AN ABSTRACT OF THE DISSERTATION OF

Yan Fang for the degree of Doctor of Philosophy in Statistics presented on May 1, 2012.

Title: Extensions to Gaussian Copula Models

Abstract approved: _____

Lisa J. Madsen Alix I. Gitelman

A copula is the representation of a multivariate distribution. Copulas are used to model multivariate data in many fields. Recent developments include copula models for spatial data and for discrete marginals. We will present a new methodological approach for modeling discrete spatial processes and for predicting the process at unobserved locations. We employ Bayesian methodology for both estimation and prediction. Comparisons between the new method and Generalized Additive Model (GAM) are done to test the performance of the prediction.

Although there exists a large variety of copula functions, only a few are practically manageable and in certain problems one would like to choose the Gaussian copula to model the dependence. Furthermore, most copulas are exchangeable, thus implying symmetric dependence. However, none of them is flexible enough to catch the tailed (upper tailed or lower tailed) distribution as well as elliptical distributions. An elliptical copula is the copula corresponding to an elliptical distribution by Sklar's theorem, so it can be used appropriately and effectively only to fit elliptical distributions. While in reality, data may be better described by a "fat-tailed" or "tailed" copula than by an elliptical copula. This dissertation proposes a novel pseudo-copula (the modified Gaussian pseudo-copula) based on the Gaussian copula to model dependencies in multivariate data. Our modified Gaussian pseudo-copula differs from the standard Gaussian copula in that it can model the tail dependence. The modified Gaussian pseudo-copula captures properties from both

elliptical copulas and Archimedean copulas. The modified Gaussian pseudo-copula and its properties are described. We focus on issues related to the dependence of extreme values. We give our pseudo-copula characteristics in the bivariate case, which can be extended to multivariate cases easily. The proposed pseudo-copula is assessed by estimating the measure of association from two real data sets, one from finance and one from insurance. A simulation study is done to test the goodness-of-fit of this new model.

Extensions to Gaussian Copula Models

by

Yan Fang

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented May 1, 2012
Commencement June 2012

Doctor of Philosophy dissertation of Yan Fang presented on May 1, 2012

APPROVED:

Co-Major Professor, representing Statistics

Co-Major Professor, representing Statistics

Chair of the Department of Statistics

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Yan Fang, Author

ACKNOWLEDGEMENTS

Academic

I would like to take this opportunity to express my thanks to those who helped me in one way or another on conducting research and the writing of this dissertation.

I want to express my deep appreciation and sincere thanks to Lisa Madsen for her guidance, understanding, patience, and most importantly, her friendship during my graduate studies at Oregon State University. The work presented in this dissertation is not possible without her support and advice. I am also thankful to my co-advisor, Alix Gitelman. I have benefited greatly from her creative ideas and perspectives. I also want to thank the useful discussion and inspiration from Lan Xue.

The members of my dissertation committee, Lisa Madsen, Alix Gitelman, Paul Murtaugh, Lan Xue, and Kevin Boston, have generously given their time and expertise to better my work. I thank them for their contribution and their good-natured support.

I gratefully acknowledge the help from many members of the OSU Statistics department as well as from other institutions. Particularly, I am indebted to Dr. Dave Birkes for his generous helps in my statistics research.

Personal

I cannot finish without saying thanks to my family, my son, my friend, Amber's family, Lisa's family and Marci Hinde, who are always my strong support in all these years. Special thanks go to my aunt and my brother, who have never lost faith in me.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF APPENDIX FIGURES

LIST OF APPENDIX TABLES

EXTENSIONS TO GAUSSIAN COPULA MODELS

1 INTRODUCTION

This dissertation forms the steps towards a fuller understanding of the copula. A construction of multivariate distributions that allows for a wide range in the choice of marginals and the type of dependence is based on copula functions. A copula is, briefly, the joint distribution function of variables which are individually uniformly distributed on $[0, 1]$. Alternative, a copula is a function which joins or "couples" a multivariate distribution function to its one-dimensional marginal distribution functions (Nelsen, 2006). Copulas are of interest to statisticians for two main reasons: first, as a way of studying scale-invariant measure of dependence, i.e, they remain unchanged under strictly increasing transformations; second, as a starting point for constructing families of joint distributions, sometimes with a view to simulation.

The importance of copulas in modeling the distribution of a multivariate random variable is justified by Sklar's Theorem (Sklar, 1959), which states: any multivariate distribution can be expressed as its copula function evaluated at its marginals; and any copula function when evaluated at any marginals is a multivariate distribution. In essence, copulas isolate the marginals from the dependence structure. Therefore, copulas can split the multivariate distribution of a random vector into the univariate marginals and the dependence structure, which allows us to easily model and estimate the distribution of random vectors by estimating marginals and copula separately. This provides substantial flexibility in developing dependence models among random variables.

In recent year, copula models have become an increasingly popular tool for modeling

dependence between random variables. The concept of copulas has been proven to be the most general and sophisticated concept of describing and modeling association or dependence between the components of a random vector. For a detailed overview of copulas, please see Joe (1997) and Nelsen (2006). Practical applications of copula techniques are found in the fields such as finance (Cherubini et al. 2004; McNeil et al. 2005), hydrology (Genest et al. 2007), public health and medical (Wang and Wells 2000), and actuarial science (Frees and Valdez 1998; Klugman and Parsa 1999).

Currently, copula-based modeling is commonly used in financial and actuarial statistics, however, it is beginning to become popular in geostatistics. Because copulas can be used to describe the joint multivariate distribution corresponding to the variables associated with the points of the studied domain, copula-based spatial models are currently an attractive tool for spatial modeling (e.g., Sang and Gelfand, 2009; Madsen, 2009 and Kazianka and Pilz, 2010).

In general, copula-based models are used to characterize the dependence between variables. However, a choice of the functional form for the copula is an open question. The limitation of the copula approach is the lack of a recommended way of checking whether the dependency structure of a data set is appropriately modeled by the chosen copula. Consequently, goodness-of-fit techniques for copulas recently gained interest, see e.g. Genest and Rivest (1993), Breymann et al. (2003), Genest and Nešlehová (2007), Genest et al. (2009), Kojadinovic (2009), and references therein.

1.1 Motivation and Background

Dependence between random variables is of fundamental importance because it may imply essential statistical relations within the real world. Modeling of dependencies among multivariate outcomes is an important and challenging aspect in probability theory and

statistical science. For bivariate cases, the Pearson product-moment correlation coefficient is commonly used to assess the strength and direction of the dependence relationship. However, it is a measure of the correlation (linear dependence) between two variables. Practically, dependencies may not be linear, nor should we be restricted to bivariate data. When multivariate data sets are analyzed, the multivariate normal is the most commonly used model. The reasons for using this model are as follows: (1) It is closed under addition; (2) Uncorrelatedness is equivalent to independence; (3) Its marginals are also normal. Although it is quite straightforward to use the multivariate normal distribution, there are flaws with using this model in many applications. Firstly, it entails rigid assumptions on the marginal and joint behaviors of the variables, since it arises from linear transformations on independent normal random variables. A random vector is said to be multivariate normally distributed if every linear combination of its components has a univariate normal distribution. Otherwise, the true distribution is not the multivariate normal. Secondly, the tails of its marginal and joint distribution may be too thin for a special application. In particular, the multivariate normal distribution may assign too small probabilities to extreme outcomes. Thirdly, the distribution exhibits a strong form of elliptical symmetry, which can become problematic in many finance and insurance applications. The dependence in the lower-left quadrant or upper-right quadrant of a bivariate distribution (i.e., tail dependence) recently has been discussed in financial applications related to market or credit risk (e.g., Hauksson et al., 2001; Embrechts et al., 2003). The tail dependence, being informative on joint extreme realizations, is identical in the multivariate normal distribution by virtue of the radially symmetric shape of the elliptical distribution. However, it may be reasonable to model asymmetry in the dependence. For example, Ang and Chen (2002) formally test the asymmetry in the dependence across stock returns and demonstrate that the correlation between Fama-French portfolio returns and the market return is significantly larger when the market return is negative than when it is positive.

Generally, separating the specification of the marginals from the dependence structure of variables in turn helps with the task of modeling the dependence under a more realistic, non-normal assumption. Hence, it is completely irrelevant whether or not the random variables are normal. Since the empirical marginals can be used instead of their explicit analogues, it is not even necessary to know the exact distribution of the variables being modeled. Clearly, copulas provide a powerful tool to model dependence structure.

As we mentioned before, copula-based model have attracted attention in spatial statistics over the past few years. In general, spatial dependence is usually described using the variogram which is strongly influenced by the univariate distribution of the random field or stochastic process. However, empirical and theoretical variogram estimates are affected by the extreme outlying observations. Hence, the variogram is not only sensitive to the outlying observations but its estimation is also based on the Gaussian assumption which may not be realistic. How can we address these problems? Can we use another method to model the spatial dependence? From the definition of copula, we can easily see copula can separate the dependence structure from the univariate marginals. This separation of modeling of the spatial dependence from the marginal behavior is particularly important when it is known that the dependence structure and the marginal properties are affected by different exogenous variables, which can be easily modeled via the parametric copula approach by letting the copula parameter depend on one variable and the marginal distribution depend on another variable. Bardossy (2006) first borrowed the idea from copulas which describe the dependence structure without the influence of the marginal distribution and proposed the copula-based spatial modeling of isotropic random fields with continuous univariate marginal distributions. Moreover, copulas can model the dependence by describing the joint multivariate distribution, the spatial dependence structure hence can be modeled by copulas. As an alternative to traditional spatial modeling and interpolation more and more researchers consider the use of copulas func-

tions in geostatistics. Nevertheless, most of copula-based spatial modeling are currently based on continuous univariate marginals. However, not all spatial data are continuous and copula-based spatial models with discrete margins are known to be difficult for both estimation and prediction beyond the bivariate case. How can we extend copula-based spatial modeling to the non-continuous case?

Except for modeling the dependence, spatial prediction is another important component of geostatistics. Spatial prediction is the process of predicting values of a target quantity at unvisited locations. When applied to a whole study area, it is also referred to as spatial interpolation or mapping. Therefore, development of generic and robust spatial interpolation techniques has been of interest. The conventional geostatistical approach for interpolation is called kriging and is based on the estimated covariance structure. Since the properties of interpolations based on an estimated covariance structure are not well known, it is common in practice to ignore the effect of the uncertainty in the covariance structure on the subsequent predictions. How can we model this existing uncertainty in a copula-based setting? Which copula is the best choice for a spatial model?

Although there exists a large variety of copula functions (Joe, 1997; Nelsen, 2006), only a few are practically manageable and often the choice in dependence modeling falls on an elliptical copula. By far the most important special case of elliptical distributions include the Gaussian copula (Li, 2000) and its convenient Student's t extension (Embrechts et al., 2001; Fang and Fang, 2002; Demarta and McMeil, 2005). Thus, we use Gaussian copula as our copula-based spatial model. What is Gaussian copula? The bivariate normal copula for bivariate (u, v) is

$$\mathrm{C}(u, v; \Sigma) = \Phi_{\Sigma}\Big(\Phi^{-1}(u), \Phi^{-1}(v)\Big), \tag{1.1}$$

where Φ_{Σ} denotes the joint distribution function of the bivariate standard normal distribution with linear correlation matrix Σ, and Φ^{-1} denotes the inverse of the distribution function of the univariate standard normal distribution. Copulas of the above form are

called Gaussian copulas. This definition can be easily extend to the n-variate case. From definition (1.1), we can see the Gaussian copula is derived from the multivariate Gaussian or normal distribution. The key advantage of Gaussian copula is that one can specify different levels of correlation between the marginals.

Because of its similarity to the Gaussian geostatistical model, Gaussian copula-based spatial model provides a convenient way to describe a relationship. However, this model is implausible for many geostatistics data sets. According to the definition of Gaussian copulas, they do not have tailed dependence, so they belong to elliptical copulas. While convenient and intuitive, elliptical copulas have a number of obvious shortcomings as a model for the real world. For example, elliptical copulas entirely lack tail dependence (see Bradley and Taqqu, 2003) and so may be inappropriate in data with extreme values. To overcome these limitations, a special class of copulas called Archimedean (for example Clayton, Frank, or Gumbel) are introduced (see Joe, 1997; Nelsen, 2006 for a review). Although Archimedean copulas are calculated over a closed-form solution and play an important role in extreme value theory, it is difficult to extend them to multivariate applications beyond two dimensions (Rachev et al., 2009). Furthermore, no matter what the dimension of the random vector is, the Archimedean family depends on only one or two dependence parameters (Hu and Kercheval, 2007). Thus, modeling pairwise dependence is limited. How should we make up the deficiency from the existing copulas?

When a copula is used to model dependence between random variables there is an immediate and obvious need to test whether the model can actually describe the data at hand accurately enough. Which copula is the right one? Among the different procedures proposed, one of the powerful tests, i.e., the multiplier goodness-of-fit test (Kojadinovic et al., 2011), is based on the empirical process comparing the empirical copula with a parametric estimate of the copula derived under the null hypothesis. In order to use this method validly and efficiently, the sample size needs to be at least 300.

1.2 Outline and Summary

There are two main contributions of this dissertation. The first contribution is to further improve copula-based spatial modeling with count marginals and to present a new methodological approach for modeling discrete spatial processes as well as predicting unobserved data. We adopt a continuous extension (that is, we extend count-valued random variables to continuous random variables) for non-continuous marginals in this new method. We apply the Gaussian copula to model the uncertainty in the covariance structure and show how it can be incorporated in a Bayesian framework. In the absence of simple analytical expressions for the joint posterior distribution, an MCMC algorithm is used to obtain the posterior samples. The posterior predictive density is approximated by averaging the plug-in predictive densities. The posterior prediction is approximated by using the median from the plug-in prediction. In order to assess the prediction performance, we compare the prediction results with those obtained by using a Generalized Additive Model (GAM).

The second main contribution aims at providing the tools for going one step further for modeling dependence: what would be the formalized dependence in the real world? Is there a way of understanding normal (i.e. Gaussian) dependence, and how can we construct models which allow to go beyond normal dependence? How can we set up a copula model which is flexible enough to capture both tailed and non-tailed dependence? To address these problems, we need to find a new model to make up for the deficiency of Gaussian copula. The Gaussian copula has a closed form probability density function and is constructed by projecting a multivariate normal distribution, so it is straightforward to estimate the unknown parameters by using maximum likelihood. Therefore, in order to overcome the aforementioned lack of flexibility of Gaussian copula, we extend the standard Gaussian copula model yet preserve tractability and computational efficiency in proposing

a new model. Hereafter we call this new model the modified Gaussian pseudo-copula. This is also one of the key motivations for this dissertation. A pseudo-copula satisfies a similar version of Sklar's theorem (Sklar, 1959), and all copulas are also pseudo-copulas. Our modified Gaussian pseudo-copula will provide properties which make up the deficiency in both elliptical and Archimedean copulas. It also effectively allows the correlations to be related to random variables as well as unknown parameters. In other words, the definition for the pairwise dependence in the modified Gaussian pseudo-copula will not only depend on unknown dependence parameters but also on the variables themselves. We supplement our analysis with a real data analysis. In addition we compare the power of the approach at distinguishing tail heaviness and skewness properties. Results show the flexibility of the new model. All of these will be presented in Chapter 5.

For simplicity, this dissertation will consider the bivariate case for the modified Gaussian pseudo-copula. It will be shown that the modified Gaussian pseudo-copula is more flexible than both Gaussian and Archimedean copulas, and that the fit of the modified Gaussian pseudo-copula to a real set is superior to the Gaussian copula and the three Archimedean copulas, i.e., Clayton, Frank and Gumbel copulas.

We would also like to understand better the extension from the discrete to continuous marginals under copula analysis, the theory on which the continuous extension is based, the behavior of multiplier goodness-of-fit tests in copula models, and the pros and cons of fitting existing copula models. As mentioned above, the main goal is the analysis of the statistical properties of the modified Gaussian pseudo-copula. We will address the properties of this new copula and illustrate the theoretical results with applications to both finance and insurance data. Further, several simulation studies are carried out to investigate the performance of the proposed copula model.

In this dissertation, all theoretical results on the modeling and measuring of association between several random variables use the concept of copulas. We consider Kendall's

τ and Spearman's ρ as the measures of association which depend on the underlying copula of the random vector only and are invariant with respect to the marginals.

As a direct function of the copula, nonparametric estimators for those measures are obtained based on the empirical copula (Deheuvels, 1979), which is derived from the multivariate empirical distribution function. When we conduct a multiplier goodness-of-fit test for the underlying copula, we compare the underlying copula with the empirical copula. If we refer to the aforementioned, we see there is the sample size limitation for multiplier goodness-of-fit test. Accordingly, we introduce the use of Akaike Information Criterion (AIC) instead of multiplier goodness-of-fit test to choose the best copula model. Certainly, AIC can not be used to do goodness-of-fit testing, but it can be used to choose the best copula model from a series of candidate families.

1.3 Detailed Outline of the Dissertation

Chapter 2 deals with the statistical modeling; estimation of regression parameters and multivariate association; and prediction of missing data for spatially dependent count data. We start by describing several properties of association in spatial data by introducing spatial correlation in Section 2.4.2. The concept of the Gaussian copula is briefly introduced in subsection 2.4.4. Since there is no unique copula form for discrete marginals, we discuss the continuous extension for count variables. We further give a brief introduction to the likelihood function of continuously extended count variables. After introducing the likelihood function, We discuss the Bayesian method used to estimate unknown parameters as well as to predict missing data in Section 2.5. In Section 2.6 we illustrate the applicability of the Bayesian prediction model via MCMC to an ecological data set. Finally, a simulation study is done to evaluate the performance of the Bayesian prediction for count data.

In Chapter 3, we give the description for both the starred copula introduced by Denuit and Lambert (2005) and the standard extension copula introduced by Schweizer and Sklar (1974) for discrete variables in Section 3.2. Both of them use bilinear interpolation to obtain a unique copula value with discrete marginals. Followed by the introduction of truncated normal distribution in Section 3.3, we derive the connection between the Gaussian copula model for binary data of Song (2000) and of Madsen and Fang (2011). Section 3.4 deals with the measure of association, Spearman's rho. Several schemes for the measure of the rank are discussed.

Since the central motivation for using copulas is to avoid rigidity in the choice of marginals and the specification of correlation structure, it is worthwhile to examine how well copula-based models fits the given data. Recently, Kojadinovic et al. (2011) suggested use of a multiplier procedure goodness-of-fit test to how well copula model fit the data. This method is based on multiplier central limit theorems and is a valid and much faster alternative to the usual parametric bootstrap-based procedure which tests that a given copula belongs to a parameterized copula family and whose validity was recently shown by Genest (2008). Although AIC cannot be used to do a goodness-of-fit hypothesis test, it can be used to choose the best copula model from a series of candidate families. In Chapter 4, a simulation study is done to compare the efficiency and accuracy of using the multiplier procedure and AIC approach to choose the true copula. After providing relevant definitions and background material in Section 4.3, the goodness-of-fit test derived from the multiplier method based the empirical process and an AIC criterion for choosing the true copula are introduced in Section 4.4. Finally, simulation results are provided in Section 4.6 to compare the multiplier procedure and AIC approach.

Chapter 5 is devoted to the statistical analysis of the new copula model, i.e. the modified Gaussian pseudo-copula. After a short motivation given in Section 5.2, an introduction to the basic theory of the modified Gaussian pseudo-copula is given in Section

5.3. We present five model functions for the association between two random variables for this new pseudo-copula. A special case of the new pseudo-copula is introduced in subsection 5.3.2 and several properties of pseudo-copulas for bivariate data are presented in subsection 5.3.3. In Section 5.4, we discuss the estimation for unknown parameters based on the pseudo-likelihood function. After providing the relevant definitions of Kendall's τ in subsection 5.4.2, we discuss the nonparametric estimation of Kendall's τ based on the empirical observations. We further give a brief introduction to the simulation used to approximate the sample version of Kendall's τ. In particular, the statistical test procedure designed to analyze the statistical properties of the copula and to detect goodness-of-fit of the copula for a given data set is developed in Section 5.5. The theoretical findings are applied to a real data set in Section 5.6. A simulation study is carried out to investigate the flexibility of the new pseudo-copula by comparing to the existing copula families in Section 5.7.

Finally, in Chapter 6, we provide the overall conclusions to the dissertation and discuss the future extensions to our work.

2 HIERARCHICAL SPATIAL MODELING FOR BOTH PREDICTING THE MISSING COUNT DATA AND ESTIMATING THE REGRESSION COEFFICIENT

Yan Fang*, Lisa Madsen** and Alix Gitelman***

Department of Statistics, Oregon State University, Corvallis, Oregon 97331, U.S.A.

*E-mail: fangya@science.oregonstate.edu

**http://www.stat.oregonstate.edu/people/lmadsen

***http://www.stat.oregonstate.edu/people/gitelman

2.1 Abstract

An intriguing problem in both ecology and natural field studies is predicting missing data at unobserved locations from discrete and spatially dependent data observed at other locations. Estimating regression parameters from these data is a related problem. For prediction, nonparametric additive model and back-fitting procedures such as Generalized Additive Model (GAM) are often used. Nevertheless, over-fitting and interpretability are sometimes problematic with nonparametric regression based on kernel and smoothing-spline estimates. On the other hand, Maximum Likelihood (ML) can be used to estimate regression parameters for discrete and spatially correlated variables, yet it lacks the mechanism to predict missing responses. This paper presents a framework for both estimating the marginal likelihood and predicting observations for spatially correlated counts data by using a Bayesian method with Markov Chain Monte Carlo (MCMC) simulation. A spatial correlation structure is incorporated in the Bayesian settings by assuming an adequate prior distribution for spatial random effects. This new method is applied to both real and simulated data to illustrate its advantages. Simulations demonstrate that the prediction by MCMC outperforms GAM, especially for large sample sizes

or the case where the missing and available data have strong dependence. In addition, we show that MCMC often yields a more accurate prediction for small count responses than does GAM.

2.2 Introduction

Madsen (2009) presented a Maximum Likelihood (ML) approach using a Gaussian copula model to analyze spatially dependent discrete data in the geostatistical framework. As the exact expected likelihood is difficult to derive in the model, an approximation is used. Generalized Estimating Equation (GEE) introduced by Liang and Zeger (1986) and Zeger and Liang (1986) can also be used to estimate unknown parameters from a discrete response variable. However, it is unsatisfactory for spatially correlated count data for several reasons discussed in Madsen (2009). Unfortunately, neither ML nor GEE can provide predictions of missing responses.

When research objectives include prediction, another procedure must be used such as non-parametric or semi-parametric additive models and the backfitting procedure (Hastie and Tibshirani, 1990). One example is the Generalized Additive Model (GAM) proposed by Hastie and Tibshirani (1984). However, GAM has its disadvantages: first, GAM cannot provide non-integer prediction for count response directly; second, GAM, being a nonparametric procedure, does not estimate regression parameters; third, GAM cannot estimate the association coefficient for spatially correlated data.

Instead of using the expected likelihood in Madsen (2009), we present a Bayesian algorithm by using Markov Chain Monte Carlo (MCMC) simulation to estimate unknown regression parameters as well as spatial association using the likelihood defined in Madsen (2009). Simultaneously, this method produces predictions at unobserved locations. Hereafter, we will refer to this method as the MCMC method.

The paper is organized as follows: In Section 2.3, we describe a Japanese beetle grub data set. Then we briefly review ML and GAM in Section 2.4. Immediately after that, the MCMC method is introduced in Section 2.5. In this section, we present the key components of the MCMC method, provide convergence diagnostics for the MCMC simulation and describe parameter estimates and how to predict missing count data. In Section 2.6 we give results for estimation and prediction with our MCMC method on the grub data set, and we compare these to the ML estimates. We also compare the predictions made using GAM and MCMC. Section 2.7 provides the analysis of a simulated data set.

2.3 Data

The Japanese beetle grub is a highly destructive plant pest of foreign origin. It was first found in the United States in a nursery in southern New Jersey in 1916. Grub dispersion patterns depend both on the locations of adult feeding aggregations and on soil properties (Dalthorp et al. 2000; Dalthorp 2004; Madsen, 2009). To study the spatial heterogeneity of grub counts, we model the grub counts collected on a golf course near Geneva, New York. More details about the data can be found in Dalthorp (2004). In this article, we restrict attention to the connection between grub counts and organic matter as well as the location determined by longitude and latitude. In addition, we predict missing grub counts based on existing data.

The grub data has 142 observations and includes the following variables: longitude, latitude, grub counts and soil organic matter. Figure 2.1 displays the grub counts (left) and soil organic matter (right). A relationship between the two measurements is notable high grub counts generally occur with somewhat lower organic matter levels.

The grub counts are discrete values range from 0 to 6 and these counts are over-dispersed in the sense of have an inflated number of zeros. Madsen (2009) suggested the

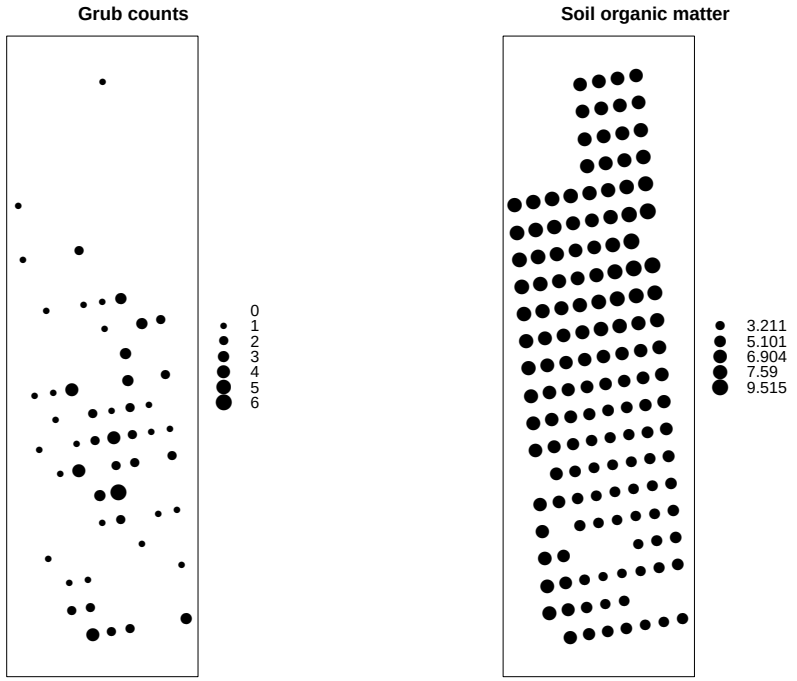


FIGURE 2.1: Views of Grub Count (left) and Soil Organic Matter (right).

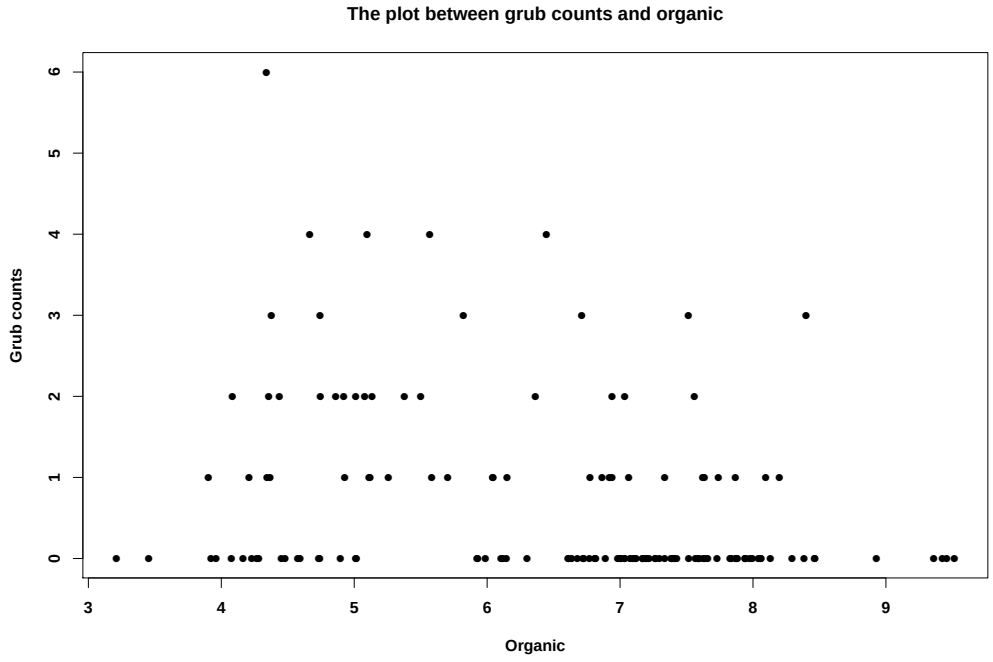


FIGURE 2.2: Scatter Plot between Grub Counts and Organic Matter.

negative binomial model for over-dispersed counts and used it to model ecological count data under different biological assumptions (Solomon, 1983). Furthermore, she suggested to use a generalized linear model to estimate regression parameters. We adapt this model to the Bayesian setting to allow estimation and prediction.

2.4 Existing Methodology

2.4.1 Over-dispersed counts

The negative binomial probability mass function of the i^{th} observation, Y_i, is:

$$p(y_i, \phi, \mu_i) = \frac{\mathbf{\Gamma}(y_i + \phi\mu_i)}{y_i!\mathbf{\Gamma}(\phi\mu_i)} \cdot \frac{\phi^{\phi\mu_i}}{(1+\phi)^{y_i+\phi\mu_i}},$$

where $\mathbf{\Gamma}(\cdot)$ is the gamma function, μ_i is the marginal mean, and ϕ is the "over-dispersion" parameter, i.e.,

$$\text{var}(Y_i) = \mu_i \cdot \frac{1+\phi}{\phi}.$$

These means are modeled as

$$\mu_i = \exp(\mathbf{X}_i^T \boldsymbol{\beta}),$$

where $\mathbf{X}_i = (1, X_{i1}, \ldots, X_{ip})^T$ is the vector of observed explanatory variables at the i^{th} location, and $\boldsymbol{\beta}$ is the vector of unknown regression parameters.

2.4.2 Spatial Correlation

Since each observation is associated with a location, we must model the spatial correlation. Assume $\boldsymbol{Y}(\cdot)$ is an isotropic random spatial process observed at location $\boldsymbol{s} = (s_1, \ldots, s_n)^T$, then the multivariate Gaussian copula can be used to model the joint distribution of $(Y_1, \ldots, Y_n) = [Y(s_1), \ldots, Y(s_n)]$ by giving the copula correlation matrix $\mathbf{\Sigma}$ a spatial form (see Madsen, 2009). Let $\rho(h)$ be an isotropic parametric correlogram (Cressie, 1993, page 67) depending on a vector of parameters $\mathbf{\Theta} = (\theta_0, \theta_1)^T$ and a distance

h. Since the exponential has the simple form while still being a valid variogram in all dimensions, the correlogram $\rho(h)$ is assumed to be exponential:

$$\rho(h) = \begin{cases} \theta_0 \exp(-h\theta_1), & h \neq 0 \\ 1, & h = 0, \end{cases} \tag{2.1}$$

where h is the distance between two locations, θ_0 is the "nugget" parameter ranging between 0 to 1, and θ_1 is the "decay" parameter.

The lower bound for θ_1 is 0, while a paractical upper bound for θ_1 is decided by the effective range R, i. e., the distance at which the correlation drops to 0.05. The formula used to get the uppper bound $(\theta_1)_{\max}$ (Cressie, 1993) is :

$$(\theta_1)_{\max} = \frac{-\log(0.05)}{\min(h)} \tag{2.2}$$

where $\min(h)$ is the minimum pairwise distance.

Accordingly, the ij^{th} entry of $n \times n$ correlation matrix, $\Sigma(\mathbf{\Theta})$, from the multivariate Gaussian copula is defined as

$$\Sigma_{ij}(\mathbf{\Theta}) = \rho(h_{ij}),$$

where h_{ij} is the distance between location s_i and s_j and $i, j \in 1, \ldots, n$. Since each entry for Σ has a spatial form, this copula models the spatial dependence among the variables.

2.4.3 Generalized Additive Models

Generalized additive models (GAM) proposed by Hastie and Tibshirani (1984) are useful in finding predictor-response relationships in many kinds of data without using a specific parametric model. These models assume that the mean of the dependent variable depends on an additive predictor through a nonlinear link function. That is, GAM can be described as

$$g\Big(E[Y]\Big) = \alpha + s_1(X_{1i}) + \ldots + s_p(X_{pi}), \tag{2.3}$$

where $g(\cdot)$ is the link function that relates the linear predictor with the expected value of the response variable Y, X_{ji} are explanatory variables, and $s_j(\cdot)$ is an unspecified

smooth function, $i = 1, \ldots, n$ and $j = 1, \ldots, p$. This is done through iterative smoothing operations and allows for various non-linear effects of the explanatory variables. GAM predictions are obtained from a fitted generalized additive model object (see, e.g., Hastie and Tibshirani, 1990 and Wood, 2006b). The statistical software R and the *mgcv* package can be used to fit GAM to data and to produce predictions given a new set of values for the model covariates or the original data used for the model fit.

2.4.4 Maximum Likelihood Models

The ML approach introduced by Madsen (2009) is an extension of the Gaussian copula that uses a spatial form, i.e., an isotropic, parametric correlogram defined in Equation (2.1) as the copula correlation matrix. To obtain a unique copula function, Madsen (2009) applied the continuous extension proposed by Denuit and Lambert (2005) when the response variables are discrete.

According to Denuit and Lambert (2005), assume Y is a count variable. Then associated with Y, a continuous random variable is defined as

$$Y^* = Y - U, \tag{2.4}$$

where U follows a continuous uniform distribution (0,1) independent of Y. Then Y^* is a continuous random variable with the distribution function with

$$\mathrm{F}^*(y) = F([y]) + (y - [y]) \times P(Y = [y + 1]), \tag{2.5}$$

and the density function

$$\mathrm{f}^*(y) = P(Y = [y + 1]), \tag{2.6}$$

where $[y]$ denotes the integer part of $y \in \mathbb{R}$.

Given discrete variables $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$ with cumulative distribution function (c.d.f) $\mathrm{F} = (\mathrm{F}_1, \ldots, \mathrm{F}_n)^T$ and density function $\mathrm{f} = (\mathrm{f}_1, \ldots, \mathrm{f}_n)^T$, where $\mathrm{F}_i = \mathrm{P}(Y_i \leq y_i)$ and $\mathrm{f}_i = \mathrm{P}(Y_i = y_i)$ with $\mathrm{P}(Y_i = y_i)$ defined in Section 2.4.1 for $i = 1, \ldots, n$. By using

the continuous extension, discrete variables are transformed to the continuous variables $\mathbf{Y}^* = (Y_1 - U_1, \ldots, Y_n - U_n)^T = (Y_1^*, \ldots, Y_n^*)^T$, where U_i is independent of Y_i and of U_j for $j \neq i$. Denuit and Lambert (2005) proved that this continuous extension preserves Kendall's τ, thus variables $\mathbf{Y}^*$ and $\mathbf{Y}$ have the same dependence relationship. Dependence among $\mathbf{Y}^* = (Y_1^*, \ldots, Y_n^*)^T$ can be modeled via a Gaussian copula

$$\mathrm{C}(y_1^*, \ldots, y_n^*; \mathbf{\Sigma}) = \Phi_\Sigma[\Phi^{-1}\{\mathrm{F}_1^*(y_1^*)\}, \ldots, \Phi^{-1}\{\mathrm{F}_n^*(y_n^*)\}], \tag{2.7}$$

where $\mathbf{\Phi_\Sigma}(\cdot)$ is the multivariate normal c.d.f with correlation matrix $\mathbf{\Sigma}$ defined in Equation (2.1), $\Phi^{-1}(\cdot)$ is the univariate normal c.d.f, and $\mathrm{F}_i^*(y_i^*) = F_i([y_i^*]) + (y_i^* - [y_i^*])P(Y_i = [y_i^* + 1])$. And the joint density function can be derived by differentiating $\mathrm{C}(y_1^*, \ldots, y_n^*; \mathbf{\Sigma})$, i. e.,

$$\mathrm{c}(y_1^*, \ldots, y_n^*; \mathbf{\Sigma}) = |\mathbf{\Sigma}|^{-1/2} \exp\left\{ -\frac{1}{2}(z^*)^T(\mathbf{\Sigma}^{-1} - \boldsymbol{I}_n)\boldsymbol{z}^* \right\} \prod_{i=1}^{n} \mathrm{f}_i^*(y_i^*), \tag{2.8}$$

where $\boldsymbol{z}^* = \left\{ \Phi^{-1}\left[\mathrm{F}_1^*(y_1^*)\right], \ldots, \Phi^{-1}\left[\mathrm{F}_n^*(y_n^*)\right] \right\}$, $\mathrm{f}_i^*(y_i^*) = P(Y_i = y_i)$, and $\boldsymbol{I}_n$ denotes the $n \times n$ identity matrix. In other words, the joint density function of $(y_1, \ldots, y_n)$ can be denoted as

$$\mathrm{c}(y_1, \ldots, y_n; \mathbf{\Sigma}) = |\mathbf{\Sigma}|^{-1/2} \exp\left\{ -\frac{1}{2}(z^*)^T(\mathbf{\Sigma}^{-1} - \boldsymbol{I}_n)\boldsymbol{z}^* \right\} \prod_{i=1}^{n} P(Y_i = y_i). \tag{2.9}$$

We can think of variable $\boldsymbol{U}$ included in both Equation (2.7) and Equation (2.9) as jitter parameters. ML proposed by Madsen (2009) used the expected joint density function with respect to $\boldsymbol{U}$ instead of the joint density function, i.e.,

$$\mathbf{c}(\boldsymbol{y}; \Sigma) = \boldsymbol{E}_{\boldsymbol{U}} \Big[|\mathbf{\Sigma}|^{-1/2} \exp\big\{ -\frac{1}{2}(z^*)^T(\mathbf{\Sigma}^{-1} - \boldsymbol{I}_n)\boldsymbol{z}^* \big\} \prod_{i=1}^{n} P(Y_i = y_i) \Big]. \tag{2.10}$$

Accordingly, unknown parameter vectors were estimated by maximizing $\log \mathbf{c}(\boldsymbol{y}; \Sigma)$. For those interested in more detail, please refer to Madsen (2009).

2.5 Bayesian Implementation

Instead of maximizing the log of the expected joint density function as proposed by Madsen (2009), we use MCMC to draw simulations from the joint posterior distribution of the parameters.We obtain high posterior density (HPD) intervals for all parameters as well as for predictions.

2.5.1 Likelihood Function

The density in Equation (2.9) forms a likelihood function for $\boldsymbol{Y}$ given the correlation parameters $\boldsymbol{\theta} = (\theta_0, \theta_1)$, regression parameter $\boldsymbol{\beta}$, a scale parameter ϕ of the marginal densities $P(Y_i = y_i)$, and the jitter parameters U_i. The likelihood function is expressed as

$$\mathrm{c}(\boldsymbol{y}; \boldsymbol{\beta}, \theta_0, \theta_1, \phi, \boldsymbol{U}) = |\boldsymbol{\Sigma}|^{-1/2} \exp\big\{ -\frac{1}{2}(z^*)^T(\boldsymbol{\Sigma}^{-1} - \boldsymbol{I}_n)\boldsymbol{z}^*\big\} \prod_{i=1}^{n} P(Y_i = y_i). \qquad (2.11)$$

There is a possibility of numerical error in calculating $z_i^* = \Phi^{-1}\{\mathrm{F}_i^*(y_i^*)\}$ at some steps of the MCMC. We might encounter situations where $\mathrm{F}_i^*(y_i^*)$ is rounded to 0 or 1. Then the inverse of $\mathrm{F}_i^*(y_i^*)$ will give negative ∞ or positive ∞ for z_i^*. To prevent this, we restrict $10^{-6} \leq \ \Phi^{-1}\{\mathrm{F}_i^*(y_i^*)\} \leq 1 - 10^{-6}$ following Pitt et al. (2006), which ensures numerical stability and adequate accuracy.

2.5.2 Priors

Without any information about the parameters, we specify non-informative priors on all the parameters using independent prior distributions. Specifically, we use $N(\mu = 0, \sigma^2 = 10^4)$ priors for the regression coefficients $\beta_0, \ldots, \beta_p$, where p is the dimension of β; a uniform (0,1) prior for the θ_0; $Gamma(0.0001, 1000)$ prior for both the nugget parameter θ_1 and the over-dispersion parameter ϕ; and a uniform (0,1) prior for the jitter parameters, U_i, with $i = 1, \ldots, n$.

2.5.3 Full conditional distributions and the MCMC update

Using Bayes's theorem, the posterior distribution is given by

$$\begin{aligned}\pi(\boldsymbol{\beta},\theta_0,\theta_1,\phi,\boldsymbol{U}|\boldsymbol{Y}) &\propto c(\boldsymbol{Y}|\boldsymbol{\beta},\theta_0,\theta_1,\phi,\boldsymbol{U})\boldsymbol{\pi}(\boldsymbol{\beta},\theta_0,\theta_1,\phi,\boldsymbol{U}) \\ &= |\boldsymbol{\Sigma}|^{-1/2}\exp\left\{-\frac{1}{2}(z^*)^T(\boldsymbol{\Sigma}^{-1}-\boldsymbol{I}_n)\boldsymbol{z}^*\right\}\prod_{i=1}^{n}P(Y_i=y_i) \\ &\quad\times\boldsymbol{\pi}(\boldsymbol{\beta})\boldsymbol{\pi}(\theta_0)\boldsymbol{\pi}(\theta_1)\boldsymbol{\pi}(\phi)\pi(\boldsymbol{U}), \qquad (2.12)\end{aligned}$$

where $\pi(\boldsymbol{\beta})$, $\boldsymbol{\pi}(\theta_0)$, $\boldsymbol{\pi}(\theta_1)$, $\boldsymbol{\pi}(\phi)$ and $\pi(\boldsymbol{U})$ are the prior distributions for the parameters β, θ_0, θ_1, ϕ and $\boldsymbol{U}$ respectively.

The posterior distribution is so complicated that the direct computation is intractable. Therefore, we will use MCMC simulation to obtain a sample from posterior $\pi(\boldsymbol{\beta},\theta_0,\theta_1,\phi,\boldsymbol{U}|\boldsymbol{Y})$, which we can use to obtain the posterior summary statistics for parameters and predictions of interest. All parameters are updated using Metropolis-Hastings steps by drawing new values from specified proposal distributions, since the full conditional distributions of all the parameters and of $\boldsymbol{U}$ are of non-standard form (see, e.g., Brooks, 1998 and King and Brooks, 2002).

In particular, we will use normal proposal densities for $\boldsymbol{\beta}$ and $\log(\phi)$ and uniform proposal densities for θ_0, θ_1, and $\boldsymbol{U}$. In MCMC simulations each successive draw uses the information from previous draw. Let t denote the current draw. Then the parameters are updated by using the proposal densities as follows:

1. Updating the parameter $\boldsymbol{\beta}$: we propose a new value $\boldsymbol{\beta}^t$ such that $\boldsymbol{\beta}^t \sim N(\boldsymbol{\beta}^{t-1}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}})$, where $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ is the variance-covariance of $\boldsymbol{\beta}$;

2. Updating θ_0: we propose a new value θ_0^t such that $\theta_0^t \sim [L_0, U_0]$, where $L_0 = \max(0, \theta_0^{t-1} - \epsilon_0)$ and $U_0 = \min(1, \theta_0^{t-1} + \epsilon_0)$;

3. Updating θ_1: the new value of θ_1^t is drawn uniformly from $[L_1, U_1]$, where $L_1 =$

$\max(0, \theta_1^{t-1} - \epsilon_1)$ and $U_1 = \min\Big((\theta_1^{t-1})_{\max}, \theta_1^{t-1} + \epsilon_1\Big)$. The value of $(\theta_1^{t-1})_{\max}$ refers to Equation (2.2);

4. Updating ϕ: randomly draw $\log(\phi^*) \sim N\Big(\log(\phi^{t-1}), \sigma_\phi^2\Big)$, take its exponential to obtain ϕ^*, and $\phi^t = \min(\phi^*, 100)$;

5. Update $\boldsymbol{U}$: the new value of $\boldsymbol{U}^t$ is drawn uniformly from $[L_U, U_U]$, where $L_U = \max(0, U^{t-1} - \epsilon_U)$ and $U_U = \min(1, U^{t-1} + \epsilon_U)$.

TABLE 2.1: The proposal distribution for the parameter

$\boldsymbol{\Sigma_\beta}$	ϵ_0	ϵ_1	σ_ϕ^2	ϵ_U
$(\mathbb{X}^T\mathbb{X})^{-1}$	0.4	0.03	0.35	0.2

Here, $\boldsymbol{\Sigma_\beta}, \epsilon_0, \epsilon_1, \sigma_\phi$ and ϵ_U are the parameter values for the proposal densities. These values need to be tuned individually in order to find the desired acceptance rates (the fraction of candidate draws that are accepted in the Metropolis-Hastings algorithm) for proposed moves. The desired acceptance rate depends on the target distribution, however it has been shown theoretically that the ideal acceptance rate is approximate 50% (see Gelman et al., 1996). For grub data set, all of these values are obtained from a short pilot simulation. Table 2.1, where $\mathbb{X}$ is the design matrix and is defined as $\mathbb{X} = (\mathbf{1}, \mathrm{x}, \mathrm{x}^2, \mathrm{x}^3)^T$ with soil organic matter as the explanatory variables x, provides the values for $\boldsymbol{\Sigma_\beta}, \epsilon_0, \epsilon_1, \sigma_\phi$ and ϵ_U respectively. The code is available upon request.

2.5.4 Checking Convergence

Five independent chains are run by using different starting values. Each chain is run for 20,000 iterations, of which the first 10,000 are treated as pre-convergence burn-in and are discarded. In the example and simulation study, we checked the adequacy of the

burn-in period by using the slightly modified Gelman-Rubin Statistic (Monahan, 2001, page 371).

$$\sqrt{\hat{R}} = \sqrt{\frac{\hat{V}(\theta)}{W}}, \tag{2.13}$$

where $\hat{V}(\theta) = (1 - \frac{1}{N})W + \frac{1}{N}B$ is estimated variance; $W = \frac{1}{l(N-1)} \sum_{j=1}^{l} \sum_{i=1}^{N} (\theta_j^i - \bar{\theta}_j)^2$ is the within chain variance; $B = \frac{N}{l-1} \sum_{j=1}^{l} (\theta_j - \bar{\theta})^2$ is the between chain variance; and $l = 5$ is the number of chains; and $N = 10,000$ is the number of retained iterations.

Once convergence is reached, $\hat{V}(\theta)$ and W should be nearly equivalent since variation within a chain and between the chains should coincide, so $\hat{R}$ should be approximately equal to one. A rule of thumb is that values of $\hat{R}$ under 1.2 (Gilks et al., 1996, page 138) indicates convergence of the Markov Chain.

2.5.5 Parameter Estimates

Parameters are estimated using means of the samples from the posterior distribution. A method for finding a posterior credible interval is by constructing the set, which is defined as

$$\boldsymbol{C} = \{\boldsymbol{\theta} \in \boldsymbol{\Theta} : p(\boldsymbol{\theta}|\boldsymbol{y}) \geq k(\boldsymbol{\alpha})\},$$

where $k(\boldsymbol{\alpha})$ is the largest constant such that $p(\boldsymbol{C}|\boldsymbol{y}) \geq \boldsymbol{\alpha}$. Here α is chosen to indicate the posterior probability of the credible interval (Banerjee et al., 2004, page 104).

2.5.6 Prediction

For missing data or for making predictions at unobserved locations, we use Bayesian kriging (Banerjee et al., 2004). Suppose that we are given a set of new locations, $\boldsymbol{s}_{new}$, and their associated explanatory variables, $\boldsymbol{X}_{new}$. We wish to predict count data, $\boldsymbol{Y}_{new}$, for these locations. Let us use $\boldsymbol{Y}_{obs}$ to denote the observed responses and $\boldsymbol{X}_{obs}$ to denote the explanatory variables corresponding to the observed responses. Our first objective is

to obtain z_{new}, which is expressed as

$$\boldsymbol{z}_{new} = \Phi^{-1}\Big(F^*_{new}(\boldsymbol{Y}^*_{new})\Big), \tag{2.14}$$

where $\boldsymbol{Y}^*_{new}$ is the continuous extension of $\boldsymbol{Y}_{new}$ and $\boldsymbol{Y}_{new} = \{F^*_{new}\}^{-1}\Big(\Phi(\boldsymbol{z}_{new})\Big)$. With the Bayesian framework, the prediction of $\boldsymbol{z}_{new}$ at $\boldsymbol{s}_{new}$ follows from the posterior predictive distribution given by

$$\begin{aligned} f(\boldsymbol{z}_{new}|\boldsymbol{z}_{obs}, \boldsymbol{X}_{obs}, \boldsymbol{X}_{new}) &= \int f(\boldsymbol{z}_{new}, \boldsymbol{\Theta}|\boldsymbol{z}_{obs}, \boldsymbol{X}_{obs}, \boldsymbol{X}_{new}) d\boldsymbol{\Theta} \\ &= \int f(\boldsymbol{z}_{new}|\boldsymbol{\Theta}, \boldsymbol{z}_{obs}, \boldsymbol{X}_{obs}, \boldsymbol{X}_{new}) f(\boldsymbol{\Theta}|\boldsymbol{z}_{obs}, \boldsymbol{X}_{obs}) d\boldsymbol{\Theta}, \end{aligned} \tag{2.15}$$

where $f(\boldsymbol{z}_{new}|\boldsymbol{\Theta}, \boldsymbol{z}_{obs}, \boldsymbol{X}_{obs}, \boldsymbol{X}_{new})$ has a conditional normal distribution arising from the joint multivariate normal distribution of $\boldsymbol{z}_{new}$ and $\boldsymbol{z}_{obs}$ with $z_{obs} = \Phi^{-1}\Big(F^*_{obs}\big(\boldsymbol{Y}^*_{obs}\big)\Big)$ and $\boldsymbol{Y}^*_{obs}$ is the continuous extension of $\boldsymbol{Y}_{obs}$.

Given target distribution $\boldsymbol{\pi}(\boldsymbol{\Theta}|\boldsymbol{Y})$ from Equation (2.12),where $\boldsymbol{\Theta} = (\boldsymbol{\beta}, \theta_0, \theta_1, \phi, \boldsymbol{U})$, the Metropolis algorithm produces a sequence of random points $(\boldsymbol{\Theta}^{(1)}, \boldsymbol{\Theta}^{(2)}, \ldots)$, which have a distribution that converges to the target distribution. The prediction process in MCMC is as follows:

1. Draw the starting points $\boldsymbol{\Theta}^{(0)}$ from the prior distribution;

2. For $m = 1, 2, \ldots$;

   a. Use Metropolis-Hastings algorithm and $\boldsymbol{Y}_{obs}$ to obtain the current value $\boldsymbol{\Theta}^{(m)}$;

   b. Given the current parameters $(\theta_0^{(m)}, \theta_1^{(m)})$, the multivariate normal distribution of $\boldsymbol{z}^{(m)} = (\boldsymbol{z}^{(m)}_{new}, \boldsymbol{z}^{(m)}_{obs})$ is

$$\begin{pmatrix} \boldsymbol{z}^{(m)}_{new} \\ \boldsymbol{z}^{(m)}_{obs} \end{pmatrix} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}^{(m)}),$$

   where

$$\boldsymbol{\Sigma}^{(m)} = \begin{bmatrix} \boldsymbol{\Sigma}^{(m)}_{new} & \boldsymbol{\Sigma}^{(m)}_{new,obs} \\ \boldsymbol{\Sigma}^{(m)}_{obs,new} & \boldsymbol{\Sigma}^{(m)}_{obs} \end{bmatrix}$$

and ij^{th} entry of the matrix $\mathbf{\Sigma}^{(m)}$ is

$$\mathbf{\Sigma}_{ij}^{(m)} = \begin{cases} \theta_0^{(m)} \exp(-h_{ij}\theta_1^{(m)}), & i \neq j \\ 1, & i = j \end{cases} . \tag{2.16}$$

In equation (2.16), h_{ij} is the distance between the locations of y_i and y_j, $\theta_0^{(m)}$ is the "nugget" parameter ranging between 0 to 1, and $\theta_1^{(m)}$ is the "decay" parameter. Hence the distribution for $\boldsymbol{z}_{new}^{(m)}$ given the observed $\boldsymbol{z}_{obs}^{(m)}$ is

$$(\boldsymbol{z}_{new}^{(m)} | \boldsymbol{z}_{obs}^{(m)}) \sim N(\mathbf{\Sigma}_{new,obs}^{(m)}(\mathbf{\Sigma}_{obs}^{(m)})^{-1}\boldsymbol{z}_{obs}^{(m)}, \quad \mathbf{\Sigma}_{new}^{(m)} - \mathbf{\Sigma}_{new,obs}^{(m)}(\mathbf{\Sigma}_{obs}^{(m)})^{-1}\mathbf{\Sigma}_{obs,new}^{(m)}) \tag{2.17}$$

c. Randomly draw $\boldsymbol{z}_{new}^{(m)}$ from the conditional distribution function defined in Equation (2.17);

d. Invert $\boldsymbol{z}_{new}^{(m)}$ back to the c.d.f of $(\boldsymbol{Y}_{new}^{*})^{(m)}$ (i.e., $\Phi(\boldsymbol{z}_{new}^{(m)})$), referring back to Equation (2.14). And then we use the Negative Binomial c.d.f with parameters $(\boldsymbol{\beta}^{(m)}, \phi^{(m)})$ to obtain $\boldsymbol{Y}_{new}^{(m)}$.

In practice, the collection $(\boldsymbol{Y}_{new}^{(m+1)}, \boldsymbol{Y}_{new}^{(m+2)}, \ldots)$ after dropping the first m burn-in iterations is a sample from the posterior predictive density. Since $\boldsymbol{Y}_{new}$ is count-valued data, we use the median instead of the average as the predicted value to avoid a non-integer value. Moreover, we can construct the 95% prediction interval for the median value. For integer values, usually we cannot obtain an exact 95% prediction interval.

2.5.7 Comparison Between GAM and MCMC Method

When prediction is the main goal of the analyses, the mean squared prediction error (MSPE) is the main way in which prediction performance is measured. Consider the problem of predicting the value of a random variable using the observed values Y. Denote

the predictor as $\hat{Y}$. Then the MSPE is

$$\text{MSPE}(\hat{Y}) = \text{E}(Y - \hat{Y})^2 = \text{E}\left[(Y - \text{E}[\hat{Y}])^2\right] + \text{var}(\hat{Y}), \tag{2.18}$$

where the expectation is taken with respect to (w.r.t.) the joint distribution of Y and $\hat{Y}$. It is the sum the squared bias of the estimator and the variance. Obviously, MSPE($\hat{Y}$) is minimized at var($\hat{Y}$) = 0. And the sample MSPE is obtained by using the following formula:

$$\text{MSPE} = \frac{1}{k}\sum_{i=1}^{k}(y_i - \hat{y}_i)^2, \tag{2.19}$$

where k is the number of predicted data points, $\hat{y}_i$ is the predicted value, and y_i is the true value.

2.6 Results

The Japanese beetle grub data, introduced in Section 2.3, were analyzed using the MCMC method. Details of the implementation and results are given here. The model for the marginal mean is

$$\mu_i = \exp(\beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3), \tag{2.20}$$

where x_i is the observed percent soil organic matter content at the i^{th} location. Generalized Estimating Equation (GEE) analysis concludes that a cubic function of organic matter is necessary (see Madsen, 2009), so we use the cubic function to model the grub data. For more information about the model of the marginal mean, refer to Dalthorp (2004)) and/or Madsen (2009)).

In our Bayesian model, we not only estimate the regression coefficients but also predict new responses. Metropolis-Hastings Algorithm (see Appendix A) is used to accomplish these purposes effectively. In Metropolis-Hastings Algorithm, we use generalized linear model (GLM) to obtain the starting values for regression coefficient parameters β

and the "over-dispersion" parameter ϕ, and set the starting value for θ_0 as 0.5 (namely,the mid-point for the range of θ_0). While for parameter θ_1, we use $\frac{(\theta_1)_{\max}+(\theta_1)_{\min}}{2}$ as the starting value with$(\theta_1)_{\max} = \frac{-log\frac{0.05}{\theta_0}}{h_{\min}}$ and $(\theta_1)_{\min} = \frac{-log\frac{0.05}{\theta_0}}{h_{\max}}$, where $h_{\min}$ and $h_{\max}$ are the minimum and maximum of pairwise distances. The starting value for U is generated randomly from $U(0,1)$. The computation intensity of this estimation procedure increase with more algorithm iterations. And the computation burden is primarily in calculating $z_i^* = \Phi^{-1}\{\mathrm{F}_i^*(y_i^*)\}$ and their derivatives. On a 3.4 GHz desktop computer, the time with 12,000 iterations for estimation is about 6.5 hours and the time with 12,000 iterations for prediction with 44% missing is about 4 hours.

In the following, we summarize the output obtained by the MCMC algorithm to the grub data. The algorithm was run for 12,000 iterations with 6,000 burn-in. The Gelman-Rubin Statistic in Equation (2.13) of all the parameters is less than 1.2 and most are less than 1.05 (see Table B.1), meaning that we have a well-defined model and the iterations are sufficiently large enough to guarantee convergence. The same statistics are also checked for cases where some count data are held out as missing, and again, the chains converge.

TABLE 2.2: Estimation of the regression coefficients for grub dataset with both MCMC and ML, where ML quantities come from Madsen (2009)

Parameter	Generalized Regression Coefficient with 95% interval	
	MCMC	ML
β_0	-22.37 (-48.31, -2.32)	-24.34(-47.6, -1.08)
β_1	10.88(0.98, 23.20)	11.96(0.54, 23.38)
β_2	-1.65(-3.58, -0.04)	-1.84(-3.66, -0.02)
β_3	0.08(0, 0.17)	0.09(-0.01, 0.19)

Point estimates of the regression coefficients, $\boldsymbol{\beta}$, from both MCMC and ML (Madsen,

2009) are given in Table 2.2. The numbers in parenthesis indicate a 95% confidence interval for ML and a 95% HPD interval for MCMC. The 95% interval of MCMC is more narrow than that corresponding to ML. All the intervals from MCMC are "significant" in the sense that they do not include zero. On the other hand, not all the intervals from ML are significant since the interval for the coefficient of the cubic of soil organic matter includes zero. Madsen (2009) has already stated that a quadric function was sufficient. Although point estimates from MCMC have smaller absolute values than the estimates from ML, the fitted mean from both MCMC and ML are quite the same. Figure 2.3 shows the data with the fitted mean functions from MCMC estimation and ML estimation. The two curves are very similar; the average squared difference in the fitted values is $\frac{1}{142}\sum_{i=1}^{142}(\hat{y}_{MCMC} - \hat{y}_{ML})^2 = 0.00053$.

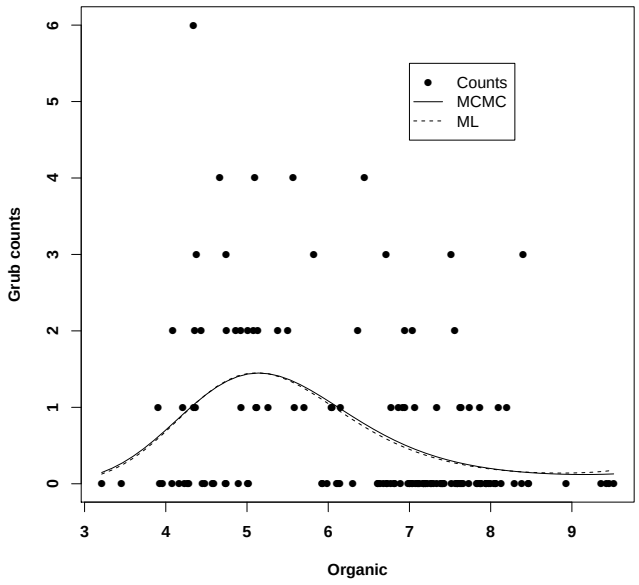


FIGURE 2.3: Plot of observed grub counts as a function of percent soil organic matter. Superimposed is the fitted mean function from both estimation procedure.

To explore the prediction problem, we randomly hold out 10%, 20% and 44% of the

data and use the remaining data to make prediction. Table 2.3 gives us the MSPE defined in Equation (2.19) between the predicted values and the true values for both MCMC and GAM.

TABLE 2.3: MSPE of MCMC and GAM, predicting 10%, 20% and 44% of missing data

	cross-validation (CV)			
		GAM		
Missing	MCMC	mean	round	median
10%	1.71	1.42	1.71	1.79
20%	1.18	1.01	1.21	1.32
44%	1.68	1.40	1.48	1.74

As shown in Table 2.3, if we just use the fitted value from GAM , the MSPE from MCMC is slightly larger than the MSPE from GAM. However, the GAM gives us a non-integer fitted value, while MCMC yields the count prediction value directly, so it is misleading to compare these two MSPEs directly. In order to get the integer prediction values, we explore two common rounding methods in GAM:

Method 1: Round the fitted value to the closest integer;

Method 2: Predict the median instead of the mean.

For simplicity, we will call method 1 and method 2 "GAM round method" and "GAM median method", respectively. Overall, MCMC and integer-valued predictions from GAM are quite similar. Table 2.3 shows that MCMC gives a more accurate prediction than GAM median method and it is similar to GAM round method. When we predict 10% of the data, the MSPE values from MCMC is the same as the value from GAM round method and smaller than the one from GAM median method. When we predict 20% of

the data, the MSPE from MCMC is smaller than the values from both two GAM rounding methods. While when we predict 44% of the data, the MSPE from MCMC is between the values from both two rounding methods. The reason may be that when the amount of the missing data is too large, we do not have enough information to get good estimates of the parameters, which leads to a large MSPE for MCMC. Nevertheless, MSPE from MCMC is always smaller than MSPE from GAM median method.

No matter the percent of the missing count data, about 50% of the counts are 0. Therefore, it seems sensible to divide the prediction of the missing count data into two groups, zero and non-zero, and then break down the comparison within these two groups. Table 2.4 gives MSPE separately from zero and non-zero.

TABLE 2.4: The decomposition of comparison between MCMC and GAM, for zero and non-zero missing data.

	Category 1: $Y = 0$				Category 2: $Y > 0$			
		GAM				GAM		
Missing	MCMC	mean	round	median	MCMC	mean	round	median
10%	0.20	0.52	0.40	0.30	5.50	4.74	5.50	5.50
20%	0.20	0.38	0.40	0.20	2.31	1.82	2.15	2.62
44%	0.14	0.30	0.46	0.09	3.67	2.81	2.78	3.89

For the non-zero count data, the Bayesian method is not clearly better or worse than GAM. However, for the zero-count data, MCMC performs much better than GAM. Even after we round GAM to the closest integer, MCMC is still much better than both two GAM rounding methods except for the case where missing percentage is 44%, Here, MSPE is 0.14, 0.30, 0.46 and 0.09 for MCMC, GAM mean, round and GAM median, respectively. GAM median MSPE agrees with MCMC for the 20% missing data case.

2.7 Simulation

We further evaluate the performance of the Bayesian prediction model against GAM with simulated data. We generate simulated data on a regular square grid with unit spacing. Two sample sizes are simulated ($n = 144$ and $n = 225$). For each sample size, two levels of spatial dependence (moderate and strong) are simulated. In our experiments, spatial dependence is specified by the effective range defined in Section 2.4. The moderate and strong dependence have effective ranges $R = 8.3$ and $R = 14$ respectively. In this simulation study, all the target means are set to be the constant exp(1). Hence, the dependence in the data is not from the spatial pattern of covariates, but from spatial proximity (Madsen, 2009). $N = 50$ simulated data sets are generated for each scenario of sample size and dependence level. As before, we randomly hold out 10%, 20% and 44% of the observations, and we use the remainder to predict them. In each scenario, the locations of the missing data are the same for all 50 simulated data sets.

The prior specification is the same as described in Section 2.5. In the proposal density definition, we set $\sigma^2 = 0.2$; $\epsilon_0 = 0.4$; $\epsilon_1 = 0.2$, $\sigma^2_\phi = 0.5$ and $\epsilon_U = 0.2$. These values are obtained on the basis of a short pilot simulation in order to find settings that give the desired acceptance rates for proposed moves. Here, 20,000 iterations with 10,000 burn-in are run for each chain.

Table 2.5 lists the prediction performance of MCMC and GAM on four scenarios of our simulated data sets. Since there are 50 data sets for each scenario, the average and the standard deviation of 50 MSPEs are given for each method. All the values from MCMC is much smaller the values from both two GAM rounding methods, MCMC prediction is thus better than integer versions of GAM prediction. Moreover, except for one case (missing percent=10%, $N = 144$, and $R = 8.3$), MCMC is better than the original GAM fitted value which uses the mean as the fitted value. The value of MSPE decreases as the

TABLE 2.5: Comparison of MCMC and GAM, predicting 10%, 20% and 44% missing from simulated data

Sample Size	Effective Range	Percent Missing	mean of CV (sd)			
				GAM		
			MCMC	mean	round	median
N=225	R=14	10%	1.28 (0.48)	1.38 (0.65)	1.45 (0.67)	1.52 (0.65)
		20%	1.27 (0.38)	1.33 (0.42)	1.42 (0.45)	1.45 (0.44)
		44%	1.44 (0.38)	1.49 (0.40)	1.57 (0.44)	1.62 (0.42)
	R=8.3	10%	2.12 (0.92)	2.17 (0.96)	2.25 (0.96)	2.26 (1.02)
		20%	2.14 (0.60)	2.14 (0.61)	2.20 (0.58)	2.25 (0.67)
		44%	2.24 (0.49)	2.25 (0.53)	2.33 (0.541)	2.33 (0.537)
N=144	R=14	10%	1.44 (0.74)	1.63 (1.01)	1.72 (1.08)	1.73 (1.18)
		20%	1.35 (0.54)	1.49 (0.63)	1.56 (0.63)	1.63 (0.66)
		44%	1.58 (0.61)	1.63 (0.65)	1.74 (0.65)	1.78 (0.72)
	R=8.3	10%	2.17 (0.92)	2.14 (0.94)	2.22 (0.98)	2.31 (1.04)
		20%	2.29 (0.95)	2.29 (0.91)	2.33 (0.96)	2.42 (1.00)
		44%	2.30 (0.60)	2.36 (0.71)	2.43 (0.72)	2.51 (0.75)

effective range increase (from $R = 8.3$ to $R = 14$). This thus suggests that MCMC may be better suited for applications with stronger spatial dependence.

As with the grub data, we divide simulated count data into two categories, one with counts less than or equal to 2 and the other with counts over 2. The reason for using 2 as the cut-off point is that the count data are typically larger than 0 in these simulated datasets. Since the GAM mean is better than both two rounding methods, we will simply use the fitted value or mean as GAM predictions here.

Again, we can see in Table 2.6 MCMC performs consistently better than GAM for

TABLE 2.6: Decomposition of comparison between MCMC and GAM for smaller and bigger counts on simulated data

Sample Size	Effective Range	Percent Missing	Category 1: $Y \in [0,2]$		Category 2: $Y > 2$	
			MCMC	GAM (mean)	MCMC	GAM (mean)
N=225	R=14	10%	0.90	1.15	1.87	1.85
		20%	1.00	1.21	1.69	1.68
		44%	1.11	1.31	2.01	1.94
	R=8.3	10%	1.38	1.70	3.01	2.83
		20%	1.61	1.87	2.82	2.56
		44%	1.65	1.95	3.00	2.74
N=144	R=14	10%	1.35	2.00	1.85	1.88
		20%	1.89	2.07	1.75	1.84
		44%	1.68	1.98	2.14	2.07
	R=8	10%	1.87	2.00	2.94	2.70
		20%	2.14	2.26	3.33	3.08
		44%	1.76	2.12	3.23	2.98

the smaller count category. In applications predicting the small-count data, MCMC may give better prediction.

2.8 Discussion

This paper discusses regression parameter estimation and missing data prediction for spatial count data, two inferential procedures that are traditionally handled separately in this setting. We present a Bayesian model that performs the two tasks simultaneously. In our model, the correlation among the counts is modeled using Gaussian copula with

the assumption that the counts have the Negative Binomial as the marginal distributions. We develop an MCMC-based approach to estimate the model and show that the method is practical. Since all the 95% intervals are narrower than the ones obtained from ML method and all of them are part of the 95% interval constructed by using ML method, the MCMC estimators presented here thus appear to be more efficient than ML estimators.

Moreover, for missing data prediction, we carried out a comprehensive comparison between the Bayesian approach and the Generalized Additive Model. With MSPE as the measurement of prediction performance, we have shown that MCMC approach usually performs better than GAM. This is especially true when the proportion of missing data is large; the count data locations have high correlations; the sample size is large; and the missing counts are small. However, using MCMC to do prediction takes more time than GAM. MCMC prediction is a time-consuming process.

Though the motivation for this research is to analyze and predict spatially correlated count data, the model is general and can be applied to other correlated count data including longitudinal measurements on a small set of subjects over a long period of time. In addition, the Gaussian copula correlation matrix is not just restricted to the spatial correlation function, other correlation models would be appropriate. The results of this paper are then relatively easy to apply for practitioners.

3 THEORETICAL RESULTS

3.1 Introduction

In probability theory and statistics, a copula can be used to describe the dependence between random variables. The motivation for using a copula model is rooted in the aim of using dependence patterns to form a multivariate non-normal distribution by combining non-normal marginal distributions. In the context of dependence, however, Sklar's theorem (Sklar, 1959) depends on the assumption of the continuity of the marginal probability density functions (p.d.f.s): if the random variables under study have continuous distribution functions, the corresponding copula is unique. Most research is devoted to this continuous situation, but much less is known about discrete cases where many desirable properties of dependence measures no longer hold. However, discrete cases exist everywhere. For example, the counts of Japanese grubs studied in Section 2.6 have non-continuous marginal distributions.

While there are a variety of copulas, the Gaussian copula is often chosen because it makes the constructed multivariate model appear to have many properties similar to the multivariate normal distribution. And this copula can be extended to be used in count-valued random variables. For example, Song et al. (2009) proposes a class of multivariate dispersion models generated from the multivariate Gaussian copula.

Consequently, we model correlated count data based on the Gaussian copula in Section 2.4.4. This model uses the expected likelihood approach proposed by Madsen (2009). There are two important components in our copula-based correlated count data model. One is using the continuous extension for count data. The other is using the Gaussian copula to join multiple univariate marginal distributions into a single multivariate distribution.

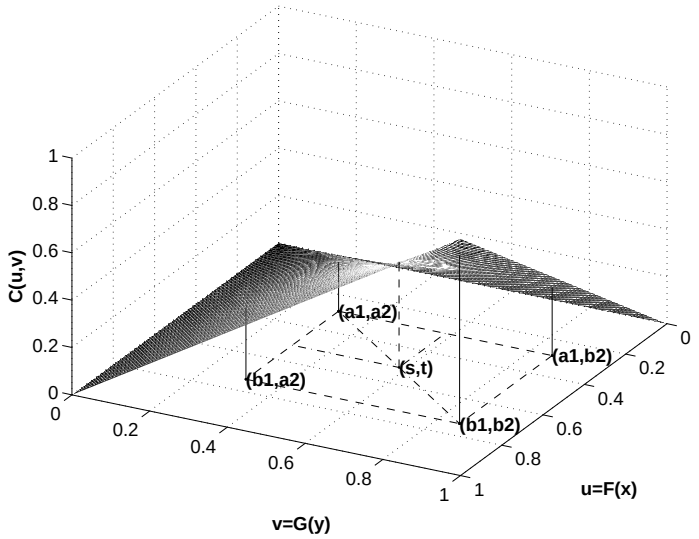


FIGURE 3.1: Bilinear Interpolation of Copula

The essence of the copula approach is that a joint distribution of random variables can be expressed as a function of the marginal distributions. A copula representing a joint distribution of discrete variables is only unique on the range of the marginal cumulative distribution functions (c.d.f.s); while a copula representing a joint distribution of continuous variables is unique on $[0, 1]$. Assume that $X \sim \mathrm{F}$ and $Y \sim \mathrm{G}$. Then the copula $C\left(\mathrm{F}(x), \mathrm{G}(y)\right)$ gives a joint distribution of X and Y. Figure 3.1 illustrates a typical copula. Suppose that F and G are discrete. Let a_1 and b_1 be adjacent points in the range of F. Similarly, let a_2 and b_2 be adjacent points in the range of G. Then C is uniquely defined at points (a_1, a_2), (a_1, b_2), (b_1, a_2) and (b_1, b_2), but not at any points (s, t) where $(s, t) \in [a_1, b_1] \times [a_2, b_2]$. C must be interpolated in the rectangle $[a_1, b_1] \times [a_2, b_2]$. We investigate two interpolations or "continuous extensions." One is the "starred copula" introduced by Denuit and Lambert (2005) and studied in Section 2.4.4. The other is the "standard extension copula" introduced by Schweizer and Sklar (1974). Denuit and Lambert (2005) assert that the starred copula and the standard extension copula are the same. In this section, one of our objectives is to investigate the details of this assertion.

Another objective is to investigate the relationship between existing methods used for modeling dependent binary random variables. Although there are several existing methods used to model correlated discrete data, we focus on only two of them. One is the expected likelihood method developed by Madsen and Fang (2011), the other is the multivariate dispersion model proposed by Song et al. (2009). Both methods use the Gaussian copula. The expected likelihood method is based on continuous marginals, while the multivariate dispersion model is based on non-continuous marginals. Our question is whether these two methods provide the same results.

The behavior of dependence among count-valued random variables, of course, is also important. For example, Genest and Nešlehová (2007) introduce a generalization of the rank correlation measure Spearman's ρ for non-continuous random variables. Gibbons (1985) also measures the degree of correspondence (again, Spearman's ρ) between rankings. Although both of these measures are obtained by using a sample version of Spearman's ρ, the relationship between the two methods is unclear. Thus, our final objective is to ascertain the connections between these two dependence measures.

The rest of the section is organized as follows. In Section 3.2, we explore the connection between the standard extension copula and the starred copula. Then we compare the two Gaussian copula models for dependent count-valued random variables in Section 3.3. At the end of the section, we examine two different estimators of Spearman's ρ.

3.2 Continuous Extension

In this section, we illustrate the two aforementioned continuous extensions to the Gaussian copula modeling dependent discrete random variables, and we explore the connections between them. The starred copula is proposed by Denuit and Lambert (2005). It is a continuous extension copula of integer-valued random variables by convolution

with unit support kernels. The standard extension copula, first introduced by Schweizer and Sklar (1974) and further developed by Nešlehová (2007), models the dependence structures in an analogous way to the unique copula in the continuous case.

Both the standard extension copula and the starred copula provide us with a unique copula for count data. Moreover, Denuit and Lambert (2005) state that both copulas use the same bilinear interpolation statistical method. However, no clear guidelines have been provided for exploring the connections between these two methods. Although both copulas use bilinear interpolations, the places where they perform these interpolations differ. In the following, we will discuss these two methods separately and find the connections between them.

3.2.1 The Standard Extension Copula

First, consider the standard extension copula. Let (X, Y) be a count-valued bivariate random variable having joint distribution function $\mathrm{H}(x, y)$ with margins F and G, respectively. F and G are arbitrary and discrete having ranges $\operatorname{ran}\mathrm{F}$ and $\operatorname{ran}\mathrm{G}$. Sklar's Theorem guarantees that there exists at least one copula C such that

$$\mathrm{H}(x, y) = \mathrm{C}(\mathrm{F}(x), \mathrm{G}(y)) \quad \text{for all} \quad x, y \in \mathbb{R}, \tag{3.1}$$

where C is uniquely determined on $\operatorname{ran}\mathrm{F} \times \operatorname{ran}\mathrm{G}$ and even on the closure of $\operatorname{ran}\mathrm{F} \times \operatorname{ran}\mathrm{G}$.

According to Nešlehová (2007), for any random variables $s, t, \in [0, 1]$ the standard extension copula is defined as:

$$\begin{aligned}\mathrm{C}^S(s, t) = (1 - \lambda_1)(1 - \lambda_2)\mathrm{C}(a_1, a_2) + (1 - \lambda_1)\lambda_2\mathrm{C}(a_1, b_2) \\ + \lambda_1(1 - \lambda_2)\mathrm{C}(b_1, a_2) + \lambda_1\lambda_2\mathrm{C}(b_1, b_2)\end{aligned} \tag{3.2}$$

with

$$\lambda_1 = \begin{cases} \frac{s - a_1}{b_1 - a_1} & \text{if } a_1 < b_1 \\ 1 & \text{if } a_1 = b_1 \end{cases}$$

and

$$\lambda_2 = \begin{cases} \frac{t-a_2}{b_2-a_2} & \text{if } a_2 < b_2 \\ 1 & \text{if } a_2 = b_2 \end{cases},$$

where a_1 , b_1 are the least and the greatest element in the closure of ran F such that $a_1 \leq s \leq b_1$; a_2, b_2 are the least and the greatest element in the closure of ran G such that $a_2 \leq t \leq b_2$.

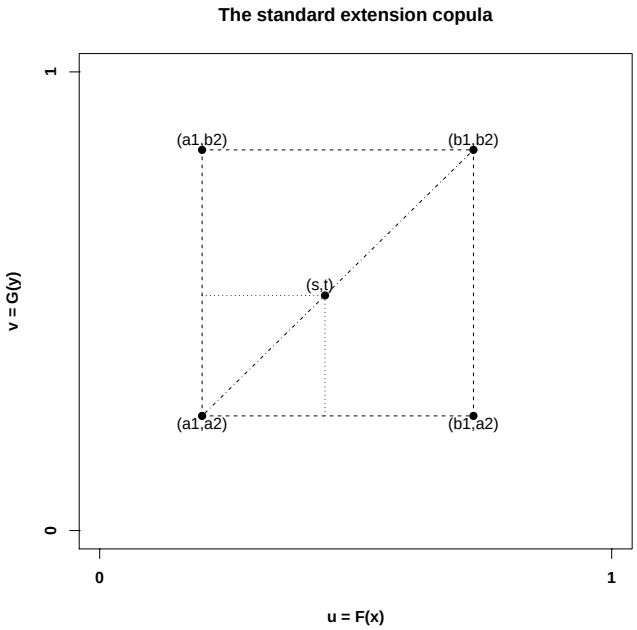


FIGURE 3.2: Plot of the Standard Extension Copula.

Figure 3.2 illustrates the standard extension copula. It also corresponds to the horizontal plane in Figure 3.1. The horizontal axis of the graph in Figure 3.2 is the marginal c.d.f of X (i.e., $u = \mathrm{F}(x)$); the vertical axis is the marginal c.d.f of Y (i.e., $v = \mathrm{G}(y)$). The four dots, $\{(a_1, b_1), (a_1, b_2), (a_2, b_1), (a_2, b_2)\} \in \operatorname{ran} F \times \operatorname{ran} G$, are the points defined above and the dot, (s, t), is the point we want to interpolate, that is, we need to define $\mathrm{C}(s.t)$. The standard extension copula performs bilinear interpolation by interpolating functions of the two marginal distributions, F and G.

3.2.2 The Starred Copula

We will now discuss the starred copula. Again, let (X, Y) be variables with marginals F and G, respectively. The corresponding non-zero probability masses are given by

$$\mathrm{f}_x = \mathrm{P}(X = x); \qquad \mathrm{g}_y = \mathrm{P}(Y = y).$$

X and Y are extended to continuous variables, X^* and Y^*, with distribution functions F^* and G^* by using independent variables, U and V, respectively. The relationships between X^* and X and Y^* and Y are:

$$X^* = X - 1 + U; \qquad Y^* = Y - 1 + V, \tag{3.3}$$

where U is a continuous random variable in $[0, 1]$, independent of X with a strictly increasing c.d.f, L_U, that shares no parameters with the distribution of X. V is a continuous random variable in $[0, 1]$, independent of Y with a strictly increasing c.d.f, L_V, that shares no parameters with the distribution of Y. Further, U and V are independent of each other. We say that X is continued by U, and Y is continued by V. The most natural choice for both U and V is the uniform distribution on $[0, 1]$, which satisfies all the constraints on L_U and L_V. We have

$$\mathrm{L}_U(u) = u; \qquad \mathrm{l}_U(u) = 1$$

and

$$\mathrm{L}_V(v) = v; \qquad \mathrm{l}_V(v) = 1,$$

for $u, v \in [0, 1]$, and l_U and l_V are the densities corresponding to L_U and L_V.

By definition of X^*, Y^*, u, and v, $X^* \leq X$ and $Y^* \leq Y$. Let $[x^*]$ denote the integer part of $x^* \in \mathbb{R}$ and $[y^*]$ denote the integer part of $y^* \in \mathbb{R}$. Then the variable u

corresponding to x^* can be denoted as

$$\begin{aligned} u = \mathrm{F}^*(x^*) &= \mathrm{P}(X^* \le x^*) = \mathrm{P}(X \le [x^*]) + \mathrm{L}_U(x^* - [x^*])\mathrm{P}(X = [x^* + 1]) \\ &= \mathrm{F}([x^*]) + \mathrm{L}_U(x^* - [x^*])\mathrm{P}(X = [x^* + 1]) \\ &= \mathrm{F}([x^*]) + (x^* - [x^*])\mathrm{P}(X = [x^* + 1]), \end{aligned} \tag{3.4}$$

Similarly,

$$v = \mathrm{G}^*(y^*) = \mathrm{G}([y^*]) + (y^* - [y^*])\mathrm{P}(Y = [y^* + 1]). \tag{3.5}$$

The p.d.f.s, f* and g*, corresponding to F* and G* are defined as

$$\mathrm{f}^*(x^*) = \mathrm{l}_U(x^* - [x^*])\mathrm{P}(X = [x^* + 1]) = \mathrm{P}(X = [x^* + 1])$$

and

$$\mathrm{g}^*(y^*) = \mathrm{l}_V(y^* - [y^*])\mathrm{P}(Y = [y^* + 1]) = \mathrm{P}(Y = [y^* + 1]).$$

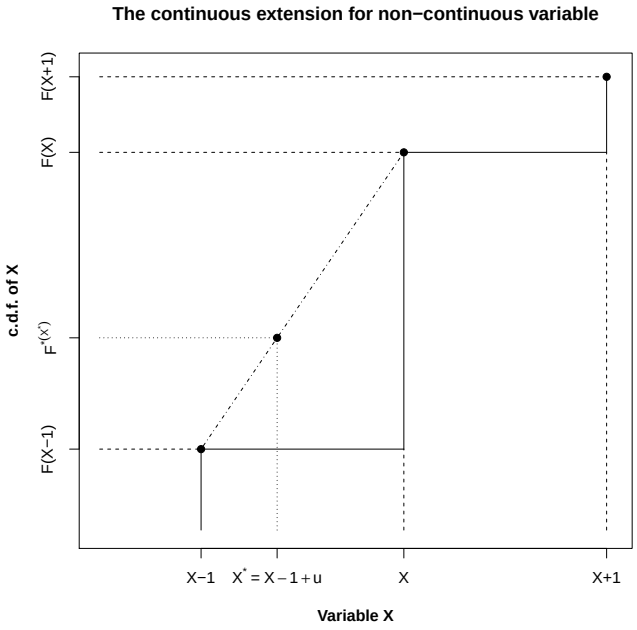


FIGURE 3.3: Plot of Continuous Extension for Count-valued Variable.

Figure 3.3 illustrates the relationship between a count-valued random variable, X, and its continuous extension, X^*. The horizontal axis shows $X-1$, X and $X+1$; the vertical axis gives the corresponding values $\mathrm{F}(X-1)$, $\mathrm{F}(X)$ and $\mathrm{F}(X+1)$. The point $(X^*, \mathrm{F}^*(X^*))$ is the continuous extension of $(X, \mathrm{F}(X))$. The value $X^* = X - 1 + U$ falls between $X-1$ and X, and $\mathrm{F}^*(X^*) = F(X-1) + u\,[F(X) - F(X-1)]$ obtained from (3.4).

According to Sklar's theorem, the joint distribution $\mathrm{H}(x, y; U, V) = \mathrm{H}(x^*, y^*) = \mathrm{P}\{X^* \le x^*, Y^* \le y^*\} = \mathrm{C}(\mathrm{F}^*(x^*), \mathrm{G}^*(y^*))$ holds for some copula C. Both distribution choices for variables U and V and the choice made for C completely determine the joint distribution, $\mathrm{H}(x, y; U, V)$. Let x_d denote the fractional part of x^* (that is, $x_d = x^* - [x^*]$, where $x^* \in \mathbb{R}$). Similarly, let y_d denote the fractional part of y^*. The joint distribution $\mathrm{H}(x, y; U, V)$ can be expressed as

$$
\begin{aligned}
\mathrm{H}(x, y; U, V) &= \mathrm{H}(x^*, y^*) = \mathrm{P}(X^* \le x^*, Y^* \le y^*) \\
&= \mathrm{P}(X \le [x^*], Y \le [y^*]) + \mathrm{L}_U(x_d)\mathrm{P}(X = [x^*]+1, Y \le [y^*]) \\
&\quad + \mathrm{L}_V(y_d)\mathrm{P}(X \le [x^*], Y = [y^*]+1) + \mathrm{L}_U(x_d)\mathrm{L}_V(y_d)\mathrm{P}(X = [x^*]+1, Y = [y^*]+1) \\
&= \mathrm{C}\left(\mathrm{F}([x^*]), \mathrm{G}([y^*])\right) \\
&\quad + \mathrm{L}_U(x_d)\left\{\mathrm{C}\left(\mathrm{F}([x^*]+1), \mathrm{G}([y^*])\right) - \mathrm{C}\left(\mathrm{F}([x^*]), \mathrm{G}([y^*])\right)\right\} \\
&\quad + \mathrm{L}_V(y_d)\left\{\mathrm{C}\left(\mathrm{F}([x^*]), \mathrm{G}([y^*]+1)\right) - \mathrm{C}\left(\mathrm{F}([x^*]), \mathrm{G}([y^*])\right)\right\} \\
&\quad + \mathrm{L}_U(x_d)\mathrm{L}_V(y_d)\left\{\mathrm{C}\left(\mathrm{F}([x^*]+1), \mathrm{G}([y^*]+1)\right) - \mathrm{C}\left(\mathrm{F}([x^*]), \mathrm{G}([y^*])\right)\right\} \\
&= \mathrm{L}_U(x_d)\left\{\mathrm{L}_V(y_d)\mathrm{C}\left(\mathrm{F}([x^*]+1), \mathrm{G}([y^*]+1)\right) + \left(1 - \mathrm{L}_V(y_d)\right)\mathrm{C}\left(\mathrm{F}([x^*]+1), \mathrm{G}([y^*])\right)\right\} \\
&\quad + \left(1 - \mathrm{L}_U(x_d)\right)\left\{\mathrm{L}_V(y_d)\mathrm{C}\left(\mathrm{F}([x^*]), \mathrm{G}([y^*]+1)\right) + \left(1 - \mathrm{L}_V(y_d)\right)\mathrm{C}\left(\mathrm{F}([x^*]), \mathrm{G}([y^*])\right)\right\} \\
&= \mathrm{L}_U(x_d)\mathrm{L}_V(y_d)\mathrm{C}\left(\mathrm{F}(X), \mathrm{G}(Y)\right) \\
&\quad + \mathrm{L}_U(x_d)(1 - \mathrm{L}_V(y_d))\mathrm{C}\left(\mathrm{F}(X), \mathrm{G}(Y-1)\right) \\
&\quad + \left(1 - \mathrm{L}_U(x_d)\right)\mathrm{L}_V(y_d)\mathrm{C}\left(\mathrm{F}(X-1), \mathrm{G}(Y)\right) \\
&\quad + \left(1 - \mathrm{L}_U(x_d)\right)\left(1 - \mathrm{L}_V(y_d)\right)\mathrm{C}\left(\mathrm{F}(X-1), \mathrm{G}(Y-1)\right) \\
&= (1-\lambda_1)(1-\lambda_2)\,\mathrm{C}\left(\mathrm{F}(X-1), \mathrm{G}(Y-1)\right) + (1-\lambda_1)\,\lambda_2\mathrm{C}\left(\mathrm{F}(X-1), \mathrm{G}(Y)\right) \\
&\quad + \lambda_1(1-\lambda_2)\mathrm{C}\left(\mathrm{F}(X), \mathrm{G}(Y-1)\right) + \lambda_1\lambda_2\mathrm{C}\left(\mathrm{F}(X), \mathrm{G}(Y)\right). \qquad (3.6)
\end{aligned}
$$

where $\lambda_1 = x_d = x^* - [x^*] = U$ and $\lambda_2 = y_d = y^* - [y^*] = V$.

Denote the unique copula corresponding to $\mathrm{H}(x^*, y^*)$ as C^*, such that

$$\mathrm{H}(x^*, y^*) = \mathrm{C}^*\left(\mathrm{F}^*(x^*), \mathrm{G}^*(y^*)\right) \quad \forall (x^*, y^*) \in \mathbb{R}^2.$$

In other words,

$$\begin{aligned}\mathrm{C}^*(s,t) = {} & (1-\lambda_1)(1-\lambda_2)\,\mathrm{C}\left(\mathrm{F}(X-1), \mathrm{G}(Y-1)\right) + (1-\lambda_1)\,\lambda_2 \mathrm{C}\left(\mathrm{F}(X-1), \mathrm{G}(Y)\right) \\ & + \lambda_1(1-\lambda_2)\mathrm{C}\left(\mathrm{F}(X), \mathrm{G}(Y-1)\right) + \lambda_1\lambda_2 \mathrm{C}\left(\mathrm{F}(X), \mathrm{G}(Y)\right), \end{aligned} \tag{3.7}$$

where $s, t \in [0, 1]$.

In this way, the copula $\mathrm{C}^*(s,t)$ is a bilinear interpolation of C at the surrounding points $(X-1, X)$ and $(Y-1, Y)$; that is, on the unit square $[X-1, X] \times [Y-1, Y]$. It extends the copula from non-continuous marginals to continuous marginals, and after this continuous extension, the copula is uniquely determined. Note that $\mathrm{C}^*(s,t)$ is independent of the distributions of U and V; distribution choices made for both U and V do not influence the starred copula.

3.2.3 Comparison of the Standard Extension Copula and the Starred Copula

Both the standard extension copula defined in (3.2) and the starred copula defined in (3.7) are based on the bilinear interpolation. The starred copula extends the (unique) values of the copula for bivariate (X, Y) defined in the range $(X-1, X) \times (Y-1, Y)$. While the standard extension copula extends the (unique) values of the copula for bivariate (X, Y) defined in the range $(\mathrm{F}(x-1), \mathrm{F}(x)) \times (\mathrm{G}(y-1), \mathrm{G}(y))$. From the derivation, we can see that the two methods are otherwise essentially the same.

3.3 Gaussian copula with count margins

In this section, we explore a Gaussian copula model for binary dependent data. Song et al. (2009) introduces a joint probability mass function by using the Gaussian

copula to model the dependence. However, it contains 2^n terms and is intractable for large n. The n-fold summation can be avoided by using a continuous extension method proposed by Denuit and Lambert (2005). Madsen and Fang (2011) propose an expected likelihood method that brings discrete distribution into the Gaussian copula framework. This method goes one step further than the Gaussian dispersion model (Song et al., 2009) to the expected likelihood method, which enables parameter estimation from high-dimensional response vectors, and it is asymptotically equivalent to Song et al.'s (2009) Gaussian copula likelihood.

In the copula-based spatially correlated model for count data in Section 2, we model the univariate marginals first; we then use the Gaussian copula to couple those marginals to the multivariate normal distribution. The multivariate normal distribution has support $\mathbb{R}^k$, where k is the number of components in the multivariate vector. However, this range will not hold when the marginals are continuous extensions of count variables. Let Y be a count-valued random variable with marginal distribution F. After the continuous extension (see 3.2.2), Y is continued to a new value Y^* by adding a perturbation, U, from a uniform distribution on [0,1]. The range for Y^* is restricted to $[Y-1, Y]$, and the range for the variable in Gaussian copula is thus restricted to $\left[\Phi^{-1}(F(Y-1)), \Phi^{-1}(F(Y))\right]$. Consequently, unlike the traditional multivariate normal distribution, the range of the variables joined through the Gaussian copula is truncated to a smaller segment.

In the following, we give a short introduction of univariate truncated normal distribution, and then we extend that to the more general multivariate truncated normal distribution. We take the continuous extension for Y and use a copula model to construct a likelihood function which corresponds to both Y and U. Then, we take the expected likelihood function (see Madsen and Fang, 2011) with respect to U instead of the likelihood function itself. Since both the method in Song et al. (2009) and the method in Madsen and Fang (2011) can be used to model count data, we want to investigate the

similarities between these two methods. Specifically, our goal is to explore the equivalence between the resulting likelihood functions using the two methods for variables with binary margins.

3.3.1 Univariate Truncated Normal Distribution

First, we want to give the univariate distribution of the quantity $\Phi^{-1}\{F^*(y^*)\}$. According to Section 3.2, Y is continued to Y^* by using a jittering variable, U, which is drawn from a continuous uniform distribution on $[0,1]$ and is independent of Y. That is, $Y^* = Y - U$. Y^* is a continuous random variable with distribution function $\mathrm{F}^*(y^*)$ given in (3.4) and density function $\mathrm{f}^*(y^*) = \mathrm{P}(Y = y)$, where $y^* \in [y-1, y]$. Let $Z = \Phi^{-1}\{\mathrm{F}^*(Y^*)\}$, where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal c.d.f. Z follows a normal distribution $N(\mu, \sigma^2)$, where $\mu = 0$ and $\sigma = 1$. From the definition of Y^*, we know that $Y^* \in [y-1, y]$. It follows that Z lies within the interval $Z \in (a, b)$, where $a = \Phi^{-1}\{\mathrm{F}([y^*])\} = \Phi^{-1}\{\mathrm{F}(y-1)\}$, $b = \Phi^{-1}\{\mathrm{F}([y^*+1])\} = \Phi^{-1}\{\mathrm{F}(y)\}$, and $-\infty \leq a \leq b \leq +\infty$. The conditional probability distribution of Z given Y is

$$
\begin{aligned}
\mathrm{F}(z|Y) &= \mathrm{P}(Z \leq z | a \leq z \leq b) \\
&= \frac{\int_a^z \mathrm{f}(z)dz}{\int_a^b \mathrm{f}(z)dz} \\
&= \frac{\int_{\Phi^{-1}\{\mathrm{F}(y-1)\}}^{z} \mathrm{f}(z)dz}{\int_{\Phi^{-1}\{\mathrm{F}(y-1)\}}^{\Phi^{-1}\{\mathrm{F}(y)\}} \mathrm{f}(z)dz} \\
&= \frac{\int_{-\infty}^{z} \mathrm{f}(z)dz \quad - \quad \int_{-\infty}^{\Phi^{-1}\{\mathrm{F}(y-1)\}} \mathrm{f}(z)dz}{\int_{-\infty}^{\Phi^{-1}\{\mathrm{F}(y)\}} \mathrm{f}(z)dz \quad - \quad \int_{-\infty}^{\Phi^{-1}\{\mathrm{F}(y-1)\}} \mathrm{f}(z)dz} \\
&= \frac{\Phi(z) - \mathrm{F}(y-1)}{\mathrm{F}(y) - \mathrm{F}(y-1)} \\
&= \frac{\Phi(z) - \mathrm{F}(y-1)}{\mathrm{P}(Y = y)},
\end{aligned}
\tag{3.8}
$$

where $\Phi(\cdot)$ is the c.d.f of a standard normal random variable.

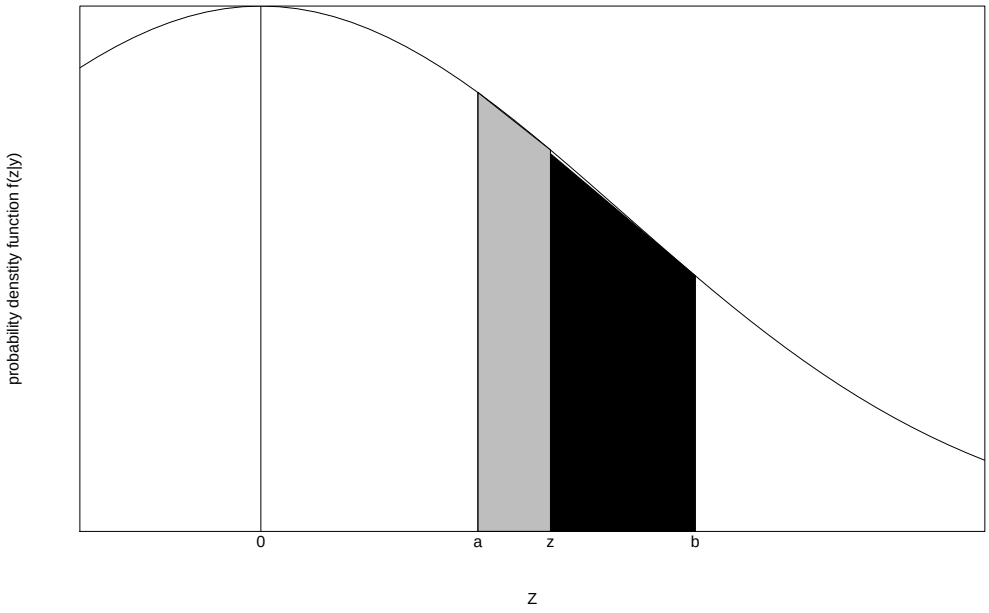


FIGURE 3.4: Truncated normal distribution

Figure 3.4 illustrates the p.d.f of the truncated normal distribution with $Z \in [a, b]$. The conditional p.d.f of Z is the proportion of the lightly shaded area to the whole shaded area. Also it can be obtained by taking the derivative of $\mathrm{F}(z|Y)$ with respect to Z; namely,

$$\begin{aligned} \mathrm{f}(z|Y = y) &= \frac{\partial \mathrm{F}(z|a \leq z \leq b)}{\partial z} \\ &= \frac{\phi(z)}{\mathrm{P}(Y = y)}, \end{aligned} \tag{3.9}$$

where $\phi(\cdot)$ denotes the standard normal p.d.f. Equation (3.9) is the p.d.f of the truncated normal distribution. In other words, $Z = \Phi^{-1}\left\{\mathrm{F}^*(y^*)\right\}$ follows a truncated normal distribution.

3.3.2 Multivariate Truncated Normal Distribution

Let $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$ be variables with marginal distributions $\mathrm{F}_1, \ldots, \mathrm{F}_n$, respectively. For each Y_i, $i = 1, \ldots, n$, we define the continuous variable

$$Y_i^* = Y_i - U_i, \tag{3.10}$$

where U_i is drawn from a continuous uniform distribution on $[0, 1]$, and is independent of both Y_i and U_j for $j \neq i$. Referring to Section 3.2, Y_i^* is a continuous random variable with c.d.f

$$\mathrm{F}_i^*(y_i^*) = \mathrm{F}_i(y_i - 1) + (y_i^* - (y_i - 1))\,\mathrm{P}(Y_i = y_i)$$

and density function

$$\mathrm{f}_i^*(y_i^*) = \mathrm{P}(Y_i = y_i),$$

where $y_i^* \in [y_i - 1, y_i]$.

If we let $Z_i = \Phi^{-1}(\mathrm{F}_i^*(y_i^*))$, then Z_i follows a truncated normal distribution. According to (3.9), the p.d.f for Z_i is

$$\mathrm{f}(Z_i = z_i | Y_i = y_i) = \frac{\phi(z_i)}{\mathrm{P}(Y_i = y_i)},$$

where $\Phi^{-1}\left\{\mathrm{F}_i([y_i^*])\right\} \leq z_i \leq \Phi^{-1}\left\{\mathrm{F}_i([y_i^* + 1])\right\}$, otherwise $\mathrm{f}(z_i | Y_i = y_i) = 0$. The first and the second moments of the truncated normal distribution are:

$$E\left[Z_i | a_i \leq Z_i \leq b_i\right] = \mu_i + \frac{\phi(\frac{a_i - \mu_i}{\sigma_i}) - \phi(\frac{b_i - \mu_i}{\sigma_i})}{\Phi(\frac{b_i - \mu_i}{\sigma_i}) - \Phi(\frac{a_i - \mu_i}{\sigma_i})}\sigma_i = \frac{\phi(a_i) - \phi(b_i)}{\Phi(b_i) - \Phi(a_i)} = \frac{\phi(a_i) - \phi(b_i)}{P(Y_i = y_i)} \tag{3.11}$$

and

$$\begin{aligned}
&\text{var}\left[Z_i | a_i \leq Z_i \leq b_i\right] \\
&= \sigma_i^2 \left[1 + \frac{\frac{a_i-\mu_i}{\sigma_i}\phi(\frac{a_i-\mu_i}{\sigma_i}) - \frac{b_i-\mu_i}{\sigma_i}\phi(\frac{b_i-\mu_i}{\sigma_i})}{\Phi(\frac{b_i-\mu_i}{\sigma_i}) - \Phi(\frac{a_i-\mu_i}{\sigma_i})} - \left(\frac{\phi(\frac{a_i-\mu_i}{\sigma_i}) - \phi(\frac{b_i-\mu_i}{\sigma_i})}{\Phi(\frac{b_i-\mu_i}{\sigma_i}) - \Phi(\frac{a_i-\mu_i}{\sigma_i})}\right)^2\right] \\
&= 1 + \frac{a_i\phi(a_i) - b_i\phi(b_i)}{\Phi(b_i) - \Phi(a_i)} - \left(\frac{\phi(a_i) - \phi(b_i)}{\Phi(b_i) - \Phi(a_i)}\right)^2 \\
&= 1 + \frac{a_i\phi(a_i) - b_i\phi(b_i)}{P(Y_i = y_i)} - \left(\frac{\phi(a_i) - \phi(b_i)}{P(Y_i = y_i)}\right)^2.
\end{aligned} \tag{3.12}$$

For more detailed information, please refer to Johnson (1994).

The joint distribution for $\mathbf{Y}$ is given by the Gaussian copula model

$$\mathrm{C}(y_1, \ldots, y_n; \Sigma) = \Phi_\Sigma\left[\Phi^{-1}\{\mathrm{F}_1^*(y_1^*)\}, \ldots, \Phi^{-1}\{\mathrm{F}_n^*(y_n^*)\}\right], \tag{3.13}$$

where Φ_Σ is the multivariate normal c.d.f with covariance matrix Σ (Joe, 2001, pages 140-141). Differentiating $\mathrm{C}(y_1, \ldots, y_n; \Sigma)$ gives the joint distribution

$$\mathrm{c}(y_1, \ldots, y_n; \Sigma) = |\Sigma|^{-1/2} \exp\left[-\frac{1}{2}\mathbf{Z}^T(\Sigma^{-1} - \mathbf{I}_n)\mathbf{Z}\right] \prod_{i=1}^{n} \mathrm{f}_i(y_i), \tag{3.14}$$

where $\mathbf{Z} = \left(\Phi^{-1}\{\mathrm{F}_1^*(y_1^*)\}, \ldots, \Phi^{-1}\{\mathrm{F}_n^*(y_n^*)\}\right)^T$, and $\mathbf{I}_n$ denotes the $n \times n$ identity matrix.

A joint probability mass function for $\mathbf{Y}$ can be found by averaging over the jitterings $\mathbf{U} = (U_1, \ldots, U_n)^T$ (see Madsen and Fang, 2011). Since $\mathbf{U}$ is part of $\mathbf{Z}$, we can obtain the joint probability mass function for $\mathbf{Y}$ by averaging over $\mathbf{Z}$ instead of $\mathbf{U}$, that is,

$$\mathrm{L}(y_1, \ldots, y_n; \Sigma) = E_{\mathbf{Z}}\left[\mathrm{c}(y_1, \ldots, y_n; \Sigma)\right] = E_{\mathbf{Z}}\left\{|\Sigma|^{-1/2} \exp\left[-\frac{1}{2}\mathbf{Z}^T(\Sigma^{-1} - \mathbf{I}_n)\mathbf{Z}\right] \prod_{i=1}^{n} \mathrm{f}_i(y_i)\right\}. \tag{3.15}$$

Song et al. (2009) give the joint probability mass function of $(Y_1, \ldots, Y_n)$ as:

$$\begin{aligned}
\mathrm{g}(y_1, \ldots, y_n; \Sigma) &= \mathrm{P}(Y_1 = y_1, \ldots, Y_n = y_n) \\
&= \sum_{j_1=1}^{2} \cdots \sum_{j_n=1}^{2} (-1)^{j_1+\ldots+j_n} \Phi_{\Sigma_\rho}\left(\Phi^{-1}\{\mu_{1j_1}\}, \ldots, \Phi^{-1}\{\mu_{nj_n}\}\right),
\end{aligned} \tag{3.16}$$

where $\mu_{i1} = \mathrm{F}_i(y_i-)$ and $\mu_{i2} = \mathrm{F}_i(y_i)$. The limit of F_i at y_i from the left. For count-valued random variables, $\mathrm{F}_i(y_i-) = F_i(y_i - 1)$.

We seek to verify the equivalence between (3.15) and (3.16). We will show equivalence for both bivariate and general cases. A Bernoulli distribution is assumed for the marginal distribution.

3.3.3 Bivariate Case

Let Y_1 and Y_2 be variables with marginal distributions F_1 and F_2, respectively. For $i = 1, 2$, Y_i is extended to Y_i^* by using a jittering variable U_i. For more information, please refer to equation (3.10).

Equation (3.13) and (3.14) can be used as the joint c.d.f and p.d.f for Y_1^* and Y_2^*, with Σ_ρ defined as

$$\Sigma_\rho = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

Suppose Y_1 and Y_2 are from a Bernoulli distribution with the success probability p_1 and p_2, respectively. That is,

$$\mathrm{f}_1(y_1; p_1) = \begin{cases} 1 - p_1 & \text{if } y_1 = 0 \\ p_1 & \text{if } y_1 = 1 \\ 0 & \text{otherwise} \end{cases}, \quad \mathrm{F}_1(y_1; p_1) = \begin{cases} 0 & \text{if } y_1 < 0 \\ 1 - p_1 & \text{if } 0 \le y_1 < 1 \\ 1 & \text{if } y_1 \ge 1 \end{cases}$$

and

$$\mathrm{f}_2(y_2; p_2) = \begin{cases} 1 - p_2 & \text{if } y_2 = 0 \\ p_2 & \text{if } y_2 = 1 \\ 0 & \text{otherwise} \end{cases}, \quad \mathrm{F}_2(y_2; p_2) = \begin{cases} 0 & \text{if } y_2 < 0 \\ 1 - p_2 & \text{if } 0 \le y_2 < 1 \\ 1 & \text{if } y_2 \ge 1 \end{cases}.$$

Notice that (Y_1, Y_2) can only take one of four possible values:

Case 1 $Y_1 = 0,\ Y_2 = 0$;

Case 2 $Y_1 = 0,\ Y_2 = 1$;

Case 3 $Y_1 = 1,\ Y_2 = 0$;

Case 4 $Y_1 = 1,\ Y_2 = 1$.

To assess the equivalence of the two methods, we arbitrarily take case 2 where $Y_1 = 0$ and $Y_2 = 1$ as an example to illustrate. $Y_1^* \in [-1, 0]$ and $Y_2^* \in [0, 1]$. The p.d.f and c.d.f for Y_1^* and Y_2^* are:

$$\mathrm{f}_1^*(y_1^*) = \mathrm{P}(Y_1 = [y_1^* + 1]) = 1 - p_1;$$

$$\mathrm{F}_1^*(y_1^*) = \mathrm{F}_1([y_1^*]) + (y_1^* - [y_1^*])\mathrm{P}(Y_1 = [y_1^* + 1]) = (1 - p_1)(y_1^* - [y_1^*]);$$

$$\mathrm{f}_2^*(y_2^*) = \mathrm{P}(Y_2 = [y_2^* + 1]) = p_2;$$

$$\mathrm{F}_2^*(y_2^*) = \mathrm{F}_2([y_2^*]) + (y_2^* - [y_2^*])\mathrm{P}(Y_2 = [y_2^* + 1]) = 1 - p_2 + p_2(y_2^* - [y_2^*]);$$

Define $Z_1 = \Phi^{-1}(\mathrm{F}_1^*(y_1^*))$ and $Z_2 = \Phi^{-1}(\mathrm{F}_2^*(y_2^*))$. Then the lower bound a_1 and the upper bound b_1 for Z_1 are given by

$$a_1 = \Phi^{-1}\left\{\mathrm{F}_1([y_1^*])\right\} = \Phi^{-1}(0) = -\infty; \qquad b_1 = \Phi^{-1}\left\{\mathrm{F}_1([y_1^* + 1])\right\} = \Phi^{-1}(1 - p_1).$$

The lower bound a_2 and the upper bound b_2 for Z_2 are given by

$$a_2 = \Phi^{-1}\left\{\mathrm{F}_2([y_2^*])\right\} = \Phi^{-1}(1 - p_2); \qquad b_2 = \Phi^{-1}\left\{\mathrm{F}_2([y_2^* + 1])\right\} = \Phi^{-1}(1) = +\infty.$$

In other words,

$$\Phi(a_1) = \Phi(-\infty) = 0; \quad \Phi(b_1) = \Phi\left(\Phi^{-1}(1 - p_1)\right) = 1 - p_1;$$

and

$$\Phi(a_2) = \Phi\left(\Phi^{-1}(1 - p_2)\right) = 1 - p_2; \quad \Phi(b_2) = \Phi(+\infty) = 1.$$

The expected likelihood method from equation (3.15) for the pair (Y_1, Y_2) is derived in Appendix C; it is:

$$\mathrm{L}(y_1, y_2; \rho) = 1 - p_1 - \Phi_{\Sigma_\rho}\Big(\Phi^{-1}(1-p_1), \Phi^{-1}(1-p_2)\Big). \tag{3.17}$$

From equation (3.16) the joint probability mass function of (Y_1, Y_2) is

$$\mathrm{g}(y_1, y_2; \Sigma_\rho) = \mathrm{P}(Y_1 = y_1, Y_2 = y_2) = \sum_{j_1=1}^{2}\sum_{j_2=1}^{2}(-1)^{j_1+j_2}\Phi_{\Sigma_\rho}\Big(\Phi^{-1}\{\mu_{1j_1}\}, \Phi^{-1}\{\mu_{2j_2}\}\Big).$$

When $Y_1 = 0$ and $Y_2 = 1$, there are 4 combinations of μ_{1j_1} and μ_{2j_2}:

Case 1 $\mu_{11} = P(Y_1 < 0) = 0, \;\; \mu_{21} = P(Y_2 \leq 0) = 1 - p_2;$

Case 2 $\mu_{11} = P(Y_1 < 0) = 0, \;\; \mu_{22} = P(Y_2 \leq 1) = 1;$

Case 3 $\mu_{12} = P(Y_1 \leq 0) = 1 - p_1, \;\; \mu_{21} = P(Y_2 \leq 0) = 1 - p_2;$

Case 4 $\mu_{12} = P(Y_1 \leq 0) = 1 - p_1, \;\; \mu_{22} = P(Y_2 \leq 1) = 1.$

Appendix D shows that the joint probability mass function of (Y_1, Y_2) is identical to the expected likelihood function. Thus far we have demonstrated that the two models are equivalent for the case when $(Y_1, Y_2) = (0, 1)$.

Similarly, when considering other cases, equation (3.15) and equation (3.16) yield the same results. We do not provide details for calculating the other cases, however we list the final results in Table 3.1 and Table 3.2.

For the Bernoulli pair (Y_1, Y_2), Table 3.1 gives the values of a_1, b_1, a_2, b_2 and $\mathrm{L}(y_1, y_2; \rho)$ needed in equation (3.15) (see Madsen and Fang, 2011). Table 3.2 gives the values of μ_{11}, μ_{12}, μ_{21}, μ_{22} and $\mathrm{g}(y_1, y_2; \Sigma_\rho)$ needed in equation (3.16) (see Song et al., 2009). From these two tables, we can see the equivalence between equation (3.15) and equation (3.16) for pairs of Bernoulli variables.

The main outcome of this analysis is the equivalence of the two approaches, namely, equation (3.15) and equation (3.16), for two Bernoulli random variables. Next, we extend the result to high dimensions, where the marginals are still assumed to be Bernoulli.

TABLE 3.1: The values of a_1, b_1, a_2, b_2, $\mathrm{L}(y_1, y_2; \rho)$ for Bernoulli bivariate (Y_1, Y_2)

(Y_1, Y_2)	a_1	b_1	a_2	b_2	$\mathrm{L}(y_1, y_2; \rho)$
(0,0)	$-\infty$	$\Phi^{-1}(1-p_1)$	$-\infty$	$\Phi^{-1}(1-p_2)$	$\Phi_{\Sigma_\rho}\Big(\Phi^{-1}(1-p_1), \Phi^{-1}(1-p_2)\Big)$
(0,1)	$-\infty$	$\Phi^{-1}(1-p_1)$	$\Phi^{-1}(1-p_2)$	$+\infty$	$\Big(1-p_1\Big) - \Phi_{\Sigma_\rho}\Big(\Phi^{-1}(1-p_1), \Phi^{-1}(1-p_2)\Big)$
(1,0)	$\Phi^{-1}(1-p_1)$	$+\infty$	$-\infty$	$\Phi^{-1}(1-p_2)$	$\Big(1-p_2\Big) - \Phi_{\Sigma_\rho}\Big(\Phi^{-1}(1-p_1), \Phi^{-1}(1-p_2)\Big)$
(1,1)	$\Phi^{-1}(1-p_1)$	$+\infty$	$\Phi^{-1}(1-p_2)$	$+\infty$	$1 - \Big(1-p_1\Big) - \Big(1-p_2\Big) + \Phi_{\Sigma_\rho}\Big(\Phi^{-1}(1-p_1), \Phi^{-1}(1-p_2)\Big)$

TABLE 3.2: The values of μ_{11}, μ_{12}, μ_{21}, μ_{22}, $\mathrm{g}(y_1, y_2; \Sigma_\rho)$ for Bernoulli bivariate (Y_1, Y_2)

(Y_1, Y_2)	μ_{11}	μ_{12}	μ_{21}	μ_{22}	$\mathrm{g}(y_1, y_2; \Sigma_\rho)$
(0,0)	0	$1-p_1$	0	$1-p_2$	$\Phi_{\Sigma_\rho}\Big(\Phi^{-1}(1-p_1), \Phi^{-1}(1-p_2)\Big)$
(0,1)	0	$1-p_1$	$1-p_2$	1	$\Big(1-p_1\Big) - \Phi_{\Sigma_\rho}\Big(\Phi^{-1}(1-p_1), \Phi^{-1}(1-p_2)\Big)$
(1,0)	$1-p_1$	1	0	$1-p_2$	$\Big(1-p_2\Big) - \Phi_{\Sigma_\rho}\Big(\Phi^{-1}(1-p_1), \Phi^{-1}(1-p_2)\Big)$
(1,1)	$1-p_1$	1	$1-p_2$	1	$1 - \Big(1-p_1\Big) - \Big(1-p_2\Big) + \Phi_{\Sigma_\rho}\Big(\Phi^{-1}(1-p_1), \Phi^{-1}(1-p_2)\Big)$

3.3.4 General Multivariate Case

We now turn to examine the case of n Bernoulli random variables for arbitrary $n \leq \infty$. Our question is whether (3.15) and (3.16) remain equivalent.

First redefine the conditional p.d.f of $\mathrm{f}(z_i|Y_i = y_i)$ for $i = 1, \ldots, n$. According to (3.9), $\mathrm{f}(z_i|Y_i = y_i)$ can be expressed as

$$\mathrm{f}(z_i|Y_i = y_i) = \frac{\phi(z_i)}{\mathrm{P}(Y_i = y_i)} = \frac{\phi(z_i)}{\int_{\Phi^{-1}\{\mathrm{F_i}(y_i-1)\}}^{\Phi^{-1}\{\mathrm{F_i}(y_i)\}} \phi(z)dz}.$$

Then, the distribution of $\mathbf{Z} = (Z_1, \ldots, Z_n)$ conditioned on $\mathbf{Y}$ as defined in (3.15) can be shown to be the product of n independent truncated standard normal distribution:

$$\begin{aligned} f(\boldsymbol{z}|\boldsymbol{Y} = \boldsymbol{y}) &= \prod_{i=1}^{n} \mathrm{f}(z_i|Y_i = y_i) \\ &= \prod_{i=1}^{n} \frac{\phi(z_i)}{\int_{\Phi^{-1}\{\mathrm{F}(y_i-1)\}}^{\Phi^{-1}\{\mathrm{F}(y_i)\}} \phi(z)dz} \\ &= \frac{\prod_{i=1}^{n} \left\{\phi(z_i)\right\}}{\prod_{i=1}^{n} \int_{\Phi^{-1}\{\mathrm{F}(y_i-1)\}}^{\Phi^{-1}\{\mathrm{F}(y_i)\}} \phi(z)dz} \\ &= \gamma^{-1}(2\pi)^{-n/2} \exp\left(-\frac{1}{2}\boldsymbol{z}^T\boldsymbol{z}\right), \end{aligned} \tag{3.18}$$

where the normalizing constant γ is given by

$$\gamma = \prod_{i=1}^{n} \int_{\Phi^{-1}\{\mathrm{F}(y_i-1)\}}^{\Phi^{-1}\{\mathrm{F}(y_i)\}} \phi(z)dz = \prod_{i=1}^{n} P(Y_i = y_i).$$

Then (3.15) is

$$
\begin{aligned}
& E_{\mathbf{Z}}\left[|\Sigma|^{-1/2}\exp\Big[-\frac{1}{2}\mathbf{Z}^T(\Sigma^{-1}-\mathbf{I}_n)\mathbf{Z}\Big]\prod_{i=1}^{n}\mathrm{f}_i(y_i)\right] \\
=& E_{\mathbf{Z}}\left[|\Sigma|^{-1/2}\exp\Big[-\frac{1}{2}\mathbf{Z}^T(\Sigma^{-1}-\mathbf{I}_n)\mathbf{Z}\Big]\prod_{i=1}^{n}\mathrm{P}(Y_i=y_i)\right] \\
=& E_{\mathbf{Z}}\left[|\Sigma|^{-1/2}\exp\Big[-\frac{1}{2}\mathbf{Z}^T(\Sigma^{-1}-\mathbf{I}_n)\mathbf{Z}\Big]\gamma\right] \\
=& \int_{\Phi^{-1}[F_1(y_1-1)]}^{\Phi^{-1}[F_1(y_1)]}\cdots\int_{\Phi^{-1}[F_n(y_n-1)]}^{\Phi^{-1}[F_n(y_n)]}\left\{|\Sigma|^{-1/2}\exp\Big[-\frac{1}{2}\mathbf{z}^T(\Sigma^{-1}-\mathbf{I}_n)\mathbf{z}\Big]\gamma\right\}f(\mathbf{z}|\boldsymbol{Y}=\boldsymbol{y})d\mathbf{z} \\
=& \int_{\Phi^{-1}[F_1(y_1-1)]}^{\Phi^{-1}[F_1(y_1)]}\cdots\int_{\Phi^{-1}[F_n(y_n-1)]}^{\Phi^{-1}[F_n(y_n)]}\left\{|\Sigma|^{-1/2}\exp\Big[-\frac{1}{2}\mathbf{z}^T(\Sigma^{-1}-\mathbf{I}_n)\mathbf{z}\Big]\gamma\right\} \\
& \times\left\{\gamma^{-1}(2\pi)^{-n/2}\exp\left(-\frac{1}{2}\mathbf{z}^T\mathbf{z}\right)\right\}d\mathbf{z} \\
=& \int_{\Phi^{-1}[F_1(y_1-1)]}^{\Phi^{-1}[F_1(y_1)]}\cdots\int_{\Phi^{-1}[F_n(y_n-1)]}^{\Phi^{-1}[F_n(y_n)]}\left\{(2\pi)^{-n/2}|\Sigma|^{-1/2}\exp\left(-\frac{1}{2}\mathbf{z}^T\Sigma^{-1}\mathbf{z}\right)\right\}d\mathbf{z},
\end{aligned}
$$

which is equal to (3.16).

We have shown the equivalence between equation (3.15) and equation(3.16) not only for bivariate vectors but also for multivariate vectors of any dimension.

3.3.5 Conclusion

The objective of this section is to compare two joint probability mass functions. One is proposed by Madsen and Fang (2011) and the other one is proposed by Song et al. (2009). They both model the dependence between discrete random variables based on Gaussian copula. In this section, we demonstrate the similarity between these two joint probability mass functions when we assume that the marginals are from Bernoulli distributions.

3.4 Measures of Association

After we introduce the marginal distributions, we are interested in a measure of dependence for count data. As we know, statisticians usually use the term "measure of association" for the measure of dependence. Kendall's τ and Spearman's ρ are examples. When using copulas to model dependence structure, we are interested in the measure of association. For continuous random variables, many dependence concepts and measures of association can be expressed in the corresponding copula only. However, this interrelationship generally fails as soon as there are discontinuities in marginal distribution functions. Genest and Nešlehová (2007) have recently introduced a generalization of rank correlation measure for non-continuous random variables. They focus on empirical distributions corresponding to bivariate random samples. In addition, Gibbons (1985) proposes a measure of Spearman's ρ for count data with tied observations, based on rank calculation. Both of the aforementioned methods try to obtain a sample version of Spearman's ρ. However, the question whether they yield the same results is not well understood.

Assume that a sample $\{X_k, Y_k\}_{k=1}^n$ of size n is from an arbitrary discrete bivariate distribution function H, with marginals F and G, respectively. H is not necessarily continuous, since ties in the observations are possible. Let $\hat{\mathrm{H}}_n, \hat{\mathrm{F}}_n$ and $\hat{\mathrm{G}}_n$ denote the empirical distribution functions. For example, $\hat{\mathrm{F}}_n(x)$ is defined as

$$\hat{\mathrm{F}}_n(x) = \frac{\#\text{ elements in the sample} \leq x}{n} = \frac{1}{n}\sum_{i=1}^{n} 1\{X_i \leq x\},$$

where $1\{X_i \leq x\} = 1$ if $X_i \leq x$, otherwise 0. Similarly, $\hat{\mathrm{G}}_n(y) = \frac{1}{n}\sum_{i=1}^{n} 1\{Y_i \leq y\}$ and $\hat{\mathrm{H}}_n(x,y) = \frac{1}{n}\sum_{i=1}^{n} 1\{X_i \leq x, Y_i \leq y\}$. Suppose that there are r distinct values of X_k's, $\xi_1 < \ldots < \xi_r$, and s distinct values of Y_k's, $\eta_1 < \ldots < \eta_s$. Furthermore, let $u_i = \#\{X_k | X_k = \xi_i\}, v_i = \#\{Y_k | Y_k = \eta_j\}$, and $w_{ij} = \#\{(X_k, Y_k) | X_k = \xi_i, Y_k = \eta_j\}$. Thus the corresponding frequencies are $p_i = \frac{u_i}{n}, q_j = \frac{v_j}{n}$ and $h_{ij} = \frac{w_{ij}}{n}$. Now, let's consider component-wise order statistics and ranks. When where $1 \leq i \leq n$, the order statistics

will be denoted by $X_{(i)}$ and $Y_{(i)}$, respectively. Since there are possible ties on both X and Y, the ranks of tied values can be defined as one of the following three forms:

First: The ordinary rank

$$\mathrm{R}(x_k) = \sum_{i=1}^{n} 1_{(x_i \le x_k)} \quad \text{and} \quad \mathrm{R}(y_k) = \sum_{i=1}^{n} 1_{(y_i \le y_k)};$$

Second: The mid-rank, for $x_k = \xi_i$,

$$\bar{\mathrm{R}}(x_k) = \begin{cases} \frac{1+\ldots+u_1}{u_1} = \frac{u_1+1}{2} & \text{if} \quad i = 1 \\ \sum_{j=1}^{i-1} u_j + \frac{u_i+1}{2} & \text{if} \quad i > 1 \end{cases}.$$

$\bar{\mathrm{R}}(y_k)$ analogously denotes the mid-rank of y_k;

Third: The random rank, where x's are ranked from smallest to largest using the integer $1, \ldots, n$, and y's are ranked likewise.

The most common practice for dealing with tied observations is to assign equal rank to tied observations. For example, both Nešlehová (2007) and Gibbons (1985) use the mid-rank for tied observations.

First, let's review the method proposed Nešlehová (2007) and derive the generalized statement for it. From equation (31) in Nešlehová (2007), Spearman's ρ corresponding to the empirical distribution function $\hat{\mathrm{H}}_n$ is equal to the sample version of Spearman's ρ,

$$\hat{\rho} = \frac{\sum_{k=1}^{n} (\bar{R}(x_k) - \bar{R}_x)(\bar{R}(y_k) - \bar{R}_y)}{\sqrt{\sum_{k=1}^{n} (\bar{R}(x_k) - \bar{R}_x)^2 \sum_{k=1}^{n} (\bar{R}(y_k) - \bar{R}_y)^2}}, \tag{3.19}$$

where $\bar{\mathrm{R}}_x$ and $\bar{\mathrm{R}}_y$ are given by $\bar{\mathrm{R}}_x = \frac{1}{n} \sum_{i=1}^{n} \bar{\mathrm{R}}(x_i)$ and $\bar{\mathrm{R}}_y = \frac{1}{n} \sum_{i=1}^{n} \bar{\mathrm{R}}(y_i)$.

Since addition is commutative, we have the following result for the sum of the rank

$$\begin{aligned}\sum_{k=1}^{n} \mathrm{R}(x_k) &= \sum_{k=1}^{r} u_k \bar{\mathrm{R}}(x_k) \\ &= \sum_{k=1}^{r} u_k [\sum_{i=1}^{k-1} u_i + \frac{1+u_k}{2}] \\ &= \sum_{k=1}^{r} u_k \frac{(\sum_{i=1}^{k-1} u_i + 1) + \ldots + (\sum_{k=1}^{k-1} u_i + u_k)}{u_k} \\ &= \sum_{i=1}^{n} i \\ &= \frac{n(n+1)}{2}.\end{aligned} \tag{3.20}$$

Hence, we have constant values for the sum of the rank from all samples. Then the average of the rank for variables $\boldsymbol{X}$ is

$$\bar{\mathrm{R}}_x = \frac{\sum_{k=1}^{n} \mathrm{R}(x_k)}{n} = \frac{n(n+1)/2}{n} = \frac{n+1}{2}. \tag{3.21}$$

Similarly, the average of the rank for variables $\boldsymbol{Y}$ is $\bar{\mathrm{R}}_y = \frac{n+1}{2}$. Now we investigate the sum of squares

$$\sum_{k=1}^{n} (R(x_k) - \bar{R}_x)^2 = \sum_{k=1}^{n} [R(x_k)]^2 - \frac{n(n+1)^2}{4}.$$

If there are u_i tied observations for ξ_i within the sample and each ξ_i is assigned the

appropriate mid-rank, then the sum of the squared rank is

$$\begin{aligned}\sum_{k=1}^{n}[\mathrm{R}(x_k)]^2 &= \sum_{k=1}^{r} u_k[\bar{\mathrm{R}}(x_k)]^2 \\ &= \sum_{k=1}^{r} u_k[\sum_{i=1}^{k-1} u_i + \frac{1+u_k}{2}]^2 \\ &= \sum_{k=1}^{r} u_k\{[\sum_{i=1}^{k-1} u_i]^2 + (u_k+1)\sum_{i=1}^{k-1} u_i + \frac{(u_k+1)^2}{4}\} \\ &= \sum_{k=1}^{r} u_k\{p_k^2 + (u_k+1)p_k + \frac{(u_k+1)^2}{4}\} \\ &= \sum_{k=1}^{r}\{u_k p_k^2 + u_k(u_k+1)p_k + \frac{u_k(u_k+1)^2}{4}\}, \end{aligned} \tag{3.22}$$

where $p_k = \sum_{i=1}^{k-1} u_i$.

Suppose that all x's take distinct values. We can still divide them into n groups in ascending order, with each group of size $u_k = 1$, for $k = 1, 2, ..., n$. Denote $x_k^{(i)}$ as the i-th smallest element in group k, then $R(x_k^{(i)}) = \sum_{i=1}^{k-1} u_i + i = p_k + i$, where $1 \le i \le u_k$. Now the corresponding sum of squared rank in the absence of ties is

$$\begin{aligned}\sum_{k=1}^{n}[\mathrm{R}(x_k)]^2 &= \sum_{k=1}^{r}\sum_{i=1}^{u_k}[\mathrm{R}(x_k^{(i)})]^2 \\ &= \sum_{k=1}^{r}\sum_{i=1}^{u_k}[p_k + i]^2 \\ &= \sum_{k=1}^{r}\sum_{i=1}^{u_k}\{p_k^2 + 2ip_k + i^2\} \\ &= \sum_{k=1}^{r}\{u_k p_k^2 + 2p_k\sum_{i=1}^{u_k} i + \sum_{i=1}^{u_k} i^2\} \\ &= \sum_{k=1}^{r}\{u_k p_k^2 + u_k(u_k+1)p_k + \frac{u_k(u_k+1)(2u_k+1)}{6}\}. \end{aligned} \tag{3.23}$$

Comparing equation (3.23) and equation (3.22), we see that this particular tie group $(u_1, u_2, ..., u_k)$ reduces the sum of squared rank by

$$\sum_{k=1}^{r}\{\frac{u_i(u_k+1)(2u_k+1)}{6} - \frac{u_k(u_k+1)^2}{4}\} = \sum_{k=1}^{r}\frac{u_k(u_k^2-1)}{12}.$$

Since the sum of squares for $R(x_k)$ in the absence of ties is

$$\begin{aligned}\sum_{k=1}^{n}[\mathrm{R}(x_k^{(k)}) - \bar{\mathrm{R}}_x]^2 &= \sum_{k=1}^{n}[\mathrm{R}(x_k^{(k)}) - \frac{n+1}{2}]^2 \\ &= \sum_{k=1}^{n} k^2 - \frac{(n+1)^2}{4} \\ &= \frac{n(n^2-1)}{12},\end{aligned} \tag{3.24}$$

the sum of squares for $\mathrm{R}(x_k)$ with ties is simply

$$\begin{aligned}\sum_{k=1}^{n}[\bar{\mathrm{R}}(x_k) - \bar{\mathrm{R}}_x]^2 &= \sum_{k=1}^{n}[\bar{\mathrm{R}}(x_k) - \frac{n+1}{2}]^2 \\ &= \frac{n(n^2-1)}{12} - \sum_{k=1}^{r}\frac{u_k(u_k^2-1)}{12}.\end{aligned} \tag{3.25}$$

According to Nešlehová (2007), $\mathrm{cov}(\tilde{X}, \tilde{Y})$ and $\mathrm{var}(\tilde{X})$ are defined as

$$\mathrm{cov}(\tilde{X}, \tilde{Y}) = \frac{1}{n^3}\sum_{k=1}^{n}(\bar{\mathrm{R}}(x_k) - \bar{\mathrm{R}}_x)(\bar{\mathrm{R}}(y_k) - \bar{\mathrm{R}}_y)$$

and

$$\mathrm{var}(\tilde{X}) = \frac{1}{n^3}\sum(\bar{\mathrm{R}}(x_k) - \bar{\mathrm{R}}_x)^2,$$

where random variables $\tilde{X}$ and $\tilde{Y}$ are defined as $\tilde{X} = \frac{\hat{\mathrm{F}}_n(\xi_i)+\hat{\mathrm{F}}_n(\xi_{i-1})}{2}$, and $\tilde{Y} = \frac{\hat{\mathrm{G}}_n(\eta_i)+\hat{\mathrm{G}}_n(\eta_{i-1})}{2}$. Then the sample version of Spearman's ρ defined in Nešlehová (2007) can be re-expressed as

$$\begin{aligned}\rho(X,Y) &= \mathrm{corr}(\tilde{X}, \tilde{Y}) \\ &= \frac{\mathrm{cov}(\tilde{X}, \tilde{Y})}{\sqrt{\mathrm{var}(\tilde{X})\,\mathrm{var}(\tilde{Y})}} \\ &= \frac{1/n^3\Big[\sum_{k=1}^{n}\bar{R}(x_k)\bar{R}(y_k) - n(n+1)^2/4\Big]}{\sqrt{\{1/n^3\Big[\frac{n(n^2-1)}{12} - \sum_{k=1}^{r}\frac{u_k(u_k^2-1)}{12}\Big]\}\{1/n^3\Big[\frac{n(n^2-1)}{12} - \sum_{k=1}^{r}\frac{v_k(v_k^2-1)}{12}\Big]\}}} \\ &= \frac{12\Big[\sum_{k=1}^{n}\bar{\mathrm{R}}(x_k)\bar{\mathrm{R}}(y_k) - n(n+1)^2/4\Big]}{\sqrt{\Big[n(n^2-1) - \sum_{k=1}^{r}u_k(u_k^2-1)\Big]\Big[n(n^2-1) - \sum_{k=1}^{r}v_k(v_k^2-1)\Big]}}.\end{aligned} \tag{3.26}$$

The expression in (3.26) is exactly the same as the equation (3.17) from page 234 in Gibbons (1985). Our results show that Gibbons (1985) and Nešlehová (2007) provide the same formula for the sample version of Spearman's ρ. Although the expressions used are different, mathematically they yield the same result.

3.5 Conclusion

In this chapter, we divided the model used in chapter 2 into three parts: marginal distribution, Gaussian copula-based dependence model, and the measure of association, and we compared these three parts to existing methods.

In Section 2.4.4, we used the starred copula to extended the non-continuous marginals to continuous marginals to obtain a unique copula; and we discussed another continuous extension, the standard extension copula. Both the starred copula and the standard extension copula are continuous extensions for count margins and used to model dependence between count data. We performed a detailed comparison of these two extensions. In essence, these two methods can be used to obtain the same unique copula. The advantage of bilinear interpolation is that it is fast and simple to implement, and both of copulas apply a bilinear interpolation. However, the places where a bilinear interpolation is used are different. Despite that, they achieve the same theoretical result.

The idea for the copula-based spatially correlated count model discussed in chapter 2 is derived from the Gaussian dispersion model suggested by Song et al. (2009). However, it is intractable when the length of the vector is more than 4. To avoid the n-fold summation as well as to obtain a unique copula, Madsen and Fang (2011) introduce the expected likelihood approach, which originates from joint truncated multivariate normal distribution. The results obtained by this method are the same as in Song et al. (2009).

In the measure of association for count data, we have used the sample version of

Spearman's ρ. The performance of Spearman's ρ has been illustrated for bivariate discrete random variables. However, ties do occur within the samples; thus, the problem of ties within ranks is considered. Currently Gibbons (1985) and Nešlehová (2007) use the sample version of Spearman's ρ. In section 3.4, we have demonstrated that the results obtained from both methods are essentially the same.

In short, the objective of this section was to explore our models used in chapter 2 and existing methods for modeling dependence structure for count data based on copula model, and measuring the sample version of Spearman's ρ for count data. We found that despite differences in how the methods appear, they provide us with the same results.

4 COMPARISON OF TWO METHODS FOR CHOOSING THE BEST COPULA

Yan Fang* and Lisa Madsen**

Department of Statistics, Oregon State University, Corvallis, Oregon 97331, U.S.A.

*E-mail: fangya@science.oregonstate.edu

**http://www.stat.oregonstate.edu/people/lmadsen

4.1 Abstract

Modeling in finance often involves the selection of the best available copula from a set of a candidate copula families. Goodness-of-fit tests are the most popular procedure to select the best copula for a particular data set. Some of them use a multi dimensional test, while others reduce the multivariate problem to the univariate one and then apply a univariate test. However, there is no recommended way to suggest which particular goodness-of-fit test should be used in a specific case. Recently some authors suggested using "blanket tests" to do goodness-of-fit test especially for large scale simulation. Kojadinovic et al. (2011) used a multiplier procedure to do a goodness-of-fit test. Although this method is shown to be asymptotically valid, it requires the sample size to be at least 300 in order to get a reliable result. This paper investigates the feasibility of using Akaike Information Criterion (AIC) to choose the best copula from a series of candidate copula models. Under appropriate conditions, if we use AIC to do copula selection, it is not necessary for the sample size to be at least 300. Furthermore, using AIC is faster than using a multiplier procedure to do goodness-of-fit. The present research will use simulation to compare the properties of these two procedures.

4.2 Introduction

Let $\mathbf{X} = (X_1, \ldots, X_p)^T$ be a random vector with continuous marginal cumulative distribution functions (c.d.f.s) $F_1, \ldots, F_p$. Sklar (1959) showed that the multivariate distribution function H of $\mathbf{X}$ can be uniquely represented as

$$\mathrm{H}(\mathbf{x}) = \mathrm{C}\left(\mathrm{F}_1(x_1), \ldots, \mathrm{F}_p(x_p)\right), \mathbf{x} \in \mathbb{R}^p, \tag{4.1}$$

where $\mathrm{C} : [0,1]^p \rightarrow [0,1]$, called a copula, is a p-dimensional c.d.f with the standard uniform margins.

Many multivariate models for dependence can be generated by parametric families C_θ of copulas. For reviews of copula models, see Joe (1997) and Nelsen (2006). Practical applications of this modeling approach are found in fields such as finance (Cherubini et al. 2004; McNeil et al. 2005), hydrology (Genest et al. 2007), public health and medical (Wang and Wells 2000), and actuarial science (Frees and Valdez 1998; Klugman and Parsa 1999).

All of the models used in the fields mentioned above are based on a parametric copula and nonparametric marginal distributions. This class of semi-parametric multivariate distributions has gained popularity in diverse fields due to its flexibility in separating the dependence structure and the marginal behaviors of a multivariate random variable. A commonly used approach to estimate the semi-parametric multivariate copula distribution is the classical maximum likelihood estimator using the two-step inference carried by Joe (1997), Shih and Louis (1995) and Genest et al. (1995).

Assume that the unknown copula C belongs to an absolutely continuous parametric family $\mathcal{C}_0 = \{\mathrm{C}_\theta : \theta \in \mathcal{O}\}$, where $\mathcal{O}$ is an open subset of $\mathbb{R}^q$ for some $q = \{1, 2, 3, \ldots\}$ and the vector of copula parameters $\theta = (\theta_1, \ldots, \theta_q)$ is estimated from the random sample $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$. When estimating the parameters for a semi-parametric copula, a natural estimation method is the pseudo-likelihood approach introduced in Genest et al. (1995)

and Shih and Louis (1995). It consists of maximizing the log pseudo-likelihood

$$\theta_n = \arg\max_{\theta} \sum_{i=1}^{n} \log \mathrm{c}_\theta\{\hat{U}_{i1}, \ldots, \hat{U}_{ip}\}, \tag{4.2}$$

where c_θ is the density function of the parametric copula $\mathrm{C}_\theta \in \mathcal{C}_0$, and the $\hat{\mathbf{U}}_i = (\hat{U}_{i1}, \ldots, \hat{U}_{ip})$ are the pseudo–observations or the rescaled empirical distribution of $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})$, namely,

$$\hat{U}_{ij} = \frac{R_{ij}}{n+1},$$

where R_{ij} is the rank of X_{ij} among $(X_{1j}, \ldots, X_{nj})$.

However, a choice of the functional form for the copula is an open question in the literature. There is no recommended method to check whether the dependency structure on which the previously discussed estimation is based is appropriately modeled by a chosen copula. Usually, we want to test the following hypothesis

$$\mathrm{H}_0 : \mathrm{C} \in \mathcal{C}_0 = \{\mathrm{C}_\theta : \theta \in \mathcal{O}\}.$$

Goodness-of-fit (GOF) tests are able to reject or fail to reject this null hypothesis. Up to now, several contributions have been made to test this hypothesis, e.g. Genest and Rivest (1993), Shih and Louis (1995), Breymann et al. (2003), Fermanian (2005), Dobrić (2007), Genest et al. (2006), Genest (2008), Genest et al. (2009), Kojadinovic (2009), Kojadinovic (2011), Kojadinovic et al. (2011), and so on . However, general guidelines and recommendations are sparse.

The Akaike Information Criterion (AIC) developed by Akaike (1974) is a measure of the relative GOF test of a statistical model. Although AIC does not provide a test of a model in the usual sense of testing a null hypothesis, AIC provides a measure for comparison among copulas, i.e. it is a tool for copula selection. One of the benefits of AIC is that it takes less time than either the multiplier method suggested by Kojadinovic (2011) or the parametric bootstrap approach used by Genest (2008). Furthermore, AIC can tell that a copula model is better than another model if it has a smaller AIC value,

whereas GOF test may fail to reject H_0 for a number of models. Although AIC may not identify best, similar AIC's suggest similar model "goodness".

The purpose of this paper is to present a critical review of AIC as a tool for choosing the best copula from a series of candidates and to compare the relative effectiveness of this procedure with the multiplier method through a simulation study involving a large number of copula alternatives and dependence conditions. We provide some basic theory regarding copulas, pseudo-observations and pseudo-likelihood in Section 4.3. Both multiplier GOF test and AIC is reviewed in Section 4.4. The simulation results for five different copula families, specifically, Clayton, Frank, Gumbel, normal and t copulas, are listed in Section 4.5.

4.3 Basic theory

For the purpose of this research we are focusing on the fit of the copula alone. We make no assumptions about the marginal distributions. We will use empirical margins to transform the observed data set into pseudo observations.

4.3.1 Copula basics

The definition of a p-dimensional copula is a multivariate distribution C with uniform $(0, 1)$ margins. According to Sklar's theorem (Sklar 1959), any multivariate distribution function H with margins c.d.f.s $F_1, \ldots, F_p$ can be written as Equation (4.1) for some copula C. If all the margins are continuous, then C is unique. So if we have a random $\mathbf{X} = (X_1, \ldots, X_p)^T$, the copula of their joint distribution function may be derived from Equation (4.1), i.e.,

$$C(u_1, \ldots, u_p) = H\left(F_1^{-1}(u_1), \ldots, F_p^{-1}(u_p)\right), \tag{4.3}$$

where the $\mathrm{F}_i^{-1}(\cdot)$'s are the inverse of the marginal c.d.f.s. The copula density is given by

$$\mathrm{c}(u_1, \ldots, u_p) = \frac{\mathrm{h}(\mathrm{F}_1^{-1}(u_1), \ldots, \mathrm{F}_p^{-1}(u_p))}{\prod_{i=1}^{p} \mathrm{f}_i(\mathrm{F}_i^{-1}(u_i))} \tag{4.4}$$

where $\mathrm{f}_i(\cdot)$ is the probability density function (p.d.f) for the variable $\mathrm{F}_i^{-1}(u_i)$ and $\mathrm{h}(\cdot)$ is the joint p.d.f for multivariate $(\mathrm{F}_1^{-1}(u_1), \ldots, \mathrm{F}_p^{-1}(u_p))$.

4.3.2 Empirical Distribution

The empirical marginal c.d.f for n observations $X_{1i}, \ldots, X_{ni}$ of a variable $\mathbf{X}_i$ is

$$\hat{\mathrm{F}}_i(x_i) = \frac{1}{n} \sum_{j=1}^{n} \mathrm{I}(X_{ji} \leq x_i) \quad i = 1, \ldots, p, \tag{4.5}$$

where $\mathrm{I}(\cdot)$ is the indicator function, taking the value 1 if $X_{ji} \leq x$ and 0 otherwise. The empirical distribution function estimates the true underlying c.d.f of the points in the sample. Using the empirical marginal c.d.f.s, the empirical copula is given by

$$\mathrm{C}_n(\mathbf{u}) = \frac{1}{n} \sum_{j=1}^{n} \mathrm{I}(\hat{\mathrm{F}}_1(X_{j1}) \leq u_1, \ldots, \hat{\mathrm{F}}_p(X_{jp}) \leq u_p) \tag{4.6}$$

with $\mathbf{u} = (u_1, \ldots, u_p)$.

4.3.3 Pseudo likelihood function

The pseudo-observation for n observations $X_{k1}, \ldots, X_{kp}$ of a variable X_k is

$$\hat{\mathbf{U}}_k = (\hat{U}_{k1}, \ldots, \hat{U}_{kp}) = (\frac{\mathrm{R}_{k1}}{n+1}, \ldots, \frac{\mathrm{R}_{kp}}{n+1}) = (\frac{n}{n+1}\hat{\mathrm{F}}_1(X_{k1}), \ldots, \frac{n}{n+1}\hat{\mathrm{F}}_p(X_{kp})), \tag{4.7}$$

where $\hat{\mathrm{F}}_i$ is the empirical c.d.f computed from $X_{1i}, \ldots, X_{ni}$ according to Equation (4.5). Then the empirical copula (4.6) can be calculated by using pseudo-observation as

$$\mathrm{C}_n(\mathbf{u}) = \frac{1}{n} \sum_{j=1}^{n} \mathrm{I}(\hat{\mathrm{U}}_{j1} \leq u_1, \ldots, \hat{\mathrm{U}}_{jp} \leq u_p). \tag{4.8}$$

Suppose that $\mathbf{X}_1, \ldots, \mathbf{X}_n$ is a random sample from c.d.f $\mathrm{C}_\theta(\mathrm{F}_1(x_1), \ldots, \mathrm{F}_p(x_p))$, where $\mathrm{F}_1, \ldots, \mathrm{F}_p$ are continuous c.d.f.s and $\mathrm{C}_\theta \in \mathcal{C}_0$ is an absolutely continuous copula such that $\theta \in \mathcal{O}$, with $\mathcal{O}$ being an open subset of $\mathbb{R}^q$ for some $q = \{1, 2, 3, \ldots\}$. For

any $i \in \{1, \ldots, n\}$, we can use the pseudo-observations $\hat{\mathbf{U}}_i$ instead of F_i in both c.d.f $\mathrm{C}_\theta(\hat{\mathrm{U}}_{i1}, \ldots, \hat{\mathrm{U}}_{ip})$ and p.d.f $\mathrm{c}_\theta(\hat{\mathrm{U}}_{i1}, \ldots, \hat{\mathrm{U}}_{ip})$. For more information, please refer to Kojadinovic et al. (2011).

We can construct the log pseudo-likelihood function (LPLF) based these pseudo-observations, more precisely,

$$l(\theta) = \sum_{i=1}^{n} \log \mathrm{c}_\theta(\hat{U}_{i1}, \ldots, \hat{U}_{ip}). \tag{4.9}$$

4.3.4 Maximum pseudo-likelihood estimation

The maximum pseudo-likelihood estimator (MPLE) θ_n (4.2) is computed from the pseudo-observation $\hat{\mathrm{U}}_1, \ldots, \hat{\mathrm{U}}_p$ by maximizing Equation (4.9).

In order to implement MPLE for a given copula family and do a GOF test, it is necessary that the vector of dependence parameters θ can be estimated from the available sample. We follow Kojadinovic (2009) and use R optim function (R Development Core Team 2009) to find the MPLE.

4.4 GOF test and model selection

We now introduce the GOF test using the multiplier method and model selection using AIC, respectively.

4.4.1 Multiplier method

This section describes a fast multiplier method proposed by Kojadinovic et al. (2011). The GOF test implemented here is based on the empirical copula (Deheuvels, 1979 and 1981b), which is a consistent estimator of the unknown copula C. The cardinal principle of this test has been studied by Deheuvels (1981a) and Genest and Rémillard (2004). The idea is to compare the empirical copula $\mathrm{C}_n(\mathbf{u})$ defined in Equation (4.8) with

an estimator $\mathrm{C}_{\theta_n}(\mathbf{u})$ of parametric copula $\mathrm{C}_\theta(\mathbf{u})$ obtained by assuming that $\mathrm{H}_0 : \mathrm{C} \in \mathcal{C}_0 = \{\mathrm{C}_\theta : \theta \in \mathcal{O}\}$ holds. Here, θ_n defined in Equation (4.2) is an estimator of θ. The natural way is to consider the distance between empirical and null hypothesis distribution functions. That is, under suitable regularity conditions, the empirical copula process is

$$\mathbb{C}_n(\mathbf{u}) = \sqrt{n}\{\mathrm{C}_n(\mathbf{u}) - \mathrm{C}_{\theta_n}(\mathbf{u})\}, \qquad \mathbf{u} = (u_1, \ldots, u_p) \in [0,1]^p. \tag{4.10}$$

The test statistic, the Cramér-von Mises Statistic, is defined as

$$\mathrm{S}_n = \int_{[0,1]^p} \mathbb{C}_n^2(\mathbf{u}) d\mathrm{C}_n(\mathbf{u}) = \sum_{i=1}^{n} \{\mathrm{C}_n(\hat{\mathbf{U}}_i) - \mathrm{C}_{\theta_n}(\hat{\mathbf{U}}_i)\}^2. \tag{4.11}$$

In many cases, the asymptotic distribution of test statistic S_n derived from the process $\mathbb{C}_n(\mathbf{u})$ depends on the unknown distribution $\mathrm{C}_{\theta_n}(\mathbf{u})$. To solve this problem, Stute et al. (1993) suggested the "parametric bootstrap" procedure. With the parametric bootstrapping, we treat the data generated from $\mathrm{C}_{\theta_n}(\mathbf{u})$ as if they are an accurate reflection of the copula $\mathrm{C}_\theta(\mathbf{u})$, and then draw many bootstrapped samples from copula $\mathrm{C}_{\theta_n}(\mathbf{u})$. Define $\mathrm{C}_n^*(\mathbf{u})$ and $\mathrm{C}_{\theta_n}^*(\mathbf{u})$ as the empirical copula and the estimator of $\mathrm{C}_\theta(\mathbf{u})$ from the bootstrapped samples, respectively. Then, $\mathrm{C}_n^*(\mathbf{u})$ and $\mathrm{C}_{\theta_n}^*(\mathbf{u})$ are analogs of $\mathrm{C}_n(\mathbf{u})$ and $\mathrm{C}_{\theta_n}(\mathbf{u})$ computed for a sample from $\mathrm{C}_\theta(\mathbf{u})$. And the empirical processes $\sqrt{n}\{\mathrm{C}_n(\mathbf{u}) - \mathrm{C}_{\theta_n}(\mathbf{u})\}$ for different samples from $\mathrm{C}_\theta(\mathbf{u})$ then converge jointly in distribution to the same limit. Accordingly, an approximate p-value for S_n can be obtained by means of the parametric bootstrap-based procedure, whose validity was shown by Genest (2008). However, there are some weaknesses in this approach. The main inconvenience is its very high computational cost, as each iteration requires both random number generation from the fitted copula and estimation of the copula parameters. As the sample size increases, the application of the parametric bootstrap-based GOF test becomes prohibitive. In order to overcome this flaw, Kojadinovic et al. (2011) proposed a fast large-sample testing procedure based on the multiplier central limit theorems inspired by Remillard and Scaillet (2009) and Kojadinovic (2009).

Let's define a sequence of the distributed processes $\{\mathbb{J}_i\}_{i=1}^n$, that is,

$$\mathbb{J}_i(\mathbf{u}) = \alpha_\theta(\mathbf{u}) - \sum_{j=1}^{p} \mathrm{C}_\theta^{[j]}(\mathbf{u})\alpha_\theta(1,\ldots,1,u_j,1,\ldots,1) - \Theta \times \dot{\mathrm{C}}_\theta(\mathbf{u})$$

with

$$\alpha_\theta(\mathbf{u}) = I(\mathbf{U}_i \leq \mathbf{u}) - \mathrm{C}_\theta(\mathbf{u}),$$

$$\mathrm{C}_\theta^{[j]}(\mathbf{u}) = \frac{\mathrm{C}_\theta(u_1,\ldots,u_{j-1},u_j+n^{-1/2},u_{j+1},\ldots,u_p) - \mathrm{C}_\theta(u_1,\ldots,u_{j-1},u_j-n^{-1/2},u_{j+1},\ldots,u_p)}{2n^{-1/2}},$$

$$\dot{\mathrm{C}}_\theta(\mathbf{u}) = \frac{\partial \mathrm{C}_\theta(\mathbf{u})}{\partial\theta},$$

and

$$\Theta = \left[E_{\mathrm{c}_\theta}\left\{\frac{\dot{\mathrm{c}}_\theta(\mathbf{u})\dot{\mathrm{c}}_\theta^T(\mathbf{u})}{\mathrm{c}_\theta^2(\mathbf{u})}\right\}\right]^{-1} \times \left[\frac{\dot{\mathrm{c}}_\theta(\mathbf{U}_i)}{\mathrm{c}_\theta(\mathbf{U}_i)} - \sum_{j=1}^{p}\int_{[0,1]^p}\{I(U_{ij}\leq u_j)-u_j\} \times \frac{\mathrm{c}_\theta^{(j)}(\mathbf{u})}{\mathrm{c}_\theta(\mathbf{u})}\frac{\dot{\mathrm{c}}_\theta(\mathbf{u})}{\mathrm{c}_\theta(\mathbf{u})}d\mathrm{C}_\theta(\mathbf{u})\right]$$

where $\dot{\mathrm{c}}_\theta(\mathbf{u}) = \frac{\partial \mathrm{c}_\theta(\mathbf{u})}{\partial\theta}$ and $\mathrm{c}_\theta^{(j)}(\mathbf{u}) = \frac{\partial \mathrm{c}_\theta(\mathbf{u})}{\partial u_j}$. Then $\mathbb{J}_1,\ldots,\mathbb{J}_n$ are independent and identically distributed (i.i.d.) processes whose form depends on the estimator of θ_n and on the hypothesized copula family C_θ. Let N be a large integer and let $Z_i^{(k)}, i \in \{1,\ldots,n\}, k \in \{1,\ldots,\mathrm{N}\}$, be i.i.d. random variables with mean 0 and variance 1 independent of the data $\mathbf{X}$. Under suitable regularity conditions, the GOF process $(\mathbb{C}_n, \mathbb{C}_\theta^{(1)},\ldots,\mathbb{C}_\theta^{(N)})$ defined in (4.10) converges weakly to $(\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbb{J}_i(\mathbf{u}), \frac{1}{\sqrt{n}}\sum_{i=1}^n Z_i^{(1)}\mathbb{J}_i(\mathbf{u}),\ldots,\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_i^{(N)}\mathbb{J}_i(\mathbf{u}))$ (refer to Kojadinovic, 2010a).

Let $\hat{\mathbb{J}}_i(\mathbf{u})$ be the version of $\mathbb{J}_i(\mathbf{u})$ in which all the unknown quantities are replaced by their estimates. Thus, the approximate p-value for the test based on the multiplier method can be obtained by means of the following procedure (see Kojadinovic and Yan 2010 and Kojadinovic et al. 2011, for more details):

Algorithm 4.4.1. ***(Testing procedure for the multiplier method)***

STEP 1 *Compute pseudo-observations* $\hat{\mathbf{U}}_1,\ldots,\hat{\mathbf{U}}_n$ *from the observed data* X *as in Equation (4.7), then use* $\hat{\mathbf{U}}_i$ *to get* C_n *by using Equation (4.8). Estimate the dependence parameter* θ *by using Equation (4.2);*

STEP 2 *Compute the test statistic* S_n *defined in Equation (4.11);*

STEP 3 *Set N to be a large integer, and repeat the following steps for every* $k \in \{1, \ldots, N\}$*:*

3.1 *Generate n i.i.d random variables* $Z_1^{(k)}, \ldots, Z_n^{(k)}$ *from the normal distribution with mean 0 and variance 1;*

3.2 *As in Kojadinovic (2010a), form an approximate independent realization of the test statistic under* H_0 *by*

$$\mathbb{C}_n^{(k)}(\mathbf{u}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i^{(k)} \hat{\mathbb{J}}_i(\mathbf{u});$$

3.3 *Compute an approximate independent realization of* S_n *under* H_0 *by*

$$\mathrm{S}_n^{(k)} = \int_{[0,1]^p} \left\{ \mathbb{C}_n^{(k)}(\mathbf{u}) \right\}^2 d\mathrm{C}_n(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \mathbb{C}_n^{(k)}(\hat{\mathbf{U}}_i) \right\}^2;$$

STEP 4 *An approximate p-value for the test is given by* $\frac{1}{N} \sum_{k=1}^{N} \mathbf{1}(\mathrm{S}_n^{(k)} \geq \mathrm{S}_n)$.

Since $\hat{\mathbb{J}}_i(\mathbf{u})$ only needs to be computed once, this procedure is fast compared to the traditional parametric bootstrap method. However, the derivation and the computation of terms $\hat{\mathbb{J}}_i(\mathbf{u})$ are complicated, since they involve partial derivatives of c.d.f and p.d.f of the hypothesized copula with respect to both variables $\mathbf{u}_i$ and parameters (see Kojadinovic and Yan 2010 and Kojadinovic et al. 2011).

4.4.2 AIC approach

This section introduces the development and application of AIC (Akaike, 1974, 1981) in copula selection. AIC derived in Akaike (1981) is defined as

$$\text{AIC} = -2(\text{ maximum log likelihood}) + 2(\text{number of free parameters}) = -2l(\hat{\theta}) + 2q, \qquad (4.12)$$

where $l(\hat{\theta})$ is the maximized value of the LPLF defined in Equation (4.9) and $\hat{\theta} = \theta_n$ defined in (4.2).

The basic idea underlying the use of AIC for copula selection is to chose the copula which is closest to the unknown true copula. We are trying to reduce the distance between the true copula and the approximate copula. Lehmann and Casella (1998) suggested using the Kullback-Leibler information as a measure of the distance between the true model and the null hypothesized model. The p.d.f of the true unknown copula and the approximate copula model under null hypothesis

for this data set are $\mathrm{c}(\mathbf{u})$ and $\mathrm{c}_\theta(\mathbf{u})$ respectively. Then the Kullback-Leibler information measure between the true p.d.f $\mathrm{c}(\mathbf{u})$ and the approximate p.d.f $\mathrm{c}_\theta(\mathbf{u})$ is defined as

$$\begin{aligned}\mathrm{K}\left(c(\mathbf{u}), c_\theta(\mathbf{u})\right) &= \int \left\{\log\left[\frac{c(\mathbf{u})}{c_\theta(\mathbf{u})}\right]\right\} c(\mathbf{u}) d\mathbf{u} \\ &= \mathrm{E}_{\mathbf{u}}\left\{\log\left[\frac{c(\mathbf{u})}{c_\theta(\mathbf{u})}\right]\right\} \\ &= \mathrm{E}_{\mathbf{u}}\left[\log c(\mathbf{u})\right] - \mathrm{E}_{\mathbf{u}}\left[\log c_\theta(\mathbf{u})\right],\end{aligned}$$

where $\mathrm{E}_{\mathbf{u}}[\cdot]$ denotes the expected value with respect to variable $\mathbf{u}$. The Kullback-Leibler information $\mathrm{K}\left(c(\mathbf{u}), c_\theta(\mathbf{u})\right) \geq 0$, or equivalently, $\mathrm{E}_{\mathbf{u}}\left[\log c(\mathbf{u})\right] \geq \mathrm{E}_{\mathbf{u}}\left[\log c_\theta(\mathbf{u})\right]$, and it equals to 0 if and only if $c(\mathbf{u}) = c_\theta(\mathbf{u})$ happens almost surely (see Chapter 2 from Kullback, 1968). The smaller the Kullback-Leibler information, the closer the approximate copula to the true copula. Alternatively, the larger the quantity $\mathrm{E}_{\mathbf{u}}\left[\log c_\theta(\mathbf{u})\right]$, the closer the function $c_\theta(\mathbf{u})$ to the true copula p.d.f $c(\mathbf{u})$.

Suppose that data $\mathbf{v}$ is generated from $\mathrm{c}(\mathbf{u})$, i.e., a realization of random variable $\mathbf{U}$, and $\hat{\theta}(\mathbf{v})$ is an estimator of θ from the empirical data $\mathbf{v}$, then $\mathrm{E}_{\mathbf{u}}\left[\log c_{\hat{\theta}(\mathbf{v})}(\mathbf{u})\right]$ is a random variable with respect to variable $\mathbf{v}$. Let $\mathrm{E}_{\mathbf{v}}\left[\mathrm{K}\left(c(\mathbf{u}), c_{\hat{\theta}(\mathbf{v})}(\mathbf{u})\right)\right]$, where $\mathrm{E}_{\mathbf{v}}[\cdot]$ denotes the expected value with respcet to variable $\mathbf{v}$, denote the expected Kullback-Leibler information. Since the expected Kullback-Leibler information is an approximately unbiased estimate of the Kullback-Leibler information for large samples, we will use the expected Kullback-Leibler information instead of the Kullback-Leibler information to measure the "distance" between $\mathrm{c}(\mathbf{u})$ and $\mathrm{c}_\theta(\mathbf{u})$. Accordingly, the Kullback-Leibler information is redefined as

$$\begin{aligned}\mathrm{K}\left(c(\mathbf{u}), c_\theta(\mathbf{u})\right) &= \mathrm{E}_{\mathbf{v}}\left[\mathrm{K}\left(c(\mathbf{u}), c_{\hat{\theta}(\mathbf{v})}(\mathbf{u})\right)\right] \\ &= \mathrm{E}_{\mathbf{v}}\left\{\mathrm{E}_{\mathbf{u}}\left[\log c(\mathbf{u})\right] - \mathrm{E}_{\mathbf{u}}\left[\log c_{\hat{\theta}(\mathbf{v})}(\mathbf{u})\right]\right\} \\ &= \mathrm{E}_{\mathbf{u}}\left[\log c(\mathbf{u})\right] - \mathrm{E}_{\mathbf{v}}\left\{\mathrm{E}_{\mathbf{u}}\left[\log c_{\hat{\theta}(\mathbf{v})}(\mathbf{u})\right]\right\}. \qquad (4.13)\end{aligned}$$

Bozdogan (1987) suggested that the AIC is an unbiased estimator of minus twice the mean expected log likelihood, i.e., $\mathrm{AIC} = -2\mathrm{E}_{\mathbf{v}}\left\{\mathrm{E}_{\mathbf{u}}\left[\log c_{\hat{\theta}(\mathbf{v})}(\mathbf{u})\right]\right\}$. Substituting into Equation (4.13), the Kullback-Leibler information can further be expressed as

$$\mathrm{K}\left(c(\mathbf{u}), c_\theta(\mathbf{u})\right) = \mathrm{E}_{\mathbf{u}}\left[\log c(\mathbf{u})\right] + \frac{1}{2}\mathrm{AIC}. \qquad (4.14)$$

Since AIC compares only a part of Kullback-Leibler information, the value of AIC itself is not meaningful. However, we can minimize AIC to minimize the Kullback-Leibler information. In

other words, the smaller the AIC, the closer the function $c_\theta(\mathbf{u})$ is to the true copula p.d.f $c(\mathbf{u})$. In addition, AIC provides a versatile procedure for statistical model identification and is free from the ambiguities inherent in the application of the conventional hypothesis testing procedures.

When there are several competing copulas, we want to know which copula fits the data best. The chosen copula model should be the one that minimizes the Kullback-Leibler information between the copula model and the true unknown copula. We can calculate the AIC for each model with the same data. The "best" model is the one with least AIC value. AIC is more computationally efficient than other copula selection methods. Though it can't do a formal GOF hypothesis test, it can be used to select the best copula from a group of copula families. The practical advantage of AIC in copula analysis will be demonstrated in next section by a simulation study comparing it to the multiplier method.

4.5 Simulation and Results

To assess the performance of the proposed AIC model-selection method, we conduct a simulation study comparing the AIC method with the multiplier method. The simulation procedure for AIC is given in Algorithm (4.5.1).

Algorithm 4.5.1. ***(Simulation procedure for AIC)***

STEP 1 *Set N to be a large integer, and repeat the following steps for every $k \in \{1, \ldots, N\}$:*

1.1 *Generate the random variables $\mathbf{X}^{(k)} = (\mathbf{X}_1^{(k)}, \ldots, \mathbf{X}_n^{(k)})$ from a given copula family;*

1.2 *Compute pseudo-observations $\hat{\mathbf{U}}_1^{(k)}, \ldots, \hat{\mathbf{U}}_n^{(k)}$ from the data set $\mathbf{X}^{(k)}$ by using Equation (4.7), then get the AIC values (4.12) by fitting each candidate copula family;*

STEP 2 *For each copula family, calculate the rate (out of N) that family achieves the least AIC.*

For the multiplier method, the nominal level was fixed at 5% throughout the whole study and the GOF test is based on 1000 iterations. The rejection proportions is calculated under a variety of alternatives.

Five one-parameter copula families were considered: Clayton, Frank, Gumbel-Hougarrd, Normal, and t with arbitrarily setting $\nu = 5$. Each copula family is used as hypothesized family

as well as data generating family. For each copula family, four dependence levels (0.2, 0.4, 0.6 and 0.8) were considered corresponding to Kendall's tau. Three different sample sizes (n=100, 300 and 500) are used. Three dimension sizes (2-variate, 3-variate and 4-variate) are tested.

There are $4 \times 3 \times 3 = 36$ scenarios. For each scenario, we will

1. conduct the hypothesis test for each of the five copula families (multiplier method) and

2. compute the AIC's for each family.

The procedure is repeated for 1000 simulated datasets to obtain the empirical levels (i.e, using the rejection proportions from 1000 replicates at level of significance 5%) for a hypothesized copula model via the multiplier method, and the rates of least AIC among the multiple copula families from 1000 replications.

the empirical level (i.e, using the rejection proportions from 1000 replicates at level of significance 5%)

According to the simulation results, the true copula used to generate the random sample gives the highest rate of least AIC. In the following, we will call this the "correct rate" (i.e. the probability that the true copula gives the least AIC in 1000 repetition). We will focus on comparing this correct rate and the empirical levels from the multiplier method.

Tables 4.1, 4.2 and 4.3 compare the performance of AIC with the multiplier method for sample size $n = 100$ with dimension $d = 2$, 3 and 4, respectively. When $n = 300$ or $n = 500$, the results are reported in Appendix E. Here, we denote copula families Gumbel by G, Clayton by C, Frank by F, Normal by N and t copula by t. For AIC, we look at the rates where AIC is minimum, whereas for multiplier method we look at the empirical levels. Evidence that the AIC method performs well are a high rate of least AIC for the true copula family and a small rate for the other families. Conversely, the multiple GOF test performs well when the empirical level is small for the true family and large for the others. For the multiplier method, the empirical level agrees with 5% nominal level well. However, the empirical levels decrease as we increase n or d (Tables 4.2 and E.1). For example, for the Clayton copula with $\tau = 0.2$ the empirical levels are 8.2%, 6.7% and 3.5% for $d = 2$, 3 and 4, respectively. Both the multiplier method and AIC work well when they are used to choose the correct copula. Nevertheless, when the dependence measure τ for the true copula is small, the true copula is almost same as the independence copula. All the

TABLE 4.1: The empirical levels and the rates of the least AIC with sample size $n = 100$ and sample dimensions $d = 2$

True	Kendall's	the rate of least AIC					the rejection proportions of H_0				
Copula	τ	G	C	F	N	t	G	C	F	N	t
G	0.2	60.1	3.3	9.3	11.5	15.8	3.0	66.5	17.7	9.5	14.2
	0.4	73.6	0.3	4.1	10.4	11.6	2.9	97.5	44.0	22.5	27.0
	0.6	79	0	2.5	8.4	10.1	2.6	100	61.8	29.7	31.7
	0.8	78.3	0	1.7	6.8	13.2	3.1	100	77.4	38.5	38.4
C	0.2	1.5	75.3	7.3	7.5	8.4	49.3	8.2	19.4	12.6	16.5
	0.4	0	93.9	1.5	2.5	2.1	91.5	8.6	58.0	41.8	47.3
	0.6	0	97.4	0.9	0.8	0.9	98.9	9.5	77.6	69.2	65.4
	0.8	0	98.5	0.9	0.1	0.5	99.8	10.2	57.4	87.0	67.6
F	0.2	14.5	13.6	44.9	19.0	8.0	18.6	52.8	5.2	9.4	12.9
	0.4	10.4	3.5	59.8	19.0	7.3	41.2	93.7	6.6	20.3	37.3
	0.6	4.8	0.9	77.9	11.3	5.1	62.6	99.5	5.8	40.9	64.1
	0.8	1.5	0	91.4	4.3	2.8	81.4	99.9	5.9	76.2	82.8
N	0.2	21.3	16.5	22.2	30.7	9.3	12.7	45.2	7.6	4.2	10.4
	0.4	16.0	7.2	11.7	52.2	12.9	20.9	83.3	13.1	3.8	12.4
	0.6	11.3	2.5	7.2	64.3	14.7	18.6	97.3	27.9	4.6	11.9
	0.8	10.1	0.3	3.9	64.3	21.4	9.9	99.7	39.4	4.0	12.1
t	0.2	14.8	11.1	3.5	5.0	65.6	8.9	33.1	9.4	8.9	3.8
	0.4	16.2	4.6	4.4	10.0	64.8	12.1	79.2	19.3	6.0	3.9
	0.6	14.7	3.2	2.5	13.0	66.6	14.0	96.6	34.3	4.8	3.2
	0.8	11.7	0.9	2.3	14.8	70.3	10.4	99.3	47.1	8.2	4.7

TABLE 4.2: The empirical levels and the rates of the least AIC with sample size $n = 100$ and sample dimensions $d = 3$

True	Kendall's	the rate of least AIC					the rejection proportions ofH_0				
Copula	τ	G	C	F	N	t	G	C	F	N	t
G	0.2	81.1	0.8	0.7	6.9	4.2	4.9	71.0	21.5	9.1	7.0
	0.4	88.4	0	4.5	2.8	4.3	2.0	98.3	49.1	22.2	17.8
	0.6	91.0	0	2.4	3.1	1.7	1.7	100	63.0	21.7	23.6
	0.8	89.0	0	1.3	2.4	0	0	100	49.3	9.9	10.2
C	0.2	4.0	87.5	4.5	4.2	3.4	59.9	6.7	34.1	16.9	30.9
	0.4	0	97.8	1.4	0.4	0.4	97.7	6.5	74.3	54.1	62.6
	0.6	0	99.1	0.5	0.2	0.2	99.4	6.0	82.0	72.4	64.5
	0.8	0	99.3	0.6	0	0.1	99.6	2.0	21.9	64.1	14.3
F	0.2	12.4	7.7	59.4	16.0	4.5	19.6	52.9	7.1	2.7	10.5
	0.4	6.1	1.8	81.0	8.5	2.6	41.9	95.6	4.9	5.1	30.1
	0.6	2.4	0	94.0	3.5	0.1	53.9	100	4.0	23.0	53.0
	0.8	0.2	0	98.3	1.0	0.5	56.4	100	0.6	48.3	47.2
N	0.2	11.2	10.9	11.2	60.0	6.7	37.6	62.3	27.0	2.7	20.1
	0.4	6.5	3.1	3.2	78.7	8.5	57.9	97.0	39.3	2.2	22.5
	0.6	3.6	0.5	1.8	82.4	11.7	44.3	99.5	54.4	2.0	11.6
	0.8	3.5	0.1	1.0	75.3	20.1	6.0	99.8	34.5	0.5	2.0
t	0.2	3.8	4.1	1.0	2.1	89.0	30.3	49.0	28.1	7.5	2.5
	0.4	3.5	1.5	0.5	3.7	90.8	43.3	93.0	43.7	4.6	2.9
	0.6	1.9	0.3	0.3	4.7	92.8	42.0	99.6	57.1	2.2	1.0
	0.8	3.3	0.9	0.5	5.8	90.4	11.3	100	38.6	0.9	0

TABLE 4.3: The empirical levels and the rates of the least AIC with sample size $n = 100$ and sample dimensions $d = 4$

True	Kendall's	the rate of least AIC					the rejection proportions of H_0				
Copula	τ	G	C	F	N	t	G	C	F	N	t
G	0.2	88.0	1.0	6.2	3.0	1.8	2.8	70.0	20.7	8.8	1.8
	0.4	94.1	0	4.0	1.4	0.5	2.6	99.1	52.9	24.4	9.3
	0.6	96.0	0	1.2	1.1	1.7	1.2	100	68.9	17.6	15.8
	0.8	93.2	0	1.1	1.1	4.6	0	100	25.8	1.4	1.9
C	0.2	2.0	95.0	1.6	1.8	1.4	62.8	3.5	36.4	17.6	35.7
	0.4	0	99.2	0.6	0.2	0	97.1	4.6	76.2	56.6	56.4
	0.6	0	99.6	0.3	0.1	0	99.5	3.7	79.1	67.0	48.9
	0.8	0	98.8	1.2	0	0	98.5	0.7	4.2	27.5	0.5
F	0.2	9.7	5.1	74.7	9.1	14.0	12.7	53.5	6.8	2.7	10.5
	0.4	4.2	0.6	91.5	3.2	0.5	31.9	96.7	5.2	5.1	30.1
	0.6	1.7	0	97.0	1.2	0.1	42.1	99.9	1.8	23.0	53.0
	0.8	0.5	0	98.8	0.4	0.3	0.3	100	0.4	48.3	47.2
N	0.2	5.9	7.4	6.0	77.8	2.9	59.0	73.8	52.0	1.1	28.0
	0.4	1.4	0.9	1.2	91.4	5.1	80.2	99.3	70.5	1.3	28.7
	0.6	1.8	0.1	0.4	89.5	8.2	61.9	100	71.1	1.2	11.3
	0.8	1.4	0	0.2	78.0	20.4	4.3	100	25.8	0	0.2
t	0.2	0.5	1.1	0.5	1.1	96.8	45.9	54.0	41.8	5.4	2.1
	0.4	0.8	0.7	0.1	1.3	97.1	68.9	96.7	64.9	3.8	2.2
	0.6	0.8	0.1	0.1	1.5	97.5	63.8	100	73.8	0.6	0.5
	0.8	0.7	0	0	2.1	97.2	9.6	100	28.9	0.3	0

copulas will fit the true copula well, then the correct rate for least AIC will be small.

When the sample size n is 100, we get a smaller correct rate for AIC and a larger value of the empirical levels for multiplier method than when sample size is larger. For example, for Frank copula with $\tau = 0.2$ and $d = 3$ the correct rate are 59.4%, 92.1% and 96.9% and the empirical levels are 7.1%, 5.3% and 4.5% for $n = 100$, 300 and 500 respectively. The reason this happens can be explained as there is not enough information from the smaller sample size.

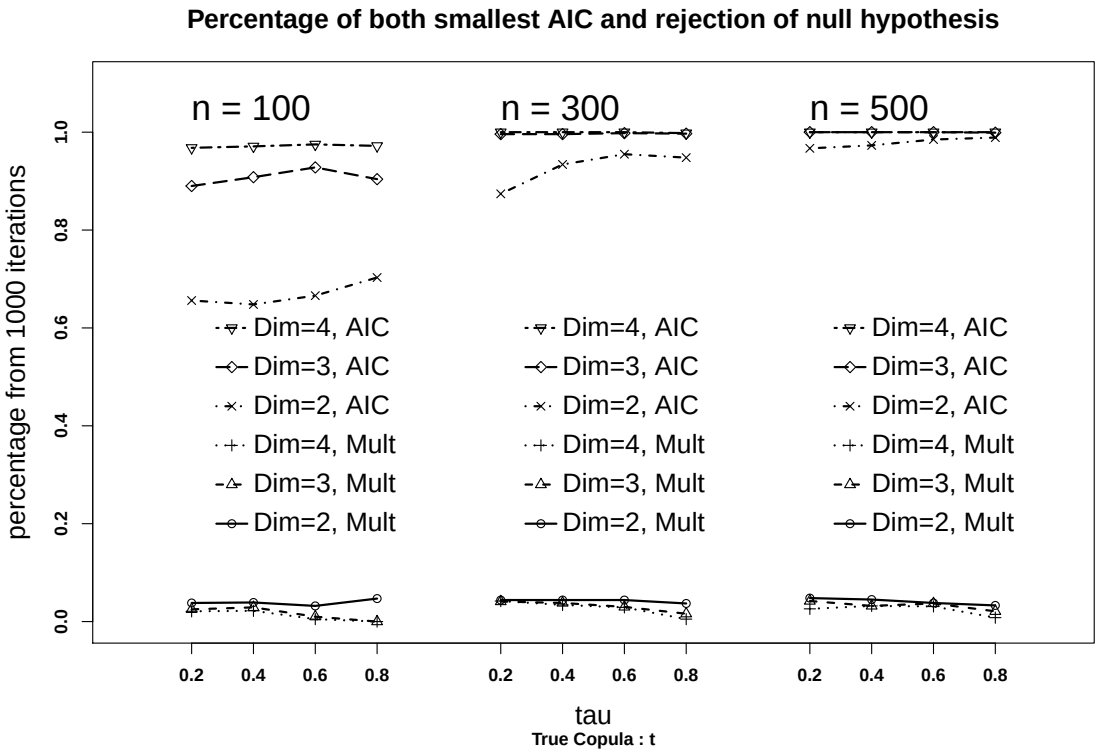


FIGURE 4.1: Plot of the correct rate of AIC and the empirical level of t copula with t copula as the generating family.

Figure 4.1 and Figures E.1, E.2, E.3, and E.4 in Appendix E summarize the simulation results for t, Clayton, Frank, Gumbel and Normal copulas, respectively. All of these figures provide the plots of the correct rates for the least AIC values in 1000 times from the copula under null hypothesis as well as the empirical levels from multiplier method.

Figure 4.1 summarizes simulation results for t copula family. The three columns from left

to right give the combination of multivariate dimension and τ for n=100, 300 and 500 respectively. In each column, the four dots on each line correspond to the four values for association parameter τ (from left to right, $\tau = 0.2, 0.4, 0.6, 0.8$). The three bottom lines give the empirical levels for GOF test using multiplier method, while the three upper lines give the AIC correct rates of the real copula family.

As the dimension increases, the correct rate for AIC increases, whereas the empirical level decreases. For example, consider the last column of Figure 4.1. As the dimension increases, the correct rate for AIC increases and approaches 1 while the empirical level decreases and approaches 0. However, for each combination of sample size and sample dimension, the correct rate does not strictly increase as the association parameter increases. For example, Figure E.4 illustrates results for the Normal copula when $n = 100$ and dimension is either $d = 3$ or $d = 4$, the correct rate of AIC increases as association parameter τ increases from 0.2 to 0.6 but it dramatically drops after $\tau = 0.6$. However, in Table 4.2, when $n = 100$ and $d = 3$, the rate of least AIC from t copula increases with τ when the true copula is the Normal copula. In reality, the larger the association parameter is, the closer Normal copula is to t copula. This explains the notable decrease of AIC correct rate from the Normal copula for $\tau > 0.6$. Even so, Normal copula still gives the highest correct rate of AIC compared to the other copula families. The empirical level of the multiplier method behaves similarly. In Figure 4.1, where $n = 100, d = 2$ and the true copula is t copula, the empirical levels are 3.8%, 3.9%, 3.2% and 4.7% for $\tau = 0.2, 0.4, 0.6, 0.8$, respectively. The empirical level is dramatically increased after $\tau = 0.6$. However, the relationship between the correct rate for AIC and the association parameter τ is unclear. For example, for t copula with $n = 100$ and $d = 2$ in Figure 4.1, the correct rate decreases slightly first, and suddenly increases after $\tau = 0.4$, especially for τ from 0.6 to 0.8. However, this plot illustrates that both AIC and multiplier method provides the same pattern of simulation results.

If the association parameter is low, then the copula is close to the independence copula and any copula will fit well, provided it can achieve the independence copula. If we use a GOF test in this case, we will get multiple copula families which will not be rejected, so we can not say which copula fits best. The same is true with the data with a very high association measure. In a GOF test, if there exist several copulas we fail to reject, then we can use AIC to choose the best. The copula with the least AIC will be one of the copulas which is not rejected by the GOF test.

4.6 Discussion

We can conclude the true copula produces the larger rate of least AIC as compared to any other copula in a series of candidates. In general, the larger the association parameter is, the higher the correct rate is. As the sample size or dimension increases, the correct rate also increases. The plots from Clayton, Frank, Gumbel, Normal and t all show the similar results. Our study supports the use of AIC to choose the best copula, provided the true copula is among those considered. However, ACI does not perform a formal GOF hypothesis test. Compared to the existing multiplier method for a GOF test, our AIC approach provides more computational efficiency. For example, for Clayton copula based on 1000 parametric bootstrap iteration with $n = 500$, $d = 2$ and $\tau = 0.8$, the time for AIC was one hour on a quad core 3.4 GHz desktop computer, while the time for GOF test is over seven hours under the same situations.

5 MODIFIED GAUSSIAN COPULA : CHARACTERISTICS AND APPLICATION TO INSURANCE AND FINANCE

Yan Fang* and Lisa Madsen**

Department of Statistics, Oregon State University, Corvallis, Oregon 97331, U.S.A.

*E-mail: fangya@science.oregonstate.edu

**http://www.stat.oregonstate.edu/people/lmadsen

5.1 Abstract

The idea of a pseudo-copula, introduced by Fermanian and Wegkamp (2004), extended the definition of copula to cover a much larger scope of situations. The main properties of copulas and Sklar's Theorem also apply for pseudo-copulas. Copula modeling, since it was first proposed for use in finance by Li (2000), has been a very important technique for modeling association in finance and insurance risk problems. However, no existing copula is flexible enough to capture tailed (upper-tailed or lower-tailed) dependence as well as elliptical dependence. Elliptical copulas are copulas of elliptically contoured distributions and fit elliptical distributions well. Real data may be better described by a "fat-tailed" or "tailed" copula than by an elliptical copula. As compared to symmetric elliptical copulas, asymmetric Archimedean copulas can capture both the upper and lower tails; howerever, in the multivariate setting, pairwise association is constrained to be the same for all two-dimensional marginals. Archimedean families depend on one or two parameters, no matter how many one-dimensional margins it might include. We propose a new approach based on the Gaussian copula to measure dependence in multivariate data. Our modified Gaussian pseudo-copula overcomes all of the drawbacks aforementioned. It captures the properties of both ellipitcal and Archimedean copulas. We will give the modified Gaussian pseudo-copula's characteristics in the bivariate case, and note that it can be extended to multivariate cases easily. The proposed pseudo-copula is assessed by estimating the measure of association from an insurance data set and conducting a simulation study comparing goodness-of-fit test statistics.

5.2 Introduction

Modeling of dependencies between different outcomes is an important and challenging aspect in statistical science. Copula functions, initially introduced by Sklar (1959), have become popular, flexible models in this field. The use of copula functions enables the specification of the marginal distributions to be separated from the dependence structure, which in turn helps with the task of modeling financial risks under more realistic, non-normal marginals. For a comprehensive review of copulas, see Nelsen (2006). Recently, copula modeling has become popular in the financial and econometric literature (Frees and Valdez 1998; Embrechts et al. 2003; Cherubini et al. 2004; McNeil et al. 2005). Although there exist a large variety of copula functions (Joe 1997; Nelsen 2006), only a few are practically manageable and often the choice in dependence modeling falls on elliptical copulas.

Nonetheless, research on relevant specifications for copulas is still in its infancy. Schmidt (2006) warns that the traditional concept of copulas refers to a static concept of dependence whereas many applications in finance are related to time series events in which a dynamic concept of dependence is needed. Patton (2002) introduces the conditional copula by adding a conditioning variable to the distributions studied and Patton (2006) points out that the conditional variable must be the same for both marginal distributions and the copula. Fermanian and Wegkamp (2004) introduces the concept of pseudo-copula and Klein (2005) mentions that it appears preferable to define a notion of conditional pseudo-copula. The definition of pseudo-copula (Fermanian and Wegkamp, 2004) satisfies all the properties of a copula except that copula $C(u_1, \ldots, u_p)$ is not necessarily u_k when all coordinates of u except u_k are equal to 1 (see Cherubini et al., 2011 for more detail). Even so, a pseudo-copula satisfies a similar version of Sklar's theorem (Sklar, 1959). Hence, all copulas are pseudo-copulas. From now on, we will call all existing copulas pseudo-copulas in the rest of the paper.

Elliptical distribution families are widely applied in statistics and econometrics, especially in finance. Important special cases of elliptical distributions include the Gaussian pseudo-copula (Li, 2000), which we will hereafter refer to as normal pseudo-copula and its convenient Student t extension (Embrechts et al. 2001; Fang and Fang 2002; Demarta and McMeil 2005). While convenient and intuitive, elliptical copulas have a number of obvious shortcomings as a model for

real data. Due to the radial symmetry, the upper-tail and the lower-tail dependence are identical, which might be undesirable in a particular application. In addition, a normal pseudo-copula has no tail dependence at all (see Bradley and Taqqu, 2003). To overcome these limitations, a special class of pseudo-copula called Archimedean (for example Clayton, Frank, or Gumbel) are introduced (see Joe 1997; Nelsen 2006 for a review). While Archimedean pseudo-copulas are calculated over a closed-form and play an important role in extreme value theory, unfortunately, they are difficult to extend to multivariate applications beyond two dimensions (Rachev et al., 2009). Furthermore, Archimedean pseudo-copulas depend on only one or two parameters (Hu and Kercheval, 2007), their usefulness in real data is limited because their properties are the same for all pairs (groups). Consequently, it is hard to describe the exact pairwise dependence.

As we know, one objective in the theory of dependence modeling and multivariate pseudo-copulas is to develop parametric families that are both flexible and appropriate as models for data with different dependence structures, including features such as fat-tail dependence and asymmetric properties. In this paper, we will focus on pseudo-copula models which capture the flexible-tail dependence as well as elliptical dependence. Since the normal pseudo-copula has a closed form probability density function (p.d.f), it is straightforward to estimate the unknown parameters as well as estimate the pairwise dependence coefficient. To overcome the aforementioned lack of flexibility of the normal pseudo-copula, we propose a modified Gaussian (MG) pseudo-copula, which shares the same form as the normal pseudo-copula except for the definition for the dependence coefficients. The pairwise dependence coefficient in the MG pseudo-copula will depend on one or two unknown parameters as well as the variables themselves, which is different from the traditional dependence coefficient ρ used in the normal pseudo-copula. In the MG pseudo-copula, ρ is a function instead of a fixed unknown number. Details will be presented in the following section.

This paper will consider the bivariate case only. It will be shown that the MG pseudo-copula is more flexible than both normal and Archimedean pseudo-copulas. Also it will be demonstrated that the fit of the MG pseudo-copula to an insurance data set is superior than the normal pseudo-copula and the three well-known Archimedean pseudo-copulas, Clayton, Frank and Gumbel pseudo-copulas.

The remainder of this paper is organized as follows. Section 5.3 describes the model and the notation for the MG pseudo-copula. Calibrations and the goodness-of-fit (GOF) tests of the

MG pseudo-copulas are discussed in Section 5.4 and Section 5.5 respectively. Examples of the application to insurance and finance are provided in Section 5.6. Section 5.7 is devoted to a simulation study. A discussion and concluding remarks are given in the last section.

5.3 Model

Based on Sklar's Theorem (Sklar, 1959), any joint function H with continuous marginal cumulative distribution functions (c.d.f) $\mathrm{F}_1, \ldots, \mathrm{F}_p$ for random variable $(X_1, \ldots, X_p)$ has a unique pseudo-copula

$$\mathrm{H}(x_1, \ldots, x_p) = \mathrm{C}\left(\mathrm{F}_1(x_1), \ldots, \mathrm{F}_p(x_p)\right),$$

where $\mathrm{C} : [0,1]^p \to [0,1]$ is a p-variate c.d.f with standard uniform margins. An example of a pseudo-copula function is the normal pseudo-copula described in Mashal and Zeevi (2002), with the following analytic expression:

$$\mathrm{C}_\Sigma(\boldsymbol{u}) = \int_0^{u_1} \cdots \int_0^{u_p} \frac{\phi_\Sigma\left(\Phi^{-1}(x_1), \ldots, \Phi^{-1}(x_p)\right)}{\prod_{i=1}^p \phi\left(\Phi^{-1}(x_i)\right)} dx_1 \ldots dx_p, \tag{5.1}$$

where $\phi_\Sigma(\cdot)$ is the standard multivariate Gaussian p.d.f with correlation matrix Σ, $\Phi^{-1}(\cdot)$ is the inverse of the standard univariate Gaussian marginal c.d.f, $\phi(\cdot)$ is the standard univariate Gaussian p.d.f, and the event relationships matrix, or mathematically correlation matrix Σ is defined as

$$\Sigma = \begin{bmatrix} 1 & \rho_{1,2} & \cdots & \rho_{1,p} \\ \rho_{1,2} & 1 & \cdots & \rho_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1,p} & \rho_{2,p} & \cdots & 1 \end{bmatrix}, \tag{5.2}$$

where $\rho_{i,j}$ is the pairwise association for each pair (i,j) with $i \neq j$ and $i, j \in (1, \ldots, p)$. The density function for this pseudo-copula is

$$\mathrm{c}_\Sigma(u_1, \ldots, u_p) = \frac{\phi_\Sigma\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_p)\right)}{\prod_{i=1}^p \phi\left(\Phi^{-1}(u_i)\right)}. \tag{5.3}$$

In the normal pseudo-copula, Σ is usually defined as an exchangeable or unstructured correlation matrixWe obtain our MG pseudo-copula by extending $\rho_{i,j}$ such that it will depend on

both u_i and u_j as well as unknown parameters a_{ij} and b_{ij}. The goal of the extension is to assure sufficient flexibility of the model. Then the correlation matrix in Equation (5.1) for the MG pseudo-copula is defined as

$$\begin{bmatrix} 1 & \rho(u_1,u_2;a_{12},b_{12}) & \cdots & \rho(u_1,u_p;a_{1p},b_{1p}) \\ \rho(u_1,u_2;a_{12},b_{12}) & 1 & \cdots & \rho(u_2,u_p;a_{2p},b_{2p}) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(u_1,u_p;a_{1p},b_{1p}) & \rho(u_2,u_p;a_{2p},b_{2p}) & \cdots & 1 \end{bmatrix}, \tag{5.4}$$

where unknown parameter $(a_{i,j}, b_{i,j})$ is used in conjunction with pair (u_i, u_j). Note that a pseudo-copula C is a true copula if and only if $\mathrm{C}(1,\ldots,u_k,\ldots,1) = u_k, \text{for every} j = 1,\ldots,p$.

As a result, the correlation matrix defined in Equation (5.2) will not be compound symmetric or auto-regressive. Hereafter, we will use Σ^* to denote the correlation matrix defined in Equation (5.4). Instead, Equation (5.1) and Equation (5.3) with Σ^* replacing Σ will be called as the MG pseudo-copula function and the density function for the MG pseudo-copula respectively. For simplicity of exposition, from now on, we restrict our attention to the bivariate case.

5.3.1 Definitions

Let (X, Y) be a pair of random variables with continuous marginal distribution F and Grespectively. Let $u = \mathrm{F}(X)$ and $v = \mathrm{G}(Y)$, thus $u \in [0,1]$ and $v \in [0,1]$. We are trying to find a function ρ of (u, v) and unknown parameters (a, b), such that these functions can define different kind of tail or non-tail dependence.

We have studied five forms of $\rho(u, v; a, b)$ given below:

Definition I: $\rho(u,v;a,b) = b(1-auv)$, where $a \in [0,1]$ and $b \in [-1,1]$;

Definition II: $\rho(u,v;a,b) = b\cos\left(\frac{\pi}{2}a(1-uv)\right)$, where $a \in [0,1]$ and $b \in [-1,1]$;

Definition III: $\rho(u,v;a,b) = b\sin\left(\frac{\pi}{2}a(1-uv)\right)$, where $a \in [0,1]$ and $b \in [-1,1]$;

Definition IV: $\rho(u,v;a,b) = b\tan\left\{\frac{\pi}{4}a(1-uv)\right\}$, where $a \in [0,1]$ and $b \in [-1,1]$;

Definition V: $\rho(u,v;a,b) = b\exp\left\{-a(1-uv)\right\}$, where $a \in [0,+\infty)$ and $b \in [-1,1]$.

In all definitions, parameter a controls the speed of the convergence of the tail and parameter b controls the shape of the tail.

5.3.2 Special case

For all five definitions, if $b = 0$, then $\rho(u, v; a, b) = 0$ and the MG pseudo-copula is reduced to the independence pseudo-copula $\mathrm{C}(u, v) = uv$. On the other hand, if $a = 0$, **I**, **II** and **V** will degenerate to the normal pseudo-copula, while **III** and **IV** will become the independence pseudo-copula.

5.3.3 Bivariate Characteristics

Define a function $\mathrm{g}(u, v; \boldsymbol{\theta})$ as

$$\mathrm{g}(u, v; \boldsymbol{\theta}) = \frac{1}{\sqrt{1 - \rho^2(u, v; \boldsymbol{\theta})}} \times \exp\left\{\frac{[\Phi^{-1}(u)]^2 + [\Phi^{-1}(v)]^2}{2}\right\} \times \exp\left\{-\frac{[\Phi^{-1}(u)]^2 - 2\rho(u, v; \boldsymbol{\theta})\Phi^{-1}(u)\Phi^{-1}(v) + [\Phi^{-1}(v)]^2}{2\left(1 - \rho^2(u, v; \boldsymbol{\theta})\right)}\right\},$$

where $(u, v) \in [0, 1]^2$, $\rho(\cdot)$ is a correlation function defined in Section 5.3.1, and $\boldsymbol{\theta} = (a, b)$ represents the unknown parameter vector. Observe that $\mathrm{g}(u, v; \boldsymbol{\theta}) = 0$ for $(u, v) \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$. From the definition of $\mathrm{g}(u, v; \boldsymbol{\theta})$, we deduce two properties of function $\mathrm{g}(u, v; \boldsymbol{\theta})$:

1. $\mathrm{g}(u, v; \boldsymbol{\theta}) \geq 0, \forall (u, v) \in [0, 1]^2$;
2. $\mathrm{g}(u, v; \boldsymbol{\theta})$ goes to zero at the far left and the far right, namely,

$$\lim_{\substack{u \to 0 \\ v \to 0}} \mathrm{g}(u, v; \boldsymbol{\theta}) = \lim_{\substack{u \to 1 \\ v \to 0}} \mathrm{g}(u, v; \boldsymbol{\theta}) = \lim_{\substack{u \to 0 \\ v \to 1}} \mathrm{g}(u, v; \boldsymbol{\theta}) = \lim_{\substack{u \to 1 \\ v \to 1}} \mathrm{g}(u, v; \boldsymbol{\theta}) = 0.$$

Then we can normalize $\mathrm{g}(u, v; \boldsymbol{\theta})$ to obtain a p.d.f, namely,

$$\mathrm{c}(u, v; \boldsymbol{\theta}) = \frac{\mathrm{g}(u, v; \boldsymbol{\theta})}{\boldsymbol{K}},$$

where $\boldsymbol{K} = \int_0^1 \int_0^1 \mathrm{g}(u, v; \boldsymbol{\theta}) du dv$.

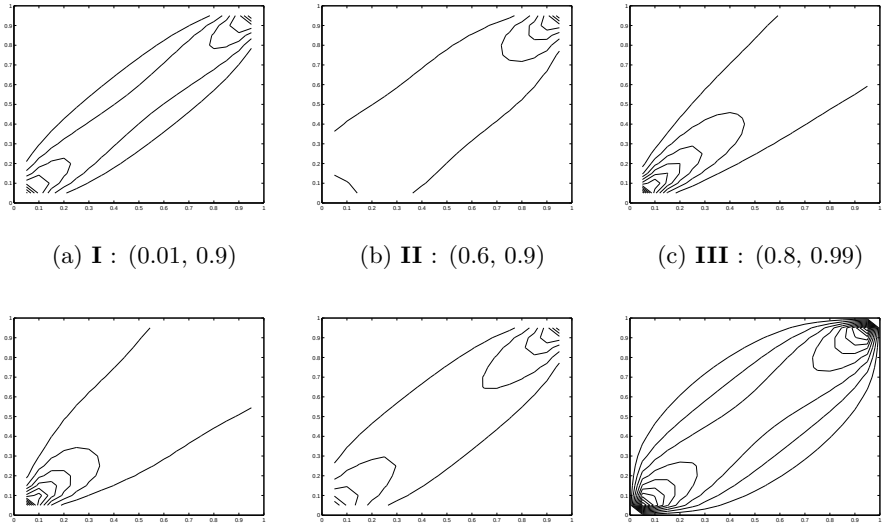
The p.d.f of the bivariate MG pseudo-copula with two parameters is given by $\mathrm{c}(u, v; \boldsymbol{\theta})$, viz,

$$\mathrm{c}(u, v; \boldsymbol{\theta}) = \frac{1}{\boldsymbol{K}} \frac{1}{\sqrt{1 - \rho^2(u, v; \boldsymbol{\theta})}} \times \exp\left\{\frac{[\Phi^{-1}(u)]^2 + [\Phi^{-1}(v)]^2}{2}\right\} \times \exp\left\{-\frac{[\Phi^{-1}(u)]^2 - 2\rho(u, v; \boldsymbol{\theta})\Phi^{-1}(u)\Phi^{-1}(v) + [\Phi^{-1}(v)]^2}{2\left(1 - \rho^2(u, v; \boldsymbol{\theta})\right)}\right\}. \tag{5.5}$$

And the bivariate MG pseudo-copula with two parameters is given by

$$\mathrm{C}(u,v;\boldsymbol{\theta}) = \frac{1}{\boldsymbol{K}} \int_0^u \int_0^v \frac{1}{\sqrt{1-\rho^2(x,y;\boldsymbol{\theta})}} \times \exp\left\{\frac{[\Phi^{-1}(x)]^2 + [\Phi^{-1}(y)]^2}{2}\right\} \times \exp\left\{-\frac{[\Phi^{-1}(x)]^2 - 2\rho(x,y;\boldsymbol{\theta})\Phi^{-1}(x)\Phi^{-1}(v) + [\Phi^{-1}(y)]^2}{2\left(1-\rho^2(x,y;\boldsymbol{\theta})\right)}\right\} dxdy. \tag{5.6}$$

The proof that $\mathrm{C}(u,v;\boldsymbol{\theta})$ is a pseudo-copula is given in Appendix F.



(a) **I** : (0.01, 0.9) (b) **II** : (0.6, 0.9) (c) **III** : (0.8, 0.99)

(d) **IV** : (0.97, 0.99) (e) **V** : (0.1, 0.9) (f) normal : 0.8

FIGURE 5.1: Contour plots for five definitions of the MG pseudo-copula and the normal pseudo-copula

In Figure 5.1, we plot the contours for the MG pseudo-copula with ρ definitions **I** through **V** and the normal pseudo-copula. It can be seen that the MG pseudo-copula with different definitions exhibits varied density shapes, as well as varied direction of tail dependence. This is in contrast to the rigid normal pseudo-copula shape shown in Figure 5.1 (f). Our experimentation shows that the density shape of the tailed dependence of different definitions can be controlled by the choice of the parameters, demonstrating the flexibility of our model. In summary, the MG pseudo-copula has the following features:

1. Distribution flexibility, including fat-tail distribution, upper-/lower- tail and elliptical distributions;
2. Manageable range of shape, which can be symmetric or asymmetric;
3. Controllable range of dependence; allowing both positive and negative dependence
4. A closed form pseudo-copula density function;

Being an extension of the normal pseudo-copula, the MG pseudo-copula can be applied to various fields by a careful choice of ρ definition and parameter tuning. In fact, many existing classes of multivariate distributions can be constructed by using this new pseudo-copula. For a given set of marginal distributions, we can construct various MG pseudo-copulas by adjusting both the parameters and the correlation function between pairs of variables. Certainly the ρ definitions need not to be limited to these five. Nevertheless, in our experiments, we found the definitions have given us enough flexibility by adjusting the parameters.

5.4 Calibration

5.4.1 Maximum pseudo-likelihood estimation

Let $((X_1, Y_1), \ldots, (X_n, Y_n))$ be a random sample from MG pseudo-copula $\mathrm{C}\left(\mathrm{F}(x), \mathrm{G}(y); a, b\right)$ and assume $\mathbf{X} = (X_1, \ldots, X_n)$ and $\mathbf{Y} = (Y_1, \ldots, Y_n)$ are random samples and F and G are continuous c.d.f.s. Generally, pseudo-copula models are used in situations where studying the association between the variables is important. In practice, it is common that the marginal c.d.f.s are unknown. Many authors adopt a parametric form for dependence while keeping the marginals unspecified. In other words, the models are based on non-parametric margins and parametric pseudo-copulas, and are referred to as semi-parametric pseudo-copulas. To estimate a semi-parametric pseudo-copula using observed data (x_i, y_i) with $i = 1, \ldots, n$, one can obtain the maximum likelihood estimator (MLE) based on pseudo observations, viz, the maximum pseudo-likelihood estimator (MPLE) (see Genest et al., 1995; Shih and Louis, 1995). The idea behind maximum likelihood parameter estimation is to determine the parameters that maximize the likelihood of the sample data. From a statistical point of view, the method of maximum likelihood yields estimators with good statistical properties. MLE methods are versatile and apply to many models and to different types of

data. In addition, they provide efficient methods for quantifying uncertainty through confidence bounds. Consequently, we can use the MPLE, which is found by maximizing (with respect to $\boldsymbol{\theta}$) the log pseudo-likelihood, to estimate the parameters in pseudo-copula. The log pseudo-likelihood is defined as

$$\log \mathrm{L}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log \mathrm{c}\{\hat{U}_{i,n}, \hat{V}_{i,n}; \boldsymbol{\theta}\}, \tag{5.7}$$

where $(\hat{U}_{i,n}, \hat{V}_{i,n})$ are pseudo-observations computed from $(\boldsymbol{X}, \boldsymbol{Y})$ by $\hat{\mathrm{U}}_{i,n} = \frac{\mathrm{R}(x_i)}{n+1}$ and $\hat{\mathrm{V}}_{i,n} = \frac{\mathrm{R}(y_i)}{n+1}$ with $\mathrm{R}(x_i)$ being the rank of x_i among $x_1, \ldots, x_n$ and $\mathrm{R}(y_i)$ being the rank of y_i among $y_1, \ldots, y_n$. This method can also be seen as a version of the inference function for the method in which the marginal c.d.f.s are estimated non-parametrically.

Let $\hat{\boldsymbol{\theta}}$ be the MPLE computed from the pseudo-observation $(\hat{U}_{i,n}, \hat{V}_{i,n})$ by maximizing Equation (5.7), namely,

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \left[\sum_{i=1}^{n} \log \mathrm{c}\{\hat{U}_{i,n}, \hat{V}_{i,n}; \boldsymbol{\theta}\} \right]. \tag{5.8}$$

To implement the MPLE for a given pseudo-copula family $\mathrm{C}(u, v; \boldsymbol{\theta})$, one needs to be able to estimate the vector of dependence parameters $\boldsymbol{\theta}$ from the available data. The corresponding MPLE can be undertaken by using the well designed R optim() function (R Development Core Team 2009) referring to Kojadinovic (2011). Conditions for the asymptotic semi-parametric efficiency of the MPLE have been investigated in Genest (2002).

If the MPLE for parameter b is 1, then we will redefined ρ as the function of parameter a, and variables u and v. That is, $\rho = \rho(u, v; a)$ and $a \in (0, 1)$. We retain 5 forms for the definition of ρ and fix $b = 1$. We then refit the pseudo-copula model and use Equation (5.8) to get MPLE for θ. Here θ is univariate, i.e., $\theta = a$.

On the other hand, if the MPLE for parameter a is 1, then we will redefined ρ as the function of parameter b, and variables u and v, except for $\mathbf{V}$. That is, $\rho = \rho(u, v; b)$ and $b \in (-1, 1)$. We retain the form for the definition of ρ and fix $b = 1$. We then refit the pseudo-copula model and use Equation (5.8) to get MPLE for θ. Here θ is univariate, i.e., $\theta = b$. For $\mathbf{V}$, the range for parameter a is $(0, +\infty)$ after removing the special case normal pseudo-copula when $a = 0$.

We wish to have the ranges of the parameters be open subsets, which will assure us that we can use the test method described in Section 5.5 to do a GOF test for the pseudo-copula.

5.4.2 Kendall's τ approximation

One popular measure of association is the population version of Kendall's τ.The sample version of Kendall's τ is given by

$$\tau = \frac{4}{n(n-1)} \sum_{i=1}^{n} \left\{ \sum_{j=1}^{n} \boldsymbol{I}(X_j \leq X_i, Y_j \leq Y_i) \right\} - 1. \tag{5.9}$$

Unlike the normal pseudo-copula, then the MG pseudo-copula does not have a closed form to calculate τ. However, we can use the mean of a series of sample τ's to estimate τ. When given a specific data set, we can adopt the following procedure :

1. Translate the original data $(\hat{\mathbf{X}}, \hat{\mathbf{Y}})$ to the pseudo-observation $(\hat{\mathbf{U}}, \hat{\mathbf{V}})$;
2. Fit the MG pseudo-copula to the pseudo observation and find the MPLE $\hat{\boldsymbol{\theta}}$ for the unknown parameters;
3. Generate N random samples from the MG pseudo-copula by using $\hat{\boldsymbol{\theta}}$ as the parameter;
4. From each random sample, obtain a sample version of Kendall's τ by using Equation (5.9), i.e., τ_i, where $i = 1, \ldots, N$;
5. An approximate Kendall's τ value is

$$\bar{\tau} = \frac{1}{N} \sum_{i=1}^{N} \tau_i$$

with standard error (se)

$$\text{se} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (\tau_i - \bar{\tau})^2}\,.$$

Please note that the above procedure also applies to other association measures, such as Spearman's ρ and Pearson's ρ.

5.5 Goodness-of-fit test

With several pseudo-copula models available, we seek a way to choose the best-fitting model. Information criteria, such as Akaike's Information Criterion (AIC), do not provide us with any understanding of the power or size of the employed decision rule. AIC method selects the model

with the minimum AIC value. GOF approaches are more powerful in deciding whether to reject or accept parametric pseudo-copulas, making them the preferred choice in empirical applications. Recently many authors have proposed GOF tests to pseudo-copula models. For example, Genest and Rivest (1993) developed an empirical method to identify the best Archimedean pseudo-copula; Fermanian (2005) approximated the underlying p.d.f by kernel smoothing of the empirical density; Berg (2005) used a GOF test which is based on the probability integral transform; Berg (2009) and Genest et al. (2009) summarized the literature on GOF tests.

The central goal of a GOF test is to investigate whether the true unknown pseudo-copula C actually belongs to some known parametric pseudo-copula family $\mathtt{C}_0 = \{\mathrm{C}(u, v; \boldsymbol{\theta}) : \boldsymbol{\theta} \in \mathcal{O}\}$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)$ and $\mathcal{O}$ is an open subset of $\mathbb{R}^q$. More formally, one wants to test

$$\mathrm{H}_0 : \mathrm{C} \in \mathtt{C}_0 \quad \text{against} \quad \mathrm{H}_1 : \mathrm{C} \notin \mathtt{C}_0.$$

According to the definition for ρ, parameter $\boldsymbol{\theta}$ is not an open set. For example, parameter b can be the boundary value, such as 1 or -1. Nevertheless, in sub-section 5.4.1, we state that if any one of the parameters is 1, we set it equal to one. Then we reduce ρ function to only one parameter function corresponding to the other parameter and refit the MG pseudo-copula to get the MPLE for one parameter pseudo-copula. Thus, we avoid getting the boundary value for parameter $\boldsymbol{\theta}$. Therefore, $\boldsymbol{\theta} \in \mathcal{O}$ in the MG pseudo-copula. As in Section 5.4, we do this in order to have a valid GOF test for our MG pseudo-copula.

There are a relatively large number of proposed testing procedures based on the empirical pseudo-copula (Deheuvels, 1981) of the data. The empirical pseudo-copula is a the consistent estimator of the unknown pseudo-copula $\mathtt{C}_0$ defined by the pseudo-observations $(\hat{U}_{i,n}, \hat{V}_{i,n})$ (see Genest et al. 2009). Traditionally, the empirical pseudo-copula $\mathrm{C}_n(u, v)$ is defined as the empirical c.d.f computed by using the pseudo-observations, i.e.,

$$\mathrm{C}_n(u, v) = \frac{1}{n+1} \sum_{j=1}^{n} \mathrm{I}\left(\hat{U}_{i,n} \leq u, \hat{V}_{i,n} \leq v\right), \qquad u, v \in [0, 1], \tag{5.10}$$

where $\mathrm{I}(\hat{U}_{i,n} \leq u, \hat{V}_{i,n} \leq v)$ is equal to 1, if $\hat{U}_{i,n} \leq u$ and $\hat{V}_{i,n} \leq v$; otherwise, it is equal to 0.

The GOF test implemented in this chapter is based on the empirical process

$$\mathbb{C}_n(u, v) = \sqrt{n} \left\{\mathrm{C}_n(u, v) - \mathrm{C}(u, v; \boldsymbol{\theta}_n)\right\}, \qquad u, v \in [0, 1], \tag{5.11}$$

where $\mathrm{C}(u, v; \boldsymbol{\theta}_n)$ is an estimator of C under the hypothesis that $\mathrm{H}_0 : \mathrm{C} \in \mathtt{C}_0$ holds. The estimator $\boldsymbol{\theta}_n$ of $\boldsymbol{\theta}$ is the same as $\hat{\theta}$ defined in Equation (5.8). We use the Cramér-Von Mises statistics (see Genest et al. 2009; Berg 2009) as the test statistic, which is defined as

$$\mathrm{S}_n = \int_{[0,1]^2} \mathbb{C}_n^2(u, v) d\mathrm{C}_n(u, v) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \mathrm{C}_n(\hat{U}_{i,n}, \hat{V}_{i,n}) - \mathrm{C}(\hat{U}_{i,n}, \hat{V}_{i,n}; \boldsymbol{\theta}_n) \right\}^2. \tag{5.12}$$

Large values of this statistic lead to the rejection of H_0. An approximate p-value can be obtained by means of a multiplier GOF test whose asymptotic validity was shown by Kojadinovic et al. (2011). Under suitable regularity conditions (see Kojadinovic et al., 2011), the multiplier method can be used do GOF test for any copula. In order to show the efficiency of the multiplier approach for our MG pseudo-copula, we give the demonstration that the MG pseudo-copula satisfies all three regularity conditions required for the multiplier GOF test in Appendix G. The multiplier GOF test procedure of multiplier GOF test is outlined in Appendix H. The multiplier test based on the maximization of the pseudo-likelihood appears to efficient when the sample size is at least 300. Since the study in chapter 4 suggests that AIC is a useful tool for copula model-select, we also calculate AIC, given by

$$AIC = -2 \ln (\text{maximized likelihood}) + 2p.$$

5.6 Applications

In this section, we illustrate the application of this new pseudo-copula on the real data sets in finance as well as in insurance. The R code is available upon request from the author.

5.6.1 Losses and ALAEs

In order to illustrate the MG pseudo-copula model, we will use the insurance data from Frees and Valdez (1998) as an example (see, e.g, Genest et al., 2009; Klugman and Parsa, 1999). There are 1500 observed claims. The two variables of interest are the indemnity payment (loss) and the corresponding allocated loss-adjustment expense (alae). For some claims, the policy limit was unknown, and we assumed there was no policy limit. There are 34 out of 1500 policies with claims equal to the policy limit and thus are considered "censored." For simplicity, most researchers ignore the censored claims (Kojadinovic et al., 2011). Since both loss and alae are not unique,

we have to face the tie-breaking problem. When computing the pseudo-observations, we assigned ranks at random in case of ties (Kojadinovic et al., 2011). The empirical Kendall's τ for the pseudo observations is 0.306.

For both the normal pseudo-copula and the MG pseudo-copula, we estimate the association by using the MPLE method. The MPLEs defined in Equation (5.7) are shown in Table 5.1. Please note that the MG pseudo-copula with ρ type **I** is the normal pseudo-copula, since a is 0. For **III**, it reduces to one parameter given $a = 1$, viz $\rho = \rho(u, v; b)$. While for **IV**, it reduces to one parameter given $b = 1$, viz $\rho = \rho(u, v; a)$.

TABLE 5.1: MPLE, Kendall's τ, p-values and AIC values for Loss Data.

Family	MPLE : $\boldsymbol{\theta}$	τ	p-value	AIC
Clayton	0.494	0.198	0	-174.291
Gumbel	1.423	0.587	0.162	-377.634 (*)
Frank	2.982	0.482	0	-317.462
normal	0.458	0.302	0	-338.242
t ($\nu = 4$)	0.433	0.285	0	-320.396
I	(0, 0.458)	0.302	0	-338.242
II	(0.648, 0.688)	0.312	0.074	-370.064 (**)
III	(1, 0.448)	0.157	0.004	-239.284
IV	(0.538, 1)	0.220	0	-211.019
V	(0.656, 0.763)	0.316	0.325	-365.174 (***)

Table 5.1 provides the estimate of parameter $\boldsymbol{\theta}$, estimated measures of association, p-values from a GOF test using multiplier method and AIC values for the traditional pseudo-copulas and the MG pseudo-copula. The AIC with one star is the smallest; AIC with two stars is the second smallest; AIC with three stars is the third smallest. Among the five given families, Clayton, Frank, Gumbel, normal and t(with degree of freedom 4), Gumbel pseudo-copula is the only one that is not rejected at a 5% significance level by using the multiplier method. Comparing AIC values, Gumbel pseudo-copula gives the smallest AIC in this five given families. Hence, we infer the data shows a positive tail.

The fourth column in Table 5.1 shows the p-value from the multiplier method. The p-value is obtained by running 1,000 iterations. From Table 5.1, we reject **I**, **III**, and **IV**. The result is corroborated by Figure 5.1. **I**, **III** and **IV** can capture lower-tail dependence, i.e., exhibiting greater dependence in the negative tail than in the positive tail. We fail to reject **II** and **V** at the 5% significant level. From Figure 5.1, we see both **II** and **V** appear to have positive tails, which match the results from Kojadinovic (2010a). Furthermore, **II** produces the second smallest AIC, and **V** produces the third smallest AIC. Moreover, AIC values from **II** and **V** are smaller than the value from Clayton, Frank, normal and t pseudo-copula. Consequently, both multiplier method and AIC values give the similar results.

5.6.2 U.S. Economic Variables

As a second illustration, we consider the U.S. economic set (Luetkpohl, 1991). The data series has 136 observations and 4 time series variables: the log of the seasonally adjusted real U.S. money (ln M1), the log of the Gross National Product (ln GNP) in 1982 dollars, the discount rate on new issues of 91-day treasury bills (rs) and the yield on long-term (20 years) treasury bonds (rl). Since this data is not independently and identically distributed, the MG model is inappropriate. Inspired by McNeil (2000), we propose a two-step procedure to study the time series data. We first fit the GARCH(1,1) model to each term series and obtain the marginal distribution of residuals. Then, we use MG, normal, t (with degree of freedom 4), Clayton, Gumbel and Frank pseudo-copulas to describe the association of the residuals. As we described in Section (5.4.1), we estimate the parameters of pseudo-copulas using MPLEs.

In this example, we want to study the dependence structure between variables rs and rl. Based on a 5% significant level, the multiplier GOF test concludes the Clayton can not be rejected, but it is very close to 5%. Therefore, this data displays sort of negative tail. While both Gumbel and Frank are rejected. Similarly, the AIC values are large by comparing the ones from normal and t. From the above example, we found that the MG pseudo-copula can capture the positive tail; on the contrary, here in this example we see that the MG pseudo-copula is flexible enough to capture the negative tail. The empirical Kendall's τ for the pseudo observations in this data set is 0.444.

Table 5.2 shows the estimators of parameter $\boldsymbol{\theta}$, measures of association, p-values from GOF

TABLE 5.2: MPLE, Kendall's τ with the standard error, p-values and AIC values: (rs, rl).

Family	MPLE : $\boldsymbol{theta}$	Kendall's τ	p-value	AIC
Clayton	1.274	0.389	0.063	-63.13
Gumbel	1.743	0.635	0.001	-63.95
Frank	4.923	0.325	0.010	-63.46
normal	0.658	0.457	0.163	-69.63 (**)
t($\nu = 4$)	0.641	0.443	0.179	-72.54 (*)
I	(0.069, 0.669)	0.435	0.807	-67.68 (***)
II	(0, 0.658)	0.457	0.163	-69.63 (**)
III	(0.962, 0.699)	0.397	0.007	-58.92
IV	(0.829, 1)	0.377	0.002	-57.39
V	(0, 0.658)	0.457	0.163	-69.63 (**)

test by using the multiplier method and AIC values between the residuals from variables rs and rl for the traditional pseudo-copulas and the MG pseudo-copula. As in table 5.1, the AIC with one star is the smallest; AIC with two stars is the second smallest; AIC with three stars is the third smallest. Please note that the time effect has been removed. For Clayton, Gumbel, Frank, normal and t and the MG pseudo-copulas, we estimate the association by using the MPLE methods and calculate the p-values from the multiplier method and the AIC values from five familiar pseudo-copula as well. All of the GOF tests were done by the methods from Section (5.5) with 1,000 iterations. Please note that the MG pseudo-copula with ρ type **II** and **V** are the normal pseudo-copula, since a is 0. For **IV**, it reduces to one parameter given $b = 1$, i.e., $\rho = \rho(u, v; a)$.

The t ($\nu = 4$) pseudo-copula gives the smallest AIC. Accordingly, the p-value is greater than 5%. The second smallest AIC is from the normal pseudo-copula with the p-value 0.163, which is greater than 5%. The third smallest AIC is from **I** with p-value 0.807. Based on 5% significant level, we fail to reject normal, t($\nu = 4$) and **I**. Again, AIC and multiplier provide similar results.

5.7 Simulation

A large-scale simulation experiment was conducted to assess the MG pseudo-copula performances and flexibility. We use the proposed GOF test to check how well the MG pseudo-copula fits the data generated from various choices of dependence structures, such as Gumbel, t and so on.

To curtail the computational effort, comparisons were restricted to the bivariate case and to five one-parameter pseudo-copula families, i.e, Clayton, Frank, Gumbel-Hougarrd, normal and t with $\nu = 4$.

Since we want to test how well the MG pseudo-copula will fit the given data from different kind of tails and shape, each pseudo-copula family except for the MG pseudo-copula is used as a data-generating family. The MG pseudo-copula will be used as a hypothesized family only. When generating the data, for each pseudo-copula family, four dependence levels are considered corresponding to Kendall's τ of 0.2, 0.4, 0.6, and 0.8. Three sample sizes (n=100, 300, and 500) were used. The number of the multiplier iterations N is fixed to 1,000. The significant level for GOF test is set to 0.05.

In summary, the data generation and the GOF test involved the following factors:

C : Five aforementioned one-parameter pseudo-copula models from which the data was generated;

τ : level of dependence, as measured by Kendall's τ (0.2, 0.4, 0.6 and 0.8);

n : size of each sample drawn from C (100, 300 and 500)

$\mathtt{C}_0$: hypothesized pseudo-copula model under H_0, 5 definitions of MG copulas, plus the five one-parameter copulas;

There are altogether $10 \times 5 \times 4 \times 3 = 600$ scenarios. For each scenario, 1,000 samples (or replications) are generated and are then used to estimate the rejection rates (i.e, the proportion of rejections of the null hypothesis under 1000 replications) of the ten null hypothesis under consideration.

TABLE 5.3: The empirical probabilities of rejecting H_0 based on 1000 replications: $n = 100$

Family	τ	Clayton	Frank	Gumbel	normal	t	**I**	**II**	**III**	**IV**	**V**
Clayton	0.2	0.077	0.187	0.481	0.121	0.155	0.090	0.085	0.090	0.140	0.095
	0.4	0.078	0.596	0.913	0.427	0.456	0.435	0.385	0.575	0.765	0.412
	0.6	0.102	0.767	0.988	0.685	0.643	0.559	0.697	0.845	0.879	0.742
	0.8	0.080	0.591	0.999	0.850	0.693	0.217	0.880	0.564	0.520	0.917
Frank	0.2	0.522	0.048	0.217	0.091	0.118	0.055	0.065	0.005	0.010	0.035
	0.4	0.918	0.063	0.416	0.173	0.326	0.073	0.120	0.020	0.090	0.094
	0.6	0.996	0.056	0.668	0.411	0.631	0.140	0.270	0.005	0.250	0.275
	0.8	0.999	0.060	0.803	0.787	0.859	0.231	0.432	0	0.900	0.596
Gumbel	0.2	0.669	0.202	0.049	0.12	0.164	0.060	0.065	0	0	0.035
	0.4	0.984	0.421	0.036	0.218	0.261	0.208	0.150	0.005	0.045	0.100
	0.6	1.00	0.611	0.025	0.322	0.355	0.281	0.093	0.005	0.225	0.090
	0.8	1.00	0.772	0.028	0.365	0.372	0.2558	0.060	0	0.838	0.245
normal	0.2	0.421	0.085	0.136	0.056	0.108	0.045	0.010	0.015	0.045	0.015
	0.4	0.814	0.137	0.19	0.03	0.136	0.021	0.006	0.060	0.205	0.010
	0.6	0.979	0.255	0.172	0.004	0.110	0.031	0	0.075	0.255	0.007
	0.8	1.00	0.398	0.088	0.037	0.094	0.019	0	0.038	0.548	0.056
t	0.2	0.384	0.125	0.122	0.106	0.054	0.076	0.165	0.025	0.03	0.05
	0.4	0.765	0.20	0.14	0.059	0.051	0.049	0.096	0.075	0.15	0.037
	0.6	0.962	0.329	0.158	0.048	0.036	0.023	0.031	0.135	0.225	0.059
	0.8	0.997	0.458	0.084	0.082	0.054	0.047	0.056	0.101	0.643	0.023

In Tables 5.3–5.5, we report the empirical probabilities of rejecting H_0, i.e, the rejection rate, based on 1,000 replications with sample size 100, 300 and 500, respectively. In all of these tables, the smaller the rejection rate, the closer the pseudo-couple under hypothesis is to the true pseudo-copula. Of course, the rejection rate from the true pseudo-copula is close to 0.

ρ type **III** fits the Frank pseudo-copula well. Except for only one case case where $\tau = 0.6$ and $n = 500$, all the rejection rates from **III** are smaller than the values from Frank pseudo-copula. The reason is that we use two parameters in the MG pseudo-copula, which allows it to fit the Frank

TABLE 5.4: The empirical probabilities of rejecting H_0 based on 1000 replications: $n = 300$

Family	τ	Clayton	Frank	Gumbel	normal	t	**I**	**II**	**III**	**IV**	**V**
Clayton	0.2	0.058	0.592	0.917	0.45	0.588	0.325	0.392	0.46	0.535	0.43
	0.4	0.063	0.995	1.00	0.955	0.975	0.86	0.953	0.825	0.930	0.974
	0.6	0.068	1.00	1.00	0.999	1.00	0.922	1.00	0.977	0.807	1.00
	0.8	0.055	1.00	1.00	1.00	1.00	0	1.00	0.786	0.545	0.988
Frank	0.2	0.873	0.041	0.578	0.195	0.471	0.14	0.111	0.015	0.055	0.085
	0.4	0.999	0.051	0.942	0.565	0.903	0.227	0.225	0.01	0.565	0.304
	0.6	1.00	0.036	0.999	0.946	0.99	0.175	0.413	0.013	0.973	0.744
	0.8	1.00	0.025	1.00	1.00	1.00	0.147	0.5	0	0.915	0.88
Gumbel	0.2	0.964	0.422	0.044	0.247	0.332	0.263	0.07	0.005	0.01	0.025
	0.4	1.00	0.84	0.032	0.515	0.581	0.487	0.157	0.16	0.395	0.061
	0.6	1.00	0.962	0.028	0.7	0.728	0.706	0.188	0	0.962	0.102
	0.8	1.00	0.999	0.021	0.726	0.731	0.755	0.15	0.023	0.835	0.214
normal	0.2	0.704	0.078	0.314	0.028	0.252	0.015	0.005	0.045	0.205	0.005
	0.4	0.997	0.294	0.583	0.041	0.347	0.022	0.007	0.435	0.955	0.021
	0.6	1.00	0.666	0.635	0.043	0.29	0.018	0	0.277	0.992	0.016
	0.8	1.00	0.921	0.503	0.027	0.212	0.184	0	0.133	0.814	0.021
t	0.2	0.66	0.264	0.185	0.114	0.039	0.117	0.127	0.015	0.04	0.035
	0.4	0.988	0.579	0.445	0.075	0.036	0.075	0.050	0.09	0.345	0.056
	0.6	1.00	0.828	0.495	0.087	0.041	0.032	0.030	0.038	0.777	0.075
	0.8	1.00	0.965	0.443	0.078	0.045	0.162	0.044	0.15	0.745	0.161

pseudo-copula better than the Frank pseudo-copula itself! While in case $\tau = 0.6$ and $n = 500$, the rejection rate is 0.06 for **III** and rejection rate is 0.047 for the Frank pseudo-copula. In addition, **III** fits the Gumbel pseudo-copula well except for case $n = 300, \tau = 0.4$ and $n = 500, \tau = 0.4$ where the rejection rates are much larger than 0.05.

When $n = 100$, the MG pseudo-copula except for **IV** fit the normal pseudo-copula well and all the rejection rates are less than 0.05. Especially for **II**, its rejection rates are smaller than the normal pseudo-copula's rejection rate for all τ's from 0.2 to 0.8. Furthermore, both **I** and **V** fit the t pseudo-copula. In addition, as Kendall's τ increases, **II** is closer to the t pseudo-copula.

TABLE 5.5: The empirical probabilities of rejecting H_0 based on 1000 replications: $n = 500$

Family	τ	Clayton	Frank	Gumbel	normal	t	**I**	**II**	**III**	**IV**	**V**
Clayton	0.2	0.045	0.861	0.998	0.723	0.868	0.6	0.73	0.71	0.8	0.75
	0.4	0.06	1.00	1.00	0.999	0.998	0.9	0.95	0.84	0.93	1.00
	0.6	0.049	1.00	1.00	1.00	1.00	0.75	0.98	0.98	0.64	1.00
	0.8	0.043	1.00	1.00	1.00	1.00	0.333	0.943	1.00	1.00	0.99
Frank	0.2	0.972	0.062	0.831	0.318	0.754	0.18	0.1	0.01	0.12	0.1
	0.4	1.00	0.049	0.999	0.848	0.997	0.36	0.36	0.03	0.78	0.46
	0.6	1.00	0.047	1.00	0.999	1.00	0.17	0.44	0.06	0.96	0.79
	0.8	1.00	0.031	1.00	1.00	1.00	0.07	0.52	0	1.00	0.87
Gumbel	0.2	0.994	0.601	0.04	0.355	0.522	0.37	0.08	0.02	0.03	0.02
	0.4	1.00	0.976	0.043	0.772	0.825	0.81	0.26	0.65	0.77	0.11
	0.6	1.00	1.00	0.032	0.856	0.876	0.86	0.32	0.09	0.99	0.11
	0.8	1.00	1.00	0.032	0.909	0.915	0.875	0.27	0.04	0.82	0.28
normal	0.2	0.891	0.13	0.498	0.041	0.464	0.03	0.01	0.08	0.3	0.02
	0.4	1.00	0.441	0.832	0.037	0.616	0	0	0.71	1.00	0
	0.6	1.00	0.88	0.886	0.047	0.502	0.01	0.01	0.94	1.00	0
	0.8	1.00	0.997	0.852	0.018	0.317	0.125	0.11	0.66	0.86	0.11
t	0.2	0.82	0.462	0.339	0.17	0.042	0.12	0.18	0.03	0.05	0.01
	0.4	1.00	0.845	0.723	0.12	0.042	0.09	0.06	0.11	0.46	0.06
	0.6	1.00	0.986	0.773	0.087	0.041	0.07	0.02	0.27	0.91	0.04
	0.8	1.00	1.00	0.695	0.094	0.041	0.23	0.18	0.76	0.67	0.11

No MG pseudo-copula fits the Clayton pseudo-copula well. This may be a problem with the multiplier GOF test. As we mentioned before, the smaller the sample size, the less power the the multiplier GOF test has. However, one can see that the smaller the Kendall's τ, the closer the MG pseudo-copula is to the Clayton pseudo-copula. Particularly, when the sample size is 100, this feature is notable. For example, when $\tau = 0.2$, rejection rates are 0.090, 0.085, 0.090, 0.140 and 0.095 for ρ type from **I** to **V**. Moreover, as the level of dependence increases for Clayton pseudo-copula, the rejection rates of **I** will increase first and then decrease. For example, when $n = 100$, the rejection rates are 0.090, 0.435, 0.559 and 0.217 for $\tau = 0.2, 0.4, 0.6, 0.8$ respectively.

This may be explained by the fact that the difference between the true pseudo-copula and the MG pseudo-copula increases first and then decreases. From Figure 5.1, the contours of $\mathbf{I}$ look like it has a lower tail, consistent with the properties of the Clayton pseudo-copula.

For most true pseudo-copulas except for Clayton family, the smaller the Kendall's τ, the smaller the rejection rates from the MG pseudo-copula. When the measure of association is small, the distribution of the data is almost the same as the distribution from independence pseudo-copula. The generating pseudo-copula type does not matter in this case, and all the pseudo-copulas under hypothesis have the same properties, i.e., no dependence and small rejection rates.

5.8 Summary and Conclusions

In this paper we introduced and studied the MG pseudo-copula. With a careful choice of definition of $\rho(u, v; a, b)$ and parameters $\boldsymbol{\theta} = (a, b)$, we demonstrated its advantage of flexibility in modeling multivariate data and capturing symmetric as well asymmetric dependence. Although the definitions should not be restricted to the proposed five, our experiments have shown that the MG pseudo-copula with five definitions are flexible enough to fit two real data sets. In contrast to the normal pseudo-copula, the MG pseudo-copula can be used for tailed data sets. Furthermore, the MG pseudo-copula can be used to fit radially symmetric data, which Archimedean pseudo-copulas usually cannot do well. Experimental results on simulated data strongly support these claims about the MG pseudo-copula.

Studies on higher dimensional cases will be carried out in further work although the computation complexity will be a concern. Furthermore, it would be interesting to find a closed form of Kendall's τ.

6 CONCLUSIONS

In this dissertation, we focused on copula models, especially on the Gaussian copula. For any copula, there should be both advantages and the disadvantages. We demonstrated an application of a spatial Gaussian copula in the data having non-continuous marginals and built a modified Gaussian pseudo-copula function to endow the Gaussian copula with more flexibility.

In Chapter 2, we used Gaussian copula-based spatial model to analyze spatially correlated count data. Madsen (2009) introduced a Maximum Likelihood method to perform a Gaussian copula-based spatial analysis. Here, we implemented Bayesian methods to simultaneously estimate the regression parameter and predict missing/unobserved count data. Using Bayesian methods not only provides parameter estimate but also narrows the 95% confidence interval for the regression parameters, compared to Maximum Likelihood (Madsen, 2009). On the other hand, Bayesian methods provide a more accurate prediction for missing values than the Generalized Additive Model (GAM). We carried out a comprehensive comparison between our model and the GAM in Section 2.7. With mean squared prediction error as the measurement, we have shown that our model usually performs better than GAM. This is especially true when the proportion of missing data is large; the count data have high correlations; the sample size is large; and the missing count data is small. In addition, GAM can not provide the count prediction directly. If the missing data are discrete, it needs to use some other schemes to do obtain count-valued predictions.

In Section 3.5, we showed the relation between: the starred copula (Denuit and Lambert, 2005) and the standard extension copula (Nešlehová, 2007), both of which are used to continuously extend discrete data; The model of Song (2000), who proposed the multivariate dispersion models generated from the multivariate Gaussian copula, and Madsen and Fang (2011), who bring discrete distributions into the Gaussian copula framework and enables the parameter estimation of high-dimensional response vectors; Spearman's ρ corresponding to the empirical distribution function introduced by Nešlehová (2007) and Spearman's ρ, which is based on rank calculations, proposed by Gibbons (1985). Despite the different forms and sources, the apparent different quantities in different settings turn out to be the same.

In Chapter 4, we demonstrated we can use AIC as a criterion for choosing the best copula

from a series of candidate copula families. According to our simulation result, using AIC as criterion for choosing copula performs well in copula selection, provided the true copula is among those considered. Compared to the existing multiplier method (Kojadinovic et al., 2011), AIC approach provides more computational efficiency.

In Part 5, we proposed new pseudo-copula, the modified Gaussian pseudo-copula. Although there existed many copula families, none of them are flexible enough to catch both elliptical and tailed distribution. The modified Gaussian pseudo-copula, which preserves almost the same formula as Gaussian copula except for the definition of the dependence structure, is shown to accommodate different tail structure. The pairwise dependence structure depends on unknown parameters as well as the variables themselves. For each pair of marginals, we used two different parameters. One controls the tail, such as lower tail, upper tail or no tail, and the other control the speed of the convergence of the tail, such as fast convergence or slow convergence. The results from Chapter 5 shows this is a valid copula model.

With a careful choice of definition and parameter tuning, it demonstrated the advantage of flexibility in modeling multivariate and capturing symmetric as well asymmetric distributions Although the definitions should not be restricted to the proposed five ρ types, our experiments have shown that the modified Gaussian pseudo-copula with five definitions are flexible enough and almost always guarantee a good fit to the given data set. In contrast to the Gaussian copula, the modified Gaussian pseudo-copula can be used for tailed data sets, i.e., positive-tail and lower-tail. Furthermore, the modified Gaussian pseudo-copula can be used to fit radially symmetric data, which Archimedean copulas usually cannot do well. Experimental results on real insurance data set and simulated data sets strongly support these features of the modified Gaussian pseudo-copula.

Although we applied this model to application in finance, it can be applied to many other fields, such as actuarial science, public health, and medicine and so on. These models are particularly attractive for high dimensional applications involving more than two marginals, as they allow the researcher to increase or decrease the flexibility of the model according the amount of data available, and, importantly, to do so in a manner that is easily interpreted and explained.

Studies on higher dimensions cases will be carried out in further work although the computation complexity might be a concern. Furthermore, it will be interesting to find the closed form to calculate the measure of the association Kendall's τ.

In sum, we introduce a Gaussian copula-based spatial model for count data. We present accessible suggestions for selecting the best copula among the series of given copulas by using AIC. Moreover, we contribute to the field of financial statistics by providing an alternative multivariate model and apply this new pseudo-copula model to real data sets.

BIBLIOGRAPHY

Ang, A., Chen, J., (2002), *Asymmetric Correlations of Equity Portfolios, Journal of Financial Economics*, 63, 443-494.

Akaike, H., 1981, *Likelihood of a model and information criteria, Journal of Econometrics*, 16, 3-14.

Akaike, H.,(1974), *A new look at the statistical model identification, IEEE Transactions on Automatic Control*, 19, (6): 716723.

Ball, D. F.(1964), Loss-on-ignition as an estimate of organic matter and organic carbon in non-calcareous soils, *Journal of Soil Science*, 15:84-92.

Banerjee, S., Carlin B.P., and Gelfand A.E.(2004), *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall/CRC.

Bardossy, A., 2006, *Copula-based geostatistical models for groundwater quality parameters, Water Resources Research*, 42:W11416.

Berg, D., Bakken, H., 2005, *A goodness-of-fit test for copulae based on the probability integral transform, Note, Norwegian Computing Center*, SAMBA/41/05. Norsk Regnesentral, Oslo, Norway.

Berg, D., (2009), *Copula goodness-of-fit testing: An overview and power comparison, The European Journal of Finance*, 15, 675–701.

Berg, D., Quessy, J.F.,(2009), *Local sensitivity analyses of goodness-of-fit tests for copulas, Scandinavian Journal of Statistics*,volume 36, 389–412.

Bozdogan, H., (1987), *Model Selection and Akaikes Information Criterion (AIC): The General Theory and Its Analytical Extensions, Psychometrika*, 52, 345370.

Bradley, B.O., Taqqu, M.S., 2003, *Financial Risk and Heavy Tails, in S.T. Rachev (ed.) Handbook of Heavy-tailed Distributions in Finance*, North Holland, pp. 35–103.

Breymann, W., Dias, A., and Embrechts, P., (2003), *Dependence structures for multivariate high-frequency data in finance, Quantitative Finance*, 3, 1-14.

Brooks, S. P.(1998), *Markov Chain Monte Carlo Method and its Application*, the Statistician, 47, 69-100.

Cherubini, G., Vecchiato, W., and Luciano, E. (2004). *Copula models in finance*. Wiley, New–York.

Cherubini, U., Mulinacci, S., Gobbi, F., and Romagnoli, S., (2011), *Dynamic Copula Methods in Finance*, 2nd ed., John Wiley and Sons, ISBN 1119954525, 9781119954521

Cressie, N. A. C. (1993), *Statistics for Spatial Data*, revised ed., John Wiley and Sons, Inc.

Dalthorp, D. (2004), The generalized linear model for spatial data: assessing the effects of environmental covariates on population density in the field, *Entomologia Experimentalis et Applicata*, 111, 117–131.

Dalthorp, D., Nyrop, J., and Villani, M. (2000), Spatial ecology of the japanese beetle, *Popillia japonica*, *Entomologia Experimentalis et Applicata*, 96, 129–139.

Deheuvels, P. (1979), *La fonction de dépendance empirique et ses propriétés: un test non paramétrique d'indépendance, Acad. Roy. Belg. Bull. Cl. Sci. 5th Ser.* , 65, 274–292.

Deheuvels, P., 1981, *A non parametric test for independence, Publ. Inst. Stat. Univ.*, Paris, 26, 29-50.

Deheuvels, P. (1981), *An asymptotic decomposition for multivariate distribution-free tests of independence, J. Multivariate Anal*, 11, 102–113.

Deheuvels, P., (1981), *A non parametric test for independence, Publ. Inst. Statist. Univ. Paris*, 26, 29–50.

Denuit, M. and Lambert, P. (2005), Constraints on concordance measures in bivariate discrete data, *Journal of Multivariate Analysis*, 93, 40–57.

Demarta, S., McNeil, A.J., 2005, *The t Copula and Related Copulas, International Statistical Review*, 73(1): 111–129

Dobrić, J., and Schmid, F., (2007), *A goodness of fit test for copulas based on Rosenblatts transformation, Computational Statistics & Data Analysis*, 51, 4633-4642.

Embrechts, P., Lindskog F., McNeil A., 2003, *Modeling Dependence with Copulas and Applications to Risk Management, In S Rachev (ed.), Handbook of Heavy Tailed Distribution in Finance*, pp, 329-384, Elsevier.

Embrechts, P., McNeil, A., Straumann, D., 2001, *Correlation and dependency in risk management: properties and pitfalls, In Risk Management: Value at Risk and Beyond*, M. Dempster & H. Moffatt, eds, http://www.math.ethz.ch/ mcneil: Cambridge University Press, pp. 176-223.

Fang, H. B., Fang, K. T., 2002, *The metaelliptical distributions with given marginals, J. Multivariate Anal*, 82, 116.

Fermanian, J.-D. and Wegkamp, M., (2004). *Time dependent copulas*, Preprint.

Fermanian, J.-D., (2005), *Goodness-of-fit tests for copulas*, *J. Multivariate Anal.*, 95, 119-152.

Flanders, H., (1973), *Differentiation under the integral sign, American Mathematical Monthly*, 80 (6): 615627.

Fréchet, M., (1951). *Sur les tableaux de corrélation dont les marges son données*, Ann. Univ. Lyon, Sect. A, 9, 53-77.

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

Gelman, A. and Rubin, D.B. (1992), Inference from iterative simulation using multiple sequences, *Statistical Science*, 7, 457-511.

Gelman, A. , John B. Carlin, Hal S. Stern and Donald B. Rubin (1995), *Bayesian Data Analysis*, Chapman & Hall/CRC.

Gelman, A., Roberts, G. O., Gilks, W. R. (1996), Efficient Metropolis jumping rules, *Bayesian Statistics*, 5, 599608.

Genest, C., Rivest, L.P., 1993, *Statistical inference procedures for bivariate archimedean copulas, Journal of the American Statistical Association*, 1034-1043.

Genest, C., Ghoudi, K., Rivest, L. P., 1995, *A semiparametric estimation procedure of dependence parameters in multivariate families of distributions*, *Biometrika*, 82, 543-552.

Genest, C., Werker, B., (2002), *Conditions for the asymptotic semiparametric efficiency of an omnibus estimator of dependence parameters in copula models, In: C. Cuadras, J. Fortiana and J.R. Lallena, Editors, Distributions With Given Marginals and Statistical Modelling, Kluwer (2002)*, pp. 103112.

Genest, C., and Rémillard, B., (2004). *Tests of independence and randomness based on the empirical copula process, Test*, 13, 335–369.

Genest, C., Favre, A.-C., Béliveau, J., and Jacques, C. (2007), *Meta-elliptical copulas and their use in frequency analysis of multivariate hydrological data, Water Resources Research*, 43, 12 pages.

Genest, C., Nešlehová, J. (2007), A primer on copulas for count data, *Astin Bulletin*, vol.37, No.2,475-515.

Genest, C., Quessy, J.-F., and Rémillard, B., (2006), Goodness-of-fit procedures for copula models based on the integral probability transformation, *Scandinavian Journal of Statistics,* 33, 337-366.

Genest, C., and Rémillard, B., 2008, *Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models*, *Annales de l'Institut Henri Poincaré: Probabilités et statistiques*, 44, 1096-1127.

Genest, C., Rémillard, B., and Beaudoin, D., 2009, *Goodness-of-fit tests for copulas: A review and a power study*, *Insurance: Mathematics and Economics*, 44, 199–214.

Gibbons, Jean Dickinson (1985), Nonparametric statistical inference, Marcel Dekker, Inc.,270 Madison Avenue, New York, New York 10016, first edition.

Gilks, W. R., Richardson, S., Spiegelhalter, D. J. and eds. (1996), Inference and monitoring convergence, *Markov Chain Monte Carlo in Practice*, Chapman and Hall/CRC, Boca Raton, Florida, 131-143.

Hadley, C. H. and Hawley, I. M.(1934), General information about the Japanese beetle in the United States, *United States Department of Agriculture Circular 332*.

Hauksson, H., Dacorogna, M., Domenig, T., Mueller, U., Samorodnitsky, G., (2001). *Multivariate Extremes, Aggregation and Risk Estimation*, Quantitative Finance 1: 7995.

Hastie, T.J. and Tibshirani, R.J. (1984), *Generalized Additive Models*, Tech. Rep. 98, Dept. of Statistics, Stanford University.

Hastie, T.J. and Tibshirani, R.J. (1990), *Generalized Additive Models*, Chapman & Hall/CRC.

Hastings, W. K. (1970), Monte Carlo Sampling Methods Using Markov Chains and Their Applications, *Biometrika*, 57, 97-109.

Hawley, I. M.(1944), Notes on the biology of the Japanese beetle, *United States Department of Agriculture, Agricultural Research Administration, Bureau of Entomology and Plant Quarantine*, E-15.

Hoff, P. D. (2007), Extending the rank likelihood for semi-parametric copula estimation, *Journal of Computational and Theoretical Nanoscience*, 1, 265–283.

Hoeffding, W., (1940), *Masstabinvariante Korrelationstheorie*, Schriften des Matematischen Instituts und des Instituts für Angewandte Matematik de Universität Berlin,5, 179-233 [Reprinted as: Scale-invariant correlation theory. In: Fisher, N. I. and Sen, P. K., editors, (1994). The Collected Works of Wassily Hoeffding, 57-107. Springer, New York.]

Hoeffding, W., (1941), *Masstabinvariante Korrelationsmasse für diskontinuierliche Verteilungen*, Arkiv fr matematischen Wirtschaften und Sozialforschung, 7, 49-70. [Reprinted as: Scale-invariant correlation measures for discontinuous distributions.In: Fisher, N. I. and Sen, P. K., editors, (1994). The Collected Works of Wassily Hoeffding, 109-133. Springer, New York.]

Hu, W., Kercheval, A. N., 2007, *The Skewed t Distribution for Portfolio Credit Risk, Advances in Econometrics*, 2007.

Joe, H., (1997), Multivariate models and dependence concepts, *Chapman and Hall/CRC,* London.

Joe, H. (2001), *Multivariate models and dependence concepts*, Chapman and Hall/CRC.

Johnson, N.L., Kotz, S., Balakrishnan, N., (1994), *Continuous Univariate Distributions,* Volume 1, Wiley. ISBN 0-471-58495-9 (Section 10.1)

Kazianka, H., Pilz, J., 2010, *Copula-based geostatistical modeling of continuous and discrete data including covariates, Stochastic Environmental Research and Risk Assessment*, 24, 661-673.

King, R. and Brooks, S. P.(2002), *Model selection for integrated recovery/recapture data*, Biometrics, 57, 97109.

Klein, E., 2005, *Capital formation, governance and banking, Nova Publishers*, Business & Economics.

Klugman, S. A., and Parsa, R., (1999), Fitting bivariate loss distributions with copulas, *Insurance Mathematicsand Economics,* 24, 139-148.

Kojadinovic, I., Yan, J., and Holmes, M., (2009), *Fast large-sample goodness-of-fit tests for copulas, Technical report 24*, Department of Statistics, The University of Connecticut.

Kojadinovic, I. and Yan, J., (2010), *Comparison of three semiparametric methods for estimating dependence parameters in copula models*, *Insurance: Mathematics and Economics*, 2010, vol. 47, issue 1, pages 52–63.

Kojadinovic, I. and Yan, J., (2010), *Modeling Multivariate Distributions with Continuous Margins Using the copula R Package*, *Journal of Statistical Software*, 2010, vol. 34, 9, pages 1–20.

Kojadinovic, I., Yan, J. and Homes, M., (2011), *Fast large sample goodness-of-fit tests for copulas,Statistica Sinica*, 21:2, in press.

Kojadinovic , I. and Yan, J., (2011), *A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems, Statistics and Computing*, 21, pages 17–30.

Kullback, S., (1968), *Information Theory and Statistics, New York : Dover Books on Mathematics*, 2nd edition.

Lehmann, E. L., Casella, G., (1998), *Theory of Point Estimation, Springer*, ISBN 0-387-98502-6, 2nd edition, page 47.

Li, David X., 2000, *On Default Correlation: A Copula Function Approach, Journal of Fixed Income*,9(4): 43-54.

Liang, Kung-Yee and Zeger, Scott (1986), Longitudinal data analysis using generalized linear models, *Biometrika*, 72, 13-22.

Luetkpohl, H., (199)1, *Introduction to multiple time series, Springer Verlag*,9(4): NY, 500-503.

Madsen, L. (2009), Maximum Likelihood Estimation of Regression Parameters With Spatially Dependent Discrete Data, *Journal of Agricultural, Biological, and Environmental Statistics*, 14, 375-391.

Madsen, L., Fang, Y., (2011), Joint Regression Analysis for Discrete Longitudinal Data, *Biometrics*.

Mashal, R., Zeevi, A., 2002, *Beyond correlation: Extreme co-movements between financial assets*,Working Paper, Columbia University.

McNeil, A.J., and Frey, R., (2000), *Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach, Journal of Empirical Finance*, 7, 271-300.

McNeil, A., Frey, R., and Embrechts, P., (2005), Quantitative risk management, *Princeton University Press,* New Jersey.

Monahan, John F. (2001), *Numerical methods of statistics* , Illustrated, Cambridge Univ Pr, 371.

Nelsen, R. B., (2006), An Introduction to Copulas, 2nd ed., Springer, New York.

Nešlehová, J. (2007), On rank correlation measures for non-continuous random variables, *Journal of Multivariate Analysis*, 98, 544-567.

Patton, Andrew J., (2002), *Applications of Copula Theory in Financial Econometrics, University of California*, San Diego. PhD dissertation.

Patton, Andrew J., (2006). *Modeling asymmetric exchange rate dependence, International Econometric Review*, 47, 527-556.

Pitt, M., Chan, D., and Kohn, R. (2006), Efficient Bayesian inference for Gaussian copula regression models, *Biometrika*, 93, 537–554.

R Development Core Team, 2009, *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna*, Austria, ISBN 3–900051–07–0. http://www.R-project.org.

Rachev, S. T., Sun, W., and Stein, M., 2009, *Copula Concepts in Financial Markets, Technical Report*, University of Karlsruhe.

Rémillard, B., Scaillet, O., 2009, *Testing for equality between two copulas, Journal of Multivariate Analysis*, 100, 377–386.

Sang, H., Gelfand, A. E., (2009), *Continuous spatial process models for spatial extreme values, Journal of Agricultural, Biological and Environmental Statistics*, 15, 4965.

Schweizer, B., Sklar, A. (1974), Operation on distribution functions not derivable from operations on random variables, *Studia Math*, 52, 43-52.

Schmidt, Thorsten, (2006). *Coping with Copulas, In: Copulas From Theory to Applications in Finance*, Risk Books.

Shih, J., Louis, T., 1995, *Inferences on the association parameter in copula models for bivariate survival data, Biometrics* , 51 (4), pp. 13841399.

Sklar, M., 1959, *Fonctions de répartition à n dimensions et leurs marges, Publications de l'Institut de Statistique de l'Université de Paris*, 8, 229-231.

Solomon, D. L., (1983), *The Spatial Distribution of Butterfly Eggs, Life Science Models*, Vol. 4, eds. H. Roberts and M. Thompson, New York: Springer-Verlag, pp. 350366.

Song, P. X.-K. (2000), Multivariate dispersion models generated from Gaussian copula, *Scandinavian Journal of Statistics*, 27, 305–320.

Song, P. X.-K. (2007), *Correlated Data Analysis*, Springer.

Song, P. X.-K., Li, M., and Yuan, Y. (2009), Joint regression analysis of correlated data using Gaussian copulas, *Biometrics*. 65, 60-68.

Stute, W., González-Manteiga, W., and Presedo-Quindimil, M., (1993), *Bootstrap based goodness-of-fit tests, Metrika*, 40, 243256, MR1235086.

Wang, W., and Wells, M. T., (2000), Model selection and semiparametric inference for bivariate failure-time data, *Journal of the American Statistical Association,* 95, 62–72.

Wood S.N. (2006b), Generalized Additive Models: An Introduction with R, *Chapman and Hall/CRC Press*.

Zeger, Scott and Liang, Kung-Yee(1986), Longitudinal Data Analysis for Discrete and Continuous Outcomes, *Biometrics*,42, 121-130.

APPENDICES

A APPENDIX Metropolis-Hastings Algorithm

1. Obtain the starting points $(\beta^{(0)}, \theta_0^{(0)}, \theta_1^{(0)}, \phi^{(0)}, U^{(0)})$;

2. For $n = 1, 2, \ldots$;

   a. update $\beta^{(n)}$;

   b. update $\theta_0^{(n)}$;

   c. update $\theta_1^{(n)}$;

   d. update $\phi^{(n)}$;

   e. update $U^{(n)}$;

   f. predict the new value based on the posterior predictive distribution $(z_{new}^{(n)}|\theta_0^{(n)}, \theta_1^{(n)}, z_{obs}^{(n)})$ and $Y_{new}^{(n)}$ is the inverse of cumulative density function for Negative binomial $(r_i^{(n)}, p^{(n)})$, where $r_i^{(n)} = \phi^{(n)} \times \exp(\beta_0^{(n)} + \beta_1^{(n)} \times x_i + \beta_2^{(n)} \times x_i^2 + \beta_3^{(n)} \times x_i^3)$ and $p^{(n)} = \frac{\phi^{(n)}}{1+\phi^{(n)}}$.

B APPENDIX Gelman-Rubin Statistics

TABLE B.1: Gelman-Rubin Statistics

Parameters	Gelman-Rubin Statistics
β_0	1.024844
β_1	1.020795
β_2	1.019808
β_4	1.01702
θ_0	1.084532
θ_1	1.005148
ϕ	1.020665

C APPENDIX The joint probability mass function from equation (3.15) for (Y_1^*, Y_2^*)

$$
\begin{aligned}
\mathrm{L}(y_1, y_2; \rho) &= E_{\mathbf{Z}}\Big[c(y_1, y_2; \Sigma_\rho)\Big] \\
&= E_{\mathbf{Z}}\Big\{|\Sigma_\rho|^{-1/2}\exp\Big[-\frac{1}{2}\mathbf{Z}^T(\Sigma_\rho^{-1}-\mathbf{I}_2)\mathbf{Z}\Big]\mathrm{f}_1(y_1)\mathrm{f}_2(y_2)\Big\} \\
&= \int_{a_1}^{b_1}\int_{a_2}^{b_2}\Big\{|\Sigma_\rho|^{-1/2}\exp\Big[-\frac{1}{2}\mathbf{Z}^T(\Sigma_\rho^{-1}-\mathbf{I}_2)\mathbf{Z}\Big]\mathrm{P}(Y_1=[y_1^*+1])\mathrm{P}(Y_2=[y_2^*+1])\Big\} \\
&\qquad \mathrm{f}(z_1,z_2)\mathrm{d}z_1\mathrm{d}z_2 \\
&= \int_{a_1}^{b_1}\int_{a_2}^{b_2}\Big\{|\Sigma_\rho|^{-1/2}\exp\Big[-\frac{1}{2}\mathbf{Z}^T(\Sigma_\rho^{-1}-\mathbf{I}_2)\mathbf{Z}\Big]\mathrm{P}(Y_1=[y_1^*+1])\mathrm{P}(Y_2=[y_2^*+1])\Big\} \\
&\qquad \frac{\phi(z_1)\phi(z_2)}{\mathrm{P}(Y_1=[y_1^*+1])\mathrm{P}(Y_2=[y_2^*+1])}\mathrm{d}z_1\mathrm{d}z_2 \\
&= \int_{a_1}^{b_1}\int_{a_2}^{b_2}\Big\{|\Sigma_\rho|^{-1/2}\exp\Big[-\frac{1}{2}\mathbf{Z}^T(\Sigma_\rho^{-1}-\mathbf{I}_2)\mathbf{Z}\Big]\frac{1}{2\pi}\exp\Big[-\frac{1}{2}\mathbf{Z}^T\mathbf{I}_2\mathbf{Z}\Big]\Big\}\mathrm{d}z_1\mathrm{d}z_2 \\
&= \int_{a_1}^{b_1}\int_{a_2}^{b_2}\Big\{\frac{1}{2\pi}|\Sigma_\rho|^{-1/2}\exp\Big[-\frac{1}{2}\mathbf{Z}^T\Sigma_\rho^{-1}\mathbf{Z}\Big]\Big\}\mathrm{d}z_1\mathrm{d}z_2 \\
&= \int_{-\infty}^{\Phi^{-1}(1-p_1)}\int_{\Phi^{-1}(1-p_2)}^{+\infty}\Big\{\frac{1}{2\pi}|\Sigma_\rho|^{-1/2}\exp\Big[-\frac{1}{2}\mathbf{Z}^T\Sigma_\rho^{-1}\mathbf{Z}\Big]\Big\}\mathrm{d}z_1\mathrm{d}z_2 \\
&= \int_{-\infty}^{\Phi^{-1}(1-p_1)}\int_{-\infty}^{+\infty}\Big\{\frac{1}{2\pi}|\Sigma_\rho|^{-1/2}\exp\Big[-\frac{1}{2}\mathbf{Z}^T\Sigma_\rho^{-1}\mathbf{Z}\Big]\Big\}\mathrm{d}z_1\mathrm{d}z_2 \\
&\quad - \int_{-\infty}^{\Phi^{-1}(1-p_1)}\int_{-\infty}^{\Phi^{-1}(1-p_2)}\Big\{\frac{1}{2\pi}|\Sigma_\rho|^{-1/2}\exp\Big[-\frac{1}{2}\mathbf{Z}^T\Sigma_\rho^{-1}\mathbf{Z}\Big]\Big\}\mathrm{d}z_1\mathrm{d}z_2 \\
&= \Phi\Big(\Phi^{-1}(1-p_1)\Big) - \int_{-\infty}^{\Phi^{-1}(1-p_1)}\int_{-\infty}^{\Phi^{-1}(1-p_2)}\Big\{\frac{1}{2\pi}|\Sigma_\rho|^{-1/2}\exp\Big[-\frac{1}{2}\mathbf{Z}^T\Sigma_\rho^{-1}\mathbf{Z}\Big]\Big\}\mathrm{d}z_1\mathrm{d}z_2 \\
&= 1-p_1 - \int_{-\infty}^{\Phi^{-1}(1-p_1)}\int_{-\infty}^{\Phi^{-1}(1-p_2)}\Big\{\frac{1}{2\pi}|\Sigma_\rho|^{-1/2}\exp\Big[-\frac{1}{2}\mathbf{Z}^T\Sigma_\rho^{-1}\mathbf{Z}\Big]\Big\}\mathrm{d}z_1\mathrm{d}z_2 \\
&= 1-p_1-\Phi_{\Sigma_\rho}\Big(\Phi^{-1}(1-p_1),\Phi^{-1}(1-p_2)\Big). \qquad \text{(C.1)}
\end{aligned}
$$

D APPENDIX The joint probability mass function from equation (3.16) for (Y_1^*, Y_2^*)

$$
\begin{aligned}
\mathrm{g}(y_1, y_2; \Sigma_\rho) &= \mathrm{P}(Y_1 = y_1, Y_2 = y_2) \\
&= \sum_{j_1=1}^{2} \sum_{j_2=1}^{2} (-1)^{j_1+j_2} \Phi_{\Sigma_\rho}\Big(\Phi^{-1}\{\mu_{1j_1}\}, \Phi^{-1}\{\mu_{2j_2}\}\Big) \\
&= (-1)^{1+1}\Phi_{\Sigma_\rho}\Big(\Phi^{-1}\{\mu_{11}\}, \Phi^{-1}\{\mu_{21}\}\Big) + (-1)^{1+2}\Phi_{\Sigma_\rho}\Big(\Phi^{-1}\{\mu_{11}\}, \Phi^{-1}\{\mu_{22}\}\Big) \\
&\quad + (-1)^{2+1}\Phi_{\Sigma_\rho}\Big(\Phi^{-1}\{\mu_{12}\}, \Phi^{-1}\{\mu_{21}\}\Big) + (-1)^{2+2}\Phi_{\Sigma_\rho}\Big(\Phi^{-1}\{\mu_{12}\}, \Phi^{-1}\{\mu_{22}\}\Big) \\
&= \Phi_{\Sigma_\rho}\Big(\Phi^{-1}(0), \Phi^{-1}(1-p_2)\Big) - \Phi_{\Sigma_\rho}\Big(\Phi^{-1}(0), \Phi^{-1}(1)\Big) \\
&\quad - \Phi_{\Sigma_\rho}\Big(\Phi^{-1}(1-p_1), \Phi^{-1}(1-p_2)\Big) + \Phi_{\Sigma_\rho}\Big(\Phi^{-1}(1-p_1), \Phi^{-1}(1)\Big) \\
&= \Phi_{\Sigma_\rho}\Big(-\infty, \Phi^{-1}(1-p_2)\Big) - \Phi_{\Sigma_\rho}\Big(-\infty, +\infty\Big) \\
&\quad - \Phi_{\Sigma_\rho}\Big(\Phi^{-1}(1-p_1), \Phi^{-1}(1-p_2)\Big) + \Phi_{\Sigma_\rho}\Big(\Phi^{-1}(1-p_1), +\infty\Big) \\
&= \Phi\Big(\Phi^{-1}(1-p_1)\Big) - \Phi_{\Sigma_\rho}\Big(\Phi^{-1}(1-p_1), \Phi^{-1}(1-p_2)\Big) \\
&= 1 - p_1 - \Phi_{\Sigma_\rho}\Big(\Phi^{-1}(1-p_1), \Phi^{-1}(1-p_2)\Big).
\end{aligned}
\tag{D.1}
$$

E APPENDIX Simulation results for AIC chapter: Table and Figure for n=300 and 500

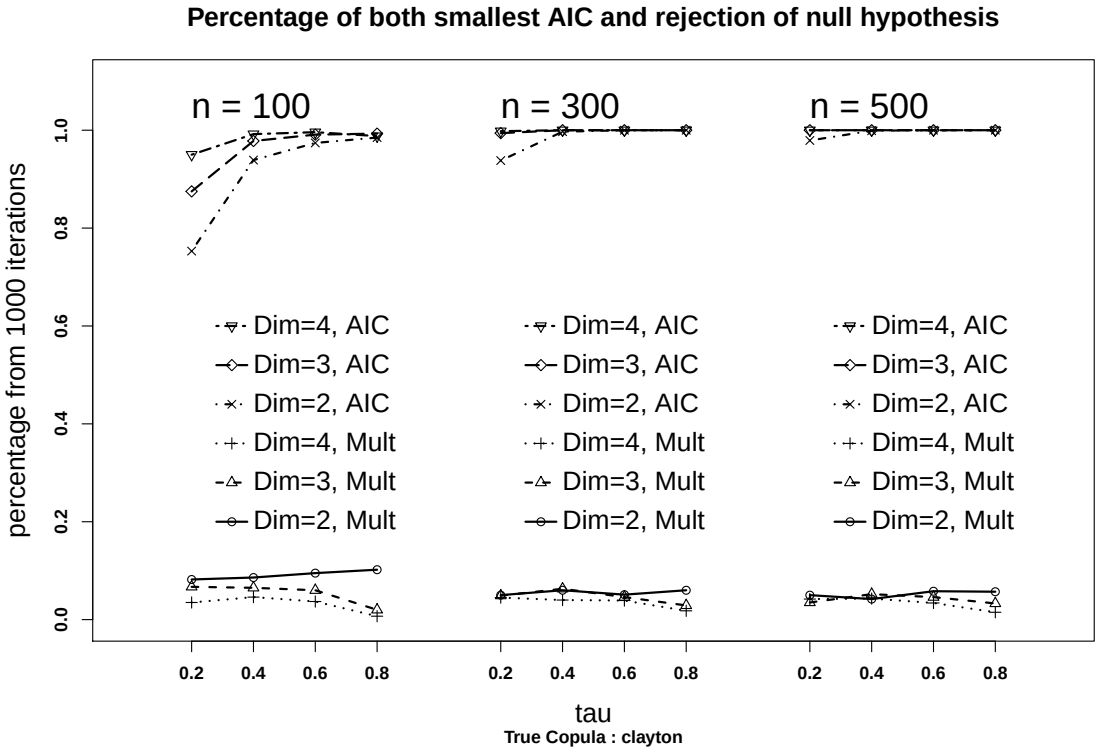


FIGURE E.1: Plot of the correct rate of AIC and the empirical level of Clayton copula with Clayton copula as the generating family.

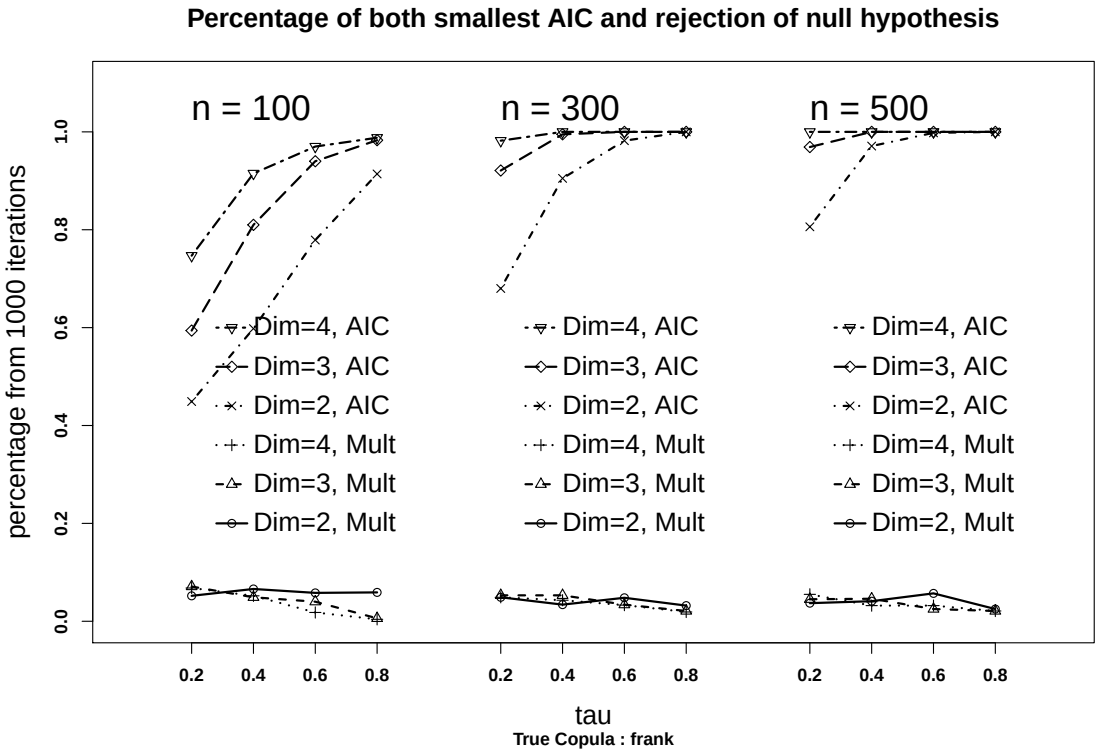


FIGURE E.2: Plot of the correct rate of AIC and the empirical level of Frank copula with Frank copula as the generating family.

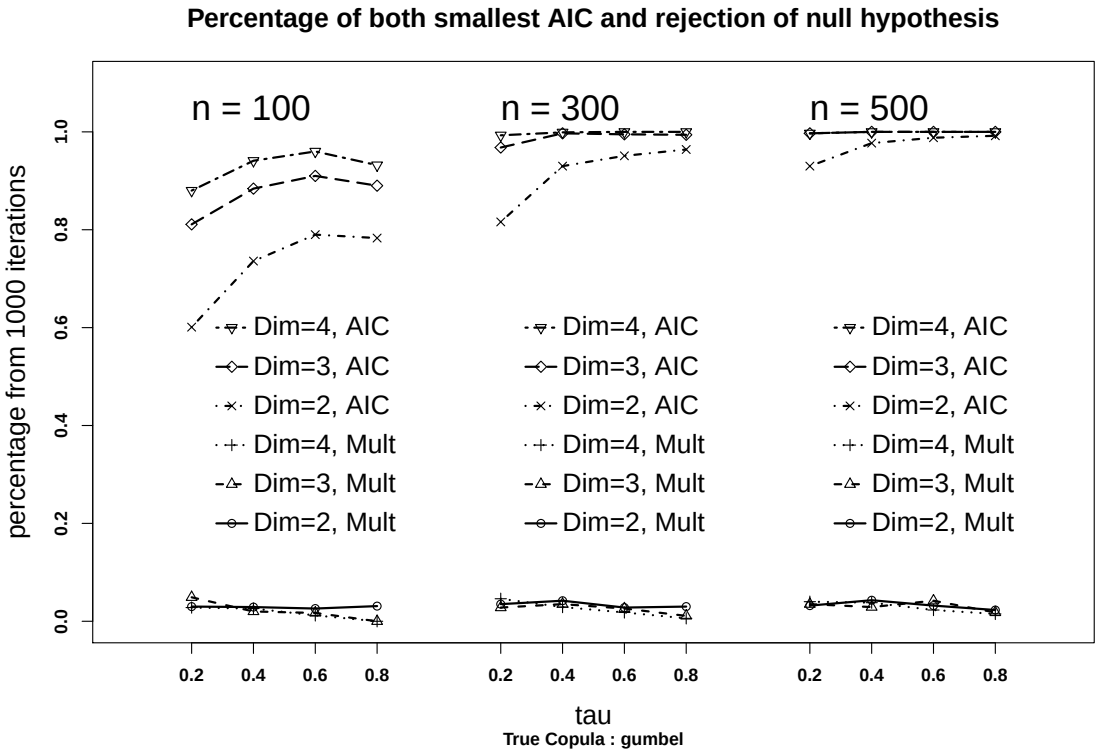


FIGURE E.3: Plot of the correct rate of AIC and the empirical level of Gumbel copula with Gumbel copula as the generating family.

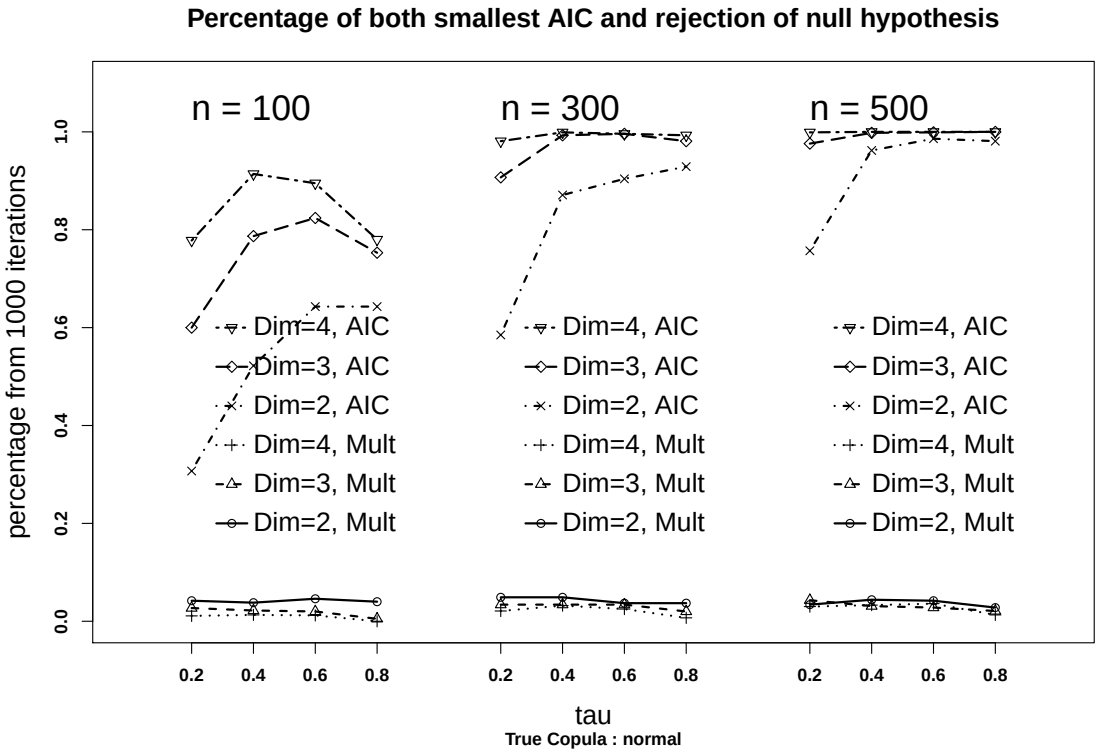


FIGURE E.4: Plot of the correct rate of AIC and the empirical level of Normal copula with Normal copula as the generating family.

TABLE E.1: The empirical levels and the rates of the least AIC with sample size $n = 300$ and sample dimensions $d = 2$

True	Kendall's	the rate of least AIC					the rejection proportions of H_0				
Copula	τ	G	C	F	N	t	G	C	F	N	t
C	0.2	0	93.8	1.9	2.7	1.6	94.0	5.0	60.0	44.0	58.0
	0.4	0	99.7	0.1	0.1	0.1	100	6.0	99.6	97.4	98.5
	0.6	0	100	0	0	0	100	5.1	100	100	100
	0.8	0	100	0	0	0	100	6.0	100	100	100
F	0.2	6.2	4.2	68.0	18.6	3.0	55.6	87.4	4.9	17.7	43.2
	0.4	0.5	0	90.5	8.3	0.7	94.5	100	3.4	58.0	90.5
	0.6	0.3	0	98.2	1.4	0.1	99.8	100	4.8	94.1	99.4
	0.8	0	0	100	0	0	100	100	3.2	99.8	100
G	0.2	81.6	0.2	3.9	8.4	5.9	3.5	96.3	41.6	24.3	33.0
	0.4	93.0	0	0.5	2.7	3.8	4.2	100	81.8	51.2	57.3
	0.6	95.1	0	0.1	1.5	3.3	2.8	100	97.3	64.5	68.4
	0.8	96.4	0	0.1	0.7	2.8	3.0	100	99.7	69.3	68.2
N	0.2	1.26	7.9	18.6	58.5	2.4	32.0	70.5	9.1	4.9	25.5
	0.4	4.8	0.8	4.2	87.1	3.1	57.9	99.7	30.0	4.9	36.8
	0.6	1.8	0	1.5	90.4	6.2	62.3	100	66.2	3.7	26.5
	0.8	1.0	0	0.1	92.9	6.0	53.8	100	94.5	3.7	19.4
t	0.4	5.5	4.2	1.6	1.3	87.4	21.7	61.0	23.8	11.4	4.4
	0.6	4.1	0.2	0.6	1.7	93.4	39.1	99.5	58.8	9.0	4.4
	0.8	2.0	0	0.2	2.3	95.5	50.7	100	83.8	7.7	4.4
	0.2	1.6	0	0.1	3.5	94.8	45.0	100	96.8	8.1	3.7

TABLE E.2: The empirical levels and the rates of the least AIC with sample size $n = 300$ and sample dimensions $d = 3$

True	Kendall's	the rate of least AIC					the rejection proportions of H_0				
Copula	τ	G	C	F	N	t	G	C	F	N	t
C	0.2	0	99.4	0.2	0.3	0.1	98.9	4.9	86.3	7.26	89.8
	0.4	0	100	0	0	0	100	6.3	100	99.5	99.9
	0.6	0	100	0	0	0	100	4.6	100	100	100
	0.8	0	100	0	0	0	100	2.9	100	100	100
F	0.2	1.6	0.5	92.1	5.7	0.1	58.0	93.1	5.3	7.2	46.8
	0.4	0.1	0	99.5	0.4	0	97.2	100	5.3	42.5	93.7
	0.6	0	0	100	0	0	99.9	100	3.4	91.0	99.7
	0.8	0	0	100	0	0	100	100	2.1	99.0	100
G	0.2	96.8	0	2.1	1.0	0.1	2.8	98.5	46.7	36.3	18.4
	0.4	99.7	0	0.2	0.1	0	3.5	100	93.3	68.8	50.9
	0.6	99.5	0	0.2	0.2	0.1	2.6	100	99.5	70.6	62.8
	0.8	99.4	0	0	0.1	0.5	1.1	100	100	59.2	55.6
N	0.2	2.2	2.3	4.7	90.7	0.1	85.3	95.0	46.9	3.4	63.8
	0.4	0.1	0	0.3	99.3	0.3	98.9	100	84.3	3.4	75.0
	0.6	0.1	0	0	99.6	0.3	99.4	100	96.1	3.4	59.6
	0.8	0.1	0	0	98.1	1.8	94.8	100	99.8	2.0	29.0
t	0.2	0.2	0.2	0	0	99.6	63.2	89.4	51.7	11.6	4.1
	0.4	0.1	0.1	0	0.2	99.6	91.6	100	89.5	8.8	3.8
	0.6	0	0	0	0.2	99.8	95.6	100	99.0	6.1	3.0
	0.8	0	0	0	0.3	99.7	90.5	100	99.9	4.0	1.6

TABLE E.3: The empirical levels and the rates of the least AIC with sample size $n = 300$ and sample dimensions $d = 4$

True	Kendall's	the rate of least AIC					the rejection proportions of H_0				
Copula	τ	G	C	F	N	t	G	C	F	N	t
C	0.2	0	99.8	0	0.2	0	99.6	4.5	93.5	88.1	96.3
	0.4	0	100	0	0	0	100	4.0	99.9	99.8	100
	0.6	0	100	0	0	0	100	3.9	100	100	100
	0.8	0	100	0	0	0	100	1.8	100	100	99.9
F	0.2	0.8	0	98.2	1.0	0	52.9	96.0	4.9	3.2	38.6
	0.4	0	0	100	0	0	96.2	100	4.3	21.7	91.0
	0.6	0	0	100	0	0	100	100	3.3	82.0	99.5
	0.8	0	0	100	0	0	100	100	1.9	99.5	99.6
G	0.2	99.3	0	0.5	0.2	0	4.6	99.4	54.7	56.3	7.5
	0.4	99.9	0	0	0.1	0	2.9	100	96.4	84.6	31.0
	0.6	100	0	0	0	0	1.8	100	99.6	79.4	55.2
	0.8	100	0	0	0	0	0.6	100	100	53.1	46.8
N	0.2	0.3	0.3	1.3	98.1	0	98.7	99.7	90.2	2.1	86.3
	0.4	0	0	0.3	99.9	0.1	100	100	99.8	3.2	92.4
	0.6	0	0	0	99.6	0.4	100	100	100	2.5	74.2
	0.8	0	0	0	99.3	0.7	99.6	100	100	0.7	27.2
t	0.2	0	0	0	0	100	90.3	95.5	85.6	13.8	4.2
	0.4	0	0	0	0	100	99.5	100	98.7	7.8	3.4
	0.6	0	0	0	0	100	99.9	100	99.9	6.9	2.9
	0.8	0	0	0	0.2	99.8	98.7	100	100	2.8	0.5

TABLE E.4: The empirical levels and the rates of the least AIC with sample size $n = 500$ and sample dimensions $d = 2$

True	Kendall's	the rate of least AIC					the rejection proportions of H_0				
Copula	τ	G	C	F	N	t	G	C	F	N	t
C	0.2	0	97.9	0.5	1.2	0.4	99.3	5.0	84.7	71.9	87.8
	0.4	0	99.9	0	0.1	0	100	4.2	100	99.9	100
	0.6	0	100	0	0	0	100	5.8	100	100	100
	0.8	0	100	0	0	0	100	5.7	100	100	100
F	0.2	1.6	1.5	80.6	16.0	0.3	84.6	97.3	3.7	28.9	77.1
	0.4	0	0	97.1	2.9	0	99.9	100	4.1	84.1	99.3
	0.6	0	0	99.8	0.1	0.1	100	100	5.7	99.8	99.9
	0.8	0	0	100	0	0	100	100	2.5	100	100
G	0.2	93.0	0	1.5	3.5	2.0	3.2	99.6	59.1	34.6	50.1
	0.4	97.7	0	0.2	1.1	1.0	4.3	100	97.0	75.0	80.6
	0.6	98.8	0	0	0.5	0.7	3.2	100	100	87.8	88.4
	0.8	99.2	0	0	0.2	0.6	2.3	100	100	89.0	89.1
N	0.2	6.5	3.6	13.6	75.5	0.6	51.7	88.2	10.2	3.4	45.2
	0.4	1.2	0.2	1.5	96.2	0.9	81.2	100	48.1	4.4	58.7
	0.6	0.1	0	0	98.6	1.3	88.5	100	88.2	4.2	51.1
	0.8	0.1	0	0.2	98.1	1.6	85.4	100	99.9	2.8	30.5
t	0.2	2.5	0.3	0.2	0.3	96.7	33.5	81.9	46.3	16.2	4.8
	0.4	1.3	0	0	1.4	97.3	66.5	100	85.0	13.0	4.5
	0.6	1.1	0	0	0.4	98.5	75.9	100	98.9	9.6	3.8
	0.8	0	0	0	1.1	98.9	73.6	100	100	8.2	3.3

TABLE E.5: The empirical levels and the rates of the least AIC with sample size $n = 500$ and sample dimensions $d = 3$

True	Kendall's	the rate of least AIC					the rejection proportions of H_0				
Copula	τ	G	C	F	N	t	G	C	F	N	t
C	0.2	0	100	0	0	0.1	100	3.5	98.7	94.2	99.6
	0.4	0	100	0	0	0.1	100	5.2	100	100	100
	0.6	0	100	0	0	0.1	100	4.6	100	100	100
	0.8	0	100	0	0	0.1	100	3.3	100	100	100
F	0.2	0.4	0	96.9	2.7	0	83.2	99.0	4.5	1.27	75.4
	0.4	0	0	100	0	0	100	100	4.6	71.0	99.8
	0.6	0	0	100	0.1	0	100	100	2.5	99.4	100
	0.8	0	0	100	0	0	100	100	2.1	100	100
G	0.2	99.7	0	0.2	0.1	0	3.5	99.8	73.5	66.3	38.4
	0.4	100	0	0	0	0	2.9	100	99.6	91.8	79.7
	0.6	100	0	0	0	0	4.2	100	100	94.5	90.5
	0.8	100	0	0	0	0	1.8	100	100	87.0	83.7
N	0.2	0.6	0.3	1.5	97.6	0	97.6	99.7	58.5	4.3	90.2
	0.4	0	0	0	99.8	0.1	100	100	97.9	3.1	96.3
	0.6	0	0	0	99.9	0.1	100	100	99.9	2.8	88.3
	0.8	0	0	0	100	0	99.7	100	100	2.1	54.5
t	0.2	0	0	0	0	100	83.5	98.0	76.1	21.2	4.2
	0.4	0	0	0	0	100	99.3	100	99.3	11.8	3.2
	0.6	0	0	0	0	100	99.8	100	100	9.6	3.7
	0.8	0	0	0	0.1	99.9	99.5	100	100	6.4	2.1

TABLE E.6: The empirical levels and the rates of the least AIC with sample size $n = 500$ and sample dimensions $d = 4$

True	Kendall's	the rate of least AIC					the rejection proportions of H_0				
Copula	τ	G	C	F	N	t	G	C	F	N	t
C	0.2	0	100	0	0	0	100	4.2	99.3	98.5	99.9
	0.4	0	100	0	0	0	100	4.3	100	100	100
	0.6	0	100	0	0	0	100	3.4	100	100	100
	0.8	0	100	0	0	0	100	1.5	100	100	100
F	0.2	0	0	100	0	0	78.5	99.7	5.5	6.6	73.1
	0.4	0	0	100	0	0	100	100	3.2	50.0	99.8
	0.6	0	0	100	0	0	100	100	3.2	98.6	100
	0.8	0	0	100	0	0	100	100	2.1	100	100
G	0.2	99.7	0	0.2	0.1	0	4.0	100	77.9	86.0	15.9
	0.4	100	0	0	0	0	3.8	100	99.8	98.2	59.7
	0.6	100	0	0	0	0	2.3	100	100	97.7	85.6
	0.8	100	0	0	0	0	1.5	100	100	85.6	79.0
N	0.2	0	0	0.1	99.9	0	100	100	97.6	3.0	99.1
	0.4	0	0	0	100	0	100	100	100	3.3	99.5
	0.6	0	0	0	100	0	100	100	100	3.6	96.8
	0.8	0	0	0	100	0	100	100	100	1.3	69.6
t	0.2	0	0	0	0	100	98.2	99.8	95.8	21.3	2.6
	0.4	0	0	0	0	100	100	100	100	13.8	3.2
	0.6	0	0	0	0	100	100	100	100	9.0	3.1
	0.8	0	0	0	0	100	100	100	100	4.8	0.8

F APPENDIX Proof of equation (5.6) to be a pseudo-copula

According to Fermanian and Wegkamp (2004), logically, pseudo-copula must satisfy three properties. Here, let us empirically demonstrate whether the function $\mathrm{C}(u, v; \theta)$ defined in equation (5.6) possesses all three properties which are needed for a pseudo-copula.

Property 1: *For every* $u, v \in [0, 1]$, $\mathrm{C}(0, v; \boldsymbol{\theta}) = 0$ *where* $\boldsymbol{\theta} = (a, b)$ *and* $\mathrm{C}(u, 0; \boldsymbol{\theta}) = 0$*;*

Proof. When $u = 0$, i. e. $\Phi^{-1}(u) = -\infty$,

$$
\begin{aligned}
&\mathrm{C}(0, v; \boldsymbol{\theta}) \\
&= \frac{1}{\boldsymbol{K}} \int_0^0 \int_0^v \frac{1}{\sqrt{1-\rho^2(x,y;\boldsymbol{\theta})}} \times \exp\left\{ \frac{[\Phi^{-1}(x)]^2 + [\Phi^{-1}(y)]^2}{2} \right\} \\
&\times \exp\left\{ - \frac{[\Phi^{-1}(x)]^2 - 2\rho(x,y;\boldsymbol{\theta})\Phi^{-1}(x)\Phi^{-1}(v) + [\Phi^{-1}(y)]^2}{2\left(1-\rho^2(x,y;\boldsymbol{\theta})\right)} \right\} dxdy \\
&= \frac{1}{\boldsymbol{K}} \int_0^v \left\{ \frac{1}{\sqrt{2\pi}} \exp\left(\frac{[\Phi^{-1}(y)]^2}{2} \right) \right\} \\
&\times \left\{ \int_0^0 \left[\frac{1}{\sqrt{2\pi[1-\rho^2(x,y;\boldsymbol{\theta})]}} \times \exp\left\{ \frac{[\Phi^{-1}(x)]^2}{2} \right\} \right. \right. \\
&\left. \left. \times \exp\left\{ - \frac{[\Phi^{-1}(x)]^2 - 2\rho(x,y;\boldsymbol{\theta})\Phi^{-1}(x)\Phi^{-1}(v) + [\Phi^{-1}(y)]^2}{2\left(1-\rho^2(x,y;\boldsymbol{\theta})\right)} \right\} \right] dx \right\} dy \\
&= 0,
\end{aligned}
$$

i. e, $\mathrm{C}(0, v; \boldsymbol{\theta}) = 0$. The proof for $\mathrm{C}(u, 0; \boldsymbol{\theta}) = 0$ is the same as the proof of $\mathrm{C}(0, v; \boldsymbol{\theta}) = 0$; □

Property 2: $\mathrm{C}(1, 1; \boldsymbol{\theta}) = 1$*;*

Proof. When both u and v are 1, then

$$\begin{aligned}
&\mathrm{C}(1,1;\boldsymbol{\theta}) \\
&= \frac{1}{\boldsymbol{K}} \int_0^1 \int_0^1 \frac{1}{\sqrt{1-\rho^2(x,y;\boldsymbol{\theta})}} \times \exp\left\{\frac{[\Phi^{-1}(x)]^2 + [\Phi^{-1}(y)]^2}{2}\right\} \\
&\times \exp\left\{-\frac{[\Phi^{-1}(x)]^2 - 2\rho(x,y;\boldsymbol{\theta})\Phi^{-1}(x)\Phi^{-1}(v) + [\Phi^{-1}(y)]^2}{2\left(1-\rho^2(x,y;\boldsymbol{\theta})\right)}\right\} dxdy \\
&= \frac{1}{\boldsymbol{K}} \int_0^1 \int_0^1 \mathrm{g}(x,y;\boldsymbol{\theta}) dxdy \\
&= \frac{1}{\boldsymbol{K}} \times \boldsymbol{K} \\
&= 1.
\end{aligned}$$

Since $\boldsymbol{K}$ is a normalized factor of $\mathrm{g}(x,y;\boldsymbol{\theta})$ defined in sub-section 5.3.3, $\int_0^1 \int_0^1 \mathrm{g}(x,y;\boldsymbol{\theta}) dxdy = \boldsymbol{K}$. Accordingly, $\mathrm{C}(1,1;\boldsymbol{\theta}) = 1$. □

Property 3: *For every $u_1, u_2, v_1, v_2 \in [0,1]$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$,*

$$\mathrm{C}(u_2,v_2;\boldsymbol{\theta}) - \mathrm{C}(u_1,v_2;\boldsymbol{\theta}) - \mathrm{C}(u_2,v_1;\boldsymbol{\theta}) + \mathrm{C}(u_1,v_1;\boldsymbol{\theta}) \geq 0. \tag{F.1}$$

Proof. With function $\mathrm{c}(x,y;\boldsymbol{\theta})$ being demonstrated as a p.d.f in sub-section 5.3.3, we have

$$\begin{aligned}
&\mathrm{C}(u,v;\boldsymbol{\theta}) \\
&= \int_0^u \int_0^v \mathrm{c}(x,y;\boldsymbol{\theta}) dxdy \\
&= \frac{1}{\boldsymbol{K}} \int_0^u \int_0^v \frac{1}{\sqrt{1-\rho^2(x,y;\boldsymbol{\theta})}} \times \exp\left\{\frac{[\Phi^{-1}(x)]^2 + [\Phi^{-1}(y)]^2}{2}\right\} \\
&\times \exp\left\{-\frac{[\Phi^{-1}(x)]^2 - 2\rho(x,y;\boldsymbol{\theta})\Phi^{-1}(x)\Phi^{-1}(v) + [\Phi^{-1}(y)]^2}{2\left(1-\rho^2(x,y;\boldsymbol{\theta})\right)}\right\} dxdy.
\end{aligned}$$

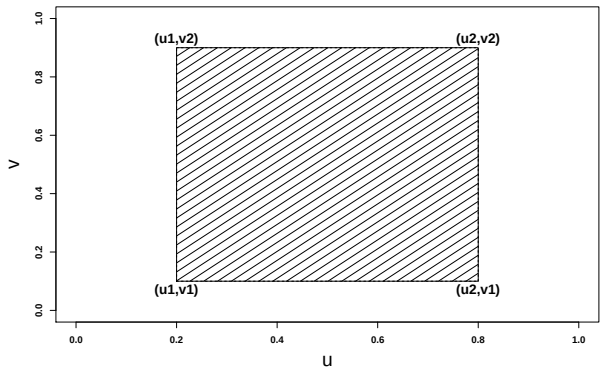


FIGURE F.1: Copula Volume

Look at the Figure F.1$\mathrm{C}(u_2, v_2; \boldsymbol{\theta}) - \mathrm{C}(u_1, v_2; \boldsymbol{\theta}) - \mathrm{C}(u_2, v_1; \boldsymbol{\theta}) + \mathrm{C}(u_1, v_1; \boldsymbol{\theta})$ defines the volume for the shaded part in the figure above with the height $\mathrm{c}(u, v; \boldsymbol{\theta}), \forall [u_1, u_2] \times [v_1, v_2]$. Since we know $\forall u, v \in [0, 1]$, $\mathrm{c}(u, v; \boldsymbol{\theta})$ is a p.d.f, which is clarified in sub-section 5.3.3. Then function $\mathrm{C}(u_2, v_2; \boldsymbol{\theta}) - \mathrm{C}(u_1, v_2; \boldsymbol{\theta}) - \mathrm{C}(u_2, v_1; \boldsymbol{\theta}) + \mathrm{C}(u_1, v_1; \boldsymbol{\theta})$ is the c.d.f on $[u_1, u_2] \times [v_1, v_2]$, i. e.,

$$
\begin{aligned}
&\mathrm{C}(u_2, v_2; \boldsymbol{\theta}) - \mathrm{C}(u_1, v_2; \boldsymbol{\theta}) - \mathrm{C}(u_2, v_1; \boldsymbol{\theta}) + \mathrm{C}(u_1, v_1; \boldsymbol{\theta}) \\
&= P\left(u_1 \leq u \leq u_2, v_1 \leq v \leq v_2\right) \\
&= \int_{u_1}^{u_2} \int_{v_1}^{v_2} \mathrm{c}(x, y; \boldsymbol{\theta}) dx dy \\
&\geq 0,
\end{aligned}
$$

$$\Rightarrow \mathrm{C}(u_2, v_2; \boldsymbol{\theta}) - \mathrm{C}(u_1, v_2; \boldsymbol{\theta}) - \mathrm{C}(u_2, v_1; \boldsymbol{\theta}) + \mathrm{C}(u_1, v_1; \boldsymbol{\theta}) \geq 0.$$

□

All of the above shows us equation (5.6) satisfies the conditions for a two-dimensional pseudo-copula. It is easy to extend C from 2-dimension to p-dimensional and C is still a pseudo-copula. The proof for p-dimensions is omitted in this paper.

G APPENDIX Demonstration for satisfying three regularity conditions

The multiplier method proposed by Kojadinovic et al. (2011) is based on Theorem 2, i.e., the multiplier central limit theorem, stated by Kojadinovic et al. (2011). That is,

Theorem 1. ***(Multiplier central limit theorem)*** *Let $\boldsymbol{\theta}_n$ be an estimate of $\boldsymbol{\theta}$ defined in equation (5.8) and, for any $k \in \{1, \ldots, N\}$ and N is a large integer, let*

$$\boldsymbol{\Theta}_n^{(k)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i^{(k)} \mathfrak{J}(U_{i,n}, V_{i,n}; \boldsymbol{\theta}_n), \tag{G.1}$$

where $Z_i^{(k)}$ with $i \in \{1, \ldots, n\}$ and $k \in \{1, \ldots, N\}$ are i.i.d. random variables with mean 0 and variance 1 independent of the data $(\boldsymbol{X}, \boldsymbol{Y})$, and

$$\begin{aligned}\mathfrak{J}(U_{i,n}, V_{i,n}; \boldsymbol{\theta}_n) = J(U_{i,n}, V_{i,n}; \boldsymbol{\theta}_n) &+ \frac{1}{n} \sum_{j=1}^{n} J^{[1]}(U_{j,n}, V_{j,n}; \boldsymbol{\theta}_n)\{\mathrm{I}\big(U_{i,n} \leq U_{j,n}\big) - \leq U_{j,n}\} \\ &+ \frac{1}{n} \sum_{l=1}^{n} J^{[2]}(U_{j,n}, V_{j,n}; \boldsymbol{\theta}_n)\{\mathrm{I}\big(V_{i,n} \leq V_{j,n}\big) - \leq V_{j,n}\}, \end{aligned} \tag{G.2}$$

with

$$J(U_{j,n}, V_{j,n}; \boldsymbol{\theta}_n) = \Big[\mathbb{E}\Big\{\frac{\dot{c}^2(U, V; \boldsymbol{\theta}_n)}{c^2(U, V; \boldsymbol{\theta}_n)}\Big\}\Big]^{-1} \frac{\dot{c}(U_{j,n}, V_{j,n}; \boldsymbol{\theta}_n)}{c(U_{j,n}, V_{j,n}; \boldsymbol{\theta}_n)},$$

and

$$J^{[1]}(U_{j,n}, V_{j,n}; \boldsymbol{\theta}_n) = \frac{\partial J(U_{j,n}, V_{j,n}; \boldsymbol{\theta}_n)}{\partial u}; \quad J^{[2]}(U_{j,n}, V_{j,n}; \boldsymbol{\theta}_n) = \frac{\partial J(U_{j,n}, V_{j,n}; \boldsymbol{\theta}_n)}{\partial v},$$

where $\dot{c}(U_{j,n}, V_{j,n}; \boldsymbol{\theta}_n) = \frac{\partial c(U_{j,n}, V_{j,n}; \boldsymbol{\theta}_n)}{\partial \boldsymbol{\theta}_n}$, and $U_{j,n} \in (0,1)$ and $V_{j,n} \in (0,1)$. From Theorem 1 proposed by Kojadinovic et al. (2011), the empirical copula process $\sqrt{n}\Big(\mathrm{C}_n(u,v) - \mathrm{C}(u,v;\boldsymbol{\theta})\Big)$ converges weakly in $l^\infty\big([0,1]^2\big)$ to the tight centered Gaussian process

$$\mathbb{C}(u, v; \boldsymbol{\theta}) = \alpha(u, v; \boldsymbol{\theta}) - \mathrm{C}^{[1]}(u, v; \boldsymbol{\theta})\alpha(u, 1; \boldsymbol{\theta}) - \mathrm{C}^{[2]}(u, v; \boldsymbol{\theta})\alpha(1, v; \boldsymbol{\theta}), \tag{G.3}$$

where $\mathrm{C}^{[1]}(u, v; \boldsymbol{\theta}) = \frac{\partial \mathrm{C}(u,v;\boldsymbol{\theta})}{\partial u}$, $\mathrm{C}^{[2]}(u, v; \boldsymbol{\theta}) = \frac{\partial \mathrm{C}(u,v;\boldsymbol{\theta})}{\partial v}$, and $\alpha(u, v; \boldsymbol{\theta})$ is a $\mathrm{C}_{\boldsymbol{\theta}}$ - Brownian bridge with covariance function $\mathbb{E}\left[\alpha(u, v; \boldsymbol{\theta})\alpha(u', v'; \boldsymbol{\theta})\right] = \mathrm{C}(u \wedge u', v \wedge v'; \boldsymbol{\theta}) - \mathrm{C}(u, v; \boldsymbol{\theta})\mathrm{C}(u', v'; \boldsymbol{\theta})$ where $u, v, u', v' \in (0, 1)$. Remillard and Scaillet (2009) suggested the consistent estimators of the partial derivatives $\mathrm{C}^{[1]}(u, v; \boldsymbol{\theta})$ and $\mathrm{C}^{[2]}(u, v; \boldsymbol{\theta})$, which are

$$\mathrm{C}_n^{[1]}(u, v) = \frac{1}{2n^{-1/2}} \Big\{\mathrm{C}_n(u + n^{-1/2}, v) - \mathrm{C}_n(u - n^{-1/2}, v)\Big\}$$

and

$$\mathrm{C}_n^{[2]}(u,v) = \frac{1}{2n^{-1/2}}\left\{\mathrm{C}_n(u, v+n^{-1/2}) - \mathrm{C}_n(u, v-n^{-1/2})\right\},$$

respectively. Then the empirical process

$$\mathbb{C}_n(u,v) = \alpha_n(u,v) - \mathrm{C}^{[1]}(u,v)\alpha_n(u,1) - \mathrm{C}^{[2]}(u,v)\alpha_n(1,v), \tag{G.4}$$

where $\alpha_n(u,v) = \frac{1}{\sqrt{n}}\sum_{i=1}^n Z_i^{(k)}\{\mathrm{I}\left(U_{i,n} \le u, V_{i,n} \le v\right) - \mathrm{C}_n(u,v)\}$*, is used as the independent copies of the weak limit* $\mathbb{C}(u,v;\boldsymbol{\theta})$ *define in equation (G.3)*

Then, under the following three regular assumptions:

A1. *For all* $\boldsymbol{\theta} \in \mathcal{O}$*, the partial derivatives* $\mathrm{C}^{[1]}(u,v;\boldsymbol{\theta})$ *and* $\mathrm{C}^{[2]}(u,v;\boldsymbol{\theta})$ *are continuous;*

A2. *For all* $\boldsymbol{\theta} \in \mathcal{O}$*,* $\sqrt{n}\left(\mathrm{C}_n(u,v) - \mathrm{C}(u,v;\boldsymbol{\theta})\right)$ *and* $\sqrt{n}(\boldsymbol{\theta}_n - \boldsymbol{\theta})$ *jointly weakly converge to* $(\mathbb{C}(u,v;\boldsymbol{\theta}), \boldsymbol{\Theta})$ *in* $l^\infty\left([0,1]^2\right) \otimes \mathbb{R}^2$*;*

A3. *For all* $\boldsymbol{\theta} \in \mathcal{O}$ *and as* $\epsilon \downarrow 0$*,*

$$\sup_{|\boldsymbol{\theta}^* - \boldsymbol{\theta}| < \epsilon} \sup_{u,v \in [0,1]} |\dot{\mathrm{C}}(u,v;\boldsymbol{\theta}^*) - \dot{\mathrm{C}}(u,v;\boldsymbol{\theta})| \to 0,$$

where $\dot{\mathrm{C}}(u,v;\boldsymbol{\theta}) = \frac{\partial \mathrm{C}(u,v;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$*,*

the process

$$\left(\sqrt{n}\big(\mathrm{C}_n(u,v) - \mathrm{C}(u,v;\boldsymbol{\theta}_n)\big), \mathbb{C}_n^{(1)}(u,v) - \boldsymbol{\Theta}_n^{(1)}\dot{\mathrm{C}}(u,v;\boldsymbol{\theta}_n), \ldots, \mathbb{C}_n^{(N)}(u,v) - \boldsymbol{\Theta}_n^{(N)}\dot{\mathrm{C}}(u,v;\boldsymbol{\theta}_n)\right)$$

converges weakly to

$$\left(\mathbb{C}(u,v;\boldsymbol{\theta}) - \boldsymbol{\Theta}\dot{\mathrm{C}}(u,v;\boldsymbol{\theta}), \mathbb{C}^{(1)}(u,v;\boldsymbol{\theta}) - \boldsymbol{\Theta}^{(1)}\dot{\mathrm{C}}(u,v;\boldsymbol{\theta}), \ldots, \mathbb{C}^{(N)}\left(u,v;\boldsymbol{\theta}\right) - \boldsymbol{\Theta}^{(N)}\dot{\mathrm{C}}(u,v;\boldsymbol{\theta})\right) \tag{G.5}$$

in $l^\infty\big([0,1]^2\big)^{\otimes(N+1)}$*, where* $\boldsymbol{\Theta}$ *is the weak limit of* $\boldsymbol{\Theta}_n = \sqrt{n}(\boldsymbol{\theta}_n - \boldsymbol{\theta})$ *and* $\big(\mathbb{C}^{(1)}(u,v;\boldsymbol{\theta}), \boldsymbol{\Theta}^{(1)}\big)$*, ...,* $\big(\mathbb{C}^{(N)}(u,v;\boldsymbol{\theta}), \boldsymbol{\Theta}^{(N)}\big)$ *are independent copies of* $(\mathbb{C}(u,v;\boldsymbol{\theta}), \boldsymbol{\Theta})$

Proof. To ensure the asymptotic validity of the multiplier method used in the GOF test for the MG pseudo-copula, we need to demonstrate whether the MG pseudo-copula satisfies all the three regularity assumptions needed in the multiplier method.

The MG pseudo-copula is defined as

$$\mathrm{C}(u,v;\boldsymbol{\theta}) = \frac{1}{\boldsymbol{K}}\int_0^u\int_0^v \frac{1}{\sqrt{1-\rho^2(x,y;\boldsymbol{\theta})}} \times \exp\left\{\frac{[\Phi^{-1}(x)]^2+[\Phi^{-1}(y)]^2}{2}\right\} \times \exp\left\{-\frac{[\Phi^{-1}(x)]^2-2\rho(x,y;\boldsymbol{\theta})\Phi^{-1}(x)\Phi^{-1}(y)+[\Phi^{-1}(y)]^2}{2\left(1-\rho^2(x,y;\boldsymbol{\theta})\right)}\right\} dxdy,$$

Then $\mathrm{C}^{[1]}(u,v;\boldsymbol{\theta})$ can be expressed as

$$\begin{aligned}\mathrm{C}^{[1]}(u,v;\boldsymbol{\theta}) &= \frac{\partial \mathrm{C}(u,v;\boldsymbol{\theta})}{\partial u} \\ &= \frac{1}{\boldsymbol{K}}\int_0^v \frac{1}{\sqrt{1-\rho^2(u,y;\boldsymbol{\theta})}} \times \exp\left\{\frac{[\Phi^{-1}(u)]^2+[\Phi^{-1}(y)]^2}{2}\right\} \\ &\quad \times \exp\left\{-\frac{[\Phi^{-1}(u)]^2-2\rho(u,y;\boldsymbol{\theta})\Phi^{-1}(u)\Phi^{-1}(y)+[\Phi^{-1}(y)]^2}{2\left(1-\rho^2(u,y;\boldsymbol{\theta})\right)}\right\} dy.\end{aligned}$$

Visually, function $\mathrm{C}^{[1]}(u,v;\boldsymbol{\theta})$ is a function corresponding to variable u and is a continuous function with respect to variable u. In the same way, function $\mathrm{C}^{[2]}(u,v;\boldsymbol{\theta})$ is a continuous function with respect to variable v. Consequently, MG pseudo-copula satisfies Assumption 1.

According to the Theorem 1 in Kojadinovic et al. (2011), if $\mathrm{C}(u,v;\theta)$ has continuous partial derivatives $\mathrm{C}^{[1]}(u,v;\boldsymbol{\theta})$ and $\mathrm{C}^{[2]}(u,v;\boldsymbol{\theta})$, then the MG pseudo-copula satisfies Assumption 2.

The density function of MG pseudo-copula is $\mathrm{c}(u,v;\rho)$, namely,

$$\begin{aligned}c(u,v;\rho) &= \frac{g(u,v;\rho)}{\boldsymbol{K}} \\ &= \frac{1}{\boldsymbol{K}}\frac{1}{\sqrt{1-\rho^2}} \\ &\quad \times \exp\left\{\frac{[\Phi^{-1}(u)]^2+[\Phi^{-1}(v)]^2}{2} - \frac{[\Phi^{-1}(u)]^2-2\rho\Phi^{-1}(u)\Phi^{-1}(v)+[\Phi^{-1}(v)]^2}{2\left(1-\rho^2\right)}\right\} \quad \text{(G.6)}\end{aligned}$$

with $\rho = \rho(u,v;\boldsymbol{\theta})$. Since this expression is the standard bivariate distribution with respect to (w.r.t) variable ρ, there exists both the first and the second derivatives w.r.t ρ. First, let's define the first and second derivative of $g(\cdot)$ with respect to ρ. They are

$$\begin{aligned}\dot{g}(u,v;\rho) &= \frac{\partial g(u,v;\rho)}{\partial \rho} \\ &= g(u,v;\rho)\left\{\frac{\rho+\Phi^{-1}(u)\Phi^{-1}(v)}{1-\rho^2} - \frac{\rho\left([\Phi^{-1}(u)]^2-2\rho\Phi^{-1}(u)\Phi^{-1}(v)+[\Phi^{-1}(v)]^2\right)}{(1-\rho^2)^2}\right\}\end{aligned}$$

and

$$\begin{aligned}\ddot{g}(u,v;\rho) &= \frac{\partial \dot{g}(u,v;\rho)}{\partial \rho} \\ &= \dot{g}(u,v;\rho)\left\{\frac{\rho+\Phi^{-1}(u)\Phi^{-1}(v)}{1-\rho^2} - \frac{\rho\left([\Phi^{-1}(u)]^2 - 2\rho\Phi^{-1}(u)\Phi^{-1}(v) + [\Phi^{-1}(v)]^2\right)}{(1-\rho^2)^2}\right\} \\ &\quad + g(u,v;\rho)\left\{\frac{\rho}{1-\rho^2}\right. \\ &\quad + \frac{2\rho\left(\rho+\Phi^{-1}(u)\Phi^{-1}(v)\right) - \left([\Phi^{-1}(u)]^2 - 4\rho\Phi^{-1}(u)\Phi^{-1}(v) + [\Phi^{-1}(v)]^2\right)}{(1-\rho^2)^2} \\ &\quad \left. - \frac{4\rho^2\left([\Phi^{-1}(u)]^2 - 2\rho\Phi^{-1}(u)\Phi^{-1}(v) + [\Phi^{-1}(v)]^2\right)}{(1-\rho^2)^3}\right\}.\end{aligned}$$

Then the first and the second derivative of $\boldsymbol{K}$ with respect to ρ are

$$\dot{\boldsymbol{K}} = \int_0^1 \int_0^1 \dot{g}(x,y;\rho)dxdy$$

and

$$\ddot{\boldsymbol{K}} = \int_0^1 \int_0^1 \ddot{g}(x,y;\rho)dxdy.$$

Therefore, the first and the second derivative of $c(u,v;\rho)$ with respect to ρ are

$$\dot{c}(u,v;\rho) = \frac{\dot{g}(u,v;\rho)}{\boldsymbol{K}} - \frac{g(u,v;\rho)\dot{\boldsymbol{K}}}{\boldsymbol{K}^2}$$

and

$$\ddot{c}(u,v;\rho) = \frac{\ddot{g}(u,v;\rho)}{\boldsymbol{K}^2} - \frac{2\dot{g}(u,v;\rho)\dot{\boldsymbol{K}} + g(u,v;\rho)\ddot{\boldsymbol{K}}}{\boldsymbol{K}^2} + \frac{2g(u,v;\rho)\dot{\boldsymbol{K}}^2}{\boldsymbol{K}^3}.$$

According to the definitions for 5 types of ρ (see Section 5.3.1), both first and second derivative exist w.r.t variable $\boldsymbol{\theta} = (a,b)$. Table G.1 gives both the first and the second derivative functions for $\rho(u,v;\boldsymbol{\theta})$ w.r.t the parameter $\boldsymbol{\theta} = (a,b)$.

Referring to the composite function rule (also known as the chain rule) of differentiation, we can define the second derivative of c$(u,v;\rho)$ w.r.t parameter a and b. They are

$$\frac{\partial^2 g(u,v;a,b)}{\partial a^2} = \ddot{g}(u,v;\rho)\Big(\frac{\partial \rho(u,v;a,b)}{\partial a}\Big)^2 + \dot{g}(u,v;\rho)\frac{\partial^2 \rho(u,v;a,b)}{\partial \boldsymbol{a}^2}$$

and

$$\frac{\partial^2 g(u,v;a,b)}{\partial b^2} = \ddot{g}(u,v;\rho)\Big(\frac{\partial \rho(u,v;a,b)}{\partial b}\Big)^2 + \dot{g}(u,v;\rho)\frac{\partial^2 \rho(u,v;a,b)}{\partial \boldsymbol{b}^2}.$$

TABLE G.1: The first and the second derivative w.r.t parameters a and b for ρ types from **I** to **V**.

Types	$\frac{\partial\rho(u,v;a,b)}{\partial a}$	$\frac{\partial\rho(u,v;a,b)}{\partial b}$	$\frac{\partial^2\rho(u,v;a,b)}{\partial a^2}$	$\frac{\partial^2\rho(u,v;a,b)}{\partial b^2}$
I	$-buv$	$1-auv$	0	0
II	$-\frac{\pi(1-uv)}{2}b\sin\big(\frac{\pi}{2}a(1-uv)\big)$	$\cos\big(\frac{\pi}{2}a(1-uv)\big)$	$-\big(\frac{\pi(1-uv)}{2}\big)^2\rho$	0
III	$\frac{\pi(1-uv)}{2}b\cos\big(\frac{\pi}{2}a(1-uv)\big)$	$\sin\big(\frac{\pi}{2}a(1-uv)\big)$	$-\big(\frac{\pi(1-uv)}{2}\big)^2\rho$	0
IV	$\frac{\pi(1-uv)}{4}b\sec^2\big(\frac{\pi}{4}a(1-uv)\big)$	$\tan\big(\frac{\pi}{4}a(1-uv)\big)$	$-2\big(\frac{\pi(1-uv)}{4}\big)^2$ $\times\sec^2\big(\frac{\pi}{4}a(1-uv)\big)\rho$	0
V	$(1-uv)\rho$	$\exp\big(-a(1-uv)\big)$	$(1-uv)^2\rho$	0

The first derivative of $\mathrm{C}(u,v;a,b)$ w.r.t parameter a is

$$\begin{aligned}\frac{\partial \mathrm{C}(u,v;a,b)}{\partial a} &= \frac{\partial}{\partial a}\Big\{\frac{1}{\boldsymbol{K}}\int_0^u\int_0^v g(x,y;a,b)dxdy\Big\}\\ &= \frac{1}{\boldsymbol{K}}\int_0^u\int_0^v \frac{\partial g(x,y;a,b)}{\partial a}dxdy - \frac{\mathrm{C}(u,v;a,b)}{\boldsymbol{K}}\frac{\partial \boldsymbol{K}}{\partial a}\\ &= \frac{1}{\boldsymbol{K}}\int_0^u\int_0^v \dot{g}(x,y;\rho)\frac{\partial\rho(x,y;a,b)}{\partial a}dxdy - \frac{\mathrm{C}(u,v;a,b)}{\boldsymbol{K}}\int_0^1\int_0^1 \dot{g}(x,y;\rho)\frac{\partial\rho(x,y;a,b)}{\partial a}dxdy\end{aligned}$$

and the second derivative of $\mathrm{C}(u,v;a,b)$ w.r.t parameter a is

$$
\begin{aligned}
\frac{\partial^2 \mathrm{C}(u,v;a,b)}{\partial a^2} &= \frac{\partial}{\partial a}\Big\{\frac{\partial \mathrm{C}(u,v;a,b)}{\partial a}\Big\} \\
&= \frac{\partial}{\partial a}\Big\{\frac{1}{\boldsymbol{K}}\int_0^u\int_0^v \frac{\partial g(x,y;a,b)}{\partial a}dxdy - \frac{\mathrm{C}(u,v;a,b)}{\boldsymbol{K}}\frac{\partial \boldsymbol{K}}{\partial a}\Big\} \\
&= \frac{1}{\boldsymbol{K}}\int_0^u\int_0^v \Big\{\frac{\partial^2 g(x,y;a,b)}{\partial a^2}\Big\}dxdy - \frac{1}{\boldsymbol{K}^2}\frac{\partial \boldsymbol{K}}{\partial a}\int_0^u\int_0^v \Big\{\frac{\partial g(x,y;a,b)}{\partial a}dxdy\Big\} \\
&\quad + \frac{\mathrm{C}(u,v;a,b)}{\boldsymbol{K}^2}\Big\{\frac{\partial \boldsymbol{K}}{\partial a}\Big\}^2 - \frac{1}{\boldsymbol{K}}\frac{\partial \mathrm{C}(u,v;a,b)}{\partial a}\frac{\partial \boldsymbol{K}}{\partial a} - \frac{\mathrm{C}(u,v;a,b)}{\boldsymbol{K}}\frac{\partial^2 \boldsymbol{K}}{\partial a^2} \\
&= \frac{1}{\boldsymbol{K}}\int_0^u\int_0^v \Big\{\ddot{g}(x,y;\rho)\Big(\frac{\partial \rho(x,y;a,b)}{\partial a}\Big)^2 + \dot{g}(x,y;\rho)\frac{\partial^2 \rho(x,y;a,b)}{\partial \boldsymbol{a}^2}\Big\}dxdy \\
&\quad - \frac{1}{\boldsymbol{K}^2}\int_0^1\int_0^1 \dot{g}(x,y;\rho)\frac{\partial \rho(x,y;a,b)}{\partial a}dxdy \int_0^u\int_0^v \dot{g}(x,y;\rho)\frac{\partial \rho(x,y;a,b)}{\partial a}dxdy \\
&\quad + \frac{\mathrm{C}(u,v;a,b)}{\boldsymbol{K}^2}\Big\{\int_0^1\int_0^1 \dot{g}(x,y;\rho)\frac{\partial \rho(x,y;a,b)}{\partial a}dxdy\Big\}^2 \\
&\quad - \frac{1}{\boldsymbol{K}^2}\int_0^u\int_0^v \dot{g}(x,y;\rho)\frac{\partial \rho(x,y;a,b)}{\partial a}dxdy \int_0^1\int_0^1 \dot{g}(x,y;\rho)\frac{\partial \rho(x,y;a,b)}{\partial a}dxdy \\
&\quad + \frac{\mathrm{C}(u,v;a,b)}{\boldsymbol{K}^2}\Big\{\int_0^1\int_0^1 \dot{g}(x,y;\rho)\frac{\partial \rho(x,y;a,b)}{\partial a}dxdy\Big\}^2 \\
&\quad - \frac{\mathrm{C}(u,v;a,b)}{\boldsymbol{K}}\int_0^1\int_0^1 \Big\{\ddot{g}(x,y;\rho)\Big(\frac{\partial \rho(x,y;a,b)}{\partial a}\Big)^2 + \dot{g}(x,y;\rho)\frac{\partial^2 \rho(x,y;a,b)}{\partial \boldsymbol{a}^2}\Big\}dxdy.
\end{aligned}
$$

Since both $g(u,v;a,b)$ and $\frac{\partial g(u,v;a,b)}{\partial a}$ are continuous, by using Leibniz integral rule (see Flanders, 1973) we obtain that $\frac{\partial}{\partial a}\Big\{\int_0^u\int_0^v g(x,y;a,b)dxdy\Big\} = \int_0^u\int_0^v \frac{\partial g(x,y;a,b)}{\partial a}dxdy$. Analogously, since both $\frac{\partial g(u,v;a,b)}{\partial a}$ and $\frac{\partial^2 g(u,v;a,b)}{\partial a^2}$ are continuous, $\frac{\partial}{\partial a}\Big\{\int_0^u\int_0^v \frac{\partial g(x,y;a,b)}{\partial a}\Big\} = \int_0^u\int_0^v \frac{\partial^2 g(x,y;a,b)}{\partial a^2}dxdy$.

Similarly, we obtain

$$
\frac{\partial^2 \mathrm{C}(u,v;a,b)}{\partial b^2} = \frac{1}{\boldsymbol{K}}\int_0^u\int_0^v \Big\{\ddot{g}(x,y;\rho)\Big(\frac{\partial \rho(x,y;a,b)}{\partial b}\Big)^2 + \dot{g}(x,y;\rho)\frac{\partial^2 \rho(x,y;a,b)}{\partial \boldsymbol{b}^2}\Big\}dxdy.
$$

Therefore, the second derivative $\ddot{\mathrm{C}}(u,v;\boldsymbol{\theta}) = (\frac{\partial^2 \mathrm{C}(u,v;a,b)}{\partial a^2}, \frac{\partial^2 \mathrm{C}(u,v;a,b)}{\partial b^2})^T$ exists. Since there exists the derivative for function $\dot{\mathrm{C}}(u,v;\boldsymbol{\theta})$, i.e., $\ddot{\mathrm{C}}(u,v;\boldsymbol{\theta})$, by using the Mean Value Theorem, one can get that $\dot{\mathrm{C}}(u,v;\boldsymbol{\theta}^*) - \dot{\mathrm{C}}(u,v;\boldsymbol{\theta}) = (\boldsymbol{\theta}^* - \boldsymbol{\theta}) \times \ddot{\mathrm{C}}(u,v;\boldsymbol{\theta}_0)$ and $\boldsymbol{\theta}_0 \in (\boldsymbol{\theta}^*, \boldsymbol{\theta})$. And because $\boldsymbol{\theta}_0 \in \mathcal{O}$

and $\mathcal{O}$ is an open set, there is $\sup_{u,v\in[0,1]} \ddot{\mathrm{C}}(u,v;\boldsymbol{\theta}_0)$ for a particular $\boldsymbol{\theta}_0$. Let W denote this value, i.e., $W = \sup_{u,v\in[0,1]} |\dot{\mathrm{C}}(u,v;\boldsymbol{\theta})$. And then we will conclude

$$
\begin{aligned}
&\sup_{|\boldsymbol{\theta}^*-\boldsymbol{\theta}|<\epsilon} \sup_{u,v\in[0,1]} |\dot{\mathrm{C}}(u,v;\boldsymbol{\theta}^*) - \dot{\mathrm{C}}(u,v;\boldsymbol{\theta})| \\
&= \sup_{|\boldsymbol{\theta}^*-\boldsymbol{\theta}|<\epsilon} \sup_{u,v\in[0,1]} |(\boldsymbol{\theta}^*-\boldsymbol{\theta}) \times \ddot{\mathrm{C}}(u,v;\boldsymbol{\theta}_0)| \\
&= \sup_{|\boldsymbol{\theta}^*-\boldsymbol{\theta}|<\epsilon} |\boldsymbol{\theta}^*-\boldsymbol{\theta}| \times \sup_{u,v\in[0,1]} |\ddot{\mathrm{C}}(u,v;\boldsymbol{\theta}_0)| \\
&= |W| \sup_{|\boldsymbol{\theta}^*-\boldsymbol{\theta}|<\epsilon} |\boldsymbol{\theta}^*-\boldsymbol{\theta}| \\
&\leq \epsilon \times |W| \\
&\to 0 \quad as \quad \epsilon \to 0.
\end{aligned}
$$

Hence,

$$\sup_{|\boldsymbol{\theta}^*-\boldsymbol{\theta}|<\epsilon} \sup_{u,v\in[0,1]} |\dot{\mathrm{C}}(u,v;\boldsymbol{\theta}^*) - \dot{\mathrm{C}}(u,v;\boldsymbol{\theta})| \to 0.$$

Therefore, the MG pseudo-copula meets Assumption 3.

In sum, MG pseudo-copula satisfies all three regularity assumptions required in Theorem 2 of Kojadinovic et al. (2011). Thus, we can use the Multiplier method to do GOF test for MG pseudo-copula. □

H APPENDIX The procedure for Multiplier method

The multiplier method for goodness-of-fit test proposed Kojadinovic et al. (2011) is done by the following procedure:

1. Compute $\mathrm{C}_n(u,v)$ from the pseudo-observations $(\hat{U}_{1,n},\hat{V}_{1,n}),\ldots,(\hat{U}_{n,n},\hat{V}_{n,n})$ and estimate the dependence parameter $\boldsymbol{\theta}$ by using $\boldsymbol{\theta}_n$ defined in equation (5.8);

2. Compute the test statistic defined in equation (5.12), viz,

   $\mathrm{S}_n = \int_{[0,1]^2} n\{\mathrm{C}_n(u,v) - \mathrm{C}(u,v;\boldsymbol{\theta}_n)\}^2 d\mathrm{C}_n(u,v) = \sum_{i=1}^{n}\{\mathrm{C}_n(\hat{U}_{i,n},\hat{V}_{i,n}) - \mathrm{C}(\hat{U}_{i,n},\hat{V}_{i,n};\boldsymbol{\theta}_n)\}^2$;

3. Repeat the following steps for every $k \in \{1,\ldots,N\}$:

   (a) Generate n i.i.d random variables $Z_i^{(k)} \sim N(0,1)$;

   (b) Form an approximate realization of the test statistic under H_0 by

   $$\begin{aligned}\mathrm{S}_n^{(k)} &= \int_{[0,1]^d}\left\{\mathbb{C}_n^{(k)}(\mathbf{u}) - \dot{\mathrm{C}}_{\theta_n}^T(\mathbf{u})\hat{\Theta}_n^{(k)}\right\}^2 d\mathrm{C}_n(\mathbf{u}) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left\{\mathbb{C}_n^{(k)}(\hat{U}_{i,n},\hat{V}_{i,n}) - \dot{\mathrm{C}}^T(\hat{U}_{i,n},\hat{V}_{i,n};\boldsymbol{\theta}_n)\boldsymbol{\Theta}_n^{(k)}\right\}^2,\end{aligned}$$

   where the definitions for terms $\mathbb{C}_n^{(k)}(\hat{U}_{i,n},\hat{V}_{i,n})$, $\dot{\mathrm{C}}^T(\hat{U}_{i,n},\hat{V}_{i,n};\boldsymbol{\theta}_n)$ and $\boldsymbol{\Theta}_n^{(k)}$ are defined in Appendix G;

4. An approximate p-value : $\frac{1}{N}\sum_{k=1}^{N}\mathbf{1}(\mathrm{S}_n^{(k)} \geq \mathrm{S}_n)$.