MODELS AND DIAGNOSTICS FOR PARSIMONIOUS DEPENDENCE WITH APPLICATIONS TO MULTIVARIATE EXTREMES

by

David Lee

B.Sc.(ActuarSc), The University of Hong Kong, 2010
M.Phil., The University of Hong Kong, 2012

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY in

The Faculty of Graduate and Postdoctoral Studies (Statistics)

THE UNIVERSITY OF BRITISH COLUMBIA (Vancouver)

December 2016

© David Lee 2016
Abstract

Statistical models with parsimonious dependence are useful for high-dimensional modelling as they offer interpretations relevant to the data being fitted and may be computationally more manageable. We propose parsimonious models for multivariate extremes; in particular, extreme value (EV) copulas with factor and truncated vine structures are developed, through (a) taking the EV limit of a factor copula, or (b) structuring the underlying correlation matrix of existing multivariate EV copulas. Through data examples, we demonstrate that these models allow interpretation of the respective structures and offer insight on the dependence relationship among variables.

The strength of pairwise dependence for extreme value copulas can be described using the extremal coefficient. We consider a generalization of the F-madogram estimator for the bivariate extremal coefficient to the estimation of tail dependence of an arbitrary bivariate copula. This estimator is tail-weighted in the sense that the joint upper or lower portion of the copula is given a higher weight than the middle, thereby emphasizing tail dependence. The proposed estimator is useful when tail heaviness plays an important role in inference, so that choosing a copula with matching tail properties is essential.

Before using a fitted parsimonious model for further analysis, diagnostic checks should be done to ensure that the model is adequate. Bivariate extremal coefficients have been used for diagnostic checking of multivariate extreme value models. We investigate the use of an adequacy-of-fit statistic based on the difference between low-order empirical and model-based features (dependence measures), including the extremal coefficient, for this purpose. The difference is computed for each of the bivariate margins and a quadratic form statistic is obtained, with large values relative to a high quantile of the reference distribution suggesting model inadequacy. We develop methods to determine the appropriate cutoff values for various parsimonious models, dimensions, dependence measures and methods of model fitting that reflect practical situations. Data examples show that these diagnostic checks are handy complements to existing model selection criteria such as the AIC and BIC, and provide the user with some idea about the quality of the fitted models.
Preface

The thesis is written under the supervision of Prof. Harry Joe. Chapter 3 is based on a submitted manuscript coauthored with the supervisor, who initiated the idea on the extension of theory about parsimonious structures to multivariate extremes. With guidance from the supervisor, the author derived the main results, identified numerical challenges in the fitting of the proposed extreme value factor copula, conducted data analysis and drafted the initial manuscript. Prof. Joe suggested revisions to the manuscript and the author implemented these revisions to bring the chapter to its present form. Ideas on Gaussian quadrature underlying the derivations in Chapter 4 were suggested by the supervisor. The author conducted most of the derivations and identified further difficulties in certain situations. The numerical procedures were implemented by the author in Fortran 90 for speed improvement. Some of the results in Chapter 4 were also included in the submitted manuscript.

The review of estimators for the extremal coefficient in Chapter 5 was conducted by the author. The supervisor and the author jointly identified the potential extension of the F-madogram to general copulas. The author derived most of the properties of the estimator, while the supervisor suggested improvements to some of the more difficult proofs. This extension is based on a manuscript in preparation coauthored with the supervisor.

The supervisor advocated the use of the difference statistic as a measure of adequacy-of-fit in Chapters 6 and 7. The author derived most of the properties, including the separability result with empirical U-statistics and maximum likelihood estimation. The data analysis and simulation studies in Chapter 7 were conducted by the author, under helpful advice and insight from the supervisor. The author developed the efficient algorithm on the evaluation of bivariate empirical distribution functions at the observed data points, as described in Appendix E, and implemented it in Fortran 90. The algorithm is based on a submitted manuscript coauthored with the supervisor.

Throughout the preparation of the thesis, Prof. Joe made ample suggestions on the improvement of presentation, motivation and big picture viewpoints as well as some technical details. Many of the proposed changes were implemented to result in the current version that fits into a unified topic with improved connection within and between chapters.
# Table of Contents

Abstract ................................................................. ii

Preface ................................................................. iii

Table of Contents .................................................. iv

List of Tables ....................................................... ix

List of Figures ....................................................... xii

List of Symbols and Notations ..................................... xv

Acknowledgements .................................................. xix

Dedication .............................................................. xx

1 Introduction ......................................................... 1
  1.1 Multivariate extreme value copula models with parsimonious dependence . 2
  1.2 Bivariate extremal coefficient and extension to general copulas .......... 4
  1.3 Adequacy-of-fit diagnostic checks for parsimonious models ............... 5
  1.4 Research contributions and organization of thesis ........................ 7

2 Preliminaries ......................................................... 9
  2.1 Copula theory and examples ........................................ 9
    2.1.1 Copulas as multivariate distributions on the unit hypercube .......... 9
    2.1.2 Factor copulas ................................................ 12
    2.1.3 Vine copulas .................................................. 14
  2.2 Extreme value theory ............................................ 16
    2.2.1 Univariate extreme value theory .................................. 16
    2.2.2 Multivariate extreme value theory .................................. 17
### 4 Numerical integration methods for extreme value factor copula models
- 4.1 Burr 1-factor extreme value copula
- 4.2 Dagum 1-factor extreme value copula

### 5 Extremal coefficient for extreme value copulas and its generalizations
- 5.1 Empirical estimators of the extremal coefficient
  - 5.1.1 Estimators assuming known margins
  - 5.1.2 Rank-based estimators
  - 5.1.3 Asymptotic efficiency of the estimators
- 5.2 Generalization of the F-Madogram estimator to non-extreme-value copulas
  - 5.2.1 Dependence properties
  - 5.2.2 Interpretation and use for general copulas
  - 5.2.3 Boundary cases
  - 5.2.4 Asymptotic normality and variance
  - 5.2.5 Extension to higher dimensions
  - 5.2.6 Potential future research

### 6 Assessing model adequacy based on empirical and fitted features
- 6.1 Introduction
- 6.1.1 Vector of differences between empirical and model-based features
- 6.1.2 Relevance to goodness-of-fit tests
- 6.2 Asymptotics of the difference vector for a correctly specified model
  - 6.2.1 Behaviour of the difference between empirical and fitted distribution functions
  - 6.2.2 Generalization to the difference for U-statistics
  - 6.2.3 Separability of variance in the asymptotic normal distribution
- 6.3 Asymptotics of the difference vector under model misspecification
- 6.4 Some comments on properties of the difference statistic
- 6.5 Decision criteria based on the adequacy-of-fit statistic

### 7 Adequacy-of-fit for multivariate copulas with parsimonious dependence
- 7.1 Motivation and background
- 7.2 Diagnostic checks based on the adequacy-of-fit statistic
  - 7.2.1 Issues and general strategies
  - 7.2.2 Strategies for different parsimonious models
  - 7.2.3 Relationship between the adequacy-of-fit statistic and SRMSR
7.3 Results for the exchangeable Gaussian distribution and some copulas with exchangeable dependence ......................................................... 156
7.3.1 Theoretical results for the exchangeable Gaussian distribution .... 156
7.3.2 Patterns for some copulas with exchangeable dependence .......... 158
7.4 Method on evaluating $\Sigma$ for factor and truncated vine models under maximum likelihood estimation ......................................................... 161
7.4.1 Empirical covariance ................................................................. 162
7.4.2 Model-based covariance ............................................................... 163
7.5 Finding an approximate critical value using a matching Gaussian copula .. 166
7.5.1 Behaviour of different copulas with the same dependence structure . 167
7.5.2 Gaussian copula approximation for factor and vine structures ....... 171
7.5.3 Improving computational efficiency ............................................... 174
7.6 Multivariate extreme value copulas — an example for pairwise likelihood estimation .................................................................................. 178
7.6.1 Finding a matching t-EV copula ...................................................... 178
7.6.2 Simulation study ............................................................................. 180
7.6.3 Implementation for parsimonious extreme value copulas ............. 183
7.7 Adequacy-of-fit for tails of copulas .................................................... 184
7.7.1 Properties of the difference statistic for a single bivariate margin ... 185
7.7.2 SRMSR critical values for parsimonious dependence structures ... 187
7.8 Model estimation methods other than the maximum likelihood ........ 189
7.8.1 Estimation of marginal distributions ............................................. 189
7.8.2 Sequential estimation for vine structures ...................................... 193
7.9 Data examples .................................................................................... 194
7.9.1 European market returns data ...................................................... 194
7.9.2 Fraser River flows data .................................................................. 199
7.9.3 United States stock returns data .................................................. 201
7.10 Summary .......................................................................................... 201
8 Conclusion ............................................................................................ 205

Bibliography ........................................................................................... 208

Appendices

A Derivations of the asymptotic properties of the rank-based F-madogram measure of dependence ............................................................ 220
B Behaviour of the asymptotic variances of the empirical Kendall’s $\tau$ and Spearman’s $\rho$ ......................................................... 223

C Maximum likelihood estimator for Gaussian vine structures when variances are estimated ......................................................... 225

D Derivations of the properties of correlation parameters for the exchangeable Gaussian distribution .................................................. 230
   D.1 Variance of the difference for one bivariate margin .................. 230
   D.2 Covariance between differences for two bivariate margins ........ 234

E An $O(N \log_2 N)$ algorithm for evaluating the bivariate empirical distribution function at the $N$ observations ...................... 237
   E.1 The modified merge sort algorithm .................................... 239
   E.2 The modified quicksort algorithm ..................................... 243
   E.3 Some comments on the algorithms .................................... 247
List of Tables

2.1 Dependence properties of some commonly used copulas
2.2 Linking copulas for each tree of a C-vine rooted at variable 1

3.1 An example of the dependence structure between observed and latent variables in a bi-factor model
3.2 Matrices of pairwise Spearman’s $\rho$ for five copulas with parsimonious dependence
3.3 Pairwise Kendall’s $\tau$ and extremal coefficient $\vartheta$ for the mixed dependence scenario
3.4 Estimated Kendall’s $\tau$ and standard errors for the Dagum 1-factor extreme value copula simulation
3.5 Correlations of the normal scores for the Fraser River flows data
3.6 Fitting results for the Fraser River flows data using normal scores and Gaussian distribution
3.7 Loadings of the Gaussian 3-factor model for the normal scores of the Fraser River flows data
3.8 Fitting results for the Fraser River flows data using multivariate extreme value copulas
3.9 Loadings of the Hüsler-Reiss 1- and 2-factor models
3.10 Correlations of the normal scores for the US stocks data
3.11 Fitting results for the US stocks data using normal scores and Gaussian distribution
3.12 Loadings of the Gaussian 1- and 2-factor models for the normal scores of the US stocks data
3.13 Fitting results for the US stocks data using multivariate extreme value copulas

5.1 Asymptotic variances of the empirical extremal coefficient $\hat{\vartheta}$ for the rank-based estimators with the independence copula
5.2 Asymptotic variances of the empirical extremal coefficient $\hat{\vartheta}$ for various estimators with the Gumbel copula
5.3 Values of $\lambda_{U,\alpha}$ and $\lambda_{L,\alpha}$ for various bivariate copulas with Kendall’s $\tau$ equal to 0.5
5.4 Comparison of the difference in $\lambda_{L,\alpha}$ for six pairs of copulas against their asymptotic standard errors for the sample size 400 ........................................ 101
5.5 Asymptotic variances of $\hat{\lambda}_{U,\alpha}$ and $\hat{\lambda}_{L,\alpha}$ for various bivariate copulas with Kendall’s $\tau$ equal to 0.5 ........................................ 105
5.6 Asymptotic variances of $\hat{\lambda}_{U,\alpha,r}$ and $\hat{\lambda}_{L,\alpha,r}$ for various bivariate copulas with Kendall’s $\tau$ equal to 0.5 ........................................ 106

6.1 Summary of empirical and model-based asymptotic variances under maximum likelihood estimation ........................................ 126
6.2 Tail properties and asymptotic variances for different parametric copula families with true Kendall’s $\tau$ equal to 0.5 ........................................ 133
6.3 Asymptotic variance estimates for Kendall’s $\tau$ using different estimators for the copula parameter, with the parametric model having Pareto margins and MTCJ copula with Kendall’s $\tau$ equal to 0.5 ........................................ 136

7.1 Matrix of pairwise differences between the empirical and model-based extremal coefficient for the Burr 1-factor EV copula and 1-factor t-EV copula with $\nu = 3$ fitted to the US stock returns example ........................................ 141
7.2 Possible approaches of obtaining critical values of the adequacy-of-fit statistic for different parsimonious dependence structures, assuming models are estimated via maximum likelihood with features being Kendall’s $\tau$ or Spearman’s $\rho$ ........................................ 154
7.3 Summary of asymptotic variances and covariances of the differences for the $d$-dimensional exchangeable Gaussian distribution, using modified correlation as the feature ........................................ 158
7.4 Simulation scenarios for factor copulas to be used for different parametric linking copula families ........................................ 168
7.5 Simulation scenarios for truncated vine copulas to be used for different parametric linking copula families ........................................ 169
7.6 Summary of conservative SRMSR critical values (for sample size 100) for Gaussian copulas with factor and truncated vine structures ........................................ 179
7.7 Pairwise Kendall’s $\tau$ and extremal coefficient for the three simulation scenarios with $d = 5$ ........................................ 181
7.8 Average pairwise Kendall’s $\tau$ and extremal coefficient for the three simulation scenarios with $d = 10$ ........................................ 182
7.9 SRMSR critical values (for sample size 100) using Hüsler-Reiss copula and t-EV copulas with matching pairwise Kendall’s $\tau$ or extremal coefficient, under weak, moderate and strong dependence scenarios for $d = 5$ and 10 ........................................ 182
7.10 Summary of SRMSR critical values (for sample size 100) using tail-weighted dependence measures with factor and truncated vine structures 189
7.11 SRMSR critical values (for sample size 100) for some exchangeable, factor and truncated vine structures based on different estimation methods with Kendall’s $\tau$ as the feature 191
7.12 SRMSR critical values (for sample size 100) for the exchangeable structure with Frank and MTCJ copulas with Kendall’s $\tau$ value of 0.5, based on copula estimation and joint estimation of marginal and dependence parameters with Pareto margins 192
7.13 SRMSR critical values (for sample size 100) for the Hüsler-Reiss copula, using copula estimation, IFM with GEV marginal distributions and the marginal ranks methods 192
7.14 SRMSR critical values (for sample size 100) for some 2-truncated C-vines with the feature being Kendall’s $\tau$ 193
7.15 Fitted models considered for the European market GARCH-filtered index returns data 196
7.16 European market returns data: SRMSR statistics and critical values for the fitted models with the feature being Kendall’s $\tau$ 197
7.17 European market returns data: SRMSR statistics and critical values for the fitted models with the features being tail-weighted dependence measures 198
7.18 European market returns data: SRMSR critical values with Kendall’s $\tau$ and tail-weighted dependence measures for sample size $n = 484$ obtained from Tables 7.6 and 7.10 199
7.19 Fraser River flows data: SRMSR statistics and critical values for some fitted models with the features being Kendall’s $\tau$ and extremal coefficient 200
7.20 US stock returns data: SRMSR statistics and critical values for some fitted models with the features being Kendall’s $\tau$ and extremal coefficient 202
# List of Figures

2.1 Dependence between observed and latent variables for the 1-factor and 2-factor copula models ............................................. 13
2.2 Vine diagram for the first tree of a C-vine rooted at variable 1 .............. 15
2.3 Locations of the gauging stations in eastern Vancouver Island ............. 21
2.4 Q-Q plots for the marginal GEV fitting of the Vancouver Island streamflows data .................................................. 22
2.5 Scatterplot of normal scores for the Vancouver Island streamflows data .... 23
2.6 Min-stable plot with exponential margin and max-stable plot with Fréchet margin for the Vancouver Island streamflows data ..................... 24
2.7 Density contours and lower tail dependence functions of the MTCJ and reflected Gumbel copulas with Kendall’s $\tau$ equal to 0.5 ..................... 26

3.1 Scatterplots of the normal scores of the data simulated from the Dagum 1-factor extreme value copula, with dependence parameters $\delta = (1, 1, 1, 1), (4, 4, 4, 4)$ and $(1, 2, 3, 4)$ for the weak, strong and mixed dependence scenarios, respectively . 57
3.2 Sampling distributions of the fitted Kendall’s $\tau$ using full and pairwise likelihood for the strong dependence scenario, with $\delta = (4, 4, 4, 4)$ for the Dagum 1-factor extreme value copula .............................................. 59
3.3 Locations of the gauging stations along the Fraser River and the pairwise scatterplot of the normal scores .............................................. 61
3.4 Vine diagram for the fitted Hüsler-Reiss 1- and 2-truncated vine models using the vine suggested from Gaussian analysis ......................... 66
3.5 Vine diagram for the fitted Hüsler-Reiss 1- and 2-truncated vine models using a D-vine following the relative positions of the stations ..................... 66
3.6 Scatterplot of the normal scores for the US stocks data (negated minimum returns) ............................................. 68
3.7 Diagram for the fitted Hüsler-Reiss 1-truncated vine model using the vine suggested from Gaussian analysis .............................. 69
5.1 Asymptotic variances of the CFG and F-madogram estimators (assuming known margins) for the Gumbel and Hüsler-Reiss copulas ........................................ 92

6.1 Asymptotic variances of the empirical, model-based and difference estimators for different parametric copula families with various strengths of dependence .............................. 132

7.1 Plots of means and variances of the reference distribution $Q$ for various dimensions, and the critical values (95% quantiles) of $Q$ expressed in terms of SRMSR for sample size $n$, i.e., $\sqrt{(\text{CV of } Q)/n}$, for $n = 100$ ........................................ 159

7.2 Plots of SRMSR (sample size 100) critical values against the common strength of dependence between variables expressed in terms of Kendall’s $\tau$, for various parametric copula families and dimensions ........................................ 160

7.3 Histograms of asymptotic empirical covariance estimates using different methods, based on a 1-truncated D-vine with 8 variables and Gumbel linking copulas with Kendall’s $\tau = 0.5$ ........................................ 163

7.4 Histograms of asymptotic model-based covariance estimates, based on a 1-truncated D-vine with 8 variables and Gumbel linking copulas with Kendall’s $\tau = 0.5$ ........................................ 165

7.5 Ratio of SRMSR critical values (for sample size 100) for various 1- and 2- factor copulas, with the feature being Kendall’s $\tau$ ........................................ 170

7.6 Ratio of SRMSR critical values (for sample size 100) for various 1- and 2-truncated C-vine copulas, with the feature being Kendall’s $\tau$ ........................................ 171

7.7 SRMSR critical values (for sample size 100) for Gaussian 1- and 2-factor copulas, with the feature being Kendall’s $\tau$ ........................................ 172

7.8 SRMSR critical values (for sample size 100) for Gaussian 1- and 2-truncated C- and D-vine copulas, with the feature being Kendall’s $\tau$ ........................................ 173

7.9 SRMSR critical values (for sample size 100) for various Gaussian factor copulas using the restricted maximum likelihood method, with the feature being Kendall’s $\tau$ ........................................ 176

7.10 SRMSR critical values (for sample size 100) for various Gaussian C-vine copulas using the restricted maximum likelihood method, with the feature being Kendall’s $\tau$ ........................................ 176

7.11 Asymptotic variances of the empirical, model-based and difference estimators using tail-weighted dependence measures for different parametric copula families with various strengths of dependence, plotted against the tail-weighted dependence measure for the respective tail ........................................ 186
7.12 SRMSR critical values (for sample size 100) for some copulas having factor or truncated C- and D-vine structures with 5 or 8 variables, with the feature being the tail-weighted dependence measure ........................................... 188

7.13 Scatterplot of the normal scores for the European market GARCH-filtered index returns data ................................................................. 195

D.1 Plots of various asymptotic variances and covariances for different dimensions and true correlations of the positive exchangeable Gaussian distribution .... 236

E.1 Illustration of the merge sort with 8 elements ................................. 239
E.2 Illustration of the modified merge sort with 8 elements .................. 241
E.3 Illustration of the quick sort with 8 elements ................................. 244
E.4 Illustration of the modified quick sort with 8 elements .................. 247
List of Symbols and Notations

The following lists the most common uses of symbols and notations in the thesis. Those not listed have one or more uses that are not the main emphasis of the thesis.

\( A, a \): Stable tail dependence function (one form of the Pickands dependence function that is homogeneous of order 1 and appears in the exponent of a multivariate extreme value copula (the "exponent function")), or sometimes a generic function

\( a_S \): Stable tail dependence function for the subset of variables \( S \)

\( A^{(S)} \): Derivative of \( A \) with respect to the subset of indices \( S \)

\( \mathbb{B} \): Brownian bridge

\( B \): One form of the Pickands dependence function, related to \( A \) as

\[
B(w_1, \ldots, w_{d-1}) = A(w_1, \ldots, w_{d-1}, 1 - \sum_{j=1}^{d-1} w_j) \text{ for } w_j \in [0, 1] \text{ and } \sum_{j=1}^{d} w_j \leq 1.
\]

For most of the thesis, we refer to \( B \) as the Pickands dependence function

\( b \): Tail dependence function

\( b_S \): Marginal tail dependence function for the subset of variables \( S \)

\( b_{S_j|S_k} \): Conditional tail dependence function with subsets of indices \( S_j \) and \( S_k \)

\( C \): Copula (i.e., multivariate distribution function with Uniform(0, 1) margins)

\( C^+ / C^- / C^\perp \): Comonotonicity / (bivariate) Countermonotonicity / Independence copula

\( \hat{C} \): Reflected or survival copula of \( C \)

\( C \): Survival function of \( C \)

\( C_S \): Marginal copula of \( C \) for the subset of variables \( S \)

\( C_{S_j|S_k} \): Conditional distribution for copulas with subsets of indices \( S_j \) and \( S_k \)

\( C_{jk:S} \): Copula for the conditional distributions \( F_{j|S} \) and \( F_{k|S} \)

\( C_{jk:S} \): Conditional distribution of \( C_{jk:S} \)

\( C_n \): Empirical copula for sample size \( n \)

\( c / c_S / c_{jk:S} \): Copula density of \( C / C_S / C_{jk:S} \)

\( \text{Cor} \): Correlation between two random variables

\( \text{Cov} \): Covariance between two random variables or elements of a vector

\( D_n \): Difference statistic (empirical minus model-based feature) for sample size \( n \)

d: Dimension or number of variables

\(\mathbb{E}\): Expectation of a random variable

\(F, G\): Distribution function, not necessarily copula

\(F_{S_j|S_k}\): Conditional distribution with subsets of indices \(S_j\) and \(S_k\)

\(\hat{F}_n\): Empirical distribution function (\(\hat{F}_{jk}\) for bivariate version for margins \((j,k)\))

\(f, g\): Density function

\(\mathbb{G}\): Gaussian process

\(\mathcal{G}\): Survival function of \(G\)

\(\mathbf{H}\): Sensitivity matrix (component of the sandwich information matrix)

\(\mathbb{I}_d\): Identity matrix of dimension \(d\)

\(I_d\): Index set \(\{1, \ldots, d\}\)

\(I\): Fisher information matrix

i.i.d.: Independent and identically distributed

\(\mathbb{J}_d\): Matrix of 1’s of dimension \(d\)

\(\mathbf{J}\): Variability matrix (component of the sandwich information matrix)

\(\ell\): Log-likelihood function

\(l\): Inference function (for estimating equation); slowly varying function

\(N\): Sample size for Monte Carlo simulation in the evaluation of integrals; Gaussian distribution if followed by arguments

\(n\): Sample size (for data, parametric bootstrap, etc.)

\(O\): Asymptotic order, e.g., \(h_1(x) = O(h_2(x))\) as \(x \to \infty\) means that, for some constants \(M\) and \(x_0\), \(|h_1(x)/h_2(x)| \leq M\) for all \(x \geq x_0\); complexity of sorting algorithms

\(O_p\): Stochastic version of \(O\), e.g., \(X_n = O_p(Y_n)\) as \(n \to \infty\) for (possibly stochastic) sequences \(X_n, Y_n\) means that \(X_n/Y_n\) is bounded in probability

\(o\): Asymptotic dominance, e.g., \(h_1(x) = o(h_2(x))\) as \(x \to \infty\) means that \(\lim_{x \to \infty} h_1(x)/h_2(x) = 0\)

\(o_p\): Stochastic version of \(o\), e.g., \(X_n = o_p(Y_n)\) as \(n \to \infty\) for (possibly stochastic) sequences \(X_n, Y_n\) means that \(X_n/Y_n\) converges to zero in probability

\(\mathbb{P}\): Probability of a set or event

\(p\): Number of factors or truncation level for vines

\(Q_n\): Quadratic form statistic for sample size \(n\), with reference distribution \(Q\)

\(\mathbb{R}^d / \mathbb{R}^d_+:\) Reals / non-negative reals in \(d\) dimensions. Superscript omitted if \(d = 1\).

\(T_{d,\nu}\): Distribution function of the \(d\)-variate \(t\) distribution with \(\nu\) degrees of freedom

\(T(F)\): Functional mapping from \(F\) to \(\mathbb{R}\) with distribution \(F \in \mathcal{F}\) as argument

\(U\): Random variable of a copula

\(V\): Latent variable for factor copula

\(\text{Var}\): Variance of a random variable
\( X, Y, Z \): Random variable
\( \beta \): Blomqvist’s \( \beta \) or sometimes a parameter
\( \delta \): Dependence parameter of copula families
\( \theta \): Dependence parameter of copula families; parameter of a general statistical model
\( \theta_0 \): True value of a parameter
\( \hat{\theta}_n \): Estimator of \( \theta \) based on a sample of size \( n \)
\( \vartheta \): Extremal coefficient
\( \vartheta_\alpha \): Extremal coefficient for general copula with exponentiation index \( \alpha \) for the F-madogram
\( \kappa_L, \kappa_U \): Lower / Upper tail order
\( \lambda_L, \lambda_U \): Lower / Upper tail dependence index
\( \lambda_\alpha \): F-madogram measure of dependence with exponentiation index \( \alpha \), with estimator
   assuming known margins \( \hat{\lambda}_\alpha \) and rank-based \( \hat{\lambda}_{\alpha,r} \). A subscript \( L \) or \( U \) preceding \( \alpha \)
   indicates lower or upper tail-weighted, respectively.
\( \mu \): Quantity related to the mean of a variable; location parameter; positive measure
\( \nu \): Degree of freedom parameter for the \( t \) and related distributions
\( \nu_\alpha \): F-madogram of a bivariate copula with exponentiation index \( \alpha \), with estimator
   assuming known margins \( \hat{\nu}_\alpha \) and rank-based \( \hat{\nu}_{\alpha,r} \)
\( \rho \): Correlation parameter
\( \rho_{jk:S} \): Partial correlation of variables \( j, k \) given the set of variables \( S \)
\( \rho_N \): Correlation of normal scores
\( \rho_S \): Spearman’s \( \rho \) or rank correlation
\( \Sigma \): Correlation / covariance matrix
\( \sigma, \sigma^2 \): Quantity related to the standard deviation or variance of a variable; scale parameter
\( \tau \): Kendall’s \( \tau \)
\( \Phi_d \): Distribution function for the standard \( d \)-variate Gaussian distribution. Subscript
   omitted if \( d = 1 \).
\( \emptyset \): Empty set
\( 1\{S\} \): Indicator function for event \( S \)
\( \succ_c \): Concordance ordering (left argument more concordant than right; \( \prec_c \) for reverse order)
\( \vee \): Maximum operator
\( \wedge \): Minimum operator
\( \triangleq \): Left argument is defined as right argument (sometimes “is denoted as”)
\( \nabla, \nabla^2 \): Gradient / Hessian
\( \mathbf{S}^\top \): Transpose of the matrix \( \mathbf{S} \)
\( \sim \): Asymptotic equality, e.g., \( h_1(x) \sim h_2(x) \) as \( x \to \infty \) means that \( \lim_{x \to \infty} h_1(x)/h_2(x) = 1 \);
   left argument is distributed as right argument
\( \xrightarrow{p} \): Convergence in probability
\( \xrightarrow{d} \): Convergence in distribution
\( h' \): Derivative of \( h \)
\( |S| \): Cardinality of a finite set \( S \)
Acknowledgements

I would like to express my deepest gratitude to my supervisor, Prof. Harry Joe, for his guidance and continuous support throughout my PhD studies. His insightful comments, ideas and approach to problems allowed me to explore many areas in the world of copula modelling that would not have been possible otherwise. Thanks to his critical but constructive comments, I have identified some inadequacies and learned various skills that will be useful in my future career.

I am grateful to the members of the supervisory committee, Prof. Lang Wu and Dr. Natalia Nolde, for their suggestions and comments. Their input has been valuable in improving the presentation and flow of ideas in the thesis. I would also like to thank Dr. Johanna Nešlehová, who provided extensive feedback as the external examiner, and Prof. Jiahua Chen and Prof. Joel Friedman for serving as the university examiners.

My heartfelt appreciation extends to other faculty members and the administrative staff of the Department of Statistics. I am in particular grateful for their thoughtful assistance which made my PhD journey a much smoother one. The support from my fellow graduate students has also been very helpful; I truly feel that we are in a big family and they are always approachable whenever I need someone to talk to. I am also thankful to the financial support from UBC in the form of the Four Year Doctoral Fellowship.

Finally, I would like to thank my parents, relatives and friends for their constant encouragement. It is never easy to start a new chapter in life in another continent; their support and help have given me the motivation and courage to go forward.
Dedication

To my parents
Chapter 1

Introduction

The modelling of multivariate observations is important as variables often depend on each other, and such dependence cannot be accounted for if each variable is modelled separately. The multivariate normal or Gaussian distribution is arguably the most commonly used model for multivariate modelling. A fully parametrized \(d\)-variate Gaussian distribution has \(d\) mean parameters, \(d\) variance parameters and \(\binom{d}{2} = d(d-1)/2\) covariance parameters, for a total model complexity of \(O(d^2)\) parameters. Despite the mathematical appeal of the Gaussian distribution, it lacks certain characteristics that make it an inappropriate choice in some situations. One aspect is tail dependence; the Gaussian distribution with correlation parameter \(\rho \in (-1, 1)\) is tail independent (Sibuya (1960)), meaning that pairs of variables are independent in the limit when they are jointly large or small. Using the Gaussian distribution to model variables that exhibit tail dependence will therefore underestimate joint tail probabilities. Another issue with the Gaussian distribution is that it is centrally or reflection symmetric, that is, if \(Y \sim N(\mu, \Sigma)\), then \(Y - \mu\) has the same distribution as \(\mu - Y\); this is an unreasonable assumption for certain types of data, such as when it is more likely to observe joint large values than joint small values.

In light of the inadequacy of the Gaussian distribution, care should be taken when modelling the dependence structure among variables. Such procedure is aided by Sklar’s theorem (Sklar (1959)), which shows that the modelling of marginal and dependence characteristics can be separated. That is, if \((X_1, \ldots, X_d) \sim F\) with marginal distributions \(F_1, \ldots, F_d\), then there exists a distribution function \(C : [0, 1]^d \rightarrow [0, 1]\) such that

\[
F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)),
\]

\(C\) being unique if \(F\) is continuous. The distribution \(C\), defined on the unit hypercube, is known as the copula of \(F\) and is a multivariate distribution function with Uniform(0,1) margins. Note that the choice of \(C\) is independent of the choice of the marginal distributions.
Multivariate copula families with properties different from the Gaussian, such as non-zero tail dependence and asymmetric behaviour, have been developed (see, e.g., Joe (1997, 2014) and Nelsen (2006)).

1.1 Multivariate extreme value copula models with parsimonious dependence

A general $d$-dimensional copula contains $O(d^2)$ pairwise bivariate margins and hence the same number of descriptions of pairwise dependence properties. Typically this means the same number of dependence parameters to be modelled as well, just like the case of a $d$-dimensional Gaussian distribution. The property that the number of parameters grows quadratically in the dimension may not be desirable for several reasons:

1. A model with too many parameters may be difficult to interpret. Although pairwise dependence summaries are useful, often the researcher is interested in the structure of the data as well. A fully parametrized, saturated model offers little insight into the dependence structure.

2. There is the risk of overfitting if the number of parameters is large compared to the sample size. When this happens, the model may pick up noise that is not part of the structure. This can affect the quality of the fit and the model’s ability to predict future behaviour of the process.

3. When the objective function depends on many parameters to be optimized over, model estimation can become computationally inefficient or even impossible. For example, in Newton’s method it is necessary to obtain the inverse of the Hessian matrix at every iteration, and the time complexity of this operation depends directly on the number of parameters.

Because of these reasons, it makes sense to consider parsimonious dependence models which try to explain patterns with a formulation as simple as possible. In multivariate statistical modelling, this means that a dependence structure is applied to an otherwise unconstrained model, in an attempt to reduce the number of parameters (model complexity) but yet allow both statistical and scientific interpretations to be made. Examples of parsimonious multivariate models include:

1. Exchangeable structure, which assumes the same dependence pattern for every pair of variables, so that the distribution is invariant to the permutation of variable indices. In the Gaussian case, the correlation matrix has identical non-diagonal elements. For
copulas, the class of Archimedean copulas (see, e.g., Chapter 4 of Nelsen (2006)) is commonly used to model the exchangeable dependence structure with a fixed number of parameters for any dimension $d$. Due to the strong assumption of exchangeability, this structure alone is typically not very useful apart from small values of $d$. For higher dimensions, the exchangeable structure may be used for a smaller subset of variables within a cluster (see, e.g., Brechmann (2014)).

2. Factor structure, which assumes that the dependence among variables is driven by one or more latent factors (see, e.g., Chapter 9 of Johnson and Wichern (2007)). With the factor structure, variables are conditionally independent given the latent factor(s). The classical Gaussian $p$-factor model decomposes a $d \times d$ correlation matrix $\Sigma_G$ as follows:

$$\Sigma_G = LL^T + \Psi,$$

where $L$ is a $d \times p$ matrix of loadings and $\Psi$ a diagonal matrix of uniquenesses. Factor copulas (Krupskii and Joe (2013)) are the copula analogue of the Gaussian factor model, where conditional independencies between the observed variables given the latent factors are preserved, while the dependence between latent and observed variables is specified by parametric bivariate copula families. The Gaussian factor model is a special case, in which all linking copulas are Gaussian with suitable correlations.

3. Markov tree and truncated vine structures (see Bedford and Cooke (2001, 2002); Brechmann et al. (2012)), in which the dependence between variables is propagated through bivariate linkages. In a Markov tree, only the dependence properties of neighbouring variables are specified. This results in variables being Markovian, in the sense that a variable is conditionally independent with another given those along the linkage. The restriction of conditional independence is relaxed in the truncated vine structure. A $p$-truncated vine has dependence properties up to order $p$ (i.e., conditioning on at most $p - 1$ variables) specified, beyond which conditional independence is assumed. For the $d$-variate Gaussian distribution, creating a Markov tree amounts to specifying the correlations of $d - 1$ acyclic pairs (i.e., no three pairs form a closed loop among three variables), while the dependence for other pairs is driven by assuming zero higher-order partial correlations. A $p$-truncated vine is constructed by allowing partial correlations of at most order $p - 1$ to be non-zero. Vine copulas or pair-copula constructions (Aas et al. (2009); Kurowicka and Joe (2011)) are generalizations of such dependence structures beyond Gaussianity; a $p$-truncated vine copula specifies parametric bivariate copula families that are applied to univariate marginal or conditional distributions of up to order (tree) $p$. 
These models are parsimonious because in each case the number of parameters has order smaller than $O(d^2)$ and the distribution is structured to allow specific interpretation. In particular, the exchangeable model has $O(1)$ parameters, while the factor and truncated vine structures have $O(d)$ parameters. Apart from continuous distributions, these modelling techniques have previously been applied to discrete and ordinal data (Panagiotelis et al. (2012, 2017); Nikoloulopoulos and Joe (2015); Stöber et al. (2015)). We extend the concept of parsimonious dependence structure to multivariate extremes. The modelling of extremes is of interest in many areas, such as finance, meteorology, hydrology and engineering, where the occurrence of extreme events can lead to catastrophic results (see, e.g., Embrechts et al. (1997); Coles (2001); Castillo et al. (2005)). Very often, joint modelling of extremes is needed when one wants to make valid inferences about the extremal behaviour of dependent random variables. We propose factor and truncated vine analogues for extreme value copula models; they are suitable for the modelling of joint extremal observations that are believed to possess such structures, for example extreme stock returns for the same sector in which a factor model is plausible, or hydrological observations along a river where stations have natural (spatial) ordering and a vine structure is a reasonable assumption. The distribution of joint extremes has the max-stability property (de Haan and Resnick (1977)), meaning that the family of distributions is closed under componentwise extrema; this is not trivially satisfied by a factor or truncated vine copula with arbitrary or even extreme value linking copulas. Direct generalization appears difficult, and therefore we take the route of obtaining the extreme value limit in the case of factor copulas, and utilizing flexible multivariate extreme value copulas (which satisfy the max-stability property) where a factor or truncated vine structure can be applied.

1.2 Bivariate extremal coefficient and extension to general copulas

To explore the relationship among variables, it is useful to consider measures of dependence strength which quantify the degree of dependence. These are typically bivariate or pairwise in nature, although we note that higher-dimensional generalizations are possible. The extremal coefficient $\vartheta$ is such a measure first proposed to describe the strength of dependence of a multivariate extreme value distribution (Sibuya (1960); Pickands (1981); Smith (1990)). For a random vector $(X_1, \ldots, X_d) \sim F$ with common marginal distribution $F_m$, the extremal coefficient measures the number of “effective” independent variables and is defined implicitly as follows:

\[ F(x, \ldots, x) = \mathbb{P}(X_1 \leq x, \ldots, X_d \leq x) = \left[ \mathbb{P}(X_1 \leq x) \right]^d = F_m^\vartheta(x). \quad (1.1) \]
For copula \( C(u, \ldots, u) \), the final term in (1.1) reduces to \( u^\vartheta \). For distributions with non-negative dependence, \( \vartheta \in [1, d] \), where the boundaries are reached at the comonotonicity \( (\vartheta = 1) \) and independence limits \( (\vartheta = d) \). Empirical estimators (i.e., based on a sample) have been proposed for the estimation of \( \vartheta \) for bivariate extreme value distributions (see, e.g., Pickands (1981); Deheuvels (1991); Capéraà et al. (1997); Hall and Tajvidi (2000); Cooley et al. (2006); Bücher et al. (2011)). We examine these estimators and propose a generalization for arbitrary copulas. This generalization allows one to put more weight at the tail portion of a copula and estimate its tail dependence properties empirically.

1.3 Adequacy-of-fit diagnostic checks for parsimonious models

When multiple models are fitted to a data set, likelihood-based model selection criteria such as the Akaike and Bayesian information criteria (Akaike (1974); Schwarz (1978)) are commonly used. However, these criteria do not measure the quality of the individual model, and it may happen that even the best model among those considered has a poor fit. To ensure that a parsimonious model is adequate, it is intuitive to compare the closeness between empirical and model-based features. A feature is mathematically a functional applied to a distribution, and provides a summary of certain characteristics of the distribution. Bivariate dependence measures such as the extremal coefficient for \( d = 2 \) have been used as diagnostics for multivariate and spatial extremes (see, e.g., Smith (1990); an application to the modelling of extreme snow depth is given in Blanchet and Davison (2011)). For general copula models, dependence measures such as Kendall’s \( \tau \) and Spearman’s \( \rho \) may be better features as they are widely used and are mathematically easier to handle. To assess the quality of tail fits, the tail-weighted dependence measures (Krupskii and Joe (2015)) may be an appropriate choice. For each of the \( \binom{d}{2} \) bivariate marginal distributions of a \( d \)-dimensional random vector, we compute the difference (or residual) between the empirical feature and the feature obtained from the assumed model based on the estimated parameters. An adequate model should be such that these differences are small in magnitude for all or most of the bivariate margins.

This type of statistics based on residuals has been investigated previously in the discrete case. A notable example of such statistics is Pearson’s \( \chi^2 \), in which a quadratic form is constructed based on the vector of residuals in each cell of a contingency table. The test based on Pearson’s \( \chi^2 \) statistic is known to be unreliable when some cells have small expected counts, or equivalently when the ratio between the sample size and number of cells is small (Cochran (1952)). When this happens, the asymptotic (or reference) distribution
can differ significantly from the distribution of the finite-sample statistic, leading to inaccurate type I error rates in conducting hypothesis tests. This is especially a problem for higher-dimensional contingency tables; in this situation, Maydeu-Olivares and Joe (2005, 2006) propose limited-information statistics that use residuals based on low-order marginal tables. This effectively reduces the number of cells with small proportions, bypassing sparsity issues in higher dimensions. This idea is further generalized to linear combinations of cell residuals in Joe and Maydeu-Olivares (2010).

The method of limited-information statistics is applicable to $\sqrt{n}$-consistent and asymptotically normal estimators, including the maximum likelihood estimator. There exist analytic expressions for the calculation of the limiting variance of the limited-information statistic, although they are not always easy to use such as when the number of low-order margins is large or when the marginal expected counts are intractable. This problem is much harder for continuous variables. In addition to the restrictions applicable to the discrete setting, here the behaviour of the limiting distribution depends on other factors such as the feature chosen as well as the method of fitting marginal and dependence parameters. Taking these challenges into consideration, we develop criteria for assessing the adequacy of a model. To facilitate interpretation, we make use of the standardized root-mean-square residuals (SRMSR) (see, e.g., Hu and Bentler (1998, 1999)) which have the same scale as the feature being considered. The most sensible approach for a particular problem depends on the feasibility of model simulation, estimation and evaluation of the features for the estimated model. As we illustrate, the simplest solution of performing a parametric bootstrap to estimate the sampling distribution of the residuals is usually not possible in high dimensions. We bypass this problem by evaluating the elements of the asymptotic covariance matrix without first estimating the model-based features if possible, or make approximate inference using surrogate models that have similar characteristics to the target model.

Our work on using residuals to determine model fit has connections to the Kolmogorov-Smirnov, Cramér-von Mises and Anderson-Darling-type statistics with estimated parameters (e.g., Chernoff and Lehmann (1954); Darling (1955) and an overview in Chapter 28 of DasGupta (2008)). Based on these statistics, goodness-of-fit tests for bivariate copulas have been developed (Genest et al. (2009); Berg (2009)). Our objective is on assessing adequacy-of-fit of parametric models, and is different from what these tests aim to offer. Goodness-of-fit procedures involve statistical tests that try to answer the question of whether the observed data come from the assumed class of models, and can be useful when this is exactly the purpose of investigation. However, there is generally no stochastic, physical or subject matter-based theoretical basis for a class of multivariate models or copulas and hence we cannot hypothesize a “correct” (null) model before data analysis. Rather
we are interested in parsimonious models that reasonably capture certain features of the
data that are the most relevant for inferences. To this end, we do not formulate statistical
hypotheses and instead use the adequacy-of-fit statistic as a diagnostic check.

Note that the Kolmogorov-Smirnov, Cramér-von Mises and Anderson-Darling-type statistics combine the information of the residuals (i.e., taking the supremum of or integrating over some functions of these residuals) in their constructions; the information about individual residuals is often lost and neglected in subsequent analysis. On the other hand, our proposed adequacy-of-fit statistic retains the residual for each bivariate margin. This has the advantage that, in the case a model is found to be inadequate, one can look for the source of misfit and make improvements accordingly. It is therefore easier to eliminate structural inadequacy using these feature-based statistics.

1.4 Research contributions and organization of thesis

To summarize, our main research contributions include the following:

- Extension of the concept of parsimonious dependence structure to the modelling of multivariate extremes (Chapters 3 and 4). These dependence structures have been previously investigated for discrete (including binary or count data), non-extreme continuous or a mixture of discrete/continuous data.

- Adaptation of the F-madogram empirical estimator of the extremal coefficient to measure the strength of tail dependence of non-extreme-value copulas (Chapter 5).

- Diagnostics for parametric models to guard against inadequacy of fit, using the vector of residuals or differences between empirical and model-based features. In certain situations, the limiting distribution of such differences has a covariance matrix that can be separated into the empirical and model-based components, allowing simpler calculations (Chapter 6).

- Assessment of adequacy-of-fit for higher-dimensional models with parsimonious dependence structure, using differences between empirical and model-based features in bivariate marginal distributions (Chapter 7).

- An efficient (in the sense of having $O(N \log_2 N)$ complexity) algorithm for the evaluation of bivariate empirical distributions at the observed values. This is given in Appendix E as an algorithmic supplement to one of the methods we use in the assessment of model adequacy-of-fit.
The rest of the thesis is structured as follows. Chapter 2 provides a review of existing methods, and also serves partly as literature review in the relevant areas. The construction of extreme value copulas with parsimonious dependence structures is dealt with in Chapter 3. We investigate two approaches to constructing extreme value copulas with factor or truncated vine structures: (a) through the extreme value limit of existing copula models, and (b) through structuring the underlying correlation matrix of flexible extreme value copulas. We illustrate through examples that these models offer intuitive interpretations. Chapter 4 includes the technical derivations pertaining to the numerical integration procedures in obtaining pairwise densities of the 1-factor extreme value copula. Chapter 5 focuses on the bivariate empirical extremal coefficient, originally designed for extreme value copulas. The bivariate extremal coefficient has been used for diagnostic purposes in modelling multivariate spatial extremes. We review some common empirical estimators and discuss the potential extension of one of these empirical estimators to the non-extreme context, and provide connections with the tail-weighted dependence measures of Krupskii and Joe (2015). Chapters 6 and 7 are devoted to the assessment of model adequacy using bivariate Kendall’s $\tau$, Spearman’s $\rho$, tail-weighted dependence measures and extremal coefficient (for extreme value copulas) for copula models with parsimonious dependence. The asymptotic properties of the vector of residuals, defined as differences between empirical and model-based features, are derived in Chapter 6. Based on residuals, this class of statistics is very general and can be applied in a variety of situations, such as a combination of different features for univariate or multivariate distributions. Chapter 7 focuses on the particular application of such statistics to assess the adequacy of a multivariate copula using residuals from bivariate marginal distributions. We demonstrate that the critical value of the adequacy-of-fit statistic depends quite heavily on the overall dependence strength of the data, with stronger dependence typically leading to smaller critical values. Various ways of determining critical values are proposed for different parsimonious dependence structures, features, dimensions and methods of model estimation. Chapter 8 contains a conclusion of the thesis and potential topics for further research. Some of the longer proofs and derivations are given in the Appendix.
Chapter 2

Preliminaries

In this chapter, we review existing results in the literature that form the basis we build upon. Section 2.1 provides an overview of copula theory and a list of some properties and parametric families of bivariate copulas. Building multivariate models with factor and vine copulas is also addressed. Section 2.2 has results in univariate and multivariate extreme value theory. Multivariate extremes are closely related to the concept of tail dependence; Section 2.3 has details on tail dependence functions and extreme value limit of copula models. We focus on model adequacy in the latter part of the thesis; to facilitate such discussion with the use of dependence measures, an overview of these measures is provided in Section 2.4. Finally, in Section 2.5 we mention several model estimation methods for multivariate copulas. Because of model complexity, maximum likelihood is not always possible or computationally efficient. Methods such as composite likelihood, inference functions for margins and marginal ranks thus emerge as viable alternatives.

2.1 Copula theory and examples

We begin by introducing the concept of copulas and several examples of common bivariate copulas. An overview of factor and vine copulas is then given.

2.1.1 Copulas as multivariate distributions on the unit hypercube

Copulas are multivariate distributions with Uniform(0,1) margins (hereafter referred to as the unit uniform distribution). For any multivariate distribution with arbitrary marginal distributions, Sklar (1959) shows that there exists a corresponding copula. Let \((X_1, \ldots, X_d) \sim F\) with marginal distributions \(F_1, \ldots, F_d\). Then there exists a copula \(C\) such that

\[
F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)).
\]
Furthermore, $C$ is unique if $F$ is continuous. The copula $C$ has unit uniform margins because of the probability integral transform in its arguments. This representation implies that the modelling of marginal and dependence components can be separated, and the choice of the dependence structure is independent of the marginal distributions.

Many parametric copula families with different tail properties exist in the literature (see, e.g., Chapter 4 of Joe (2014)). We list some commonly used parametric bivariate copula families below; they are used in various places of this thesis. The support of all these copulas is on $[0, 1]^2$.

- **Gaussian (or normal) copula:**
  \[
  C(u, v; \rho) = \Phi_2\left(\Phi^{-1}(u), \Phi^{-1}(v); \rho\right), \quad \rho \in [-1, 1],
  \]
  where $\Phi$ and $\Phi_2$ are the univariate and bivariate standard Gaussian distribution functions, respectively.

- **Student t copula:**
  \[
  C(u, v; \rho, \nu) = T_{2, \nu}\left(T^{-1}_{1, \nu}(u), T^{-1}_{1, \nu}(v); \rho\right), \quad \nu \in (0, \infty); \rho \in [-1, 1],
  \]
  where $T_{d, \nu}$ is the distribution function of the $d$-dimensional Student t distribution with $\nu$ degrees of freedom.

- **Frank copula (Frank (1979)):**
  \[
  C(u, v; \delta) = -\frac{1}{\delta} \log \left[\frac{1 - e^{-\delta} - (1 - e^{-\delta u})(1 - e^{-\delta v})}{1 - e^{-\delta}}\right], \quad \delta \in (-\infty, \infty).
  \]

- **Bivariate Mardia-Takahashi-Clayton-Cook-Johnson (MTCJ) copula (Mardia (1962); Takahasi (1965); Clayton (1978); Cook and Johnson (1981)):**
  \[
  C(u, v; \delta) = (u^{-\delta} + v^{-\delta} - 1)^{-1/\delta}, \quad \delta \in [0, \infty).
  \]

- **Gumbel copula (Gumbel (1960)):**
  \[
  C(u, v; \delta) = \exp\left\{-\left[(- \log u)^\delta + (- \log v)^\delta\right]^{1/\delta}\right\}, \quad \delta \in [1, \infty).
  \]
  It is an extreme value copula, to be defined in Section 2.2.

- **Hüsler-Reiss copula (Hüsler and Reiss (1989)):**
  \[
  C(u, v; \delta) = \exp\left\{-w\Phi\left(\frac{1}{\delta} + \frac{\delta}{2} \log \left(\frac{w}{x}\right)\right) - x\Phi\left(\frac{1}{\delta} + \frac{\delta}{2} \log \left(\frac{x}{w}\right)\right)\right\}, \quad \delta \in [0, \infty),
  \]
  where $w = -\log u$ and $x = -\log v$. It is also an extreme value copula.
• t-EV copula (Demarta and McNeil (2005)):

\[
C(u, v; \rho, \nu) = \exp \left\{ -wT_{1,\nu + 1} \left[ \frac{\nu + 1}{\sqrt{1 - \rho^2}} \left[ \left( \frac{w}{x} \right)^{1/\nu} - \rho \right] \right] \right\}
\]

where \( w = -\log u \) and \( x = -\log v \). It is also an extreme value copula.

• BB1 copula (Joe and Hu (1996)):

\[
C(u, v; \theta, \delta) = \left( 1 + \left[ (\theta - 1)^{\delta} + (\theta - 1)^{-\delta} \right]^{1/\theta} \right)^{-1/\theta}, \quad \theta \in (0, \infty); \delta \in [1, \infty).
\]

This is a two-parameter family with potentially different levels of upper and lower tail dependence, to be addressed later in this subsection.

• Independence copula, \( C^\perp \):

\[
C^\perp = C^\perp(u, v) \triangleq uv.
\]

As the name suggests, the copula corresponding to any bivariate distribution with independent variables is the independence copula.

• Comonotonicity copula, \( C^+ \):

\[
C^+ = C^+(u, v) \triangleq \min(u, v).
\]

The two variables of a comonotonicity copula are perfectly positively dependent, i.e., if \((U, V) \sim C^+\), then \( U = V \) almost surely.

• Countermonotonicity copula, \( C^- \):

\[
C^- = C^-(u, v) \triangleq \max(0, u + v - 1).
\]

The two variables of a countermonotonicity copula are perfectly negatively dependent, i.e., if \((U, V) \sim C^-\), then \( U = 1 - V \) almost surely.

Copulas can be differentiated based on their dependence properties. Some of them include\(^1\):

• **Symmetry.** There are various types of symmetry properties. A copula has permutation symmetry (or exchangeability) if \( C(u, v) = C(v, u) \) for all \((u, v) \in [0, 1]^2\), and reflection symmetry if \( \hat{C}(u, v) \triangleq \bar{C}(1 - u, 1 - v) = u + v - 1 + C(1 - u, 1 - v) \) is the same copula as \( C(u, v) \). Here \( \bar{C} \) is the survival function of \( C \) and \( \hat{C} \) is known as the

\(^1\)We focus on bivariate copulas in the following discussion of dependence properties.
reflected or survival copula of \( C \). If \((U, V) \sim C\), then \( \widehat{C} \) is the distribution function for the pair \((1 - U, 1 - V)\). Graphically, a copula is permutation symmetric if the bivariate density is symmetric along the \((0, 0) - (1, 1)\) diagonal.

- **Tail dependence.** Roughly speaking, a copula has lower (resp. upper) tail dependence if the probability of one variable being very small (resp. large), given the other one is very small (resp. large), is non-zero. The extreme value limit of a copula without tail dependence is the independence copula. Section 2.3 has a detailed overview of tail dependence and extreme value limits.

- **Quadrant dependence.** A copula has positive quadrant dependence (PQD) if \( C(u, v) \geq uv \) for all \((u, v) \in [0, 1]^2\). That is, a copula is PQD if variables are more likely to be jointly large or small compared to the independence copula. Reversing the inequality gives negative quadrant dependence (NQD).

- **Stochastically increasing positive dependence.** For \((U, V) \sim C\), \( V \) is said to be stochastically increasing (SI) in \( U \) if the conditional distribution \( C_{V|U}(v|u) \) is decreasing in \( u \) for all \( v \). That is, as \( U \) increases, the conditional probability of \( V \leq v \) drops for fixed \( v \) (\( V \) is likely to be larger). Note that SI positive dependence implies PQD. Reversing the inequality results in stochastically decreasing random variables.

- **Concordance ordering.** A copula family \( C(u, v; \theta) \) is increasing (resp. decreasing) in concordance ordering if \( \theta_2 \geq \theta_1 \) implies \( C(u, v; \theta_2) \geq C(u, v; \theta_1) \) (resp. \( C(u, v; \theta_2) \leq C(u, v; \theta_1) \)) for all \((u, v) \in [0, 1]^2\). A copula that is increasing in concordance ordering with respect to \( \theta \) is more likely to yield joint small or large values for larger \( \theta \).

The properties of the copulas mentioned above are summarized in Table 2.1.

### 2.1.2 Factor copulas

The factor copula model is a generalization of the classical factor analysis to non-Gaussian dependence structures. It assumes the observed variables \( U_1, \ldots, U_d \) depend on one or more latent variables \( V_1, \ldots, V_p \). Figure 2.1 presents a schematic diagram of the 1- and 2-factor copulas. With a 1-factor copula, the observed variables are assumed to depend on one common latent variable; whereas in a 2-factor copula dependence arises from two latent variables which are assumed to be independent of each other.

Each edge in Figure 2.1 represents one bivariate linking copula with the distribution indicated. For example, \( V_1U_1 \) indicates that the bivariate copula links the distributions \( F_{V_1} \) and \( F_{U_1} \), while \( V_2U_2; V_1 \) indicates that the copula links the conditional distributions
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Student t</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Frank</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>MTCJ</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Gumbel</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Hüsler-Reiss</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>t-EV</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>BB1</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Table 2.1: Dependence properties of some commonly used copulas. Unless specified, properties correspond to the copulas with non-boundary parameter values.

Figure 2.1: Dependence between observed and latent variables for the 1-factor (left) and 2-factor (right) copula models. The observed variables are denoted by $U_1, \ldots, U_d$ while the latent variables are denoted by $V_1$ and $V_2$.

$F_{V_2|V_1}$ and $F_{U_2|V_1}$. As in the usual case for copulas, we assume that $U_1, \ldots, U_d$ have been transformed to have unit uniform margins, and the latent variables $V_i$ are independent unit uniform variables. To specify a 1-factor copula model, one needs $d$ bivariate copulas to link the observed variables to the latent factor, i.e., $\{C_{U_i,V_1}: i = 1, \ldots, d\}$. The joint distribution function of the observed variables, $C(u_1, \ldots, u_d)$, can be found by integrating
out the latent variable:

\[ C(u_1, \ldots, u_d) = \int_0^1 C_{U_1, \ldots, U_d|V_1}(u_1, \ldots, u_d|v_1) f_{V_1}(v_1) \, dv_1 = \int_0^1 \prod_{i=1}^d C_{U_i|V_1}(u_i|v_1) \, dv_1, \quad (2.2) \]

due to the assumed independence among the conditional distributions, where \( C_{U_i|V_1} \) is the conditional distribution of \( U_i \) given \( V_1 \) and \( f_{V_1}(v_1) = 1 \) is the density function of \( V_1 \) when \( v_1 \in [0, 1] \). The copula density is

\[ c(u_1, \ldots, u_d) = \frac{\partial^d C(u_1, \ldots, u_d)}{\partial u_1 \cdots \partial u_d} = \int_0^1 \prod_{i=1}^d c_{U_i,V_1}(u_i, v_1) \, dv_1 \]

as \( C_{U_i|V_1}(u_i|v_1) = \partial C_{U_i,V_1}(u_i, v_1)/\partial v_1 \) for \( i = 1, \ldots, d \) and further differentiation with respect to \( u_i \) gives the copula density for \( U_i \) and \( V_1 \).

For the 2-factor copula model, in addition to the \( d \) copulas \( \{C_{U_i,V_1} : i = 1, \ldots, d\} \), another \( d \) are used to link the observed variables to the second latent variable given the first one. That is, a copula in this second level links the conditional distributions \( C_{U_i|V_1} \) and \( C_{V_2|V_1} \) (which is unit uniform), denoted by \( C_{U_i,V_2;V_1} \). The set of copulas to be specified is therefore \( \{C_{U_i,V_1}, C_{U_i,V_2;V_1} : i = 1, \ldots, d\} \). In this case, the joint distribution function of the observed variables is

\[
C(u_1, \ldots, u_d) = \int_0^1 \int_0^1 C_{U_1, \ldots, U_d|V_1,V_2}(u_1, \ldots, u_d|v_1, v_2) f_{V_1,V_2}(v_1, v_2) \, dv_1 dv_2
= \int_0^1 \int_0^1 \prod_{i=1}^d C_{U_i|V_2,V_1}(C_{U_i|V_1}(u_i|v_1)|v_2) \, dv_1 dv_2, \quad (2.3)
\]

where \( f_{V_1,V_2}(v_1, v_2) = 1 \) is the joint density of \( V_1 \) and \( V_2 \) and \( C_{U_i|V_2,V_1} \) is the conditional distribution of \( U_i|V_1 \) given \( V_2|V_1 \). The derivation of (2.3) is presented in Krupskii and Joe (2013). The copula density function, necessary for inference, is given by

\[ c(u_1, \ldots, u_d) = \int_0^1 \int_0^1 \prod_{i=1}^d \left[ c_{U_i,V_2;V_1}(C_{U_i|V_1}(u_i|v_1), v_2) \cdot c_{U_i,V_1}(u_i, v_1) \right] \, dv_1 dv_2. \]

Numerical integration is usually needed for the evaluation of the integral. The factor copula model includes the Gaussian factor model as a special case; it can be constructed by taking the linking copulas as suitable Gaussian copulas.

### 2.1.3 Vine copulas

The pair-copula construction (or vine copula) approach (Bedford and Cooke (2001, 2002); Aas et al. (2009)) allows one to build multivariate copulas hierarchically using only bivariate
linking copulas in each hierarchy or tree, directly linking variables together. In the first
tree of a $d$-dimensional vine copula, the $d$ variables are linked together through the use
of $d - 1$ bivariate copulas. In each subsequent tree, the copulas used link the conditional
distributions implied by the copulas in the previous tree. Therefore a full regular vine on $d$
variables has $d - 1$ trees and is completely specified by $\binom{d}{2}$ bivariate copulas. An example
is that of a C-vine (canonical vine); Table 2.2 lists the linking copulas in each tree of such a
vine rooted at the first variable, where each entry in the table corresponds to one bivariate
copula linking the variables indicated. Variables after the semicolon are those conditioned
upon, e.g., 34:12 means that the copula models the dependence between the conditional
distributions $F_{3|12}$ and $F_{4|12}$. Meanwhile, Figure 2.2 displays the vine diagram for the
first tree of the same vine copula. By choosing the location of the connecting edges and
parametric copula families appropriately, one may introduce different dependence structures
among the variables.

<table>
<thead>
<tr>
<th>Tree</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>$\ldots$</th>
<th>$d - 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>12</td>
<td>23;1</td>
<td>34;12</td>
<td>$(d - 1,d);(123,\ldots,d - 2)$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>24;1</td>
<td>35;12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linking variables</td>
<td>14</td>
<td>25;1</td>
<td>$\vdots$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>$\vdots$</td>
<td>3$d$;12</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1$d$</td>
<td>$\vdots$</td>
<td>2$d$;1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.2: Linking copulas for each tree of a C-vine rooted at variable 1. For brevity, some
commas between variables are omitted. Variables after the semicolon are those conditioned
upon.

Figure 2.2: Vine diagram for the first tree of a C-vine rooted at variable 1

The copula density of a $d$-dimensional regular vine is given by the product of all con-
stituent bivariate copula densities, i.e.,
\[ c(u_1, \ldots, u_d) = \prod_{[i_1, i_2] \in \mathcal{E}(\mathcal{V})} c_{i_1, i_2; S(i_1, i_2)}(u_{i_1 | u_{S(i_1, i_2)}}, C_{i_2 | S(i_1, i_2)}(u_{i_2 | u_{S(i_1, i_2)}})), \]

where \([i_1, i_2] \in \mathcal{E}(\mathcal{V})\) is an edge and \(\mathcal{E}(\mathcal{V})\) is the set of all edges of the vine \(\mathcal{V}\) (see equation (3.41) of Joe (2014)). The distribution function is in general a \(d\)-dimensional integral, obtained by integrating (2.4) with respect to \(u_1, \ldots, u_d\). The existence of an explicit density function allows statistical inference using the likelihood function, but quantities involving the distribution function may be difficult to obtain and numerical integration is often required.

The dependence structure is sometimes well approximated by the first few trees. If high-order residual dependence can be ignored, a parsimonious vine model can be obtained by truncation. A \(p\)-truncated vine with \(p < d - 1\) is one with all linking copulas beyond the \(p\)th tree being the independence copula. A truncated vine has fewer parameters and model estimation is less costly. The special case of a 1-truncated vine is also known as a Markov tree; it is so named due to its conditional independence property given variables along the path. For example, the vine in Figure 2.2 is a Markov tree if the shown linkages are the only ones with non-independence copulas. In this case, variable \(i\) is independent of variable \(j\) given variable 1, for \(2 \leq i < j \leq d\). Note that the \(p\)-factor copula model is an implementation of a \(p\)-truncated C-vine rooted at the latent variables (Krupskii and Joe (2013)); all linking copulas beyond tree \(p\) are assumed to be independence copulas.

### 2.2 Extreme value theory

This section provides a brief overview on univariate and multivariate extreme value theory. An example is then given to demonstrate that a multivariate extreme value distribution may not be an appropriate model for an arbitrary combination of extreme variables.

#### 2.2.1 Univariate extreme value theory

Let \(X_1, \ldots, X_n\) be \(n\) independent and identically distributed random variables from the distribution \(F\), and \(M_n \triangleq \max_{i=1}^n X_i\) be the maximum. Fisher and Tippett (1928) and Gnedenko (1943) show that, if there exist location and scale parameters \(a_n \in \mathbb{R}\) and \(b_n > 0\) such that \((M_n - a_n)/b_n\) converges to a non-degenerate distribution as \(n \to \infty\), then the limiting distribution must either be Gumbel\(^2\) (light-tailed), Weibull (finite upper end point) or Fréchet (heavy-tailed). These three families of distributions can be condensed into the

---

\(^2\)The Gumbel distribution here is different from the Gumbel copula mentioned in Section 2.1.
generalized extreme value (GEV) distribution (von Mises (1954); Jenkinson (1955)) with distribution function

\[ G(x; \theta) = \begin{cases} \exp \left\{ - [1 + \gamma \left( \frac{x-\mu}{\sigma} \right)]^{-1/\gamma} \right\}, & \gamma \neq 0; \\ \exp \left\{ - \exp \left( - \frac{x-\mu}{\sigma} \right) \right\}, & \gamma = 0, \end{cases} \]

where \([y]_+ = \max(y, 0)\) and \(\theta = \{\mu, \sigma, \gamma\}\) is the set of parameters. In particular, the Gumbel, Weibull and Fréchet distributions are retrieved as \(\gamma = 0, \gamma < 0\) and \(\gamma > 0\), respectively. The above convergence in distribution result somewhat resembles the central limit theorem for the mean, in which suitable standardization of the sample mean by its corresponding location and scale parameters (i.e., mean and standard deviation) leads to the convergence to the standard normal distribution as \(n \to \infty\). Since the minimum \(m_n \triangleq \bigwedge_{i=1}^n X_i = - \bigvee_{i=1}^n (-X_i)\), it suffices to consider the theory for maxima.

Because each sample contains only one maximum, it is customary to use the block maxima approach for inference (see, e.g., Coles (2001); Beirlant et al. (2004)). Observations are divided into blocks of sufficiently large size or based on natural separation in the measurement unit such as months or years, and the maximum is obtained in each block. Apart from maximum likelihood estimation of the parameters (Prescott and Walden (1980, 1983); Hosking (1985)), alternatives such as the probability-weighted moments (Greenwood et al. (1979); Hosking (1985)) and elemental percentile (Castillo and Hadi (1995)) methods have been proposed.

2.2.2 Multivariate extreme value theory

Some early literature on multivariate extremes includes de Haan and Resnick (1977) and Pickands (1981). Let \((X_{i1}, \ldots, X_{id}) \sim F\) be the \(i\)th replicate of a random vector of dimension \(d\), \(i = 1, \ldots, n\). The vector of componentwise maxima is defined as \((M_{n1}, \ldots, M_{nd}) \triangleq (\bigvee_{i=1}^n X_{i1}, \ldots, \bigvee_{i=1}^n X_{id})\). Suppose there exist standardizing constants \(a_{nj} \in \mathbb{R}, b_{nj} > 0\) for \(j = 1, \ldots, d\) such that, for \(y_j = a_{nj} + b_{nj} x_j\), we have

\[ P(M_{n1} \leq y_1, \ldots, M_{nd} \leq y_d) = F^n(y_1, \ldots, y_d) \triangleq H_n(y_1, \ldots, y_d) \to H(y_1, \ldots, y_d) \]

for some non-degenerate distribution function \(H\) as \(n \to \infty\). It can be shown that \(H\) has GEV margins. It is customary to transform \(H\) to have some standardized margins to focus on the modelling of dependence structure. Let \(G_i\) be the marginal GEV distribution of the \(i\)th variable of \(H\), and \(H(y_1, \ldots, y_d) = C(G_1(y_1), \ldots, G_d(y_d))\) for some copula \(C\). Write

\[ C(u_1, \ldots, u_d) = \exp \{ - A(w_1, \ldots, w_d) \}, \quad (2.5) \]
where \( w_j = -\log u_j, j = 1, \ldots, d \). The representation of de Haan and Resnick (1977) and Pickands (1981) is that the exponent function can be written as

\[
A(w_1, \ldots, w_d) = \int_{S_d} \bigg( \bigwedge_{j=1}^d (w_j \omega_j) \bigg) \, d\mu(\omega),
\]

where \( S_d = \{ \omega \in \mathbb{R}^d : ||\omega|| = 1 \} \) is the unit simplex on \( \mathbb{R}^d \) and \( \mu \) is a positive measure on \( S_d \) that satisfies the mean constraints

\[
\int_{S_d} \omega_j \, d\mu(\omega) = 1, \quad j = 1, \ldots, d.
\]

The exponent function \( A \) is homogeneous of order 1 (see, e.g., Section 3.15 of Joe (2014)). The copula \( C \) constructed above is max-stable, defined as

\[
C^n(u_1^{1/n}, \ldots, u_d^{1/n}) = C(u_1, \ldots, u_d)
\]

for every positive integer \( n \). A copula is an extreme value copula if and only if it satisfies (2.6) (Galambos (1987); Joe (1997)). In this case, \( G(w_1, \ldots, w_d) \) is a min-stable survival function with unit exponential margins. The definition \( G \) is useful in the derivations below.

The density function of an extreme value copula is obtained by differentiating (2.5) with respect to its arguments. Using Theorem 8.46 of Joe (2014) and the chain rule, the density of an extreme value copula \( C_{EV}(u_1, \ldots, u_d) \) is given by

\[
c_{EV}(u_1, \ldots, u_d) = \frac{(-1)^d}{\prod_{i=1}^d u_i} \cdot \frac{\partial^d G(w_1, \ldots, w_d)}{\prod_{i=1}^d \partial w_i}, \quad w_i = -\log u_i,
\]

with the mixed partial derivatives of \( \overline{G} \) of order \( k \leq d \) equal to

\[
(-1)^k \frac{\partial^k \overline{G}}{\prod_{i=1}^k \partial w_i} = e^{-A} \cdot \left[ \sum_{P=(S_1, \ldots, S_{|P|})} (-1)^{k-|P|} \prod_{i=1}^{|P|} A^{(S_i)} \right],
\]

where the arguments of \( \overline{G} \) are omitted for brevity, \( A^{(S)} = \frac{\partial^{|S|} A}{\prod_{j \in S} \partial w_j} \) for a non-empty subset \( S \) of the index set \( \{1, \ldots, d\} \), and the summation is over all possible partitions of the index set. For \( d = 1, 2, 3 \), we can readily obtain the following expressions:

\[
-\frac{\partial \overline{G}}{\partial w_1} = e^{-A} \left( \frac{\partial A}{\partial w_1} \right) = e^{-A} \left( A^{(1)} \right);
\]

\[
-\frac{\partial^2 \overline{G}}{\partial w_1 \partial w_2} = e^{-A} \left( \frac{\partial^2 A}{\partial w_1 \partial w_2} \right) = e^{-A} \left( A^{(1)} A^{(1)} - A^{(12)} \right);
\]

\[
-\frac{\partial^3 \overline{G}}{\partial w_1 \partial w_2 \partial w_3} = e^{-A} \left( \frac{\partial^3 A}{\partial w_1 \partial w_2 \partial w_3} \right) = e^{-A} \left( A^{(1)} A^{(1)} A^{(3)} - A^{(12)} A^{(3)} - A^{(13)} A^{(1)} - A^{(23)} A^{(1)} + A^{(123)} \right).
\]
Since the summation in (2.7) enumerates all possible partitions of the index set, the number of terms in the partial derivative (and hence the copula density) grows according to the Bell numbers (Bell (1934)).

The case with \( d = 2 \) has been studied in detail. Examples of commonly used bivariate extreme value distributions include the Gumbel or logistic (Gumbel (1960)), asymmetric logistic (Tawn (1988)), Galambos or negative logistic (Galambos (1975)) and its asymmetric variant (Joe (1993)), and the Hüsler-Reiss (Hüsler and Reiss (1989)) model. However, not all of these have a direct multivariate generalization. The Hüsler-Reiss distribution is one that allows so: A random vector \((U_1, \ldots, U_d)\), with unit uniform margins, follows the Hüsler-Reiss distribution if the distribution function is given by

\[
C_{U_1,\ldots,U_d}(u_1, \ldots, u_d; \{\delta_{jk}\}) = \exp \left\{ -\sum_{j=1}^{d} w_j \Phi_{d-1, \Gamma_j}(v_j) \right\},
\]

(2.8)

where \( v_j = (v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_d) \) with \( v_k = \delta_{jk}^{-1} + \delta_{jk} \log (u_j/u_k) / 2 \), \( w_j = -\log u_j \) and \( \Phi_{d-1, \Gamma_j} \) is the \((d-1)\)-dimensional Gaussian distribution with zero mean and correlation matrix

\[
\Gamma_j = (\gamma_{ik})_{i, k \neq j}, \quad \gamma_{ik} = \frac{\delta_{ij}^{-2} + \delta_{ij}^{-2} - \delta_{ik}^{-2}}{2\delta_{ij}^{-1} \delta_{kj}^{-1}}.
\]

(2.9)

The \( \left(\begin{smallmatrix} d \\ 2 \end{smallmatrix}\right) \) parameters, one for each bivariate margin, are \( \delta_{ij} \geq 0 \) with \( \delta_{ji} = \delta_{ij} \); for completeness, \( \delta_{ii}^{-1} \) is defined to be 0 for all \( i \). The distribution function in the compact form (2.8) is attributed to Nikoloulopoulos et al. (2009). The Hüsler-Reiss distribution is generalizable to higher dimensions because of the particular way it is constructed. In the bivariate case, it is obtained as the extreme value limit of a bivariate Gaussian vector whose correlation approaches 1 at a suitable rate, as the sample size tends to infinity. A direct generalization is obtained by letting all pairwise correlations of a multivariate Gaussian vector tend to 1. It is easy to verify that the distribution (2.8) satisfies the max-stability condition (2.6).

The Hüsler-Reiss distribution is used in the field of spatial extremes (see, e.g., Smith (1990)) and arises from a completely different framework known as the class of max-stable processes. It generalizes finite multivariate extreme value distributions to a continuous state space. In particular, consider a Poisson process \( \{\xi_i, U_i \in N\} \) on \( \mathbb{R}_+ \times \mathbb{R}^d \) with intensity measure \( \xi^{-2} d\xi \times \nu(du) \), where \( \nu \) is a positive measure. Let \( g : (\mathbb{R}^d \times \mathbb{R}^d) \to \mathbb{R}_+ \) be the Gaussian density \( g(u, t) = (2\pi)^{-d/2} |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2}(u - t)^\top \Sigma^{-1}(u - t) \right\} \) centred at \( u \) and with covariance matrix \( \Sigma \). The process defined by

\[
Z(t) = \max_{i=1,\ldots,n} \{\xi_i g(U_i, t)\}
\]

(2.10)

has finite-dimensional distributions being Hüsler-Reiss (Genton et al. (2011)). The process (2.10) is now known as the Smith model. The joint distribution of \( Z(t_1) \) and \( Z(t_2) \) for
$t_1, t_2 \in \mathbb{R}^d$ has dependence parameter \( \delta_{12} = 2 \left[ (t_1 - t_2) \Sigma^{-1}(t_1 - t_2) \right]^{-1/2} \). Smith (1990) gives the process (2.10) a storm profile interpretation: Consider an infinite number of storms with intensities \( \xi_i \) and centres \( u_i \). The function \( g(u, t) \) reflects the effect, such as precipitation, at location \( t \) induced by a storm centred at \( u \). The overall effect depends on the size of the storm \( \xi_i \) in a multiplicative manner. The finite-dimensional representation of the Smith process is a parsimonious model of the Hüsler-Reiss distribution because the dependence parameters \( \delta_{ij} \) are completely specified by the covariance matrix \( \Sigma \). In other words, \( \Sigma \) determines the joint distribution of an arbitrary number of random variables in this random field.

A generalization of the multivariate Hüsler-Reiss distribution is the t-EV distribution (Demarta and McNeil (2005); Nikoloulopoulos et al. (2009)), obtained by taking the extreme value limit of the t distribution. A random vector \((U_1, \ldots, U_d)\), with unit uniform margins, follows the t-EV distribution if the distribution function is given by

\[
C(u_1, \ldots, u_d; \Omega, \nu) = \exp \left\{ - \sum_{j=1}^d w_j T_{d-1, \nu+1}(s_j; \Omega_j) \right\}, \tag{2.11}
\]

where \( s_j = (s_1, \ldots, s_{j-1}, s_{j+1}, \ldots, s_d) \) with \( s_k = \sqrt{\nu + 1} \left[ (w_j/w_k)^{1/\nu} - \rho_{jk} \right] / \sqrt{1 - \rho_{jk}^2} \), \( w_j = -\log u_j \), \( \nu \) is the degrees of freedom parameter, \( \Omega_j \) is the correlation matrix \( \Omega = (\rho_{rs}) \) with the \( j \)th variable partialled out (that is, the \( \rho_{rs} \)'s are partial correlations), and \( T_{d-1, \nu+1}(\cdot; \Omega_j) \) is the \((d - 1)\)-variate t distribution function with \( \nu + 1 \) degrees of freedom and correlation matrix \( \Omega_j \) (Section 4.16 of Joe (2014)). The t-EV distribution is the finite-dimensional distribution of the max-stable process known as extremal-t (Opitz (2013), see also Ribatet (2013)). The Hüsler-Reiss distribution is obtained by taking the limit \( \nu \to \infty \).

### 2.2.3 Relationship between marginal and joint extremes

Through a data set, we illustrate in this section that the class of multivariate extreme value copulas may not be an appropriate family for an arbitrary collection of marginally extreme variables. In this example, the variables of interest are annual maxima of daily streamflows at 6 gauging stations in eastern Vancouver Island. The locations of the stations are shown in Figure 2.3. There are around 50 observations for each station except Mill Bay, which has 34. Inspection of the data set reveals that, in many cases, the extremes for different stations are recorded on vastly different dates. For instance, in 2003 extreme streamflows were recorded in mid-October for three of the stations, mid-March for two of them and early January for the remaining one. Given this feature, the componentwise maxima (with block sizes of one year) are unlikely to correspond to actual sample points from a multivariate extreme value distribution.
To analyze the data set, we first fit GEV distributions to the margins; the resulting Q-Q plots (Figure 2.4) show acceptable fit. Next, we observe the pairwise dependence structure by means of normal scores plots, in which the data are transformed to standard normal margins using the empirical quantiles. The normal scores plot is preferred over a plot of the raw data because variables may have different scales. It is also preferred over a plot of data transformed to unit uniform margins because the tail behaviour is more easily identified in a normal scores plot. From the pairwise normal scores plots in Figure 2.5, it is clear that the lower tail has stronger dependence than the upper tail for most pairs. Joint modelling using multivariate extreme value copulas is inappropriate as such copulas for the maximum can only have upper tail dependence (see Section 2.3.3). Another way to illustrate the unsuitability of multivariate extreme value copulas is to produce empirical max-stable or min-stable plots as in Figure 2.6. These plots are based on the principle of max- or min-stability of the corresponding weighted maxima/minima: For a multivariate extreme value copula $C_{EV}(u_1, \ldots, u_d) = \exp \{-A(w_1, \ldots, w_d)\} = G(w_1, \ldots, w_d)$, let $W_1, \ldots, W_d$ be unit
exponential random variables with joint survival function \( G \). Define \( W^* = \bigwedge_{i=1}^{d} W_i/\omega_i \), where the \( \omega_i \)'s are positive constants. Then,

\[
P(W^* > w) = P(W_1 > w\omega_1, \ldots, W_d > w\omega_d) = \exp\{-wA(\omega_1, \ldots, \omega_d)\}
\]

(2.12)
due to the homogeneity property of \( A \). Hence \( W^* \) is exponentially distributed with rate \( A(\omega_1, \ldots, \omega_d) \). An empirical min-stable plot is a Q-Q plot of the realizations of \( W^* \) versus the theoretical quantiles of an exponential distribution, for some arbitrarily chosen weights \( \omega_i \)'s. The realizations of \( W^* \) are obtained from the empirical probability integral transform, so that each margin has unit exponential distribution. The mean of the exponential distribution (2.12) can be estimated using maximum likelihood method (i.e., the sample mean in this case). If the data can be reasonably approximated by a multivariate extreme value distribution, the Q-Q plot should show no substantial departure from a straight line. Similarly, for the max-stable plot, we let \( V_i = 1/W_i, \ i = 1, \ldots, d \), be unit Fréchet random variables with joint distribution \( F(v_1, \ldots, v_d) = \overline{G}(1/v_1, \ldots, 1/v_d) \). We use \( (\omega_1, \ldots, \omega_d) = (1, \ldots, 1) \) for the plots in Figure 2.6. The departure from the straight line in these plots suggests that a multivariate extreme value copula is not appropriate here.

![Figure 2.4: Q-Q plots for the marginal GEV fitting of the Vancouver Island streamflows data](image)

The reason for the mismatch in the dates is that the extremes for various stations are initiated by different mechanisms. Unlike the Fraser River example we will introduce in Section 3.7, snowmelt is not the main driving force behind extreme streamflows in eastern
Vancouver Island. Even though the stations are only tens of kilometres apart, heavy rainfall tends to be more localized and thus an episode of extreme precipitation is less likely to affect all stations at the same time. As a result, joint extremes may not exhibit upper tail dependence and multivariate extreme value copulas are unsuitable. For such kind of data, it is possible to conduct inference by fitting GEV distributions to the margins but using general copulas for the multivariate dependence structure. Examples of this hybrid modelling technique can be found in the hydrology literature, e.g., Klein et al. (2010) and Requena et al. (2013).
2.3 Tail dependence functions and extreme value limit of copula models

In this section, we discuss the concept of tail dependence and outline its relationship with extreme value copulas. Some other quantities related to the dependence properties of a bivariate extreme value copula are then given.

2.3.1 Tail dependence functions

Tail dependence functions describe the behaviour of a copula at the joint upper or lower tail. The concept of tail dependence is introduced formally in Nikoloulopoulos et al. (2009); see also Joe et al. (2010) for its use in vine copulas. We first consider continuous bivariate copulas \( C \); these concepts can be extended to higher dimensions easily.

To begin, the tail dependence index (or coefficient) is a commonly used measure in copula and extreme value theory to summarize the strength of dependence as one approaches the limit of a multivariate distribution (see, e.g., Joe (1997); Coles (2001)). The lower tail dependence index is defined as

\[
\lambda_L \triangleq \lim_{u \to 0^+} \frac{C(u, u)}{u} = \lim_{u \to 0^+} \frac{P(U_1 \leq u | U_2 \leq u)}{u} = \lim_{u \to 0^+} \frac{P(U_2 \leq u | U_1 \leq u)}{u},
\]  

(2.13)
provided the limit exists. Similarly, the upper tail dependence index is defined as
\[
\lambda_U \triangleq \lim_{u \to 1-} \frac{\bar{C}(u,u)}{1-u} = \lim_{u \to 1-} \mathbb{P}(U_1 > u|U_2 > u) = \lim_{u \to 1-} \mathbb{P}(U_2 > u|U_1 > u),
\]
where \( \bar{C}(u,u) = \mathbb{P}(U_1 > u, U_2 > u) = 1 - 2u + C(u,u) \) is the survival function of \((U_1, U_2)\) at \((u,u)\). It is easy to see that \( \lambda_L, \lambda_U \in [0,1] \); when \( \lambda_L \) (resp. \( \lambda_U \)) is 0, we say that the bivariate copula has no lower (resp. upper) tail dependence. The strength of tail dependence increases with the magnitude of \( \lambda_L \) or \( \lambda_U \).

One limitation of the tail dependence index is that it only summarizes dependence along the main diagonal of the copula. The tail dependence function (Nikoloulopoulos et al. (2009); Joe et al. (2010)) is a generalization of the tail dependence index. For a bivariate copula \( C \), the lower tail dependence function is given by
\[
b(w_1, w_2; C) \triangleq \lim_{u \to 0^+} \frac{C(uw_1, uw_2)}{u},
\]
while the upper tail dependence function is
\[
b^*(w_1, w_2; C) \triangleq \lim_{u \to 0^+} \frac{\bar{C}(1 - uw_1, 1 - uw_2)}{u} = \lim_{u \to 0^+} \frac{\hat{C}(uw_1, uw_2)}{u},
\]
where \( \hat{C}(v_1, v_2) = \bar{C}(1 - v_1, 1 - v_2) \) is the reflected copula of \( C \) and thus the upper tail dependence function of \( C \) is equal to the lower tail dependence function of \( \hat{C} \). The lower and upper tail dependence indices are then given by \( b(1,1; C) \) and \( b^*(1,1; C) \), respectively. By varying the ratio \( w_1/w_2 \), we can obtain a measure of the tail dependence of \( C \) along different directions. Figure 2.7 plots the lower tail dependence functions for two bivariate copulas: MTCJ and reflected Gumbel. For each copula, the plot on the left shows the density contour of the variables \((X_1, X_2)\) with standard Gaussian margins, while the plot on the right is the contour of the corresponding lower tail dependence function \( b(w_1, w_2; C) \). The lower tail dependence index is also shown. Both copulas have Kendall’s \( \tau \) equal to 0.5, but note that the MTCJ copula has stronger lower tail dependence than the reflected Gumbel copula.

The \( b \) and \( b^* \) functions can be extended to higher dimensions in a straightforward manner. For a \( d \)-dimensional copula \( C_d \), we have
\[
b(w_1, \ldots, w_d; C_d) = \lim_{u \to 0^+} \frac{C_d(uw_1, \ldots, uw_d)}{u};
\]
\[
b^*(w_1, \ldots, w_d; C_d) = \lim_{u \to 0^+} \frac{\bar{C}_d(1 - uw_1, \ldots, 1 - uw_d)}{u}.
\]

A related quantity is the marginal tail dependence function for a subset of variables \( S = \{k_1, \ldots, k_m\} \subset \{1, \ldots, d\} \), where \( m \leq d \). The definitions for the lower and upper tails
are

\[ b_S(w_{k_1}, \ldots, w_{k_m}; C_d) = \lim_{u \to 0^+} \frac{C_S(u w_{k_1}, \ldots, u w_{k_m})}{u}, \]

\[ b^*_S(w_{k_1}, \ldots, w_{k_m}; C_d) = \lim_{u \to 0^+} \frac{C_S(1 - u w_{k_1}, \ldots, 1 - u w_{k_m})}{u}, \]

where \( C_S \) is the distribution function of \( U_{k_1}, \ldots, U_{k_m} \). In the following, the copula argument in the tail dependence function will be omitted when there is no ambiguity.

Another related quantity is the conditional tail dependence function, involved in the computation of tail dependence functions for factor and vine copulas. It is defined as

\[ b_{k_1|k_2, \ldots, k_m}(w_{k_1}|w_{k_2}, \ldots, w_{k_m}) = \lim_{u \to 0^+} C_{U_{k_1}|U_{k_2}, \ldots, U_{k_m}}(u w_{k_1}|u w_{k_2}, \ldots, u w_{k_m}), \]

where \( k_i \in \{1, \ldots, d\} \) and \( k_i \neq k_j \) if \( i \neq j \). When \( m = 2 \), \( b_{k_1|k_2}(w_{k_1}|w_{k_2}) \) can be obtained by differentiating \( b_{k_1,k_2}(w_{k_1}, w_{k_2}) \) with respect to \( w_{k_2} \) (Theorem 8.58 of Joe (2014)).

Figure 2.7: Density contours (left) and lower tail dependence functions (right) of the MTCJ and reflected Gumbel copulas with Kendall’s \( \tau \) equal to 0.5. The lower tail dependence index is highlighted as the value of \( b(1, 1) \).
The tail dependence function is related to the joint probability of each component being small/large. A similar quantity useful in extreme value context is related to the probability of some of the components being small/large, defined here as the $a(\cdot)$ and $a^*(\cdot)$ functions for the lower and upper tail, respectively:

$$a(w_1, \ldots, w_d) = \lim_{u \to 0^+} \frac{\mathbb{P}\left\{ \bigcup_{i=1}^d \{U_i \leq uw_i\} \right\}}{u};$$

$$a^*(w_1, \ldots, w_d) = \lim_{u \to 0^+} \frac{\mathbb{P}\left\{ \bigcup_{i=1}^d \{U_i > 1 - uw_i\} \right\}}{u}.$$

Note that $a(\cdot)$ and $a^*(\cdot)$ are known as stable tail dependence functions in the extreme value literature (Huang (1992); Drees and Huang (1998)). By the inclusion-exclusion principle,

$$a(w_1, \ldots, w_d) = \lim_{u \to 0^+} \frac{\sum_{S \subset I_d} (-1)^{|S|-1} \mathbb{P}\left\{ \bigcap_{i \in S} \{U_i \leq uw_i\} \right\}}{u} = \sum_{S \subset I_d, S \neq \emptyset} (-1)^{|S|-1} b_S(w_i, i \in S),$$

(2.16)

where $I_d = \{1, \ldots, d\}$ is the index set and $S$ is a non-empty subset of $I_d$. A similar relationship holds between $a^*(\cdot)$ and $b^*(\cdot)$. As we will see below, the quantity $a(\cdot)$ appears in the extreme value limit of a copula; equation (2.16) therefore connects the tail properties of a copula to its extreme value limit.

### 2.3.2 Extreme value limit of copula models

Since extreme value copulas satisfy the max-stability condition (2.6), we can construct an extreme value copula from an arbitrary copula $C$ by computing its extreme value limit, i.e., $\lim_{n \to \infty} C_n(u_1^{1/n}, \ldots, u_d^{1/n})$ for the upper tail or $\lim_{n \to \infty} \hat{C}_n(u_1^{1/n}, \ldots, u_d^{1/n})$ for the lower tail, provided that it exists and is not the distribution function of a degenerate distribution. In the following, we quote the procedure outlined in Joe (2014) for the evaluation of such limit with the help of tail dependence functions.

The lower extreme value limit\(^3\) of $C$, $C_{LEV}$, is equivalent to the upper extreme value

\(^3\)Here we mean that $C_{LEV}$ has the tail properties of the lower extreme value limit of $C$. In terms of sampling, this means componentwise minima are taken from observations sampled from $C$. However, we focus on maxima for ease of presentation, and thus the $C_{LEV}$ constructed in the subsequent text is actually reflected and has upper tail dependence instead. For modelling purposes, componentwise minima are negated so that they exhibit upper tail dependence.
limit of \( \hat{C} \), i.e., \( \lim_{n \to \infty} \hat{C}^n(u_1^{1/n}, \ldots, u_d^{1/n}) \). First, we have from definition

\[
\hat{C}^n(u_1^{1/n}, \ldots, u_d^{1/n}) = \left[ \mathbb{P} \left( 1 - U_1 \leq u_1^{1/n}, \ldots, 1 - U_d \leq u_d^{1/n} \right) \right]^n \\
= \left[ \mathbb{P} \left( 1 - U_1 \leq 1 - (-\log u_1)/n, \ldots, 1 - U_d \leq 1 - (-\log u_d)/n \right) \right]^n \\
= \left[ \mathbb{P} \left( \bigcap_{i=1}^d \{ U_i > -\log u_i/n \} \right) \right]^n,
\]

where the approximation follows from \( u_i^{1/n} = \exp\{n^{-1} \log u_i\} \sim 1 - (-\log u_i)/n \) as \( n \to \infty \). Recall that, as \( t \downarrow 0 \),

\[
t \cdot a(w_1, \ldots, w_d) \sim \mathbb{P} \left( \bigcup_{i=1}^d \{ U_i \leq tw_i \} \right) = 1 - \mathbb{P} \left( \bigcap_{i=1}^d \{ U_i > tw_i \} \right)
\]

so that \( \mathbb{P} \left( \bigcap_{i=1}^d \{ U_i > tw_i \} \right) \sim 1 - t \cdot a(w_1, \ldots, w_d) \). Substituting \( t = 1/n \) (which approaches zero from above as \( n \to \infty \)) and \( w_i = -\log u_i \), we obtain

\[
\hat{C}^n(u_1^{1/n}, \ldots, u_d^{1/n}) \sim \left[ 1 - a(-\log u_1, \ldots, -\log u_d)/n \right]^n \to \exp \{ -a(-\log u_1, \ldots, -\log u_d) \}
\]
as \( n \to \infty \). Hence, \( C_{LEV}(u_1, \ldots, u_d) = \exp \{ -a(-\log u_1, \ldots, -\log u_d) \} \). The upper extreme value limit of \( C \) can be obtained via a similar argument, with \( C_{UEV}(u_1, \ldots, u_d) = \exp \{ -a^*(-\log u_1, \ldots, -\log u_d) \} \).

An insight from this result is that the extreme value limit of a given copula \( C \) can be expressed in terms of the exponent functions \( a(\cdot) \) and \( a^*(\cdot) \), which are in turn related to the tail dependence functions \( b(\cdot) \) and \( b^*(\cdot) \) via (2.16). A class of potentially useful extreme value copula models can thus be derived from the original copula.

### 2.3.3 Measures of dependence for bivariate extreme value copulas

There exist several measures that describe the strength of dependence of a bivariate extreme value copula \( C_{EV} \). One widely used measure is the Pickands dependence function (Pickands (1981)). Write

\[
C_{EV}(u_1, u_2) = \exp \{ -A(-\log u_1, -\log u_2) \}.
\] (2.17)

The Pickands dependence function is given by \( B(w) = A(w, 1 - w) \) with \( 0 \leq w \leq 1 \). It can be shown that \( B(0) = B(1) = 1 \), \( \max(w, 1 - w) \leq B(w) \leq 1 \) and \( B \) is convex. The Pickands dependence function also serves to simplify the calculation of other dependence measures such as Kendall’s \( \tau \) and Spearman’s \( \rho \), avoiding the need to evaluate a two-dimensional integral as is usually needed for general bivariate distributions (see Theorem 8.44 of Joe (2014)). A summary of the strength of dependence using a single number, known as the
extremal coefficient, has origins dating back to Sibuya (1960) but is coined in Smith (1990). It is defined as the parameter $\vartheta$ that satisfies
\[ C_{EV}(u, u) = u^{\vartheta}. \]

The possible range of $\vartheta$ is $[1, 2]$, with $\vartheta = 2$ at independence and $\vartheta = 1$ at comonotonicity. The extremal coefficient can be interpreted as the “effective” number of independent variables. Since $A$ in (2.17) is homogeneous of order 1,
\[ \vartheta = A(1, 1) = 2B(1/2). \]

There is a one-to-one mapping between the extremal coefficient and the upper tail dependence index, while the lower tail dependence index of an extreme value copula is always zero except for the comonotonicity limit. This can be derived from the definition of the various quantities:
\[ \lambda_{EV}^{UL} = \lim_{u \to 0^+} \frac{C_{EV}(u, u)}{u} = \lim_{u \to 0^+} u^{\vartheta - 1} = \begin{cases} 0 & \text{if } \vartheta \in (1, 2], \\ 1 & \text{if } \vartheta = 1; \end{cases} \]
\[ \lambda_{EV}^{EU} = \lim_{u \to 1^-} \frac{C_{EV}(u, u)}{1 - u} = \lim_{v \to 0^+} \frac{2v - 1 + (1 - v)^{\vartheta}}{v} = \lim_{v \to 0^+} \frac{2v - 1 + 1 - \vartheta v + o(v)}{v} \]
\[ = 2 - \vartheta \text{ or } 2 - A(1, 1) \text{ or } 2 \left[ 1 - B(1/2) \right]. \]

### 2.4 Dependence measures for general bivariate copulas

In this section, we state the definitions of several dependence measures used in the thesis. Dependence measures (or measures of concordance) are quantities that summarize the strength of association among variables. The Pearson correlation coefficient is not invariant under monotone transformations (see, e.g., Section 2.12 of Joe (2014) for this and other properties of the Pearson correlation coefficient) and therefore not desirable for use in copula modelling, as the dependence structure does not depend on marginal distributions.

Throughout this section, we let $U_i = (U_{i1}, U_{i2}) \top$, $i = 1, \ldots, n$, be an i.i.d. sample from a bivariate copula $C$ with realizations $u_i = (u_{i1}, u_{i2}) \top$, and $r_{ij}$ be the rank of $u_{ij}$ among $u_{1j}, \ldots, u_{nj}$, $j = 1, 2$. The subscript $i$ is dropped when we refer to the distributional properties of the random variables. Note that these measures are also defined for distributions whose margins are not $U(0,1)$.

1. Kendall’s $\tau$ (based on the number of concordant and discordant pairs in a sample), with population version given by
\[ \tau = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1 \]
and the empirical version (without ties) given by

\[ \hat{\tau} = \frac{4}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 1 \{(u_{i1} - u_{j1})(u_{i2} - u_{j2}) > 0\} - 1. \]

2. Spearman’s \( \rho \) (based on the correlation of marginal ranks), with population and empirical (without ties) versions given by

\[ \rho_S = 12 \int_0^1 \int_0^1 uv \, dC(u,v) - 3 = \text{Cor}(U_1, U_2); \quad \hat{\rho}_S = \frac{\sum_{i=1}^{n} r_{i1}r_{i2} - n [(n+1)/2]^2}{n(n^2-1)/12} \]

3. Blomqvist’s \( \beta \) (based on the centre of the distribution), with population version \( \beta = 4C(0.5, 0.5) - 1 \). The most efficient way of estimating \( \beta \) from data (see Section 2.12.3 of Joe (2014)) is through the ranks:

\[ \hat{\beta} = \frac{2}{n} \sum_{i=1}^{n} 1 \left\{ \left( r_{i1} - \frac{n+1}{2} \right) \left( r_{i2} - \frac{n+1}{2} \right) \geq 0 \right\} - 1. \]

4. Correlation of normal scores \( \rho_N \) (based on the normal scores of the observations), with population version \( \rho_N = \text{Cor} [\Phi^{-1}(U_1), \Phi^{-1}(U_2)] \). Let

\[ s_{i1} = (r_{i1} - 0.5)/n; \quad s_{i2} = (r_{i2} - 0.5)/n \quad (2.18) \]
be the ranks adjusted to \([0, 1]\) scale. The empirical version is given by

\[ \hat{\rho}_N = \text{Cor} [\Phi^{-1}(s_{i1}), \Phi^{-1}(s_{i2})], \]
where \( \text{Cor} \) denotes the sample correlation.

5. Tail-weighted dependence measures (Krupskii and Joe (2015)), defined as the correlation of a functional transform of the observations in a restricted region of the sample space. For continuous function \( h(\cdot) : [0, 1] \to (0, \infty) \) and truncation level \( p \in (0, 0.5] \), the population version is given by

\[ g_L(h, p) = \text{Cor} \left[ h \left( 1 - \frac{U_1}{p} \right), h \left( 1 - \frac{U_2}{p} \right) \mid U_1 < p, U_2 < p \right]; \]
\[ g_U(h, p) = \text{Cor} \left[ h \left( 1 - \frac{1-U_1}{p} \right), h \left( 1 - \frac{1-U_2}{p} \right) \mid 1 - U_1 < p, 1 - U_2 < p \right], \]

where the subscripts \( L \) and \( U \) denote the lower and upper measures, respectively. The empirical version is obtained by replacing the correlation function with the sample correlation, and the random variables \( U \)'s with the adjusted ranks \( s_{i1} \) and \( s_{i2} \) defined in (2.18). Under certain regularity conditions, the empirical estimator is asymptotically normal.
All five measures defined above are invariant under monotone transformations. In addition, the first four measures, i.e., Kendall’s $\tau$, Spearman’s $\rho$, Blomqvist’s $\beta$ and correlation of normal scores $\rho_N$, satisfy all desirable properties for measures of concordance suggested in Scarsini (1984) (see Definition 2.8 of Joe (2014)). These four measures can be considered as central or global dependence measures as they summarize the dependence strength of the whole copula. For some applications, it may be of interest to focus only on the tails. The tail dependence index is one such measure, but its definition (2.13) or (2.14) involves a limit argument and cannot be estimated reliably from data except when one considers extreme value copulas, in which case the (upper) tail dependence index can be estimated via the extremal coefficient. The tail-weighted dependence measures serve as an alternative, for which the empirical version can be easily estimated from data. Note that the tail-weighted dependence measures do not satisfy all of the criteria in Scarsini (1984); for example, they are not defined for the countermonotonicity copula.

2.5 Model estimation methods for multivariate copulas

Despite the desirable properties of maximum likelihood estimation, it is not always feasible for parametric multivariate copula models due to model complexity. In this section, we provide an overview of several alternative approaches for inference.

2.5.1 Composite likelihood

Maximum likelihood estimation requires the joint density function to be known. In high-dimensional problems, it is often hard or even impossible to obtain the joint density. Composite likelihood methods (Lindsay (1988); Cox and Reid (2004); Varin et al. (2011)) provide an attractive alternative whereby likelihood contributions involve marginal or conditional densities of lower dimensions. Let $F(\cdot, \theta)$ be the distribution function of the $d$-dimensional i.i.d. random vectors $Y_1, \ldots, Y_n$ with parameter $\theta$, $f_{A_1}(y_{A_1}; \theta), \ldots, f_{A_K}(y_{A_K}; \theta)$ be a set of density or mass functions corresponding to marginal distributions indexed by $A_1, \ldots, A_K$, and $f_{B_1|C_1}(y_{B_1}|y_{C_1}; \theta), \ldots, f_{B_K|C_K}(y_{B_K}|y_{C_K}; \theta)$ be a set of density or mass functions corresponding to conditional distributions with variables indexed by $B_1, \ldots, B_K$ left of the conditioning operator and $C_1, \ldots, C_K$ right of it. The composite likelihood can be defined as the product of these marginal density or mass functions

$$L_{C, \text{marg}}(\theta; y_1, \ldots, y_n) = \prod_{i=1}^{n} \prod_{k=1}^{K} \left( f_{A_k}(y_{A_i}; \theta) \right)^{w_k},$$
or that of the conditional density or mass functions

\[
\mathcal{L}_{C,\text{cond}}(\theta; y_1, \ldots, y_n) = \prod_{i=1}^{n} \prod_{k=1}^{K} \left[ f_{B_k|C_k}(y_i, B_k|y_i, C_k; \theta) \right]^{w_k},
\]

where \( y_1, \ldots, y_n \) are the observed values (with subsets \( y_{i,A_k}, y_{i,B_k} \) or \( y_{i,C_k} \) as in the above expressions) and \( w_k \) are some non-negative weights. That is, the distributions \( f_{A_k} \)'s or \( f_{B_k|C_k} \)'s are effectively treated as independent. Special emphasis is on pairwise likelihood, obtained by using bivariate marginal densities in the likelihood construction. We use pairwise likelihoods in Chapter 3 as it is relatively easy to obtain bivariate densities of extreme value copulas. For simplicity, we also let all the weights be equal in the sequel, as our focus is not on the optimal choice of weights. Let \( c_{jk} \) be the marginal copula density for the observed variables \( j \) and \( k \). The pairwise composite log-likelihood of the copula model is

\[
\ell_P(\theta; u) = \sum_{i=1}^{n} \sum_{j<k} \log c_{jk}(u_{ij}, u_{ik}; \hat{\theta}_P),
\]

(2.19)

where \( \theta \) is the collection of all parameters and \( u = (u_{rs}) \) for \( 1 \leq r \leq n \) and \( 1 \leq s \leq d \) is the collection of all data in the unit square, assumed to have unit uniform margins. The pairwise composite likelihood estimator \( \hat{\theta}_P \) is obtained by maximizing (2.19) with respect to \( \theta \). Given the usual regularity conditions, \( \hat{\theta}_P \) is consistent and has asymptotic distribution

\[
\sqrt{n} \left( \hat{\theta}_P - \theta_0 \right) \xrightarrow{d} N \left( 0, G^{-1}(\theta_0) \right)
\]
as \( n \to \infty \), where \( \theta_0 \) is the true value and \( G(\theta) = H(\theta)J^{-1}(\theta)H(\theta) \) is the Godambe or sandwich information matrix (Godambe (1960)); here \( H(\theta) = E[-\nabla^2 \ell_P(\theta; u_1)] \) and \( J(\theta) = \text{Var}[\nabla \ell_P(\theta; u_1)] \) are the sensitivity and variability matrices, respectively, and \( u_1 = (u_{i1}, \ldots, u_{id}) \) is the first observation. If \( \ell_P \) were the full log-likelihood, \( H(\theta) = J(\theta) \) and \( G(\theta) \) would reduce to the usual Fisher information matrix. Estimation using composite likelihood has lower efficiency in the sense that the variance of the estimator does not attain the Cramér-Rao lower bound.

The sensitivity and variability matrices can be estimated using the sample average:

\[
\widehat{H}(\hat{\theta}_P) = -\frac{1}{n} \sum_{i=1}^{n} \sum_{j<k} \frac{\partial^2 \log c_{jk}(u_{ij}, u_{ik}; \hat{\theta}_P)}{\partial \theta \partial \theta^T};
\]

\[
\widehat{J}(\hat{\theta}_P) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j<k} \left[ \frac{\partial \log c_{jk}(u_{ij}, u_{ik}; \hat{\theta}_P)}{\partial \theta} \right] \left[ \frac{\partial \log c_{jk}(u_{ij}, u_{ik}; \hat{\theta}_P)}{\partial \theta^T} \right].
\]

The matrix \( \widehat{H}(\hat{\theta}_P) \) is obtained easily by modern numerical optimization packages used to maximize the composite log-likelihood.
2.5.2 Inference function for margins

Multivariate model fitting via copulas typically deals with both marginal and dependence modelling. Let \( Y_i = (Y_{i1}, \ldots, Y_{id})^\top \) be i.i.d. random vectors from the distribution

\[
G(y; \zeta_1, \ldots, \zeta_d, \delta) = C(G_1(y_1; \zeta_1), \ldots, G_d(y_d; \zeta_d); \delta),
\]

where \( \zeta_j \) is the marginal parameter vector for the \( j \)th margin with distribution function \( G_j \), and \( \delta \) is the copula parameter vector. The full (or joint) maximum likelihood method involves fitting the parameters \( (\zeta_1, \ldots, \zeta_d, \delta) \) at the same time, using the joint density \( g \). Alternatively, if the marginal distributions are (assumed) known, then one can also regard the fitting of \( \delta \) using the copula density \( c \) as through maximum likelihood.

When the number of variables is large, a computationally more efficient approach is to first estimate the marginal parameters individually, obtaining maximum likelihood estimates \( \tilde{\zeta}_1, \ldots, \tilde{\zeta}_d \), and then estimate the copula parameter using the copula density \( c(G_1(y_{i1}; \tilde{\zeta}_1), \ldots, G_d(y_{id}; \tilde{\zeta}_d); \delta) \). This amounts to solving the set of equations

\[
(\partial\ell_1/\partial\zeta_1^\top, \ldots, \partial\ell_d/\partial\zeta_d^\top, \partial\ell/\partial\delta^\top) = 0^\top,
\]

where \( \ell_1, \ldots, \ell_d \) are the marginal log-likelihoods and \( \ell \) is the log-likelihood for the joint distribution. This is known as the method of inference function for margins (IFM) (Joe and Xu (1996); Joe (2005)) or more simply as the two-stage approach. The estimates from this method are typically different from the joint maximum likelihood estimates\(^4\); in particular, the copula parameter estimator \( \tilde{\delta} \) is generally less efficient than the maximum likelihood estimator. The vector of estimator \( \tilde{\theta} = (\tilde{\zeta}_1, \ldots, \tilde{\zeta}_d, \tilde{\delta})^\top \) is asymptotically normal; Joe (2005) provides a decomposition of the asymptotic covariance matrix.

2.5.3 Marginal ranks

When one does not want to assume a parametric family for the margins, it is possible to use only the marginal ranks as input data for copula parameter estimation, bypassing parametric modelling for the margins. Specifically, the copula parameter \( \delta \) is estimated using the copula density \( c(s_{i1}, \ldots, s_{id}; \delta) \), where \( s_{ij} \) are ranks adjusted to the \([0, 1]\) scale as in (2.18). The marginal ranks method is also known as the semiparametric method as the margins are "estimated" nonparametrically. Genest et al. (1995) demonstrate that the resulting estimator of \( \delta \) is asymptotically normal but less efficient than the maximum likelihood estimator.

\(^4\)The two estimates can be the same in certain cases; see Joe (2005).
Note that the IFM method works for discrete margins or when covariates are involved in marginal modelling (e.g., univariate regression models), but the marginal ranks method works only for continuous margins without covariates.
Chapter 3

Parsimonious multivariate extreme value copula models

In multivariate statistics based on the Gaussian distribution, the factor model is a plausible parsimonious model when the dependence among the observed variables can be explained by latent variables. In the more general multivariate modelling based on copulas, factor copula models have been developed in recent years, as well as parsimonious copula models based on Markov tree structures and their vine extensions. For a $d$-dimensional distribution, each of these structures can be described by a total of $O(d)$ parameters instead of the $O(d^2)$ parameters needed in a saturated dependence model. In this chapter, we develop factor and vine analogues for extreme value observations, and use the theory from copula modelling to construct multivariate extreme value copulas with such dependence structures. We show that these models offer insightful interpretations using data examples.

3.1 Introduction

To model the relationship among multivariate observations, the two simplest structures are factor and tree dependence (Markov trees). A 1-factor model assumes that variables are linked to a common, single latent factor, through which dependence is generated. One example is the returns on stocks in the same sector. Meanwhile, a Markov tree structure can be considered when the dependence cannot be adequately explained by latent factors. In a Markov tree, the $d$ observed variables are connected through $d-1$ bivariate acyclic linkages, and non-neighbouring variables are assumed conditionally independent given the variables along the tree path. An example of Markov tree for time-ordered observations is the autoregressive model of order 1. In spatial applications where data are recorded from
stations at different geographic locations, the tree can be drawn according to their locations so that nearest neighbours are linked.

The factor structure can be extended to general $p$-factor models ($p \geq 2$), as well as the bi-factor model and its generalizations where variables are linked to both common and group-specific factors. The Markov tree structure can be extended by adding layers of trees which relax the conditional independence assumption; one thus adds conditional dependence parameters to such additional layers. These dependence structures have been studied outside the Gaussian context through a copula approach, see, for example, Krupskii and Joe (2013) for factor copulas and Brechmann et al. (2012) for truncated vine copulas, where the multiple tree dependence known as vines comes from Bedford and Cooke (2001, 2002). The cases with discrete or mixed continuous/discrete response types have also been dealt with (see Sections 3.9 and 3.10 of Joe (2014) and the references contained therein).

In this chapter, we extend these parsimonious concepts to multivariate extreme value copulas. The conditional independencies implied by the factor and tree structures are hard or may even be impossible to replicate in the extreme value context, as extreme value copula models must satisfy the max-stability property. Instead, classes of multivariate extreme value models are constructed through: (a) extreme value limits of multivariate parametric copula families with specified parsimonious dependence structures, and (b) parsimonious representations of existing extreme value copulas with $\binom{d}{2}$ dependence parameters, by expressing each as functions of other parameters of order $O(d)$. For (b), we consider, in particular, the Hüsler-Reiss distribution (Hüsler and Reiss (1989)), obtained as a non-standard extreme value limit of the multivariate Gaussian distribution. This distribution is chosen for its flexibility in imposing various dependence structures.

The rest of this chapter is organized as follows. We introduce the extreme value factor copula model in Section 3.2. It is the extreme value limit of factor copulas, a class of structured copulas that assume that dependence between observed variables is driven by one or more latent variables. Another class of models, the structured Hüsler-Reiss copula, allows for both factor and vine structures among the variables and is discussed in Section 3.3. A comparison of these models is given in Section 3.4. The inference procedures via composite likelihood methods are addressed in Section 3.5, while Section 3.6 contains a simulation study. We illustrate the use and interpretation of these models in Section 3.7 with two data examples.
3.2 Extreme value factor copula model

The factor copula model described in Section 2.1.2 (Krupskii and Joe (2013)) can be considered as an extension of the Gaussian factor model to general copulas. In this section, we extend the idea of factor structure to the field of extremes using a copula approach. We first derive the extreme value limits of factor copulas; this is followed by a description of the dependence properties of the resulting copulas. Examples of 1- and 2-factor extreme value copula models with specific conditional tail dependence functions or linking copulas between the observed variables and latent factors are then given. We defer the technical details regarding numerical integration methods for likelihood calculation to Chapter 4.

3.2.1 Construction of extreme value factor copulas

By taking the extreme value limit of the factor copula using the procedure outlined in Section 2.3.2, we obtain a class of models (which we name as extreme value factor copulas) suitable for data of extremes with a factor structure. In this subsection, we demonstrate the construction of various extreme value factor copulas. For brevity, we replace the subscript $U_i$ by $i$ in the following discussion, for instance $C_i|V \triangleq C_{U_i|V}$ and $b_i|V \triangleq b_{U_i|V}$. All copulas are assumed to be continuous.

A 1-factor extreme value copula

To obtain the lower extreme value limit of the 1-factor copula model $C$ given by (2.2), we first obtain the lower tail dependence function $b(\cdot)$ of $C$ as follows:

$$b(w_1, \ldots, w_d) = \lim_{u \to 0^+} u^{-1} \int_0^1 \prod_{i=1}^d C_i|V_1(uw_i|v_1) dv_1 = \lim_{u \to 0^+} \int_0^{1/u} \prod_{i=1}^d C_i|V_1(uw_i|uz_1) dz_1$$

$$= \int_0^\infty \lim_{u \to 0^+} \left[ 1\{z_1 \leq u^{-1}\} \cdot \prod_{i=1}^d C_i|V_1(uw_i|uz_1) \right] dz_1$$

$$= \int_0^\infty \prod_{i=1}^d b_i|V_1(w_i|z_1) dz_1, \quad (3.1)$$

in which the substitution $v_1 = uz_1$ is applied. The validity of exchanging the limit and integral operators may be achieved through Lebesgue’s dominated convergence theorem. Alternatively, using results from Joe et al. (2010), a sufficient condition is that the $b_i|V_1$’s are proper distributions on $[0, \infty)$ (see Section 3.16 of Joe (2014); this condition is satisfied for the parametric families described in Section 3.2.4 below). Similarly, we have $b_S(w_i, i \in S) = \int_0^\infty \prod_{i \in S} b_i|V_1(w_i|z_1) dz_1$ for any non-empty subset $S$ of $I_d$. By letting $m_i = b_i|V_1(w_i|z_1)$, we
can arrive at an expression for \( a(\cdot) \) using the relationship (2.16):

\[
a(w_1, \ldots, w_d) = \sum_i b_i(w_i) - \sum_{i<j} b_{i,j}(w_i, w_j) + \cdots + (-1)^{d-1} b_{1,\ldots,d}(w_1, \ldots, w_d)
\]

\[
= \int_0^\infty \left( \sum_i m_i - \sum_{i<j} m_im_j + \cdots + (-1)^{d-1}m_1 \cdots m_d \right) dz_1
\]

\[
= \int_0^\infty \left[ 1 - \prod_{i=1}^d (1 - m_i) \right] dz_1.
\]

(3.2)

The equality between the second and third lines can be established by considering a set of independent events \( E_1, \ldots, E_d \) with respective occurrence probabilities \( m_1, \ldots, m_d \) (note that \( m_i \in [0,1] \), as it is the limit of a conditional probability). Then the inclusion-exclusion principle \( \mathbb{P}\left( \bigcup_{i=1}^d E_i \right) = \sum_i \mathbb{P}(E_i) - \sum_{i<j} \mathbb{P}(E_i \cap E_j) + \cdots + (-1)^{d-1} \mathbb{P}\left( \bigcap_{i=1}^d E_i \right) \) implies that

\[
1 - \prod_{i=1}^d (1 - m_i) = \sum_i m_i - \sum_{i<j} m_im_j + \cdots + (-1)^{d-1}m_1 \cdots m_d.
\]

The copula of the lower extreme value limit of \( C \) is then given by

\[
C_{LEV}(u_1, \ldots, u_d) = \exp \left\{ -\int_0^\infty \left[ 1 - \prod_{i=1}^d (1 - b_i(u_i|z_1)) \right] dz_1 \right\},
\]

where \( w_i = -\log u_i \). Note that the extreme value limit depends only on the conditional tail dependence functions of the linking copulas. Similarly, the upper extreme limit of \( C \) can be obtained as the lower extreme limit of its reflected copula \( \widehat{C} \).

**B 2-factor extreme value copula and higher-order generalization**

A similar technique can be applied to the 2-factor copula (2.3). By setting \( v_1 = uz_1 \), the tail dependence function is

\[
b(w_1, \ldots, w_d) = \lim_{u \to 0^+} u^{-1} \int_0^1 \int_0^1 \prod_{i=1}^d C_{i|V_2:V_1} \left( C_{i|V_1}(uw_i|v_1) | v_2 \right) dv_1dv_2
\]

\[
= \int_0^1 \int_0^\infty \lim_{u \to 0^+} \left[ 1\{z_1 \leq u^{-1}\} \prod_{i=1}^d C_{i|V_2:V_1} \left( C_{i|V_1}(uw_i|uz_1) | v_2 \right) \right] dz_1dv_2
\]

\[
= \int_0^1 \int_0^\infty \prod_{i=1}^d C_{i|V_2:V_1} \left( b_{i|V_1}(w_i|z_1) | v_2 \right) dz_1dv_2,
\]

(3.3)

as \( b_{i|V_1}(w_i|z_1) = \lim_{u \to 0^+} C_{i|V_1}(uw_i|uz_1) \), and the conditions needed for exchanging the limit and integral operators are similar to the 1-factor case. Note how (3.3) resembles (3.1), the

---

5The relationship \( w_i = -\log u_i \) emphasizes the min-stable representation of the distribution function; same for other occurrences in this chapter.
only differences being now a two dimensional integral and the conditional tail dependence functions are themselves arguments for the conditional distribution functions of the second layer copulas. Similarly, the marginal tail dependence functions are given by \( b_S(w_i, i \in S) = \int_0^1 \prod_{i \in S} C_{i|V_2;V_1} \left( b_{i|V_1}(w_i|z_1)|v_2 \right) dz_1dv_2 \). The relationship between \( a(\cdot) \) and \( b(\cdot) \) can be obtained in the same manner as (3.2) with \( m_i = C_{i|V_2;V_1} \left( b_{i|V_1}(w_i|z_1)|v_2 \right) \). Here we only list the final expression:

\[
a(w_1, \ldots, w_d) = \int_0^1 \int_0^\infty \left( 1 - \prod_{i=1}^d \left[ 1 - C_{i|V_2;V_1} \left( b_{i|V_1}(w_i|z_1)|v_2 \right) \right] \right) dz_1dv_2, \quad (3.4)
\]

and the lower extreme value limit is \( C_{LEV}(u_1, \ldots, u_d) = \exp \{ -a(w_1, \ldots, w_d) \} \). The limit thus involves (a) the conditional tail dependence functions for the copulas linked to the first latent factor, and (b) the (conditional distributions of) copulas that link the conditional distributions of observed variables given the first latent factor and that of the second latent factor given the first.

These results can be generalized to \( p \)-factor copulas with \( p \geq 3 \). With \( d \) observed variables, the specification of a \( p \)-factor copula requires \( d \times p \) linking copulas. Note that the conditional distribution function of \( U_i \) given all latent factors can be written as

\[
C_{i|V_1,\ldots,V_p}(u_i|v_1,\ldots,v_p) = P(U_i \leq u_i|V_1 = v_1,\ldots,V_p = v_p) = \frac{\partial}{\partial v_p} P(U_i \leq u_i, V_p \leq v_p|V_1 = v_1,\ldots,V_{p-1} = v_{p-1}) = \frac{\partial}{\partial v_p} C_{i|V_p;V_1,\ldots,V_{p-1}} \left( C_{i|V_1,\ldots,V_{p-1}}(u_i|v_1,\ldots,v_{p-1}), v_p \right) = C_{i|V_p;V_1,\ldots,V_{p-1}} \left( C_{i|V_1,\ldots,V_{p-1}}(u_i|v_1,\ldots,v_{p-1})|v_p \right).
\]

This recursion allows us to express \( C_{i|V_1,\ldots,V_p}(u_i|v_1,\ldots,v_p) \) in terms of the conditional distributions of the linking copulas at each level. For example, when \( p = 3 \), we have

\[
C_{i|V_1,V_2,V_3}(u_i|v_1,v_2,v_3) = C_{i|V_1,V_2,V_3}(C_{i|V_2,V_3}(C_{i|V_1}(u_i|v_1)|v_2)|v_3).
\]

For general \( p \), we have

\[
C_{i|V_1,\ldots,V_p}(u_i|v_1,\ldots,v_p) = C_{i|V_p,V_1,\ldots,V_{p-1}} \left( \cdots C_{i|V_2,V_1} \left( C_{i|V_1}(u_i|v_1)|v_2 \right) \cdots |v_p \right).
\]

The expression of \( C_{i|V_1,\ldots,V_k}(u_i|v_1,\ldots,v_p) \) in terms of \( C_{i|V_1}(u_i|v_1) \), the only term that contains \( v_1 \), allows us to obtain the corresponding extreme value factor copula using a similar approach for the 1- and 2-factor extreme value copulas\(^6\). The general \( p \)-factor copula is

\(^6\)This assumes that the linking copulas \( C_{i|V_k;V_1,\ldots,V_{k-1}}(u_i|v_k) \) do not involve \( v_1,\ldots,v_{k-1} \), which is often the simplifying assumption used in vine copula construction.
and the extreme value limit is
\[ b(w_1, \ldots, w_d) = \lim_{u \to 0^+} u^{-1} \int_0^1 \cdots \int_0^1 \prod_{i=1}^d C_{i|V_p:V_1,\ldots,V_{p-1}} \left( \cdots C_{i|V_2:V_1} \left( C_{i|V_1} (u w_i | v_1) | v_2 \right) \cdots | v_p \right) \, dv_1 \cdots dv_p, \]

assuming that the conditions for the dominated convergence theorem, or that for the tail dependence functions \( b_{i|V_1} \), are satisfied.

By setting \( m_i = C_{i|V_p:V_1,\ldots,V_{p-1}} \left( \cdots C_{i|V_2:V_1} \left( b_{i|V_1} (w_i | z_1) | v_2 \right) \cdots | v_p \right) \), we obtain the \( a(\cdot) \) function as
\[
a(w_1, \ldots, w_d) = \int_0^1 \cdots \int_0^\infty \left[ 1 - \prod_{i=1}^d (1 - m_i) \right] \, dz_1 dv_2 \cdots dv_p
= \int_0^1 \cdots \int_0^\infty \left[ 1 - \prod_{i=1}^d \left( 1 - C_{i|V_p:V_1,\ldots,V_{p-1}} \left( \cdots C_{i|V_2:V_1} \left( b_{i|V_1} (w_i | z_1) | v_2 \right) \cdots | v_p \right) \right) \right] \, dz_1 dv_2 \cdots dv_p,
\]

and the extreme value limit is \( C_{LEV}(u_1, \ldots, u_d) = \exp \{ -a(w_1, \ldots, w_d) \} \). Although quite flexible, numerical evaluation of the \( p \) integrals can be challenging even with Gaussian quadrature. Less complex subsets of the \( p \)-factor extreme value copula, such as the bi-factor model to be discussed below, can offer computational efficiency and may even improve interpretability of the model when used appropriately.

\section*{C Bi-factor extreme value copula}

The bi-factor model with \( G \) groups of non-overlapping variables is a special case of the general \( p \)-factor model with \( p = G + 1 \), and is often a reasonable assumption in practice when not all variables depend on every latent factor. Each observed variable is related to the common latent variable, denoted now as \( V_0 \). Each variable in group \( g \) (for \( g = 1, \ldots, G \)) is linked to latent variable \( V_g \). An example would be stocks in different sectors; a latent factor common to all stocks can be attributed to the overall state of the economy, while
other latent factors are sector-specific whose effects are more local to the sectors concerned. Table 3.1 shows a possible dependence structure among the observed and latent variables in a bi-factor model. The presence of a checkmark indicates dependence between the elements of that pair, conditional on the previous latent variables. For example, the checkmark in the cell linking \( U_4 \) and \( V_2 \) means that \( U_4|V_0, V_1 \) and \( V_2|V_0, V_1 \) are dependent. Note that each observed variable depends on exactly two latent variables, one of them being \( V_0 \).

<table>
<thead>
<tr>
<th>Observed variable</th>
<th>Latent variable accessor</th>
<th>Latent variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>( U_1 )</td>
<td>( V_0 )</td>
<td>( V_1 )</td>
</tr>
<tr>
<td>( U_2 )</td>
<td>( \checkmark )</td>
<td>( \checkmark )</td>
</tr>
<tr>
<td>( U_3 )</td>
<td>( \checkmark )</td>
<td>( \checkmark )</td>
</tr>
<tr>
<td>( U_4 )</td>
<td>( \checkmark )</td>
<td>( \checkmark )</td>
</tr>
<tr>
<td>( \vdots )</td>
<td>( \checkmark )</td>
<td>( \checkmark )</td>
</tr>
<tr>
<td>( U_d )</td>
<td>( \checkmark )</td>
<td>( \checkmark )</td>
</tr>
</tbody>
</table>

Table 3.1: An example of the dependence structure between observed and latent variables in a bi-factor model. A checkmark indicates dependence between the elements of that pair, conditional on the previous latent variables.

The latent variables \( V_0, V_1, \ldots, V_G \) are assumed to be independent. Two variables in group \( g \) are conditionally independent given \( V_0, V_g \) and two variables in different groups are conditionally independent given \( V_0 \). Suppose the indices for group \( g \) variables are \( k_g-1, \ldots, k_g \), for \( g = 1, \ldots, G \) where \( k_0 = 0 \) and \( k_G = d \). Then, \( U_{k_g-1}, \ldots, U_{k_g} \) (conditionally) depend only on \( V_0 \) and \( V_g \), \( g = 1, \ldots, G \). Note that, if \( U_i|V_0, \ldots, V_{g-1} \) and \( V_g|V_0, \ldots, V_{g-1} \) are independent (i.e., no checkmark in the \( U_i-V_g \) cell in Table 3.1), then \( C_{i|V_g;V_1,\ldots,V_{g-1}}(x|y) = x \) for \( x \in [0,1] \) as \( F_{U_i|V_1,\ldots,V_{g-1}} \) is unit uniform. This implies

\[
C_{i|V_G;V_0, V_1, \ldots, V_{G-1}}(\cdots C_{i|V_3;V_0}(C_{i|V_0}(u_i|v_0)|v_1) \cdots |v_G) = C_{i|V_G;V_0}(C_{i|V_0}(u_i|v_0)|v_G),
\]

where \( q_i \geq 1 \) is the index of the latent factor that \( U_i \) is dependent upon conditionally, i.e., \( q_i = g \) if and only if \( k_{g-1} + 1 \leq i \leq k_g \). Hence the bi-factor copula is

\[
C(u_1, \ldots, u_d) = \prod_{i=0}^{G+1} C_{i|V_{q_i};V_0}(C_{i|V_0}(u_i|v_0)|v_{q_i}) \, dv_0 \cdots dv_G
\]

\[
= \prod_{g=1}^{G} \left[ \prod_{i=k_g-1+1}^{k_g} C_{i|V_g;V_0}(C_{i|V_0}(u_i|v_0)|v_g) \, dv_g \right] dv_0.
\]

The \((G+1)\)-dimensional integral is thus decomposed into a 1-dimensional outer integral with a product of 1-dimensional integrals as integrand, so that numerically, this has the
complexity of a 2-dimensional nested integral. For example, consider the case where there are 3 latent factors and 7 observed variables such that, in addition to $V_0$, the first 3 observed variables depend on $V_1$ only while the rest depend on $V_2$ only. Then the bi-factor copula is

$$C_{\text{bifact}}(u_1, \ldots, u_7) = \int_0^1 \left[ \int_0^1 \prod_{i=1}^3 C_{i|V_1;V_0}(C_{i|V_0}(u_i|v_0)|v_1) \, dv_1 \right] \left[ \int_0^1 \prod_{i=4}^7 C_{i|V_2;V_0}(C_{i|V_0}(u_i|v_0)|v_2) \, dv_2 \right] \, dv_0,$$

whereas the full 3-factor copula is

$$C_{\text{full}}(u_1, \ldots, u_7) = \int_0^1 \int_0^1 \int_0^1 \prod_{i=1}^d C_{i|V_3;V_0}(C_{i|V_0}(u_i|v_0)|v_1) \, dv_0 \, dv_1 \, dv_2.$$

It can be seen that the bi-factor model is a simpler formulation. Since the bi-factor copula is a special case of the general $(G+1)$-factor copula model, so is the tail dependence function. We have

$$b(w_1, \ldots, w_d) = \int_0^1 \cdots \int_0^\infty \prod_{i=1}^d C_{i|V_0,q_i}(b_{i|V_0}(w_i|z_0)|v_{q_i}) \, dz_0 \, dv_1 \cdots dv_G$$

$$= \int_0^\infty \prod_{g=1}^G \left( \int_{i=k_{g-1}+1}^1 \prod_{i=1}^{k_g} C_{i|V_0,q_i}(b_{i|V_0}(w_i|z_0)|v_{q_i}) \, dv_{q_i} \right) \, dz_0,$$

and therefore

$$a(w_1, \ldots, w_d) = \int_0^1 \cdots \int_0^\infty \left[ 1 - \prod_{i=1}^d \left( 1 - C_{i|V_0,q_i}(b_{i|V_0}(w_i|z_0)|v_{q_i}) \right) \right] \, dz_0 \, dv_1 \cdots dv_G$$

$$= \int_0^\infty \left[ 1 - \prod_{g=1}^G \left( \int_{i=k_{g-1}+1}^1 \prod_{i=1}^{k_g} \left( 1 - C_{i|V_0,q_i}(b_{i|V_0}(w_i|z_0)|v_{q_i}) \right) \, dv_{q_i} \right) \right] \, dz_0,$$

and the extreme value factor copula is $C_{\text{LEV}}(u_1, \ldots, u_d) = \exp \{-a(w_1, \ldots, w_d)\}$. Again, using the above 7-variable example, we obtain the $b(\cdot)$ and $a(\cdot)$ functions as

$$b(w_1, \ldots, w_7) = \int_0^\infty \left[ \int_0^1 \prod_{i=1}^3 C_{i|V_1;V_0}(b_{i|V_0}(w_i|z_0)|v_1) \, dv_1 \right] \left[ \int_0^1 \prod_{i=4}^7 C_{i|V_2;V_0}(b_{i|V_0}(w_i|z_0)|v_2) \, dv_2 \right] \, dz_0;$$

$$a(w_1, \ldots, w_7) = \int_0^\infty \left[ 1 - \left( \int_0^1 \prod_{i=1}^3 C_{i|V_1;V_0}(b_{i|V_0}(w_i|z_0)|v_1) \right) \left( \int_0^1 \prod_{i=4}^7 C_{i|V_2;V_0}(b_{i|V_0}(w_i|z_0)|v_2) \right) \right] \, dz_0.$$
3.2.2 Extreme value limit of vine copulas

The extreme value limit of an arbitrary (truncated) vine copula usually involves a high-dimensional integral that depends on the number of variables, except for the case of a $p$-truncated C-vine whose extreme value limit is a $p$-dimensional integral. The $p$-factor copula model is simply a $p$-truncated C-vine rooted at the latent variable(s) and thus inherits this nice property. For an illustration, we derive the distribution function of the 1-truncated model is simply a $p$-truncated C-vine whose extreme value limit is a $p$-dimensional integral. Therefore, although a similar derivation of the extreme value limit applies to the general vine, it is computationally challenging to do likelihood inference. This is the motivation behind the structured Hüsler-Reiss model in Section 3.3.

\[
C(u_1, \ldots, u_d) = \int_0^{u_1} \cdots \int_0^{u_d} \prod_{j=2}^d c_{1j}(v_1, v_j) \, dv_d \cdots dv_1 = \int_0^{u_1} \prod_{j=2}^d \left[ \int_0^{u_j} c_{1j}(v_1, v_j) \, dv_j \right] \, dv_1
\]

and the distribution function of the 2-truncated C-vine rooted at variable 1 in the first tree and the edge $\{1, 2\}$ in the second tree:

\[
C(u_1, \ldots, u_d)
\]

\[
= \int_0^{u_1} \cdots \int_0^{u_d} \prod_{j=2}^d c_{1j}(v_1, v_j) \prod_{k=3}^d c_{2k;1} \left( C_{2|1}(v_2|v_1), C_{k|1}(v_k|v_1) \right) \, dv_d \cdots dv_1
\]

\[
= \int_0^{u_1} \int_0^{u_2} c_{12}(v_1, v_2) \prod_{k=3}^d \left[ \int_0^{u_k} c_{2k;1} \left( C_{2|1}(v_2|v_1), C_{k|1}(v_k|v_1) \right) \right] \, dv_2 \, dv_1
\]

\[
= \int_0^{u_1} \int_0^{u_2} c_{12}(v_1, v_2) \prod_{k=3}^d \left[ \int_0^{C_{k|1}(v_k|v_1)} c_{k|2;1} \left( C_{2|1}(v_2|v_1) \right) \, dz_k \right] \, dv_2 \, dv_1
\]

\[
= \int_0^{u_1} \int_0^{u_2} c_{12}(v_1, v_2) \prod_{k=3}^d C_{k|2;1} \left( C_{k|1}(u_k|v_1)|C_{2|1}(v_2|v_1) \right) \, dv_2 \, dv_1,
\]

where the marginal copula density $c_{2;1} \left( C_{2|1}(v_2|v_1) \right) = 1$ and the substitution $z_k = C_{k|1}(v_k|v_1)$ is applied so that $dz_k = c_{1k}(v_1, v_k) \, dv_k$. A similar derivation generalizes to a $p$-truncated C-vine. The separability of the integrand into a product of tractable integrals allows a simple form for the copula, as well as the $b(\cdot)$ and $a(\cdot)$ functions.

In complete generality, a regular vine copula has $O(d)$ intractable integrals. Section 3.15 of Joe (2014) has an example of the $d$-dimensional 1-truncated D-vine copula. The tail dependence function, obtained by applying (2.15) to the individual components of the integrand, is a $(d-2)$-dimensional integral. Therefore, although a similar derivation of the extreme value limit applies to the general vine, it is computationally challenging to do likelihood inference. This is the motivation behind the structured Hüsler-Reiss model in Section 3.3.
3.2.3 Bivariate dependence properties

The stable tail dependence function for the \((i, j)\) bivariate marginal distribution of a \(p\)-factor extreme value copula is

\[
a_{ij}(w_i, w_j) = \int_{0}^{1} \cdots \int_{0}^{\infty} (m_i + m_j - m_i m_j) \, dz_1 \cdots dv_p = w_i + w_j - b_{ij}(w_i, w_j),
\]

where \(m_i = C_{i|V_{p-1}} \cdots C_{i|V_1} (b_{i|V_1}(w_i|z_1)|v_2) \cdots |v_p)\) and \(b_{ij}\) is the bivariate marginal tail dependence function of the parent factor copula. This relationship allows us to extend some dependence results for the factor copula to its extreme value limit easily. The following list gives some bivariate dependence properties applicable to the extreme value factor copulas.

- **Concordance ordering.** A bivariate distribution \(F\) is more concordant than \(G\), written as \(F \succ_c G\), if \(F \geq G\) or equivalently \(\overline{F} \geq \overline{G}\) pointwise. Increasing in the concordance ordering means that a larger dependence parameter of a copula leads to stronger dependence, i.e., the copula becomes more concordant. For the 1-factor case, it is given in Krupskii and Joe (2013) that, assuming (a) the linking copula \(C_{j|V_1}\) is fixed and \(C_{j|V_1}\) is stochastically increasing (i.e., \(1 - C_{j|V_1}(u|v) = \mathbb{P}(U_j > u|V_1 = v)\) is increasing in \(v\) for all \(u \in (0, 1)\)), and (b) \(C_{i|V_1}\) increases in the concordance ordering, then the factor copula \(C_{ij}\) increases in the concordance ordering. By definition, its corresponding tail dependence function \(b_{ij}\) is then increasing in the parameter(s) of \(C_{i|V_1}\), meaning that the extreme value limit \(C_{EV} = u_i u_j \exp\{b_{ij}(w_i, w_j)\}\) increases in the concordance ordering.

- **1-Factor dependence structure.** Consider variables \(i, j, k\) linked to a latent variable \(V_1\) in the 1-factor copula in (2.2), and assume that the 1-factor copula has lower tail dependence. Suppose that each linking copula with \(V_1\) is stochastically increasing. If \(C_{i|V_1} \succ_c C_{j|V_1} \succ_c C_{k|V_1}\) are ordered in the concordance ordering \(\succ_c\), then the preceding item implies that \(C_{ij} \succ_c C_{ik} \succ_c C_{jk}\) for the bivariate margins of the 1-factor copula, and the same concordance ordering holds for the corresponding bivariate margins of the extreme value limit. If the variables are indexed so that the strength of dependence (concordance) is decreasing as \(i\) increases from 1 to \(d\), then the bivariate dependence of the \((i, j)\) margin for the 1-factor copula and its extreme value limit decreases as \(i, j\) increase. This is the typical pattern of dependence for the 1-factor structure. By a matrix of empirical bivariate dependence measures, it is possible to assess whether the 1-factor structure is a good approximation. If not, one could consider 2- or bi-factor structures assuming there are plausible latent variables.
• **Dependence measures.** The tail dependence function of the extreme value factor copula is the same as that of the corresponding factor copula, and the tail dependence index is $b_{ij}(1,1)$. The Pickands dependence function is $B(w) = a_{ij}(w,1-w) = 1 - b_{ij}(w,1-w)$ and the extremal coefficient is $\vartheta = 2 - b_{ij}(1,1)$. The tail dependence function of the factor copula thus contains sufficient information on these quantities.

• **Dependence boundaries.** The full $d$-dimensional extreme value factor copula becomes the independence (resp. comonotonicity) copula when all linking copulas are tail independent (resp. comonotonicity). The bivariate marginal copula attains these limits when all linking copulas related to the variables concerned do. The extreme value limit is not joint independence if there is tail dependence for every linking copula connected to the first latent factor.

Therefore, to construct parametric extreme value copulas with factor structure and interpretable parameters, we start with factor copulas where the bivariate linking copulas are in parametric families that increase in concordance, cover the range of independence to comonotonicity, and are stochastically increasing. These properties are satisfied by 1-parameter families of bivariate copulas that are mentioned in the next subsection.

### 3.2.4 Examples of 1-factor and 2-factor extreme value copulas

Here we present examples of the 1- and 2-factor extreme value copulas. In each case $d$ is the number of observed variables. To distinguish between the roles of different copulas, we label the linking copulas, factor copulas and their associated (lower) extreme value factor copulas as $\tilde{C}$, $C$ and $C_{EV}$, respectively. A similar convention is applied to marginal and conditional tail dependence functions, so that $\tilde{b}$ are those for the linking copulas and $b$ the factor copulas.

#### A 1-factor with Dagum (inverse Burr) conditional tail dependence functions

The 1-factor extreme value copula is characterized by the conditional tail dependence function, and thus we name such copulas based on these functions. Let $b_{i|V_1}(w_i|z_1) = [1 + (w_i/z_1)^{-\delta_i}]^{-1/\delta_i - 1}$ with $\delta_i > 0$, $i = 1,\ldots,d$. This is a special case of the Dagum (Dagum (1975)) or inverse Burr distribution and can be derived as the lower conditional tail dependence function of the MTCJ copula, or the upper one for the Galambos copula with dependence parameters $\delta_i$ that increase in the concordance ordering.

In the following, we derive the various quantities assuming that the linking copulas $C_{i, V_1}$, $i = 1,\ldots,d$, are MTCJ:
• Linking copula:

\[ \tilde{C}_{i,V_1}(u_i, v_1) = (u_i^{-\delta_i} + v_1^{-\delta_i} - 1)^{-1/\delta_i} \]

where \( \delta_i > 0 \) controls the strength of dependence of the linking copula.

• Tail dependence function between \( U_i \) and \( V_1 \):

\[ \tilde{b}_{i,V_1}(w_i, z_1) = \lim_{u \to 0^+} u^{-1} \left( (uw_i)^{-\delta_i} + (uz_1)^{-\delta_i} - 1 \right)^{-1/\delta_i} = \left( w_i^{-\delta_i} + z_1^{-\delta_i} \right)^{-1/\delta_i} . \]

• Conditional tail dependence function:

\[ \tilde{b}_{i|V_1}(w_i|z_1) = \frac{\partial}{\partial z_1} \tilde{b}_{i,V_1}(w_i, z_1) = \left[ 1 + (w_i/z_1)^{-\delta_i} \right]^{-1/\delta_i-1} . \]

With \( \tilde{b}_{i|V_1} \) derived, we obtain the \( b(\cdot) \) and \( a(\cdot) \) functions for the factor copula as

\[ b(w_1, \ldots, w_d) = \int_0^\infty \prod_{i=1}^d \tilde{b}_{i|V_1}(w_i|z_1) \, dz_1 = \int_0^\infty \prod_{i=1}^d \left[ 1 + (w_i/z_1)^{-\delta_i} \right]^{-1/\delta_i-1} \, dz_1 ; \]
\[ a(w_1, \ldots, w_d) = \int_0^\infty \left[ 1 - \prod_{i=1}^d \left( 1 - \left[ 1 + (w_i/z_1)^{-\delta_i} \right]^{-1/\delta_i-1} \right) \right] \, dz_1 , \]

using (3.1) and (3.2). Note that this is valid as \( \tilde{b}_{i|V_1} \) is a proper distribution on \([0, \infty)\).

Finally, \( C_{EV}(u_1, \ldots, u_d) = \exp\left\{ -a(w_1, \ldots, w_d) \right\} \). Numerically stable evaluation of these types of integrals is discussed in Chapter 4.

B 1-factor with Burr (Singh-Maddala) conditional tail dependence functions

Alternatively, let \( \tilde{b}_{i|V_1}(w_i|z_1) = 1 - [(w_i/z_1)^{\theta_i} + 1]^{1/\theta_i-1} \) with \( \theta_i > 1, i = 1, \ldots, d \). This is a special case of the Burr Type XII (Burr (1942)) or Singh-Maddala (Singh and Maddala (1976)) distribution and can be derived as the upper conditional tail dependence function of the Gumbel or Joe/B5 (Joe (1993)) copulas with dependence parameters \( \theta_i \), again increasing in the concordance ordering. The following exposition uses reflected Gumbel linking copulas and the lower extreme value limit is derived.

• Linking copula:

\[ \tilde{C}_{i,V_1}(u_i, v_1) = \overline{C}_{Gum}(1 - u_i, 1 - v_1) \]
\[ = u_i + v_1 - 1 + \exp\left\{ - \left[ (-\log (1 - u_i))^{\theta_i} + (-\log (1 - v_1))^{\theta_i} \right]^{1/\theta_i} \right\} , \]

where \( \theta_i > 1 \) controls the strength of dependence of the linking copula, and \( \overline{C}_{Gum} \) is the survival function of the Gumbel copula.
• Tail dependence function between $U_i$ and $V$:

$$ \tilde{b}_{i,V}(w_i, z_1) $$

$$ = \lim_{u \to 0^+} u^{-1} \left[ uw_i + uz_1 - 1 + \exp \left\{ - \left[ (-\log (1 - uw_i))^\theta_i + (-\log (1 - uz_1))^\theta_i \right]^{1/\theta_i} \right\} \right] $$

$$ = w_i + z_1 + \lim_{u \to 0^+} \left[ u^{-1} \cdot \exp \left\{ -u \left( w_i^{\theta_i} + z_1^{\theta_i} \right)^{1/\theta_i} \right\} \right] = w_i + z_1 - \left( w_i^{\theta_i} + z_1^{\theta_i} \right)^{1/\theta_i}. $$

• Conditional tail dependence function:

$$ \tilde{b}_{i|V}(w_i|z_1) = \frac{\partial}{\partial z_1} \tilde{b}_{i,V}(w_i, z_1) = 1 - \left[ (w_i/z_1)^{\theta_i} + 1 \right]^{1/\theta_i - 1}. $$

The conditional tail dependence function $\tilde{b}_{i|V}(w_i|z_1)$ is again a proper distribution on $[0, \infty)$.

Using (3.1) and (3.2), the $b(\cdot)$ and $a(\cdot)$ functions for this factor copula are

$$ b(w_1, \ldots, w_d) = \int_0^\infty \prod_{i=1}^d \left( 1 - \left[ (w_i/z_1)^{\theta_i} + 1 \right]^{1/\theta_i - 1} \right) dz_1; $$

$$ a(w_1, \ldots, w_d) = \int_0^\infty \left( 1 - \prod_{i=1}^d \left[ (w_i/z_1)^{\theta_i} + 1 \right]^{1/\theta_i - 1} \right) dz_1, $$

and $C_{EV}(u_1, \ldots, u_d) = \exp \left\{ -a(w_1, \ldots, w_d) \right\}.$

C 2-factor with Dagum conditional tail dependence functions for factor 1

Here we provide an example of 2-factor extreme value copula with Dagum conditional tail dependence functions and dependence parameters $\delta_i$, $i = 1, \ldots, d$, for the first factor. From (3.3) and (3.4), the $b(\cdot)$ and $a(\cdot)$ functions are

$$ b(w_1, \ldots, w_d) = \int_0^1 \int_0^\infty \prod_{i=1}^d \bar{C}_{i|V_2;V_1} \left( \left[ 1 + (w_i/z_1)^{-\delta_i} \right]^{-1/\delta_i - 1} \left| v_2; \theta_i \right. \right) dz_1 dv_2; $$

$$ a(w_1, \ldots, w_d) = \int_0^1 \int_0^\infty \left( 1 - \prod_{i=1}^d \left[ 1 - \bar{C}_{i|V_2;V_1} \left( \left[ 1 + (w_i/z_1)^{-\delta_i} \right]^{-1/\delta_i - 1} \left| v_2; \theta_i \right. \right. \right) \right) dz_1 dv_2, $$

where $\bar{C}_{i|V_2;V_1}$ is the conditional distribution function of the linking copula between the $i$th observed variable and the second latent factor with dependence parameter $\theta_i$. Numerical techniques similar to the 1-factor case can be used to compute $a(\cdot).$
3.3 Structured Hüsler-Reiss model

In this section, we propose another class of parsimonious extreme value dependence models based on the multivariate Hüsler-Reiss copula (Hüsler and Reiss (1989)). This is an alternative to extreme value limits of vine copulas, which involve high-dimensional integrals even in bivariate margins and are computationally intractable. For parsimonious submodels, we impose a structure on the correlation matrix of the Hüsler-Reiss copula according to the desired dependence structure. Although this is proposed mainly for vine structures, the factor structures are also well covered. Note that parsimonious forms of the Hüsler-Reiss copula have been developed in the context of multivariate spatial extremes (see, e.g., Smith (1990); Davison et al. (2012)).

3.3.1 Hüsler-Reiss copula with parsimonious dependence

The Hüsler-Reiss copula (2.8) with parameters \( \delta_{ij} \) is derived as the extreme value limit of the multivariate Gaussian distribution, with the assumption that the pairwise correlations approach 1 when the sample size \( n \to \infty \) in the following fashion:

\[
[1 - \rho_{ij}(n)] \log n \to \delta_{ij}^{-2} \in (0, \infty), \quad 1 \leq i \neq j \leq d,
\]

(3.5)

where \( \Sigma(n) = (\rho_{ij}(n))_{1 \leq i,j \leq d} \) is the correlation matrix, dependent on \( n \), of a \( d \)-variate Gaussian distribution with zero mean and unit variance for all variables. Nikoloulopoulos et al. (2009) show that the Hüsler-Reiss copula can be obtained as the limit of the t-EV copula when the dispersion matrix \( \Sigma(\nu) = (\rho_{ij}(\nu)) \) is such that \( \rho_{ij}(\nu) = 1 - 2\delta_{ij}^{-2}/\nu \), with the limit \( \nu \to \infty \). This and (3.5) provide a link between the correlations \( \rho_{ij} \) of the underlying Gaussian variates and the parameters of the Hüsler-Reiss distribution \( \delta_{ij} \). We propose an extreme value model in which the \( \rho_{ij} \)'s are structured according to the dependence pattern assumed for the data. This structure is translated to the parameters \( \delta_{ij} \) and is respected during the model fitting procedure. Specifically, (3.5) suggests that \( \delta_{ij} = \gamma(1 - \rho_{ij})^{-1/2} \) for \( n \) large, where \( \gamma > 0 \) is a proportionality constant. Different structures can be applied to \( \rho_{ij} \), for example:

- **Factor structure.** For a given number of factors \( p \), define the \( d \times p \) matrix of parameters

\[
L = \begin{bmatrix}
\alpha_{11} & \alpha_{12} & \cdots & \alpha_{1p} \\
\alpha_{21} & \alpha_{22} & \cdots & \alpha_{2p} \\
\vdots & \vdots & \ddots & \vdots \\
\alpha_{d1} & \alpha_{d2} & \cdots & \alpha_{dp}
\end{bmatrix},
\]

(3.6)
where $\alpha_{ij} \in [-1, 1]$ is the parameter for the $i$th variable and the $j$th factor. Its role is analogous to the parameters in classical factor analysis. The correlations are related to the parameters via the relationship $\rho_{ij} = \sum_{k=1}^{p} \alpha_{ik} \alpha_{jk}$ for $i \neq j$, and $\rho_{ii} = 1$. For practical model fitting, the following alternative parametrization is more useful:

$$L = \begin{bmatrix}
\alpha_{11} & \alpha_{12}\sqrt{1 - \alpha_{11}^2} & \alpha_{13}\sqrt{(1 - \alpha_{11}^2)(1 - \alpha_{12}^2)} & \cdots & \alpha_{1p}\sqrt{\prod_{j=1}^{p-1} (1 - \alpha_{1j}^2)} \\
\alpha_{21} & \alpha_{22}\sqrt{1 - \alpha_{21}^2} & \alpha_{23}\sqrt{(1 - \alpha_{21}^2)(1 - \alpha_{22}^2)} & \cdots & \alpha_{2p}\sqrt{\prod_{j=1}^{p-1} (1 - \alpha_{2j}^2)} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\alpha_{d1} & \alpha_{d2}\sqrt{1 - \alpha_{d1}^2} & \alpha_{d3}\sqrt{(1 - \alpha_{d1}^2)(1 - \alpha_{d2}^2)} & \cdots & \alpha_{dp}\sqrt{\prod_{j=1}^{p-1} (1 - \alpha_{dj}^2)} 
\end{bmatrix}.$$

With this construction, each $\alpha_{ij}$ can algebraically take the range $[-1, 1]$ independently and the parametrization is better suited for numerical optimization. Refer to Section 6.16 of Joe (2014) for a partial correlation interpretation. Note that this model is different from the extreme value limit of factor copulas introduced in Section 3.2, with stable tail dependence functions being integrals that do not generally correspond to the sum of Gaussian distribution functions for the Hüsler-Reiss copula.

- **Vine/tree structure.** The number of parameters (or correlations) needed to specify a Markov tree is $d - 1$, each corresponding to an edge of the associated graphical model. A 2-truncated vine uses $d - 2$ additional partial correlation parameters to connect the $d - 1$ edges in the Markov tree, and similarly for higher-order truncated vines (see Section 3.9 of Joe (2014)). A $p$-truncated vine has $(d - 1) + (d - 2) + \cdots + (d - p) = p [d - (p + 1)/2]$ parameters, and so the number of parameters grows at the rate $O(d^2)$ for fixed $p$. The parameters can be labelled according to the Gaussian pair-copula it represents. For example, the parameters for a $d$-dimensional D-vine (linear in tree 1) are $\{\{\alpha_{12}, \alpha_{23}, \ldots, \alpha_{(d-1),d}\}, \{\alpha_{13:2}, \alpha_{24:3}, \ldots, \alpha_{(d-2),d:(d-1)}\}, \ldots, \{\alpha_{1:d-1:23,}(d-1)\}\}$. By construction of vine copulas, the $\alpha$’s are within $[-1, 1]$ and are algebraically independent.

With this relationship between the dependence parameters $\delta_{ij}$ and the correlation parameters $\rho_{ij}$ of the underlying Gaussian variables, it is possible to specify certain $\delta_{ij}$’s instead and retrieve the corresponding $\alpha$ parameters and/or $\gamma$. For example, with the 1-truncated D-vine, the correlation or dependence parameters for neighbouring variables are related as $\delta_{i,i+1} = \gamma (1 - \alpha_{i,i+1}^2)^{-1/2}$, $i = 1, \ldots, d - 1$. Hence, by specifying $\delta_{12}, \ldots, \delta_{d-1,d}$, we can obtain $\alpha_{i,i+1}$ given any fixed $\gamma$. In this situation, it is possible to specify an extra $\delta$ parameter for the estimation of $\gamma$. The parameter set $\{\alpha_{12}, \ldots, \alpha_{d-1,d}, \gamma\}$ determines the correlations and hence dependence parameters for the other pairs. A similar argument generalizes to
any single tree or 1-truncated vine, i.e., specifying the $\delta$ parameters corresponding to the variables linked in the first tree determines all other dependence parameters given a fixed $\gamma$. Likewise, for a $p$-truncated vine, one will need to specify the $\delta$ parameters corresponding to the variables of the bivariate linking copulas in the $p$ trees.

Such relationship is more complex for the factor model as every $\rho_{ij}$ depends on more than one $\alpha$ parameter. As an example, for the 1-factor model, we obtain the $\binom{d}{2}$ equations $\delta_{ij} = \gamma(1 - \alpha_i \alpha_j)^{-1/2}$ for $i < j$ using the parametrization (3.6). For a given $\gamma$, we can solve for the $\alpha$’s (unique up to sign) if $\delta_{12}, \ldots, \delta_{1d}$ and one other $\delta$ parameter is known. A $p$-factor model has a total of $pd + 1$ parameters in our parametrization, and can only be solved if the number of given dependence parameters is no more than this. Even so, the system of equations may not have real solutions, even if the $\delta$’s are compatible in the sense that they constitute a valid Hüsler-Reiss distribution.

The Hüsler-Reiss copula has been generalized to the t-EV copula (2.11) (Demarta and McNeil (2005); Nikoloulopoulos et al. (2009)) and their skewed counterparts (Padoan (2011)). A multivariate t copula with factor structure has been defined with the correlation matrix having factor structure (Klüppelberg and Kuhn (2009)). It is possible to impose a factor structure directly on the correlation matrix of a t-EV copula. However, the bivariate strength of dependence corresponding to a correlation parameter of zero depends on the degree of freedom $\nu$ used (Davison et al. (2012); Ribatet (2013)); it approaches the independence limit as $\nu \rightarrow \infty$. This may complicate interpretation of the model. Nevertheless, it may be useful when the dependence strengths among all variables are rather strong, so that the aforementioned boundary is not a major concern. The stock returns data example in Section 3.7 contains an illustration of the 1-factor t-EV model.

### 3.3.2 Bivariate dependence properties

Since the structured Hüsler-Reiss model is a subset of the unconstrained Hüsler-Reiss copula, its dependence properties are restricted by those of the latter. In particular, using (2.16), the bivariate tail dependence function is given by

$$b_{ij}(w_i, w_j) = w_i \Phi \left( \frac{\delta_{ij}}{2} \log \frac{w_i}{w_i} - \frac{1}{\delta_{ij}} \right) + w_j \Phi \left( \frac{\delta_{ij}}{2} \log \frac{w_i}{w_j} - \frac{1}{\delta_{ij}} \right),$$

which is symmetric in $(w_i, w_j)$ and thus the bivariate marginal copula is permutation symmetric. The tail dependence index is given by $b_{ij}(1, 1) = 2 \Phi(-1/\delta_{ij})$, and the extremal coefficient is $2 - b_{ij}(1, 1) = 2 \Phi(1/\delta_{ij})$. The Pickands dependence function is

$$B(w) = w \Phi \left( \frac{1}{\delta_{ij}} + \frac{\delta_{ij}}{2} \log \frac{w}{1 - w} \right) + (1 - w) \Phi \left( \frac{1}{\delta_{ij}} + \frac{\delta_{ij}}{2} \log \frac{1 - w}{w} \right).$$
The structured Hüsler-Reiss model is increasing in concordance ordering with respect to both \( \alpha \) and \( \gamma \): Increasing \( \alpha \) and/or \( \gamma \) leads to an increase in some or all of the \( \rho \)'s (and thus the \( \delta \)'s), leading to stronger dependence.

For the \( d \)-dimensional copula, independence is obtained when all \( \delta \) parameters are zero, i.e., when \( \gamma = 0 \). Note that, even when \( \rho_{ij} = -1 \) (the lower bound), \( \delta_{ij} \) will not be zero unless \( \gamma = 0 \). Therefore care must be exercised when interpreting the fitted correlation parameters. Meanwhile, the comonotonicity copula is the limit as \( \delta_{ij} \to \infty \) for every \((i, j)\), or equivalently when \( \rho_{ij} = 1 \) for every pair. For the factor model, this happens when all the parameters for the first factor are 1. For the vine model, this happens when all \( \alpha \) parameters are 1. Regardless of the dependence structure, we obtain the same limit by letting \( \gamma \to \infty \).

### 3.4 Comparison between the extreme value factor copula and the structured Hüsler-Reiss copula

Both the structured Hüsler-Reiss copula and the extreme value factor copula are potentially useful parsimonious models for multivariate extremes. In this section, we contrast some differences between the two classes of models.

- The extreme value factor copula model is suitable for data exhibiting a latent or observed factor structure, whereas the structured Hüsler-Reiss model is applicable to any parsimonious dependence structure that can be parametrized and represented in terms of the correlation matrix of a Gaussian distribution.

- Simulation of the extreme value factor copula can be approximated by taking the componentwise maxima of the corresponding factor copula. However, there is currently no simple way to simulate from the multivariate Hüsler-Reiss copula in general.

- The parameters between different linking copulas in the extreme value factor copula model are algebraically independent. For the Hüsler-Reiss model, the collection of parameters have to be such that all \( d \) constituent \((d-1)\)-dimensional correlation matrices in (2.9) are positive definite. It is however not difficult to implement parameter verification in numerical optimization procedures. Some parameters may approach the boundary, i.e., 1 or \(-1\), during model fitting of the structured Hüsler-Reiss model, especially when the sample size is small. If that occurs, one can fit again with such parameters fixed at the boundary.
• For the factor copula model, the strength of dependence is controlled by both the choice of linking copulas and the parameter values, while the dependence structure is fixed (i.e., factor). On the other hand, the Hüsler-Reiss model does not involve the choice of copula families. The dependence structure can be specified, and the parameter values affect the strength of dependence. The $\alpha$ parameters mentioned above control the strength of dependence for specific pairs and other variables linked to them, while the proportionality constant $\gamma$ controls the overall strength. Since $\gamma$ acts multiplicatively on all $\delta_{ij}$'s, increasing $\gamma$ leads to an increase in the strength of dependence among all variables, keeping other parameters constant.

• The Hüsler-Reiss copula with factor structure does not coincide with the class of extreme value factor copula in terms of the pairwise strength of dependence, but with careful choice of parameters, the pairwise extremal coefficients can be well approximated. In general, the approximation is better when the $\rho_{ij}$'s are close to 1; this is not surprising given the limit argument in deriving the Hüsler-Reiss copula.

• Model estimation, via pairwise likelihood to be described in the Section 3.5, is faster and generally more stable for the structured Hüsler-Reiss copula as Gaussian densities and distribution functions are relatively easy to evaluate. With the extreme value factor copula model, one must obtain the exponent function and its derivatives through numerical integration methods (see Chapter 4).

The models can also be compared through their pairwise dependence characteristics. We give a representative example for illustration. Table 3.2 displays the Spearman’s $\rho$ for each bivariate margin of different parsimonious models for 5 observed variables. The top left panel corresponds to the 1-factor copula with MTCJ linkages and dependence parameters $(\delta_1, \ldots, \delta_5) = (1, 2, 3, 4, 5)$ (dependence is stronger as the index increases), from which the factor structure is apparent in the sense that the strength of dependence is similarly ordered across columns and rows. This property is carried over to its extreme value limit shown in the top right panel. In this case, the Spearman’s $\rho$’s are somewhat stronger than its parent factor copula. This extreme value 1-factor copula can be closely approximated by the Hüsler-Reiss 1-factor structure with parameters $(\tanh^{-1} \alpha_{11}, \ldots, \tanh^{-1} \alpha_{15}, \gamma) = (4.25, 4.83, 5.19, 5.45, 5.67, 0.0297)$.

The bottom panel of Table 3.2 suggests that the dependence pattern of the structured Hüsler-Reiss copula follows that of the tree imposed on the underlying correlation matrix, with weaker dependence for variables farther apart in the tree. The matrix of pairwise Spearman’s $\rho$ for a linear tree dependence structure (1-truncated D-vine) with correlation parameters $\alpha$ all equal to 0.5 and $\gamma = 1$ is given in the bottom left panel, and it shows that
adjacent variables are the most correlated with dependence strength tapering off between variables that are farther apart. The bottom middle panel results from the same vine but with all $\alpha = 0.875$ and $\gamma = 0.5$ instead, so that the dependence between adjacent variables remains the same. We can see that a smaller $\gamma$ results in a more flexible coverage of the strength of dependence. Finally, the bottom right panel represents the dependence structure of a tree with variables 1 to 4 linearly connected, but variable 5 is connected to 3 instead. The parameters are all $\alpha = 0.875$ and $\gamma = 0.5$. Since variables 4 and 5 are symmetric, they have the same bivariate dependence with the other three variables. Because relative pairwise dependence strengths are preserved, one may thus apply the assumed structure directly to the proposed models. The magnitudes of the parameters, however, should not be interpreted directly. For example, even if $\alpha < 0$ implying possible negative values in the underlying correlation matrix, the actual correlation of the observed extremes is still positive, albeit with a smaller magnitude.

### 3.5 Statistical inference via composite likelihood methods

As mentioned in Section 2.2.2, it is impractical to consider the full likelihood for multivariate extreme value copulas as the number of terms in the density grows rapidly with the
dimension. In this section, we describe the use of composite likelihood methods (Section 2.5.1) for the statistical inference of the two classes of parsimonious extreme value copulas introduced in this chapter. In particular, we focus on the pairwise likelihood with equal weight for each bivariate marginal density.

We suggest using the two-stage estimation procedure (i.e., the inference functions for margins method, see Section 2.5.2) for model fitting, where the marginal distributions are first estimated for each variable using the generalized extreme value distribution. The observations are then transformed to unit uniform and are fitted by the copula.

For extreme value copulas, the bivariate marginal densities in (2.19) have the form
\[ c_{12}(u_1, u_2) = e^{-A(A^{(1)}A^{(2)} - A^{(12)})}/u_1 u_2, \]
where \( A^{(1)}, A^{(2)} \) and \( A^{(12)} \) are the partial derivatives of the exponent function \( A(w_1, w_2) \), with respect to the arguments according to the order indicated. The pairwise density for the Hüsler-Reiss copula involves only univariate Gaussian densities and distribution functions and is thus easy to compute (see Section 4.10 of Joe (2014)). For the 1- and 2-factor extreme value copulas, the partial derivatives of \( A \) with respect to the non-empty set \( S \subset \{1, 2\} \) are as follows, using (3.2) and (3.4):

\[
A_{1\text{-fact}}^{(S)} = \int_0^\infty (-1)^{|S|-1} \prod_{j \in S} b'_j|V_1(w_j|z_1) \prod_{i \notin S} \left[ 1 - b_i|V_1(w_i|z_1) \right] \, dz_1,
\]
\[
A_{2\text{-fact}}^{(S)} = \int_0^1 \int_0^\infty (-1)^{|S|-1} \left\{ \prod_{j \in S} \left[ c_{j,V_2|V_1} \left( b_{j|V_1}(w_j|z_1), v_2 \right) \cdot b'_j|V_1(w_j|z_1) \right] \right\} \prod_{i \notin S} \left[ 1 - c_{i|V_2|V_1} \left( b_{i|V_1}(w_i|z_1)|v_2 \right) \right] \, dz_1 \, dv_2,
\]
where \( b'_j|V_1(w_j|z_1) = \partial b_{j|V_1}(w_j|z_1)/\partial w_j \). Similarly, for \( p \geq 3 \), we have

\[
A_{p\text{-fact}}^{(S)} = \int_0^1 \cdots \int_0^\infty (-1)^{|S|-1} \left\{ \prod_{j \in S} \left[ \prod_{k=3}^p c_{j,V_k|V_1,...,V_{k-1}} \left( \cdots c_{j|V_2|V_1} \left( b_{j|V_1}(w_j|z_1)|v_2 \right) \cdots , v_k \right) \right] \cdot c_{j,V_2|V_1} \left( b_{j|V_1}(w_j|z_1), v_2 \right) \cdot b'_j|V_1(w_j|z_1) \right\} \prod_{i \notin S} \left[ 1 - c_{i|V_p|V_1,...,V_{p-1}} \left( \cdots c_{i|V_2|V_1} \left( b_{i|V_1}(w_i|z_1)|v_2 \right) \cdots |v_p \right) \right] \, dz_1 \, dv_2 \cdots dv_p.
\]

\(^7\)Standardization to unit Fréchet margin is also a popular choice in the extreme value literature. However, as we are fitting a copula model, the unit uniform is a more natural choice here.

\(^8\)Here the quantity \( \exp\{- A(w_1, w_2)\} \) assumes the role of \( G \), the min-stable survival function with unit exponential margins. The \( A(\cdot) \) function is identical to \( a(\cdot) \) in previous sections. In multivariate extreme value theory, it is customary to use the capital letter \( A \) to denote the exponent function; we adopt this convention here.
In most cases, $A^{(S)}$ has to be evaluated numerically. The pairwise likelihood reduces the accumulation of rounding errors, as it does not require as many terms as in a higher-dimensional marginal density. In Chapter 4, we discuss the issues on the numerical evaluation of $A$ and $A^{(S)}$, and stabilize (and speed up) such evaluations through Gaussian quadrature and transformations of the integrands. With these techniques, we can obtain the accuracy required for the estimation of standard errors through the sensitivity and variability matrices.

3.6 Simulation Study

We conduct a simulation study on the Dagum 1-factor extreme value factor model, based on the extreme value limit of a 1-factor MTCJ copula with parameters $\delta = (\delta_1, \ldots, \delta_d)$, with the following intentions: (a) to verify that the numerical evaluation of the likelihood of the model is valid; (b) to compare the statistical efficiency between full and composite (pairwise) likelihood estimation in dimension $d = 4$, and; (c) to show that the method of pairwise likelihood yields accurate model-based standard errors. Note that with $d \geq 5$, maximum likelihood with the $d$-dimensional density is impractical in terms of computational time and getting an analytic expression for the density. We consider 3 sets of parameters for $\delta$ with $d = 4$ and let $\zeta_i = \delta_i / (\delta_i + 2) \in (0, 1)$ be a transform of $\delta_i$ that corresponds to the Kendall’s $\tau$ value for a bivariate MTCJ copula with parameter $\delta_i$. The parameter sets are:

- **Weak dependence:** $\delta = (1, 1, 1, 1)$ or $\zeta = (\zeta_1, \zeta_2, \zeta_3, \zeta_4) = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$. This corresponds to a Kendall’s $\tau$ of 0.268 between variables of the extreme value factor copula, and an extremal coefficient of 1.667.

- **Strong dependence:** $\delta = (4, 4, 4, 4)$ or $\zeta = \left(\frac{2}{3}, \frac{2}{3}, \frac{2}{3}, \frac{2}{3}\right)$. This corresponds to a Kendall’s $\tau$ of 0.707 between variables of the extreme value factor copula, and an extremal coefficient of 1.227.

- **Mixed dependence:** $\delta = (1, 2, 3, 4)$ or $\zeta = \left(\frac{1}{3}, \frac{1}{2}, \frac{3}{7}, \frac{2}{7}\right)$. The pairwise Kendall’s $\tau$ and extremal coefficients implied by this parameter set are listed in Table 3.3.

<table>
<thead>
<tr>
<th>Pair</th>
<th>$\tau$</th>
<th>$\vartheta$</th>
<th>Pair</th>
<th>$\tau$</th>
<th>$\vartheta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1,2)</td>
<td>0.358</td>
<td>1.565</td>
<td>(2,3)</td>
<td>0.562</td>
<td>1.358</td>
</tr>
<tr>
<td>(1,3)</td>
<td>0.387</td>
<td>1.534</td>
<td>(2,4)</td>
<td>0.588</td>
<td>1.334</td>
</tr>
<tr>
<td>(1,4)</td>
<td>0.399</td>
<td>1.521</td>
<td>(3,4)</td>
<td>0.667</td>
<td>1.263</td>
</tr>
</tbody>
</table>

Table 3.3: Pairwise Kendall’s $\tau$ and extremal coefficient $\vartheta$ for the mixed dependence scenario.
Some scatterplots of the normal scores of the data simulated according to the three dependence patterns given above are in Figure 3.1.

To explore the finite-sample performance, for each set of parameters we consider sample sizes \( n = 100 \) and 500, and number of replications \( R = 1,000 \), where each simulated sample is fitted by both full and pairwise likelihood. In each case we compute the average values of the Kendall’s \( \tau \) of the estimated parameter and its associated estimated variance, obtained from the observed information matrix for the fitting using full likelihood and the sandwich estimator for the fitting using pairwise likelihood. The sample variance is also obtained from the estimated Kendall’s \( \tau \) for each fitting method and compared against the estimated variance. All variability estimates are reported in the following as standard errors.

Table 3.4 summarizes the results of the simulation, while Figure 3.2 displays the sampling distribution for the strong dependence scenario. Estimates are reported in terms of the Kendall’s \( \tau \) corresponding to the parameters to reduce the right-skewness of the sampling distributions. We observe that both estimators have small bias even with \( n = 100 \), and that the sampling distribution is skewed when sample size is small. From the efficiency estimates, we observe that the pairwise likelihood estimators are in general more variable than their full likelihood counterparts. The difference is especially pronounced when the variables are strongly dependent; an explanation is that there is a bigger loss of information by using pairwise likelihood when there is strong dependence between variables. As expected, an increase in the sample size leads to a decrease of the variability in all cases. With regard to the variability estimates, we observe that there is some departure between the model-based and sampling standard errors for both methods when the sample size is small, but not so when \( n = 500 \).

In summary, estimation using pairwise likelihood results in acceptable performance, and the model-based standard errors are an accurate representation of the sampling variability when the sample size is reasonably large. With small sample sizes, however, we note that a more stable estimate of the variance under composite likelihood may be obtained through resampling methods such as the jackknife (see, e.g., Zhao and Joe (2005)), but this will incur a significant computational cost since it requires refitting the model multiple times.

Finally, simulation based on the structured Hüsler-Reiss model is not attempted as the composite likelihood method is known to be feasible. In the context of spatial extremes with the Smith and Brown-Resnick parsimonious representation of the Hüsler-Reiss model, pairwise and triplewise likelihood estimators have been compared in Genton et al. (2011) and Huser and Davison (2013). The performance of pairwise likelihood estimators is satisfactory except for cases of the Smith model where spatial locations and the parametrization of the Hüsler-Reiss model leads to matrices that are singular or nearly singular.
Figure 3.1: Scatterplots of the normal scores of the data simulated from the Dagum 1-factor extreme value copula, with dependence parameters $\delta = (1, 1, 1, 1), (4, 4, 4, 4)$ and $(1, 2, 3, 4)$ for the weak, strong and mixed dependence scenarios, respectively.
<table>
<thead>
<tr>
<th>N</th>
<th>Full likelihood</th>
<th></th>
<th>Pairwise likelihood</th>
<th></th>
<th>Efficiency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \hat{\zeta}_1 )</td>
<td>( \hat{\zeta}_2 )</td>
<td>( \hat{\zeta}_3 )</td>
<td>( \hat{\zeta}_4 )</td>
<td>( \hat{\zeta}_1 )</td>
</tr>
<tr>
<td>100</td>
<td>.338</td>
<td>.342</td>
<td>.339</td>
<td>.337</td>
<td>.338</td>
</tr>
<tr>
<td></td>
<td>(.048)</td>
<td>(.048)</td>
<td>(.048)</td>
<td>(.047)</td>
<td>(.048)</td>
</tr>
<tr>
<td>500</td>
<td>.333</td>
<td>.335</td>
<td>.334</td>
<td>.333</td>
<td>.334</td>
</tr>
<tr>
<td></td>
<td>(.020)</td>
<td>(.021)</td>
<td>(.021)</td>
<td>(.020)</td>
<td>(.021)</td>
</tr>
</tbody>
</table>

Weak dependence: \( \delta = (1, 1, 1, 1) \) and \( \zeta = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}) \)

<table>
<thead>
<tr>
<th>N</th>
<th>Full likelihood</th>
<th></th>
<th>Pairwise likelihood</th>
<th></th>
<th>Efficiency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \hat{\zeta}_1 )</td>
<td>( \hat{\zeta}_2 )</td>
<td>( \hat{\zeta}_3 )</td>
<td>( \hat{\zeta}_4 )</td>
<td>( \hat{\zeta}_1 )</td>
</tr>
<tr>
<td>100</td>
<td>.664</td>
<td>.667</td>
<td>.666</td>
<td>.665</td>
<td>.669</td>
</tr>
<tr>
<td></td>
<td>(.025)</td>
<td>(.025)</td>
<td>(.025)</td>
<td>(.025)</td>
<td>(.035)</td>
</tr>
<tr>
<td>500</td>
<td>.665</td>
<td>.665</td>
<td>.665</td>
<td>.665</td>
<td>.667</td>
</tr>
<tr>
<td></td>
<td>(.011)</td>
<td>(.011)</td>
<td>(.011)</td>
<td>(.011)</td>
<td>(.014)</td>
</tr>
</tbody>
</table>

Strong dependence: \( \delta = (4, 4, 4, 4) \) and \( \zeta = (\frac{2}{3}, \frac{2}{3}, \frac{2}{3}, \frac{2}{3}) \)

<table>
<thead>
<tr>
<th>N</th>
<th>Full likelihood</th>
<th></th>
<th>Pairwise likelihood</th>
<th></th>
<th>Efficiency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \hat{\zeta}_1 )</td>
<td>( \hat{\zeta}_2 )</td>
<td>( \hat{\zeta}_3 )</td>
<td>( \hat{\zeta}_4 )</td>
<td>( \hat{\zeta}_1 )</td>
</tr>
<tr>
<td>100</td>
<td>.336</td>
<td>.505</td>
<td>.607</td>
<td>.657</td>
<td>.334</td>
</tr>
<tr>
<td></td>
<td>(.032)</td>
<td>(.033)</td>
<td>(.038)</td>
<td>(.042)</td>
<td>(.031)</td>
</tr>
<tr>
<td>500</td>
<td>.335</td>
<td>.502</td>
<td>.602</td>
<td>.656</td>
<td>.333</td>
</tr>
<tr>
<td></td>
<td>(.014)</td>
<td>(.015)</td>
<td>(.017)</td>
<td>(.019)</td>
<td>(.014)</td>
</tr>
</tbody>
</table>

Mixed dependence: \( \delta = (1, 2, 3, 4) \) and \( \zeta = (\frac{1}{3}, \frac{2}{3}, \frac{2}{3}, \frac{2}{3}) \)

<table>
<thead>
<tr>
<th>N</th>
<th>Full likelihood</th>
<th></th>
<th>Pairwise likelihood</th>
<th></th>
<th>Efficiency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \hat{\zeta}_1 )</td>
<td>( \hat{\zeta}_2 )</td>
<td>( \hat{\zeta}_3 )</td>
<td>( \hat{\zeta}_4 )</td>
<td>( \hat{\zeta}_1 )</td>
</tr>
<tr>
<td>100</td>
<td>.054</td>
<td>.048</td>
<td>.051</td>
<td>.049</td>
<td>.055</td>
</tr>
<tr>
<td></td>
<td>(.020)</td>
<td>(.020)</td>
<td>(.020)</td>
<td>(.020)</td>
<td>(.020)</td>
</tr>
<tr>
<td>500</td>
<td>.022</td>
<td>.022</td>
<td>.022</td>
<td>.021</td>
<td>.037</td>
</tr>
<tr>
<td></td>
<td>(.010)</td>
<td>(.010)</td>
<td>(.010)</td>
<td>(.010)</td>
<td>(.015)</td>
</tr>
</tbody>
</table>

Table 3.4: Estimated Kendall’s \( \tau \) and standard errors for the Dagum 1-factor extreme value copula simulation. Standard errors are shown in brackets. Efficiency is calculated as the ratio of sampling variances between the estimators using full and pairwise likelihood, and is capped at 100%. The Kendall’s \( \tau \) transform \( \zeta = \delta/(\delta + 2) \) is used because the sampling distribution of \( \hat{\delta} \) is strongly skewed for the sample size \( n = 100 \).
Figure 3.2: Sampling distributions of the fitted Kendall’s $\tau$ using full and pairwise likelihood for the strong dependence scenario, with $\delta = (4, 4, 4, 4)$ for the Dagum 1-factor extreme value copula

### 3.7 Data examples

The proposed parsimonious extreme value copulas are applied to two data sets where there is a plausible latent factor: (a) the Fraser River flows data where a common source of streamflow and spatial dependence among gauge stations is likely, and (b) the United States stock returns data for selected listed companies in the same sector. For example (a), the tree dependence structure is also plausible from neighbouring stations along the river.
3.7.1 Fraser River flows data

The Fraser River is the largest river in British Columbia (Canadian Heritage Rivers System (2016)) which originates from the Rocky Mountains near the British Columbia–Alberta border. It continues downstream through Prince George and turns south, eventually reaching the Lower Mainland and discharging into the Strait of Georgia south of Richmond. The rate or volume of the flow along the river is highly dependent on the climate conditions in central and southeastern BC. Snow accumulates during the winter and melts in the spring and summer, causing an increase in river flows. This is compounded by possible rainstorms in the late spring and early summer along the Fraser River; it is the time of the year when most extreme streamflows are recorded.

To study the dependence characteristics of river flows along the Fraser River, we have downloaded such data at 8 hydrometric gauging stations from WaterOffice, Environment Canada; their locations are plotted in the left panel of Figure 3.3. The observations are annual maxima of daily streamflows until 2013. We check that the autocorrelations are insignificant. Most stations commenced operations in the 1950s; the only exceptions are Hope (6) whose records date back to 1912, and Mission (8) with earliest observation in 1965. The pairwise likelihood approach thus allows us to utilize information between 1950s and 1964 when data for most but one or two stations are available.

<table>
<thead>
<tr>
<th>Station</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1.000</td>
<td>0.749</td>
<td>0.659</td>
<td>0.570</td>
<td>0.546</td>
<td>0.580</td>
<td>0.694</td>
<td>0.768</td>
</tr>
<tr>
<td>2</td>
<td>0.749</td>
<td>1.000</td>
<td>0.781</td>
<td>0.696</td>
<td>0.701</td>
<td>0.752</td>
<td>0.713</td>
<td>0.663</td>
</tr>
<tr>
<td>1</td>
<td>0.659</td>
<td>0.781</td>
<td>1.000</td>
<td>0.925</td>
<td>0.886</td>
<td>0.895</td>
<td>0.842</td>
<td>0.671</td>
</tr>
<tr>
<td>4</td>
<td>0.570</td>
<td>0.696</td>
<td>0.925</td>
<td>1.000</td>
<td>0.934</td>
<td>0.897</td>
<td>0.817</td>
<td>0.669</td>
</tr>
<tr>
<td>5</td>
<td>0.546</td>
<td>0.701</td>
<td>0.886</td>
<td>0.934</td>
<td>1.000</td>
<td>0.932</td>
<td>0.855</td>
<td>0.715</td>
</tr>
<tr>
<td>7</td>
<td>0.580</td>
<td>0.752</td>
<td>0.895</td>
<td>0.897</td>
<td>0.932</td>
<td>1.000</td>
<td>0.923</td>
<td>0.758</td>
</tr>
<tr>
<td>6</td>
<td>0.694</td>
<td>0.713</td>
<td>0.842</td>
<td>0.817</td>
<td>0.855</td>
<td>0.923</td>
<td>1.000</td>
<td>0.920</td>
</tr>
<tr>
<td>8</td>
<td>0.768</td>
<td>0.663</td>
<td>0.671</td>
<td>0.669</td>
<td>0.715</td>
<td>0.758</td>
<td>0.920</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Table 3.5: Correlations of the normal scores for the Fraser River flows data. The station labels are the same as in Figure 3.3, but are reordered according to the sequence of stations from source to mouth.

The right panel of Figure 3.3 shows a pairwise scatterplot of the data, transformed to standard normal margins, while Table 3.5 gives the correlations of the normal scores. Dependence among stations is evident from the plot. Both factor and vine are plausible structural assumptions for the data. For a factor interpretation, the latent factor could be rainfall in the interior of BC: Higher rainfall contributes to increased river flows at all stations. Meanwhile, there is a natural ordering of the stations, namely from the source
area such as Red Pass (3) and McBride (2) to downstream area such as Hope (6) and Mission (8). A vine structure is thus also possible as stations close to each other are likely to exhibit stronger dependence than those far apart.

We conduct an exploratory analysis using the correlation matrix of normal scores after probability integral transforms to standard Gaussian margins. This entails fitting the transformed data using the classical factor analysis and a Gaussian vine. With 8 variables, we can fit at most a 4-factor model and a 5-truncated vine\(^9\). A summary of the results is given in Table 3.6. For each fitted model, we obtain several measures that are helpful to assess the distance between the sample correlation matrix \(R_{obs}\) and the fitted correlation matrix \(R_{mod}\). They include the maximum and average absolute difference across the elements of the matrices, and the \(D\) measure of discrepancy given by

\[
D = \log |R_{mod}| - \log |R_{obs}| + \text{tr} \left( R_{mod}^{-1} R_{obs} \right) - d,
\]

where \(d = 8\) is the number of variables (see, e.g., Chapter 15 of Mulaik (2009)). Smaller values for these measures mean that \(R_{mod}\) is closer to \(R_{obs}\). The values of the Akaike

\(^9\)The sequential minimum spanning tree algorithm is used for fitting vine models. It is computationally feasible to run an exhaustive search for the best vine on 8 variables, but this is not attempted here.
and Bayesian information criteria are also reported. We observe that the performance of the best factor model (4-factor) and the best truncated vine model (2-truncated vine) are similar in terms of the BIC, but the factor models generally have smaller values for the discrepancy measures than the vine models. There is no clear winner in this case.

The loadings for the 3-factor model (Table 3.7) suggest that the stations can be divided into three groups, i.e., (a) those close to the source (stations 2 and 3), (b) those along the midstream (stations 1, 4, 5 and 7), and (c) those located downstream (stations 6 and 8). This supports the assertion regarding spatial dependence among the stations. For the vine model, the best 1- and 2-truncated Gaussian vines turn out to be D-vines. The first tree can be drawn as 2–1–4–5–7–6–8–3, which follows exactly the order of the stations from upstream to downstream except Red Pass (3). This also hints at a spatial dependence structure where stations closer together exhibit stronger dependence.

<table>
<thead>
<tr>
<th>Model</th>
<th># param.</th>
<th>$D$</th>
<th>Discrepancy: $R_{obs}$ and $R_{mod}$ Average</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-factor</td>
<td>8</td>
<td>3.608</td>
<td>0.251</td>
<td>0.058</td>
<td>717.8</td>
</tr>
<tr>
<td>2-factor</td>
<td>15</td>
<td>1.830</td>
<td>0.215</td>
<td>0.028</td>
<td>634.6</td>
</tr>
<tr>
<td>3-factor</td>
<td>21</td>
<td>0.890</td>
<td>0.035</td>
<td>0.009</td>
<td>595.3</td>
</tr>
<tr>
<td>4-factor</td>
<td>26</td>
<td>0.419</td>
<td>0.022</td>
<td>0.003</td>
<td>579.6</td>
</tr>
<tr>
<td>1-truncated</td>
<td>7</td>
<td>1.832</td>
<td>0.339</td>
<td>0.056</td>
<td>618.8</td>
</tr>
<tr>
<td>2-truncated</td>
<td>13</td>
<td>1.297</td>
<td>0.392</td>
<td>0.047</td>
<td>601.5</td>
</tr>
<tr>
<td>3-truncated</td>
<td>18</td>
<td>1.102</td>
<td>0.337</td>
<td>0.029</td>
<td>600.9</td>
</tr>
<tr>
<td>4-truncated</td>
<td>22</td>
<td>1.016</td>
<td>0.296</td>
<td>0.021</td>
<td>604.2</td>
</tr>
<tr>
<td>5-truncated</td>
<td>25</td>
<td>0.841</td>
<td>0.306</td>
<td>0.018</td>
<td>600.6</td>
</tr>
</tbody>
</table>

Table 3.6: Fitting results for the Fraser River flows data using normal scores and Gaussian distribution

<table>
<thead>
<tr>
<th>Station</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.251</td>
<td><strong>0.907</strong></td>
<td>0.330</td>
</tr>
<tr>
<td>2</td>
<td>0.569</td>
<td>0.578</td>
<td>0.246</td>
</tr>
<tr>
<td>1</td>
<td><strong>0.830</strong></td>
<td>0.393</td>
<td>0.284</td>
</tr>
<tr>
<td>4</td>
<td><strong>0.878</strong></td>
<td>0.290</td>
<td>0.264</td>
</tr>
<tr>
<td>5</td>
<td><strong>0.853</strong></td>
<td>0.233</td>
<td>0.365</td>
</tr>
<tr>
<td>7</td>
<td><strong>0.807</strong></td>
<td>0.237</td>
<td>0.491</td>
</tr>
<tr>
<td>6</td>
<td>0.606</td>
<td>0.336</td>
<td><strong>0.718</strong></td>
</tr>
<tr>
<td>8</td>
<td>0.356</td>
<td>0.471</td>
<td><strong>0.761</strong></td>
</tr>
</tbody>
</table>

Table 3.7: Loadings of the Gaussian 3-factor model for the normal scores of the Fraser River flows data. Coefficients larger than 0.7 in absolute value are shown in boldface.

After conducting this exploratory investigation, we fit the data to the extreme value
factor copula models and the structured Hüsler-Reiss models. Each margin is first fitted by the GEV distribution and the observations are transformed to unit uniform through the probability integral transform using the estimated parameters. Diagnostic checks (not shown) suggest that the GEV distribution fits well to the 8 univariate series. We consider the Dagum and Burr 1-factor extreme value copulas; for the structured Hüsler-Reiss model, we fit the 1- and 2-factor structures as well as 1- and 2-truncated vines. The vine obtained from the Gaussian analysis above is used. Since there is a natural ordering of the stations, we additionally fit 1- and 2-truncated D-vines that take such information into consideration.

The results of the fitting are given in Table 3.8. For each fitted model we present the parameter estimates as well as their standard errors estimated from the Godambe information matrix, which is conditional on the GEV univariate estimation stage. We also obtain the composite likelihood information criterion (CLIC, Varin and Vidoni (2005)) and the composite likelihood Bayesian information criterion (CLBIC, Gao and Song (2010)), which are the composite likelihood analogues of the AIC and BIC. We observe that the two classes of models perform similarly in terms of the CLBIC. The Hüsler-Reiss model with a 2-truncated vine structure performs the best, where the vine is the one obtained from the minimum spanning tree algorithm (indicated in the table as “Gauss. vine”). However both the Burr 1-factor extreme value copula and the Hüsler-Reiss model with 1-truncated D-vine structure based on the location of the stations (indicated in the table as “Stn. order”) are close competitors. It is therefore useful to analyze the results for both classes of models.

Both 1-factor extreme value copula models suggest that sites 4, 5 and 7 have the strongest dependence with the latent factor. This is in agreement with the strong pairwise correlations among these three sites observed in Figure 3.3 and Table 3.5. These three stations are in the midstream of the Fraser river and this suggests that the river flows there are likely to be indicative of the overall flow. Meanwhile, sites 2, 3 and 8 have the weakest dependence with the latent factor. These stations are near the ends of the river and the flows there are more likely to be influenced by other factors.

For the Hüsler-Reiss models, there are cases where the $\alpha$ parameters reach the boundary (1 or $-1$). In this situation we optimize the pairwise likelihood with those parameter(s) fixed at the boundary. This can potentially have impact on the magnitude of $\text{tr}(H^{-1}J)$ and consequently the information criteria; investigation of such effects could be a future area of research and is not carried out in the present work. We transform the parameters of the Hüsler-Reiss factor models back to the loading matrix for easier interpretation (Table 3.9). Both factor models point to a contrast between the midstream stations (1, 4, 5, 7) and the source station 3. The 2-factor model suggests that downstream stations 6 and 8 are related to each other, but this is not apparent from the 1-factor model. The 1-
<table>
<thead>
<tr>
<th>Model</th>
<th>Estimates (SE)</th>
<th>( \theta_i, \delta_i \text{ or } \alpha_i )</th>
<th>( \gamma )</th>
<th>( \text{tr}(\mathbf{H}^{-1}\mathbf{J}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>EV1f (Burr)</td>
<td></td>
<td>( \begin{bmatrix} 3.928 &amp; 2.147 &amp; 1.854 &amp; 4.618 &amp; 5.237 &amp; 3.613 &amp; 5.484 &amp; 2.492 \end{bmatrix} )</td>
<td>( \begin{bmatrix} 0.489 &amp; 0.222 &amp; 0.200 &amp; 0.995 &amp; 2.076 &amp; 0.468 &amp; 1.690 &amp; 0.280 \end{bmatrix} )</td>
<td>( \begin{bmatrix} 25.24 &amp; -790 &amp; -1530 &amp; -1480 \end{bmatrix} )</td>
</tr>
<tr>
<td>EV1f (Dagum)</td>
<td></td>
<td>( \begin{bmatrix} 3.206 &amp; 1.422 &amp; 1.125 &amp; 3.884 &amp; 4.440 &amp; 2.895 &amp; 4.820 &amp; 1.770 \end{bmatrix} )</td>
<td>( \begin{bmatrix} 0.513 &amp; 0.219 &amp; 0.201 &amp; 1.088 &amp; 2.119 &amp; 0.502 &amp; 1.894 &amp; 0.284 \end{bmatrix} )</td>
<td>( \begin{bmatrix} 27.14 &amp; -788 &amp; -1521 &amp; -1467 \end{bmatrix} )</td>
</tr>
<tr>
<td>HR1f</td>
<td></td>
<td>( \begin{bmatrix} 0.690 &amp; -0.225 &amp; -1.000 &amp; 0.831 &amp; 0.870 &amp; 0.648 &amp; 0.870 &amp; 0.192 &amp; 1.855 \end{bmatrix} )</td>
<td>( \begin{bmatrix} 0.105 &amp; 0.179 &amp; (0.108) &amp; (0.119) &amp; (0.177) &amp; (0.139) &amp; (0.290) &amp; (0.180) \end{bmatrix} )</td>
<td>( \begin{bmatrix} 33.05 &amp; -785 &amp; -1503 &amp; -1437 \end{bmatrix} )</td>
</tr>
<tr>
<td>HR2f</td>
<td></td>
<td>( \begin{bmatrix} 0.060 &amp; 0.140 &amp; -0.009 &amp; 0.272 &amp; -0.081 &amp; -0.824 &amp; -0.406 &amp; -1.000 \end{bmatrix} )</td>
<td>( \begin{bmatrix} 0.258 &amp; 0.287 &amp; 0.303 &amp; 0.244 &amp; 0.182 &amp; 0.058 &amp; 0.161 &amp; — \end{bmatrix} )</td>
<td>( \begin{bmatrix} 35.21 &amp; -808 &amp; -1546 &amp; -1476 \end{bmatrix} )</td>
</tr>
<tr>
<td>HR1t (Gauss. vine)</td>
<td>T1</td>
<td>( \begin{bmatrix} 0.826 &amp; -0.074 &amp; 0.820 &amp; 0.845 &amp; 0.862 &amp; 0.834 &amp; 0.344 &amp; 1.491 \end{bmatrix} )</td>
<td>( \begin{bmatrix} 0.066 &amp; 0.714 &amp; 0.053 &amp; 0.074 &amp; 0.078 &amp; 0.082 &amp; 0.346 &amp; 0.252 \end{bmatrix} )</td>
<td>( \begin{bmatrix} 31.70 &amp; -795 &amp; -1527 &amp; -1464 \end{bmatrix} )</td>
</tr>
<tr>
<td>HR2t (Gauss. vine)</td>
<td>T1</td>
<td>( \begin{bmatrix} 0.841 &amp; 0.105 &amp; 0.811 &amp; 0.783 &amp; 0.788 &amp; 0.761 &amp; 0.297 \end{bmatrix} )</td>
<td>( \begin{bmatrix} 0.047 &amp; 0.295 &amp; 0.045 &amp; 0.068 &amp; 0.093 &amp; 0.081 &amp; 0.153 \end{bmatrix} )</td>
<td>( \begin{bmatrix} 34.23 &amp; -810 &amp; -1551 &amp; -1483 \end{bmatrix} )</td>
</tr>
<tr>
<td></td>
<td>T2</td>
<td>( \begin{bmatrix} 0.816 &amp; 0.100 &amp; 0.800 &amp; 0.782 &amp; 0.788 &amp; 0.760 &amp; 0.297 \end{bmatrix} )</td>
<td>( \begin{bmatrix} 0.047 &amp; 0.295 &amp; 0.045 &amp; 0.068 &amp; 0.093 &amp; 0.081 &amp; 0.153 \end{bmatrix} )</td>
<td>( \begin{bmatrix} 34.23 &amp; -810 &amp; -1551 &amp; -1483 \end{bmatrix} )</td>
</tr>
<tr>
<td>HR1t (Stn. order)</td>
<td>T1</td>
<td>( \begin{bmatrix} 0.625 &amp; 0.588 &amp; 0.892 &amp; 0.912 &amp; 0.902 &amp; 0.887 &amp; 0.887 &amp; 1.212 \end{bmatrix} )</td>
<td>( \begin{bmatrix} 0.129 &amp; 0.229 &amp; 0.057 &amp; 0.056 &amp; 0.056 &amp; 0.057 &amp; 0.067 &amp; 0.282 \end{bmatrix} )</td>
<td>( \begin{bmatrix} 29.44 &amp; -798 &amp; -1538 &amp; -1479 \end{bmatrix} )</td>
</tr>
<tr>
<td>HR2t (Stn. order)</td>
<td>T1</td>
<td>( \begin{bmatrix} 0.335 &amp; 0.380 &amp; 0.772 &amp; 0.798 &amp; 0.791 &amp; 0.818 &amp; 0.843 \end{bmatrix} )</td>
<td>( \begin{bmatrix} 0.149 &amp; 0.160 &amp; 0.082 &amp; 0.086 &amp; 0.064 &amp; 0.045 &amp; 0.046 \end{bmatrix} )</td>
<td>( \begin{bmatrix} 35.68 &amp; -809 &amp; -1547 &amp; -1476 \end{bmatrix} )</td>
</tr>
<tr>
<td></td>
<td>T2</td>
<td>( \begin{bmatrix} -0.832 &amp; -0.905 &amp; 0.027 &amp; 0.213 &amp; -0.461 &amp; -1.000 &amp; 1.743 \end{bmatrix} )</td>
<td>( \begin{bmatrix} 0.541 &amp; 0.352 &amp; 0.219 &amp; 0.263 &amp; 0.331 &amp; — &amp; 0.149 \end{bmatrix} )</td>
<td>( \begin{bmatrix} — &amp; — &amp; — &amp; — &amp; — &amp; — &amp; — \end{bmatrix} )</td>
</tr>
</tbody>
</table>

**Table 3.8:** Fitting results for the Fraser River flows data using multivariate extreme value copulas. The models are EV: Extreme value factor copula and HR: Structured Hüsler-Reiss copula, where 1f and 2f indicate the number of factors, and 1t and 2t the level of truncation for vines. The saturated Hüsler-Reiss copula (last row) is the unstructured distribution where all \( \frac{8^2}{2} = 28 \) parameters are allowed to vary independently. The SEs are based on the pairwise log-likelihood with the univariate parameters held fixed from the first stage GEV estimation. The last 4 columns are respectively: (1) Trace of the matrix \( \mathbf{H}^{-1}\mathbf{J} \) which can be considered as the penalty on the pairwise likelihood; (2) Negative pairwise log-likelihood; (3) Composite likelihood information criterion; (4) Composite likelihood Bayesian information criterion.
factor Hüsler-Reiss model has the highest negative pairwise log-likelihood as well as both CLIC and CLBIC, which are even higher than that of the saturated model; the 1-factor Hüsler-Reiss model may thus be inferior to the others for this data set. Meanwhile, the vine diagrams for the Hüsler-Reiss vine models are plotted in Figures 3.4 and 3.5. The former is fitted using the vine obtained from the Gaussian analysis, while the latter is fitted using a D-vine that follows the order of the stations from upstream to downstream. In Figure 3.4, we observe that the dependence strengths between stations 8 and 3, as well as stations 2 and 1, are quite a lot weaker than the other neighbours. In Figure 3.5, where the stations are in their natural order, we observe somewhat weaker dependence between stations 3 and 2, as well as 2 and 1. This pattern is in agreement with the observation in Table 3.5. It should be noted that the correlations implied by the truncated vine models are not directly comparable to the correlations of the normal scores of the original data, due to the limit argument in the derivation of the Hüsler-Reiss copula and the role of the $\rho_{ij}$’s as correlations of the underlying normal variates instead. This can be illustrated through the fitted coefficients of the second tree, where multiple strong negative partial correlations are obtained. The partial correlations of the observed normal scores are $(r_{13;2}, r_{24;1}, r_{15;4}, r_{47;5}, r_{56;7}, r_{78;6}) = (0.181, -0.110, 0.158, 0.204, -0.031, -0.603)$. These values are generally larger than the coefficients of the second tree in Figure 3.5, but the ranks are similar with the exception of $r_{13;2}$, for which the fitted coefficient is strongly negative.

Finally, we observe that the saturated Hüsler-Reiss model is competitive against some of the poorer parsimonious models. It appears that the dependence structure of our data could be more complex than a 1-factor or 1-truncated vine structure can handle, but the gain of using a parsimonious model still outweighs the loss of information for some structures.

<table>
<thead>
<tr>
<th>Station</th>
<th>Factor 1</th>
<th>Station</th>
<th>Factor 1</th>
<th>Factor 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>-1.000</td>
<td>3</td>
<td>-0.009</td>
<td>-1.000</td>
</tr>
<tr>
<td>2</td>
<td>-0.225</td>
<td>2</td>
<td>0.140</td>
<td>-0.195</td>
</tr>
<tr>
<td>1</td>
<td>0.690</td>
<td>1</td>
<td>0.060</td>
<td>0.720</td>
</tr>
<tr>
<td>4</td>
<td>0.831</td>
<td>4</td>
<td>0.272</td>
<td>0.962</td>
</tr>
<tr>
<td>5</td>
<td>0.870</td>
<td>5</td>
<td>-0.081</td>
<td>0.832</td>
</tr>
<tr>
<td>7</td>
<td>0.870</td>
<td>7</td>
<td>-0.406</td>
<td>0.841</td>
</tr>
<tr>
<td>6</td>
<td>0.648</td>
<td>6</td>
<td>-0.824</td>
<td>0.549</td>
</tr>
<tr>
<td>8</td>
<td>0.192</td>
<td>8</td>
<td>-1.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table 3.9: Loadings of the Hüsler-Reiss 1- and 2-factor models. Coefficients larger than 0.7 in absolute value are shown in boldface.
3.7.2 United States stock returns data

The second example we consider is on the returns of selected stocks traded on the US stock exchanges. Extreme value theory is widely used in the financial sector, one reason being that it provides the theoretical foundation for the modelling of value-at-risk (VaR) on stock returns. The VaR is an extreme quantile of the distribution for the returns; usually the quantile corresponding to a loss is of interest. Data for the extreme tails are often sparse, making empirical quantiles unreliable. Tsay (2010) illustrates the use of univariate extreme value theory on monthly minima of daily log returns. The log return for day $t$ is defined as $r_t = \log \left( \frac{P_t}{P_{t-1}} \right)$ where $P_t$ is the closing stock price on day $t$. When $P_t/P_{t-1}$ is close to 1, $r_t$ is close to the percentage change $(P_t - P_{t-1})/P_{t-1}$.

To highlight the use of the extreme value factor copula model, we selected 7 major stocks in the pharmaceutical sector. They are: (1) GlaxoSmithKline PLC (GSK); (2) Johnson & Johnson (JNJ); (3) Eli Lilly and Co (LLY); (4) Merck & Co., Inc. (MRK); (5) Mylan NV (MYL); (6) Novartis AG (NVS), and; (7) Pfizer Inc. (PFE). The observations are bimonthly minima of daily log returns between January 1997 and October 2016, for a
total of 119 observations. We choose bimonthly minima as the resulting series show weaker autocorrelations than monthly minima. Returns prior to 1997 are not used to reduce the effect of potential nonstationarity. For example, the advent of modern computing technology in the 1980s has seen a rapid growth in trading volume and frequency (and thus volatility), and is thought to be a potential factor underlying the 1987 stock market crash (Carlson (2007)). A joint extreme treatment of stock returns may be of use when the weights of different assets are to be selected to achieve an objective, such as maximizing expected returns, subject to restrictions on the VaR of the portfolio (Smith (2002)).

We negate the minimum returns in the subsequent analysis. They generally exhibit upper tail dependence, i.e., extreme losses tend to occur together. The pairwise scatterplots of the normal scores are shown in Figure 3.6, while the correlations are given in Table 3.10.

We then fit Gaussian factor and vine models to the normal scores; the results are shown in Table 3.11. Unlike the preceding example, here the factor models clearly outperform the vine models. The discrepancy between $R_{obs}$ and $R_{mod}$ for the 1-factor model with 7 parameters is comparable to that for the 2-truncated vine model with 11 parameters. This
Figure 3.6: Scatterplot of the normal scores for the US stocks data (negated minimum returns)

is also reflected by the superior AIC and BIC values for the factor models; they suggest that a 1- or 2-factor structure may be good enough. The results here are consistent with intuition — there is no obvious ordering or tree structure among the stocks, but the returns of stocks in the same sector can be thought of driven by some common factors such as the general economic environment and sector-specific drivers. The loadings of the 1- and 2-factor models are given in Table 3.12 and they seem to be consistent with the above explanation.

Next, we fit univariate GEV distribution to each margin and transform the observations to unit uniform before applying the extreme value factor copula and structured Hülsler-Reiss models. The truncated vines used for the Hülsler-Reiss model are those suggested from the
Table 3.12: Loadings of the Gaussian 1- and 2-factor models for the normal scores of the US stocks data. Coefficients larger than 0.7 in absolute value are shown in boldface.

<table>
<thead>
<tr>
<th>Stock</th>
<th>1-factor</th>
<th>2-factor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Factor 1</td>
<td>Factor 1</td>
</tr>
<tr>
<td>GSK</td>
<td>0.682</td>
<td>0.640</td>
</tr>
<tr>
<td>JNJ</td>
<td>0.823</td>
<td>0.685</td>
</tr>
<tr>
<td>LLY</td>
<td>0.650</td>
<td>0.498</td>
</tr>
<tr>
<td>MRK</td>
<td>0.610</td>
<td>0.488</td>
</tr>
<tr>
<td>MYL</td>
<td>0.597</td>
<td>0.561</td>
</tr>
<tr>
<td>NVS</td>
<td>0.771</td>
<td>0.784</td>
</tr>
<tr>
<td>PFE</td>
<td>0.735</td>
<td>0.336</td>
</tr>
</tbody>
</table>

Partial correlation vines from the correlation matrix of normal scores. We additionally fit a 1-factor t-EV model with 3 degrees of freedom. The parameter estimates, standard errors (conditional on the GEV univariate estimation stage) and various auxiliary quantities are reported in Table 3.13. For this data set, the 1-factor extreme value copulas fit better than the Hüsler-Reiss models, as can be seen from its smaller CLIC and CLBIC values. However, the 1-factor t-EV model appears to be superior. From the fitted parameters of the 1-factor models, the returns of Johnson & Johnson are more strongly correlated to the latent factor. Among the Hüsler-Reiss models, the 1-truncated vine (Figure 3.7) appears to be the best. Both Novartis and Pfizer connect to three other stocks due to their stronger dependence in extreme returns.

We run separate analyses on stocks in the banking and consumer staples sectors and find that the factor structure fits better than the vine structure in general. These results suggest that the factor structure for daily returns is a plausible modelling assumption among major stocks in the same sector, and hence diversification within one sector is less useful. With the fitted marginal and dependence models, it is then possible to arrive at a model-based VaR for a portfolio consisting of the stocks involved. The dependence model also allows us to gain insight on the joint tail behaviour, for all or a subset of the stocks.

![Diagram](image)

Figure 3.7: Diagram for the fitted Hüsler-Reiss 1-truncated vine model using the vine suggested from Gaussian analysis.
<table>
<thead>
<tr>
<th>Model</th>
<th>Estimates (SE)</th>
<th>$\gamma$</th>
<th>$\text{tr}(H^{-1}J)$</th>
<th>npllk</th>
<th>CLIC</th>
<th>CLBIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>EV1f (Burr)</td>
<td>F1 1.827 2.744 1.673 1.578 1.589 2.179 2.017</td>
<td>(0.185) (0.758) (0.188) (0.165) (0.134) (0.283) (0.247)</td>
<td>22.34</td>
<td>−331</td>
<td>−617</td>
<td>−555</td>
</tr>
<tr>
<td>EV1f (Dagum)</td>
<td>F1 1.093 1.857 0.936 0.830 0.853 1.470 1.277</td>
<td>(0.189) (0.613) (0.204) (0.179) (0.140) (0.288) (0.250)</td>
<td>24.29</td>
<td>−321</td>
<td>−594</td>
<td>−526</td>
</tr>
<tr>
<td>HR1f</td>
<td>F1 0.720 0.891 0.580 0.500 0.540 0.824 0.773 0.768</td>
<td>(0.224) (0.089) (0.447) (0.451) (0.475) (0.183) (0.245) (0.308)</td>
<td>24.81</td>
<td>−306</td>
<td>−562</td>
<td>−493</td>
</tr>
<tr>
<td>HR2f</td>
<td>F1 −0.148 −0.428 −0.293 −1.000 −0.292 −0.394 −0.485</td>
<td>(0.288) (0.261) (0.161) — (0.230) (0.259) (0.164)</td>
<td>28.09</td>
<td>−306</td>
<td>−555</td>
<td>−477</td>
</tr>
<tr>
<td></td>
<td>F2 0.768 0.856 0.508 0.000 0.459 0.785 0.692 0.775</td>
<td>(0.179) (0.046) (0.154) — (0.298) (0.108) (0.064) (0.151)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tEV1f ($\nu = 3$)</td>
<td>F1 0.790 0.915 0.749 0.720 0.692 0.859 0.841</td>
<td>(0.058) (0.041) (0.069) (0.087) (0.077) (0.041) (0.047)</td>
<td>17.31</td>
<td>−345</td>
<td>−656</td>
<td>−608</td>
</tr>
<tr>
<td>HR1t (Gauss. vine)</td>
<td>T1 0.641 0.600 0.565 0.406 0.237 0.289 0.918</td>
<td>(0.122) (0.130) (0.128) (0.164) (0.265) (0.283) (0.117)</td>
<td>21.26</td>
<td>−307</td>
<td>−572</td>
<td>−513</td>
</tr>
<tr>
<td>HR2t (Gauss. vine)</td>
<td>T1 0.593 0.452 0.742 0.382 0.680 0.477</td>
<td>(0.269) (0.347) (0.172) (0.363) (0.181) (0.290)</td>
<td>29.79</td>
<td>−309</td>
<td>−558</td>
<td>−475</td>
</tr>
<tr>
<td></td>
<td>T2 0.470 0.534 0.460 0.309 0.261 0.754</td>
<td>(0.153) (0.120) (0.242) (0.158) (0.213) (0.222)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saturated HR</td>
<td>There are a total of 21 parameters and are omitted here</td>
<td></td>
<td>34.88</td>
<td>−310</td>
<td>−549</td>
<td>−452</td>
</tr>
</tbody>
</table>

Table 3.13: Fitting results for the US stocks data using multivariate extreme value copulas. The models are EV: Extreme value factor copula, HR: Structured Hüsler-Reiss copula, and tEV: Structured t-EV copula, where 1f and 2f indicate the number of factors, and 1t and 2t the level of truncation for vines. The saturated Hüsler-Reiss copula (last row) is the unstructured distribution where all $\binom{21}{2} = 21$ parameters are allowed to vary independently. The SEs are based on the pairwise log-likelihood with the univariate parameters held fixed from the first stage GEV estimation. The last 4 columns are respectively: (1) Trace of the matrix $H^{-1}J$ which can be considered as the penalty on the pairwise likelihood; (2) Negative pairwise log-likelihood; (3) Composite likelihood information criterion; (4) Composite likelihood Bayesian information criterion.
Chapter 4

Numerical integration methods for extreme value factor copula models

This chapter is devoted to the technical details regarding the numerical integration involved in computing the bivariate densities of the 1-factor extreme value copula. The bivariate density for general extreme value copulas is given by

\[ c_{12}(u_1, u_2) = e^{-A(\frac{A^{(1)} - A^{(12)}}{u_1 u_2})} \]

where \( A \) is the exponent function and \( A^{(S)} \) is the partial derivative of \( A \) with respect to the element(s) in the set \( S \). For the 1-factor extreme value copula, we have the following results in Section 3.5:

\[ A = A(w_1, w_2) = \int_0^\infty \left[ 1 - (1 - b_{1|V}(w_1|x)) (1 - b_{2|V}(w_2|x)) \right] dx; \]

\[ A^{(1)} = A^{(1)}(w_1, w_2) = \int_0^\infty b'_{1|V}(w_1|x) (1 - b_{2|V}(w_2|x)) dx; \]

\[ A^{(2)} = A^{(2)}(w_1, w_2) = \int_0^\infty b'_{2|V}(w_2|x) (1 - b_{1|V}(w_1|x)) dx; \]

\[ A^{(12)} = A^{(12)}(w_1, w_2) = -\int_0^\infty b'_{1|V}(w_1|x)b'_{2|V}(w_2|x) dx, \]

where \( b' \) is the derivative of \( b \) with respect to the first argument, and \( w_i = -\log u_i, i = 1, 2 \). Although these integrals can be evaluated by numerical adaptive integration, they are not precise enough to allow accurate evaluation of the gradient and Hessian of the composite likelihood and thus the parameter and standard error estimates are unstable. This is partly due to the fact that the integrand often has a somewhat heavy tail, especially for \( A \), and partly because of the algebraic operations on \( A \) and its derivatives that magnify numerical imprecision.

A first step to enhancing stability is to transform \( w_1 \) and \( w_2 \) so that they always sum to 1. For any \( w_1 \) and \( w_2 \), obtain \( w^*_1 = w_1/(w_1 + w_2) \) and \( w^*_2 = w_2/(w_1 + w_2) \), which are used for
calculation. Because \( A_i \), \( A_i(w_1, w_2) = (w_1 + w_2)A_i(w_1^*, w_2^*) \), \( A_i(w_1, w_2) = A_i(w_1^*, w_2^*) \) for \( i = 1, 2 \), and \( A_{12}(w_1, w_2) = (w_1 + w_2)^{-1}A_{12}(w_1^*, w_2^*) \). Such transformation has the best effect when \( w_1 \) and \( w_2 \) are close to 0, i.e., when \( u_1 \) and \( u_2 \) are close to 1, since this will bring them away from the boundary where numerical difficulties are the most likely to happen. In the rest of the discussion, we assume that \( w_1 \) and \( w_2 \) have already been transformed in this manner.

To further improve the stability of integral computation, we split the integral into the regions \( x \in [0, 1) \) and \( x \in [1, \infty) \), guided by the difference in asymptotic behaviour of the integrands as \( x \to 0^+ \) and \( x \to \infty \). The Gauss-Laguerre quadrature (see, e.g., Salzer and Zucker (1949) and page 890 of Abramowitz and Stegun (1970)) is applied to these integrals. Gaussian quadrature methods attempt to approximate integrals of the form \( \int h(x)f(x) \, dx \), where \( f(x) \) is a density function, as \( \sum_{k=1}^{n_q} w_k h(x_k) \), where the \( x_k \)'s are the \( n_q \) quadrature nodes and the \( w_k \)'s are the corresponding quadrature weights (Stroud and Secrest (1966)). Such evaluation is exact if \( h \) is a polynomial of at most degree \( n_q \).

For integrand \( h(x) \) with the transformation \( x = e^{\eta y} \) for constant \( \eta > 0 \), we obtain

\[
\int_0^\infty h(x) \, dx = \int_0^1 h(x) \, dx + \int_1^\infty h(x) \, dx = \int_{-\infty}^0 h(e^{\eta y}) \cdot \eta e^{\eta y} \, dy + \int_0^\infty h(e^{\eta y}) \cdot \eta e^{\eta y} \, dy = \int_0^\infty h(e^{-\eta y}) \cdot \eta e^{-(\eta-1)y} \cdot e^{-y} \, dy + \int_0^\infty h(e^{\eta y}) \cdot \eta e^{(\eta+1)y} \cdot e^{-y} \, dy, \tag{4.1}
\]

so that each integral becomes an expectation of a standard exponential random variable for the application of the Gauss-Laguerre quadrature. Choosing \( \eta \) appropriately can stabilize integral evaluation; different choices of \( \eta \) can be made for the two integrals. One practical challenge associated with the transformation (4.1) is that \( e^{(\eta+1)y} \) can grow rapidly, especially when \( \eta \) is large. Numerical overflow is very likely to happen when the nodes are large enough, say \( y > 100 \). In this situation, a first-order approximation is applied directly to the term \( h(e^{\eta y}) \cdot \eta e^{(\eta+1)y} \). It is also found that, for the integral \( x \in [1, \infty) \) in the original integrand, stability can be further improved by another transformation that turns the limits of the integration into \([0, 1]\) and allows the use of the Gauss-Jacobi or Gauss-Legendre quadrature methods.

In the following, we provide the details on the transformations and discuss the resulting behaviour for the Burr and Dagum 1-factor extreme value copulas in Sections 4.1 and 4.2, respectively. Bivariate copulas are assumed throughout the chapter.
4.1 Burr 1-factor extreme value copula

For the 1-factor extreme value copula with Burr conditional tail dependence functions and parameters \( \theta_1, \theta_2 > 1 \), we have, for \( i = 1, 2 \), that

\[
\hat{b}_{i|V}(w_i|x) = 1 - \left[ \frac{w_i}{x} \right]^{\theta_i} + 1 \right]^{1/\theta_i - 1}; \quad \hat{b}'_{i|V}(w_i|x) = \frac{\theta_i - 1}{x} \left( \frac{w_i}{x} \right)^{\theta_i - 1} \left[ \left( \frac{w_i}{x} \right)^{\theta_i} + 1 \right]^{1/\theta_i - 2}.
\]

The \( A \) function. The \( A \) function is given by

\[
A(w_1, w_2) = \int_0^\infty \left( 1 - \left[ \left( \frac{w_1}{x} \right)^{\theta_1} + 1 \right]^{1/\theta_1 - 1} \left[ \left( \frac{w_2}{x} \right)^{\theta_2} + 1 \right]^{1/\theta_2 - 1} \right) \, dx. \tag{4.2}
\]

Let \( h(x) \) be the integrand of (4.2). First assume \( \theta_1 < \theta_2 \); results for the reverse case are readily obtained by interchanging the subscripts. We first consider the second integral of (4.1), corresponding to the region \( x \in [1, \infty) \). When \( x \to \infty \), we apply the Taylor series expansion to \( 1 - \hat{b}_{i|V}(w_i|x) \) and obtain

\[
h(x) \approx 1 - \prod_{i=1}^2 \left[ 1 + \left( \frac{1}{\theta_i} - 1 \right) \left( \frac{w_i}{x} \right)^{\theta_i} + \frac{1}{2} \left( \frac{1}{\theta_i} - 1 \right) \left( \frac{1}{\theta_i} - 2 \right) \left( \frac{w_i}{x} \right)^{2\theta_i} + o \left( x^{-2\theta_i} \right) \right]
= \left( 1 - \frac{1}{\theta_1} \right) \left( \frac{w_1}{x} \right)^{\theta_1} + o \left( x^{-\theta_1} \right). \tag{4.3}
\]

Hence the tail of \( h(x) \) behaves like \( x^{-\theta_1} \). The transformation \( \eta e^{\eta y} \cdot \eta e^{(\eta+1)y} \) therefore behaves like \( C_0 e^{(\eta+1-y\theta_1)y} \), where \( C_0 \) is a quantity\(^\text{10}\) not involving \( y \). This transformation is non-increasing at the limit (i.e., as \( y \to \infty \)) if and only if \( \eta + 1 - \eta \theta_1 \leq 0 \), or \( \eta \geq (\theta_1 - 1)^{-1} \).

By setting \( \eta = (\theta_1 - 1)^{-1} \), we can ensure that \( h(e^{\eta y} \cdot \eta e^{(\eta+1)y}) \) tends to a constant as \( y \to \infty \), while at the same time reducing the possibility of numerical overflow for the part \( e^{(\eta+1)y} \).

Note that this value of \( \eta \) also works when \( \theta_1 = \theta_2 \) as the analogous expansion of (4.3) has the same dominating term \( x^{-\theta_1} \).

Next, for the first integral of (4.1) which corresponds to the region \( x \in [0,1) \) and transformed \( y > 0 \), note that

\[
h(e^{-\eta y} \cdot \eta e^{-(\eta-1)y}) = \left( 1 - \left[ \left( w_1 e^{\eta y} \right)^{\theta_1} + 1 \right]^{1/\theta_1 - 1} \left[ \left( w_2 e^{\eta y} \right)^{\theta_2} + 1 \right]^{1/\theta_2 - 1} \right) \eta e^{-(\eta-1)y}
\leq \eta e^{-(\eta-1)y},
\]

as \( 0 \leq \left[ \left( w_1 e^{\eta y} \right)^{\theta_1} + 1 \right]^{1/\theta_1 - 1} \left[ \left( w_2 e^{\eta y} \right)^{\theta_2} + 1 \right]^{1/\theta_2 - 1} \leq 1 \). Using \( \eta = 1 \) will ensure that \( 0 \leq h(e^{-\eta y} \cdot \eta e^{-(\eta-1)y}) \leq 1 \), improving stability in the calculation of the integral.

\(^\text{10}\)Hereafter, we use \( C_i \)'s to denote various constants not depending on \( x \) or \( y \).
The first derivatives $A^{(1)}$ and $A^{(2)}$. The $A^{(1)}$ function is given by

$$A^{(1)}(w_1, w_2) = \int_0^\infty \left\{ \frac{\theta_1 - 1}{x} \left( \frac{w_1}{x} \right)^{\theta_1 - 1} \left[ \left( \frac{w_1}{x} \right)^{\theta_1} + 1 \right]^{1/\theta_1 - 2} \left[ \left( \frac{w_2}{x} \right)^{\theta_2} + 1 \right]^{1/\theta_2 - 1} \right\} \, dx. \quad (4.4)$$

Let $h_1(x)$ be this integrand and $\theta_1 \leq \theta_2$. For the region $x \in [1, \infty)$, the Taylor series expansion result above implies that, as $x \to \infty$,

$$h_1(x) \approx C_1 x^{-\theta_1} \left[ 1 + C_2 \left( \frac{w_1}{x} \right)^{\theta_1} + o \left( x^{-\theta_1} \right) \right] \left[ 1 + C_3 \left( \frac{w_2}{x} \right)^{\theta_2} + o \left( x^{-\theta_2} \right) \right] = C_4 x^{-\theta_1} + o \left( x^{-\theta_1} \right).$$

Hence the tail behaviour of $h_1(x)$ is the same as that of $h(x)$; we can thus use $\eta = (\theta_1 - 1)^{-1}$.

For $x \in [0, 1)$ with transformed $y > 0$, the transformation $h_1(e^{-\eta y}) \cdot \eta e^{-(\eta-1)y}$ is equal to

$$(\theta_1 - 1)^{-1} \cdot e^{\eta y} \cdot \left( e^{\eta y} \right)^{\theta_1} \cdot \left( e^{\eta y} \right)^{\theta_2} = e^{\eta y (\theta_1 - \theta_2 + 1)} y.$$ 

As $y \to \infty$, the tail of this expression behaves like

$$e^{\eta y} \cdot e^{\eta y (\theta_1 - \theta_2 + 1)} \cdot e^{\eta y (\theta_2 - 1)} \cdot e^{-(\eta-1)y} = e^{\eta y (1 - \theta_1 - \theta_2 + 1)} y.$$ 

Since $2 - \theta_1 - \theta_2 \leq 0$, using $\eta = 1$ will ensure that $h_1(e^{-\eta y}) \cdot \eta e^{-(\eta-1)y}$ is non-increasing in $y$ in the limit.

As for $A^{(2)}$, a swap of indices in (4.4) implies $h_2(x) = C_5 x^{-\theta_2} + o \left( x^{-\theta_2} \right)$, where $h_2(x)$ is the integrand for $A^{(2)}$. As $x \to \infty$ (i.e., $y \to \infty$), the transformation $h_2(e^{\eta y}) \cdot \eta e^{(\eta+1)y}$ has the same tail behaviour as $e^{(\eta+1)y \eta \theta_2}$. The choice $\eta = (\theta_1 - 1)^{-1}$ can again be used as $\eta + 1 - \eta \theta_2 = (\theta_1 - \theta_2)/(\theta_1 - 1) \leq 0$ by the assumption that $\theta_1 \leq \theta_2$. The transformed integrand is therefore non-increasing in $y$ in the limit.

For $x \in [0, 1)$, the same result (4.5) applies as the expression is symmetric in $\theta_1$ and $\theta_2$.

The second derivative $A^{(12)}$. The second derivative is given by

$$A^{(12)}(w_1, w_2) = - \int_0^\infty \prod_{i=1}^2 \left\{ \frac{\theta_i - 1}{x} \left( \frac{w_i}{x} \right)^{\theta_i - 1} \left[ \left( \frac{w_i}{x} \right)^{\theta_i} + 1 \right]^{1/\theta_i - 2} \right\} \, dx. \quad (4.6)$$

Let $h_{12}(x)$ be the integrand of (4.6). As $x \to \infty$, a similar Taylor series expansion results in

$$h_{12}(x) = C_6 x^{-(\theta_1 + \theta_2)} + o \left( x^{-(\theta_1 + \theta_2)} \right).$$

Hence the transformation $h_{12}(e^{\eta y}) \cdot \eta e^{(\eta+1)y}$ has the same tail behaviour as $e^{\eta y (1 - \theta_1 - \theta_2 + 1)} y$. Either $\eta = (\theta_1 - 1)^{-1}$ or 1 will work here; the former choice leads to a more rapid decline of the transformed integrand when $\theta_1 < 2$, but is more prone to numerical overflow for the
term $e^{(\eta+1)y}$. However, this can be overcome through a first-order approximation of the whole integrand.

As $x \to 0^+$, we have $h(x) \approx C_7 x^{-(\theta_1+\theta_2)} \cdot x^{2\theta_1-1} \cdot x^{2\theta_2-1} = C_7 x^{\theta_1+\theta_2-2}$. The transformation $h_{12}(e^{-\eta y}) \cdot \eta e^{-(\eta-1)y}$ has the tail behaviour of $e^{[\eta(1-\theta_1-\theta_2)+1]y}$, the same as that for the transformation for the part $x \in [1, \infty)$. The choices of $\eta$ in the preceding paragraph are therefore also applicable to the part $x \in [0, 1)$.

**First-order approximation of** $h(e^{\eta y}) \cdot \eta e^{(\eta+1)y}$ **as** $y \to \infty$. As noted above, when evaluating the integral for the part $x \in [1, \infty)$, the quantity $e^{(\eta+1)y}$ may be too big to handle numerically when $y$ is large. When this occurs, a first-order tail approximation of the whole transformed integrand is necessary. The following shows the derivation for each integrand.

- For $A$, the expression (4.3) for $h(x)$ implies that, if $\theta_1 < \theta_2$ and $x = e^y$ is large,

$$h(e^{\eta y}) \cdot \eta e^{(\eta+1)y} \approx \left(1 - \frac{1}{\theta_1}\right) \omega_1 e^{-\theta_1 y} \cdot \eta e^{(\eta+1)y} = \frac{\theta_1 - 1}{\theta_1} \eta e^{\eta(1-\theta_1)+1}y = \frac{\theta_1}{\theta_1}$$

with the choice of $\eta = (\theta_1 - 1)^{-1}$. If $\theta_1 = \theta_2 = \theta$, we have

$$h(x) \approx \left(1 - \frac{1}{\theta}\right) \left(\frac{\omega_1^\theta + \omega_2^\theta}{x^{\theta}}\right),$$

and the approximation becomes $h(e^{\eta y}) \cdot \eta e^{(\eta+1)y} \approx (\omega_1^\theta + \omega_2^\theta)/\theta$.

- For $A^{(1)}$, the expansion of $h_1(x)$ (integrand of (4.4)) is

$$h_1(x) = (\theta_1 - 1) \omega_1^{\theta_1-1} x^{-\theta_1} [1 + o(1)] [1 + o(1)] \approx (\theta_1 - 1) \omega_1^{\theta_1-1} x^{-\theta_1}.$$

The resulting approximation is

$$h_1(e^{\eta y}) \cdot \eta e^{(\eta+1)y} \approx (\theta_1 - 1) \omega_1^{\theta_1-1} \eta e^{\eta(1-\theta_1)+1}y = \omega_1^{\theta_1-1}$$

with the same choice of $\eta$ as that for approximating $A$.

- For $A^{(2)}$, the transformed integrand is decreasing in $y$. At the range of values where $e^{(\eta+1)y}$ overflows, $h_2(e^{\eta y}) \cdot \eta e^{(\eta+1)y}$ is very close to zero for all practical purposes if $\theta_1 < \theta_2$. If $\theta_1 = \theta_2 = \theta$, the approximation for the integrand of $A^{(1)}$ remains the same, while that for $A^{(2)}$ becomes $\omega_2^\theta$ instead of zero.

- For $A^{(12)}$, $h_{12}(e^{\eta y}) \cdot \eta e^{(\eta+1)y}$ has the same tail behaviour as $e^{[\eta(1-\theta_1-\theta_2)+1]y}$, where $h_{12}$ is the integrand of (4.6). Since $\eta(1-\theta_1-\theta_2) + 1 < 0$ when $\eta = (\theta_1 - 1)^{-1}$ or 1, the approximation is again zero. It remains zero even when $\theta_1 = \theta_2$. 75
In terms of programming implementation, for \(i = 1, 2\) we check whether \(\theta_i = \min(\theta_1, \theta_2)\). If this is true, we increment the approximation for \(A\) and \(A^{(i)}\) (for large \(y\)) by \(w_i^{\theta_i}/\theta_i\) and \(w_i^{\theta_i-1}\), respectively.

**Gauss-Jacobi quadrature.** Although a first-order approximation is sufficient in most cases, the facts that \(h(e^{\eta y}) \to 0\) (or \(h_i(e^{\eta y}) \to 0\)) and \(e^{(\eta+1)y} \to \infty\) as \(y \to \infty\) may cause numerical instability, especially when \(e^{(\eta+1)y}\) is very large but does not overflow. Here we present another transformation that may improve the evaluation of the integral for the region \(x \in [1, \infty)\).

Consider the transformation \(s = x^{-\theta_1}\), so that \(dx = -(s^{-(\theta_1+1)/\theta_1}/\theta_1) \, ds\). Let \(A_+, A_+^{(1)}, A_+^{(2)}\) and \(A_+^{(12)}\) be the respective integrals with lower limit equal to 1 instead of 0. Then we have

\[
A_+^{(1)} = \int_1^\infty \left\{ (\theta_1-1) w_1^{\theta_1-1} x^{-\theta_1} (w_1^{\theta_1} x^{-\theta_1} + 1)^{1/\theta_1-2} \left( w_2^{\theta_2} x^{-\theta_2} + 1 \right)^{1/\theta_2-1} \right\} dx

= \int_0^1 \left\{ \frac{\theta_1-1}{\theta_1} w_1^{\theta_1-1} s^{-1/\theta_1} (w_1^{\theta_1} s + 1)^{1/\theta_1-2} \left( w_2^{\theta_2} s^{\theta_2/\theta_1} + 1 \right)^{1/\theta_2-1} \right\} ds

= B \left( 1 - \frac{1}{\theta_1}, 1 \right) \cdot \frac{\theta_1-1}{\theta_1} \frac{w_1^{\theta_1-1}}{w_1^{\theta_1}} \left[ \left( w_1^{\theta_1} B_1 + 1 \right)^{1/\theta_1-2} \left( w_2^{\theta_2} B_1^{\theta_2/\theta_1} + 1 \right)^{1/\theta_2-1} \right]

= w_1^{\theta_1-1} \left[ \left( w_1^{\theta_1} B_1 + 1 \right)^{1/\theta_1-2} \left( w_2^{\theta_2} B_1^{\theta_2/\theta_1} + 1 \right)^{1/\theta_2-1} \right],
\]

where \(B(\cdot, \cdot)\) is the Beta function and \(B_1\) is a Beta \((1 - \theta_1^{-1}, 1)\) random variable. The expectation can be evaluated via Gauss-Jacobi quadrature. Using this transformation, the likelihood of numerical overflow or imprecision due to rounding is greatly reduced, as the range of integration is \([0, 1]\) rather than \([0, \infty)\) in the Gauss-Laguerre quadrature, and the integrand \((w_1^{\theta_1} B_1 + 1)^{1/\theta_1-2} \left( w_2^{\theta_2} B_1^{\theta_2/\theta_1} + 1 \right)^{1/\theta_2-1}\) is always in \([0, 1]\). Since there is no assumption about the relative magnitudes of \(\theta_1\) and \(\theta_2\) in the above derivation, we can simply swap the indices in (4.7) to obtain an expression for \(A_+^{(2)}\). The expectation is then taken with respect to a Beta \((1 - \theta_2^{-1}, 1)\) random variable instead.

Meanwhile, for the second derivative \(A_+^{(12)}\), we have

\[
A_+^{(12)} = -\int_1^\infty \left\{ (\theta_1-1) (\theta_2-1) w_1^{\theta_1-1} w_2^{\theta_2-1} x^{-(\theta_1+\theta_2)} (w_1^{\theta_1} x^{-\theta_1} + 1)^{1/\theta_1-2} \left( w_2^{\theta_2} x^{-\theta_2} + 1 \right)^{1/\theta_2-2} \right\} dx

= -\int_0^1 \left\{ \frac{\theta_1-1}{\theta_1} \frac{(\theta_2-1)}{\theta_2} w_1^{\theta_1-1} w_2^{\theta_2-1} s^{-(\theta_1+\theta_2)/\theta_1} (w_1^{\theta_1} s + 1)^{1/\theta_1-2} \left( w_2^{\theta_2} s^{\theta_2/\theta_1} + 1 \right)^{1/\theta_2-2} \right\} ds

= (1 - \theta_2) w_1^{\theta_1-1} w_2^{\theta_2-1} \left[ B_1^{\theta_2/\theta_1} \left( w_1^{\theta_1} B_1 + 1 \right)^{1/\theta_1-2} \left( w_2^{\theta_2} B_1^{\theta_2/\theta_1} + 1 \right)^{1/\theta_2-2} \right],
\]
which involves an integral with bounded integrand and range.

A similar derivation can be performed on \( A_+ \) as follows:

\[
A_+ = \int_1^\infty \left[ 1 - (w_1^\theta x^{-\theta_1} + 1)^{1/\theta_1-2} (w_2^\theta x^{-\theta_2} + 1)^{1/\theta_2-1} \right] dx
\]

\[
= \int_0^1 \frac{1}{\theta_1} s^{-1-1/\theta_1} \left[ 1 - (w_1^\theta s + 1)^{1/\theta_1-2} (w_2^\theta s^{\theta_2/\theta_1} + 1)^{1/\theta_2-1} \right] ds.
\]

However, this integral cannot be condensed into an expectation with finite range and finite integrand everywhere in that range. Experimentation with this transformation does not produce better results than the Gauss-Laguerre transformation or the adaptive integration function integrate in \( \mathbb{R} \).

**Behaviour of the exponent function and its derivatives at the independence limit.** The transformations above can help us discover the behaviour of \( A \) and its derivatives at the independence limit, i.e., when \( \theta_1, \theta_2 \to 1^+ \). These results can also serve to verify the validity of the transformations. To simplify calculations, we assume \( \theta_1 = \theta_2 = \theta \to 1^+ \) in the following (a similar result applies to distinct \( \theta \)'s by letting \( \theta_2 = k\theta_1 \) for \( k > 0 \)).

For the region \( x \in [0, 1) \), observe that \( 0 \leq h(x) \leq 1 \). For the first derivative, the integrand of \( A^{(1)} \) (see (4.4)) is

\[
h_1(x) = (\theta - 1)w_1^\theta -1 x^{-\theta} (w_1^\theta x^{-\theta} + 1)^{1/\theta_1-2} (w_2^\theta x^{-\theta} + 1)^{1/\theta_2-1}
\]

i.e., \( 0 \leq h_1(x) \leq (\theta - 1)w_1^\theta w_2^\theta \). A similar result holds for \( h_2(x) \), the integrand of \( A^{(2)} \). For the second derivative, the integrand of \( A^{(12)} \) is

\[
h_{12}(x) = (\theta - 1)^2 w_1^\theta -1 w_2^\theta x^{-2\theta} (w_1^\theta x^{-\theta} + 1)^{1/\theta_1-2} (w_2^\theta x^{-\theta} + 1)^{1/\theta_2-2}
\]

i.e., \( 0 \leq h_{12}(x) \leq (\theta - 1)^2 w_1^\theta w_2^\theta \). Since the bounds of these integrands are all integrable, the dominated convergence theorem applies. Hence

\[
\lim_{\theta \to 1^+} \int_0^1 h(x; \theta) \, dx = \int_0^1 \lim_{\theta \to 1^+} h(x; \theta) \, dx = 0;
\]

\[
\lim_{\theta \to 1^+} \int_0^1 h_i(x; \theta) \, dx = \int_0^1 \lim_{\theta \to 1^+} h_i(x; \theta) \, dx = 0, \quad i = 1, 2;
\]

\[
\lim_{\theta \to 1^+} \int_0^1 h_{12}(x; \theta) \, dx = \int_0^1 \lim_{\theta \to 1^+} h_{12}(x; \theta) \, dx = 0,
\]

where the \( h \) functions are written in this way to emphasize their dependence on \( \theta \).

For the region \( x \in [1, \infty) \), we have the following results:
Combining the limits for the two regions, we obtain

\[ A_+ = \int_0^1 \frac{1}{\theta} s^{-1-1/\theta} \left[ 1 - (w_1^\theta s + 1)^{1/\theta - 1} \right] ds \]

\[ = \frac{1}{\theta} B \left( 1 - \frac{1}{\theta}, 1 \right) \mathbb{E} \left[ \frac{1 - \left( w_1^\theta B_1 + 1 \right)^{1/\theta - 1} \left( w_2^\theta B_1 + 1 \right)^{1/\theta - 1}}{B_1} \right] \]

\[ = \mathbb{E} \left[ \frac{1 - \left( w_1^\theta B_1 + 1 \right)^{1/\theta - 1} \left( w_2^\theta B_1 + 1 \right)^{1/\theta - 1}}{(\theta - 1)B_1} \right], \quad (4.8) \]

where \( B_1 \sim \text{Beta}(1 - \theta^{-1}, 1) \). As \( \theta \) decreases towards 1, \( B_1 \) becomes increasingly concentrated at zero, eventually becoming a point mass there. If we expand the terms inside the expectation (4.8) about zero, we observe

\[ 1 - \left( w_1^\theta B_1 + 1 \right)^{1/\theta - 1} \left( w_2^\theta B_1 + 1 \right)^{1/\theta - 1} \]

\[ \frac{(\theta - 1)B_1}{1 - \left[ 1 + (\theta^{-1} - 1) w_1^\theta B_1 + o(B_1) \right] \left[ 1 + (\theta^{-1} - 1) w_2^\theta B_1 + o(B_1) \right]} \]

\[ = \frac{- (\theta^{-1} - 1) (w_1^\theta + w_2^\theta) B_1 + o(B_1)}{(\theta - 1)B_1} \to w_1 + w_2 \]

as \( B_1 \to 0^+ \). Hence, \( A_+(w_1, w_2) \to w_1 + w_2 \) as \( \theta \to 1^+ \).

• For \( A^{(1)} \), we obtained \( A_+^{(1)} = \frac{1}{\theta} \mathbb{E} \left[ \left( w_1^\theta B_1 + 1 \right)^{1/\theta - 2} \left( w_2^\theta B_1 + 1 \right)^{1/\theta - 1} \right] \) above. Since \( \left( w_1^\theta B_1 + 1 \right)^{1/\theta - 2} \left( w_2^\theta B_1 + 1 \right)^{1/\theta - 1} \to 1 \) as \( B_1 \to 0^+ \), \( A_+^{(1)}(w_1, w_2) \to w_1^\theta = 1 \) as \( \theta \to 1^+ \). This is also true for \( A_+^{(2)} \).

• For \( A^{(12)} \), we obtained \( A_+^{(12)} = (1 - \theta) \frac{1}{\theta} w_1^{\theta - 1} w_2^{\theta - 1} \mathbb{E} \left[ B_1 \left( w_1^\theta B_1 + 1 \right)^{1/\theta - 2} \left( w_2^\theta B_1 + 1 \right)^{1/\theta - 2} \right] \) above. As \( B_1 \to 0^+ \), \( B_1 \left( w_1^\theta B_1 + 1 \right)^{1/\theta - 2} \left( w_2^\theta B_1 + 1 \right)^{1/\theta - 2} \to 0 \). Hence \( A_+^{(12)}(w_1, w_2) \to 0 \) as \( \theta \to 1^+ \).

Combining the limits for the two regions, we obtain

\[ A(w_1, w_2) \to w_1 + w_2; \quad A^{(i)}(w_1, w_2) \to 1, \ i = 1, 2; \quad A^{(12)}(w_1, w_2) \to 0 \]

as \( \theta \to 1^+ \). The limit of \( A \) is consistent with the construction of the extreme value factor copula; if the linking copulas between the latent and observed variables are independence copulas, the resulting bivariate extreme value distribution of the observed variables \( U_1 \) and \( U_2 \) should be the product of the marginal distributions, i.e., \( F_{U_1U_2}(u_1, u_2) = u_1 u_2 = \exp \{ \log u_1 + \log u_2 \} = \exp \{ -(w_1 + w_2) \} \), so that \( A(w_1, w_2) = w_1 + w_2 \).
4.2 Dagum 1-factor extreme value copula

For the 1-factor extreme value copula with Dagum conditional tail dependence functions and parameters \(\delta_1, \delta_2 > 0\), we have, for \(i = 1, 2\), that

\[
b_{i|V}(w_i|x) = \left[ 1 + \left( \frac{w_i}{x} \right)^{-\delta_i} \right]^{-1/\delta_i - 1}; \quad b'_{i|V}(w_i|x) = \frac{1 + \delta_i}{x} \left( \frac{w_i}{x} \right)^{-\delta_i - 1} \left[ 1 + \left( \frac{w_i}{x} \right)^{-\delta_i} \right]^{-1/\delta_i - 2}.
\]

The \(A\) function. The \(A\) function is given by

\[
A(w_1, w_2) = \int_0^\infty \left\{ 1 - \left( 1 - \left[ 1 + \left( \frac{w_1}{x} \right)^{-\delta_1} \right]^{-1/\delta_1 - 1} \right) \left( 1 - \left[ 1 + \left( \frac{w_2}{x} \right)^{-\delta_2} \right]^{-1/\delta_2 - 1} \right) \right\} \, dx.
\]

First, we assume \(\delta_1 < \delta_2\) in the following discussion. Let \(h(x)\) be the integrand of (4.9). As \(x \to \infty\),

\[
h(x) \approx 1 - \left( 1 - \left[ 1 + \left( \frac{w_1}{x} \right)^{-\delta_1} \right]^{-1/\delta_1 - 1} \right) \left( 1 - \left[ 1 + \left( \frac{w_2}{x} \right)^{-\delta_2} \right]^{-1/\delta_2 - 1} \right)
\]

i.e., the tail of \(h(x)\) behaves like \(x^{-1-\delta_1}\). Consider the operations in (4.1). For the region \(x \in [1, \infty)\) and \(y > 0\), the transformation \(h(e^{\eta y}) \cdot \eta e^{(\eta+1)y}\) therefore behaves like \(e^{(1-\eta\delta_1)y}\), which is non-increasing at the limit (i.e., as \(y \to \infty\)) when \(1 - \eta\delta_1 \leq 0\), or \(\eta \geq \delta_1^{-1}\).

Setting \(\eta = \delta_1^{-1}\) implies that \(h(e^{\eta y}) \cdot \eta e^{(\eta+1)y}\) tends to a constant as \(y \to \infty\), and minimizes the possibility of numerical overflow for the part \(e^{(\eta+1)y}\). Note that this also works when \(\delta_1 = \delta_2\), and the subscripts can be interchanged if \(\delta_1 > \delta_2\).

A Taylor series expansion reveals that, as \(x \to 0^+\), we have

\[
h(x) = 1 + \sum_{i=1}^2 \left[ \frac{1 + \delta_i}{\delta_i} \left( \frac{w_i}{x} \right)^{-\delta_i} + o \left( x^{\delta_i} \right) \right] = 1 + \left( \frac{1 + \delta_1}{\delta_1} \right) \left( \frac{1 + \delta_2}{\delta_2} \right) w_1^{-\delta_1} w_2^{-\delta_2} x^{\delta_1 + \delta_2} + o \left( x^{\delta_1 + \delta_2} \right).
\]

Hence for the region \(x \in [0, 1)\) and \(y > 0\),

\[
h(e^{-\eta y}) \cdot \eta e^{-(\eta-1)y} = \left[ 1 + \left( \frac{1 + \delta_1}{\delta_1} \right) \left( \frac{1 + \delta_2}{\delta_2} \right) w_1^{-\delta_1} w_2^{-\delta_2} e^{-\eta(\delta_1 + \delta_2)y} + o \left( e^{-\eta(\delta_1 + \delta_2)y} \right) \right] \eta e^{-(\eta-1)y}.
\]

The choice of \(\eta\) is fairly flexible here and makes little difference to the evaluation of this integral. When \(\eta = 1\), the transformed integrand tends to 1 as \(y \to \infty\).
The first derivatives $A^{(1)}$ and $A^{(2)}$. The $A^{(1)}$ function is given by

$$A^{(1)}(w_1, w_2) = \int_0^\infty \left\{ \frac{1 + \delta_1}{x} \left( \frac{w_1}{x} \right)^{-\delta_1 - 1} \left[ 1 + \left( \frac{w_1}{x} \right)^{-\delta_1} \right]^{-1/\delta_1 - 2} \left( 1 - \left[ 1 + \left( \frac{w_2}{x} \right)^{-\delta_2} \right]^{-1/\delta_2 - 1} \right) \right\} dx.$$ 

Let $h_1(x)$ be this integrand. As $x \to \infty$,

$$h_1(x) \approx (1 + \delta_1) w_1^{-\delta_1 - 1} x^{\delta_1} \left[ 1 - \left( \frac{w_1}{x} \right)^{1+\delta_1} \right] \leq (1 + \delta_1) w_1^{-\delta_1 - 1}. \quad (4.11)$$

The tail behaviour is thus the same as that of $h(x)$ when $\delta_1 < \delta_2$, allowing the use of the same $\eta$ for the transformed integrand for the region $x \in [1, \infty)$.

As for $x \in [0, 1)$, we have

$$h_1(x) = (1 + \delta_1) w_1^{-\delta_1 - 1} x^{\delta_1} \left[ 1 - \frac{2\delta_1 + 1}{\delta_1} \left( \frac{w_1}{x} \right)^{-\delta_1} + o(x^{\delta_1}) \right] \left[ \frac{\delta_2 + 1}{\delta_2} \left( \frac{w_2}{x} \right)^{-\delta_2} + o(x^{\delta_2}) \right] = C_9 x^{\delta_1 + \delta_2} + o(x^{\delta_1 + \delta_2}).$$

Hence, as $y \to \infty$ (i.e., $x \to 0^+$ with $x = e^{-\eta y}$), the tail of the transformed integrand $h(e^{-\eta y}) e^{-(\eta-1)y}$ behaves like $e^{[1-\eta(\delta_1+\delta_2+1)]y}$. We can choose $\eta = 1$, so that $1 - \eta(\delta_1 + \delta_2 + 1) \leq 0$ for all $\delta_1, \delta_2 > 0$.

For $A^{(2)}$ with $x \in [1, \infty)$, exchanging indices reveals that $h_2(x)$ has the same tail behaviour as $x^{-\delta_2 - 1}$, so that the transformation $h(x^{\eta y}) \cdot e^{(\eta+1)y}$ behaves like $e^{\eta y}$. With the same choice $\eta = \delta_1^{-1}$, $1 - \eta \delta_2 < 0$ and the transformed integrand is decreasing in $y$ (or non-increasing, in the case $\delta_1 = \delta_2$) in the limit. For $x \in [0, 1)$, the derivation is the same as above, since the dominating term $x^{\delta_1 + \delta_2}$ is symmetric in $\delta_1$ and $\delta_2$. By interchanging subscript, a similar result applies when $\delta_1 > \delta_2$.

The second derivative $A^{(12)}$. The second derivative is given by

$$A^{(12)}(w_1, w_2) = -\int_0^\infty \prod_{i=1}^2 \left\{ \frac{1 + \delta_1}{x} \left( \frac{w_1}{x} \right)^{-\delta_1 - 1} \left[ 1 + \left( \frac{w_1}{x} \right)^{-\delta_1} \right]^{-1/\delta_1 - 2} \right\} dx.$$ 

Let $h_{12}(x)$ be this integrand. As $x \to \infty$,

$$h_{12}(x) \approx C_9 \prod_{i=1}^2 x^{\delta_i} \cdot x^{-2\delta_i - 1} = C_9 x^{-\delta_1 - \delta_2 - 2}.$$ 

Hence the transformation $h(e^{\eta y}) \cdot e^{(\eta+1)y}$ behaves like $e^{[1-\eta(\delta_1+\delta_2+1)]y}$. With $\eta = \delta_1^{-1}$, $1 - \eta(\delta_1 + \delta_2 + 1) = -(\delta_2 + 1)/\delta_1 < 0$ and the transformed integrand tends
to zero as $y \to \infty$. As in the case for Burr conditional tail dependence functions, $\eta = 1$ also works and the integrand also declines to zero as $y \to \infty$.

When $x \to 0^+$, we have

$$h_{12}(x) \approx \prod_{i=1}^{2} \left\{ (1 + \delta_i) w_i^{-\delta_i - 1} x^{\delta_i} \cdot \left[ 1 - \frac{2\delta_i + 1}{\delta_i} \left( \frac{w_i}{x} \right)^{-\delta_i} + o\left( x^{\delta_i} \right) \right] \right\} = C_{10} x^{\delta_1 + \delta_2} + o\left( x^{\delta_1 + \delta_2} \right).$$

The dominating term is the same as for $h_1(x)$, and the conclusion there applies to this case as well.

**First-order approximation of** $h(e^{\eta y}) \cdot \eta e^{(\eta+1)y}$ **as** $y \to \infty$. The following shows the appropriate first-order approximation of the transformed integrand for $x \in [1, \infty)$.

- For $A$, we obtain from (4.10) that, if $\delta_1 < \delta_2$, $h(e^{\eta y}) \cdot \eta e^{(\eta+1)y} \approx w_1^{\delta_1 + 1} \eta e^{(1-\eta \delta_1)y} = w_1^{\delta_1 + 1}/\delta_1$ as $y \to \infty$ with $\eta = \delta_1^{-1}$. If $\delta_1 = \delta_2 = \delta$, $h(x) \approx (w_1^{\delta_1 + 1} + w_2^{\delta_1 + 1})x^{-\delta_1 - 1}$ and the approximation becomes $h(e^{\eta y}) \cdot \eta e^{(\eta+1)y} \approx (w_1^{\delta_1 + 1} + w_2^{\delta_1 + 1})/\delta$.

- For $A^{(1)}$, we obtain from (4.11) that $h_1(e^{\eta y}) \cdot \eta e^{(\eta+1)y} \approx (1 + \delta_1) w_1^{\delta_1} \eta e^{(1-\eta \delta_1)y} = (1 + \delta_1) w_1^{\delta_1}/\delta_1$ with $\eta = \delta_1^{-1}$ if $\delta_1 < \delta_2$.

- For $A^{(2)}$, the transformed integrand is decreasing to zero as $y \to \infty$ when $\delta_1 < \delta_2$. If $\delta_1 = \delta_2 = \delta$, the first-order approximation becomes $(1 + \delta)w_2^{\delta_1}/\delta$.

- For $A^{(12)}$, the transformed integrand is decreasing to zero as $y \to \infty$ regardless of the magnitudes of $\delta_1$ and $\delta_2$, as the dominating term of $h_{12}(x)$ is $x^{-\delta_1 - \delta_2 - 2} = e^{-y(\delta_1 + \delta_2 - 2)y}$.

In terms of implementation, for $i = 1, 2$ we check whether $\delta_i = \min(\delta_1, \delta_2)$. If this is true, we increment the approximation for $A$ and $A^{(i)}$ (for large $y$) by $w_i^{\delta_i + 1}/\delta_i$ and $(1 + \delta_i)w_i^{\delta_i}/\delta_i$, respectively.

**Gauss-Jacobi quadrature.** The transformation $s = x^{-\delta_1}$ or $dx = -\left( s^{-(\theta_1+1)/\theta_1} / \theta_1 \right) ds$ can be applied to the derivatives of $A$, so that the integrals are transformed to the unit interval with a bounded integrand. Let $A_+, A_+^{(1)}, A_+^{(2)}$ and $A_+^{(12)}$ be the respective integrals.
with lower limit equal to 1 instead of 0. Then we have

\[ A_{+}^{(1)} = \int_{0}^{1} \left\{ \frac{1 + \delta_{1}}{\delta_{1}} w_{1}^{-\delta_{1}-1}s^{-1/\delta_{1}}(w_{1}^{\delta_{1}}s + 1)^{1-\delta_{1}} \left[ 1 - (1 + w_{2}^{-\delta_{2}}s^{-\delta_{2}/\delta_{1}})^{-1/\delta_{2}-1} \right] \right\} ds = \frac{1 + \delta_{1}}{\delta_{1}} w_{1}^{\delta_{1}} \left[ \int_{0}^{1} (w_{1}^{\delta_{1}}s + 1)^{-1/\delta_{1}-2} ds - \int_{0}^{1} (w_{1}^{\delta_{1}}s + 1)^{-1/\delta_{1}-2} \left[ 1 + w_{2}^{-\delta_{2}}s^{-\delta_{2}/\delta_{1}} \right]^{-1/\delta_{2}-1} ds \right] = 1 - (w_{1}^{\delta_{1}} + 1)^{-1/\delta_{1}-1} - \frac{1 + \delta_{1}}{\delta_{1}} w_{1}^{\delta_{1}} w_{2}^{\delta_{2}+1} \int_{0}^{1} s^{(\delta_{2}+1)/\delta_{1}} (w_{1}^{\delta_{1}}s + 1)^{-1/\delta_{1}-2} \left[ 1 + w_{2}^{\delta_{2}}s^{\delta_{2}/\delta_{1}} \right]^{-1/\delta_{2}-1} ds \]

where \( B_{2} \sim \text{Beta}(1 + \delta_{1}^{-1},1) \). The quantity inside the expectation is bounded between 0 and 1 as \( B_{2} \) takes values in the same range. Swapping indices result in an analogous formula for \( A_{+}^{(2)} \).

For \( A_{+}^{(12)} \), we obtain

\[ A_{+}^{(12)} = -\int_{0}^{1} \left\{ \frac{(1 + \delta_{1}) (1 + \delta_{2})}{\delta_{1}} w_{1}^{\delta_{1}} w_{2}^{\delta_{2}/\delta_{1}} (w_{1}^{\delta_{1}}s + 1)^{-1/\delta_{1}-2} \left[ 1 + w_{2}^{\delta_{2}}s^{\delta_{2}/\delta_{1}} + 1 \right]^{-1/\delta_{2}-2} \right\} ds = -(1 + \delta_{2}) w_{1}^{\delta_{1}} w_{2}^{\delta_{2}+1} \left[ B_{2}^{\delta_{2}/\delta_{1}} (w_{1}^{\delta_{1}}B_{2} + 1)^{-1/\delta_{1}-2} \left( w_{2}^{\delta_{2}}B_{2}^{\delta_{2}/\delta_{1}} + 1 \right)^{-1/\delta_{2}-2} \right]. \]

In this way, the term inside the expectation is bounded between 0 and 1.

For \( A_{+} \), we have

\[ A_{+} = \int_{0}^{1} \frac{1}{\delta_{1}} s^{-1/\delta_{1}-1} \left[ 1 - (1 + w_{1}^{-\delta_{1}}s^{-1})^{-1/\delta_{1}-1} \right] \left[ 1 - (1 + w_{2}^{-\delta_{2}}s^{-\delta_{2}/\delta_{1}})^{-1/\delta_{2}-1} \right] ds = \int_{0}^{1} \frac{1}{\delta_{1}} w_{1}^{\delta_{1}+1} (w_{1}^{\delta_{1}}s + 1)^{-1/\delta_{1}-1} ds + \int_{0}^{1} \frac{1}{\delta_{1}} w_{2}^{\delta_{2}+1} s^{\delta_{2}/\delta_{1}-1} (w_{2}^{\delta_{2}}s^{\delta_{2}/\delta_{1}} + 1)^{-1/\delta_{2}-1} ds - \int_{0}^{1} \frac{1}{\delta_{1}} w_{1}^{\delta_{1}+1} w_{2}^{\delta_{2}+1} s^{(1+\delta_{2})/\delta_{1}} (w_{1}^{\delta_{1}}s + 1)^{-1/\delta_{1}-1} (w_{2}^{\delta_{2}}s^{\delta_{2}/\delta_{1}} + 1)^{-1/\delta_{2}-1} ds = w_{1} \left[ 1 - (w_{1}^{\delta_{1}} + 1)^{-1/\delta_{1}} \right] + \frac{1}{\delta_{1}} w_{2}^{\delta_{2}+1} E \left[ B_{2}^{(\delta_{2}-1)/\delta_{1}} (w_{2}^{\delta_{2}}B_{2}^{\delta_{2}/\delta_{1}} + 1)^{-1/\delta_{2}-1} \right] - \frac{1}{\delta_{1}} w_{1}^{\delta_{1}+1} w_{2}^{\delta_{2}+1} E \left[ B_{2}^{\delta_{2}/\delta_{1}} (w_{1}^{\delta_{1}}B_{2} + 1)^{-1/\delta_{1}-1} (w_{2}^{\delta_{2}}B_{2}^{\delta_{2}/\delta_{1}} + 1)^{-1/\delta_{2}-1} \right]. \]

Unlike the Burr case, here we can write the transformed integral into expectations that involve a finite range and bounded integrand. For \( \delta_{1} \leq \delta_{2} \), \( B_{2}^{(\delta_{2}-1)/\delta_{1}} \) is bounded when \( B_{2} \in [0,1] \). If \( \delta_{1} > \delta_{2} \), we can swap the indices so that the random variables for which expectation is taken are all distributed as Beta\((1+\delta_{2}^{-1},1)\) instead, and that the \( B_{2}^{(\delta_{1}-1)/\delta_{2}} \) term inside one of the expectation operators will remain bounded.
Behaviour of the exponent function and its derivatives at the independence limit. To obtain the limits of the various integrals when \( \delta_1, \delta_2 \to 0^+ \), for simplicity we assume \( \delta_1 = \delta_2 = \delta \) (a similar result applies to distinct \( \delta \)'s by letting \( \delta_2 = k\delta_1 \) for \( k > 0 \)). For \( x \in [0, 1) \), it is obvious that \( 0 \leq h(x) \leq 1 \). Also, for the integrands of \( A^{(1)} \) and \( A^{(12)} \),

\[
\begin{align*}
h_1(x) &= (1 + \delta)w_1^{-\delta-1}x^\delta \left( 1 + w_1^{-\delta}x^\delta \right)^{-1/\delta -2} \left[ 1 - (1 + w_2^{-\delta}x^\delta)^{-1/\delta -1} \right] \\
&\leq (1 + \delta)w_1^{-\delta-1}; \\
h_{12}(x) &= (1 + \delta)^2w_1^{-\delta-1}w_2^{-\delta-1}x^{2\delta} \left( 1 + w_1^{-\delta}x^\delta \right)^{-1/\delta -2} \left( 1 + w_2^{-\delta}x^\delta \right)^{-1/\delta -2} \\
&\leq (1 + \delta)^2w_1^{-\delta-1}w_2^{-\delta-1}. 
\end{align*}
\]

The bounds are again integrable, and hence we can exchange the order of limit and integration. Since \( (1 + w_i^{-\delta}x^\delta)^{-1/\delta -1} \to 0 \) as \( \delta \to 0^+ \), \( i = 1, 2 \), we have

\[
\begin{align*}
\lim_{\delta \to 0^+} \int_0^1 h_1(x; \theta) \, dx &= \int_0^1 \lim_{\delta \to 0^+} h_1(x; \theta) \, dx = 0; \\
\lim_{\delta \to 0^+} \int_0^1 h_i(x; \theta) \, dx &= \int_0^1 \lim_{\delta \to 0^+} h_i(x; \theta) \, dx = 0, \quad i = 1, 2; \\
\lim_{\delta \to 0^+} \int_0^1 h_{12}(x; \theta) \, dx &= \int_0^1 \lim_{\delta \to 0^+} h_{12}(x; \theta) \, dx = 0.
\end{align*}
\]

Meanwhile, for \( x \in [1, \infty) \), the derivations are listed as follows:

- For \( A \), note that the first two terms of (4.12) are the same when \( \delta_1 = \delta_2 \), except that the first term contains \( w_1 \) while the second \( w_2 \). Hence

\[
\begin{align*}
A_+ = w_1 \left[ 1 - (w_1^\delta + 1)^{-1/\delta} \right] + w_2 \left[ 1 - (w_2^\delta + 1)^{-1/\delta} \right] \\
- \frac{1}{1 + \delta} (w_1w_2)^{\delta+1} E \left[ B_2 \left( w_1^\delta B_2 + 1 \right)^{-1/\delta -1} (w_2^\delta B_2 + 1)^{-1/\delta -1} \right], 
\end{align*}
\]

where \( B_2 \sim \text{Beta}(1 + \delta^{-1}, 1) \). When \( \delta \to 0^+ \), the first two terms of (4.13) tend to \( w_1 \) and \( w_2 \), respectively, while \( B_2 \) becomes increasingly concentrated at 1, eventually becoming a point mass there. At \( B_2 = 1 \), \( B_2 \left[ (w_1^\delta B_2 + 1) (w_2^\delta B_2 + 1) \right]^{-1/\delta -1} = \left[ (w_1^\delta + 1) (w_2^\delta + 1) \right]^{-1/\delta -1} \to 0 \) as \( \delta \to 0^+ \). Therefore \( A_+(w_1, w_2) \to w_1 + w_2 \) as \( \delta \to 0^+ \).

- For \( A^{(1)} \), we obtained

\[
A^{(1)}_+ = 1 - (w_1^\delta + 1)^{-1/\delta -1} - (w_2^\delta w_1^{\delta+1} E) \left[ B_2 \left( w_1^\delta B_2 + 1 \right)^{-1/\delta -2} (1 + w_2^\delta B_2)^{-1/\delta -1} \right].
\]

As \( \delta \to 0^+ \),\( 1 - (w_1^\delta + 1)^{-1/\delta -1} \to 1 \) and \( B_2 \left( w_1^\delta B_2 + 1 \right)^{-1/\delta -2} (1 + w_2^\delta B_2)^{-1/\delta -1} \to 0 \) as \( B_2 \to 1^- \). Hence \( A^{(1)}_+(w_1, w_2) \to 1 \) as \( \delta \to 0^+ \). A similar argument applies to \( A^{(2)}_+ \).
• For $A^{(12)}$, the result is

$$A^{(12)}_+ = -(1 + \delta) w_1^\delta w_2^\delta \left[ B_2 \left( w_1^\delta B_2 + 1 \right)^{-1/\delta - 2} \left( w_2^\delta B_2 + 1 \right)^{-1/\delta - 2} \right].$$

As $\delta \to 0^+$, $B_2$ converges to a point mass at 1, at which the value inside the expectation tends to zero. Hence $A^{(12)}_+(w_1, w_2) \to 0$ as $\delta \to 0^+$.

In summary, we obtain the same limits as in the Burr case.
Chapter 5

Extremal coefficient for extreme value copulas and its generalizations

The extremal coefficient $\vartheta$ is a measure of the strength of dependence used to study the dependence properties of extreme value copulas. As noted in Chapter 1, it can be interpreted as the number of effective independent variables of a multivariate copula (see (1.1)). A $d$-dimensional extreme value copula can be written as

\[ C(u_1, \ldots, u_d) = \exp\left\{ -A(-\log u_1, \ldots, -\log u_d) \right\}, \]

where the exponent function $A$ is homogeneous of order 1 (Section 2.2.2). Because of this property, one has

\[ C(u, \ldots, u) = \exp\left\{ \log u \cdot A(1, \ldots, 1) \right\} = u^{A(1, \ldots, 1)} \]

and thus $\vartheta = A(1, \ldots, 1)$. The possible range of $\vartheta$ is $[1, d]$; it is equal to 1 if $C$ is the comonotonicity copula and $d$ if $C$ is the independence copula.

Due to the importance of this summary measure for extreme value copulas, there exists many estimators of $\vartheta$ with different properties. For the rest of this chapter, we focus on the bivariate case as most of such estimators were proposed for bivariate copulas. Also, the matrix of extremal coefficients for all bivariate margins of a multivariate distribution is used in the diagnostic checks in Chapters 6 and 7, where model adequacy-of-fit is considered.

In this chapter, we review the properties of the different empirical estimators of $\vartheta$. We focus on a particular type of estimators known as the F-madogram, and demonstrate its relationship with general copulas. Through a generalization of the F-madogram, it is...
possible to construct a class of estimators for general copulas that put more weight in the tail portion, arriving at a tail-weighted measure of dependence. Properties of this estimator are given, together with a note on the potential extension to higher dimensions.

5.1 Empirical estimators of the extremal coefficient

We first review several empirical estimators of the extremal coefficient in the literature. The study of the dependence between variables of a bivariate extreme value distribution dates back to Sibuya (1960), but the first proposal of estimation of the dependence properties from data came much later in Pickands (1981), where the notion of the Pickands dependence function is introduced. Empirical estimators are generally developed for the estimation of the whole Pickands dependence function \( B(w) = A(w, 1 - w) \) for \( w \in [0, 1] \), where \( A \) is the exponent function (stable tail dependence function) of the extreme value distribution, written in copula form as

\[
C(u, v) = \exp \{-A(-\log u, -\log v)\}. \tag{5.1}
\]

The Pickands dependence function is also commonly expressed as the constituent of a min-stable survival function with unit exponential margins:

\[
\mathcal{G}(x, y) = \exp \left\{ -(x + y)B \left( \frac{x}{x+y} \right) \right\}, \tag{5.2}
\]

or that of a max-stable distribution function with unit Fréchet margins:

\[
G(x, y) = \exp \left\{ -\left( \frac{1}{x} + \frac{1}{y} \right) B \left( \frac{y}{x+y} \right) \right\}.
\]

The extremal coefficient is a multiple of the value at a point of the Pickands dependence function, namely \( \vartheta = 2B(1/2) \). The Pickands dependence function of an extreme value distribution satisfies the boundary condition \( \max(w, 1-w) \leq B(w) \leq 1 \) (which implies \( B(0) = B(1) = 1, B'(0) \in [-1, 0] \) with \( B'(w) = dB(w)/dw, B'(1) \in [0, 1] \) and \( B(1/2) \in [1/2, 1] \)) and is convex. This is not necessarily satisfied by the empirical estimator, and some estimators proposed in the literature are in fact modifications of the one in Pickands (1981) so that some of these conditions are automatically satisfied. Although these restrictions are not directly related to the estimation of the extremal coefficient, the modifications may result in a reduction in the variability of the estimators and are thus desirable.

Most of the earlier estimators were proposed assuming known margins (i.e., each margin is transformed to a standard distribution using the known marginal distribution). Although derivation of the asymptotic properties is easier, in most situations marginal distributions
are unknown. One variant uses the empirical counterpart, i.e., the adjusted or scaled ranks for each margin, as input. For bivariate observations \( Y_i = (Y_{i1}, Y_{i2})^\top, i = 1, \ldots, n \), the scaled ranks \( R_i = (R_{i1}, R_{i2})^\top \) are defined as

\[
R_{ik} = \frac{1}{n+1} \sum_{j=1}^{n} 1(Y_{jk} \leq Y_{ik}),
\]

for \( k = 1, 2 \). The denominator uses \( n+1 \) to bypass boundary issues; other choices exist but they do not affect asymptotic properties. It is important to differentiate between these two situations because the resulting estimators can have quite different asymptotic variances, as pointed out in a thorough study in Genest and Segers (2009). In the following, we first provide the definitions and properties for the known margin version of each estimator, and then proceed to the rank-based version. Throughout the discussion, we assume \( Y_i = (Y_{i1}, Y_{i2})^\top, i = 1, \ldots, n \), is an i.i.d. sequence of bivariate random vectors that have survival function (5.2) (i.e., exponential margins) and \( U_i = (U_{i1}, U_{i2})^\top, i = 1, \ldots, n \), is an i.i.d. sequence of bivariate random vectors from the extreme value copula (5.1). The subscript \( i \) is dropped when we refer to the distributional properties of the vectors.

### 5.1.1 Estimators assuming known margins

Several existing empirical estimators for the extremal coefficient, when the margins are assumed known, are:

1. **Pickands estimator** (Pickands (1981)). This is the first proposed estimator for the Pickands dependence function. The corresponding estimator for the extremal coefficient is

\[
\hat{\vartheta}_P^{-1} = \frac{1}{n} \sum_{i=1}^{n} \min (Y_{i1}, Y_{i2}) .
\]

The Pickands estimator for \( B(w) \) is not necessarily convex, and it does not satisfy the boundary conditions. The estimator \( \hat{\vartheta}_P \) need not be exactly 1 at comonotonicity.

2. **Deheuvels estimator** (Deheuvels (1991)). This estimator utilizes an adjustment to make sure the endpoint conditions \( B(0) = B(1) = 1 \) are satisfied. The extremal coefficient is estimated by

\[
\hat{\vartheta}_D^{-1} = \frac{1}{n} \sum_{i=1}^{n} \left[ \min (Y_{i1}, Y_{i2}) - \frac{1}{4} (Y_{i1} + Y_{i2}) + \frac{1}{2} \right] .
\]

Similar to the Pickands estimator, \( \hat{\vartheta}_D \) need not be 1 at comonotonicity.
3. **Hall and Tajvidi’s (HT) estimator** (Hall and Tajvidi (2000)). Instead of applying an additive correction, the HT estimator uses a multiplicative correction. This construction allows the resulting estimator for \( B(w) \) to satisfy the boundary condition
\[
\max(w, 1 - w) \leq B(w) \leq 1
\]
although still not necessarily convex. The extremal coefficient is estimated by
\[
\hat{\vartheta}_{HT}^{-1} = \frac{1}{n} \sum_{i=1}^{n} \min \left( \frac{Y_{i1}}{\overline{Y}_1}, \frac{Y_{i2}}{\overline{Y}_2} \right),
\]
where \( \overline{Y}_k = n^{-1} \sum_{i=1}^{n} Y_{ik} \) for \( k = 1, 2 \). With the multiplicative correction, \( \hat{\vartheta}_{HT} = 1 \) when \((Y_1, Y_2)\) is comonotonic.

4. **Capéraà-Fougères-Genest (CFG) estimator** (Capéraà et al. (1997)). The CFG estimator is based on the observation that the random variable
\[
Z = \log \frac{U_1}{U_1 U_2}
\]
has distribution function
\[
H(z) = z + z(1 - z) \frac{B'(z)}{B(z)}
\]
for \( z \in [0, 1) \). This implies that the Pickands dependence function can be written as
\[
B(w) = \exp \left\{ \int_0^w \frac{H(z) - z}{z(1 - z)} \, dz \right\} = \exp \left\{ - \int_w^1 \frac{H(z) - z}{z(1 - z)} \, dz \right\}.
\]
Replacing \( H(z) \) by the empirical version \( \hat{H}_n(z) \), the authors define their estimator as
\[
\log \hat{B}_{CFG}(w; p) = p(w) \int_0^w \frac{\hat{H}_n(z) - z}{z(1 - z)} \, dz - [1 - p(w)] \int_w^1 \frac{\hat{H}_n(z) - z}{z(1 - z)} \, dz
\]
for suitable choices of the weight function \( p(w) \). Beirlant et al. (2004) provide a much simpler method to motivate the CFG estimator, in the sense that
\[
\mathbb{E} \left[ \log \min \left( \frac{Y_1}{1 - w}, \frac{Y_2}{w} \right) \right] = - \log B(w) - \gamma,
\]
where \( \gamma = - \int_0^{\infty} \log(x)e^{-x} \, dx \) is Euler’s constant (see Segers (2007)). The CFG estimator can alternatively be written as
\[
\log \hat{B}_{CFG}(w; p) = - \frac{1}{n} \sum_{i=1}^{n} \left[ \log \min \left( \frac{Y_{i1}}{1 - w}, \frac{Y_{i2}}{w} \right) - p(w) \log Y_{i1} - [1 - p(w)] \log Y_{i2} \right].
\]
It satisfies the endpoint conditions if \( p \) is such as \( p(0) = 1 \) and \( p(1) = 0 \), and in practice the choice \( p(w) = 1 - w \) is satisfactory (Segers (2007)). This leads to the following estimator for the extremal coefficient:
\[
\log \hat{\vartheta}_{CFG} = - \frac{1}{n} \sum_{i=1}^{n} \left[ \log \min (Y_{i1}, Y_{i2}) - \frac{1}{2} \left( \log Y_{i1} + \log Y_{i2} \right) \right].
\]
5. **F-madogram estimator** (Cooley et al. (2006)). This estimator is inspired by the madogram used in spatial statistics as a measure of the dependence between two sites. The general empirical F-madogram is defined as:

\[
\hat{\nu}_\alpha = \frac{1}{2n} \sum_{i=1}^{n} |U_{i1}^\alpha - U_{i2}^\alpha|.
\]

(5.5)

For bivariate extreme value copulas, it can be shown that \(\mathbb{E}(\hat{\nu}_\alpha) = \alpha/(\alpha+1) - \alpha/(\alpha+\vartheta)\) and thus the estimator for the extremal coefficient is given by

\[
\hat{\vartheta}_\alpha = \frac{\alpha + \alpha(1 + \alpha)\hat{\nu}_\alpha}{\alpha - (1 + \alpha)\hat{\nu}_\alpha}.
\]

This estimator is also always 1 at comonotonicity. The original formulation by Cooley et al. (2006) has \(\alpha = 1\), while Naveau et al. (2009) generalize the estimator to the estimation of the whole Pickands dependence function with the powers in (5.5) being \(1-\beta\) and \(\beta\) for \(\beta \in (0, 1)\); their estimator for the extremal coefficient has \(\alpha = 1/2 = \beta\). More recently, Fonseca et al. (2015) consider the case where the powers in (5.5) can be any numbers \(\alpha, \beta > 0\).

### 5.1.2 Rank-based estimators

Because marginal distributions are rarely known, in practice one is not usually equipped with the observations \(Y_i\) or \(U_i\) that follow the stipulated marginal distributions. The rank-based approach uses instead the scaled ranks as a proxy of uniformly distributed observations in \([0, 1]\). Before we review the rank-based estimators, define the empirical copula as

\[
C_n(u_1, u_2) = \frac{1}{n} \sum_{i=1}^{n} 1(R_{i1} \leq u_1, R_{i2} \leq u_2)
\]

for \(u_1, u_2 \in [0, 1]\). This representation is useful in determining the behaviour of the estimators.

1. **Estimators by Pickands / Deheuvels / Hall and Tajvidi.** Genest and Segers (2009) show that the rank-based versions of the endpoint-corrected Pickands estimators, i.e., the Deheuvels and HT estimators, are asymptotically equivalent to the rank-based Pickands estimator. Letting \(\tilde{R}_{ik} = -\log R_{ik}\) for \(i = 1, \ldots, n\) and \(k = 1, 2\), the rank-based Pickands estimator for the extremal coefficient can be written as

\[
\hat{\vartheta}_{P,r}^{-1} = \frac{1}{n} \sum_{i=1}^{n} \min(\tilde{R}_{i1}, \tilde{R}_{i2}) = \frac{1}{2} \int_0^1 \frac{1}{u} C_n(u^{1/2}, u^{1/2}) \, du.
\]
2. Capéraà-Fougères-Genest (CFG) estimator. Genest and Segers (2009) also show that the rank-based, endpoint-corrected version of the CFG estimator is asymptotically equivalent to the uncorrected one, i.e., the direct empirical analogue of (5.3). For asymptotic properties, it is easier to work with the uncorrected version, with corresponding estimator for the extremal coefficient given by

\[
\log \hat{\vartheta}_{CFG,r} = -\gamma - \frac{1}{n} \sum_{i=1}^{n} \log \min(\hat{R}_{i1}, \hat{R}_{i2}) = \log 2 - \gamma + \int_{0}^{1} C_n(u^{1/2}, u^{1/2}) \frac{1(u > e^{-1})}{u \log u} \, du.
\]

3. F-madogram estimator. The rank-based F-madogram estimator for the extremal coefficient can be obtained by replacing the \(U\)'s in (5.5) by \(R\)'s:

\[
\hat{\nu}_{\alpha,r} = \frac{1}{2n} \sum_{i=1}^{n} |R_{i1}^\alpha - R_{i2}^\alpha|; \quad \hat{\vartheta}_{\alpha,r} = \frac{\alpha + \alpha(1 + \alpha)\hat{\nu}_{\alpha,r}}{\alpha - (1 + \alpha)\hat{\nu}_{\alpha,r}}. \tag{5.6}
\]

To write \(\hat{\vartheta}_{\alpha,r}\) in terms of the empirical copula, we can proceed in a similar fashion as in Appendix A of Genest and Segers (2009), using \(|a - b|/2 = \max(a, b) - (a + b)/2:

\[
\hat{\nu}_{\alpha,r} = \frac{1}{n} \sum_{i=1}^{n} \left[ \max(R_{i1}^\alpha, R_{i2}^\alpha) - \frac{1}{2} R_{i1}^\alpha - \frac{1}{2} R_{i2}^\alpha \right]
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{1} \left( 1 - 1\{R_{i1} \leq u^{1/\alpha}, R_{i2} \leq u^{1/\alpha} \} \right) \, du
\]

\[
- \frac{1}{2n} \sum_{i=1}^{n} \int_{0}^{1} \left( 1 - 1\{R_{i1} \leq u^{1/\alpha}, R_{i2} \leq 1 \} \right) \, du
\]

\[
- \frac{1}{2n} \sum_{i=1}^{n} \int_{0}^{1} \left( 1 - 1\{R_{i1} \leq 1, R_{i2} \leq u^{1/\alpha} \} \right) \, du
\]

\[
= \frac{1}{2} \int_{0}^{1} \left[ C_n(u^{1/\alpha}, 1) + C_n(1, u^{1/\alpha}) - 2C_n(u^{1/\alpha}, u^{1/\alpha}) \right] \, du, \tag{5.7}
\]

and \(\hat{\nu}_{\alpha,r}\) is given by (5.6).

4. Minimum distance estimator (Bücher et al. (2011)). This estimator for the Pickands dependence function is proposed as the minimizer of a weighted \(L^2\) distance, so that \(\hat{B}_{MD} = \arg \min_{B} M_h(B, C)\), where

\[
M_h(B, C) = \int_{0}^{1} \int_{0}^{1} \left[ \log C(u^{1-w}, u^{w}) - \log(u)B(t) \right]^2 h(u) \, du \, dw
\]
for copula $C$ with positive quadrant dependence, with $h(u)$ being a continuous nonnegative weight function. Bücher et al. (2011) show that the minimizer $\hat{B}_{MD}(w)$, which satisfies the boundary condition, is given by

$$\hat{B}_{MD}(w) = H_h^{-1} \int_0^1 \log C(u^{1-w}, u^w) \log(u) h(u) \, du,$$

with $H_h = \int_0^1 \log^2(u) h(u) \, du$ being a normalizing constant. The authors point out the particular class of weight functions $h_k(u) = -u^k / \log u$, for $k \geq 0$, yields the estimator

$$\hat{B}_{MD}(w, k) = -(k + 1)^2 \int_0^1 \log C(u^{1-w}, u^w) u^k \, du. \quad (5.8)$$

The rank-based estimator is obtained by replacing $C$ in (5.8) by $C_n$. For the extremal coefficient, this results in

$$\hat{\vartheta}_{MD,r}(k) = -2(k + 1)^2 \int_0^1 \log C_n(u^{1/2}, u^{1/2}) u^k \, du.$$

More sophisticated rank-based estimators for the Pickands dependence function (that satisfy the boundary and convexity conditions) exist in the literature, such as the projection estimator of Fils-Villetard et al. (2008) and one based on a transformation of the pair $(u_1, u_2) \mapsto (\log(u_2) / \log(u_1u_2), \log(C(u_1, u_2)) / \log(u_1u_2))$ and B-splines smoothing, as illustrated in Cormier et al. (2014).

### 5.1.3 Asymptotic efficiency of the estimators

Under certain regularity conditions, each of the estimators for the extremal coefficient mentioned is $\sqrt{n}$-consistent and asymptotically normal. However their asymptotic variances can be quite different, as observed in some articles that review the asymptotic properties of estimators for the Pickands dependence function, e.g., Naveau et al. (2009), Genest and Segers (2009) and Bücher et al. (2011). For most extremal coefficient estimators assuming known margins, it is possible to derive expressions for the asymptotic variance or at least use Monte Carlo simulations to estimate it (without first obtaining a sampling distribution); the estimator by Hall and Tajvidi is the only one for which these methods appear difficult. Because the Pickands and Deheuvels estimators are not necessarily 1 under comonotonicity, we do not consider them in the sequel due to the higher variability of these estimators when the dependence is strong, compared to those that evaluate to 1 under comonotonicity. The expression for the asymptotic variance of the F-madogram estimator is derived in Section 5.2.4, while the representation of the CFG estimator as a transform of the sum of i.i.d. random variables (5.4) allows quick estimation of the asymptotic variance via simulation:
The variance of \( \log \min(Y_{i1}, Y_{i2}) - \frac{1}{2}(\log Y_{i1} + \log Y_{i2}) \) in (5.4) can be estimated with a single sample of sufficient size, and then converted to the asymptotic variance of the CFG estimator using the delta method.

Figure 5.1 displays the asymptotic variances of the CFG and F-madogram estimators with known margins, using the Gumbel and Hüsler-Reiss copulas. For the F-madogram estimator, we consider the cases with \( \alpha = 1 \) (original formulation by Cooley et al. (2006)) and \( \alpha = 1/2 \) (modification by Naveau et al. (2009)). The results show that the CFG and F-madogram \( (\alpha = 1) \) estimators have very similar performance with the CFG estimator having slightly smaller asymptotic variance at moderate dependence but larger near independence. The F-madogram estimator with \( \alpha = 1/2 \) has noticeably higher asymptotic variance than for \( \alpha = 1 \) in both cases.

Figure 5.1: Asymptotic variances of the CFG and F-madogram estimators (assuming known margins) for the Gumbel and Hüsler-Reiss copulas

Analogous computations for the rank-based estimators are more difficult. Writing the estimators as functions of \( C_n \) facilitates calculations as a suitably scaled \( C_n \) converges in distribution to a Gaussian process \( \mathbb{G}_C \) (Fermanian et al. (2004)):

\[
\sqrt{n} [C_n(u_1, u_2) - C(u_1, u_2)] \overset{d}{\to} \mathbb{G}_C(u_1, u_2),
\]

where

\[
\mathbb{G}_C(u_1, u_2) = \mathbb{B}_C(u_1, u_2) - \mathbb{B}_C(u_1, 1) \frac{\partial C}{\partial u_1}(u_1, u_2) - \mathbb{B}_C(1, u_2) \frac{\partial C}{\partial u_2}(u_1, u_2)
\]
\[ B_C(u_1, u_2) - B_C(u_1, 1)C_{2|1}(u_2|u_1) - B_C(1, u_2)C_{1|2}(u_1|u_2), \quad (5.10) \]

in which \( B_C(u_1, u_2) \) is a Brownian bridge with covariance function \( \mathbb{E} [B_C(u_1, u_2)B_C(u_3, u_4)] = C(u_1 \wedge u_3, u_2 \wedge u_4) - C(u_1, u_2)C(u_3, u_4) \). The proof in Fermanian et al. (2004) requires that the copula has continuous partial derivatives on the closed square \([0, 1]^2\). This condition is not satisfied by many common parametric families (such as those with tail dependence), but the assumptions required for the convergence have been weakened over time (see, e.g., Tsukahara (2005); Omelka et al. (2009); Segers (2012)). Fermanian et al. (2004) also show that, if \( J : [0, 1]^2 \rightarrow \mathbb{R} \) is of bounded variation, continuous from above and with discontinuities of the first kind (Neuhaus (1971)), then

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ J(R_{i1}, R_{i2}) - \mathbb{E}[J(R_{i1}, R_{i2})] \right\} \xrightarrow{d} \int_0^1 \int_0^1 G_C(u_1, u_2) dJ(u_1, u_2), \quad (5.11)
\]

with the limiting distribution being Gaussian. These results permit the numerical computation of the asymptotic variance of the limiting Gaussian distribution. Expressions for the asymptotic variance are generally 2-dimensional integrals, but there is a simplification with the independence copula \( C(u_1, u_2) = u_1u_2 \). The asymptotic variances for the rank-based extremal coefficient estimators when \( C \) is the independence copula are listed in Table 5.1. Because of the lack of analytic expressions, it is often hard to tell which estimator dominates another. However, based on simulations in Genest and Segers (2009), the rank-based CFG estimator usually has smaller asymptotic variance than the rank-based Pickands estimator, although this is not always true. The simulation study in Naveau et al. (2009) uses transformed observations from fitted marginal GEV distributions rather than ranks; their results on the Gumbel or logistic model suggest that the CFG and F-madogram estimators have similar performance. The minimum distance estimator has asymptotic variance that depends on the value of \( k \); a study of its performance with the asymmetric Galambos or negative logistic model can be found in Bücher et al. (2011), using \( k = 1 \) or 5. None of these results in asymptotic variances much smaller than the rank-based CFG estimator.

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Asymptotic variance</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pickands</td>
<td>1.333</td>
<td>Genest and Segers (2009)</td>
</tr>
<tr>
<td>CFG</td>
<td>0.595</td>
<td></td>
</tr>
<tr>
<td>F-madogram</td>
<td>( \frac{(2 + \alpha)^2}{(1 + \alpha)(3 + 2\alpha)} )</td>
<td>See Section 5.2.4</td>
</tr>
<tr>
<td>Minimum distance</td>
<td>( \frac{4(1+k)^2}{(1+2k)(3+4k)} )</td>
<td>Bücher et al. (2011)</td>
</tr>
</tbody>
</table>

Table 5.1: Asymptotic variances of the empirical extremal coefficient \( \hat{\vartheta} \) for the rank-based estimators with the independence copula.
Finally, it is interesting to note that the rank-based estimators can sometimes have smaller asymptotic variances than their known margin counterparts. Genest and Segers (2009) show that this is always true for the Pickands and the uncorrected CFG estimators. This is however not always the case for the endpoint-corrected versions of the Pickands or CFG estimator, or the F-madogram estimator. Table 5.2 shows some simulation results that compare the asymptotic variances of different empirical estimators for the extremal coefficient, for two Gumbel copulas with different strengths of dependence. The Pickands and Deheuvels estimators are not considered for the same reason as mentioned at the beginning of the subsection. It is clear that none of the estimators dominates another, although the HT estimator has somewhat higher asymptotic variances than the rest when the dependence is weak.

<table>
<thead>
<tr>
<th>$\vartheta = 1.7$ (Kendall’s $\tau = 0.23$)</th>
<th>$\vartheta = 1.3$ (Kendall’s $\tau = 0.62$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Estimator</strong></td>
<td><strong>Known</strong></td>
</tr>
<tr>
<td>HT</td>
<td>0.92</td>
</tr>
<tr>
<td>CFG</td>
<td>0.61</td>
</tr>
<tr>
<td>F-madogram ($\alpha = 1$)</td>
<td>0.57</td>
</tr>
<tr>
<td>F-madogram ($\alpha = 0.5$)</td>
<td>0.72</td>
</tr>
<tr>
<td>Min. dist. ($k = 1$)</td>
<td>—</td>
</tr>
</tbody>
</table>

Table 5.2: Asymptotic variances of the empirical extremal coefficient $\hat{\vartheta}$ for various estimators with the Gumbel copula. “Known” stands for the estimator assuming known margins and “Rank” the rank-based estimator.

With regard to the use of the extremal coefficient as a dependence measure for adequacy-of-fit diagnostics in Chapters 6 and 7, the rank-based F-madogram estimator is chosen primarily because of its simplicity in computation compared to the minimum distance estimator and the competitive performance with the CFG estimator as suggested from the simulation study of Naveau et al. (2009). The original formulation by Cooley et al. (2006) with $\alpha = 1$ is selected as it appears to be less variable than that of Naveau et al. (2009) with $\alpha = 1/2$, based on the results in Section 5.2.4.

### 5.2 Generalization of the F-Madogram estimator to non-extreme-value copulas

Most of the estimators in the preceding section were originally motivated in the context of extreme value distributions. In this section, we show that one particular class, the F-madogram estimators, can be extended to a dependence measure that describes the tail properties of an arbitrary bivariate copula.
Consider the class of F-madogram estimators indexed by the parameter $\alpha > 0$ as in (5.5). This form of $\hat{\nu}_\alpha$ has recently been considered in Fonseca et al. (2015), but their focus is also restricted to extreme value distributions.

Observe that the population version of $\hat{\nu}_\alpha$ is given by
\[
\nu_\alpha = \frac{1}{2}\mathbb{E}[U_1^\alpha - U_2^\alpha] = \frac{1}{2}\mathbb{E}[2\max(U_1^\alpha, U_2^\alpha) - U_1^\alpha - U_2^\alpha]
\]
\[
= \mathbb{E}[\max(U_1^\alpha, U_2^\alpha)] - \frac{1}{\alpha + 1}
\]
(5.12)
as $U_1^\alpha \sim \text{Beta}(\alpha - 1, 1)$ and $\mathbb{E}(U_1^\alpha) = (\alpha + 1)^{-1}$, and similarly for $U_2^\alpha$. The distribution for the maximum can be obtained as follows:
\[
\mathbb{P}[\max(U_1^\alpha, U_2^\alpha) \leq u] = \mathbb{P}(U_1 \leq u^{1/\alpha}, U_2 \leq u^{1/\alpha}) = C(u^{1/\alpha}, u^{1/\alpha}) \triangleq F_M(u),
\]
and hence the expectation is
\[
\mathbb{E}[\max(U_1^\alpha, U_2^\alpha)] = \int_0^1 u \, dF_M(u) = \int_0^1 [1 - F_M(u)] \, du = 1 - \int_0^1 C(u^{1/\alpha}, u^{1/\alpha}) \, du.
\]
Let $\gamma_\alpha = \gamma_\alpha(C) = \int_0^1 C(u^{1/\alpha}, u^{1/\alpha}) \, du$. This means that
\[
\nu_\alpha = \frac{\alpha}{\alpha + 1} - \gamma_\alpha.
\]
(5.13)
Note that $\gamma_\alpha = \int_0^1 u^{\vartheta/\alpha} \, du = \alpha/(\alpha + \vartheta)$ for extreme value copulas with extremal coefficient $\vartheta$. In general, $\vartheta$ is dependent on $\alpha$ (written as $\vartheta_\alpha$ below; i.e., $\gamma_\alpha = \alpha/(\alpha + \vartheta_\alpha)$) and is related to these quantities by the relationship
\[
\vartheta_\alpha = \frac{\alpha + (1 + \alpha)\nu_\alpha}{\alpha - (1 + \alpha)\nu_\alpha} = \alpha(\gamma_\alpha^{-1} - 1).
\]
(5.14)
For the comonotonicity copula, $\vartheta_\alpha = 1$, $\gamma_\alpha = \alpha/(\alpha + 1)$ and $\nu_\alpha = 0$; for the independence copula, $\vartheta_\alpha = 2$, $\gamma_\alpha = \alpha/(\alpha + 2)$ and $\nu_\alpha = \alpha/[(\alpha + 1)(\alpha + 2)]$.

We define the population version of the F-madogram measure of dependence as
\[
\lambda_\alpha = \lambda_\alpha(C) \triangleq 2 - \vartheta_\alpha = 2 + \alpha(1 - \gamma_\alpha^{-1}),
\]
(5.15)
such that $\lambda_\alpha \in [0, 1]$ for $C$ with positive dependence. The empirical version is
\[
\hat{\lambda}_\alpha = 2 - \hat{\vartheta}_\alpha,
\]
(5.16)
where $\hat{\vartheta}_\alpha$ is obtained from (5.14), replacing $\nu_\alpha$ by its sample counterpart $\hat{\nu}_\alpha$ in (5.5) when the margins are assumed known. The rank-based version $\hat{\nu}_{\alpha,r}$ is obtained by replacing the $U$’s in (5.5) by the $R$’s, the scaled marginal ranks. From the relationship (5.12), it is
possible to define \( \hat{\nu}_\alpha \) in a different way, with 
\[ \tilde{\nu}_\alpha \equiv n^{-1} \sum_{i=1}^n \max(U_{1i}^\alpha, U_{2i}^\alpha) - (\alpha + 1)^{-1} \]
This is however not as desirable as \( \hat{\nu}_\alpha \) because, unlike \( \hat{\nu}_\alpha \), \( \tilde{\nu}_\alpha \) need not be 0 for observations from the comonotonicity copula.

Using the definition (5.15), \( \lambda_\alpha \) is a decreasing function of \( \hat{\nu}_\alpha \) and has a higher value when the copula is more strongly correlated. This direction is in agreement with other commonly used measures of dependence such as Kendall’s \( \tau \). In particular, for extreme value copulas, \( \lambda_\alpha \) coincides with the tail dependence index (for the appropriate tail) for any value of \( \alpha > 0 \).

Note that, when \( \alpha = 1 \), the F-madogram \( \nu_\alpha \) (for non-extreme-value copulas) is a linear transformation of a measure of association known as Spearman’s footrule (Spearman (1904, 1906)), with sample version being
\[
\hat{\phi}_r = 1 - \frac{3}{n^2 - 1} \sum_{i=1}^n |S_{i1} - S_{i2}| = 1 - \frac{3}{n - 1} \sum_{i=1}^n |R_{i1} - R_{i2}|,
\]
where \( S_{i1} \) and \( S_{i2} \) are the marginal ranks of the \( i \)th observation, and the corresponding population version \( \varphi = 1 - 3\mathbb{E}|U_1 - U_2| = 1 - 6\nu_1 \). The sample and asymptotic distributional properties of \( \hat{\phi}_r \) have been previously studied in Genest et al. (2010). Equations (5.14) and (5.15) imply
\[
\lambda_1 = 2 + (1 - \gamma_1^{-1}) = 2 - \left( \frac{1 + 2\nu_1}{1 - 2\nu_1} \right) = \frac{1 - 6\nu_1}{1 - 2\nu_1} = \frac{\varphi}{1 - 2\nu_1},
\]
i.e., both \( \lambda_1 \) and \( \varphi \) are 0 at independence and 1 at comonotonicity, but \( \lambda_1 \geq \varphi \) (resp. \( \lambda_1 \leq \varphi \)) for all copulas with positive (resp. negative) dependence as \( 1 - 2\nu_1 \leq 1 \). At the countermonotonicity limit, we have \( \lambda_1 = -1 \) (see Section 5.2.1) and \( \varphi = -1/2 \).

### 5.2.1 Dependence properties

We first investigate the behaviour of \( \lambda_\alpha \) as a measure of concordance of a bivariate copula. Scarsini (1984) has a list of criteria that a measure of concordance should satisfy; these are summarized in Definition 2.8 of Joe (2014). We check each of these items:

1. **Domain**: Satisfied as \( \lambda_\alpha \) is defined for all bivariate pairs with copula \( C \).

2. **Symmetry (permutation)**: Satisfied as \( \gamma_\alpha = \int_0^1 C(u^{1/\alpha}, u^{1/\alpha}) \, du \) is symmetric in the arguments. See below for comments on reflection symmetry.

3. **Coherence**: Satisfied as \( C_1(u_1, u_2) \prec_c C_2(u_1, u_2) \) (i.e., \( C_2 \) is larger than \( C_1 \) in the concordance ordering or equivalently \( C_2 \geq C_1 \) pointwise) implies \( \gamma_\alpha \) is larger for \( C_2 \) than \( C_1 \), and so is \( \lambda_\alpha \).
4. **Range**: The measure is constructed such that $\lambda_\alpha = 1$ at comonotonicity. We show below that $\lambda_\alpha$ is not necessarily $-1$ at countermonotonicity, and hence this item is not completely satisfied in general.

5. **Independence**: Satisfied as $\lambda_\alpha = 0$ for the independence copula.

6. **Sign reversal**: As the range condition is not generally satisfied, it is impossible for $\lambda_\alpha$ for all $(U_1, U_2)$ to be the negation of that for $(-U_1, U_2)$.

7. **Continuity**: Satisfied as $\lambda_\alpha$ is defined based on the copula.

8. **Invariance**: Satisfied as monotonic marginal transformation does not affect the copula.

Thus, $\lambda_\alpha$ satisfies many desirable properties for a measure of concordance. For reflection symmetry (i.e., same value of $\lambda_\alpha$ for a copula $C$ and its reflection $\hat{C}$), note that

$$
\gamma_\alpha(\hat{C}) = \int_0^1 \hat{C}(u^{1/\alpha}, u^{1/\alpha}) \, du = \int_0^1 (2u^{1/\alpha} - 1 + C(1 - u^{1/\alpha}, 1 - u^{1/\alpha})) \, du \\
= \frac{\alpha - 1}{\alpha + 1} + \int_0^1 C(v^{1/\alpha}, v^{1/\alpha})(v^{-1/\alpha} - 1)^{\alpha-1} \, dv,
$$

where $\gamma_\alpha(\hat{C})$ denotes the value of $\gamma_\alpha$ for the copula $\hat{C}$. This quantity is equal to $\gamma_\alpha(C)$ if $C$ is reflection symmetric (then it holds for any $\alpha > 0$), or when $\alpha = 1$ (then it holds for any copula). Otherwise, it is not generally true that $\gamma_\alpha(\hat{C}) = \gamma_\alpha(C)$.

For item 4 (range), because $\lambda_\alpha$ is a coherent dependence measure, the lower bound of its range can be obtained by considering the countermonotonicity copula, i.e., the Fréchet lower bound of a bivariate copula. The countermonotonicity copula is given by $C^-(u_1, u_2) = \max(0, u_1 + u_2 - 1)$ and thus

$$
\gamma^-_\alpha = \int_0^1 \max(0, 2u^{1/\alpha} - 1) \, du = \int_{2^{-\alpha}}^1 (2u^{1/\alpha} - 1) \, du = \frac{2^{-\alpha} + \alpha - 1}{1 + \alpha},
$$

where the minus sign at the superscript denotes the value for countermonotonicity copula. This implies

$$
\lambda^-_\alpha = 2 + \alpha \left(1 - \frac{1}{\gamma^-_\alpha}\right) = \frac{2^{-\alpha}(\alpha + 2) - 2}{2^{-\alpha} + \alpha - 1},
$$

an increasing function of $\alpha$. When $\alpha \to 0^+$, the limit of $\lambda^-_\alpha$ can be obtained using L’Hôpital’s rule:

$$
\lim_{\alpha \to 0^+} \lambda^-_\alpha = \lim_{\alpha \to 0^+} \frac{2^{-\alpha}(\alpha + 2) - 2}{2^{-\alpha} + \alpha - 1} = \lim_{\alpha \to 0^+} \frac{2^{-\alpha}(1 - \alpha \log 2 - \log 4)}{1 - 2^{-\alpha} \log 2} = \frac{1 - \log 4}{1 - \log 2} \approx -1.259.
$$

When $\alpha \to \infty$, $\lambda^-_\alpha \to 0$, and $\lambda^-_\alpha = -1$ when $\alpha = 1$. We thus observe that the range requirement of being $-1$ at countermonotonicity is only satisfied in the original formulation of the F-madogram, and can be slightly less than $-1$ when $\alpha \in (0, 1)$. 


5.2.2 Interpretation and use for general copulas

Interpretation of $\lambda_\alpha$ in (5.15) for bivariate copulas can be made by focusing on the behaviour of $\gamma_\alpha$. For $\alpha = 1$, it is an integral along the diagonal of the copula at equal increment $du$. When $\alpha > 1$, $u^{1/\alpha} > u$ and more emphasis is on the distribution function at the joint upper tail, whereas the opposite is true when $\alpha \in (0, 1)$. The quantity $\lambda_\alpha$ can thus be thought of as a tail-weighted summary that puts different weights on the strength of dependence of a copula (in terms of the magnitude of $C(u_1, u_2)$ along the diagonal) at different locations.

The use of this estimator in this regard can be compared to that of the tail-weighted dependence measures proposed in Krupskii and Joe (2015).

To illustrate this, we compute the value of $\lambda_\alpha$ for several parametric copula families with a constant value of Kendall’s $\tau$ at 0.5. They include the Gaussian (symmetric with no tail dependence, i.e., the tail dependence index is zero for both tails), Frank (symmetric with even lighter tails than Gaussian), Hüsler-Reiss (asymmetric with upper tail dependence, extreme value copula), Gumbel (asymmetric with upper tail dependence, extreme value copula), reflected MTCJ (very asymmetric with upper tail dependence), $t$ (symmetric with dependence at both tails), and BB1 (asymmetric with dependence at both tails) copulas.

The BB1 family has two parameters; we set the upper tail dependence index to 0.5 as well in order to arrive at a unique set of parameters. This results in a lower tail dependence index of 0.303. For a copula $C$, $\lambda_\alpha(C)$ puts more weight on the upper tail when $\alpha$ is large. For the reflected copula $\tilde{C}$, $\lambda_\alpha(\tilde{C})$ puts more weight on the lower tail of $C$ when $\alpha$ is large. Hereafter, when necessary we make the distinction that $\lambda_{U,\alpha} = \lambda_\alpha(C)$ and $\lambda_{L,\alpha} = \lambda_\alpha(\tilde{C})$, respectively. The results are shown in Table 5.3; note that the $\lambda_\alpha$ values are not very different among copulas with the same Kendall’s $\tau$ when $\alpha$ is small. The difference becomes more pronounced as $\alpha$ increases and, more importantly, the magnitudes are indeed indicative of the strength of tail dependence expected of these copulas: For the panel with $\lambda_{U,\alpha}$ values, note that the reflected MTCJ has the heaviest upper tail, and is followed by the Hüsler-Reiss and Gumbel copulas; the BB1 copula and $t$ copula with small degrees of freedom come next; they have moderate upper tail dependence. The Gaussian copula and $t$ copula with large degrees of freedom have even lighter upper tails, and the Frank copula has the lightest tails of all the parametric copula families considered. As for the lower tails, only the $t$ and BB1 copulas have lower tail dependence and this is also reflected by the values of $\lambda_{L,\alpha}$ for $\alpha$ large. By the construction of $\lambda_{U,\alpha}$, it does not change with $\alpha$ when $C$ is the extreme value copula (i.e., $C(u, u) = u^\vartheta$ for some $\vartheta$ that does not depend on $u$); in this case $\lambda_{U,\alpha} = \lambda_U$, the upper tail dependence index, for all $\alpha$. This can be seen for the Hüsler-Reiss and Gumbel copulas in the upper panel of Table 5.3. For general copulas, we show in Section 5.2.3 that $\lambda_\alpha$ converges to the tail dependence index for the appropriate
tail as $\alpha \to \infty$.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>Gaussian</th>
<th>Frank</th>
<th>HR</th>
<th>Gumbel</th>
<th>rMTCJ</th>
<th>t ($\nu = 3$)</th>
<th>t ($\nu = 20$)</th>
<th>BB1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.61</td>
<td>0.59</td>
<td>0.58</td>
<td>0.59</td>
<td>0.54</td>
<td>0.64</td>
<td>0.62</td>
<td>0.62</td>
</tr>
<tr>
<td>0.5</td>
<td>0.60</td>
<td>0.59</td>
<td>0.58</td>
<td>0.59</td>
<td>0.56</td>
<td>0.62</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>1</td>
<td>0.58</td>
<td>0.58</td>
<td>0.59</td>
<td>0.59</td>
<td>0.60</td>
<td>0.58</td>
<td>0.59</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.55</td>
<td>0.56</td>
<td>0.58</td>
<td>0.59</td>
<td>0.62</td>
<td>0.57</td>
<td>0.55</td>
<td>0.56</td>
</tr>
<tr>
<td>5</td>
<td>0.50</td>
<td>0.47</td>
<td>0.58</td>
<td>0.59</td>
<td>0.65</td>
<td>0.53</td>
<td>0.50</td>
<td>0.54</td>
</tr>
<tr>
<td>10</td>
<td>0.45</td>
<td>0.37</td>
<td>0.58</td>
<td>0.59</td>
<td>0.68</td>
<td>0.51</td>
<td>0.46</td>
<td>0.52</td>
</tr>
<tr>
<td>20</td>
<td>0.40</td>
<td>0.26</td>
<td>0.58</td>
<td>0.59</td>
<td>0.69</td>
<td>0.49</td>
<td>0.42</td>
<td>0.51</td>
</tr>
<tr>
<td>50</td>
<td>0.34</td>
<td>0.14</td>
<td>0.58</td>
<td>0.59</td>
<td>0.70</td>
<td>0.48</td>
<td>0.36</td>
<td>0.50</td>
</tr>
<tr>
<td>100</td>
<td>0.29</td>
<td>0.08</td>
<td>0.58</td>
<td>0.59</td>
<td>0.70</td>
<td>0.47</td>
<td>0.33</td>
<td>0.50</td>
</tr>
<tr>
<td>$\lambda_U$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.58</td>
<td>0.59</td>
<td>0.71</td>
<td>0.45</td>
<td>0.07</td>
<td>0.50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>Gaussian</th>
<th>Frank</th>
<th>HR</th>
<th>Gumbel</th>
<th>rMTCJ</th>
<th>t ($\nu = 3$)</th>
<th>t ($\nu = 20$)</th>
<th>BB1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.61</td>
<td>0.59</td>
<td>0.66</td>
<td>0.66</td>
<td>0.69</td>
<td>0.64</td>
<td>0.62</td>
<td>0.64</td>
</tr>
<tr>
<td>0.5</td>
<td>0.60</td>
<td>0.59</td>
<td>0.62</td>
<td>0.63</td>
<td>0.69</td>
<td>0.64</td>
<td>0.62</td>
<td>0.61</td>
</tr>
<tr>
<td>1</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.59</td>
<td>0.59</td>
<td>0.60</td>
<td>0.58</td>
<td>0.59</td>
</tr>
<tr>
<td>2</td>
<td>0.55</td>
<td>0.56</td>
<td>0.52</td>
<td>0.53</td>
<td>0.50</td>
<td>0.57</td>
<td>0.55</td>
<td>0.55</td>
</tr>
<tr>
<td>5</td>
<td>0.50</td>
<td>0.47</td>
<td>0.43</td>
<td>0.44</td>
<td>0.35</td>
<td>0.53</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>10</td>
<td>0.45</td>
<td>0.37</td>
<td>0.36</td>
<td>0.36</td>
<td>0.24</td>
<td>0.51</td>
<td>0.46</td>
<td>0.47</td>
</tr>
<tr>
<td>20</td>
<td>0.40</td>
<td>0.26</td>
<td>0.29</td>
<td>0.30</td>
<td>0.15</td>
<td>0.49</td>
<td>0.42</td>
<td>0.44</td>
</tr>
<tr>
<td>50</td>
<td>0.34</td>
<td>0.14</td>
<td>0.21</td>
<td>0.22</td>
<td>0.07</td>
<td>0.48</td>
<td>0.36</td>
<td>0.40</td>
</tr>
<tr>
<td>100</td>
<td>0.29</td>
<td>0.08</td>
<td>0.17</td>
<td>0.17</td>
<td>0.04</td>
<td>0.47</td>
<td>0.33</td>
<td>0.38</td>
</tr>
<tr>
<td>$\lambda_L$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.45</td>
<td>0.07</td>
<td>0.30</td>
</tr>
</tbody>
</table>

Table 5.3: Values of $\lambda_{U,\alpha}$ and $\lambda_{L,\alpha}$ for various bivariate copulas with Kendall’s $\tau$ equal to 0.5. “HR” and “rMTCJ” stand for the Hüsler-Reiss and reflected MTCJ copulas, respectively. The values for $\lambda_U$ and $\lambda_L$ in the final rows are the upper and lower tail dependence index, respectively.

The choice of $\alpha$ is important in using the F-madogram measure of dependence to assess the tail characteristics of a distribution. It is desirable to find a value of $\alpha$ such that the difference of $\lambda_\alpha$ between a model with tail dependence and one without is large relative to the variability of the empirical estimator, so as to obtain a higher differentiating power among copula models. To explore the range of desirable values of $\alpha$, we run a simulation study similar to that of Table 2 of Krupskii and Joe (2015). Six scenarios are considered; in each of these scenarios, we estimate the difference between the $\lambda_{L,\alpha}$ values for a lower tail dependent copula (either t with 4 degrees of freedom or reflected Gumbel) versus one without tail dependence (for the lower tail) to high accuracy using large samples (20,000
replications each of size 20,000) and the rank-based version of the estimator. These are compared against their variability (in terms of standard errors), for a sample size of 400 which reflects realistic scenarios. Copula parameters are chosen so that the Spearman’s \( \rho \) is at 0.5 or 0.7; these simulation settings are the same as those in Krupskii and Joe (2015) so as to allow comparisons between the two measures. Results of the simulation are given in Table 5.4; they appear to suggest that a value of \( \alpha \) between 15 and 20 is more effective in differentiating between copulas with and without tail dependence. As with the tail-weighted dependence measures of Krupskii and Joe (2015), a sample size of 400 seems to be insufficient in detecting the difference between the tails of a Gaussian copula and a t copula with 4 degrees of freedom, but marginally sufficient for copulas with more different lower tails. For \( \alpha \) less than 50, we can also see that the standard errors of the differences are typically smaller than those of the tail-weighted dependence measures of Krupskii and Joe (2015), with their chosen functions and truncation level.

5.2.3 Boundary cases

We investigate the properties of the F-madogram measure of dependence \( \lambda_\alpha = \lambda_{U,\alpha} = \lambda_\alpha(C) \) as \( \alpha \) approaches the lower or upper limit, i.e., as \( \alpha \to 0^+ \) and \( \alpha \to \infty \), and show that \( \lambda_\alpha \) tends to the upper tail dependence index \( \lambda_U \) in the latter case.

Limit of \( \lambda_\alpha \) as \( \alpha \to 0^+ \). When \( \alpha \to 0^+ \), the integrand of \( \gamma_\alpha, C(u^{1/\alpha}, u^{1/\alpha}) \), tends to zero everywhere for \( u \in [0,1) \) and is only 1 at \( u = 1 \). The integrand is also bounded in \([0,1]\), and thus the exchange of limit and integral is valid. This yields \( \lim_{\alpha \to 0^+} \gamma_\alpha = 0 \) and thus

\[
\lim_{\alpha \to 0^+} \lambda_\alpha = 2 - \lim_{\alpha \to 0^+} \frac{\alpha}{\gamma_\alpha} = 2 - \lim_{\alpha \to 0^+} \left[ \frac{1}{\alpha} \int_0^1 C(u^{1/\alpha}, u^{1/\alpha}) \, du \right]^{-1}
= 2 - \lim_{\alpha \to 0^+} \left[ \int_0^1 v^{\alpha-1} C(v, v) \, dv \right]^{-1}
= 2 - \left[ \int_0^1 v^{-1} C(v, v) \, dv \right]^{-1}.
\]

The limit exists for non-comonotonicity copula as \( C(v, v) \leq v \) and \( C(v, v)v^{\alpha-1} \leq v^{-1} \).

Limit of \( \lambda_\alpha \) as \( \alpha \to \infty \) when the upper tail of \( C \) is well-behaved so that the upper tail dependence index \( \lambda_U \) exists. When \( \alpha \to \infty \), \( C(u^{1/\alpha}, u^{1/\alpha}) \) tends to 1 everywhere for \( u \in (0,1] \) but is undefined at \( u = 0; \lim_{\alpha \to \infty} \gamma_\alpha = 1 \) and thus

\[
\lim_{\alpha \to \infty} \lambda_\alpha = 2 + \lim_{\alpha \to \infty} \frac{\alpha(\gamma_\alpha - 1)}{\gamma_\alpha} = 2 + \lim_{\alpha \to \infty} \alpha(\gamma_\alpha - 1).
\]

11Operationally, asymptotic variances \( \sigma^2 \) are estimated from the simulation result, and then converted to the standard error for sample size 400 via the relationship \( \sigma/\sqrt{400} \).
<table>
<thead>
<tr>
<th>α</th>
<th>( t_4^- ) Gaussian</th>
<th>( r\text{Gumbel}^- ) Gaussian</th>
<th>( t_4^- ) Gumbel</th>
<th>( r\text{Gumbel}^- ) Gumbel</th>
<th>( t_4^- ) Frank</th>
<th>( r\text{Gumbel}^- ) Frank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.03 (0.05)</td>
<td>0.02 (0.05)</td>
<td>0.02 (0.05)</td>
<td>0.00 (0.05)</td>
<td>0.02 (0.05)</td>
<td>0.01 (0.05)</td>
</tr>
<tr>
<td>3</td>
<td>0.03 (0.05)</td>
<td>0.06 (0.05)</td>
<td>0.07 (0.05)</td>
<td>0.10 (0.05)</td>
<td>0.04 (0.05)</td>
<td>0.07 (0.05)</td>
</tr>
<tr>
<td>5</td>
<td>0.04 (0.06)</td>
<td>0.09 (0.06)</td>
<td>0.09 (0.06)</td>
<td>\textbf{0.14 (0.06)}</td>
<td>0.06 (0.06)</td>
<td>\textbf{0.11 (0.06)}</td>
</tr>
<tr>
<td>10</td>
<td>0.07 (0.07)</td>
<td>\textbf{0.14 (0.07)}</td>
<td>0.13 (0.07)</td>
<td>\textbf{0.20 (0.07)}</td>
<td>0.12 (0.07)</td>
<td>\textbf{0.19 (0.07)}</td>
</tr>
<tr>
<td>15</td>
<td>0.08 (0.08)</td>
<td>\textbf{0.17 (0.08)}</td>
<td>0.15 (0.08)</td>
<td>\textbf{0.24 (0.08)}</td>
<td>0.16 (0.08)</td>
<td>\textbf{0.24 (0.08)}</td>
</tr>
<tr>
<td>20</td>
<td>0.09 (0.09)</td>
<td>\textbf{0.19 (0.09)}</td>
<td>0.17 (0.09)</td>
<td>\textbf{0.26 (0.09)}</td>
<td>0.18 (0.08)</td>
<td>\textbf{0.27 (0.08)}</td>
</tr>
<tr>
<td>50</td>
<td>0.13 (0.13)</td>
<td>0.24 (0.13)</td>
<td>0.20 (0.12)</td>
<td>\textbf{0.32 (0.12)}</td>
<td>\textbf{0.24 (0.11)}</td>
<td>0.35 (0.11)</td>
</tr>
<tr>
<td>100</td>
<td>0.15 (0.17)</td>
<td>0.28 (0.17)</td>
<td>0.22 (0.16)</td>
<td>\textbf{0.35 (0.16)}</td>
<td>0.26 (0.15)</td>
<td>\textbf{0.39 (0.15)}</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>α</th>
<th>( t_4^- ) Gaussian</th>
<th>( r\text{Gumbel}^- ) Gaussian</th>
<th>( t_4^- ) Gumbel</th>
<th>( r\text{Gumbel}^- ) Gumbel</th>
<th>( t_4^- ) Frank</th>
<th>( r\text{Gumbel}^- ) Frank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.03 (0.03)</td>
<td>0.02 (0.04)</td>
<td>0.01 (0.04)</td>
<td>0.00 (0.04)</td>
<td>0.02 (0.03)</td>
<td>0.01 (0.03)</td>
</tr>
<tr>
<td>3</td>
<td>0.03 (0.04)</td>
<td>0.07 (0.04)</td>
<td>0.06 (0.04)</td>
<td>\textbf{0.10 (0.04)}</td>
<td>0.04 (0.04)</td>
<td>0.07 (0.04)</td>
</tr>
<tr>
<td>5</td>
<td>0.04 (0.05)</td>
<td>\textbf{0.10 (0.05)}</td>
<td>0.09 (0.05)</td>
<td>\textbf{0.15 (0.05)}</td>
<td>0.07 (0.05)</td>
<td>\textbf{0.13 (0.05)}</td>
</tr>
<tr>
<td>10</td>
<td>0.06 (0.06)</td>
<td>\textbf{0.14 (0.06)}</td>
<td>\textbf{0.14 (0.06)}</td>
<td>\textbf{0.22 (0.06)}</td>
<td>0.15 (0.06)</td>
<td>\textbf{0.23 (0.06)}</td>
</tr>
<tr>
<td>15</td>
<td>0.08 (0.07)</td>
<td>\textbf{0.17 (0.07)}</td>
<td>\textbf{0.17 (0.07)}</td>
<td>\textbf{0.26 (0.07)}</td>
<td>\textbf{0.20 (0.07)}</td>
<td>\textbf{0.29 (0.07)}</td>
</tr>
<tr>
<td>20</td>
<td>0.09 (0.08)</td>
<td>\textbf{0.19 (0.08)}</td>
<td>\textbf{0.19 (0.08)}</td>
<td>\textbf{0.29 (0.08)}</td>
<td>\textbf{0.23 (0.08)}</td>
<td>\textbf{0.34 (0.08)}</td>
</tr>
<tr>
<td>50</td>
<td>0.13 (0.13)</td>
<td>\textbf{0.25 (0.12)}</td>
<td>0.24 (0.13)</td>
<td>\textbf{0.37 (0.12)}</td>
<td>\textbf{0.33 (0.11)}</td>
<td>0.46 (0.10)</td>
</tr>
<tr>
<td>100</td>
<td>0.15 (0.18)</td>
<td>0.30 (0.17)</td>
<td>0.28 (0.17)</td>
<td>\textbf{0.42 (0.16)}</td>
<td>\textbf{0.38 (0.15)}</td>
<td>\textbf{0.52 (0.13)}</td>
</tr>
</tbody>
</table>

Table 5.4: Comparison of the difference in \( \lambda_{L,\alpha} \) for six pairs of copulas against their asymptotic standard errors for the sample size 400 (in brackets). The upper panel uses copulas with Spearman’s \( \rho \) equal to 0.5 and the bottom panel 0.7; values that are significant at 5% level are shown in boldface.

Note that

\[
\alpha(\gamma_{\alpha} - 1) = \int_0^1 \alpha \left[ C(u^{1/\alpha}, u^{1/\alpha}) - 1 \right] du = \int_0^1 \alpha \left[ 2u^{1/\alpha} - 2 + C(u^{1/\alpha}, u^{1/\alpha}) \right] du
\]

\[
= \int_0^1 \alpha C(u^{1/\alpha}, u^{1/\alpha}) du - \frac{2\alpha}{\alpha + 1},
\]

and thus we need to find

\[
\lim_{\alpha \to \infty} \int_0^1 \alpha C(u^{1/\alpha}, u^{1/\alpha}) du.
\]

First, note that the integral is bounded as \( C(u^{1/\alpha}, u^{1/\alpha}) = 1 - 2u^{1/\alpha} + C(u^{1/\alpha}, u^{1/\alpha}) \leq 1 - u^{1/\alpha} \), meaning that

\[
\int_0^1 \alpha C(u^{1/\alpha}, u^{1/\alpha}) du \leq \int_0^1 \alpha (1 - u^{1/\alpha}) du = \frac{\alpha}{\alpha + 1} \leq 1
\]
for any positive \( \alpha \), and tends to 1 as \( \alpha \to \infty \). Then, consider the tail expansion of the survival function of \( C \) using the definition of the upper tail dependence index \( \lambda_U \):

\[
\overline{C}(1-v, 1-v) \sim v\lambda_U
\]

as \( v \to 0^+ \), where the operator \( \sim \) is such that \( f_1(v) \sim f_2(v) \) means \( \lim_{v \to 0^+} f_1(v)/f_2(v) = 1 \). Suppose \( 0 < \lambda_U < 1 \). Then for every (small) \( \epsilon > 0 \), there exists \( \delta > 0 \) such that for every \( 0 \leq v < \delta \) we have

\[
v(\lambda_U - \epsilon) \leq \overline{C}(1-v, 1-v) \leq v(\lambda_U + \epsilon).
\]

Also, there exists some \( \alpha^* \) such that for every \( \alpha > \alpha^* \), \( 1 - u^{1/\alpha} < \delta \) for every \( u > \epsilon \). Write

\[
\int_0^1 \alpha \overline{C}(u^{1/\alpha}, u^{1/\alpha}) \, du = \int_0^\epsilon \alpha \overline{C}(u^{1/\alpha}, u^{1/\alpha}) \, du + \int_\epsilon^1 \alpha \overline{C}(u^{1/\alpha}, u^{1/\alpha}) \, du \triangleq h_1(\epsilon, \alpha) + h_2(\epsilon, \alpha).
\]

We treat each integral separately. For \( h_1(\epsilon, \alpha) \),

\[
0 \leq h_1(\epsilon, \alpha) \leq \int_0^\epsilon \alpha (1 - u^{1/\alpha}) \, du = \alpha \epsilon \left( 1 - \frac{\alpha}{\alpha + 1} \epsilon^{1/\alpha} \right).
\]

Because the upper limit tends to \( \epsilon \) \( (1 - \log \epsilon) \) as \( \alpha \to \infty \), there exists some \( \alpha^{**} \) and constant \( M > 1 \) such that for all \( \alpha > \alpha^{**} \) we have

\[
0 \leq h_1(\epsilon, \alpha) \leq M\epsilon \left(1 - \log \epsilon \right).
\]

For \( h_2(\epsilon, \alpha) \), since \( 1 - u^{1/\alpha} < \delta \) for all \( u > \epsilon \) and \( \alpha > \alpha^* \), we use (5.18) to establish the bounds

\[
(\lambda_U - \epsilon) \int_\epsilon^1 \alpha (1 - u^{1/\alpha}) \, du \leq h_2(\epsilon, \alpha) \leq (\lambda_U + \epsilon) \int_\epsilon^1 \alpha (1 - u^{1/\alpha}) \, du
\]

\[
\Rightarrow (\lambda_U - \epsilon) \left[ \frac{\alpha}{\alpha + 1} - \alpha \epsilon \left( 1 - \frac{\alpha}{\alpha + 1} \epsilon^{1/\alpha} \right) \right] \leq h_2(\epsilon, \alpha) \leq (\lambda_U + \epsilon),
\]

where the upper limit uses the relationship \( \int_\epsilon^1 \alpha (1 - u^{1/\alpha}) \, du \leq \int_0^1 \alpha (1 - u^{1/\alpha}) \, du \leq 1 \). Thus

\[
(\lambda_U - \epsilon) \left[ \frac{\alpha}{\alpha + 1} - M\epsilon \left(1 - \log \epsilon \right) \right] \leq h_2(\epsilon, \alpha) \leq (\lambda_U + \epsilon)
\]

for all \( \alpha > \max(\alpha^*, \alpha^{**}) \), and, as \( \alpha \to \infty \),

\[
(\lambda_U - \epsilon) \left[ 1 - M\epsilon \left(1 - \log \epsilon \right) \right] \leq h_1(\epsilon, \infty) + h_2(\epsilon, \infty) \leq M\epsilon \left(1 - \log \epsilon \right) + (\lambda_U + \epsilon),
\]

where \( h_j(\epsilon, \infty) = \lim_{\alpha \to \infty} h_j(\epsilon, \alpha), j = 1, 2 \). Since \( \epsilon > 0 \) can be arbitrarily small,

\[
\lim_{\alpha \to \infty} \int_0^1 \alpha \overline{C}(u^{1/\alpha}, u^{1/\alpha}) \, du = \lambda_U.
\]
The proof applies to $\lambda_U = 1$ or 0 by taking (5.18) as $v(\lambda_U - \epsilon) \leq \overline{C}(1 - v, 1 - v) \leq v$ or $0 \leq \overline{C}(1 - v, 1 - v) \leq v(\lambda_U + \epsilon)$, respectively.

Putting this result back into (5.17), we obtain

$$\lim_{\alpha \to \infty} \lambda_\alpha = \lambda_U,$$

thus reinforcing the interpretation of $\lambda_\alpha$ that more weight is put on the upper tail as $\alpha$ increases, eventually coinciding with the upper tail dependence index when $\alpha \to \infty$.

### 5.2.4 Asymptotic normality and variance

In this subsection, we explore the asymptotic properties of the F-madogram measure of dependence estimator as the sample size $n \to \infty$. We deal with the two versions of the estimator separately: The estimator assuming known margins $\hat{\lambda}_\alpha$, and the one using ranks $\hat{\lambda}_{\alpha,r}$. We focus on the treatment of $\lambda_\alpha = \lambda_{U,\alpha} = \lambda_\alpha(C)$.

**Estimator assuming known margins.** For the estimator assuming known margins, the asymptotic normality of $\sqrt{n} (\hat{\nu}_\alpha - \nu_\alpha)$ is immediate noting from (5.5) that it is a sum of i.i.d. random variables with finite variance. We thus have

$$\sqrt{n} (\hat{\nu}_\alpha - \nu_\alpha) \xrightarrow{d} N(0, \sigma_\nu^2),$$

where $\sigma_\nu^2 = \text{Var} (|U_{1}^\alpha - U_{2}^\alpha|/2)$ is obtained as follows:

$$\sigma_\nu^2 = \frac{1}{4} \left\{ \mathbb{E} \left( (U_{1}^\alpha - U_{2}^\alpha)^2 \right) - \left( \mathbb{E} |U_{1}^\alpha - U_{2}^\alpha| \right)^2 \right\} = \frac{1}{2} \mathbb{E} (U_{1}^{2\alpha}) - \frac{1}{2} \mathbb{E} [(U_{1}U_{2})^\alpha] - \nu_\alpha^2$$

$$= \frac{1}{2(1 + 2\alpha)} - \frac{1}{2} \rho_\alpha - \nu_\alpha^2,$$

where $\rho_\alpha \triangleq \mathbb{E} [(U_{1}U_{2})^\alpha]$. To obtain the asymptotic variance of $\hat{\lambda}_\alpha$, we refer to (5.14) with $\lambda_\alpha = 2 - \vartheta_\alpha$ and let

$$h(x) = 2 - \frac{\alpha + \alpha(1 + \alpha)x}{\alpha - (1 + \alpha)x} = \frac{\alpha - (1 + \alpha)(2 + \alpha)x}{\alpha - (1 + \alpha)x}.$$ 

This yields

$$h'(\nu_\alpha) = -\frac{\alpha(1 + \alpha)^2}{[\alpha(\nu_\alpha - 1) + \nu_\alpha]^2} = -\frac{(\alpha + 2 - \lambda_\alpha)^2}{\alpha}$$

using the relationship

$$\nu_\alpha = \frac{\alpha(1 - \lambda_\alpha)}{(1 + \alpha)(\alpha + 2 - \lambda_\alpha)}$$

from (5.13) and (5.14). Then by the delta method,

$$\sqrt{n}(\hat{\lambda}_\alpha - \lambda_\alpha) \xrightarrow{d} N \left( 0, \sigma^2 \right),$$

$$103$$
where
\[
\sigma^2 = \left[h'(\nu_\alpha)\right]^2 \frac{\sigma^2}{\alpha^2} = \frac{(\alpha + 2 - \lambda_\alpha)^4}{\alpha^2} \left[\frac{1}{2(1 + 2\alpha)} - \frac{1}{2\alpha} - \nu_\alpha^2\right]. \tag{5.19}
\]

Table 5.5 tabulates the values of asymptotic variances for a combination of various parametric copula families and values of \(\alpha\); these are the same scenarios as those considered in Table 5.3. Note that \(\mathbb{E}[(U_1 U_2)^\alpha]\) can be obtained by the following relationship, in which \(V_1\) and \(V_2\) are i.i.d. unit uniform random variables and are independent of \((U_1, U_2) \sim C:
\[
\begin{align*}
&\int_0^1 \int_0^1 (u_1 u_2)^\alpha \, dC(u_1, u_2) \\
&= \int_0^1 \int_0^1 \mathbb{P}(V_1^{1/\alpha} \leq u_1, V_2^{1/\alpha} \leq u_2 | U_1 = u_1, U_2 = u_2) \, du_1 du_2 \\
&= \mathbb{P}(V_1^{1/\alpha} \leq U_1, V_2^{1/\alpha} \leq U_2) \\
&= \int_0^1 \int_0^1 \mathbb{P}(U_1 > v_1^{1/\alpha}, U_2 > v_2^{1/\alpha} | V_1 = v_1, V_2 = v_2) \, dv_1 dv_2 \\
&= \int_0^1 \int_0^1 C(v_1^{1/\alpha}, v_2^{1/\alpha}) \, dv_1 dv_2.
\end{align*}
\]

The final integral can be evaluated numerically and is stable as both the integrand and the limits of integration are bounded by \([0, 1]\); this is preferred over the evaluation of \(\int_0^1 \int_0^1 (u_1 u_2)^\alpha c(u_1, u_2) \, du_1 du_2\) directly as copula densities may asymptote to \(\infty\) near \((0, 0)\) or \((1, 1)\).

From Table 5.5, we note that for most of these families, the asymptotic variances are the smallest when \(\alpha\) is small (e.g., 1 or less). This is in agreement with the behaviour of \(\lambda_\alpha\) we observed earlier, that it tends to be similar among copulas with the same overall strength of dependence. For given \(\alpha\), the asymptotic variance appears to be smaller with a larger value of \(\lambda_\alpha\).

**Rank-based estimator.** Equation (5.7) expresses the rank-based estimator \(\hat{\nu}_{\alpha,r}\) in terms of the empirical copula \(C_n\); the population version \(\nu_\alpha\) can be obtained by replacing \(C_n\) with \(C\) in (5.7). The convergence result of the empirical copula process (5.9) then implies that
\[
\sqrt{n}(\hat{\nu}_{\alpha,r} - \nu_\alpha) \overset{d}{\to} X,
\]
where
\[
X = \frac{1}{2} \int_0^1 \mathbb{G}_C(u^{1/\alpha}, 1) \, du + \frac{1}{2} \int_0^1 \mathbb{G}_C(1, u^{1/\alpha}) \, du - \int_0^1 \mathbb{G}_C(u^{1/\alpha}, u^{1/\alpha}) \, du. \tag{5.20}
\]

The normality of \(X\) is given by (5.11) with \(J(u, v) = |u^\alpha - v^\alpha|/2\). For the asymptotic distribution of \(\hat{\lambda}_{\alpha,r}\), note that from (5.13) and (5.15) we have
\[
\lambda_\alpha = 2 + \alpha \left(1 - \frac{1}{\alpha/(\alpha + 1) - \nu_\alpha}\right); \quad \hat{\lambda}_{\alpha,r} = 2 + \alpha \left(1 - \frac{1}{\alpha/(\alpha + 1) - \hat{\nu}_{\alpha,r}}\right),
\]
104
Asymptotic variance of $\hat{\lambda}_{U,\alpha}$ — Upper tail-weighted for large $\alpha$

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>Gaussian</th>
<th>Frank</th>
<th>HR</th>
<th>Gumbel</th>
<th>rMTCJ</th>
<th>$t (\nu = 3)$</th>
<th>$t (\nu = 20)$</th>
<th>BB1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.23</td>
<td>0.32</td>
<td>0.33</td>
<td>0.34</td>
<td>0.49</td>
<td>0.25</td>
<td>0.23</td>
<td>0.24</td>
</tr>
<tr>
<td>0.5</td>
<td>0.19</td>
<td>0.21</td>
<td>0.23</td>
<td>0.24</td>
<td>0.30</td>
<td>0.22</td>
<td>0.19</td>
<td>0.20</td>
</tr>
<tr>
<td>1</td>
<td>0.18</td>
<td>0.18</td>
<td>0.19</td>
<td>0.20</td>
<td>0.20</td>
<td>0.22</td>
<td>0.19</td>
<td>0.20</td>
</tr>
<tr>
<td>2</td>
<td>0.24</td>
<td>0.25</td>
<td>0.21</td>
<td>0.22</td>
<td>0.17</td>
<td>0.28</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>5</td>
<td>0.54</td>
<td>0.67</td>
<td>0.36</td>
<td>0.37</td>
<td>0.24</td>
<td>0.52</td>
<td>0.54</td>
<td>0.48</td>
</tr>
<tr>
<td>10</td>
<td>1.14</td>
<td>1.61</td>
<td>0.65</td>
<td>0.67</td>
<td>0.40</td>
<td>0.97</td>
<td>1.11</td>
<td>0.89</td>
</tr>
<tr>
<td>20</td>
<td>2.50</td>
<td>3.78</td>
<td>1.25</td>
<td>1.29</td>
<td>0.73</td>
<td>1.91</td>
<td>2.39</td>
<td>1.73</td>
</tr>
<tr>
<td>50</td>
<td>7.07</td>
<td>10.92</td>
<td>3.06</td>
<td>3.13</td>
<td>1.71</td>
<td>4.80</td>
<td>6.62</td>
<td>4.28</td>
</tr>
<tr>
<td>100</td>
<td>15.38</td>
<td>23.24</td>
<td>6.08</td>
<td>6.21</td>
<td>3.34</td>
<td>9.67</td>
<td>14.22</td>
<td>8.52</td>
</tr>
</tbody>
</table>

Asymptotic variance of $\hat{\lambda}_{L,\alpha}$ — Lower tail-weighted for large $\alpha$

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>Gaussian</th>
<th>Frank</th>
<th>HR</th>
<th>Gumbel</th>
<th>rMTCJ</th>
<th>$t (\nu = 3)$</th>
<th>$t (\nu = 20)$</th>
<th>BB1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.23</td>
<td>0.32</td>
<td>0.15</td>
<td>0.16</td>
<td>0.11</td>
<td>0.25</td>
<td>0.23</td>
<td>0.21</td>
</tr>
<tr>
<td>0.5</td>
<td>0.19</td>
<td>0.21</td>
<td>0.15</td>
<td>0.17</td>
<td>0.14</td>
<td>0.22</td>
<td>0.19</td>
<td>0.19</td>
</tr>
<tr>
<td>1</td>
<td>0.18</td>
<td>0.18</td>
<td>0.19</td>
<td>0.20</td>
<td>0.20</td>
<td>0.22</td>
<td>0.19</td>
<td>0.20</td>
</tr>
<tr>
<td>2</td>
<td>0.24</td>
<td>0.25</td>
<td>0.30</td>
<td>0.30</td>
<td>0.36</td>
<td>0.28</td>
<td>0.25</td>
<td>0.27</td>
</tr>
<tr>
<td>5</td>
<td>0.54</td>
<td>0.67</td>
<td>0.72</td>
<td>0.72</td>
<td>0.97</td>
<td>0.52</td>
<td>0.54</td>
<td>0.55</td>
</tr>
<tr>
<td>10</td>
<td>1.14</td>
<td>1.61</td>
<td>1.54</td>
<td>1.53</td>
<td>2.11</td>
<td>0.97</td>
<td>1.11</td>
<td>1.08</td>
</tr>
<tr>
<td>20</td>
<td>2.50</td>
<td>3.78</td>
<td>3.36</td>
<td>3.34</td>
<td>4.52</td>
<td>1.91</td>
<td>2.39</td>
<td>2.23</td>
</tr>
<tr>
<td>50</td>
<td>7.07</td>
<td>10.92</td>
<td>9.36</td>
<td>9.29</td>
<td>11.94</td>
<td>4.80</td>
<td>6.62</td>
<td>5.87</td>
</tr>
<tr>
<td>100</td>
<td>15.38</td>
<td>23.24</td>
<td>20.00</td>
<td>19.86</td>
<td>24.41</td>
<td>9.67</td>
<td>14.22</td>
<td>12.20</td>
</tr>
</tbody>
</table>

Table 5.5: Asymptotic variances of $\hat{\lambda}_{U,\alpha}$ and $\hat{\lambda}_{L,\alpha}$ for various bivariate copulas with Kendall’s $\tau$ equal to 0.5. The estimators are defined in (5.16), with the population version of $\hat{\lambda}_{U,\alpha}$ being $\lambda_{\alpha}(C)$ for copula $C$, and that of $\hat{\lambda}_{L,\alpha}$ being $\lambda_{\alpha}(\hat{C})$. “HR” and “rMTCJ” stand for the Hüsler-Reiss and reflected MTCJ copulas, respectively.

and therefore

$$\sqrt{n} \left( \hat{\lambda}_{\alpha,r} - \lambda_{\alpha} \right) = -\sqrt{n} \alpha \left( \frac{1}{\alpha/(\alpha + 1) - \hat{\nu}_{\alpha,r}} - \frac{1}{\alpha/(\alpha + 1) - \nu_{\alpha}} \right)$$

$$= -\alpha \left( \frac{\sqrt{n} (\hat{\nu}_{\alpha,r} - \nu_{\alpha})}{[\alpha/(\alpha + 1) - \hat{\nu}_{\alpha,r}][\alpha/(\alpha + 1) - \nu_{\alpha}]} \right)$$

$$= r_{\alpha} \left[ \sqrt{n} (\hat{\nu}_{\alpha,r} - \nu_{\alpha}) \right] + o_p(1),$$

where $r_{\alpha} = -\alpha[\alpha/(\alpha + 1) - \nu_{\alpha}]^{-2} = -\alpha \gamma_{\alpha}^{-2}$. As a result,

$$\sqrt{n} \left( \hat{\lambda}_{\alpha,r} - \lambda_{\alpha} \right) \overset{d}{\to} N \left( 0, \frac{(\alpha + 2 - \lambda_{\alpha})^4}{\alpha^2 \text{Var}(X)} \right),$$

where the multiplicative constant in the variance is the same as $[h'(\nu_{\alpha})]^2$ in (5.19).
As we remarked in Section 5.1.3, the asymptotic variance is usually a 2-dimensional integral that must be evaluated numerically. Table 5.6 displays the asymptotic variances of the rank-based estimators \( \hat{\lambda}_{U,\alpha,r} \) and \( \hat{\lambda}_{L,\alpha,r} \), evaluated using the expressions in Appendix A. Compared to the estimator assuming known margins, it appears that the general pattern of the asymptotic variances attaining the smallest values with \( \alpha \approx 1 \) continues to hold. With smaller \( \alpha \), the asymptotic variance of the rank-based estimators is higher than that of the estimator assuming known margins, while the reverse is true when \( \alpha \) is larger (e.g., \( \alpha \geq 5 \)).

Table 5.6: Asymptotic variances of \( \hat{\lambda}_{U,\alpha,r} \) and \( \hat{\lambda}_{L,\alpha,r} \) for various bivariate copulas with Kendall’s \( \tau \) equal to 0.5. The estimators are defined in (5.16) but are based on marginal ranks; the population version of \( \hat{\lambda}_{U,\alpha,r} \) is \( \lambda_{\alpha}(C) \) for copula \( C \), and that of \( \hat{\lambda}_{L,\alpha,r} \) is \( \lambda_{\alpha}(\hat{C}) \). “HR” and “rMTCJ” stand for the Hüsler-Reiss and reflected MTCJ copulas, respectively.

The expressions for the asymptotic variance can be greatly simplified if \( C \) is the independence copula. Here we outline the derivations; the more tedious parts of the proof are relegated to Appendix A. When \( C(u_1, u_2) = u_1 u_2 \), we have

\[
G_C(u_1, u_2) = \mathbb{E}_C(u_1, u_2) - u_2 \mathbb{E}_C(u_1, 1) - u_1 \mathbb{E}_C(1, u_2),
\]
where

\[ E[B_C(u_1, u_2)B_C(u_3, u_4)] = \left( u_1 \wedge u_3 \right) \left( u_2 \wedge u_4 \right) - u_1 u_2 u_3 u_4. \]

With this, the covariance function of \( G_C \) is given by

\[ E[G_C(u_1, u_2)G_C(u_3, u_4)] = \left( u_1 \wedge u_3 - u_1 u_3 \right) \left( u_2 \wedge u_4 - u_2 u_4 \right), \]

which allows the computation of the variance of \( X \) (after some simplification), as

\[
\text{Var}(X) = E(X^2) = \int_0^1 \int_0^1 E \left[ G_C \left( u^{1/\alpha}, u^{1/\alpha} \right) G_C \left( v^{1/\alpha}, v^{1/\alpha} \right) \right] \, du \, dv
\]

so that

\[
\sqrt{n} \left( \hat{\lambda}_{\alpha,r} - \lambda_{\alpha} \right) \overset{d}{\to} N \left( 0, \sigma^2 \right),
\]

where

\[
\sigma^2 = \frac{(2 + \alpha)^4}{\alpha^2} \cdot \frac{\alpha^2}{(2 + \alpha)^2 (3 + 5\alpha + 2\alpha^2)} = \frac{(2 + \alpha)^2}{(1 + \alpha)(3 + 2\alpha)}.
\]

Interestingly, this asymptotic variance is a decreasing function in \( \alpha \), from \( 4/3 \) when \( \alpha \to 0^+ \) to \( 1/2 \) as \( \alpha \to \infty \). From this, we can see that the (rank-based) F-madogram measure of dependence has smaller variability than the tail-weighted dependence measures of Krupskii and Joe (2015) under independence, where it is proved that the asymptotic variance of the latter is \( 1/p^2 \), with \( 0 < p \leq 0.5 \) being the truncation threshold.

### 5.2.5 Extension to higher dimensions

One advantage of the F-madogram estimator is that it can be easily extended to higher dimensions. Note that

\[
\hat{\nu}_\alpha^{(2)} = \frac{1}{2n} \sum_{i=1}^n |U_{i1}^\alpha - U_{i2}^\alpha| = \frac{1}{n} \sum_{i=1}^n \left[ \max(U_{i1}^\alpha, U_{i2}^\alpha) - \frac{1}{2}(U_{i1}^\alpha + U_{i2}^\alpha) \right],
\]

where the bracketed superscript on \( \hat{\nu} \) denotes the dimension of the underlying data. This latter form allows a generalization to higher dimensions, considered in Marcon et al. (2016). Let \( m_i \) be the index of the largest observation in the vector \( (U_{i1}, \ldots, U_{id}) \), then define

\[
\hat{\nu}_\alpha^{(d)} = \frac{1}{dn} \sum_{i=1}^n \sum_{j=1}^d (U_{i,m_i}^\alpha - U_{i,j}^\alpha) = \frac{1}{n} \sum_{i=1}^n \left[ \max(U_{i1}^\alpha, \ldots, U_{id}^\alpha) - \frac{1}{d}(\sum_{j=1}^d U_{ij}^\alpha) \right].
\]

This can be interpreted as the average distance between the maximum and other components of the vector. For an arbitrary \( d \)-dimensional copula \( C \), the population version of \( \hat{\nu}_\alpha^{(d)} \)
is given by
\[
\nu^{(d)}_\alpha = \mathbb{E} \left[ \max(U_{i1}^{\alpha}, \ldots, U_{id}^{\alpha}) - \frac{1}{d} \sum_{j=1}^{d} U_{ij}^{\alpha} \right] = \frac{\alpha}{\alpha + 1} - \int_{0}^{1} C(u^{1/\alpha}, \ldots, u^{1/\alpha}) \, du
\]
\[
\triangleq \frac{\alpha}{\alpha + 1} - \gamma^{(d)}_\alpha
\]
similar to (5.13). Note that \( \gamma^{(d)}_\alpha = \alpha / (\alpha + d) \) for the independence copula and by defining
\[
\lambda^{(d)}_\alpha = \frac{d + \alpha \left[ 1 - 1/\gamma^{(d)}_\alpha \right]}{d - 1},
\]
a measure of dependence can be obtained; it satisfies the conditions \( \lambda^{(d)}_\alpha = 0 \) for the independence copula and \( \lambda^{(d)}_\alpha = 1 \) for the comonotonicity copula.

5.2.6 Potential future research

With a similar derivation, it is also possible to express the CFG estimator as an integral of the copula and its survival function along the diagonal. It may thus be possible to utilize the estimator as a dependence measure for general bivariate copulas. However, the expressions are more involved and are not as easily interpretable as that for the F-madogram.

A direction for future work could be to modify estimators for the extremal coefficient to general measures of dependence that reveal other characteristics of a copula. For example, by considering different powers of exponentiation for each margin of the F-madogram estimator (e.g., \( \alpha \) for the first margin, \( \beta \) for the second), \( \gamma \) becomes an integral not along the diagonal but \( (u^{1/\alpha}, u^{1/\beta}) \), for \( u \in [0, 1] \). It may thus be possible to devise a measure on permutation symmetry by considering the difference along \( (u^{1/\alpha}, u^{1/\beta}) \) and \( (u^{1/\beta}, u^{1/\alpha}) \).
Chapter 6

Assessing model adequacy based on empirical and fitted features

An important part of data modelling is to ensure that the fitted parsimonious models, which impose certain restrictions on the relationship between variables, capture sufficient characteristics expressed by the data. In Chapters 6 and 7, we turn our attention to model evaluation. In particular, we make use of various features that compare the characteristics exhibited by the data and those implied by the fitted models. We demonstrate the limiting behaviour of this comparison statistic in this chapter and illustrate how we can construct a range of flexible statistics that measure the amount and source of misfit in terms of the features desired. This forms the theoretical background for the methods described in Chapter 7, where we consider the evaluation of the adequacy-of-fit of a general multivariate copula.

6.1 Introduction

We first provide an overview of and the intuition behind the discrepancy statistic based on the difference between the empirical and fitted parametric distributions, and their functional extensions. A connection to some existing work on goodness-of-fit procedures using the Kolmogorov-Smirnov and Cramér-von Mises-type tests is then made.

6.1.1 Vector of differences between empirical and model-based features

A statistical model is constructed as a tractable mathematical representation of a system that allows one to gain understanding on its mechanism and, in some cases, to assist in making reasonable prediction in the future. With this objective in mind, one way to assess
model adequacy is to compare how close the fitted model is to the data being modelled. This is especially relevant when the researcher thinks a decent model has been obtained, for instance through exploring the structure of the data or from subject knowledge, but would like to seek guidance as to whether it is good enough.

For a sample of (possibly multivariate) i.i.d. observations \( Y_1, \ldots, Y_n \sim F \) and parametric model \( G \) with parameter vector \( \theta \), differences of the form \( T(\hat{F}_n) - T\left[ G(\cdot; \hat{\theta}_n) \right] \) are an intuitive measure of discrepancy between the data and the fitted model, where \( \hat{F}_n \) is the empirical distribution of \( F \), \( \hat{\theta}_n \) is a model-based estimate of the parameter vector for \( G \), and \( T \) is a functional taking a distribution function as its argument. This difference can also be called residual as it measures the lack-of-fit between the data and the model. We are interested in the difference statistic

\[
D_n = \begin{pmatrix}
T_1(\hat{F}_n) - T_1\left[ G(\cdot; \hat{\theta}_n) \right] \\
T_2(\hat{F}_n) - T_2\left[ G(\cdot; \hat{\theta}_n) \right] \\
\vdots \\
T_m(\hat{F}_n) - T_m\left[ G(\cdot; \hat{\theta}_n) \right]
\end{pmatrix} \triangleq T_{n,\text{emp}} - T_{n,\text{mod}},
\]

(6.1)

where \( m \) denotes the number of functionals or features\(^\text{12}\) being compared. In evaluating the adequacy of dependence modelling, some possible choices of \( T \) are measures of dependence such as Kendall’s tau for \((j,k)\) margin, \( \tau_{jk} \), with \( T(F) = 4 \int_{\mathbb{R}^2} F_{jk}(y) \, dF_{jk}(y) - 1 \) and Spearman’s rho for \((j,k)\) margin, \( \rho_{jk} \), with \( T(F) = 12 \int_{\mathbb{R}^2} F_j(y_1) F_k(y_2) \, dF_{jk}(y_1, y_2) - 3 \), where \( F_S \) is the marginal distribution of the variable(s) in the set \( S \). Other choices of \( T \) that may be helpful in evaluating model adequacy include the \( p \)th moment with \( T(F) = \int_{\mathbb{R}^d} y^p \, dF(y) \), or the value of distribution function at a given point \( t \) with \( T(F) = F(t) \).

In the discrete case of contingency tables with a total of \( m \) cells, we can take \( T_j(\hat{F}_n) \) as the sample proportion for cell \( j \) and \( T_j\left[ G(\cdot; \hat{\theta}_n) \right] \) as its model-based counterpart, \( j = 1, \ldots, m \). This makes \( D_n \) the difference between observed and expected cell probabilities; it is the building block of Pearson’s \( \chi^2 \) statistic. An adequate or good parametric statistical model should be such that the magnitude of the difference is small. In higher-dimensional problems, Pearson’s \( \chi^2 \) statistic becomes impractical due to the sparsity of observations, resulting in small expected cell probabilities. Maydeu-Olivares and Joe (2005, 2006) propose the use of low-order marginal tables to bypass sparsity issues. In this case, each \( T_j \) is the proportion or probability for one cell of a low-order (typically bivariate) marginal table. This concept is further extended in Joe and Maydeu-Olivares (2010) where linear combinations of the (marginal) cell probabilities are considered.

\(^{12}\)The functionals \( T_j \) can be considered as the aspects of the distribution one wants to examine, and hence they can also be referred to as features.
The limiting distributions of those statistics (based on low-order marginal cell probabilities and their linear combinations) have been derived. In particular, there are expressions for the computation of the limiting covariance matrix of the difference statistic. However, it is not always computationally easy to obtain this matrix, for example if the dimension of the table is high resulting in a large number of low-order margins (i.e., when $m$ is large). The implementation of the theory is impractical for high-dimensional models that are not closed under margins, such that low-order marginal model-based cell probabilities cannot be obtained without aggregating the high-dimensional cell probabilities. These issues carry over to the continuous case, but there are extra factors that make this problem much harder for general continuous distributions. For instance, expressions for the limiting covariance matrix depend on the features $T_j$ used (these are no longer cell probabilities) and the method of model estimation. It is generally undesirable to discretize continuous distributions, as important features (such as tail properties) can be lost and the resulting statistic may lose its ability to signal model inadequacy. Similar to the discrete case, we want to use features that take low-dimensional distributions as inputs to bypass sparsity in higher dimensions. When the model is not closed under margins, it can be computationally intensive to obtain the model-based feature due to the lack of a tractable low-dimensional distribution.

In this chapter, we show that $\sqrt{n}D_n \xrightarrow{d} N(0, \Sigma)$ for the class of features being U-statistics under correct model specification, when $\theta$ is estimated using maximum likelihood as well as the more general case of an $\sqrt{n}$-consistent estimator that is the solution of a set of estimating equations. This is a fundamental result for us to investigate methods to assess adequacy-of-fit of multivariate distributions in Chapter 7. If $\hat{\theta}_n$ is the maximum likelihood estimator, then the asymptotic covariance matrix $\Sigma$ depends on the Fisher information matrix and the gradient of the model-based features with respect to the parameter. It can also be shown that the asymptotic covariance matrix of $\sqrt{n}T_{n, emp}$ is “larger” than that of $\sqrt{n}T_{n, mod}$, in the sense that the difference is equal to $\Sigma$, a positive semi-definite matrix. In plain terms, this means that the empirical feature is more variable than both the model-based counterpart as well as the difference, whose asymptotic variance can be broken down into the difference of the respective asymptotic variances. This is an important observation because, by not having to consider the cross covariance, the derivation of asymptotic properties for the difference statistic is greatly simplified when the model is correctly specified. It is useful to obtain the asymptotic behaviour of the differences because this gives us an idea as to what variability we can expect if the data generating mechanism reasonably follows the assumed model, thereby providing a guide as to whether the model is adequate.

Usually, obtaining the limiting behaviour of the vector of differences is an intermediate
step in model evaluation. Using the vector $D_n$, we can construct quadratic form statistics $Q_n$ that summarize the discrepancies for different features. The following lists some possible constructions:

1. The first possibility is to simply use the square of each element, so that $Q_n = \frac{n}{m} D_n^\top D_n$ after scaling by the number of observations and features. This is the (scaled) mean squared discrepancy among the features considered, and gives equal weight to each element. A closely related quantity is $\sqrt{m^{-1} D_n^\top D_n}$, known as the standardized root mean squared residual (SRMSR) which is explained in more detail in Section 7.1. With this formulation, it is necessary to approximate the limiting distribution (to which $Q_n$ converges as $n \to \infty$) using methods such as moment matching.

2. If $\Sigma$ is known or can be computed accurately, then another choice is $Q_n = n D_n^\top \Sigma^{-1} D_n$ or $Q_n = n D_n^\top \Sigma^{-} D_n$ for singular $\Sigma$, where $\Sigma^{-}$ is its generalized inverse. This statistic converges to a $\chi^2$ distribution as $n \to \infty$. In addition to having a convenient limiting distribution, this formulation is also useful when the features are of substantially different magnitudes, so that standardization is desirable.

3. Alternatively, one may only scale the individual terms by considering the statistic $Q_n = n D_n^\top [\text{diag}(\Sigma)]^{-1} D_n$; one such example is in Bartholomew and Leung (2002). The limiting distribution can be written as a mixture of independent chi-squared variables. Moment matching can be used, or the tail probabilities of the limiting distribution can be evaluated numerically, see, e.g., Rice (1980).

Much of the exposition in this chapter focuses on the behaviour of $D_n$; we develop decision criteria based on the asymptotic distribution of $\sqrt{n} D_n$. The quadratic form statistic $Q_n$ will be revisited in Section 6.5 where its use in evaluating model adequacy is discussed. Regarding the construction of $D_n$, although the theoretical derivations in this chapter mainly focus on the class of U-statistics with the parametric model fitted by means of estimating equations that yield $\sqrt{n}$-consistent and asymptotically normal estimators, we consider a wider class of empirical features and model estimation methods in Chapter 7:

- Empirical features: (A1) U-statistics; (A2) rank-based F-madogram estimator of the extremal coefficient, and; (A3) rank-based tail-weighted dependence measures.

- Model estimation methods: (B1) Maximum likelihood; (B2) estimating equations; (B3) estimation of dependence parameters with marginal ranks.

In this chapter, we demonstrate the asymptotic normality of the combination (A1) and (B1), as well as (A1) and (B2), by decomposing the U-statistic and the expansion of the
model-based estimator into sums of i.i.d. random variables. Some of the results in Chapter 7 depend also on the normality of other combinations; we acknowledge that this is still open to further research, and in Section 6.4 we mention some technical details and make conjectures for results not yet rigorously demonstrated in the current work.

6.1.2 Relevance to goodness-of-fit tests

Statistics used for evaluating goodness-of-fit based on the difference between distribution functions have a long history. The Kolmogorov-Smirnov, Cramér-von Mises and Anderson-Darling statistics all belong to this category. For testing the simple hypothesis that a univariate data set belongs to a completely specified model $F_0(t)$, these statistics are defined as

$$\text{KS}_n = \sup_{t \in \mathbb{R}} |\hat{F}_n(t) - F_0(t)|;$$
$$\text{CvM}_n = \int_{-\infty}^{\infty} \left[\hat{F}_n(t) - F_0(t)\right]^2 dF_0(t);$$
$$\text{AD}_n = \int_{-\infty}^{\infty} \left[\hat{F}_n(t) - F_0(t)\right]^2 F_0(t) \left[1 - F_0(t)\right] dF_0(t),$$

respectively. If $F_0$ is continuous, it can be proved that the sampling distributions of these statistics do not depend on $F_0$, and thus they are “distribution-free”. However, in practice $F_0(t)$ is usually unknown, and therefore it is often more practical to consider the composite hypothesis that the data follow a certain parametric family of distributions. Direct modification to the test statistics above amounts to changing the completely specified $F_0(t)$ to the fitted distribution $G(t; \hat{\theta}_n)$, as defined at the beginning of this section. However, the resulting asymptotic distributions under correct model specification generally depend on $G$ and $\theta$. Some early work tackling goodness-of-fit tests with estimated parameters includes Chernoff and Lehmann (1954) in the discrete setting for $\chi^2$ tests, and Darling (1955) which has the result for general absolutely continuous distribution; see also Chapter 7 of Durbin (1973). Because these statistics often have intractable limiting distributions, they were not very useful at that time when computational power was limited; results in the form of tables of quantiles were only available for specific models of interest, see, e.g., Durbin (1975) and Stephens (1976). Chapter 28 of DasGupta (2008) has a summary of the relevant texts and developments on testing composite null hypotheses.

Goodness-of-fit tests for multivariate copulas are generally not distribution-free. Nevertheless, due to the increasing emphasis on multivariate modelling and the advent of technology which made analogous procedures viable, tests based on Kolmogorov-Smirnov and
Cramér-von Mises-type statistics as well as Rosenblatt’s transformation have been studied, see, e.g., Genest et al. (2009, 2013), Berg (2009) and the references therein.

By taking $T_i$ as the distribution function at $t_i$, (6.1) can be written as a vector of differences of distribution functions at various points on its support. The vector $D_n$ is thus connected to these goodness-of-fit statistics when the model is estimated. The connection is the most obvious with the Kolmogorov-Smirnov statistic given by $\sup_{t \in \mathbb{R}} |\hat{F}_n(t) - G(t; \hat{\theta}_n)|$; this is the supremum of the absolute value of $D_n$ when viewed as a process, i.e., with $D_n$ having infinite number of elements within the support of the distributions. Meanwhile, with the Cramér-von Mises statistic, we consider its generalization by applying a weight function $\tilde{w}(t)$ to the integrand. For absolutely continuous $G$ with density $g$, this generalization is given by

$$GCV_M = \int_{-\infty}^{\infty} \tilde{w}(t) \left[ \hat{F}_n(t) - G(t; \hat{\theta}_n) \right]^2 dG(t; \hat{\theta}_n) = \int_{-\infty}^{\infty} w(t; \hat{\theta}_n) \left[ \hat{F}_n(t) - G(t; \hat{\theta}_n) \right]^2 dt$$

where $w(t; \hat{\theta}_n) = \tilde{w}(t)g(t; \hat{\theta}_n)$, $\Delta t_{im} = (t_{im} - t_{i-1,m})/m$ and $\{t_{0m}, t_{1m}, \ldots\}$ is an increasing sequence of points on $\mathbb{R}$ that gets denser as $m$ increases. Written this way, (6.2) becomes a weighted sum of infinitely many values of the squared differences at the $t_{im}$’s; in matrix notation, this is equal to $D_n^\top W_n D_n$, a quadratic form in $D_n$, where $W_n$ is a diagonal matrix with elements $w(t_{1m}; \hat{\theta}_n)\Delta t_{1m}, \ldots, w(t_{mm}; \hat{\theta}_n)\Delta t_{mm}$. Since the Anderson-Darling statistic is itself a generalized Cramér-von Mises statistic, this decomposition applies as well.

Note how this resembles the quadratic form statistic $Q_n$ illustrated above, in particular the third construction where here $W_n$ takes the place of diag($\Sigma$). This also illustrates some differences between our current work and the Cramér-von Mises statistic, where the dimensionality of $D_n$ is infinite in the latter while typically finite when elements of $D_n$ are some features under consideration. Also, for the Cramér-von Mises statistic, the weight matrix $W_n$ depends on the estimator $\hat{\theta}_n$ (and hence the sample size $n$), whereas the $Q_n$ considered above is constructed using the asymptotic covariance matrix $\Sigma$. A note valid to all three classical statistics is that they are not suggestive of direction for model improvement in case the null hypothesis is rejected, as the locations and patterns of maximal differences are lost by construction of the statistics.

Although much of the literature in this area has focused on goodness-of-fit, in this and the next chapters we draw our attention to a computationally similar, but conceptually different idea which we call adequacy-of-fit. With the former, the emphasis is on testing whether the data belong to a particular parametric family. This could be useful when the model is scientifically driven or when a gold standard exists, for example in approximate
simulations where one wants to check if the simulated data set is close to the target model. More often, however, there is no target model that could be deemed the truth and the parametric model is merely constructed to help explain the underlying structure or mechanism of the data. In this case, our focus is whether the model is adequate for the intended purpose. An application of this philosophy in the time series context can be found in Tsay (1992); see, e.g., Gelman et al. (1996); Ray and Lindsay (2008) for similar remarks. As a result, we will not suggest hypothesis tests aimed at proving the validity (or invalidity) of statistical models. Instead, we develop guidelines to intuitively assess model adequacy based on the discrepancy between empirical and model-based features.

The rest of this chapter is organized as follows. We derive the asymptotic behaviour of the difference statistic under correct model specification in Section 6.2. We first start with distribution functions and then extend the results to functionals, specifically the class of U-statistics. The properties under model misspecification are dealt with in Section 6.3. Some comments on the difference statistic are given in Section 6.4. In Section 6.5, we develop a decision criteria for model adequacy based on the difference statistic and discuss factors that may affect its quality for this purpose. This section serves as a bridge between Chapters 6 and 7.

6.2 Asymptotics of the difference vector for a correctly specified model

In this section, we obtain the asymptotic properties of $D_n$, the vector of differences defined in (6.1). We start with the difference between the empirical and fitted distribution functions when the model is correctly specified, and then generalize the results to the difference of a wider class of functionals $T$ (U-statistics). Throughout this section, we let $Y_1, \ldots, Y_n$ be a random sample in $\mathbb{R}^d$ from the distribution $F$, and $G(\cdot; \theta)$ be a parametric model to which the sample is fitted, with density $g$ and vector parameter $\theta \in \Theta$. Denote $\hat{F}_n(t) = n^{-1} \sum_{i=1}^n 1\{Y_i \leq t\}$ as the empirical distribution at $t$ and $\ell(\theta; y) = \log g(y; \theta)$ as the log-likelihood function for one observation. Unless otherwise specified, comparison operators on vectors are applied elementwise. As for $\hat{\theta}_n$, the maximum likelihood estimator is considered first, followed by $\sqrt{n}$-consistent estimators resulting from solving a system of estimating equations.

The following lists the assumptions to be used in various proofs in this chapter.

**Assumption A1.** The parametric model $G$ is correctly specified, in the sense that there exists $\theta_0 \in \text{int}(\Theta)$ such that $F(t) = G(t; \theta_0)$ for all $t \in \mathbb{R}^d$. 
Assumption A2 (R1 in Serfling (1980), page 144). For every $y \in \mathbb{R}^d$, the log-likelihood function is thrice differentiable with respect to every $\theta \in \text{int}(\Theta)$, i.e.,

$$
\begin{align*}
\frac{\partial \ell(\theta; y)}{\partial \theta_i}, \quad & \frac{\partial^2 \ell(\theta; y)}{\partial \theta_i \partial \theta_j}, \quad & \frac{\partial^3 \ell(\theta; y)}{\partial \theta_i \partial \theta_j \partial \theta_k}
\end{align*}
$$

exist for all values of $\theta_i, \theta_j, \theta_k$, three (potentially identical) elements of $\theta$.

Assumption A3. The maximum likelihood estimator, $\hat{\theta}_n = \arg \max_{\theta} \sum_{i=1}^n \ell(\theta; y_i)$, solves the score equation

$$
\frac{\partial}{\partial \theta} \sum_{i=1}^n \ell(\hat{\theta}_n; y_i) = 0.
$$

Assumption A4 (R2 in Serfling (1980), page 144). For every $\theta_i, \theta_j$ and $\theta_k$, three elements in the interior of $\Theta$, there exists a neighbourhood $N(\theta_i), N(\theta_j)$ and $N(\theta_k)$ such that the functions

$$
\begin{align*}
\left| \frac{\partial g(y; \theta)}{\partial \theta_i} \right| & \leq k_1(y), \quad & \left| \frac{\partial^2 g(y; \theta)}{\partial \theta_i \partial \theta_j} \right| & \leq k_2(y), \quad & \left| \frac{\partial^3 \ell(\theta; y)}{\partial \theta_i \partial \theta_j \partial \theta_k} \right| & \leq k_3(y)
\end{align*}
$$

for all $y$, and that $\int k_1(y) \, dy < \infty$, $\int k_2(y) \, dy < \infty$ and $\int k_3(y) g(y; \theta) \, dy < \infty$ within these neighbourhoods.

Assumption A5 (R3 in Serfling (1980), page 145). For every $\theta \in \text{int}(\Theta)$, the expectation

$$
\mathbb{E}_\theta \left[ \frac{\partial \ell(\theta; Y)}{\partial \theta} \frac{\partial \ell(\theta; Y)}{\partial \theta^\top} \right]
$$

exists and is non-singular, where $\mathbb{E}_\theta$ stresses that the expectation is taken with respect to the distribution $Y \sim G(\cdot; \theta)$.

The assumptions above guarantee the existence of the Taylor series expansion of the maximum likelihood estimator and ensure that various differentiation procedures under the integral sign are valid. In the rest of this chapter, we use the following lemma to demonstrate asymptotic normality of the difference statistic and obtain its asymptotic variance.

Lemma 6.1. Suppose we can decompose a $d$-dimensional random vector $X_n$ into

$$
X_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n S_i + \zeta_n,
$$

where $S_1, \ldots, S_n$ are i.i.d. random vectors with mean $\mu$ and covariance matrix $\Sigma_S$, and $\zeta_n = o_p(1)$. Then $X_n \xrightarrow{d} N(\mu, \Sigma_S)$.

Proof. The result follows directly from the central limit theorem and Slutsky’s theorem. \qed
6.2.1 Behaviour of the difference between empirical and fitted distribution functions

In the following, we establish the limiting behaviour of the difference process \( \hat{F}_n(t) - G(t; \hat{\theta}_n) \). Note that this has been considered in Darling (1955), but we include the details here as they contain definitions of notations and illustrate methods that we will use in later sections. To simplify notation, let \( \ell'() \triangleq \frac{\partial \ell(\cdot)}{\partial \theta} \) and \( \ell''() \triangleq \frac{\partial^2 \ell(\cdot)}{\partial \theta \partial \theta^\top} \), and let

\[
I \triangleq I(\theta_0) = -\mathbb{E}[\ell''(\theta_0; Y_1)]
\]

be the Fisher information matrix.

Theorem 6.2. Let \( Y_1, \ldots, Y_n \) be a random sample from \( G(\cdot; \theta_0) \). Assume A1 to A5. For any given vectors \( t_1, t_2, \ldots, t_m \in \mathbb{R}^d \), the vector

\[
\sqrt{n} \begin{bmatrix} \hat{F}_n(t_1) - G(t_1; \hat{\theta}_n) \\ \vdots \\ \hat{F}_n(t_m) - G(t_m; \hat{\theta}_n) \end{bmatrix} \xrightarrow{d} N(0, \Sigma),
\]

where \( \Sigma = (\sigma_{jk}) \) contains the elements

\[
\sigma_{jk} = \sigma_{kj} = G(t_j; \theta_0) [1 - G(t_k; \theta_0)] - \frac{\partial G}{\partial \theta^\top}(t_j; \theta_0) I^{-1} \frac{\partial G}{\partial \theta}(t_k; \theta_0), \quad 1 \leq j \leq k \leq m,
\]

with \( I \) defined in (6.4).

Proof. The expansion of the score function about the true value \( \theta_0 \) gives

\[
\sqrt{n} (\hat{\theta}_n - \theta_0) = I^{-1} Z(\theta_0) + O_p(n^{-1/2}),
\]

where \( Z(\theta_0) = n^{-1/2} \sum_{j=1}^{n} \ell'(\theta_0; Y_j) = O_p(1) \) and will be denoted by \( Z \) when there is no ambiguity. By expanding \( G(t; \hat{\theta}_n) \) about \( \theta_0 \), we obtain

\[
G(t; \hat{\theta}_n) = G(t; \theta_0) + \frac{\partial G}{\partial \theta^\top}(t; \theta_0) (\hat{\theta}_n - \theta_0) + \tilde{R}_n = G(t; \theta_0) \left( \frac{1}{\sqrt{n}} \frac{\partial G}{\partial \theta^\top}(t; \theta_0) I^{-1} Z + R_n, \right.
\]

where \( \tilde{R}_n \) and \( R_n \) are both \( o_p(n^{-1/2}) \). We can then rewrite the left hand side of (6.5) as

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{bmatrix} 1 \{ Y_i \leq t_1 \} - G(t_1; \theta_0) - \frac{\partial G}{\partial \theta^\top}(t_1; \theta_0) I^{-1} \ell'(\theta_0; Y_i) \\ \vdots \\ 1 \{ Y_i \leq t_m \} - G(t_m; \theta_0) - \frac{\partial G}{\partial \theta^\top}(t_m; \theta_0) I^{-1} \ell'(\theta_0; Y_i) \end{bmatrix} - \sqrt{n} \begin{bmatrix} R_{n1} \\ \vdots \\ R_{nm} \end{bmatrix},
\]

where \( \sqrt{n} (R_{n1}, \ldots, R_{nm})^\top = o_p(1) \). Let the term within the square brackets after the summation sign be denoted as \( S_i = (S_{i1}, \ldots, S_{im})^\top \). Then we have \( \mathbb{E}(S_i) = 0 \), while
Cov($S_i$) has elements

$$
\text{Cov}(S_{ij}, S_{ik}) = \text{Cov} \left[ 1 \{ Y_1 \leq t_j \}, \frac{\partial G}{\partial \theta^j}(t_j; \theta_0)I^{-1}\ell'(\theta_0; Y_1), 1 \{ Y_1 \leq t_k \} - \frac{\partial G}{\partial \theta^k}(t_k; \theta_0)I^{-1}\ell'(\theta_0; Y_1) \right]
$$

$$
= \text{Cov} \left[ 1 \{ Y_1 \leq t_j \}, 1 \{ Y_1 \leq t_k \} \right] - \text{Cov} \left[ 1 \{ Y_1 \leq t_j \}, \frac{\partial G}{\partial \theta^j}(t_j; \theta_0)I^{-1}\ell'(\theta_0; Y_1) \right]
$$

$$
- \text{Cov} \left[ 1 \{ Y_1 \leq t_k \}, \frac{\partial G}{\partial \theta^k}(t_k; \theta_0)I^{-1}\ell'(\theta_0; Y_1) \right]
$$

$$
+ \text{Cov} \left[ \frac{\partial G}{\partial \theta^j}(t_j; \theta_0)I^{-1}\ell'(\theta_0; Y_1), \frac{\partial G}{\partial \theta^k}(t_k; \theta_0)I^{-1}\ell'(\theta_0; Y_1) \right]. \quad (6.9)
$$

Now, Cov $\left[ 1 \{ Y_1 \leq t_j \}, 1 \{ Y_1 \leq t_k \} \right] = P(Y_1 \leq t_j \wedge t_k) - G(t_j; \theta_0)G(t_k; \theta_0)$ and is equal to $G(t_j; \theta_0) [1 - G(t_k; \theta_0)]$ when $j \leq k$. The second term is

$$
\text{Cov} \left[ 1 \{ Y_1 \leq t_j \}, \frac{\partial G}{\partial \theta^j}(t_j; \theta_0)I^{-1}\ell'(\theta_0; Y_1) \right]
$$

$$
= \frac{\partial G}{\partial \theta^j}(t_j; \theta_0)I^{-1}E \left[ 1 \{ Y_1 \leq t_j \} \ell'(\theta_0; Y_1) \right]
$$

$$
= \frac{\partial G}{\partial \theta^j}(t_j; \theta_0)I^{-1} \int_{-\infty}^{t_j} \frac{\partial g(y; \theta_0)}{\partial \theta_0} \cdot g(y; \theta_0) \, dy
$$

$$
= \frac{\partial G}{\partial \theta^j}(t_j; \theta_0)I^{-1} \frac{\partial G}{\partial \theta^j}(t_j; \theta_0),
$$

and is equal to the third term. The final term is also equal to the second term as

$$
\text{Cov} \left[ \frac{\partial G}{\partial \theta^j}(t_j; \theta_0)I^{-1}\ell'(\theta_0; Y_1), \frac{\partial G}{\partial \theta^k}(t_k; \theta_0)I^{-1}\ell'(\theta_0; Y_1) \right] = \frac{\partial G}{\partial \theta^j}(t_j; \theta_0)I^{-1} \frac{\partial G}{\partial \theta^k}(t_k; \theta_0).
$$

Hence, for $j \leq k$,

$$
\text{Cov}(S_{ij}, S_{ik}) = G(t_j; \theta_0) [1 - G(t_k; \theta_0)] - \frac{\partial G}{\partial \theta^j}(t_j; \theta_0)I^{-1} \frac{\partial G}{\partial \theta^k}(t_k; \theta_0).
$$

The proof is completed using Lemma 6.1.

The convergence to a mean zero Gaussian distribution can be proved for the more general case where $\hat{\theta}_n$ is a $\sqrt{n}$-consistent estimator that is the solution of a system of estimating equations. We first replace the assumptions in the maximum likelihood context to those applicable to estimating equations:

**Assumption B1.** Instead of the log-likelihood function, the Assumptions A2 to A5 are satisfied for a vector of inference functions $l_i$, such that the estimator $\hat{\theta}_n$ is a root of the estimating equation

$$
\frac{1}{n} \sum_{i=1}^{n} l(\theta; Y_i) = 0,
$$

118
with $\mathbb{E}[l(\theta_0; Y_i)] = 0$. With this, assumption A5 is modified such that $\mathbb{E}_\theta [l(\theta; Y)l^\top(\theta; Y)]$ and $\mathbb{E}_\theta [\partial l(\theta; Y)/\partial \theta^\top]$ exist and are not singular.

This more general estimator covers several model fitting methods, which are especially relevant to multivariate modelling, that will be addressed in Sections 6.4 and 7.8. One example of these is the composite likelihood estimator; in this case the inference functions $l$ are the sum of marginal or conditional score functions.

The estimating equation analogue of Theorem 6.2 is given below as Theorem 6.2′:

**Theorem 6.2′.** Let $Y_1, \ldots, Y_n$ be a random sample from $G(\cdot; \theta_0)$. Assume A1 and B1. For any given vectors $t_1, t_2, \ldots, t_m \in \mathbb{R}^d$, the vector

$$
\sqrt{n} \begin{bmatrix}
\hat{F}_n(t_1) - G(t_1; \hat{\theta}_n) \\
\vdots \\
\hat{F}_n(t_m) - G(t_m; \hat{\theta}_n)
\end{bmatrix} \xrightarrow{d} N(0, \Sigma_e),
$$

where $\Sigma_e$ is an $m \times m$ covariance matrix.

**Proof.** For $\sqrt{n}$-consistent estimators coming from estimating equations, a Taylor series expansion gives

$$
\sqrt{n} \left( \hat{\theta}_n - \theta_0 \right) = H^{-1} Z + O_p \left( n^{-1/2} \right),
$$

where $Z = n^{-1/2} \sum_{i=1}^n l(\theta_0; Y_i)$ and $H = \mathbb{E}[\partial l(\theta_0; Y_1)/\partial \theta^\top]$. The proof then follows in a similar fashion as in the proof of Theorem 6.2, in particular the analogous version of (6.8) becomes

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^n \begin{bmatrix}
1 \{ Y_i \leq t_1 \} - G(t_1; \theta_0) - \frac{\partial G}{\partial \theta^\top} (t_1; \theta_0) l(\theta_0; Y_i) \\
\vdots \\
1 \{ Y_i \leq t_m \} - G(t_m; \theta_0) - \frac{\partial G}{\partial \theta^\top} (t_m; \theta_0) l(\theta_0; Y_i)
\end{bmatrix} - \sqrt{n} \begin{bmatrix}
R_{n1} \\
\vdots \\
R_{nm}
\end{bmatrix},
$$

to which Lemma 6.1 is applied. However, in this case the elements of the covariance matrix cannot be written as (6.6) because the mixed empirical-fitted covariance terms (i.e., the second and third terms in (6.9)) are in general not the same as the model-based covariance (i.e., the last term in (6.9)).

These results can be easily extended to transformations of the distribution function; this is useful for features that are functions of them with the sample version estimated by a direct transformation of the empirical distribution function. The proof is given in Corollary 6.3.

**Corollary 6.3.** Assume A1 to A5 (or A1 and B1 in the case of estimating equations). For fixed vectors $t_1, t_2, \ldots, t_m \in \mathbb{R}^d$ and function $h : [0, 1]^m \rightarrow \mathbb{R}^m$ with continuous first derivative $\nabla h$ at $(G(t_1; \theta_0), \ldots, G(t_m; \theta_0))$, the vector

$$
\sqrt{n} \left[ h(\hat{F}_n) - h(G_{\theta_0}) \right] \xrightarrow{d} N(0, (\nabla h)^\top \Sigma (\nabla h)),
$$

where $\Sigma$ is an $m \times m$ covariance matrix.
where $\Sigma$ is the covariance matrix of the limiting distribution of $\sqrt{n} \left( \hat{F}_n - G_{\theta_0} \right)$, $\hat{F}_n = (\hat{F}_n(t_1), \ldots, \hat{F}_n(t_m))^\top$ and $G_{\theta} = (G(t_1; \theta), \ldots, G(t_m; \theta))^\top$.

**Proof.** The proof is similar to that of the delta method. From the mean value theorem, we have

$$h(\hat{F}_n) = h(G_{\theta_0}) + \nabla h(G_{\theta_0})^\top \left[ \hat{F}_n - G_{\theta_0} \right] + o_p(n^{-1/2})$$

and

$$h(G_{\theta_n}) = h(G_{\theta_0}) + \nabla h(G_{\theta_0})^\top \left[ G_{\theta_n} - G_{\theta_0} \right] + o_p(n^{-1/2}).$$

Hence $\sqrt{n} \left[ h(\hat{F}_n) - h(G_{\theta_n}) \right] = \sqrt{n} [\nabla h(G_{\theta_0})]^\top \left( \hat{F}_n - G_{\theta_n} \right) + o_p(1)$ and the result follows.

---

**6.2.2 Generalization to the difference for U-statistics**

Here, we show that a result analogous to Theorem 6.2 can be extended to the difference of functionals applied to the empirical and fitted distributions, with empirical functionals being U-statistics (Hoeffding (1948)). Let $T : F \to \mathbb{R}$ be a functional that maps from a multivariate distribution $F \in \mathcal{F}$ to a real number. For this result to hold, we need to apply the following smoothness assumption on $T$.

**Assumption A6.** The functional $T$ is twice differentiable with respect to every $\theta \in \text{int}(\Theta)$, i.e.,

$$\frac{\partial T}{\partial \theta_i} \quad \text{and} \quad \frac{\partial^2 T}{\partial \theta_i \partial \theta_j}$$

exist for all values of $\theta_i$ and $\theta_j$, two (potentially identical) elements of $\theta$.

If A6 is satisfied, then $T$ admits the Taylor series expansion

$$T \left[ G(:, \hat{\theta}_n) \right] = T \left[ G(:, \theta_0) \right] + \frac{1}{\sqrt{n}} \frac{\partial T}{\partial \theta} \left[ G(:, \theta_0) \right] \mathbf{I}^{-1} \mathbf{Z} + R_n$$

under maximum likelihood estimation, similar to (6.7), where $R_n$ is $o_p(n^{-1/2})$ and $\mathbf{I}$ is defined in (6.4).

**Theorem 6.4.** Let $Y_1, \ldots, Y_n$ be a random sample from $G(:, \theta_0)$. Assume A1 to A6 and suppose $T(\hat{F}_n)$ is a U-statistic, estimable of degree $r$, that can be written as

$$T(\hat{F}_n) = \left( \begin{array}{c} n \\ r \end{array} \right)^{-1} \sum_{\beta \in B} h(\mathbf{Y}_{\beta_1}, \ldots, \mathbf{Y}_{\beta_r}),$$

where $h$ is a symmetric kernel that is square-integrable with non-zero limiting variance, and $B$ is a set of cardinality $\left( \begin{array}{c} n \\ r \end{array} \right)$, with each element $\beta$ being one of the possible subsets of $r$.
integers from \{1, \ldots, n\}. Then

\[
\sqrt{n} \left( T(\hat{F}_n) - T \left[ G(:, \hat{\theta}_n) \right] \right) \overset{d}{\to} N(0, a-b),
\]  

(6.10)

where

\[
a = r^2 \text{Cov} [h(Y_1, \ldots, Y_r), h(Y_1, Y_{r+1}, \ldots, Y_{2r-1})];
\]

(6.11)

\[
b = \frac{\partial T}{\partial \theta^T} [G(:, \theta_0)] I^{-1} \frac{\partial T}{\partial \theta} [G(:, \theta_0)].
\]

(6.12)

**Proof.** By the definition of U-statistics, \( \mathbb{E} [T(\hat{F}_n)] = T[G(:, \theta_0)] \overset{d}{=} \gamma \). Define

\[
V_n^* = \frac{r}{n} \sum_{i=1}^{n} [h_1(Y_i) - \gamma]
\]

as in Lemma 3.3.8 of Randles and Wolfe (1979), where \( h_1(x) = \mathbb{E} [h(x, Y_2, \ldots, Y_r)] \) and the expectation is taken with respect to \( Y_2, \ldots, Y_r \). Theorem 3.3.13 in Randles and Wolfe (1979) has that \( n \mathbb{E} \left[ (T(\hat{F}_n) - \gamma - V_n^*)^2 \right] \to 0 \) as \( n \to \infty \); this implies the convergence in probability \( \sqrt{n} \left[ T(\hat{F}_n) - \gamma - V_n^* \right] = o_p(1) \), and the following decomposition into a sum of i.i.d. random variables, to which Lemma 6.1 is applied:

\[
\sqrt{n} \left( T(\hat{F}_n) - T \left[ G(:, \hat{\theta}_n) \right] \right) = \sqrt{n} \left[ V_n^* - \left( T \left[ G(:, \hat{\theta}_n) \right] - \gamma \right) \right] + o_p(1)
\]

\[
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ rh_1(Y_i) - r\gamma - \frac{\partial T}{\partial \theta^T} [G(:, \theta_0)] I^{-1} \ell'(\theta_0; Y_i) \right] + o_p(1)
\]

\[
\overset{d}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} S_i + o_p(1).
\]

It is obvious that \( \mathbb{E}(S_i) = 0 \). The variance of \( S \) is

\[
\text{Var}(S_i) = r^2 \text{Var} [h_1(Y_i)] + \frac{\partial T}{\partial \theta^T} [G(:, \theta_0)] I^{-1} \frac{\partial T}{\partial \theta} [G(:, \theta_0)]
\]

\[
-2r \frac{\partial T}{\partial \theta^T} [G(:, \theta_0)] I^{-1} \text{Cov} [h_1(Y_i), \ell'(\theta_0; Y_i)],
\]

where \( \text{Var} [h_1(Y_i)] = \text{Cov} [h(Y_1, \ldots, Y_r), h(Y_1, Y_{r+1}, \ldots, Y_{2r-1})] \) from equation (3.3.7) of Randles and Wolfe (1979), and

\[
\text{Cov} [h_1(Y_i), \ell'(\theta_0; Y_i)] = \mathbb{E} [h_1(Y_i) \ell'(\theta_0; Y_i)]
\]

\[
= \mathbb{E} [h(Y_1, \ldots, Y_r) \ell'(\theta_0; Y_1)] = \frac{1}{r} \mathbb{E} \left[ h(Y_1, \ldots, Y_r) \sum_{i=1}^{r} \ell'(\theta_0; Y_i) \right]
\]

\[
= \frac{1}{r} \int_{-\infty}^{\infty} h(y_1, y_2, \ldots, y_r) \prod_{i=1}^{r} g(y_i; \theta_0) \prod_{i=1}^{r} g(y_i; \theta_0) \int_{-\infty}^{\infty} \text{d}y_1 \cdots \text{d}y_r
\]

\[
= \frac{1}{r} \frac{\partial T}{\partial \theta} [G(:, \theta_0)],
\]

121
where the third equality uses the fact that \( h \) is symmetric about its arguments. Hence

\[
\text{Var}(S_i) = r^2 \text{Cov} [h(Y_1, \ldots, Y_r), h(Y_1, Y_{r+1}, \ldots, Y_{2r-1})] - \frac{\partial T}{\partial \theta} [G(\cdot; \theta_0)] I^{-1} \frac{\partial T}{\partial \theta} [G(\cdot; \theta_0)].
\]

\(\square\)

**Example 6.1.** We illustrate the result (6.10) with a simple example where \( G \) is the Poisson distribution with true parameter \( \theta_0 \), and the functional is the second moment \( E(Y^2) \). The functional transform is thus \( T[G(\cdot; \theta)] = \theta^2 + \theta \) with sample version \( T(\hat{F}_n) = n^{-1} \sum_{i=1}^n Y_i^2 \) (and thus with degree \( r = 1 \)). The maximum likelihood estimator is \( \hat{\theta}_n = n^{-1} \sum_{i=1}^n Y_i \stackrel{d}{=} \bar{Y} \), the score function is \( \ell'(\theta_0; Y) = \theta_0^{-1} Y - 1 \) and the Fisher information is \( I = \theta_0^{-1} \). The partial derivatives of the functional are

\[
\frac{\partial T}{\partial \theta} [G(\cdot; \theta)] = 2 \theta + 1; \quad \frac{\partial^2 T}{\partial \theta^2} [G(\cdot; \theta)] = 2; \quad \frac{\partial^k T}{\partial \theta^k} [G(\cdot; \theta)] = 0
\]

for \( k \geq 3 \). Since \( \sqrt{n} \left( \hat{\theta}_n - \theta_0 \right) = I^{-1} Z(\theta_0) \), we have

\[
\sqrt{n} \left( T(\hat{F}_n) - T \left[ G(\cdot; \hat{\theta}_n) \right] \right) = \sqrt{n} \left[ \frac{1}{n} \sum_{i=1}^n Y_i^2 - \bar{Y} (\bar{Y} + 1) \right]
\]

\[
= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[ Y_i^2 - \theta_0 (\theta_0 + 1) - (2\theta_0 + 1) \theta_0 \left( \frac{Y_i}{\theta_0} - 1 \right) \right] - \sqrt{n} (\bar{Y} - \theta_0)^2
\]

\[
\triangleq \frac{1}{\sqrt{n}} \sum_{i=1}^n S_i - \zeta_n. \tag{6.13}
\]

In this case, we can evaluate the variance of (6.13) directly:

\[
\text{Var} \left( \sqrt{n} \left[ T(\hat{F}_n) - T \left[ G(\cdot; \hat{\theta}_n) \right] \right] \right) = \text{Var}(S_1) + \text{Var}(\zeta_n) - 2 \sqrt{n} \text{Cov}(S_1, \zeta_n).
\]

By (6.11) and (6.12), the variance of \( S_i \) is \( a - b \) where \( a = \text{Var}(Y_i^2) = \theta_0 (1 + 6\theta_0 + 4\theta_0^2) \), \( b = \theta_0 (2\theta_0 + 1)^2 \) and \( a - b = 2\theta_0^2 \). For the variance of \( \zeta_n \), we have \( Y_+ \triangleq \sum_{i=1}^n Y_i \sim \text{Poisson}(n\theta_0) \) and thus

\[
\text{Var}(\zeta_n) = \text{Var} \left( \sqrt{n} (\bar{Y} - \theta_0)^2 \right) = n^{-3} \left\{ \mathbb{E} \left[ (Y_+ - n\theta_0)^4 \right] - \mathbb{E} \left[ (Y_+ - n\theta_0)^2 \right]^2 \right\}
\]

\[
= n^{-3} (n^2 \theta_0^2) \left( 3 + \frac{1}{n\theta_0} - 1 \right) = O(n^{-1}),
\]

and the covariance term is

\[
\text{Cov}(S_1, \zeta_n) = n^{-3/2} \mathbb{E} \left[ S_1 (Y_+ - n\theta_0)^2 \right] = n^{-3/2} \mathbb{E} \left[ S_1 \left( Y_1^2 + 2Y_1 \sum_{i=2}^n Y_i - 2n\theta_0 Y_1 \right) \right]
\]

\[
= n^{-3/2} \mathbb{E} \left[ S_1 Y_1^2 + 2(n - 1)\theta_0 S_1 Y_1 - 2n\theta_0 S_1 Y_1 \right] = O(n^{-3/2}),
\]

122
where we make use of the properties that $\mathbb{E}(S_1) = 0$ and that $S_1$ is independent of $Y_j$, $j \neq 1$. This means that both $\text{Var}(\zeta_n)$ and $2\sqrt{n}\text{Cov}(S_1, \zeta_n)$ are of order $O(n^{-1})$ and converge to zero as $n \to \infty$. Therefore $\lim_{n \to \infty} \text{Var} \left[ \frac{1}{\sqrt{n}} \left( T(\hat{F}_n) - T \left[ G(; \hat{\theta}_n) \right] \right) \right] = \text{Var}(S_1) = 2\theta_0^2$.

**Example 6.2.** Let $Y_i = (Y_{i1}, Y_{i2})^T$, $i = 1, \ldots, n$, be a random sample with distribution function $G(\cdot; \theta_0)$ and density function $g$. We obtain expressions for the limiting distribution in Theorem 6.4 with $T$ being Kendall’s $\tau$ or Spearman’s $\rho$, denoted below as $T_\tau = 4 \int G(y_1, y_2) dG(y_1, y_2) - 1$ and $T_{\rho_S} = 12 \int G_1(y_1)G_2(y_2) dG(y_1, y_2) - 3$, respectively, where $G_1$ and $G_2$ are the (known) marginal distribution functions of $G$. Let the true values of the measures be $\tau$ and $\rho_S$, the marginal densities be $g_m$ for $m = 1, 2$, $1_{ij}^* = 1 \{Y_{i1} > Y_{j1}, Y_{i2} > Y_{j2}\}$, $1_{ijk}^* = 1 \{Y_{i1} < Y_{k1}, Y_{j2} < Y_{k2}\}$, and assume there are no ties in the data. The empirical functionals can be written as

$$
T_\tau(\hat{F}_n) = \left( \frac{n}{2} \right)^{-1} \sum_{i<j} \left[ 2(1_{ij}^* + 1_{ji}^*) - 1 \right];
$$

$$
T_{\rho_S}(\hat{F}_n) = \left( \frac{n}{3} \right)^{-1} \sum_{i<j<k} \left[ 2(1_{ijk}^* + 1_{ikj}^* + 1_{jik}^* + 1_{kij}^* + 1_{ki}^* + 1_{kj}^*) - 3 \right] + O_p(n^{-1}),
$$

i.e., Kendall’s $\tau$ and Spearman’s $\rho$ have $r = 2$ and 3, respectively, and the $h$ functions are the terms inside the square brackets. The quantity $a$ in (6.11) can be obtained by considering the projection $V_n^*$

$$
V_n^* = \frac{r}{n} \sum_{i=1}^{n} \mathbb{E}[h(Y_1, Y_2, \ldots, Y_r) - \gamma | Y_1].
$$

The conditional expectations for Kendall’s $\tau$ are

$$
\mathbb{E}(1_{ij}^* | Y_i) = \mathbb{E}[1 \{Y_{i1} > Y_{j1}, Y_{i2} > Y_{j2}\} | Y_i] = G(Y_i; \theta_0); \quad \mathbb{E}(1_{ji}^* | Y_i) = \mathbb{E}[1 \{Y_{j1} > Y_{i1}, Y_{j2} > Y_{i2}\} | Y_i] = G(Y_i; \theta_0),
$$

and hence

$$
V_n^* = \frac{4}{n} \sum_{i=1}^{n} \left[ G(Y_i; \theta_0) + G(Y_i; \theta_0) \right] - 2(\tau + 1); \quad a = \text{Var} \left( 4 \left[ G(Y_i; \theta_0) + G(Y_i; \theta_0) \right] \right)
$$

$$
= 16 \int_{-\infty}^{\infty} \left[ G(y; \theta_0) + G(y; \theta_0) \right]^2 dG(y; \theta_0) - 4(\tau + 1)^2.
$$
For Spearman’s $\rho$, the conditional expectations are

$$
\mathbb{E}\left(1_{ijk}^{PS}|Y_i\right) = \mathbb{E}\left[1 \{ Y_{i1} < Y_{k1}, Y_{j2} < Y_{k2}\} | Y_i \right] = \mathbb{E}\left[\mathbb{E}(1 \{ Y_{i1} < Y_{k1}, Y_{j2} < Y_{k2}\} | Y_i, Y_j) | Y_i \right]
$$

$$
= \mathbb{E}\left[\mathcal{G}(Y_{i1}, Y_{j2}; \theta_0) | Y_i \right] = \int_{-\infty}^{\infty} \mathcal{G}(Y_{i1}, w; \theta_0)g_2(w) \, dw = \mathbb{E}\left(1_{ikj}^{PS}|Y_i\right);
$$

$$
\mathbb{E}\left(1_{ijk}^{PS}|Y_i\right) = \mathbb{E}\left[1 \{ Y_{j1} < Y_{i1}, Y_{k2} < Y_{i2}\} | Y_i \right]
$$

$$
= \mathbb{E}\left[\mathcal{G}(Y_{i1}, Y_{j2}; \theta_0) | Y_i \right] = \int_{-\infty}^{\infty} \mathcal{G}(w, Y_{i2}; \theta_0)g_1(w) \, dw = \mathbb{E}\left(1_{ikj}^{PS}|Y_i\right);
$$

$$
\mathbb{E}\left(1_{ikl}^{PS}|Y_i\right) = \mathbb{E}\left[1 \{ Y_{j1} < Y_{i1}, Y_{k2} < Y_{i2}\} | Y_i \right]
$$

$$
= \mathbb{E}\left[\mathcal{G}(Y_{i1}, Y_{j2}; \theta_0) | Y_i \right] = \int_{-\infty}^{\infty} \mathcal{G}(w, Y_{i2}; \theta_0)g_1(w) \, dw = \mathbb{E}\left(1_{ikj}^{PS}|Y_i\right).
$$

Let $B_1(Y_i; \theta_0) = \int_{-\infty}^{\infty} \mathcal{G}(Y_{i1}, w; \theta_0)g_2(w) \, dw$ and $B_2(Y_i; \theta_0) = \int_{-\infty}^{\infty} \mathcal{G}(w, Y_{i2}; \theta_0)g_1(w) \, dw$. Then

$$V_n^* = \frac{3}{n} \sum_{i=1}^{n} \left\{ 4[B_1(Y_i; \theta_0) + B_2(Y_i; \theta_0) + G_1(Y_{i1})G_2(Y_{i2})] - (\rho_S + 3) \right\}$$

$$\triangleq \frac{12}{n} \sum_{i=1}^{n} B(Y_i; \theta_0) - 3(\rho_S + 3)$$

$$a = \text{Var}[12B(Y_i; \theta_0)] = 144 \int_{-\infty}^{\infty} B^2(y; \theta_0) \, dG(y; \theta_0) - 9(\rho_S + 3)^2.$$

Meanwhile, for the partial derivative that appears in the quantity $b$ in (6.12), we obtain

$$
\frac{\partial T_{12}}{\partial \theta} \left[ G(\cdot; \theta_0) \right] = 4 \frac{\partial}{\partial \theta} \left[ \int G(y; \theta_0)g(y; \theta_0) \, dy \right]
$$

$$= 4 \int \left[ \frac{\partial G}{\partial \theta}(y; \theta_0) \cdot g(y; \theta_0) + G(y; \theta_0) \cdot \frac{\partial g}{\partial \theta}(y; \theta_0) \right] \, dy
$$

$$= 4 \int \left[ \frac{\partial G}{\partial \theta}(y; \theta_0) + G(y; \theta_0)\ell'(\theta_0; y) \right] \, dG(y; \theta_0), \quad \ell' = \partial g/\partial \theta;
$$

$$
\frac{\partial T_{PS}}{\partial \theta} \left[ G(\cdot; \theta_0) \right] = 12 \frac{\partial}{\partial \theta} \left[ \int G_1(y_1)G_2(y_2) \, dG(y; \theta_0) \right]
$$

$$= 12 \int G_1(y_1)G_2(y_2)\ell'(\theta_0; y) \, dG(y; \theta_0).$$

For copulas $U = (U_1, U_2) \sim C$, $\partial T_{PS} \left[ G(\cdot; \theta_0) \right] / \partial \theta$ reduces to $12 \int u_1u_2\ell'(\theta_0; u) \, dC(u; \theta_0)$ as the marginal distribution functions are $G_m(u) = u$ for $0 \leq u \leq 1$, $m = 1, 2$.

Next, we provide the vector version of Theorem 6.4, as well as the one applicable to estimators as solutions of estimating equations.
Theorem 6.5. Let $Y_1, \ldots, Y_n$ be a random sample from $G(\cdot; \theta_0)$. Assume A1 to A6. Let

$$T_l(\hat F_n) = \left( \begin{array}{c} \frac{1}{n} \sum_{\beta \in B} h_1(Y_{\beta_1}, \ldots, Y_{\beta_r}), \quad l = 1, \ldots, m \end{array} \right)$$

be U-statistics estimable of the same degree $r$ and estimated based on the same sample, where $h_l$ is the symmetric kernel for the $l$-th functional that is square-integrable with non-zero limiting variance. Then

$$\sqrt{n} \left( \begin{array}{c} T_1(\hat F_n) - T_1[G(\cdot; \hat \theta_n)] \\ \vdots \\ T_m(\hat F_n) - T_m[G(\cdot; \hat \theta_n)] \end{array} \right) \xrightarrow{d} N(0, \Sigma),$$

where $\Sigma$ has $(j, k)$ entry

$$r^2 \kappa_{1}^{(j, k)} - \frac{\partial T_l}{\partial \theta} [G(\cdot; \theta_0)] I^{-1} \frac{\partial T_k}{\partial \theta} [G(\cdot; \theta_0)]$$

with $\kappa_{1}^{(j, k)} = \text{Cov} [h_j(Y_1, \ldots, Y_r), h_k(Y_1, Y_{r+1}, \ldots, Y_{2r-1})]$ and $I$ defined in (6.4).

Proof. The projection argument is a direct generalization of that in Theorem 6.4 and Theorem 3.3.13 in Randles and Wolfe (1979). Once the U-statistics are projected to the space of sums of i.i.d. random variables, the multivariate central limit theorem can be applied as in Theorem 6.2.

Theorem 6.5'. Let $Y_1, \ldots, Y_n$ be a random sample from $G(\cdot; \theta_0)$. Assume A1, A6 and B1. Let

$$T_l(\hat F_n) = \left( \begin{array}{c} \frac{1}{n} \sum_{\beta \in B} h_l(Y_{\beta_1}, \ldots, Y_{\beta_r}), \quad l = 1, \ldots, m \end{array} \right)$$

be U-statistics estimable of the same degree $r$ and estimated based on the same sample, where $h_l$ is the symmetric kernel for the $l$th functional that is square-integrable with non-zero limiting variance. Then

$$\sqrt{n} \left( \begin{array}{c} T_1(\hat F_n) - T_1[G(\cdot; \hat \theta_n)] \\ \vdots \\ T_m(\hat F_n) - T_m[G(\cdot; \hat \theta_n)] \end{array} \right) \xrightarrow{d} N(0, \Sigma_e),$$

where $\Sigma_e$ is a $m \times m$ covariance matrix.
6.2.3 Separability of variance in the asymptotic normal distribution

We showed in the previous subsections that, when the parameter is estimated via maximum likelihood, there are cases where the scaled difference statistic (one element of \( \sqrt{n}D_n \)) has an asymptotic normal distribution with mean zero and a variance that can be written as a difference. In each case, this asymptotic variance is that of the empirical feature minus that of the model-based feature. Note that, given Assumptions A1 to A6, the following results concerning the maximum likelihood estimator hold, with \( I \) being the Fisher information matrix defined in (6.4):

\[
\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{d}{\to} N(0, I^{-1}(\theta_0));
\]

\[
\sqrt{n}\left[ G(t; \hat{\theta}_n) - G(t; \theta_0) \right] \overset{d}{\to} N\left(0, \frac{\partial G}{\partial \theta}^{-1}(t; \theta_0)I^{-1}\frac{\partial G}{\partial \theta}(t; \theta_0)\right), \text{ fixed } t;
\]

\[
\sqrt{n}\left(T \left[ G(\cdot; \hat{\theta}_n) \right] - T \left[ G(\cdot; \theta_0) \right] \right) \overset{d}{\to} N\left(0, \frac{\partial T}{\partial \theta}^{-1}\left[ G(\cdot; \theta_0) \right]I^{-1}\frac{\partial T}{\partial \theta} \left[ G(\cdot; \theta_0) \right] \right).
\]

Meanwhile, the asymptotic variances for the empirical distribution function and U-statistics are available in standard non-parametric theory. Results corresponding to the two features in the previous subsections are summarized in Table 6.1.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Empirical AVar</th>
<th>Model AVar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distribution function, fixed ( t )</td>
<td>( G(1 - G) )</td>
<td>( \frac{\partial G}{\partial \theta}^{-1}\frac{\partial G}{\partial \theta} )</td>
</tr>
<tr>
<td>U-statistics, ( T(F) = \binom{n}{r}^{-1} \sum_{\beta \in B} \frac{h(Y_{\beta_1}, \ldots, Y_{\beta_r})}{r^2 \alpha_1} )</td>
<td>( r^2 \alpha_1 ) (*see below)</td>
<td>( \frac{\partial T}{\partial \theta}^{-1}\frac{\partial T}{\partial \theta} )</td>
</tr>
</tbody>
</table>

*For U-statistics, \( \alpha_1 = \text{Cov}[h(Y_1, \ldots, Y_r), h(Y_1, Y_{r+1}, \ldots, Y_{2r-1})] \).


With the separability property, one can avoid the non-trivial calculation of the cross covariance between the empirical and model-based features. Note that the separability property may not hold in the more general case of estimating equations. Theorem 6.6 below proves that the variance of the scaled difference statistic approaches the variance of the limiting distribution, given certain conditions on the behaviour of the moments for the remainder term \( \zeta_n \) in (6.3). This has implications in the parametric bootstrap where variances are estimated from the sampling distribution. In the vector case as in the proof, this can be generalized to the covariance matrix of the scaled differences.

**Theorem 6.6.** In the decomposition (6.3), let \( \zeta_n = n^{-1/2}W_n \), where \( W_n = O_p(1) \). If \( \text{Cov}(W_n) = O(1) \) as \( n \to \infty \) and \( \text{E}(S_1) = 0 \), then

\[
\lim_{n \to \infty} \text{Cov}(X_n) - \text{Cov}(S_1) = 0.
\]
Proof. Since \( X_n = n^{-1/2} \sum_{i=1}^n S_i + n^{-1/2} W_n \), we have

\[
\text{Cov}(X_n) = \text{Cov}(S_1) + \frac{1}{n} \text{Cov}(W_n) + 2 \text{Cov} \left( \frac{1}{n} \sum_{i=1}^n S_i, W_n \right) .
\] (6.14)

The middle term \( n^{-1} \text{Cov}(W_n) \to 0 \) as \( n \to \infty \) by assumption, and \( E \left[ (n^{-1} \sum_{i=1}^n S_{ij})^2 \right] = n^{-1} \text{Var}(S_{ij}) \to 0 \) as \( n \to \infty \) for \( j = 1, \ldots, d \), where \( S_i^T = (S_{i1}, \ldots, S_{id}) \). Letting \( W_n^T = (W_{n1}, \ldots, W_{nd}) \), \( S_j = \sum_{i=1}^n S_{ij} / n \) and using the Cauchy-Schwarz inequality, the \((j,k)\) entry of the last covariance matrix of (6.14) is then given by

\[
\text{Cov} \left( S_j, W_{nk} \right) = E \left[ S_j W_{nk} \right] \leq \sqrt{E \left( S_j^2 \right) E(W_{nk}^2)} \to 0
\]
as \( n \to \infty \) because \( E(W_{nk}^2) \) is bounded, and hence \( \lim_{n \to \infty} [\text{Cov}(X_n) - \text{Cov}(S_1)] = 0 \). \( \square \)

Under the regularity conditions, the remainder term \( \zeta_n \) is of order \( O_p(n^{-1/2}) \) for maximum likelihood estimators and those coming from estimating equations with \( \sqrt{n} \)-consistency. However, the condition that \( \text{Cov}(W_n) \), or \( \text{Var}(W) \) for the single-element case, is bounded as \( n \to \infty \) is not so trivial to verify. For practical applications, we can check the magnitudes of \( n \text{Var} \left[ \hat{F}_n(t) - G(t; \hat{\theta}_n) \right] \) or \( n \text{Var} \left( T(\hat{F}_n) - T \left[ G(\cdot; \hat{\theta}_n) \right] \right) \) for several values of \( n \). If they do not seem to grow as \( n \) increases, that may be a sufficient indication that the variance of the scaled difference converges to the variance of the limiting distribution.

6.3 Asymptotics of the difference vector under model misspecification

The properties shown in Section 6.2 are only valid when the parametric model used for fitting coincides with the generating model. In this section, we mention the characteristics of the same statistic when the fitted model is misspecified. In particular, we show that the statistic is still asymptotically normal, but it has a mean that grows with the sample size. The Kullback-Leibler (KL) divergence (Kullback and Leibler (1951)), which can be considered as a distance measure between two distributions, is central to the subsequent discussion. The KL divergence of a density \( g \) from \( f \) is defined as

\[
KL(g||f) = \int f(t) \log \frac{f(t)}{g(t)} \, dt \geq 0.
\]

When the model is misspecified, there is no \( \theta_0 \) such that \( F(t) = G(t; \theta_0) \) for all \( t \in \mathbb{R}^d \). We will need the following assumption:
Assumption A7. For every $\theta \in \text{int}(\Theta)$, the expectations
\[
E_F \left[ \frac{\partial \ell(\theta; Y)}{\partial \theta} \frac{\partial \ell(\theta; Y)}{\partial \theta^\intercal} \right], \quad E_F \left[ \frac{\partial^2 \ell(\theta; Y)}{\partial \theta \partial \theta^\intercal} \right]
\]
exist and are not singular, where $E_F$ stresses that the expectation is taken with respect to the true distribution $F$.

We first state the classical result on maximum likelihood estimators when the model is misspecified, as Theorem 6.7 below.

Theorem 6.7. Let $Y_1, \ldots, Y_n$ be a random sample from $F$ with density $f$ and $G(\cdot; \theta)$ be the fitted parametric model with density $g$. Assume A2–A4 and A7. Then we have
\[
\hat{\theta} \overset{p}{\to} \tilde{\theta}; \quad (6.15)
\]
\[
\sqrt{n} \left( \hat{\theta} - \tilde{\theta} \right) \overset{d}{\to} N \left( 0, H^{-1}(\tilde{\theta})J(\tilde{\theta})H^{-1}(\tilde{\theta}) \right), \quad (6.16)
\]
where $\tilde{\theta}$ is the parameter value that minimizes the Kullback-Leibler divergence of $g(\cdot; \theta)$ from $f$, i.e.,
\[
\tilde{\theta} = \arg \min_{\theta} \int f(t) \log \left( \frac{f(t)}{g(t; \theta)} \right) dt,
\]
with $\tilde{\theta}$ in the interior of the parameter space, and
\[
J(\tilde{\theta}) = E_F \left[ \frac{\partial \ell(\tilde{\theta}; Y_1)}{\partial \theta} \frac{\partial \ell(\tilde{\theta}; Y_1)}{\partial \theta^\intercal} \right]; \quad H(\tilde{\theta}) = -E_F \left[ \frac{\partial^2 \ell(\tilde{\theta}; Y_1)}{\partial \theta \partial \theta^\intercal} \right].
\]

Proof. This is the standard result for maximum likelihood estimation of misspecified models, see, e.g., Huber (1967) and White (1982).
Theorem 6.8. Let $Y_1, \ldots, Y_n$ be a random sample from $F$ and $G(\cdot; \theta)$ be the fitted parametric model. Assume $A2$–$A4$ and $A7$, so that (6.15) and (6.16) hold. Let $\eta = F(t) - G(t; \hat{\theta})$. Then

$$\sqrt{n} \left[ \hat{F}_n(t) - G(t; \hat{\theta}_n) - \eta \right] \overset{d}{\to} N(0, \sigma^2),$$

where

$$\sigma^2 = \text{Var} \left[ 1 \{Y_i \leq t\} - \frac{\partial G}{\partial \theta}(t; \hat{\theta}) H^{-1} \ell'(\hat{\theta}; Y_i) \right]$$

$$\neq F(t) [1 - F(t)] - \frac{\partial G}{\partial \theta}(t; \hat{\theta}) J^{-1} \frac{\partial G}{\partial \theta}(t; \hat{\theta})$$

in general.

Proof. The decomposition of this difference is

$$\sqrt{n} \left[ \hat{F}_n(t) - G(t; \hat{\theta}_n) - \eta \right]$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ 1 \{Y_i \leq t\} - G(t; \hat{\theta}) - \eta - \frac{\partial G}{\partial \theta}(t; \hat{\theta}) H^{-1} \ell'(\hat{\theta}; Y_i) \right] - \sqrt{n} R_n$$

$$\equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} S_i + o_p(1).$$

where $R_n = o_p(n^{-1/2})$. Now $E(S_i) = F(t) - G(t; \hat{\theta}) - \eta = 0$, while

$$\text{Var}(S_i) = \text{Var} \left[ 1 \{Y_i \leq t\} + \frac{\partial G}{\partial \theta}(t; \hat{\theta}) H^{-1} \frac{\partial G}{\partial \theta}(t; \hat{\theta}) \right]$$

$$- 2 \text{Cov} \left[ 1 \{Y_i \leq t\}, \frac{\partial G}{\partial \theta}(t; \hat{\theta}) H^{-1} \ell'(\hat{\theta}; Y_i) \right]$$

as $\text{Cov} [\ell'(\hat{\theta}; Y_i)] = E \left( [\ell'(\hat{\theta}; Y_i)]; [\ell'(\hat{\theta}; Y_i)]^\top \right) = J$. The covariance term is

$$\text{Cov} \left[ 1 \{Y_i \leq t\}, \frac{\partial G}{\partial \theta}(t; \hat{\theta}) H^{-1} \ell'(\hat{\theta}; Y_i) \right] = \frac{\partial G}{\partial \theta}(t; \hat{\theta}) H^{-1} \int_{-\infty}^{t} \frac{\partial g(y; \hat{\theta})}{\partial \theta} \cdot \frac{f(y)}{g(y; \theta)} \, dy.$$

The integral

$$\int_{-\infty}^{t} \frac{\partial g(y; \hat{\theta})}{\partial \theta} \cdot \frac{f(y)}{g(y; \theta)} \, dy \neq J H^{-1} \frac{\partial G}{\partial \theta}(t; \hat{\theta})$$

in general. Note that, when the model is correctly specified, we have $J = H$ and

$$\int_{-\infty}^{t} \frac{\partial g(y; \hat{\theta})}{\partial \theta} \cdot \frac{f(y)}{g(y; \theta)} \, dy = \frac{\partial G}{\partial \theta}(t; \hat{\theta}),$$

retrieving (6.17). \qed
Theorem 6.9. Let $Y_1, \ldots, Y_n$ be a random sample from $F$ and $G(\cdot; \theta)$ be the fitted parametric model. Assume A2–A4 and A6–A7, and that $T_k(F_n)$ is a U-statistic outlined in Theorem 6.4, $k = 1, \ldots, m$. Let $\eta_k = T_k(F) - T_k(G(\cdot; \hat{\theta}))$. Then

$$\sqrt{n} \begin{pmatrix} T_1(F_n) - T_1(G(\cdot; \hat{\theta}_n)) - \eta_1 \\ \vdots \\ T_m(F_n) - T_m(G(\cdot; \hat{\theta}_n)) - \eta_m \end{pmatrix} \xrightarrow{d} N(0, \Sigma_{\text{mis}}),$$

for some covariance matrix $\Sigma_{\text{mis}}$.

Proof. The proof for one element, i.e., $\sqrt{n} \left( T_k(F_n) - T_k(G(\cdot; \hat{\theta}_n)) \right) - \eta_k$, follows from the projection argument and Theorem 6.8. This is readily extended to the vector case as in Theorem 6.2.

Note that Theorem 6.9 is also valid for $\hat{\theta}_n$ being the solution of estimating equations, with the assumptions changed to A6 and B1.

Theorems 6.8 and 6.9 contain two important observations: (a) that the scaled difference $\sqrt{n} \left( \hat{F}_n(t) - G(t; \hat{\theta}_n) \right)$ or $\sqrt{n} \left( T(\hat{F}_n) - T(G(\cdot; \hat{\theta}_n)) \right)$ has a mean that grows at rate $O(n^{1/2})$ if $\eta \neq 0$, and (b) that the asymptotic variance of the difference is no longer separable when the model is misspecified, even under maximum likelihood estimation.

Estimation via composite likelihood is one instance of modelling via estimating equations. In this case $\hat{\theta} = \theta_0$, the true parameter value, under some assumptions such as $\theta$ is identifiable from the margins being used. The asymptotic covariance matrix is still given by $H^{-1}(\theta_0)J(\theta_0)H^{-1}(\theta_0)$. The asymptotic variance of the scaled difference cannot be decomposed into the difference of the respective asymptotic variances in general.

### 6.4 Some comments on properties of the difference statistic

In this section, we provide some comments on the difference $T(\hat{F}_n) - T(G(\cdot; \hat{\theta}_n))$ that are relevant to its practical use as a diagnostic statistic.

- As Section 6.2 suggests, under correct model specification, maximum likelihood estimation and subject to regularity conditions, the variance of the limiting distribution of $\sqrt{n} \left( T(\hat{F}_n) - T(G(\cdot; \hat{\theta}_n)) \right)$ can be decomposed into the difference of the limiting variances of the empirical and model-based features. This means that, in the limit, the variability of the scaled difference cannot be larger than that of the empirical estimator, and goes in the opposite direction as that of the model-based estimator; a more precise maximum likelihood estimator leads to a smaller limiting variance of the model-based feature, and a larger variance of the difference.
• There are cases where the empirical and model-based features are identical and hence the difference is always zero. Such features are not suitable for use in assessing model adequacy. One example is when $T(F)$ is the mean of $F$ and the parametric model is one with the maximum likelihood estimator for the mean given by the sample mean, for example the exponential, geometric and Poisson distributions, and the binomial distribution with the success probability being the only parameter. For the multivariate Gaussian distribution, the maximum likelihood estimators for the mean and covariance parameters are their sample counterparts. However, if data are generated from a multivariate Gaussian distribution with zero mean and given marginal variances, and only the correlations are fitted, the resulting maximum likelihood estimators are no longer the sample correlations (see Section 7.5.3). Because of this, fitting a (saturated) multivariate Gaussian copula to data on the unit hypercube will generally result in correlation parameter estimates slightly different from the sample correlations.

• For $T$ being a central dependence measure for bivariate copulas (or bivariate monotone association), properties of the empirical estimators are available in Section 2.12 of Joe (2014). Here we only consider $T$ being Kendall’s $\tau$ or Spearman’s $\rho$, as they are invariant to marginal standardization and their empirical estimators have smaller asymptotic variance than Blomqvist’s $\beta$. For both empirical and model-based estimators, the asymptotic variance generally decreases as a function of the strength of dependence. This is however not a monotone relationship; for example, the asymptotic variance of the empirical Kendall’s $\tau$ for Gumbel or MTCJ copula is slightly higher at $\tau = 0.1$ than $\tau = 0$. As the strength of dependence approaches comonotonicity, both the empirical and model-based measures converge to 1 and the asymptotic variance approaches zero. The asymptotic variance for the difference is generally not a monotone function of the dependence strength; for many copulas it is increasing up to moderate dependence (true Kendall’s $\tau$ around 0.5), before eventually decreasing again. Finally, when $T$ is Kendall’s $\tau$, the empirical, model-based and difference estimators typically all have smaller variances than when $T$ is Spearman’s $\rho$ unless the strength of dependence is very strong. A demonstrative plot of the magnitudes for representative parametric copula families with different tail properties is given in Figure 6.1.

• Figure 6.1 also suggests that the asymptotic variances for various copulas can be quite different, even when the overall dependence strength is fixed. There is evidence that this behaviour, especially for the empirical estimator, can partly be attributed to the
Figure 6.1: Asymptotic variances of the empirical (solid line with circles), model-based (dotted line with circles) and difference estimators (solid, thick line without circles) for different parametric copula families with various strengths of dependence. The features are Kendall’s $\tau$ in the top row and Spearman’s $\rho$ in the bottom. The x-axis plots the dependence strength (in Kendall’s $\tau$) for each copula family, and the lines are smoothed to reduce the effect of sampling variability. The Frank copula is plotted to a Kendall’s $\tau$ value of 0.75 due to numerical instability beyond this value.

Tail properties of the copula; the variability gets larger as tail dependence increases or tail order decreases. A heuristic demonstration of the empirical asymptotic variance is given in Appendix B. Table 6.2 shows the asymptotic variance for some copulas with the same Kendall’s $\tau$ being set at 0.5. Frank and Gaussian copulas have weak dependence in the joint tails (with Frank weaker than Gaussian) and smaller empirical asymptotic variance. Hüsler-Reiss, Gumbel and reflected MTCJ (rMTCJ) copulas are asymmetric with strong dependence in one tail only. Among these, the rMTCJ copula has the strongest upper tail dependence with a given value of $\tau$. These parametric copula families generally result in larger variances. Finally, t copulas with small degrees of freedom have strong tail dependence in both tails and the largest variance among the parametric copula families considered.

---

13The tail order (see Hua and Joe (2011)) also describes the tail behaviour of a copula. It must be at least 1, with higher values indicating weaker dependence.
<table>
<thead>
<tr>
<th>Copula</th>
<th>Copula parameter</th>
<th>Tail order</th>
<th>Tail dependence index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Lower</td>
<td>Upper</td>
</tr>
<tr>
<td>Gaussian</td>
<td>0.71</td>
<td>1.17</td>
<td>1.17</td>
</tr>
<tr>
<td>t ($\nu = 1$)</td>
<td>0.71</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>t ($\nu = 3$)</td>
<td>0.71</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>t ($\nu = 10$)</td>
<td>0.71</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>t ($\nu = 30$)</td>
<td>0.71</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Hüsler-Reiss</td>
<td>1.81</td>
<td>1.42</td>
<td>1</td>
</tr>
<tr>
<td>Gumbel</td>
<td>2</td>
<td>1.41</td>
<td>1</td>
</tr>
<tr>
<td>rMTCJ</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Frank</td>
<td>5.74</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 6.2: Tail properties (above) and asymptotic variances (below) for different parametric copula families with true Kendall’s $\tau$ equal to 0.5

- For copula fitting with maximum likelihood using variables $U_1, \ldots, U_n$ i.i.d. from copula $C$, the scaled difference statistic $\sqrt{n}D_n$ is asymptotically normally distributed and the separability property of the asymptotic variance holds when the empirical feature is a U-statistic. However, in practice data do not usually lie in the unit hypercube and marginal modelling is relevant. Consider data $Y_i = (Y_{i1}, \ldots, Y_{id})^T$ i.i.d. from distribution

$$G(y; \zeta_1, \ldots, \zeta_d, \delta) = C(G_1(y_1; \zeta_1), \ldots, G_d(y_d; \zeta_d); \delta),$$  \hspace{1cm} (6.18)

where $\zeta_j$ is the marginal parameter vector for the $j$th margin with distribution function $G_j$, and $\delta$ is the copula parameter vector. Several different estimators of $\delta$ exist:

1. If all the $G_j$’s and $\zeta_j$’s are known, this reduces to the copula problem as $\hat{\delta}$, the maximum likelihood estimator of $\delta$ using the combined model, is the same as
the maximizer of the copula \( C(u; \delta) \) using the vector of transformed data \( U_i = (G_1(Y_{i1}; \zeta_1), \ldots, G_d(Y_{id}; \zeta_d))^\top \). Therefore this case can be reduced to the (pure) copula fitting approach mentioned above, for which the separability property holds.

2. If the \( G_j \)'s are known (or assumed correctly specified) but the \( \zeta_j \)'s are not, and the full likelihood consisting of densities of the form

\[ g(y_i; \zeta_1, \ldots, \zeta_d; \delta) = c(G_1(y_{i1}; \zeta_1), \ldots, G_d(y_{id}; \zeta_d); \delta) \prod_{j=1}^d g_j(y_{ij}; \zeta_j), \]

where \( c \) and \( g_j \) are respectively the copula and \( j \)th marginal density, is maximized with respect to the whole parameter vector \( \theta = (\zeta_1^\top, \ldots, \zeta_d^\top, \delta^\top)^\top \) jointly, the resulting estimators will still conform to the maximum likelihood framework (subject to regularity conditions) and thus the scaled difference statistic arising from this fit will be asymptotically normal with separable asymptotic variances. However, this method is only practical when the dimension \( d \) is small.

3. If the \( G_j \)'s are known (or assumed correctly specified) but the \( \zeta_j \)'s are not, and it is not practical to maximize the likelihood with respect to the whole parameter vector at the same time, we can first estimate the marginal parameters individually, obtaining maximum likelihood estimates \( \hat{\zeta}_1, \ldots, \hat{\zeta}_d \), and then estimate the copula parameter given these marginal parameters. This is the method of inference function for margins (IFM) (Joe and Xu (1996); Joe (2005)), mentioned briefly in Section 2.5.2. The IFM copula estimator \( \hat{\delta} \) is typically different from and is less efficient than the maximum likelihood estimator \( \hat{\delta} \). In this case, the scaled difference is asymptotically normal, as the estimator is \( \sqrt{n} \)-consistent arising from the solution of a set of unbiased estimating equations, thus satisfying Assumption B1. However, the separability property does not hold in this case. We observe that the variability of the difference statistic appears to be smaller than that obtained using a purely copula (or known margin) approach in many situations; see Section 7.8 for details.

4. When one does not want to assume a certain parametric form for the \( G_j \)'s, the copula parameter can be estimated via the marginal ranks method, in which the log-likelihood consisting of densities of the form \( c(s_{i1}, \ldots, s_{id}; \delta) \) is maximized, where \( s_{ij} = n^{-1} (\sum_{k=1}^n 1\{y_{kj} \leq y_{ij}\} - 0.5) \) is the adjusted rank of observation \( y_{ij} \). An outline of the proof on asymptotic normality of the resulting estimator \( \hat{\delta} \) is given in Genest et al. (1995), using results in multivariate rank statistics (see, e.g., Ruymgaart et al. (1972); Ruymgaart (1974); Rüschendorf (1976)).
essence, they argue that the quantity $\sqrt{n}(\hat{\delta} - \delta_0)$, where $\delta_0$ is the true value, can still be written as a sum of i.i.d. random variables plus negligible terms, so that asymptotic normality can be established; Genest et al. (1995) show that the necessary assumptions are satisfied by a large number of one-parameter bivariate copula families. Although not rigorously proved in this thesis, we believe that such decomposition into a sum of i.i.d. random variables will allow a proof similar to the one for $\sqrt{n}$-consistent estimators as solutions to estimating equations, so that asymptotic normality of the scaled difference statistic can be established (such convergence is assumed for some results in Chapter 7 to hold). In general, the separability property does not hold in this case.

The following illustrative simulation set-up provides an example of the behaviour of the asymptotic variances for the various copula parameter estimators described above. Here we restrict our attention to a bivariate random vector $Y = (Y_1, Y_2)^T$ with distribution function (6.18), where $C$ is the MTCJ copula $C(u_1, u_2; \delta) = (u_1^{-\delta} + u_2^{-\delta} - 1)^{-1/\delta}$ and $G_k$ is the two-parameter Pareto distribution $G_k(y_k; \alpha_k, \sigma_k) = 1 - (1 + y_k/\sigma_k)^{-\alpha_k}$ for $y_k > 0$, $k = 1, 2$. The parameters chosen are $(\alpha_k, \sigma_k) = (4, 3)$ for $k = 1$ and $(2, 1)$ for $k = 2$, and $\delta = 2$ so that the copula has Kendall’s $\tau$ equal to 0.5. Samples of size $n = 20,000$ are generated from $G$, from which the empirical Kendall’s $\tau$ and the model-based Kendall’s $\tau$ based on the estimators above (i.e., known marginal distributions, joint maximum likelihood, IFM and marginal ranks) are obtained. The procedure is repeated 2,000 times to obtain a sampling distribution, from which asymptotic variances are estimated. The results are displayed in Table 6.3. When the margins are known or joint maximum likelihood is performed, the estimated asymptotic variance of the (scaled) difference statistic is close to the difference between the empirical and model-based asymptotic variance estimates. This is no longer true for models fitted with IFM or marginal ranks. Also, the model-based asymptotic variance estimates using marginal ranks are much larger than the other methods. In this case the known margin and IFM approaches have similar model-based asymptotic variances, but as we noted above the variance using IFM is typically larger. The estimated model-based asymptotic variance for joint maximum likelihood estimation is the smallest, and results in the largest variance for the difference statistic. Finally, when margins are known, the estimated asymptotic variance for the difference is larger than those using IFM and marginal ranks. Although a theoretical proof appears difficult, this seems to be the general pattern we see for more complex parsimonious structures. The

\[^{14}\text{These estimates are compared against those arising from some other values of } n \text{ to ensure that Theorem 6.6 is applicable.}\]
behaviour for these structures and its implications on using the difference statistic as a diagnostic tool are discussed in Section 7.8.

<table>
<thead>
<tr>
<th>Method</th>
<th>Asymptotic variance</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Empirical</td>
<td>Model</td>
<td>$E - M$</td>
<td>Difference</td>
</tr>
<tr>
<td>Known margins</td>
<td>0.13</td>
<td>0.15</td>
<td>0.15</td>
<td>0.15</td>
</tr>
<tr>
<td>Joint MLE</td>
<td>0.12</td>
<td>0.16</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>IFM</td>
<td>0.13</td>
<td>0.15</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>Marginal ranks</td>
<td>0.25</td>
<td>0.03</td>
<td>0.10</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.3: Asymptotic variance estimates for Kendall’s $\tau$ using different estimators for the copula parameter, with the parametric model having Pareto margins and MTCJ copula with Kendall’s $\tau$ equal to 0.5. The column labelled “$E - M$” is the difference between the empirical and model-based asymptotic variances, and should be close to the asymptotic variance of the difference (last column) if the separability property holds.

- When the functional is the rank-based F-madogram empirical estimator of the extremal coefficient or tail-weighted dependence measures, a different treatment is required as they cannot be naturally expressed as U-statistics. In both cases, the empirical estimators are asymptotically normal using the theory of the empirical copula process, while the asymptotic normality of the model-based counterparts (for the model estimation methods in the preceding item) is a result of the delta method. However, it has not been proved that the scaled difference statistic is asymptotically normal, although extensive simulations suggest that this may be true. This is our conjecture in Chapter 7, upon which some of the results are based.

6.5 Decision criteria based on the adequacy-of-fit statistic

In the previous sections, we demonstrated the properties of the difference statistic $D_n$ when the model is correctly specified or misspecified. An important message is that $\sqrt{n}D_n$ is approximately normal for large $n$ under some mild conditions on the functionals $T_1, \ldots, T_m$; the limiting distribution has mean zero when the model is correct, and the mean of $\sqrt{n}D_n$ grows at the rate $O(n^{1/2})$ when the model is misspecified. Thus $D_n$ satisfies the intuitive requirement of a “distance” measure between the empirical and assumed distributions. A decision of model adequacy can be made based on the adequacy-of-fit statistic $Q_n = nD_n^\top D_n$; other possible formulations were outlined at the beginning of this chapter. A fitted model can be seen as too parsimonious if $Q_n$ exceeds the 100$(1 - \alpha)$% quantile of its limiting (or reference) distribution. We refer to this quantile as a critical value.
This critical value depends on the matrix $\Sigma$ which in turn depends on the feature being used and the assumed structure in the case of a multivariate distribution; details on the determination of an appropriate critical value will be given in Chapter 7. Based on the convergence results of a misspecified model in Section 6.3, we have the following insights:

- The distance (with respect to the chosen features) between the true distribution and the distribution closest to the truth within the family of assumed distributions, i.e., $T(F) - T\left[G(\cdot; \tilde{\theta})\right]$, affects the ability of the statistic to detect model departure from the data. It is therefore important to choose the features carefully; they should reflect the purpose of model fitting for a particular problem. For instance, if the objective is to find a model that can adequately represent the overall strength of dependence between two variables, then Kendall’s $\tau$ and Spearman’s $\rho$ are choices to consider. If tail properties are of interest, features more specific to this purpose, such as tail-weighted dependence measures (Krupskii and Joe (2015)), may be better options.

- Increasing the sample size improves the detection of departure from the true model. This is essentially the same statement that relates the sample size to the power of a hypothesis test. The emphasis here is that statistical models are merely a tool to assist the researcher in explaining data patterns (“all models are wrong, but some are useful” (Box (1979))). In practical applications, the 95% level critical value will likely be exceeded with a sufficiently large sample size. In this case it is constructive to examine the actual magnitudes of departure, i.e., those of $D_n$, and ask whether such differences are scientifically or practically significant.

- The magnitude of the asymptotic covariance matrix under model misspecification, i.e., $\Sigma_{mis}$ in Theorem 6.9, may affect how likely one will obtain a small value for the adequacy-of-fit statistic when the model is misspecified. This and the asymptotic covariance matrix under correct model specification (i.e., $\Sigma$) are affected by the precision of the empirical estimator. Using an efficient estimator for the features of interest will therefore improve the performance of the statistic.

We consider the adequacy-of-fit statistic a guide as to whether further model improvement is necessary. This comes into play because model selection criteria like AIC and BIC do not provide an indication as to whether a particular model is representative enough. Nevertheless, the adequacy-of-fit statistic should not be used blindly to evaluate the relative strengths of competing models without reference to model selection criteria. An overparametrized model will likely lead to a smaller value than one with fewer parameters due to a better fit, but this does not imply that the latter is inadequate.
Chapter 7

Adequacy-of-fit for multivariate copulas with parsimonious dependence

The theoretical properties of the difference between empirical and model-based features were discussed in Chapter 6. We suggested the possible use of a quadratic form statistic based on the vector of differences for model diagnostics, with the intuition that model inadequacy will lead to a larger value of the statistic. In this chapter, we extend this idea to the adequacy-of-fit of multivariate copulas using the differences for pairwise margins. In particular, the functional or features to be considered are measures of dependence for each of the \( \binom{d}{2} \) bivariate margins of a \( d \)-dimensional copula, including Kendall’s \( \tau \), Spearman’s \( \rho \), tail-weighted dependence measures, and extremal coefficient for extreme value copulas. Sections 7.1 and 7.2 serve as an introduction to this chapter; we motivate the study of such a statistic using bivariate margins in Section 7.1 and make connections to previous work in this area. Section 7.2 provides the big picture viewpoints of the challenges we encounter and suggest appropriate strategies for each situation, and a brief overview of the contents in subsequent sections.

7.1 Motivation and background

Let \( F \) be the true distribution for \( d \)-dimensional i.i.d. observations \( Y_i = (Y_{i1}, \ldots, Y_{id})^T \), \( i = 1, \ldots, n \), and \( G(\cdot; \theta) \) be a model with parameter \( \theta \in \Theta \) to which the data are fitted, resulting in the estimator \( \hat{\theta}_n \). The empirical distribution function is denoted as \( \hat{F}_n \). Suppose \( G \) is correctly specified in the sense that \( F(t) = G(t; \theta_0) \) for all \( t \in \mathbb{R}^d \), for some \( \theta_0 \in \Theta \). We
suggested in the previous chapter that the vector of difference statistic $\hat{F}_n(t) - G(t; \hat{\theta}_n)$ or $T(\hat{F}_n) - T\left[G(\cdot; \hat{\theta}_n)\right]$ for functional $T$ can be used as a discrepancy measure of the adequacy of model fit. However, observations are sparse when $d$ is large and, as a result, $\hat{F}_n$ becomes an increasingly unreliable estimate as more variables are considered in the model. This affects the use of the difference statistic because the limiting (or reference) distribution may not be representative of the finite-sample behaviour.

This problem has been previously studied in the discrete context. The Pearson’s $\chi^2$ statistic is widely used to inspect the validity of a model for item response data, typically presented in the form of cell counts for each category of each item. The total number of cells grows exponentially with the number of items and can easily exceed the sample size. This leads to sparse contingency tables, i.e., with many cells having zero observed counts and small expected probabilities, even for a moderate number of items. In this case, the asymptotic distribution of the $\chi^2$ statistic can be substantially different from its empirical distribution, meaning that tests and decisions based on the asymptotic distribution can be invalid. The issue of sparsity is alleviated by using limited-information methods through low-order marginal tables; this strategy is advocated in, e.g., Reiser and VandenBerg (1994); Reiser (1996); Reiser and Lin (1999); Bartholomew and Leung (2002); Maydeu-Olivares and Joe (2005, 2006). Let $p$ be the vector of observed cell proportions and $p_i$ be the vector of observed proportions for each cell of all order $i$ marginal tables (one category for each margin is left out as the corresponding proportion can be determined from the other entries of the vector), with corresponding model-based probabilities $\pi(\theta)$ and $\pi_i(\theta)$, respectively. Let $p_r = (p_1^\top, \ldots, p_r^\top)^\top$ be the vector of such proportions for all marginal tables of order $r$ or fewer, and similarly $\pi_r(\theta)$ be the model-based counterpart. Under certain regularity conditions, Maydeu-Olivares and Joe (2006) show that

$$\sqrt{n}\left[p_r - \pi_r(\hat{\theta}_n)\right] \overset{d}{\to} N(0, \Sigma_r) \quad (7.1)$$

for $\hat{\theta}_n$ being a $\sqrt{n}$-consistent and asymptotically normal estimator, including the maximum likelihood estimator, and the covariance matrix $\Sigma_r$ of the limiting or reference distribution has known expressions. The authors define the limited information statistic of order $r$ as $Q_n = n\left[p_r - \pi_r(\hat{\theta}_n)\right]^\top V_r(\hat{\theta}_n) \left[p_r - \pi_r(\hat{\theta}_n)\right]$, where $V_r \triangleq V_r(\theta_0) = V_r \Sigma_r V_r$, so that $\Sigma_r$ is a generalized inverse of $V_r$. This quadratic form statistic converges to a chi-squared distribution with appropriate degrees of freedom as $n \to \infty$. For limited sample sizes, it is demonstrated that the empirical distribution of this limited information statistic matches its asymptotic (or reference) distribution better than the Pearson’s $\chi^2$ statistic does. In Joe and Maydeu-Olivares (2010), this theory is extended to the quadratic forms in arbitrary linear combinations of the cell residuals $p_r - \pi_r(\hat{\theta}_n)$. 139
In this chapter, we extend the idea of examining adequacy of model fit using lower-order marginals to general copulas, as a way to bypass the sparsity issue in high dimensions. This is achieved through the theory developed in Chapter 6 regarding the asymptotic behaviour of differences that are analogous to the cell residuals in (7.1). We will mainly focus on bivariate marginal differences; these methods are applicable to trivariate and higher-order margins with suitably defined functionals (features) and sufficient computational power.

Previous work on assessing the quality of fit of parsimonious models involves comparing the discrepancy between a $d \times d$ observed correlation matrix $R_{\text{obs}}$ and the estimated, model-based correlation matrix $R_{\text{mod}}$, and has its roots in structural equation modelling. The statistic that is relevant to our current work is the standardized root mean squared residual (SRMSR) (see, e.g., Hu and Bentler (1998)), defined by

$$SRMSR = \left[ \sum_{1 \leq j < k \leq d} \frac{(r_{jk} - \hat{\rho}_{jk})^2}{d(d-1)/2} \right]^{1/2},$$

where $r_{jk}$ and $\hat{\rho}_{jk}$ are the entries of $R_{\text{obs}}$ and $R_{\text{mod}}$, respectively. The difference $r_{jk} - \hat{\rho}_{jk}$ is known as the residual for the bivariate margin $(j, k)$. The SRMSR has the same scale as the correlations being considered and thus offers an intuitive interpretation as an average deviation between the observed and fitted correlations. Traditional research in this area focuses on establishing cutoff values for various fit indices; these cutoffs typically correspond to quantiles of a reference distribution and are used as guidelines to determine whether the chosen parametric model is too parsimonious or underparametrized. A summary of such cutoffs is given by Hooper et al. (2008). These studies are mostly simulation-based, see, e.g., Hu and Bentler (1998) and Hu and Bentler (1999). In this regard, our work can be considered as an attempt to formalize the study of SRMSR to general multivariate parsimonious models.

In the following, we illustrate the questions we try to answer in this chapter with two examples.

**Example 7.1.** In Section 3.7 we fitted structured extreme value copulas to two data sets. For the US stock returns example with sample size $n = 119$, the matrices of absolute pairwise differences between the empirical and model-based extremal coefficient based on the best two models using BIC are listed in Table 7.1. Also listed in the table are the maximum absolute difference and the average root-mean-square difference for each model. In this chapter, we investigate whether the differences are large enough to suggest that the model(s) may not be adequate for the data.

---

15The formulation in Hu and Bentler (1998) uses the covariance matrix and includes the diagonal elements, but the idea is similar.
Table 7.1: Matrix of pairwise differences between the empirical and model-based extremal coefficient for the Burr 1-factor EV copula (top) and 1-factor t-EV copula with $\nu = 3$ (bottom) fitted to the US stock returns example

<table>
<thead>
<tr>
<th>Model</th>
<th>Matrix of absolute pairwise differences: Extremal coefficient</th>
<th>Maximum absolute difference:</th>
<th>Average RMS difference:</th>
</tr>
</thead>
<tbody>
<tr>
<td>EV1f (Burr)</td>
<td>$\begin{pmatrix} 0 &amp; .053 &amp; .028 &amp; .078 &amp; .003 &amp; .051 &amp; .055 \ 0 &amp; .034 &amp; .054 &amp; .040 &amp; .004 &amp; .057 \ 0 &amp; .058 &amp; .047 &amp; .012 &amp; .110 \ 0 &amp; .111 &amp; .085 &amp; .123 \ 0 &amp; .055 &amp; .071 \ 0 &amp; .012 \ 0 \end{pmatrix}$</td>
<td>0.123</td>
<td>0.064</td>
</tr>
<tr>
<td>tEV1f ($\nu = 3$)</td>
<td>$\begin{pmatrix} 0 &amp; .049 &amp; .002 &amp; .032 &amp; .012 &amp; .051 &amp; .037 \ 0 &amp; .019 &amp; .018 &amp; .034 &amp; .013 &amp; .048 \ 0 &amp; .024 &amp; .031 &amp; .002 &amp; .084 \ 0 &amp; .067 &amp; .055 &amp; .089 \ 0 &amp; .057 &amp; .057 \ 0 &amp; .009 \ 0 \end{pmatrix}$</td>
<td>0.089</td>
<td>0.045</td>
</tr>
</tbody>
</table>

One may consider the parametric bootstrap on the fitted model to gain insight on the behaviour of the differences assuming correct model specification. We discuss the challenges underlying this approach, for instance the unavailability of accurate simulation methods for some models.

**Example 7.2.** Brechmann and Joe (2015) consider approaches to select the number of truncation levels of a vine copula based on fit indices. The proposed methods are applied to the GARCH-filtered return time series of 19 assets and indices of a Norwegian market portfolio, as well as those of 15 largest German companies represented in the DAX index. In each case, the authors identify better 4- to 6-truncated vine copulas in terms of BIC values than those previously reported. For each fitted model, the maximum and average absolute deviations between empirical and model-based Kendall’s $\tau$ are computed. The model-based Kendall’s $\tau$ values are estimated by simulating 10,000 samples from the fitted models. The authors compare these deviations with the typical variances based on the empirical Kendall’s $\tau$ of a sample from a bivariate copula with moderate strength of dependence, and suggest that the deviations are within sampling variability.

For this example, our results based on maximum likelihood estimation in Chapter 6
suggest that the variability of the empirical Kendall’s $\tau$ is higher than that of the difference between the empirical and model-based Kendall’s $\tau$. This difference can sometimes be substantial (Figure 6.1). Using the variance of the empirical Kendall’s $\tau$ as the basis of comparison will thus likely overestimate the range within which deviations can be attributed to sampling variability. It is then natural to consider the parametric bootstrap to get an idea of what range of deviations for the difference is reasonable. However, evaluating the model-based Kendall’s $\tau$ requires obtaining the bivariate marginal copulas; for general vine copulas they are usually intractable. It is also impractical to estimate the model-based Kendall’s $\tau$ by simulating from the fitted model for the bootstrap sample, since this has to be done for each of the many samples. In this chapter, we explore ways to obtain critical values without going through this computationally intensive procedure.

An additional challenge in developing cutoff or critical values for diagnostic statistics is that these values may be sensitive to how the model is fitted. For the sake of computational efficiency, fitting of complex copula models can be performed using the inference functions for margins (IFM) approach (Joe and Xu (1996); Joe (2005)) or marginal ranks approach (Genest et al. (1995)), where dependence parameters are fitted using pseudo observations obtained based on the fitted marginal distributions or ranks. Meanwhile, composite likelihood can be used when the full density is hard to obtain, such as in the case of multivariate extreme value copulas. For vine copulas, yet another possibility is sequential fitting for each tree in an attempt to reduce computational complexity. In each case, the resulting estimates are generally different from those based on maximum likelihood estimation of the copula, and this may result in somewhat different behaviour of these diagnostic statistics. Our goal is to obtain a conservative (upper) bound for critical values that are applicable to most practical situations. Note that a conservative estimate is sufficient if such bound is within the practical significance for the problem.

The adequacy-of-fit statistic to be introduced in this chapter is meant to be a diagnostic tool (see Krupskii and Joe (2015)), and therefore much of the comments in Section 6.5 applies to here as well. Unlike in the discrete case of Maydeu-Olivares and Joe (2005, 2006) and Joe and Maydeu-Olivares (2010), where a model is theorized using subject knowledge in psychology, the choice of parametric copula family and structure is not guided by scientific theory in general; they are instead chosen to best represent the structure of the data and permit further inference based on the model. As a result, the purpose we suggest is different from goodness-of-fit procedures such as those in Genest et al. (2009), Huang and Prokhorov (2014) and Schepsmeier (2014), where hypothesis tests are conducted in order to approve or disprove the validity of a model. The pathway of statistical modelling we adopt can be summarized as follows:
1. Potential models are selected based on simple summaries, such as pairwise correlations of the normal scores, that reveal data pattern.

2. The models are fitted and compared using model selection criteria such as AIC or BIC, or procedures like cross-validation and stepwise methods.

3. Adequacy-of-fit diagnostic checks are applied to the final chosen models. Those deemed adequate (i.e., with diagnostic statistic smaller than a cutoff value based on a reference distribution) are to be used for further inference and as basis for decision making. If the checks suggest that the models are inadequate, one should revise the models and the whole procedure should be repeated.

An additional advantage of combining information from lower-order marginal distributions is that the resulting statistic provides directions on model improvement. For example, if the discrepancy is found to be substantially larger for features involving a particular variable, one should analyze the source of misfit accordingly. Systematic misfit such as trends in the differences should be removed prior to using the statistic as a means to assess model adequacy (see Maydeu-Olivares and Joe (2014)). As we mentioned at the beginning of Chapter 6, procedures that retain only the global summary of the distance between distributions, such as Kolmogorov-Smirnov and Cramér-von Mises-type statistics and their copula extensions, do not typically offer guidance for model improvement. Likelihood ratio tests in general also fall into this latter category (Section 7.8 of Cox and Wermuth (1996)).

7.2 Diagnostic checks based on the adequacy-of-fit statistic

In this section, we give an overview of the big picture viewpoints regarding the use of the adequacy-of-fit statistic arising as a quadratic form of the vector of differences, and suggest appropriate strategies for each situation. The connection of such statistic to the SRMSR is then made.

7.2.1 Issues and general strategies

In addition to the notations defined at the beginning of this chapter, we denote $F_{jk}$ as the bivariate marginal distribution for the $(j,k)$ margin with empirical estimate $\hat{F}_{jk}$. The subscript $n$ is dropped to prevent cluttering of indices. The adequacy-of-fit statistic $Q_n$ is defined as

$$Q_n = n \left( \frac{d}{2} \right)^{-1} \sum_{1 \leq j < k \leq d} w_{jk} \left( T(\hat{F}_{jk}) - T\left[ G_{jk}(\cdot; \hat{\theta}_n) \right] \right)^2,$$  

(7.3)
where \( w_{jk} \) is the weight associated with the difference for the margin \((j,k)\). The statistic \( Q_n/n \) can therefore be interpreted as the weighted average pairwise squared error with respect to \( T \). The following box makes concrete the classes of empirical features \( T(\hat{P}_{jk}) \) and model estimation methods for \( \hat{\theta}_n \) we consider in this chapter:

**Empirical features**

1. U-statistics, including Kendall’s \( \tau \) and Spearman’s \( \rho \);
2. Rank-based F-madogram estimator of the extremal coefficient, and;
3. Rank-based tail-weighted dependence measures.

**Model estimation methods**

1. Maximum likelihood;
2. Estimating equations (including the methods of inference functions for margins, composite likelihood and sequential vines), and;
3. Marginal ranks or semiparametric estimation.

The relevant combinations from the categories above are: (a) U-statistics with any model estimation method; (b) extremal coefficient for extreme value copulas with estimating equations or marginal ranks method, and; (c) tail-weighted dependence measures with any model estimation method. To obtain the moment properties of the limiting distribution of \( Q_n \), we rely on the result that the vector of differences is asymptotically normal. It is known that the vector of empirical features (those listed above) is asymptotically normal, and similarly for the corresponding vector of model-based features. However, that the vector of differences is asymptotically normal is only proved for the combination of empirical U-statistics with modelling via estimating equations (including maximum likelihood) in Chapter 6. For the other cases, asymptotic normality of the vector of differences is assumed. A rigorous proof involves non-trivial details and remains an open problem; a representation that allows the application of Lemma 6.1 is needed.

For U-statistics with the estimating equations method, the result of asymptotic normality applied to the bivariate marginal features is stated in Theorem 7.1.

**Theorem 7.1.** Let \( Y_1, \ldots, Y_n \) be a random sample from \( F \) and \( G \) be the fitted parametric model. Suppose \( G \) is correctly specified, i.e., there exists \( \theta_0 \) such that \( F(t) = G(t; \theta_0) \) for every \( t \in \mathbb{R}^d \). Let \( \hat{\theta}_n \) be the solution of a set of estimating equations, satisfying Assump-
tion B1 in Section 6.2, such that \( \hat{\theta}_n \) is \( \sqrt{n} \)-consistent and asymptotically normal. If the functional \( T \) is a U-statistic that satisfies the assumptions in Theorem 6.4, then

\[
\sqrt{n} D_n \triangleq \sqrt{n} \begin{bmatrix}
T(\hat{F}_{12}) - T\left[G_{12}(\cdot; \hat{\theta}_n)\right] \\
T(\hat{F}_{13}) - T\left[G_{13}(\cdot; \hat{\theta}_n)\right] \\
\vdots \\
T(\hat{F}_{d-1,d}) - T\left[G_{d-1,d}(\cdot; \hat{\theta}_n)\right]
\end{bmatrix} \xrightarrow{d} N(0, \Sigma), \tag{7.4}
\]

where \( \Sigma \) is a square matrix of dimension \( \binom{d}{2} \). Let \( 1 \leq i_{pq} \leq \binom{d}{2} \) be the index of pairwise margin \( (p, q) \), \( 1 \leq p < q \leq d \). If further \( \hat{\theta}_n \) is the maximum likelihood estimator, then the \( (i_{jk}, i_{lm}) \) entry of \( \Sigma \), where \( j < k \), \( l < m \), is:

\[
r^2 \text{Cov} \left[ h_{jk}(Y_1, \ldots, Y_r), h_{lm}(Y_1, Y_{r+1}, \ldots, Y_{2r-1}) \right] - \frac{\partial T}{\partial \theta_i} \left[ G_{jk}(\cdot; \theta_0) \right] I^{-1} \frac{\partial T}{\partial \theta_i} \left[ G_{lm}(\cdot; \theta_0) \right],
\]

using the notation in Theorem 6.5.

However, if \( G \) is misspecified, then we have

\[
\sqrt{n}(D_n - \Delta) \xrightarrow{d} N(0, \Sigma^*) \tag{7.5}
\]

as \( n \to \infty \), where \( \Delta \) is a non-zero vector with at least one \( O(1) \) element provided that at least one of \( T(\hat{F}_{jk}) - T\left[G_{jk}(\cdot; \hat{\theta})\right] = O(1) \), with \( G(\cdot; \hat{\theta}) \) minimizing the Kullback-Leibler divergence of \( g \) from \( f \).

**Proof.** This theorem is an application of the relevant theorems (Theorems 6.5, 6.5′ and 6.9), where each of the \( m = \binom{d}{2} \) functionals here uses only partial information of \( F \) (or \( G \)), namely a bivariate marginal distribution.

With the distributional properties (7.4) and (7.5) for the combination of empirical U-statistics and modelling via estimating equations, or the assumed analogous results for the other combinations, the moments of the limiting distribution of \( Q_n \) are stated in Corollary 7.2.

**Corollary 7.2.** Let \( m = \binom{d}{2} \) be the total number of bivariate pairs and \( \sigma_{pq} \) be the elements of \( \Sigma \) in (7.4), \( 1 \leq p, q \leq m \). Let \( \Gamma = \text{diag}(w_{12}, \ldots, w_{d-1,d}) \) be the diagonal matrix of weights with entries indexed by \( \gamma_{pq} \), so that the \( Q_n \) defined in (7.3) is equal to \( nD_n^\top \Gamma D_n / m \). If (7.4) or (7.5) applies, then we have the following:
1. When $G$ is correctly specified, $Q_n \overset{d}{\to} Q$ with

$$
\mathbb{E}(Q) = \frac{\text{tr}(\Gamma \Sigma)}{m} = \frac{1}{m} \sum_{p=1}^{m} \gamma_{pp} \sigma_{pp};
$$

(7.6)

$$\text{Var}(Q) = \frac{2\text{tr}(\Gamma \Sigma \Gamma \Sigma)}{m^2} = \frac{2}{m^2} \sum_{p=1}^{m} \sum_{q=1}^{m} \gamma_{pp} \gamma_{qq} \sigma_{pq}^2.
$$

(7.7)

2. When $G$ is misspecified, $n (D_n - \Delta)^\top \Gamma (D_n - \Delta) / m \overset{d}{\to} Q^*$ with

$$
\mathbb{E}(Q^*) = \frac{\text{tr}(\Gamma \Sigma^*)}{m};
$$

$$\text{Var}(Q^*) = \frac{2\text{tr}(\Gamma \Sigma^* \Gamma \Sigma^*)}{m^2}.
$$

The expressions in Corollary 7.2 are derived directly from the moments of quadratic forms of normal random variables, see, e.g., Mathai and Provost (1992). When the model is correctly specified, the reference distribution $Q$ has a constant mean; otherwise the mean of an appropriate approximating distribution grows at $O(n)$. We thus seek an upper quantile (“critical value” hereafter) of $Q$ under the assumed model, beyond which model improvement is recommended. To approximate the quantile of $Q$, we match the first two moments (7.6) and (7.7) to that of a gamma distribution (such as in Bartholomew and Leung (2002) where approximation via a chi-squared distribution is employed; see also Maydeu-Olivares and Joe (2008)), and then obtain the quantile of this approximating distribution.

In practice, however, the computation of $Q_n$ or $\Sigma$ can be challenging; the following issues highlight the difficulties encountered:

**Issues of assessing model adequacy-of-fit using the quadratic form statistic**

1. For the dependence measures we consider, the model-based feature is only easy to obtain when the bivariate marginal distributions are numerically tractable. This happens when the parametric model is closed under margins.

2. As seen above, the asymptotic distribution of the quadratic form statistic (or the standardized version based on the SRMSR, to be described in Section 7.2.3) depends on the covariance matrix $\Sigma$ associated with the residual vector.

3. It is generally not easy to compute $\Sigma$; one particular case where this may be evaluated is when the empirical feature is a U-statistic and the model is estimated using maximum likelihood.
4. When $\Sigma$ cannot be easily computed or reliably estimated, one approach is to use a surrogate model to approximate or obtain bounds on the critical value.

5. We examine the performance of using surrogate models in low dimensions, where results can be compared to those of the target model, via maximum likelihood estimation of the copula.

In light of these challenges, the strategies we propose are based on the practicality for different combinations of empirical features and model estimation methods, and are summarized as follows. Some methods involve the use of specific functions or a reference table (i.e., Table 7.6); these are mentioned as software products below.

**User-oriented pathways to obtain critical values; software products**

1. Parametric bootstrap can be used whenever feasible (theoretically and computationally). This involves the following steps:
   a) Fit parametric copula model (using maximum likelihood, inference functions for margins or marginal ranks method) and obtain parameter estimates;
   b) Repeatedly simulate (with large sample sizes) from the fitted model, obtain empirical and model-based features and hence a sampling distribution of the quadratic form statistic;
   c) Obtain critical value as a high quantile of this sampling distribution and convert it to the SRMSR version for easier interpretation.

When parametric bootstrap is infeasible (model-based feature is difficult to compute or model simulation/fitting is computationally expensive), we consider the following pathways:

2. Kendall's $\tau$ and Spearman's $\rho$ for non-extreme-value copulas:
   a) Fit parametric copula model as in 1(a) above.
   b) For maximum likelihood estimation, it may be possible to evaluate $\Sigma$, the covariance matrix of the reference distribution, by separating the empirical and model-based components (Section 7.4). After getting $\Sigma$, use gamma approximation with the first two moments (based on Corollary 7.2) to obtain
the critical value as a high quantile of this distribution and convert it to the SRMSR version.

c) Alternatively, use a parsimonious Gaussian copula model with matching Kendall’s $\tau$ value of the linking copulas (for factor and truncated vine structures) as surrogate (Section 7.5). Obtain an estimate of $\Sigma$ for this surrogate model, or otherwise obtain a crude estimate of the critical value from Table 7.6; in the latter case, step (d) may be skipped.

d) Based on $\Sigma$, use gamma approximation with the first two moments to obtain critical value as a high quantile of this distribution and convert it to the SRMSR version.

e) With the results based on a Gaussian surrogate, apply a multiplying factor based on the tail heaviness of the parametric copula families of the target model if necessary. In Section 7.5, we assess the magnitude of multiplying factors by computing the critical values of the actual copula model (using the $\Sigma$ matrix under maximum likelihood estimation) and comparing these to the Gaussian critical values.

3. Kendall’s $\tau$, Spearman’s $\rho$ and extremal coefficient for extreme value copulas:

   a) Fit parametric copula model using composite likelihood and obtain parameter estimates.
   
   b) If the fitted model is easy to simulate from, proceed to (c). Otherwise (e.g., Hüsler-Reiss copula and t-EV copula with large degrees of freedom), use a t-EV copula with small degrees of freedom (and matching pairwise Kendall’s $\tau$ or extremal coefficient) as a surrogate model (Section 7.6).
   
   c) Perform parametric bootstrap on the target or surrogate model from (b) and obtain critical values as in 1(b) and 1(c) above. This is possible as the models are closed under margins.
   
   d) Based on a limited simulation study, a multiplying factor of 1.2 should be applied if the t-EV copula is used as a surrogate for the Hüsler-Reiss copula (and t-EV copula with large degrees of freedom). If the parametric bootstrap is performed on the target model, no multiplying factor is necessary.

4. Tail-weighted dependence measure for specific tails of non-extreme-value copulas:

   a) Fit parametric copula model as in 1(a) above.
b) Repeatedly simulate from the fitted model and obtain the vector of empirical measures for each sample. From these, estimate the limiting covariance matrix for the empirical measure $\Sigma_{emp}$. In Section 7.7, we suggest that $\Sigma_{emp}$ is only slightly “larger” than $\Sigma$ for the residual vector, and that great simplifications result by considering $\Sigma_{emp}$.

c) Use gamma approximation as in 2(d) above, but with $\Sigma_{emp}$ instead.

Software products as functions or reference tables for interpolation in R:

I. Function for the evaluation of $\Sigma$ for U-statistics and under maximum likelihood as in 2(b) above, suitable for models whose simulation and fitting are scalable to large sample sizes.

II. Tabulated Gaussian critical values for different structures and average dependence strengths as in the last part of 2(c) above.

III. Fast algorithms for restricted maximum likelihood estimation for the Gaussian surrogate copulas (see Section 7.5.3).

IV. Function for the conversion of $\Sigma$ to critical values using gamma approximation.

In several cases where parametric bootstrap\footnote{Note the deviation here from the typical parametric bootstrap that simulates samples of the same size as the observed data. Here, we use these (larger) simulated samples to obtain a better estimate of the sampling distribution of the quadratic form statistic, or the covariance matrix of the reference distribution.} is possible, we check that the critical values based on the gamma approximation are similar to those from the sampling distribution. Meanwhile, different estimation methods may have an effect on the critical values. Although some of the work in this chapter (e.g., analysis of the surrogate Gaussian model in Section 7.5) focuses on copula estimation assuming known margins, we argue in Section 7.8 that the resulting critical values may be considered as conservative (upper) bounds for other estimation methods for the margins, such as the IFM and marginal ranks methods; the diagonal elements of $\Sigma$ (i.e., the variance terms) appear to be smaller in these cases when the feature is Kendall’s $\tau$ or Spearman’s $\rho$.

A very general argument that applies irrespective of the specific dependence structure is that the critical value tends to drop as bivariate dependence increases, especially when the variables are strongly correlated. This is because the elements of $\Sigma$, being the variances and covariances of combinations of $T(\hat{F}_{jk}) - T(\hat{G}_{jk}(\cdot; \hat{\theta}_n))$, decrease in magnitude when the bivariate distribution becomes more concordant. This in turn leads to a smaller mean
and variance of the statistic under correct model specification. In the extreme case of comonotonicity, \( T(\hat{F}_{jk}) = T\left[ G_{jk}(\cdot; \hat{\theta}_n) \right] \) and the statistic drops to zero.

In the rest of the chapter, we consider the case where all weights are 1 (i.e., \( w_{jk} \) in (7.3)), with the only exception of the vine structure where partitioning of the bivariate margins for separate treatment is recommended; this corresponds to setting some of the weights to 1 and the rest to 0 (see Section 7.2.2). All critical values correspond to the 95% quantile of the reference distribution.

### 7.2.2 Strategies for different parsimonious models

This subsection summarizes the parsimonious dependence structures, and the suggested strategies for them. It serves as an elaboration of some of the pathways described above. The methods described can in principle be applied to other structures not considered here:

1. **Exchangeable:** The distribution of an exchangeable model is invariant to permutation of variable indices. Therefore, it may be a suitable model when all variables have similar pairwise dependence strength with each other.

2. **Factor:** The factor structure may be plausible when the strengths of pairwise dependence are similarly ordered across columns and rows. For example, the correlation matrix of a \( d \)-dimensional 1-factor Gaussian model is given by

   \[
   \begin{pmatrix}
   1 & \alpha_1\alpha_2 & \alpha_1\alpha_3 & \cdots & \alpha_1\alpha_{d-1} & \alpha_1\alpha_d \\
   1 & \alpha_2\alpha_3 & \cdots & \alpha_2\alpha_{d-1} & \alpha_2\alpha_d \\
   \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\
   \cdots & \cdots & \alpha_{d-1}\alpha_d & 1
   \end{pmatrix},
   \]

   where \( \alpha_1, \ldots, \alpha_d \) are the factor loadings. The dependence strengths are similarly ordered in the sense that one row (column) is a multiple of another. Note that the pattern may not be as apparent in higher-order factor models.

3. **Truncated vine:** Variables in a vine structure form an acyclic network; dependence is the strongest between variables close together in the network, and weakens as one goes further out. A 1-truncated vine is also known as a Markov tree. As an example, the correlation matrices of a Gaussian 1-truncated C-vine rooted at the first variable and 1-truncated D-vine, with \( \rho_{ij} \) being the correlation between variables \( i \) and \( j \), are
as follows:

\[
\begin{pmatrix}
1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1d} \\
\rho_{12} & 1 & \rho_{12}\rho_{13} & \cdots & \rho_{12}\rho_{1d} \\
\vdots & & \ddots & \ddots & \ddots \\
\rho_{1,d-1}\rho_{1d} & \cdots & \cdots & 1 \\
\end{pmatrix}
\]

\[
\begin{pmatrix}
1 & \rho_{12} & \rho_{12}\rho_{23} & \cdots & \prod_{i=2}^{d} \rho_{i-1,i} \\
\rho_{23} & 1 & \rho_{23}\rho_{24} & \cdots & \prod_{i=3}^{d} \rho_{i-1,i} \\
\vdots & & \ddots & \ddots & \ddots \\
\rho_{d-1,d} & \cdots & \cdots & 1 \\
\end{pmatrix}
\]

A higher-order vine can be constructed by relaxing the conditional independence assumption. A Gaussian \(p\)-truncated vine only restricts partial correlations conditioning on \(k\) variables to be zero, for all \(k \geq p\). Note that the factor structure can be written as a C-vine if the latent variables are included.

4. Structured extreme value copulas: This category includes extreme value copulas with parsimonious dependence structures, such as those studied in Chapter 3. It is not mutually exclusive with the categories above; for example, the exchangeable Gumbel copula is an extreme value copula and is at the same time one with an exchangeable structure.

With these structures in mind, the issues and strategies mentioned in Section 7.2.1 are elaborated as follows:

1. The empirical features are in all cases easily computable, but not so for the model-based version. Models that are closed under margins have numerically tractable bivariate marginal distributions, so that bivariate features can be evaluated\(^\text{17}\). Features of a Gaussian copula with any dependence structure can be easily obtained as long as the model-based correlation matrix is available. For general copulas with various dependence structures, the feasibility can be summarized as follows:

- **Exchangeable**: Easily computable;
- **1- and 2-factor**: Computable with numerical integration;
- **1- and 2-truncated C-vine**: Easily computable for pairs linked in tree 1; computable with numerical integration for pairs linked in trees 2 and beyond;
- **Other truncated vines**: Easily computable for pairs linked in tree 1; computable with numerical integration for pairs linked in trees 2. Pairs beyond tree 2 involve higher-order integrations.

\(^\text{17}\)In the case of truncated C-vines, this is interpreted in the sense of including also the rooted variables.
• Structured extreme value: All structured models in Chapter 3 have model-based
dependence measures either easily computable or computable with numerical
integration.

In general, as long as simulation is feasible, the features can be estimated via Monte
Carlo simulation. This is however impractical in our case because such procedure has
to be repeated for every sample in the parametric bootstrap.

2. For empirical U-statistics with models fitted by maximum likelihood, the decision
path (to obtain a sampling distribution of $Q_n$, or the asymptotic covariance matrix
$\Sigma$, or an approximation of the critical value of the adequate-of-fit statistic) can be
summarized as follows:

a) Parametric bootstrap may be conducted to obtain the sampling distribution of
the adequacy-of-fit statistic when model simulation and fitting are efficient and
the fitted features can be easily computed. This method bypasses the evaluation
of $\Sigma$, and no approximation of the distribution of $Q$ is necessary. Technically,
one may need to check that the magnitude of the sample covariance matrix of
scaled differences (i.e., multiplied by $\sqrt{n}$) does not grow with the sample size $n$,
to ensure that the assumption in Theorem 6.6 is valid empirically. However, in
our subsequent applications and simulations, no such violation is detected and
thus we will treat the sample covariance matrix, for sufficiently large $n$, as a
reasonable estimate of the covariance matrix of the reference distribution.

b) If simulation is quick but model fitting is computationally costly, or the model-
based features are hard to obtain, two alternative approaches can be considered:

i. Evaluate the asymptotic covariance matrix $\Sigma$ indirectly (Section 7.4). We
develop methods to evaluate $\Sigma$ without having to fit the assumed model
repeatedly. It will still be necessary to perform maximum likelihood esti-
mation once, typically with a sample size larger than would be required for
parametric bootstrap, to obtain an estimate of the Fisher information ma-
trix and various quantities to sufficient precision. This method bypasses the
computation of model-based features, but may not scale very well to high
dimensions with many parameters.

ii. Use the Gaussian copula with the same dependence structure and compara-
ble parameter values (such as by matching pairwise Kendall’s $\tau$), and then
apply a multiplying factor to obtain a conservative bound (Section 7.5). We

152
observe that models with the same dependence structure but different linking copulas have critical values that behave similarly, and the difference in their magnitudes seems to be related to their tail behaviour. The Gaussian copula is used because simulation, fitting and obtaining model-based features are all efficient. A suitable multiplying factor can be applied to the resulting critical value to obtain a conservative bound for the target distribution. When the dimension is high, we further propose a simplified process for speed improvement.

Table 7.2 lists the possible methods to obtain the critical value for some parsimonious dependence structures, using Kendall’s $\tau$ or Spearman’s $\rho$ as dependence measures and under maximum likelihood estimation. Some methods described in the table allow us to estimate critical values for copula models with arbitrary linking copulas; these critical values are then compared against those for the Gaussian copula to arrive at a sensible value of the multiplying factors.

For $p$-truncated vine models, we suggest that the covariance matrix be partitioned according to whether the pair is modelled in the first $p$ trees (hereafter we refer to these pairs as “before” truncation level\(^{18}\), and the rest “after”). That is, we reorder the vector in (7.4) to read

$$
\sqrt{n} \begin{pmatrix}
  T(\hat{F}_{b_1}) - T\left[ G_{b_1}(\cdot; \hat{\theta}_n) \right] \\
  \vdots \\
  T(\hat{F}_{b_r}) - T\left[ G_{b_r}(\cdot; \hat{\theta}_n) \right] \\
  T(\hat{F}_{a_1}) - T\left[ G_{a_1}(\cdot; \hat{\theta}_n) \right] \\
  \vdots \\
  T(\hat{F}_{a_s}) - T\left[ G_{a_s}(\cdot; \hat{\theta}_n) \right]
\end{pmatrix} \overset{d}{\to} N\left( \mathbf{0}, \begin{pmatrix}
  \Sigma_{bb} & \Sigma_{ba} \\
  \Sigma_{ab} & \Sigma_{aa}
\end{pmatrix} \right),
$$

(7.8)

where $b_1, \ldots, b_r$ (resp. $a_1, \ldots, a_s$) are each an ordered pair indexing the bivariate margins before (resp. after) the truncation level, $r+s = \binom{d}{2}$ and the asymptotic covariance matrix is partitioned with $b$ and $a$ indexing “before” and “after”, respectively. This is necessary because the magnitudes of the differences are quite different between these two groups. When one fits a $p$-truncated vine, one expects the pairs up to the truncation level to be fitted better than those after because the latter involves certain independence copulas in the model hierarchy (restrictive assumption). As a result, the magnitudes of the differences for the “before” pairs are likely smaller than

\(^{18}\)More precisely, it refers to “before and including” truncation level.
the “after” pairs, as is observed in Section 7.5 for the critical values. Appendix C contains an example on the Gaussian vine with estimated variances, for which it is possible to prove that the empirical and model-based correlations are the same for the “before” pairs. Such partitioning allows one to inspect different aspects of the model fit: The “before” part for the adequacy-of-fit of the assumed structure (e.g., a $p$-truncated C-vine versus D-vine) and the “after” part for the discrepancy of residual dependence. Note that this partitioning is an instance of allocating different weights to each component of the difference vector, such that $w_{b_1} = \cdots = w_{b_r} = 1$ (resp. 0) and $w_{a_1} = \cdots = w_{a_s} = 0$ (resp. 1) for the “before” (resp. “after”) statistic.

<table>
<thead>
<tr>
<th>Structure</th>
<th>$d^*$</th>
<th>Strategy for models estimated via ML; Feature chosen is Kendall’s $\tau$ or Spearman’s $\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p$-factor</td>
<td>Small</td>
<td>Model-based features can only be obtained efficiently if $p$ is small (1 or 2), where parametric bootstrap is possible. If $p$ is larger, one may consider evaluating $\Sigma$ (Section 7.4).</td>
</tr>
<tr>
<td></td>
<td>Large</td>
<td>Model fitting may be slow; approximation of critical values using Gaussian surrogate model (Section 7.5) should be considered.</td>
</tr>
<tr>
<td>$p$-truncated vine</td>
<td>Small</td>
<td>Model fitting is relatively quick, but model-based features cannot be obtained efficiently in most cases (except for C-vines with small $p$). Evaluating $\Sigma$ is recommended.</td>
</tr>
<tr>
<td></td>
<td>Large</td>
<td>Model fitting may be slow; approximation of critical values using Gaussian surrogate model should be considered.</td>
</tr>
<tr>
<td>Structured EV</td>
<td>All</td>
<td>ML is generally not possible. In Section 7.6, we seek surrogate models and perform parametric bootstrap.</td>
</tr>
</tbody>
</table>

* “Large” and “small” values of $d$ depend on the level of speed and accuracy desired.

Table 7.2: Possible approaches of obtaining critical values of the adequacy-of-fit statistic for different parsimonious dependence structures, assuming models are estimated via maximum likelihood (ML) with features being Kendall’s $\tau$ or Spearman’s $\rho$. Except for the last row, these approaches form the basis on which analysis is conducted in Sections 7.4 and 7.5.

3. The class of extreme value copulas (with feature being Kendall’s $\tau$, Spearman’s $\rho$ or extremal coefficient) has to be treated separately as models are typically fitted by composite (pairwise) likelihood due to the intractability of multivariate densities. Parametric bootstrap is possible as bivariate distribution functions are available. A surrogate model may be considered when the target model is difficult to simulate from, and in this case a multiplying factor for the critical value may be needed (Section 7.6).
4. For empirical feature being tail-weighted dependence measures, we mention in Section 7.7 that the asymptotic variance for the empirical measure is usually much higher than the model-based measure, and hence the asymptotic behaviour of the difference can be approximated by that of the empirical measure alone. Because no model-based feature is involved, its associated challenges do not apply in this case.

5. Complex multivariate models may be fitted by methods other than the maximum likelihood, and critical values obtained from models estimated via these alternative methods can be different. Obviously parametric bootstrap is still a choice if feasible, but we suggest in Section 7.8 that critical values pertinent to the maximum likelihood estimation may serve as a conservative (upper) bound. This is especially relevant when the sample size is large; see Section 7.2.3.

7.2.3 Relationship between the adequacy-of-fit statistic and SRMSR

The feature or functional analogue of the SRMSR in (7.2) can be defined as

\[
\text{SRMSR} = \left[ \left( \frac{d}{2} \right)^{-1} \sum_{1 \leq j < k \leq d} w_{jk} \left( T(\hat{F}_{jk}) - T \left[ G_{jk}(\cdot; \hat{\theta}_n) \right] \right)^2 \right]^{1/2} = \sqrt{\frac{Q_n}{n}},
\]

where \( w_{jk} \) is the weight associated with the difference for the margin \((j, k)\) and \( Q_n \) is as defined in (7.3). Although distributional properties are derived for \( Q_n \), the SRMSR has the same unit of the feature and thus allows a direct interpretation in terms of the magnitude of the differences. Expressing critical values in terms of the SRMSR also allows the researcher to relate them to the practical significance for the specific problem at hand. For instance, it is possible that the reference distribution suggests a critical value much smaller than the limit of the tolerance that is deemed scientifically significant based on subject knowledge, especially when the sample size is large. In this case, even if the fitted model results in an SRMSR greater than the critical value, it may still be acceptable to use this model for inference. For this reason, when it is hard to obtain exact critical values, conservative bounds may suffice if they are still smaller than what is deemed practically significant.

The magnitude of the SRMSR is dependent on the sample size as both \( T(\hat{F}_{jk}) \) and \( T \left[ G_{jk}(\cdot; \hat{\theta}_n) \right] \) converge to \( T(F_{jk}) \) when the model is correctly specified. In order to compare the critical values for different models in simulation studies, it is convenient to use a common sample size, which is set at 100 in the rest of this chapter. For example, if a model has an SRMSR (with sample size 100) critical value of 0.08, we consider it inadequate for a sample of size \( n_1 \) when a sample SRMSR value is above 0.08. The SRMSR critical value for another sample size \( n_1 \), \( CV_{n_1} \), can be obtained through the relationship \( CV_{n_1} = \sqrt{100/n_1} \times CV_{100} \).
It is important not to confuse the SRMSR sample size and the simulation sample size; the latter is much larger than 100 in all cases.

Hooper et al. (2008) quote the results obtained in Hu and Bentler (1999) on a factor model, based on the empirical and model covariance matrices. They report a SRMSR critical value of 0.08 to be suitable for various sample sizes (ranging from 150 to 5,000).

### 7.3 Results for the exchangeable Gaussian distribution and some copulas with exchangeable dependence

This section is on properties of the adequacy-of-fit statistics for models with an exchangeable structure, based on central dependence measures and maximum likelihood estimation. We investigate the behaviour of the adequacy-of-fit statistic based on the exchangeable Gaussian distribution in Section 7.3.1, for which \( \Sigma \) can be obtained in closed form. It offers a good starting point to study how changes in the properties of the distribution (such as strength of dependence and dimension) will affect various quantities. In Section 7.3.2, we consider other exchangeable copula families whose limiting variances and covariances do not have closed-form expressions. We describe patterns to be expected and provide some comments.

Because of the restrictive assumption, exchangeable models have limited use in practice, but they are analytically or numerically more manageable. Also, some of the patterns shown here are also applicable to other dependence structures we consider later in this chapter.

#### 7.3.1 Theoretical results for the exchangeable Gaussian distribution

Let \( Y_1, \ldots, Y_n \) be \( d \)-dimensional i.i.d. variables following the exchangeable Gaussian distribution with density

\[
g(y; \mu, \sigma^2, \rho) = \frac{1}{(2\pi)^{d/2}|R|^{1/2}} \exp \left\{ -\frac{(y - \mu_1d)'R^{-1}(y - \mu_1d)}{2} \right\}, \quad R = \sigma^2 [(1 - \rho)I_d + \rho J_d],
\]

where \( \mu \) and \( \sigma^2 \) are respectively the common mean and variance, \( \rho \) is the common pairwise correlation, \( \mathbf{1}_d \) is a \( d \)-vector of 1’s, \( I_d \) is the \( d \)-dimensional identity matrix and \( J_d \) is a \( d \times d \) matrix of 1’s. In the following, we focus on non-negative dependence with \( \rho \geq 0 \).

The maximum likelihood estimators for this model (e.g., Joe and Lee (2009)) are

\[
\hat{\mu} = \frac{1}{nd} \sum_{i=1}^{n} \sum_{j=1}^{d} y_{ij}; \quad \hat{\sigma}^2 = \frac{s_1^2}{d}; \quad \hat{\rho} = \frac{1}{d-1} \left( \frac{s_2^2}{s_1^2} - 1 \right),
\]

156
where $y_{ij}$ are the realizations of $Y_i = (Y_{i1}, \ldots, Y_{id})^T$, $1 \leq i \leq n; 1 \leq j \leq d$,

$$s_1^2 = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{d} (y_{ij} - \hat{\mu})^2; \quad s_2^2 = \frac{1}{n} \sum_{i=1}^{n} \left( \sum_{j=1}^{d} y_{ij} - d\hat{\mu} \right)^2.$$ 

We are interested in the estimator $\hat{\rho}$ which specifies the dependence structure of the variables. Because the maximum likelihood estimator of $\rho$ is asymptotically independent with that of $\mu$, we henceforth consider the centred data $z_{ij} = y_{ij} - \mu$ and the corresponding random variables $Z_{ij}$'s, so that $\mu = \hat{\mu} = 0$ for the shifted observations.

When $d = 2$, or when we are considering only the $(j,k)$ margin of the distribution, the estimator for $\rho$ becomes

$$\hat{\rho}_{jk} = \frac{\sum_{i=1}^{n} (z_{ij} + z_{ik})}{\sum_{i=1}^{n} (z_{ij}^2 + z_{ik}^2)} - 1 = \frac{\sum_{i=1}^{n} z_{ij}z_{ik}}{\frac{1}{2} \left( \sum_{i=1}^{n} z_{ij}^2 + \sum_{i=1}^{n} z_{ik}^2 \right)}, \quad (7.10)$$

We use (7.10) as the “empirical” estimator for each bivariate margin. Note that the sample correlation coefficient is $r_{jk} = \sum_{i=1}^{n} z_{ij}z_{ik}/\sqrt{\sum_{i=1}^{n} z_{ij}^2 \sum_{i=1}^{n} z_{ik}^2}$, and it follows from the arithmetic mean–geometric mean inequality that $\hat{\rho}_{jk} \leq r_{jk}$. However, as $n \to \infty$, $r_{jk}$, $\hat{\rho}_{jk}$ and $\hat{\rho}$ all converge in probability to $\rho$.

With this setting, the difference statistic for the $(j,k)$ margin is given by $\hat{\rho}_{jk} - \hat{\rho}$, and the adequacy-of-fit statistic can be defined as

$$Q_n = n \left( \begin{array}{c} d \\ 2 \end{array} \right)^{-1} \sum_{j<k} (\hat{\rho}_{jk} - \hat{\rho})^2.$$ 

**Properties of the adequacy-of-fit statistic.** The derivation of the asymptotic covariance matrix for the difference vector $\sqrt{n}[\hat{\rho}_{12} - \hat{\rho}, \ldots, (\hat{\rho}_{d-1,d} - \hat{\rho})]^T$ is given in Appendix D; these quantities and the corresponding limits as $n, d \to \infty$ are summarized in Table 7.3. The asymptotic variance of $\hat{\rho}_{jk} - \hat{\rho}$ decreases with $\rho$ but increases with $d$; the reason for the latter trend is that the empirical variance does not change with $d$, but the model estimator $\hat{\rho}$ benefits from having more data available and thus its variance decreases with $d$, resulting in a higher variance for the difference. Meanwhile, the asymptotic covariances of the differences are generally not monotonic functions of $\rho$ because of two competing factors: (a) negative dependence due to a smaller (resp. larger) empirical estimate for a certain bivariate margin being compensated by a larger (resp. smaller) estimate for a different margin, while the model-based estimate $\hat{\rho}$ tends to be more stable, and (b) positive dependence due to a shared margin. Factor (b) is only relevant to the entries with shared margins, and thus those with distinct margins tend to have smaller covariances.

Next, we make use of (7.6) and (7.7) to compute approximate critical values for the adequacy-of-fit statistic $Q_n$ by matching the first two moments of the reference distribution.
Table 7.3: Summary of asymptotic variances and covariances of the differences for the $d$-dimensional exchangeable Gaussian distribution, using modified correlation as the feature $Q$ to that of a gamma distribution, invoking Theorem 6.6 for the convergence of asymptotic covariances. We compare the critical values from the approximating gamma distribution against those from the sampling distribution of $Q_n$ based on samples of size $n = 10,000$ using the limiting Gaussian distribution, and they match closely with each other. The mean of $Q$ is equal to $\lim_{n \to \infty} n \text{Var}(\hat{\rho}_{jk} - \hat{\rho})$ for any $j \neq k$; it decreases in $\rho$ and increases in $d$. The variance of $Q$ depends on the number of covariance terms with and without shared margins. For any $d$, the asymptotic covariance matrix $\Sigma$ contains $d(d-1)/2$ diagonal elements of variances and $d(d-1)(d-2)$ covariances with a shared margin or index. When $d \geq 4$, the rest or $d(d-1)(d-2)(d-3)/4$ elements are covariances without shared margins. The left and middle panels of Figure 7.1 plot the mean and variance of $Q$ for different $\rho$ and $d$, while the right panel shows the critical value based on the reference distribution $Q$, expressed in the scale of the SRMSR (with sample size 100).

From Figure 7.1, we observe that as $d$ increases, the increase in the mean of $Q$ is compensated by a rapid drop of its variance, leading to a smaller critical value. Also, at moderate to strong dependence, the critical value does not change much with $d$. It is also clear that the critical value drops as a function of $\rho$ and approaches zero as $\rho \to 1$.

### 7.3.2 Patterns for some copulas with exchangeable dependence

The maximum likelihood estimator for exchangeable copula models is in general analytically intractable, and so is $\Sigma$. Here we consider some exchangeable parametric copula families and explore the behaviour of the adequacy-of-fit statistic as a function of pairwise dependence or dimension. Since maximum likelihood fit is available for $d$ in the order of several dozens, we implement parametric bootstrap directly to obtain the SRMSR critical values$^{19}$.

$^{19}$For consistency with other methods described in the following sections, we obtain the covariance matrix of the sampling distribution of the difference vector and approximate the critical value by moment matching using a gamma distribution. In most cases, this approximation is very close to the critical values obtained...
Figure 7.1: Plots of means and variances of the reference distribution $Q$ for various dimensions (left and middle), and the critical values (95% quantiles) of $Q$ expressed in terms of SRMSR for sample size $n$, i.e., $\sqrt{(\text{CV of } Q)}/n$, for $n = 100$ (right).

Figure 7.2 shows the critical value of the SRMSR for some exchangeable copulas, including Gaussian (tail symmetric; no tail dependence), Frank (tail symmetric; lighter tails than Gaussian), Gumbel (tail asymmetric; upper tail dependence), MTCJ (very tail asymmetric; lower tail dependence) and $t$ (tail symmetric; both lower and upper tail dependence) with 3 and 20 degrees of freedom, at various dependence strengths and dimensions. The features considered are Kendall’s $\tau$ and Spearman’s $\rho$. Each parametric bootstrap uses 1,000 replications with a sample size of 1,000. An additional simulation set for the Gumbel exchangeable copula is run in order to get some idea on the variability of the SRMSR estimates; in most cases the differences are within 5% or 0.003 in magnitude. From the plots, we see that the general pattern that increasing dependence leads to a smaller critical value continues to hold. For the Gumbel and MTCJ copulas with higher dimensions, there is an interval of increasing critical value as a function of true Kendall’s $\tau$, before it drops eventually. The critical values with feature being Kendall’s $\tau$ are in most cases smaller than those for Spearman’s $\rho$, except when the dependence is very strong (Kendall’s $\tau$ beyond around 0.8). The variation among different parametric copula families is not very big but visible — generally smaller for Frank and Gaussian and higher for Gumbel, MTCJ and $t$ with small degrees of freedom. This agrees with the general trend mentioned in Section 6.4 that the asymptotic variance for the difference $T(\hat{F}_n) - T \left[ G(\cdot; \hat{\theta}_n) \right]$ tends to be larger for copulas with tail dependence. Finally, the critical values are decreasing in $d$, and are quite stable when $d \geq 10$. It appears that the critical values for small values of $d$ can be used as a conservative bound for higher dimensions.

directly from the sampling distribution of the SRMSR. With an SRMSR sample size of 100, the difference is mostly less than 0.003.
Figure 7.2: Plots of SRMSR (sample size 100) critical values against the common strength of dependence between variables expressed in terms of Kendall’s $\tau$, for various parametric copula families and dimensions. The features are Kendall’s $\tau$ (top two rows) and Spearman’s $\rho$ (bottom two rows).
7.4 Method on evaluating $\Sigma$ for factor and truncated vine models under maximum likelihood estimation

When it is hard to perform the parametric bootstrap due to costly model fitting procedure or model-based feature calculation, it may still be possible to evaluate $\Sigma$ under the maximum likelihood framework. Examples of parsimonious models that fall into this category are factor copulas with more than two factors or truncated vines, except for the C-vine with low truncation level. Model-based features may be estimated via simulation, but this is prohibitively costly since it is nested within the loop for the parametric bootstrap.

In this section, we discuss how the elements of the matrix $\Sigma$ in Theorem 7.1 can be evaluated when the copula model is estimated via maximum likelihood. Once the matrix is obtained, part 1 of Corollary 7.2 allows us to compute the moments of the statistic. Under the maximum likelihood framework, much of the computation needed to evaluate $\Sigma$ is simplified through the separability of variance/covariance property shown in Chapter 6. Since this property is only proved for the class of U-statistics, the following calculations pertain to empirical features being Kendall’s $\tau$ and Spearman’s $\rho$.

Under maximum likelihood, the convergence result (7.4) and the statement of the asymptotic distribution in Theorem 6.5 imply that the $(i_{jk}, i_{lm})$ element of $\Sigma$, i.e., the covariance for margins $(j, k)$ and $(l, m)$ in (7.4), is given by

$$
\sigma_{i_{jk}i_{lm}} = \sigma_{i_{jk}i_{lm}, \text{emp}} - \frac{\partial T}{\partial \theta^T} \left[ G_{jk}(\cdot; \theta_0) \right] I^{-1} \frac{\partial T}{\partial \theta} \left[ G_{lm}(\cdot; \theta_0) \right],
$$

(7.11)

where $\sigma_{i_{jk}i_{lm}, \text{emp}}$ is the asymptotic covariance of the empirical features for the margins concerned. Since we are dealing directly with the asymptotic distribution in this case, the assumption in Theorem 6.6 does not apply. The key to the evaluation of $\Sigma$ is that we can compute the two terms in the right hand side of (7.11) separately. To unify the following discussion, we henceforth consider the case when $G$ is a copula, denoted by $C$ with density $c$ and observations $U \sim C$ below. We will still use $\hat{F}$ to denote the empirical distribution, so as to avoid confusion with the reflected copula already denoted by $\hat{C}$.

The method described in this section allows us to estimate the critical value for truncated vine models in Section 7.5. We also intend to provide computer code for the evaluation of $\Sigma$; it can be applied as long as simulation from the parametric model and maximum likelihood estimation can be scaled to large sample sizes within reasonable computational cost.
7.4.1 Empirical covariance

For the feature $T$ being Kendall’s $\tau$, an expression of the asymptotic empirical covariance, which is equal to the limit of the scaled finite-sample covariance as $n \to \infty$, can be found in Genest et al. (2011):

\[
\lim_{n \to \infty} n \text{Cov} \left[ T(\hat{F}_{jk}), T(\hat{F}_{lm}) \right] = 16 \text{Cov} \left[ C_{jk}(U_j, U_k) + C_{jk}(U_j, U_k), C_{lm}(U_l, U_m) + C_{lm}(U_l, U_m) \right]
\]

\[
= 16 \int_0^1 (C_{jk} + C_{jk})(C_{lm} + C_{lm}) \, dC - 4(\tau_{jk} + 1)(\tau_{lm} + 1),
\]

(7.12)

where $\tau_{rs} = T[G_{rs}(\cdot; \theta_0)]$. This limit can be evaluated numerically in at least two ways:

1. The integral, equal to the expectation of the quantity $(C_{jk} + C_{jk})(C_{lm} + C_{lm})$ with respect to $C$, can be evaluated via Monte Carlo (MC) simulation from $C$. If the bivariate distribution functions cannot be efficiently computed, one may substitute this with the empirical estimate, i.e., $\hat{F}_{jk}(u_j, u_k) = N^{-1} \sum_{i=1}^{N} 1\{U_{ij} \leq u_j, U_{ik} \leq u_k\}$ and similarly for the survival functions, where $N$ is the MC sample size. The “naïve” way of evaluating this quantity, i.e., by comparing $O(N)$ values for each of the $N$ simulated observations, requires $O(N^2)$ comparisons and could be very slow when $N$ is large. Appendix E contains a more efficient algorithm that uses only $O(N \log_2 N)$ comparisons modified from the merge sort and quicksort algorithms. Given a large sample, $\tau_{jk}$ and $\tau_{lm}$ can be approximated by their sample versions.

2. Alternatively, since this limit does not depend on the parametric model fit, one may simply construct the sampling distribution of $(T(\hat{F}_{jk}), T(\hat{F}_{lm}))$. Using the algorithm in Knight (1966), the sample Kendall’s $\tau$ can be computed at complexity $O(n \log_2 n)$ for a sample of size $n$.

A simulation run suggests that method 1 requires a much larger sample size to achieve comparable accuracy as method 2. This is not surprising given that the latter requires many replications while the former evaluates the integral with one MC simulation sample. Approximating the bivariate distribution and survival functions by their empirical counterparts introduces extra uncertainty for method 1. Figure 7.3 shows the result of the simulation for a 1-truncated D-vine with 8 observed variables and Gumbel linking copulas, all having a Kendall’s $\tau$ of 0.5. The histogram in the left panel shows the aggregate distribution of the estimated asymptotic empirical covariance using method 1, between the margins $(12, 13), (23, 24), (34, 35), (45, 46), (56, 57)$ and $(67, 68)$ which have the same theoretical value. The MC sample size is 1,000,000 and the whole procedure is repeated 100 times. Meanwhile,
the right panel shows an analogous histogram using method 2 with sample size 10,000 and 1,000 replications in each run, for a total of 100 runs.

Figure 7.3: Histograms of asymptotic empirical covariance estimates using different methods, based on a 1-truncated D-vine with 8 variables and Gumbel linking copulas with Kendall’s τ = 0.5. The covariances considered are those for the Kendall’s τ feature between bivariate margins (12, 13), (23, 24), (34, 35), (45, 46), (56, 57) and (67, 68). “MCN” refers to the Monte Carlo simulation sample size.

With the feature being Spearman’s ρ, method 2 is more desirable because the integral involves more complicated expressions. It is also preferred in terms of coding as one simply needs to replace the computation of the sample Kendall’s τ by that of Spearman’s ρ, requiring minimal effort.

7.4.2 Model-based covariance

The model-based covariance, i.e., the second term in the right hand side of (7.11), contains the gradient of the functional and the Fisher information matrix. The Fisher information matrix can be estimated to desired accuracy by fitting the model to a large simulated sample, independent of the functional being used. For the gradient, we have the following:

- When $T$ is Kendall’s τ:
  \[
  \frac{\partial T}{\partial \theta} [C_{jk}(\cdot; \theta_0)] = 4 \int_0^1 \left[ C_{jk}(u; \theta_0) + \overline{C}_{jk}(u; \theta_0) \right] \ell_{jk}(\theta_0; u) \, dC_{jk}(u; \theta_0), \quad (7.13)
  \]
where $\ell^\prime_{jk}(\theta_0; u) = \partial \log c_{jk}(u; \theta_0) / \partial \theta$ and $c_{jk}$ is the copula density. The proof is as follows, where for brevity we drop the subscripts denoting the bivariate margins. Since $T[C(\cdot; \theta_0)] = 4 \int_0^1 C(u; \theta_0) dC(u; \theta_0) - 1$, its derivative with respect to $\theta$ is

$$\frac{\partial T}{\partial \theta}[C(\cdot; \theta_0)] = 4 \int_0^1 \left[ \frac{\partial C}{\partial \theta}(u; \theta_0)c(u; \theta_0) + C(u; \theta_0) \frac{\partial c}{\partial u}(u; \theta_0) \right] du.$$

As $\int_0^1 C(u; \theta_0) \frac{\partial}{\partial u}(u; \theta_0) du = \int_0^1 \overline{C}(u; \theta_0) \ell'(u; \theta_0) dC(u; \theta_0)$, it suffices to prove that

$$\int_0^1 \frac{\partial C}{\partial \theta}(u; \theta_0)c(u; \theta_0) du = \int_0^1 \overline{C}(u; \theta_0) \ell'(u; \theta_0) dC(u; \theta_0),$$

which is derived below using integration by parts:

$$\int_0^1 \overline{C}(u; \theta_0) \ell'(u; \theta_0) dC(u; \theta_0)
= \int_0^1 \int_0^1 \overline{C}(u_1, u_2; \theta_0) \frac{\partial^2 C}{\partial u_2 \partial \theta}(u_1, u_2; \theta_0) du_1 du_2
= \int_0^1 \left\{ \overline{C}(u_1, u_2; \theta_0) \frac{\partial^2 C}{\partial u_2 \partial \theta}(u_1, u_2; \theta_0) \right\} du_1
= -\int_0^1 \int_0^1 \frac{\partial^2 C}{\partial u_2 \partial \theta}(u_1, u_2; \theta_0) \frac{\partial C}{\partial u_1}(u_1, u_2; \theta_0) du_1 du_2
= -\int_0^1 \int_0^1 \left\{ \frac{\partial C}{\partial u_1}(u_1, u_2; \theta_0) \frac{\partial C}{\partial \theta}(u_1, u_2; \theta_0) \right\} du_1
= \int_0^1 \frac{\partial C}{\partial \theta}(u; \theta_0)c(u; \theta_0) du.$$

The two differences in square brackets are zero, as (a) $\overline{C}(1, u_2; \theta_0) = 0$ for all $u_2$; (b) $C(0, u_2; \theta_0) = 0$ for all $u_2$ and $\theta_0$; (c) $\overline{C}(u_1, 1; \theta_0) = 0$ for all $u_1$, and; (d) $C(u_1, 0; \theta_0) = 0$ for all $\theta_0$.

- When $T$ is Spearman's $\rho$:

$$\frac{\partial T}{\partial \theta}[C_{jk}(\cdot; \theta_0)] = 12 \frac{\partial}{\partial \theta} \int_0^1 \int_0^1 u_1 u_2 dC_{jk}(u_1, u_2; \theta_0).
= 12 \int_0^1 \int_0^1 u_1 u_2 \ell'_{jk}(\theta_0; u_1, u_2) dC_{jk}(u_1, u_2; \theta_0). \quad (7.14)$$

In each case, $\partial T[C_{jk}(\cdot; \theta_0)] / \partial \theta$ can be written as an expectation where the integrands can be easily evaluated or estimated for a given observation: $\ell'_{jk}(\theta_0; u)$ can be approximated using a numerical gradient, while $C_{jk}(u; \theta_0)$ and $\overline{C}_{jk}(u; \theta_0)$ can be replaced by the empirical counterparts. The expressions (7.13) and (7.14) can thus be evaluated by MC simulation.
without first obtaining $T[C_{jk}(\cdot; \theta_0)]$. This is useful for bivariate margins that do not have analytic forms.

Using the 1-truncated D-vine example in the previous subsection, we evaluate the derivative of the model-based Kendall’s $\tau$ using (7.13) with an MC sample size of 10,000 and 100,000. Together with the Hessian matrix from the model fit, we arrive at estimates of the asymptotic model-based covariance between the margins (12, 13), (23, 24), (34, 35), (45, 46), (56, 57) and (67, 68). The aggregated frequencies are charted in Figure 7.4, and suggest that the estimates are too variable with MC sample size 10,000. For 1- and 2-truncated vines, a sample size of 100,000 can be fitted quickly for $d = 8$. The scalability of model fitting for $p$-factor copula is not as good as its density is a $p$-dimensional integral. When computational efficiency is a concern, it may be better to approximate or get a bound for the critical value using the Gaussian copula, to be described in the Section 7.5.

![Histograms of asymptotic model-based covariance estimates](image)

Figure 7.4: Histograms of asymptotic model-based covariance estimates, based on a 1-truncated D-vine with 8 variables and Gumbel linking copulas with Kendall’s $\tau = 0.5$. The covariances considered are those for the Kendall’s $\tau$ feature between bivariate margins (12, 13), (23, 24), (34, 35), (45, 46), (56, 57) and (67, 68). “MCN” refers to the Monte Carlo simulation sample size.

Because the MC procedures for the empirical and model-based covariances are completely separate, there is no guarantee that the estimated $\sigma_{i,j,k,l}$, i.e., the diagonal elements of $\Sigma$, are positive. This may happen when the dependence is strong, so that the true asymptotic empirical and model-based variances are close to each other. It is also possible to obtain non-positive-definite covariance matrices. When this occurs, it may be necessary
to increase the sample size (of parametric bootstrap simulation and/or MC evaluation of integrals) to obtain more precise estimates.

7.5 Finding an approximate critical value using a matching Gaussian copula

When parametric bootstrap is impractical and the asymptotic covariance matrix $\Sigma$ of the vector of differences cannot be reliably evaluated, it may be constructive to at least explore crude patterns that can be expected. This can be achieved by using the Gaussian copula as a surrogate model, with dependence structure and strengths matching those of the target copula. In this section, we focus on matching pairwise Kendall’s $\tau$ values; this matching Gaussian copula can be obtained in the following manner:

1. Obtain model-based Kendall’s $\tau$ of the bivariate linking copulas for the multivariate parametric copula model $C$ in question;

2. Replace the linking copulas of $C$ by Gaussian ones, preserving the values of Kendall’s $\tau$ as well as the structure of the model. The resulting copula is the matching Gaussian copula.

Using the Gaussian copula as a surrogate model is an attractive option as its simulation and fitting procedures are available for many dependence structures, and in most cases these have been optimized for speed. Furthermore, model-based features can be easily obtained from the fitted correlation matrix. In this section, we investigate the behaviour of the critical value for Gaussian copulas with factor and vine dependence structures, when the copula is fitted via maximum likelihood. The subsections are organized as follows:

1. In Section 7.5.1, we investigate the effects on the critical values of replacing linking copulas of structured models with Gaussian ones. The investigation focuses on copulas with low dimensions and small number of factors/truncation level as parametric bootstrap or the evaluation of $\Sigma$ is efficient in this case. The results allow us to suggest sensible values for the multiplying factors for the conversion of critical values based on Gaussian copulas to those of arbitrary copulas.

2. In Section 7.5.2, we explore the trends of the critical values as a function of the average pairwise dependence of the structured Gaussian copula, for factor and truncated vine models with various dimensions. We observe that the critical values do not vary much with the dimension. Also, the critical values for Gaussian copulas having a
certain structure depend much on the average pairwise dependence. This justifies the summary of critical values in Table 7.6 based on average dependence strengths.

3. In Section 7.5.3, we introduce methods to improve the computational efficiency of model fitting for Gaussian copulas. This subsection aims to: (a) explore patterns for higher dimensions or number of factors/truncation levels, and (b) verify that these alternative methods yield critical values similar to those in Section 7.5.2 for cases that can be compared.

The results of this section, i.e., critical values for Kendall’s $\tau$ and Spearman’s $\rho$ measures based on the Gaussian copula and multiplying factors for conversion to models with other linking copulas, are summarized in Table 7.6.

7.5.1 Behaviour of different copulas with the same dependence structure

Results in Section 6.4 point to larger asymptotic variances (and thus critical values) when the tail dependence of the linking copulas is strong. We conduct some simulations on factor and truncated vine copula models to observe the magnitude of this difference relative to the Gaussian copula; their results, with the feature being Kendall’s $\tau$, are shown in Figures 7.5 and 7.6. For ease in interpretation, the figures plot the ratio of SRMSR critical values using Gaussian as the baseline. The SRMSR critical value for each Gaussian copula is given in the axis label. The simulation scenarios (indexed using Roman numerals in the figures) are listed in Tables 7.4 and 7.5.

For truncated vines, we focus on C- and D-vines as they can be considered as the two boundaries of possible vine structures (Kurowicka and Joe (2011)). Results are shown only for C-vines as those for D-vines are similar; we will briefly comment on the differences in the following analysis. Two other random vine structures are attempted but there is no special behaviour worthy of mention.

Most of the parametric copula families attempted have one dependence parameter, including the t copula where the degrees of freedom parameter is fixed in model fitting. The only exception is the BB1 copula which has two parameters: $\theta > 0$ and $\delta > 1$. Here we fix $\delta$ at 1.2, so that the upper tail dependence index $\lambda_U$ is fixed at $2 - 2^{1/\delta} = 0.22$, and find the value of $\theta$ that matches the stipulated Kendall’s $\tau$ value.\(^{20}\)

From the figures, we observe that the critical values obtained for the Gaussian copula are in the middle of the spectrum. The Frank copula or Frank/Frank combination (lighter tails than Gaussian) generally yields the smallest critical values. For factor models, the

\(^{20}\)Another value of $\delta$ is attempted but nothing special is observed.
**1-factor:** Parametric bootstrap is used with a sample size of 10,000 for Gaussian linking copulas and 1,000 for all others; 1,000 replications throughout.

<table>
<thead>
<tr>
<th>$d$</th>
<th>Scenario</th>
<th>Kendall’s $\tau$ values of linking copulas</th>
<th>Avg. pairwise $\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5/8</td>
<td>I/II/III</td>
<td>Same throughout (0.3/0.5/0.8)</td>
<td>0.13/0.33/0.72*</td>
</tr>
<tr>
<td>5</td>
<td>IV</td>
<td>Equally spaced between 0.3 and 0.8</td>
<td>0.36</td>
</tr>
<tr>
<td>8</td>
<td>V</td>
<td>Equally spaced between 0.3 and 0.8</td>
<td>0.37</td>
</tr>
</tbody>
</table>

**2-factor:** Parametric bootstrap is used with a sample size of 10,000 and 1,000 replications for Gaussian linking copulas; sample size 1,000 and 100 replications for all others. Linking copulas and strengths of dependence for each factor are chosen to mimic practical situations: Copulas for the second factor typically have weak tail dependence (Frank or t with large degrees of freedom**) and smaller Kendall’s $\tau$ value.

<table>
<thead>
<tr>
<th>$d$</th>
<th>Scenario</th>
<th>Kendall’s $\tau$ values of linking copulas</th>
<th>Avg. pairwise $\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>I</td>
<td>1st factor: 0.20–0.40; 2nd factor: 0.10–0.15</td>
<td>0.15</td>
</tr>
<tr>
<td>5</td>
<td>II</td>
<td>1st factor: 0.50–0.70; 2nd factor: 0.10–0.30</td>
<td>0.47</td>
</tr>
<tr>
<td>8</td>
<td>III</td>
<td>1st factor: 0.20–0.35; 2nd factor: 0.10–0.15</td>
<td>0.13</td>
</tr>
<tr>
<td>8</td>
<td>IV</td>
<td>1st factor: 0.55–0.70; 2nd factor: 0.10–0.30</td>
<td>0.51</td>
</tr>
</tbody>
</table>

* Values are only approximate as they can differ slightly among parametric copula families; likewise for other average pairwise $\tau$ listed in this section.

** We include the Gumbel/Gumbel combination to observe the behaviour when all linking copulas have somewhat stronger tail dependence.

Table 7.4: Simulation scenarios for factor copulas to be used for different parametric linking copula families

Gumbel copula (asymmetric with upper tail dependence) and t copula with 3 degrees of freedom (symmetric with tail dependence) have larger critical values than Gaussian, while the t copula with 20 degrees of freedom behaves similarly as the Gaussian copula. The Gumbel/Gumbel, BB1/Frank and Gumbel/Frank combinations (BB1 copula is asymmetric with dependence in both tails) may have considerably larger critical values than Gaussian (a factor of 1.2 or more). For the truncated vine models, we observe that there is more variability in the “before” critical values than the “after” ones. The “before” critical values are usually much smaller than the “after” ones, as mentioned in Section 7.2. The behaviour with respect to the strength of tail dependence of the linking copulas is similar to the factor copulas, with combinations involving the MTCJ copula (asymmetric with strong lower tail dependence) yielding the largest critical values. Not shown here, the “after” critical values for the D-vines (in the region between 0.8 and 0.95) are larger than those for the C-vines (between 0.6 and 0.9), while the spread among models is somewhat smaller.

Based on the simulation results, we propose the following multiplying factors (MF) for
1-truncated*

<table>
<thead>
<tr>
<th>d</th>
<th>Scenario</th>
<th>Kendall’s τ values of linking copulas</th>
<th>Avg. pairwise τ</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>I/II/III</td>
<td>Same throughout (0.3/0.5/0.7)</td>
<td>C-vine: 0.20/0.40/0.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>D-vine: 0.17/0.38/0.61</td>
</tr>
<tr>
<td>5</td>
<td>IV</td>
<td>Equally spaced between 0.3 and 0.8</td>
<td>C/D-vine: 0.43/0.41</td>
</tr>
<tr>
<td>8</td>
<td>V/VI/VII</td>
<td>Same throughout (0.3/0.5/0.7)</td>
<td>C-vine: 0.17/0.38/0.61</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>D-vine: 0.12/0.28/0.53</td>
</tr>
<tr>
<td>8</td>
<td>VIII</td>
<td>Equally spaced between 0.3 and 0.8</td>
<td>C/D-vine: 0.41/0.33</td>
</tr>
</tbody>
</table>

2-truncated*: We assume weaker dependence for the linking copulas in the second tree**

<table>
<thead>
<tr>
<th>d</th>
<th>Scenario</th>
<th>Kendall’s τ values of linking copulas</th>
<th>Avg. pairwise τ</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>I</td>
<td>1st tree: 0.20–0.35; 2nd tree: 0.10–0.15</td>
<td>C/D-vine: 0.21/0.20</td>
</tr>
<tr>
<td>5</td>
<td>II</td>
<td>1st tree: 0.55–0.70; 2nd tree: 0.10–0.30</td>
<td>C/D-vine: 0.57/0.57</td>
</tr>
<tr>
<td>8</td>
<td>III</td>
<td>1st tree: 0.20–0.35; 2nd tree: 0.10–0.15</td>
<td>C/D-vine: 0.20/0.17</td>
</tr>
<tr>
<td>8</td>
<td>IV</td>
<td>1st tree: 0.55–0.70; 2nd tree: 0.10–0.25</td>
<td>C/D-vine: 0.53/0.50</td>
</tr>
</tbody>
</table>

* The method in Section 7.4 is used to evaluate Σ. The empirical asymptotic covariance is estimated from the sampling distribution with sample size 100,000, while for the model-based covariance, the integral is evaluated with an MC sample size of 100,000.

** We include the MTCJ/MTCJ combination to observe potential extreme behaviour.

Table 7.5: Simulation scenarios for truncated vine copulas to be used for different parametric linking copula families

the conversion of SRMSR critical values for the Gaussian copula to other families. This factor is not necessary if every linking copula has weaker tail dependence than Gaussian.

- For 1-factor copula, a MF of 1.2 appears sufficient in most cases. For 2-factor copulas, the MF should be set at around 1.3, or 1.4 if all linking copulas have stronger tail dependence (e.g., Gumbel) and when the average pairwise Kendall’s τ is small.

- For truncated vines, the MF’s are different for the “before” and “after” critical values:
  - “Before”: Unless the linking copulas in the first tree are strongly tail dependent, a MF of 1.2 is recommended. If the overall pairwise dependence is weak but a copula with strong tail asymmetry such as the MTCJ is used, a larger MF such as 1.6 should be applied.
  - “After”: There is much less variation and a MF of 1.2 should be sufficient.

The SRMSR critical values (sample size 100) for the Gaussian copulas attempted are mostly between 0.04 and 0.06 for factor copulas and the “before” statistic of truncated vine cop-
Figure 7.5: Ratio of SRMSR critical values (for sample size 100) for various 1- and 2-factor copulas, with the feature being Kendall’s \( \tau \). Text left of each panel indicates the dimension, scenario number and the critical value for a Gaussian copula model of that scenario. The scenarios are listed in Table 7.4.

ulas, and between 0.06 and 0.09 for the “after” statistic. With the multiplying factors, a conservative approximation of the critical value is around 0.08 for factor models and the “before” statistic of truncated vines, and 0.11 for the “after” statistic. This means that we consider a factor or truncated vine model adequate if the average discrepancy (root mean square deviation) between the pairwise empirical and model-based Kendall’s \( \tau \) is less than 0.08 (or 0.11 for the “after” statistic), based on a sample of size 100. This is close to the cutoff value in Hooper et al. (2008), but our critical value is smaller when the sample size is larger than 100. For example, with a sample size of 1,000, the SRMSR critical values are around 0.025 for the factor models and the “before” statistic of truncated vines, and 0.035 for the “after” statistic, using the square root rule.

The trends and the multiplying factors for Spearman’s \( \rho \) are similar and are thus not repeated here. We however note that the critical values based on Spearman’s \( \rho \) are generally larger than those based on Kendall’s \( \tau \) unless the overall pairwise dependence is very strong. This applies to the next two subsections on Gaussian copulas as well; for the sake of brevity,
Figure 7.6: Ratio of SRMSR critical values (for sample size 100) for various 1- and 2-truncated C-vine copulas, with the feature being Kendall’s \( \tau \). Text left of each panel indicates the dimension, scenario number and the critical value for a Gaussian copula model of that scenario. The scenarios are listed in Table 7.5.

we will thus focus on the discussion based on Kendall’s \( \tau \) in the remainder of this section.

7.5.2 Gaussian copula approximation for factor and vine structures

With higher dimensions, model estimation becomes more demanding. It may not be possible to evaluate model-based bivariate marginal Kendall’s \( \tau \) or Spearman’s \( \rho \) using their defining integration operations with more factors and higher truncation levels. In this and the next
subsections, we focus on the patterns of the critical values based on Gaussian copulas, as the model-based features can be conveniently obtained as a transformation of fitted correlations. We aim to establish conservative bounds of the critical values for various Gaussian copulas.

Figures 7.7 and 7.8 show the SRMSR critical values for the Gaussian copula with factor and truncated vine structures, respectively, using Kendall’s $\tau$ as the feature. In each case, a parametric bootstrap is conducted with $d = 5, 8$ or 11. For vines, we separate the statistic into the “before” and “after” components. The points (with various shapes) represent the scenarios with realistic dependence strengths (dependence in the second layer weaker than the first); such scenarios have Kendall’s $\tau$ of each linking copula randomly (uniformly) generated within the following regions:

- 1-factor and 1-truncated vine: All parameters between $(0.1, 0.3); (0.3, 0.5); (0.5, 0.7); (0.7, 0.9)$ and between $(0.1, 0.9);

- 2-factor and 2-truncated vine: Parameters in the two layers between $\{(0.3, 0.5), (0.1, 0.3)\}$; $\{(0.5, 0.7), (0.3, 0.5)\}$; $\{(0.7, 0.9), (0.5, 0.7)\}$ and $\{(0.7, 0.9), (0.1, 0.3)\}$.

In each case, a sample size of 1,000 is used with 500 replications. All scenarios are repeated (2 to 3 times) to provide insight on the variability within each simulation setting.

![Figure 7.7: SRMSR critical values (for sample size 100) for Gaussian 1- and 2-factor copulas, with the feature being Kendall’s $\tau$. Lines: Scenarios with identical linking copulas; Points: With random realistic linking copulas.](image-url)
Figure 7.8: SRMSR critical values (for sample size 100) for Gaussian 1- and 2-truncated C- and D-vine copulas, with the feature being Kendall’s \( \tau \). Lines: Scenarios with identical linking copulas; Points: With random realistic linking copulas.

In addition to these points, we run simulations where all linking copulas are identical; these are run with sample size 10,000 and 1,000 replications and the resulting SRMSR critical values are shown as line segments in the figures.

We can see that the dimension has a relatively small effect on the critical value, especially when the dimension is already large, similar to the trend observed in Figure 7.2. For both structures, the separation between the lines for \( d = 8 \) and 11 is smaller than that between the lines for \( d = 5 \) and 8. This suggests that if high-dimensional simulation or model fitting is impractical, it may still be constructive to look at the behaviour with lower dimensions.

For truncated vines, the critical value of the “before” statistic generally peaks at moderate dependence, while that of the “after” statistic appears to be monotonically decreasing as a function of pairwise dependence. Such observation provides further support to separating the statistic for vines based on the truncation level.

Finally, we note that the locations of the points are well described by the lines which assume identical linking copulas. This suggests that the critical value for a given scenario can be well approximated by one that has only one dependence parameter (average pairwise Kendall’s \( \tau \)). This is especially useful when there are more factors or the truncation level.
is high. The conservative critical values in Table 7.6 are based on this principle. There are several points that are quite a bit lower than the corresponding lines for the 1-factor and 1-truncated vines; these are mostly the cases with largely different parameter values. In this case, the lines can be interpreted as a conservative bound for models having the same structure and pairwise Kendall’s $\tau$, but potentially different parameters.

Even with the Gaussian copula, model fitting can become prohibitively costly for large $d$. In the next subsection, we propose an approach to further reduce the computational burden. This allows us to explore patterns that have not yet been addressed in this subsection.

### 7.5.3 Improving computational efficiency

When there are many parameters, we propose the following modification to improve the efficiency of fitting the Gaussian model. First, we estimate each correlation parameter using only the data for the bivariate margin concerned. Let $Y_1, \ldots, Y_n$, with $Y_i = (Y_{1i}, Y_{2i})^\top$, denote a (centred) sample from the Gaussian distribution with zero mean and covariance matrix $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, $-1 \leq \rho \leq 1$. The conventional maximum likelihood estimator for the correlation is \( \hat{\rho} = \frac{S_{12}}{\sqrt{S_{11}S_{22}}} \), where $S_{pq} = n^{-1} \sum_{i=1}^n Y_{ip}Y_{iq}$; it corresponds to the one where the whole covariance matrix is estimated. That is to say, it is the full maximum likelihood estimator when both marginal parameter (variance) and dependence parameter (correlation) are estimated at the same time. To estimate only the dependence parameter as in fitting a copula, we fix the variances at 1 and obtain the log-likelihood function

$$
\ell(\rho, y) = -\frac{n}{2} \log |\Sigma| - \frac{1}{2} \sum_{i=1}^n y_i^\top \Sigma^{-1} y = -\frac{n}{2} \log(1 - \rho^2) - \frac{n}{2} \frac{1}{1 - \rho^2} (S_{11} - 2\rho S_{12} + S_{22}) .
$$

When $\rho \neq \pm 1$, setting $\partial \ell / \partial \rho$ to zero yields

$$
-\rho^3 + \rho^2 S_{12} + \rho(1 - S_{11} - S_{22}) + S_{12} = 0 , \quad (7.15)
$$
a cubic equation in $\rho$. The real root of this equation, i.e., the maximum likelihood estimator of $\rho$ with variance fixed\(^{21}\), is generally not the sample correlation. When the true correlation is 1 (resp. $-1$), the log-likelihood function asymptotes to infinity as $\rho \to 1$ (resp. $-1$) and therefore this yields the same estimate as the sample correlation. It is theoretically possible for (7.15) to have three real roots in $(-1, 1)$, but this only happens with unrealistic (yet compatible) combinations of $S_{11}, S_{12}$ and $S_{22}\(^{22}\)$. In practical application, we can thus\(^{21}\) strictly speaking, one has to check that the solution $\hat{\rho}$ is indeed a maximum. The proof appears difficult; we numerically experiment with various sensible configurations and find that $\hat{\rho}$ maximizes $\ell$ in each case, and so for practical purposes it is reasonable to assume that this holds.

\(^{22}\) As an illustration, a combination of $S_{11} = S_{22} = 0.3$ and $S_{12} = 0.05$ yields the three real roots ($-0.528, -0.133, 0.711$). It is however practically impossible to obtain $S_{11} = S_{22} = 0.3$ with typical simulation sample sizes from a Gaussian distribution with unit marginal variances.
assume there is only one real root and this is the maximum likelihood estimator fixing the marginal variances at 1.

After estimating all pairwise correlations, denoted as $\hat{\Sigma} = (\hat{\rho}_{jk})$, we proceed in the following manner:

- **Factor model:** Conduct factor analysis using $\hat{\Sigma}$ as the correlation matrix (e.g., `factanal` in R). The fitted correlation matrix is then converted to a matrix of Kendall’s $\tau$ or Spearman’s $\rho$. By using $\hat{\Sigma}$ and not the sample correlation matrix, we ensure that no marginal information is transferred to subsequent estimation procedures. The implementation of `factanal` in R is fast enough to handle higher-dimensional data. This two-stage approach is not maximum likelihood estimation as $\hat{\Sigma}$ does not correspond to the covariance matrix from a statistical model, and the critical values need not be equal to those from the copula-only approach. Comparisons using the scenarios in Section 7.5.2 indicate that a further multiplying factor of 1.1 is appropriate for the conversion of critical values based on this approach (henceforth referred to as restricted maximum likelihood estimation) to those using the copula approach.

- **Truncated vine model:** Convert $\hat{\Sigma}$ to a partial correlation matrix using the specified vine structure. Truncate all higher-order partial correlations to zero and convert the resulting matrix back to correlations. This correlation matrix can then be converted to a matrix of Kendall’s $\tau$ or Spearman’s $\rho$. Simulations using this method result in critical values within 0.001 or 0.002 of those from the copula-only approach (well within sampling variability), and we suggest that no further adjustment is needed in this case.

In essence, for the factor structure we make use of the fast `factanal` function with an adjustment to bypass the fitting of marginal variances, while for the vine structure we avoid joint optimization of the parameters altogether. This provides considerable time savings: As an illustration, with a 2-truncated D-vine with 11 variables and a sample size of 1,000, 100 replications of the parametric bootstrap takes around 10 seconds with this restricted maximum likelihood method, compared to around 2 minutes with a full maximum likelihood fit on the Gaussian copula\(^\text{23}\).

Using this method, we run simulations analogous to those in the previous subsection but with more variables, factors and higher truncation levels. When the feature is Kendall’s $\tau$, we show in Figures 7.9 and 7.10 the critical values for some factor and truncated C-vine models, with $d = 10, 15, 20$ and 25 and up to 4 factors / 4-truncated vines. The realistic scenarios (plotted as points) have the same configuration as in the previous subsection for

---

\(^{23}\)Run with an Intel Core i5-2450M CPU (2.5 GHz) with 6 GB of RAM.
Figure 7.9: SRMSR critical values (for sample size 100) for various Gaussian factor copulas using the restricted maximum likelihood method, with the feature being Kendall’s $\tau$. Lines: Scenarios with identical linking copulas; Points: With random realistic linking copulas.

Figure 7.10: SRMSR critical values (for sample size 100) for various Gaussian C-vine copulas using the restricted maximum likelihood method, with the feature being Kendall’s $\tau$. Lines: Scenarios with identical linking copulas; Points: With random realistic linking copulas.

1- and 2-factor and truncated vine models, and for higher-order models we consider the following:

- 3-factor and 3-truncated vine: Parameters (expressed in Kendall’s $\tau$) in the three layers between $\{(0.5, 0.7), (0.3, 0.5), (0.1, 0.3)\}$; $\{(0.7, 0.9), (0.5, 0.7), (0.3, 0.5)\}$;
\{(0.7, 0.9), (0.3, 0.5), (0.1, 0.3)\} and \{(0.3, 0.5), (0.1, 0.3), (0.1, 0.3)\};

• 4-factor and 4-truncated vine: Parameters (expressed in Kendall’s \(\tau\)) in the four layers between \{(0.7, 0.9), (0.5, 0.7), (0.3, 0.5), (0.1, 0.3)\}; \{(0.7, 0.9), (0.3, 0.5), (0.1, 0.3), (0.1, 0.3)\}; \{(0.5, 0.7), (0.3, 0.5), (0.1, 0.3), (0.1, 0.3)\} and \{(0.3, 0.5), (0.1, 0.3), (0.1, 0.3), (0.1, 0.3)\}.

The lines plot the critical values for scenarios that assume identical linking copulas. Note that results for the factor models have not been adjusted using the multiplying factor mentioned above. Results for D-vines are similar to those of C-vines (except for a slightly larger “after” critical value as mentioned in the previous subsection) and are not shown here. Random R-vines are also attempted but there are no special observations.

These simulations provide further evidence that increase in dimension has small effect on the critical values. The points for realistic scenarios are well approximated by the lines that assume identical linking copulas, even for 3- or 4-factor and truncated vines. Based on these observations, this restricted maximum likelihood method is a viable approach for higher-dimensional copulas when reduction of computation time is needed.

From the results in this and the preceding subsections, we utilize the following combinations of scenarios to obtain conservative bounds of the SRMSR critical values for Gaussian models:

• Factor copula: SRMSR critical values for 1-factor copula with 5 variables, applicable to all dimensions;

• “Before” statistic for truncated vine copulas: The maximum of SRMSR critical values for 1/2-truncated C/D-vines with 5 variables, and 1/4-truncated C/D-vines with 20 variables. It is also applicable to all dimensions.

• “After” statistic for truncated vine copulas: The maximum of SRMSR critical values for 1/2-truncated C/D-vines with 5 variables to be used for lower dimensions (fewer than 20); the maximum of SRMSR critical values for 1-truncated C/D-vines with 20 variables to be used for higher dimensions (20 or more).

The resulting values, as well as multiplying factors for the conversion to other parametric copula families suggested in Section 7.5.1, are summarized in Table 7.6. As the simulation sets above have different average pairwise Kendall’s \(\tau\), these values are smoothed using cubic splines. They are useful when only a crude critical value is needed.

To summarize, when it is possible (both analytically and computationally) to conduct parametric bootstrap or evaluate \(\Sigma\), they are the preferred methods as the resulting critical value is the most relevant to the model being studied. Otherwise, the Gaussian copula can
be used as a surrogate model, from which a critical value is obtained and a multiplying
factor applied (if necessary) to arrive at a conservative bound for the target model. When
the dimension is high, speed improvement for the Gaussian copula is possible using the
restricted maximum likelihood method.

7.6 Multivariate extreme value copulas — an example for pairwise likelihood estimation

Separate consideration for multivariate extreme value copulas is needed as these models
are typically fitted by composite likelihood methods. In addition, simulation from certain
families can be unreliable or even impossible, precluding direct parametric bootstrap. In
this section, we investigate the behaviour of the critical value using the t-EV copula (2.11)
as a surrogate for parametric bootstrap when simulation is difficult. We also consider the
use of the extremal coefficient in addition to Kendall’s $\tau$ and Spearman’s $\rho$, where the
empirical extremal coefficient is estimated using the rank-based F-madogram (5.6) with
the exponentiation parameter $\alpha = 1$.

Using the t-EV copula has two advantages:

1. It is an extreme value copula and the dependence properties may be closer to other
classes of extreme value copulas than the Gaussian copula is. When the feature
used is the extremal coefficient, the model-based estimate can be easily obtained as
a transformed value of the distribution function for extreme value copulas.

2. Its parametrization involves a correlation matrix $\Omega$, on which parsimonious depen-
dence structures can be imposed.

We first describe how one can obtain a matching t-EV copula for extreme value models that
cannot be easily simulated from. A simulation study is then carried out to investigate the
performance of this surrogate in terms of the critical values obtained. Finally, we suggest
how this procedure can be used for general parsimonious extreme value copulas; this is
relevant to the data examples we studied in Section 3.7.

7.6.1 Finding a matching t-EV copula

For extreme value copulas, we are mainly interested in the three parametric copula families
in Chapter 3, namely extreme value factor copula, Hüsler-Reiss copula and t-EV copula.
Because of their stochastic representations, the extreme value factor copula and t-EV copu-
ula can be approximately simulated by taking componentwise maxima of their underlying
### Measure: Kendall’s $\tau$

<table>
<thead>
<tr>
<th>Avg. pair. $\tau$</th>
<th>Factor</th>
<th>Vine</th>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All $d$</td>
<td>Low $d$</td>
<td>High $d$</td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.064</td>
<td>0.102</td>
<td>0.076</td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>0.060</td>
<td>0.094</td>
<td>0.079</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>0.061</td>
<td>0.092</td>
<td>0.080</td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>0.063</td>
<td>0.087</td>
<td>0.078</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.061</td>
<td>0.081</td>
<td>0.074</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>0.054</td>
<td>0.075</td>
<td>0.066</td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>0.044</td>
<td>0.060</td>
<td>0.054</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.031</td>
<td>0.042</td>
<td>0.041</td>
<td></td>
</tr>
</tbody>
</table>

### Measure: Spearman’s $\rho$

<table>
<thead>
<tr>
<th>Avg. pair. $\tau$</th>
<th>Factor</th>
<th>Vine</th>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All $d$</td>
<td>Low $d$</td>
<td>High $d$</td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.096</td>
<td>0.073</td>
<td>0.113</td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>0.088</td>
<td>0.070</td>
<td>0.140</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>0.086</td>
<td>0.074</td>
<td>0.133</td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>0.082</td>
<td>0.073</td>
<td>0.121</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.072</td>
<td>0.066</td>
<td>0.106</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>0.056</td>
<td>0.056</td>
<td>0.086</td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>0.036</td>
<td>0.048</td>
<td>0.058</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.019</td>
<td>0.038</td>
<td>0.030</td>
<td></td>
</tr>
</tbody>
</table>

### Multiplying factors for other copula families

<table>
<thead>
<tr>
<th>Structure</th>
<th>Property</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-factor</td>
<td>All*</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td><strong>2-factor</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moderate to strong tail dependence in factor 1 only (e.g., Gumbel, BB1, t)</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>Strong tail dependence in both factors, small average pairwise $\tau$</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td><strong>1- and 2-truncated vine (before)</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moderate to strong tail dependence (e.g., Gumbel, BB1, t)</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>Strong tail dependence and asymmetry, small average pairwise $\tau$ (e.g., MTCJ)</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td><strong>1- and 2-truncated vine (after)</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>All*</td>
<td>1.2</td>
</tr>
</tbody>
</table>

* No adjustment needed if the copula has similar or lighter tails than Gaussian.

### Notes on usage:

1. Calculate average pairwise $\tau$ (for all pairs) of fitted target model and obtain critical value for the chosen structure; interpolate if needed. For the “after” statistic of vines, use “Low $d$” if there are fewer than 20 variables and “High $d$” otherwise.

2. Depending on the tail properties of the target model, apply multiplying factors as needed.

3. Convert to SRMSR critical value for sample size $n$ with the factor $\sqrt{100/n}$.

4. Gaussian copulas with 1–4 factors and 1–4 truncation levels for vines with up to $d = 25$ were studied in the derivation of the above critical values; we expect similar patterns for other factor/vine structures. For other dependence structures, it may be possible to obtain conservative bounds using a similar technique.

Table 7.6: Summary of conservative SRMSR critical values (for sample size 100) for Gaussian copulas with factor and truncated vine structures. Tables are: SRMSR critical values as a function of average pairwise Kendall’s $\tau$ when the feature is Kendall’s $\tau$ (top left) and Spearman’s $\rho$ (top right); multiplying factors for other parametric copula families suggested in Section 7.5.1 (middle); notes on usage (bottom).
models, i.e., factor copula and t copula, respectively. There is however no equivalent procedure for the Hüsler-Reiss copula, and thus one of the emphases in this section is to explore methods whereby approximate critical values can be obtained for Hüsler-Reiss copulas.

The use of surrogate models is also necessary for t-EV models with large degrees of freedom \( \nu \) as the bivariate dependence can be weak even when the correlation parameter \( \rho \) is very high. For example, when \( \nu = 50 \), even a value of \( \rho = 0.95 \) yields a Kendall’s \( \tau \) of only 0.2. For practical data with moderate dependence, this means that \( \Omega \) mostly contains entries close to 1 when \( \nu \) is sufficiently large, in which case approximate simulation by taking componentwise maxima of the t copula is relatively poor. Note that the Hüsler-Reiss copula can be obtained as the limit of a t-EV copula as \( \nu \to \infty \) (Nikoloulopoulos et al. (2009)); this reinforces the idea that a t-EV copula “close to” Hüsler-Reiss cannot be simulated accurately by taking componentwise maxima.

Because of these reasons, we suggest the use of t-EV copulas with small values of \( \nu \) as surrogate models. In the following, we choose \( \nu = 3 \) and we expect that the critical values for other reasonable choices of \( \nu \) will not be very different. To obtain a matching t-EV copula, we match pairwise extremal coefficients by finding \( \rho \) for a given \( \nu \). This is preferred over matching Kendall’s \( \tau \) as it is easier to obtain model-based extremal coefficients. This is repeated for each bivariate margin in order to get \( \Sigma_T \), the correlation matrix for the matching t-EV copula. If \( \Sigma_T \) is not positive definite, we search for the closest positive definite matrix \( \Sigma_T^+ \) (with a certain threshold on the smallest eigenvalue for numerical purposes); this is easily implemented in R using the nearPD function in the package Matrix, based on Higham (2002). It should be checked that the features of interest implied by \( \Sigma_T^+ \) are close to those of the target copula, in order to ensure that we are still approximating the target copula by a model that roughly preserves pairwise dependence strengths. After obtaining \( \Sigma_T^+ \), approximate simulation of the t-EV copula can be achieved by taking componentwise maxima of a sufficiently large sample with size \( n_t \) (1,000 for the simulation study below) generated from the t copula with \( \nu \) degrees of freedom and correlation matrix \( \Sigma_T^+ \). The componentwise maxima are then transformed to unit uniform using the Beta\((n_t, 1)\) distribution function, since they are maxima over \( n_t \) observations.

### 7.6.2 Simulation study

We conduct a simulation to study the performance of approximating the Hüsler-Reiss copula using a t-EV copula with \( \nu = 3 \). Although the general unconstrained Hüsler-Reiss copula cannot be simulated, a subset which coincides with the finite-dimensional distribution of the Smith or Brown-Resnick max-stable processes can be approximately simulated to satisfactory accuracy (Ribatet (2013)). We note that a goodness-of-fit test based on
the empirical and model-based extremal coefficients is considered in Shang (2013) in the
context of max-stable processes using parametric bootstrap.

For our purpose, we choose the Smith model with dimension \( d = 5 \) or 10:

1. Spatial locations \( \{Z_1^T, \ldots, Z_5^T\} = \{(0,0), (1.9,0.2), (1.1,1.0), (0.9,0.3), (0.4,1.1)\}\), corresponding to a 5-dimensional Hüsler-Reiss copula;

2. Spatial locations \( \{Z_1^T, \ldots, Z_{10}^T\} = \{(0,0), (0.7,0.7), (0.1,1.6), (1.3,0.0), (1.6,1.5), (1.8,0.7), (0.8,1.6), (0.4,1.1), (1.1,1.2), (1.3,0.5)\}\), corresponding to a 10-dimensional Hüsler-Reiss copula.

In each case, the dependence parameter \( \delta_{jk} \) in (2.1) of the Hüsler-Reiss copula, where the subscripts are added to indicate the bivariate pair, is related to the spatial locations through the equation \( 2/\delta_{jk} = \sqrt{(Z_j - Z_k)^T \Sigma_Z^{-1} (Z_j - Z_k)} \). Here \( \Sigma_Z \) is a deterministic matrix that regulates the strength and direction of spatial dependence among stations. We choose \( \Sigma_Z = (\frac{1}{\rho_Z} \rho_Z I) \) for three values of \( \rho_Z \): 0.25, 1 and 4, corresponding to weak, moderate and strong dependence, respectively. The pairwise Kendall’s \( \tau \) and extremal coefficients implied by these parameter sets for the 5-dimensional case are indicated in Table 7.7, and the average values for \( d = 10 \) are in Table 7.8.

<table>
<thead>
<tr>
<th></th>
<th>Kendall’s ( \tau )</th>
<th>Extremal coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Weak (( \rho_Z = 0.25 ))</strong></td>
<td>[Table entry]</td>
<td>[Table entry]</td>
</tr>
<tr>
<td></td>
<td>1 0.04 0.11 0.28 0.19</td>
<td>1 1.94 1.86 1.66 1.76</td>
</tr>
<tr>
<td></td>
<td>1 0.20 0.25 0.06</td>
<td>1 1.74 1.69 1.92</td>
</tr>
<tr>
<td></td>
<td>1 0.39 0.40</td>
<td>1 1.53 1.52</td>
</tr>
<tr>
<td></td>
<td>1 0.28</td>
<td>1 1.65</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>Average: 0.28</strong></td>
<td></td>
<td><strong>Average: 1.73</strong></td>
</tr>
<tr>
<td><strong>Moderate (( \rho_Z = 1 ))</strong></td>
<td>[Table entry]</td>
<td>[Table entry]</td>
</tr>
<tr>
<td></td>
<td>1 0.27 0.38 0.56 0.48</td>
<td>1 1.66 1.54 1.36 1.44</td>
</tr>
<tr>
<td></td>
<td>1 0.49 0.54 0.31</td>
<td>1 1.43 1.38 1.62</td>
</tr>
<tr>
<td></td>
<td>1 0.64 0.65</td>
<td>1 1.28 1.28</td>
</tr>
<tr>
<td></td>
<td>1 0.56</td>
<td>1 1.36</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>Average: 0.49</strong></td>
<td></td>
<td><strong>Average: 1.44</strong></td>
</tr>
<tr>
<td><strong>Strong (( \rho_Z = 4 ))</strong></td>
<td>[Table entry]</td>
<td>[Table entry]</td>
</tr>
<tr>
<td></td>
<td>1 0.56 0.64 0.76 0.71</td>
<td>1 1.37 1.29 1.19 1.23</td>
</tr>
<tr>
<td></td>
<td>1 0.71 0.74 0.59</td>
<td>1 1.22 1.20 1.34</td>
</tr>
<tr>
<td></td>
<td>1 0.81 0.81</td>
<td>1 1.14 1.14</td>
</tr>
<tr>
<td></td>
<td>1 0.76</td>
<td>1 1.19</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>Average: 0.71</strong></td>
<td></td>
<td><strong>Average: 1.23</strong></td>
</tr>
</tbody>
</table>

Table 7.7: Pairwise Kendall’s \( \tau \) and extremal coefficient for the three simulation scenarios with \( d = 5 \)

For each scenario, simulation from the Smith model is done using the package SpatialExtremes in R. We check the goodness-of-fit of each bivariate margin to the target
Table 7.8: Average pairwise Kendall’s τ and extremal coefficient for the three simulation scenarios with $d = 10$

<table>
<thead>
<tr>
<th>Weak ($\rho_Z = 0.25$)</th>
<th>Moderate ($\rho_Z = 1$)</th>
<th>Strong ($\rho_Z = 4$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. Kendall’s τ</td>
<td>0.25</td>
<td>0.51</td>
</tr>
<tr>
<td>Avg. extremal coefficient</td>
<td>1.70</td>
<td>1.42</td>
</tr>
</tbody>
</table>

Hüsker-Reiss copula on some selected scenarios, using the function `gofCopula` from the `copula` package; no substantial anomaly is found. Parametric bootstrap (with 1,000 samples of size 1,000) is performed on the true model, the t-EV copula with matching pairwise extremal coefficients, and another one with matching Kendall’s τ. Results of the simulation are shown in Table 7.9. In addition, we consider the $t_3$ and Gaussian copulas by matching Kendall’s τ, and obtain critical values of the SRMSR for the features Kendall’s τ and Spearman’s ρ. However their performance is worse than the t-EV copula (critical values are farther away from those using the true model) and therefore these two families are omitted.

Table 7.9: SRMSR critical values (for sample size 100) using Hüsker-Reiss copula and t-EV copulas with matching pairwise Kendall’s τ or extremal coefficient (in square brackets), under weak (W), moderate (M) and strong (S) dependence scenarios for $d = 5$ (above) and 10 (below). The average pairwise Kendall’s τ and extremal coefficients for these scenarios are in Tables 7.7 and 7.8.

<table>
<thead>
<tr>
<th>$d = 5$</th>
<th>Kendall’s τ</th>
<th>Spearman’s ρ</th>
<th>Extremal coefficient $\vartheta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copula</td>
<td>W M S</td>
<td>W M S</td>
<td>W M S</td>
</tr>
<tr>
<td>Hüsker-Reiss</td>
<td>0.055 0.060 0.047</td>
<td>0.078 0.072 0.042</td>
<td>0.072 0.063 0.043</td>
</tr>
<tr>
<td>t-EV ($\nu = 3$) $[\vartheta]$</td>
<td>0.051 0.055 0.045</td>
<td>0.074 0.068 0.044</td>
<td>0.063 0.055 0.038</td>
</tr>
<tr>
<td>t-EV ($\nu = 3$) $[\tau]$</td>
<td>0.053 0.055 0.046</td>
<td>0.076 0.068 0.043</td>
<td>0.064 0.054 0.038</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$d = 10$</th>
<th>Kendall’s τ</th>
<th>Spearman’s ρ</th>
<th>Extremal coefficient $\vartheta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copula</td>
<td>W M S</td>
<td>W M S</td>
<td>W M S</td>
</tr>
<tr>
<td>Hüsker-Reiss</td>
<td>0.055 0.057 0.044</td>
<td>0.074 0.069 0.040</td>
<td>0.067 0.060 0.040</td>
</tr>
<tr>
<td>t-EV ($\nu = 3$) $[\vartheta]$</td>
<td>0.047 0.052 0.043</td>
<td>0.066 0.064 0.043</td>
<td>0.056 0.052 0.037</td>
</tr>
<tr>
<td>t-EV ($\nu = 3$) $[\tau]$</td>
<td>0.048 0.052 0.043</td>
<td>0.067 0.064 0.042</td>
<td>0.057 0.052 0.036</td>
</tr>
</tbody>
</table>

It can be seen that most of the SRMSR critical values obtained using the t-EV copula are somewhat smaller than those of the Hüsker-Reiss copula. This difference appears to decrease with the overall strength of dependence, for both $d = 5$ and 10. Using the t-EV copula as a surrogate should provide a reasonable approximation of the critical value, but we suggest a multiplying factor of 1.2 for a conservative bound for the Hüsker-Reiss copula. It should however be noted that the scenarios considered above are limited and future research can focus on obtaining a more accurate estimate of the critical value for
Hüsler-Reiss copulas. Results of the t-EV copula based on matching Kendall’s τ show little difference with those based on matching extremal coefficients; this suggests that the choice of the feature to match does not matter much from a practical point of view. Note also that the largest critical values for Kendall’s τ occur with moderate dependence, a phenomenon we observed in Section 7.5. Finally, the critical values for $d = 5$ are generally larger than those for $d = 10$. Although we assume different spatial locations in the two cases, the average Kendall’s τ and extremal coefficients are similar. This observation appears to be consistent with that in Section 7.5 as well.

7.6.3 Implementation for parsimonious extreme value copulas

The above simulation study suggests that using the t-EV copula with $\nu = 3$ as a surrogate model for the Hüsler-Reiss copula may allow us to obtain comparable SRMSR critical values. We are unable to conduct a simulation study with structured Hüsler-Reiss copulas (such as the class of models discussed in Section 3.3) because they are generally not finite-dimensional distributions of max-stable processes and thus cannot be simulated using that framework. However, we can use procedures similar to that in Section 7.6.1 to obtain a surrogate t-EV copula, from which approximate simulation is possible.

For general parsimonious extreme value copulas models, we adopt the following strategy in finding an approximate critical value for the adequacy-of-fit statistic or SRMSR:

- If the model has a stochastic representation that allows (approximate) simulation, such as by taking componentwise maxima, and model fitting via composite likelihood methods is possible, we suggest the use of parametric bootstrap on the fitted model. Because the convergence to the extreme value limit as $n \to \infty$ may be slow, one may need to perform a preliminary check on the simulated data to see if they are representative of the target model, for example by comparing empirical extremal coefficients with the model-based counterpart.

- If approximate simulation is not possible or of poor quality, such as with the structured Hüsler-Reiss copula or t-EV copula with large degrees of freedom, we suggest the use of a t-EV copula with small degrees of freedom (e.g., $\nu = 3$) as a surrogate model. With this method, one should match the pairwise extremal coefficient (or other dependence measures) of the target model to the surrogate model as in Section 7.6.1. From here, there are two slightly different ways to proceed:

  1. Samples are simulated from this matching t-EV surrogate model and fitted with the corresponding structure imposed on the t-EV copula, and then the empirical and model-based features are computed.
2. Alternatively, impose the target structure on the correlation matrix of the t-EV copula first. The resulting correlation matrix will most likely be slightly different from that of the matching copula. For the factor structure, parameters of this model can be obtained by performing factor analysis on the correlation matrix, while for truncated vines with given vine structure and truncation level, one can find parameters that yield the “closest” correlation matrix with respect to some metric, such as the Frobenius norm of the matrix of differences. Parametric bootstrap is conducted on this conforming surrogate model.

In the first method, the generating model is closest to the target distribution in terms of pairwise dependence measures, while the second method ensures that the generating model is the same as the fitting model. There is little difference between the two if the data possess (approximately) the structure being fitted, but our experience with some data sets suggests that critical values may be inflated substantially if this is not the case. We recommend using the second method to ensure that critical values are obtained from fitting a correctly specified model. As in the simulation study above, a multiplying factor of 1.2 should be applied.

An illustration of SRMSR critical value estimation is given in the second and third examples in Section 7.9.

### 7.7 Adequacy-of-fit for tails of copulas

In certain applications, it is crucial to use copulas with matching tail properties in order to obtain valid inferences. Examples include the calculation of joint exceedance probabilities in extreme value theory, or the value-at-risk of a portfolio that contains assets with dependent returns. Since Kendall’s $\tau$ and Spearman’s $\rho$ summarize dependence over the whole range of the copula, they are not desirable measures for describing tail dependence. As demonstrated in Section 5.7.3 of Joe (2014), model-based estimation of Kendall’s $\tau$ and Spearman’s $\rho$ can often be insensitive to the parametric copula family fitted unless the model is badly misspecified. The tail dependence indices given by (2.13) and (2.14) measure the strength of tail dependence, but they are defined by limits and have no direct sample counterpart except for the case of extreme value copulas. For general copulas, tail-weighted dependence measures (Krupskii and Joe (2015)) have sample versions that are readily available and thus emerge as a viable tool for assessing adequacy-of-fit in the tails. In this section, we explore some characteristics of the adequacy-of-fit statistic (or SRMSR) with features being tail-weighted dependence measures.
7.7.1 Properties of the difference statistic for a single bivariate margin

In the following, we use the weighting function \( a(u) = u^6 \) and quantile \( p = 0.5 \) as suggested in Krupskii and Joe (2015), so that the lower and upper tail-weighted dependence measures for \((U_1, U_2) \sim C\) are defined respectively as

\[
\rho_L = \text{Cor}[(1 - 2U_1)^6, (1 - 2U_2)^6 | U_1 < 0.5, U_2 < 0.5];
\]

\[
\rho_U = \text{Cor}[(2U_1 - 1)^6, (2U_2 - 1)^6 | U_1 > 0.5, U_2 > 0.5],
\]

with empirical version for observations \( \{(X_{i1}, X_{i2}); i = 1, \ldots, n\} \) being the sample correlation of adjusted ranks \( R_{ij} = n^{-1} \left[ \text{Rank}_i(X_{ij}) - 0.5 \right], j = 1, 2, \) that lie within the appropriate quadrant. The model-based tail-weighted dependence measures are obtained using numerical integration methods (Gauss-Legendre quadrature).

As in Section 6.4, we check the behaviour of the asymptotic variance for the empirical and model-based features for different bivariate parametric copula families. Figure 7.11 displays the results for six families: Gaussian and Frank copulas are symmetric and have no tail dependence; t copula is symmetric and has dependence in both tails; Gumbel copula is asymmetric and has upper tail dependence, while MTCJ copula is asymmetric and has lower tail dependence. For ease in comparison, we reflect the MTCJ copula so that the upper tail is tail dependent instead. Bivariate copulas with Kendall’s \( \tau \) of \( 0.1, 0.2, \ldots, 0.8 \) are considered for each family. In each case, samples of size 100,000 are simulated and fitted to the generating family by maximum likelihood. We obtain the sampling distribution of the difference statistic (and hence asymptotic variances) using 1,000 replications. These variances are plotted against the true values of the tail-weighted dependence measures because Kendall’s \( \tau \) is not indicative of the tail properties of a copula.

We observe from Figure 7.11 that the asymptotic variance of the empirical feature is typically much higher than the model-based counterpart when the model is correctly specified, regardless of the copula family and especially for weak and moderate dependence. This is not surprising because the model-based feature is obtained from the parameter estimate, which utilizes information from the whole sample. Meanwhile, the empirical feature is estimated with a subset of the observations, and thus there is much less information available. For example, with the independence copula, there are on average only \( 1/4 \) of the observations in a specific quadrant for the estimation of the empirical feature; this fraction increases to around \( 1/2 \) in the case of the comonotonicity copula. This property can also be observed from Figure 7.11, in the sense that the ratio of the asymptotic variances between the empirical and model-based features decreases along the \( x \)-axis.

As the strength of dependence tends towards comonotonicity, both the empirical and model-based features tend to 1 and the asymptotic variances are eventually decreasing. At
the independence limit, Krupskii and Joe (2015) show that the asymptotic variance of the empirical feature is equal to $1/p^2$, or 4 in this case with $p = 0.5$. With the exception of $t$ copulas, the other four parametric copula families considered in Figure 7.11 approach the independence limit as the Kendall’s $\tau$ value tends to zero. From this, we can infer that for some parametric copula families, the asymptotic variance of the empirical measure may increase slightly with strength of dependence when the latter is weak, before eventually dropping. This is clearly the case for the upper tails of the Gumbel and reflected MTCJ copulas. Such behaviour resembles that of the asymptotic variance of the empirical Kendall’s $\tau$ for these two copulas, as noted in Section 2.12.6 of Joe (2014).

Finally, comparing the behaviour for the two tails, Figure 7.11 also suggests that different parametric copula families, and possibly the different tails, behave similarly beyond moderate strength of tail dependence (i.e., for a given value of the measure). This suggests that the analysis of lower and upper tails may be combined.

Our simulation suggests that the asymptotic variance of the difference between empirical
and model-based tail-weighted dependence measures is only slightly smaller than that of
the empirical measure for a wide range of dependence strengths. Therefore, when assessing
adequacy-of-fit based on the difference statistic, it is practical to drop the model part and
simply estimate the variability of the empirical measure. This removes the main burdens
associated with the parametric bootstrap, and no surrogate model is needed as long as the
fitted model can be accurately simulated from.

7.7.2 SRMSR critical values for parsimonious dependence structures

Here we investigate the behaviour of the SRMSR critical values using tail-weighted depen-
dence measures for the factor and truncated vine structures. With the observation in the
preceding subsection in mind, we compute only the empirical features (for both tails) for
each simulated sample. The models and parameters are the same as those used in Sec-
tion 7.5.1, except that the MTCJ and BB1 copulas are reflected so that the upper tail has
stronger tail dependence, and several combinations are added.

Each simulation scenario uses 1,000 replications of sample size 100,000 (or 10,000 for
the more time-consuming 2-factor copula). The covariance matrix is used for matching
the first two moments of a reference gamma distribution, from which the critical value is
obtained and converted to that of SRMSR with a sample size of 100.

In Figure 7.12, we plot the critical values against the average pairwise tail-weighted
dependence measure for the tail being examined. Several dimensions of the plot have been
condensed, noting that critical values for the 1- and 2-factor copulas exhibit similar patterns,
and so do 1- and 2-truncated vines after separating into “before” and “after” statistics.
Results for the lower and upper tails, as well as C- and D-vines, are also combined. The
distinction between results for 1 and 2 layers of factor or vine structure is retained through
the use of a different colour for each layer. The aspects preserved in the figure include
the families of linking copulas (represented by symbols), dimension and the “before”/“after”
partitioning for truncated vines.

It can be clearly seen that the SRMSR critical values using tail-weighted dependence
measures are several times larger than those using Kendall’s $\tau$ and Spearman’s $\rho$ due to the
sample size reduction (as in the bivariate setting) and dropping the model-based estimation.
Regardless of the structure, these critical values appear to increase slightly as the average
pairwise tail-weighted dependence measure increases from zero, peak at moderate values
of the measure and decrease steadily afterwards. There is some spread among different
linking copulas especially at moderate tail dependence. Critical values for the Gaussian
and Frank copulas are respectively in the middle and lower end of the spectrum, resembling
the behaviour for Kendall’s $\tau$ and Spearman’s $\rho$. The plots for 8 variables have the same
general pattern as those for 5 variables and are almost a downward translation of the latter. These results also suggest that the “before” and “after” statistics for truncated vines should be kept separate.

The smoothed critical values for \( d = 5 \) are displayed in Table 7.10 as a function of the average feature. Given the observed trends, we expect the critical values for higher dimensions to be smaller than those for \( d = 5 \) and thus this table offers a conservative bound of the typical critical values. For both tails of the t copula and the dependent tail of the Gumbel and MTCJ copulas, a multiplying factor of 1.1 (for factor and the “before” part of truncated vines) to 1.2 (for the “after” part of truncated vines) is needed as the plots suggest higher critical values than the smoothed fit. It should be emphasized that the values given in Table 7.10 serve as a guide only. Because there are no model-based estimates involved, it is a straightforward task to obtain a sampling distribution of the covariance matrix for pairwise empirical tail-weighted dependence measures when simulation is efficient.
### Table 7.10: Summary of SRMSR critical values (for sample size 100) using tail-weighted dependence measures (TWDM) with factor and truncated vine structures. Multiply by $\sqrt{100/n}$ to obtain the corresponding critical value for sample size $n$.

<table>
<thead>
<tr>
<th>Average TWDM</th>
<th>Factor</th>
<th>Vine Before</th>
<th>Vine After</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.27</td>
<td>0.29</td>
<td>0.28</td>
</tr>
<tr>
<td>0.2</td>
<td>0.28</td>
<td>0.29</td>
<td>0.29</td>
</tr>
<tr>
<td>0.3</td>
<td>0.29</td>
<td>0.29</td>
<td>0.31</td>
</tr>
<tr>
<td>0.4</td>
<td>0.29</td>
<td>0.27</td>
<td>0.32</td>
</tr>
<tr>
<td>0.5</td>
<td>0.27</td>
<td>0.24</td>
<td>0.31</td>
</tr>
<tr>
<td>0.6</td>
<td>0.25</td>
<td>0.21</td>
<td>0.28</td>
</tr>
<tr>
<td>0.7</td>
<td>0.21</td>
<td>0.17</td>
<td>0.25</td>
</tr>
<tr>
<td>0.8</td>
<td>0.15</td>
<td>0.13</td>
<td>0.19</td>
</tr>
<tr>
<td>0.9</td>
<td>0.09</td>
<td>0.08</td>
<td>0.12</td>
</tr>
</tbody>
</table>

**7.8 Model estimation methods other than the maximum likelihood**

In much of this chapter, we obtain the covariance matrix $\Sigma$ of the reference distribution and SRMSR critical values based on copulas fitted to the so-called copula data, i.e., data simulated directly from the copula. However, real data sets seldom lie within the unit hypercube and model fitting typically involves manipulating univariate margins either before or together with the dependence modelling. In this section, we address some issues regarding the effect of different model fitting methods have on the critical values of the diagnostic statistic. Through simulation studies, we observe that: (a) critical values obtained from the maximum likelihood method on the copula are typically larger than when margins are estimated separately (inference function and marginal ranks methods in Sections 2.5.2 and 2.5.3), and (b) for truncated vine models, critical values based on the joint modelling of dependence parameters are similar to those based on the sequential estimation method. Therefore, for practical purposes, the critical values suggested in the preceding sections can be considered conservative (upper) bounds for other practical estimation methods. It remains a research problem as to whether the phenomenon in (a) can be proved for specific cases.

**7.8.1 Estimation of marginal distributions**

In Section 6.4 we commented that fitting a copula to data generated from it generally results in the largest variance for the difference statistic, compared to two common alternatives
including: (a) first fitting marginal parameters, then dependence parameters with marginal
parameters fixed at the fitted values (the inference function for margins or IFM method),
and (b) converting each margin to the $[0,1]$ interval using ranks, then fitting dependence
parameters with this transformed data set. We perform parametric bootstrap simulation
for some selected dependence structures, copulas and marginal distributions to explore how
the critical values may change under different model estimation methods. We only mention
the results for the Kendall’s $\tau$ feature as the patterns for Spearman’s $\rho$ are similar. For the
exchangeable and factor structures, three marginal distributions are considered:

1. Pareto distribution with parameters $(\alpha_k, \sigma_k)$ for margin $k$ and density function

$$g_k(y) = \frac{\alpha_k}{\sigma_k} \left(1 + \frac{y}{\sigma_k}\right)^{-(\alpha_k+1)}, \quad y > 0.$$ 

2. t distribution with parameters $(\mu_k, \sigma_k, \nu_k)$ for margin $k$ and density function

$$g_k(y) = \frac{1}{\sigma_k} h_{\nu_k} \left(\frac{y - \mu_k}{\sigma_k}\right), \quad y \in \mathbb{R},$$

where $h_{\eta}(x)$ is the density of the standard t distribution with $\eta$ degrees of freedom
evaluated at $x$.

3. Weibull distribution with parameters $(\alpha_k, \beta_k)$ for margin $k$ and density function

$$g_k(y) = \frac{\alpha_k}{\beta_k^\alpha} y^{\alpha_k-1} \exp \left\{ - \left(\frac{y}{\beta_k}\right)^{\alpha_k} \right\}, \quad y > 0.$$ 

These scenarios cover various degrees of tail heaviness and symmetry. Various parameter
values are chosen for different margins to mimic the situation where each margin may have
a different measurement unit or scale. Note that the choice of margins is irrelevant in
the marginal ranks method as ranks are preserved under monotonic transformations. We
revisit the scenarios used in Sections 7.3 and 7.5; for exchangeable dependence structure,
the simulation is performed on the Frank and MTCJ families, two with vastly different
tail behaviours, with Kendall’s $\tau$ controlled at 0.5 among variables. For 1-factor model,
we choose the one with Gumbel linking copulas and weak dependence, that is scenario I
in Table 7.4. We note that the behaviour for other dependence strengths is similar. For
2-factor model, we use all four scenarios in the same table with Gumbel/Frank linking
copulas. As for truncated vines, there is one more layer of tuning in terms of structures and
for brevity we consider only Pareto marginal distributions with $d = 8$. We only consider
Gaussian vines as parametric bootstrap is efficient in this case.
Table 7.11: SRMSR critical values (for sample size 100) for some exchangeable, factor and truncated vine structures based on different estimation methods with Kendall’s \( \tau \) as the feature. The columns labelled “Cop” give the critical values for copula estimation assuming known margins, “P”/“T”/“W” give the critical values based on IFM estimation with Pareto/t/Weibull margins, and “Rank” give the critical values based on marginal ranks estimation. Roman numerals indicate scenarios used in Tables 7.4 and 7.5.

The results of this simulation study are given in Table 7.11. For all the cases considered, the critical values using only copula estimation are larger than both the IFM and marginal ranks methods, with the marginal ranks critical values the smallest in most cases. This observation is similar to that in Section 6.4 and provides some evidence towards using the critical values suggested in the preceding sections for a conservative bound. Given the tendency of these critical values being smaller when margins are estimated, exceedance of the conservative bound serves as a strong signal for the need of model improvement.

Next, we briefly comment on the behaviour under joint maximum likelihood estimation of marginal and dependence parameters. As mentioned in Section 6.4, this is only practically relevant in low dimensions. The method in Section 7.4 is applicable for features being U-statistics due to the separability property under maximum likelihood estimation, or parametric bootstrap can be used if fitting is fast. Since joint estimation does not scale well to high dimensions and complicated structures, we only perform some simulations on the exchangeable structure with Pareto margins using sample size 2,000 and 500 replications.
The results in Table 7.12 show that, unlike in the bivariate case, critical values obtained using joint maximum likelihood estimation are not necessarily larger than those based on known margins.

<table>
<thead>
<tr>
<th>Copula family</th>
<th>$d = 3$</th>
<th>$d = 7$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cop</td>
<td>Joint</td>
</tr>
<tr>
<td>Frank</td>
<td>.058</td>
<td>.055</td>
</tr>
<tr>
<td>MTCJ</td>
<td>.081</td>
<td>.077</td>
</tr>
</tbody>
</table>

Table 7.12: SRMSR critical values (for sample size 100) for the exchangeable structure with Frank and MTCJ copulas with Kendall’s $\tau$ value of 0.5, based on copula estimation (“Cop”) and joint estimation of marginal and dependence parameters (“Joint”) with Pareto margins.

For extreme value copulas, we considered pairwise likelihood estimation of the copula in Section 7.6. Methods for marginal modelling include the IFM with GEV margins and the marginal ranks method. We repeat the simulation scenarios for the Hüsler-Reiss copula in Table 7.9, but using these two alternative estimation methods instead. Various parameter values are chosen for the GEV margins in the IFM method. The resulting critical values based on Kendall’s $\tau$, Spearman’s $\rho$ and extremal coefficient are shown in Table 7.13; we observe a similar pattern as in the non-extreme case, where the copula estimation assuming known margin yields the largest critical values. The IFM and marginal ranks methods have similar but smaller critical values. Note that the (assumed) asymptotic normality of the difference statistic is invoked in the estimation of these critical values.

<table>
<thead>
<tr>
<th>$d = 5$ Method</th>
<th>Kendall’s $\tau$</th>
<th>Spearman’s $\rho$</th>
<th>Extremal coefficient $\vartheta$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>W</td>
<td>M</td>
<td>S</td>
</tr>
<tr>
<td>Copula</td>
<td>0.055</td>
<td>0.060</td>
<td>0.047</td>
</tr>
<tr>
<td>IFM</td>
<td>0.045</td>
<td>0.039</td>
<td>0.028</td>
</tr>
<tr>
<td>Rank</td>
<td>0.048</td>
<td>0.041</td>
<td>0.029</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$d = 10$ Method</th>
<th>Kendall’s $\tau$</th>
<th>Spearman’s $\rho$</th>
<th>Extremal coefficient $\vartheta$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>W</td>
<td>M</td>
<td>S</td>
</tr>
<tr>
<td>Copula</td>
<td>0.053</td>
<td>0.057</td>
<td>0.044</td>
</tr>
<tr>
<td>IFM</td>
<td>0.043</td>
<td>0.036</td>
<td>0.030</td>
</tr>
<tr>
<td>Rank</td>
<td>0.045</td>
<td>0.039</td>
<td>0.029</td>
</tr>
</tbody>
</table>

Table 7.13: SRMSR critical values (for sample size 100) for the Hüsler-Reiss copula, using copula estimation, IFM with GEV marginal distributions and the marginal ranks methods. The average pairwise Kendall’s $\tau$ and extremal coefficients for the weak (W), moderate (M) and strong (S) dependence scenarios are in Tables 7.7 and 7.8.
7.8.2 Sequential estimation for vine structures

For truncated vine structures, sequential estimation is computationally more efficient than maximum likelihood. In sequential estimation, copulas in tree 1 are first fitted with all higher-order linking copulas ignored. The estimated parameters are then used as fixed inputs to the fitting of tree 2 copulas, and this process continues until tree \( p \), where \( p \leq d \) is the truncation level. Similar to the IFM method, this amounts to solving a set of estimating equations. The separability property of asymptotic variances does not hold and thus evaluation of \( \Sigma \) using the method in Section 7.4 is not applicable. A small-scale simulation is conducted here to gain insight on the effect of critical values compared to maximum likelihood estimation of vine parameters. We choose 2-truncated C-vines with all linking copulas being Frank or Gumbel, and one scenario with Gumbel on the first tree and Frank on the second. The C-vine structure is chosen because it has numerically tractable margins and allows parametric bootstrap. The results are shown in Table 7.14; they seem to suggest that there is little difference between the critical values obtained using these two methods, except for the Frank case where the “before” critical value may be somewhat smaller under the sequential method.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Copula</th>
<th>( d = 5 ) Before</th>
<th>After</th>
<th>( d = 8 ) Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Cop</td>
<td>Seq</td>
<td>Cop</td>
<td>Seq</td>
</tr>
<tr>
<td>Weak</td>
<td>Frank</td>
<td>.032</td>
<td>.030</td>
<td>.060</td>
<td>.056</td>
</tr>
<tr>
<td>(5d: I; Gumbel)</td>
<td>.060</td>
<td>.058</td>
<td>.093</td>
<td>.096</td>
<td>.055</td>
</tr>
<tr>
<td>8d: III) Gumbel/Frank</td>
<td>.050</td>
<td>.046</td>
<td>.096</td>
<td>.095</td>
<td>.050</td>
</tr>
<tr>
<td>Strong</td>
<td>Frank</td>
<td>.033</td>
<td>.026</td>
<td>.089</td>
<td>.088</td>
</tr>
<tr>
<td>(5d: II; Gumbel)</td>
<td>.054</td>
<td>.054</td>
<td>.077</td>
<td>.078</td>
<td>.056</td>
</tr>
<tr>
<td>8d: IV) Gumbel/Frank</td>
<td>.057</td>
<td>.053</td>
<td>.079</td>
<td>.076</td>
<td>.052</td>
</tr>
</tbody>
</table>

Table 7.14: SRMSR critical values (for sample size 100) for some 2-truncated C-vines with the feature being Kendall’s \( \tau \). The columns labelled “Cop” give the critical values for copula estimation assuming known margins, and “Seq” give the critical values based on sequential estimation. Roman numerals indicate scenarios used in Table 7.5.

It should be noted that the restricted maximum likelihood method used in Section 7.5.3 to speed up the fitting of Gaussian vines results in the same estimates for the first tree as sequential estimation because the objective functions for both methods consist of the same copula densities\(^{24}\). The estimates for other trees are in general different. However, in practice tree 1 usually contains the strongest correlations (and guides the placement of edges in the associated graphical model), and thus the fitted pairwise correlations based on

\(^{24}\)By fixing the diagonal elements of the covariance matrix to 1 in the restricted maximum likelihood method, we are in effect fitting a Gaussian copula.
the restricted maximum likelihood method are close to those based on sequential estimation. Therefore, the critical values based on these two estimation methods are typically similar. We rerun several scenarios considered in Section 7.5.3 using sequential estimation and in each case the difference of the SRMSR critical values (for sample size 100) is less than 0.001. Based on these observations, we believe that the critical values for vines resulting from copula estimation, or in the Gaussian case the restricted maximum likelihood method, are representative of those resulting from sequential estimation.

7.9 Data examples

We have covered the use of the adequacy-of-fit statistic for model diagnostics under a wide variety of situations, including different types of dependence structures (exchangeable, factor and truncated vines), nature of the data (directly observed continuous observations and multivariate extrema), features (Kendall’s $\tau$, Spearman’s $\rho$, extremal coefficient and tail-weighted dependence measures), estimation methods (known margins, IFM and rank-based) and dimensions. Here we implement the various strategies discussed on some data examples. The first example is on (non-extremal) European market returns; it illustrates the use of diagnostic statistics for general factor and vine copula models. The second and third examples correspond to the two multivariate extremes data sets in Section 3.7.

7.9.1 European market returns data

In this example, we consider the returns for $d = 7$ European market indices during the recession years 2008–2009. The market indices included in the modelling include: (1) OSE (Norway); (2) FTSE (UK); (3) AEX (Netherlands); (4) CAC (France); (5) SMI (Switzerland); (6) DAX (Germany), and; (7) ATX (Austria), for which daily returns of length $n = 484$ are available. This data set has been previous studied in Joe (2014) as an illustration of the modelling methods for various parsimonious dependence structures. In their analysis, an AR(1)–GARCH(1,1) model with standardized t-distributed innovations is first fitted to each univariate series to model serial dependence, and the resulting GARCH-filtered data are used for dependence modelling. A normal scores correlations and semi-correlations matrix, given in Table 7.20 of Joe (2014), suggests that the overall pairwise dependence among variables is quite strong, and that the joint lower tail is slightly heavier than the joint upper tail for most pairs; a pairwise scatterplot of the normal scores is given in Figure 7.13. Various factor and truncated vine models with different linking copulas are then fitted; the results of these fits are presented in Table 7.22 of Joe (2014). A further check using tail-weighted dependence measures is conducted to inspect whether the modelled tails
match the observed ones, and an informal comparison between empirical and model-based Kendall’s $\tau$ and tail-weighted dependence measures is made.

Figure 7.13: Scatterplot of the normal scores for the European market GARCH-filtered index returns data

To illustrate the use of the adequacy-of-fit statistic, we focus on the fitted models listed in Table 7.15. In this example, the factor and truncated vine structures have comparable performance; the difference lies in the choice of copula (light-tailed Frank versus heavy-tailed t), where models with $t$ linking copulas fare better in terms of both AIC and BIC values. Joe (2014) suggests that the 2-factor and 2-truncated vine models may be adequate as the improvement from 2-truncated to 3-truncated vines is much less than that from
1-truncated to 2-truncated, or from 1-factor to 2-factor copulas.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Linking copula</th>
<th># param.</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-factor</td>
<td>Frank</td>
<td>7</td>
<td>−4714</td>
<td>−4685</td>
</tr>
<tr>
<td>1-factor</td>
<td>$t_{15}$</td>
<td>8</td>
<td>−5085</td>
<td>−5051</td>
</tr>
<tr>
<td>2-factor</td>
<td>Frank</td>
<td>14</td>
<td>−5011</td>
<td>−4952</td>
</tr>
<tr>
<td>2-factor</td>
<td>$t_{15}/t_{10}$</td>
<td>16</td>
<td>−5171</td>
<td>−5104</td>
</tr>
<tr>
<td>1-truncated</td>
<td>Frank</td>
<td>6</td>
<td>−4610</td>
<td>−4585</td>
</tr>
<tr>
<td>1-truncated</td>
<td>$t_{10}$</td>
<td>7</td>
<td>−5041</td>
<td>−5012</td>
</tr>
<tr>
<td>2-truncated</td>
<td>BB1/Frank</td>
<td>17</td>
<td>−5041</td>
<td>−4970</td>
</tr>
<tr>
<td>2-truncated</td>
<td>$t_{10}/t_{10}$</td>
<td>13</td>
<td>−5188</td>
<td>−5133</td>
</tr>
<tr>
<td>3-truncated</td>
<td>Frank</td>
<td>15</td>
<td>−4883</td>
<td>−4820</td>
</tr>
<tr>
<td>3-truncated</td>
<td>$t_{15}/t_{10}/t_{15}$</td>
<td>18</td>
<td>−5210</td>
<td>−5135</td>
</tr>
</tbody>
</table>

Table 7.15: Fitted models considered for the European market GARCH-filtered index returns data

For each of the 10 fitted models listed in Table 7.15, we obtain the SRMSR statistic for the Kendall’s $\tau$ feature and compare it with the critical value for sample size $n = 484$. The results are shown in Table 7.16. For the factor models, the parametric bootstrap is used to estimate the critical value of the reference distribution. For the truncated vine models, we evaluate the elements of $\Sigma$ via Monte Carlo simulation described in Section 7.4. The 1-factor model with $t_{15}$ linking copulas is the only factor model for which the SRMSR statistic is larger than the critical value, but the difference is small and does not suggest substantial departure. The 2-factor model with $t_{15}/t_{10}$ linking copulas has a much smaller SRMSR statistic than the critical value. It appears that all the listed factor models can be considered adequate, or nearly so in the case of the 1-factor model with $t_{15}$ linking copulas.

Meanwhile, both 1-truncated vine copulas have “before” statistics smaller than the corresponding critical values, but their “after” statistics are much larger. This signals that the fitting of 1-truncated vines heavily favours matching the strength of dependence between the variables linked in the first tree, and the match for other pairs is substantially worse. For this reason, 1-truncated vine models are likely inadequate for the data. The 2-truncated vine with BB1/Frank linking copulas is probably not a good choice given that both the “before” and “after” statistics are substantially larger than the critical values, although as mentioned below the tail fit appears adequate. On the other hand, the 2-truncated vine with $t_{10}$ linking copulas has a “before” statistic smaller than the critical value, but an “after” statistic slightly larger; it can be considered as a marginal case. For the 3-truncated vine with Frank copula, both “before” and “after” statistics are larger than the critical values and hence it is not a desirable model. The 3-truncated vine with $t_{15}/t_{10}/t_{15}$ copulas can be deemed adequate, but it has the largest number of parameters (18) among those considered.
<table>
<thead>
<tr>
<th>Structure</th>
<th>Linking copula</th>
<th>Before SRMSR</th>
<th>Before CV</th>
<th>After SRMSR</th>
<th>After CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-factor</td>
<td>Frank</td>
<td>.021</td>
<td>.021</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-factor</td>
<td>t₁₅</td>
<td>.026</td>
<td>.023</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-factor</td>
<td>Frank</td>
<td>.015</td>
<td>.017</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-factor</td>
<td>t₁₅/t₁₀</td>
<td>.011</td>
<td>.022</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7.16: European market returns data: SRMSR statistics and critical values (CV) for the fitted models with the feature being Kendall’s τ. When the value of the statistic is greater than the critical value, both numbers are boldfaced.

A second adequacy-of-fit check is performed on the tails of the copulas using tail-weighted dependence measures\(^{25}\); the results are given in Table 7.17. Following the observations in Section 7.7, we obtain the critical values by considering only the empirical asymptotic covariance matrix. Here it is evident that none of the models with only Frank linking copulas is adequate. On the other hand, all models with only t linking copulas, and the 2-truncated vine with BB1/Frank linking copulas, can be considered adequate; we do not find substantial evidence against their adequacy in either tail.

To summarize the results for both Kendall’s τ and tail-weighted dependence measures, we find that the 2-factor and 3-truncated vine models with t linking copulas are adequate, while the 1-factor and 2-truncated vine models with t linking copulas are marginally adequate. In this case, the diagnostic check points to similar results with the choice based on AIC and BIC, except for the 2-truncated vine with BB1/Frank linking copulas where the clear inadequacy of the match of pairwise Kendall’s τ is not very well reflected by the latter criteria (note the 2-truncated BB1/Frank combination has even smaller AIC/BIC values than the 3-truncated vine with all Frank linking copulas). The exact choice among these depends on how one wants to interpret the fitted model and the degree of parsimony desired. All models with only Frank linking copulas, as well as the 2-truncated vine with BB1/Frank linking copulas, are inadequate even though some of them have decent fits for the central strength of dependence. This example highlights the importance of considering

\(^{25}\)With default tuning function and parameter as stated in Section 7.7.
<table>
<thead>
<tr>
<th>Structure</th>
<th>Linking copula</th>
<th>Lower tail SRMSR</th>
<th>CV</th>
<th>Upper tail SRMSR</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-factor</td>
<td>Frank</td>
<td>.301</td>
<td>.101</td>
<td>.236</td>
<td>.100</td>
</tr>
<tr>
<td>1-factor</td>
<td>t_15</td>
<td>.046</td>
<td>.091</td>
<td>.074</td>
<td>.093</td>
</tr>
<tr>
<td>2-factor</td>
<td>Frank</td>
<td>.188</td>
<td>.098</td>
<td>.124</td>
<td>.099</td>
</tr>
<tr>
<td>2-factor</td>
<td>t_15/t_10</td>
<td>.045</td>
<td>.092</td>
<td>.081</td>
<td>.092</td>
</tr>
</tbody>
</table>

Table 7.17: European market returns data: SRMSR statistics and critical values (CV) for the fitted models with the features being tail-weighted dependence measures. When the value of the statistic is greater than the critical value, both numbers are boldfaced. Due to sampling variability, the critical values for the lower and upper tail-weighted dependence measures for the tail symmetric models could be slightly different.

both central and tail dependence: If we had only considered Kendall’s $\tau$, we would have concluded that the factor models with Frank linking copulas are adequate when in fact they match the tails poorly, resulting in inaccurate inferences concerning the tails such as the value-at-risk of a portfolio.

Next, we can compare the critical values obtained above with the conservative bounds suggested in the preceding sections. Matrices of pairwise empirical Kendall’s $\tau$ and tail-weighted dependence measures for the data are given in Table 7.24 of Joe (2014), yielding an average Kendall’s $\tau$ of 0.649, and average lower/upper tail-weighted dependence measures of 0.693 and 0.625, respectively. Critical values are obtained via linear interpolation of the values in Tables 7.6 and 7.10, and are given in Table 7.18. Note that a sample size multiplying factor of $\sqrt{\frac{100}{484}} = 0.455$ has been applied when converting from the tabulated values that use $n = 100$.

For factor copulas, these numbers match quite closely with those obtained via parametric bootstrap presented above. There is more variability for truncated vines; with Kendall’s $\tau$ measure, critical values for truncated vines with Frank linking copulas are generally smaller than the values suggested in Tables 7.6 and 7.10. Even without applying a multiplying factor, the critical values for truncated vines with t linking copulas are slightly smaller as

Table 7.18: European market returns data: SRMSR critical values with Kendall’s $\tau$ and tail-weighted dependence measures (TWDM) for sample size $n = 484$ obtained from Tables 7.6 and 7.10. Copula-specific multiplying factors have not been applied to these numbers.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Factor</th>
<th>Vine Before</th>
<th>Vine After</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kendall’s $\tau$</td>
<td>.022</td>
<td>.022</td>
<td>.031</td>
</tr>
<tr>
<td>Lower TWDM</td>
<td>.095</td>
<td>.077</td>
<td>.114</td>
</tr>
<tr>
<td>Upper TWDM</td>
<td>.109</td>
<td>.091</td>
<td>.124</td>
</tr>
</tbody>
</table>

well. With the tail-weighted dependence measures, there seems to be some underestimation of the critical value for the lower tail “before” statistic for Frank copulas (that for t copulas appears to be satisfactory when a multiplying factor of 1.1 is applied), while the others appear to be conservative enough. This discrepancy may be due to the mismatch between the empirical and model-based tail behaviours; this mismatch is bigger when the light-tailed Frank copula is fitted to the data which exhibit tail dependence.

Note that the original analysis in Joe (2014) uses the uniform scores of the GARCH-filtered returns for the estimation of dependence parameters. Based on the comments in Section 7.8, we believe the above critical values are conservative bounds so that exceedance is a strong signal for the need of improvement.

7.9.2 Fraser River flows data

For the Fraser River flows data, extreme value modelling was carried out in Section 3.7. In the following illustration on adequacy-of-fit, all models are considered except for the Hüsler-Reiss truncated vine model, where we choose those with a D-vine structure according to the relative locations of the gauging stations for better interpretability.

In addition to these models, we fit two max-stable processes to the data. Max-stable processes are spatial, infinite-dimensional generalizations of multivariate extremes with a particular structure imposed, such that the strength of dependence decreases with the distance between observation sites. The max-stable process due to Smith (1990) is a special case of the Brown-Resnick model (Brown and Resnick (1977)) and they have finite-dimensional distributions being Hüsler-Reiss. The Brown-Resnick model offers more flexibility in the parametrization of the covariance function which controls spatial dependence patterns (see, e.g., Huser and Davison (2013)). Schlather (2002) proposes another class of max-stable processes with finite-dimensional distributions being t-EV with 1 degree of freedom. Without covariates, the Brown-Resnick model has 2 parameters (range and smoothness). For the Schlather model, we choose the powered exponential covariance function, which has 3
parameters (nugget, range and smoothness) using the implementation in the \texttt{SpatialExtremes} package in \texttt{R}. It is common to use geographical coordinates, i.e., latitude, longitude and altitude (if applicable), and the Euclidean distance\textsuperscript{26} as a measure of separation between stations. However, as the variable concerned in this case is the discharge volume of water, we use the distance along river as the measure of spatial separation.

We compute the SRMSR statistic for each fitted model, with features being Kendall’s $\tau$ and the extremal coefficient. A parametric bootstrap is then conducted to estimate critical values. Because there are missing values in the data set and the sample size is rather small, we simulate the exact missing pattern and data size in the bootstrap samples. The critical value is directly obtained as the 95\% quantile of the sampling distribution for the SRMSR statistic. The t-EV copula with 3 degrees of freedom is used as a surrogate for the Hüsler-Reiss copula by first matching pairwise extremal coefficients, and then imposing the appropriate structure on the t-EV correlation matrix. A multiplying factor of 1.2 is applied to the resulting SRMSR critical values.

<table>
<thead>
<tr>
<th>Structure</th>
<th># param.</th>
<th>Kendall’s $\tau$ SRMSR CV</th>
<th>Extremal coefficient SRMSR CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>EV 1-factor (Burr)</td>
<td>8</td>
<td>0.095 0.079</td>
<td>0.086 0.086</td>
</tr>
<tr>
<td>EV 1-factor (Dagum)</td>
<td>8</td>
<td>0.095 0.081</td>
<td>0.086 0.088</td>
</tr>
<tr>
<td>HR 1-factor</td>
<td>9</td>
<td>0.080 0.100</td>
<td>0.066 0.104</td>
</tr>
<tr>
<td>HR 2-factor</td>
<td>17</td>
<td>0.059 0.094</td>
<td>0.048 0.096</td>
</tr>
<tr>
<td>Max-stable (Schlather)</td>
<td>3</td>
<td>0.135 0.098</td>
<td>0.111 0.097</td>
</tr>
<tr>
<td>Max-stable (Brown-Resnick)</td>
<td>2</td>
<td>0.121 0.088</td>
<td>0.111 0.090</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Structure</th>
<th># param.</th>
<th>Overall Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>SRMSR CV</td>
<td>SRMSR CV</td>
</tr>
<tr>
<td>HR 1-truncated D-vine</td>
<td>8</td>
<td>0.075 0.104</td>
<td>0.062 0.079</td>
</tr>
<tr>
<td>HR 2-truncated D-vine</td>
<td>14</td>
<td>0.062 0.111</td>
<td>0.063 0.079</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Structure</th>
<th># param.</th>
<th>Overall Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>SRMSR CV</td>
<td>SRMSR CV</td>
</tr>
<tr>
<td>HR 1-truncated D-vine</td>
<td>8</td>
<td>0.067 0.103</td>
<td>0.044 0.072</td>
</tr>
<tr>
<td>HR 2-truncated D-vine</td>
<td>14</td>
<td>0.052 0.114</td>
<td>0.050 0.075</td>
</tr>
</tbody>
</table>

Table 7.19: Fraser River flows data: SRMSR statistics and critical values (CV) for some fitted models with the features being Kendall’s $\tau$ and extremal coefficient. When the value of the statistic is greater than the critical value, both numbers are boldfaced.

\textsuperscript{26}Provided that the stations are not too far apart, or otherwise other choices such as the great circle distance may be a more relevant measure.
The comparisons are given in Table 7.19. The SRMSR statistics based on Kendall’s $\tau$ for both extreme value 1-factor models are larger than their respective critical values, while those based on extremal coefficient are close. It is worth noting that the Burr model has the second smallest CLBIC value among the models fitted in Table 3.8. On the other hand, all structured Hüsler-Reiss models are deemed adequate, even for the 1-factor model which has the largest CLIC and CLBIC values. The SRMSR statistic for the “before” pairs is very close to or larger than that for the “after” pairs, for the Hüsler-Reiss model with 2-truncated vine structure. Because of such behaviour, we include also the statistic computed from all pairs (labelled as “overall”). The max-stable models stand out as clearly inadequate here, due to their lack of parameters and flexibility in terms of the structure. Considering both parsimony and adequacy-of-fit, the Hüsler-Reiss copula with 1-truncated D-vine structure is recommended for further inferences. The Hüsler-Reiss copula with 2-factor or 2-truncated vine structure are also acceptable, but these are less parsimonious models.

Finally, we perform a parametric bootstrap that takes into account univariate margins for the 1-truncated D-vine Hüsler-Reiss model, using the fitted marginal GEV distribution parameters and the IFM method for estimation, and a t-EV surrogate. The critical values (with multiplying factor of 1.2) based on Kendall’s $\tau$ are 0.086, 0.073 and 0.094 for the overall, before and after statistics, respectively, while those based on the extremal coefficient are 0.080, 0.061 and 0.089. The 1-truncated D-vine Hüsler-Reiss model can still be considered adequate using this set of stricter critical values.

### 7.9.3 United States stock returns data

A similar comparison is made for the extreme (minimum) returns of 7 stocks traded on the US stock exchanges, whose fitting results are in Table 3.13. The Burr 1-factor extreme value copula and the t-EV 1-factor copula are the better models according to the CLIC and CLBIC. The SRMSR statistics and the associated critical values are tabulated in Table 7.20. Quite a number of models are flagged as inadequate here; the structured Hüsler-Reiss copulas appear inferior to the other classes of models. Based on this result, both the Burr 1-factor extreme value copula and the t-EV 1-factor copula are suitable models for further inferences or analysis.

### 7.10 Summary

In this chapter, we develop strategies for assessing the adequacy-of-fit of multivariate models with parsimonious dependence structures. In particular, we make use of a high quantile (critical value) of the limiting or reference distribution of the quadratic form statistic based
<table>
<thead>
<tr>
<th>Structure</th>
<th># param.</th>
<th>Kendall’s τ</th>
<th>Extremal coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>SRMSR</td>
<td>CV</td>
</tr>
<tr>
<td>EV 1-factor (Burr)</td>
<td>7</td>
<td>.059</td>
<td>.060</td>
</tr>
<tr>
<td>EV 1-factor (Dagum)</td>
<td>7</td>
<td>.062</td>
<td>.061</td>
</tr>
<tr>
<td>HR 1-factor</td>
<td>8</td>
<td>.077</td>
<td>.071</td>
</tr>
<tr>
<td>HR 2-factor</td>
<td>15</td>
<td>.076</td>
<td>.063</td>
</tr>
<tr>
<td>t-EV 1-factor</td>
<td>7</td>
<td>.044</td>
<td>.060</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Structure</th>
<th></th>
<th>Overall</th>
<th>Kendall’s τ</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>SRMSR</td>
<td>CV</td>
<td>SRMSR</td>
</tr>
<tr>
<td>HR 1-truncated vine</td>
<td>7</td>
<td>.077</td>
<td>.075</td>
<td>.063</td>
</tr>
<tr>
<td>HR 2-truncated vine</td>
<td>12</td>
<td>.074</td>
<td>.067</td>
<td>.065</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Structure</th>
<th></th>
<th>Overall</th>
<th>Extremal coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>SRMSR</td>
<td>CV</td>
</tr>
<tr>
<td>HR 1-truncated vine</td>
<td>7</td>
<td>.093</td>
<td>.089</td>
</tr>
<tr>
<td>HR 2-truncated vine</td>
<td>12</td>
<td>.090</td>
<td>.081</td>
</tr>
</tbody>
</table>

Table 7.20: US stock returns data: SRMSR statistics and critical values (CV) for some fitted models with the features being Kendall’s τ and extremal coefficient. When the value of the statistic is greater than the critical value, both numbers are boldfaced.

on the vector of differences between empirical and model-based features. These features are applied to each of the \( \binom{d}{2} \) bivariate margins of the \( d \)-dimensional distribution; Kendall’s τ and Spearman’s ρ are mainly used as measures for central dependence while the tail-weighted dependence measures are more suitable to assess the match in the tails. The extremal coefficient is used for extreme value copula models. For truncated vine models, we further separate the treatment of pairs directly modelled by the vine (“before” pairs) and those that include higher-order independence copulas (“after” pairs). Critical values based on the reference distributions are converted to those of the SRMSR (for the desired sample size) for easier interpretation and comparison with the practical significance needed for the problem at hand. We consider a model adequate if the value of the SRMSR statistic is smaller than its critical value, and inadequate otherwise. Inadequate models can then be improved with an emphasis on the variables that are poorly fitted; information on this can be conveniently obtained from the SRMSR statistic due to its feature-based nature.

The critical values are not always easy to obtain, and therefore we propose different methods depending on the complexity in model simulation, fitting and obtaining model-based features. When all of these are computationally manageable, the parametric boot-
strap can be used and this bypasses the calculation of the covariance matrix of the reference distribution. Otherwise, we need to consider feasible alternatives. A summary is provided below, grouped by the dependence measure used.

1. For Kendall’s $\tau$ and Spearman’s $\rho$ of non-extreme-value copulas:

   a) The most general approach is to use a surrogate Gaussian copula to obtain the asymptotic covariance matrix $\Sigma$ and approximate the critical value, via a gamma approximation to the distribution of the limiting distribution (i.e., $Q$). A multiplying factor can be applied depending on the tail properties of the linking copulas of the target model. We observe that the critical values depend mostly on the average pairwise dependence strength, the structure and copula family being considered, but less so on the dimension, especially when it is large. The products from this research include tabulated conservative bounds for Gaussian factor and vine models; these can be interpolated for other average dependence strengths, and also the faster algorithms that are designed for structured Gaussian copulas of higher dimensions (e.g., more than 10).

   b) The parametric bootstrap is a feasible method for simple dependence structures for which simulation/fitting of the model and the computation of model-based dependence measures are manageable. Examples include exchangeable copulas, low-dimensional factor and truncated C-vine copulas.

   c) Under maximum likelihood estimation, we can make use of the separability property of the asymptotic covariance matrix of the scaled differences, and evaluate the asymptotic covariance of the empirical and model-based portions separately using Monte Carlo simulation. We intend to provide a program for such evaluation when model simulation and fitting scale efficiently to large sample sizes.

   d) For other estimation methods such as inference function for margins (IFM), marginal ranks or sequential estimation for truncated vines, a limited simulation study is carried out in Section 7.8 using various copulas with exchangeable, 1- and 2-factor, as well as 1- and 2-truncated vine structures. We comment that the critical values obtained based on maximum likelihood estimation can often be used as a conservative bound for these cases.

2. For Kendall’s $\tau$, Spearman’s $\rho$ and extremal coefficient of extreme value copulas:

   a) Current research is limited to the use of the extremal coefficient for extreme value copulas, but we note from the development in Chapter 5 that it is possible to extend its use for general copulas as a tail-weighted dependence measure. For
structured extreme value copulas, the approach for all the three measures is to conduct parametric bootstrap on either the target model if it can be efficiently simulated from, or otherwise on a (possibly structured) surrogate model.

b) The simulation study in Section 7.6 suggests that the t-EV copula may be used as a surrogate for the Hülsler-Reiss copula; a multiplying factor of 1.2 on the SRMSR critical value appears sufficient, although the results are based on a limited set of scenarios as there are currently no efficient algorithms to simulate from a general multivariate Hülsler-Reiss copula. We remark that the critical values based on Gaussian or t surrogate copulas match poorly to those based on the Hülsler-Reiss copula, and are not recommended in this case.

c) For other estimation methods, a simulation study in Section 7.8 suggests that the critical values obtained based on copula estimation (i.e., assuming known margins) may be larger than those based on IFM or marginal ranks methods, and may be used as conservative bounds.

3. For tail-weighted dependence measures:

   a) The parametric bootstrap can be used to estimate the asymptotic covariance matrix of the vector of empirical measures. This is then converted to a conservative critical value via gamma approximation of the limiting distribution. This method is practical as long as model simulation is feasible.

   b) The simulation study in Section 7.7 suggests that the empirical tail-weighted dependence measure is generally much more variable than its model-based counterpart. The asymptotic covariance is therefore dominated by the behaviour of the empirical measure; we suggest that fitting of the parametric bootstrap samples can be omitted altogether to ease the computational burden and avoid the evaluation of model-based tail-weighted dependence measures.

Due to the diverse possibilities of dependence measures, structures and model fitting procedures, it is an extremely challenging, if not impossible, task to devise a strategy that is applicable to every possible combination. The more pragmatic pathway is to exploit the properties relevant to the scenario faced, and develop methods targeted to them. This is the approach we have taken in this chapter, and we believe that some of these ideas and strategies can be applied to parsimonious dependence structures not considered here as well.
Chapter 8

Conclusion

Statistical models with parsimonious dependence structures are practical in reducing computational complexity and offering plausible interpretation of the data. We demonstrate how factor or truncated vine structures may be incorporated into the setting of multivariate extremes. The extreme value limit of a general multivariate copula may not be numerically tractable, and therefore structuring the underlying correlation matrix of an existing multivariate extreme copula can be a useful alternative approach. Statistical inference is possible using composite likelihood methods as these models are closed under margins, allowing low-order (typically pairwise) densities to be computed readily. The data examples show that, with $O(d)$ parameters for $d$ variables, these parsimonious extreme value copulas offer intuitive interpretation and may outperform a fully saturated model with $O(d^2)$ parameters. Potential future work in this area includes:

- **The development of a more efficient algorithm in obtaining the bivariate density of the extreme value factor model.** As the technical details in Chapter 4 suggest, the derivations on the transformation of integrals for numerical considerations are not trivial and instabilities are possible in boundary cases. It is thus useful to explore the possibility of extreme value factor copulas for which the bivariate distribution has a simpler form. This is especially important in extending the numerical methods to the 2-factor case.

- **Investigation of the connection between general factor/truncated vine models and the class of extreme value copulas.** It is of interest to establish conditions of a general factor or truncated vine copula that guarantee max-stability. This may provide another pathway to the development of new classes of structured extreme value copulas that are easier to implement, as there are existing algorithms for the modelling of factor and truncated vine copulas.
• **Implementation of other parsimonious structures.** One such possibility is the combined factor-vine model proposed in Brechmann and Joe (2014). This may be a suitable model when a factor structure is plausible, but there is still substantial residual dependence conditional on the latent factors. A truncated vine structure may then be imposed to account for such dependence. This can be easily accommodated in the structured Häusler-Reiss or t-EV models.

Regarding bivariate monotone association for extreme value copulas, we review various empirical estimators in the literature. Rank-based estimators have received more attention in recent years because marginal distributions are seldom known in practice, and thus the asymptotic properties for rank-based estimators are relevant. An extension of the estimator based on F-madograms to general parametric copula families is proposed; we suggest that the estimator can be used as a measure of tail-weighted dependence strength, in a similar fashion as the tail-weighted dependence measures proposed in Krupskii and Joe (2015).

The dependence measure based on the F-madogram has the advantage that it is simple to compute numerically by integrating along the diagonal of a copula, in contrast to the measure by Krupskii and Joe (2015) where a conditional correlation has to be computed. It also satisfies many of the desirable properties for measures of bivariate monotone association. As suggested in Chapter 5, it is possible to modify this estimator to measure the degree of permutation symmetry of a copula. Another direction is to consider the high-dimensional extension of the estimator. For extreme value copulas, Schlather and Tawn (2003) investigate the conditions for extremal coefficients of various dimensions to be self-consistent, i.e., that there exists a valid extreme value copula with such set of extremal coefficients. It would be useful to analyze the F-madogram estimator in this regard, and possible modifications if these conditions are not satisfied.

When conducting statistical inference using models with parsimonious dependence structures, one should be vigilant against underfitting, i.e., using a model that is too parsimonious and that ignores potentially important trends in the data. The concept of adequacy-of-fit provides the theoretical basis for the choice of parsimonious models using the difference between empirical and model-based features. We use pairwise dependence measures as features to avoid sparsity issues with high-dimensional models. For copulas with an arbitrary dependence structure, it may be challenging to obtain the properties (and critical values) of a reference distribution that is used for the assessment of model adequacy; we propose various ways of determining appropriate critical values for several dependence structures. When the parametric bootstrap is impractical due to intractable bivariate distributions or the computational burden on model fitting, alternatives such as evaluating the asymptotic covariance of the reference distribution or approximating critical values using surrogate
models can be considered.

The development of these alternative methods is complicated by the fact that there is more than one way to fit a copula model, especially when marginal fitting has to be considered. The choice of empirical estimators may also influence the asymptotic behaviour of the difference statistic. Results on the moments of the reference distribution rest on the assumption that the difference statistic is asymptotically normal; we have proved normality for one specific situation (i.e., U-statistics with model fitted via estimating equations), but there are other combinations whose proofs remain an open problem. Based on simulation studies, we suggest that critical values obtained based on maximum likelihood estimation may be used as conservative bounds for those based on other estimation methods. In this regard, a future research direction is on computing more precise estimates of critical values in these situations, when the parametric bootstrap cannot be easily implemented. It is also worth looking into whether the critical values (or elements of the $\Sigma$ matrix for the reference distribution) are indeed theoretically larger for maximum likelihood in certain situations. Solving the complications associated with model simulation from the general Hüsler-Reiss copula is another possible topic.

Finally, the algorithm for efficient sequential sorting in Appendix E is useful when a bivariate distribution function is intractable and is estimated using the sample counterpart in Monte Carlo methods. It is our plan to extend the method to higher dimensions $d$, with emphasis on optimizing for speed and temporary storage requirements. Although the algorithm is asymptotically more efficient than the naive method, the practical gain diminishes as $d$ increases due to the increasing complexity of the algorithm. The search for simpler implementations is thus a necessity for higher-dimensional problems.


Frank, M. J. (1979). On the simultaneous associativity of \( f(x, y) \) and \( x + y - f(x, y) \). Aequationes Mathematicae, 19, 194–226.


Appendix A

Derivations of the asymptotic properties of the rank-based F-madogram measure of dependence

In this appendix, we derive expressions relevant to the asymptotic variance of the rank-based F-madogram measure of dependence, and show the relevant simplifications for the case with the independence copula.

For general copulas, the asymptotic variance of the rank-based estimator usually involves 2-dimensional integrals that must be evaluated numerically. In this case, with $X$ given in (5.20), we have

$$
\text{Var}(X) = \mathbb{E}(X^2)
= \frac{1}{4} \int_0^1 \int_0^1 \mathbb{E} \left[ G_C(u^{1/\alpha}, 1) G_C(v^{1/\alpha}, 1) \right] du dv + \frac{1}{4} \int_0^1 \int_0^1 \mathbb{E} \left[ G_C(1, u^{1/\alpha}) G_C(1, v^{1/\alpha}) \right] du dv
+ \int_0^1 \int_0^1 \mathbb{E} \left[ G_C(u^{1/\alpha}, u^{1/\alpha}) G_C(v^{1/\alpha}, v^{1/\alpha}) \right] du dv + \frac{1}{2} \int_0^1 \int_0^1 \mathbb{E} \left[ G_C(u^{1/\alpha}, 1) G_C(1, v^{1/\alpha}) \right] du dv
- \int_0^1 \int_0^1 \mathbb{E} \left[ G_C(u^{1/\alpha}, 1) G_C(v^{1/\alpha}, v^{1/\alpha}) \right] du dv - \int_0^1 \int_0^1 \mathbb{E} \left[ G_C(1, u^{1/\alpha}) G_C(v^{1/\alpha}, 1) \right] du dv,
$$

(A.1)

where $G_C$ is the limiting Gaussian process in (5.9), with covariance function that involves the Brownian bridge $B_C$ (see (5.10)):

$$
\mathbb{E} [G_C(u_1, u_2) G_C(u_3, u_4)]
= \mathbb{E} \{ [B_C(u_1, u_2) - B_C(u_1, 1) C_{2|1}(u_2|u_1) - B_C(1, u_2) C_{1|2}(u_1|u_2)]
$$

220
\[ \cdot [B_C(u_3, u_4) - B_C(u_3, 1)C_{2|1}(u_4|u_3) - B_C(1, u_4)C_{1|2}(u_3|u_4)]\]

\[ = \mathbb{E}[B_C(u_1, u_2)B_C(u_3, u_4)] - C_{2|1}(u_2|u_1)\mathbb{E}[B_C(u_1, 1)B_C(u_3, u_4)] \]

\[ - C_{1|2}(u_1|u_2)\mathbb{E}[B_C(1, u_2)B_C(u_3, u_4)] - C_{2|1}(u_4|u_3)\mathbb{E}[B_C(u_1, u_2)B_C(u_3, 1)]\]

\[ + C_{2|1}(u_2|u_1)C_{2|1}(u_4|u_3)\mathbb{E}[B_C(u_1, 1)B_C(u_3, 1)] + C_{1|2}(u_1|u_2)C_{2|1}(u_4|u_3)\mathbb{E}[B_C(1, u_2)B_C(u_3, 1)]\]

\[ - C_{1|2}(u_3|u_4)\mathbb{E}[B_C(u_1, u_2)B_C(1, u_4)] + C_{1|2}(u_3|u_4)C_{2|1}(u_2|u_1)\mathbb{E}[B_C(u_1, 1)B_C(1, u_4)]\]

\[ + C_{1|2}(u_1|u_2)C_{1|2}(u_3|u_4)\mathbb{E}[B_C(1, u_2)B_C(1, u_4)], \]

(A.2)

in which the expectations involving products of Brownian bridges are:

\[ \mathbb{E}[B_C(u_1, u_2)B_C(u_3, u_4)] = C(u_1 \land u_3, u_2 \land u_4) - C(u_1, u_2)C(u_3, u_4) \]

\[ \mathbb{E}[B_C(u_1, 1)B_C(u_3, u_4)] = C(u_1 \land u_3, u_4) - u_1C(u_3, u_4) \]

\[ \mathbb{E}[B_C(1, u_2)B_C(u_3, u_4)] = C(u_3, u_2 \land u_4) - u_2C(u_3, u_4) \]

\[ \mathbb{E}[B_C(u_1, u_2)B_C(u_3, 1)] = C(u_1 \land u_3, u_2) - u_3C(u_1, u_2) \]

\[ \mathbb{E}[B_C(u_1, 1)B_C(u_3, 1)] = u_1 \land u_3 - u_1u_3 \]

\[ \mathbb{E}[B_C(1, u_2)B_C(u_3, 1)] = C(u_3, u_2) - u_2u_3 \]

\[ \mathbb{E}[B_C(u_1, u_2)B_C(1, u_4)] = C(u_1, u_2 \land u_4) - u_4C(u_1, u_2) \]

\[ \mathbb{E}[B_C(u_1, 1)B_C(1, u_4)] = C(u_1, u_4) - u_1u_4 \]

\[ \mathbb{E}[B_C(1, u_2)B_C(1, u_4)] = u_2 \land u_4 - u_2u_4. \]

The integrals can be evaluated once the expressions of \( C, C_{1|2} \) and \( C_{2|1} \) are given. When \( C \) is the independence copula, i.e., \( C(u_1, u_2) = u_1u_2 \), we show the following two results:

1. \( \mathbb{E}[G_C(u_1, u_2)G_C(u_3, u_4)] = (u_1 \land u_3 - u_1u_3)(u_2 \land u_4 - u_2u_4), \) and;

2. \( \text{Var}(X) = \int_0^1 \int_0^1 \mathbb{E}[G_C(u^{1/\alpha}, u^{1/\alpha})G_C(v^{1/\alpha}, v^{1/\alpha})] \, du \, dv, \) where \( X \) is defined in (5.20).

The first result is immediate using (A.2):

\[ \mathbb{E}[G_C(u_1, u_2)G_C(u_3, u_4)] \]

\[ = [(u_1 \land u_3)(u_2 \land u_4) - u_1u_2u_3u_4] - u_2 [(u_1 \land u_3)u_4 - u_1u_3u_4] - u_1 [(u_2 \land u_4)u_3 - u_2u_3u_4] \]

\[ - u_4 [(u_1 \land u_3)u_2 - u_1u_2u_3] + u_2u_4 (u_1 \land u_3 - u_1u_3) + 0 \]

\[ - u_3 [(u_2 \land u_4)u_1 - u_1u_2u_4] + 0 + u_1u_3 (u_2 \land u_4 - u_2u_4) \]

\[ = (u_1 \land u_3 - u_1u_3)(u_2 \land u_4 - u_2u_4). \]
For the second result, note the following simplifications for the integrands of (A.1):

\[
\begin{align*}
\mathbb{E}\left[ G_C\left( u^{1/\alpha}, 1 \right) G_C\left( v^{1/\alpha}, 1 \right) \right] &= \mathbb{E}\left[ G_C\left( 1, u^{1/\alpha} \right) G_C\left( 1, v^{1/\alpha} \right) \right] \\
&= \left[ (u \wedge v)^{1/\alpha} - (uv)^{1/\alpha} \right] \cdot (1 - 1) = 0;
\end{align*}
\]

\[
\begin{align*}
\mathbb{E}\left[ G_C\left( u^{1/\alpha}, 1 \right) G_C\left( 1, v^{1/\alpha} \right) \right] &= \left( u^{1/\alpha} - u^{1/\alpha} \right) \left( v^{1/\alpha} - v^{1/\alpha} \right) = 0;
\end{align*}
\]

\[
\begin{align*}
\mathbb{E}\left[ G_C\left( u^{1/\alpha}, 1 \right) G_C\left( v^{1/\alpha}, v^{1/\alpha} \right) \right] &= \mathbb{E}\left[ G_C\left( 1, u^{1/\alpha} \right) G_C\left( v^{1/\alpha}, v^{1/\alpha} \right) \right] \\
&= \left[ (u \wedge v)^{1/\alpha} - (uv)^{1/\alpha} \right] \left( v^{1/\alpha} - v^{1/\alpha} \right) = 0.
\end{align*}
\]

Thus \( \text{Var}(X) = \int_0^1 \int_0^1 \mathbb{E}\left[ G_C\left( u^{1/\alpha}, u^{1/\alpha} \right) G_C\left( v^{1/\alpha}, v^{1/\alpha} \right) \right] \, du \, dv. \)
Appendix B

Behaviour of the asymptotic variances of the empirical Kendall’s $\tau$ and Spearman’s $\rho$

In this appendix, we make use of the expansions of a copula near the boundary to show that the empirical asymptotic variances of both Kendall’s $\tau$ and Spearman’s $\rho$ depend in part on the tail properties of the copula.

The definition of tail dependence has been given in Section 2.3. For the purpose of this illustration, we assume that $C$ is an absolutely continuous bivariate copula and the joint upper and lower tails are well behaved, in the sense that the lower and upper tail orders $\kappa_L$ and $\kappa_U$ for a bivariate copula $C$ can be defined as the quantities that satisfy

$$C(u, u) \sim u^{\kappa_L} l(u); \quad C(1-u, 1-u) \sim u^{\kappa_U} l^*(u) \quad \text{as } u \to 0^+,$$

where $l$ and $l^*$ are slowly varying functions at $0^+$ (Hua and Joe (2011)). The tail order must be at least 1. Tail dependence is said to be strong if $\kappa = 1$, in which case the corresponding tail dependence parameter can be larger than 0. A bivariate copula has intermediate tail dependence if $1 < \kappa < 2$ (or if $\kappa = 1$ but the slowly varying function for the corresponding tail approaches zero as $u \to 0^+$) and tail quadrant independence if $\kappa = 2$ and the slowly varying function is bounded. Given some regularity conditions, $c(u, u) \sim u^{\kappa_L-2} l^{**}(u)$ for some slowly varying $l^{**}$ and similarly for the upper tail order (see, e.g., Section 2.16 of Joe (2014)).

For the asymptotic variance of $\sqrt{n} \hat{\tau}_{\text{emp}}$, the scaled empirical estimator of Kendall’s $\tau$ with true value denoted as $\tau$, we have

$$\lim_{n \to \infty} n \text{Var}(\hat{\tau}_{\text{emp}}) = 16 \int_0^1 \int_0^1 \left[ C(u, v) + \overline{C}(u, v) \right]^2 dC(u, v) - 4(\tau + 1)^2 \quad \text{(B.1)}$$
from Example 6.2. For a given $\tau$, the asymptotic variance is affected by the magnitude of the integral in (B.1), which can be written as

$$
\int_0^1 \int_0^1 \left[ C(u, v) + \overline{C}(u, v) \right]^2 dC(u, v) = \int_0^1 \int_0^1 [1 - u - v + 2C(u, v)]^2 c(u, v) dudv.
$$

Focusing on what happens near the corner $(0,0)$ by letting $u = v$, note that

$$
[1 - 2u + 2C(u, u)]^2 c(u, u) = \left[ \underbrace{(1 - 2u)^2 + 4C^2(u, u) - 4(1 - 2u)C(u, u)}_{\to u^{\kappa L - 2}l_3(u)} \right] c(u, u),
$$

where the behaviour as $u \to 0$ for each term is as indicated with the $l$’s being some (assumed bounded) slowly varying functions. Therefore, the integrand behaves like $u^{\kappa L - 2}$ in the limit. This asymptotes to 1 under tail quadrant independence and infinity for intermediate or strong tail dependence, the rate of which increases with decreasing $\kappa$; this observation helps explain the higher asymptotic variances associated with stronger tail dependence of the copula. A similar argument using $\overline{C}$ and $\kappa_U$ applies to the joint upper tail.

For the asymptotic variance of $\sqrt{n}\hat{\rho}_{\text{S,emp}}$, the scaled empirical estimator of Spearman’s $\rho$ with true value denoted as $\rho_S$, we have

$$
\lim_{n \to \infty} n \text{Var}(\hat{\rho}_{S,\text{emp}}) = 144 \int_0^1 \int_0^1 \left[ \int_0^1 \overline{C}(u, w) dw + \int_0^1 \overline{C}(x, v) dx + uv \right]^2 dC(u, v) - 9(\rho_S + 3)^2
$$

from Example 6.2. Again, we are interested in the behaviour of the integrand when $u \to 0$ and $v \to 0$. Set $u = v$, in which case the integrand becomes

$$
\left[ \int_0^1 \overline{C}(u, w) dw + \int_0^1 \overline{C}(x, u) dx + u^2 \right]^2 c(u, u).
$$

As $u \to 0$, both of the integrals approach $\int_0^1 \overline{C}(0, w) dw = \int_0^1 (1 - w) dw = 1/2$, where the boundedness of both the integrand and the integration limits ensures the order of integration and taking limits can be exchanged. Therefore (B.2) behaves like $c(u, u) \sim u^{\kappa L - 2}l_3(u)$ when $u \to 0$. The aforementioned behaviour for Kendall’s $\tau$ thus applies to Spearman’s $\rho$ as well.
Appendix C

Maximum likelihood estimator for Gaussian vine structures when variances are estimated

In this appendix, we show that the maximum likelihood estimator of the correlation parameters coincides with the sample correlations for the pairs before and including the truncation level of truncated Gaussian vines. This is an extreme case of the partitioning suggested in Section 7.2, where the vector on the left hand side of (7.8) contains zeros for all “before” pairs and non-zero entries for the “after” pairs.

In the derivations below, we assume the mean vector of the Gaussian distribution is zero and the data are centred to have sample means zero as well. This takes away the mean component of the Gaussian distribution (whose estimation is separate from that of the covariance matrix) and simplifies notation. The data are however not standardized to have unit variance, as the variance components in the covariance matrix are also estimated.

We begin with a $d$-dimensional 1-truncated Gaussian C-vine, whose upper triangular portion of the covariance matrix is given by

\[
\Sigma = \begin{pmatrix}
\sigma_1^2 & \sigma_1\sigma_2\rho_{12} & \sigma_1\sigma_3\rho_{13} & \cdots & \sigma_1\sigma_d\rho_{1d} \\
\sigma_2^2 & \sigma_2\sigma_3\rho_{13} & \sigma_2\sigma_4\rho_{14} & \cdots & \sigma_2\sigma_d\rho_{1d} \\
\sigma_3^2 & \cdots & \sigma_3\sigma_d\rho_{1d} \\
\vdots & \ddots & \vdots \\
\sigma_d^2 & & & & 
\end{pmatrix},
\]

i.e., the $j$th diagonal is $\sigma_j^2$, the $(1,j)$ and $(j,1)$ entries are $\sigma_1\sigma_j\rho_{1j}$ for $j \neq 1$, the $(j,k)$ and $(k,j)$ entries are $\sigma_j\sigma_k\rho_{1j}\rho_{1k}$ for $1 \neq j < k$. The precision matrix $\Sigma^{-1} = (\varsigma_j^k)$ is such that
ς^{jk} = 0 unless either j or k is equal to 1, or j = k, when \( \rho_{12}, \ldots, \rho_{1d} \) are non-zero; this can be shown using a stochastic representation using d i.i.d. standard Gaussian random variables: Let \( X = (X_1, \ldots, X_d)^\top \sim N(0, \Sigma) \) and \( Z = (Z_1, \ldots, Z_d)^\top \sim N(0, I_d) \), where \( Z \) is independent of \( X \) and \( I_d \) is the identity matrix of dimension \( d \). For simplicity, assume \( \sigma_1 = \ldots = \sigma_d = 1 \) as the marginal variances do not affect the location of zeros in the precision matrix. Then we can represent the random vector \( X \) as

\[
X_1 = Z_1 \\
X_2 = \phi_2 X_1 + \psi_2 Z_2 \\
\vdots \\
X_d = \phi_d X_1 + \psi_d Z_d,
\]

where \( \phi_j = \rho_{1j} \) and \( \psi_j = \sqrt{1 - \phi_j^2} \), \( j = 2, \ldots, d \) (equation (6.10) of Joe (2014)). The density function of \( X \) is

\[
f_X(x) \propto \exp \left\{ -\frac{1}{2} x^\top \Sigma^{-1} x \right\} = \exp \left\{ -\frac{1}{2} \sum_{j=1}^d \sum_{k=1}^d \varsigma^{jk} x_j x_k \right\}, \tag{C.1}
\]

where \( x = (x_1, \ldots, x_d)^\top \). The density can also be decomposed into a product of conditional densities:

\[
f_X(x) = f_{X_1}(x_1) \cdot f_{X_2|X_1}(x_2|x_1) \cdots f_{X_d|X_1,\ldots,X_{d-1}}(x_d|x_1,\ldots,x_{d-1})
\]

\[
\propto \exp \left\{ -\frac{1}{2} x_1^2 \right\} \cdot \exp \left\{ -\frac{1}{2 \psi_2^2} (x_2 - \phi_2 x_1)^2 \right\} \cdots \exp \left\{ -\frac{1}{2 \psi_d^2} (x_d - \phi_d x_1)^2 \right\}. \tag{C.2}
\]

We can then locate the zeros of \( \Sigma^{-1} \) by comparing the terms in the exponent of (C.1) and (C.2); the terms with non-zero coefficients are \( x_1^2, \ldots, x_d^2, x_1 x_2, \ldots, x_1 x_d \).

Letting \( \theta \) be the collection of all parameters, the density of \( X \sim N(0, \Sigma) \) at \( x = (x_1, \ldots, x_d)^\top \) can be written as an exponential family representation:

\[
f_X(x; \theta) = (2\pi)^{-d/2} |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} x^\top \Sigma^{-1} x \right\}
\]

\[
= (2\pi)^{-d/2} \exp \left\{ -\frac{1}{2} \sum_{j=1}^d \varsigma^{jj} x_j^2 - \sum_{j=2}^d \varsigma^{1j} x_1 x_j - \frac{1}{2} \log |\Sigma| \right\}
\]

\[
\triangleq h(x) \exp \{ \eta^\top(\theta) T(x) - A(\theta) \},
\]

226
where

$$h(x) = (2\pi)^{-d/2};$$

$$\eta^\top(\theta) = -\left(\frac{\varsigma_{11}}{2}, \ldots, \frac{\varsigma_{dd}}{2}, \varsigma_{12}, \ldots, \varsigma_{1d}\right);$$

$$T^\top(x) = (x_1^2, \ldots, x_d^2, x_1x_2, \ldots, x_1x_d);$$

$$A(\theta) = \frac{1}{2} \log |\Sigma|.$$

Given an i.i.d. sample, the maximum likelihood estimators of the parameters in an exponential family distribution can be found by equating the sample average of the sufficient statistics \(T(x)\) for all observations to the mean of \(T(X)\) (see, e.g., Chapter 2 of Bickel and Doksum (2001)). Let \(x_i = (x_{i1}, \ldots, x_{id}), 1 \leq i \leq n\). The transformed score equation

$$\mathbb{E}[T(X_1)] = \frac{1}{n} \sum_{i=1}^{n} T(x_i)$$

$$\implies (\hat{\sigma}_1^2, \ldots, \hat{\sigma}_d^2, \hat{\sigma}_1\hat{\sigma}_2\hat{\rho}_{12}, \ldots, \hat{\sigma}_1\hat{\sigma}_d\hat{\rho}_{1d}) = \frac{1}{n} \sum_{i=1}^{d} (x_{i1}^2, \ldots, x_{id}^2, x_{i1}x_{i2}, \ldots, x_{i1}x_{id}).$$

It is then easy to see that

$$\hat{\sigma}_j^2 = \frac{1}{n} \sum_{i=1}^{d} x_{ij}^2, \ 1 \leq j \leq d; \ \hat{\rho}_{1k} = \frac{1}{n\hat{\sigma}_1\hat{\sigma}_k} \sum_{i=1}^{d} x_{i1}x_{ik}, \ 2 \leq k \leq d.$$ 

This means that \(\hat{\rho}_{1j} = r_{1j}\), the sample correlations as the data have been centred. However, for \(2 \leq j < k \leq d\), note that \(\hat{\rho}_{jk} = \hat{\rho}_{1j}\hat{\rho}_{1k}\) which is not equal to \(r_{jk} = \sum_{i=1}^{d} x_{ij}x_{ik}/(n\hat{\sigma}_j\hat{\sigma}_k)\) in general.

The preceding proof can be modified easily for other 1-truncated Gaussian vines by rewriting the quantities in terms of the variable pairs directly connected in the corresponding graphical model.

Next, we consider a \(d\)-dimensional 2-truncated Gaussian C-vine, whose covariance matrix \(\Sigma\) contains the following entries:

- \((j, j), 1 \leq j \leq d: \sigma_j^2;\)
- \((1, k)\) and \((k, 1), 2 \leq k \leq d: \sigma_1\sigma_k\rho_{1k};\)
- \((2, m)\) and \((m, 2), 3 \leq m \leq d: \sigma_2\sigma_m [\rho_{12}\rho_{1m} + \rho_{2m};1\sqrt{(1 - \rho_{12}^2)(1 - \rho_{1m}^2)}];\)
- \((j, k), 2 \leq j < k \leq d: \sigma_j\sigma_k [\rho_{1j}\rho_{1k} + \rho_{2j};1\rho_{2k};1\sqrt{(1 - \rho_{1j}^2)(1 - \rho_{1k}^2)}].\)
The precision matrix $\Sigma^{-1}$ has entries $\varsigma^{jk} = \varsigma^{kj} = 0$ for $2 \leq j < k \leq d$ (if all correlation and partial correlation parameters above are non-zero), and can again be shown using a stochastic representation; in this case, assuming $\sigma_1 = \ldots = \sigma_d = 1$, the representation is

\[
\begin{align*}
X_1 &= Z_1 \\
X_2 &= \phi_{21}X_1 + \psi_2Z_2 \\
X_3 &= \phi_{31}X_1 + \phi_{32}X_2 + \psi_3Z_3 \\
\vdots \\
X_d &= \phi_{d1}X_1 + \phi_{d2}X_2 + \psi_dZ_d,
\end{align*}
\]

where $\phi_{j2} = \rho_{2j1}\sqrt{(1-\rho_{1j}^2)/(1-\rho_{12}^2)}$ for $j = 3, \ldots, d$, $\phi_{j1} = \rho_{1j} - \phi_{j2}\rho_{12}$ and $\psi_j = \sqrt{1 - \sum_{l=1}^{j-1} \sum_{k=1}^{j-1} \phi_{jl}\phi_{jk}\rho_{lk}}$ for $j = 2, \ldots, d$ (equations (6.8) and (6.11) of Joe (2014)), $\phi_{rs} = 0$ if not present in the above stochastic representation, and $\rho_{rs}$ are the entries of $\Sigma$). The density function of $X$ is

\[
f_X(x) = f_{X_1}(x_1) \cdot f_{X_2|X_1}(x_2|x_1) \cdots f_{X_d|X_1,\ldots,X_{d-1}}(x_d|x_1, \ldots, x_{d-1}) \\
\propto \exp\left\{-\frac{1}{2}x_1^2\right\} \cdot \exp\left\{-\frac{1}{2\psi_2^2}(x_2 - \phi_{21}x_1)^2\right\} \cdots \exp\left\{-\frac{1}{2\psi_d^2}(x_d - \phi_{d1}x_1 - \phi_{d2}x_2)^2\right\}.
\]

(C.4)

The terms in the exponent of (C.4) are $x_1^2, \ldots, x_d^2, x_1x_2, \ldots, x_1x_d, x_2x_3, \ldots, x_2x_d$; these correspond to the non-zero positions in $\Sigma^{-1}$.

For the 2-truncated C-vine, the exponential family representation of the $N(0, \Sigma)$ density is such that

\[
\eta^\top(\theta)T(x) = -\frac{1}{2} \sum_{j=1}^{d} \varsigma^{jj}x_j^2 - \sum_{j=1}^{2} \sum_{k=j+1}^{d} \varsigma^{jk}x_jx_k,
\]

with

\[
\eta^\top(\theta) = -\left(\frac{\varsigma^{11}}{2}, \ldots, \frac{\varsigma^{dd}}{2}, \varsigma^{12}, \ldots, \varsigma^{1d}, \ldots, \varsigma^{23}, \ldots, \varsigma^{2d}\right); \\
T^\top(x) = (x_1^2, \ldots, x_d^2, x_1x_2, \ldots, x_1x_d, x_2x_3, \ldots, x_2x_d).
\]

For the collection of observations $x = (x_1, \ldots, x_n)$ with $x_i = (x_{i1}, \ldots, x_{id})$, $1 \leq i \leq n$,
solving the score equation (C.3) yields

\[ \hat{\sigma}_j^2 = \frac{1}{n} \sum_{i=1}^{d} x_{ij}^2, \quad 1 \leq j \leq d; \]

\[ \hat{\rho}_{1k} = \frac{1}{n\hat{\sigma}_1\hat{\sigma}_k} \sum_{i=1}^{d} x_{i1}x_{ik}, \quad 2 \leq k \leq d; \]  \hspace{1cm} (C.5)

\[ \hat{\rho}_{2m} = \frac{1}{n\hat{\sigma}_2\hat{\sigma}_m} \sum_{i=1}^{d} x_{i2}x_{im}, \quad 3 \leq m \leq d. \]  \hspace{1cm} (C.6)

Here the maximum likelihood estimator of \( \rho_{2m;1} \) is \( (\hat{\rho}_{2m} - \hat{\rho}_{12}\hat{\rho}_{1m}) / \sqrt{(1 - \hat{\rho}_{12}^2)(1 - \hat{\rho}_{1m}^2)} \).

In this case, \( \hat{\rho}_{1k} \) and \( \hat{\rho}_{2m} \) are equal to the respective sample correlations, while for the other pairs we have

\[ \hat{\rho}_{jk} = \hat{\rho}_{ij}\hat{\rho}_{ik} + \hat{\rho}_{2j;1}\hat{\rho}_{2k;1}\sqrt{(1 - \hat{\rho}_{1j}^2)(1 - \hat{\rho}_{1k}^2)} \]

\[ \neq \frac{1}{n\hat{\sigma}_j\hat{\sigma}_k} \sum_{i=1}^{d} x_{ij}x_{ik} = r_{jk} \]

in general.

The proof above can similarly be modified to other 2-truncated vine structures by identifying the pairs connected in trees 1 and 2; the cross terms in the vectors \( \eta(\theta) \) and \( T(x) \) will involve these pairs, while equations (C.5) and (C.6) will solve for these pairwise correlations instead. Using an analogous argument, the above procedure can be repeated for \( p \)-truncated vines, where \( p \leq d \), to show that the maximum likelihood estimates of the correlation parameters for pairs connected before and including tree \( p \) are the same as the sample correlations, while those after tree \( p \) are not necessarily so.

The case of fitting a Gaussian copula with data already marginally transformed to \( U(0,1) \) is equivalent to fitting the corresponding normal scores to the multivariate Gaussian distribution with variances fixed at 1. This is different from the situation above and the maximum likelihood estimates of the correlation parameters for the “before” pairs are close to, but not exactly the same as the sample correlations.
Appendix D

Derivations of the properties of correlation parameters for the exchangeable Gaussian distribution

The following derivations are about the exchangeable Gaussian distribution in Section 7.3.1. We first list the results of some moments concerning correlated Gaussian random variables that are useful in subsequent derivations. Suppose \((X_j, X_k, X_l, X_m)\) jointly follows (7.9) with \(\mu = 0\) and \(\sigma^2 = 1\). Then the fourth moments are given by

\[
E(X_j^4) = 3, \quad E(X_j^3X_k) = 3\rho, \quad E(X_j^2X_k^2) = 1 + 2\rho^2, \quad E(X_j^2X_kX_l) = \rho(1 + 2\rho) \quad \text{and} \quad E(X_jX_kX_lX_m) = 3\rho^2.
\]

This can be obtained from the Cholesky decomposition of the correlation matrix, resulting in the stochastic representation

\[
\begin{aligned}
X_j &= W_j; \\
X_k &= \rho W_j + \sqrt{1 - \rho^2} W_k; \\
X_l &= \rho W_j + \frac{\rho(1 - \rho)}{\sqrt{1 - \rho^2}} W_k + \frac{\sqrt{(1 + 2\rho)(1 - \rho)}}{\sqrt{1 + \rho}} W_l; \\
X_m &= \rho W_j + \frac{\rho(1 - \rho)}{\sqrt{1 - \rho^2}} W_k + \frac{\rho(1 - \rho)}{\sqrt{(1 + 2\rho)(1 - \rho^2)}} W_l + \frac{\sqrt{(1 + 2\rho - 3\rho^2)}}{\sqrt{1 + 2\rho}} W_m,
\end{aligned}
\]

where \(W_j, W_k, W_l\) and \(W_m\) are independent standard Gaussian random variables.

D.1 Variance of the difference for one bivariate margin

We first obtain the diagonal elements of the asymptotic covariance matrix by deriving the asymptotic variance of \(\hat{\rho}_{jk} - \hat{\rho}\), and show that separability holds in this case. In what follows we take \(j = 1\) and \(k = 2\), since the results for the other margins are the same. We use the
letter $Z$ to denote the centred random variables of $Y$ in Section 7.3.1. Let $S_i = (Z_{i1} + Z_{i2})^2$, $T_i = Z_{i1}^2 + Z_{i2}^2$, $U_i = \left( \sum_{j=1}^{d} Z_{ij} \right)^2$, $V_i = \sum_{j=1}^{d} Z_{ij}^2$, and the corresponding first observations (together with those for the $Z$’s) be denoted by their respective letters without subscript $i$. Their sample means are denoted by $\overline{S}$, $\overline{T}$, $\overline{U}$ and $\overline{V}$ with expectations $\mu_S \triangleq \mathbb{E}(\overline{S}) = 2(1 + \rho)$, $\mu_T \triangleq \mathbb{E}(\overline{T}) = 2$, $\mu_U \triangleq \mathbb{E}(\overline{U}) = d[1 + (d - 1)\rho]$ and $\mu_V \triangleq \mathbb{E}(\overline{V}) = d$, respectively. Then, by letting $h_1(s, t) = s/t$; $h_2(u, v) = u/v$ and performing Taylor series expansions of $h$ about $(\mu_S, \mu_T)$ and $(\mu_U, \mu_V)$, we obtain

\[
n \text{Var}(\hat{\rho}_{12}) = n \text{Var}(\overline{S}/\overline{T} - \mu_S/\mu_T) \n \quad \rightarrow \left( \frac{\partial h_1}{\partial s} \right)^2 \text{Var}(S) + \left( \frac{\partial h_1}{\partial t} \right)^2 \text{Var}(T) + 2 \frac{\partial h_1}{\partial s} \frac{\partial h_1}{\partial t} \text{Cov}(S, T) \n \quad = \frac{1}{\mu_T^2} \text{Var}(S) + \frac{\mu_S^2}{\mu_T^2} \text{Var}(T) - \frac{2 \mu_S}{\mu_T^2} \text{Cov}(S, T); \n \text{Var}(\hat{\rho}) = n \frac{1}{(d - 1)^2} \text{Var}(\overline{U}/\overline{V} - \mu_U/\mu_V) \n \quad \rightarrow \frac{1}{(d - 1)^2} \left[ \left( \frac{\partial h_2}{\partial u} \right)^2 \text{Var}(U) + \left( \frac{\partial h_2}{\partial v} \right)^2 \text{Var}(V) + 2 \frac{\partial h_2}{\partial u} \frac{\partial h_2}{\partial v} \text{Cov}(U, V) \right] \n \quad = \frac{1}{(d - 1)^2} \left[ \frac{1}{\mu_V^2} \text{Var}(U) + \frac{\mu_U^2}{\mu_V^2} \text{Var}(V) - \frac{2 \mu_U}{\mu_V^2} \text{Cov}(U, V) \right]. \]

The variances and covariances are as follows:

- **Var($S$) = Var($(Z_1 + Z_2)^2$):**
  \[
  \text{Var} \left[ (Z_1 + Z_2)^2 \right] = \mathbb{E} \left[ (Z_1 + Z_2)^4 \right] - \mu_S^2 = 2\mathbb{E}(Z_j^4) + 8\mathbb{E}(Z_j^2Z_k) + 6\mathbb{E}(Z_j^2Z_l) - \mu_S^2 = 8(1 + \rho^2). \]

- **Var($T$) = Var$(Z_1^2 + Z_2^2)$:**
  \[
  \text{Var} \left( Z_1^2 + Z_2^2 \right) = \mathbb{E} \left[ (Z_1^2 + Z_2^2)^2 \right] - \mu_T^2 = 2\mathbb{E}(Z_j^4) + 2\mathbb{E}(Z_j^2Z_k)^2 - \mu_T^2 = 4(1 + \rho^2). \]

- **Var($U$) = Var$((Z_1 + \cdots + Z_d)^2$):**
  \[
  \text{Var} \left[ (Z_1 + \cdots + Z_d)^2 \right] = \mathbb{E} \left[ (Z_1 + \cdots + Z_d)^4 \right] - \mu_U^2 \n  = d\mathbb{E}(Z_j^4) + 4d(d - 1)\mathbb{E}(Z_j^2Z_k) + 3d(d - 1)\mathbb{E}(Z_j^2Z_l) + 6d(d - 1)(d - 2)\mathbb{E}(Z_jZ_kZ_l) \n  + d(d - 1)(d - 2)(d - 3)\mathbb{E}(Z_jZ_kZ_lZ_m) - d^2 \left[ 1 + (d - 1)\rho \right]^2 \n  = 2d^2 \left[ 1 + (d - 1)\rho \right]^2. \]

- **Var($V$) = Var$(Z_1^2 + \cdots + Z_d^2)$:**
  \[
  \text{Var} \left( Z_1^2 + \cdots + Z_d^2 \right) = d\mathbb{E}(Z_j^4) + d(d - 1)\mathbb{E}(Z_j^2Z_k)^2 - d^2 = 2d \left[ 1 + (d - 1)\rho \right]^2. \]
\[ \text{Cov}(S, T) = \text{Cov} \left[ (Z_1 + Z_2)^2, Z_1^2 + Z_2^2 \right] : \]
\[
\text{Cov} \left[ (Z_1 + Z_2)^2, Z_1^2 + Z_2^2 \right] = 2\mathbb{E}(Z_1^4) + 4\mathbb{E}(Z_1^3 Z_k) + 2\mathbb{E}(Z_1^2 Z_k^2) - \mu_S \mu_T = 4(1 + \rho)^2. 
\]

\[ \text{Cov}(U, V) = \text{Cov} \left[ (Z_1 + \cdots + Z_d)^2, Z_1^2 + \cdots + Z_d^2 \right] : \]
\[
\text{Cov} \left[ (Z_1 + \cdots + Z_d)^2, Z_1^2 + \cdots + Z_d^2 \right] = \text{d} \mathbb{E} \left[ Z_1^2 \left( Z_1 + \cdots + Z_d\right)^2 \right] - d^2 [1 + (d-1)\rho] \\
= \text{d} \left[ \mathbb{E}(Z_1^4) + 2(d-1)\mathbb{E}(Z_1^3 Z_k) + (d-1)\mathbb{E}(Z_1^2 Z_k^2) + (d-1)(d-2)\mathbb{E}(Z_1^2 Z_k Z_l) \right] \\
- d^2 [1 + (d-1)\rho] \\
= 2d [1 + (d-1)\rho]^2. 
\]

Hence
\[
n\text{Var}(\hat{\rho}_{12}) \rightarrow \frac{8(1 + \rho)^2}{4} + \frac{16(1 + \rho)^2(1 + \rho^2)}{16} - \frac{16(1 + \rho)(1 + \rho)^2}{8} = (1 - \rho^2)^2, 
\]
and
\[
n\text{Var}(\hat{\rho}) \\
\rightarrow \frac{1}{(d-1)^2} \left( \frac{2d^2 [1 + (d-1)\rho]^2}{d^2} + \frac{2d^3 [1 + (d-1)\rho]^2 [1 + (d-1)\rho^2]}{d^3} - \frac{4d^2 [1 + (d-1)\rho]^3}{d^3} \right) \\
= \frac{2(1 - \rho)^2 [1 + (d-1)\rho]^2}{d(d-1)} 
\]
after some simplification. Note that \( \lim_{n \to \infty} n\text{Var}(\hat{\rho}_{12}) \) is the same as the asymptotic variance of the sample correlation.

Next, note that
\[
n\text{Var}(\hat{\rho}_{12} - \hat{\rho}) = n\text{Var} \left[ \frac{S}{T} - (d-1)^{-1} \frac{U}{V} \right] \\
= n\text{Var} \left[ \frac{S}{T} - (d-1)^{-1} \frac{U}{V} \right] - \left[ \frac{\mu_S}{\mu_T} - (d-1)^{-1} \frac{\mu_U}{\mu_V} \right]. 
\]

Consider the function \( h_3(s, t, u, v) = s/t - (d-1)^{-1} u/v \) and its derivative at \( (\mu_S, \mu_T, \mu_U, \mu_V) \). A similar technique yields
\[
n\text{Var}(\hat{\rho}_{12} - \hat{\rho}) \\
\rightarrow \frac{\text{Var}(S)}{\mu_T^2} + \frac{\mu_S^2 \text{Var}(T)}{\mu_T^4} - 2\mu_S \text{Cov}(S, T) \\
\quad + \frac{1}{(d-1)^2} \left[ \frac{\text{Var}(U)}{\mu_V^2} + \frac{\mu_U^2 \text{Var}(V)}{\mu_V^4} - 2\mu_U \text{Cov}(U, V) \right] \\
\quad + \frac{2}{(d-1)} \left[ \frac{-\text{Cov}(S, U)}{\mu_T \mu_V} + \frac{\mu_U \text{Cov}(S, V)}{\mu_T^2 \mu_V} + \frac{\mu_S \text{Cov}(T, U)}{\mu_T \mu_V} - \frac{\mu_S \mu_U \text{Cov}(T, V)}{\mu_T^2 \mu_V} \right]. \quad (D.1) 
\]
The first two lines of (D.1) correspond to the limits of $n\text{Var}(\hat{\rho}_{12})$ and $n\text{Var}(\hat{\rho})$, while the third line is equal to the limit of $-2n\text{Cov}(\hat{\rho}_{12}, \hat{\rho})$ and contains another 4 covariances to be calculated. They are:

- $\text{Cov}(S, U) = \text{Cov} \left[ (Z_1 + Z_2)^2, (Z_1 + \cdots + Z_d)^2 \right]$
  \[
  = 2 \left[ \mathbb{E}(Z_1^4) + 2(d - 1)\mathbb{E}(Z_1^2Z_k) + (d - 1)\mathbb{E}(Z_1^2Z_k^2) + (d - 1)(d - 2)\mathbb{E}(Z_1^2Z_kZ_l) \right]
  + 4\mathbb{E}(Z_1^2Z_k^2) + 4\mathbb{E}(Z_1^2Z_k^2) + 10(d - 2)\mathbb{E}(Z_1^2Z_kZ_l) + 2(d - 2)(d - 3)\mathbb{E}(Z_1Z_kZ_l)Z_m)
  - 2(d + \rho)(1 + \rho) \right]
  = 8[1 + (d - 1)\rho]^2,\]
  where (D.2) is the result of splitting $\mathbb{E}\left[ (Z_1 + Z_2)^2(Z_1 + \cdots + Z_d)^2 \right]$ into $2\mathbb{E}[Z_1^2(Z_1 + \cdots + Z_d)^2] + 2\mathbb{E}[Z_1Z_2(Z_1 + \cdots + Z_d)^2]$.

- $\text{Cov}(S, V) = \text{Cov} \left[ (Z_1 + Z_2)^2, Z_1^2 + \cdots + Z_d^2 \right]$
  \[
  = \mathbb{E} \left[ (Z_1 + Z_2)^2(Z_1^2 + \cdots + Z_d^2) \right] - \mu_S\mu_V
  = 2 \left[ \mathbb{E}(Z_1^4) + (d - 1)\mathbb{E}(Z_1^2Z_k^2) \right] + 4\mathbb{E}(Z_1^2Z_k^2) + 2(d - 2)\mathbb{E}(Z_1^2Z_kZ_l) - 2d(1 + \rho)
  = 4 \left[ 1 + 2\rho + (2d - 3)\rho^2 \right].
  \]

- $\text{Cov}(T, U) = \text{Cov} \left[ Z_1^2 + Z_2^2, (Z_1 + \cdots + Z_d)^2 \right]$
  \[
  = 2 \left[ \mathbb{E}(Z_1^4) + 2(d - 1)\mathbb{E}(Z_1^2Z_k) + (d - 1)\mathbb{E}(Z_1^2Z_k^2) + (d - 1)(d - 2)\mathbb{E}(Z_1^2Z_kZ_l) \right]
  - 2d[1 + (d - 1)\rho]
  = 4[1 + (d - 1)\rho]^2.
  \]

- $\text{Cov}(T, V) = \text{Cov} \left[ Z_1^2 + Z_2^2, Z_1^2 + \cdots + Z_d^2 \right]$
  \[
  \text{Cov} \left( Z_1^2 + Z_2^2, Z_1^2 + \cdots + Z_d^2 \right) = 2 \mathbb{E}(Z_1^4) + (d - 1)\mathbb{E}(Z_1^2Z_k^2) - 2d = 4 \left[ 1 + (d - 1)\rho^2 \right].
  \]

Hence

\[
\frac{n\text{Cov}(\hat{\rho}_{12}, \hat{\rho})}{(d - 1)} \rightarrow \frac{1}{2d} \left( \frac{8[1 + (d - 1)\rho]^2}{d} - \frac{4d[1 + (d - 1)\rho][1 + 2\rho + (2d - 3)\rho^2]}{2d^2} 
  - \frac{8(1 + \rho)[1 + (d - 1)\rho]^2}{4d}
  + \frac{8d(1 + \rho)[1 + (d - 1)\rho][1 + (d - 1)\rho^2]}{4d^2} \right)
  = \frac{2(1 - \rho)^2[1 + (d - 1)\rho]^2}{d(d - 1)} = \lim_{n \to \infty} n\text{Var}(\hat{\rho}),
\]

(D.3)
which also proves that \( \lim_{n \to \infty} n \text{Var}(\hat{\rho}_{12} - \hat{\rho}) = \lim_{n \to \infty} n [\text{Var}(\hat{\rho}_{12}) - \text{Var}(\hat{\rho})]. \)

### D.2 Covariance between differences for two bivariate margins

The off-diagonal elements of the asymptotic covariance matrix can be further divided into two subgroups: (1) Covariances with a shared margin and (2) Covariances without any shared margins. Without loss of generality, we obtain the limits for \( n \text{Cov}(\hat{\rho}_{12} - \hat{\rho}, \hat{\rho}_{13} - \hat{\rho}) \) in subgroup (1) and \( n \text{Cov}(\hat{\rho}_{12} - \hat{\rho}, \hat{\rho}_{34} - \hat{\rho}) \) in subgroup (2). Note, in the special case of \( d = 3 \), that no element belongs to subgroup (2). Either case, the covariance can be written as

\[
\text{Cov}(\hat{\rho}_{jk}, \hat{\rho}_{lm}) = n \left[ \text{Cov}(\hat{\rho}_{jk}, \hat{\rho}_{lm}) - \text{Cov}(\hat{\rho}_{jk}, \hat{\rho}) - \text{Cov}(\hat{\rho}_{lm}, \hat{\rho}) + \text{Var}(\hat{\rho}) \right] \quad (D.4)
\]

where \( j = l \) for subgroup (1) and \( j \neq l \) for subgroup (2), and \( k \neq m \) (and are neither \( j \) nor \( l \)) in both cases. The equality comes from the fact that the last three terms in the right hand side of (D.4) are the same, as proved in (D.3). Thus the only additional task needed is to compute the limit of \( n \text{Cov}(\hat{\rho}_{jk}, \hat{\rho}_{lm}) \), which varies depending on whether \( j = l \).

We obtain an expression for \( \text{Cov}(\hat{\rho}_{jk}, \hat{\rho}_{lm}) \) indirectly through expanding \( n \text{Var}(\hat{\rho}_{jk} - \hat{\rho}_{lm}) \). Define \( S_{jk} = \frac{1}{n} \sum_{i=1}^{n} (Z_{ij} + Z_{ik})^2, T_{jk} = \frac{1}{n} \sum_{i=1}^{n} (Z_{ij}^2 + Z_{ik}^2) \) with corresponding first observations \( S_{jk} \) and \( T_{jk} \), respectively. Then we have

\[
n \text{Var}(\hat{\rho}_{jk} - \hat{\rho}_{lm}) = n \text{Var}(S_{jk}/T_{jk} - S_{lm}/T_{lm}) = n \left[ \text{Var}(\hat{\rho}_{jk}) + \text{Var}(\hat{\rho}_{lm}) - 2\text{Cov}(\hat{\rho}_{jk}, \hat{\rho}_{lm}) \right].
\]

Let \( h_{4}(s_{jk}, t_{jk}, s_{lm}, t_{lm}) = s_{jk}/t_{jk} - s_{lm}/t_{lm} \). From the Taylor series expansion of \( S_{jk}/T_{jk} - S_{lm}/T_{lm} \) about \( (\mu_s, \mu_T, \mu_s, \mu_T) \), the respective expectations of the four sample averages, we obtain

\[
n \text{Var}(S_{jk}/T_{jk} - S_{lm}/T_{lm}) \to \text{Var} \left( \frac{S_{jk}}{\mu_T} - \frac{\mu_s T_{jk}}{\mu_T^2} - \frac{S_{lm}}{\mu_T} + \frac{\mu_s T_{lm}}{\mu_T^2} \right).
\]

Subtracting terms that belong to \( n \text{Var}(\hat{\rho}_{jk}) \) and \( n \text{Var}(\hat{\rho}_{lm}) \), we arrive at the following:

\[
n \text{Cov}(\hat{\rho}_{jk}, \hat{\rho}_{lm}) \to \text{Cov}(S_{jk}, S_{lm}) - \frac{\mu_s}{\mu_T^3} \text{Cov}(S_{jk}, T_{lm}) + \frac{\mu_s^2}{\mu_T^4} \text{Cov}(T_{jk}, T_{lm}).
\]

For subgroup (1), we obtain:

- \( \text{Cov}(S_{12}, S_{13}) = \text{Cov} \left[ (Z_1 + Z_2)^2, (Z_1 + Z_3)^2 \right] \):
  \[
  \text{Cov} \left[ (Z_1 + Z_2)^2, (Z_1 + Z_3)^2 \right] = \mathbb{E}(Z_1^2) + 4\mathbb{E}(Z_1^2 Z_k) + 3\mathbb{E}(Z_2^2 Z_k) + 8\mathbb{E}(Z_2^2 Z_k Z_l) - 4(1 + \rho)^2
  \]
  \[
  = 2(1 + 3\rho)^2.
  \]
Hence

\[ \text{Cov}(S_{12}, T_{13}) = \text{Cov} \left[ (Z_1 + Z_2)^2, (Z_1^2 + Z_3^2) \right] : \]

\[ \begin{align*}
\text{Cov} \left[ (Z_1 + Z_2)^2, (Z_1^2 + Z_3^2) \right] &= \mathbb{E}(Z_j^4) + 2\mathbb{E}(Z_j^3 Z_k) + 3\mathbb{E}(Z_j^2 Z_k^2) + 2\mathbb{E}(Z_j Z_k Z_l) - 4(1 + \rho) \\
&= 2(1 + 2\rho + 5\rho^2),
\end{align*} \]

which is the same as \text{Cov}(T_{12}, S_{13}).

\[ \text{Cov}(T_{12}, T_{13}) = \text{Cov} \left( Z_1^2 + Z_2^2, Z_1^2 + Z_3^2 \right) : \]

\[ \text{Cov} \left( Z_1^2 + Z_2^2, Z_1^2 + Z_3^2 \right) = \mathbb{E}(Z_j^4) + 3\mathbb{E}(Z_j^2 Z_k^2) - 4 = 2(1 + 3\rho^2). \]

Hence

\[ \begin{align*}
n\text{Cov}(\hat{\rho}_{12}, \hat{\rho}_{13}) &\to \frac{2(1 + 3\rho)^2}{4} - \frac{8(1 + \rho)(1 + 2\rho + 5\rho^2)}{8} + \frac{8(1 + \rho)^2(1 + 3\rho^2)}{16} \\
&= \frac{1}{2\rho}(1 - \rho)^2(2 + 3\rho).
\end{align*} \]

For subgroup (2), we obtain:

\[ \text{Cov}(S_{12}, S_{34}) = \text{Cov} \left[ (Z_1 + Z_2)^2, (Z_3 + Z_4)^2 \right] : \]

\[ \begin{align*}
\text{Cov} \left[ (Z_1 + Z_2)^2, (Z_3 + Z_4)^2 \right] &= 4\mathbb{E}(Z_j^2 Z_k^2) + 8\mathbb{E}(Z_j Z_k Z_l) + 4\mathbb{E}(Z_j Z_k Z_l Z_m) - 4(1 + \rho)^2 \\
&= 32\rho^2.
\end{align*} \]

\[ \text{Cov}(S_{12}, T_{34}) = \text{Cov} \left[ (Z_1 + Z_2)^2, Z_3^2 + Z_4^2 \right] : \]

\[ \text{Cov} \left[ (Z_1 + Z_2)^2, Z_3^2 + Z_4^2 \right] = 4\mathbb{E}(Z_j^2 Z_k^2) + 4\mathbb{E}(Z_j Z_k Z_l) - 4(1 + \rho) = 16\rho^2. \]

which is the same as \text{Cov}(T_{12}, S_{34}).

\[ \text{Cov}(T_{12}, T_{34}) = \text{Cov} \left( Z_1^2 + Z_2^2, Z_3^2 + Z_4^2 \right) : \]

\[ \text{Cov} \left( Z_1^2 + Z_2^2, Z_3^2 + Z_4^2 \right) = 4\mathbb{E}(Z_j^2 Z_k^2) - 4 = 8\rho^2. \]

Hence

\[ \begin{align*}
n\text{Cov}(\hat{\rho}_{12}, \hat{\rho}_{34}) &\to \frac{32\rho^2}{4} - \frac{64\rho^2(1 + \rho)}{8} + \frac{32\rho^2(1 + \rho)^2}{16} = 2\rho^2(1 - \rho)^2.
\end{align*} \]

Finally, with \( k, m, j \) distinct and \( k, m, l \) distinct, we have

\[ n\text{Cov}(\hat{\rho}_{jk} - \hat{\rho}_{lm} - \hat{\rho}) = \begin{cases}
\frac{(1 - \rho)^2 [(d - 1)(d - 4)\rho(2 - \rho) - 4]}{2d(d - 1)}, & \text{if } j = l; \\
\frac{2(1 - \rho)^2 [(d - 1)\rho(2 - \rho) + 1]}{d(d - 1)}, & \text{if } j \neq l.
\end{cases} \]

Figure D.1 plots the various quantities derived above against \( \rho \), for different dimensions \( d \).
Figure D.1: Plots of various asymptotic variances and covariances for different dimensions and true correlations of the positive exchangeable Gaussian distribution. The thick line in the top left panel plots the empirical asymptotic variance, and is the upper bound of the model-based asymptotic variance.
Appendix E

An $O(N \log_2 N)$ algorithm for evaluating the bivariate empirical distribution function at the $N$ observations

Let $F(y_1, \ldots, y_d)$ be a high-dimensional distribution (large values of $d$) with a parsimonious structure imposed, e.g., factor or truncated vine. We mentioned in Chapter 7 that the bivariate marginal distributions of $F$ may not have analytic expressions, and even numerical evaluation could be impractical. For example, with a general (regular) vine copula, the $(j, k)$ bivariate marginal distribution involves an $(m - 1)$-dimensional integral if $(j, k)$ is linked in tree $m$, and simplifications are only possible in certain special cases such as a low-order truncated C-vine. This prohibits the evaluation of quantities of the form

$$\int \int h(F_{jk}(y_j, y_k)) \ dF_{jk}(y_j, y_k)$$  \hspace{1cm} (E.1)

for the $(j, k)$ bivariate margin, where $h$ is a possibly non-linear function (or more generally, integrals with respect to the whole distribution with integrand being functions of bivariate distributions $\int h(F_{12}, F_{13}, \ldots, F_{d-1,d}) \ dF$, such as that in (7.12)). If it is computationally easy to simulate from $F$, it is then practical to estimate (E.1) by the sample counterpart

$$\frac{1}{N} \sum_{i=1}^{N} h(\hat{F}_{N,jk}(Y_{ij}, Y_{ik})) ,$$  \hspace{1cm} (E.2)
where $N$ is the sample size of the Monte Carlo (MC) simulation and

$$\hat{F}_{N,jk}(Y_{ij}, Y_{ik}) = N^{-1} \sum_{m=1}^{N} 1 \{ Y_{mj} \leq Y_{ij}, Y_{mk} \leq Y_{ik} \}$$

is the bivariate empirical distribution function at $(Y_{ij}, Y_{ik})$ based on the sample \{(Y_{1j}, Y_{1k}), \ldots, (Y_{Nj}, Y_{Nk})\}. In the following, assume $h$ is continuous and integrable with respect to $F$, so that the convergence of (E.2) to (E.1) in probability as $N \to \infty$ is guaranteed via the Glivenko-Cantelli theorem and the law of large numbers.

To compute (E.2), it is necessary to evaluate $\hat{F}_{N,jk}$ at each of the $N$ points\(^{27}\). The naive way to do this is to make the $O(N)$ comparisons in (E.3) for each of the $N$ observations, for a total of $O(N^2)$ comparison operations. This is inefficient for large MC sample sizes. Because the bivariate empirical distribution has to be obtained for each of the $(d^2)$ bivariate margins, it is crucial to be able to evaluate the bivariate empirical distribution function efficiently.

We use a procedure modified from either the merge sort or quicksort algorithm (see, e.g., Knuth (1998); Cormen et al. (2009)). These two are common sorting algorithms that are efficient in the sense that on average they have an $O(N \log_2 N)$ complexity (i.e., number of comparisons). The merge sort has the extra advantage that the worst case complexity is also $O(N \log_2 N)$; it is possible for the quicksort algorithm to take $O(N^2)$ comparisons, although this can be practically avoided with a careful design of the sorting mechanism. The proposed modifications use an extra layer of indices as storage and preserve the $O(N \log_2 N)$ complexity of these two algorithms. Without loss of generality, let $j = 1$ and $k = 2$ in (E.3). For our purpose, assume that each of the vectors \{Y_{11}, \ldots, Y_{N1}\} and \{Y_{12}, \ldots, Y_{N2}\} is a scrambled sequence of the integers \{1, \ldots, N\} (as only their ranks are relevant\(^{28}\)), and the vector \{Y_{11}, \ldots, Y_{N1}\} has been sorted by an efficient algorithm (i.e., with $O(N \log_2 N)$ complexity) to result in the sorted sequence \{Y_{k_1,1}, \ldots, Y_{k_N,1}\}, where $k_j$ is the index of the $j$th smallest observation, $j = 1, \ldots, N$, so that the problem amounts to finding the sequential rank with respect to the second index. That is, for each of the $Y_{k_j,2}$’s, we find its rank among the elements \{Y_{k_1,2}, \ldots, Y_{k_j,2}\}, as the rest \{Y_{k_j+1,2}, \ldots, Y_{k_N,2}\} are such that $Y_{k_j+1,1}, \ldots, Y_{k_n,1}$ are all larger than $Y_{k_j,1}$ and hence the indicator function in (E.3) evaluates to zero. In the following, we describe the algorithms to achieve this, and for

\(^{27}\)For our purpose in Section 7.4, we also need to evaluate the bivariate empirical survival function at the $N$ observations. However, since $1 \{ Y_{mj} \geq Y_{ij}, Y_{mk} \geq Y_{ik} \} = 1 \{ -Y_{mj} \leq -Y_{ij}, -Y_{mk} \leq -Y_{ik} \}$, the survival function can be obtained at the same order of complexity as $\hat{F}_{N,jk}$. We therefore only focus on the evaluation of $\hat{F}_{N,jk}$ in this appendix.

\(^{28}\)We assume $F$ is an absolutely continuous distribution and there are no ties in the simulated values. A comment for the case with ties is given at the end of Section E.1.
simplicity in notation we assume \{Y_{11}, \ldots, Y_{N1}\} is the sorted sequence with corresponding second variable \{Y_{12}, \ldots, Y_{N2}\}.

### E.1 The modified merge sort algorithm

The merge sort algorithm divides elements into pairs and makes comparison for each pair. Pairs are then merged, two at a time, to yield sorted subsequences each of length 4. They are then merged to yield longer sorted subsequences until the whole sequence is sorted. Figure E.1 illustrates the merge sort with \(N = 8\) elements. There are \(O(N)\) comparisons in each layer, but there are only \(O(\log_2 N)\) layers as each merge reduces the number of groups by half. This results in a total of \(O(N \log_2 N)\) comparisons to be made regardless of the initial degree of “sortedness” of the data.

![Figure E.1: Illustration of the merge sort with 8 elements](image)

Algorithm 1 displays the pseudocode of this algorithm; it recursively splits elements into half and these shorter subsequences are sorted and merged. The operation indicated by the curly braces in Figure E.1 is the essence of the `Merge` function. The two vectors being merged are scanned from left to right; with each scan the smaller element is inserted into the output vector, and the index of the vector to which the smaller element belongs is incremented so that the next scan compares the element to its right with the same element in the other vector.

We modify the merge sort by adding a counter associated with each element. In each merge operation, the counter associated with an element \(x\) in the “right” vector (i.e., \(y_2\) in Algorithm 1) is incremented by the number of elements from the “left” vector (i.e., \(y_1\) in Algorithm 1) already inserted into the output vector. The reasoning is that these elements...
Algorithm 1 The merge sort algorithm

1: function Sort(y)  \Comment{The main merge sort function; recursive}
2: \hspace{1em} \textbf{N} \leftarrow \text{length}(y)
3: \hspace{1em} \textbf{if} \ N = 1 \textbf{then}
4: \hspace{2em} \textbf{return} \ y
5: \hspace{1em} \textbf{else}
6: \hspace{2em} \textbf{m} \leftarrow \lfloor \text{N}/2 \rfloor
7: \hspace{2em} y_1 \leftarrow \text{Sort}(y[1:\text{m}])
8: \hspace{2em} y_2 \leftarrow \text{Sort}(y[(\text{m} + 1):\text{N}])
9: \hspace{2em} y \leftarrow \text{Merge}(y_1, y_2)
10: \hspace{1em} \textbf{return} \ y
11: \hspace{1em} \textbf{end if}
12: \hspace{1em} \textbf{end function}

13: function Merge(y_1, y_2)  \Comment{The function to sort and merge two subsequences}
14: \hspace{1em} \textbf{N}_1 \leftarrow \text{length}(y_1); \textbf{N}_2 \leftarrow \text{length}(y_2); \textbf{N} \leftarrow \textbf{N}_1 + \textbf{N}_2
15: \hspace{1em} \textbf{if} \ y_1[\textbf{N}_1] \leq y_2[1] \textbf{then}
16: \hspace{2em} y \leftarrow \text{concatenate} \ y_1, y_2
17: \hspace{2em} \textbf{return} \ y
18: \hspace{1em} \textbf{else}
19: \hspace{2em} \textbf{initialize} \ i \leftarrow 1, j \leftarrow 1, k \leftarrow 1
20: \hspace{2em} \textbf{initialize} \ y \text{ of length } \textbf{N}
21: \hspace{1em} \textbf{while} \ k \leq \textbf{N} \textbf{do}
22: \hspace{2em} \textbf{if} \ y_1[i] \leq y_2[j] \textbf{then}
23: \hspace{3em} y[k] \leftarrow y_1[i]; \ i \leftarrow i + 1
24: \hspace{2em} \textbf{else}
25: \hspace{3em} y[k] \leftarrow y_2[j]; \ j \leftarrow j + 1
26: \hspace{2em} \textbf{end if}
27: \hspace{2em} \textbf{if} \ i > \textbf{N}_1 \textbf{ or } j > \textbf{N}_2 \textbf{ then}  \Comment{Done if one vector is exhausted}
28: \hspace{2em} \text{fill in the rest of } y \text{ with remaining content of the non-exhausted vector}
29: \hspace{2em} \textbf{break while}
30: \hspace{2em} \textbf{end if}
31: \hspace{2em} k \leftarrow k + 1
32: \hspace{2em} \textbf{end while}
33: \hspace{1em} \textbf{return} \ y
34: \hspace{1em} \textbf{end if}
35: \hspace{1em} \textbf{end function}
are smaller than \( x \), but have not been counted in the previous merge operations. Elements from the “right” vector inserted before \( x \) are not counted even if they are smaller than \( x \) as these elements have already been counted in a previous merge operation. The counters for elements in the “left” vector are left untouched, as (locally) there are no elements preceding them. Figure E.2 shows this procedure graphically. A number is added to the top right corner of a digit if its counter has to be incremented; for easier understanding, that number is the increment arising from that merge operation only, and the table on the right keeps track of the increments. The row labelled “total” gives the total number of lesser elements preceding each of them; to match the definition (E.3), 1 should be added to each final count to include the element itself, and then the count should be divided by \( N \) to obtain the empirical distribution function at \((Y_{i1}, Y_{i2})\).

![Figure E.2](image)

**Figure E.2**: Illustration of the modified merge sort with 8 elements. Elements in grey are those coming from the “left” vector and are thus eligible for counting if a larger element from the “right” vector enters after them. The smaller digit at the top right corner of each box is the counter for that merge operation. The row for “total” indicates the number of smaller elements to the left of each \( Y_{i2} \); the bivariate empirical distribution function at \((Y_{i1}, Y_{i2})\) is obtained by adding 1 to the “total” and then dividing the number by 8.

The pseudocode of this modified algorithm is displayed in Algorithm 2. The main function that splits elements by half, \texttt{Sort}, is unchanged, while an indicator counting the number of “left” elements entered is introduced in the \texttt{Merge} function. Note that, after
Algorithm 2 The modified merge sort algorithm for sequential ranks. The modification consists of lines 2, 17, 21, 23 and 32.

1: Input is the matrix \((Y_{11}, Y_{12}), \ldots, (Y_{N1}, Y_{N2})\), where \(Y_1 = Y_{11}, \ldots, Y_{N1}\) are sorted and \(Y_2 = Y_{12}, \ldots, Y_{N2}\) are paired accordingly. Convert \(Y_1\) and \(Y_2\) to the marginal ranks \(1, \ldots, N\). In the following, the vector \(Y_2\) is fed to \(y\).

2: initialize global variable \(z\) a vector of zeros of the same length as \(y\)
3: (comment: \(z\) will be modified by the following functions)

4: function Sort\((y)\)
   \(\triangleright\) The main merge sort function; recursive
   5: \(N \leftarrow \text{length}(y)\)
   6: if \(N = 1\) then
      7: return \(y\)
   8: else
      9: \(m \leftarrow \lfloor N/2 \rfloor\)
   10: \(y_1 \leftarrow \text{Sort}(y[1:m]); y_2 \leftarrow \text{Sort}(y[(m+1):N]); y \leftarrow \text{Merge}(y_1, y_2)\)
   11: return \(y\)
   12: end if

13: end function

14: function Merge\((y_1, y_2)\)
   \(\triangleright\) The function to sort and merge two subsequences
   15: \(N_1 \leftarrow \text{length}(y_1); N_2 \leftarrow \text{length}(y_2); N \leftarrow N_1 + N_2\)
   16: initialize \(i \leftarrow 1, j \leftarrow 1, k \leftarrow 1\)
   17: initialize \(b \leftarrow 0\) \(\triangleright\) Counts the number of “left” vector elements entered
   18: initialize \(y\) of length \(N\)
   19: while \(i \leq N_1\) and \(j \leq N_2\) do
      20: if \(y_1[i] \leq y_2[j]\) then
         21: \(y[k] \leftarrow y_1[i]; b \leftarrow b + 1; i \leftarrow i + 1\)
      22: else
         23: \(y[k] \leftarrow y_2[j]; z[y_2[j]] \leftarrow z[y_2[j]] + b; j \leftarrow j + 1\)
      24: end if
      25: \(k \leftarrow k + 1\)
   26: end while
   27: if \(i \leq N_1\) then \(\triangleright\) If “left” vector not yet exhausted
      28: fill in the rest of \(y\) with remaining content of the “left” vector
   29: end if
   30: if \(j \leq N_2\) then \(\triangleright\) If “right” vector not yet exhausted
      31: fill in the rest of \(y\) with remaining content of the “right” vector
      32: increment \(z\) (at indices of remaining “right” elements) by \(b\)
   33: end if
   34: return \(y\)
35: end function

36: Increment every element of \(z\) by 1, and then divide by \(N\) to obtain the empirical distribution function at the \(N\) observations; the \(r\)th element of \(z\) corresponds to the \(s\)th observation where \(Y_{s,2} = r\).
the function is executed, the $i$th entry of the vector $z$ contains the sequential rank for the element $i$ (not for the $i$th element). This can however be rearranged to the order of appearance of the elements at a computational complexity of $O(N)$. Because the vector $y$ used for sorting contains distinct integers, each observation (with value $y_1[i]$ or $y_2[j]$ in the algorithm) maps uniquely to one index of the $z$ vector. If ties are possible, for example in the case of discrete bivariate distributions, a second array should be implemented for the mapping of the observations to the $z$ vector, so that identical values in $y$ map to different indices in $z$. In addition, there are several possibilities that must be considered:

- If ties occur in the first key (i.e., $Y_{m1}$) but not the second (i.e., $Y_{m2}$, elements of the vector being fed into the sequential rank algorithm), then the second key should act as the tiebreaker when sorting the first key.

- If ties occur in the second key but not the first, then the algorithm is still applicable after implementing the aforementioned change of unique mapping to the $z$ vector.

- If there are identical entries (same first and second keys), it is necessary to accommodate for them using alternative methods, such as pre-populating the $z$ vector with non-zero entries. Since the sequential rank algorithm does not search beyond the current observation, it cannot detect the presence of identical entries that will lead to $1\{Y_{m1} \leq Y_{i1}, Y_{m2} \leq Y_{i2}\} = 1$ for some $m$ after the current observation $i$.

### E.2 The modified quicksort algorithm

An alternative to merge sort is the quicksort algorithm. Although it has a worst case complexity of $O(N^2)$, this can be avoided with a careful design and quicksort can sometimes be more efficient than merge sort (Skiena (2008)). At each iteration, a pivot is chosen and each element is compared to this pivot; all smaller elements are brought to one side of the pivot and the larger elements on the other side, thus forming two subsequences and completing the sorting of the pivot (i.e., the pivot is at its final position). This is repeated for each subsequence until every element has been sorted. The choice of the pivot is important; ideally it should roughly split the elements into two subsequences of similar lengths, so that the number of iterations has order $O(\log_2 N)$. Consider an already sorted list\(^{29}\) and always choosing the first or last element as the pivot. In this case, each step reduces the length of the longer sequence by 1 and thus $O(N)$ iterations are needed, resulting in the worst case complexity of $O(N^2)$. In statistical applications, we frequently encounter bivariate

\(^{29}\)An example of this is when observations are simulated from the comonotonicity or countermonotonicity copula.
observations that are related to each other (in the sense that a trend, not necessarily monotone, can be detected on a scatterplot); a fixed-location pivot is not desirable for the above reason. We thus adopt a random pivot (Section 7.3 of Cormen et al. (2009)) where each element of a subsequence has the same probability to be chosen.

Figure E.3 demonstrates the quicksort algorithm for the same 8 elements as in the preceding section. For simplicity in illustration, the middle element (or the \((N/2)\)-th element if the subsequence is of an even length \(N\)) is chosen as the pivot and highlighted in grey. Each step places elements sequentially to the left of the pivot if they are smaller, or to the right otherwise. The pivot is then highlighted in black, signifying that it has been sorted. Algorithm 3 has the pseudocode of the quicksort algorithm; it is the version that results in a stable sort, i.e., preserving the order in case of ties, at the expense of extra storage space. It is possible to perform an in-place quicksort with minimal additional storage, but this sort is not stable and is much harder to modify to suit our needs.

![Diagram of quicksort algorithm](image)

Figure E.3: Illustration of the quick sort with 8 elements. Elements in grey are the pivots chosen for the next step; they are highlighted in black in the next step as they rest in their final locations.
Algorithm 3 The quicksort algorithm

1: function Sort(y)  
2:     N ← length(y)  
3:     if N ≤ 1 then  
4:         return y  
5:     else  
6:         m ← a random integer in 1:N  
7:     initialize k ← 1  
8:     initialize empty vectors y_1, y_2, y_m  
9:     while k ≤ N do  
10:        if y[k] < y[m] then push y[k] to the end of y_1; end if  
11:        if y[k] = y[m] then push y[k] to the end of y_m; end if  
12:        if y[k] > y[m] then push y[k] to the end of y_2; end if  
13:         k ← k + 1  
14:     end while  
15:     y_1 ← Sort(y_1)  
16:     y_2 ← Sort(y_2)  
17:     y ← concatenate y_1, y_m, y_2  
18:     return y  
19: end if  
20: end function

In our modification, a counter associated with each element is again introduced. During the scanning of an input vector and placement of its elements into an output vector, we keep track of the number of elements whose values are smaller than the pivot, denoted as b. When an element larger than the pivot is scanned, we increment the associated counter by b; when the pivot itself is scanned, we increment its counter by b and then increase b by 1, so that subsequent elements that are larger than the pivot will count the pivot as a smaller preceding number. This ensures that no increment is missed and that comparisons leading to the increment of a counter will not be made again at a later stage of the sorting. An example of the modified quicksort in action is shown in Figure E.4; the pseudocode is given in Algorithm 4. As with the modified merge sort, 1 is to be added to every element of the counter after the whole procedure, and then the count be divided by N to obtain the empirical distribution function at (Y_1, Y_2). The same comment on the merge sort in the case of ties also applies here, as this version of the quicksort is stable and preserves the order of tied observations.
Algorithm 4 The modified quicksort algorithm for sequential ranks. The modification consists of lines 2, 11, 16, 20 and 24.

1: Input is the matrix $(Y_{11}, Y_{12}), \ldots, (Y_{N1}, Y_{N2})$, where $Y_1 = Y_{11}, \ldots, Y_{N1}$ are sorted and $Y_2 = Y_{12}, \ldots, Y_{N2}$ are paired accordingly. Convert $Y_1$ and $Y_2$ to the marginal ranks $1, \ldots, N$. In the following, the vector $Y_2$ is fed to $y$.

2: initialize global variable $z$ a vector of zeros of the same length as $y$
3: (comment: $z$ will be modified by the following function)
4: function Sort($y$) $\triangleright$ Recursive function
5: $N \leftarrow$ length($y$)
6: if $N \leq 1$ then
7: return $y$
8: else
9: $m \leftarrow$ a random integer in $1:N$
10: initialize $k \leftarrow 1$
11: initialize $b \leftarrow 0$
12: initialize empty vectors $y_1, y_2, y_m$
13: while $k \leq N$ do
14: if $y[k] < y[m]$ then
15: push $y[k]$ to the end of $y_1$
16: $b \leftarrow b + 1$
17: end if
18: if $y[k] = y[m]$ then
19: push $y[k]$ to the end of $y_m$
20: $z[y[k]] \leftarrow z[y[k]] + b; b \leftarrow b + 1$
21: end if
22: if $y[k] > y[m]$ then
23: push $y[k]$ to the end of $y_2$
24: $z[y[k]] \leftarrow z[y[k]] + b$
25: end if
26: $k \leftarrow k + 1$
27: end while
28: $y_1 \leftarrow$ Sort($y_1$)
29: $y_2 \leftarrow$ Sort($y_2$)
30: $y \leftarrow$ concatenate $y_1, y_m, y_2$
31: return $y$
32: end if
33: end function

34: Increment every element of $z$ by 1, and then divide by $N$ to obtain the empirical distribution function at the $N$ observations; the $r$th element of $z$ corresponds to the $s$th observation where $Y_{s2} = r$. 

246
Figure E.4: Illustration of the modified quick sort with 8 elements. Elements in grey are the pivots chosen for the next step; they are highlighted in black in the next step as they rest in their final locations. The smaller digit at the top right corner of each box is the counter for that scan operation. The smaller digit at the bottom left corner is the order of scanning for that element, and is used to assist in counting the increment for the pivot and elements to the right of it. The row for “total” indicates the number of smaller elements to the left of each $Y_{i2}$; the bivariate empirical distribution function at $(Y_{i1}, Y_{i2})$ is obtained by adding 1 to the “total” and then dividing the number by 8.

### E.3 Some comments on the algorithms

Because the two modified algorithms do not add any loops or layers of sorting within the existing algorithms, they both have a computational complexity of $O(N \log_2 N)$. The difference in computational time between these algorithms and the naive method with $O(N^2)$ complexity is not large when $N$ is small (such as in the hundreds), mainly because the overhead associated with them dominates algorithmic complexity. However, as $N$ increases, this becomes very noticeable as doubling the sample size quadruples the running time of the
naive method on average\textsuperscript{30}, while those of the modified algorithms are only slightly more than doubled\textsuperscript{31}. For the sample size needed in the MC simulation, the modified algorithms become a necessity.

In terms of the actual running time, we do not observe much difference between the two modified algorithms. Nevertheless, we suggest the use of the modified merge sort as each merge step reduces the number of subsequences by almost exactly one half, making the total number of merges constant for a particular sample size. For the modified quicksort algorithm, the random nature of the pivot introduces uncertainty in the number of splits required, and thus the running time is more variable. The modified merge sort is also easier to understand conceptually; by adding more layers of indices associated with each element, it is possible to extend this method to evaluate empirical distribution functions of higher dimensions $d$ at a complexity of $O\left(N \left(\log_2 N\right)^{d-1}\right)$, and is asymptotically better than the $O((d-1)N^2)$ naive algorithm for any given $d$.

\textsuperscript{30}Strictly speaking, the $O(\cdot)$ notation only stipulates the upper bound of the asymptotic behaviour of the algorithm. However, in this case the average complexity is also of the same order.

\textsuperscript{31}Note that $(2N \log_2(2N))/(N \log_2 N) \to 2$ as $N \to \infty$. 