Composite Bernstein Copulas

Jingping Yang* Zhijin Chen† Fang Wang‡ Ruodu Wang§

December 10, 2014

Abstract

Copula function has been widely used in insurance and finance for modeling interdependency between risks. Inspired by the Bernstein copula (BC) put forward by Sancetta and Satchell (2004), we introduce a new class of multivariate copulas, the composite Bernstein copula (CBC), generated from a composition of two copulas. This new class of copula functions is able to capture tail dependence, and it has a reproduction property for the three important dependency structures: comonotonicity, countermonotonicity and independence. We introduce an estimation procedure based on the empirical composite Bernstein copula (ECBC) which incorporates both prior information and data into the estimation. Simulation studies and an empirical study on financial data illustrate the advantages of the ECBC estimation method, especially in capturing tail dependence.

Key-words: composite Bernstein copula; copula construction; tail dependence; non-parametric estimation.

---

*LMEQF, Department of Financial Mathematics, Peking University, Beijing, 100871, China. Email address: yangjp@math.pku.edu.cn

†Department of Financial Mathematics, School of Mathematical Sciences and Center for Statistical Sciences, Peking University, Beijing, 100871, China, Email address:bjzhijin@126.com

‡Corresponding author, School of Mathematical Sciences, Capital Normal University, Beijing, 100048, China, Email address: fang72_wang@cnu.edu.cn

§Department of Statistics and Actuarial Science, University of Waterloo, Waterloo N2L 3G1, Canada, Email address: wang@uwaterloo.ca

1 Introduction

A copula (or a copula function) is a multivariate distribution function with uniform [0,1] marginal distributions. Sklar's Theorem shows that for each joint distribution function H with marginal distributions $F_1, \ldots, F_n$, there exists an n-dimensional copula C such that

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

and the copula C is unique when the marginal distributions $F_1, \ldots, F_n$ are continuous. For a detailed introduction of copulas, the readers are referred to the introductory book Nelsen (2006). Copulas have been widely used in insurance and finance during the past few decades, and the related research is developing extensively in statistics, probability and other quantitative fields, including financial mathematics and actuarial science in particular. The reader is also referred to the books Cherubini et al (2004) for copula methods in finance and McNeil et al (2005) for copula methods in quantitative risk management.

New constructions of copulas have become an important research direction for the past few years, including recently introduced copula families such as the vine copula (Czado, 2009) and the nested copula (Hofert, 2009). In order to provide a general approach in statistical estimation and to study the properties of some parametric copulas with complicated forms, Sancetta and Satchell (2004) introduced a new family of copulas called the Bernstein copula (BC). For a given copula C, based on the Bernstein polynomials, the BC is defined as

$$C_B(u_1, \ldots, u_n) = \sum_{v_1=0}^{m_1} \cdots \sum_{v_n=0}^{m_n} C(\frac{v_1}{m_1}, \ldots, \frac{v_n}{m_n}) P_{v_1,m_1}(u_1) \ldots P_{v_n,m_n}(u_n), \tag{1}$$

where $P_{v_j,m_j}(u_j) := \binom{m_j}{v_j} u_j^{v_j}(1-u_j)^{m_j-v_j}$ and $m_1, \ldots, m_n$ are positive integers. More discussions on the motivation of the BC can be found in Sancetta and Satchell (2004).

Among many recent research papers on the Bernstein copula, the readers are referred to Janssen et al (2012), Baker (2008), Dou et al (2013), Sancetta (2007), and Weiβ and Scheffer (2012) from probabilistic and statistical perspectives, and Diers et al (2012) and Tavin (2012) from the perspective of applications in non-life insurance and finance.

Inspired by the BC, we will construct a new family of copulas. By looking at the BC from another prospective, a new construction will be revealed. Let $F_{Bin(m,u)}$ be the binomial distribution function with parameters (m,u), $m \in \mathbb{N}$, $u \in [0,1]$ and denote by $F^{-1}_{Bin(m,u)}$ the left-continuous inverse function of $F_{Bin(m,u)}$. Note that $P_{v_j,m_j}(u_j) = \mathbb{P}(F^{-1}_{Bin(m,u_j)}(U) = v_j)$ for a random variable $U \sim \mathrm{U}[0,1]$, thus the expression (1) can be written in another form as

$$C_B(u_1,\ldots,u_n) = \mathbb{E}\left[C\left(\frac{F^{-1}_{Bin(m_1,u_1)}(U_1)}{m_1},\ldots,\frac{F^{-1}_{Bin(m_n,u_n)}(U_n)}{m_n}\right)\right],$$

where $u_i \in [0,1]$, $i=1,\ldots,n$, and $U_1,\ldots,U_n$ are independent uniform [0,1] random variables. Using this representation, a natural generalization would be

$$C_{m_1,\ldots,m_n}(u_1,\ldots,u_n|C,D) := \mathbb{E}\left[C\left(\frac{F^{-1}_{Bin(m_1,u_1)}(U_1^D)}{m_1},\ldots,\frac{F^{-1}_{Bin(m_n,u_n)}(U_n^D)}{m_n}\right)\right] \quad (2)$$

for $(u_1,\ldots,u_n) \in [0,1]^n$, where D is a copula function, and $(U_1^D,\ldots,U_n^D)$ is a random vector with distribution $\bar{D}$, the survival copula of D. The reason why we use $\bar{D}$ instead of D will be revealed later in Section 2. When D is chosen as the independent copula, i.e.

$$D(u_1,u_2,\ldots,u_n) = \bar{D}(u_1,u_2,\ldots,u_n) = \prod_{i=1}^n u_i, \quad u_i \in [0,1],\ i=1,\ldots,n,$$

(2) becomes the Bernstein copula (1). Note that the above expression (2) involves two copula functions C and D. Here the copula function C is called the *target copula* and the copula D is called the *base copula*. (2) can be used to construct new family of copulas. For example, with a given target copula C, by choosing different copula functions D one obtains a family of copulas. The generalization (2) leads to interesting properties that are not shared by the BC, such as capturing tail dependence, as explained later.

This paper will focus on the function (2). First we will prove that for each copula function D and given positive integers $m_1, \ldots, m_n$, the function $C_{m_1,\ldots,m_n}(u_1, \ldots, u_n|C, D)$ is a copula function. The function is called a *composite Bernstein copula* (CBC) since it is based on a composition of the target copula C and the base copula D. The properties of CBC for fixed $m_1, \ldots, m_n$ and for $m_i \to \infty,\ i = 1, \ldots, n$ will be discussed. It will be shown that the CBC converges to the target copula C as $m_i \to \infty,\ i = 1, \ldots, n$, regardless of the base copula D. We will also prove that for finite $m_1, \ldots, m_n$ the CBC is equal to the target copula with some special choices of target copulas and base copulas, such as Fréchet upper copula $M(u_1, \ldots, u_n) = \min\{u_i, i = 1, \ldots, n\},\ u_1, \ldots, u_n \in [0, 1]$, the independent copula $\Pi(u_1, \ldots, u_n) = \prod_{i=1}^n u_i,\ u_1, \ldots, u_n \in [0, 1]$ and the bivariate Fréchet lower copula $W(u, v) = \max\{u + v - 1, 0\},\ u, v \in [0, 1]$. The above reproduction property is very important for application in insurance and finance, since Fréchet upper copula, the independent copula and the bivariate Fréchet lower copula corresponds to the three important dependency structures in insurance: comonotonicity, independence and countermonotonicity (Dhaene et al, 2002a, 2002b).

As pointed out in Sancetta and Satchell (2004), a limitation of the Bernstein copula is that it fails to capture extreme tail behavior, a relevant and challenging issue in insurance and finance (Donnelly and Embrechts, 2010). Fortunately, the CBC allows us to exhibit the tail dependence by choosing proper base copulas D. We will show that the tail dependence coefficient of CBC is given by a combination of the tail dependence coefficient of the base copula and that of the target copula.

Based on CBC, we will also provide a copula estimation method using the empirical CBC (ECBC). The new method is a flexible non-parametric estimation as it incorporates both prior information and data. A simulation study will show how the choices of base copula affect the estimation results and an empirical study for financial data highlights the features of the new method.

The rest of the paper is organized as follows. In Section 2, we define the CBC and discuss the theoretical properties of CBC, concerning monotonicity, continuity, symmetry, reproduction and tail dependence. In Section 3, we introduce the ECBC, provide an estimation method based on the ECBC and show its asymptotic properties. Simulation studies and a real data analysis are provided to show the advantage of the new estimation method proposed in Section 4. In Section 5, we draw a conclusion. Some proofs are put in the Appendix.

2 General Theory of Composite Bernstein Copulas

Throughout, let $F_{Bin(m,u)}$ be the binomial distribution function with parameter (m, u), $m \in \mathbb{N}$, $u \in [0, 1]$ and $F^{-1}_{Bin(m,u)}$ be the left-continuous inverse function of $F_{Bin(m,u)}$, that is, $F^{-1}_{Bin(m,u)}(v) := \inf\{x \in \mathbb{R} : F_{Bin(m,u)}(x) \geq v\}$, $v \in [0, 1]$. Moreover, let $N_{m,u}$ be a binomial random variable with distribution $F_{Bin(m,u)}$. For any copula C, its survival copula is denoted as $\bar{C}$:

$$\bar{C}(u_1, \ldots, u_n) = \mathbb{P}(1 - V_i \leq u_i,\ i = 1, \ldots, n),$$

where $(V_1, \ldots, V_n)$ is a random vector with distribution function C.

2.1 Definition of the composite Bernstein copula

For a given n-copula C, by incorporating the information of another n-copula D, with the given positive integers $m_i,\ i = 1, \ldots, n$, we can construct a new function $C_{m_1,\ldots,m_n}(u_1, \ldots, u_n|C, D)$, $u_i \in [0, 1],\ i = 1, \ldots, n$ as follows,

$$\begin{aligned}
&C_{m_1,\ldots,m_n}(u_1, \ldots, u_n|C, D)\\
=&\mathbb{E}\left[C\left(\frac{F^{-1}_{Bin(m_1,u_1)}(U_1^D)}{m_1}, \ldots, \frac{F^{-1}_{Bin(m_n,u_n)}(U_n^D)}{m_n}\right)\right]\\
=&\sum_{l_1=0}^{m_1}\cdots\sum_{l_n=0}^{m_n} C(\frac{l_1}{m_1}, \ldots, \frac{l_n}{m_n})\mathbb{P}(F^{-1}_{Bin(m_1,u_1)}(U_1^D) = l_1, \ldots, F^{-1}_{Bin(m_n,u_n)}(U_n^D) = l_n), \quad (3)
\end{aligned}$$

where $(U_1^D, \ldots, U_n^D)$ is a random vector with distribution function $\bar{D}$. As mentioned in the introduction, (3) can be seen as a generalization of the Bernstein copula, with possibly different features.

The function $C_{m_1,\ldots,m_n}(u_1, \ldots, u_n|C, D)$ can also be written in an alternative form. By letting $(V_1, \ldots, V_n)$ be a random vector with distribution function C, we have

$$\begin{aligned}
& C_{m_1,\ldots,m_n}(u_1, \ldots, u_n|C, D) \\
= \;& \mathbb{E}\left[\mathbb{P}\left(V_i \leq \frac{F^{-1}_{Bin(m_i,u_i)}(U_i^D)}{m_i}, i = 1, \ldots, n|U_1^D, \ldots, U_n^D\right)\right] \\
= \;& \mathbb{E}[\mathbb{P}(F_{Bin(m_i,u_i)}(m_i V_i) \leq U_i^D, i = 1, \ldots, n|V_1, \ldots, V_n)] \\
= \;& \mathbb{E}[D(1 - F_{Bin(m_1,u_1)}(m_1 V_1), \ldots, 1 - F_{Bin(m_n,u_n)}(m_n V_n))]. \qquad (4)
\end{aligned}$$

Note that (3) and (4) are equivalent. Throughout the paper, in different places we will use either (3) or (4), whichever is more convenient.

Remark 2.1. We can see that as long as C is a distribution function, (3) is properly defined and (4) holds. Definition 4 will be used in Section 3 for the estimation purpose, where C is replaced by the empirical copula, which is not a copula function in general.

We will first show that $C_{m_1,\ldots,m_n}(u_1, \ldots, u_n|C, D)$ is a copula function with nice properties as long as C, D are copulas.

Theorem 2.1. *Suppose C and D are two n-copulas. Then the following holds:*

(i) $C_{m_1,\ldots,m_n}(u_1, \ldots, u_n|C, D)$, $u_1, \ldots, u_n \in [0, 1]$ *is a copula function.*

(ii) $C_{1,\ldots,1}(u_1, \ldots, u_n|C, D) = D(u_1, \ldots, u_n)$ *for* $u_1, \ldots, u_n \in [0, 1]$.

(iii) As $\underline{m} := \min\{m_1, \ldots, m_n\} \to \infty$, *for* $u_1, \ldots, u_n \in [0, 1]$ *we have that*

$$C_{m_1,\ldots,m_n}(u_1, \ldots, u_n|C, D) \to C(u_1, \ldots, u_n) \qquad (5)$$

uniformly and the convergence rate is bounded by

$$|C_{m_1,\ldots,m_n}(u_1, \ldots, u_n|C, D) - C(u_1, \ldots, u_n)| \leq \sum_{i=1}^{n} \sqrt{\frac{u_i(1 - u_i)}{m_i}}. \qquad (6)$$

(iv) $C_{m_1,\ldots,m_n}(\cdot|C,D)$ *admits a density on* $[0,1]^n$ *if* D *admits a density on* $[0,1]^n$.

Proof. (i) Let $1 \geq u_{2,i} \geq u_{1,i} \geq 0,\ i = 1,\ldots,n$. Then from (4) we know that for $l_i = 0$ or $l_i = 1,\ i = 1,2,\ldots,n$,

$$\begin{aligned}
& C_{m_1,\ldots,m_n}(u_{1,1} + l_1(u_{2,1} - u_{1,1}),\ldots,u_{1,n} + l_n(u_{2,n} - u_{1,n})|C,D) \\
= \ & \mathbb{P}(1 - U_1 \leq 1 - F_{Bin(m_1,u_{1,1}+l_1(u_{2,1}-u_{1,1}))}(m_1V_1),\ldots, \\
& 1 - U_n \leq 1 - F_{Bin(m_n,u_{1,n}+l_n(u_{2,n}-u_{1,n}))}(m_nV_n)).
\end{aligned}$$

Thus

$$\begin{aligned}
& \sum_{l_1=0}^{1}\cdots\sum_{l_n=0}^{1}(-1)^{l_1+\cdots+l_n} \\
& \times C_{m_1,\ldots,m_n}(u_{1,1} + l_1(u_{2,1} - u_{1,1}),\ldots,u_{1,n} + l_n(u_{2,n} - u_{1,n})|C,D) \\
= \ & \mathbb{E}[\sum_{l_1=0}^{1}\cdots\sum_{l_n=0}^{1}(-1)^{l_1+\cdots+l_n}\mathbb{P}(1 - U_1 \leq 1 - F_{Bin(m_1,u_{1,1}+l_1(u_{2,1}-u_{1,1}))}(m_1V_1),\ldots, \\
& 1 - U_n \leq 1 - F_{Bin(m_n,u_{1,n}+l_n(u_{2,n}-u_{1,n}))}(m_nV_n))] \\
= \ & \mathbb{P}(1 - F_{Bin(m_1,u_{1,1})}(m_1V_1) \leq 1 - U_1^D \leq 1 - F_{Bin(m_1,u_{1,2})}(m_1V_1),\ldots, \\
& 1 - F_{Bin(m_n,u_{n,1})}(m_nV_n) \leq 1 - U_n^D \leq 1 - F_{Bin(m_n,u_{n,2})}(m_nV_n)) \geq 0,
\end{aligned}$$

due to the property that for fixed k the function $F_{Bin(m_i,x)}(k)$ is decreasing about x. Also, we can easily verify that

$$\begin{aligned}
& C_{m_1,\ldots,m_i,\ldots,m_n}(1,\ldots,1,u_i,1,\ldots,1|C,D) \\
= \ & \mathbb{E}[C(1,\ldots,1,\frac{F^{-1}_{Bin(m_i,u_i)}(U_i^D)}{m_i},1,\ldots,1)] = \mathbb{E}\left[\frac{F^{-1}_{Bin(m_i,u_i)}(U_i^D)}{m_i}\right] = u_i.
\end{aligned}$$

Thus $C_{m_1,\ldots,m_n}(u_1,\ldots,u_n|C,D)$ is a copula function.

(ii) Using (4) we have

$$\begin{aligned}
C_{1,\ldots,1}(u_1,\ldots,u_n|C,D) &= \mathbb{E}[D(1 - F_{Bin(1,u_1)}(V_1),\ldots,1 - F_{Bin(1,u_n)}(V_n))] \\
&= D(u_1,\ldots,u_n).
\end{aligned}$$

(iii) Using the definition (3) one can verify that

$$
\begin{aligned}
& |C_{m_1,\ldots,m_n}(u_1,\ldots,u_n|C,D) - C(u_1,\ldots,u_n)| \\
\leq\ & \mathbb{E}|C(\frac{F^{-1}_{Bin(m_1,u_1)}(U_1^D)}{m_1},\ldots,\frac{F^{-1}_{Bin(m_n,u_n)}(U_n^D)}{m_n}) - C(u_1,\ldots,u_n)| \\
\leq\ & \sum_{i=1}^{n}\mathbb{E}|\frac{F^{-1}_{Bin(m_i,u_i)}(U_i^D)}{m_i} - u_i| \\
\leq\ & \sum_{i=1}^{n}\frac{\sqrt{\mathrm{Var}(F^{-1}_{Bin(m_i,u_i)}(U_i^D))}}{m_i} = \sum_{i=1}^{n}\sqrt{\frac{u_i(1-u_i)}{m_i}}.
\end{aligned}
$$

Thus the inequality (6) follows. (5) is implied by (6).

(iv) Note that for fixed $v \in (0,1)$ the function $F_{Bin(m,u)}(mv)$ is differentiable with respect to $u \in [0,1]$. Thus, if D admits a bounded density, then we know that for each $v_1,\ldots,v_n \in (0,1)$, the function $D(1-F_{Bin(m_1,u_1)}(m_1v_1),\ldots,1-F_{Bin(m_n,u_n)}(m_nv_n))$ has a bounded $\frac{\partial^n}{\partial u_1\ldots\partial u_n}$ derivative for $u_1,\ldots,u_n \in [0,1]$. Thus, by (4) we can see that $C(\cdot|C,D)$ also has a bounded density.

□

The copula $C_{m_1,\ldots,m_n}(u_1,\ldots,u_n|C,D)$ defined in (3) becomes the Bernstein copula when D is the independent copula. Hence, in this paper, we call $C_{m_1,\ldots,m_n}(u_1,\ldots,u_n|C,D)$ a composite Berstein copula (CBC), as a generalization of the Bernstein copula. By Theorem 2.1, the copula function $C_{m_1,\ldots,m_n}(\cdot|C,D)$ is close to C as $\min\{m_i, i=1,\ldots,n\} \to \infty$, hence the copula $C_{m_1,\ldots,m_n}(\cdot|C,D)$ can be used to approximate the copula C, as mentioned in Sancetta and Satchell (2004). On the other hand, when $m_1,\ldots,m_n$ are close to 1, the defined CBC $C_{m_1,\ldots,m_n}(\cdot|C,D)$ is close to the copula D. For the above reasons, we call C a *target copula* and D a *base copula*.

Remark 2.2. By the proof of Theorem 2.1 we know that even if C is not continuous, $C_{m_1,\ldots,m_n}(\ldots|C,D)$ is continuous as long as D is continuous. This would provide a good tool for density estimation. From the proof of the theorem we can also find that the marginal density of $C_{m_1,\ldots,m_n}(\cdot|C,D)$ exists if the corresponding marginal density of D

exists. On the contrary, since the function $F^{-1}_{Bin(m,u)}(v)$ in (3) is not continuous with respect to u, the condition that C admits a density is not sufficient for $C_{m_1,\ldots,m_n}(\cdot|C,D)$ to admit a density.

The following proposition shows that for a CBC with the base copula D and the target copula C, every marginal distribution of the CBC can also be expressed as a CBC, where the corresponding base copula and target copula can be chosen as the marginal distribution of the base copula D and the target copula C. This simple property is essential to a copula family.

Proposition 2.1. *For any n-copulas C, D and each $i = 1, \ldots, n$,*

$$
\begin{aligned}
&C_{m_1,\ldots,m_n}(u_1, \ldots, u_{i-1}, 1, u_{i+1}, \ldots, u_n|C, D) \\
&= C_{m_1,\ldots,m_{i-1},m_{i+1},\ldots,m_n}(u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_n|C_i, D_i), \quad u_j \in [0, 1], j \neq i,
\end{aligned}
$$

where

$$C_i(u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_n) = C(u_1, \ldots, u_{i-1}, 1, u_{i+1}, \ldots, u_n), \quad u_j \in [0, 1], j \neq i,$$

and

$$D_i(u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_n) = D(u_1, \ldots, u_{i-1}, 1, u_{i+1}, \ldots, u_n), \quad u_j \in [0, 1], j \neq i,$$

are the $(n-1)$-marginal copulas of C and D respectively.

Proof. It can be verified directly with (3). □

The above proposition states the relationship between the marginal copulas of a CBC and the marginal distributions of the corresponding target copula and base copula.

2.2 Properties of the composite Bernstein copula

In this section, we study several properties of CBC concerning continuity, linearity and symmetry.

Proposition 2.2.

(i) *If the sequence of copulas C_k converges to a copula C uniformly as k goes to infinity, then $C_{m_1,\ldots,m_n}(\cdot|C_k, D)$ converges to $C_{m_1,\ldots,m_n}(\cdot|C, D)$ uniformly.*

(ii) *If the sequence of copulas D_k converges to a copula D uniformly as k goes to infinity, then $C_{m_1,\ldots,m_n}(\cdot|C, D_k)$ converges to $C_{m_1,\ldots,m_n}(\cdot|C, D)$ uniformly.*

(iii) *Suppose that two target copulas C_1 and C_2 satisfy that $C_1 \leq C_2$, then we have that $C_{m_1,\ldots,m_n}(\cdot|C_1, D) \leq C_{m_1,\ldots,m_n}(\cdot|C_2, D)$.*

(iv) *Suppose that two base copulas D_1 and D_2 satisfy that $D_1 \leq D_2$, then we have that $C_{m_1,\ldots,m_n}(\cdot|C, D_1) \leq C_{m_1,\ldots,m_n}(\cdot|C, D_2)$.*

Proof. For any two copula functions C_1 and C_2, applying equation (3) we have

$$\begin{aligned}
& C_{m_1,\ldots,m_n}(u_1,\ldots,u_2|C_1,D) - C_{m_1,\ldots,m_n}(u_1,\ldots,u_2|C_2,D) \\
= \; & \mathbb{E}[C_1(\frac{F^{-1}_{Bin(m_1,u_1)}(U_1^D)}{m_1},\ldots,\frac{F^{-1}_{Bin(m_n,u_n)}(U_n^D)}{m_n})] \\
& -\mathbb{E}[C_2(\frac{F^{-1}_{Bin(m_1,u_1)}(U_1^D)}{m_1},\ldots,\frac{F^{-1}_{Bin(m_n,u_n)}(U_n^D)}{m_n})].
\end{aligned}$$

Thus part (i) and part (iii) can be proved directly from the above equality. For any two copula functions D_1 and D_2, applying equation (4) we have

$$\begin{aligned}
& C_{m_1,\ldots,m_n}(u_1,\ldots,u_n|C,D_1) - C_{m_1,\ldots,m_n}(u_1,\ldots,u_n|C,D_2) \\
= \; & \mathbb{E}[D_1(1-F_{Bin(m_1,u_1)}(m_1V_1),\ldots,1-F_{Bin(m_n,u_n)}(m_nV_n))] \\
& -\mathbb{E}[D_2(1-F_{Bin(m_1,u_1)}(m_1V_1),\ldots,1-F_{Bin(m_n,u_n)}(m_nV_n))].
\end{aligned}$$

Thus part (ii) and part (iv) can be proved directly from the above equality. □

From the above proposition, we can see that the CBC is quite robust with respect to the target and base copulas. For a given target (base) copula, different base (target) copulas can be chosen to adjust the value of CBC. Moreover, a linear combination of base copulas can be chosen to further adjust the value of CBC conveniently, as shown in the following proposition. It is a straightforward consequence of (3) and (4), so we omit the proof here.

Proposition 2.3. *Suppose $\lambda \in [0,1]$ is a constant.*

(i) Suppose C_1, C_2 are two n-copulas and $C = \lambda C_1 + (1-\lambda)C_2$, then for any base copula D,

$$C_{m_1,\ldots,m_n}(\cdot|C,D) = \lambda C_{m_1,\ldots,m_n}(\cdot|C_1,D) + (1-\lambda)C_{m_1,\ldots,m_n}(\cdot|C_2,D).$$

(ii) Suppose D_1, D_2 are two n-copulas and $D = \lambda D_1 + (1-\lambda)D_2$, then for any target copula C,

$$C_{m_1,\ldots,m_n}(\cdot|C,D) = \lambda C_{m_1,\ldots,m_n}(\cdot|C,D_1) + (1-\lambda)C_{m_1,\ldots,m_n}(\cdot|C,D_2).$$

Remark 2.3. We can see that $C_{m_1,\ldots,m_n}$ is a mapping from $\mathcal{C}_n \times \mathcal{C}_n$ to $\mathcal{C}_n$, where $\mathcal{C}_n$ is the space of n-copulas. The above proposition shows that the CBC admits linearity in terms of base copulas and target copulas. In summary, $C_{m_1,\ldots,m_n} : \mathcal{C}_n \times \mathcal{C}_n \to \mathcal{C}_n$ is a monotone, bi-linear and continuous functional.

The next proposition studies the symmetry of the CBC. An n-copula C is symmetric, if $C(\mathbf{u}) = C(\sigma(\mathbf{u}))$ for all $\mathbf{u} \in [0,1]^n$ where σ is any n-permutation. And an n-copula C is radially symmetric, if $C(\mathbf{u}) = \bar{C}(\mathbf{u})$ for all $\mathbf{u} \in [0,1]^n$ where $\bar{C}$ is the survival copula of C.

Proposition 2.4. *(i) If C and D are both symmetric n-copulas, then $C_{m,\ldots,m}(\cdot|C,D)$ is also symmetric;*

(ii) For any n-copulas C and D, we have

$$\bar{C}_{m_1,\ldots,m_n}(\cdot|C,D) = C_{m_1,\ldots,m_n}(\cdot|\bar{C},\bar{D}).$$

In particular, if C and D are both radially symmetric, then $C_{m_1,\dots,m_n}(\cdot|C,D)$ is also radially symmetric.

Proof. (i) We only show the case $n = 2$ as the general case $n \geq 3$ is similar. Let (U_1^D, U_2^D) be a random vector with distribution function $\bar{D}$. By the definition of CBC and the symmetry of C and D (and hence $\bar{D}$) we have that

$$\begin{aligned} C_{m,m}(u_2, u_1|C,D) &= \mathbb{E}[C(\frac{F^{-1}_{Bin(m,u_2)}(U_1^D)}{m}, \frac{F^{-1}_{Bin(m,u_1)}(U_2^D)}{m})] \\ &= \mathbb{E}[C(\frac{F^{-1}_{Bin(m,u_2)}(U_2^D)}{m}, \frac{F^{-1}_{Bin(m,u_1)}(U_1^D)}{m})] \\ &= \mathbb{E}[C(\frac{F^{-1}_{Bin(m,u_1)}(U_1^D)}{m}, \frac{F^{-1}_{Bin(m,u_2)}(U_2^D)}{m})] = C_{m,m}(u_1, u_2|C,D), \end{aligned}$$

thus $C_{m,m}(u_2, u_1|C,D)$ is symmetric.

(ii) For n fixed and any n-copula C, we define a linear operator $S[C] = 1 + \sum_{i=1}^n (-1)^i S_i(C)$ where $S_i(C)$ is the sum of all i-marginal copulas of C, that is,

$$S_i(C)(u_1, \dots, u_n) = \sum_{1 \leq j_1 < j_2 < \dots < j_i \leq n} \mathbb{P}(V_{j_1} \leq u_{j_1}, \dots, V_{j_i} \leq u_{j_i}),$$

where $(V_1, \dots, V_n)$ follows copula C. Using Poincare Formula it is easy to check that $S[C](1-u_1, \dots, 1-u_n) = \bar{C}(u_1, \dots, u_n)$ holds for any copula C. Let $(U_1^D, \dots, U_n^D)$ follow copula $\bar{D}$ and denote

$$g_i := \frac{F^{-1}_{Bin(m_i,1-u_i)}(U_i^D)}{m_i}, \quad h_i := \frac{F^{-1}_{Bin(m_i,u_i)}(1-U_i^D)}{m_i}.$$

Note that it is easy to verify $g_i + h_i = 1$ almost surely.

By the definition of CBC we know that $\mathbb{E}[C(g_1, \dots, g_n)] = C_{m_1,\dots,m_n}(1-u_1, \dots, 1-u_n|C,D)$. We first verify that

$$\mathbb{E}[S[C](g_1, \dots, g_n)] = S[C_{m_1,\dots,m_n}(\cdot|C,D)](1-u_1, \dots, 1-u_n).$$

This can be seen from Proposition 2.1, by noting that each term of the form

$$C_{m_1,\dots,m_n}(1-u_1, \dots, 1-u_{i-1}, 1, 1-u_{i+1}, \dots, 1-u_n|C,D)$$

in $S[C_{m_1,\ldots,m_n}(\cdot|C,D)](1-u_1,\ldots,1-u_n)$ is equal to the term

$$\begin{aligned}&\mathbb{E}[C_i(g_1,\ldots,g_{i-1},g_{i+1},\ldots,g_n)]\\&=C_{m_1,\ldots,m_{i-1},m_{i+1},\ldots,m_n}(1-u_1,\ldots,1-u_{i-1},1-u_{i+1},\ldots,1-u_n|C_i,D_i)\end{aligned}$$

in $\mathbb{E}[S[C](g_1,\ldots,g_n)]$, where C_i and D_i are defined in Proposition 2.1. Other marginal copula terms are similar. Therefore,

$$\begin{aligned}&C_{m_1,\ldots,m_n}(u_1,\ldots,u_n|\bar{C},\bar{D})\\=&\mathbb{E}[\hat{C}(h_1,\ldots,h_n)]=\mathbb{E}[S(C(1-h_1,\ldots,1-h_n))]\\=&\mathbb{E}[S[C](g_1,\ldots,g_n)]\\=&S[C_{m_1,\ldots,m_n}(\cdot|C,D)](1-u_1,\ldots,1-u_n)=\hat{C}_{m_1,\ldots,m_n}(u_1,\ldots,u_n|C,D).\end{aligned}$$

□

A typical radially symmetric family is elliptically contoured distributions (Fang et al, 1990), including multivariate normal distributions, Student-t distributions, and multivariate symmetric stable distributions. Sometimes radial symmetry restricts the use of elliptically contoured distributions in finance or insurance (Frahm et al, 2003). By Proposition 2.4, choosing different base copulas leads to a CBC with or without this symmetry.

2.3 Reproduction property

For a given target copula C, it is interesting to see whether there exists base copula D such that the corresponding CBC can reproduce the target copula C, i.e.

$$C_{m_1,\ldots,m_n}(\cdot|C,D)=C$$

holds for some positive integers $m_1,\ldots,m_n$. For the simplest case $m_1=\cdots=m_n=1$, $D=C$ is equivalent to $C_{1,\ldots,1}(\cdot|C,D)=C$ by Theorem 2.1(ii). However, for the other values of $m_1,\ldots,m_n$, $D=C$ is not sufficient for $C_{m_1,\ldots,m_n}(\cdot|C,D)=C$ in general.

We find that in the case $m_1 = \cdots = m_n = m$, for the three fundamental copula functions: Fréchet upper copula $M(u_1, \ldots, u_n) = \min\{u_i, i = 1, \ldots, n\}, u_1, \ldots, u_n \in [0,1]$, independent copula $\Pi(u_1, \ldots, u_n) = \prod_{i=1}^n u_i, u_1, \ldots, u_n \in [0,1]$ and Fréchet lower copula $W(u,v) = \max\{u+v-1, 0\},\ u, v \in [0,1]$ (Fréchet lower copula is a copula only in the bivariate case), the condition $D = C$ is sufficient.

Proposition 2.5. *In (i) and (ii), all copulas are n-copulas. In (iii), all copulas are 2-copulas.*

(i) $C_{m,\ldots,m}(\cdot|M, D) = M$ *if* $D = M$*;*

(ii) $C_{m_1,\ldots,m_n}(\cdot|\Pi, D) = \Pi$ *if* $D = \Pi$*;*

(iii) $C_{m,m}(\cdot|W, D) = W$ *if* $D = W$*.*

Proof. (ii) can be directly verified and so we only verify that of (i) and (iii).

(i) We have

$$\begin{aligned} C_{m\ldots,m}(u_1, \ldots, u_n|C, D) &= \mathbb{E}[C(\frac{F^{-1}_{Bin(m,u_1)}(U)}{m}, \ldots, \frac{F^{-1}_{Bin(m,u_n)}(U)}{m})] \\ &= \mathbb{E}[\min\{\frac{F^{-1}_{Bin(m,u_i)}(U)}{m}, i = 1, \ldots, n\}] \\ &= \mathbb{E}[\frac{F^{-1}_{Bin(m,\min\{u_1,\ldots,u_n\})}(U)}{m}] \\ &= \frac{m \min\{u_1, \ldots, u_n\}}{m} = \min\{u_1, \ldots, u_n\}. \end{aligned}$$

The above proof process uses the fact that $F^{-1}_{Bin(m,u)}(x)$ is increasing about u when x is fixed since the function $F_{Bin(m,u)}(x)$ is strictly decreasing about u when x is fixed.

(iii) Consider the case $C(u_1, u_2) = D(u_1, u_2) = \max\{u_1 + u_2 - 1, 0\}$. Note that the left-continuous inverse function $F^{-1}_{Bin(m,u)}$ can be expressed as

$$F^{-1}_{Bin(m,u)}(x) = k, \quad F_{Bin(m,u)}(k-1) < x \leq F_{Bin(m,u)}(k),\ k = 0, 1, 2, \ldots, m,$$

where $F_{Bin(m,u)}(-1) \equiv 0$. It is easy to see that for $k = 0, 1, 2, \ldots, m$,

$$F_{Bin(m,u)}(k) + F_{Bin(m,1-u)}(m-k-1) = 1. \tag{7}$$

So by the equation (7), if $F_{Bin(m,u)}(k-1) < x < F_{Bin(m,u)}(k)$, $k = 0, 1, 2, \ldots, m$, we have that

$$F_{Bin(m,1-u)}(m-k-1) < 1 - x < F_{Bin(m,1-u)}(m-k)$$

and

$$F^{-1}_{Bin(m,1-u)}(1-x) = m - k = m - F^{-1}_{Bin(m,u)}(x), \qquad u \in [0,1]. \tag{8}$$

Thus for $x \neq F_{Bin(m,u_i)}(k)$, $k = -1, 0, 1, 2, \ldots, m, i = 1, 2$,

$$F^{-1}_{Bin(m,u_1)}(x) + F^{-1}_{Bin(m,u_2)}(1-x) - m = F^{-1}_{Bin(m,u_1)}(x) - F^{-1}_{Bin(m,1-u_2)}(x)$$

according to the equation (8). Note that the fact that $F^{-1}_{Bin(m,u)}(x)$ is increasing about u when x is fixed, we finally have

$$\begin{aligned}
& C_{m,m}(u_1, u_2|C, D) \\
= \;& \mathbb{E}[\max\{\frac{F^{-1}_{Bin(m,u_1)}(U)}{m} + \frac{F^{-1}_{Bin(m,u_2)}(1-U)}{m} - 1, 0\}] \\
= \;& \mathbb{E}[\max\{\frac{F^{-1}_{Bin(m,u_1)}(U) + F^{-1}_{Bin(m,u_2)}(1-U) - m}{m}, 0\} \left(I_{\{u_1 \geq 1-u_2\}} + I_{\{u_1 < 1-u_2\}}\right)] \\
= \;& \mathbb{E}[\frac{F^{-1}_{Bin(m,u_1)}(U) + F^{-1}_{Bin(m,u_2)}(1-U) - m}{m}] I_{\{u_1+u_2-1 \geq 0\}} \\
= \;& \frac{mu_1 + mu_2 - m}{m} I_{\{u_1+u_2-1 \geq 0\}} = \max\{u_1 + u_2 - 1, 0\}.
\end{aligned}$$

□

Remark 2.4. Proposition 2.5 states that CBC has the reproduction property for Fréchet upper copula, the independent copula and the bivariate Fréchet lower copula, which correspond to the three important dependency structures in insurance and finance: comonotonicity, independence and countermonotonicity (Dhaene et al, 2002a, 2002b). Thus CBC shows its advantage for modeling these special dependency structures.

2.4 Bivariate tail dependence

Tail dependence (see e.g. Joe, 1997) describes the significance of dependence in the tail of a bivariate distribution; see also Schmidt (2004). The lower tail dependence coefficient of a copula C is defined as $\lambda_L^C := \lim_{u \downarrow 0} \frac{C(u,u)}{u}$ and upper tail dependence coefficient of a copula C is defined as $\lambda_U^C := \lim_{u \downarrow 0} \frac{\bar{C}(u,u)}{u}$. Gaussian copula is widely applied in finance due to its relatively simple estimation procedure and computational ease. However, it is often criticized for not being able to characterize tail dependence between assets because of its tail independence property. As mentioned in Sancetta and Satchell (2004), the BC is also unable to capture tail dependence (note that $|C_m - C| \to 0$ uniformly as $m \to \infty$ does not imply $\lambda_L^{C_m} \to \lambda_L^C$ or $\lambda_U^{C_m} \to \lambda_U^C$).

CBC is able to capture tail dependence by choosing appropriate base copulas. We have the following theorem.

Theorem 2.2. *Assume that the tail dependence coefficients λ_L^D and λ_U^D of the base copula D exist.*

(i) The lower tail dependence coefficient

$$\lambda_L^{C_{m,m}(\cdot|C,D)} = m \times C(\frac{1}{m}, \frac{1}{m}) \times \lambda_L^D.$$

(ii) The upper tail dependence coefficient

$$\lambda_U^{C_{m,m}(\cdot|C,D)} = m \times \bar{C}(\frac{1}{m}, \frac{1}{m}) \times \lambda_U^D.$$

(iii) Assume that the tail dependence coefficients λ_L^C and λ_U^C of the target copula C also exist. Then, as $m \to \infty$,

$$\lambda_L^{C_{m,m}(\cdot|C,D)} \to \lambda_L^C \lambda_L^D, \quad \lambda_U^{C_{m,m}(\cdot|C,D)} \to \lambda_U^C \lambda_U^D.$$

Proof. First, write

$$\begin{aligned}\lambda_L^{C_{m,m}(\cdot|C,D)} &= \lim_{u\downarrow 0}\frac{C_{m,m}(u,u|C,D)}{u} \\ = \quad &\lim_{u\downarrow 0}\frac{1}{u}\left[\sum_{n_1=1}^{m}\sum_{n_2=1}^{m} C(\frac{n_1}{m},\frac{n_2}{m})\mathbb{P}(F^{-1}_{Bin(m,u)}(U_1^D)=n_1, F^{-1}_{Bin(m,u)}(U_2^D)=n_2)\right],\end{aligned}$$

where (U_1^D, U_2^D) follows $\bar{D}$.

If $n_1 > 1$ or $n_2 > 1$, we have

$$\begin{aligned}&\lim_{u\downarrow 0}\frac{1}{u}\mathbb{P}(F^{-1}_{Bin(m,u)}(U_1^D)=n_1, F^{-1}_{Bin(m,u)}(U_2^D)=n_2) \\ \leq \quad &\lim_{u\downarrow 0}\left[\frac{\binom{m}{n_1}u^{n_1}(1-u)^{m-n_1}}{u}+\frac{\binom{m}{n_2}u^{n_2}(1-u)^{m-n_2}}{u}\right]=0. \end{aligned} \tag{9}$$

As a result, we have

$$\lambda_L^{C_{m,m}(\cdot|C,D)} = C(\frac{1}{m},\frac{1}{m})\lim_{u\downarrow 0}\frac{1}{u}\mathbb{P}(F^{-1}_{Bin(m,u)}(U_1^D)=1, F^{-1}_{Bin(m,u)}(U_2^D)=1).$$

Observe that

$$\begin{aligned}&\mathbb{P}(U_1^D > (1-u)^m, U_2^D > (1-u)^m) - \mathbb{P}(F^{-1}_{Bin(m,u)}(U_1^D)=1, F^{-1}_{Bin(m,u)}(U_2^D)=1) \\ \leq \quad & 2(1-F_{Bin(m,u)}(1)) = o(u),\end{aligned}$$

and

$$\begin{aligned}&\lim_{u\downarrow 0}\frac{\mathbb{P}(U_1^D > (1-u)^m, U_2^D > (1-u)^m)}{u} \\ = \quad &\lim_{u\downarrow 0}\frac{\mathbb{P}(U_1^D > 1-mu+o(u), U_2^D > 1-mu+o(u))}{u} = m\times\lambda_L^D. \end{aligned} \tag{10}$$

Then we obtain $\lambda_L^{C_{m,m}(\cdot|C,D)} = m\times C(\frac{1}{m},\frac{1}{m})\times\lambda_L^D$.

For (ii), note that by Proposition 2.4 (ii), $\bar{C}_{m,m}(\cdot|C,D) = C_{m,m}(\cdot|\bar{C},\bar{D})$. Thus

$$\lambda_U^{C_{m,m}(\cdot|C,D)} = \lambda_L^{\bar{C}_{m,m}(\cdot|C,D)} = m\times\bar{C}(\frac{1}{m},\frac{1}{m})\times\lambda_L^{\bar{D}} = m\times\bar{C}(\frac{1}{m},\frac{1}{m})\times\lambda_U^D.$$

(iii) is directly implies by (i) and (ii).

□

Remark 2.5. Theorem 2.2 implies the fact that the BC always have zero tail dependence coefficients since the base copula D of a BC is the independent copula, with $\lambda_U^D = \lambda_L^D = 0$. In the CBC family, we can choose base copulas D with $\lambda_U^D = \lambda_L^D = 1$ (such as the Fréchet upper copula M) to preserve the tail dependence coefficients asymptotically.

2.5 Numerical example

In this section, we provide a numerical example for CBC as $m \to \infty$ to exhibit the influence of base copulas on the difference between CBC and its target copula. Since the expression for CBC is not explicit in general, we use a Monte-Carlo simulation with sample size 10000 for the definition (3) to approximate CBC in this section.

From Theorem 2.1, we know that no matter which base copula we choose, the CBC $C_{m,m}(u_1, u_2|C, D)$ will converge to $C(u_1, u_2)$ as $m \to \infty$. In the following, we choose C^ρ, a Gaussian copula with correlation parameter ρ, as the target copula, and the Fréchet upper copula M, the Fréchet lower copula W, the independent copula Π and the target copula C^ρ itself are chosen as the base copula to report the numerical values when m is finite. We would like to see how close CBC is to the target copula.

The absolute distance is approximated by

$$T(C_{m,m}(\cdot|C, D)) = \frac{1}{K^2} \sum_{i=1}^{K} \sum_{j=1}^{K} |C(\frac{i}{K}, \frac{j}{K}) - C^*_{m,m}(\frac{i}{K}, \frac{j}{K}|C, D)|, \tag{11}$$

here $C^*_{m,m}(\cdot|C, D)$ is the Monte-Carlo simulation of $C_{m,m}(\cdot|C, D)$, which we treat as the true value of $C_{m,m}(\cdot|C, D)$.

From Table 2.1, we find that when the target copula has a highly positive correlation, the Fréchet upper copula as a base copula leads to faster convergence than the independent and the Fréchet lower copula copulas. Opposite observation can be found in the case of negative correlation. At the same time, it is clear that the target copula itself as base copula leads to the fastest convergence. However when m is large enough the effect of M or W is similar to that of the target copula itself.

Table 2.1: Approximate distance measured in (11)

Target Copula: Gaussian ($\rho = 0.7$)					Target Copula: Gaussian ($\rho = -0.7$)				
m	Base copula				m	Base copula			
	M	Π	W	$C^{0.7}$		M	Π	W	$C^{-0.7}$
10	0.0039	0.0112	0.0252	0.0024	10	0.0238	0.0100	0.0052	0.0025
20	0.0027	0.0059	0.0132	0.0022	20	0.0118	0.0057	0.0033	0.0023
30	0.0025	0.0042	0.0091	0.0021	30	0.0078	0.0042	0.0030	0.0022
40	0.0022	0.0034	0.0071	0.0020	40	0.0058	0.0035	0.0025	0.0021
100	0.0021	0.0023	0.0034	0.0021	100	0.0032	0.0024	0.0022	0.0020
200	0.0020	0.0021	0.0026	0.0020	200	0.0023	0.0020	0.0020	0.0021

3 Empirical Composite Bernstein Copula

3.1 Definition of empirical composite Bernstein copula

In this section we discuss statistical inference using the composite Bernstein copula. An estimation procedure will be provided below. We propose to estimate a copula C based on CBC.

As discussed in Sancetta and Satchell (2004), the Bernstein copulas can be used to estimate unknown copulas by constructing the empirical Bernstein copulas (EBC). The CBC serves in the same procedure with even more flexibility by allowing to choose the base copula D. In what follows, we will introduce the *empirical composite Bernstein copula* (ECBC).

Let $C_N(\mathbf{u})$, $\mathbf{u} \in [0,1]^n$ be the empirical copula of sample data $\mathbf{V}_1, \ldots, \mathbf{V}_N \in [0,1]^n$ from a copula C, i.e.,

$$C_N(\mathbf{u}) = \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}_{\{\mathbf{V}_j \leq \mathbf{u}\}},$$

where the "$\leq$" is a component-wise inequality.

Remark 3.1. Here we assume that the data are sampled from a copula, and hence the marginal distributions are known to be U$[0,1]$. If the marginal distributions are unknown, the empirical copula should be defined as $\frac{1}{N}\sum_{j=1}^{N}\mathbf{1}_{\{\tilde{\mathbf{V}}_j\le\mathbf{u}\}}$, where $\tilde{\mathbf{V}}_i = (\tilde{V}_{i,1},\ldots,\tilde{V}_{i,n})$ with $\tilde{V}_{i,j} = F_{Nj}(V_{i,j})$ and F_{Nj} is the marginal empirical distribution function of the jth component.

C_N is not a copula for finite N; it is only a copula asymptotically (as $N\to\infty$). Note that the empirical copula C_N does not have a density. As mentioned in Remark 2.1, the definition (3) does not require C to be a copula. Hence, the following empirical composite Bernstein copula (ECBC) is properly defined as

$$\tilde{C}_{m_1,\ldots,m_n}(\mathbf{u}|N,D) := C_{m_1,\ldots,m_n}(\mathbf{u}|C_N,D), \quad \mathbf{u}\in[0,1]^n.$$

Note that this definition only involves the information of C_N on the points $(\frac{v_1}{m_1},\ldots,\frac{v_n}{m_n})$ for $v_i\in\{0,\ldots,m_i\}$, $i=1,\ldots,n$.

From (4) we can see that the function $\tilde{C}_{m_1,\ldots,m_n}(\mathbf{u}|N,D)$ can be easily calculated by using

$$\begin{aligned}&\tilde{C}_{m_1,\ldots,m_n}(u_1,\ldots,u_n|N,D)\\&=\frac{1}{N}\sum_{j=1}^{N}D(1-F_{Bin(m_1,u_1)}(m_1V_{j1}),\ldots,1-F_{Bin(m_n,u_n)}(m_nV_{jn})),\end{aligned}\tag{12}$$

where $\mathbf{V}_j=(V_{j1},\ldots,V_{jn}),\ j=1,\ldots,N$ are the sample data. Moreover, we can express $\tilde{C}_{m_1,\ldots,m_n}(u_1,\ldots,u_n|N,D)$ as

$$\begin{aligned}&\tilde{C}_{m_1,\ldots,m_n}(u_1,\ldots,u_n|N,D)\\&=\sum_{l_1=0}^{m_1}\cdots\sum_{l_n=0}^{m_n}C_N(\frac{l_1}{m_1},\ldots,\frac{l_n}{m_n})\mathbb{P}(F^{-1}_{Bin(m_1,u_1)}(U_1)=l_1,\ldots,F^{-1}_{Bin(m_n,u_n)}(U_n)=l_n).\end{aligned}\tag{13}$$

Remark 3.2. ECBC defined in this paper is one generalization of the empirical Bernstein copula in Sancetta and Satchell (2004). When the base copula D is chosen as the independent copula, ECBC becomes the empirical Bernstein copula. As for the density of copula function, by Theorem 2.1 we know that the ECBC always has a density if D has a density. Thus the density of $\tilde{C}_{m,\ldots,m}(\mathbf{u}|N,D)$, denoted as $\tilde{c}_{CB}$, exists whenever D has a density.

3.2 Limit theorems for the empirical composite Bernstein copula

The following asymptotic property holds for ECBC. A proof will be given in the Appendix.

Theorem 3.1. *Denote* $\underline{m} = \min\{m_i : i = 1, \ldots, n\}$.

(1) We have

$$\lim_{\underline{m}\to\infty} \lim_{N\to\infty} \tilde{C}_{m_1,\ldots,m_n}(\mathbf{u}|N,D) = C(\mathbf{u}) \text{ a.s., } \mathbf{u} \in [0,1]^n. \tag{14}$$

(2) As $\underline{m} \to \infty$ *and* $N \to \infty$,

$$\sup_{\mathbf{u}\in[0,1]^n} |\tilde{C}_{m_1,\ldots,m_n}(\mathbf{u}|N,D) - C(\mathbf{u})| = O_P\left(\frac{1}{\min\{\sqrt{N}, \sqrt{\underline{m}}\}}\right). \tag{15}$$

The above theorem provides the influence of the sample size N and the parameters m_i, $i = 1, \ldots, n$ on the convergence rate of ECBC.

In the following we consider the asymptotic normality of the ECBC. For simplicity we consider the bivariate case with $m_1 = m_2 = m$. For $u_1, u_2 \in (0,1)$, we denote

$$\begin{aligned}
\sigma^2(u_1,u_2) &= C(u_1,u_2)(1 - C(u_1,u_2)), \\
V(u_1,u_2) &= \frac{2}{\sqrt{\pi}}\left(\frac{\partial C(u_1,u_2)}{\partial u_1}\sqrt{u_1(1-u_1)} + \frac{\partial C(u_1,u_2)}{\partial u_2}\sqrt{u_2(1-u_2)}\right)
\end{aligned}$$

and

$$\begin{aligned}
b(u_1,u_2) &= \frac{1}{2}\left[\frac{\partial^2 C(u_1,u_2)}{\partial u_1^2}u_1(1-u_1) + \frac{\partial^2 C(u_1,u_2)}{\partial u_2^2}u_2(1-u_2)\right] \\
&\quad + \frac{\partial^2 C(u_1,u_2)}{\partial u_1 \partial u_2}\sqrt{u_1(1-u_1)}\sqrt{u_2(1-u_2)}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} st\; d\bar{D}(\Phi(s),\Phi(t)),
\end{aligned}$$

where we assume that the corresponding partial derivatives exist.

Theorem 3.2. *Assume that* $C(u_1,u_2),(u_1,u_2)\in(0,1)^2$ *has uniformly bounded third order partial derivatives and* $N^{1/2}m^{-1}\to a\in[0,\infty)$ *as* $m\to\infty$. *Then for* $(u_1,u_2)\in(0,1)^2$,

$$\mathrm{Var}(N^{1/2}(\tilde{C}_{m,m}(u_1,u_2|N,D)-C(u_1,u_2)))=\sigma^2(u_1,u_2)-\frac{1}{\sqrt{m}}V(u_1,u_2)+o\left(\frac{1}{\sqrt{m}}\right),$$

as $m\to\infty$. *Moreover, for* $(u_1,u_2)\in(0,1)^2$,

$$N^{1/2}(\tilde{C}_{m,m}(u_1,u_2|N,D)-C(u_1,u_2))\xrightarrow{d}N(ab(u_1,u_2),\sigma^2(u_1,u_2)),\quad m\to\infty.$$

Remark 3.3. (1) Note that

$$\mathrm{Var}(N^{1/2}(\tilde{C}_{m,m}(u_1,u_2|N,D)-C(u_1,u_2)))<\sigma^2(u_1,u_2)$$

if N is large enough. Thus comparing to the empirical copula function we can reduce the error of the estimator.

(2) In the case $a=0$, the limiting distribution does not depend on the base copula D. If $a>0$, the mean of the limit distribution depends on D, whereas the asymptotic variance $\sigma^2(u_1,u_2)$ does not depend on D.

(3) By comparing with the empirical copula function, ECBC can reduce the estimation error and has the same asymptotic variance. In the case $N^{1/2}m^{-1}\to a>0$, ECBC leads to some bias $ab(u_1,u_2)$.

Remark 3.4. Similarly to the discussion in Janssen et al. (2012, Remark 4), different choices of m will influence the estimation effect. For instance, the optimal m could be chosen to minimize the following asymptotic mean squared error (AMSE)

$$\mathrm{AMSE}(\tilde{C}_{m,m})=N^{-1}\sigma(u_1,u_2)-m^{-1/2}N^{-1}V(u_1,u_2)+m^{-2}b^2(u_1,u_2),$$

leading to an optimal choice

$$m(N)=\left(\frac{4b^2(u_1,u_2)}{V(u_1,u_2)}\right)^{2/3}N^{2/3}.$$

Therefore, one may choose $m(N)=cN^{2/3}$ for some $c>0$.

3.3 Some remarks on ECBC

The most important advantage of the estimation procedure using ECBC is that we are able to incorporate both prior information and data into estimation by choosing the base copula D, which allows flexibility in the estimation. For examples, we can use five scenarios for choosing the base copula D:

(i) If we have a good guess of the real copula C of the data, we can use it as the base copula.

(ii) If we do not have a good guess for the real copula, but we guess the real copula is in a parametric family, we can first perform MLE or other classic estimation method to find the parametric estimation, and use it as the base copula.

(iii) If we do not have a parametric family, but we observe that each vector of the data is positively correlated, then we can use M as the base copula. On the other hand, when $n = 2$, if we observe that the components of each data point are negatively correlated, we can use W as the base copula.

(iv) If we do not have any information, then we can use the empirical copula as the base copula, which gives a completely non-parametric estimation.

(v) When $n = 2$, if we want to capture the tail dependence information, we can choose a copula with a large tail dependence coefficient as the base copula.

An accurate guess or prior information of the real copula, chosen as the base copula, will enhance the estimation significantly (see Section 4 below). Even if the prior guess is wrong, Theorem 3.1 shows that the ECBC still converges to the real copula as $\underline{m}, N \to \infty$.

In the above estimation procedure, m_i needs to be determined beforehand; see Sancetta and Satchell (2004) and Janssen et al (2012). Some suggestions on choosing the optimal m_i are given in Remark 3.4. To simplify the procedure, as in Sancetta and

Satchell (2004) and Janssen et al (2012), we often assume that $m_1 = m_2 = \cdots = m_n = m$ in our simulation studies. The parameter m measures the preference between the base copula and the data. If m is small, the base copula is trusted. As m increases, the data is more trusted (see also Remark 2.3). Thus, it is reasonable to choose $m \to \infty$ as $N \to \infty$. It is worth pointing out that this logic coincides with the classic Bayesian statistics: the more data we have, the more we trust the data; the less data we have, the more we trust the prior. This also provides an explanation for the assumption that $m \to \infty$ as $N \to \infty$ in the EBC estimation in Sancetta and Satchell (2004). From Theorem 3.2 we can see that as $N^{1/2}m^{-1} \to 0$, ECBC converges to the target copula C. Thus in the statistical estimation, the parameters $m_i,\ i = 1, \ldots, n$ can be chosen as a function of the sample N, such as satisfying $N = O(m_i^{3/2})$ as suggested in Remark 3.4.

In summary, one can choose the copula function D first, which shows the prior opinion about our consideration, and then choose the numbers $m_i,\ i = 1, \ldots, n$ based on the size of the sample, and finally the statistical estimation of the parameters in estimator can be carried out.

4 Simulation Studies and Real Data Analysis

4.1 Simulation studies

In this section, we carry out some simulation studies in the bivariate case to compare the empirical copula estimator $C_N(u_1, u_2)$ with the empirical composite Bernstein copula estimator $\tilde{C}_{m,m}(u_1, u_2|N, D)$ for different choices of base copula D, parameter m , including the empirical Bernstein copula estimator (i.e., $D = \Pi$). Choices of the base copula D and the parameter m allow flexibility in the estimation.

In the study, the estimation quality is evaluated by empirically calculating their mean discrete L_1-norm between the real copula C and an estimator C^E, denoted by

$$R(C^E, C) = \mathbb{E}\left\{\frac{1}{K^2}\sum_{i=1}^{K}\sum_{j=1}^{K}|C(\frac{i}{K},\frac{j}{K}) - C^E(\frac{i}{K},\frac{j}{K})|\right\}. \tag{16}$$

Here we take $K = 100$, the repetition $r = 10000$. The candidates for C^E include

(a) The empirical copula C_N.

(b) The parametric MLE of a Gaussian copula: $\hat{C}^G$.

(c) The ECBC based on M, Π, W: $\tilde{C}_{m,m}(\cdot|N, M)$, $\tilde{C}_{m,m}(\cdot|N, \Pi)$, $\tilde{C}_{m,m}(\cdot|N, W)$. Note that $\tilde{C}_{m,m}(\cdot|N, \Pi)$ is the empirical Bernstein copula (EBC) introduced in Sancetta and Satchell (2004).

(d) The ECBC based on the estimated Gaussian copula: $\tilde{C}_{m,m}(\cdot|N, \hat{C}^G)$.

(e) The ECBC based on the real copula C: $\tilde{C}_{m,m}(\cdot|N, C)$.

(f) The ECBC based on the empirical copula C_N: $\tilde{C}_{m,m}(\cdot|N, C_N)$.

In the first simulation, we focus on the influence of the base copula. We choose the above estimators C^E (a)-(f), and generate N ($N = 50, 200$) iid random vectors from a bivariate Gaussian copula C with $\theta = 0.7$. The Gaussian copula with parameter $\theta \in (-1, 1)$ is defined as: for $(u, v) \in [0, 1]^2$,

$$C_\theta(u, v) = \Phi_\theta(\Phi^{-1}(u), \Phi^{-1}(v)),$$

where Φ is the standard Normal distribution function, and Φ_θ is a two-dimensional normal distribution function with mean zero and covariance matrix $\begin{pmatrix} 1 & \theta \\ \theta & 1 \end{pmatrix}$.

In the second simulation, we simulate samples from different copula families. We choose the above estimators C^E (a)-(d), and generate $N = 50$ iid random vectors from Gumbel, t- and Clayton copulas with different parameters.

- The Clayton copula with parameter $\theta \in [-1, \infty)\backslash\{0\}$ is defined as: for $(u, v) \in [0, 1]^2$,

$$C_\theta(u, v) = \left[\max\left(u^{-\theta} + v^{-\theta} - 1, 0\right)\right]^{-\frac{1}{\theta}}.$$

- The t-copula with parameters $\rho \in [-1, 1]$ and $\nu > 0$ is defined as: for $(u, v) \in [0, 1]^2$,

$$C(u, v) = \int_{-\infty}^{t_\nu^-(u)} \int_{-\infty}^{t_\nu^-(v)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \{1 + \frac{x^2 - 2\rho xy + y^2}{\nu(1-\rho^2)}\}^{-(\nu+2)/2}\, dydx,$$

  where t_ν is the distribution function of a t-distribution with ν degrees of freedom and t_ν^- denotes the generalized inverse function of t_ν.

- The Gumbel copula with parameter $\theta \in [1, \infty)$ is defined as: for $(u, v) \in [0, 1]^2$,

$$C_\theta(u, v) = \exp\left\{-\left[(-\ln u)^\theta - (-\ln v)^\theta\right]^{\frac{1}{\theta}}\right\}.$$

The results from the first simulation are reported in Table 4.2 and the results from the second simulation are reported in Table 4.3.

We observe some interesting facts:

(i) The parametric MLE $\hat{C}^G$ performs the best for all four samples. This confirms that a parametric Gaussian estimation can be a quite good approximation to the three parametric copula families.

(ii) Among the non-parametric methods, ECBC and EBC generally perform better than the empirical copula C_N.

(iii) From Table 4.2, it is clear that $\tilde{C}_{m,m}(\cdot|N, C)$ outperforms the other ECBC type of estimators, and $\tilde{C}_{m,m}(\cdot|N, \hat{C}^G)$ also preforms quite well. This suggests that a more accurate base copula leads to a better estimation.

(iv) $\tilde{C}_{m,m}(\cdot|N, W)$ performs poorly for small m because W is very far away from the real copula C. $\tilde{C}_{m,m}(\cdot|N, M)$ performs pretty well, almost always better than the EBC $\tilde{C}_{m,m}(\cdot|N, \Pi)$ since our choices of real copula all have a positive ρ.

Table 4.2: $R(C^E, C_\theta)$ for different choices of C^E. Here we omit N in the ECBC.

$N = 50$						
m	$\tilde{C}_{m,m}(\cdot\|M)$	$\tilde{C}_{m,m}(\cdot\|\Pi)$	$\tilde{C}_{m,m}(\cdot\|W)$	$\tilde{C}_{m,m}(\cdot\|\hat{C}^G)$	$\tilde{C}_{m,m}(\cdot\|C_\rho)$	$\tilde{C}_{m,m}(\cdot\|C_N)$
3	0.0250	0.0374	0.0718	0.0234	0.0227	0.0408
7	0.0304	0.0327	0.0442	0.0301	0.0297	0.0402
14	0.0343	0.0347	0.0383	0.0342	0.0340	0.0407
30	0.0374	0.0373	0.0381	0.0373	0.0372	0.0415
50	0.0387	0.0386	0.0390	0.0387	0.0386	0.0418
	$R(C_N, C_\theta) = 0.0438$			$R(\hat{C}^G, C_\theta) = 0.0051$		

$N = 200$						
m	$\tilde{C}_{m,m}(\cdot\|M)$	$\tilde{C}_{m,m}(\cdot\|\Pi)$	$\tilde{C}_{m,m}(\cdot\|W)$	$\tilde{C}_{m,m}(\cdot\|\hat{C}^G)$	$\tilde{C}_{m,m}(\cdot\|C_\rho)$	$\tilde{C}_{m,m}(\cdot\|C_N)$
3	0.0151	0.0337	0.0713	0.0123	0.0120	0.0216
7	0.0166	0.0209	0.0364	0.0157	0.0155	0.0216
14	0.0180	0.0189	0.0243	0.0177	0.0176	0.0217
30	0.0192	0.0193	0.0206	0.0191	0.0190	0.0217
50	0.0198	0.0198	0.0202	0.0197	0.0197	0.0216
	$R(C_N, C_\theta) = 0.0221$			$R(\hat{C}^G, C_\theta) = 0.0028$		

(v) Recall that m measures the preference between the data and the base copula. We observe that when we have a good guess (such as $M, C, \hat{C}^G$) as the base copula, small m leads to a better estimation. When we have a bad guess as the base copula (such as W), small m leads to a worse estimation. Based on this observation, it is reasonable to let m increase as N increases, since a larger N leads to a more convincing non-parametric estimation.

(vi) From Table 4.3, among the three ECBCs, $\tilde{C}_{m,m}(\cdot|N, \Pi)$ performs best when the Spearman's ρ is low, $\tilde{C}_{m,m}(\cdot|N, M)$ performs best when the Spearman's ρ is high.

Table 4.3: $R(C^E, C_\theta)$ for different choices of C^E. Here we omit N in the ECBC. The parameters θ are in parentheses and the Spearman's ρ are also listed.

	Clayton(1), $\rho = 0.479$ $R(\hat{C}^G, C_\theta) = 0.0109$, $R(C_N, C_\theta) = 0.0440$				Clayton(3), $\rho = 0.788$ $R(\hat{C}^G, C_\theta) = 0.0093$, $R(C_N, C_\theta) = 0.0445$			
m	$\tilde{C}_{m,m}(\cdot\|M)$	$\tilde{C}_{m,m}(\cdot\|\Pi)$	$\tilde{C}_{m,m}(\cdot\|W)$	$\tilde{C}_{m,m}(\cdot\|\hat{C}^G)$	$\tilde{C}_{m,m}(\cdot\|M)$	$\tilde{C}_{m,m}(\cdot\|\Pi)$	$\tilde{C}_{m,m}(\cdot\|W)$	$\tilde{C}_{m,m}(\cdot\|\hat{C}^G)$
3	0.0309	0.0317	0.0634	0.0263	0.0258	0.0430	0.0771	0.0264
7	0.0338	0.0324	0.0404	0.0326	0.0321	0.0367	0.0489	0.0325
14	0.0361	0.0351	0.0372	0.0356	0.0356	0.0367	0.0408	0.0358
30	0.0382	0.0377	0.0379	0.0381	0.0387	0.0389	0.0400	0.0388
50	0.0396	0.0393	0.0393	0.0395	0.0400	0.0400	0.0405	0.0400
	t(0.5,5), $\rho = 0.473$ $R(\hat{C}^G, C_\theta) = 0.0096$, $R(C_N, C_\theta) = 0.0437$				t(0.8,5), $\rho = 0.777$ $R(\hat{C}^G, C_\theta) = 0.0052$, $R(C_N, C_\theta) = 0.0474$			
m	$\tilde{C}_{m,m}(\cdot\|M)$	$\tilde{C}_{m,m}(\cdot\|\Pi)$	$\tilde{C}_{m,m}(\cdot\|W)$	$\tilde{C}_{m,m}(\cdot\|\hat{C}^G)$	$\tilde{C}_{m,m}(\cdot\|M)$	$\tilde{C}_{m,m}(\cdot\|\Pi)$	$\tilde{C}_{m,m}(\cdot\|W)$	$\tilde{C}_{m,m}(\cdot\|\hat{C}^G)$
3	0.0285	0.0325	0.0662	0.0248	0.0258	0.0443	0.0784	0.0257
7	0.0317	0.0330	0.0441	0.0315	0.0346	0.0389	0.0505	0.0348
14	0.0349	0.0356	0.0394	0.0353	0.0388	0.0400	0.0441	0.0389
30	0.0375	0.0379	0.0390	0.0378	0.0416	0.0419	0.0431	0.0416
50	0.0390	0.0392	0.0398	0.0392	0.0429	0.0431	0.0436	0.0430
	Gumbel(1.5), $\rho = 0.478$ $R(\hat{C}^G, C_\theta) = 0.0078$, $R(C_N, C_\theta) = 0.0420$				Gumbel(2.5), $\rho = 0.789$ $R(\hat{C}^G, C_\theta) = 0.0052$, $R(C_N, C_\theta) = 0.0477$			
m	$\tilde{C}_{m,m}(\cdot\|M)$	$\tilde{C}_{m,m}(\cdot\|\Pi)$	$\tilde{C}_{m,m}(\cdot\|W)$	$\tilde{C}_{m,m}(\cdot\|\hat{C}^G)$	$\tilde{C}_{m,m}(\cdot\|M)$	$\tilde{C}_{m,m}(\cdot\|\Pi)$	$\tilde{C}_{m,m}(\cdot\|W)$	$\tilde{C}_{m,m}(\cdot\|\hat{C}^G)$
3	0.0299	0.0290	0.0631	0.0223	0.0267	0.0465	0.0816	0.0266
7	0.0298	0.0293	0.0399	0.0280	0.0351	0.0404	0.0533	0.0352
14	0.0325	0.0320	0.0354	0.0319	0.0388	0.0405	0.0450	0.0388
30	0.0353	0.0350	0.0358	0.0351	0.0417	0.0423	0.0436	0.0417
50	0.0368	0.0366	0.0370	0.0367	0.0432	0.0434	0.0441	0.0432

The ECBC based on the Gaussian copula outperforms the other ECBCs in most cases.

In summary, the ECBC estimation procedure provides a new method which incorporates both the prior information and the data, and it shows its advantage by comparing with the above non-parametric methods when an appropriate base copula is chosen. It is very flexible to choose different base copulas.

4.2 Simulation for different m_1, m_2

In the next we study the the impact of (m_1, m_2) for $m_1 \neq m_2$. Table 4.4 reports the L_1-error of ECBC with various choices of (m_1, m_2). The sample is simulated from a Gumbel copula with parameter $\theta = 2.5$, and the sample size is 500.

We observe that for ECBC with M as the base copula, the estimation error is minimized when m_1 is close to m_2, whereas for EBC, such a trend is not observed. However, optimal choices of m_1, m_2 are not easy to obtain.

4.3 Financial data analysis

In financial practice, Gaussian copula is often chosen as the benchmark correlation structure. However, it is well-known that Gaussian copula has zero tail dependence coefficient and the tail property of real financial data cannot be captured. In this empirical study, we tried to solve this issue by ECBC. From proposition 2.2, we understand that if the base copula M is chosen, ECBC has positive tail dependence coefficient. Thus, we can try to capture the tail property of financial data by ECBC with base copula M and check the impact on overall and tail error.

We use SPY 500 and NASDAQ daily return data, from Jan, 29th, 1993 to Jan, 9th, 2013.

We first use $AR(1)$ model to filter the return series to avoid auto-correlation. In the Durbin Watson test, the DW statistic for filtered SPY and NASDAQ return is 2.0005 and 2.0001, with a p value of 0.9964 and 0.9862. Based on the filtered time series, half of the samples are randomly chosen as the test data set. Parametric MLE of Gaussian

Table 4.4: L_1-error of ECBC and EBC with different m_1, m_2

Empirical composite Bernstein copula with M as the base copula

$m_1 \backslash m_2$	3	5	10	20	30	40	50	200
3	0.0095	0.0094	0.0111	0.0133	0.0149	0.0161	0.0161	0.0213
5	0.0089	0.0107	0.0106	0.0108	0.0127	0.0116	0.0118	0.0147
10	0.0106	0.0118	0.0107	0.0110	0.0118	0.0118	0.0117	0.0119
20	0.0136	0.0112	0.0113	0.0117	0.0119	0.0117	0.0122	0.0132
30	0.0148	0.0116	0.0119	0.0117	0.0132	0.0126	0.0125	0.0124
40	0.0151	0.0124	0.0119	0.0124	0.0127	0.0128	0.0133	0.0127
50	0.0166	0.0116	0.0118	0.0126	0.0131	0.0124	0.0125	0.0136
200	0.0215	0.0153	0.0126	0.0119	0.0130	0.0117	0.0137	0.0139

Empirical Bernstein copula

$m_1 \backslash m_2$	3	5	10	20	30	40	50	200
3	0.0371	0.0327	0.0292	0.0282	0.0267	0.0262	0.0256	0.0260
5	0.0341	0.0264	0.0212	0.0204	0.0198	0.0174	0.0179	0.0180
10	0.0300	0.0224	0.0183	0.0151	0.0147	0.0145	0.0149	0.0137
20	0.0280	0.0205	0.0151	0.0129	0.0126	0.0122	0.0128	0.0138
30	0.0270	0.0191	0.0153	0.0124	0.0137	0.0133	0.0132	0.0127
40	0.0258	0.0196	0.0145	0.0135	0.0132	0.0132	0.0135	0.0127
50	0.0262	0.0178	0.0147	0.0131	0.0133	0.0129	0.0129	0.0137
200	0.0263	0.0187	0.0143	0.0124	0.0132	0.0119	0.0138	0.0139

copula $\hat{C}^G$, parametric MLE of Gumbel copula $\hat{C}^{Gu}$, and ECBC with base copula M and Π are obtained from the test data. The distance between these estimators and the empirical copula C_N based on the training data are shown in the following Table 4.5.

L_1-error is applied here as the error measure:

$$R^{\alpha}(C^E, C_N) = \mathbb{E}\left\{\frac{1}{(K-1)^2}\sum_{i=1}^{K-1}\sum_{j=1}^{K-1}\left|C^E(\frac{i}{K}\alpha, \frac{j}{K}\alpha) - C_N(\frac{i}{K}\alpha, \frac{j}{K}\alpha)\right|\right\}, \tag{17}$$

where $K = 100$.

Table 4.5: L_1-error for different estimators, including parametric MLE of Gaussian copula $\hat{C}^G$, parametric MLE of Gumbel copula $\hat{C}^{Gu}$, and ECBC with base copula M and Π. Distance shown in the unit of 10^{-3}

	$\alpha = 1$		$\alpha = 0.05$		$\alpha = 0.01$	
	$R^1(\hat{C}^G, C_N) = 1.6292$		$R^{0.05}(\hat{C}^G, C_N) = 1.1262$		$R^{0.01}(\hat{C}^G, C_N) = 1.3118$	
	$R^1(\hat{C}^{Gu}, C_N) = 4.7681$		$R^{0.05}(\hat{C}^{Gu}, C_N) = 3.6895$		$R^{0.01}(\hat{C}^{Gu}, C_N) = 1.3997$	
m	$\tilde{C}_{m,m}(\cdot\|M)$	$\tilde{C}_{m,m}(\cdot\|\Pi)$	$\tilde{C}_{m,m}(\cdot\|M)$	$\tilde{C}_{m,m}(\cdot\|\Pi)$	$\tilde{C}_{m,m}(\cdot\|M)$	$\tilde{C}_{m,m}(\cdot\|\Pi)$
3	3.9409	37.8040	2.3443	10.2103	0.3606	2.3568
5	3.3739	24.4224	1.9946	9.4493	0.3079	2.3218
10	2.7408	12.4731	1.3858	8.0679	0.2501	2.2511
20	2.0906	6.0320	0.9791	6.1239	0.2936	2.1234
30	1.8766	3.8730	0.8258	5.0689	0.4261	2.0348
40	1.7323	2.8096	0.7716	4.2333	0.5217	1.9614
50	1.6297	2.2997	0.7702	3.8635	0.6431	1.9114
200	1.5406	1.6491	0.8837	1.6745	0.9355	1.4716

It can be clearly observed that by applying ECBC with M as base copula, not only we obtain a smaller overall error compared to EBC but also much better tail distribution estimation, beating the Gaussian MLE.

5 Conclusion

Based on Bernstein copula (BC) presented by Sancetta and Satchell (2004), this paper studied one new class of copula functions: the composite Bernstein copulas (CBC). A CBC is constructed by mixing the information of a base copula and a target copula.

The CBC converges to the target copula as $m_i \to \infty$ and the copula has nice theoretical properties. The CBC can also be used to model tail dependence. The empirical CBC (ECBC) was introduced as a non-parametric estimation procedure, and its asymptotic properties were shown. The ECBC is able to incorporate prior information flexibly with difference choices of base copulas and the parameter m. Simulation study and empirical analysis of financial data showed the advantage of the new estimation method, especially in capturing tail dependence. We remark that in the ECBC estimation procedure, the optimal choice of m is still unclear and is a possible research direction for future study.

Acknowledgments. We thank two referees for helpful suggestions. J. Yang's research was partly supported by the Key Program of National Natural Science Foundation of China (Grants No. 11131002) and the National Natural Science Foundation of China (Grants No. 11271033). F. Wang's research was supported by the National Natural Science Foundation of China (Grant No.11471222) and Foundation of Beijing Education Bureau (Grant No.201510028002). R. Wang's research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

[1] R. Baker (2008). *An order-statistics-based method for constructing multivariate distributions with fixed marginals.* Journal of Multivariate Analysis, Vol.99, 2312-2327.

[2] U. Cherubini, E. Luciano and W. Vecchiato (2004). Copula methods in finance. John Wiley & Sons, England.

[3] C. Czado (2009). *Pair-Copula constructions of multivariate copulas.* In Copula theorey and its applications, 93-110, Ed. P. Jaworski, F. Durante, W. Häedle and T. Rychlik. Springer, Berlin.

[4] J. Dhaene, M. Denuit, M.J. Goovaerts, R. Kaas, and D. Vyncke (2002a). *The concept of comonotonicity in actuarial science and finance: Theory* . Insurance: Mathematics and Economics, Vol.31(1), 3-33.

[5] J. Dhaene, M. Denuit, M.J. Goovaerts, R. Kaas, and D. Vyncke (2002b). *The concept of comonotonicity in actuarial science and finance: Application.* Insurance: Mathematics and Economics, Vol.31(2), 133-161.

[6] C. Donnelly and P. Embrechts. *The Devil is in the Tails: Actuarial Mathematics and the Subprime Mortgage Crisis.* ASTIN Bulletin, Vol.40(1), 1-33.

[7] D. Diers, M. Eling and S.D. Marek (2012). *Dependence modeling in non-life insurance using the Bernstein copula.* Insurance: Mathematics and Economics, Vol.50, 430-436.

[8] X. L. Dou, S. Kuriki and G. D. Lin (2013). *EM algorithms for estimating the Bernstein copula function.* Preprint available at
`http://arxiv.org/pdf/1301.2677v1.pdf`

[9] K. T. Fang, S. Kotz and K. Ng (1990). Symmetric multivariate and related distributions. Chapman and Hall, London.

[10] G. Frahm, M. Junker and A. Szimayer (2003). *Elliptical copulas: applicability and limitations.* Statistics & Probability letters, Vol.63(3), 275-286.

[11] M. Hofert (2009). *Construction and sampling of nested Archimedean copulas.* In Copula theorey and its applications, 147-160, Ed. P. Jaworski, F. Durante, W. Häedle and T. Rychlik. Springer, Berlin.

[12] P. Janssen, J. Swanepoel and N. Veraverbeke (2012). *Large sample behavior of the Bernstein copula estimator.* Journal of Statistical Planning and Inference, Vol.142, 1189-1197.

[13] H. Joe (1997). Multivariate models and dependence concepts. Chapman and Hall, London.

[14] A. J. McNeil, R. Frey, and P. Embrechts (2005). Quantitative risk management: concepts, techniques, tools. Princeton, NJ: Princeton University Press.

[15] R. B. Nelsen (2006). An introduction to copulas. Second Edition. Springer Science+Business Media, Inc.

[16] A. Sancetta (2007). *Online forecast combinations of distributions: Worst case bounds*. Journal of Ecomometrics, Vol. 141, 621-651.

[17] A. Sancetta and S. Satchell (2004). *The Bernstein copula and its applications to modeling and approximations of multivariate distributions*. Ecomometric Theory, Vol.20, 535-562.

[18] R. Schmidt (2004). *Tail dependence*. In Statistical tools in finance and insurance, Ed. W. Häedle, P. Cizek and R. Weron, Springer Verlag, 65-91.

[19] B. Tavin (2012). *Detection of arbitrage in a market with multi-asset derivatives and known risk-neutral marginals*. Preprint available at
`http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2080047`

[20] G. N. F. Weiβ and M. Scheffer (2012). *Smooth nonparametric Bernstein Vine copulas*. Preprint available at `http://arxiv.org/pdf/1210.2043.pdf`

A Proof of Theorem 3.1

Part (1) is implied by the Law of Large Numbers and Theorem 2.1 (iii). In the following we show part (2).

Note that

$$\begin{aligned}
&\sup_{0\le u_i\le 1, i\le n} |\tilde{C}_{m_1,\dots,m_n}(u_1,\dots,u_n|N,D) - C(u_1,\dots,u_n)| \\
&\le \sup_{0\le u_i\le 1, i\le n} |\tilde{C}_{m_1,\dots,m_n}(u_1,\dots,u_n|N,D) - C_{m_1,\dots,m_n}(u_1,\dots,u_n|C,D)| \\
&\quad + \sup_{0\le u_i\le 1, i\le n} |C_{m_1,\dots,m_n}(u_1,\dots,u_n|C,D) - C(u_1,\dots,u_n)|. \qquad (18)
\end{aligned}$$

For the first term of the right-hand of the inequality (18), from (3) and (13) we get

$$\begin{aligned}
&\sup_{0\le u_i\le 1, i\le n} |\tilde{C}_{m_1,\dots,m_n}(u_1,\dots,u_n|N,D) - C_{m_1,\dots,m_n}(u_1,\dots,u_n|C,D)| \\
&= \sup_{0\le u_i\le 1, i\le n} |\sum_{l_1=0}^{m_1}\cdots\sum_{l_n=0}^{m_n}(C_N(\frac{l_1}{m_1},\dots,\frac{l_n}{m_n}) - C(\frac{l_1}{m_1},\dots,\frac{l_n}{m_n})) \\
&\qquad \times\mathbb{P}(F^{-1}_{Bin(m_1,u_1)}(U_1^D) = l_1,\dots,F^{-1}_{Bin(m_n,u_n)}(U_n^D) = l_n)| \\
&\le \sup_{0\le l_i\le m_i, i\le n} |C_N(\frac{l_1}{m_1},\dots,\frac{l_n}{m_n}) - C(\frac{l_1}{m_1},\dots,\frac{l_n}{m_n})| \\
&\le \sup_{0\le u_i\le 1, i\le n} |C_N(u_1,\dots,u_n) - C(u_1,\dots,u_n))|.
\end{aligned}$$

Thus we obtain that

$$\begin{aligned}
&\sup_{0\le u_i\le 1, i\le n} |\tilde{C}_{m_1,\dots,m_n}(u_1,\dots,u_n|N,D) - C_{m_1,\dots,m_n}(u_1,\dots,u_n|C,D)| \\
&\le \sup_{0\le u_i\le 1, i\le n} |C_N(u_1,\dots,u_n) - C(u_1,\dots,u_n))|), \quad \text{a.s.} \qquad (19)
\end{aligned}$$

For the second term of the right-hand of the inequality (18), note that

$$\begin{aligned}
&|C_{m_1,\dots,m_n}(u_1,\dots,u_n|C,D) - C(u_1,\dots,u_n)| \\
&= |\mathbb{E}[C(\frac{F^{-1}_{Bin(m_1,u_1)}(U_1)}{m_1},\dots,\frac{F^{-1}_{Bin(m_n,u_n)}(U_n))}{m_n}) - C(u_1,\dots,u_n)]| \\
&\le \sum_{i=1}^n \mathbb{E}|\frac{F^{-1}_{Bin(m_1,u_1)}(U_1)}{m_i} - u_i| \\
&\le \sum_{i=1}^n \sqrt{\mathrm{Var}(\frac{F^{-1}_{Bin(m_i,u_i)}(U_i)}{m_i})} = O(\frac{1}{\sqrt{\underline{m}}}). \qquad (20)
\end{aligned}$$

Combining (19) and (20), we can get

$$
\begin{aligned}
& \sup_{0\le u_i\le 1, i\le n} |\tilde{C}_{m_1,\ldots,m_n}(u_1,\ldots,u_n|N,D) - C(u_1,\ldots,u_n)| \\
\le & \sup_{0\le u_i\le 1, i\le n} |C_N(u_1,\ldots,u_n) - C(u_1,\ldots,u_n))|) + O(\frac{1}{\sqrt{\underline{m}}}) \\
= & O_P(\frac{1}{\sqrt{N}}) + O(\frac{1}{\sqrt{\underline{m}}}) = O_P(\frac{1}{\min\{\sqrt{N},\sqrt{\underline{m}}\}}). \quad \square
\end{aligned} \tag{21}
$$

B Proof of Theorem 3.2

For $i \le N$, denote

$$
\begin{aligned}
Y_{mi} \quad = \quad & \sum_{l_1=0}^{m}\sum_{l_2=0}^{m}\{(I(V_{i,1}\le \frac{l_1}{m}, V_{i,2}\le \frac{l_2}{m}) - C(\frac{l_1}{m},\frac{l_2}{m})) \\
& \quad \mathbb{P}(F^{-1}_{Bin(m,u_1)}(U^D_1) = l_1, F^{-1}_{Bin(m,u_2)}(U^D_2) = l_2)\}.
\end{aligned}
$$

Then

$$
\begin{aligned}
& N^{1/2}(\tilde{C}_{m,m}(u_1,u_2|N,D) - C_{m,m}(u_1,u_2|C,D)) \\
= \quad & N^{-1/2}\sum_{i=1}^{N}\sum_{l_1=0}^{m}\sum_{l_2=0}^{m}\{(I(V_{i,1}\le \frac{l_1}{m}, V_{i,2}\le \frac{l_2}{m}) - C(\frac{l_1}{m},\frac{l_2}{m})) \\
& \quad \times\mathbb{P}(F^{-1}_{Bin(m,u_1)}(U^D_1) = l_1, F^{-1}_{Bin(m,u_2)}(U^D_2) = l_2)\} \\
= \quad & N^{-1/2}\sum_{i=1}^{N} Y_{mi}.
\end{aligned} \tag{22}
$$

Note that $Y_{mi}, i \le N$ are independent and identically distributed random variables. For simplicity, we denote

$$
\begin{aligned}
N_{1,1}(u_1) = F^{-1}_{Bin(m,u_1)}(U^D_{1,1}), N_{1,2}(u_2) = F^{-1}_{Bin(m,u_2)}(U^D_{1,2}), \\
N_{2,1}(u_1) = F^{-1}_{Bin(m,u_1)}(U^D_{2,1}), N_{1,2}(u_2) = F^{-1}_{Bin(m,u_2)}(U^D_{2,2}),
\end{aligned}
$$

where $(U^D_{1,1}, U^D_{1,2})$ and $(U^D_{2,1}, U^D_{2,2})$ are independent random vectors with distribution $\bar{D}$. For $i = 1, 2$ and $j = 1, 2$,

$$
\frac{N_{i,j}(u_j)}{m} - u_j = \frac{N_{i,j}(u_j) - mu_j}{\sqrt{m}}\frac{1}{\sqrt{m}}.
$$

Note that

$$\frac{N_{i,j}(u_j) - mu_j}{\sqrt{m}} \xrightarrow{d} N(0, u_j(1-u_j))$$

and as $m \to \infty$,

$$\begin{aligned}
&\mathbb{P}(\frac{N_{i,1}(u_1) - mu_1}{\sqrt{mu_1(1-u_1)}} \le s, \frac{N_{i,2}(u_2) - mu_2}{\sqrt{mu_2(1-u_2)}} \le t) \\
&= \bar{D}(\mathbb{P}(\frac{N_{i,1}(u_1) - mu_1}{\sqrt{mu_1(1-u_1)}} \le s), \mathbb{P}(\frac{N_{i,2}(u_2) - mu_2}{\sqrt{mu_2(1-u_2)}} \le t)) \\
&\to \bar{D}(\Phi(s), \Phi(t)).
\end{aligned}$$

Thus we have

$$\begin{aligned}
& \mathrm{Var}(Y_{mi}) \\
= \;& \mathbb{E}[\mathbb{E}\{(I(V_{i,1} \le \frac{N_{1,1}(u_1)}{m}, V_{i,2} \le \frac{N_{1,2}(u_2)}{m}) - C(\frac{N_{1,1}(u_1)}{m}, \frac{N_{1,2}(u_2)}{m})) \\
& \times(I(V_{i,1} \le \frac{N_{2,1}(u_1)}{m}, V_{i,2} \le \frac{N_{2,2}(u_2)}{m}) - C(\frac{N_{2,1}(u_1)}{m}, \frac{N_{2,2}(u_2)}{m}))|V_{i,1}, V_{i,2}\}] \\
= \;& \mathbb{E}[(I(V_{i,1} \le \frac{N_{1,1}(u_1) \wedge N_{2,1}(u_1)}{m}, V_{i,2} \le \frac{N_{1,2}(u_2) \wedge N_{2,2}(u_2)}{m}) \\
& -I(V_{i,1} \le \frac{N_{1,1}(u_1)}{m}, V_{i,2} \le \frac{N_{1,2}(u_2)}{m}) \times C(\frac{N_{2,1}(u_1)}{m}, \frac{N_{2,2}(u_2)}{m}) \\
& -C(\frac{N_{1,1}(u_1)}{m}, \frac{N_{1,2}(u_2)}{m}) I(V_{i,1} \le \frac{N_{2,1}(u_1)}{m}, V_{i,2} \le \frac{N_{2,2}(u_2)}{m}) \\
& +C(\frac{N_{1,1}(u_1)}{m}, \frac{N_{1,2}(u_2)}{m}) C(\frac{N_{2,1}(u_1)}{m}, \frac{N_{2,2}(u_2)}{m})] \\
= \;& \mathbb{E}[C(\frac{N_{1,1}(u_1) \wedge N_{2,1}(u_1)}{m}, \frac{N_{1,2}(u_2) \wedge N_{2,2}(u_2)}{m}) \\
& -C(\frac{N_{1,1}(u_1)}{m}, \frac{N_{1,2}(u_2)}{m}) C(\frac{N_{2,1}(u_1)}{m}, \frac{N_{2,2}(u_2)}{m})] \\
= \;& C(u_1, u_2) - C(u_1, u_2)^2 + \frac{2}{\sqrt{m}} \frac{\partial C(u_1, u_2)}{\partial u_1} \sqrt{u_1(1-u_1)} \mathbb{E}(Z_1 \wedge Z_2) \\
& \qquad + \frac{2}{\sqrt{m}} \frac{\partial C(u_1, u_2)}{\partial u_2} \sqrt{u_2(1-u_2)}) \mathbb{E}(Z_1 \wedge Z_2) + o(\frac{1}{\sqrt{m}}),
\end{aligned}$$

where Z_1 and Z_2 are independent $\mathrm{N}(0,1)$ random variables. It is easy to verify that

$$\mathbb{E}(Z_1 \wedge Z_2) = -\frac{1}{\sqrt{\pi}}.$$

Finally,

$$
\begin{aligned}
& \mathrm{Var}(Y_{mi}) \\
= \; & C(u_1,u_2) - C(u_1,u_2)^2 - \frac{2}{\sqrt{m}\sqrt{\pi}} \frac{\partial C(u_1,u_2)}{\partial u_1} \sqrt{u_1(1-u_1)} \\
& \quad - \frac{2}{\sqrt{m}\sqrt{\pi}} \frac{\partial C(u_1,u_2)}{\partial u_2} \sqrt{u_2(1-u_2)} + o(\frac{1}{\sqrt{m}}) \\
= \; & \sigma(u_1,u_2) - \frac{1}{\sqrt{m}} V(u_1,u_2) + o(\frac{1}{\sqrt{m}}).
\end{aligned}
$$

Thus from (22) we know that

$$
N^{1/2}(\tilde{C}_{m,m}(u_1,u_2|N,D) - C_{m,m}(u_1,u_2|C,D)) \xrightarrow{d} N(0,\sigma^2(u_1,u_2)). \tag{23}
$$

On the other-hand, under the condition of Theorem 3.2, we have

$$
\begin{aligned}
& C_{m,m}(u_1,u_2|C,D) - C(u_1,u_2) \\
& = \mathbb{E}[C(\frac{F^{-1}_{Bin(m,u_1)}(U^D_1)}{m}, \frac{F^{-1}_{Bin(m,u_2)}(U^D_2)}{m}) - C(u_1,u_2)] \\
& = \frac{1}{2}\frac{\partial^2 C(u_1,u_2)}{\partial u_1^2} \mathbb{E}(\frac{F^{-1}_{Bin(m,u_1)}(U^D_1)}{m} - u_1)^2 + \frac{1}{2}\frac{\partial^2 C(u_1,u_2)}{\partial u_2^2} E(\frac{F^{-1}_{Bin(m,u_2)}(U^D_2)}{m} - u_2)^2 \\
& \quad + \frac{\partial^2 C(u_1,u_2)}{\partial u_1 \partial u_2} \mathbb{E}[(\frac{F^{-1}_{Bin(m,u_1)}(U^D_1)}{m} - u_1)(\frac{F^{-1}_{Bin(m,u_2)}(U^D_2))}{m} - u_2)] + o(\frac{1}{m}) \\
& = \frac{1}{2m}\{\frac{\partial^2 C(u_1,u_2)}{\partial u_1^2} u_1(1-u_1) + \frac{\partial^2 C(u_1,u_2)}{\partial u_2^2} u_2(1-u_2)\} \\
& \quad + \frac{1}{m^2}\frac{\partial^2 C(u_1,u_2)}{\partial u_1 \partial u_2} \mathrm{Cov}(F^{-1}_{Bin(m,u_1)}(U^D_1), F^{-1}_{Bin(m,u_2)}(U^D_2)) + o(m^{-1}).
\end{aligned} \tag{24}
$$

Note that

$$
(\frac{F^{-1}_{Bin(m,u_1)}(U^D_1) - mu_1}{\sqrt{mu_1(1-u_1)}}, \frac{F^{-1}_{Bin(m,u_2)}(U^D_2) - mu_2}{\sqrt{mu_2(1-u_2)}}) \xrightarrow{d} \bar{D}(\Phi(s),\Phi(t)),
$$

thus

$$
\begin{aligned}
& \mathrm{Cov}(\frac{F^{-1}_{Bin(m,u_1)}(U^D_1) - mu_1}{\sqrt{mu_1(1-u_1)}}, \frac{F^{-1}_{Bin(m,u_2)}(U^D_2) - mu_2}{\sqrt{mu_2(1-u_2)}}) \\
& = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} st d\bar{D}(\Phi(s),\Phi(t)) + o(1).
\end{aligned}
$$

Thus from (24) we get

$$
C_{m,m}(u_1,u_2|C,D) - C(u_1,u_2) = m^{-1} b(u_1,u_2) + o(m^{-1}). \tag{25}
$$

Combining equation (23) and equation (25), we get

$$
\begin{aligned}
&N^{1/2}(\tilde{C}_{m,m}(u_1,u_2|N,D) - C(u_1,u_2|C,D)) \\
&= N^{1/2}(\tilde{C}_{m,m}(u_1,u_2|N,D) - C_{m,m}(u_1,u_2|C,D)) + \frac{N^{1/2}}{m}(b(u_1,u_2) + o(1)) \\
&\xrightarrow{d} N(ab(u_1,u_2), \sigma^2(u_1,u_2)). \quad \square
\end{aligned}
$$