The Bernstein Copula and Its Applications to Modeling and Approximations of Multivariate Distributions
Author(s): Alessio Sancetta and Stephen Satchell
Source: *Econometric Theory*, Vol. 20, No. 3 (Jun., 2004), pp. 535-562
Published by: Cambridge University Press
Stable URL: http://www.jstor.org/stable/3533531
Accessed: 19-01-2018 15:42 UTC

Econometric Theory, **20**, 2004, 535–562. Printed in the United States of America.
DOI: 10.1017/S026646660420305X

THE BERNSTEIN COPULA AND ITS APPLICATIONS TO MODELING AND APPROXIMATIONS OF MULTIVARIATE DISTRIBUTIONS

ALESSIO SANCETTA
Trinity College, University of Cambridge

STEPHEN SATCHELL
Trinity College, University of Cambridge

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 copulas. Rates of consistency when the Bernstein copula density is estimated empirically are given. In order of magnitude, this estimator has variance equal to the square root of the variance of common nonparametric estimators, e.g., kernel smoothers, but it is biased as a histogram estimator.

1. INTRODUCTION

The task of modeling multivariate distributions has always been challenging. For this reason, it is very common to use elliptic distributions because of 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 is returns distributions in financial economics and the complex range of dependence that they exhibit. Among many references, the reader should look at Embrechts, McNeil, and Straumann (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. In practice we may have a clear idea of what the marginals are, although knowledge of their joint dependence could be limited. Therefore, being able to use a two-step procedure may be advantageous. Further, computer estimation can be speeded up. Though this may result in a loss of efficiency, corrections for the estimators have been proposed (e.g., White, 1994).

We would thank Mark Salmon for interesting us in the copula function and Peter Phillips, an associate editor, and the referees for many valuable comments. All remaining errors are our sole responsibility. Address correspondence to: Stephen Satchell, Trinity College, Cambridge CB2 1TQ; e-mail: ses11@hermes.cam.ac.uk.

 0266-4666/04 $12.00

For example, in the first step we may model unconditional marginal distributions or may model conditional distributions as in the case of generalized autoregressive conditional heteroskedasticity (GARCH) models. Then, cross dependence of the marginal distributions could be modeled in the second step. The same approach is valid in the case of stationary time series when we want to model their joint distribution given their unconditional marginals (when the series is Markovian, see Darsow, Nguyen, and Olsen, 1992).

Although the copula function is a fairly new concept in econometrics, there has been growing interest in financial econometrics (Hu, 2002; Bouyé, Gaussel, and Salmon, 2001; Patton, 2001; Rockinger and Jondeau, 2001; Longin and Solnik, 2001). 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 for studying extreme values (for a list of extreme value copulas, see Joe, 1997).

We introduce a family of copulas defined in terms of Bernstein polynomials. We call a copula from this family a Bernstein copula (BC). After defining a new class of copulas, we show how the BC can be used as an approximation to known and unknown copulas. When the approximation is used as an approximation for known copulas, we call it the approximate Bernstein copula (ABC) representation. This new representation leads to a general approach in estimation and also to simplification of many operations whenever a parametric copula is given. Moreover, for parametric copulas, the Bernstein representation is useful in studying properties of the copula function itself. When the same approach is used to approximate unknown copulas, we call this the empirical Bernstein copula (EBC).

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. It is important that Bernstein polynomials are closed under differentiation; for simple restrictions on the coefficients they always lead to a proper copula function, and when used as empirical estimators, they have lower variance than commonly used nonparametric estimators.

The plan for the paper is as follows. Section 2 introduces the BC and derives some of its mathematical and statistical properties. In Section 3 we introduce the important extension of the Bernstein representation of given copulas. Issues related to making the polynomial representation operational are considered in Section 4, where, as mentioned previously, it is shown that empirical estimation of the BC density has remarkably good properties in terms of its asymptotic variance. Moreover, a short simulation is provided to highlight some of the theoretical results in a finite sample. Other estimation procedures are possible, and we discuss some, but by no means all, of the issues in Section 5. Proofs of some results appear in the Appendix.

2. THE BERNSTEIN COPULA

We first recall Sklar's representation for multivariate distributions. Let H be a k-dimensional distribution function with one-dimensional marginals $F_1,\dots,F_k$; then there exists a function C from the unit k-cube to the unit interval such that

$$H(x_1,\dots,x_k) = C(F_1(x_1),\dots,F_k(x_k));$$

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

We list some properties of the copula function C.

(1) C is nondecreasing in all its arguments;
(2) C satisfies the Fréchet bounds

$$\min(0,u_1+\cdots+u_k-(k-1)) \le C(u_1,\dots,u_k) \le \min(u_1,\dots,u_k),$$

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

(3) $\prod_{j=1}^k u_j$ is a copula for independent random variables, which is called the product copula;
(4) C is Lipschitz with constant one; i.e.,

$$|C(x_1,\dots,x_k) - C(y_1,\dots,y_k)| \le \sum_{j=1}^k |x_j - y_j|.$$

Properties (1) and (2) are necessary and sufficient for the definition of the copula function.

Let $\alpha(v_1/m_1,\dots,v_k/m_k)$ be a real-valued constant indexed by $(v_1,\dots,v_k)$, such that $0 \le v_j \le m_j \in \mathbb{N}$. We could also use $\alpha_{v_1,\dots,v_k}$, $0 \le v_j \le m_j$, $\forall j$. Let

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

Define the following map, $C_B\colon [0,1]^k \to [0,1]$, where

$$C_B(u_1,\dots,u_k) = \sum_{v_1=0}^{m_1}\dots\sum_{v_k=0}^{m_k} \alpha\left(\frac{v_1}{m_1},\dots,\frac{v_k}{m_k}\right) P_{v_1,m_1}(u_1)\cdots P_{v_K,m_k}(u_k). \qquad \textbf{(2)}$$

Under specific conditions on $\alpha(v_1/m_1,\dots,v_k/m_k)$, (2) can be shown to be a copula. In particular, (2) is defined in terms of k-dimensional Bernstein polynomials. We state the following definition.

DEFINITION 1. *A Bernstein polynomial approximation to* $f \in C_{[0,1]^k}$ *(*$C_{[0,1]^k}$ *is the space of continuous bounded functions on* $[0,1]^k$*) is obtained by applying a linear operator* B_m^k *to* $f \in C_{[0,1]^k}$ *such that*

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

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

$$(B_m^k f)(\mathbf{x}) = \int_0^1 \cdots \int_0^1 f(t_1,..,t_k)\, d_{t_1} K_{m_1}(x_1,t_1)\cdots d_{t_k} K_{m_k}(x_k,t_k), \tag{4}$$

for the kernel

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

$$K_m(x,0) \equiv 0,$$

which is constant for $v/m \le t < (v+1)/m$ and has jumps of $\binom{m}{v} x^v (1-x)^{m-v}$ at points $t = v/m$.

The singular integral representation establishes some clear parallels to kernel density estimation in statistics. One important property of multivariate Bernstein polynomials is given next (e.g., DeVore and Lorentz, 1993, p. 10).

LEMMA 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 $C_{[0,1]^k}$.*

Further, the derivatives of a Bernstein polynomial approximation converge to the approximand whenever this is differentiable (see, e.g., (12) and (13), which follow). This concept and its usefulness will become clear in Section 2.2.

Throughout the paper, we reserve the symbols C_B, C_n, and C for the BC, the empirical copula based on n observations, and a general copula, respectively, and use c_B for the BC density; definitions will be given in due course. The letters m and n will only be used to define the order of polynomial and the sample size, respectively.

2.1. The Definition of Bernstein Copula

To restrict (2) to be a copula function, we need to impose some conditions on $\alpha(v_1/m_1,\ldots,v_k/m_k)$.

THEOREM 1. *$C_B(u_1,\ldots,u_k)$ is a copula function if*

$$\sum_{l_1=0}^{1} \cdots \sum_{l_k=0}^{1} (-1)^{l_1+\cdots+l_k} \alpha\left(\frac{v_1+l_1}{m_1},\ldots,\frac{v_k+l_k}{m_k}\right) \ge 0, \tag{5}$$

letting $0 \le v_j \le m_j - 1, j = 1,\ldots,k$, *and*

$$\min\left(0, \frac{v_1}{m_1} + \ldots + \frac{v_k}{m_k} - (k-1)\right) \le \alpha\left(\frac{v_1}{m_1},\ldots,\frac{v_k}{m_k}\right) \le \min\left(\frac{v_1}{m_1},\ldots,\frac{v_k}{m_k}\right), \tag{6}$$

in particular

$$\lim_{v_j \to 0} \alpha\left(\frac{v_1}{m_1},\ldots,\frac{v_k}{m_k}\right) = 0, \quad \forall j = 1,\ldots,k, \tag{7}$$

and

$$\alpha\left(1,\ldots,1,\frac{v_j}{m_j},1,\ldots,1\right) = \frac{v_j}{m_j}, \quad \forall j = 1,\ldots,k. \tag{8}$$

Proof. See the Appendix. ∎

As a result of (7) in Theorem 1, we could take $1 \le v_j \le m_j$, $\forall j$, in (2). Theorem 1 gives sufficient conditions that are necessary only if $m \to \infty$.

Therefore, we can define the object to be studied in this paper.

DEFINITION 2. *If* $\alpha(v_1/m_1,\ldots,v_k/m_k)$ *satisfies (5) and (6), we call* $C_B(u_1,\ldots,u_k)$ *in (2) the BC.*

From Theorem 1 we see that a BC is a genuine copula, and by Lemma 1, we know that the BC may serve as an approximation to a given arbitrary parametric copula. By the properties of Bernstein polynomials, the coefficients of the BC have a direct interpretation as the points of some arbitrary approximated copula. In this approach, we use the terms ABC or EBC (see Sections 3 and 4).

The BC belongs to the family of polynomial copulas. Polynomial copulas are special cases of copulas with polynomial sections in one or more variables (details on copulas with polynomial sections can be found in Nelsen, 1998, Ch. 3).

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

$$\begin{aligned} C_B(u_1,\ldots,u_k) &= u_1 \cdots u_k + \sum_{v_1=0}^{m_1} \ldots \sum_{v_k=0}^{m_k} \gamma\left(\frac{v_1}{m_1},\ldots,\frac{v_k}{m_k}\right) P_{v_1,m}(u_1) \cdots P_{v_K,m}(u_k) \\ &= \sum_{v_1=0}^{m} \ldots \sum_{v_k=0}^{m} \alpha\left(\frac{v_1}{m_1},\ldots,\frac{v_k}{m_k}\right) P_{v_1,m}(u_1) \cdots P_{v_K,m}(u_k), \end{aligned} \tag{9}$$

where

$$\gamma\left(\frac{v_1}{m_1},\ldots,\frac{v_k}{m_k}\right) = \alpha\left(\frac{v_1}{m_1},\ldots,\frac{v_k}{m_k}\right) - \frac{v_1}{m_1}\cdots\frac{v_k}{m_k}. \tag{10}$$

The equality follows from the fact that

$$u_1\cdots u_k = \sum_{v_1=0}^{m_1}\cdots\sum_{v_k=0}^{m_k}\left(\frac{v_1}{m_1},\ldots,\frac{v_k}{m_k}\right)P_{v_1,m}(u_1)\cdots P_{v_K,m}(u_k).$$

2.2. The Bernstein Density

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

$$\Delta_{1,\ldots,k}\alpha\left(\frac{v_1}{m},\ldots,\frac{v_k}{m}\right) \equiv \sum_{l_1=0}^{1}\cdots\sum_{l_k=0}^{1}(-1)^{k+l_1+\cdots+l_k}\alpha\left(\frac{v_1+l_1}{m},\ldots,\frac{v_k+l_k}{m}\right), \tag{11}$$

where for the ease of notation, we set $m_j = m\ \forall j$. By the properties of Bernstein polynomials, the BC density has the following appealing structure:

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

where $c_B = (\partial^k C_B)/\partial u_1\cdots\partial u_k$ and the expression is obtained by direct differentiation of (2) with respect to each variable and rearranging (for the univariate case, see Lorentz, 1953). Differentiating, a term in the summation is lost, and the coefficients of the polynomial are written in difference form that is directly linked to the k-dimensional rectangle inequality in (5); i.e., the copula density is always positive.

For convenience, we use the following definition for the BC density:

$$c_B(u_1,\ldots,u_k) = \sum_{v_1=0}^{m}\cdots\sum_{v_k=0}^{m}\beta\left(\frac{v_1}{m},\ldots,\frac{v_k}{m}\right) \times\prod_{j=1}^{k}\binom{m}{v_j}u_j^{v_j}(1-u_j)^{m-v_j}, \tag{12}$$

where $\beta(v_1/m,\ldots,v_k/m)$ is defined accordingly; i.e.,

$$\beta\left(\frac{v_1}{m},\ldots,\frac{v_k}{m}\right) \equiv (m+1)^k\Delta_{1,\ldots,k}\alpha\left(\frac{v_1}{m+1},\ldots,\frac{v_k}{m+1}\right). \tag{13}$$

Notice that $\Delta\alpha(v/m) = \alpha(v/m + 1/m) - \alpha(v/m)$ so that

$$\lim_{m\to\infty} \Delta\alpha\left(\frac{v}{m}\right)m = (\partial/\partial v)\alpha\left(\frac{v}{m}\right)$$

(when the derivative exists) and similarly in higher dimensions; i.e.,

$$\lim_{m\to\infty} \beta\left(\frac{v_1}{m},\ldots,\frac{v_k}{m}\right) = (\partial^k/\partial v_1\cdots\partial v_k)\alpha\left(\frac{v_1}{m+1},\ldots,\frac{v_k}{m+1}\right)$$

(when this exists).

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

We consider the moments of the BC. Many operations find convenient representation in terms of hypergeometric functions. For an introduction to hypergeometric functions for economists and many of the symbols we use, the reader is referred to Abadir (1999).

The copula is

$$C_B(u_1,\ldots,u_k) = \sum_{v_1=0}^{m}\cdots\sum_{v_k=0}^{m} \alpha\left(\frac{v_1}{m},\ldots,\frac{v_k}{m}\right) \times \prod_{j=1}^{k}\binom{m}{v_j}u_j^{v_j}(1-u_j)^{n_j-v_j},$$

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

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

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 $\frac{1}{12}$,

$$\rho_S = 12\mathrm{cov}(u,v) = 12E(uv) - 3,$$

where

$$E(uv) = \int uvdC(u,v) = \int (1-u-v+C(u,v))\,dudv,$$

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 (for further discussion, see Schweizer and Wolff, 1981). Random variables, which have zero covariance, could have nonzero ρ_S. The use of ρ_S in financial economics could be advocated on the basis of the documented nonlinearities and its simple estimation. For the BC ρ_S can be derived as follows:

$$\begin{aligned}\rho_S &= 12\int_0^1\int_0^1 [1-u_1-u_2+C_B(u_1,u_2,1,\dots,1)]\,du_1\,du_2-3\\&=12\sum_{v_1=0}^{m}\sum_{v_2=0}^{m}\gamma\left(\frac{v_1}{m},\frac{v_2}{m},1,\dots,1\right)\\&\quad\times\prod_{j=1,2}\binom{m}{v_j}\int_0^1\int_0^1 u_j^{v_j}(1-u_j)^{m-v_j}\,du_1\,du_2\\&=12\sum_{v_1=0}^{m}\sum_{v_2=0}^{m}\gamma\left(\frac{v_1}{m},\frac{v_2}{m},1,\dots,1\right)\\&\quad\times\prod_{j=1,2}\binom{m}{v_j}B(v_j+1,m+1-v_j),\end{aligned}$$

where γ was defined in (10) and $B(a,b)$ is the beta function. The first equality follows by writing the BC as the sum of the product copula and the perturbation term. Notice that

$$12\int_0^1\int_0^1 (1-u_1-u_2+u_1u_2)\,du_1\,du_2=3.$$

Even when the BC is used as an approximation (see the next section), the preceding 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 for evaluating complicated integrals.

We derive the moment generating function (m.g.f.)of the density in (12). We do it for the one variable case, and then we just extend it to the k-dimensional case.

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

where $\beta(v/m)$ is given by (13) 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 $a \equiv v+1$, $c \equiv m+2$, and $z \equiv t$ this implies

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

Therefore,

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

To obtain the m.g.f. for the k-dimensional BC density, we replace the univariate result in the multivariate definition. Defining $M_c(\mathbf{t})$ as the m.g.f. of (12), for $\mathbf{t}$ a $k \times 1$ vector,

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

where $\bullet$ is the inner product. These results can be used to further investigate the properties of the BC and its approximations. Deriving results for the joint moments of the BC is quite easy in virtue of its incomplete beta function representation. The joint moments are important for studying the scale-free dependence properties of the variables.

3. ABC: BERNSTEIN REPRESENTATION OF ARBITRARY COPULAE

Although the BC 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 BC can be used in place of the original copula. As mentioned in Section 1, we call this the approximate Bernstein copula (ABC) representation. By the approximation 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 2, which follows). 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).

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

$$C(u,v) \equiv (u^{-\theta} + v^{-\theta} - 1)^{-(1/\theta)}, \qquad \theta \geq 0. \tag{14}$$

Notice that this copula has unbounded density. Figure 1 shows the three-dimensional graph of the KS copula density.

In Table 1, we report the values of Spearman's rho as a function of the dependence parameter θ in the approximation for $m = 10, 30, 50, 100, 200, 300$, and the corresponding values for the KS copula. Because of computational difficulties we do not calculate the approximation when the dependence parameter achieves its limiting value (∞). Figure 2 shows the contour plot of the two copulas when $\theta = 1.06$ and $m = 30$. These seem to be indistinguishable.

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 Spearman's rho for the KS copula can only be done numerically.

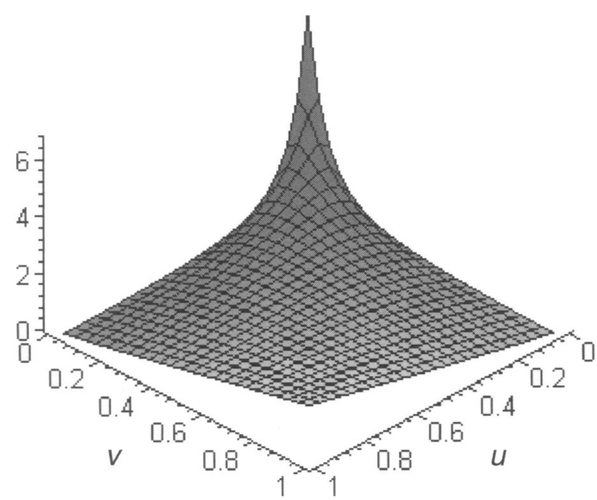

FIGURE 1. KS copula density, $\theta = 1.06$.

TABLE 1. Spearman's rho for different values of the dependence parameter θ

θ	0	0.14	0.31	0.51	0.76	1.06	1.51	2.14	3.19	5.56	∞
$\rho_S(KS)$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$\rho_S(B_{10})$	0	0.08	0.16	0.24	0.32	0.4	0.48	0.57	0.65	0.73	*
$\rho_S(B_{30})$	0	0.09	0.19	0.28	0.37	0.46	0.56	0.65	0.75	0.84	*
$\rho_S(B_{50})$	0	0.09	0.19	0.29	0.38	0.48	0.58	0.67	0.77	0.86	*
$\rho_S(B_{100})$	0	0.1	0.2	0.29	0.39	0.49	0.59	0.69	0.78	0.88	*
$\rho_S(B_{200})$	0	0.1	0.2	0.3	0.39	0.49	0.59	0.69	0.79	0.89	*
$\rho_S(B_{300})$	0	0.1	0.2	0.3	0.4	0.49	0.6	0.7	0.8	0.89	*

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

THEOREM 2. *Let $f \in C_{[0,1]^k}$ and $\partial f/\partial x_j$ be Lipschitz $\forall j$; then*

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

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

Proof. See the Appendix. ■

This rate of approximation can be improved using linear combinations of Bernstein polynomials (see Butzer, 1953). Let $f^{(2l)}$ (i.e., the $2l$ derivative of f) be Lipschitz of order γ; then Butzer (1953) shows that one can construct a linear combination of one-dimensional Bernstein polynomials of order m such that the error is $O(m^{-l-\gamma})$ compared with $O(n^{-2l-\gamma})$ for the best approximating polynomials of order m (Jackson, 1930, p. 18). However, the coefficients of Bernstein polynomials have clear interpretations in the case of both the copula and the copula density. Further, all the derivatives of Bernstein polynomials converge to the true derivatives of the approximated function when the latter is differentiable. Other optimal properties of Bernstein polynomials in connection with the copula function and joint continuity of Markov operators are discussed in Li, Mikusiński, and Taylor (1998) and Kulpa (1999).

A limitation of the BC is that it cannot be used to model extreme tail behavior defined in terms of the coefficient of tail dependence.[1] This happens because convergence under the sup norm is not sufficient for ensuring that the BC and its approximand converge to an arbitrary limit at the same speed. Nevertheless, the BC can capture increasing dependence as we move to the tails. In fact, the KS copula in (14) exhibits lower tail dependence, and the BC can approximate arbitrarily well the tail behavior of this copula inside the cube.

4. EMPIRICAL BERNSTEIN COPULA: AN ESTIMATION PROCEDURE

There are several possible estimation procedures that can be employed to make the BC operational. However, the BC might also be used as an estimator for

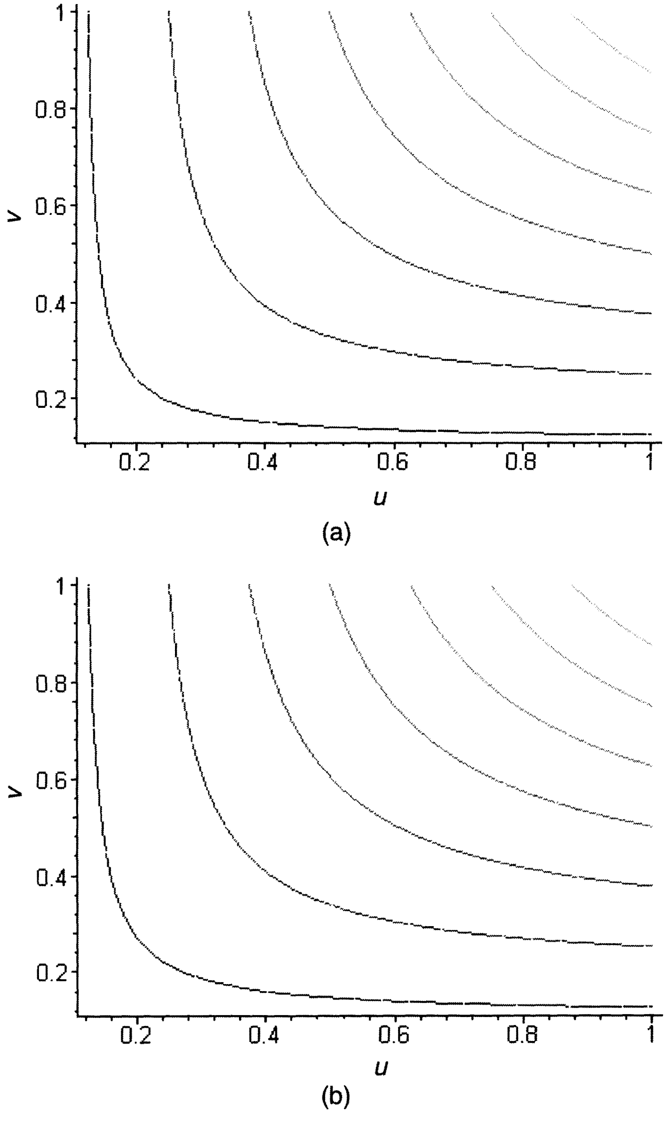

FIGURE 2. (a) KS copula density, $\theta = 1.06$. (b) ABC for the KS copula ($\theta = 1.06$, $m = 30$).

unknown copulas. In this case, we would use the BC density as an estimator, where the coefficients of the copula are given by the empirical copula and the order of polynomial would be related to the smoothing properties of the estimator. As mentioned in Section 1, we call this the empirical Bernstein copula (EBC) density.

4.1. The Empirical Bernstein Copula Density

Let $C_n(v_1/m,\dots,v_k/m)$ be the empirical copula at $(v_1/m,\dots,v_k/m)$, i.e.,

$$\frac{1}{n}\sum_{s=1}^{n} I\left\{\bigcap_{j=1}^{k}[u_{js}\le t_{v_j}]\right\}, \tag{15}$$

where $I_{\{\mathcal{A}\}}$ is the indicator of the set $\mathcal{A}$ and $t_{v_j}\equiv v_j/m$. By construction, the empirical copula is a valid distribution function. However, it has marginals that are uniform only asymptotically; i.e.,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{s=1}^{n} I\{u_{js}\le u_j\}=u_j,$$

for $j=1,\dots,k$. Therefore it is a valid copula only asymptotically.[2] The EBC, say, $\tilde{C}_B(\mathbf{u})$, is defined as the usual BC where

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

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

$$\begin{aligned}\tilde{c}_B &= \sum_{v_1=0}^{m-1}\cdots\sum_{v_k=0}^{m-1}\triangle_{1,\dots,k}\left(\frac{m^k}{n}\sum_{s=1}^{n} I\left\{\bigcap_{j=1}^{k}[u_{js}\le t_{v_j}]\right\}\right)\\ &\quad\times\prod_{j=1}^{k}\binom{m-1}{v_j}u_j^{v_j}(1-u_j)^{m-1-v_j},\end{aligned} \tag{16}$$

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

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

The optimal choice of m depends on the topology we use. We choose m to minimize the mean square error (MSE) of the density; i.e., $\min\|\tilde{c}_B - c\|_2^2$ where $\|\dots\|_2$ 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 Mean Square Error of the Estimated Bernstein Copula Density

We want to choose (16) such that it is optimal under the L_2 norm. The asymptotic properties of this estimator under the L_2 norm have been studied in detail in Sancetta (2003) under the following condition.

Condition 1. $\mathbf{u}_1,\ldots,\mathbf{u}_n$ is a sequence of independent strictly stationary uniform $[0,1]^k$ random vectors with copula $C(\mathbf{u})$ and copula density $c(\mathbf{u})$ that has finite first derivatives everywhere in the k-cube.

Remark. The independence condition is not required, but we use it to shorten the proof as much as possible. From the proof of Theorem 3 it can be seen that the results are still valid under appropriate mixing conditions by use of known coupling results.

We use $\lesssim$ to indicate greater or equal up to a multiplicative absolute constant; e.g., $a \lesssim b$ implies $\exists C$, $0 < C < \infty$, such that $a \leq Cb$. (This notation is commonly used in the empirical process literature; see, e.g., van der Vaart and Wellner, 2000.)

THEOREM 3. *Let $\tilde{c}_B$ be the k-dimensional* BC *density. Under Condition 1,*

i. $Bias(\tilde{c}_B) \lesssim m^{-1}$.
ii. *Let* $\lambda_j \equiv [u_j(1-u_j)]^{1/2}$;
(a) for $u_j \in (0,1), \forall j$,

$$\operatorname{var}(\tilde{c}_B) \lesssim \left(n \prod_{j=1}^{k} \lambda_j\right)^{-1} m^{k/2}(1+m^{-1}),$$

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

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

iii.

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

in MSE
(a) for $u_j \in (0,1), \forall j$, *if* $m^{k/2}/n \to 0$ *as* $m,n \to \infty$;
(b) for $u_j = 0,1, \forall j$, *if* $m^k/n \to 0$ *as* $m,n \to \infty$.
iv. The optimal order of the polynomial in an MSE *sense is*
(a) $m \lesssim n^{2/(k+4)}$ *if* $u_j \in (0,1), \forall j$;
(b) $m \lesssim n^{1/(k+2)}$ *if* $u_j = 0,1, \forall j$.
v. If $m - k \geq 2$, *then* $\tilde{c}_B(\mathbf{u})$ *and* $\tilde{C}_B(\mathbf{u})$ *are Donsker in* $(0,1)^k$; *i.e.,* $z_B(\mathbf{u}) \equiv \sqrt{m^{-k/2}n}[\tilde{c}_B(\mathbf{u}) - E\tilde{c}_B(\mathbf{u})]$ *and* $\tilde{Z}_B(\mathbf{u}) \equiv \sqrt{n}[\tilde{C}_B(\mathbf{u}) - E\tilde{C}_B(\mathbf{u})]$ *converge to a zero mean Gaussian process with continuous sample paths and covariance function* $E[z_B(\mathbf{u}_1)z_B(\mathbf{u}_2)]$ *and* $E[\tilde{Z}_B(\mathbf{u}_1)\tilde{Z}_B(\mathbf{u}_2)]$, *respectively.*

The proof of Theorem 3 is given in the next section. The weak limit implies that convergence is uniform: if a class of functions is Donsker, then it is Glivenko–Cantelli; i.e., convergence holds uniformly over the class (see, e.g., van der Vaart and Wellner, 2000, p. 82). Therefore, the results of Theorem 3 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 bias for the histogram estimator. In this respect, kernel smoothers would lead to a bias not higher than $O(m^{-2})$, as opposed to $O(m^{-1})$ in our case. Notice that it is not possible to reduce the bias to $O(m^{-2})$ by shifting the histogram, i.e., using frequency polygons (for details on frequency polygons, see, e.g., Scott, 1992, Ch. 4). In this case the first term in the expansion would vanish, but other terms of the same order would not. Further, the $O(k/m)$ rate of approximation of k-dimensional Bernstein polynomials imposes a lower bound on the bias unless linear combinations of Bernstein polynomials are used (see the discussion at the end of Section 3).

Although the bias is of the same order as the histogram estimator, the variance is of smaller order (except at the edges of the hypercube): $\mathrm{var}(\tilde{c}_B) = O(m^{k/2})$ instead of $O(m^k)$ 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 common 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 $O(m^{1/2})$ for all the coordinates inside the k-hypercube, whereas for the coordinates on the boundaries the contribution to the variance is $O(m)$.

As m and n go to infinity, it follows that this estimator has a rate of consistency $m^{k/2}/n \to 0$ inside the hypercube versus $m^k/n \to 0$ for other common nonparametric estimators, e.g., the Gaussian kernel. Inside the hypercube, the optimal order of smoothing is $m = O(n^{2/(k+4)})$ in MSE sense versus $m = O(n^{1/(k+2)})$ for the histogram and $m = O(n^{1/(k+4)})$ for a first-order kernel.

This implies that the Bernstein polynomials require very little smoothing (i.e., a large order of polynomial, which implies a less smooth graph). This is due to the fact that Bernstein polynomials are fairly slow to adjust, as already mentioned in Section 3.

4.3. Simulation

In this section, we use some short simulations to investigate the finite sample properties of the EBC density. We choose the KS copula in (14) as the true one. In particular, we let $\theta = 0.6$. This corresponds to $\rho_S = 0.34$. Figure 3 shows the plot for the copula density of the KS copula with dependence parameter $\theta = 0.6$. The KS copula is singular at the origin. As a condition in Theorem 3 we used the fact that the copula density is nonsingular. Therefore, comparing this copula density with the EBC will be of interest for several reasons.

We notice that this copula density exhibits lower tail dependence. This means specifically that random variables become more dependent on the left side of their support. This appears to be an important property when modeling joint financial returns: asset returns are more dependent when they are negative (for the case of daily financial returns, see Silvapulle and Granger, 2001; Fortin

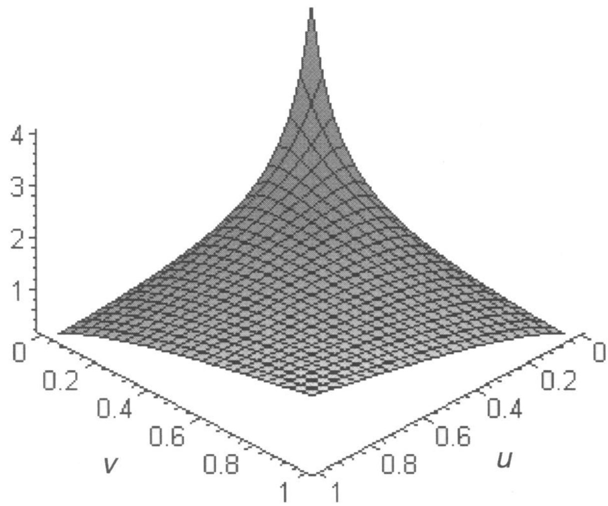

FIGURE 3. KS copula density, $\theta = 0.6$.

and Kuzmics, 2002). By the properties of the copula function we are not concerned with the marginal distributions: these either could be known or could be estimated using a parametric specification or the empirical distribution function. Assuming the marginals are correctly specified, then, by definition, the copula is the joint distribution of uniform $[0,1]^k$ marginals. For example, if we were to model daily financial returns parametrically, the modified Weibull distribution or the double gamma could be a good choice (e.g., Knight et al., 1995; Laherrère and Sornette, 1998). Therefore, the copula allows us to divide the estimation problem into two steps, which are independent. In this simulation, we only consider the performance of our estimator, so that the marginals are uniform. This corresponds to the case of marginals being known. Carrying out a simulation of the whole joint distribution (i.e., copula and marginals) might be of interest also. However, in this case, the error in the parametric estimation of the marginal would confound the error in the estimation of the copula. On the other hand, the purpose of this simulation is to provide a brief direct illustration of the small sample properties of our estimator and their link to the theoretical asymptotic results of Theorem 3.

The realism of the simulation needs to be discussed. We have arbitrary marginals that could reflect skewness and kurtosis and a copula that has, as mentioned previously, lower tail dependence in line with current empirical research on financial returns. It is the latter we are focusing on in our simulation.

We generate 100 samples of $n = 500$ observations from the KS copula with $\theta = 0.6$; the value of n corresponds to the number of our daily observations (for details on simulating data from multivariate distributions and copulas, see

Li, Scarsini, and Shaked, 1996). We consider estimation of the BC density for $m = 4,6,\ldots,40$. This boils down to finding the histogram estimator with bin width equal to m^{-1} and then applying the Bernstein operator to it. We plot the estimated copula for $m = 12$ in Figure 4.

To study the global performance of our estimator, we should look at the integrated mean square error (IMSE); i.e.,

$$\int_{[0,1]^2} [E(c(u,v) - \tilde{c}_B(u,v))^2]\, dudv$$
$$= \int_{[0,1]^2} [\mathrm{Bias}(\tilde{c}_B(u,v))]^2\, dudv + \int_{[0,1]^2} \mathrm{var}(\tilde{c}_B(u,v))\, dudv.$$

For each m, we use our 100 samples of 500 observations from the KS copula to compute an approximate bias squared and variance, say, $\overline{\mathrm{Bias}}(\tilde{c}_B(u,v))^2$ and $\overline{\mathrm{var}}(\tilde{c}_B(u,v))$. Both $\overline{\mathrm{Bias}}(\tilde{c}_B(u,v))^2$ and $\overline{\mathrm{var}}(\tilde{c}_B(u,v))$ are obtained by evaluating $\tilde{c}_B(u,v)$ at 10,000 uniform $[0,1]^2$ pseudorandom points (for each m, we just use the same points). Therefore, we approximate the preceding integral by averaging over these 10,000 points; i.e.,

$$\int_{[0,1]^2} [\mathrm{Bias}(\tilde{c}_B(u,v))]^2\, dudv \simeq \sum_{j=1}^{10,000} \frac{\overline{\mathrm{Bias}}(\tilde{c}_B(u_j,v_j))^2}{10,000},$$
$$\int_{[0,1]^2} \mathrm{var}(\tilde{c}_B(u,v))\, dudv \simeq \sum_{j=1}^{10,000} \frac{\overline{\mathrm{var}}(\tilde{c}_B(u_j,v_j))}{10,000},$$

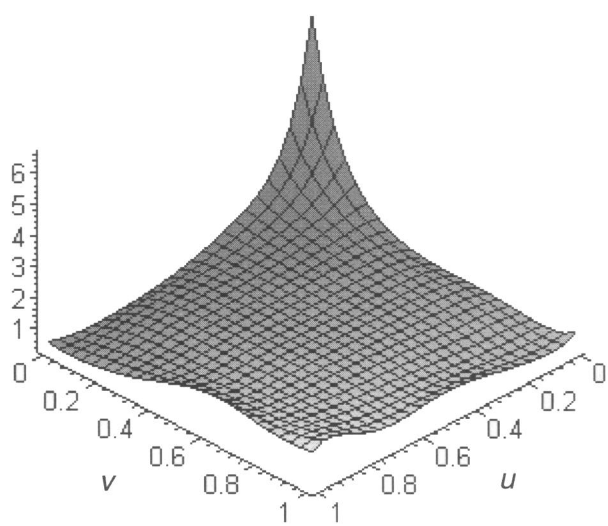

FIGURE 4. EBC density, $m = 12$.

where $\{(u_j, v_j)\}_{j=1}^{10,000}$ is our uniform $[0,1]^2$ sample of points where both the KS copula density and the EBC density are evaluated. We summarize our results in Table 2.[3] We also report the same results when the copula is estimated using the two-dimensional histogram estimator. Because the EBC density is obtained from the histogram, it is natural to make comparisons with the two-dimensional histogram.

The main result of Table 2 is that applying the Bernstein operator to the two-dimensional histogram estimator allows us to decrease the size of the mesh and, consequently the bias, keeping the uncertainty (i.e., the variance) low. Further, the results for the ECB do not seem to be particularly sensitive to m as opposed to the histogram.

The results agree with the asymptotic results in Theorem 3 in the sense that the variance grows linearly in m as opposed to the histogram estimator, whose variance appears to be quadratic in m. The biases also agree with Theorem 3 (the EBC and the histogram have similar biases), though not as accurately in small samples. In Table 3 we report the results from regressing the log of the integrated variance of the two estimators (the EBC and the histogram) on the

TABLE 2. IMSE for simulation

	EBC			Histogram		
m	Int. Bias^2	Int. Var.	IMSE	Int. Bias^2	Int. Var.	IMSE
2	0.307	0.002	0.308	0.300	0.007	0.306
4	0.242	0.005	0.247	0.235	0.031	0.266
6	0.207	0.009	0.216	0.205	0.070	0.275
8	0.186	0.012	0.198	0.187	0.126	0.313
10	0.169	0.015	0.185	0.171	0.196	0.367
12	0.157	0.019	0.176	0.159	0.294	0.453
14	0.147	0.023	0.169	0.148	0.390	0.538
16	0.138	0.026	0.164	0.143	0.512	0.656
18	0.130	0.030	0.159	0.137	0.640	0.777
20	0.121	0.033	0.154	0.128	0.792	0.920
22	0.114	0.037	0.150	0.127	0.969	1.095
24	0.109	0.040	0.149	0.114	1.151	1.265
26	0.104	0.043	0.147	0.112	1.327	1.439
28	0.098	0.047	0.145	0.105	1.547	1.653
30	0.093	0.051	0.145	0.103	1.787	1.890
32	0.088	0.055	0.143	0.103	2.030	2.133
34	0.082	0.058	0.141	0.102	2.291	2.393
36	0.079	0.062	0.141	0.099	2.586	2.685
38	0.075	0.066	0.140	0.097	2.853	2.950
40	0.072	0.069	0.141	0.090	3.171	3.260

Note: Figures are rounded to three decimal places.

TABLE 3. Regression of log(Int. var) on log(m)

EBC Estimator			Histogram Estimator		
Coefficient	Value	Std. Error	Coefficient	Value	Std. Error
(Intercept)	−6.715	0.0154	(Intercept)	−6.238	0.0083
log(m)	1.1003	0.0051	log(m)	2.0054	0.0027
Multiple R-squared:		0.9992	Multiple R-squared:		0.9999

log of m. We use $m = 4,5,\dots,41$ with the corresponding estimated variances (i.e., 38 observations). The slopes of the regression support the conclusions of Theorem 3. For the sake of conciseness, we do not report results from the regression of the biases because the coefficients of $\log m$ are similar for the two estimators (as implied by Theorem 3) but very sensitive to the choice of range of m due to the influence of higher order terms.

4.4. Proof of Theorem 3

The proof of part ii of Theorem 3 is based on the normal approximation to the binomial distribution (e.g., see Stuart and Ord, 1994, pp. 138–140). In particular, let

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

and

$$\mathcal{P}_{v,m}(v) \equiv (2\pi u(1-u)m)^{-1/2} \exp\left\{-\frac{m}{2u(1-u)}\left(\frac{v}{m}-u\right)^2\right\}. \tag{17}$$

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 $z = (v/m - u)$ to prove that the error is uniform. Taking squares of the two distributions (i.e., the binomial and the Gaussian),

$$\sum_{v=0}^{m} f\left(\frac{v}{m}\right) (P_{v,m}(u))^2 \simeq \int_{-\infty}^{\infty} f\left(\frac{v}{m}\right) (\mathcal{P}_{v,m}(v))^2\, dv, \tag{18}$$

where again the error holds uniformly. With this remark, we can prove Theorem 3.

Notation. *We use* $\mathbf{u}_1,\ldots,\mathbf{u}_n$ *to indicate n copies of k-dimensional uniform random vectors in* $[0,1]^k$. *On the other hand* $\mathbf{u}=(u_1,\ldots,u_k)$ *denotes a fixed, but arbitrary, k-dimensional vector. Following Abadir and Magnus (2000), we define* $D_i c(\mathbf{u}) \equiv (\partial c(\mathbf{u}_s)/\partial u_i)|_{\mathbf{u}_s=\mathbf{u}}$.

Proof of Theorem 3. $\text{Bias}(\tilde{c}_B) \equiv E(\tilde{c}_B) - c(\mathbf{u})$, which we can rewrite as

$$E\int c_n(\mathbf{t})\, d_{\mathbf{t}} K_m(\mathbf{u},\mathbf{t}) - c(\mathbf{u}).$$

Clearly,

$$\begin{aligned} E\int c_n(\mathbf{t})\, d_{\mathbf{t}} K_m(\mathbf{u},\mathbf{t}) - c(\mathbf{u}) &\le E\int [c_n(\mathbf{t}) - c(\mathbf{t})]\, d_{\mathbf{t}} K_m(\mathbf{u},\mathbf{t}) \\ &\quad + E\left(\int c(\mathbf{t})\, d_{\mathbf{t}} K_m(\mathbf{u},\mathbf{t}) - c(\mathbf{u})\right) \\ &\le E\int [c_n(\mathbf{t}) - c(\mathbf{t})]\, d_{\mathbf{t}} K_m(\mathbf{u},\mathbf{t}) + O\left(\frac{k}{m}\right), \end{aligned}$$

where the second inequality follows from the same argument as in the previous proof. Because the Bernstein operator is bounded, by Fubini's theorem, it is sufficient to consider

$$E[c_n(\mathbf{t}) - c(\mathbf{t})],$$

i.e., the bias for a histogram estimator. Therefore (e.g., Scott, 1992, p. 81; or just use (22), which follows, where no Taylor expansion is required),

$$E[c_n(\mathbf{t}) - c(\mathbf{t})] = O(m^{-1}).$$

Therefore,

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

For the variance, notice that the probability of one observation falling inside a subset of the hypercube is equal to the probability of a 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

$$\text{var}(\tilde{c}_B) = \sum_{v_1=0}^{m-1} \cdots \sum_{v_k=0}^{m-1} \left(\frac{m^{2k}}{n^2} \sum_{s=1}^{n} (p_{s,v_1\cdots v_k} - (p_{s,v_1\cdots v_k})^2)\right) \prod_{j=1}^{k} (P_{v_j,m-1}(u_j))^2, \tag{19}$$

where

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

Consider the following simple identity for any differentiable function f:

$$\int_t^{t+(1/m)} f(u)\, du = \frac{f(t)}{m} - \int_t^{t+(1/m)} \left(u - t - \frac{1}{m}\right) df(u), \tag{20}$$

where the left-hand side can be recovered by simple integration of the second term on the right-hand side. By direct application of (20) we have

$$\begin{aligned} p_{s,v_1\cdots v_k} = \int_{t_{v_k}}^{t_{v_k}+(1/m)} \cdots \int_{t_{v_2}}^{t_{v_2}+(1/m)} \Bigg(& \frac{c(t_{v_1}, u_{2s}, \ldots, u_{ks})}{m} \\ & - \int_{t_{v_1}}^{t_{v_1}+(1/m)} \left(u_{1s} - t_{v1} - \frac{1}{m}\right) \\ & \times D_1 c(u_{1s}, u_{2s}, \ldots, u_{ks})\, du_{1s} \Bigg) du_{2s} \cdots du_{ks}. \end{aligned} \tag{21}$$

Now, by Condition 1, $D_1 c(u_{1s}, u_{2s}, \ldots, u_{ks}) \le M$ for some $M < \infty$. Therefore, the last term in (21) is bounded by

$$M \int_{t_{v_1}}^{t_{v_1}+(1/m)} \left(t_{v1} + \frac{1}{m} - u_{1s}\right) du_{1s} = O(m^{-2}).$$

Substituting in (21),

$$p_{s,v_1\cdots v_k} = \int_{t_{v_k}}^{t_{v_k}+(1/m)} \cdots \int_{t_{v_2}}^{t_{v_2}+(1/m)} \left(\frac{c(t_{v_1}, u_{2s}, \ldots, u_{ks})}{m} + O(m^{-2}) \right) du_{2s} \cdots du_{ks}.$$

Applying (20) repeatedly, we have

$$p_{s,v_1\cdots v_k} = \frac{c(t_{v_1}, \ldots, t_{v_k})}{m^k} + O(m^{-(k+1)}). \tag{22}$$

Because $p_{s,v_1\cdots v_k} \le 1$, it follows that $(p_{s,v_1\cdots v_k})^2 = o(p_{s,v_1\cdots v_k})$. Therefore, substituting (22) in (19), we have that

$$\operatorname{var}(\tilde{c}_B) = \frac{m^{2k}}{n} \sum_{v_1=0}^{m-1} \cdots \sum_{v_k=0}^{m-1} \left(\frac{c(t_{v_1}, \ldots, t_{v_k})}{m^k} + O(m^{-(k+1)}) \right) \prod_{j=1}^{k} (P_{v_j, m-1}(u_j))^2. \tag{23}$$

We use (18) to approximate (23). By Condition 1, $c(t_{v_1}, \ldots, t_{v_k})$ is bounded, say, by $M < \infty$. Recall that $t_{v_j} \equiv v_j/m$, $j = 1, \ldots, k$. Consequently, solve the following type of integral:

$$\Gamma_j = \int_{\mathbb{R}} c\left(\frac{v_1}{m}, \ldots, \frac{v_k}{m}\right) \frac{\exp\left\{-\frac{(m-1)}{u_j(1-u_j)}\left(\frac{v_j}{m-1} - u_j\right)^2\right\}}{[2\pi(m-1)u_j(1-u_j)]} dv_j$$

$$\leq M \int_{\mathbb{R}} \frac{\exp\left\{-\frac{(m-1)}{u_j(1-u_j)}\left(\frac{v_j}{m-1} - u_j\right)^2\right\}}{[2\pi(m-1)u_j(1-u_j)]} dv_j.$$

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

$$\Gamma_j \leq \int_{\mathbb{R}} \frac{\exp\{-x_j^2\}}{2\pi\sqrt{(m-1)u_j(1-u_j)}} dx_j$$

$$= \frac{1}{\sqrt{2\pi(m-1)u_j(1-u_j)}}.$$

Therefore, $\Gamma_j = O(m^{-1/2})$. This shows that integration leads to a drop in asymptotic magnitude equal to $m^{-1/2}$ for each dimension. Let $\lambda_j \equiv [u_j(1-u_j)]^{1/2}$; then

$$\operatorname{var}(\tilde{c}_B) \lesssim \frac{m^{2k}}{n}\left([4\pi(m-1)]^{k/2}\prod_{j=1}^{k}\lambda_j\right)^{-1}\left(\frac{M}{m^k} + m^{-(k+1)}\right)$$

$$\lesssim \left(n\prod_{j=1}^{k}\lambda_j\right)^{-1} m^{k/2}(1+m^{-1}).$$

At the edges of the hypercube, i.e., $u = 0,1$, $(P_{v_j,m-1}(u))^2 = P_{v_j,m-1}(u)$; then

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

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

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

$$\left(\frac{\partial}{\partial m}\right)MSE \lesssim \left(\frac{\partial}{\partial m}\right)\left(m^{-2} + \frac{m^{k/2}}{n}\right)$$

$$\lesssim \left(-m^{-3} + \frac{km^{(k/2)-1}}{2n}\right)$$

$$= 0,$$

which implies $m^{(k+4)/2} = O(n)$. Similarly, for the density at the boundaries of the k-cube,

$$\begin{aligned}\left(\frac{\partial}{\partial m}\right) MSE &\lesssim \left(\frac{\partial}{\partial m}\right)\left(m^{-2} + \frac{m^k}{n}\right)\\ &\lesssim \left(-m^{-3} + \frac{km^{k-1}}{2n}\right)\\ &= 0,\end{aligned}$$

which implies $m^{k+2} = O(n)$.

The finite-dimensional distributions of the EBC 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 i.i.d. are clearly sufficient for the central limit theorem). But the BC 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 properties just mentioned has finite ε bracketing numbers of order $\exp\{\varepsilon^{-(k/(m-1))}\}$. It follows that their entropy integral with bracketing is finite. This is enough to show (for the i.i.d. case, see Ossiander, 1987; for generalizations, see Pollard, 2001) that the BC density converges to a Gaussian process with continuous sample paths. The $\sqrt{m^{-(k/2)}n}$ term is required for the leading term in the variance expansion to be independent of n; i.e., $m \to \infty$ as $n \to \infty$ (as usual for nonparametric estimators). The same condition applies to the copula because it is m times differentiable together with the same properties of the density. Clearly, the simple root-n standardization is employed in this case (integration absorbs the smoothing parameter). ■

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 $O(m^{-1/2})$. 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; recall the singular integral representation in (4). 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 exactly local so that we just recover the properties of the histogram estimator.

5. CONCLUSION AND SOME FURTHER EXTENSIONS

We studied a new object in multivariate analysis called the Bernstein copula (BC). Furthermore, we showed that, subject to regularity, any copula can be represented (approximated) by some BC. This copula representation should allow us to take advantage of the properties of the copula function whenever multivariate normality is not a good assumption.

We made the procedure operational by providing an empirical estimation procedure with its rates of consistency and an interesting result for the variance. The result for the variance may provide a partial solution to the curse of dimensionality, particularly in semiparametric estimation, i.e., when the marginals are known.

The study of the BC led us to consider many topics all at once. It is clear that much has been omitted from this paper. We did not discuss joint continuity of the BC under the *-product defined by Darsow et al. (1992) (see Kulpa, 1999; Li et al., 1998). The *-product is a powerful tool that allows us to define Markov processes and general time series dependence concepts. The BC is closed under this operation.

The consistency results for the EBC have been derived under the condition of i.i.d. observations. Although the more general case of stationary random variables could be dealt with (under suitable mixing conditions), for the sake of conciseness we refrained from doing so. We restricted m to be the same for all coordinates. This is not necessary, but the notation would have been cumbersome. What seems more relevant in practice is the actual choice of m (possibly different in each coordinate) in empirical work. We did not discuss this issue because it is a common problem for similar (nonparametric) estimators to the EBC. As shown in our example, one may consider some metric and minimize it in terms of m. The results in Table 2 show that the error is not particularly sensitive to the choice of m as opposed to the histogram estimator. In practice we may use cross validation or some modified version of it (the true estimator would be approximated using the jackknife). For more details on this and related limitations of this approach, the reader is referred to Scott (1992). We could have compared our estimator with the Genest and Rivest (1993) nonparametric procedure for selecting the best bivariate archimedean copulas (e.g., for details on archimedean copulas, see Joe, 1997, p. 86). These authors derive an estimator, say, $K_n(v)$, for a function $K(v)$, which is closely related to the generator of an archimedean copula. Therefore, their approach could be used to estimate empirically the generator of a bivariate archimedean copula and from that to obtain an estimator for a bivariate archimedean copula. However, using the words of these authors: their approach "often proves more convenient in application to use . . . as a tool for identifying the parametric family of Archimedean copulas that provides the best possible fit to the data"; further, "one may be tempted to determine directly from $K_n(v)$ the Archimedean copula" (p. 1035). "Though this would be formally possible, whenever $v - K_n(v^-) < 0$ for all $0 < v < 1$, it generally will be theoretically more meaningful—as well as computationally more convenient—to use K_n as a tool to help identify the parametric family of Archimedean copulas that provides the best possible fit to the data" (p. 1038). Indeed, we experienced computational difficulties estimating empirically an archimedean copula using this approach. Although the KS copula used in our simulation as the true model belongs to the class of bivariate archimedean copulas, our estimator is not restricted to this class, has wider application, and is easier to compute.

Finally, alternative estimation procedures have not been considered in detail. Although the paper provides a promising result for the variance of the empirical 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., $\mathcal{D}^1 c_B = \sum_{j=1}^{k} \partial c_B / \partial u_j$, and $c_B = \partial^k C_B / (\partial u_1 \cdots \partial u_k)$. However, unlike the case of the EBC 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 BC. Some of the issues omitted from this paper are the subject of current research.

NOTES

1. Lower and upper tail dependence are, respectively, defined as

$$\lambda_L = \lim_{u \to 0} \Pr(u_1 < u \mid u_2 < u)$$

and

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

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

2. Clearly we could easily impose the boundary condition

$$C_n(1, \ldots, 1, u_j, 1, \ldots, 1) = u_j, \; 1 \le j \le k.$$

However, this will not ensure that C_n is monotonically increasing in a finite sample.

3. The results in Table 2 should be considered purely illustrative. A larger number of Monte Carlo simulations would be required in order to achieve more accurate convergence. Unfortunately, our calculations were labor-intensive, and a larger number of simulations would require more investment in programming to obtain results within a reasonable time. Another problem is caused by the use of Monte Carlo integration in place of integration over the unit square. In this

case, more advanced approaches could have been used, such as quasi-random numbers to bound the error by a quantity $O(\ln(n)^k/n)$ (for $k = 2$ dimensions in our case) instead of the usual $O(1/\sqrt{n})$ for the Monte Carlo integration approach (for further details, see, e.g., Spanier and Maize, 1994).

REFERENCES

Abadir, K.M. (1999) An introduction to hypergeometric functions for economists. *Econometric Reviews* 18, 287–330.

Abadir, K.M. & J.R. Magnus (2002) Notation in econometrics: A proposal for a standard. *Econometrics Journal* 5, 76–90.

Berens, H. & R. DeVore (1980) A characterization of Bernstein polynomials. In E.W. Cheney (ed.), *Approximation Theory*, vol. 3, pp. 213–219. Academic Press.

Bouyé, E., N. Gaussel, & M. Salmon (2001) Investigating Dynamic Dependence Using Copulae. Working paper, Financial Econometrics Research Centre.

Butzer, P.L. (1953) Linear combinations of Bernstein polynomials. *Canadian Journal of Mathematics* 5, 559–567.

Darsow, W.F., B. Nguyen, & E.T. Olsen (1992) Copulas and Markov processes. *Illinois Journal of Mathematics* 36, 600–642.

Devore, R.A. & G.G. Lorentz (1993) *Constructive Approximations*. Springer.

Embrechts, P., A. McNeil, & D. Straumann (1999) Correlation: Pitfalls and alternatives. *Risk* 5, 69–71.

Fortin, I. & C.A. Kuzmics (2002) Tail-dependence in stock-return pairs. *International Journal of Intelligent Systems in Accounting, Finance, and Management* 11, 89–107.

Genest, C. & L.P. Rivest (1993) Statistical inference procedures for bivariate archimedean copulas. *Journal of the American Statistical Association* 88, 1034–1043.

Hu, L. (2002) Dependence Patterns across Financial Markets: Methods and Evidence. Preprint, Department of Economics, Yale University.

Jackson, D. (1930) *The Theory of Approximation*. American Mathematical Society.

Joe, H. (1997) *Multivariate Models and Dependence Concepts*. Chapman and Hall.

Knight, J., S. Satchell, & K. Tran (1995) Statistical modelling of asymmetric risk in asset returns. *Applied Mathematical Finance* 2, 155–172.

Kulpa, T. (1999) On approximation of copulas. *International Journal of Mathematics and Mathematical Science* 22, 259–269.

Laherrère, J. & D. Sornette (1998) Stretched exponential distributions in nature and economy: "Fat tails" with characteristic scales. *European Physical Journal B* 2, 525–539.

Li, H., M. Scarsini, & M. Shaked (1996) Linkages: A tool for the construction of multivariate distributions with given nonoverlapping multivariate marginals. *Journal of Multivariate Analysis* 56, 20–41.

Li, X., P. Mikusiński, & M.D. Taylor (1998) Strong approximation of copulas. *Journal of Mathematical Analysis and Applications* 225, 608–623.

Longin, F. & B. Solnik (2001) Extreme correlation of international equity markets. *Journal of Finance* 56, 651–678.

Lorentz, G.G. (1953) *Bernstein Polynomials*. University of Toronto Press.

Marichev, O.I. (1983) *Handbook of Integral Transforms of Higher Transcendental Functions*. Ellis Horwood.

Nelsen, R.B. (1998) *An Introduction to Copulas*. Lecture Notes in Statistics 139. Springer-Verlag.

Ossiander, M. (1987) A central limit theorem under metric entropy with L_2 bracketing. *Annals of Probability* 15, 897–919.

Patton, A.J. (2001) Modelling Time Varying Exchange Rate Dependence Using the Conditional Copula. Discussion Paper Department of Economics, University of California, San Diego.

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

Phillips, P.C.B. (1983) ERA's: A new approach to small sample theory. *Econometrica* 51, 1505–1525.

Pollard, D. (2001) General Bracketing Inequalities. Preprint, Department of Statistics, Yale University.

Rockinger, M. & E. Jondeau (2001) Conditional Dependency of Financial Series: An Application of Copulas. Preprint.

Sancetta, A. (2003) Nonparametric Estimation of Multivariate Distributions with Given Marginals: L_2 Theory. Cambridge Working Papers in Economics 0320.

Schweizer, B. & E.F. Wolff (1981) On nonparametric measures of dependence for random variables. *Annals of Statistics* 9, 879–885.

Scott, D.W. (1992) *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley.

Silvapulle, P. & C. Granger (2001) Large returns, conditional correlation, and portfolio diversification: A value at risk approach. *Quantitative Finance* 1, 542–551.

Sklar, A. (1973) Random variables, joint distribution functions, and copulas. *Kybernetika* 9, 449–460.

Spanier, J. & E.H. Maize (1994) Quasi-random methods for estimating integrals using relatively small samples. *SIAM Review* 36, 18–44.

Stuart A. & J.K. Ord (1994) *Kendall's Advanced Theory of Statistics*. Edward Arnold.

Van der Vaart, A. & J.A. Wellner (2000) *Weak Convergence of Empirical Processes*. Springer Series in Statistics. Springer.

White, H. (1994) *Estimation, Inference, and Specification Analysis.* Cambridge University Press.

APPENDIX: PROOFS

Proof of Theorem 1. Consider the BC $C_B(u_1,\dots,u_k)$ as an approximation to a copula $C(u_1,\dots,u_k)$, i.e.,

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

Then (5) and (6) are sufficient for $C(v_1/m,\dots,v_k/m)$ to be a copula. Because Bernstein polynomials do not interpolate exactly, (5) and (6) are not necessary for $C_B(u_1,\dots,u_k)$ to be a copula. ■

Proof of Theorem 2.

$$\begin{aligned}(B_m^k f)(\mathbf{x}) - f(\mathbf{x}) &= \sum_{v_1=0}^{m_1}\cdots\sum_{v_k=0}^{m_k} P_{v_1,m_1}(x_1)\cdots P_{v_k,m_k}(x_k) \\ &\quad\times\left[f\left(\frac{v_1}{m_1},\dots,\frac{v_k}{m_k}\right) - f(x_1,\dots,x_k)\right] \\ &= \sum_{v_1=0}^{m_1}\cdots\sum_{v_k=0}^{m_k} P_{v_1,m_1}(x_1)\cdots P_{v_K,m_K}(x_k)\int_{(x_1,\dots,x_k)}^{(v_1/m_1,\dots,v_k/m_k)} \nabla f dr,\end{aligned}$$

$\nabla f \equiv [f'^{1}(s_1,\dots,s_k),\dots,f'^{k}(s_1,\dots,s_k)]$, where $f'^{j}(s_1,\dots,s_k) \equiv \partial f(s_1,\dots,s_k)/\partial s_j$ 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, and so the path of integration is irrelevant.

The preceding 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

$$\int_{(x_1,\ldots,x_k)}^{(v_1/m_1,\ldots,v_k/m_k)} \nabla f dr = \int_{x_1}^{v_1/m_1} f'^{1}(s_1, x_2, x_3, \ldots, x_k)\, ds_1 + \ldots + \int_{x_k}^{v_k/m_k} f'^{k}\left(\frac{v_1}{m_1}, \frac{v_2}{m_2}, \ldots, \frac{v_{k-1}}{n_{k-1}}, s_k\right) ds_k.$$

Now, considering the jth term,

$$\begin{aligned}
&\int_{x_j}^{v_j/m_j} f'^{j}\left(\frac{v_1}{m_1}, \frac{v_2}{m_2}, \ldots, s_j, \ldots, x_k\right) ds_j \\
&= f'^{j}\left(\frac{v_1}{m_1}, \ldots, x_j, x_{j+1}, \ldots, x_k\right)\left(\frac{v_j}{m_j} - x_j\right) \\
&\quad - \int_{x_j}^{v_j/m_j}\left(s_j - \frac{v_j}{m_j}\right) df'^{j}\left(\frac{v_1}{m_1}, \frac{v_2}{m_2}, \ldots, s_j, \ldots, x_k\right). \qquad \textbf{(A.1)}
\end{aligned}$$

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:

$$|f'^{1}(s_1, \ldots, s_j + h_j, \ldots, s_k) - f'^{1}(s_1, \ldots, s_j, \ldots, s_k)| \leq M_j |h_j|.$$

It follows that the last integral in (A.1) does not exceed $M_j \int_{x_j}^{v_j/n_j} (s_j - x_j)\, ds_j = \frac{1}{2} M_j (v_j/m_j - x_j)^2$. Therefore,

$$\begin{aligned}
|(B_m^k f)(\mathbf{x}) - f(\mathbf{x})| &\leq \sum_{v_1=0}^{m_1} \ldots \sum_{v_k=0}^{m_k} P_{v_1,m_1}(x_1) \cdots P_{v_k,m_k}(x_k) \\
&\quad \times \frac{1}{2} \sum_{j=1}^{k} M_j \left(\frac{v_j}{m_j} - x_j\right)^2 \\
&= \frac{1}{2}\left[M_1 \frac{x_1(1-x_1)}{m_1} + \cdots + M_k \frac{x_k(1-x_k)}{m_k}\right]
\end{aligned}$$

for any X, where the first term on the right-hand side of (A.1) is exactly zero when the Bernstein operator is applied to $(v_j/m_j - x_j)$. ■