Analysis of directional dependence using asymmetric copula-based regression models

Daeyoung Kim & Jong-Min Kim

Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA, 01003, USA
Statistic Discipline, Division of Science and Mathematics, University of Minnesota-Morris, Morris, MN, 56267, USA


To cite this article: Daeyoung Kim & Jong-Min Kim (2013): Analysis of directional dependence using asymmetric copula-based regression models, Journal of Statistical Computation and Simulation, DOI:10.1080/00949655.2013.779696

To link to this article: http://dx.doi.org/10.1080/00949655.2013.779696

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Analysis of directional dependence using asymmetric copula-based regression models

Daeyeong Kim\textsuperscript{a} and Jong-Min Kim\textsuperscript{b}\textsuperscript{*}

\textsuperscript{a}Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA 01003, USA; \textsuperscript{b}Statistic Discipline, Division of Science and Mathematics, University of Minnesota-Morris, Morris, MN 56267, USA

(Received 8 January 2013; final version received 21 February 2013)

The directional dependence between variables using asymmetric copula regression has drawn much attention in recent years. There are, however, some critical issues which have not been properly addressed in regards to the statistical inference of the directional dependence. For example, the previous use of asymmetric copulas failed to fully capture the dependence patterns between variables, and the method used for the parameter estimation was not optimal. In addition, no method was considered for selecting a suitable asymmetric copula or for computing the general measurements of the directional dependence when there are no closed-form expressions. In this paper, we propose a generalized multiple-step procedure for the full inference of the directional dependence in joint behaviour based on the asymmetric copula regression. The proposed procedure utilizes several novel methodologies that have not been considered in the literature of the analysis of directional dependence. The performance and advantages of the proposed procedure are illustrated using two real data examples, one from biological research on histone genes, and the other from developmental research on attention deficit hyperactivity disorder.

Keywords: asymmetric copula; directional dependence; regression function

1. Introduction

Copula has been increasingly popular as a flexible tool for modelling the dependence of multivariate data in many fields of application such as biostatistics, medical research, econometrics, finance, actuarial science, hydrology, etc. The copula arises from Sklar’s theorem,\cite{1} which proves that for any $q$ continuous random variables, there exists a unique function, copula, that couples $q$ univariate marginal distributions into the $q$-dimensional distribution function. The main advantages of using the copula method are as follows: (1) one can model the dependence structure of the joint distribution and its marginal distributions separately and (2) when the variables are transformed by increasing transformations such as rank, the copula describing these transformed variables is the same as the copula for the original variables. Standard references for a detailed overview of copula theory and applications include the books by Joe \cite{2} and Nelsen \cite{3}.

Due to these advantages of the copula, the directional dependence between the variables using copulas has recently drawn some attention in the literature.\cite{4–9} By directional dependence we mean the likely direction of influence between two variables, such as recursive influence (the
influence is from one variable to the other) or non-recursive influence (two variables influence each other). The determination of directional dependence is an interesting but difficult research issue in observational studies when the directional dependence is known to theoretically exist and empirical justification is required.

Directional dependence has been studied from different viewpoints. Dodge and Rousson [10] and Muddapur [11] investigated directional dependence to determine the direction of a regression line. They showed that, assuming that the error term in a linear regression model is symmetric, the skewness of the two variables and the Pearson’s correlation can be used to determine which variable is an independent one and which is a dependent one through a regression line. While the methods proposed by Dodge and Rousson [10] is simple to use, one needs to check that the underpinning assumptions (a regression function is linear and the error term is symmetric) are valid. Furthermore, their methods study directional dependence stemming from only the marginal behaviour of the variables, not the joint behaviour, and only work for the skewed data.[4]

Sungur [4,5] argued that the best way of understanding the directional dependence between variables is to study the dependence structure through the copula regression model. This is because the copula regression approach can model the joint dependence between the variables, independently from the choice of the marginal distributions, by using the normalized ranks of data. Note that the normalized rank of data is a monotone increasing transformation of the marginals through which the joint dependence is invariant. This is particularly important for the data analysis because it is common that the marginal distributions are not known in practice. Sungur [4,5] defined directional dependence stemming from not only marginal behaviour of variables, but also joint behaviour of them using asymmetric copulas under the copula regression perspective and studied their properties. They also considered the general measurements of the directional dependence between the two variables. Thus, one can not only detect the existence of the directional dependence between the variables but also quantify the degree of the directional dependence.

There have been several papers that have tried to apply the theories developed by Sungur [4,5] to real data sets. Jung et al. [6] and Uhm et al. [9] used an asymmetric type of Farlie–Gumbel–Morgenstern (FGM) copula in the form of the Rodríguez-Lallena and Úbeda-Flores family of copulas to study asymmetry of financial data. Kim et al. [7,8] employed Rodríguez-Lallena and Úbeda-Flores copula family and a survival truncated FGM modification copula to examine directional dependence of gene data.

However, we have found in the papers by Jung et al. [6], Kim et al. [7,8], and Uhm et al. [9] several critical issues which have not been properly addressed. First, an asymmetric type of FGM copula used in those papers is of limited usage in real applications because it is only appropriate when the dependence between variables is modest in magnitude. Second, those papers did not use any of the well-known estimation methods in the literature for parameter estimation. Instead, they used a profile approach that maximizes the likelihood function over the dependence parameter of interest given the fixed values of the other parameters. Third, they mainly focused the estimation of the amount of the directional dependence without empirically verifying the existence of the directional dependence in the data.

The goals of this paper are two-fold: (1) to address several critical problems found in Jung et al. [6], Kim et al. [7,8], and Uhm et al. [9], and (2) to propose a generalized multiple-step procedure for the full inference of the directional dependence in joint behaviour based on the asymmetric copula regression. The proposed procedure utilizes several novel methodologies that have not been considered in the literature of the analysis of directional dependence. We first propose using various asymmetric copulas designed to have the flexibility of capturing various dependence structures. We also illustrate how to use a semiparametric estimation method that does not require knowledge of the marginal distributions of the variables. Furthermore, we show how to use a Monte Carlo method to compute the measurements for identifying and quantifying the directional dependence between the variables when their closed forms are not available.
This article is organized as follows. In Section 2, we briefly review the concept of the copula, its properties, and the directional dependence based on copula regression function in Sungur.[4,5] Section 3, with its four subsections, describes the proposed generalized multiple-step procedure using asymmetric copula regression for the analysis of directional dependence. To illustrate the proposed procedure, Section 4 provides two empirical data examples. We then end this article with a discussion in Section 5.

2. Directional dependence in the copula regression

2.1. Copulas and Sklar's theorem

We here briefly review the copula and its properties. For a detailed overview of copula theory, please see Joe [2] and Nelsen [3].

Let $X$ and $Y$ be two continuous random variables defined on the same probability space such that $F_X$ and $F_Y$ are the univariate distribution functions of $X$ and $Y$, respectively, and $F_{X,Y}$ is the joint distribution function of $(X, Y)$. One can view $F_X$ and $F_Y$ as transformed variables, denoted by $U$ and $V$, respectively, which transform the values of $X$ and $Y$ into values in the real unit interval $[0, 1]$: $u = F_X(x)$ and $v = F_Y(y)$, where $x, y, u,$ and $v$ are the values of the corresponding random variables $X, Y, U,$ and $V$.

A copula is the distribution function expressing the dependence of a random vector $(X, Y)$. Specifically, a bivariate copula is a function $C : [0, 1]^2 \to [0, 1]$ satisfying the following properties:

1. $C(u, 0) = C(0, v) = 0$, $C(u, 1) = u$ and $C(1, v) = v$ for all $u, v \in [0, 1]$.
2. $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$ for all $u_1, u_2, v_1, v_2 \in [0, 1]$ for which $u_1 \leq u_2$ and $v_1 \leq v_2$.

Due to Sklar’s theorem,[1] for any continuous random variables $X$ and $Y$, there exists a unique bivariate copula $C : [0, 1]^2 \to [0, 1]$ such that

$$F_{XY}(x, y) = C(F_X(x), F_Y(y)) = C(u, v).$$

This theorem implies three important results. First, any bivariate distribution function of $X$ and $Y$, $F_{X,Y}$, can be represented as a function of its marginals, $F_X$ and $F_Y$, through a copula $C$. Second, any monotone strictly increasing transformation of $X$ and $Y$ (such as $U = F_X(X)$ and $V = F_Y(Y)$) will leave the joint dependence between them unchanged, and thus the copula associated with a pair $(X, Y)$ will be the same as the copula with transformed variables $(U, V)$. Third, we can describe the dependence structure between $X$ and $Y$ by considering only the pairs of ranks of $X$ and $Y$. In other words, given $n$ pairs of data points $\{(X_i, Y_i)\}$ from a pair of continuous random variables $(X, Y)$, we can approximate the corresponding couple $(U_i, V_i) = (F_X(X_i), F_Y(Y_i))$ using the pairs of ranks $\{(R_i, S_i)\}$; $i = 1, \ldots, n$ where $R_i$ is the rank of $X_i$ among $X_1, \ldots, X_n$ and $S_i$ is the rank of $Y_i$ among $Y_1, \ldots, Y_n$:

$$u_i = \frac{R_i}{n+1} \quad \text{and} \quad v_i = \frac{S_i}{n+1}.$$  \hfill (1)

Using these normalized ranks, one can construct the empirical distribution functions for $X$ and $Y$. Since the marginal distributions of $U$ and $V$ are uniform on $[0, 1]$ and thus parameter-free, this rank-based approach allows us to compute joint probabilities without knowing marginal distributions.
We close this subsection by defining one important concept playing a particular role in the
directional dependence that will be introduced in the following section. Let $U$ and $V$ be uniform
random variables in $[0, 1]$ and $C$ is their copula. Then $C$ is symmetric if and only if $C(u, v) =
C(v, u)$ for every $u$ and $v$ in $[0, 1]$.

2.2. Concept and measurement of directional dependence

Sungur [4,5] defined two types of directional dependence using the copula regression setting and
considered the general measurements of the directional dependence.

Let $(U, V)$ be a random pair with uniform marginals on the $[0, 1]$ and their copula $C$. Then the
conditional distribution function for $V$ given $U = u$, denoted by, $C_u(v)$, is

$$C_u(v) \equiv P(V \leq v \mid U = u) = \frac{\partial C(u, v)}{\partial u} \tag{2}$$

and the copula regression function of $V$ on $U$, denoted by $r_{V|U}^C(u)$, is

$$r_{V|U}^C(u) \equiv E_C(V \mid U = u) = 1 - \int_0^1 C_u(v) \, dv. \tag{3}$$

One can define the conditional distribution function for $U$ given $V = v$ and the copula regression
function of $U$ on $V$ in a similar fashion:

$$C_v(u) \equiv P(U \leq u \mid V = v) = \frac{\partial C(u, v)}{\partial v}, \tag{4}$$

$$r_{U|V}^C(v) \equiv E_C(U \mid V = v) = 1 - \int_0^1 C_v(u) \, du. \tag{5}$$

Note that if $U$ and $V$ are independent, $C(u, v) = uv$ and thus $r_{V|U}^C(u)$ in Equation (3) and $r_{U|V}^C(v)$
in Equation (5) are equal to 0.5. The functional forms of the copula regressions in Equations (3)
and (5) depend on the choice of the copula function. For example, for the FGM copula (i.e.
$C(u, v) = uv + \theta uv(1 - u)(1 - v)$ for each $\theta \in [−1, 1]$), the copula regression function is linear.

Sungur [4] proposed two types of directional dependence, ‘direction of dependence’ and ‘directional
dependence’. The first stems from the marginal behaviour of the variables and the latter
from the joint behaviour of them, i.e. copula.

DEFINITION 1 [4]

1. The random pair $(U, V)$ is directionally dependent in joint behaviour if $r_{V|U}^C(w)$ in Equation (3)
and $r_{U|V}^C(w)$ in Equation (5) differ.

2. The random pair $(X, Y)$ is directionally dependent in marginals if $r_{V|U}^C(w) = r_{U|V}^C(w) = r^C(w)$
and $r_{X|Y}^C(z) \neq r_{Y|X}^C(z)$, where $r_{X|Y}^C(x) = E_C[Y \mid X = x] = F_Y^{-1}(r^C(F_X(x)))$ and $r_{Y|X}^C(y) =
E_C[X \mid Y = y] = F_X^{-1}(r^C(F_Y(y)))$.

There are a few important observations from Definition 1.[4,5] First, if there is no directional
dependence in joint behaviour and the marginals for $X$ and $Y$ are the same, then it is not possible
to identify directional dependence in marginals (i.e. no direction of dependence). Second, it is
well-known that copula is a way of studying scale-free measures of dependence. For copula
analysis, one needs to consider the transformed pair, $(U, V)=(F_X(X), F_Y(Y))$, that are uniformly
distributed in $[0, 1]$. Then, the marginals $U$ and $V$ are the same, and thus the difference in the copula
regression functions (one from a given value of $U$ to $V$ and the other from a given value of $V$ to $U$)
results from the joint behaviour of the variables, not of the marginal behaviour. Third, if both of the copula regression functions \( r^C_{U|V}(w) \) and \( r^C_{V|U}(w) \) are linear, then the random pair \((U, V)\) cannot be directionally dependent in joint behaviour. Fourth, for the class of symmetric copulas satisfying \( C(u, v) = C(v, u) \) for every \((u, v) \in [0, 1]^2\), we can investigate only the directional dependence in marginals, not the directional dependence in joint behaviour. If one is interested in a directional dependence in joint behaviour of two variables, one should start with an ‘asymmetric’ copula that can offer a better modelling of the joint dependence structure. The asymmetric copula functions will be discussed in detail in Section 3.1.

Note that Sungur [4,5] provided a few examples where two random variables are directionally dependent in marginals but not in joint behaviour, and vice versa. In this paper, our main interest is the directional dependence in \textit{joint behaviour}, rather than one in marginals, because the marginal distributions are typically unknown in applications and thus the normalized ranks of the observed data defined in Equation (1) will be used for the analysis of directional dependence.

Sungur [4] also proposed the general measures for the directional dependence in joint behaviour:

\[
\rho^2_{U \to V} = \frac{\text{Var}(r^C_{U|V}(U))}{\text{Var}(V)} = \frac{E[(r^C_{U|V}(U) - 1/2)^2]}{1/12} = 12E[(r^C_{U|V}(u))^2] - 3, \quad (6)
\]

\[
\rho^2_{V \to U} = \frac{\text{Var}(r^C_{V|U}(V))}{\text{Var}(U)} = \frac{E[(r^C_{V|U}(V) - 1/2)^2]}{1/12} = 12E[(r^C_{V|U}(v))^2] - 3. \quad (7)
\]

Note that \( \rho^2_{U \to V} \) can be interpreted as the proportion of total variation of \( V \) that can be explained by the copula regression of \( V \) on \( U \). One can view \( \text{Var}(r^C_{U|V}(U)) \) as the expected square distance of the copula regression function from independence \( (r^C_{U|V}(u) = r^C_{V|U}(v) = 0.5 \) for independence case with \( C(u, v) = uv \) \) in the direction of \( U \) to \( V \). Furthermore, the comparison between \( \text{Var}(r^C_{U|V}(V)) \) and \( \text{Var}(r^C_{V|U}(U)) \) tells us which copula regression has higher predictive power.

### 3. Multiple-step procedure for the inference of the directional dependence

In this section, we propose a generalized multiple-step procedure for the inference of the directional dependence in joint behaviour based on the asymmetric copula regression that has not been used in the area of the analysis of directional dependence. A summary of the proposed procedure is given at the end of this section.

#### 3.1. Asymmetric copula

As we explained in Section 2.2, we need to consider a family of asymmetric copulas indexed by the parameters as a statistical model for the directional dependence in joint behaviour.

Kim et al. [7], Jung et al. [6], Kim et al. [8], and Uhm et al. [9] used a type of generalized FGM (GFGM) copula, a specific form of the Rodríguez-Lallena and Úbeda-Flores copula family, [12] to study directional dependence of gene data and financial data:

\[
C(u, v; \phi) = uv + f(u)g(v) = uv + \theta uv(1-u)^{\alpha}(1-v)^{\beta},
\]

where \( u, v \in [0, 1], \phi = (\theta, \alpha, \beta), \alpha > 1, \beta > 1, f(u) = \sqrt{\theta}u(1-u)^{\alpha}, \text{ and } g(v) = \sqrt{\theta}v(1-v)^{\beta}. \) Here the parameter \( \theta \), measuring the dependence between \( u \) and \( v \), satisfies \(-1 \leq \theta \leq \min\{(1+\alpha)/(\alpha-1), (1+\beta)/(\beta-1), (\alpha-1)/(\alpha+1), (\beta-1)/(\beta+1)\}\). The parameters \( \alpha \) and \( \beta \) are the asymmetry parameters. However, as will be shown in Section 4, the GFGM copula in Equation (8) is of limited usage in real applications because...
the range of $\theta$ is restricted by asymmetry parameters and this copula can only model relatively weak dependence.

In this paper, we propose using a large family of asymmetric copulas which is more flexible than the GFGM copula of Equation (8) in capturing various types of dependence structures in the data. Liebscher [13] and Durante [14] introduced a general formula for new families of copulas whose members are asymmetric by multiplying two symmetric copulas:

$$C(u,v; \phi) = C_1(u^\alpha, v^\beta) C_2(u^{\tilde{\alpha}}, v^{\tilde{\beta}}),$$  \hspace{1cm} (9)

where $\alpha, \beta \in (0, 1)$, $\alpha \neq 1/2$ or $\beta \neq 1/2$, $\alpha + \tilde{\alpha} = 1$, $\beta + \tilde{\beta} = 1$, and $C_1$ and $C_2$ are two symmetric copula families of choice. Here, $\phi$ represents the set of parameters including the asymmetry parameters, $\alpha$ and $\beta$, and the dependence parameters in either $C_1$ or $C_2$. For notational convenience, we suppress the dependence parameters in $C_1$ and $C_2$. If $\alpha$ and $\beta$ approach either 0 or 1 together, the symmetric copula is supported by the data, and there is no directional dependence in joint behaviour. Note that Liebscher [13] and Durante [14] studied the theoretical properties of the families of asymmetric copulas $C(u,v; \phi)$ in Equation (9) in terms of Kendall’s $\tau$ and tail dependence, but did not apply them in application problems.

Table 1 presents the asymmetric copulas that will be considered in the data analysis of this paper. The asymmetric copulas, (M1, M2, M3, and M4), are members of the Khoudraji family [15] that uses the independence copula for $C_1$ and the Archimedean copula for $C_2$. The copula model, M6, is an asymmetric extreme value copula where $C_1$ and $C_2$ are members of the so-called Gumbel–Hougaard family of copulas with the same dependence parameter. Note that the copula models, (M1, M2, M3, M4, M5, and M6), have one dependence parameter, $\theta$, and the copula models (M7, M8, and M9) have two dependence parameters, $\theta_1$ and $\theta_2$, one for each $C_1$ and $C_2$, respectively.

Where the independence copula is $C(u,v) = uv$, the Clayton copula is

$$C(u,v; \theta) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta} \text{ with } \theta \in (0, \infty),$$

the Gumbel copula is

$$C(u,v; \theta) = \exp\{-[(-\log(u))^{\theta} + (-\log(v))^{\theta}]^{1/\theta}\} \text{ with } \theta \in [1, \infty),$$

the Frank copula is

$$C(u,v; \theta) = -\frac{1}{\theta} \log\left\{1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}\right\} \text{ with } \theta \in \mathbb{R}\{0\},$$

**Table 1.** Choice of symmetric copulas for $C_1$ and $C_2$ in asymmetric copula of Equation (9). Note that $\theta, \theta_1,$ and $\theta_2$ represent the dependence parameters.

<table>
<thead>
<tr>
<th>$C(u,v; \phi)$</th>
<th>$C_1(u^\alpha, v^\beta)$</th>
<th>$C_2(u^{\tilde{\alpha}}, v^{\tilde{\beta}})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1 Independence</td>
<td>Clayton ($\theta$)</td>
<td>Gumbel ($\theta$)</td>
</tr>
<tr>
<td>M2 Independence</td>
<td>Gumbel ($\theta$)</td>
<td>Frank ($\theta$)</td>
</tr>
<tr>
<td>M3 Independence</td>
<td>Frank ($\theta_1$)</td>
<td>Clayton ($\theta_2$)</td>
</tr>
<tr>
<td>M4 Independence</td>
<td>AMH (Ali–Mikhail–Haq) ($\theta$)</td>
<td>Plackett ($\theta$)</td>
</tr>
<tr>
<td>M5 Independence</td>
<td>Plackett ($\theta$)</td>
<td>Gumbel ($\theta$)</td>
</tr>
<tr>
<td>M6 Independence</td>
<td>Gumbel ($\theta$)</td>
<td>Gumbel ($\theta$)</td>
</tr>
<tr>
<td>M7 Clayton ($\theta_1$)</td>
<td>Clayton ($\theta_2$)</td>
<td>Clayton ($\theta_2$)</td>
</tr>
<tr>
<td>M8 Gumbel ($\theta_1$)</td>
<td>Gumbel ($\theta_2$)</td>
<td>Gumbel ($\theta_2$)</td>
</tr>
<tr>
<td>M9 Frank ($\theta_1$)</td>
<td>Frank ($\theta_2$)</td>
<td>Gumbel ($\theta_2$)</td>
</tr>
</tbody>
</table>
the AMH copula is
\[
C(u, v; \theta) = \frac{uv}{1 - \theta(1 - u)(1 - v)} \quad \text{with } \theta \in [-1, 1],
\]
and the Plackett copula is
\[
C(u, v; \theta) = \frac{[1 + (\theta - 1)(u + v)] - \sqrt{[1 + (\theta - 1)(u + v)]^2 - 4\theta(\theta - 1)uv}}{2(\theta - 1)}
\]
with \(\theta > 0, \theta \neq 1\).

### 3.2. Parameter estimation

In this section, we discuss the estimation of the parameters for the asymmetric copula in Table 1 given a sample \((x_i, y_i)\) in a data set with \(n\) data points \(\{(x_i, y_i), i = 1, \ldots, n\}\).

For the estimation for the copula function, the following three methods are commonly used: the maximum likelihood (ML) method, the inference functions for margins (IFM) method,[2] and the maximum pseudo-likelihood (MPL) method.[16–18] While the ML and IFM methods are the well-known parametric methods due to their optimality properties and easy implementation (for the IFM method), they have the risk of being sensitive to the choice of the models selected for the marginals. For example, Kim et al. [19] indicates that the ML and IFM methods are nonrobust against misspecification of the marginals. On the other hand, the MPL method, which uses the normalized ranks in Equation (1) and thus provides the rank-based estimators, is not affected by the choice of the marginal distributions.

However, Jung et al. [6], Kim et al. [7,8], and Uhm et al. [9] used none of the methods described above in the analysis of the directional dependence. Instead, they used a profile approach that maximizes the likelihood function over the dependence parameter given fixed values of the asymmetry parameters.

In this paper, we propose using the MPL method for the inference of directional dependence in joint behaviour based on asymmetric copula regression because in practice it is common that the marginal distributions of the variables are unknown. We first compute the ranks \(R_i\) and \(S_i\) of \(x_i\) and \(y_i\) as shown in Equation (1): \(u_i = R_i/(n + 1)\) and \(v_i = S_i/(n + 1)\). Given \(n\) paired points \(\{(u_i, v_i), i = 1, \ldots, n\}\) and an asymmetric copula \(C(u, v; \phi)\), a pseudo log-likelihood function is
\[
\ell(\phi) = \log \prod_{i=1}^{n} c(u_i, v_i; \phi) = \sum_{i=1}^{n} \log c(u_i, v_i; \phi),
\]
where \(\phi = (\theta, \alpha, \beta)\) or \((\theta_1, \theta_2, \alpha, \beta)\) and \(c(u, v; \phi)\) is a copula density given by
\[
c_{\alpha, \beta}(u, v; \theta) = \frac{\partial^2 C(u, v; \phi)}{\partial u \partial v}.
\]
Note that the asymmetric density functions \(c(u, v; \phi)\) for asymmetric copulas in Table 1 are given in Appendix. Then, we obtain the maximum pseudo-likelihood estimators (MPLE) for the parameters by maximizing the pseudo log-likelihood of Equation (10) with respect to \(\phi\).

The pseudo log-likelihood in Equation (10) is nonlinear in the parameters, and the parameters are usually constrained (e.g. \(\alpha, \beta \in (0, 1), \alpha \neq 1/2 \text{ or } \beta \neq 1/2\), and the range of \(\theta\) is also constrained, depending on the copula function). Thus, we obtain the MPLE for \(\phi\) by using constrained nonlinear algorithms such as interior-point algorithm [20] and active-set algorithm [21] provided by the MATLAB function, ‘fmincon’.

In order to find the global maximum in the pseudo log-likelihood, we start with interior-point algorithm using multiple starting values randomly selected from the parameter space, keep the estimate with the largest pseudo log-likelihood and then use it as a starting value to run the
active-set algorithm. Note that the estimates computed from the interior-point algorithm satisfy constraints at all iterations and the active-set algorithm, which is relatively faster, is effective when constraints are not smooth. A package of Matlab codes implemented for this paper is available upon request.

For the confidence intervals for the dependence parameter estimates, we consider a parametric bootstrap approach.[22] This is because the bootstrap approach can take into account the finite sample variation of the estimated parameters, and it can still produce the valid confidence intervals even when the parameters are constrained.

### 3.3. Copula model selection

A subsequent important problem is the selection of the copula that appropriately fits the data at hand. To select a suitable asymmetric copula among a set of candidate copula models, we recommend first choosing an asymmetric copula function with the highest values of the directional dependence measures, $\rho_{U \rightarrow V}^2$ in Equation (6) and $\rho_{V \rightarrow U}^2$ in Equation (7) (i.e. use $\rho_{U \rightarrow V}^2$ and $\rho_{V \rightarrow U}^2$ as a prior criteria for copula selection) and then testing a goodness-of-fit (GOF) to evaluate whether a chosen copula function is a good fit to the data. There are several reasons behind this recommendation. First, the higher value of $\rho_{U \rightarrow V}^2$ ($\rho_{V \rightarrow U}^2$) for a specific copula function indicates the higher predictive power of that copula regression function in the sense that the regression function of $v$ ($u$) based on the selected copula should explain most of the variation of $u$ ($v$). Second, it is possible that there are more than one copula function that fits to the data, and then in such a case, we need to consider other criteria rather than the GOF test. Last, most GOF tests in the literature require a computer-intensive bootstrap method. For each bootstrap sample, a strategy of multiple starting values in a numerical algorithm is required to find the global maximum in the pseudo log-likelihood for the parameters, and thus, this makes the computation of the $p$-value from the GOF test very computationally intensive.

In the literature, there are several rank-based procedures for testing the GOF of any class of copulas such as two tests based on the empirical copula [23] and two tests based on the Kendall transform.[24] For a detailed review of these GOF tests, please see Genest and Favre[25] and Genest et al. [26]. In this paper, we will use a GOF test based on the empirical copula which is a rank-based version of the Cramér–von Mises statistic,

$$CM_n = n \sum_{i=1}^{n} (C_n(u_i, v_i) - C(u_i, v_i; \hat{\phi}))^2,$$

(11)

where $u_i$ and $v_i$ are the normalized ranks defined in Equation (1), $C_n(u, v)$ is the empirical copula and $\hat{\phi}$ is the MPLE for $\phi$. Following the recommendation by Genest and Favre,[25] we will implement the GOF test based on Equation (11) through the parametric bootstrap.

### 3.4. Statistical inference on the directional dependence

In this section, we describe computation of the copula regression functions and the directional dependence measures for a given copula needed to detect and measure the directional dependence.

#### 3.4.1. Computation of the copula regression function and the directional dependence measures

For the GFGM copula in Equation (8), Jung et al. [6] presented the closed-form representation of the copula regression functions in Equations (3) and (5), and the general measures for the
directional dependence in joint behaviour of Equations (6) and (7):

\[ r_{U|V}^C(v) = 0.5 - \theta \beta \text{Beta}(2, \alpha + 1)(1 - v)^{\beta - 1}[1 - (1 + \beta)v], \]
\[ r_{V|U}^C(u) = 0.5 - \theta \beta \text{Beta}(2, \beta + 1)(1 - u)^{\alpha - 1}[1 - (1 + \alpha)u], \]
\[ \rho_{U \rightarrow V}^2 = \rho_c^2 \frac{\int_0^1 (f'(u))^2 \, du}{12\int_0^1 f(u) \, du^2}, \quad \rho_{V \rightarrow U}^2 = \rho_c^2 \frac{\int_0^1 (g'(v))^2 \, dv}{12\int_0^1 g(v) \, dv^2}, \]

where \( f(u) = \sqrt{\beta}u(1 - u)^\alpha, \ g(v) = \sqrt{\beta}v(1 - v)^\beta \) in \( C(u, v; \phi) \) of Equation (8), and Beta is the beta function, \( \text{Beta}(a, b) = \int_0^1 t^{a-1}(1 - t)^{b-1} \, dt \). For detailed formula for \( \rho_{U \rightarrow V}^2 \) and \( \rho_{V \rightarrow U}^2 \), please see the Appendix.

When one employs the asymmetric copulas given in Table 1, it is not a simple task to find the closed forms of the copula regression functions, \( r_{V|U}^C(u) \) in Equation (3) and \( r_{U|V}^C(v) \) in Equation (5), and the general measures for the directional dependence in joint behaviour, \( \rho_{U \rightarrow V}^2 \) in Equation (6) and \( \rho_{V \rightarrow U}^2 \) in Equation (7):

\[ r_{V|U}^C(u) = 1 - \int_0^1 C_v(u) \, dv, \quad r_{U|V}^C(v) = 1 - \int_0^1 C_u(u) \, du, \]
\[ \rho_{U \rightarrow V}^2 = 12E[(r_{V|U}^C(u))^2] - 3, \quad \rho_{V \rightarrow U}^2 = 12E[(r_{U|V}^C(v))^2] - 3, \]

where \( C_v(u) = \partial C(u, v)/\partial u, \ C_u(u) = \partial C(u, v)/\partial v \). This is because, as shown in Appendix, the forms of the conditional distribution functions \( C_u(v) \) and \( C_v(u) \) under the asymmetric copulas of Table 1 are quite complicated. Note that the conditional distribution functions \( C_u(v) \) and \( C_v(u) \) for asymmetric copulas in Table 1 are given in Appendix.

Therefore, we compute the copula regression function \( r_{V|U}^C(u) \) and \( r_{U|V}^C(v) \), and the measures for the directional dependence \( \rho_{U \rightarrow V}^2 \) and \( \rho_{V \rightarrow U}^2 \) by using the quasi-Monte Carlo method based on the quasi-random points.\([27]\) Note that the computations of the copula regression functions and the directional dependence measures are the integrations over the one-dimensional unit interval \([0, 1]\) and the two-dimensional unit square \([0, 1]^2\), respectively. We first generate the two independent sets of Sobol quasi-random points generated from \([0, 1]\), evaluate these integrals (the copula regression function and the directional dependence measures) at the generated points and approximate them by the empirical averages:

\[ r_{V|U}^C(u) \approx \tilde{r}_{V|U}^C(u) = 1 - \frac{1}{S} \sum_{s=1}^S C_v(u_s), \quad r_{U|V}^C(v) \approx \tilde{r}_{U|V}^C(v) = 1 - \frac{1}{S} \sum_{s=1}^S C_u(u_s), \]
\[ \rho_{U \rightarrow V}^2 \approx \tilde{\rho}_{U \rightarrow V}^2 = \frac{12}{S} \sum_{s=1}^S (\tilde{r}_{V|U}^C(u_s))^2 - 3, \quad \rho_{V \rightarrow U}^2 \approx \tilde{\rho}_{V \rightarrow U}^2 = \frac{12}{S} \sum_{s=1}^S (\tilde{r}_{U|V}^C(v_s))^2 - 3, \]

where \((u_1, \ldots, u_S)\) and \((v_1, \ldots, v_S)\) are two sets of Sobol quasi-random points from \([0, 1]\), \(C_v(v_s) = \partial C(u, v; \phi)/\partial u|_{u=v_s} \) for a fixed value of \( u \), and \( C_u(u_s) = C(u, v; \phi)/\partial v|_{u=u_s} \) for a fixed value of \( v \). Note that \( \theta, \alpha, \) and \( \beta \) in Equations (12) and (13) are replaced by the corresponding MPLE’s.

A naturally rising question is that why we use the quasi-Monte Carlo method based on quasi-random points rather than the classical Monte Carlo based on pseudo-random points. Several studies in the literature have shown evidence that pseudo-random draws often do not explore the sample space well.\([28]\) Note that we also found in the data analysis of Section 4 that the size of points (= \( S \)) should be very large to get stable and accurate estimates of Equations (12) and (13), resulting in a long computation time. In order to resolve this limitation of the pseudo-random draw, the quasi-Monte Carlo method, a deterministic method designed to provide quasi-random
points that better cover the sample space, has been developed. Several theoretical and empirical studies have shown that the quasi-Monte Carlo method can significantly increase the accuracy of the estimate of the integral over the classical Monte Carlo method.\[29,30\] In this paper, we use the MATLAB function ‘sobolset’ to obtain Sobol quasi-random numbers.

3.4.2. Detection and quantification of the directional dependence

Given the approximate copula regression functions in Equation (12), we can use Definition 1 in order to empirically detect directional dependence resulting from the joint behaviour of \((U, V)\). In other words, we can simply construct a plot of \(\tilde{r}_{V|U}(w)\) vs. \(\tilde{r}_{U|V}(w)\) in Equation (12) with a 45° reference line and see if there is any departure from the reference line. We call this plot the CR-plot. If a plot of \(\tilde{r}_{V|U}(w)\) vs. \(\tilde{r}_{U|V}(w)\) follows the reference line, the random pair \((U, V)\) cannot be directionally dependent in joint behaviour. Any departure from the reference line in the CR-plot is a sign of directional dependence. By checking whether the curve is located above or below the line \(v = u\), one may suspect the form of directional dependence in the data.

Once we empirically identify the existence of directional dependence between \(U\) and \(V\), we then obtain the approximate directional dependence measures in Equation (13). For interpretation of these measures, please see Section 2.2. Note that if the random pair \((U, V)\) is not directionally dependent, \(\tilde{\rho}_{U\rightarrow V}^2\) and \(\tilde{\rho}_{V\rightarrow U}^2\) would be the same numerically because \(U\) and \(V\) are both uniformly distributed from \([0, 1]\). However, the same value of \(\tilde{\rho}_{U\rightarrow V}^2\) and \(\tilde{\rho}_{V\rightarrow U}^2\) does not imply no directional dependence. Therefore, one needs to always check the CR-plot.

Given the estimates of \(\tilde{\rho}_{U\rightarrow V}^2\) and \(\tilde{\rho}_{V\rightarrow U}^2\) based on an asymmetric copula regression function suitable to the data at hand, we can use a parametric bootstrap procedure and obtain the confidence intervals needed to examine the uncertainty in the estimates.

We close this section by summarizing the procedure for the inference of the directional dependence proposed in this Section 3.

\textit{Step 1.} Obtain the normalized ranks of the observed data, Equation (1).

\textit{Step 2.} Construct various candidates of asymmetric copulas from Equation (9), as shown in Table 1, and given the pairs of normalized ranks, obtain the MPL estimates for the parameters in each asymmetric copula.

\textit{Step 3.} Calculate the general measurements for the directional dependence in joint behaviour (Equation (13)) for each copula considered and use them as a priori criteria to select an asymmetric copula function with highest predictive power.

\textit{Step 4.} Perform a formal GOF test to assess whether the selected asymmetric copula is suitable for the data.

\textit{Step 5.} If the copula model selected in Steps 3 and 4 is appropriate for the data, construct the CR-plot to empirically detect directional dependence in joint behaviour. If the CR-plot indicates a sign of directional dependence, examine the values of the general measurements for the directional dependence computed in Step 3 to determine the directional dependence in joint behaviour. One can also do a parametric bootstrap to obtain the confidence intervals for the dependence parameters and the directional dependence measurements.

4. Numerical examples

This section presents two empirical data examples, one from yeast cell cycle data analysed in Kim et al. [7] and the other from developmental research analysed in von Eye and DeShon,[31] to illustrate the methodologies described in Section 3.
4.1. Histone genes

The first example concerns expression data of histone genes involved in the cell cycle of yeast cells.[32,33] The data set consists of measurements on 6221 genes observed across 80 time points, and 800 genes regulated by cell cycle were identified. The data includes 79\((= n)\) observations for eight histone genes with known interaction patterns: HHT1, HHT2, HHF1, HHF2, HTA1, HTA2, HTB1, and HTB2. These eight genes encode the four histones (H2A, H2B, H3, and H4), which package DNA into chromosomes. Chromosomes need to be replicated before cell division, and expression of the histone genes should be tightly regulated for the proper functioning of the replication process. To detect a directional dependence of these histone genes, Kim et al. [7] proposed using the GFGM copula in Equation (8).

In this example, for illustrative purposes we only consider two histone genes, (HTB1 and HTB2), to show a poor fit of the GFGM copula in Equation (8), and the usefulness of the copula models listed in Table 1, for investigating directional dependence between two histone genes. Note that from the gene network in Chen et al. [33], HTB1 has direct connection to HTB2.

Figure 1 shows the scatter plots for the data of HTB1 and HTB2, and the corresponding normalized ranks. Note that we obtained the normalized ranks for HTB1 (denoted by \(U\)) and HTB2 (denoted by \(V\)) using Equation (1) of Section 2.1, \{(\(u_i, v_i\); \(i = 1, \ldots, 79\). We observe from Figure 1 the positive association between two histone genes. The computed Pearson’s, Kendall’s, and Spearman’s correlation coefficients were 0.8623, 0.6926, and 0.8523, respectively, and the approximate \(p\)-values of the associated tests of independence were all very small (close to 0).

In order to model the directional dependence resulting from the joint behaviour of the two genes, we used the inferential procedure given at the end of Section 3. As the candidates of the asymmetric copulas, the GFGM copula in Equation (8) and the nine families of asymmetric copulas given in Table 1 of Section 3.1 were considered.

Table 2 provides the MPLE for the parameters in each of the 10 asymmetric copula models and the corresponding directional dependence measures. We first observe that, among the 10 models considered, the model M3 (product of independence copula and Frank copula) has the highest values of \(\tilde{\rho}_{V \rightarrow U}^2\) and \(\tilde{\rho}_{U \rightarrow V}^2\). But, unlike other asymmetric copulas, the GFGM copula has the smallest values of the directional dependence measures. From Figure 2(a), showing estimated copula regressions based on two copula models, the M3 and GFGM copula, we can see that the GFGM copula fitted the data very poorly compared with the model M3.

As explained in Section 3.3, the bootstrap-based GOF test using Equation (11) was applied for the M3 and then the approximate \(p\)-value based on 1000 parametric bootstrap samples was large.
Table 2. MPLE and the directional dependence measures, $\tilde{\rho}_{U\rightarrow V}^2$ and $\tilde{\rho}_{V\rightarrow U}^2$ for HTB1 and HTB2. Note that $\phi$ for the GFGM copula and M1–M6 is $\phi = (\theta, \alpha, \beta)$ and $\phi$ for M7–M9 is $\phi = (\theta_1, \theta_2, \alpha, \beta)$.

<table>
<thead>
<tr>
<th>C(u, v; $\phi$)</th>
<th>$\phi$</th>
<th>$\tilde{\rho}_{V\rightarrow U}^2$</th>
<th>$\tilde{\rho}_{U\rightarrow V}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFGM</td>
<td>(1.6793, 1.2320, 1.2320)</td>
<td>0.0645</td>
<td>0.0645</td>
</tr>
<tr>
<td>M1</td>
<td>(3.5514, 1.0000e−04, 0.0523)</td>
<td>0.6467</td>
<td>0.6377</td>
</tr>
<tr>
<td>M2</td>
<td>(3.2772, 1.0000e−04, 0.0779)</td>
<td>0.6875</td>
<td>0.6933</td>
</tr>
<tr>
<td>M3</td>
<td>(12.1596, 1.0112e−04, 0.0525)</td>
<td>0.7526</td>
<td>0.7535</td>
</tr>
<tr>
<td>M4</td>
<td>(1.0000, 1.0000e−04, 1.0000e−04)</td>
<td>0.2642</td>
<td>0.2642</td>
</tr>
<tr>
<td>M5</td>
<td>(35.9508, 1.0000e−04, 0.0525)</td>
<td>0.7104</td>
<td>0.7135</td>
</tr>
<tr>
<td>M6</td>
<td>(3.5089, 0.0259, 0.1381)</td>
<td>0.7346</td>
<td>0.7372</td>
</tr>
<tr>
<td>M7</td>
<td>(3.8619, 59.5775, 0.7494, 0.6661)</td>
<td>0.6752</td>
<td>0.6814</td>
</tr>
<tr>
<td>M8</td>
<td>(3.3470, 25.4312, 0.9888, 0.9061)</td>
<td>0.7090</td>
<td>0.7118</td>
</tr>
<tr>
<td>M9</td>
<td>(9.2397, 20.3734, 0.3789, 0.3169)</td>
<td>0.7346</td>
<td>0.7372</td>
</tr>
</tbody>
</table>

In order to check if two genes are directionally dependent, we constructed the CR-plot described in Section 3.4.2, a plot of $\tilde{r}_{V|U}(w)$ vs. $\tilde{r}_{U|V}(w)$ in Equation (12). Figure 2(b) is the CR-plot for the copula model M3. We see that directional dependence between the genes exists only when the values of two variables are both large or both small, which explains the small difference between $\tilde{\rho}_{V\rightarrow U}^2(=0.7526)$ and $\tilde{\rho}_{U\rightarrow V}^2(=0.7535)$ for the M3.

Table 3 presents 95% bootstrap bias-corrected percentile confidence intervals for the dependence parameters and the directional dependence measures. That is, given the MPLE for the parameters in the model M3, we generated 1000 (= B) random samples of size $n (= 79)$ from the same copula model, estimated the MPLE of the parameters for each bootstrap sample, and computed the directional dependence measures, $\rho_{V\rightarrow U}^2$ and $\rho_{U\rightarrow V}^2$. Given $B$ bootstrap estimates for the dependence parameters and the directional dependence measures, we then obtained 95% bootstrap confidence intervals using the bias-corrected percentile method.[22] We see that the bootstrap confidence interval for the dependence parameter in the M3 does not include zero (note that a zero value of $\theta$ in the Frank copula represents independence between two genes).

From the results shown above, we see that there is empirical evidence that the two genes are directionally dependent, and the proportional of total variability of HTB1 (HTB2) that is explained by the copula regression using HTB2 (HTB1) is 75.26% (75.35%).

4.2. Attention deficit hyperactivity disorder and blood lead content

The second example, investigated by Nigg et al. [34] and von Eye and DeShon,[31] concerns the relation between attention deficit hyperactivity disorder (ADHD) and blood lead concentration. Nigg et al. [34] discussed the direction of the dependence between blood lead level and ADHD because (1) lead may affect brain development or (2) children with ADHD cannot stay focused enough to keep away from lead-tainted objects. Note that they found that children with ADHD have higher levels of lead in their blood than children without this disorder.

von Eye and DeShon [31] analysed the same data to determine the direction of dependence using the linear regression method proposed by Dodge and Rousson.[10] Assuming that the error term in a linear regression model is symmetric, the cube of the Pearson’s correlation between two variables $X$ and $Y$ (denoted by $\rho_{XY}^3$) is equal to the ratio of the skewness of $Y$ (denoted by $\gamma_Y$) to the skewness of $X$ (denoted by $\gamma_X$): $(\rho_{XY}^3)^2 = \gamma_Y^2/\gamma_X^2$. Since $(\rho_{XY}^3)^2 \leq 1$, $\gamma_Y^2 \leq \gamma_X^2$, $Y$ is linearly...
dependent on $X$. However, as pointed out in Sungur,[4] the approach suggested by Dodge and Rousson [10] is designed to determine directional dependence in the marginal behaviour of the variables, not the joint behaviour. We will investigate directional dependence in joint behaviour through the copula regression function.

The data set has three variables measured from 150 ($= n$) children, child blood lead level ($\mu g/dL$), the number of diagnostic and statistical manual of mental disorders fourth edition (DSM-IV) inattentive symptoms, and the number of DSM-IV hyperactive-impulsive symptoms. For illustration purposes, we here investigate the directional dependence between child blood lead level and the number of DSM-IV hyperactive-impulsive symptoms.

The scatterplot of lead blood level with hyperactive-impulsiveness is shown in the leftmost plot of Figure 3. Table 4 represents the descriptive statistics for these two variables, and the values
Table 3. Estimates for the dependence parameter in the model M3 and directional dependence measures, and the corresponding 95% bootstrap bias-corrected percentile confidence intervals.

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Bootstrap confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta$</td>
<td>12.1596 (8.9624, 15.4642)</td>
</tr>
<tr>
<td>$\tilde{\rho}_{V \rightarrow U}$</td>
<td>0.7526 (0.6345, 0.8508)</td>
</tr>
<tr>
<td>$\tilde{\rho}_{U \rightarrow V}$</td>
<td>0.7535 (0.6360, 0.8501)</td>
</tr>
</tbody>
</table>

Figure 3. Scatter plots of blood lead level and the number of DSM-IV hyperactive-impulsive symptoms – the leftmost plot is for the raw data, the middle plot is for the continuous version of the data, and the rightmost plot is for the normalized ranks for the continuous version.

Table 4. Descriptive statistics for lead blood level, hyperactive-impulsiveness measured from 150(=n) children, and continuous version of hyperactive-impulsiveness.

<table>
<thead>
<tr>
<th></th>
<th>Lead level</th>
<th>Hyperactive-impulsiveness</th>
<th>Continuous version of hyperactive-impulsiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1.0337</td>
<td>4.1567</td>
<td>3.6589</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.5420</td>
<td>3.4257</td>
<td>3.4232</td>
</tr>
<tr>
<td>Skewness</td>
<td>2.1564</td>
<td>0.2227</td>
<td>0.2241</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>8.1580</td>
<td>1.4323</td>
<td>1.4649</td>
</tr>
</tbody>
</table>

of skewness and kurtosis indicate that both variables are far from being normally distributed. To quantify the degree of dependence between the two variables, we calculate two well-known nonparametric measures of dependence: the Kendall’s $\tau$ and Spearman’s $\rho$ coefficients. The computed values were 0.241 for Kendall’s $\tau$ and 0.353 for Spearman’s $\rho$, and the approximate $p$-values of the associated tests of independence were $2.566e-05$ and $9.527e-06$, respectively. Since we need to capture the joint dependence structure between the two variables free from the marginal influence of each variable, we will use a copula model based on the normalized ranks of each variable, as shown in Equation (1) of Section 2.1.

Before obtaining the ranks for each variable, however, notice that the hyperactive-impulsiveness is a discrete variable valued in a subset of the set of nonnegative integers, and there are ties in the data. Note that the presence of ties in the data may substantially affect the copula estimation.[35] In order to fully facilitate the good properties of the copula, one needs to deal with the discreteness of the variable and the presence of ties in an appropriate way. Thus, we obtain a continuous extension of the hyperactive-impulsiveness by using the randomization technique proposed by [36]: given integer-valued $X_i$, consider a continuous random variable $X_i^* = X_i + (U_i - 1)$ where $U_i$ is uniform on $(0, 1)$ and independent of $X_i$. As shown in,[36] the original variable can be recovered from its continuous extension, and the distribution function of the original variable is exactly the same as that of its continuous extension. Furthermore, this approach randomly breaks the ties in the data.
Note that Kojadinovic and Yan [37] verified that the randomization (designed to randomly break the ties) does not change the results for the copula inference.

The middle plot of Figure 3 shows the scatter plots of lead blood level with a continuous version of hyperactive-impulsiveness. From Table 4, we can see that the continuous extension does not affect the descriptive statistics of hyperactive-impulsiveness much. Note that the sample values of Kendall’s $\tau$ and Spearman’s $\rho$ between lead level and continuous version of hyperactive-impulsiveness were 0.227 and 0.346, respectively, and the associated $p$-values of tests of independence were $4.123\times10^{-5}$ and $1.443\times10^{-5}$, respectively.

In order to model the directional dependence between lead blood level and the continuous version of hyperactive-impulsiveness, we first get the normalized ranks for each variable using Equation (1) of Section 2.1, denoted by $\{(u_i, v_i), i = 1, \ldots, 150(= n)\}$ (see the rightmost plot of Figure 3 for a scatter plot of the normalized ranks for lead blood level, denoted by $U$, and the continuous version of hyperactive-impulsiveness, denoted by $V$). We then fitted the data to nine families of asymmetric copulas given in Table 1 of Section 3.1 and the GFGM copula in Equation (8). For the estimation of the parameter in a copula model, we used the method of MPL explained in Section 3.4 that provides the MPLE for the parameters. We then computed the approximate directional dependence measures introduced in Section 3.4.1, $\tilde{\rho}^2_{U \rightarrow V}$ and $\tilde{\rho}^2_{V \rightarrow U}$ in Equation (13).

Table 5 gives the MPLE for the parameters in each of the 10 asymmetric copula models and the corresponding directional dependence measures. We first observe that the GFGM copula model has the smallest values of the directional dependence measures among the 10 models. Note that the estimate of the dependence parameter in the GFGM copula occurred at the boundary of the parameter space, which is the upper limit of the possible range, $[-1, 1.3726]$. Among the models considered, the model M7 (product of two Clayton copulas) and M9 (product of two Frank copulas) have the larger values of $\tilde{\rho}^2_{U \rightarrow V}$ and $\tilde{\rho}^2_{V \rightarrow U}$ than those values based on other copula models. We also applied the bootstrap-based GOF test using the Cramér–von Mises statistic in Equation (11) for the two models, M7 and M9, and we could not reject them at even the 10% level due to the large $p$-values. Therefore, we focus on these two models M7 and M9 for further analysis. Note that Figure 4 displays the estimated copula regression curves using three copulas M7, M9, and GFGM.

Given the estimated copula models, we investigated whether or not a pair of variables, blood lead level, and hyperactive-impulsiveness, is directionally dependent by constructing the CR-plot described in Section 3.4.2, a plot of $\tilde{r}_{V|U}(w)$ vs. $\tilde{r}_{U|V}(w)$ in Equation (12). Figure 5 is the CR-plot for M7 and M9. We see that directional dependence exists for both models because there appears to be departure from the reference line and the departure in the model M7 is more noticeable.

<table>
<thead>
<tr>
<th>Model</th>
<th>$(\theta_1, \alpha, \beta)$</th>
<th>$\tilde{\rho}^2_{U \rightarrow V}$</th>
<th>$\tilde{\rho}^2_{V \rightarrow U}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFGM</td>
<td>(1.3726, 1.1059, 1.2392)</td>
<td>0.0450</td>
<td>0.0523</td>
</tr>
<tr>
<td>M1</td>
<td>(1.8042, 0.2781, 0.4188)</td>
<td>0.1072</td>
<td>0.1012</td>
</tr>
<tr>
<td>M2</td>
<td>(1.3320, 1.0004e−04, 0.1903)</td>
<td>0.1116</td>
<td>0.1183</td>
</tr>
<tr>
<td>M3</td>
<td>(2.3842, 1.0000e−04, 0.1248)</td>
<td>0.1151</td>
<td>0.1149</td>
</tr>
<tr>
<td>M4</td>
<td>(0.7901, 1.0001e−04, 0.0436)</td>
<td>0.1127</td>
<td>0.1117</td>
</tr>
<tr>
<td>M5</td>
<td>(2.6876, 1.0417e−04, 1.0017e−04)</td>
<td>0.1016</td>
<td>0.1016</td>
</tr>
<tr>
<td>M6</td>
<td>(1.3324, 1.0000e−04, 0.1922)</td>
<td>0.1119</td>
<td>0.1184</td>
</tr>
<tr>
<td>M7</td>
<td>(1.5953, 61.2448, 0.3891, 0.9396)</td>
<td>0.1205</td>
<td>0.1337</td>
</tr>
<tr>
<td>M8</td>
<td>(1.2359, 60.2369, 0.7957, 0.9771)</td>
<td>0.0869</td>
<td>0.0936</td>
</tr>
<tr>
<td>M9</td>
<td>(3.9240, 18.2006, 0.3541, 0.9117)</td>
<td>0.1294</td>
<td>0.1323</td>
</tr>
</tbody>
</table>
Figure 4. Estimated copula regressions based on three copula models, M7 (red line), M9 (green line), and GFGM (blue line).

Figure 5. CR-plot – a plot of $\tilde{r}_{V|U}(w)$ vs. $\tilde{r}_{U|V}(w)$ in Equation (12) for M7 (left) and M9 (right).

The results for the two models M7 and M9 shown in Table 5 and Figure 5 provide empirical justification that blood lead level and hyperactive-impulsiveness are directionally dependent. The proportion of total variation of lead level that is explained by the copula regression using hyperactive-impulsiveness is 12.05% for M7 and 12.94% for M9. The proportion of total variation of hyperactive-impulsiveness that is explained by the copula regression using blood lead level is 13.37% for M7 and 13.23% for M9. Though the two proportions in each model appear not to differ substantially, it is slightly more edged for hyperactive-impulsiveness to be explained by blood lead level rather than vice versa.

Table 6 presents 95% bootstrap bias-corrected percentile confidence intervals for the dependence parameters and the approximate directional dependence measures. It appears that the model M7 shows more significant results than the model M9. That is, the 95% confidence intervals for the dependence parameters in the M7 do not include zero, but the 95% confidence intervals for the first Frank copula dependence parameter $\theta_1$ in the M9 included zero. Note that a zero value of $\theta$ in the Clayton and Frank copula represents independence between two variables. Moreover, the 95% confidence intervals for $\rho^2_{U\rightarrow V}$ and $\rho^2_{V\rightarrow U}$ in the model M7 are shorter than those of the model M9. Note that we found that the maximized pseudo log-likelihood value for the model M7 (11.4761) was higher than that of the model M9 (10.6734).
Table 6. Estimates for the dependence parameters and directional dependence measures in the two models (M7 and M9), and the corresponding 95% bootstrap bias-corrected percentile confidence intervals.

<table>
<thead>
<tr>
<th></th>
<th>M7</th>
<th>M9</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\theta_1)</td>
<td>1.5953</td>
<td>3.9240</td>
</tr>
<tr>
<td></td>
<td>(0.2571, 19.1502)</td>
<td>(−91.0763, 11.8828)</td>
</tr>
<tr>
<td>(\theta_2)</td>
<td>61.2448</td>
<td>18.2006</td>
</tr>
<tr>
<td></td>
<td>(2.4288, 245.1837)</td>
<td>(4.9347, 26.7412)</td>
</tr>
<tr>
<td>(\rho_{V\rightarrow U}^2)</td>
<td>0.1205</td>
<td>0.1294</td>
</tr>
<tr>
<td></td>
<td>(0.0419, 0.1908)</td>
<td>(0.0480, 0.2314)</td>
</tr>
<tr>
<td>(\rho_{U\rightarrow V}^2)</td>
<td>0.1337</td>
<td>0.1323</td>
</tr>
<tr>
<td></td>
<td>(0.0568, 0.2013)</td>
<td>(0.0504, 0.2310)</td>
</tr>
</tbody>
</table>

5. Discussion

In this article, we have presented a generalized procedure for the full inference of the directional dependence in joint behaviour based on the copula regression and demonstrated it using real data examples. We focused on the directional dependence stemming from the joint behaviour of the two variables. As the reviewer pointed out, it would be worthy to undertake research about the conditions under which the method based on marginal distributions and the method based on the joint distribution of the variables yield different results in terms of directional dependence.

In applications of the proposed procedure for the analysis of directional dependence, we considered bivariate data. A valuable extension of this research would be to construct multivariate asymmetric copulas and generalize the inference procedure to the multivariate data. Finally, we assumed the directional dependence between two variables with no extraneous variables. An extraneous variable is an unobserved variable affecting the two variables of interest. In practice, there often exists such an extraneous variable, and the consideration of an extraneous variable may reveal a spurious effect between two variables of interest. Future work will consider development of the copula-based regression model which takes into account the adjustment of the effect of an extraneous variable in the context of directional dependence.

Acknowledgements

We are thankful to the respected editor, the associate editor, and an anonymous learned referee for the valuable comments on the original version of this manuscript which led to substantial improvement. We appreciate professor Alexander von Eye’s willingness to provide us the data set (ADHD and blood lead content in Section 4.2) used in our paper. We also appreciate professor Christian Genest for giving us several comments about the copula GOF test for asymmetric copula family by email.

References

Appendix

In this appendix, we first provide the explicit analytical expressions of the general measurement for the directional dependence in the Rodríguez-Lallena and Úbeda-Flores copula family of Equation (8). Note that the formulas given in Jung et al. [6] are incorrect. The second part of the appendix derives the formulas of the copula density, the conditional copula distribution function and its derivatives needed to compute the MPLE for the parameters maximizing Equation (10),
the copula regression functions in Equation (12) and the directional dependence measurements in Equation (13) for each of the asymmetric copulas given in Table 1.

A.1. The general measures for directional dependence in the Rodríguez-Lallena and Úbeda-Flores copula family of Equation (8)

The detailed formula for \( \rho_{U \rightarrow V}^2 \) and \( \rho_{V \rightarrow U}^2 \) in Rodríguez-Lallena and Úbeda-Flores copula family of Equation (8) are as follows:

\[
\rho_{U \rightarrow V}^2 = \frac{\int_0^1 (f'(u))^2 \, du}{12 \int_0^1 f(u) \, du^2}, \\
\rho_{V \rightarrow U}^2 = \frac{\int_0^1 (g'(v))^2 \, dv}{12 \int_0^1 g(v) \, dv^2},
\]

(A1)

where

\[
\rho_c = 12 \int_0^1 \int_0^1 C(u, v) \, du \, dv - 3 = 12 \theta \text{Beta}(2, \alpha + 1) \text{Beta}(2, \beta + 1),
\]

\[
\int_0^1 (f'(u))^2 \, du = \theta \alpha^2 \text{Beta}(1, 2 \alpha - 1), \quad \left[ \int_0^1 f(u) \, du \right]^2 = \theta \left[ \text{Beta}(1, \alpha + 1) \right]^2,
\]

\[
\int_0^1 (g'(v))^2 \, dv = \theta \beta^2 \text{Beta}(1, 2 \beta - 1), \quad \left[ \int_0^1 g(v) \, dv \right]^2 = \theta \left[ \text{Beta}(1, \beta + 1) \right]^2.
\]

A.2. Conditional distribution functions and density functions for asymmetric copulas in Table 1, \( C(u, v; \phi) = C_1(u^\alpha, v^\beta)C_2(u^\tilde{\alpha}, v^\tilde{\beta}) \)

For notational convenience we suppress the dependence parameter \( \theta \) in \( C_1 \) and \( C_2 \). The conditional distribution functions for \( V \) given \( U \), \( C_0(u) \), and for \( U \) given \( V \), \( C_1(u) \), are

\[
C_0(v) = \frac{\partial}{\partial u} C(u, v; \phi) = \left( \frac{\partial}{\partial u} C_1(u^\alpha, v^\beta) \right) C_2(u^\tilde{\alpha}, v^\tilde{\beta}) + C_1(u^\alpha, v^\beta) \left( \frac{\partial}{\partial v} C_2(u^\tilde{\alpha}, v^\tilde{\beta}) \right),
\]

\[
C_1(u) = \frac{\partial}{\partial v} C(u, v; \phi) = \left( \frac{\partial}{\partial v} C_1(u^\alpha, v^\beta) \right) C_2(u^\tilde{\alpha}, v^\tilde{\beta}) + C_1(u^\alpha, v^\beta) \left( \frac{\partial}{\partial u} C_2(u^\tilde{\alpha}, v^\tilde{\beta}) \right),
\]

and the density function for asymmetric copula, \( c(u, v; \phi) \), is

\[
c(u, v; \phi) = \frac{\partial^2 C(u, v; \phi)}{\partial u \partial v} = \frac{\partial^2}{\partial u \partial v} \left[ C_1(u^\alpha, v^\beta)C_2(u^\tilde{\alpha}, v^\tilde{\beta}) \right]
\]

\[
= \frac{\partial}{\partial u} \left[ \left( \frac{\partial}{\partial v} C_1(u^\alpha, v^\beta) \right) C_2(u^\tilde{\alpha}, v^\tilde{\beta}) + C_1(u^\alpha, v^\beta) \left( \frac{\partial}{\partial v} C_2(u^\tilde{\alpha}, v^\tilde{\beta}) \right) \right]
\]

\[
= \left( \frac{\partial^2}{\partial u \partial v} C_1(u^\alpha, v^\beta) \right) C_2(u^\tilde{\alpha}, v^\tilde{\beta}) + \left( \frac{\partial}{\partial v} C_1(u^\alpha, v^\beta) \right) \left( \frac{\partial}{\partial u} C_2(u^\tilde{\alpha}, v^\tilde{\beta}) \right)
\]

\[
+ \left( \frac{\partial}{\partial u} C_1(u^\alpha, v^\beta) \right) \left( \frac{\partial}{\partial v} C_2(u^\tilde{\alpha}, v^\tilde{\beta}) \right) + C_1(u^\alpha, v^\beta) \left( \frac{\partial^2}{\partial u \partial v} C_2(u^\tilde{\alpha}, v^\tilde{\beta}) \right),
\]

where

\[
\frac{\partial}{\partial u} C_1(u^\alpha, v^\beta) = \alpha u^{\alpha-1} \left( \frac{\partial}{\partial u(u^\alpha)} C_1(u^\alpha, v^\beta) \right), \quad \frac{\partial}{\partial u} C_2(u^\tilde{\alpha}, v^\tilde{\beta}) = \tilde{\alpha} u^{\tilde{\alpha}-1} \left( \frac{\partial}{\partial u(u^\tilde{\alpha})} C_2(u^\tilde{\alpha}, v^\tilde{\beta}) \right),
\]

\[
\frac{\partial}{\partial v} C_1(u^\alpha, v^\beta) = \beta v^{\beta-1} \left( \frac{\partial}{\partial v(v^\beta)} C_1(u^\alpha, v^\beta) \right), \quad \frac{\partial}{\partial v} C_2(u^\tilde{\alpha}, v^\tilde{\beta}) = \tilde{\beta} v^{\tilde{\beta}-1} \left( \frac{\partial}{\partial v(v^\tilde{\beta})} C_2(u^\tilde{\alpha}, v^\tilde{\beta}) \right),
\]

\[
\frac{\partial^2}{\partial u \partial v} C_1(u^\alpha, v^\beta) = \alpha \beta u^{\alpha-1} v^{\beta-1} \left( \frac{\partial^2}{\partial (u^\alpha) \partial (v^\beta)} C_1(u^\alpha, v^\beta) \right),
\]

\[
\frac{\partial^2}{\partial u \partial v} C_2(u^\tilde{\alpha}, v^\tilde{\beta}) = \tilde{\alpha} \tilde{\beta} u^{\tilde{\alpha}-1} v^{\tilde{\beta}-1} \left( \frac{\partial^2}{\partial (u^{\tilde{\alpha}}) \partial (v^{\tilde{\beta}})} C_2(u^\tilde{\alpha}, v^\tilde{\beta}) \right).
\]
A.2.3. Gumbel copula for $C$

For the independence copula in M1–M5 of Table 1, $C_1(u^\alpha, v^\beta) = u^\alpha v^\beta$,

$$\frac{\partial}{\partial u} C_1(u^\alpha, v^\beta) = \alpha u^{\alpha-1} v^\beta, \quad \frac{\partial}{\partial v} C_1(u^\alpha, v^\beta) = \beta u^\alpha v^{\beta-1} \quad \text{and} \quad \frac{\partial^2}{\partial u \partial v} C_1(u^\alpha, v^\beta) = \alpha \beta u^{\alpha-1} v^{\beta-1}. $$

In order to compute $C_u(v)$, $C_i(u)$, and $c(u, v; \phi)$ in the other copula functions of Table 1, it is enough to have the closed forms of

$$C_2(u^\alpha, v^\beta), \quad \frac{\partial}{\partial (u^\alpha)} C_2(u^\alpha, v^\beta), \quad \frac{\partial}{\partial (v^\beta)} C_2(u^\alpha, v^\beta) \quad \text{and} \quad \frac{\partial^2}{\partial (u^\alpha)(v^\beta)} C_2(u^\alpha, v^\beta).$$

A.2.1. Clayton copula for $C_2$ in M1

The Clayton copula is

$$C_2(u^\alpha, v^\beta; \theta) = ((u^\alpha)^{-\theta} + (v^\beta)^{-\theta} - 1)^{-1/\theta} \quad \text{with} \quad \theta \in (0, \infty),$$

$$\frac{\partial}{\partial (u^\alpha)} C_2(u^\alpha, v^\beta; \theta) = \frac{(u^\alpha)^{-\theta-1}}{[1 - (u^\alpha)^{-\theta} + (v^\beta)^{-\theta}]^{2/\theta}}, \quad \frac{\partial}{\partial (v^\beta)} C_2(u^\alpha, v^\beta; \theta) = \frac{(v^\beta)^{-\theta-1}}{[1 - (u^\alpha)^{-\theta} + (v^\beta)^{-\theta}]^{2/\theta}};$$

$$\frac{\partial^2}{\partial (u^\alpha)(v^\beta)} C_2(u^\alpha, v^\beta; \theta) = \frac{(\theta + 1)(u^\alpha)^{-\theta-1}(v^\beta)^{-\theta-1}}{[1 - (u^\alpha)^{-\theta} + (v^\beta)^{-\theta}]^{3/\theta}}.$$

A.2.2. Gumbel copula for $C_2$ in M2

The Gumbel copula is

$$C_2(u^\alpha, v^\beta; \theta) = \exp\{\left(-\log(u^\alpha)^\theta + (\log(v^\beta)^\theta\right)^{1/\theta}) \quad \text{with} \quad \theta \in [1, \infty),$$

$$\frac{\partial}{\partial (u^\alpha)} C_2(u^\alpha, v^\beta; \theta) = C_2(u^\alpha, v^\beta; \theta) A^{(1/\theta)-1} \left(-\log(u^\alpha)^\theta\right)^{\theta-1} \frac{1}{u^\alpha};$$

$$\frac{\partial}{\partial (v^\beta)} C_2(u^\alpha, v^\beta; \theta) = C_2(u^\alpha, v^\beta; \theta) A^{(1/\theta)-1} \left(-\log(v^\beta)^\theta\right)^{\theta-1} \frac{1}{v^\beta};$$

$$\frac{\partial^2}{\partial (u^\alpha)(v^\beta)} C_2(u^\alpha, v^\beta; \theta) = \frac{(-\log(u^\alpha)^\theta)^{\theta-1}}{u^\alpha} \left[\left(\frac{\partial}{\partial (v^\beta)} C_2(u^\alpha, v^\beta; \theta)\right) A^{(1/\theta)-1}\right.\right.$$

$$\left. + C_2(u^\alpha, v^\beta; \theta) \left(\frac{1}{\theta} - 1\right) A^{(1/\theta)-2} (-\log(v^\beta)^\theta)^{\theta-1} \left(-\frac{1}{v^\beta}\right)\right],$$

where $A = (-\log(u^\alpha)^\theta + (-\log(v^\beta)^\theta)^{1/\theta}$.

A.2.3. Frank copula for $C_2$ in M3

The Frank copula is

$$C_2(u^\alpha, v^\beta; \theta) = -\frac{1}{\theta} \log \left\{ 1 + \frac{(\exp(-\theta u^\alpha) - 1)(\exp(-\theta v^\beta) - 1)}{\exp(-\theta) - 1} \right\} \quad \text{with} \quad \theta \in R \setminus \{0\},$$

$$\frac{\partial}{\partial (u^\alpha)} C_2(u^\alpha, v^\beta; \theta) = \frac{\exp(-\theta u^\alpha) - 1 - \exp(-\theta v^\beta)}{A}, \quad \frac{\partial}{\partial (v^\beta)} C_2(u^\alpha, v^\beta; \theta) = \frac{\exp(-\theta v^\beta) - 1 - \exp(-\theta u^\alpha)}{A},$$

$$\frac{\partial^2}{\partial (u^\alpha)(v^\beta)} C_2(u^\alpha, v^\beta; \theta) = \frac{\theta \exp(-\theta u^\alpha + v^\beta)(1 - \exp(-\theta))}{A^2},$$

where $A = (\exp(-\theta) - 1 + (\exp(-\theta u^\alpha) - 1)(\exp(-\theta v^\beta) - 1)$. 

A.2.4. AMH copula for $C_2$ in M4

The AMH copula is

$$C_2(u^\alpha, v^\beta; \theta) = \frac{u^\alpha v^\beta}{1 - \theta(1 - u^\alpha)(1 - v^\beta)}$$

with $\theta \in [-1, 1]$.

\[
\frac{\partial}{\partial(u^\alpha)} C_2(u^\alpha, v^\beta; \theta) = \frac{v^\beta(1 - \theta + \theta v^\beta)}{A^2}, \quad \frac{\partial}{\partial(v^\beta)} C_2(u^\alpha, v^\beta; \theta) = \frac{u^\alpha(1 - \theta + \theta u^\alpha)}{A^2},
\]

\[
\frac{\partial^2}{\partial(u^\alpha)(v^\beta)} C_2(u^\alpha, v^\beta; \theta) = \frac{\theta^2(1 - u^\alpha)(1 - v^\beta) + \theta(-2 + u^\alpha + v^\beta + u^\alpha v^\beta) + 1}{A^3},
\]

where $A = 1 - \theta(1 - u^\alpha)(1 - v^\beta)$.

A.2.5. Plackett copula for $C_2$ in M5

The Plackett copula is

$$C_2(u^\alpha, v^\beta; \theta) = \frac{A - \sqrt{A^2 - 4\theta(\theta - 1)u^\alpha v^\beta}}{2(\theta - 1)}$$

with $\theta > 0, \theta \neq 1$.

\[
\frac{\partial}{\partial(u^\alpha)} C_2(u^\alpha, v^\beta; \theta) = 0.5 - 0.5(A^2 - 4\theta(\theta - 1)u^\alpha v^\beta)^{-0.5}(A - 2\theta v^\beta),
\]

\[
\frac{\partial}{\partial(v^\beta)} C_2(u^\alpha, v^\beta; \theta) = 0.5 - 0.5(A^2 - 4\theta(\theta - 1)u^\alpha v^\beta)^{-0.5}(A - 2\theta u^\alpha),
\]

\[
\frac{\partial^2}{\partial(u^\alpha)(v^\beta)} C_2(u^\alpha, v^\beta; \theta) = \theta[1 + (\theta - 1)(u^\alpha + v^\beta - 2u^\alpha v^\beta)]
\]

\[
\frac{(A^2 - 4\theta(\theta - 1)u^\alpha v^\beta)^{3/2}}{(A^2 - 4\theta(\theta - 1)u^\alpha v^\beta)^{3/2}},
\]

where $A = 1 + (\theta - 1)(u^\alpha + v^\beta)$.

A.2.6. Asymmetric copulas in M6–M9

One can construct the asymmetric copulas in M6–M9 by using $C_u(v), C_v(u),$ and $c(u, v; \phi)$ given in Appendix A.2.1–A.2.5.