Insurance: Mathematics and Economics 77 (2017) 38–48

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

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

Remarks on composite Bernstein copula and its application to credit risk analysis

Nan Guo [a], Fang Wang [b,*], Jingping Yang [c]

[a] Department of Financial Mathematics, School of Mathematical Sciences, Peking University, Beijing, 100871, China
[b] School of Mathematical Sciences, Capital Normal University, Beijing, 100048, China
[c] LMEQF, Department of Financial Mathematics and Center for Statistical Sciences, Peking University, Beijing, 100871, China

ARTICLE INFO

Article history:
Received April 2017
Received in revised form August 2017
Accepted 20 August 2017
Available online 31 August 2017

Keywords:
Composite Bernstein copula
EM algorithm
Probabilistic structure
Credit risk analysis
Baker's distribution

ABSTRACT

The composite Bernstein copula (CBC) (Yang et al., 2015) is a copula function generated from a composition of two copulas. This paper first shows that some well-known copulas belong to the CBC family with desirable properties. An EM algorithm for estimating the CBC is proposed, and it is applied for a real dataset to show the fitting result of the CBC in modeling dependence. The probabilistic structure for the CBC family is presented, which is useful for generating random numbers from the CBC. Finally, the probabilistic structure of the CBC is applied to credit risk analysis of collateralized debt obligations to show its advantage in empirical analysis.

1. Introduction

A copula function is a multivariate distribution function with uniform [0,1] marginal distributions. It is known that for each joint distribution function H with marginal distributions $F_1, \dots, F_n$, there exists an n-dimensional copula C such that

$$H(x_1, x_2, \dots, x_n) = C(F_1(x_1), F_2(x_2), \dots, F_n(x_n)).$$

The copula function C is unique when the marginal distributions are continuous. See Nelsen (2006) for detailed introduction on copula functions. Now copula functions have been widely used in insurance and finance. Please see Shi and Frees (2011) for copula methods in loss reserving of non-life insurance and Avanzi et al. (2011) for modeling dependence in insurance claims using Lévy copulas. See also Cherubini et al. (2004) for copula methods in finance and McNeil et al. (2015) for copula methods in quantitative risk management.

Constructions of copula functions have become an important research area for the past few years. The Bernstein copula (BC) is a family of copula functions introduced by Sancetta and Satchell (2004) for approximating copula functions. Following Sancetta and Satchell (2004)'s work, Janssen et al. (2012), Baker (2008), Dou et al. (2013), Dou et al. (2016), Sancetta (2007) and Weiss and Scheffer (2012) discussed the BC from probabilistic and statistical perspectives, and Diers et al. (2012) and Tavin (2015) focused on the applications of the BC in non-life insurance and finance. Recently, Yang et al. (2015) introduced a new family of multivariate copulas, the composite Bernstein copula (CBC), generated from a composition of two copulas. For fixed $m \in \mathbb{N}$, let $F_{B(m,u)}$ be the binomial distribution function with parameter (m, u), $u \in [0, 1]$, and $F^{\leftarrow}_{B(m,u)}$ be the left-continuous inverse function of $F_{B(m,u)}$. Given n-copulas C and D and the positive integers m_i, $i = 1, \dots, n$, the CBC is defined as

$$C_{m_1,\dots,m_n}(u_1, \dots, u_n|C, D) := \mathbb{E}[C(\frac{F^{\leftarrow}_{B(m_1,u_1)}(U_1^D)}{m_1}, \dots, \frac{F^{\leftarrow}_{B(m_n,u_n)}(U_n^D)}{m_n})],$$

where $(U_1^D, \dots, U_n^D)$ is a random vector with distribution function $\overline{D}$, here $\overline{D}$ is the survival copula of the copula function D. And C is called the target copula and D is called the base copula. One interesting property for the CBC is that

$$C_{1,\dots,1}(\cdot|C, D) = D(\cdot), \quad \lim_{m_i \to \infty, i=1,\dots,n} C_{m_1,\dots,m_n}(\cdot|C, D) = C(\cdot). \tag{1}$$

Moreover, the CBC is able to capture the tail dependence, and it has a reproduction property for the three important dependence structures: comonotonicity, countermonotonicity and independence. The CBC can also incorporate both prior information and data into the statistical estimation. Please refer to Yang et al. (2015) for details.

In this paper, we discuss the CBC from its generality, its statistical estimation, its probability structure and its application in

* Corresponding author.
E-mail addresses: nan.guo@pku.edu.cn (N. Guo), fang72_wang@cnu.edu.cn (F. Wang), yangjp@math.pku.edu.cn (J. Yang).

http://dx.doi.org/10.1016/j.insmatheco.2017.08.007
0167-6687/

portfolio credit risk. We first present three equivalent statements for the CBC, then we illustrate that some well-known copulas, such as Bernstein copulas in Sancetta and Satchell (2004), Baker's distributions in Baker (2008), Baker's Type II BB distributions in Bayramoglu and Bayramoglu (Bairamov) (2014) and mixture copulas in Baker (2014), belong to the CBC family. An EM algorithm for estimating the CBC is proposed, and an analysis for the consomic mouse data (Dou et al., 2016) shows that the CBC is superior to the BC and the Gaussian copula for this dataset according to the Akaike information criterion (AIC). The probabilistic structure for the CBC family is presented, where the relationship between the CBC and its base and target copulas is described from probabilistic viewpoint. Finally, the probabilistic structure for the CBC family is applied in the modeling of portfolio credit risk to show its advantage.

The remainder of the paper is organized as follows. In Section 2, we present three equivalent statements for the CBC and illustrate that some well-known copulas belong to the CBC family. In Section 3, an EM algorithm is proposed for estimating the CBC. In Section 4, we give the probabilistic structure of the CBC and apply it to analyze Spearman's rho of the copula family. In Section 5, we apply the probabilistic structure of the CBC in portfolio credit risk analysis. The conclusions are drawn in Section 6. Some proofs are put in Appendices.

2. The families of the CBC

2.1. Basic introduction about the CBC

In the following, we will give some equivalent statements for the CBC, which illustrate the copula family from different viewpoints.

As we introduced before, for fixed $m \in \mathbb{N}$, let $F_{B(m,u)}$ be the binomial distribution function with parameter (m, u), $u \in [0, 1]$. For $j = 0, 1, \ldots, m-1$ and $u \in [0, 1]$, we denote $\overline{F}_{B(m,u)}(j) = 1 - F_{B(m,u)}(j)$. Then it holds that

$$\overline{F}_{B(m,u)}(j) = \sum_{i=j+1}^{m} \binom{m}{i} u^i (1-u)^{m-i} = mB_{j,m-1}(u), \tag{2}$$

where for $k = 0, 1, \ldots, m$,

$$b_{k,m}(t) = \binom{m}{k} t^k (1-t)^{m-k},\ t \in [0, 1]; \quad B_{k,m}(u) = \int_0^u b_{k,m}(t)dt.$$

It is also known that $\overline{F}_{B(m,u)}(j)$ is the distribution function of the $(j+1)$th smallest order statistic of m i.i.d. Uniform [0,1] random variables. For detailed introduction about $\overline{F}_{B(m,u)}(j)$ and order statistics, please see Reiss (2012).

Theorem 2.1. *For the given copula functions C and D, and the positive integers m_i, $i = 1, \ldots, n$, the following statements are equivalent:*

(1) *For a random vector $(U_1^D, \ldots, U_n^D)$ with the distribution function D,*

$$\begin{aligned} &C_{m_1,\ldots,m_n}(u_1, \ldots, u_n|C, D) \\ &= \mathbb{E}[C(\frac{F^{\leftarrow}_{B(m_1,u_1)}(U_1^D)}{m_1}, \ldots, \frac{F^{\leftarrow}_{B(m_n,u_n)}(U_n^D)}{m_n})]. \end{aligned} \tag{3}$$

(2) *For a random vector $(V_1, \ldots, V_n)$ with the distribution function C,*

$$\begin{aligned} &C_{m_1,\ldots,m_n}(u_1, \ldots, u_n|C, D) \\ &= \mathbb{E}[D(\overline{F}_{B(m_1,u_1)}(m_1V_1), \ldots, \overline{F}_{B(m_n,u_n)}(m_nV_n))]. \end{aligned} \tag{4}$$

(3) *For an n-discrete random vector $(K_1, \ldots, K_n)$ defined on the grids $\prod_{i=1}^n \{1, 2, \ldots, m_i\}$ with the uniform marginal laws,*

$$\begin{aligned} &C_{m_1,\ldots,m_n}(u_1, \ldots, u_n|C, D) \\ &= \sum_{k_1=1}^{m_1} \cdots \sum_{k_n=1}^{m_n} D(m_1B_{k_1-1,m_1-1}(u_1), \ldots, m_nB_{k_n-1,m_n-1}(u_n)) \\ &\quad \times \mathbb{P}(K_1 = k_1, \ldots, K_n = k_n), \end{aligned} \tag{5}$$

where $K_i = [m_iV_i] + 1$, $i = 1, \ldots, n$, and $(V_1, \ldots, V_n)$ is a random vector with the distribution C.

Proof. First we assume that (3) holds. Since $(V_1, \ldots, V_n)$ is a random vector with the distribution function C, we have

$$\begin{aligned} &C_{m_1,\ldots,m_n}(u_1, \ldots, u_n|C, D) \\ &= \mathbb{E}[\mathbb{P}(V_i \le \frac{F^{\leftarrow}_{B(m_i,u_i)}(U_i^D)}{m_i}, i = 1, \ldots, n|U_1^D, \ldots, U_n^D)] \\ &= \mathbb{E}[\mathbb{P}(F_{B(m_i,u_i)}(m_iV_i) \le U_i^D, i = 1, \ldots, n|V_1, \ldots, V_n)] \\ &= \mathbb{E}[D(\overline{F}_{B(m_1,u_1)}(m_1V_1), \ldots, \overline{F}_{B(m_n,u_n)}(m_nV_n))]. \end{aligned}$$

Thus (4) follows.

Similarly, from (4) we can prove that (3) holds.

Next we assume that (4) holds. Note that $(V_1, \ldots, V_n)$ has the distribution C, thus we have

$$\begin{aligned} &C_{m_1,\ldots,m_n}(u_1, \ldots, u_n|C, D) \\ &= \mathbb{E}[D(\overline{F}_{B(m_1,u_1)}(m_1V_1), \ldots, \overline{F}_{B(m_n,u_n)}(m_nV_n))] \\ &= \sum_{k_1=0}^{m_1-1} \cdots \sum_{k_n=0}^{m_n-1} D(\overline{F}_{B(m_1,u_1)}(k_1), \ldots, \overline{F}_{B(m_n,u_n)}(k_n)) \\ &\quad \times \mathbb{P}([m_1V_1] = k_1, \ldots, [m_nV_n] = k_n) \\ &= \sum_{k_1=1}^{m_1} \cdots \sum_{k_n=1}^{m_n} D(\overline{F}_{B(m_1,u_1)}(k_1-1), \ldots, \overline{F}_{B(m_n,u_n)}(k_n-1)) \\ &\quad \times \mathbb{P}([m_1V_1] = k_1-1, \ldots, [m_nV_n] = k_n-1). \end{aligned}$$

For $K_i = [m_iV_i] + 1$, $\mathbb{P}(K_i = k_i) = \mathbb{P}(k_i - 1 \le m_iV_i < k_i) = \frac{1}{m_i}$. Then

$$\begin{aligned} &C_{m_1,\ldots,m_n}(u_1, \ldots, u_n|C, D) \\ &= \sum_{k_1=1}^{m_1} \cdots \sum_{k_n=1}^{m_n} D(m_1B_{k_1-1,m_1-1}(u_1), \ldots, m_nB_{k_n-1,m_n-1}(u_n)) \\ &\quad \times \mathbb{P}(K_1 = k_1, \ldots, K_n = k_n), \end{aligned}$$

due to the fact that $\overline{F}_{B(m,u)}(k-1) = mB_{k-1,m-1}(u)$ by Eq. (2). Thus (5) follows.

Similarly, from (5) we can get (4). □

For the CBC in Theorem 2.1, C is called the target copula and D is called the base copula.

Remark 2.1. Theorem 2.1 presents three equivalent statements for the CBC. Either one of them can be applied for discussing the CBC and connecting the CBC with some known copula families in the literature:

(1) Expression (3) can be seen by adding noise on the target copula C. Note that we can express (3) as follows:

$$\begin{aligned} &C_{m_1,\ldots,m_n}(u_1, \ldots, u_n|C, D) \\ &= \mathbb{E}[C(u_1 + \frac{F^{\leftarrow}_{B(m_1,u_1)}(U_1^D)}{m_1} - u_1, \ldots, u_n + \frac{F^{\leftarrow}_{B(m_n,u_n)}(U_n^D)}{m_n} - u_n)]. \end{aligned}$$

Thus the random noise vector

$$(\frac{F^{\leftarrow}_{B(m_1,u_1)}(U_1^D)}{m_1} - u_1, \ldots, \frac{F^{\leftarrow}_{B(m_n,u_n)}(U_n^D)}{m_n} - u_n),$$