The boxplots centered at true values represent the estimates of the fixed parameters in each model.

distribution. We simulated a matrix of true parameters ( $\kappa, \sigma$  and  $\xi$ ) through orthogonal polynomials of order 2 and 5 by supplying  $x$  and  $\beta = (20, 10, 10)$ . Further, we generated DEGPD responses for each model using true parameters with some modifications. For instance, the DEGPD response variable for  $M_1$  is generated using a varying true parameter for  $\sigma$ , keeping ( $\kappa = 10$ ) and ( $\xi = 0.2$ ) constant. Similarly, the DEGPD response variable for  $M_2$  is generated using varying  $\kappa$  parameters by keeping the remaining parameters constant. In contrast, the DEGPD response variable for  $M_3$  is generated using

varying  $\xi$  parameters by keeping the remaining constant. The sample size was set to  $n = 15000$ , which is consistent with real applications. We fitted models based on the maximum likelihood estimator defined in 3.3. For simulation, the number of replication was set to 500.

Figure 3.6 shows the estimates of the non-parametric effect of covariates individually on the parameters of DEGPD type (i). Boxplots centered at true values in each model are constructed for constant parameters. By looking at plots of varying  $\sigma$  parameters in the top panel, corresponding to model  $M_1$ , we see that the non-parametric effect is estimated correctly when compared with the true parameter. Also, the boxplots centered at true values show that constant parameters  $\kappa$  and  $\xi$  are estimated correctly. Considering the covariate in the  $\kappa$  parameter (i.e.,  $M_2$ ), the estimated model seems reasonable for both cases, and constant parameters are also estimated correctly. On the other side, by looking at the bottom plots, corresponding to model  $M_3$ , the non-parameter effects are not estimated correctly when we use the shape parameter  $\xi$  as a function of covariate  $x$  with a complex function. Also, the boxplots of the constant parameters deviate from true values. This result supports the argument that we find in the existing literature that the shape parameter is difficult to estimate when using covariates in GAM (Ranjbar *et al.*, 2022). Overall, the model  $M_1$  performs much better than the others, even when we have more complex non-parametric forms. Similarly, the DEGPD (ii)-(iv) and ZIDEGPD (i)-(iii) are hard to estimate correctly when estimating parameters  $p, \delta, \kappa_2$  and  $\xi$  as a function of covariate  $x$ .

### 3.6 Bivariate discrete extended generalized Pareto distribution

We carried out a simulation study to assess the robustness of the BDEGPD proposed in Section (3.4). In order to simulate realizations from the proposed BDEGPD with different settings of parameters of each type, one can use the following algorithm. The steps of the algorithm to be applied are the following:

- a. Simulate a random variable  $E_1$  and  $E_2$  from exponential distribution with rate parameter is 1. That is,  $E_1 \sim Exp(1)$  and  $E_2 \sim Exp(1)$ .
- b. Simulate  $\Lambda$  variable from Gamma distribution with unit scale and shape parameters. That is  $\Lambda \sim Gamma(1, 1)$
- c. Set  $X_1 = \frac{E_1}{\Lambda}$  and  $X_2 = \frac{E_2}{\Lambda}$ , the resulting variable  $X_1$  and  $X_2$  marginally follow GPD with shape parameter is equal to 1.

- d. Set  $U_1 = H_1(X_1)$  and  $U_2 = H_1(X_2)$ , where  $H_1(X_i), i = 1, 2$  denotes the CDF of GPD with shape parameter equal to 1.
- e. Simulate  $Y_i = \sigma H_\xi^{-1}[G^{-1}(U_i)] \quad i = 1, 2$
- f. Set  $y_1 = \lfloor Y_1 \rfloor$  and  $y_2 = \lfloor Y_2 \rfloor$ , where  $\lfloor \cdot \rfloor$  denotes the floor function.

Furthermore, different parameter settings have tried to evaluate the performance of the maximum likelihood estimator. For instance, we permanently fix the scale and dependence parameters ( $\sigma = 1, \rho = 0.9$ ) for all models associated with G families. In addition, we try the upper shape parameter  $\xi = 0.1, 0.2$ . Again, the sample size  $n = 500$  with  $10^4$  replications are used to calculate RMSEs for each BDEGPD model. As for the remaining parameters related to G families, we use the same values as those used for DEGPD in subsection 3.5.1.

To summarize the simulation study results related to BDEGPD, Table 3.2 reports the RMSEs of MLEs. We are interested in checking the overall fitting of each model and investigating how different settings of the upper shape parameter influence the other parameters. Here, we use the complete dependence model with dependence parameter  $\rho = 1$ . We do not need to report the RMSE of  $\rho$  parameter anymore. It can be observed from Table 3.2 that the RMSEs of Model type (ii), type (iii), and type(iv) show more variability in parameters, especially in  $\delta$  and  $\kappa_2$  when shape parameter  $\xi = 0.1$ . Reasons behind the variable in skewness parameters (i.e.,  $\delta$  and  $\kappa_2$ ) have been discussed prior (see, for instance, subsection 3.5.1). Furthermore, when we increase the shape parameter  $\xi = 0.1$  to  $\xi = 0.2$  with the same specification of other parameters, the RMSEs corresponding to parameters relatively increased for all four models. Again, the same variability is noted in  $\delta$  and  $\kappa_2$ . Overall, the simulation study suggested that all four BDEGPD models are flexible. In addition, BDEGPD corresponding  $G(u; \psi) = u^\kappa$  is considered more favorable for modeling the entire range of bivariate discrete count data.

TABLE 3.2: Root mean square errors of the parameter estimates of BDEGPD found from  $10^4$  independent data sets of size  $n = 500$ . The cells contain RMSEs of parameters  $\kappa/\sigma/\xi$  for BDEGPD associated with family  $G(u; \psi) = u^\kappa$ ,  $\delta/\sigma/\xi$  for BDEGPD associated with family  $G(u; \psi) = 1 - D_\delta\{(1 - u)^\delta\}$ ,  $\delta/\kappa/\sigma/\xi$  for BDEGPD associated with  $G(u; \psi) = [1 - D_\delta\{(1 - u)^\delta\}]^{\kappa/2}$ , and  $p/\kappa_1/\kappa_2/\sigma/\xi$  for BDEGPD linked with  $G(u; \psi) = pu^{\kappa_1} + (1 - p)u^{\kappa_2}$ , respectively.

Model type (i) $G(u; \psi) = u^\kappa$		
$\kappa$	$\xi$	
	0.1	0.2
1	0.48/0.27/0.09	0.71/0.31/0.10
2	0.56/0.23/0.07	0.79/0.27/0.07
5	1.17/0.21/0.05	1.40/0.25/0.05
10	4.32/0.24/0.05	6.06/0.29/0.05
Model type (ii) $G(u; \psi) = 1 - D_\delta\{(1 - u)^\delta\}$		
$\delta$	$\xi$	
	0.1	0.2
0.5	4.83/0.26/0.07	4.17/0.34/0.06
1	10.44/0.20/0.06	14.54/0.23/0.06
2	24.26/0.26/0.06	25.48/0.28/0.06
5	58.47/0.30/0.08	35.35/0.33/0.08
Model type (iii) $G(u; \psi) = [1 - D_\delta\{(1 - u)^\delta\}]^{\kappa/2}$		
$(\delta, \kappa)$	$\xi$	
	0.1	0.2
(0.5, 1)	1.65/0.36/0.25/0.08	1.65/0.38/0.31/0.09
(1, 2)	5.84/0.98/0.29/0.08	3.89/0.90/0.32/0.07
(2, 5)	4.57/1.92/0.29/0.05	8.79/2.57/0.31/0.06
(5, 10)	7.25/2.81/0.25/0.05	7.25/3.41/0.29/0.05
Model type (iv) $G(u; \psi) = pu^{\kappa_1} + (1 - p)u^{\kappa_2}$		
$(\kappa_1, \kappa_2)$	$\xi$	
	0.1	0.2
(1, 2)	0.33/0.59/1.37/0.27/0.07	0.35/0.58/1.31/0.31/0.08
(2, 5)	0.39/1.06/2.48/0.24/0.06	0.39/1.09/2.77/0.29/0.06
(5, 10)	0.38/1.90/4.07/0.24/0.05	0.38/1.84/3.72/0.27/0.05
(10, 20)	0.37/2.96/10.95/0.21/0.04	0.36/2.94/6.36/0.24/0.04

### 3.7 Real data applications

This section discusses the real applications of the proposed models in sections 3.2, 3.3, and 3.4. Firstly we shall consider a dataset on automobile insurance claims. Then we consider the avalanches data of the Haute-Maurienne massif of the French Alps with environmental variables as covariates. For BDEGPD application, we use the avalanches data of the Haute-Maurienne and the Maurienne region of the French Alps.

#### 3.7.1 Discrete extended generalized Pareto distribution

We apply the proposed DEGPD models to the automobile insurance claims data of the companies of New York City recorded from 2009 to 2020. The data is recorded under the Department of Financial Services (DFS) rank of automobile insurance companies running a business in New York State based on the number of consumer complaints upheld against them as a percentage of their total business over two years. Complaints typically include problems like delays in the payment of no-fault claims and nonrenewal of policies. Insurers with the least upheld complaints per million dollars of premiums stand at the top. The data is freely available on the given website <https://www.ny.gov/programs/open-ny>. The frequency distribution of the data (1942 observations) is depicted in Figure 3.7.

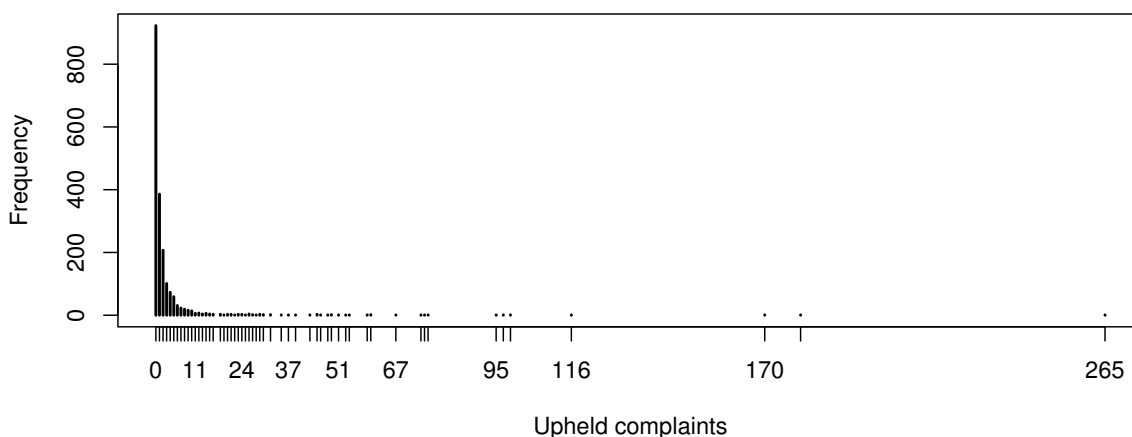


FIGURE 3.7: Frequency distribution of upheld complaints of automobile insurance companies in New York City (2009-2020).

Moreover, the DEGPD and ZIDEGPD based on parametric families (i), (ii), (iii), and (iv) are fitted to the upheld complaints count data. Results of the fitted DEGPD models with their standard errors and bootstrap confidence intervals are given in Table 3.3. AIC and BIC associated with fitted DEGPD and ZIDEGPD models and Chi-square goodness-of-fit test statistic along with p-values are reported in Table 3.4. According to

AIC and BIC values given Table 3.4, the DEGPD (i) and ZIDEGPD (i) with  $G(u; \psi) = u^\kappa$  perform better for this specific data example. DEGPD type (ii) fitting is also quite reasonable to the upheld complaints data with a smaller estimate of the parameter  $\delta$ . However, model type (ii) has restricted flexibility on its lower tail. This is one of the disadvantages of model type (ii) (Naveau *et al.*, 2016). In case of zero-inflation in the data the ZIDEGPD based on  $G(u; \psi) = 1 - D_\delta\{(1 - u)^\delta\}$  may perform better by the reason that it has an additional parameter which represents the zero proportion separately.

TABLE 3.3: Estimated parameters for extended versions of discrete Pareto distribution with all four parametric families fitted to insurance complaints data of New York City. Standard errors are reported between parenthesis. The bootstrap confidence intervals at level 95% are reported between square brackets.

$G(u; \psi) = u^\kappa$				
$\kappa$	$\sigma$	$\xi$		
1.41	0.80	0.73		
(0.37)	(0.20)	(0.05)		
[1.00, 2.35]	[0.43, 1.22]	[0.61, 0.83]		
$G(u; \psi) = 1 - D_\delta\{(1 - u)^\delta\}$				
$\delta$	$\sigma$	$\xi$		
0.006	0.36	0.65		
(0.86)	(0.15)	(0.04)		
[0.00, 1.46]	[0.33, 0.61]	[0.60, 0.80]		
$G(u; \psi) = [1 - D_\delta\{(1 - u)^\delta\}]^{\kappa/2}$				
$\kappa$	$\delta$	$\sigma$	$\xi$	
1.61	0.11	0.49	0.65	
(0.42)	(0.65)	(0.24)	(0.07)	
[1.05, 2.79]	[0.00, 1.68]	[0.27, 1.20]	[0.55, 0.83]	
$G(u; \psi) = pu^{\kappa_1} + (1 - p)u^{\kappa_2}$				
$p$	$\kappa_1$	$\kappa_2$	$\sigma$	$\xi$
0.11	0.01	2.08	0.63	0.73
(0.12)	(0.44)	(2.59)	(0.16)	(0.05)
[0.00, 0.46]	[0.00, 1.59]	[1.44, 10.66]	[0.20, 0.84]	[0.64, 0.83]

The fitting of DEGPD type (iii) with  $G(u; \psi) = [1 - D_\delta\{(1 - u)^\delta\}]^{\kappa/2}$  and DEGPD type (iv) with  $G(u; \psi) = pu^{\kappa_1} + (1 - p)u^{\kappa_2}$  is also quite sensible with lower AIC and BIC value as compared to ZIDEGPD type (iii) and ZIDEGPD type (iv). But the  $\kappa_2$

parameter of DEGPD type (iv) has gained more variability, which is also pointed by Naveau *et al.* (2016) in the continuous framework. In addition, q-q plots given in Figure 3.8 show that all types of DEGPD are fitted reasonably well to upheld complaints data of New York City. Furthermore, the p-values of chi-square test statistic corresponding to each of DEGPD and ZIDEGPD indicate that the fitting of the models proposed in section 3.2 is pretty good for this specific real data example. Based on AIC and BIC, we prefer the DEGPD of type (i) for upheld complaints data.

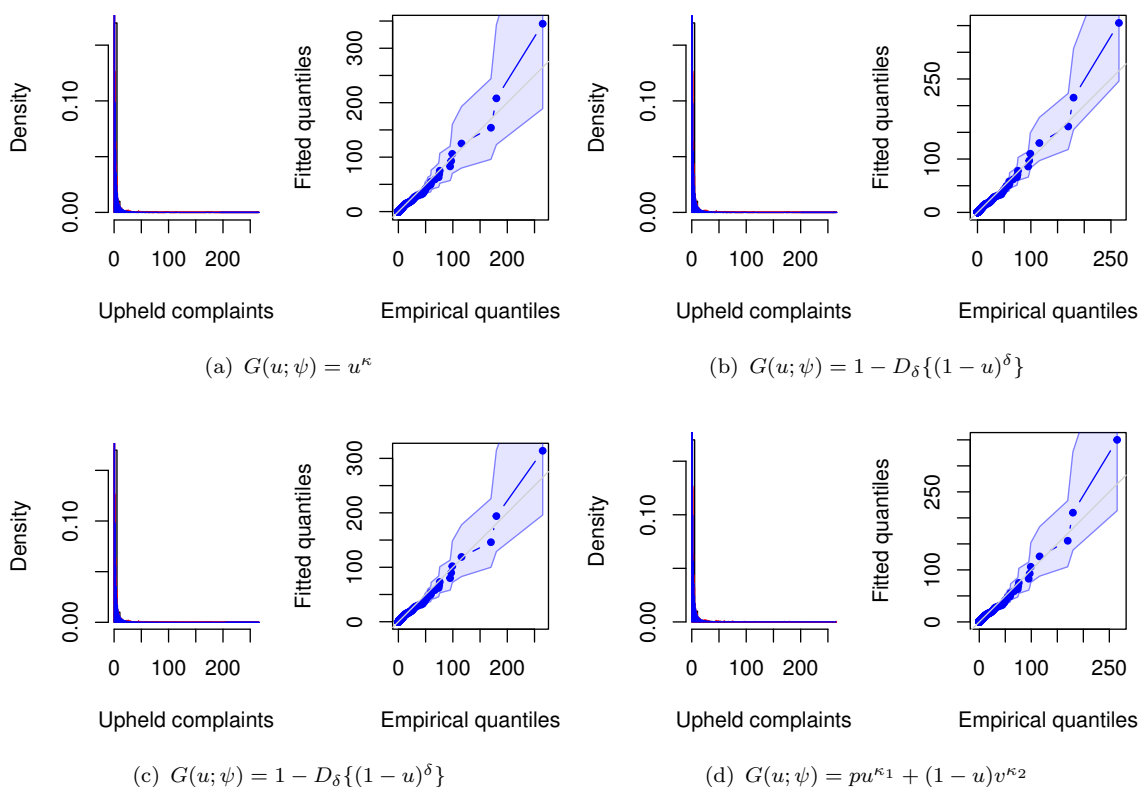


FIGURE 3.8: Quantile-quantile plots of the fitted models with 95% bootstrap-based confidence intervals

### 3.7.2 GAM forms applications of DEGPD and ZIDEGPD to avalanches data

In the Alpine regions, extreme frequency or magnitude snow avalanches are considered a life-threatening hazard. Avalanches are usually caused by severe storms that bring high snowfalls coupled with snow drifting, but strong variations of environmental factors (e.g., temperature, wind, humidity, and precipitation, etc.) causing snow melt and/or fluctuations of the freezing point can also be involved (Evin *et al.*, 2021). It is crucial to anticipate future avalanches activity in the short-term and long-term management.

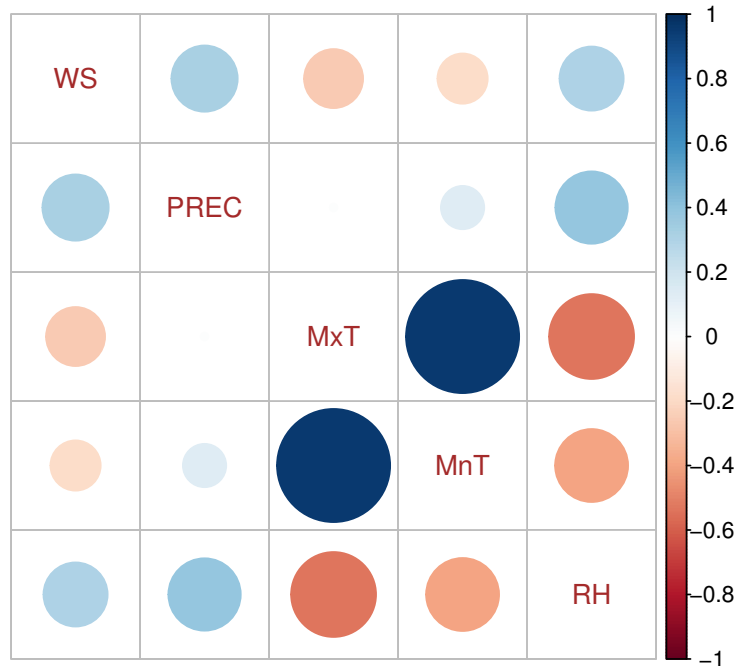


FIGURE 3.9: Correlation among covariates for avalanches data.

Since extreme events have potentially terrible consequences, it is crucial to assess their statistical characteristics correctly. To this end, we try to highlight avalanches events over a short period of time with the help of newly proposed extreme value models. In particular, we intend to quantify how weather-related variables affect the probability of avalanche occurrence each day.

TABLE 3.4: AIC and BIC associated with the fitted DEGPD and ZIDEGPD along with Chi-square goodness-of-fit test. P-values are reported between parenthesis.

Model	AIC		BIC		Chi-square	
	DEGPD	ZIDEGPD	DEGPD	ZIDEGPD	DEGPD	ZIDEGPD
(i)	7290.93	7291.40	7307.65	7313.69	0.20 (0.99)	0.18 (0.99)
(ii)	7291.88	7293.05	7308.60	7315.34	0.19 (0.99)	0.20 (0.99)
(iii)	7293.36	7294.56	7315.64	7322.42	0.20 (0.99)	0.20 (0.99)
(iv)	7294.45	7297.56	7322.30	7330.99	0.20 (0.99)	0.22 (0.99)

The *Enquête Permanente sur les Avalanches* (EPA) collected avalanche data from the French Alps, which has monitored about 3900 paths since the early 20th century



(see Mougín, 1922; Evin *et al.*, 2021). Quantitative (run-out elevations, deposit volumes, etc.) and qualitative (flow regime, snow quality, etc.) information are collected for each event. It varies in quality occasionally, depending on the local observers (mostly forestry rangers). Natural avalanche activity is also uncertain because records tend to record paths visible from valleys, so high-elevation activity may be underestimated.

TABLE 3.5: Detailed information of covariates

Name	Definition
WS	maximum wind speed at 10 meters (m/s)
PREC	precipitation (mm/day)
MxT	maximum temperature at 2 Meters ( $^{\circ}\text{C}$ )
MnT	minimum temperature at 2 Meters ( $^{\circ}\text{C}$ )
RH	relative humidity at 2 Meters (%).

We consider the dataset in Dkengne *et al.* (2016) and the three-day moving sum of the daily number of avalanche events recorded from February 1982 to April 2021. Environmental covariates (see Table 3.5) have been downloaded from <https://power.larc.nasa.gov/data-access-viewer/> by specifying latitude and longitude information. Then, the moving median of the previous three days was considered for each of them.

Figure 3.9 displays a correlation plot among the covariates, highlighting maximum temperature (MxT) and minimum temperature (MnT) are positively strong correlated, while precipitation (PREC) has no significant correlation with MxT. On the other hand, relative humidity (RH) has a moderate positive correlation with wind speed (WS) and PREC, while it has a weak negative correlation with temperature variables. Further, wind speed and precipitation have a weak correlation with minimum and maximum temperature variables. Backward variable selection procedures based on AIC were performed for the DEGPD model under the GAM form, using `evgam` function with our own developed code. A preliminary study showed that a constant model is numerically preferred for lower shape parameters ( $\kappa$  and  $\kappa_1$ ), threshold tuning parameters ( $\delta$  and  $\kappa_2$ ), and upper shape parameter  $\xi$ .

After comparing different combinations of the covariates, we found WS, MxT, PREC and RH are more appropriate to use as covariates. It turned out that the models with

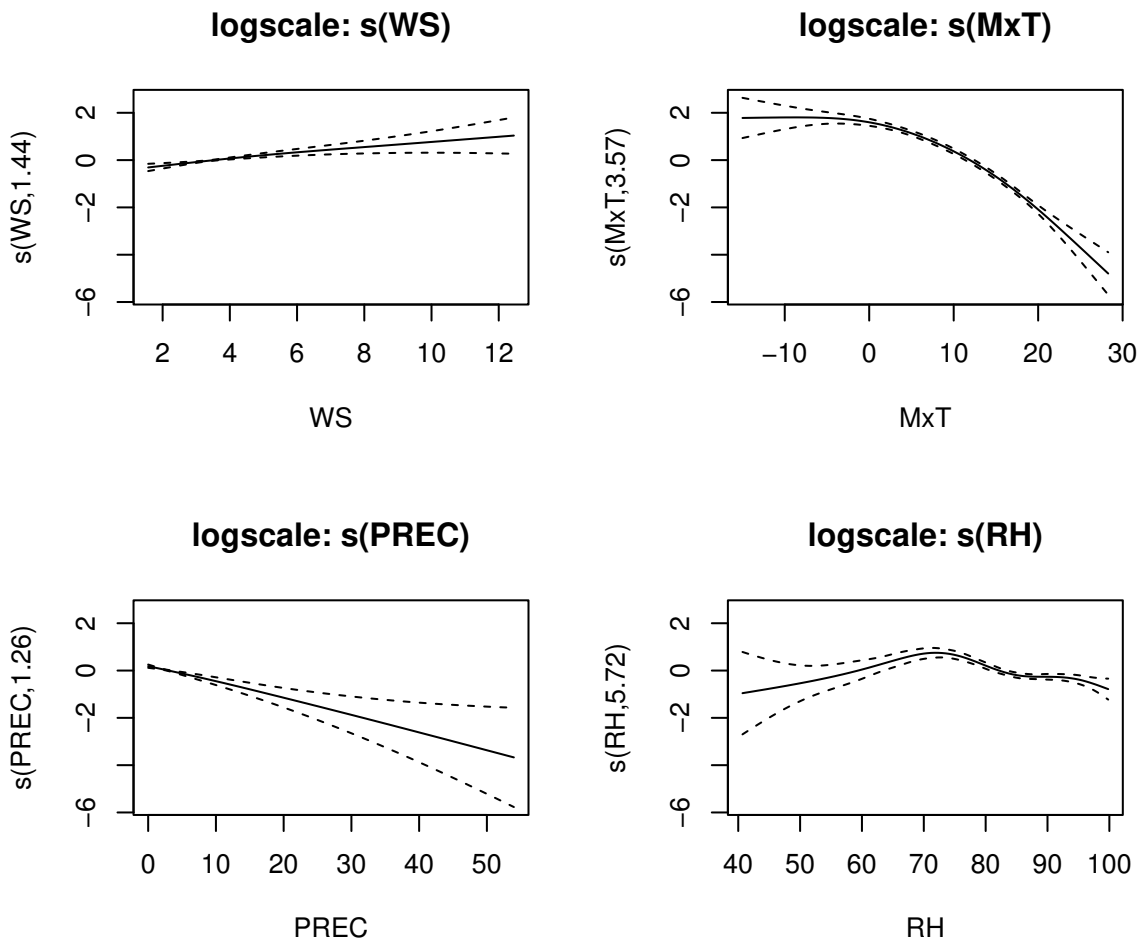


FIGURE 3.10: Estimated non-parametric effects of covariates in the  $\sigma$  component of DEGPD model type (i).

the lowest AIC are

$$\begin{aligned}
 \text{Model type (i)} & : \kappa = cst; \quad \sigma = s(WS) + s(MxT) + s(PREC) + s(RH); \quad \xi = cst \\
 \text{Model type (ii)} & : \delta = cst; \quad \sigma = s(WS) + s(MxT) + s(PREC) + s(RH); \quad \xi = cst \\
 \text{Model type (iii)} & : \kappa = cst; \quad \delta = cst; \quad \sigma = s(WS) + s(MxT) + s(PREC) + s(RH); \\
 & \quad \xi = cst \\
 \text{Model type (iv)} & : p = cst; \quad \kappa_1 = cst; \quad \kappa_2 = cst; \\
 & \quad \sigma = s(WS) + s(MxT) + s(PREC) + s(RH); \quad \xi = cst
 \end{aligned}$$

where  $s(\cdot)$  indicates the smoothed predictor.

We fitted all four DEGPD GAM form models to the response variable (i.e., avalanches counts) with the above-selected covariates. Table 3.6 shows the fitted model results. It can be observed from Table 3.6 that parametric and nonparametric terms for DEGPD GAM form models are statistically significant except for the shape parameter in DEGPD

type (iii). In addition, during estimation, much variability has been seen in constant parameters  $\delta$  and  $\kappa_2$ . This may be due to the appearance of  $\delta$  and  $\kappa_2$  as skewness parameters of the model (Naveau *et al.*, 2016).

Figure 3.10 shows the corresponding estimated functions of DEGPD type (i) model for the regressors that are included as nonparametric terms in the model for the effect of the environmental variables. To this end, the included nonparametric term is significant and has similar behavior for all models. Based on the results of all four models, a broad interpretation of the finding that the temperature and relative humidity seems to better explain the avalanches occurrence as compared to wind and precipitation. The fluctuation in temperature may cause more avalanches coupled with snow drifting. We also fitted ZIDEGPD GAM form models looking at possible zero inflation in the avalanches data.

A slight improvement is noted in ZIDEGPD types (i) and (iii). We found that parameter  $\pi$  in ZIDEGPD type (ii) is insignificant. This is possible due to much variation gained by  $\delta$  parameter. The results of the ZIDEGPD model can be found in Table S2 of "Appendix B".

Further, when comparing our proposed models, the GAM forms DEGPD and ZIDEGPD type (i) and type (iii) overall performed well for the avalanches data. It may be possible the other proposed models perform better to other real data examples.

To assess the overall adequacy of GAM form DEGPD or ZIDEGPD models, we also fitted other existing competitor distributions such as DGPD (1.6) and Poisson distribution. Moreover, we fitted DGPD without threshold selection because the density behavior of the avalanches response is similar to DGPD density. The goodness-of-fit assessment uses the randomized residuals (Dunn and Smyth, 1996; Chiogna and Gaetan, 2007) defined as

$$r_i = \Phi^{-1}((1 - u_i)F(k_i - 1; \hat{\theta}_i) + u_i F(k_i; \hat{\theta}_i)) \quad (3.30)$$

where  $\Phi$  is a standard normal distribution function,  $u_i$  is drawn from a uniform distribution, and  $F(.; \hat{\theta})$  is the parametric estimate of the CDF of the fitted model. Randomization allows obtaining continuous residuals even if the response variable is discrete.

Randomization allows to achieve continuous residuals even if the response variable is discrete. Aside from sampling variation in the parameter estimates, the randomized residuals appear to be exactly normal if the fitted model is correctly specified. Figure 3.11 shows normal quantile-quantile plots of randomized residuals of the proposed GAM form DEGPD type (i) to DEGPD type (iv) and competing models. Graphical representation of residuals of ZIDEGPD type (i) to ZIDEGPD type (iii) models is given in

Figure 3.12. The randomized residuals derived from our proposed models show no apparent departure from normality detected, while randomized residuals based on DGPD and Poisson models departed from normality at the lower and upper tail, respectively.

TABLE 3.6: Estimated coefficients and smooth terms for GAM form DEGPD models fitted to avalanches data.

Model type (i) $G(u; \psi) = u^\kappa$				
Parameter (intercept)	Estimate	Std.Error	t value	P-value
$\log(\kappa)$	-1.83	0.06	-32.38	<2e-16
$\log(\sigma)$	0.2	0.1	2.06	0.0197
$\log(\xi)$	-0.54	0.08	-6.88	2.93e-12
** Smooth terms for $\log(\sigma)$ **				
$\log(\sigma)$	edf	max.df	Chi.sq	Pr(>  t )
s(WS)	1.44	4	22.01	8.66e-06
s(MxT)	3.57	4	663.24	<2e-16
s(PREC)	1.26	4	33.94	2.91e-08
s(RH)	5.72	9	79.14	9.9e-15
Model type (ii) $G(u; \psi) = 1 - D_\delta\{(1 - u)^\delta\}$				
Parameter (intercept)	Estimate	Std.Error	t value	P-value
$\log(\delta)$	4.72	1.28	3.7	0.000109
$\log(\sigma)$	-2.52	0.07	-34.84	<2e-16
$\log(\xi)$	-0.14	0.03	4.94	3.89e-07
** Smooth terms for $\log(\sigma)$ **				
$\log(\sigma)$	edf	max.df	Chi.sq	Pr(>  t )
s(WS)	1.90	4	31.12	3.75e-07
s(MxT)	3.52	4	528.98	<2e-16
s(PREC)	1.27	4	38.59	1.07e-08
s(RH)	6.30	9	59.21	8.61e-11
Model type (iii) $G(u; \psi) = [1 - D_\delta\{(1 - u)^\delta\}]^{\kappa/2}$				
Parameter (intercept)	Estimate	Std.Error	t value	P-value
$\log(\kappa)$	-0.62	0.18	-3.4	0.000339
$\log(\delta)$	4.82	0.67	7.24	2.3e-13
$\log(\sigma)$	-0.62	0.29	-2.15	0.0158
$\log(\xi)$	-0.13	0.09	-1.39	0.0819
** Smooth terms for $\log(\sigma)$ **				
s(WS)	1.78	4	29.29	4.66e-07
s(MxT)	3.61	4	562.36	<2e-16
s(PREC)	1.31	4	49.93	5.2e-10
s(RH)	6.07	9	47.71	1.3e-08

*Cont...*

TABLE 3.6: (*Cont...*) Estimated coefficients and smooth terms for GAM form DEGPD models fitted to avalanches data.

Model type (iv) $G(u; \psi) = pu^{\kappa_1} + (1 - p)u^{\kappa_2}$				
Parameter (intercept)	Estimate	Std.Error	t value	P-value
logit( $p$ )	4.89	0.33	14.69	<2e-16
log( $\kappa_1$ )	-1.84	0.06	-31.88	<2e-16
log( $\kappa_2$ )	3.5	0.41	8.51	<2e-16
log( $\sigma$ )	0.17	0.1	1.66	0.0484
log( $\xi$ )	-0.86	0.13	-6.78	6.07e-12
** Smooth terms for log( $\sigma$ ) **				
s(WS)	1.67	4	19.68	3.08e-05
s(MxT)	3.57	4	623.83	<2e-16
s(PREC)	1.05	4	35.18	4.71e-09
s(RH)	5.67	9	80.50	5.31e-15

We further check the normality of randomized residuals by using Kolmogorov–Smirnov test. Table 3.7 indicates that the newly proposed models considered for GAM form modeling deliver a good fit for the avalanches. among the different types, DEGPD type (i) and DEGPD type (iv) clearly outperform than others, as shown in Figure 3.11 as well. Furthermore, the randomized residuals from GPD and Poisson distribution did not meet the normality assumption.

TABLE 3.7: Kolmogorov–Smirnov (KS) test statistics (p-values between parentheses) of the proposed DEGPD type (i) to type (iv) and competitor (DGPD and Poisson) distributions.

Model	DEGPD	DGPD	Poisson
(i)	0.0038 (0.9855)	0.0132 (0.0130)	0.21048 (2.2e-16)
(ii)	0.0106 (0.0779)		
(iii)	0.0098 (0.1958)		
(iv)	0.0066 (0.5590)		

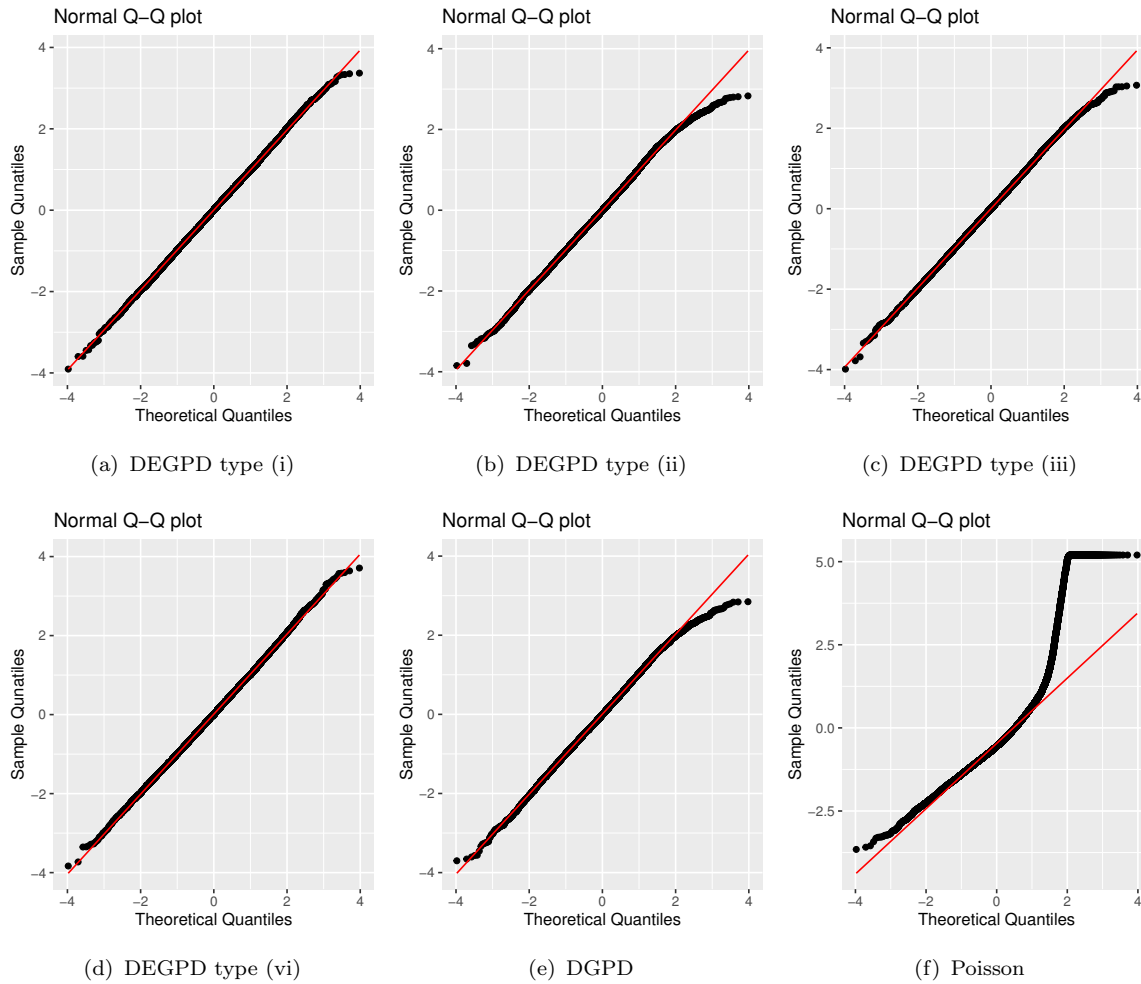


FIGURE 3.11: Diagnostic plots of residual quantiles of the proposed DEGPD type (i) to type (iv) and competitor (DGPD and Poisson) distributions.

### 3.8 Bivariate discrete extended generalized Pareto distribution

We fitted four different BDEGPD models described in section 3.4 to the data of Haute-Maurienne and Maurienne massifs of the French Alps using the MLE procedure. Results of the fitted models with AIC and BIC are summarized in Table 3.8. By comparing AIC and BIC, the model associated with  $G(u; \psi) = [1 - D_\delta\{(1 - u)^\delta\}]^{\kappa/2}$  in (3.29) performs pretty good overall. Similarly, the Model type (iv)  $G(u; \psi) = pu^{\kappa_1} + (1 - p)u^{\kappa_2}$  fits quite well, but it has five parameters. As for the model based on  $G(u; \psi) = 1 - D_\delta\{(1 - u)^\delta\}$  seems to be the poorest fit, it also lacks flexibility in the lower tail (Naveau *et al.*, 2016), as we also note in a simulation study. Model type (i) with  $G(u; \psi) = u^\kappa$  in (3.29) is frequently comparable.

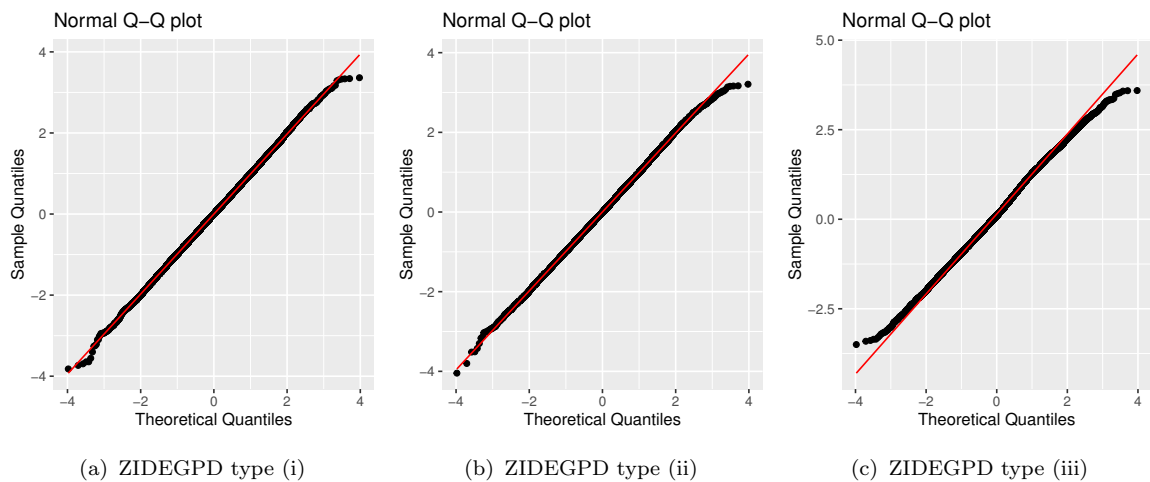


FIGURE 3.12: Diagnostic plots of residual quantiles for the proposed ZIDEGPD type (i) to ZIDEGPD type (iii) models.

### 3.9 Final remarks

This chapter proposes different versions of DEGPD and ZIDEGPD models to demonstrate that it can jointly model the entire range of count data without selecting a threshold. Further, the DEGPD and ZIDEGPD GAM form models are developed and implemented. The flexibility of these models and their many practical advantages in discrete nature data make them very attractive. A few parameters make it simple to implement, interpret, and comply with discrete EVT for both upper and lower tails. The inference is performed through the MLE procedure, which shows more adequacy in results. Compared to ZIDEGPD, the fitted DEGPD appears more straightforward, robust, and genuinely represents zero proportion in the upheld complaints data of NYC. We observed that our proposed ZIDEGPD models are more flexible and robust for the data with zero inflation, and the remaining observations have risen in the lower tail up to structural mode and exponential decay at the upper heavier tail. As noted in the simulation study, the parameters  $\delta$  and  $\kappa_2$  gained more variability when estimated through the MLE method; Bayesian analysis with informative priors may improve the estimates of these parameters.

In addition, we developed and implemented the GAM forms methodology of our proposed models that allow for non-identically distributed discrete extremes. This methodology was implemented in `evgam` using the author's written R functions. The response variable of interest (three-day moving sum of daily avalanches at Haute-Maurienne massif of French Alps) is statistically explained by other environmental variables (e.g., temperature, wind, precipitation and humidity). GAM form models proposed in this study

TABLE 3.8: Estimated parameters for bivariate extended discrete generalized Pareto distribution with all four parametric families fitted to Haute-Maurienne and Maurienne massifs of French Alps. Standard errors are reported between parenthesis. The bootstrap confidence intervals at level 95% are reported between square brackets.

Model type (i) $G(u; \psi) = u^\kappa$						
$\kappa$	$\sigma$	$\xi$	-	-	AIC	BIC
2.79	1.02	0.52	-	-	2624.19	2640.41
(0.77)	(0.42)	(0.05)				
[1.85, 4.02]	[0.02, 1.51]	[0.45, 0.70]				
Model type (ii) $G(u; \psi) = 1 - D_\delta\{(1 - u)^\delta\}$						
$\delta$	$\sigma$	$\xi$	-	-		
1.09	1.42	0.52	-	-	2647.22	2663.45
(0.43)	(0.15)	(0.04)				
[0.42, 1.83]	[1.15., 1.67]	[0.42, 0.60]				
Model type (iii) $G(u; \psi) = [1 - D_\delta\{(1 - u)^\delta\}]^{\kappa/2}$						
$\kappa$	$\delta$	$\sigma$	$\xi$	-		
0.64	6.68	0.40	0.62	-	<b>2587.10</b>	<b>2607.38</b>
(0.42)	(3.65)	(0.38)	(0.07)			
[0.01, 1.79]	[2.44, 14.68]	[0.08, 1.47]	[0.46, 0.71]			
Model type (iv) $G(u; \psi) = pu^{\kappa_1} + (1 - p)u^{\kappa_2}$						
$p$	$\kappa_1$	$\kappa_2$	$\sigma$	$\xi$		
0.01	2.16	6.82	0.46	0.61	2587.14	2611.48
(0.13)	(5.13)	(7.07)	(0.35)	(0.06)		
[0.00, 0.30]	[0.00, 14.32]	[2.13, 20.43]	[0.19, 1.38]	[0.42, 0.65]		

allows parametric non-parametric functional forms, which would most likely be required for larger datasets. Our models (especially DEGPD and ZIDEGPD type (i)) also show more flexibility and a good fit for avalanches data with the effect of environmental conditions as covariates than other competing models (i.e., DGPD, negative binomial, and Poisson). It is worth mentioning that GAM form DEGPD models may perform better than the other real data example. Again, GAM form ZIDEGPD models are more flexible and adequate when the response variable has zero inflation, and the remaining observations have an exponential rise in the lower tail till mode and then decay at the upper heavier tail.



# Conclusions

## Discussion

In this dissertation, we have considered the problem of capturing different types of tail behavior in discrete extreme values, arguing that more tools are needed to support an approach where one may want to take into account the asymptotic dependence of extremes, including asymptotic independence. We have focused on hierarchical models with Gamma variables or stationary Gamma processes with Gamma marginals to induce simple or temporal dependence among discrete threshold exceedances. Optimal threshold selection that defines exceedances remains a problematic issue. Considering this issue also, we have focused on the idea of using a smooth transition between the two tails (lower and upper) to force large and small discrete extreme values to comply with EVT. This thesis further focuses on the models with an additional parameter representing the proportion of zero values in the data in the case of zero inflation.

In particular, chapter 1 proposes a model that allows us to assess the changes in the extremal dependence structure over the tail. We introduce different types of dependence in the proposed model using hierarchical settings and copula-based constructions. This framework is the first step toward modeling bivariate discrete extremes with asymptotic dependence or independence. Based on the tail dependence measure, we can achieve asymptotic dependence and asymptotic independence in bivariate and copula-based models with an appropriate choice of the hierarchical model having Gamma marginals. Generally speaking, asymptotic dependence and asymptotic independence can be tackled with our proposal explained in chapter 1. In addition, the proposed model was applied to the avalanches count data of two massifs of the French Alps.

Model-based and empirical estimates of the tail dependence measure (developed in section 1.3 of the chapter 1) were compared based on conditional probabilities for the estimated models, namely CDM, GM, KM, TCM, CCDM, CGM, CKM, and CTGM. A decreasing degree of dependence can be seen in empirical values, and this corresponds to convergence to asymptotic independence as  $k \rightarrow \infty$ . In the proposed formulation, the copula-based model is preferable over simple models for the avalanches data of the

French Alps; however, all models produce more stable estimates than the empirical counterparts as  $k$  increases.

Chapter 2 is intended to discuss a class of models for time series with discrete nature extremes. We developed a discrete extreme value model based on the latent process framework to obtain a flexible dependence structure. As a result of this construction, the marginal DGPD is preserved with a positive shape parameter, which is reliable for modeling extreme value exceedances. In time series modeling, the latent Markov chains with a hierarchical setting were used to induce temporal dependence in exceedances. In the developed models, one notable feature is that the temporal dependence structure varies considerably as one moves further into the right tail of the marginal distribution. To this end, the theoretical constructions of the tail dependence measure demonstrated that both asymptotic dependence and asymptotic independence of discrete time series extremes are possibly achievable with an appropriate choice of the latent process  $\Lambda_t$ . Moreover, a measure of tail dependence ( $\chi$ ) is developed to evaluate the extreme dependence of the proposed models.

Chapter 3 proposes different versions of DEGPD and ZIDEGPD models to demonstrate that they can jointly model the entire range of count data (low and high extreme) without selecting a threshold. In addition, GAM models for DEGPD and ZIDEGPD are developed and implemented. Discrete data models developed in chapter 3 are advantageous due to their flexibility and practicality. Fewer parameters make them simple to implement, interpret, and comply with discrete EVT for both upper and lower tails. It appears that DEGPD fits the upheld complaints data of NYC more robustly and straightforwardly than ZIDEGPD. Accordingly, we found that ZIDEGPD models are more flexible and robust for the data with zero inflation.

GAM form models proposed for the entire range of discrete extremes allows parametric non-parametric functional forms, which would most likely be required for larger datasets. Models (especially DEGPD and ZIDEGPD type (i)) show more flexibility and a good fit for avalanches data with the effect of environmental conditions as covariates than other competing models (i.e., DGPD, negative binomial, and Poisson). It is worth mentioning here that the other GAM form DEGPD models may perform better than existing models when applied to other real data examples.

Hence, our proposed models can be applied to discrete count data with extreme observations. Further, GAM forms proposals are flexible regarding spatial modeling of discrete extremes.

## Future directions of research

Interestingly, the model proposed in chapter 1 can be extended to the multivariate case (Bacro *et al.*, 2023), but it is challenging. On the other side, dependence can be considered through Gaussian graphical models (Hitz and Evans, 2016) when working with high dimensional data, which are more feasible with inverse covariance matrix to reduce dimensionality and to visualize the dependence structure as a graph. Following are some other possibilities for continuing the proposed models.

**Proposition:** *Let  $\Lambda = (\Lambda_1, \Lambda_2)$  have a bivariate distribution with gamma margins. Suppose, given  $\Lambda = (\Lambda_1, \Lambda_2)$ ,  $Y_i, i = 1, 2$  are independent geometric random variables with parameter  $q_i = 1 - e^{-(\omega + \lambda_i)/\beta}, i = 1, 2$ , where  $\beta > 0$ . Then  $\mathbf{Y} = (Y_1, Y_2)$  follows a bivariate distribution with New geometric Discrete Pareto distribution marginals, which was introduced by Bhati and Bakouch (2019). Bivariate LTs of bivariate gamma distributions are supportive of defining the bivariate distribution.*

**Proposition:** *Let  $\Lambda = (\Lambda_1, \Lambda_2)$  have a bivariate distribution with gamma margins. Suppose, given  $\Lambda = (\lambda_1, \lambda_2)$ ,  $Y_i, i = 1, 2$  are independent Zero Modified geometric random variables (i.e ZMG( $\pi_i, q_i$ ), where  $q_i = 1 - e^{-\lambda_i/\beta}, i = 1, 2$ , with  $\beta > 0$ . Then  $\mathbf{Y} = (Y_1, Y_2)$  follows a bivariate distribution with Zero Modified Discrete Pareto distribution marginals which was introduced by Constantinescu et al. (2019). Again, bivariate LTs of bivariate Gamma distributions are useful to define the bivariate distribution.*

On the other side, models based on chapter 2 can further be modified in different ways; see, for instance, section 2.7 of chapter 2.

Models proposed in chapter 3 can be applied to the variables with discrete count data having extreme observation. Further, GAM forms proposals are flexible regarding spatial-temporal modeling of discrete extremes. In addition, it may be possible to extend DEGPD theoretically using r-Pareto processes (Huser and Wadsworth, 2020) and use it to modeling of discrete spatial extremes.



# Appendix

## Appendix A

### A.1 Proofs of propositions

*Proof. of Proposition 1.1:* Let  $Y_i, i = 1, 2$  are independent Geometric random variables with parameter  $q_i = 1 - e^{-\Lambda_i/\beta}, i = 1, 2$ , where  $\beta > 0$ . The SF of Geometric of distribution when  $Y_i, i = 1, 2$  are independent Geometric random variables

$$P(Y_i > k_i | \Lambda) = S(k_i | \Lambda) = e^{-\frac{\Lambda k_i}{\beta}}, \quad i = 1, 2 \quad (\text{A.1})$$

where  $\Lambda$  follows standard Gamma marginal distribution having PMF is

$$f(\lambda, \alpha, \beta) = [\beta^\alpha / \Gamma(\alpha)] \lambda^{\alpha-1} e^{-\beta\lambda}. \quad (\text{A.2})$$

Without loss of generality, the joint SF function we have

$$P(Y_1 > k_1, Y_2 > k_2 | \Lambda) = S(k_1, k_2 | \Lambda) = E \left( e^{-\frac{\Lambda k_1}{\beta} - \frac{\Lambda k_2}{\beta}} \right) \quad (\text{A.3})$$

$$S(k_1, k_2) = \int_0^\infty \int_0^\infty e^{-\frac{k_1 \lambda}{\beta} - \frac{k_2 \lambda}{\beta}} f(\lambda) d\lambda = L^{(1)}(s) \Big|_{(s_1 = \frac{k_1 + k_2}{\beta})} \quad (\text{A.4})$$

where  $S(k_1, k_2)$  is the joint SF function of a bivariate distribution that has DGPD margins and the expression  $L^{(1)}(s)$  is the univariate Laplace transform of  $\Lambda = \Lambda$ .

The PMF of the bivariate distribution having DGPD margins is written by using the joint SF as

$$P(Y_1 = k_1, Y_2 = k_2) = S(k_1, k_2) - S(k_1 + 1, k_2) - S(k_1, k_2 + 1) + S(k_1 + 1, k_2 + 1) \quad (\text{A.5})$$

$$P(k_1, k_2) = L^{(1)}(s)|_{(s=\frac{k_1+k_2}{\beta})} - L^{(1)}(s)|_{(s=\frac{k_1+1+k_2}{\beta})} - L^{(1)}(s)|_{(s=\frac{k_1+k_2+1}{\beta})} + L^{(1)}(s)|_{(s=\frac{k_1+1+k_2+1}{\beta})} \quad (\text{A.6})$$

This completes the proof.  $\square$

*Proof. of Proposition 1.2* Let  $Y_i, i = 1, 2$  are independent Geometric random variables with parameter  $q_i = 1 - e^{-\Lambda_i/\beta}, i = 1, 2$ , where  $\beta > 0$ . The SF of Geometric of distribution when  $Y_i, i = 1, 2$  are independent Geometric random variables

$$P(Y_i > k_i | \Lambda_i) = S(k_i | \Lambda_i) = e^{-\frac{\lambda_i k_i}{\beta}}, \quad i = 1, 2 \quad (\text{A.7})$$

where  $\Lambda_i$  follow Gamma marginal distribution having probability mass function given in A.2. Without loss of generality, the joint SF function we have

$$P(Y_1 > k_1, Y_1 > k_1 | \Lambda_1, \Lambda_2) = S(k_1, k_2 | \Lambda_1, \Lambda_2) = E\left(e^{-\frac{\Lambda_1 k_1}{\beta} - \frac{\Lambda_2 k_2}{\beta}}\right) \quad (\text{A.8})$$

$$S(k_1, k_2) = \int_0^\infty \int_0^\infty e^{-\frac{k_1 \lambda_1}{\beta} - \frac{k_2 \lambda_2}{\beta}} f(\lambda_1, \lambda_2) d\lambda_1 d\lambda_2 = L^{(2)}(s_1, s_2)|_{(s_1=\frac{k_1}{\beta}, s_2=\frac{k_2}{\beta})} \quad (\text{A.9})$$

where  $S(k_1, k_1)$  is the joint SF function of a bivariate distribution that has DGPD margins and the expression  $L^{(2)}(s_1, s_2)$  is the bivariate Laplace Transform of  $\Lambda = (\Lambda_1, \Lambda_2)$ . The bivariate PMF can be obtained using the definition (A.5).

$$P(k_1, k_2) = L^{(2)}(s_1, s_2)|_{(s_1=\frac{k_1}{\beta}, s_2=\frac{k_2}{\beta})} - L^{(2)}(s_1, s_2)|_{(s_1=\frac{k_1+1}{\beta}, s_2=\frac{k_2}{\beta})} - L^{(2)}(s_1, s_2)|_{(s_1=\frac{k_1}{\beta}, s_2=\frac{k_2+1}{\beta})} + L^{(2)}(s_1, s_2)|_{(s_1=\frac{k_1+1}{\beta}, s_2=\frac{k_2+1}{\beta})}$$

This completes the proof.  $\square$

*Proof. of Proposition 1.3* Let  $Y_i, i = 1, 2$  are independent Geometric random variables with parameter  $q_i = 1 - e^{-\frac{\Lambda}{\beta}}, i = 1, 2$ . Given  $\Lambda = \lambda$  follow a standard Gamma distribution with PDF given in (A.2). Thus, the SFs of Geometric distribution corresponding random  $Y_1$  and  $Y_2$  are  $S(Y_1 > k_1 | \Lambda = \lambda) = e^{-\frac{\lambda k_1}{\beta}}$  and  $S(Y_2 > k_2 | \Lambda = \lambda) = e^{-\frac{\lambda k_2}{\beta}}$ . The joint SF of bivariate Geometric distribution developed through FGMC is given by

$$S(k_1^*, k_2^* | \lambda) = \left[ e^{-\left(\frac{\lambda k_1}{\beta} + \frac{\lambda k_2}{\beta}\right)} \right] \left[ 1 + \phi \left( 1 - e^{-\frac{\lambda k_1}{\beta}} - e^{-\frac{\lambda k_2}{\beta}} + e^{-\left(\frac{\lambda k_1}{\beta} + \frac{\lambda k_2}{\beta}\right)} \right) \right] \quad (\text{A.10})$$

$$S(k_1^*, k_2^* | \lambda) = \left[ e^{-\left(\frac{\lambda k_1}{\beta} + \frac{\lambda k_2}{\beta}\right)} + \phi \left( e^{-\left(\frac{\lambda k_1}{\beta} + \frac{\lambda k_2}{\beta}\right)} - e^{-\left(\frac{2\lambda k_1}{\beta} + \frac{\lambda k_2}{\beta}\right)} - e^{-\left(\frac{\lambda k_1}{\beta} + \frac{2\lambda k_2}{\beta}\right)} + e^{-\left(\frac{2\lambda k_1}{\beta} + \frac{2\lambda k_2}{\beta}\right)} \right) \right] \quad (\text{A.11})$$

Using definition A.3, and we get

$$S(k_1, k_1) = \int_0^\infty \left[ e^{-\left(\frac{\lambda k_1}{\beta} + \frac{\lambda k_2}{\beta}\right)} + \phi \left( e^{-\left(\frac{\lambda k_1}{\beta} + \frac{\lambda k_2}{\beta}\right)} - e^{-\left(\frac{2\lambda k_1}{\beta} + \frac{\lambda k_2}{\beta}\right)} - e^{-\left(\frac{\lambda k_1}{\beta} + \frac{2\lambda k_2}{\beta}\right)} + e^{-\left(\frac{2\lambda k_1}{\beta} + \frac{2\lambda k_2}{\beta}\right)} \right) \right] f(\lambda) d\lambda \quad (\text{A.12})$$

Later we write in form of LTs

$$S(k_1, k_1) = L^{(1)}(s)|_{(s=\frac{k_1+k_2}{\beta})} + \phi \left( L^{(1)}(s)|_{(s=\frac{k_1+k_2}{\beta})} - L^{(1)}(s)|_{(s=\frac{2k_1+k_2}{\beta})} - L^{(1)}(s)|_{(s=\frac{k_1+2k_2}{\beta})} + L^{(1)}(s)|_{(s=\frac{2k_1+2k_2}{\beta})} \right) \quad (\text{A.13})$$

where  $\phi$  is the copula dependence parameter and  $L^{(1)}(\cdot)$  is the LT of random variable  $\Lambda$ .  $S(k_1, k_2)$  is a joint SF of copula-based bivariate distribution with DGPD marginals. The bivariate PMF can be obtained using the definition (A.5). This completes the proof.  $\square$

*Proof. of Proposition 1.4* Let  $Y_i$  (for  $i = 1, 2$ ) are independent Geometric random variables with parameter  $q_i = 1 - e^{-\frac{\Lambda_i}{\beta}}, i = 1, 2$ . Given  $\mathbf{\Lambda} = (\Lambda_1, \Lambda_2)$  follow a bivariate Gamma distribution with Gamma margins. Thus, the SFs of Geometric distribution corresponding random  $Y_1$  and  $Y_2$  are  $S(Y_1 > k_1 | \Lambda_1 = \lambda_1) = e^{-\frac{\lambda_1 k_1}{\beta}}$  and  $S(Y_2 > k_2 | \Lambda_2 = \lambda_2) = e^{-\frac{\lambda_2 k_2}{\beta}}$ . The joint SF of bivariate Geometric distribution developed via FGMC is given by

$$S(k_1^*, k_1^* | \lambda_1, \lambda_2) = \left[ e^{-\left(\frac{\lambda_1 k_1}{\beta} + \frac{\lambda_2 k_2}{\beta}\right)} + \phi \left( e^{-\left(\frac{\lambda_1 k_1}{\beta} + \frac{\lambda_2 k_2}{\beta}\right)} - e^{-\left(\frac{2\lambda_1 k_1}{\beta} + \frac{\lambda_2 k_2}{\beta}\right)} - e^{-\left(\frac{\lambda_1 k_1}{\beta} + \frac{2\lambda_2 k_2}{\beta}\right)} + e^{-\left(\frac{2\lambda_1 k_1}{\beta} + \frac{2\lambda_2 k_2}{\beta}\right)} \right) \right] \quad (\text{A.14})$$

Using (A.9) and we get the joint SF of copula based bivariate distribution which as DGPD marginals

$$S(k_1, k_1) = L^{(2)}(s_1, s_2)|_{(s_1=\frac{k_1}{\beta}, s_2=\frac{k_2}{\beta})} + \phi \left( L^{(2)}(s_1, s_2)|_{(s_1=\frac{k_1}{\beta}, s_2=\frac{k_2}{\beta})} - L^{(2)}(s_1, s_2)|_{(s_1=\frac{2k_1}{\beta}, s_2=\frac{k_2}{\beta})} - L^{(2)}(s_1, s_2)|_{(s_1=\frac{k_1}{\beta}, s_2=\frac{2k_2}{\beta})} + L^{(2)}(s_1, s_2)|_{(s_1=\frac{2k_1}{\beta}, s_2=\frac{2k_2}{\beta})} \right) \quad (\text{A.15})$$

where  $\phi$  is the copula dependence parameter and where  $L^{(2)}(\cdot, \cdot)$  is the bivariate Laplace transform of  $\mathbf{\Lambda} = (\Lambda_1, \Lambda_2)$ .  $S(k_1, k_2)$  is a joint SF of copula-based bivariate distribution with DGPD marginals. The bivariate PMF can be obtained using the definition (A.5). This completes the proof.  $\square$

## A.2 Properties of conditional tail dependence

### A.2.1 Gaver model

To evaluate the dependence between the exceedances of  $Y_1$  and  $Y_2$  variables, we use the definition of limiting measure ( $\chi$ ) found in Coles *et al.* (1999). That is

$$\chi = \lim_{k \rightarrow \infty} \frac{P(Y_1 \geq k+1, Y_2 \geq k+1)}{P(Y_1 \geq k+1)} \quad (\text{A.16})$$

To obtain the limiting measure  $\chi$  for GM, we solve (A.16) by incorporating univariate and bivariate LTs of Gamma distribution associated with GM, provided in section 1.2 of chapter 1 of the thesis, that is,

$$\begin{aligned} \chi &= \lim_{k \rightarrow \infty} \frac{L^{(2)}(s_1, s_2)|_{(s_1=\frac{k+1}{\beta}, s_2=\frac{k+1}{\beta})}}{L^{(1)}(s)|_{(s=\frac{k+1}{\beta})}} \\ &= \lim_{k \rightarrow \infty} \frac{\left[ \frac{(\beta+\rho)\beta}{(\beta+k+1)(\beta+k+1+\rho k+1)} \right]^\alpha}{\left[ \frac{\beta}{\beta+k+1} \right]^\alpha} \\ &= \lim_{k \rightarrow \infty} \left[ \frac{\beta(k+1+\beta)\{\beta+\rho(k+1)\}}{\beta(k+1+\beta)(\beta+k+1+\rho(k+1))} \right]^\alpha \\ &= \lim_{k \rightarrow \infty} \left[ \frac{\{\beta+\rho(k+1)\}}{(\beta+(k+1)+\rho(k+1))} \right]^\alpha \\ &= \lim_{k \rightarrow \infty} \left[ \frac{\beta+\rho k+\rho}{\beta+k+1+\rho k+\rho} \right]^\alpha \\ &= \lim_{k \rightarrow \infty} \left[ \frac{\mathcal{K}(\frac{\beta}{k}+\rho+\frac{\rho}{k})}{\mathcal{K}(\frac{\beta}{k}+1+\frac{1}{k}+\rho+\frac{\rho}{k})} \right]^\alpha \\ &= \left[ \frac{\rho}{1+\rho} \right]^\alpha \end{aligned} \quad (\text{A.17})$$



In order to gain a finer characterization of the joint tail decay rate under asymptotic independence, the  $\bar{\chi}$  index is defined by following the definition of Coles *et al.* (1999) as

$$\begin{aligned}
\bar{\chi} &= \lim_{k \rightarrow \infty} \frac{2 \log P(Y_1 \geq k+1)}{\log P(Y_1 \geq k+1, Y_2 \geq k+1)} - 1 \\
&= \lim_{k \rightarrow \infty} \frac{2 \log L^{(1)}(s)|_{(s=\frac{k+1}{\beta})}}{\log L^{(2)}(s_1, s_2)|_{(s_1=\frac{k+1}{\beta}, s_2=\frac{k+1}{\beta})}} - 1 \\
&= \lim_{k \rightarrow \infty} \frac{2 \log \left[ \frac{\beta}{(k+1+\beta)} \right]^\alpha}{\log \left[ \frac{\beta\{\beta+\rho(k+1)\}}{(k+1+\beta)(\beta+(k+1)+\rho(k+1))} \right]^\alpha} - 1 \\
&= \lim_{k \rightarrow \infty} \frac{2 [\log(\beta) - \log(k+1+\beta)]}{\log(\beta) + \log\{\beta + \rho(k+1)\} - \log(k+1+\beta) - \log\{\beta + (k+1) + \rho(k+1)\}} - 1 \\
&= \lim_{k \rightarrow \infty} \frac{2 [\log(\beta) - \log(k+1+\beta)]}{\log(\beta) - \log(k+1+\beta) + \log\{\beta + \rho(k+1)\} - \log\{\beta + (k+1) + \rho(k+1)\}} - 1 \\
&= \lim_{k \rightarrow \infty} \frac{2 [\log(\beta) - \log(k+1+\beta)]}{[\log(\beta) - \log(k+1+\beta)] \left[ 1 + \frac{\log\{\beta+\rho(k+1)\}}{\log(\beta)-\log(k+1+\beta)} - \frac{\log\{\beta+(k+1)+\rho(k+1)\}}{\log(\beta)-\log(k+1+\beta)} \right]} - 1 \\
&= \lim_{k \rightarrow \infty} \left[ \frac{2}{1 + \frac{\log(\beta+\rho(k+1))}{\log(\beta)-\log(k+1+\beta)} - \frac{\log(\beta+(k+1)+\rho(k+1))}{\log(\beta)-\log(k+1+\beta)}} \right] - 1 \\
&= \frac{2}{\lim_{k \rightarrow \infty} \left[ 1 + \frac{\log(k)+\log(\frac{\beta}{k}+\rho+\frac{\rho}{k})}{\log(\beta)-\log(k)-\log(1+\frac{1}{k}+\frac{\beta}{k})} - \frac{\log(k)+\log(\frac{\beta}{k}+1+\frac{1}{k}+\rho+\frac{\rho}{k})}{\log(\beta)-\log(k)+\log(1+\frac{1}{k}+\frac{\beta}{k})} \right]} - 1 \\
&= \lim_{k \rightarrow \infty} \left[ 1 + \frac{1+\frac{1}{\log(k)} \left[ \log(\frac{\beta}{k}+\rho+\frac{\rho}{k}) \right]}{-1+\frac{1}{\log(k)} \left[ \log(\beta)-\log(1+\frac{1}{k}+\frac{\beta}{k}) \right]} - \frac{1+\frac{1}{\log(k)} \left[ \log(\frac{\beta}{k}+1+\frac{1}{k}+\rho+\frac{\rho}{k}) \right]}{-1+\frac{1}{\log(k)} \left[ \log(\beta)+\log(1+\frac{1}{k}+\frac{\beta}{k}) \right]} \right] - 1 \\
&= 2 - 1 = 1
\end{aligned} \tag{A.18}$$

The positive limit of  $\chi = [\rho/1 + \rho]^\alpha$  and  $\bar{\chi} = 1$  indicate that the discrete random variables ( $Y_1$  and  $Y_2$ ) are said to be asymptotically dependent when one work with GM.

## A.2.2 Copula-based Gaver model

By using the definition given in A.16, we have

$$\chi = \lim_{k \rightarrow \infty} \frac{\left( \begin{array}{l} L^{(2)}(s_1, s_2) | (s_1 = \frac{k+1}{\beta}, s_2 = \frac{k+1}{\beta}) + \phi \left[ L^{(2)}(s_1, s_2) | (s_1 = \frac{k+1}{\beta}, s_2 = \frac{k+1}{\beta}) \right. \\ \left. - L^{(2)}(s_1, s_2) | (s_1 = \frac{2(k+1)}{\beta}, s_2 = \frac{k+1}{\beta}) - L^{(2)}(s_1, s_2) | (s_1 = \frac{k+1}{\beta}, s_2 = \frac{2(k+1)}{\beta}) \right. \\ \left. + L^{(2)}(s_1, s_2) | (s_1 = \frac{2(k+1)}{\beta}, s_2 = \frac{2(k+1)}{\beta}) \right] \end{array} \right)}{L^{(1)}(s) | (s = \frac{k+1}{\beta})} \quad (\text{A.19})$$

where  $L^{(2)}(\cdot, \cdot)$  is the bivariate LT of GM defined in (1.29). Thus, we have

$$\chi = \lim_{k \rightarrow \infty} \frac{\left( \begin{array}{l} \left[ \frac{(\beta+\rho)\beta}{(\beta+k+1)(\beta+k+1+\rho(k+1))} \right]^\alpha + \phi \left[ \left[ \frac{(\beta+\rho)\beta}{(\beta+k+1)(\beta+k+1+\rho(k+1))} \right]^\alpha \right. \\ \left. - \left[ \frac{(\beta+\rho)\beta}{(\beta+k+1)(\beta+2(k+1)+\rho(k+1))} \right]^\alpha - \left[ \frac{(\beta+\rho)\beta}{(\beta+2(k+1)(\beta+k+1+2\rho(k+1))} \right]^\alpha \right. \\ \left. + \left[ \frac{(\beta+\rho)\beta}{(\beta+2(k+1)(\beta+2(k+1)+2\rho(k+1))} \right]^\alpha \right] \end{array} \right)}{\left[ \frac{\beta}{\beta+k+1} \right]^\alpha}$$

$$\chi = \lim_{k \rightarrow \infty} (A) + \phi \left[ \lim_{k \rightarrow \infty} (B) - \lim_{k \rightarrow \infty} (C) - \lim_{k \rightarrow \infty} (D) + \lim_{k \rightarrow \infty} (E) \right] \quad (\text{A.20})$$

Similar to GM, we would like to prove all limits of A.20 as

$$\lim_{k \rightarrow \infty} (A) = \lim_{y \rightarrow \infty} (B) = \lim_{y \rightarrow \infty} \frac{\left[ \frac{\{\beta+\rho(y+1)\}\beta}{(\beta+y+1)\{\beta+y+1+\rho(y+1)\}} \right]^\alpha}{\left[ \frac{\beta}{\beta+y+1} \right]^\alpha} = \left[ \frac{\rho}{1+\rho} \right]^\alpha$$

$$\lim_{y \rightarrow \infty} (C) = \lim_{k \rightarrow \infty} \frac{\left[ \frac{\{\beta+\rho(k+1)\}\beta}{(\beta+k+1)\{\beta+2(k+1)+\rho(k+1)\}} \right]^\alpha}{\left[ \frac{\beta}{\beta+k+1} \right]^\alpha} = \left[ \frac{\rho}{2+\rho} \right]^\alpha$$

$$\lim_{k \rightarrow \infty} (D) = \lim_{k \rightarrow \infty} \frac{\left[ \frac{\{\beta+2\rho(k+1)\}\beta}{\{\beta+2(k+1)\}\{\beta+k+1+2\rho(k+1)\}} \right]^\alpha}{\left[ \frac{\beta}{\beta+k+1} \right]^\alpha} = \left[ \frac{\rho}{2+\rho} \right]^\alpha$$

$$\lim_{k \rightarrow \infty} (E) = \lim_{k \rightarrow \infty} \frac{\left[ \frac{\{\beta+2\rho(k+1)\}\beta}{\{\beta+2(k+1)\}\{\beta+2(k+1)+2\rho(k+1)\}} \right]^\alpha}{\left[ \frac{\beta}{\beta+k+1} \right]^\alpha} = \left[ \frac{\rho}{2+2\rho} \right]^\alpha$$

After putting all limits in A.20, we get

$$\chi = \left(\frac{\rho}{1+\rho}\right)^\alpha + \phi \left[ \left(\frac{\rho}{1+\rho}\right)^\alpha - \left(\frac{\rho}{2+\rho}\right)^\alpha - \left(\frac{\rho}{(1+2\rho)}\right)^\alpha + \left(\frac{\rho}{(2+2\rho)}\right)^\alpha \right] \quad (\text{A.21})$$

In addition, the  $\bar{\chi}$  corresponding to CGM is derived by using the definition given in A.18 as

$$\bar{\chi} = \lim_{k \rightarrow \infty} \frac{2 \log \left[ \frac{\beta}{\beta+k+1} \right]^\alpha}{\log \left[ \begin{array}{l} \left[ \frac{\{\beta+\rho(k+1)\}\beta}{(\beta+k+1)\{\beta+k+1+\rho(k+1)\}} \right]^\alpha + \phi \left[ \frac{\{\beta+\rho(k+1)\}\beta}{(\beta+k+1)\{\beta+k+1+\rho(k+1)\}} \right]^\alpha \\ - \left[ \frac{\{\beta+\rho(k+1)\}\beta}{(\beta+k+1)\{\beta+2(k+1)+\rho(k+1)\}} \right]^\alpha - \left[ \frac{\{\beta+2\rho(k+1)\}\beta}{\{\beta+2(k+1)\}\{\beta+k+1+2\rho(k+1)\}} \right]^\alpha \\ + \left[ \frac{\{\beta+2\rho(k+1)\}\beta}{\{\beta+2(k+1)\}\{\beta+2(k+1)+2\rho(k+1)\}} \right]^\alpha \end{array} \right]} - 1 \quad (\text{A.22})$$

$$\bar{\chi} = \lim_{k \rightarrow \infty} \frac{2 \log \left[ \frac{\beta}{\beta+k+1} \right]^\alpha}{\log \left[ \begin{array}{l} \left[ \frac{\{\beta+2\rho(k+1)\}\beta}{\{\beta+2(k+1)\}\{\beta+2(k+1)+2\rho(k+1)\}} \right]^\alpha \left[ 1 + \phi \left( 1 - \frac{\left[ \frac{\{\beta+\rho(k+1)\}\beta}{(\beta+k+1)\{\beta+2(k+1)+\rho(k+1)\}} \right]^\alpha - \left[ \frac{\{\beta+2\rho(k+1)\}\beta}{\{\beta+2(k+1)\}\{\beta+k+1+2\rho(k+1)\}} \right]^\alpha}{\left[ \frac{\{\beta+\rho(k+1)\}\beta}{(\beta+k+1)\{\beta+k+1+\rho(k+1)\}} \right]^\alpha} \right. \right. \\ \left. \left. + \frac{\left[ \frac{\{\beta+2\rho(k+1)\}\beta}{\{\beta+2(k+1)\}\{\beta+2(k+1)+2\rho(k+1)\}} \right]^\alpha}{\left[ \frac{\{\beta+\rho(k+1)\}\beta}{(\beta+k+1)\{\beta+k+1+\rho(k+1)\}} \right]^\alpha} \right) \right] \end{array} \right]} - 1$$

$$\bar{\chi}_c = \lim_{k \rightarrow \infty} \frac{2 \log \left[ \frac{\beta}{\beta+k+1} \right]^\alpha}{\log \left[ \left[ \frac{\{\beta+2\rho(k+1)\}\beta}{\{\beta+2(k+1)\}\{\beta+2(k+1)+2\rho(k+1)\}} \right]^\alpha \left[ 1 + \phi \left( 1 - A - B + C \right) \right] \right]} - 1$$

where

$$A = \frac{\left[ \frac{\{\beta+\rho(k+1)\}\beta}{(\beta+k+1)\{\beta+2(k+1)+\rho(k+1)\}} \right]^\alpha}{\left[ \frac{\{\beta+\rho(k+1)\}\beta}{(\beta+k+1)\{\beta+k+1+\rho(k+1)\}} \right]^\alpha} = \frac{\left[ \frac{\{\beta+k+1+\rho(k+1)\}\{\beta+2\rho(k+1)\}}{\{\beta+2(k+1)+\rho(k+1)\}\{\beta+\rho(k+1)\}} \right]^\alpha}{1}$$

$$\begin{aligned}
B &= \frac{\left[ \frac{\{\beta+2\rho(k+1)\}\beta}{\{\beta+2(k+1)\}\{\beta+k+1+2\rho(k+1)\}} \right]^\alpha}{\left[ \frac{\{\beta+\rho(k+1)\}\beta}{(\beta+k+1)\{\beta+k+1+\rho(k+1)\}} \right]^\alpha} \\
&= \left[ \frac{\{\beta+(k+1)\}\{\beta+k+1+\rho(k+1)\}\{\beta+2\rho(k+1)\}}{\{\beta+2(k+1)\}\{\beta+k+1+2\rho(k+1)\}\{\beta+\rho(k+1)\}} \right]^\alpha
\end{aligned}$$

$$\begin{aligned}
C &= \frac{\left[ \frac{\{\beta+2\rho(k+1)\}\beta}{\{\beta+2(k+1)\}\{\beta+2(k+1)+2\rho(k+1)\}} \right]^\alpha}{\left[ \frac{\{\beta+\rho(k+1)\}\beta}{(\beta+k+1)\{\beta+k+1+\rho(k+1)\}} \right]^\alpha} \\
&= \left[ \frac{\{\beta+(k+1)\}\{\beta+k+1+\rho(k+1)\}\{\beta+2\rho(k+1)\}}{\{\beta+2(k+1)\}\{\beta+2(k+1)+2\rho(k+1)\}\{\beta+\rho(k+1)\}} \right]^\alpha
\end{aligned}$$

$$\bar{\chi} = \lim_{k \rightarrow \infty} \frac{2\alpha [\log(\beta) - \log(\beta + k + 1)]}{\alpha \log \left[ \frac{\beta\{\beta+\rho(k+1)\}}{\{\beta+k+1\}\{\beta+k+1+\rho(k+1)\}} \right] + \log \left\{ 1 + \phi(1 - A - B + C) \right\}} - 1 \quad (\text{A.23})$$

$$= \lim_{k \rightarrow \infty} \frac{2\alpha [\log(\beta) - \log(\beta + k + 1)]}{\alpha [\log(\beta) + \log\{\beta + \rho(k+1)\} - \log(\beta + k + 1) - \log(\beta + k + 1 + \rho x)] + D} - 1 \quad (\text{A.24})$$

where,  $D = \log \left\{ 1 + \phi(1 - A - B + C) \right\}$

$$\begin{aligned}
\bar{\chi} &= \lim_{k \rightarrow \infty} \frac{2\alpha [\log(\beta) - \log(\beta + k + 1)]}{\alpha [\log(\beta) - \log(\beta + k + 1)] + \alpha [\log\{\beta + \rho(k+1)\} - \log\{\beta + k + 1 + \rho(k+1)\}] + D} - 1 \\
&= \lim_{k \rightarrow \infty} \frac{2\alpha [\log(\beta) - \log(\beta + k + 1)]}{\alpha [\log(\beta) - \log(\beta + k + 1)] \left[ 1 + \frac{\alpha [\log\{\beta + \rho(k+1)\} - \log\{\beta + k + 1 + \rho(k+1)\}]}{\alpha [\log(\beta) - \log(\beta + k + 1)]} \right] + \frac{1}{\alpha [\log(\beta) - \log(\beta + k + 1)]} D} - 1 \\
&= \lim_{k \rightarrow \infty} \frac{2}{\left[ 1 + \frac{[\log\{\beta + \rho(k+1)\} - \log\{\beta + k + 1 + \rho(k+1)\}]}{[\log(\beta) - \log(\beta + k + 1)]} \right] + \frac{1}{\alpha [\log(\beta) - \log(\beta + k + 1)]} D} - 1 \\
&= 1
\end{aligned}$$

The limiting measure  $\chi > 0$  and  $|\bar{\chi}| = 1$  show that the CGLM also allow the asymptotic dependence between  $Y_1$  and  $Y_2$  over tail rather it may weak.

### A.2.3 Kibble model

To derive the joint tail dependence measures  $\chi$  and  $\bar{\chi}$  for KM, we use the definitions given in (A.16) and (A.18) with bivariate LT associated with KM. Thus, the limiting

measure  $\chi$  is

$$\begin{aligned}
\chi &= \lim_{k \rightarrow \infty} \frac{L^{(2)}(s_1, s_2)|_{(s_1=\frac{k+1}{\beta}, s_2=\frac{k+1}{\beta})}}{L^{(1)}(s)|_{(s=\frac{k+1}{\beta})}} \\
&= \lim_{k \rightarrow \infty} \frac{\left[ \frac{\beta^2}{(k+1+\beta)(k+1+\beta) - \rho\{(k+1)(k+1)\}} \right]^\alpha}{\left[ \frac{\beta}{\beta+k+1} \right]^\alpha} \\
&= \lim_{k \rightarrow \infty} \left[ \frac{(k+1+\beta)\beta}{(k+1+\beta)(k+1+\beta) - \rho(k+1)^2} \right]^\alpha \\
&= \lim_{k \rightarrow \infty} \left[ \frac{\beta}{(k+1+\beta) \left[ 1 - \frac{\rho(k+1)^2}{(k+1+\beta)^2} \right]} \right]^\alpha \\
&= 0
\end{aligned} \tag{A.25}$$

and, the  $\bar{\chi}$  is

$$\begin{aligned}
\bar{\chi} &= \lim_{k \rightarrow \infty} \frac{2 \log L^{(1)}(s)|_{(s=\frac{k+1}{\beta})}}{\log L^{(2)}(s_1, s_2)|_{(s_1=\frac{k+1}{\beta}, s_2=\frac{k+1}{\beta})}} - 1 \\
&= \lim_{k \rightarrow \infty} \frac{2 \log \left[ \frac{\beta}{(k+1+\beta)} \right]^\alpha}{\log \left[ \frac{\beta^2}{(k+1+\beta)^2 - \rho(k+1)^2} \right]^\alpha} - 1 \\
&= \lim_{k \rightarrow \infty} \left[ \frac{2 \{ \log(\beta) - \log(k+1+\beta) \}}{2 \log(\beta) - \log \{ (k+1+\beta)^2 - \rho(k+1)^2 \}} \right] - 1 \\
&= \lim_{k \rightarrow \infty} \left[ \frac{2 \log(\beta) - 2 \log(k+1+\beta)}{2 \log(\beta) - \log \left[ (k+1+\beta)^2 \left\{ 1 - \frac{\rho(k+1)^2}{(k+1+\beta)^2} \right\} \right]} \right] - 1 \\
&= \lim_{k \rightarrow \infty} \left[ \frac{2 \log(\beta) - 2 \log(k+1+\beta)}{2 \log(\beta) - 2 \log(k+1+\beta) - \log \left\{ 1 - \frac{\rho(k+1)^2}{(k+1+\beta)^2} \right\}} \right] - 1 \\
&= \lim_{k \rightarrow \infty} \left[ \frac{2 \log(\beta) - 2 \log(k+1+\beta)}{2 \log(\beta) - 2 \log(k+1+\beta) \left[ 1 - \frac{\log \left\{ 1 - \frac{\rho(k+1)^2}{(k+1+\beta)^2} \right\}}{2 \log(\beta) - 2 \log(k+1+\beta)} \right]} \right] - 1 \\
&= \lim_{k \rightarrow \infty} \left[ \frac{1}{1 - \frac{\log \left\{ 1 - \frac{\rho(k+1)^2}{(k+1+\beta)^2} \right\}}{2 \log(\beta) - 2 \log(k+1+\beta)}} \right] - 1 \\
&= 0
\end{aligned} \tag{A.26}$$

The limiting measures  $\chi = \bar{\chi} = 0$  show no local clustering at an extreme level of Y1 and Y2 variables.

### A.2.4 Copula based Kibble model

To prove the asymptotic properties of  $\chi$  and  $\bar{\chi}$  associated with CWM, we use the usual definition given in (A.19) and (A.22) with bivariate LT linked with WM. Thus, the limiting measure  $\chi$  is

$$\chi = \lim_{k \rightarrow \infty} \frac{\left( \begin{aligned} & \left[ \frac{\beta^2}{(k+1+\beta)^2 - \rho(k+1)^2} \right]^\alpha + \phi \left[ \frac{\beta^2}{(k+1+\beta)^2 - \rho(k+1)^2} \right]^\alpha \\ & - \left[ \frac{\beta^2}{(2k+2+\beta)(k+1+\beta) - 2\rho(k+1)^2} \right]^\alpha - \left[ \frac{\beta^2}{(k+1+\beta)(2k+2+\beta) - 2\rho(k+1)^2} \right]^\alpha \\ & \qquad \qquad \qquad + \left[ \frac{\beta^2}{(2k+2+\beta)^2 - 4\rho(k+1)^2} \right]^\alpha \end{aligned} \right)}{\left[ \frac{\beta}{\beta+k+1} \right]^\alpha}$$

$$\chi = \lim_{k \rightarrow \infty} (A) + \phi \left[ \lim_{k \rightarrow \infty} (B) - \lim_{k \rightarrow \infty} (C) - \lim_{k \rightarrow \infty} (D) + \lim_{k \rightarrow \infty} (E) \right] \quad (\text{A.27})$$

Similar to WM, we would like to prove all limits of (A.27) as

$$\lim_{k \rightarrow \infty} (A) = \lim_{k \rightarrow \infty} (B) = \lim_{k \rightarrow \infty} \frac{\left[ \frac{\beta^2}{(k+1+\beta)^2 - \rho(k+1)^2} \right]^\alpha}{\left[ \frac{\beta}{\beta+k+1} \right]^\alpha} = 0$$

$$\lim_{k \rightarrow \infty} (C) = \lim_{k \rightarrow \infty} (D) = \lim_{k \rightarrow \infty} \frac{\left[ \frac{\beta^2}{(2k+2+\beta)(k+1+\beta) - 2\rho(k+1)^2} \right]^\alpha}{\left[ \frac{\beta}{\beta+k+1} \right]^\alpha} = 0$$

$$\lim_{k \rightarrow \infty} (E) = \lim_{k \rightarrow \infty} \frac{\left[ \frac{\beta^2}{(2k+2+\beta)^2 - 4\rho(k+1)^2} \right]^\alpha}{\left[ \frac{\beta}{\beta+y+1} \right]^\alpha} = 0$$

By putting all limits in (A.27), we get

$$\chi = 0 \quad (\text{A.28})$$

and  $\bar{\chi}$  is

$$\bar{\chi} = \lim_{k \rightarrow \infty} \frac{2 \log \left[ \frac{\beta}{\beta+k+1} \right]^\alpha}{\log \left[ \left[ \frac{\beta^2}{(k+1+\beta)^2 - \rho(k+1)^2} \right]^\alpha + \phi \left[ \left[ \frac{\beta^2}{(k+1+\beta)^2 - \rho(k+1)^2} \right]^\alpha - \left[ \frac{\beta^2}{(2k+2+\beta)(k+1+\beta) - 2\rho(k+1)^2} \right]^\alpha + \left[ \frac{\beta^2}{(2k+2+\beta)^2 - 4\rho(k+1)^2} \right]^\alpha \right]} - 1 \quad (\text{A.29})$$

$$\bar{\chi} = \lim_{k \rightarrow \infty} \frac{2 \log \left[ \frac{\beta}{\beta+k+1} \right]^\alpha}{\log \left[ \left[ \frac{\beta^2}{(k+1+\beta)^2 - \rho(k+1)^2} \right]^\alpha \left[ 1 + \phi \left( 1 - \frac{\left[ \frac{\beta^2}{(2k+2+\beta)(k+1+\beta) - 2\rho(k+1)^2} \right]^\alpha}{\left[ \frac{\beta^2}{(k+1+\beta)^2 - \rho(k+1)^2} \right]^\alpha} - \frac{\left[ \frac{\beta^2}{(2k+2+\beta)(k+1+\beta) - 2\rho(k+1)^2} \right]^\alpha}{\left[ \frac{\beta^2}{(k+1+\beta)^2 - \rho(k+1)^2} \right]^\alpha} + \frac{\left[ \frac{\beta^2}{(2k+2+\beta)(k+1+\beta) - 2\rho(k+1)^2} \right]^\alpha}{\left[ \frac{\beta^2}{(k+1+\beta)^2 - \rho(k+1)^2} \right]^\alpha} \right) \right]} - 1$$

$$\bar{\chi} = \lim_{k \rightarrow \infty} \frac{2\alpha \log [\log(\beta) - \log(\beta+k+1)]}{\log \left[ \frac{\beta^2}{(k+1+\beta)^2 - \rho(k+1)^2} \right]^\alpha + \log \left[ 1 + \phi \left( 1 - \frac{\left[ \frac{\beta^2}{(2k+2+\beta)(k+1+\beta) - 2\rho(k+1)^2} \right]^\alpha}{\left[ \frac{\beta^2}{(k+1+\beta)^2 - \rho(k+1)^2} \right]^\alpha} - \frac{\left[ \frac{\beta^2}{(2k+2+\beta)(k+1+\beta) - 2\rho(k+1)^2} \right]^\alpha}{\left[ \frac{\beta^2}{(k+1+\beta)^2 - \rho(k+1)^2} \right]^\alpha} + \frac{\left[ \frac{\beta^2}{(2k+2+\beta)(k+1+\beta) - 2\rho(k+1)^2} \right]^\alpha}{\left[ \frac{\beta^2}{(k+1+\beta)^2 - \rho(k+1)^2} \right]^\alpha} \right) \right]} - 1$$

$$\begin{aligned} \bar{\chi} &= \lim_{k \rightarrow \infty} \frac{2\alpha [\log(\beta) - \log(\beta+k+1)]}{\log \left[ \frac{\beta^2}{(k+1+\beta)^2 - \rho(k+1)^2} \right]^\alpha + D} - 1 \\ &= \lim_{k \rightarrow \infty} \frac{2\alpha \{\log(\beta) - \log(\beta+k+1)\}}{2\alpha \{\log(\beta) - \log(\beta+k+1)\} \left\{ 1 - \frac{\log \left( 1 - \frac{\rho(k+1)^2}{(\beta+k+1)^2} \right)}{2 \log(\beta) - 2 \log(\beta+k+1)} \right\} + D} - 1 \\ &= \lim_{k \rightarrow \infty} \frac{2\alpha \{\log(\beta) - \log(\beta+k+1)\}}{2\alpha \{\log(\beta) - \log(\beta+k+1)\} \left\{ \left\{ 1 - \frac{\log \left( 1 - \frac{\rho(k+1)^2}{(\beta+k+1)^2} \right)}{2 \log(\beta) - 2 \log(\beta+k+1)} \right\} + \frac{1}{2\alpha \{\log(\beta) - \log(\beta+k+1)\}} D \right\}} - 1 \end{aligned}$$

$$\begin{aligned}
&= \lim_{k \rightarrow \infty} \frac{1}{\left[ \left\{ 1 - \frac{\log\left(1 - \frac{\rho(k+1)^2}{(\beta+k+1)^2}\right)}{2\log(\beta) - 2\log(\beta+k+1)} \right\} + \frac{1}{2\alpha\{\log(\beta) - \log(\beta+k+1)\}} D \right]} - 1 \\
&= \frac{1}{\left[ \left\{ 1 - 0 \right\} + 0 \right]} - 1 \\
&= 0
\end{aligned} \tag{A.30}$$

Again, the limiting value of  $\chi$  and  $\bar{\chi}$  tend to zero, indicating no local clustering at the extreme level of  $Y_1$  and  $Y_2$  variables when generated from CWM.

### A.2.5 Thinned Gamma model

By using the usual definitions of  $\chi$  and  $\bar{\chi}$ , we derive the asymptotic properties of tail dependence measure  $\chi$  and  $\bar{\chi}$  correspond to TGM. Moreover, the limiting measure  $\chi$  is

$$\begin{aligned}
\chi &= \lim_{k \rightarrow \infty} \frac{L^{(2)}(s_1, s_2)|_{(s_1=\frac{k+1}{\beta}, s_2=\frac{k+1}{\beta})}}{L^{(1)}(s)|_{(s=\frac{k+1}{\beta})}} \\
&= \lim_{k \rightarrow \infty} \frac{\left[ \frac{\beta^{\alpha(2-\rho)}}{(\beta+k+1)^{2\alpha(1-\rho)}(\beta+2k+2)^{\alpha\rho}} \right]}{\left[ \frac{\beta}{\beta+k+1} \right]^\alpha} \\
&= \lim_{k \rightarrow \infty} \left[ \frac{\beta^{\alpha(2-\rho)}(\beta+k+1)^\alpha}{\beta^\alpha(\beta+k+1)^{2\alpha(1-\rho)}\{\beta+2(k+1)\}^{\alpha\rho}} \right] \\
&= \lim_{k \rightarrow \infty} \left[ \frac{\beta^{\alpha(1-\rho)}}{(\beta+k+1)^{\alpha(1-2\rho)}\{\beta+2(k+1)\}^{\alpha\rho}} \right] \\
&= 0
\end{aligned}$$

$$\begin{aligned}
\bar{\chi} &= \lim_{k \rightarrow \infty} \frac{2\log L^{(1)}(s)|_{(s=\frac{k+1}{\beta})}}{\log L^{(2)}(s_1, s_2)|_{(s_1=\frac{k+1}{\beta}, s_2=\frac{k+1}{\beta})}} - 1 \\
&= \lim_{k \rightarrow \infty} \frac{2\log \left[ \frac{\beta}{\beta+k+1} \right]^\alpha}{\log \left[ \frac{\beta^{\alpha(2-\rho)}}{(\beta+k+1)^{2\alpha(1-\rho)}(\beta+2k+2)^{\alpha\rho}} \right]} - 1 \\
&= \lim_{k \rightarrow \infty} \left[ \frac{2\alpha\{\log(\beta) - \log(k+1+\beta)\}}{2\alpha[\log(\beta) - \log(\beta+k+1)] - \alpha\rho[\log(\beta) - 2\log(\beta+k+1) + \log(\beta+2k+2)]} \right] - 1 \\
&= \lim_{k \rightarrow \infty} \left[ \frac{2\alpha[\log(\beta) - \log(k+1+\beta)]}{2\alpha[\log(\beta) - \log(k+1+\beta)] \left[ 1 - \frac{\rho[\log(\beta) - 2\log(\beta+k+1) + \log(\beta+2k+2)]}{2[\log(\beta) - \log(k+1+\beta)]} \right]} \right] - 1 \\
&= \lim_{k \rightarrow \infty} \left[ \frac{1}{\left[ 1 - \frac{\rho[\log(\beta) - 2\log(\beta+k+1) + \log(\beta+2k+2)]}{2[\log(\beta) - \log(k+1+\beta)]} \right]} \right] - 1
\end{aligned}$$



$$\begin{aligned}
&= \frac{1}{\left[1 - \frac{\rho[0-2+0+1+0]}{2(0-1-0)}\right]} - 1 \\
&= \frac{\rho}{2 - \rho}
\end{aligned} \tag{A.31}$$

We conclude on the basis of  $\chi$  and  $\bar{\chi}$  that the  $Y_1$  and  $Y_2$  variables are asymptotically independent when generated from TGM. A larger value of  $\rho$  in  $\bar{\chi}$  may lead to stronger dependence.

### A.2.6 Copula-based Thinned Gamma model

To prove the asymptotic properties of  $\chi$  and  $\bar{\chi}$  associated with CTGM, we use the usual definition given in (A.19) and (A.22) with bivariate LT linked with CTGM. Thus, the limiting measure  $\chi$  is

$$\begin{aligned}
\chi &= \lim_{k \rightarrow \infty} \frac{\left( \left[ \frac{\beta^{\alpha(2-\rho)}}{(\beta+k+1)^{2\alpha(1-\rho)}(\beta+2k+2)^{\alpha\rho}} \right] + \phi \left[ \left[ \frac{\beta^{\alpha(2-\rho)}}{(\beta+k+1)^{2\alpha(1-\rho)}(\beta+2k+2)^{\alpha\rho}} \right] \right. \right. \\
&\quad \left. \left. - 2 \left[ \frac{\beta^{2\alpha-\alpha\rho}}{(\beta+2k+2)^{\alpha-\alpha\rho}(\beta+k+1)^{\alpha-\alpha\rho}(\beta+3k+3)^{\alpha\rho}} \right] \right. \right. \\
&\quad \left. \left. + \left[ \frac{\beta^{2\alpha-\alpha\rho}}{(\beta+2k+2)^{2\alpha-2\alpha\rho}(\beta+4k+4)^{\alpha\rho}} \right] \right) \right)}{\left[ \frac{\beta}{\beta+k+1} \right]^\alpha} \\
\chi &= \lim_{k \rightarrow \infty} (A) + \phi \left[ \lim_{k \rightarrow \infty} (B) - 2 \lim_{k \rightarrow \infty} (C) + \lim_{k \rightarrow \infty} (D) \right]
\end{aligned} \tag{A.32}$$

Similar to TGM, we would like to prove all limit of (A.32) separately

$$\begin{aligned}
\lim_{k \rightarrow \infty} (A) &= \lim_{k \rightarrow \infty} (B) = \lim_{k \rightarrow \infty} \frac{\left[ \frac{\beta^{2\alpha-\alpha\rho}}{(\beta+k+1)^{2\alpha-2\alpha\rho}(\beta+2k+2)^{\alpha\rho}} \right]}{\left[ \frac{\beta}{\beta+k+1} \right]^\alpha} = 0 \\
\lim_{k \rightarrow \infty} (C) &= \lim_{k \rightarrow \infty} \frac{\left[ \frac{\beta^{2\alpha-\alpha\rho}}{(\beta+2k+2)^{\alpha-\alpha\rho}(\beta+k+1)^{\alpha-\alpha\rho}(\beta+3k+3)^{\alpha\rho}} \right]}{\left[ \frac{\beta}{\beta+k+1} \right]^\alpha} = 0 \\
\lim_{k \rightarrow \infty} (D) &= \lim_{k \rightarrow \infty} \frac{\left[ \frac{\beta^{2\alpha-\alpha\rho}}{(\beta+2k+2)^{2\alpha-2\alpha\rho}(\beta+4k+4)^{\alpha\rho}} \right]}{\left[ \frac{\beta}{\beta+k+1} \right]^\alpha} = 0
\end{aligned}$$

By incorporating all the limits in (A.32), we get

$$\chi = 0 \tag{A.33}$$

and  $\bar{\chi}$  is

$$\begin{aligned}
\bar{\chi} &= \lim_{k \rightarrow \infty} \frac{2 \log \left[ \frac{\beta}{\beta+k+1} \right]^\alpha}{\log \left( \left[ \frac{\beta^\alpha(2-\rho)}{(\beta+k+1)^{2\alpha(1-\rho)}(\beta+2k+2)^{\alpha\rho}} \right] + \phi \left[ \frac{\beta^\alpha(2-\rho)}{(\beta+k+1)^{2\alpha(1-\rho)}(\beta+2k+2)^{\alpha\rho}} \right] \right.} - 1 \\
&\quad \left. - 2 \left[ \frac{\beta^{2\alpha-\alpha\rho}}{(\beta+2k+2)^{\alpha-\alpha\rho}(\beta+k+1)^{\alpha-\alpha\rho}(\beta+3k+3)^{\alpha\rho}} \right] \right. \\
&\quad \left. + \left[ \frac{\beta^{2\alpha-\alpha\rho}}{(\beta+2k+2)^{2\alpha-2\alpha\rho}(\beta+4k+4)^{\alpha\rho}} \right] \right) \\
&= \lim_{k \rightarrow \infty} \frac{2 \log \left[ \frac{\beta}{\beta+k+1} \right]^\alpha}{\log \left( \left[ \frac{\beta^\alpha(2-\rho)}{(\beta+k+1)^{2\alpha(1-\rho)}(\beta+2k+2)^{\alpha\rho}} \right] \left[ 1 + \phi \left( 1 \right. \right. \right.} - 1 \\
&\quad \left. \left. - \frac{2 \left[ \frac{\beta^{2\alpha-\alpha\rho}}{(\beta+2k+2)^{\alpha-\alpha\rho}(\beta+k+1)^{\alpha-\alpha\rho}(\beta+3k+3)^{\alpha\rho}} \right]}{\left[ \frac{\beta^\alpha(2-\rho)}{(\beta+k+1)^{2\alpha(1-\rho)}(\beta+2k+2)^{\alpha\rho}} \right]} \right. \right. \\
&\quad \left. \left. + \left[ \frac{\beta^{2\alpha-\alpha\rho}}{(\beta+2k+2)^{2\alpha-2\alpha\rho}(\beta+4k+4)^{\alpha\rho}} \right] \right) \right]} \\
&= \lim_{k \rightarrow \infty} \frac{2 \log \left[ \frac{\beta}{\beta+k+1} \right]^\alpha}{\log \left[ \frac{\beta^\alpha(2-\rho)}{(\beta+k+1)^{2\alpha(1-\rho)}(\beta+2k+2)^{\alpha\rho}} \right] + \log \left[ 1 + \phi \left( 1 \right. \right.} - 1 \\
&\quad \left. \left. - \frac{2 \left[ \frac{\beta^{2\alpha-\alpha\rho}}{(\beta+2k+2)^{\alpha-\alpha\rho}(\beta+k+1)^{\alpha-\alpha\rho}(\beta+3k+3)^{\alpha\rho}} \right]}{\left[ \frac{\beta^\alpha(2-\rho)}{(\beta+k+1)^{2\alpha(1-\rho)}(\beta+2k+2)^{\alpha\rho}} \right]} \right. \right. \\
&\quad \left. \left. + \left[ \frac{\beta^{2\alpha-\alpha\rho}}{(\beta+2k+2)^{2\alpha-2\alpha\rho}(\beta+4k+4)^{\alpha\rho}} \right] \right) \right]} \\
&= \lim_{k \rightarrow \infty} \frac{2\alpha [\log(\beta) - \log(\beta+k+1)]}{2\alpha [\log(\beta) - \log(\beta+k+1)] + \alpha\rho [2 \log(\beta+k+1) - \log(\beta) - \log(\beta+2k+2)] + D} - 1 \\
&= \lim_{k \rightarrow \infty} \frac{1}{1 + \frac{\alpha\rho [2 \log(\beta+k+1) - \log(\beta) - \log(\beta+2k+2)]}{2\alpha [\log(\beta) - \log(\beta+k+1)]} + \frac{1}{2\alpha [\log(\beta) - \log(\beta+k+1)]} D} - 1 \\
&= \frac{1}{1 - \frac{\rho}{2} + 0} - 1 \\
&= \frac{\rho}{2 - \rho}
\end{aligned} \tag{A.34}$$

The limiting value of  $\chi = 0$  and  $\bar{\chi} = \rho/2 - \rho$  show that the  $Y_1$  and  $Y_2$  variables are asymptotically independent when generated from CTGM. A larger value of  $\rho$  in  $\bar{\chi}$  may lead to dependence.

### A.2.7 Complete dependence model

To evaluate the dependence between the exceedances of  $Y_1$  and  $Y_2$  variables, we opt usual definition for  $\chi$  and  $\bar{\chi}$  found in Coles *et al.* (1999). To obtain the limiting measure  $\chi$  for CDM, we solve (A.16) by incorporating LTs of Gamma distribution corresponding to CDM. That is, the limiting measure  $\chi$  is

$$\begin{aligned}
 \chi &= \lim_{k \rightarrow \infty} \frac{L^{(1)}(s)|_{(s=\frac{k+1+k+1}{\beta})}}{L^{(1)}(s)|_{(s=\frac{k+1}{\beta})}} \\
 &= \lim_{k \rightarrow \infty} \frac{\left[\frac{\beta}{2k+2+\beta}\right]^\alpha}{\left[\frac{\beta}{\beta+k+1}\right]^\alpha} \\
 &= \left[\frac{1}{2}\right]^\alpha
 \end{aligned} \tag{A.35}$$

and the  $\bar{\chi}$  is

$$\begin{aligned}
 \bar{\chi} &= \lim_{k \rightarrow \infty} \frac{2 \log L^{(1)}(s)|_{(s=\frac{k+1}{\beta})}}{\log L^{(1)}(s)|_{(s=\frac{k+1+k+1}{\beta})}} - 1 \\
 &= \lim_{k \rightarrow \infty} \frac{2 \log \left[\frac{\beta}{(k+1+\beta)}\right]^\alpha}{\log \left[\frac{\beta}{(2k+2+\beta)}\right]^\alpha} - 1 \\
 &= \lim_{k \rightarrow \infty} \frac{2 [\log(\beta) - \log(k+1+\beta)]}{\log(\beta) - \log(2k+2+\beta)} - 1 \\
 &= \lim_{k \rightarrow \infty} \frac{2 \left[ \frac{\log(\beta)}{\log(k)} - 1 - \frac{1}{\log(k)} \log\left(1 + \frac{1}{k} + \frac{\beta}{k}\right) \right]}{\left[ \frac{\log(\beta)}{\log(k)} - 1 - \frac{1}{\log(k)} \log\left(2 + \frac{2}{k} + \frac{\beta}{k}\right) \right]} - 1 \\
 &= 2 - 1 = 1
 \end{aligned} \tag{A.36}$$

The positive limit of  $\chi = [\rho/1 + \rho]^\alpha$  and  $\bar{\chi} = 1$  indicate that the discrete random variables ( $Y_1$  and  $Y_2$ ) are said to be asymptotically dependent when one work with GLM.

### A.2.8 Copula-based complete dependence model

Again, we use the usual definition of the limiting measure  $\chi$  and  $\bar{\chi}$  to prove the asymptotic properties of tail dependence for CCDM. Thus, the copula-based tail dependence measure for CCDM is

$$\begin{aligned} \chi &= \lim_{k \rightarrow \infty} \frac{\left( \left[ \frac{\beta}{2k+2+\beta} \right]^\alpha + \phi \left[ \left[ \frac{\beta}{2k+2+\beta} \right]^\alpha - \left[ \frac{\beta}{3k+3+\beta} \right]^\alpha - \left[ \frac{\beta}{3k+3+\beta} \right]^\alpha \right. \right. \\ &\quad \left. \left. + \left[ \frac{\beta}{4k+4+\beta} \right]^\alpha \right) \right)}{\left[ \frac{\beta}{\beta+k+1} \right]^\alpha} \\ &= A + \phi[B - 2C + D] \end{aligned} \tag{A.37}$$

where,

$$\lim_{k \rightarrow \infty} (A) = \lim_{k \rightarrow \infty} \frac{\left[ \frac{\beta}{2k+2+\beta} \right]^\alpha}{\left[ \frac{\beta}{\beta+k+1} \right]^\alpha} = \left[ \frac{1}{2} \right]^\alpha = \lim_{k \rightarrow \infty} (B)$$

$$\lim_{k \rightarrow \infty} (C) = \lim_{k \rightarrow \infty} \frac{\left[ \frac{\beta}{3k+3+\beta} \right]^\alpha}{\left[ \frac{\beta}{\beta+k+1} \right]^\alpha} = \left[ \frac{1}{3} \right]^\alpha$$

$$\lim_{k \rightarrow \infty} (C) = \lim_{k \rightarrow \infty} \frac{\left[ \frac{\beta}{4k+4+\beta} \right]^\alpha}{\left[ \frac{\beta}{\beta+k+1} \right]^\alpha} = \left[ \frac{1}{4} \right]^\alpha$$

$$\chi = \left( \frac{1}{2} \right)^\alpha + \phi \left[ \left( \frac{1}{2} \right)^\alpha - 2 \left( \frac{1}{3} \right)^\alpha + \left( \frac{1}{4} \right)^\alpha \right] \tag{A.38}$$

and, the  $\bar{\chi}$  is

$$\bar{\chi} = \lim_{k \rightarrow \infty} \frac{2 \log \left[ \frac{\beta}{\beta+k+1} \right]^\alpha}{\left( \log \left[ \frac{\beta}{2k+2+\beta} \right]^\alpha + \phi \left[ \left[ \frac{\beta}{2k+2+\beta} \right]^\alpha - \left[ \frac{\beta}{3k+3+\beta} \right]^\alpha - \left[ \frac{\beta}{3k+3+\beta} \right]^\alpha \right. \right. \\ \left. \left. + \left[ \frac{\beta}{4k+4+\beta} \right]^\alpha \right) \right)} - 1$$

$$\begin{aligned}
&= \lim_{k \rightarrow \infty} \frac{\alpha [2 \log(\beta) - 2 \log(\beta + k + 1)]}{\left( \log \left[ \frac{\beta}{2k+2+\beta} \right]^\alpha + \log \left[ 1 + \phi \left\{ 1 - \frac{\left[ \frac{\beta}{3k+3+\beta} \right]^\alpha}{\left[ \frac{\beta}{2k+2+\beta} \right]^\alpha} - \frac{\left[ \frac{\beta}{3k+3+\beta} \right]^\alpha}{\left[ \frac{\beta}{2k+2+\beta} \right]^\alpha} + \frac{\left[ \frac{\beta}{4k+4+\beta} \right]^\alpha}{\left[ \frac{\beta}{2k+2+\beta} \right]^\alpha} \right\} \right]} - 1 \\
&= \lim_{k \rightarrow \infty} \frac{\alpha [2 \log(\beta) - 2 \log(\beta + k + 1)]}{\left( \alpha \log(\beta) - \alpha \log(2k + 2 + \beta) + \log \left[ 1 + \phi \left\{ 1 - 2 \left[ \frac{2k+2+\beta}{3k+3+\beta} \right]^\alpha + \left[ \frac{2k+2+\beta}{4k+4+\beta} \right]^\alpha \right\} \right]} - 1 \\
&= \lim_{k \rightarrow \infty} \frac{[2 \log(\beta) - 2 \log(k) - 2 \log(\frac{\beta}{k} + 1 + \frac{1}{k})]}{\left( \log(\beta) - \log(k) - \log(2 + \frac{2}{k} + \frac{\beta}{k}) + \frac{1}{\alpha} \log \left[ 1 + \phi \left\{ 1 - 2 \left[ \frac{2k+2+\beta}{3k+3+\beta} \right]^\alpha + \left[ \frac{2k+2+\beta}{4k+4+\beta} \right]^\alpha \right\} \right]} - 1 \\
\bar{\chi} &= \lim_{k \rightarrow \infty} \frac{\log(k) \left[ \frac{2 \log(\beta)}{\log(k)} - 2 - \frac{2 \log(\frac{\beta}{k} + 1 + \frac{1}{k})}{\log(k)} \right]}{\log(k) \left[ \frac{\log(\beta)}{\log(k)} - 1 - \frac{\log(2 + \frac{2}{k} + \frac{\beta}{k})}{\log(k)} + \frac{1}{\log(k)} D \right]} - 1 = \frac{0 - 2 - 0}{0 - 1 - 0 + 0} - 1 = 1 \text{ (A.39)}
\end{aligned}$$

where  $D = \frac{1}{\alpha} \left[ 1 + \phi \left\{ 1 - 2 \frac{2k+2+\beta}{3k+3+\beta} + \frac{2k+2+\beta}{4k+4+\beta} \right\} \right]$ . The limiting measure  $\chi > 0$  and  $\bar{\chi} = 1$  indicate that the CCDM has asymptotic dependent behavior.



# Appendix

## Appendix B

### B.1 Maximum likelihood procedure of DEGPD

(i) The PMF of DEGPD corresponding to  $G(u; \psi) = u^\kappa$ ,  $\psi = \kappa > 0$  is written as

$$P(Y = k) = \left[ \left\{ 1 - \left( 1 + \frac{\xi(n+1)}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^\kappa - \left\{ 1 - \left( 1 + \frac{\xi n}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^\kappa \right] \quad (\text{B.1})$$

The log-likelihood function is defined as

$$l(\kappa, \sigma, \xi) = \sum_{j=1}^n \log \left[ \left\{ 1 - \left( 1 + \frac{\xi(k_j+1)}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^\kappa - \left\{ 1 - \left( 1 + \frac{\xi(k_j)}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^\kappa \right] \quad (\text{B.2})$$

(ii) First, we need to define the PMF based on

$$G(u; \psi) = 1 - D_\delta \{(1-u)^\delta\}, \quad \psi = \delta > 0$$

where  $D_\delta$  is CDF of beta distribution with parameters  $1/\delta$  and 2. By definition

$$H(x) = G \left\{ F_\xi \left( \frac{x}{\sigma} \right) \right\}$$

where  $F_\xi(\cdot)$  is the CDF of GPD. So,

$$\begin{aligned} H(x) &= 1 - D_\delta \left[ \left\{ 1 - F_\xi \left( \frac{x}{\sigma} \right) \right\}^\delta \right] \\ H(x) &= 1 - D_\delta \left[ \left\{ \bar{F}_\xi \left( \frac{x}{\sigma} \right) \right\}^\delta \right] \end{aligned} \quad (\text{B.3})$$

We solve,  $D_\delta \left[ \left\{ \bar{F}_\xi \left( \frac{x}{\sigma} \right) \right\}^\delta \right] = D_\delta [z]$ , where  $z = \left\{ \bar{F}_\xi \left( \frac{x}{\sigma} \right) \right\}^\delta$

The CDF of beta distribution with above specific parameters (i.e.,  $1/\delta$  and 2) is written as

$$D_\delta(z, 1/\delta, 2) = \frac{B(z, 1/\delta, 2)}{B(1/\delta, 2)} \quad (\text{B.4})$$

After solving the Beta functions, we get

$$H(x) = 1 + \frac{1}{\delta} \left[ \bar{F}_\xi \left( \frac{x}{\sigma} \right) \right]^{\delta+1} - \frac{\delta+1}{\delta} \left[ \bar{F}_\xi \left( \frac{x}{\sigma} \right) \right]$$

The PMF of DEGPD corresponding to  $G(u; \psi) = 1 - D_\delta\{(1-u)^\delta\}$  is written as

$$P(Y = k) = \frac{1}{\delta} \left\{ \left( 1 + \frac{\xi(k+1)}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^{\delta+1} - \frac{\delta+1}{\delta} \left\{ \left( 1 + \frac{\xi(k+1)}{\sigma} \right)^{-\frac{1}{\xi}} \right\} \\ - \frac{1}{\delta} \left\{ \left( 1 + \frac{\xi k}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^{\delta+1} + \frac{\delta+1}{\delta} \left\{ \left( 1 + \frac{\xi k}{\sigma} \right)^{-\frac{1}{\xi}} \right\} \quad (\text{B.5})$$

The log likelihood function is defined as

$$l(\delta, \sigma, \xi) = \sum_{j=1}^n \log \left[ \frac{1}{\delta} \left\{ \left( 1 + \frac{\xi(k_j+1)}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^{\delta+1} - \frac{\delta+1}{\delta} \left\{ \left( 1 + \frac{\xi(k_j+1)}{\sigma} \right)^{-\frac{1}{\xi}} \right\} \right. \\ \left. - \frac{1}{\delta} \left\{ \left( 1 + \frac{\xi k_j}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^{\delta+1} + \frac{\delta+1}{\delta} \left\{ \left( 1 + \frac{\xi k_j}{\sigma} \right)^{-\frac{1}{\xi}} \right\} \right] \quad (\text{B.6})$$

**(iii)** Similar to model **(ii)**, the CDF of EGPD corresponding to

$$G(u; \psi) = [1 - D_\delta\{(1-u)^\delta\}]^{\kappa/2}; \psi = (\delta, \kappa) > 0$$

is written as

$$H(x) = \left[ F_\xi \left( \frac{x}{\sigma} \right) + \frac{1}{\delta} \left[ \bar{F}_\xi \left( \frac{x}{\sigma} \right) \right]^{\delta+1} - \frac{1}{\delta} \left[ \bar{F}_\xi \left( \frac{x}{\sigma} \right) \right] \right]^{\frac{\kappa}{2}}$$

The PMF of DEGPD corresponding to  $G(u; \psi) = [1 - D_\delta\{(1-u)^\delta\}]^{\kappa/2}$  can be written as

$$P(Y = k) = \left[ 1 - \left( 1 + \frac{\xi(k+1)}{\sigma} \right)^{-\frac{1}{\xi}} + \frac{1}{\delta} \left\{ \left( 1 + \frac{\xi(k+1)}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^{\delta+1} \right]^{\frac{\kappa}{2}}$$



$$\begin{aligned}
& - \frac{1}{\delta} \left\{ \left( 1 + \frac{\xi(k+1)}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^{\frac{\kappa}{2}} \\
& - \left[ 1 - \left( 1 + \frac{\xi k}{\sigma} \right)^{-\frac{1}{\xi}} + \frac{1}{\delta} \left\{ \left( 1 + \frac{\xi k}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^{\delta+1} - \frac{1}{\delta} \left\{ \left( 1 + \frac{\xi k}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^{\frac{\kappa}{2}} \right] \quad (\text{B.7})
\end{aligned}$$

The log likelihood function is defined as

$$\begin{aligned}
l(\kappa, \delta, \sigma, \xi) = \sum_{j=1}^n \log & \left[ \left[ 1 - \left( 1 + \frac{\xi(k_j+1)}{\sigma} \right)^{-\frac{1}{\xi}} + \frac{1}{\delta} \left\{ \left( 1 + \frac{\xi(k_j+1)}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^{\delta+1} \right. \right. \\
& - \frac{1}{\delta} \left\{ \left( 1 + \frac{\xi(k_j+1)}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^{\frac{\kappa}{2}} \left. \right]^{\frac{\kappa}{2}} - \left[ 1 - \left( 1 + \frac{\xi k_j}{\sigma} \right)^{-\frac{1}{\xi}} + \frac{1}{\delta} \left\{ \left( 1 + \frac{\xi k_j}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^{\delta+1} \right. \\
& \left. \left. - \frac{1}{\delta} \left\{ \left( 1 + \frac{\xi k_j}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^{\frac{\kappa}{2}} \right] \right] \quad (\text{B.8})
\end{aligned}$$

(iv) The PMF corresponding to  $G(u; \psi) = pu^{\kappa_1} + (1-p)u^{\kappa_2}$ ,  $\psi = (p, \kappa_1, \kappa_2) > 0$  with  $\kappa_1 \leq \kappa_2$  is written as

$$\begin{aligned}
P(Y = k) = p & \left[ \left\{ 1 - \left( 1 + \frac{\xi(k+1)}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^{\kappa_1} - \left\{ 1 - \left( 1 + \frac{\xi k}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^{\kappa_1} \right] \\
& + (1-p) \left[ \left\{ 1 - \left( 1 + \frac{\xi(k+1)}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^{\kappa_2} - \left\{ 1 - \left( 1 + \frac{\xi k}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^{\kappa_2} \right] \quad (\text{B.9})
\end{aligned}$$

The likelihood function is defined as

$$\begin{aligned}
l(p, \kappa_1, \kappa_2, \sigma, \xi) = \sum_{j=1}^n \log & \left[ p \left[ \left\{ 1 - \left( 1 + \frac{\xi(k_j+1)}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^{\kappa_1} - \left\{ 1 - \left( 1 + \frac{\xi k_j}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^{\kappa_1} \right] \right. \\
& \left. + (1-p) \left[ \left\{ 1 - \left( 1 + \frac{\xi(k_j+1)}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^{\kappa_2} - \left\{ 1 - \left( 1 + \frac{\xi k_j}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^{\kappa_2} \right] \right] \quad (\text{B.10})
\end{aligned}$$

To find the estimates of unknown parameters of above model, we solve the derivatives of log likelihood function of model (i), (ii), (iii) and (iv) numerically.

## B.2 Maximum likelihood procedure of ZIDEGPD

(i) The PMF of ZIDEGPD based on  $G(u; \psi) = u^\kappa$ ,  $\psi = \kappa > 0$  is written as

$$P(Z = m) = \begin{cases} \pi + (1 - \pi) \left[ 1 - \left( 1 + \frac{\xi}{\sigma} \right)^{-\frac{1}{\xi}} \right]^\kappa, & m = 0 \\ (1 - \pi) \left[ \left\{ 1 - \left( 1 + \frac{\xi(m+1)}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^\kappa - \left\{ 1 - \left( 1 + \frac{\xi(m)}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^\kappa \right], & m > 0 \end{cases} \quad (\text{B.11})$$

Thus, the likelihood function is defined as

$$L(\theta) = \left[ \pi + (1 - \pi) G \left\{ F_\xi \left( \frac{1}{\sigma} \right) \right\} \right]^r \prod_{j=1; m_j \neq 0}^n (1 - \pi) \left[ G \left\{ F_\xi \left( \frac{m_j + 1}{\sigma} \right) \right\} - G \left\{ F_\xi \left( \frac{m_j}{\sigma} \right) \right\} \right] \quad (\text{B.12})$$

The log likelihood function is

$$\begin{aligned} l(\pi, \kappa, \sigma, \xi) &= r \log \left[ \pi + (1 - \pi) \left\{ 1 - \left( 1 + \frac{\xi}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^\kappa \right] + (n - r) \log(1 - \pi) \\ &+ \sum_{j=1}^n \log \left[ \left\{ 1 - \left( 1 + \frac{\xi(m_j + 1)}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^\kappa - \left\{ 1 - \left( 1 + \frac{\xi(m_j)}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^\kappa \right] \end{aligned} \quad (\text{B.13})$$

(ii) The PMF of ZIDEGPD obtained via  $G(u; \psi) = 1 - D_\delta \{(1 - u)^\delta\}$ ,  $\psi = \delta > 0$

$$P(Z = m) = \begin{cases} \pi + (1 - \pi) \left[ \frac{1}{\delta} \left\{ \left( 1 + \frac{\xi}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^{\delta+1} - \frac{\delta+1}{\delta} \left\{ \left( 1 + \frac{\xi}{\sigma} \right)^{-\frac{1}{\xi}} \right\} + 1 \right], & m = 0 \\ (1 - \pi) \left[ \frac{1}{\delta} \left\{ \left( 1 + \frac{\xi(m+1)}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^{\delta+1} - \frac{\delta+1}{\delta} \left\{ \left( 1 + \frac{\xi(m+1)}{\sigma} \right)^{-\frac{1}{\xi}} \right\} \right. \\ \left. - \frac{1}{\delta} \left\{ \left( 1 + \frac{\xi(m)}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^{\delta+1} + \frac{\delta+1}{\delta} \left\{ \left( 1 + \frac{\xi(m)}{\sigma} \right)^{-\frac{1}{\xi}} \right\} \right], & m > 0 \end{cases} \quad (\text{B.14})$$

The log-likelihood function is defined as

$$\begin{aligned} l(\pi, \delta, \sigma, \xi) &= r \log \left[ \pi + (1 - \pi) \left[ \frac{1}{\delta} \left\{ \left( 1 + \frac{\xi}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^{\delta+1} - \frac{\delta+1}{\delta} \left\{ \left( 1 + \frac{\xi}{\sigma} \right)^{-\frac{1}{\xi}} \right\} + 1 \right] \right] \\ &+ (n - r) \log(1 - \pi) + \sum_{j=1}^n \log \left[ \frac{1}{\delta} \left\{ \left( 1 + \frac{\xi(m_j + 1)}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^{\delta+1} \right. \\ &\left. - \frac{\delta+1}{\delta} \left\{ \left( 1 + \frac{\xi(m_j + 1)}{\sigma} \right)^{-\frac{1}{\xi}} \right\} - \frac{1}{\delta} \left\{ \left( 1 + \frac{\xi(m_j)}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^{\delta+1} \right. \\ &\left. + \frac{\delta+1}{\delta} \left\{ \left( 1 + \frac{\xi(m_j)}{\sigma} \right)^{-\frac{1}{\xi}} \right\} \right] \end{aligned} \quad (\text{B.15})$$

(iii) The PMF of ZIDEGPD corresponding to  $G(u; \psi) = [1 - D_\delta \{(1 - u)^\delta\}]^{\kappa/2}$ ;  $\psi = (\delta, \kappa) > 0$  is

$$P(Z = m) = \begin{cases} \pi + (1 - \pi) \left[ \left[ 1 - \left(1 + \frac{\xi}{\sigma}\right)^{-\frac{1}{\xi}} + \frac{1}{\delta} \left\{ \left(1 + \frac{\xi}{\sigma}\right)^{-\frac{1}{\xi}} \right\}^{\delta+1} - \frac{1}{\delta} \left\{ \left(1 + \frac{\xi}{\sigma}\right)^{-\frac{1}{\xi}} \right\} \right]^{\frac{\kappa}{2}} \right], & m = 0 \\ (1 - \pi) \left[ \left[ 1 - \left(1 + \frac{\xi(m+1)}{\sigma}\right)^{-\frac{1}{\xi}} + \frac{1}{\delta} \left\{ \left(1 + \frac{\xi(m+1)}{\sigma}\right)^{-\frac{1}{\xi}} \right\}^{\delta+1} - \frac{1}{\delta} \left\{ \left(1 + \frac{\xi(m+1)}{\sigma}\right)^{-\frac{1}{\xi}} \right\} \right]^{\frac{\kappa}{2}} \right. \\ \left. - \left[ 1 - \left(1 + \frac{\xi m}{\sigma}\right)^{-\frac{1}{\xi}} + \frac{1}{\delta} \left\{ \left(1 + \frac{\xi m}{\sigma}\right)^{-\frac{1}{\xi}} \right\}^{\delta+1} - \frac{1}{\delta} \left\{ \left(1 + \frac{\xi m}{\sigma}\right)^{-\frac{1}{\xi}} \right\} \right]^{\frac{\kappa}{2}} \right], & m > 0 \end{cases} \quad (\text{B.16})$$

The log-likelihood function is defined as

$$\begin{aligned} l(\pi, \kappa, \delta, \sigma, \xi) = r \log & \left[ \pi + (1 - \pi) \left[ \left[ 1 - \left(1 + \frac{\xi}{\sigma}\right)^{-\frac{1}{\xi}} + \frac{1}{\delta} \left\{ \left(1 + \frac{\xi}{\sigma}\right)^{-\frac{1}{\xi}} \right\}^{\delta+1} - \frac{1}{\delta} \left\{ \left(1 + \frac{\xi}{\sigma}\right)^{-\frac{1}{\xi}} \right\} \right]^{\frac{\kappa}{2}} \right] \right] \\ & + (n - r) \log(1 - \pi) + \sum_{j=1}^n \log \left[ \left[ 1 - \left(1 + \frac{\xi(m_j + 1)}{\sigma}\right)^{-\frac{1}{\xi}} \right. \right. \\ & \left. \left. + \frac{1}{\delta} \left\{ \left(1 + \frac{\xi(m_j + 1)}{\sigma}\right)^{-\frac{1}{\xi}} \right\}^{\delta+1} - \frac{1}{\delta} \left\{ \left(1 + \frac{\xi(m_j + 1)}{\sigma}\right)^{-\frac{1}{\xi}} \right\} \right]^{\frac{\kappa}{2}} - \right. \\ & \left. \left[ 1 - \left(1 + \frac{\xi m_j}{\sigma}\right)^{-\frac{1}{\xi}} + \frac{1}{\delta} \left\{ \left(1 + \frac{\xi m_j}{\sigma}\right)^{-\frac{1}{\xi}} \right\}^{\delta+1} - \frac{1}{\delta} \left\{ \left(1 + \frac{\xi m_j}{\sigma}\right)^{-\frac{1}{\xi}} \right\} \right]^{\frac{\kappa}{2}} \right] \end{aligned} \quad (\text{B.17})$$

(iv) The PMF of ZIDGPD corresponding to  $G(u; \psi) = pu^{\kappa_1} + (1 - p)u^{\kappa_2}$ ,  $\psi = (p, \kappa_1, \kappa_2) > 0$  is

$$P(Z = m) = \begin{cases} \pi + (1 - \pi) \left[ p \left\{ 1 - \left(1 + \frac{\xi}{\sigma}\right)^{-\frac{1}{\xi}} \right\}^{\kappa_1} + (1 - p) \left[ \left\{ 1 - \left(1 + \frac{\xi}{\sigma}\right)^{-\frac{1}{\xi}} \right\}^{\kappa_2} \right] \right], & m = 0 \\ (1 - \pi) \left[ p \left[ \left\{ 1 - \left(1 + \frac{\xi(m+1)}{\sigma}\right)^{-\frac{1}{\xi}} \right\}^{\kappa_1} - \left\{ 1 - \left(1 + \frac{\xi m}{\sigma}\right)^{-\frac{1}{\xi}} \right\}^{\kappa_1} \right] \right. \\ \left. + (1 - p) \left[ \left\{ 1 - \left(1 + \frac{\xi(m+1)}{\sigma}\right)^{-\frac{1}{\xi}} \right\}^{\kappa_2} - \left\{ 1 - \left(1 + \frac{\xi m}{\sigma}\right)^{-\frac{1}{\xi}} \right\}^{\kappa_2} \right] \right], & x > 0 \end{cases} \quad (\text{B.18})$$

The likelihood function is defined as

$$\begin{aligned} l(\pi, p, \kappa_1, \kappa_2, \sigma, \xi) = r \log & \left[ \pi + (1 - \pi) \left[ p \left\{ 1 - \left(1 + \frac{\xi}{\sigma}\right)^{-\frac{1}{\xi}} \right\}^{\kappa_1} + (1 - p) \left[ \left\{ 1 - \left(1 + \frac{\xi}{\sigma}\right)^{-\frac{1}{\xi}} \right\}^{\kappa_2} \right] \right] \right] \\ & + (n - r) \log(1 - \pi) + \sum_{j=1}^n \log \left[ p \left[ \left\{ 1 - \left(1 + \frac{\xi(m_j + 1)}{\sigma}\right)^{-\frac{1}{\xi}} \right\}^{\kappa_1} - \left\{ 1 - \left(1 + \frac{\xi m_j}{\sigma}\right)^{-\frac{1}{\xi}} \right\}^{\kappa_1} \right] \right. \\ & \left. + (1 - p) \left[ \left\{ 1 - \left(1 + \frac{\xi(m_j + 1)}{\sigma}\right)^{-\frac{1}{\xi}} \right\}^{\kappa_2} - \left\{ 1 - \left(1 + \frac{\xi m_j}{\sigma}\right)^{-\frac{1}{\xi}} \right\}^{\kappa_2} \right] \right] \end{aligned}$$

$$+ (1-p) \left[ \left\{ 1 - \left( 1 + \frac{\xi(m_j+1)}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^{\kappa_2} - \left\{ 1 - \left( 1 + \frac{\xi m_j}{\sigma} \right)^{-\frac{1}{\xi}} \right\}^{\kappa_2} \right] \quad (\text{B.19})$$

Similar to DEGPD, the unknown parameters of ZIDEGPD models, we solve the derivatives of log likelihood function of model (i), (ii), (iii) and (iv) numerically.

TABLE S1: Root mean square errors of parameter estimates ZIDEGPD found from  $10^4$  independent data sets of size  $n = 1000$ .

Model type (i) $G(u; \psi) = u^\kappa$											
$\pi$		$\kappa$		$\sigma$		$\xi$		-		-	
TRUE	RMSE	TRUE	RMSE	TRUE	RMSE	TRUE	RMSE	TRUE	RMSE	TRUE	RMSE
0.2	0.04	5	3.00	1	0.28	0.2	0.06				
0.2	0.01	10	3.19	1	0.20	0.2	0.04				
0.5	0.03	5	3.83	1	0.34	0.2	0.07				
0.5	0.02	10	3.97	1	0.25	0.2	0.05				
Model type (ii) $G(u; \psi) = 1 - D_\delta\{(1-u)^\delta\}$											
$\pi$		$\delta$		$\sigma$		$\xi$		-		-	
TRUE	RMSE	TRUE	RMSE	TRUE	RMSE	TRUE	RMSE	TRUE	RMSE	TRUE	RMSE
0.2	0.09	1	4.18	1	0.22	0.20	0.06				
0.2	0.10	5	3.08	1	0.17	0.20	0.06				
0.5	0.06	1	4.10	1	0.20	0.20	0.07				
0.5	0.09	5	1.72	1	0.20	0.20	0.08				
Model type (iii) $G(u; \psi) = [1 - D_\delta\{(1-u)^\delta\}]^{\kappa/2}$											
$\pi$		$\kappa$		$\delta$		$\sigma$		$\xi$		-	
TRUE	RMSE	TRUE	RMSE	TRUE	RMSE	TRUE	RMSE	TRUE	RMSE	TRUE	RMSE
0.2	0.03	5	6.09	1	1.90	1	0.29	0.20	0.05		
0.2	0.02	10	4.20	5	4.10	1	0.27	0.20	0.05		
0.5	0.03	5	4.49	1	1.63	1	0.38	0.20	0.07		
0.5	0.02	10	5.41	5	4.80	1	0.34	0.20	0.07		
Model type (iv) $G(u; \psi) = pu^{\kappa_1} + (1-p)u^{\kappa_2}$											
$\pi$		$p$		$\kappa_1$		$\kappa_2$		$\sigma$		$\xi$	
TRUE	RMSE	TRUE	RMSE	TRUE	RMSE	TRUE	RMSE	TRUE	RMSE	TRUE	RMSE
0.2	0.14	0.5	0.29	1	4.03	5	4.16	1	0.34	0.20	0.07
0.2	0.07	0.5	0.30	5	2.24	10	9.40	1	0.25	0.20	0.05
0.5	0.23	0.5	0.35	1	0.66	5	2.14	1	0.37	0.20	0.08
0.5	0.09	0.5	0.33	5	2.33	10	7.64	1	0.27	0.2	0.06

TABLE S2: Estimated coefficients and smooth terms for GAM form ZIDEGPD models fitted to Avalanches data of Haute-Maurienne massif of French Alps

Model type (i) $G(u; \psi) = u^\kappa$				
** Parametric terms **				
Parameter (intercept)	Estimate	Std.Error	t value	P-value
$\log(\kappa)$	-1.83	0.06	-32.25	<2e-16
$\log(\sigma)$	0.2	0.1	1.99	0.0232
$\log(\xi)$	-0.53	0.08	-6.84	4.04e-12
$\text{logit}(\pi)$	-52.06	0.09	-16.84	6.02e-13
** Smooth terms **				
	edf	max.df	Chi.sq	Pr(>  t )
$\log(\sigma)$				
s(WS)	1.02	4	23.56	1.29e-06
s(MxT)	3.99	4	652.73	<2e-16
s(PREC)	1.00	4	36.41	1.6e-09
s(RH)	4.19	9	70.35	7.59e-14
Model type (ii) $G(u; \psi) = 1 - F_\delta\{(1 - u)^\delta\}$				
** Parametric terms **				
Parameter (intercept)	Estimate	Std.Error	t value	P-value
$\log(\delta)$	4.67	599.22	0.01	0.497
$\log(\sigma)$	-0.73	0.08	-8.92	<2e-16
$\log(\xi)$	-0.3	0.05	-6.17	3.44e-10
$\text{logit}(\pi)$	0.48	9.04	0.05	0.479
** Smooth terms **				
	edf	max.df	Chi.sq	Pr(>  t )
$\log(\sigma)$				
s(WS)	1.71	4	22.34	1.3e-05
s(MxT)	2.54	4	692.76	<2e-16
s(PREC)	1.03	4	33.96	7.51e-09
s(RH)	5.50	9	85.89	3.46e-16

*Cont...*

TABLE S2 : (*Cont...*) Estimated coefficients and smooth terms for GAM form ZIDEGPD models fitted to Avalanches data of Haute-Maurienne massif of French Alps

Model type (iii) $G(u; \psi) = [1 - D_\delta \{(1 - u)^\delta\}]^{\kappa/2}$				
** Parametric terms **				
Parameter (intercept)	Estimate	Std.Error	t value	P-value
$\log(\kappa)$	-1.23	0.21	-5.96	1.27e-09
$\log(\delta)$	2.1	0.43	4.86	5.83e-07
$\log(\sigma)$	0.52	0.01	37.92	<2e-16
$\log(\xi)$	-0.65	0.07	-9.43	<2e-16
$\text{logit}(\pi)$	-0.96	0.59	-1.62	0.0528
** Smooth terms **				
$\log(\sigma)$				
s(WS)	1.00	4	610.20	<2e-16
s(MxT)	3.95	4	11340.79	<2e-16
s(PREC)	1.01	4	792.57	<2e-16
s(RH)	7.89	9	1701.40	<2e-16







# Bibliography

- Anderson, C. (1970) Extreme value theory for a class of discrete distributions with applications to some stochastic processes. *Journal of Applied Probability* **7**(1), 99–113.
- Anderson, C. (1980) Local limit theorems for the maxima of discrete random variables. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 88, pp. 161–165.
- Bacro, J.-N., Gaetan, C., Opitz, T. and Toulemonde, G. (2020) Hierarchical space-time modeling of asymptotically independent exceedances with an application to precipitation data. *Journal of the American Statistical Association* **115**(530), 555–569.
- Bacro, J.-N., Gaetan, C., Opitz, T. and Toulemonde, G. (2023) Multivariate peaks-over-threshold with latent variable representations of generalized Pareto vectors. *Preprint* .
- Barbiero, A. (2019) A bivariate Geometric distribution allowing for positive or negative correlation. *Communications in Statistics-Theory and Methods* **48**(11), 2842–2861.
- Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. L. (2004) *Statistics of Extremes: Theory and Applications*. John Wiley & Sons.
- Bhati, D. and Bakouch, H. S. (2019) A new infinitely divisible discrete distribution with applications to count data modeling. *Communications in Statistics-Theory and Methods* **48**(6), 1401–1416.
- Biller, B. and Corlu, C. G. (2012) Copula-based multivariate input modeling. *Surveys in Operations Research and Management Science* **17**(2), 69–84.
- Billingsley, P. (1995) *Probability and measure*. John Wiley & Sons.
- Bopp, G. P. and Shaby, B. A. (2017) An Exponential–Gamma mixture model for extreme santa ana winds. *Environmetrics* **28**(8), e2476.

- Bortot, P. and Gaetan, C. (2014) A latent process model for temporal extremes. *Scandinavian Journal of Statistics* **41**(3), 606–621.
- Bortot, P. and Gaetan, C. (2022) A model for space-time threshold exceedances with an application to extreme rainfall. *Statistical Modelling* p. 1471082X221098224.
- Bortot, P. and Tawn, J. A. (1998) Models for the extremes of markov chains. *Biometrika* **85**(4), 851–867.
- Boutsikas, M. and Koutras, M. (2002) Modeling claim exceedances over thresholds. *Insurance: Mathematics and Economics* **30**(1), 67–83.
- Buddana, A. and Kozubowski, T. J. (2014) Discrete Pareto distributions. *Economic Quality Control* **29**(2), 143–156.
- Carlstein, E. (1986) The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *The Annals of Statistics* pp. 1171–1179.
- Carreau, J. and Bengio, Y. (2009) A hybrid Pareto model for asymmetric fat-tailed data: the univariate case. *Extremes* **12**(1), 53–76.
- Carrer, N. L. and Gaetan, C. (2022) Distributional regression models for Extended Generalized Pareto distributions. *arXiv preprint arXiv:2209.04660* .
- de Carvalho, M., Pereira, S., Pereira, P. and de Zea Bermudez, P. (2022) An extreme value Bayesian lasso for the conditional left and right tails. *Journal of Agricultural, Biological and Environmental Statistics* **27**(2), 222–239.
- Casson, E. and Coles, S. (1999) Spatial regression models for extremes. *Extremes* **1**(4), 449–468.
- Chakraborty, S. and Chakravarty, D. (2012) Discrete Gamma distributions: Properties and parameter estimations. *Communications in Statistics-Theory and Methods* **41**(18), 3301–3324.
- Chavez-Demoulin, V. and Davison, A. (2012) Modelling time series extremes. *REVSTAT-Statistical Journal* **10**(1), 109–133.
- Chavez-Demoulin, V. and Davison, A. C. (2005) Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **54**(1), 207–222.

- Cherubini, U., Luciano, E. and Vecchiato, W. (2004) *Copula Methods in Finance*. John Wiley & Sons.
- Chiogna, M. and Gaetan, C. (2007) Semiparametric zero-inflated Poisson models with application to animal abundance studies. *Environmetrics* **18**(3), 303–314.
- Choulakian, V. and Stephens, M. A. (2001) Goodness-of-fit tests for the generalized Pareto distribution. *Technometrics* **43**(4), 478–484.
- Coles, S. (2001) *An Introduction to Statistical Modeling of Extreme Values*. Springer, London.
- Coles, S., Heffernan, J. and Tawn, J. (1999) Dependence measures for extreme value analyses. *Extremes* **2**(4), 339–365.
- Coles, S. G. and Tawn, J. A. (1991) Modelling extreme multivariate events. *Journal of the Royal Statistical Society: Series B (Methodological)* **53**(2), 377–392.
- Coles, S. G. and Tawn, J. A. (1994) Statistical methods for multivariate extremes: an application to structural design. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **43**(1), 1–31.
- Constantinescu, C. D., Kozubowski, T. J. and Qian, H. H. (2019) Probability of ruin in discrete insurance risk model with dependent Pareto claims. *Dependence Modeling* **7**(1), 215–233.
- Courseau, V. and Veraart, A. E. (2022) Asymptotic theory for the inference of the latent trawl model for extreme values. *Scandinavian Journal of Statistics* **49**(4), 1448–1495.
- Das, B. and Resnick, S. I. (2011) Detecting a conditional extreme value model. *Extremes* **14**(1), 29–61.
- Davis, R. A. and Mikosch, T. (2009) The extremogram: A correlogram for extreme events. *Bernoulli* **15**(4), 977–1009.
- Davis, R. A., Mikosch, T. and Cribben, I. (2012) Towards estimating extremal serial dependence via the bootstrapped extremogram. *Journal of Econometrics* **170**(1), 142–152.
- Davison, A. C. and Smith, R. L. (1990) Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B (Methodological)* **52**(3), 393–425.

- De Carvalho, M. and Ramos, A. (2012) Bivariate extreme statistics, ii. *REVSTAT-Statistical Journal* **10**(1), 83–107.
- De Haan, L. and De Ronde, J. (1998) Sea and wind: multivariate extremes at work. *Extremes* **1**(1), 7–45.
- De Haan, L. and Resnick, S. I. (1977) Limit theory for multivariate sample extremes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **40**(4), 317–337.
- De Waal, D. and Van Gelder, P. (2005) Modelling of extreme wave heights and periods through copulas. *Extremes* **8**(4), 345–356.
- Dey, D. K. and Yan, J. (2016) *Extreme Value Modeling and Risk Analysis: Methods and Applications*. Chapman & Hall/CRC.
- Dkengne, P. S., Eckert, N. and Naveau, P. (2016) A limiting distribution for maxima of discrete stationary triangular arrays with an application to risk due to avalanches. *Extremes* **19**(1), 25–40.
- Dunn, P. K. and Smyth, G. K. (1996) Randomized quantile residuals. *Journal of Computational and Graphical Statistics* **5**(3), 236–244.
- Dupuis, D. J. (1999) Exceedances over high thresholds: A guide to threshold selection. *Extremes* **1**(3), 251–261.
- Dutfoy, A., Parey, S. and Roche, N. (2014) Multivariate extreme value theory—a tutorial with applications to hydrology and meteorology. *Dependence Modeling* **2**(1), 30–48.
- Eastoe, E., Koukoulas, S. and Jonathan, P. (2013) Statistical measures of extremal dependence illustrated using measured sea surface elevations from a neighbourhood of coastal locations. *Ocean Engineering* **62**(1), 68–77.
- Embrechts, P., Kluppelberg, C. and Mikosch, T. (1999) Modelling extremal events. *British Actuarial Journal* **5**(2), 465–465.
- Embrechts, P., McNeil, A. and Straumann, D. (2002) Correlation and dependence in risk management: properties and pitfalls. *Risk Management: Value at Risk and Beyond* **1**, 176–223.
- Evin, G., Sielenou, P. D., Eckert, N., Naveau, P., Hagenmuller, P. and Morin, S. (2021) Extreme avalanche cycles: Return levels and probability distributions depending on snow and meteorological conditions. *Weather and Climate Extremes* **33**(1), 100344.

- Ewans, K. and Jonathan, P. (2014) Evaluating environmental joint extremes for the offshore industry using the conditional extremes model. *Journal of Marine Systems* **130**, 124–130.
- Farlie, D. J. (1960) The performance of some correlation coefficients for a general bivariate distribution. *Biometrika* **47**(3/4), 307–323.
- Fisher, R. A. and Tippett, L. H. C. (1928) Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Proc. Cambridge Philos. Soc.*, *24*., number 2, pp. 180–190.
- Frigessi, A., Haug, O. and Rue, H. (2002) A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extremes* **5**(3), 219–235.
- Furrer, R. and Naveau, P. (2007) Probability weighted moments properties for small samples. *Statistics and Probability Letters* **77**(1), 190–195.
- Gaetan, C. and Grigoletto, M. (2004) Smoothing sample extremes with dynamic models. *Extremes* **7**(3), 221–236.
- Gamet, P. and Jalbert, J. (2022) A flexible extended generalized Pareto distribution for tail estimation. *Environmetrics* **33**(6), e2744.
- Gardes, L. and Girard, S. (2010) Conditional extremes from heavy-tailed distributions: An application to the estimation of extreme rainfall return levels. *Extremes* **13**(2), 177–204.
- Gaver, D. P. and Lewis, P. (1980) First-order autoregressive Gamma sequences and point processes. *Advances in Applied Probability* **12**(3), 727–745.
- Genest, C. and Nešlehová, J. (2007) A primer on copulas for count data. *ASTIN Bulletin: The Journal of the IAA* **37**(2), 475–515.
- Gorgi, P. (2020) Beta–Negative Binomial auto-regressions for modelling integer-valued time series with extreme observations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **82**(5), 1325–1347.
- Gudendorf, G. and Segers, J. (2010) Extreme-value Copulas. In *Copula Theory and its Applications*, pp. 127–145. Springer Berlin Heidelberg.
- Gumbel, E. J. (1960) Bivariate Exponential distributions. *Journal of the American Statistical Association* **55**(292), 698–707.

- Gumbel, E. J. (1961) Bivariate Logistic distributions. *Journal of the American Statistical Association* **56**(294), 335–349.
- Gumbel, E. J. and Mustafi, C. K. (1967) Some analytical properties of bivariate extremal distributions. *Journal of the American Statistical Association* **62**(318), 569–588.
- de Hann, L. and Ferreira, A. (2006) *Extreme Value Theory: An Introduction*. Springer-Verlag, New York.
- Hastie, T. and Tibshirani, R. (1990) *Generalized Additive Models*. New York: Chapman & Hall, CRC Press.
- Heffernan, J. E. and Tawn, J. A. (2004) A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66**(3), 497–546.
- Hitz, A., Davis, R. and Samorodnitsky, G. (2017) Discrete extremes. *arXiv preprint arXiv:1707.05033* .
- Hitz, A. S. and Evans, R. J. (2016) Modeling website visits. *arXiv preprint arXiv:1611.01024* .
- Hu, Y. and Scarrott, C. (2018) evmix: An r package for extreme value mixture modeling, threshold estimation and boundary corrected kernel density estimation. *Journal of Statistical Software* **85**(5), 1–27.
- Huerta, G. and Sansó, B. (2007) Time-varying models for extreme values. *Environmental and Ecological Statistics* **14**(1), 285–299.
- Huser, R. and Wadsworth, J. L. (2020) Advances in statistical modeling of spatial extremes. *Wiley Interdisciplinary Reviews: Computational Statistics* p. e1537.
- Inouye, D. I., Yang, E., Allen, G. I. and Ravikumar, P. (2017) A review of multivariate distributions for count data derived from the Poisson distribution. *Wiley Interdisciplinary Reviews: Computational Statistics* **9**(3), e1398.
- Jonathan, P., Flynn, J. and Ewans, K. (2010) Joint modelling of wave spectral parameters for extreme sea states. *Ocean Engineering* **37**(11), 1070–1080.
- Katz, R. W., Parlange, M. B. and Naveau, P. (2002) Statistics of extremes in hydrology. *Advances in Water Resources* **25**(8), 1287–1304.

- Kibble, W. (1941) A two-variate Gamma type distribution. *Sankhyā: The Indian Journal of Statistics* pp. 137–150.
- Kotz, S. and Nadarajah, S. (2000) *Extreme Value Distributions: Theory and Applications*. Imperial College Press, London.
- Kozubowski, T. J., Panorska, A. K. and Forister, M. L. (2015) A discrete truncated Pareto distribution. *Statistical Methodology* **26**, 135–150.
- Krishna, H. and Pundir, P. S. (2009) Discrete Burr and discrete Pareto distributions. *Statistical Methodology* **6**(2), 177–188.
- Lai, C. D. and Balakrishnan, N. (2009) *Continuous bivariate distributions*. Springer, New York.
- Lambert, D. (1992) Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* **34**(1), 1–14.
- Le Gall, P., Favre, A.-C., Naveau, P. and Prieur, C. (2022) Improved regional frequency analysis of rainfall data. *Weather and Climate Extremes* **36**(1), 1–13.
- Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1983) *Extremes and Related Properties of Random Sequences and Processes*. Springer, Berlin.
- Ledford, A. W. and Tawn, J. A. (1996) Statistics for near independence in multivariate extreme values. *Biometrika* **83**(1), 169–187.
- Ledford, A. W. and Tawn, J. A. (1997) Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **59**(2), 475–499.
- Lindsay, B. G. (1988) Composite likelihood methods. *Contemporary Mathematics* **80**(1), 221–239.
- MacDonald, A., Scarrott, C. J., Lee, D., Darlow, B., Reale, M. and Russell, G. (2011) A flexible extreme value mixture model. *Computational Statistics & Data Analysis* **55**(6), 2137–2157.
- Marshall, A. W. and Olkin, I. (1988) Families of multivariate distributions. *Journal of the American Statistical Association* **83**(403), 834–841.
- McNeil, A. J., Frey, R. and Embrechts, P. (2015) *Quantitative risk management: concepts, techniques and tools-revised edition*. Princeton university press.

- Morgenstern, D. (1956) Einfache Beispiele zweidimensionaler Verteilungen. *Mitteilungsblatt für Mathematische Statistik* **8**(1), 234–235.
- Mougin, P. (1922) Avalanches in Savoy, vol. iv. *Ministry of Agriculture, General Directorate of Water and Forests, Department of Great Hydraulic Forces, Paris*.
- Naghattini, M. (2017) *Fundamentals of Statistical Hydrology*. Springer International Publishing Switzerland.
- Naveau, P., Huser, R., Ribereau, P. and Hannart, A. (2016) Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. *Water Resources Research* **52**(4), 2753–2769.
- Nelsen, R. B. (2007) *An Introduction to Copulas*. Springer, New York.
- Nikoloulopoulos, A. K. (2013a) Copula-based models for multivariate discrete response data. In *Copulae in Mathematical and Quantitative Finance*, pp. 231–249. Springer.
- Nikoloulopoulos, A. K. (2013b) On the estimation of normal copula discrete regression models using the continuous extension and simulated likelihood. *Journal of Statistical Planning and Inference* **143**(11), 1923–1937.
- Nikoloulopoulos, A. K. and Karlis, D. (2009) Modeling multivariate count data using copulas. *Communications in Statistics-Simulation and Computation* **39**(1), 172–187.
- Noven, R. C., Veraart, A. E. and Gandy, A. (2018) A latent trawl process model for extreme values. *Journal of Energy Markets* **11**(3), 1–24.
- Pauli, F. and Coles, S. (2001) Penalized likelihood inference in extreme value analyses. *Journal of Applied Statistics* **28**(5), 547–560.
- Pedeli, X. and Varin, C. (2020) Pairwise likelihood estimation of latent autoregressive count models. *Statistical Methods in Medical Research* **29**(11), 3278–3293.
- Peres, M. V. d. O., Achcar, J. A. and Martinez, E. Z. (2018) Bivariate modified Weibull distribution derived from Farlie-Gumbel-Morgenstern copula: a simulation study. *Electronic Journal of Applied Statistical Analysis* **11**(2), 463–488.
- Pickands, J. (1975) Statistical inference using extreme order statistics. *The Annals of Statistics* pp. 119–131.
- Pickands, J. (1981) Multivariate extreme value distributions. *Bulletin of the International Statistical Institute* pp. 859–878.



- Prieto, F., Gómez-Déniz, E. and Sarabia, J. M. (2014) Modelling road accident blackspots data with the discrete generalized Pareto distribution. *Accident Analysis & Prevention* **71**, 38–49.
- Ramos, A. and Ledford, A. (2009) A new class of models for bivariate joint tails. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71**(1), 219–241.
- Ranjbar, S., Cantoni, E., Chavez-Demoulin, V., Marra, G., Radice, R. and Jatón-Ogay, K. (2022) Modelling the extremes of seasonal viruses and hospital congestion: The example of flu in a swiss hospital. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **71**(4), 884–905.
- Reich, B. J., Shaby, B. A. and Cooley, D. (2014) A hierarchical model for serially-dependent extremes: A study of heat waves in the western us. *Journal of Agricultural, Biological, and Environmental Statistics* **19**(1), 119–135.
- Ribatet, M. and Sedki, M. (2013) Extreme value copulas and max-stable processes. *Journal de la Société Française de Statistique* **154**(1), 138–150.
- Ribereau, P., Masiello, E. and Naveau, P. (2016) Skew generalized extreme value distribution: Probability-weighted moments estimation and application to block maxima procedure. *Communications in Statistics-Theory and Methods* **45**(17), 5037–5052.
- Rigby, R. A. and Stasinopoulos, D. M. (2005) Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **54**(3), 507–554.
- Sang, H. and Gelfand, A. E. (2009) Hierarchical modeling for extreme values observed over space and time. *Environmental and Ecological Statistics* **16**(3), 407–426.
- Sartori, N. (2006) Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions. *Journal of Statistical Planning and Inference* **136**(12), 4259–4275.
- Scarrott, C. and MacDonald, A. (2012) A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT—Statistical Journal* **10**(1), 33–60.
- Schlather, M. and Tawn, J. A. (2003) A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika* **90**(1), 139–156.
- Scotto, M. G., Weiß, C. H., Möller, T. A. and Gouveia, S. (2018) The max-inar (1) model for count processes. *Test* **27**(4), 850–870.

- Shimura, T. (2012) Discretization of distributions in the maximum domain of attraction. *Extremes* **15**, 299–317.
- Sibuya, M. (1960) Bivariate extreme statistics, i. *Annals of the Institute of Statistical Mathematics* **11**(3), 195–210.
- Simpson, E. S. and Wadsworth, J. L. (2021) Conditional modelling of spatio-temporal extremes for red sea surface temperatures. *Spatial Statistics* **41**, 100482.
- Sklar, M. (1959) Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris* **8**, 229–231.
- Smith, R. L., Tawn, J. A. and Coles, S. G. (1997) Markov chain models for threshold exceedances. *Biometrika* **84**(2), 249–268.
- Stasinopoulos, M. D., Rigby, R. A. and Bastiani, F. D. (2018) GAMLSS: A distributional regression approach. *Statistical Modelling* **18**, 248–273.
- Tawn, J. A. (1988) Bivariate extreme value theory: models and estimation. *Biometrika* **75**(3), 397–415.
- Tencaliec, P., Favre, A., Naveau, P., Prieur, C. and Nicolet, G. (2019) Flexible semi-parametric generalized Pareto modeling of the entire range of rainfall amount. *Environmetrics* **env.2582**.
- Varin, C. and Vidoni, P. (2005) A note on composite likelihood inference and model selection. *Biometrika* **92**(3), 519–528.
- Walker, S. G. (2000) A note on the innovation distribution of a Gamma distributed autoregressive process. *Scandinavian journal of statistics* **27**(3), 575–576.
- Warren, D. (1992) A multivariate Gamma distribution arising from a markov model. *Stochastic Hydrology and Hydraulics* **6**(3), 183–190.
- Wolpert, R. L. (2021) Lecture notes on stationary Gamma processes. *arXiv preprint arXiv:2106.00087* .
- Wood, S. N. (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**(1), 3–36.
- Wood, S. N. (2017) *Generalized Additive models: an Introduction with R*. Second edition. New York: Chapman and Hall/CRC.

- Wood, S. N., Pya, N. and Säfken, B. (2016) Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association* **111**(516), 1548–1563.
- Yadav, R., Huser, R. and Opitz, T. (2020) Spatial hierarchical modeling of threshold exceedances using rate mixtures. *Environmetrics* **32**(3), e2662.
- Yee, T. W. and Stephenson, A. G. (2007) Vector generalized linear and additive extreme value models. *Extremes* **10**(1), 1–19.
- Youngman, B. D. (2019) Generalized additive models for exceedances of high thresholds with an application to return level estimation for US wind gusts. *Journal of the American Statistical Association* **114**(528), 1865–1879.
- Youngman, B. D. (2020) evgam: an R package for generalized additive extreme value models. *arXiv preprint:2003.04067* .
- Zachary, S., Feld, G., Ward, G. and Wolfram, J. (1998) Multivariate extrapolation in the offshore environment. *Applied Ocean Research* **20**(5), 273–295.



# Touqeer Ahmad

## CURRICULUM VITAE

### Contact Information

---

University of Padova  
Department of Statistics  
via Cesare Battisti, 241-243  
35121 Padova. Italy.

Tel. +39 320 380 7270

e-mail: [touqeer.ahmad@studenti.unipd.it](mailto:touqeer.ahmad@studenti.unipd.it)

Personal webpage: <https://www.stat.unipd.it/ricerca/touqeer-ahmad>

### Current Position

---

#### PhD Student

*Since December 2019; (expected completion: May 2023)*

Department of Statistical Sciences, University of Padova, Italy.

*Thesis title: On the modeling of discrete extreme values*

Supervisor: Prof. Carlo Gaetan

Co-supervisor: Dr. Philippe Naveau

### Research interests

---

- Development of new discrete and continuous extreme value models with different kinds of dependence (asymptotic dependence or asymptotic independence) over the tail. Dependence could be a temporal or Spatio-temporal.
- Optimal threshold selection is not an easy task, modeling the entire range of discrete and continuous extremes is an interesting topic to improve the prediction of environmental extremes.
- Modeling with Generalized Additive model forms of the discrete and continuous extremes.
- Spatial modelling of discrete and continuous extremes.

### Education

---

*Septemeber 2015 – August 2017*

#### MS Statistics

International Islamic University, Islamabad, Pakistan, Faculty of Basic and Applied Sciences

Title of dissertation: “Rainfall Frequency Analysis in Pakistan using Bayesian Approach ”

Supervisor: Prof. Ishfaq Ahmad

Final mark: CGPA 3.85/4.00 (Gold medal+ distinction)

*Septemeber 2013 – August 2015*

#### Masters of Science in Statistics .

International Islamic University, Islamabad, Pakistan, Faculty of Basic and Applied Sciences

Final mark: CGPA 3.77/4.00

## Visiting periods

---

*March 2022 – July 2022*

Le Laboratoire des Sciences du Climat et de l'Environnement (LSCE), Paris, France.  
 Université de Versailles Saint-Quentin-en-Yvelines - UVSQ, Paris, France.  
 Supervisors: Prof. Philippe Naveau and Prof. Julien Worms

*December 2022 – February 2023*

Research center for statistics, University of Geneva, Switzerland.  
 Supervisor: Prof. Sebastian Engelke

## Work experience

---

*February 2018 – November 2019*

**Govt. of Punjab Higher Education Department, Pakistan.**  
 Lecturer Statistics.

*July 2016 – March 2018*

**Department of Statistics, AIOU, Islamabad, Pakistan.**  
 Teaching Assistant.

## Awards and Scholarship

---

*December 2019 – January 2023*

Won a fully funded scholarship grant from "Fondazione Cassa di Risparmio di Padua e Rovigo (CARIPARO)" for PhD in Statistical Sciences at the University of Padua, Italy.

*March 2019*

Received Gold Medal by standing at top in MS Statistics degree.

*August 2015*

Received Laptop from Prime Minister laptop scheme by standing at the top in Masters of Science in Statistics degree.

## Computer skills

---

- R language, Python
- C++
- SPSS, Minitab, Eviews
- Latex, MS Office

## Language skills

---

Urdu, Punjabi: native;  
 English: Advance (written/spoken)  
 Italian: basic (spoken);

## Publications

---

### Journals articles

**Ahmad, T.**, Ahmad, I., Arshad, I. A., & Almanjahie, I. M. (2022). An efficient Bayesian modelling of extreme winds in the favor of energy generation. *Energy Reports* **9**(1), 2980–2992. <https://doi.org/10.1016/j.egyr.2023.01.093>

**Ahmad, T.**, Ahmad, I., Arshad, I. A., & Bianco, N. (2021). A comprehensive study on the Bayesian modelling of extreme rainfall: a case study from Pakistan. *International Journal of Climatology* **42**(1), 208–224. <https://doi.org/10.1002/joc.7240>

Noor, F., Masood, S., Sabar, Y., Shah, S. B. H., **Ahmad, T.**, Abdollahi, A., & Sajid, A. (2021). Bayesian analysis of cancer data using a 4-component exponential mixture model. *Computational and Mathematical Methods in Medicine* **2021**, <https://doi.org/10.1155/2021/6289337/>.

Cheema, A. R., Firdous, S., **Ahmad, T.**, & Imran, M. (2021). Family planning and fertility reduction in Pakistan. *Ilkogretim Online*, **20**(5), 3617–3627. <https://www.ilkogretim-online.org/fulltext/218-1616651153.pdf>

Ahmad, I., **Ahmad, T.**, & Almanjahie, I. M. (2019). Modelling of extreme rainfall in Punjab, Pakistan using Bayesian and frequentist approach. *Applied Ecology and Environmental Research* **17**(6), 13729–13748. [https://doi.org/10.15666/aeer/1706\\_1372913748](https://doi.org/10.15666/aeer/1706_1372913748).

### Articles submitted to Journals

Ahmad, I., **Ahmad, T.**, Almanjahie, I. M., & Athar M. A., (2022). An estimation of regional and at-site quantiles of extreme winds under flood index procedure. *Theoretical and Applied Climatology* (Under review).

**Ahmad, T.**, Gaetan, C., & Naveau P., (2022). Modelling of discrete extremes through extended versions of discrete generalized Pareto distribution. *Computational Statistics & Data Analysis* <https://doi.org/10.48550/arXiv.2210.15253> (Under review).

### Working papers

**Ahmad, T.**, Gaetan, C., (2022). Bivariate dependent modelling of discrete avalanches extremes. *Ready for submission*.

**Ahmad, T.**, Gaetan, C., (2022). A latent process model for discrete temporal extremes. *Ready for submission*.

Ahmad, I., **Ahmad, T.**, Shafiq, U.R. (2022). A detailed study on quantification and modelling of drought characteristics using different copula families. *Ready for submission*.

### Presentations

---

**Ahmad, T.**, (2022). Some new versions of discrete extreme models. (Yearly) *Department of statistical science University of Padova, Italy*. Date, 17 February 2022.

**Ahmad, T.**, (2022). Some new versions of discrete extreme models. (Invited speaker) *Laboratoire de Mathématiques de Versailles*, Versailles, France. Date, 19 April 2022.

**Ahmad, T.**, (2022). Modelling the entire range of discrete extreme data. *International Conference on Recent Trends in Statistics & Data Analytics*, National University of Science and Technology, Islamabad. Date, 23 September 2022.

**Ahmad, T.**, Gaetan, C., & Naveau P., (2022). Modelling of discrete extremes through extended versions of discrete generalized Pareto distribution. 15th International Conference of the ERCIM Working Group on Computational and Methodological Statistics (CMStatistics) King's College London, UK. Date 17-19 December 2022.

## References

---

### **Prof. Carlo Gaetan**

Institution: Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, Italy

Address: office Z.A12 (Zeta A building, 1st floor), Campus Scientifico via Torino

Phone: +3941 234 8404

e-mail: gaetan@unive.it

### **Dr. Philippe Naveau**

Institution: Laboratoire des Sciences du Climat et l'Environnement (LSCE), Paris, France

Address: Building 701 Room 60A, Laboratoire des Sciences du Climat et l'Environnement (LSCE) CNRS Orme des Merisiers / Bat. 701 CE Saclay 91191 Gif-sur-Yvette FRANCE

Phone: (+33) 1 69 08 77 16

e-mail: philippe.naveau@lsce.ipsl.fr



