The Bernstein Copula and its Applications to Modelling and Approximations of Multivariate Distributions*

A. Sancetta and S. E. Satchell

Trinity College, Cambridge

Abstract

We define the Bernstein copula and study its statistical properties in terms of both distributions and densities. We also develop a theory of approximation for multivariate distributions in terms of Bernstein copulae To further motivate the introduction of this new object, we present a simulated example in the context of portfolio optimization. Rates of consistency when the Bernstein copula density is estimated nonparametrically are given. In order of magnitude, this estimator has variance equal to the square root of other nonparametric estimators, e.g. kernel smothers, but it is biased as an histogram estimator.

---

*We would like to thank Mark Salmon for interesting us to the copula function, and Peter Phillips, an associate editor and the referees for many valuable comments. All remaining errors are our sole responsability.

1 Introduction

The task of modelling multivariate distributions has always been a challenging one. For this reason, it is very common to use elliptic distributions due to the fact that these are simple to characterize. However, in many cases of interest in economics, it is found that this simple characterization contrasts with empirical evidence. A typical example are returns distributions in financial economics and the complex range of dependence that they exhibit. Among many references, the reader should look at Embrechts et al. (1999).

When dealing with vectors of random variables, the copula function becomes a very useful object because it allows us to model the dependence between the variables separately from their marginals. While the copula function is a fairly new concept in econometrics, there has been a growing interest in financial econometrics in these very recent years: for example, Bouye et al. (2001), Patton (2001), Rockinger and Jondeau (2001), and Longin and Solnik (2000). The last authors deal with multivariate extreme value theory in the context of financial assets and use a dependence function equivalent to the copula function. The copula has been considered elsewhere to study extreme values; see Joe (1997) for a list of extreme value copulae.

We introduce a family of copulae defined as Bernstein polynomials. This new representation leads to a general approach in estimation as well as simplifications of many operations whenever a parametric copula is given. Moreover, for parametric copulae, the Bernstein representation is useful in studying properties of the copula function itself.

There are many possible representations of continuous functions in terms of polynomials, e.g. Hermite polynomials (the Edgeworth expansion) and Padé approximations (the extended rational polynomials in Phillips, 1982, 1983). However, none of these polynomial representations share the same properties of Bernstein polynomials in the context of the copula function. We make this clear in the text. However, just to give a hint, we can say that Bernstein polynomials are closed under differentiation; for simple restrictions on the coefficients they always lead to a proper copula function; when used for nonparametric estimation, their rate of convergence in mean square error has lower variance than other nonparametric estimators.

The plan for the paper is as follows. Section 2 introduces the Bernstein copula and derives some of its mathematical and statistical properties. In section 2.4 we introduce the important extension of Bernstein representation of given copulae. In order to provide further motivation for our definition, the Bernstein copula is applied to a particular instance of portfolio optimization where the marginals are Weibull distributions; this is done in Section 3. Issues related to making the polynomial representation operational are deferred to section 4 where, as mentioned above, it is shown that the nonparametric Bernstein copula density may provide some solution to the curse of dimensionality. Other estimation procedures are possible, and discuss some, but by no means all of the issues, in Section 5.

2 The Bernstein Copula

For expositional convenience, we recall Sklar's representation for multivariate distributions. Let H be a k dimensional distribution function with 1-dimensional margins $F_1, ..., F_k$, then there exists a function C from the unit k-cube to the unit cube such that

$$H(x_1, ..., x_k) = C\left(F_1(x_1), ..., F_k(x_k)\right);$$

C is referred to as the k-Copula. If each F_j is continuous, the copula is unique. For more details and a proof see Sklar (1973).

Let $\alpha\left(\frac{v_1}{m_1}, ..., \frac{v_k}{m_k}\right)$ be a real valued constant indexed by $(v_1, ..., v_k)$, $v_j \in \mathbb{N}_+$, such that $0 \le v_j \le m_j$. While we could simply use $\alpha_{v_1,...,v_k}$, $0 \le v_j \le m_j$, $\forall j$, for conceptual convenience we do not do so. Now, we define the object to be studied in this paper.

Definition 1. *Let*

$$P_{v_j,m_j}(u_j) \equiv \binom{m_j}{v_j} u_j^{m_j - v_j}\left(1 - u_j\right)^{m_j - v_j}. \tag{1}$$

If $C_B : [0,1]^k \to [0,1]$, where

$$C_B\left(u_1, ..., u_k\right) = \sum_{v_1} ... \sum_{v_k} \alpha\left(\frac{v_1}{m_1}, ..., \frac{v_k}{m_k}\right) P_{v_1,m_j}(u_1) \cdots P_{v_K,m_j}(u_k), \tag{2}$$

satisfies the properties of the copula function, then C_B is a Bernstein copula for any $m_j \ge 1$.

The Bernstein copula generalizes families of polynomial copulae. Polynomial copulae are special cases of copulae with polynomial sections in one or more variables. For the two dimensional case,

$$C(u_1, u_2) = u_1^3 a(u_2) + u_1^2 b(u_2) + u_1 c(u_2) + d(u_2)$$

is a proper copula for suitable choice of functions $a(...)$, $b(...)$, $c(...)$ and $d(...)$. This copula has cubic sections and the just mentioned functions can be polynomials. Then, they can be exactly written as a Bernstein copula. More details on copulae with polynomial sections can be found in Nelsen (1998, ch. 3).

In order to study the properties of the Bernstein copula, it is convenient to recall some properties of Bernstein polynomials.[1] We state the following as a theorem because we could not find a reference in the literature apart from the well known univariate case. Without loss of generality we consider the unit hypercube instead of arbitrary compact subsets of $\mathbb{R}^k$.[2]

Theorem 1. *Let $C_{[0,1]^k}$ be the space of bounded continuous functions in the k dimensional hypercube $[0,1]^k$. Then, the set of Bernstein polynomials defined in (3) is dense in $C_{[0,1]^k}$.*

Proof. See the Appendix. ■

We encourage the reader to read the proof of Theorem 1 in the Appendix before proceeding any further in order to become familiar with some properties of Bernstein polynomials that will be assumed below. In particular, Bernstein polynomials can be represented as a linear operator B_m^k such that given $f \in C_{[0,1]^k}$, then

$$(B_m^k f)(X) \equiv \sum_{v_1=0}^{m_1} ... \sum_{v_k=0}^{m_k} f\left(\frac{v_1}{m_1}, ..., \frac{v_k}{m_k}\right) P_{v_1,m_1}(x_1) \cdots P_{v_k,m_k}(x_k), \tag{3}$$

or in the more general singular integral representation via the Stieltjes integral

$$(B_m^k f)(X) = \int_0^1 \cdots \int_0^1 f(t_1, .., t_k) d_{t_1} K_{m_1}(x_1, t_1) \cdots d_{t_k} K_{m_k}(x_k, t_k),$$

for the kernel

$$\begin{aligned} K_m(x,t) &\equiv \sum_{v \le mt} \binom{m}{v} x^v (1-x)^{m-v}, \\ K_m(x,0) &\equiv 0, \end{aligned}$$

which is constant for $\frac{v}{m} \le t < \frac{v+1}{m}$ and has jumps of $\binom{m}{v} x^v (1-x)^{m-v}$ at points $t = \frac{v}{m}$. This representation establishes some clear parallels to kernel density estimation in statistics. This idea is implicitly exploited in Section 4.

Throughout, we reserve the symbols C_B, C_n and C for the Bernstein copula, the empircal copula based on n observations, and a general copula, respectively, and use c_B for the Bernstein copula density; definitions will be given in due course. For simplicity, $m_j = m \; \forall j$. Therefore, the letters m and n are only used to define the order of polynomial and the sample size, respectively.

2.1 Some Properties of the Bernstein Copula

We list some properties of C_B in common with all other copulae:

(1) C_B is increasing in all its arguments. Notice that throughout the paper, we use increasing to mean nondecreasing;

(2) C_B satisfies the Fréchet bounds, i.e.

$$\min\left(0, u_1 + ... + u_k - (k-1)\right) \le C_B\left(u_1, ..., u_k\right) \le \min\left(u_1, ..., u_k\right),$$

which implies C_B is grounded: i.e. $C_B(u_1, ..., u_k) = 0$ if $u_j = 0$ for at least one j, and $C_B(1, ..., 1, u_j, 1, ..., 1) = u_j$, $\forall j$;

(3) $\prod_{j=1}^{k} u_j$ is a copula for independent random variables, i.e. the product copula;

(4) C_B is Lipschitz, i.e.

$$\left|C_B\left(x_1, ..., x_k\right) - C_B\left(y_1, ..., y_k\right)\right| \leq \sum_{j=1}^{k} \left|x_j - y_j\right|.$$

In light of these properties, the next result shows related properties specific to the Bernstein copula

Theorem 2. *$C_B\left(u_1, ..., u_k\right)$ is a Bernstein copula if and only if*

$$\sum_{l_1=0}^{1} .... \sum_{l_k=0}^{1} (-1)^{l_1+....+l_k} \alpha\left(\frac{v_1+l_1}{m}, ..., \frac{v_k+l_k}{m}\right) \geq 0 \tag{4}$$

and

$$\min\left(0, \frac{v_1}{m} + ... + \frac{v_k}{m} - (k-1)\right) \leq \alpha\left(\frac{v_1}{m}, ..., \frac{v_k}{m}\right) \leq \min\left(\frac{v_1}{m}, ..., \frac{v_k}{m}\right),$$

in particular

$$\lim_{v_j \to 0} \alpha\left(\frac{v_1}{m}, ..., \frac{v_k}{m}\right) = 0, \ \forall j = 1, ..., k, \tag{5}$$

and

$$\alpha\left(1, ..., 1, \frac{v_j}{m}, 1, ..., 1\right) = \frac{v_j}{m}, \ \forall j = 1, ..., k. \tag{6}$$

Proof. Consider the Bernstein copula $C_B\left(u_1, ..., u_k\right)$ as an approximation to a copula $C\left(u_1, ..., u_k\right)$, i.e.

$$C\left(\frac{v_1}{m}, ..., \frac{v_k}{m}\right) = \alpha\left(\frac{v_1}{m}, ..., \frac{v_k}{m}\right).$$

Then Theorem 2 follows by the definition of copula function. ■

It is clear that the coefficients of the Bernstein copula have a direct interpretation as the points of some arbitrary approximated copula. However, in the context of Theorem 2, the Bernstein copula should not be understood as an approximation but a generalization of polynomial families of copulae in virtue of Theorem 1.

In some cases, it is useful to consider the following representation of the Bernstein copula as the sum of the product copula and a perturbation term,

$$\begin{aligned} C_B(u_1,...,u_k) &= u_1\cdots u_k + \sum_{v_1=0}^{m}...\sum_{v_k=0}^{m}\gamma\left(\frac{v_1}{m},...,\frac{v_k}{n}\right)P_{v_1,m}(u_1)\cdots P_{v_K,m}(u_k) \\ &= \sum_{v_1=0}^{m}...\sum_{v_k=0}^{m}\alpha\left(\frac{v_1}{m},...,\frac{v_k}{m}\right)P_{v_1,m}(u_1)\cdots P_{v_K,m}(u_k), \end{aligned} \tag{7}$$

where

$$\gamma\left(\frac{v_1}{m},...,\frac{v_k}{m}\right) = \alpha\left(\frac{v_1}{m},...,\frac{v_k}{m}\right) - \frac{v_1}{m}\cdots\frac{v_k}{m}. \tag{8}$$

The equality follows from the fact that

$$u_1\cdots u_k = \sum_{v_1}...\sum_{v_k}\left(\frac{v_1}{n}\cdots\frac{v_k}{n}\right)P_{v_1,n}(u_1)\cdots P_{v_K,n}(u_k);$$

e.g. see the proof of Theorem 1. This leads to the following important decomposition.

Theorem 3. *Any copula $C(u_1,...,u_k)$ can be written as $u_1\cdots u_k + G(u_1,...,u_k)$, where $u_1\cdots u_k$ is the product copula and $G(u_1,...,u_k)$ is a perturbation term containing all information about the dependence of $(u_1,...,u_k)$.*

Proof. Consider (7), then use uniform convergence of Bernstein polynomials. ■

Notice that $G(u_1,...,u_k)$ is the distance of the copula from the product copula. This is bounded above and below by the Fréchet bounds. For a 2-copula, the Fréchet bounds define a skewed quadrilateral where the product copula is the paraboloid inside it.

2.2 The Bernstein Density

Any Bernstein copula has a copula density; this is because the Bernstein copula is absolutely continuous. Define $\Delta_{1,\ldots,k}$ as the k dimensional forward difference operator, i.e.

$$\Delta_{1,\ldots,k}\alpha\left(\frac{v_1}{m},...,\frac{v_k}{m}\right) \equiv \sum_{l_1=0}^{1}...\sum_{l_k=0}^{1}(-1)^{k+l_1+\ldots+l_k}\,\alpha\left(\frac{v_1+l_1}{m},...,\frac{v_k+l_k}{m}\right).$$

Due to the convexity preserving properties of Bernstein polynomials, the Bernstein copula density has the following appealing structure

$$\begin{aligned} c_B\left(u_1,...,u_k\right) &= m^k\sum_{v_1=0}^{m-1}...\sum_{v_k=0}^{m-1}\Delta_{1,\ldots,k}\alpha\left(\frac{v_1}{m},...,\frac{v_k}{m}\right) \\ &\quad \times P_{v_1,m-1}(u_1)\cdots P_{v_K,m-1}(u_k), \end{aligned}$$

where $c_B = \frac{\partial^k C_B}{\partial u_1\cdots\partial u_k}$ and the expression is obtained by direct differentiation of (2) with respect to each variable and rearranging; see Lorentz (1953) for the univariate case. Differentiating, a term in the summation is lost, and the coefficients of the polynomial are written in difference form which is directly linked to the k dimensional rectangle inequality in (4), i.e. the copula density is always positive.

For convenience, we use the following definition for the Bernstein copula density,

$$
\begin{aligned}
c_B(u_1, ..., u_k) &= \sum_{v_1=0}^{m} ... \sum_{v_k=0}^{m} \beta\left(\frac{v_1}{m}, ..., \frac{v_k}{m}\right) \\
&\quad \times \prod_{j=1}^{k} \binom{m}{v_j} u_j^{v_j} (1-u_j)^{m-v_j}, \qquad (9)
\end{aligned}
$$

where $\beta\left(\frac{v_1}{m}, ..., \frac{v_k}{m}\right)$ is defined accordingly, i.e.

$$
\beta\left(\frac{v_1}{m}, ..., \frac{v_k}{m}\right) \equiv (m+1)^k \Delta_{1,...,k}\alpha\left(\frac{v_1}{m+1}, ..., \frac{v_k}{m+1}\right). \qquad (10)
$$

2.3 Spearman's Rho and the Moment Generating Function of the Bernstein Copula

We consider the moments of the Bernstein copula. Many operations find convenient representation in terms of hypergeometric functions; see Abadir (1999) for an introduction to economists and many of the symbols we use.

The copula is

$$
\begin{aligned}
C_B(u_1, ..., u_k) &= \sum_{v_1=0}^{m} ... \sum_{v_k=0}^{m} \alpha\left(\frac{v_1}{m}, ..., \frac{v_k}{m}\right) \\
&\quad \times \prod_{j=1}^{k} \binom{m}{v_j} u_j^{v_j} (1-u_j)^{n_j-v_j},
\end{aligned}
$$

and its bivariate marginal distribution, say for u_1 and u_2, is

$$
\begin{aligned}
C_B(u_1, u_2, 1, ..., 1) &= \sum_{v_1=0}^{m} \sum_{v_2=0}^{m} \alpha\left(\frac{v_1}{m}, \frac{v_2}{m}, 1, ..., 1\right) \\
&\quad \times \prod_{j=1,2} \binom{m}{v_j} u_j^{v_j} (1-u_j)^{m-v_j}.
\end{aligned}
$$

We now calculate Spearman's rho (ρ_S). Using well known properties of the uniform distributions on $[0,1]$, namely that it has mean $\frac{1}{2}$ and variance $\frac{1}{12}$,

$$
\begin{aligned}
\rho_S &= 12cov\left(u,v\right) \\
&= 12E\left(uv\right) - 3,
\end{aligned}
$$

where

$$
\begin{aligned}
E\left(uv\right) &= \int uvdC\left(u,v\right) \\
&= \int \left(1-u-v+C\left(u,v\right)\right)dudv,
\end{aligned}
$$

using integration by parts. It should be noted that ρ_S is independent of the definition of the marginals whereas Pearson's correlation coefficient (i.e. conventional correlation) does depend upon the marginal distributions; see Schweizer and Wolff (1981) for further discussion. Random variables which have zero covariance, could have non-zero ρ_S. The use of ρ_S in financial economics could be advocated on the basis of the documented non-linearities and its simple estimation. For the Bernstein copula ρ_S is equal to,

$$
\begin{aligned}
\rho_S &\equiv 12\int_0^1\int_0^1 \left[1-u_1-u_2+C_B(u_1,u_2,1,...,1)\right]du_1du_2 - 3 \\
&= 12\sum_{v_1=0}^{m}\sum_{v_2=0}^{m}\gamma\left(\frac{v_1}{m},\frac{v_2}{m},1,...,1\right) \\
&\quad \times \prod_{j=1,2}\binom{m}{v_j}\int_0^1\int_0^1 u_j^{v_j}\left(1-u_j\right)^{m-v_j}du_1du_2 \\
&= 12\sum_{v_1=0}^{m}\sum_{v_2=0}^{m}\gamma\left(\frac{v_1}{m},\frac{v_2}{m},1,...,1\right) \\
&\quad \times \prod_{j=1,2}\binom{m}{v_j}B\left(v_j+1,m+1-v_j\right),
\end{aligned}
$$

where γ was defined in (8) and $B(a,b)$ is the beta function. The first equality follows by writing the Bernstein copula as the sum of the product copula and the perturbation term. Notice that

$$12\int_0^1\int_0^1 (1-u_1-u_2+u_1u_2)\,du_1du_2 = 3.$$

As shown in Proposition 2, all dependency information is contained in the perturbation term. Even when the Bernstein copula is used as an approximation (see next subsection), the above Spearman's rho can be used as an approximation to the true Spearman's rho of any copula. If enough terms of our proposed Bernstein approximation are included, Spearman's rho can be easily found to any degree of accuracy without the need of evaluating complicated integrals.

For the sake of completeness the moment generating function of the density in (9) is found. We do it for the one variable case. Then we just extend it to the k dimensional case.

$$\begin{aligned} M_u(t) &= \int_0^1 \exp\{tu\}\, c(u)du \\ &= \sum_{v=0}^{n} \beta\left(\frac{v}{m}\right)\binom{m}{v}\int_0^1 \exp\{tu\}\, u^v(1-u)^{m-v}du, \end{aligned}$$

where $\beta\left(\frac{v}{m}\right)$ is given by (10) for $k=1$. Before proceeding any further, we notice the following (see Marichev, 1983, p. 87),

$${}_1F_1\,(a;c;z)\,B\,(a,c) = \int_0^1 \exp\{z\tau\}\,\tau^{a-1}(1-\tau)^{c-a-1}d\tau,$$

Re $c >$Re $a > 0$, where ${}_1F_1(a;c;z)$ is Kummer's confluent hypergeometric function and $\Gamma\,(c)$ is the gamma function. For $a \equiv v+1$, $c \equiv n+2$, and $z \equiv t$ this implies

$$\int_0^1 \exp\{tu\}\, u^v(1-u)^{n-v}du = {}_1F_1\,(v+1;n+2;t)\,B\,(v+1,n-v+1)\,.$$

Therefore,

$$M_u(t) = \sum_{v=0}^{n} \beta\left(\frac{v}{n}\right)\binom{n}{v} {}_1F_1\left(v+1; n+2; t\right) B\left(v+1, n-v+1\right)$$

To obtain the moment generating function for the k dimensional Bernstein approximation, we replace the univariate result in the multivariate definition,

$$\begin{aligned} M_u(t) &= \int_0^1 \cdots \int_0^1 \exp\left\{t\left(u_1 + ... + u_k\right)\right\} \sum_{v_1=0}^{m} ... \sum_{v_k=0}^{m} \beta\left(\frac{v_1}{m}, ..., \frac{v_k}{m}\right) \\ &\quad \times \prod_{j=1}^{k} \binom{m}{v_j} u^{v_j} (1-u_j)^{m-v_j} du_1 \cdots du_k \\ &= \sum_{v_1=0}^{m} ... \sum_{v_k=0}^{m} \beta\left(\frac{v_1}{m}, ..., \frac{v_k}{m}\right) \\ &\quad \times \prod_{j=1}^{k} {}_1F_1\left(v_j+1; m+2; t\right) B\left(v_j+1, m-v_j+1\right). \end{aligned}$$

These results can be used to further investigate the properties of the Bernstein copula and its approximations. Deriving results on the joint moments of the Bernstein copula is quite easy in virtue of its incomplete Beta function representation. The joint moments are important to study the scale free dependence properties of the variables.

2.4 Bernstein Representation of Arbitrary Copulae

While the Bernstein copula should be regarded as a copula in its own right, it is particularly suited to problems where a parametric copula is available but in a very complicated form. In this case, the Bernstein copula can be used in place of the original copula. By the approximating properties of Bernstein polynomials, the coefficients are simple to find. One may object that Bernstein polynomials have a slower rate of convergence as compared to other polynomial approximations (see Theorem 4, below).[2] However, they have the best rate of convergence within the class of all operators with the same shape preserving property; see Berens and DeVore (1980). Given the properties of the copula, we could not hope for anything better.

To give an example of the viability of the Bernstein approximation and its range of dependance we approximate the Kimeldorf and Sampson copula (see e.g. Joe (1997) p. 141), which is equal to

$$C(u,v) \equiv \left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1}{\theta}}. \tag{11}$$

Figure I shows the 3 dimensional graph of the Kimeldorf and Sampson copula density. We report the values of Spearman's rho as a function of the dependance parameter θ in the approximation for m of order 10, 30, 50 and the corresponding ones for the Kimeldorf and Sampson copula (KS). Figure II and III show the contourplot of the two copulae when $\theta = 1.06$ and $m = 30$. In Table I, values for Spearman's rho in KS are from Joe (1997), values for the Bernstein copula were calculated by the authors. Because of computational difficulties the limit of the dependence parameter to infinity was not calculated for the approximation.

All differences are due to polynomials being fairly slow in adjusting at turning points. Improvements can be achieved by increasing the order of the polynomial, keeping all computations manageable and straightforward. Integral evaluation for the computation of Sperman's rho for the Kimeldorf and Sampson copula could not be performed on a PC using Maple.

To lend some rigor to this numerical example, we state the following.

Theorem 4. *Let* $f \in C_{[0,1]^k}$ *and* $\frac{\partial f}{\partial x_j}$ *be Lipschitz* $\forall j$, *then*

$$\left|\left(B_m^k f\right)(x_1, ..., x_k) - f(x_1, ..., x_k)\right| \leq M \sum_{j=1}^{k} \frac{x_j (1 - x_j)}{2m},$$

where B_m^k *is the* k *dimensional Bernstein operator and* M *is a constant.*

Proof. See the Appendix. ■

Notice that a more precise statement on the constant M in Theorem 4 can be given, but for simplicity we just give a version that allows us to define the speed of convergence.

There are properties of some copulae derived by limiting operations, i.e. tail dependence.[4] For the case of tail dependence, Bernstein polynomials always have limit equal to zero, that is no tail dependence. Just to give a hint about the reason for this to happen, notice that convergence under the sup norm is not sufficient for assuring that the Bernstein copula and its approximand converge to an arbitrary limit at the same speed. Fortunately, it is the case that a Bernstein copula can capture increasing dependence as we move to the tails. Notice that the Kimeldorf and Sampson copula in our example above exhibits lower tail dependence.

3 An Example

In order to make the above discussion less abstract, we provide an example interesting in its own right and to clarify ideas. We consider two random variables (i.e. log returns) with highly nonlinear dependence and an agent with negative exponential utility function who wants to find the optimal portfolio weights.

Let z_1 and z_2 be log returns on each assets, where the weights constrained to add up to one simplify to the term w. Then, we face the following problem

$$\sup_{w} -E \exp\left\{-\gamma\left[wz_1 + (1-w)\, z_2\right]\right\}.$$

Define $W = wz_1 + (1-w)\, z_2$, then, the Bernoulli utility function $-\exp\{-\gamma W\}$ is characterized by a constant Arrow-Pratt coefficient of risk aversion equal to γ. As one referee mentioned, constant absolute risk aversion may be questionable in this context, but is nevertheless central to much applied work due to its properties. Notice that

$$-E \exp\{-\gamma W\} \tag{12}$$

is just equal to minus the Laplace transform of the wealth's (W) probability density function.

In particular, we simulate two series, z_1, z_2. we use the following data generating process (DGP):

(1) for the marginals we use

$$pdf\,(z_j) = a_j \exp\left\{-a_j \left|z_j - \mu_j\right|\right\},$$

which is a double exponential with mean μ_j;

(2) for the copula function we use the KS copula which was given in (11), i.e.

$$C\,(u, v) \equiv \left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1}{\theta}},$$

which is known to generate lower tail dependence. The parameters for the DGP are given in Table II.

Using this DGP we generate 5000 observations. The correlation for z_1 and z_2 is equal to 0.3468 and other sample characteristics are summarized in Table III. The data are defined over the real line and can be thought as log differences in prices. Although in portfolio problems it is appropriate to use arithmetic returns, we use geometric returns. Arithmetic returns have the irritating prospect of being bounded below. For pedagogical purposes we will compare our estimates with the misspecified assumption of normality, then we need both models to have the same range. The economic rationale for not bounding returns below is to jettison the free disposal assumption. Thus we could interpret a model of an investor who considers optimizing a portfolio of forward contracts, ownership of which confers a liability on the holder.

Our investigator will make the following assumptions:

(1) To model the marginals, she uses a Weibull density of the following form

$$\frac{a_j b_j \left(\left|z_j - \mu_j\right|\right)^{b_j - 1}}{2} \exp\left\{-a_j \left(\left|z_j - \mu_j\right|\right)^{b_j}\right\},\ j = 1, 2;$$

notice that the exponential can be embedded in the Weibull for $b_j = 1$. It is noticed that Weibull distributions are frequently used in a range of practical financial problems, especially as models for loss and risk.[5]

(2) To model the dependence between the marginals, she uses, incorrectly, a Placket copula, i.e.

$$\begin{aligned} C\left(u, v; \delta\right) \quad \equiv \quad & \frac{1}{2}\{\left(\delta - 1\right) 1 + \left(\delta - 1\right)\left(u + v\right) \\ & - \left[\left(1 + \left(\delta - 1\right)\left(u + v\right)\right)^2 - 4\delta\left(\delta - 1\right) uv\right]^{\frac{1}{2}}\}; \end{aligned}$$

see Joe (1997) for details.

The parameters a_j, b_j, μ_j and δ are estimated from the simulated data. In particular, μ_j was directly calculated as the mean from these data. As common in estimation of multivariate distributions when the copula is used, the parameters were estimated in a two step procedure. The likelihoods for the univariate marginals were separately maximized. Using the estimated parameters from the univariate likelihoods, the likelihood for the copula was optimized with respect to the dependence parameter; see Joe (1997) p. 299-301 for details. The parameters estimates are given in Table IV. For estimation of b_j, we use the constraint $b_j \geq 1$ for both series. We used $b_j \geq 1$ as constraint because, otherwise, the integral does not converge.[6]

The parametric specification gives rise to an expression that is too complex to evaluate as compared to a neat bivariate normal distribution. Therefore, we resort to the Bernstein copula as a viable alternative for our calculations. In this case, the coefficients do not need to be estimated because we are using the Bernstein copula as an approximation to the erroneous copula. The Bernstein copula approximation will allow us to find an approximation to optimal portfolio weights by simple integral maximization.

Using the result in Appendix A, the joint pdf of z_1 and z_2 using the Bernstein copula density is given by the following,

$$pdf\left(z_1, z_2\right) = c_B(F_1\left(z_1\right), F_2\left(z_2\right))$$

$$
\begin{aligned}
&= \sum_{v_1=0}^{m}\sum_{v_2=0}^{m} c\left(\frac{v_1}{m},\frac{v_2}{m};\theta\right) \\
&\quad \times \prod_{j=1,2}\sum_{s_j=0}^{v_j}\binom{m}{v_j}\frac{(-v_j)_{s_j}}{\Gamma(s_j+1)} \\
&\quad \times a_j b_j \left|x_j-\mu_j\right|^{b_j-1}\left(\frac{\exp\left\{-a_j\left|x_j-\mu_j\right|^{b_j}\right\}}{2}\right)^{m-v_j+1+s_j},
\end{aligned}
$$

$$\text{if } z_j \quad \geq \quad \mu_j,$$

$$
\begin{aligned}
&= \sum_{v_1=0}^{m}\sum_{v_2=0}^{m} c\left(\frac{v_1}{m},\frac{v_2}{m};\theta\right) \\
&\quad \times \prod_{j=1,2}\sum_{s_j=0}^{m-v_j}\binom{m}{v_j}\frac{(v_j-m)_{s_j}}{\Gamma(s_j+1)} \\
&\quad \times a_j b_j \left|x_j-\mu_j\right|^{b_j-1}\left(\frac{\exp\left\{-a_j\left|x_j-\mu_j\right|^{b_j}\right\}}{2}\right)^{v_j+1+s_j},
\end{aligned}
$$

$$\text{if } z_j \quad < \quad \mu_j,$$

where $c(u,v;\theta) = \frac{\partial^2 C(u,v;\theta)}{\partial u \partial v}$, and $(...)_s$ is Pochhammer's symbol and $\Gamma(...)$ is the gamma function. Therefore, the problem is reduced to the evaluation of the Laplace transform of a simple Weibull density function and this can be done by expanding the exponential in order to get a series in terms of gamma integrals. Details of the calculations can be found in Sancetta and Satchell (2001a). The function was maximized with respect to the weight w which is associated with the first asset. In order to check if our procedure outperforms simple alternatives, the same calculation was carried out assuming normality. Our true benchmark was derived optimizing the empirical Laplace transform $\frac{1}{n}\sum_{t=1}^{n}\exp\{-\gamma W_t\}$ using a simulated sample of 30,000 observations from the same joint distribution. It can be shown that $\exp\{-\gamma W_t\}$ is Glivenko-Cantelli, i.e.

$$\frac{1}{n}\sum_{t=1}^{n}\exp\{-\gamma W_t\} \to E\left(\exp\{-\gamma W_t\}\right), \text{ a.s.}$$

uniformly in γ and w; e.g. see Andrews (1987). The coefficients chosen for the expected utility function as defined in (12) together with the results are given in Table V.[7] The order of the Bernstein polynomial in the approximation was $m = 15$. The optimal weight using a nonparametric specification, was also computed: in this case a nonparametric Bernstein copula density which is defined in (14), i.e. a Bernstein copula with generator $\alpha\left(\frac{v_1}{m}, \frac{v_2}{m}\right) = C_n\left(\frac{v_1}{m}, \frac{v_2}{m}\right)$, where $C_n\left(\frac{v_1}{m}, \frac{v_2}{m}\right)$ is the empirical copula and was chosen with $m = 25$.

It is interesting to note that despite the fact that the parametric copula we estimate is misspecified, the results are fairly good. Also the nonparametric approach provides reasonable answers. We did not investigate the change of optimal weight with respect to different m. The normality assumption leads to results that can be poor. It is necessary to realize that the poor performance is mainly due to the nonlinear dependence. Asymmetric data should be expected to lead to results that are even worse under normality. Asymmetry using our framework could be accounted for, but this is not in the scope of the paper. Personal calculations by the authors show that results can be reasonably good even when the copula is misspecified as long as dependence is not too high.

4 An Estimation Procedure

The study of the Bernstein copula would erroneously lead us to think that it is only a tool for numerical analysis if we did not address the problem of estimation. The main result of this section is Theorem 5.[8] There are several possible estimation procedure that can be employed to make the Bernstein copula operational. For the sake of conciseness, we will only discuss one of the many possible ones and provide consistency results: nonparametric Bernstein copula density estimation where the coefficients of the copula are given by the empirical copula.

4.1 Nonparametric Bernstein Copula Density

Recall the functional form of the Bernstein copula:

$$\begin{aligned} C_B(u_1, ..., u_k) &= \sum_{v_1=0}^{m} ... \sum_{v_k=0}^{m} \alpha\left(\frac{v_1}{m}, ..., \frac{v_k}{m}\right) \\ &\quad \times \prod_{j=1}^{k} \binom{m}{v_j} u_j^{v_j} (1-u_j)^{m-v_j}. \end{aligned}$$

Let $C_n\left(\frac{v_1}{m}, ..., \frac{v_k}{m}\right)$ be the empirical copula at $\left(\frac{v_1}{m}, ..., \frac{v_k}{m}\right)$, i.e.

$$\frac{1}{n}\sum_{s=1}^{n} I\left\{\bigcap_{j=1}^{k}\left[u_{js} \leq t_{v_j}\right]\right\}, \tag{13}$$

where $I_{\{A\}}$ is the indicator of the set A. The nonparametric Bernstein copula, say $\tilde{C}_B(\mathbf{u})$, is defined as the usual Bernstein copula where

$$\alpha\left(\frac{v_1}{m}, ..., \frac{v_k}{m}\right) = C_n\left(\frac{v_1}{m}, ..., \frac{v_k}{m}\right).$$

Differentiating it is easy to see that the coefficients of the polynomial are equivalent to a k dimensional histogram estimator (see Scott, 1992, for details on the historgram estimator),

$$
\begin{aligned}
\tilde{c}_B &= \sum_{v_1=0}^{m-1} .... \sum_{v_k=0}^{m-1} \triangle_{1,....,1} \left(\frac{m^k}{n} \sum_{s=1}^{n} I \left\{ \bigcap_{j=1}^{k} \left[u_{js} \leq t_{v_j} \right] \right\} \right) \\
& \prod_{j=1,2} \binom{m-1}{v_j} u_j^{v_j} (1-u_j)^{m-1-v_j}, \qquad (14)
\end{aligned}
$$

where we use $\tilde{c}_B$ to stress that it is a particular estimator and $\triangle_{1,....,1}$ is the k dimensional difference operator, i.e.

$$
\triangle_{1,....,1} I \left\{ \bigcap_{j=1}^{k} \left[u_{js} \leq t_{v_j} \right] \right\} = \sum_{l_1=0}^{1} .... \sum_{l_k=0}^{1} (-1)^{l_1+....+l_k} \; I \left\{ \bigcap_{j=1}^{k} \left[u_{js} \leq t_{v_j} + \frac{l_j}{m} \right] \right\}.
$$

The optimal choice of m depends on the topology we use. We choose m to minimize the mean square error of the density, i.e. $\min \|\tilde{c}_B - c\|_2^2$ where $\|...\|_2$ is the L_2 norm under the true probability measure, and c is the true copula density. Just increasing m will reduce the bias but increase the variance of $\tilde{c}_B$.

4.2 Consistency in MSE of the Nonparametric Bernstein Copula

We want to choose (14) such that it is optimal under the L_2 norm. There is a lot to be said on this and on the properties of Bernstein polynomials in this case. We state the following condition.

Condition 1. $\{\mathbf{u}\}$ $(k \times 1)$ *is a sequence of independent strictly stationary uniform* $[0,1]$ *random vectors with copula* $C(\mathbf{u})$ *and copula density* $c(\mathbf{u})$ *which has finite first derivative everywhere in the k-cube.*

Remark. *The independence condition is not required, but we use it in order* *to shorten the proof as much as possible. From the proof of Theorem 5 it can be seen that the results are still valid under appropriate mixing conditions.*

Therefore, we content ourself stating the following.

Theorem 5. *Let $\tilde{c}_B$ be the k dimensional Bernstein copula density. Under Condition 1*

i. $Bias(\tilde{c}_B) = O\left(m^{-1}\right)$*;*

ii.

$(a.)$ *for* $u_j \in (0,1), \forall j$,

$$var\left(\tilde{c}_B\right) \simeq \frac{m^{\frac{k}{2}}\left(c\left(\mathbf{u}\right) + O\left(m^{-1}\right)\right)}{n\left(4\pi\right)^{\frac{k}{2}}\prod_{j=1}^{k}\lambda_j},$$

$(b.)$ *for* $u_j = 0, 1, \forall j$,

$$var\left(\tilde{c}_B\right) = \frac{m^k}{n}c\left(\mathbf{u}\right) + O\left(\frac{m^{k-1}}{n}\right);$$

iii.

$$\tilde{c}_B\left(\mathbf{u}\right) \to c\left(\mathbf{u}\right)$$

in mean square error:

$(a.)$ *for* $u_j \in (0,1), \forall j$, *if* $\frac{m^{\frac{k}{2}}}{n} \to 0$ *as* $m, n \to \infty$;

$(b.)$ *for* $u_j = 0, 1, \forall j$, *if* $\frac{m^k}{n} \to 0$ *as* $m, n \to \infty$;

iv. The optimal order of polynomial in a mean square error sense is:

$(a.)$ $m = O\left(n^{\frac{2}{k+4}}\right)$ *if* $u_j \in (0,1), \forall j$;

$(b.)$ $m = O\left(n^{\frac{1}{k+2}}\right)$ *if* $u_j \in (0,1), \forall j$;

iv. If $m-k \geq 2$, $\tilde{c}_B\left(\mathbf{u}\right)$ *and* $\tilde{C}_B\left(\mathbf{u}\right)$ *are Donsker, i.e.* $z_B\left(\mathbf{u}\right) \equiv \left(\tilde{c}_B\left(\mathbf{u}\right) - E\tilde{c}_B\left(\mathbf{u}\right)\right)$

and $\tilde{Z}_B(\mathbf{u}) \equiv \left[\tilde{C}_B(\mathbf{u}) - E\tilde{C}_B(\mathbf{u})\right]$ *converge to a zero mean Gaussian process with continuous sample paths and covariance function*

$$E\left[z_B(\mathbf{u}_1)\, z_B(\mathbf{u}_2)\right],$$

and

$$E\left[\tilde{Z}_B(\mathbf{u}_1)\, \tilde{Z}_B(\mathbf{u}_2)\right],$$

respectively.

The proof of Theorem 5 is given in the next subsection. The weak limit of the infinite dimensional distribution of the nonparametric Bernstein copula density and the empirical Bernstein copula are given because these can be used to devise test statistics for independence based on some norm of the limiting Gaussian process. The limiting distribution of the norm would not be known, but the bootstrap can be used in this case. Recall that if a class is Dosker, then it is Glivenko-Cantelli, i.e. convergence holds uniformly over the class. Therefore, the results of Theorem 5 hold uniformly.

For comparison purposes, let $h \equiv m^{-1}$ be the smoothing factor in the usual sense. The bias is of the same order as the one for the histogram estimator. In this respect, kernel smoothers would lead to a bias not higher than $O\left(m^{-2}\right)$. The reason for not calculating the constant is that in order to find the term that is $O\left(m^{-2}\right)$ it is required to take a Taylor series at least to third order. The result of taking a Taylor series up to second order does not seem to be a rewarding exercise. Details are available upon request. Notice that it is not possible to reduce the bias to $O\left(m^{-2}\right)$ by shifting the histogram. In this case the first term in the expansion would vanish, but other terms of same order would not.

While the bias is of the same order as the histogram estimator, the variance is of smaller order (except at the edges of the cube): $var\left(\tilde{c}_B\right) = O\left(m^{\frac{k}{2}}\right)$ instead of $O\left(m^k\right)$ as is the case for the histogram and kernel estimators. On the other hand, for $u_j = 0, 1$, for all j's, the variance is of the same order as for these other nonparametric estimators. The case $u_j = 0, 1$ for only some j is not included because the result is just a mixture of the two extreme cases: the variance goes down by a factor that is $O\left(m^{\frac{1}{2}}\right)$ for all the coordinates inside the k-hypercube while for the coordinates on the boundaries the contribution to the variance is $O\left(m\right)$. This just follows from the fact that Bernstein polynomials define a Tensor product space; see Cheney and Ward (2000) for a discussion in this context.

As m and n go to infinity, it follows that this estimator has rate of consistency $\frac{m^{\frac{k}{2}}}{n} \rightarrow 0$ inside the hypercube, versus $\frac{m^k}{n} \rightarrow 0$ for other nonparametric estimators. Inside the hypercube, the optimal order of smoothing is $m = O\left(n^{\frac{2}{k+4}}\right)$ in mean square error sense, versus $m = O\left(n^{\frac{1}{k+2}}\right)$ for the histogram and $m = O\left(n^{\frac{1}{k+4}}\right)$ for a first order kernel.

This implies that the Bernstein polynomials require very little smoothing (i.e. a large order of polynomial). This is due to the fact that Bernstein polynomials are fairly slow to adjust as already mentioned in Section 2.

4.3 Proof of the Theorem

The proof of part *ii.* in Theorem 5 is based on the normal approximation to the binomial distribution, e.g. see Stuart and Ord (1994, p. 138-140). In particular, let

$$P_{v,m}(u) \equiv \binom{m}{v} u^v (1-u)^{m-v},$$

and

$$\mathcal{P}_{v,m}(v) \equiv (2\pi u(1-u)m)^{-\frac{1}{2}} \exp\left\{-\frac{m}{2u(1-u)}\left(\frac{v}{m}-u\right)^2\right\}, \tag{15}$$

then

$$\sum_{v=0}^{m} f\left(\frac{v}{m}\right) P_{v,m}(u) \simeq \int_{-\infty}^{\infty} f\left(\frac{v}{m}\right) \mathcal{P}_{v,m}(v)\, dv.$$

A formal proof may be given through the Edgeworth expansion for $z = \left(\frac{v}{m} - u\right)$ in order to prove that the error is uniform. Taking squares of the two distributions (i.e. the binomial and the Gaussian),

$$\sum_{v=0}^{m} f\left(\frac{v}{m}\right) \left(P_{v,m}(u)\right)^2 \simeq \int_{-\infty}^{\infty} f\left(\frac{v}{m}\right) \left(\mathcal{P}_{v,m}(v)\right)^2 dv, \tag{16}$$

where again the error holds uniformly. With this reminder, we prove Theorem 5.

Notation. *Here, we use* $\mathbf{u}_s$ *to indicate the vector of rv's. On the other hand* $\mathbf{u}$ *will denote a fixed, but arbitrary value. This generates no confusion as long as one is willing to look at a function as a point in the space. Moreover,* $\partial_i c(\mathbf{u}) \equiv \frac{\partial c(\mathbf{u}_s)}{\partial u_i}|_{\mathbf{u}_s = \mathbf{u}}$.

Proof of Theorem 5. $\text{Bias}(\tilde{c}_B) \equiv E(\tilde{c}_B) - c(\mathbf{u})$. Let $t_{v_j} \equiv \frac{v_j}{m}$. Now

$$E(\tilde{c}_B) = \sum_{v_1=0}^{m-1} \cdots \sum_{v_k=0}^{m-1} \left(\frac{m^k}{n} \sum_{s=1}^{n} p_{s,v_1 \cdots v_k} \right) \prod_{j=1}^{k} \binom{m-1}{v_j} u_j^{v_j} (1-u_j)^{m-1-v_j}, \tag{17}$$

where

$$p_{s,v_1 \cdots v_k} \equiv \int_{t_{v_k}}^{t_{v_k}+\frac{1}{m}} \cdots \int_{t_{v_1}}^{t_{v_1}+\frac{1}{m}} c(\mathbf{u}_s)\, du_{1s} \cdots du_{ks}.$$

By the mean value theorem,

$$p_{s,v_1 \cdots v_k} = \int_{t_{v_k}}^{t_{v_k}+\frac{1}{m}} \cdots \int_{t_{v_1}}^{t_{v_1}+\frac{1}{m}} \left(c(\mathbf{u}) + \sum_{j=1}^{k} \partial_j c(\mathbf{u}^*) (u_{js} - u_j) \right) du_{1s} \cdots du_{ks},$$

where $\|\mathbf{u}^* - \mathbf{u}\| \leq 1$ and $\|...\|$ is the Euclidean distance. By Condition 1, $\max \partial_j c(\mathbf{u}) \leq q$ for some $q < \infty$, then

$$\begin{aligned}
& \int_{t_{v_k}}^{t_{v_k}+\frac{1}{m}} \cdots \int_{t_{v_1}}^{t_{v_1}+\frac{1}{m}} \partial_j c(\mathbf{u}^*) (u_{js} - u_j)\, du_{1s} \cdots du_{ks} \\
\leq\ & \max_{\mathbf{u}} \partial_j c(\mathbf{u}) \int_{t_{v_k}}^{t_{v_k}+\frac{1}{m}} \cdots \int_{t_{v_1}}^{t_{v_1}+\frac{1}{m}} (u_{js} - u_j)\, du_{1s} \cdots du_{ks} \\
=\ & q \left(\frac{t_j - u_j}{m^k} + O\left(\frac{1}{m^{k+1}} \right) \right).
\end{aligned}$$

By independence, this implies

$$\frac{1}{n} \sum_{s=1}^{n} p_{s,v_1 \cdots v_k} = \frac{c(\mathbf{u})}{m^k} + q \sum_{j=1}^{k} \frac{t_j - u_j}{m^k} + O\left(m^{-(1+k)} \right). \tag{18}$$

Recall that $c(\mathbf{u})$ is some arbitrary, but constant value. Bernstein polynomials preserve the constant, i.e. $B_{m-1}^k c(\mathbf{u}) = c(\mathbf{u})$, on the other hand, by direct calculation

$$\sum_{v_j=0}^{m-1}(t_j-u_j)\binom{m-1}{v_j}u_j^{v_j}(1-u_j)^{m-1-v_j}=\frac{u_j}{m},$$

that is $B_{m-1}^{k}(t_j-u_j)=\frac{u_j}{m}$; recall that $t_j=\frac{v_j}{m}$. It follows that substituting in (17),

$$E(\tilde{c}_B)=c(\mathbf{u})+O\left(m^{-1}\right),$$

remembering to multiply each term in (18) by m^k. Therefore,

$$Bias(\tilde{c}_B)=O\left(\frac{1}{m}\right).$$

For the variance, notice that the probability of one observation falling inside a subset of the hypercube is equal to the probability of success in a Bernoulli trial. We know that the probability of n successes, where n is the sample size, is given by a binomial distribution. By the variance of n independent Bernoulli trials

$$\begin{aligned} var(\tilde{c}_B) &= \sum_{v_1=0}^{m-1}\cdots\sum_{v_k=0}^{m-1}\left(\frac{m^{2k}}{n^2}\sum_{s=1}^{n}p_{s,v_1\cdots v_k}\left(1-p_{s,v_1\cdots v_k}\right)\right)\\ &\quad\times\prod_{j=1}^{k}\left(P_{v_j,m-1}(u)\right)^2\\ &\approx \sum_{v_1=0}^{m-1}\cdots\sum_{v_k=0}^{m-1}\left(\frac{m^{2k}}{n^2}\sum_{s=1}^{n}p_{s,v_1\cdots v_k}\right)\prod_{j=1}^{k}\left(P_{v_j,m-1}(u)\right)^2\\ &= \frac{m^{2k}}{n}\sum_{v_1=0}^{m-1}\cdots\sum_{v_k=0}^{m-1}\left(\frac{c(\mathbf{u})}{m^k}+q\sum_{j=1}^{k}\frac{t_j-u_j}{m^k}+O\left(m^{-(1+k)}\right)\right)\prod_{j=1}^{k}\left(P_{v_j,m-1}(u)\right)^2, \end{aligned}$$

where $\approx$ means asymptotic equality, i.e.

$$var(\tilde{c}_B)=\frac{m^{2k}}{n}\sum_{v_1=0}^{m-1}\cdots\sum_{v_k=0}^{m-1}\left(\frac{c(\mathbf{u})}{m^k}+q\sum_{j=1}^{k}\frac{t_j-u_j}{m^k}+O\left(m^{-(1+k)}\right)\right)\prod_{j=1}^{k}\left(P_{v_j,m-1}(u)\right)^2, \tag{19}$$

where the $O\left(m^{-(1+k)}\right)$ term is clearly different from the previous display. Notice that we have $mt_j = v_j$, but $v_j = 0, ..., m-1$, i.e. $t_j - \mu_j = \frac{m-1}{m}\left(\frac{v_j}{m-1} - \frac{m}{m-1}\mu_j\right)$. Use (16) to approximate (19). Consequently, solve the following type of integral,

$$\begin{aligned}\Gamma_j &= \int_{\mathbb{R}} \left(t_{v_j} - u_j\right) \frac{\exp\left\{-\frac{(m-1)}{u_j(1-u_j)}\left(\frac{v_j}{m-1} - u_j\right)^2\right\}}{\left[2\pi\left(m-1\right)u_j\left(1-u_j\right)\right]} dv_j \\ &= \int_{\mathbb{R}} \frac{m-1}{m}\left(\frac{v_j}{m-1} - \frac{m}{m-1}\mu_j\right) \\ &\quad \times \frac{\exp\left\{-\frac{(m-1)}{u_j(1-u_j)}\left(\frac{v_j}{m-1} - u_j\right)^2\right\}}{\left[2\pi\left(m-1\right)u_j\left(1-u_j\right)\right]} dv_j.\end{aligned}$$

Simply make the following change of variable, $x_j = \sqrt{\frac{(m-1)}{u_j(1-u_j)}}\left(\frac{v_j}{m-1} - u_j\right)$, with Jacobian $\sqrt{\left(m-1\right)u_j\left(1-u_j\right)}$. Then

$$\begin{aligned}\Gamma_j &= \int_{\mathbb{R}} \frac{m-1}{m}\left(x_j + \mu_j\right) \\ &\quad \times \frac{\exp\left\{-x_j^2\right\}}{2\pi\sqrt{\left(m-1\right)u_j\left(1-u_j\right)}} dv_j \\ &= \frac{\sqrt{m-1}}{2m\sqrt{\pi u_j\left(1-u_j\right)}} u_j.\end{aligned}$$

This shows that the integration results in a drop in asymptotic magnitude equal to $m^{-\frac{1}{2}}$ for each dimension. Let $\lambda_j \equiv \left[u_j\left(1-u_j\right)\right]^{\frac{1}{2}}$, then

$$\begin{aligned}var\left(\tilde{c}_B\right) &\simeq \frac{m^{2k}}{n}\left(\left[4\pi\left(m-1\right)\right]^{\frac{k}{2}} \prod_{j=1}^{k} \lambda_j\right)^{-1} \left(\frac{c\left(\mathbf{u}\right)}{m^k} + O\left(m^{-(k+1)}\right)\right) \\ &\simeq \frac{m^{\frac{k}{2}}\left(c\left(\mathbf{u}\right) + O\left(m^{-1}\right)\right)}{n\left(4\pi\right)^{\frac{k}{2}} \prod_{j=1}^{k} \lambda_j}.\end{aligned}$$

Due to integrability of the reminder, one could use Hölder inequality to show that the orders of magnitude in the above display are correct. At at the opposite edges of the hypercube, i.e. $u = 0, 1$, $\left(P_{v_j,m-1}(u)\right)^2 = P_{v_j,m-1}(u)$, then

$$\begin{aligned} var\left(\tilde{c}_B\right) &= \frac{m^{2k}}{n}\left(\frac{c\left(\mathbf{u}\right)}{m^k} + O\left(m^{-(1+k)}\right)\right) \\ &= \frac{m^k}{n}c\left(\mathbf{u}\right) + O\left(\frac{m^{k-1}}{n}\right). \end{aligned}$$

The mean square error (MSE) convergence just follows by considering the leading terms for the square bias and the variance for the two distinct cases: $MSE = Bias\left(\tilde{c}_B\right)^2 + Var\left(\tilde{c}_B\right).$ The optimal order of the polynomial follows by minimization of asymptotic MSE with respect to m.

The finite dimensional distributions of the nonparametric Bernstein copula density converge to a normal distribution. This follows from the fact that it is the sum of bounded random variables and Condition 1 (weaker conditions than iid are clearly sufficient for the central limit theorem). But the Bernstein copula density has $m-1$ bounded derivatives (recall that Bernstein polynomials are closed under differentiation) and any Bernstein polynomial is Lipschitz. By Theorem 2.7.1 in van der Vaart and Wellner (2000, p. 155) the class of functions that satisfy the just mentioned properties has finite ε bracketing numbers of order $\exp\left\{\varepsilon^{-\frac{k}{m-1}}\right\}$. It follows that their entropy integral with bracketing is finite. This is enough to show (see Ossiander ,1987, for the iid case or Pollard, 2001, for generalizations) that the Bernstein copula density converges to a Gaussian process with continuous sample paths. The same condition applies to the copula because it is m times differentiable together with the same properties of the density. ■

From the proof it is clear that what drives the variance down is the fact that approximating the square of $P_{v,m-1}(u)$ leads to a normal approximation times an extra term that is $O\left(m^{-\frac{1}{2}}\right)$. In order to provide more intuition on this result and the difference between the edges of the box and the points inside it, we provide the following heuristic explanation. Bernstein polynomials average the information about the function throughout its support. Therefore, while slow at adjusting, they have a behavior that is not just local as is the case for other nonparametric estimators. On the other hand, the result at the corners of the hypercube is clear: the approximation at these points is exact and it is not influenced by the behavior of the function in its domain, i.e. it is just local.

5 Conclusion and Some Further Extensions

We studied a new object in multivariate analysis called the Bernstein copula. Furthermore, we showed that, subject to regularity, any copula can be represented (approximated) by some Bernstein copula. This copula representation should allow us to take advantage of the properties of the copula function whenever multivariate normality is not a good assumption. Our optimal portfolio example had the pedagogical purpose of showing how the Bernstein copula could be used as an approximation to a copula and why one may want to discard the simple normality alternative. Further, we made the procedure operational by providing a nonparametric estimation procedure with its rates of consistency and an interesting result for the variance.

This study of the Bernstein copula led us to consider many topics all at once. It is clear that a lot has been left out from this paper. We did not discuss joint continuity of the Bernstein copula under the *-product defined by Darsow et al. (1992); see Kulpa (1997) and Li et al. (1998). The *-product is a powerful tool that allows us to define Markov processes and general time series dependence concepts. The Bernstein copula is closed under this operation. The link of the perturbation term and Spearman's rho seems to suggest some nonparametric estimation procedure which as now, we have yet to explore. Our example for portfolio optimization could be generalized. Moreover, one could take advantage of the simple closed form of the moment generating function of the Bernstein copula and devise new measures of risk in portfolio optimization so as to replace mean variance optimization. The kernel smother representation has only been used indirectly in the proof of Theorem 5. Finally, alternative estimation procedures have not been considered in detail. While the paper provided a promising result for the variance of the nonparametric estimator, this is one among many others that could be studied. For example, one could look at the following estimation problems

$$\max \mathbf{P}_n \left[\ln c_B - \lambda_n \int (\mathcal{D}^\alpha c_B)^2 \right],$$

or

$$\min \mathbf{P}_n \left[(C_n - C_B)^2 + \lambda_n \int (\mathcal{D}^\alpha c_B)^2 \right],$$

where $\mathbf{P}_n$ is the empirical measure, C_n is the empirical copula, λ_n is a smoothing parameter going to zero as $n \to \infty$, the unqualified integral is a Lebesgue integral, $\mathcal{D}^\alpha$ is the differential operator of order α, i.e. $\mathcal{D}^1 c_B = \sum_{j=1}^{k} \frac{\partial c_B}{\partial u_j}$, and

$c_B = \frac{\partial^k C_B}{\partial u_1 \cdots \partial u_k}$. However, unlike the case of the nonparametric Bernstein copula density, these estimators will not automatically lead to a copula unless constraints are imposed. Nevertheless, under suitable constraints, it seems plausible that the study of these estimators may lead to analogous results in virtue of the kernel representation of the Bernstein copula. Some of the issues left out of this paper are the subject of current research.

Notes

1. Our definition of k dimensional Bernstein polynomials is a generalization of 2 dimensional Bernstein polynomials, to our knowledge, first given in Butzer (1953).

2. By simple transformation: $x \in [a, b] \to t \in [0, 1]$, $t \equiv \frac{x-a}{b-a}$. In general, we can define a transformation that makes the real line isomorphic to the unit interval: $x \in \mathbb{R} \to t \in [0, 1]$, $t \equiv \frac{x}{1-x} - \frac{1-x}{x}$.

3. The simple Bernstein approximation can be improved by taking linear combinations; see Butzer (1952b). Let $f^{(2l)} \in Lip\gamma$ be the $2l$ derivative of f, then, Buzter (1952b) shows that his liner combination of one dimensional Bernstein polynomials (equation (10)) has error $O\left(n^{-l-\gamma}\right)$ compared to $O\left(n^{-2l-\gamma}\right)$ for the best polynomials of order n.

4. Lower and upper tail dependence are respectively defined as

$$\lambda_L = \lim_{u \to 0} \Pr\left(u_1 < u | u_2 < u\right),$$

and

$$\lambda_U = \lim_{u \to 1} \Pr\left(u_1 > u | u_2 > u\right),$$

where λ_L and λ_U are between zero and one. No tail dependence corresponds to theses probabilities being exactly zero.

5. Laherrère and Sornette (1998) show evidence that a Weibull distribution provides adequate fit for daily financial returns. Calculations by the authors confirm this. Moreover, extensions to model asymmetric returns are possible and can be used in the context of the Bernstein copula.

6. The integral does not converge for $b < 1$ and it is only conditionally convergent for $b = 1$.

7. The use of negative wealth implies that the coefficient of risk aversion cannot be interpreted in the same way as in the case of negative wealth. In this scenario, $\gamma > .75$ implies a very high degree of risk aversion due to the fact that the Bernoulli utility decreases exponentially fast for negative wealth. For this reason, the results become particularly data dependent on the left tail.

8. Theorem 5 is an adapted version of the results in Sancetta and Satchell (2001b) where nonparametric estimation is discussed at length.

References

[1] Abadir, K.M. (1999) An Introduction to Hypergeometric Functions for Economists. *Econometric Reviews* 18, 287-330.

[2] Andrews, D.W.K. (1987) Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers. *Econometrica* 55, 1465-1471.

[3] Berens, H. & R. DeVore (1980) A Characterization of Bernstein Polynomials. In E.W.Cheney (ed.), *Approximation Theory III*, 213-219. New York: Academic Press.

[4] Buoyé, E., N. Gaussel and M. Salmon (2001) Investigating Dynamic Dependence Using Copulae. *FERC Working Paper.*

[5] Butzer, P.L. (1953) On Two Dimensional Bernstein Polynomials. *Canadian Journal of Mathematics* 5, 107-113.

[6] Darsow, W.F., B. Nguyen and E.T. Olsen (1992) Copulas and Markov Processes. *Illinois Journal of Mathematics* 36, 600-642.

[7] Embrechts, P., A. McNeil & D. Straumann (1999) Correlation: Pitfalls and Alternatives. *Risk* 5, 69-71.

[8] Joe, H. (1997) *Multivariate Models and Dependence Concepts.* London: Chapman & Hall Ltd.

[9] Kulpa, T. (1999) On Approximation of Copulas. *International Journal of Mathematics and Mathematical Science* 22, 259-269.

[10] Laherrère, J. & D. Sornette (1998) Stretched Exponential Distributions in Nature and Economy: "Fat Tails" with Characteristic Scales. *The European Physical Journal* B 2, 525-539.

[11] Li, X., P. Mikusiński and M.D. Taylor (1998) Strong Approximation of Copulas. *Journal of Mathematical Analysis and Applications* 225, 608-623.

[12] Rockinger, M. and E. Jondeau (2001) Conditional Dependency of Financial Series: An Application of Copulas. *Preprint.*

[13] Lorentz, G.G. (1953) *Bernstein Polynomials.* Toronto: University of Toronto Press.

[14] Marichev, O.I. (1983) *Handbook of Integral Transforms of Higher Transcendental Functions.* Chichester, England: Ellis Horwood Ltd.

[15] Nelsen, R.B. (1998) *An Introduction to Copulas.* Lecture Notes in Statistics 139. New York: Springer Verlag.

[16] Ossiander, M. (1987) A Central Limit Theorem under Metric Entropy with L_2 Bracketing. *The Annals of Probability* 15, 897-919.

[17] Patton, A.J. (2001) Modelling Time Varying Exchange Rate Dependence Using the Conditional Copula. *Discussion Paper Department of Economics, UCSD.*

[18] Phillips, P.C.B. (1982) Best Uniform and Modified Padé Approximations of Probability Densities in Econometrics. In W. Hildenbrand (ed.), *Advances in Econometrics*, 123-167. Cambridge: Cambridge University Press.

[19] Phillips, P.C.B. (1983) ERA's: A New Approach to Small Sample Theory. *Econometrica* 51, 1505-1525.

[20] Pollard, D. (2001) General Bracketing Inequalitites. *Department of Statistics, Yale University.*

[21] Sancetta, A. and S. Satchell (2001a) Bernstein Approximations to Copula Function and Portfolio Optimization. *DAE Discussion Paper, University of Cambridge.*

[22] Sancetta, A. and S. Satchell (2001b) The Nonparametric Bernstein Copula Density: L_2 Theory. *Preprint.*

[23] Schweizer, B and E.F. Wolff (1981) On Nonparametric Measures of Dependence for Random Variables. *The Annals of Statistics* 9, 879-885.

[24] Scott, D.W. (1992) *Multivariate Density Estimation: Theory, Practice and Visualization.* New York: John Wiley & Sons Inc.

[25] Sklar, A. (1973) Random Variables, Joint Distribution Functions, and Copulas. *Kybernetika* 9, 449-460.

[26] Stuart A. and .Ord (1994) *Kendall's Advanced Theory of Statistics.* London: Edward Arnold.

[27] Van der Vaart, A. and J.A. Wellner (2000) *Weak Convergence of Empirical Processes.* Springer Series in Statistics. New York: Springer.

A Derivation of the Joint Density of the Portfolio's Assets

For the sake of generality, we allow for k assets. Let

$$u_j \equiv 1 - \frac{\exp\left\{-a_j \left|x_j - \mu_j\right|^{b_j}\right\}}{2}, \text{ if } x_j \geq \mu_j$$

$$u_j \equiv \frac{\exp\left\{-a_j\left|x_j-\mu_j\right|^{b_j}\right\}}{2}, \text{ if } x_j < \mu_j,$$

and $\prod_{j=1}^{k}\left(a_j b_j\left|x_j-\mu_j\right|^{b_j-1}\right)\exp\left\{-a_j\left|x_j-\mu_j\right|^{b_j}\right\}$ be the Jacobian of $u_j \to x_j$. Substituting in (9) and using the relation $(1-u)^v = \sum_{s=0}^{v}(-1)^s\binom{v}{s}u^s$, we can write the joint density as

$$\begin{aligned} c(x_1,...,x_k) &= \sum_{v_1=0}^{m}...\sum_{v_k=0}^{m}\beta\left(\frac{v_1}{m},...,\frac{v_k}{m}\right) \\ &\times\prod_{j=1}^{k}\sum_{s_j=0}^{v_j}\binom{m}{v_j}\frac{(-v_j)_{s_j}}{\Gamma(s_j+1)} \\ &\times\left(a_j b_j\left|x_j-\mu_j\right|^{b_j-1}\right)\left(\frac{\exp\left\{-a_j\left|x_j-\mu_j\right|^{b_j}\right\}}{2}\right)^{m-v_j+1+s_j}, \text{ if } x_j \geq \mu_j \end{aligned}$$

$$\begin{aligned} c(x_1,...,x_k) &= \sum_{v_1=0}^{m}...\sum_{v_k=0}^{m}\beta\left(\frac{v_1}{m},...,\frac{v_k}{m}\right) \\ &\times\prod_{j=1}^{k}\sum_{s_j=0}^{m-v_j}\binom{m}{v_j}\frac{(v_j-m)_{s_j}}{\Gamma(s_j+1)} \\ &\times\left(a_j b_j\left|x_j-\mu_j\right|^{b_j-1}\right)\left(\frac{\exp\left\{-a_j\left|x_j-\mu_j\right|^{b_j}\right\}}{2}\right)^{v_j+1+s_j}, \text{ if } x_j < \mu_j, \end{aligned}$$

where $(-1)^s\binom{v}{s} = \frac{(-v)_s}{\Gamma(s+1)}$.

B Proofs

Proof of Theorem 2. Consider the following k dimensional Bernstein linear operator B_m^k such that

$$(B_m^k f)(X) \equiv \sum_{v_1=0}^{m_1}...\sum_{v_K=0}^{m_k} f\left(\frac{v_1}{m_1},...,\frac{v_k}{m_k}\right)P_{v_1,m_1}(x_1)\cdots P_{v_K,m_K}(x_k),$$

where $f \in C_{[0,1]^k}$ $(X \in [0,1]^k)$ and

$$P_{v_j,m_j}(x_j) \equiv \binom{m_j}{v_j} x_j^{v_j}(1-x_j)^{m_j-v_j}.$$

By the theorem on linear monotone operators (see Sancetta and Satchell, 2001, for its statement in k dimensions) it is sufficient to show uniform convergence to f for the following cases: $f(X) = 1, x_j, x_j^2$, $1 \le j \le k$. Now,

$$(B_m^k 1)(X) \equiv \sum_{v_1}^{m_1} ... \sum_{v_K}^{m_K} P_{v_1,m_1}(x_1) \cdots P_{v_K,m_K}(x_k) = 1,$$

by the binomial theorem.

$$\begin{aligned}
(B_m^k x_j)(X) &\equiv \sum_{v_1}^{m_1} ... \sum_{v_j}^{m_j} ... \sum_{v_K}^{m_K} \left(\frac{v_j}{m_j}\right) P_{v_1,m_1}(x_1) \cdots P_{v_j,m_j}(x_j) \cdots P_{v_K,m_K}(x_k) \\
&= \sum_{v_j}^{m_j} \left(\frac{v_j}{m_j}\right) \binom{m_j}{v_j} x_j^{v_j}(1-x_j)^{m_j-v_j} \\
&= x_j \sum_{v_j=1}^{m_j} \frac{m_j - 1!}{(m_j - v_j)!(v_j - 1)!} x_j^{v_j-1}(1-x_j)^{m_j-v_j} \\
&= x_j \sum_{v_j=0}^{m_j-1} \binom{m_j - 1}{v_j} x_j^{v_j-1}(1-x_j)^{m_j-v_j-1} \\
&= x_j[x_j + (1-x_j)]^{m-1} = x_j,
\end{aligned}$$

where the first equality follows by the binomial theorem.

$$
\begin{aligned}
(B_m^k x_j^2)(X) &\equiv \sum_{v_1}^{m_1} \dots \sum_{v_j}^{m_j} \dots \sum_{v_K}^{m_K} \left(\frac{v_j}{m_j}\right)^2 P_{v_1,m_1}(x_1)\cdots P_{v_j,m_j}(x_j)\cdots P_{v_K,m_K}(x_k) \\
&= \sum_{v_j}^{m_j} \left(\frac{v_j}{m_j}\right)^2 \binom{m_j}{v_j} x_j^{v_j}(1-x_j)^{m_j-v_j} \\
&= \sum_{v_j=1}^{m_j} \left(\frac{v_j}{m_j}\right) \binom{m_j-1}{v_j-1} x_j^{v_j}(1-x_j)^{m_j-v_j} \\
&= \frac{m_j-1}{m_j} \sum_{v_j=1}^{m_j} \left(\frac{v_j-1}{m_j-1}\right) \binom{m_j-1}{v_j-1} x_j^{v_j}(1-x_j)^{m_j-v_j} \\
&\quad + \frac{1}{m_j} \sum_{v_j=1}^{m_j} \binom{m_j-1}{v_j-1} x_j^{v_j}(1-x_j)^{m_j-v_j} \\
&= \frac{m_j-1}{m_j} x_j^2 + \frac{1}{m_j} x_j \to x_j^2.
\end{aligned}
$$

■

Proof of Theorem 3.

$$
\begin{aligned}
(B_m^k f)(X) - f(X) &= \sum_{v_1}^{m_1} \dots \sum_{v_k}^{m_k} P_{v_1,m_1}(x_1)\cdots P_{v_k,m_k}(x_k) \\
&\quad \times \left[f\left(\frac{v_1}{m_1}, \dots, \frac{v_k}{m_k}\right) - f(x_1, \dots, x_k) \right] \\
&= \sum_{v_1}^{m_1} \dots \sum_{v_k}^{m_k} P_{v_1,m_1}(x_1)\cdots P_{v_K,m_K}(x_k) \int_{(x_1,\dots,x_k)}^{\left(\frac{v_1}{m_1},\dots,\frac{v_k}{m_k}\right)} \nabla f dr
\end{aligned}
$$

$\nabla f \equiv [f'^{1}(s_1, ..., s_k), ..., f'^{k}(s_1, ..., s_k)]$, where $f'^{j}(s_1, ..., s_k) \equiv \frac{\partial f(s_1,\dots,s_k)}{\partial s_j}$, and r is a vector valued function that defines the path between the end points of the integral. By definition, ∇f is a conservative vector field, so the path of integration is irrelevant. The above line integral can be split into k integrals along any paths parallel to the axis and perpendicular to each other. For example, we can write

$$\begin{aligned}\int_{(x_1,\dots,x_k)}^{\left(\frac{v_1}{m_1},\dots,\frac{v_k}{m_k}\right)} \nabla f dr &= \int_{x_1}^{\frac{v_1}{m_1}} f'^{_1}\left(s_1, x_2, x_3, \dots, x_k\right) ds_1 + \dots \\ &+ \int_{x_j}^{\frac{v_j}{m_j}} f'^{_j}\left(\frac{v_1}{m_1}, \frac{v_2}{m_2}, \dots, s_j, \dots, x_k\right) ds_j + \dots \\ &+ \int_{x_k}^{\frac{v_k}{m_k}} f'^{_k}\left(\frac{v_1}{m_1}, \frac{v_2}{m_2}, \dots, \frac{v_{k-1}}{n_{k-1}}, s_k\right) ds_k\end{aligned}$$

Now,

$$\begin{aligned}\int_{x_j}^{\frac{v_j}{m_j}} f'^{_j}\left(\frac{v_1}{m_1}, \frac{v_2}{m_2}, \dots, s_j, \dots, x_k\right) ds_j &= f'^{_j}\left(\frac{v_1}{m_1}, \dots, x_j, x_{j+1}, \dots, x_k\right)\left(\frac{v_j}{m_j} - x_j\right) \\ &- \int_{x_j}^{\frac{v_j}{m_j}} \left(s_j - \frac{v_j}{m_j}\right) df'^{_j}\left(\frac{v_1}{m_1}, \frac{v_2}{m_2}, \dots, s_j, \dots, x_k\right).\end{aligned} \tag{20}$$

From here the crude result of the Theorem can be obtained assuming that $f'^{_j} \in Lip_{M_j}1$, i.e. $f'^{_j}$ satisfies the Lipschitz condition with constant M_j and exponent 1:

$$\left|f'^{_1}\left(s_1, \dots, s_j + h_j, \dots, s_k\right) - f'^{_1}\left(s_1, \dots, s_j, \dots, s_k\right)\right| \leq M_j \left|h_j\right|.$$

It follows that the last integral in (20) does not exceed $M_j \int_{x_j}^{\frac{v_j}{n_j}} \left(s_j - x_j\right) ds_j =$

$\frac{1}{2}M_j\left(\frac{v_j}{m_j}-x_j\right)^2$. Therefore,

$$\begin{aligned}\left|(B_m^k f)(X)-f(X)\right| &\leq \sum_{v_1}^{m_1}...\sum_{v_K}^{m_K} P_{v_1,m_1}(x_1)\cdots P_{v_K,m_K}(x_k) \\ &\quad \times\frac{1}{2}\sum_{j=1}^{k} M_j\left(\frac{v_j}{m_j}-x_j\right)^2 \\ &= \frac{1}{2}\left[M_1\frac{x_1(1-x_1)}{m_1}+...+M_k\frac{x_k(1-x_k)}{m_k}\right]\end{aligned}$$

for any X, where the first term in the right hand side of (20) is exactly zero when the Bernstein operator is applied to $\left(\frac{v_j}{n_j}-x_j\right)$; see Proof of Theorem 2.

■

C Tables and Graphs

Table I. Spearman's rho for different values of the dependance parameter θ

θ	0	.14	.31	.51	.76	1.06	1.51	2.14	3.19	5.56	∞
$\rho_S(KS)$	0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1
$\rho_S(B_{10})$	0	.08	.16	.24	.32	.4	.48	.57	.65	.73	*
$\rho_S(B_{30})$	0	.09	.19	.28	.37	.46	.56	.65	.75	.84	*
$\rho_S(B_{50})$	0	.09	.19	.29	.38	.48	.58	.67	.77	.86	*

Table II Parameters for the DGP

a_1	μ_1	a_2	μ_2	θ
1	0.5	2	0.1	0.6

Table III. Descriptive Statistics (n=5000)

	Mean	Variance	Kurtosis
Z_1	0.4968	1.9442	5.6247
Z_2	0.0970	0.4899	6.0084

Table IV. Parameter Estimates

a_1	b_1	a_2	b_2	δ
1.000750	1.015783	2.026432	1.000000	2.789959

Table V. Optimal portfolio weight for a negative exponential utility function

	Empirical	Placket Copula	Nonparametric	Multivariate
γ	Laplace $(n = 30,000)$	B. Approximation	Bernstein Colpula	Normal
.25	0.9641	0.9303	0.9835	0.9963
.5	0.4172	0.4459	0.4502	0.5412
.75	0.2809	0.2999	0.2820	0.3899
1	0.1676	0.2447	0.2136	0.3137

Figure I. Kimeldorf and Sampson (KS) copula density (θ=1.06)

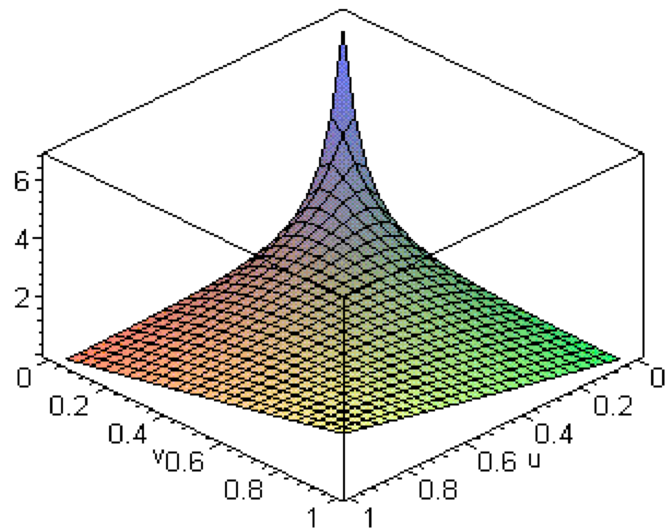

Figure II. KS copula (θ=1.06), contour plot

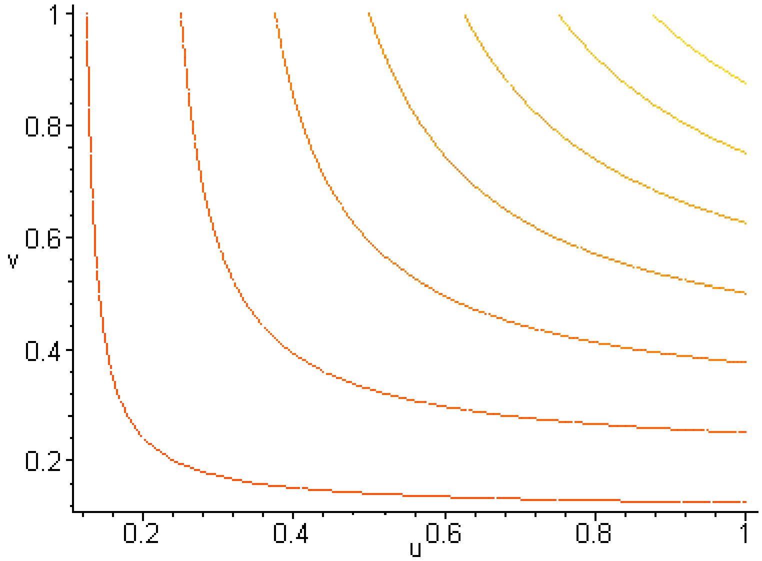

Figure III. Bernstein approximation to the KS copula (θ=1.06, n=30), contour plot