



Nonparametric bivariate copula estimation based on shape-restricted support vector regression

Yongqiao Wang^{a,*}, He Ni^a, Shouyang Wang^b

^a School of Finance, Zhejiang Gongshang University, Hangzhou 310018, Zhejiang, China

^b Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100080, China

ARTICLE INFO

Article history:

Received 8 December 2011
Received in revised form 9 April 2012
Accepted 8 May 2012
Available online 17 May 2012

Keywords:

Support vector regression
Copula
Nonparametric estimation
Dependence
Shape-restriction

ABSTRACT

Copula has become a standard tool in describing dependent relations between random variables. This paper proposes a nonparametric bivariate copula estimation method based on shape-restricted ϵ -support vector regression (ϵ -SVR). This method explicitly supplements the classical ϵ -SVR with constraints related to three shape restrictions: grounded, marginal and 2-increasing, which are the necessary and sufficient conditions for a bivariate function to be a copula. This nonparametric method can be reformulated to a convex quadratic programming, which is computationally tractable. Experiments on both five artificial data sets and three international stock indexes clearly showed that it could achieve significantly better performance than common parametric models and kernel smoother.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Since [1], copula has become a standard tool of dependence modeling in multivariate statistical analysis [2,3]. Copula summarizes all dependence information between random variables and separates marginal components of a joint distribution from its distribution structure. In the last 20 years, copula has been widely applied in a variety of areas, such as engineering, finance, insurances, economics, etc. See, for example, the recent monographs [4–6] (and references therein).

In what follows we focus on bivariate copula only for simplicity. Let X_1 and X_2 be two continuous random variables of interest with joint distribution function H and marginal distributions F_1 and F_2 , respectively. When the marginal distributions F_1 and F_2 are continuous, Sklar's theorem [1] ensures that there exists a unique copula function $C: [0, 1]^2 \rightarrow [0, 1]$, which satisfies

$$H(x_1, x_2) = C(F_1(x_1), F_2(x_2)), \quad \forall (x_1, x_2) \in \mathbb{R}^2. \quad (1)$$

By copula technique, the estimation of the joint distribution H can be separated by two steps: marginal distributions construction and copula estimation. Assume we should estimate H based on the i.i.d observations from the distribution $H: (x_{i1}, x_{i2}) \in \mathbb{R}^2, i = 1, \dots, T$. The first step is to estimate the two marginal distributions, F_1 and F_2 , based

on the data sets $x_{i1}, i = 1, \dots, T$, and $x_{i2}, i = 1, \dots, T$, respectively. Assume \hat{F}_1 and \hat{F}_2 are the estimated cumulative distributions of X_1 and X_2 , respectively. The second step is to estimate the copula function C based on the data set $(\hat{F}_1(x_{i1}), \hat{F}_2(x_{i2})), i = 1, \dots, T$. Let \hat{C} denote the estimation of C , then the estimation for H is

$$\hat{H}(x_1, x_2) = \hat{C}(\hat{F}_1(x_1), \hat{F}_2(x_2)), \quad \forall (x_1, x_2) \in \mathbb{R}^2. \quad (2)$$

Due to the fact that the univariate cumulative distribution estimation problem has been extensively researched in statistics, this paper focuses on copula estimation and neglects the first step.

Most often copula is obtained by Maximum Likelihood Estimation (MLE). Assume C comes from a copula family indexed by a real-valued parameter θ . The MLE for θ can be obtained by

$$\hat{\theta} = \arg \max_{\theta} \prod_{i=1}^T C(\hat{F}_1(x_{i1}), \hat{F}_2(x_{i2}); \theta). \quad (3)$$

Commonly used copula families in practice are:

- Gaussian

$$C(u_1, u_2; \rho) = \Phi_{\rho}(\Phi^{-1}(u_1), \Phi^{-1}(u_2)) \quad (4)$$

where $\rho \in (-1, 1)$, $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, $\Phi_{\rho}(\cdot, \cdot)$ is the bivariate normal distribution function with marginal distributions standard normal and correlation ρ .

* Corresponding author. Tel.: +86 571 28877720; fax: +86 571 28877705.

E-mail addresses: wangyq@zjgsu.edu.cn (Y. Wang), nihe@zjgsu.edu.cn (H. Ni), swang@iss.ac.cn (S. Wang).

- Student's t

$$C(u_1, u_2; \rho, v) = t_{\rho, v}(t_v^{-1}(u_1), t_v^{-1}(u_2)) \quad (5)$$

where $\rho \in (-1, 1)$, $v \in \mathbb{N}$, $t_v(\cdot)$ is the Student's t distribution function with degree of freedom v , $t_{\rho, v}$ is the bivariate Student's t distribution function with correlation ρ and degree of freedom v .

- Clayton

$$C(u_1, u_2; \alpha) = \max \left[(u_1^{-\alpha} + u_2^{-\alpha} - 1)^{-1/\alpha}, 0 \right] \quad (6)$$

where $\alpha \in [-1, 0) \cup (0, +\infty)$.

- Frank

$$C(u_1, u_2; \alpha) = -\frac{1}{\alpha} \ln \left(1 + \frac{(e^{-\alpha u_1} - 1)(e^{-\alpha u_2} - 1)}{e^{-\alpha} - 1} \right) \quad (7)$$

where $\alpha \in \mathbb{R}$, $\alpha \neq 0$.

- Gumbel

$$C(u_1, u_2; \alpha) = \exp\{-[(-\ln u_1)^\alpha + (-\ln u_2)^\alpha]^{1/\alpha}\} \quad (8)$$

where $\alpha \in [1, +\infty)$.

Though the MLE method has tractable computational complexity and nice asymptotic statistical properties, its performance severely depends on its guess on the copula family. For example, financial risk will be greatly underestimated if dependence between financial returns is assumed to be the Gaussian copula, which has zero lower tail dependence. For example, the Gaussian copula assumption in Collateralized Debt Obligations (CDOs) pricing [7] has been criticized to be one of the key reasons behind the global 2008–2009 Subprime Crisis.¹

To overcome model specification error of MLE, some nonparametric methods have been proposed to estimate the underlying copula. Empirical copula, introduced by [8,9], extends the idea of empirical distribution for univariate variable to copula as follows

$$\hat{C}(u_1, u_2) = \frac{1}{T} \sum_{j=1}^T I\{\hat{F}_1(x_{1j}) \leq u_1, \hat{F}_2(x_{2j}) \leq u_2\}. \quad (9)$$

Because empirical copula is highly discontinuous and wiggly, several methods have been proposed to smooth empirical copula, including kernel smoother [10], spline [11–14] and wavelets [15–17]. However, nonparametric copula estimation with explicit 2-increasing shape restriction can hardly be found.

One obvious obstacle of nonparametric methods for copula estimation is that their estimators are often invalid to be a copula function. The necessary and sufficient conditions for a bivariate function $C: [0, 1]^2 \rightarrow [0, 1]$ to be a copula are.

1. grounded: $\forall u_1, u_2 \in [0, 1]$, C satisfies

$$C(u_1, 0) = 0, \quad C(0, u_2) = 0. \quad (10)$$

2. marginal: $\forall u_1, u_2 \in [0, 1]$, C satisfies

$$C(u_1, 1) = u_1, \quad C(1, u_2) = u_2. \quad (11)$$

3. 2-increasing: $\forall u_1, u_2, v_1, v_2 \in [0, 1]$ such that $u_1 \leq v_1$ and $u_2 \leq v_2$, C satisfies

$$C(v_1, v_2) - C(v_1, u_2) - C(u_1, v_2) + C(u_1, u_2) \geq 0. \quad (12)$$

Different from MLE, whose assumed functional form ensures its estimator to be valid, nonparametric methods should explicitly append constraints related to the above three shape restrictions.

These shape restrictions can be regarded as prior knowledge, which can be exploited to improve fitting performance. Its contribution is especially obvious when the size of the training data is small, in which case the common nonparametric estimators have great possibility of violating these shape restrictions.

Shape-restricted regression dates back to the literature on isotonic regression [18,19]. There exists a large literature on the problem of estimating monotone, concave or convex regression functions. Because some of these estimators are not smooth, many efforts have been devoted to the search of a simple, smooth and efficient estimation of a shape-restricted regression function. Typical applications of shape-restricted regression include the study of dose response experiments in medicine and the study of utility functions, product functions, profit functions and cost functions in microeconomics. Shape-restricted regression has been incorporated into wavelets [20], spline [21,22] and Bernstein polynomials [23,24] etc.

In machine learning area, shape restrictions also have been incorporated into support vector regression, such as monotone least squares support vector regression [25], monotone kernel quantile regression [26], boundary derivatives kernel regression [27], convex support vector regression [28]. Among them, [28] can tackle multivariate regression, while others are applicable only to univariate regression. [29] solved a support vector regression when one has some prior knowledge. However, it can only handle monotonicity or concavity of certain points.

The main contribution of this paper is to propose a nonparametric copula estimation based on shape-restricted ϵ -support vector regression. In this paper, the method is called Kernel Copula Regression (KCR). In this method the estimator is obtained by fitting a bivariate function with ϵ -support vector regression based on samples (\mathbf{u}_i, C_i) , $i = 1, \dots, T$, where $\mathbf{u}_i = (\hat{F}_1(x_{1i}), \hat{F}_2(x_{2i}))$ and C_i is its corresponding empirical copula (9). To make the estimator satisfy the three shape restrictions (10)–(12), additional constraints are explicitly imposed on the classical ϵ -support vector regression [30,31]. To make the KCR estimator grounded and marginal, equality constraints are imposed on points that are spaced evenly on the four boundaries. To make it 2-increasing, nonnegative second-order mixed derivatives constraints are imposed on equidistant grid points of $[0, 1]^2$. We expect to obtain better estimation by spanning a network of grounded, marginal and 2-increasing points.

The advantages of KCR are multifold. First it is a nonparametric estimation method, which can handle any complex dependence structure. Second this estimator is smooth, which is an obvious superiority over the empirical copula estimator. Third this estimator satisfies the three shape-restrictions of copula: grounded, marginal and 2-increasing, provided that sufficient constraints related to shape restrictions are appended. Fourth its training involves a convex quadratic programming, which is computationally tractable.

The structure of the paper is as follows. Section 2 introduces how to apply the classical ϵ -support vector regression (ϵ -SVR) to estimate copula. A toy data set is used to show its qualitative shortcomings. Section 3 presents the novel nonparametric method for copula estimation. We will detail how to impose additional constraints related to the three shape restrictions and how to transform it to a convex quadratic programming. Section 4 presents the numerical comparison results of KCR with other state-of-the-art methods. The paper is concluded in Section 5.

2. ϵ -Support vector regression estimator

This section discusses how to apply the classical ϵ -support vector regression to copula estimation based on observations (x_{1i}, x_{2i}) ,

¹ Please refer to the two reports: Jones Sam, The formula that felled Wall Street, Financial Times, April 24, 2009; Salmon Felix, Recipe for disaster: the formula that killed Wall Street, Wired Magazine, March 2009.

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات