Journal of Multivariate Analysis 141 (2015) 147–167

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

# Generalized additive models for conditional dependence structures

Thibault Vatter *, Valérie Chavez-Demoulin

*Faculty of Business and Economics, University of Lausanne, 1015 Lausanne, Switzerland*

ARTICLE INFO

*Article history:*
Received 20 October 2014
Available online 21 July 2015

*AMS subject classification:*
62F
62G
62H
62J

*Keywords:*
Conditional rank correlations
Copula
Penalized log-likelihood
Regression splines
Semiparametric modeling
Intraday financial returns

ABSTRACT

We develop a generalized additive modeling framework for taking into account the effect of predictors on the dependence structure between two variables. We consider dependence or concordance measures that are solely functions of the copula, because they contain no marginal information: rank correlation coefficients or tail-dependence coefficients represent natural choices. We propose a maximum penalized log-likelihood estimator, derive its $\sqrt{n}$-consistency and asymptotic normality, discuss details of the estimation procedure and the selection of the smoothing parameter. Finally, we present the results from a simulation study and apply the new methodology to a real dataset. Using intraday asset returns, we show that an intraday dependence pattern, due to the cyclical nature of market activity, is shaped similarly to the individual conditional second moments.



## 1. Introduction

Generalized additive models [22] are a natural extension of linear and generalized linear models. Built on roughness penalty smoothing, a generalized additive model (GAM) is a flexible data analysis tool in a traditionally univariate context. Consider the following research questions, however:

- There are relationships between a population's life expectancy and the country's GDP, as well as between male and female life expectancy in a given country. Can we measure the effect of the GDP on the later while controlling for the former?
- Cellular biologists study how predictor genes coordinate the expression of target genes. Can we further quantify how the association between the targets depends on the predictors?
- Volatilities of intraday asset returns show periodicities due to the cyclical nature of market activity and macroeconomic news releases. Is this also true for their dependence structure?

To obtain a statistically sound answer, we need to extend GAMs to the dependence structure between random variables. In this context, we distinguish between two closely related concepts, namely dependence and concordance. For instance, Pearson's correlation is often used as a measure of concordance and its absolute value as a measure of dependence. It detects linear relationships between variables, but it depends on their margins and lacks robustness to outliers. Borrowing from [27], two desirable properties of a dependence (respectively concordance) measure are:

* Correspondence to: University of Lausanne, Anthropole Building, Office 3016, 1015 Lausanne, Switzerland.
*E-mail addresses:* thibault.vatter@unil.ch (T. Vatter), valerie.chavez@unil.ch (V. Chavez-Demoulin).

http://dx.doi.org/10.1016/j.jmva.2015.07.003
0047-259X/

- invariance to monotone increasing transformations of the margins (up to a sign change if one of the transformations is monotone decreasing);
- the existence and uniqueness of a minimum (a zero), which is attained whenever the variables are independent.

Essentially, the first property states that such a measure should depend on the copula only. This is the case for rank correlation coefficients (concordance) or tail-dependence coefficients (dependence). Furthermore, convenient mappings between such measures and the parameters of common copulas often exist. As such, many conditional dependence and concordance measures, although unobservable, can be modeled directly.

While copulas are well studied [24,27], their formal developments to conditional distributions have rather recent origins. Patton [30] first extended the standard theory by imposing a mutual conditioning algebra for each margin and the copula. Fermanian and Wegkamp [13] relax this mutual algebra requirement and develop the concept of pseudo-copulas under strong mixing assumptions.

However, the investigation of modeling with the dependence structure as function of covariates has only been recently explored. As often in statistics, it is useful to distinguish between distribution-free and parametric methods. In [17], the authors suggest two kernel-based estimators of conditional copulas and corresponding conditional association measures. While their estimators are useful as descriptive statistics, their framework is limited to a single covariate without formal inference. On the parametric side, Acar et al. [2] consider a copula parameter that varies with a single covariate. The authors estimate their model using the concept of local likelihood and further suggest a testing framework in [3]. In [10], the authors develop Bayesian inference tools for a bivariate copula, conditional on a single covariate, coupling mixed or continuous outcomes. It is extended to multiple covariates in the continuous case by Sabeti et al. [32].

Compared to existing methods, the framework that we develop in this paper benefits directly from the complete GAM toolbox. We consider models where the dependence structure varies with an arbitrary set of covariates in a parametric, nonparametric or semiparametric way.

The structure of the paper is as follows: In Section 2, we develop the theoretical framework of generalized additive models for the dependence structure. We present the general model for the conditional dependence or concordance measure in Section 2.1. In Section 2.2, we state some asymptotic properties of the penalized log-likelihood estimator, namely its $\sqrt{n}$-consistency and asymptotic normality, assuming either known or unknown margins. In Section 2.3, we recast the penalized likelihood estimation as an iteratively reweighted generalized ridge regression. We close our theoretical considerations in Section 2.4, by discussing a measure of the penalized model's effective dimension and the selection of smoothing parameters. In Section 3, we present a simulation study and an application using a real dataset. We analyze the results of the simulation study in Section 3.1. We study the cross-sectional dynamics of intraday asset returns in Section 3.2. We conclude and suggest directions for further work in Section 4.

## 2. Generalized additive models for conditional dependence structures

In this section, we detail the approach to model the dependence structure between two random variables as a function of an arbitrary set of exogenous predictors (covariates).

To set up the notations, we use uppercase (boldface) letters for scalar random variables (random vectors and matrices) and lowercase (boldface) letters for scalars (vectors and matrices). We differentiate scalar functions with identical names by their arguments, and similarly for boldfaced vectors and matrices functions. We use vectors columns, $\|\cdot\|^2$ for the Euclidean norm, subscripts for the elements of a given matrix or vector and superscripts for either independent copies of a random quantity or realized observations. We denote real intervals (or Cartesian products thereof) by double-struck capital letters, except for the usual $\mathbb{N}, \mathbb{Z}, \mathbb{R}$, etc. For $k, l \in \mathbb{N}$ and $\mathbb{W} \subseteq \mathbb{R}^l$, we use $C^k(\mathbb{W})$ for the space of functions with $k$ continuous (partial) derivatives on the interior of $\mathbb{W}$.

Let $\mathbf{Y} \in \mathbb{Y} \subseteq \mathbb{R}^2$ be the random vector (responses) of interest, $\mathbf{X} \in \mathbb{X} \subseteq \mathbb{R}^q$ be a vector of $q$ covariates (predictors). For $\mathbf{y} \in \mathbb{Y}, \mathbf{x} \in \mathbb{X}$ and $i \in \{1, 2\}$, we denote by $F_{\mathbf{Y}_i|\mathbf{X}}(\mathbf{y}_i \mid \mathbf{x}) = P(\mathbf{Y}_i \leq \mathbf{y}_i \mid \mathbf{X} = \mathbf{x})$ the conditional margins and by $\mathbf{U}$ the random vector of conditional probability integral transforms with $\mathbf{U}_i = F_{\mathbf{Y}_i|\mathbf{X}}(\mathbf{Y}_i \mid \mathbf{X})$. Assume further that all variables are continuous and have strictly positive densities. In what follows, we rely on the conditional equivalent to Sklar's theorem [33,30]: for all $\mathbf{x} \in \mathbb{X}$, there exists a unique conditional copula $C(\cdot \mid \mathbf{x})$ which is the conditional distribution of $\mathbf{U} \mid \mathbf{X} = \mathbf{x}$. In other words, for $\mathbf{u} \in [0, 1]^2$, we have that

$$C(\mathbf{u} \mid \mathbf{x}) = P(\mathbf{U} \leq \mathbf{u} \mid \mathbf{X} = \mathbf{x}) = F_{\mathbf{Y}|\mathbf{X}} \left\{ F^{-1}_{\mathbf{Y}_1|\mathbf{X}}(\mathbf{u}_1 \mid \mathbf{x}), F^{-1}_{\mathbf{Y}_2|\mathbf{X}}(\mathbf{u}_2 \mid \mathbf{x}) \mid \mathbf{x} \right\}. \quad (1)$$

**Remark.** In [30], the conditioning vector is the same for the two margins and copula. In time series, conditioning algebras are usually augmented with past observations. Therefore, the concept of conditional copulas can be rather restrictive (see [13]). However, conditional copulas and exogeneity of the predictors are sufficient to develop a regression-like theory for the dependence structure.