Journal of Multivariate Analysis 159 (2017) 111–133

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

# Score tests for covariate effects in conditional copulas

Irène Gijbels [a], Marek Omelka [b,*], Michal Pešta [b], Noël Veraverbeke [c,d]

[a] Department of Mathematics and Leuven Statistics Research Center (LStat), KU Leuven, Celestijnenlaan 200B, Box 2400, B-3001 Leuven (Heverlee), Belgium

[b] Department of Probability and Statistics, Faculty of Mathematics and Physics, Charles University, Sokolovská 83, 186 75 Praha 8, Czech Republic

[c] Center for Statistics, Hasselt University, Agoralaan-building D, B-3590 Diepenbeek, Belgium

[d] Unit for BMI, North-West University, Potchefstroom, South Africa

ARTICLE INFO

*Article history:*
Received 27 September 2016
Available online 10 May 2017

*AMS subject classifications:*
62H15
62F05
62G10

*Keywords:*
Conditional copula
Covariate effect
Parametric dependence structure
Rao score test
Specification test

ABSTRACT

We consider copula modeling of the dependence between two or more random variables in the presence of a multivariate covariate. The dependence parameter of the conditional copula possibly depends on the value of the covariate vector. In this paper we develop a new testing methodology for some important parametric specifications of this dependence parameter: constant, linear, quadratic, etc. in the covariate values, possibly after transformation with a link function. The margins are left unspecified. Our novel methodology opens plenty of new possibilities for testing how the conditional copula depends on the multivariate covariate and also for variable selection in copula model building. The suggested test is based on a Rao-type score statistic and regularity conditions are given under which the test has a limiting chi-square distribution under the null hypothesis. For small and moderate sample sizes, a permutation procedure is suggested to assess significance. In simulations it is shown that the test performs well (even under misspecification of the copula family and/or the dependence parameter structure) in comparison to available tests designed for testing for constancy of the dependence parameter. The test is illustrated on a real data set on concentrations of chemicals in water samples.



## 1. Introduction

Conditional copulas provide a convenient way to model the dependence between random variables whose dependence structure is possibly influenced by covariates. See Patton [18] for an early reference, Acar et al. [2] and Abegaz et al. [1] for semiparametric estimation, and Veraverbeke et al. [26] and Gijbels et al. [14] for nonparametric estimation of conditional copulas, among others.

A crucial fact in conditional copula modeling is that, in general, the covariates influence the conditional copula on two levels: the copula itself (the dependence structure) may change with the value of the covariate vector, and the margins may be influenced by the covariate vector. An important simplification occurs when the dependence structure remains unchanged whatever the realized value of the covariate vector. This simplification is often referred to as 'the simplifying assumption'; see, e.g., Hobæk Haff et al. [15], Acar et al. [4], Stöber et al. [23] and Gijbels et al. [11].

* Corresponding author.
*E-mail address:* omelka@karlin.mff.cuni.cz (M. Omelka).

http://dx.doi.org/10.1016/j.jmva.2017.05.001

In this paper we contribute to testing for covariate effects on conditional copulas, and this within a parametric copula setting. The literature on such testing problems is quite limited. It includes a semiparametric likelihood ratio type of test – not assuming any structure on the functional dependence parameter in a given copula – that was proposed and studied in Acar et al. [3]; and nonparametric tests, leaving the copula dependence structure as well as the margins fully unspecified. In contrast to these semiparametric and nonparametric approaches, this paper presents a new approach, in which the starting point is to model parametrically the copula, but also its functional dependence parameter. The margins are left unspecified. Despite this parametric framework for both the copula and the dependence parameter, it turns out that the test proposed herein continues to have a very good performance under misspecification of one or both of these parts. This is an important first advantage. A second, more practical, advantage is that the proposed test does not require any choice of smoothing parameter (due to its major parametric setting). Thirdly, the developed test methodology can be applied for several testing problems: (i) testing for no covariate effect; (ii) testing for specific effects of a selection of covariates; (iii) testing for specific effects of all covariates (such as linear versus quadratic).

The paper is further organized as follows. In Section 2 we present the statistical framework, and briefly review semiparametric and nonparametric tests that are available in the literature. The new test methodology is exposed in Section 3, in which the essential elements of the derivation of the test and its asymptotic behavior are presented. Details about the theoretical results, including their proofs, are provided in the Appendix. The test methodology is applicable to several testing settings, as is explained in Section 3. In Sections 4 and 5 the test methodology is illustrated in various testing problems, in a univariate as well as a multivariate covariate setting. The use of the test in statistical analysis is demonstrated in a real data example in Section 6.

## 2. Statistical framework and state-of-the-art

In this section we first introduce the statistical framework, the main testing problem of interest, and briefly indicate major testing procedures available in the literature.

### 2.1. Statistical framework

Suppose we have $n$ independent and identically distributed observations $(Y_{11}, Y_{21}, \mathbf{X}_1), \ldots, (Y_{1n}, Y_{2n}, \mathbf{X}_n)$ from a random vector $(Y_1, Y_2, \mathbf{X})$, where $Y_1$ and $Y_2$ are univariate random variables and $\mathbf{X}$ is a $d$-dimensional random vector. Let $H(y_1, y_2, \mathbf{x})$ be the cumulative distribution function of $(Y_1, Y_2, \mathbf{X})$. Denote the joint and marginal distribution functions of $(Y_1, Y_2)$, conditionally on $\mathbf{X} = \mathbf{x}$, as

$$H_{\mathbf{x}}(y_1, y_2) = \Pr(Y_1 \le y_1, Y_2 \le y_2 \mid \mathbf{X} = \mathbf{x}),$$
$$F_{1\mathbf{x}}(y_1) = \Pr(Y_1 \le y_1 \mid \mathbf{X} = \mathbf{x}), \qquad F_{2\mathbf{x}}(y_2) = \Pr(Y_2 \le y_2 \mid \mathbf{X} = \mathbf{x}).$$

If $F_{1\mathbf{x}}$ and $F_{2\mathbf{x}}$ are continuous, then by Sklar's theorem (see, e.g., Sklar [22], Nelsen [17]) applied to the conditional probability distribution setting, there exists a unique copula $C_{\mathbf{x}}$ that links the conditional margins into the conditional joint distribution through the relation

$$H_{\mathbf{x}}(y_1, y_2) = C_{\mathbf{x}}\left\{F_{1\mathbf{x}}(y_1), F_{2\mathbf{x}}(y_2)\right\}.$$

The function $C_{\mathbf{x}}$ is called a *conditional copula*.

A first interest in this paper is to test whether the conditional copula $C_{\mathbf{x}}$ really depends on $\mathbf{x}$. More formally, we want to test the hypothesis

$$\mathcal{H}_0 : \forall_{\mathbf{x},\mathbf{x}' \in \mathbf{R}_{\mathbf{X}}} \; C_{\mathbf{x}} = C_{\mathbf{x}'} \tag{1}$$

versus the alternative

$$\mathcal{H}_A : \exists_{\mathbf{x},\mathbf{x}' \in \mathbf{R}_{\mathbf{X}}} \; C_{\mathbf{x}} \neq C_{\mathbf{x}'},$$

where $\mathbf{R}_{\mathbf{X}}$ denotes the domain of the covariate $\mathbf{X}$.

### 2.2. State-of-the-art

In what follows, suppose for a moment that we can observe $U_{1i} = F_{1\mathbf{X}_i}(Y_{1i})$, $U_{2i} = F_{2\mathbf{X}_i}(Y_{1i})$ from the conditional copula $C_{\mathbf{X}_i}$. If these observations are not available (i.e., $F_{1\mathbf{X}}$ and $F_{2\mathbf{X}}$ are unknown), then one needs to estimate these, via estimation of the conditional margins, and work with pseudo-observations, denoted by $(\widehat{U}_{1i}, \widehat{U}_{2i})$.

#### 2.2.1. Semiparametric approach of Acar et al. [3]

In Acar et al. [3] the conditional copula function is modeled as

$$C_{\mathbf{x}}(u_1, u_2) = C(u_1, u_2; \theta(\mathbf{x})), \tag{2}$$