A simple estimator of two-dimensional copulas, with applications*

Eddie Anderson† **Artem Prokhorov**‡ **Yajing Zhu**§

26 Mar 2020

Abstract

Copulas are distributions with uniform marginals. Nonparametric copula estimates may violate the uniformity condition in finite samples. We look at whether it is possible to obtain valid piecewise linear copula densities by triangulation. The copula property imposes strict constraints on design points, making an equi-spaced grid a natural starting point. However, the mixed-integer nature of the problem makes a pure triangulation approach impractical on fine grids. As an alternative we study the ways of approximating copula densities with triangular functions which guarantees that the estimator is a valid copula density. The family of resulting estimators can be viewed as a non-parametric MLE of B-spline coefficients on possibly non-equally spaced grids under simple linear constraints. As such, it can be easily solved using standard convex optimization tools and allows for a degree of localization. A simulation study shows an attractive performance of the estimator in small samples and compares it with some of the leading alternatives. We demonstrate empirical relevance of our approach using three applications. In the first application we investigate how the body mass index of children depends on that of parents. In the second application, we construct a bivariate copula underlying the Gibson paradox from macroeconomics. In the third application, we show the benefit of using our approach in testing the null of independence against the alternative of an arbitrary dependence pattern.

*The authors acknowledge the Sydney Informatics Hub and the University of Sydneys high performance computing cluster Artemis for providing the high performance computing resources that have contributed to the research results reported within this paper. Research for this paper was supported by a grant from the Russian Science Foundation (Project No. 16-18-10432).

†University of Sydney Business School, Australia, and Imperial College London, UK; email: edward.anderson@sydney.edu.au

‡University of Sydney Business School, Australia, and St.Petersburg State University, Russia; email: artem.prokhorov@sydney.edu.au

§Statistics Canada, Ottawa; email: yajing.s.zhu@gmail.com

1 Introduction

Copulas are multivariate distributions of marginal probability integral transforms. They are widely used for modelling dependence between the marginals and they have found many applications in economics and finance (see, e.g., Fan and Patton, 2014, for a survey). A key feature of copulas is that all marginals of a copula are uniform on $[0, 1]$. This condition is what distinguishes copula density estimation from estimation of general densities on a hypercube.

In the context of nonparametric copula estimation, the uniformity condition is often difficult to impose. It involves integral equations based on the copula estimator, which translate into restrictions on the parameter space and the kernel or basis functions used. For example, the Bernstein copula estimator (see, e.g., Sancetta and Satchell, 2004) can be shown to satisfy these conditions if and only if the matrix of copula parameters is doubly stochastic. The penalized exponential series copula estimator (see, e.g., Gao et al., 2015) satisfies these conditions under a set of non-linear constraints on parameters and basis functions. Due to such difficulties, estimators that do impose the uniformity condition may exhibit computational issues (see, e.g., Qu et al., 2009; Qu and Yin, 2012).

In this paper we propose a class of copula density estimators obtained using B-splines over a possibly sparse grid. The uniform marginals property imposes certain constraints on the density surface that are easiest to handle on an equi-spaced grid using a triangular basis function. We work out the constraints and explore the difficulties arising from a direct application of triangulation in approximating bivariate copula densities with piecewise linear surfaces while guaranteeing the uniform marginal property. The main difficulty is that such copula constructions require a mixed integer optimization which is hard to work with.

We then develop a straightforward spline-based method using a specific basis function, which reduces this problem to a convex non-parametric maximum likelihood estimation, subject to linear equality constraints – an easy problem to handle in most available software packages. The estimator also has well known statistical properties, being a member of a constrained MLE family of estimators (see, e.g., Aitchison and Silvey, 1958). We provide sufficient conditions for the copula property to hold inside grid cells provided it holds at grid knots. We then generalize the spline estimator to higher degree splines and non-equi-spaced grids. The latter contribution is important because it provides a natural but often overlooked way of imposing a finer grid at a corner and sparser elsewhere. That is to say, our estimator has a localization property. We do not pursue uneven grids in the paper but we use B-splines of a higher order in simulations to illustrate the additional computational burden associated with this generalization.

The use of B-splines is not new in statistics literature on copula modelling. Kauerman et al. (2013) propose a spline-based multivariate copula estimator with a focus on dimensionality reduction permitted by a sparsity pattern. Shen et al. (2008) consider linear splines as a means of improving over the empirical copula and Erdely (2016) shows that the linear spline estimators are actually a checkerboard copula of Li et al. (1997). The paper is also related to the literature on methods of constructing bivariate copulas using the copula values at some points in the unit square, such as a copula obtained from a diagonal section (see, e.g., Nelsen et al., 2008), a horizontal section (see, e.g., Klement et al., 2007) and using rectangles (see, e.g., Durante et al., 2009). Just like these other papers, we offer a copula construction method but our focus is on developing a spline-based estimator imposing the uniformity of marginals.

We compare our estimator with a battery of commonly-used copula density estimators, including the empirical beta copula, Bernstein polynomials, exponential series, data-mirror and naive kernel estimators. The list is by no means complete but representative of what is used in practice. Our estimator performs very well overall and, in particular, we show in a simulation study that, for various strengths of dependence and various sample sizes, the estimator is able to capture key features of the true copula no worse than the competitors.

As applications, we provide new insights into several well-studied econometric data sets and into a recent independence test. First, we reconsider the dependence between intergenerational body mass indices and uncover a stronger dependence at the upper end of the partent-child BMI distribution. Second, we use our estimators to nonparametrically model a macroeconomic phenomenon known as Gibson's paradox, which has up to now been modelled using only restrictive parametric distributions. Finally, we look at the power of a copula-based test for arbitrary dependence between two random variables proposed by Belalia et al. (2017). The test power is visibly increased when the copula propertly is imposed and our estimator dominates the Bernstein polynomial estimator initially used by Belalia et al. (2017). The key observation in all our applications is that by imposing the copula property the new estimators provide confidence in dependence-based measures and tests.

The paper is organized as follows. Section 2 discusses the uniform marginals property and what it implies for the construction of a piecewise linear surface and defines the estimator we propose. Section 3 lists selected copula density estimators which serve as benchmarks. Section 4 discusses simulation results. Section 5 discusses the applications to intergenerational body mass index estimation and to Gibson's paradox. Section 6 discusses the application to independence testing. Section 7 concludes.

2 Copula property and proposed estimator

Let H denote an absolutely continuous bivariate distribution function with one-dimensional marginals F_1, F_2. A copula function C: $[0,1]^2 \to [0,1]$ can be obtained from the equation

$$H(x_1, x_2) = C(F_1(x_1), F_2(x_2))$$

by inversion as follows

$$C(\mathbf{u}) = H(F_1^{-1}(u_1), F_2^{-1}(u_2)),$$

where $\mathbf{u} = (u_1, u_2) \in [0,1]^2$ and F_j^{-1} is the generalized inverse of $F_j, j = 1, 2$. The Sklar (1959) theorem states that C is unique for continuous distributions. When it exists, the copula density $c(\mathbf{u})$ is defined as $\frac{\partial^2}{\partial u_1 \partial u_2} C(\mathbf{u})$.

A key property of $c(\mathbf{u})$ is that it has uniform marginals (see, e.g., Ibragimov and Prokhorov, 2017, Definition 1.1). This property can be equivalently stated in terms of the distribution function

$$C(1, u) = C(u, 1) = u, \text{ for all } u \in [0, 1],$$

or in terms of the density function

$$\int c(u_1, u_2) du_1 = \int c(u_1, u_2) du_2 = 1, \text{ for all } u_j \in (0, 1), j = 1, 2.$$

This property is what makes copula estimation different from a generic density estimation on a hypercube. Violations of the copula properties may lead to invalid plug-in estimates of copula-based functionals, e.g., a plug-in estimator of tail-dependence coefficient can be shown to take negative values when the copula condition is violated.

Practical examples of when this is critical include a variety of actuarial and risk management applications. A valid estimate of the tail dependence coefficient is crucial in actuarial applications because higher tail dependence increases the underwriter's risk and leads to higher insurance premiums. For example, Bozic et al. (2014) show how copula-based tail-dependence estimates can be used in an agricultural risk management tool to provide dairy producers with individualized protection against catastrophic financial losses. Zhang et al. (2013) use copula-based tail-dependence estimates to characterize the extreme dependence between the returns of individual assets and the market index.

Additionally, traditional kernel-based nonparametric copula density estimation inherits the well-known issues of nonparametric density estimation. For example, the methods that rely on symmetric kernels are not consistent on the boundaries of $[0, 1]^2$ and the boundary bias is substantial (see, e.g., Gijbels and Mielniczuk, 1990; Omelka et al., 2009). To address such issues, clever techniques and innovative kernels have been proposed, for example, the data-mirror estimator which we will define shortly. However, addressing the boundary issue often comes at the expense of the copula property in that the new kernels and adjustment techniques make it harder to impose it.

The estimator we propose is based on approximating copula densities using spline surfaces. This eliminates the boundary biases associated with traditional nonparametric estimators based on symmetric kernels. Moreover, this allows a degree of localization, where we can either use the sparsity patterns in the data to make the grid finer in areas containing more information, or we can control the boundary bias by making the grid finer near corners, while still preserving the copula property.

The most basic approach would be to take a piecewise linear surface on the unit square, however we need to preserve the copula property. Suppose that $c(u, v)$ is known at points $(u, v) \in \Gamma = \{(u_1, v_1), \ldots, (u_k, v_k)\}$, where this set contains the corners of the unit square and some other points inside it. Then, we can define a piecewise linear surface elsewhere on the unit square through a triangulation from any three points in the set. For example, we can fix a set of triangles $T_1, \ldots, T_k$, where each T_i is the convex hull of three points in Γ and the T_i's do not overlap and cover the whole region.

It turns out that the copula property is quite restrictive for the set of the design points Γ and will imply that for each design point $(u, v) \in \Gamma$, there must be two others to guarantee that the copula property holds, and one of the points must have the same u-coordinate and the other point must have the same v-coordinate. This is explored in more detail in Appendix A. This makes it natural to define Γ as a set of points lying on a grid. The simple version of a grid we start with is an equi-spaced grid on the unit square with grid points

$$\Gamma = \{(i/k, j/k)\}_{i,j\in\{0,1,\ldots,k\}}$$

as the design points.

Now suppose we want to construct by a triangulation a piecewise linear surface on $[0, 1] \times [0, 1]$ with the values of $c(u, v)$ given at each point in Γ. In this case we need to define for each square on the grid which of the two diagonals will be used in the triangulation. In this case it turns out that the copula property will be satisfied provided that they are satisfied on the grid lines and in addition

$$\int_0^1 c\left(\frac{0.5+i}{k}, v\right) dv = 1, i = 0, 1, ..., k-1, \tag{1}$$

$$\int_0^1 c\left(u, \frac{0.5+j}{k}\right) du = 1, j = 0, 1, ..., k-1. \tag{2}$$

This follows because if $c(u, v)$ is piecewise linear then $\int_0^1 c(u, v)dv$ is a quadratic function of u for the range $u \in (i/k, (i+1)/k)$. Hence ensuring that the integral has the same value at the half way point as at the end points will guarantee that the quadratic term vanishes and hence $\int_0^1 c(u, v)dv$ is constant on $(i/k, (i+1)/k)$.

Because of the piecewise linearity of c, each of the conditions (1)-(2) translates into a straightforward equation involving $c(i/k, j/k)$, $j = 0, 1, ...k$, and $c((i+1)/k, j/k)$, $j = 0, 1, ...k$. The exact form of this depends on the direction of the diagonals in the triangulation. Thus we can obtain some simple linear restrictions that the c values must satisfy, a total of $2k + 1$ in both directions giving $4k + 2$ in total.

However the question of which diagonal to use makes this quite hard to work with. For example, we can have lower left triangle and upper right triangle to form a cover of the cell, or have lower right triangle and upper left triangle to form a cover. Such a method of copula construction would therefore require a mixed integer optimization, something we wish to avoid. For that reason we turn next to a more straightforward approach using B-splines.

The alternative we pursue is to follow the logic of the empirical copula based estimators and write a basis expansion of $c(u,v)$ as follows

$$c(u,v) = \sum_{i=0}^{k}\sum_{j=0}^{k} w_{ij} b_{ij}(u,v),$$

where $\{w_{ij}\}_{i,j\in\{0,1,\ldots,k\}}$ are copula density estimates at the equi-spaced grid points Γ, akin to the Bernstein copula weights $\boldsymbol{\omega}$ (to be defined in Section 3), and $b_{ij}(u,v)$, $i,j \in \{0,1,\ldots,k\}$ are certain basis functions. The basis functions are meant to smooth the density estimate in between the grid points and so they have the following properties:

$$b_{ij}(i/k, j/k) = 1,$$
$$b_{ij}(u,v) = 0 \quad \text{if} \quad \left|u - \frac{i}{k}\right| \geq 1/k, \text{ or } \left|v - \frac{j}{k}\right| \geq 1/k.$$

This also allows for the values w_{ij} of the copula density at the grid points Γ.

For reasons that will become clear shortly, the choice of b_{ij} we propose is

$$b_{ij}(u,v) = g(u - i/k)g(v - j/k),$$

where g is the triangular function $g(z) = (1 - k|z|)_+$. Because this estimator uses a B-spline we call it a spline estimator (SE). Even though $g(\cdot)$ is linear, this will not produce a piecewise linear basis in two dimensions.

Proposition 1 *The SE with the basis function*

$$b_{ij}(u,v) = g(u - i/k)g(v - j/k), \text{ where } g(z) = (1 - k|z|)_+,$$

satisfies the copula property at every point $(u,v) \in [0,1]^2$ *provided it holds on the grid knots* $\Gamma = \{(i/k, j/k)\}_{i,j\in\{0,1,\ldots,k\}}$.

Proof. It follows from the definion of SE that if the v component is fixed then $c(u,v)$ will be linear between grid points since it is the sum of the four terms from the adjacent grid knots, and each of these terms is linear in the u component if v is held fixed. Then, for $0 \leq \alpha, \beta \leq 1$, it is easy to see that

$$c((\alpha + i)/k, (\beta + j)/k) = (1-\alpha)c(i/k, (\beta + j)/k) + \alpha c((i+1)/k, (\beta + j)/k)$$

Thus

$$\int_0^1 c((\alpha + i)/k, y)dy = (1-\alpha)\int_0^1 c(i/k, v)dv + \alpha \int_0^1 c((i+1)/k, v)dv,$$

and with a similar argument

$$\int_0^1 c(u, (\beta + j)/k)du = (1-\beta)\int_0^1 c(u, j/k)du + \beta \int_0^1 c(u, (j+1)/k)du.$$

The result follows given that the copula property holds on Γ. ■

We propose a maximum likelihood SE of w_{ij} which imposes the copula property on Γ and hence achieves the copula property at any point within the grid. More precisely, given an iid sample (u_i, v_i), $i = 1, 2, ...n$, the SE of the copula density $\hat{c}(u, v) = \sum_{i=0}^{k} \sum_{j=0}^{k} w_{ij}^* b_{ij}(u, v)$ is obtained using the solution w_{ij}^* to the following optimization problem

$$\max_{w_{ij}, i,j \in \{0,1,...k\}} \sum_{i=1}^{n} \log c(u_i, v_i), \tag{3}$$

subject to the constraints

$$w_{i0} + 2\sum_{j=1}^{k-1} w_{ij} + w_{ik} = 2k, \quad i = 0, 1, ..., k, \tag{4}$$

$$w_{0j} + 2\sum_{i=1}^{k-1} w_{ij} + w_{kj} = 2k, \quad j = 0, 1, ..., k, \tag{5}$$

$$w_{ij} \geq 0,$$

where the constraints (4)-(5) ensure that the copula property holds on the grid points Γ.

The above estimation problem uses a concave objective function and linear equality constraints, so it is easy to solve numerically using most available software packages. In fact the basis function b_{ij} can be interpreted as a tensor product of linear splines, also known as a tensor-product sieve. The low order of the spline contributes to computational efficiency of this estimator.

Next we generalize this estimator to a family of tensor product splines over a sparse grid and to higher degrees. Key to the generalization is that the estimator's property to preserve the uniformity of marginals does not hinge on linearity of $g(\cdot)$ or equally spaced grids. The next proposition shows that a tensor product spline surface generated by two sets of generic univariate B-splines of arbitrary degrees can preserve the uniform marginal property of the bivariate copula under mild assumptions on the knot vectors $\boldsymbol{t}$ and $\boldsymbol{s}$, on the spline and on the coefficients $\{w_{ij}\}$.

Proposition 2 *Let $\{B_{i,d_t,\boldsymbol{t}}\}_{i=1}^{k_t}$ and $\{B_{j,d_s,\boldsymbol{s}}\}_{j=1}^{k_s}$ denote generic B-splines of degrees d_t and d_s, respectively, which are defined on knot vectors $\boldsymbol{t} = (t_i)_{i=1}^{k_t+d_t+1}$ and $\boldsymbol{s} = (s_j)_{j=1}^{k_s+d_s+1}$, respectively. Assume $t_1 = s_1 = 0$ and $t_{k_t+d_t+1} = s_{k_s+d_s+1} = 1$. Assume that the knot vectors $\boldsymbol{s}$ and $\boldsymbol{t}$ are (d_s+1)-regular and (d_t+1)-regular, respectively.*

Then, the tensor product spline surface

$$c(u, v) = \sum_{i=1}^{k_t} \sum_{j=1}^{k_s} w_{ij} \ B_{i,d_t,\boldsymbol{t}}(u) \ B_{j,d_s,\boldsymbol{s}}(v) \tag{6}$$

satisfies the copula property if

$$\sum_{j=1}^{k_s} \frac{s_{j+d_s+1} - s_j}{d_s + 1} w_{ij} = 1, \;\; i = 1, 2, \ldots, k_t, \tag{7}$$

$$\sum_{i=1}^{k_t} \frac{t_{i+d_t+1} - t_i}{d_t + 1} w_{ij} = 1, \;\; j = 1, 2, \ldots, k_s, \tag{8}$$

$$w_{ij} \geq 0, \;\; i = 1, 2, \ldots, k_t, \;\; j = 1, 2, \ldots, k_s.$$

Proof. We are going to prove that $\forall u \in [0, 1)$, $\int_0^1 c(u, v)\mathrm{d}v = 1$. The proof for integration in the other dimension is similar.

$$\begin{aligned}
\int_0^1 c(u, v)\mathrm{d}v &= \int_0^1 \sum_{i=1}^{k_t} \sum_{j=1}^{n_s} w_{ij} \; B_{i,d_t,\boldsymbol{t}}(u) \; B_{j,d_s,\boldsymbol{s}}(v)\mathrm{d}v \\
&= \int_0^1 \sum_{i=1}^{k_t} \; B_{i,d_t,\boldsymbol{t}}(u) \sum_{j=1}^{n_s} w_{ij} \; B_{j,d_s,\boldsymbol{s}}(v)\mathrm{d}v \\
&= \sum_{i=1}^{k_t} \; B_{i,d_t,\boldsymbol{t}}(u) \sum_{j=1}^{k_s} w_{ij} \int_0^1 B_{j,d_s,\boldsymbol{s}}(v)\mathrm{d}v \\
&= \sum_{i=1}^{k_t} \; B_{i,d_t,\boldsymbol{t}}(u) \sum_{j=1}^{k_s} w_{ij} \int_{s_1}^{s_{n_s+d_s+1}} B_{j,d_s,\boldsymbol{s}}(v)\mathrm{d}v \\
&= \sum_{i=1}^{k_t} \; B_{i,d_t,\boldsymbol{t}}(u) \sum_{j=1}^{k_s} w_{ij} \frac{s_{j+d_s+1} - s_j}{d_s + 1} \\
&= \sum_{i=1}^{k_t} \; B_{i,d_t,\boldsymbol{t}}(u) \\
&= 1.
\end{aligned}$$

The last two equations follow from the definition of a $(d + 1)$-regular knot vector and a result on B-spline integration due to Bhatti and Brachen (2006). Both the definition and the result are provided in Appendix B. ∎

As before, the generalized spline estimator has the advantage of satisfying the copula property anywhere inside the grid given that it holds on the grid knots. The latter condition is imposed implicitly via constraints (4)-(5) and (7)-(8). However, the arbitrary spline degree allows us to model a wider class of dependence in finite samples than B-splines of degree one.

An important improvement over the equi-spaced grid offered by Proposition 2 is to permit knot locations to be chosen in an adaptive way, while preserving the copula property. This localization property of our estimator is particulary useful when the copula density shows strong sparsity pattern (that is, strong concentration of data points in certain regions of the support, e.g., at the corners) that are consistent with the copula property and data are not abundant. In such cases, an irregular grid will have more knots in areas with higher density and fewer in sparse areas. This means SE can use local data properties to adapt the degree of the spline flexibility.

For any finite spline degree, SE has a similar drawback to the Bernstein copula density estimator. It cannot model extreme tail behavior in finite samples as it uses a finite degree polynomial approximation on all grid cells including the corners. Nevertheless, it can capture increasing dependence as we move to the corners and it generally performs very well against the competitors, which can be seen in simulations in Section 4.

It is worth discussing the asymptotic and finite sample properties of the proposed estimator. As mentioned in the introduction, it falls within the family of constrained MLE and hence inherits the well-established properties of this family. Results concerning its asymptotic behavior go back at least to Aitchison and Silvey (1958, Section 5). More recently, comprehensive results, both theoretical and simulation-based, have been obtained in the literature on sieve MLE estimation (see, e.g., Chen, 2007; Lu, 2010). For example, in the context of partially linear models, Lu (2010, Section 3) shows consistency and asymptotical normality of the sieve MLE based on a B-spline likelihood subject to linear constraints. It is not hard to see that the proofs go through for our estimator if we simply replace what Lu (2010) denotes by $\psi(\mathbf{Z})$ with $c(\mathbf{u})$ and drop the linear part of the model denoted by $\mathbf{X}'\beta$, provided that standard regularity conditions (smoothness of the true density and a moderate rate of growth of the number of knots as $n \to \infty$) hold (see conditions C1 and C2 of Lu, 2010).

It is well established in the literature that convergence of spline-based estimators is slowed by increased non-smoothness of the true function and by slower convergence (to zero) of the maximum spacing between knots. In our setting, commonly used copulas are known to satisfy strong smoothness properties (see, e.g., Siburg and Stoimenov, 2008) and the number of knots is chosen by cross-validation. Evidence of desirable large- and small-sample behaviour of the spline MLE estimator is provided, for example, by Hua and Huang (2010), Xue and Liang (2010) and Lu (2010, Section 4).

We note that in this context, the iid-ness property can be relaxed to various forms of dependence, e.g. α-mixing. This generalization has been widely used in the literature on splines for time series (see, e.g., Wang and Yang, 2009; Bouezmarni et al., 2010; Wang and Wang, 2015; Shao and Yang, 2017) and applies in our setting.

3 Other available estimators

We now list some of the commonly used copula density estimators which we will later compare to our estimator.

The naive kernel estimator (NKE). This early estimator has its roots in the multivariate density estimation and is based on using a smoothing kernel (see, e.g., Gijbels and Mielniczuk, 1990). The simplest of such methods is to define the copula density at a point (u_1, u_2) as follows

$$c(\mathbf{u}) = \frac{1}{4h^2} \lim_{h \to 0} \mathcal{P}(|U_1 - u_1| \leq h, |U_2 - u_2| \leq h)$$

and to estimate it by dropping the limit and replacing the probability with relative frequency over the small region of width h as follows

$$\hat{c}_h(\mathbf{u}) = \frac{1}{4h^2 n} \#(i \; ; \; |U_{1i} - u_1| \leq h, |U_{2i} - u_2| \leq h),$$

where n is the sample size and $\#(\cdot)$ means the number of observations satisfying the property in $(\cdot)$. A smoothed version of this is to use one of the standard nonparametric kernel functions instead of the indicator function. It is well-known that many such kernels exhibit a large boundary bias as they assign a considerable mass outside the unit square. It is unclear how to impose the copula property on this estimator.

The data-mirror estimator (DME). This estimator is meant to provide an adjustment to the naive kernel estimators that fixes the boundary problem by reflecting each data point around the edges and corners of the unit square (see, e.g., Schuster, 1985) and using the kernel estimator based on the artificial data. Specifically, instead of using the original observations (u_{1i}, u_{2i}), we enlarge the data set by including the images of (u_{1i}, u_{2i}), that is $(\pm u_{1i}, \pm u_{2i})$, $(\pm u_{1i}, 2 - u_{2i})$, $(2 - u_{1i}, \pm u_{2i})$ and $(2 - u_{1i}, 2 - u_{2i})$.

The data-mirror estimator is then given by

$$\begin{aligned}
\hat{c}_h(\mathbf{u}) = & \frac{1}{nh^2} \sum_{i=1}^{n} \left\{ k\left(\frac{u_1 - U_{1i}}{h}\right) k\left(\frac{u_2 - U_{2i}}{h}\right) + k\left(\frac{u_1 + U_{1i}}{h}\right) k\left(\frac{u_2 - U_{2i}}{h}\right) \right. \\
& + k\left(\frac{u_1 - U_{1i}}{h}\right) k\left(\frac{u_2 + U_{2i}}{h}\right) + k\left(\frac{u_1 + U_{1i}}{h}\right) k\left(\frac{u_2 + U_{2i}}{h}\right) \\
& + k\left(\frac{u_1 - U_{1i}}{h}\right) k\left(\frac{u_2 + U_{2i} - 2}{h}\right) + k\left(\frac{u_1 + U_{1i}}{h}\right) k\left(\frac{u_2 + U_{2i} - 2}{h}\right) \\
& + k\left(\frac{u_1 + U_{1i} - 2}{h}\right) k\left(\frac{u_2 - U_{2i}}{h}\right) + k\left(\frac{u_1 + U_{1i} - 2}{h}\right) k\left(\frac{u_2 + U_{2i}}{h}\right) \\
& \left. + k\left(\frac{u_1 + U_{1i} - 2}{h}\right) k\left(\frac{u_2 + U_{2i} - 2}{h}\right) \right\},
\end{aligned}$$

where $k(\cdot)$ can be any symmetric kernel. This is known to correct for the boundary biases but this also leads to a convergence rate of $\mathcal{O}(h)$ near the boundaries, which is slower than the usual rate of $\mathcal{O}(h^2)$ obtained in the interior. Similar to KNE, it is unclear how to impose the copula property.

The penalized exponential series estimator (ESE). This estimator proposed by Chui and Wu (2009) and Gao et al. (2015) is similar to the series density estimator but does not produce non-positive density. The idea is to approximate the log copula density function by a linear combination of basis functions, penalized by a roughness penalty to balance between goodness-of-fit and parsimony. In essence this is a penalized maximum-likelihood estimation (MLE) in which one can use such tools as AIC, BIC and cross-validation to do model selection. However, leave-one-out cross validation is rather expensive for ESE with a large number of basis functions so Gao et al. (2015) worked out an approximate cross-validated log-likelihood which requires obtaining the ESE based on the full sample only once.

Let $\phi_k(u_1, u_2), k = 1 \dots K$ be a series of linearly independent basis functions defined on the unit square and approximate $c(\mathbf{u})$ by

$$\hat{c}(\mathbf{u}) = \frac{exp(g(\mathbf{u}))}{\int exp(g(\mathbf{u}))du_1 du_2},$$

where $g(\mathbf{u}) = a'\phi(\mathbf{u})$ with $a = (a_1, \dots, a_K)'$ and $\phi(\mathbf{u}) = (\phi_1(\mathbf{u}_i), \dots, \phi_K(\mathbf{u}_i))'$. The penalized MLE objective function is given by

$$\mathcal{Q} = \frac{1}{n}\sum_{i=1}^{n} a'\phi(\mathbf{u}_i) - \ln \int exp(g(\mathbf{u}))du_1 du_2 - \frac{\lambda}{2}a'Wa,$$

where W is a positive definite weight matrix and λ is the smoothing parameter. Imposing the copula property on this estimator amounts to a set of non-linear constraints on parameters and basis functions which leads to an intractible estimator.

As discussed by Gao et al. (2015), leave-one-out cross-validation using the above criterion function directly is computationally impractical so we use the following cross-validated log-likelihood approximation, which is a modified version of the approximation proposed by Gao et al. (2015),

$$\mathcal{L}_{-} \approx \mathcal{L} - \frac{1}{n(n-1)}\text{trace}(\Phi \hat{H}^{-1}\Phi') + \frac{1}{n^2(n-1)}(\iota'\Phi)\hat{H}^{-1}(\Phi'\iota),$$

where H denotes the Hessian matrix of $\mathcal{Q}$ and $\mathcal{L}$ denotes the quasi-likelihood function

$$\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n} a'\phi(\mathbf{u}_i) - \ln \int exp(\mathbf{u}_i)du_1 du_2,$$

and Φ is a $n \times K$ matrix with the i-th row being $\big(\phi_1(\mathbf{u}_i), \dots, \phi_K(\mathbf{u}_i)\big)$ and ι is an $n \times 1$ vector with every element equal to unity. The modification reflects what we believe to be a typo in equation (8) of Gao et al. (2015) – we provide details of that modification in Appendix C.

Bernstein polynomial-based estimators. The basic version of the Bernstein copula as developed by Sancetta and Satchell (2004) for iid data and extended by Bouezmarni et al. (2010, 2013) to dependent data and to unbounded densities is defined as follows

$$C_{J,n}(\mathbf{u}) = \sum_{v_1=0}^{J}\sum_{v_2=0}^{J} C_n\left(\frac{v_1}{J}, \frac{v_2}{J}\right)\prod_{d=1}^{2} P_{v_d,J}(u_d), \tag{9}$$

where $C_n(\cdot)$ is the empirical copula and $P_{v_d,J}(u) = \binom{J}{v_d}u^{v_d}(1-u)^{J-v_d}$ is the binomial probability mass function with parameter v_d and J, and J is an integer that serves as a smoothing parameter. We can define the empirical copula as follows (see, e.g., Deheuvels, 1979)

$$C_n(\mathbf{u}) = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{2}\mathbb{I}\{V_{ij} \le u_j\},$$

where $\mathbb{I}(\cdot)$ is an indicator function and $\{(V_{i1}, V_{i2})\}_{i=1}^n$ is an iid sample.

The corresponding (empirical) Bernstein copula density estimator can be written as follows

$$c_{J,n}(\mathbf{u}) = J^2 \sum_{v_1=0}^{J-1} \sum_{v_2=0}^{J-1} \omega_{(v_1,v_2)} \prod_{d=1}^{2} P_{v_d,J-1}(u_d),$$

where $\omega_v = \omega_{(v_1,v_2)} = \frac{1}{n}\sum_{i=1}^n \mathbb{I}\{V_{i,1} \in A_{v_1}\}\mathbb{I}\{V_{i,2} \in A_{v_2}\}$ with $A_{v_d} = \left(\frac{v_d}{J}, \frac{v_d+1}{J}\right]$, $d = 1, 2$. Here, the coefficients ω_v form the copula histogram based on the sample $\{(V_{i,1}, V_{i,2})\}_{i=1}^n$ and so this estimator can be interpreted as a smoothed histogram using the products of binomial probabilities $P_{v_d,J-1}(\cdot)$ as smoothing weights (see, e.g., Burda and Prokhorov, 2014).

An alternative version of this estimator can be obtained by estimating the coefficients $\boldsymbol{\omega} = \left\{\omega_{(v_1,v_2)}\right\}_{v_1,v_2=0}^{J-1}$ through a Bernstein sieve maximum likelihood estimation (SMB) (see, e.g., Panchenko and Prokhorov, 2016). Given a sample $\{(V_{i,1}, V_{i,2})\}_{i=1}^n$ and a fixed number $J \in \mathbb{N}$, the SMB can be written as follows

$$\begin{aligned} &\arg\max_{\boldsymbol{\omega}} \sum_{i=1}^n \log c_J(V_{i,1}, V_{i,2}; \boldsymbol{\omega}), \\ &\text{s.t.} \qquad \omega_{(v_1,v_2)} \geq 0; \quad v_1, v_2 = 0, \ldots, J-1 \\ &\qquad\qquad \text{and other possible constraints on } \boldsymbol{\omega} \end{aligned} \tag{10}$$

where $c_J(\mathbf{u}; \boldsymbol{\omega}) = J^2 \sum_{v_1=0}^{J-1} \sum_{v_2=0}^{J-1} \omega_{(v_1,v_2)} \prod_{d=1}^2 P_{v_d,J-1}(u_d)$, and $\boldsymbol{\omega} = \left\{\omega_{(v_1,v_2)}\right\}_{v_1,v_2=0}^{J-1}$. It is not difficult to see that in theory the proposed B-spline estimator and the Bernstein polynomial estimator are members of the sieve MLE family. In fact, given the grid points A_{v_d}, the Bernstein polynomial estimator can be viewed as a special case of the B-spline estimator, which is different from ours in its basis function and in the way it locates knots.

The additional constraints on $\boldsymbol{\omega}$ usually include $\sum \omega_{(v_1,v_2)} = 1$, which guarantees that $c_J(\cdot, \cdot\,; \boldsymbol{\omega})$ is a genuine joint density function. A further set of constraints are required to show that $c_J(\cdot, \cdot\,; \boldsymbol{\omega})$ is a copula density. In the following proposition we provide a sufficient condition for the SMB to satisfy the copula property. To our knowledge, this is a new result.

Proposition 3 *Given $J \in \mathbb{N}$, the Bernstein copula approximator*

$$c_J(u_1, u_2; \boldsymbol{\omega}) = J^2 \sum_{v_1=0}^{J-1} \sum_{v_2=0}^{J-1} \omega_{(v_1,v_2)} \prod_{d=1}^{2} P_{v_d,J-1}(u_d)$$

satisfies the copula property if

$$\begin{aligned} &\sum_{v_1=0}^{J-1} \omega_{(v_1,v_2)} = \frac{1}{J}, \ \textit{for } v_2 = 0, \ldots, J-1, \\ &\sum_{v_2=0}^{J-1} \omega_{(v_1,v_2)} = \frac{1}{J}, \ \textit{for } v_1 = 0, \ldots, J-1, \\ &\omega_{(v_1,v_2)} \geq 0, \ \textit{for } v_1, v_2 = 0, \ldots, J-1. \end{aligned} \tag{11}$$

Proof. For $\forall u_1 \in [0,1]$,

$$\int_{[0,1]} c_J(u_1,u_2;\boldsymbol{\omega})\ \mathrm{d}u_2 = J^2 \sum_{v_1=0}^{J-1}\sum_{v_2=0}^{J-1} \omega_{(v_1,v_2)} P_{v_1,J-1}(u_1) \int_{[0,1]} P_{v_2,J-1}(u_2)\ \mathrm{d}u_2 \tag{12}$$

$$= J^2 \sum_{v_1=0}^{J-1}\sum_{v_2=0}^{J-1} \omega_{(v_1,v_2)} P_{v_1,J-1}(u_1) \frac{1}{J} \tag{13}$$

$$= \sum_{v_1=0}^{J-1} P_{v_1,J-1}(u_1) \left(J \sum_{v_2=0}^{J-1} \omega_{(v_1,v_2)} \right). \tag{14}$$

If $\sum_{v_2=0}^{J-1} \omega_{(v_1,v_2)} = \frac{1}{J}$, $v_1 = 0,\dots,J-1$, then

$$\int_{[0,1]} c_J(u_1,u_2;\boldsymbol{\omega})\ \mathrm{d}u_2 = 1,$$

because $\sum_{v_1=0}^{J-1} P_{v_1,J-1}(u_1) = 1$, being the sum of binomial probabilities. Similar argument applies to $\int_{[0,1]} c_J(u_1,u_2;\boldsymbol{\omega})\ \mathrm{d}u_1$. Hence, conditions (11) guarantee that the copula property holds. ■

We now briefly reconsider the empirical Bernstein copula density estimator, where $\omega_{(v_1,v_2)} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\{V_{i,1} \in A_{v_1}\}\mathbb{I}\{V_{i,2} \in A_{v_2}\}$ and $A_{v_d} = (\frac{v_d}{J}, \frac{v_d+1}{J}]$, $d = 1,2$. Let $V_{i,1} = F_{1;n}(X_{1i})$ and $V_{i,2} = F_{2;n}(X_{2i})$ be the empirical cumulative distribution function for X_1 and X_2, respectively. If $X_{21},\dots,X_{2n}$ have no ties, then there are no equal ranks and $(V_{1,2},\dots,V_{n,2})$ will be a permutation of $(\frac{1}{n},\dots,\frac{n}{n})$. Then,

$$\sum_{v_1=0}^{J-1} \omega_{(v_1,v_2)} = \sum_{v_1=0}^{J-1} \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\{V_{i,1} \in A_{v_1}\}\mathbb{I}\{V_{i,2} \in A_{v_2}\} \tag{15}$$

$$= \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\{V_{i,2} \in A_{v_2}\} \tag{16}$$

which is either $\frac{1}{n}\lfloor \frac{n}{J} \rfloor$ or $\frac{1}{n}\lceil \frac{n}{J} \rceil$, where $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ denote the nearest smaller and greater integers, respectively.

Therefore the empirical Bernstein copula density estimator satisfies the copula property if J is a factor of n and there are no equal rankings in either dimension. It is worth noting, however, that the empirical copula can never be a genuine copula function, because, as a step function, it cannot satisfy the copula property $C(1,u) = C(u,1) = u$ for all $u \in [0,1]$.

The empirical beta copula (EBC). This estimator proposed by Segers et al. (2017) has been shown to satisfy the copula property. It is based on the ranks $R_{i,j}^{(n)}$ of $X_{i,j}$ among $X_{1,j},\dots,X_{n,j}$, $j = 1,2$, defined as follows

$$R_{i,j}^{(n)} = \sum_{k=1}^{n} \mathbb{I}\{X_{k,j} \le X_{i,j}\}.$$

The EBC estimator is defined by

$$\mathbb{C}_n^{\beta}(\mathbf{u}) = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{2} F_{n,R_{i,j}^{(n)}}(u_j),$$

where $F_{n,r}(u)$ is the beta distribution $\mathcal{B}(r, n+1-r)$, $u \in [0,1]$ and $r \in \{1,\dots,n\}$.

The corresponding EBC density estimatior can be written as follows

$$\mathbb{c}_n^{\beta}(\mathbf{u}) = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{2} f_{n,R_{i,j}^{(n)}}(u_j),$$

where $f_{n,r}(u)$ is the beta density function, $u \in [0,1]$ and $r \in \{1,\ldots,n\}$. In the presence of ties, the condition for $\mathbb{c}_n^{\beta}$ to be a copula density is to asign the ranks $R_{i,j}^{(n)}$ randomly to the ties.

4 Simulation study

In order to compare the proposed spline estimator with the competitors described in Section 3, we conduct a series of Monte Carlo simulations. We set the number of observations at $n \in \{100,\ 150\}$ and $\tau \in \{0.25,\ 0.50,\ 0.75\}$, and consider four commonly used bivariate parametric copulas: Gaussian, Frank, Clayton and Gumbel. We also examine the performance of selected estimators at the corners.

One practical problem is the choice of the grid size k which serves as the tuning parameter for the spline estimator (it is also denoted by J for the Bernstein sieve estimator). We use AIC, BIC and leave-one-out cross-validation (CV) for model selection. However, the theoretical implications of using these techniques in our setting are not well explored.

We compare the performance of the spline estimators (SE) in various cases to the data mirror estimator (DME), the naive kernel estimator (NKE), the penalized exponential estimator (PESE), the sieve MLE with Bernstein polynomials without copula density constraint (SMB), the sieve MLE with Bernstein polynomials with copula density constraint (SMB$_{CDC}$) and the empirical beta copula density estimator (EBCE).

We let the grid parameter k for the spline estimator range from 1 to $9-d$, where the degree of the B-splines is $d = \{1, 2\}$, in order to make sure that the sample size is greater than the number of parameters to estimate. The number of estimates to obtain for each k is $(k+d)^2$. The grid parameter J for the sieve MLE with Bernstein polynomials ranges from 1 to 9, to make sure that the sample size is greater than the number of parameters to estimate. We use AIC and BIC to choose J for these estimators.

For PESE, the truncated power series we use is given by

$$\phi(x) = [1, x, x^2, \ldots, x^r, (x - x_1^*)_+^r, \ldots, (x - x_k^*)_+^r],$$

where $(x)_+ = \max(0, x)$ and $x_1^*, \ldots, x_k^*$ are the knots of the spline basis functions. This truncated power series performs relatively well in Gao et al. (2015). We set $r = 2, k = 2$ with $x_1^* = 1/3$ and $x_2^* = 2/3$. The tensor product contains a total of 24 basis functions, which implies that 24 parameters are to be estimated for each smoothing parameter λ. In the simulation, we pick among three values of $\lambda = \{2, 5, 10\}$. For DME and NKE estimators we use CV for bandwidth selection.

Table 1 contains the simulation results for Kendall's $\tau = 0.5$. Every entry is based on averages of the mean squared error (MSE) or mean absolute deviation (MAD) over 100 repetitions of density estimation, evaluated on a 29-by-29 equally spaced grid on the unit square $(0,1)^2$. The results in Table 1 suggest that all spline estimators perform better in terms of MSE and MAD than the kernel estimators, and all spline estimators perform better in terms of computational time compared to PESE. $\mathrm{SE}_{d=1}^{AIC}$ performs better in most cases than SMB^{AIC} both in accuracy and time. $\mathrm{SE}_{d=2}^{AIC}$ dominates PESE both in accuracy and time. $\mathrm{SE}_{d=2}^{AIC}$ performs better that SMB^{AIC} in terms of MSE.

Figures 1-2 visualize copula density estimators for the case where the sample is generated from the Gaussian copula. Figure 1 plots the copula density estimates interpolated over a 29-by-29 equally spaced grid using a single replication, while Figure 2 is based on averages over 10 simulations. It is clear that the spline estimators better capture the key features of the true copula than the other estimators. Interestingly, EBCE and NKE are noticeably undersmoothed – an observation for which we do not have a theoretical explanation.

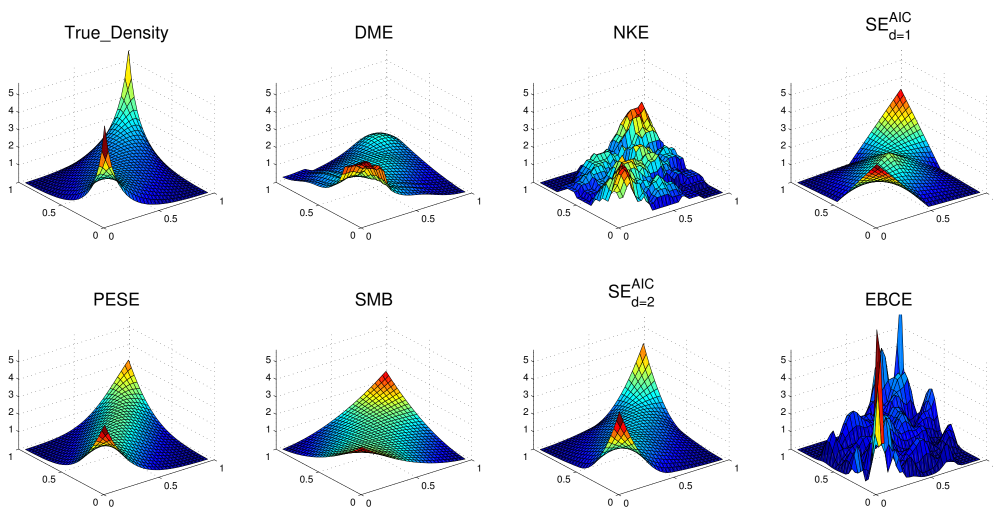


Figure 1: Plots of copula density estimates over a 29-by-29 equi-spaced grid, single simulation; Gaussian copula with Kendall's $\tau = 0.5$.

Table 1: MSE (and MAD) over a 29-by-29 equi-spaced grid; Kendall's $\tau = 0.5$.

True Copula	DME^{CV}	NKE^{CV}	PESE^{CV}	EBCE	SMB^{AIC}	SMB^{BIC}	SMB^{AIC}_{CDC}	SMB^{BIC}_{CDC}	$\text{SE}^{AIC}_{d=1}$	$\text{SE}^{BIC}_{d=1}$	$\text{SE}^{CV}_{d=1}$	$\text{SE}^{AIC}_{d=2}$	$\text{SE}^{BIC}_{d=2}$
						$n = 100$							
Frank	0.323	0.641	0.072	0.383	0.091	0.103	0.089	0.105	0.031	0.105	0.134	0.016	0.087
	(0.426)	(0.615)	(0.158)	(0.441)	(0.160)	(0.167)	(0.156)	(0.175)	(0.137)	(0.416)	(0.153)	(0.107)	(0.160)
Clayton	0.308	0.713	0.330	0.431	0.384	0.391	0.379	0.386	0.374	0.445	0.246	0.234	0.256
	(0.336)	(0.572)	(0.281)	(0.425)	(0.272)	(0.279)	(0.256)	(0.267)	(0.241)	(0.302)	(0.234)	(0.185)	(0.197)
Gumbel	0.302	0.921	0.114	0.428	0.156	0.149	0.147	0.139	0.200	0.175	0.127	0.110	0.123
	(0.358)	(1.145)	(0.172)	(0.436)	(0.244)	(0.252)	(0.178)	(0.235)	(0.201)	(0.187)	(0.134)	(0.154)	(0.185)
Gaussian	0.580	0.627	0.105	0.378	0.074	0.112	0.072	0.102	0.072	0.238	0.106	0.061	0.084
	(0.397)	(0.726)	(0.114)	(0.433)	(0.101)	(0.129)	(0.100)	(0.121)	(0.166)	(0.372)	(0.104)	(0.146)	(0.180)
Time	1.3s	3.3s	615.5s	0.01s	1.7s	1.8s	1.3s	1.3s	1.6s	1.6s	168.1s	71.0s	71.0s
						$n = 150$							
Frank	0.323	0.641	0.072	0.383	0.091	0.103	0.089	0.092	0.031	0.105	0.134	0.016	0.087
	(0.426)	(0.615)	(0.158)	(0.441)	(0.160)	(0.167)	(0.156)	(0.160)	(0.137)	(0.416)	(0.153)	(0.107)	(0.160)
Clayton	0.308	0.713	0.330	0.431	0.384	0.391	0.379	0.386	0.374	0.445	0.246	0.234	0.230
	(0.336)	(0.572)	(0.281)	(0.425)	(0.272)	(0.279)	(0.256)	(0.267)	(0.241)	(0.302)	(0.234)	(0.185)	(0.179)
Gumbel	0.302	0.921	0.114	0.428	0.156	0.149	0.147	0.139	0.200	0.175	0.127	0.110	0.102
	(0.358)	(1.145)	(0.172)	(0.436)	(0.244)	(0.240)	(0.178)	(0.235)	(0.201)	(0.187)	(0.134)	(0.154)	(0.166)
Gaussian	0.580	0.627	0.105	0.378	0.074	0.112	0.072	0.102	0.072	0.238	0.106	0.061	0.084
	(0.397)	(0.726)	(0.114)	(0.433)	(0.101)	(0.129)	(0.100)	(0.121)	(0.166)	(0.372)	(0.104)	(0.146)	(0.180)
Time	1.8s	3.7s	619.8s	0.02s	1.9s	1.9s	1.5s	1.5s	1.6s	1.6s	210.7s	69.6s	70.4s

All values are averaged over 100 simulations. MAD is given in parentheses.

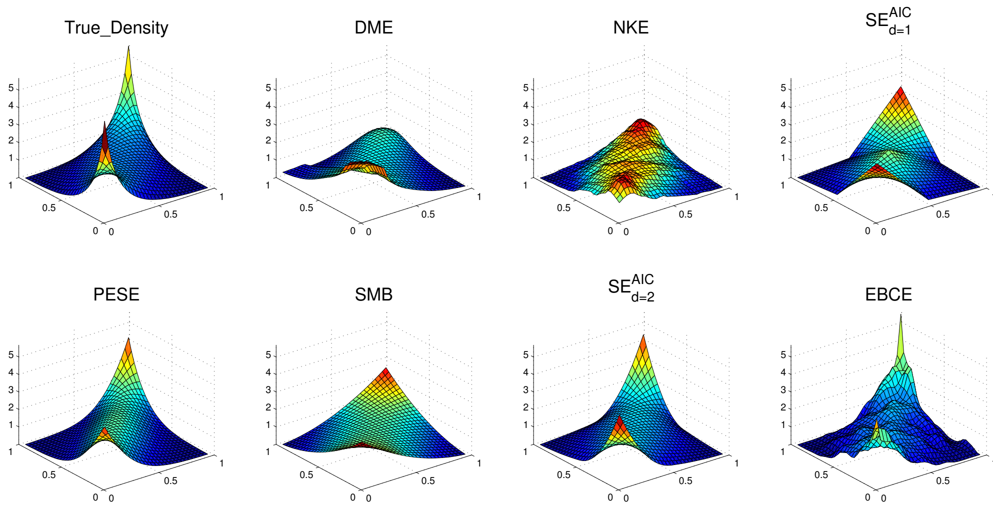


Figure 2: Plots of copula density estimates over a 29-by-29 equi-spaced grid, averaged over 10 simulations; Gaussian copula with Kendall's $\tau = 0.5$.

Table 2 and 3 contain the simulation results for Kendall's $\tau = 0.75$ and $\tau = 0.25$, respectively. Every entry is based on an average of MSE and MAD of 100 estimated densities, evaluated on a 29-by-29 equally spaced grid on $[0,1]^2$. Similar conclusions as in Table 1 seem appropriate for the case of both stronger and weaker dependence.

In order to investigate the behaviour of the estimators near the boundary we follow Bouezmarni et al. (2013) and focus on upper right and lower left corners of the square. Specifically, we define two regions S_1 and S_2 of $[0,1]^2$ as follows

$$S_1 = S \cap \{(u_{1i}, u_{2i}) : \sqrt{u_{1i}^2 + u_{2i}^2} < 0.56\},$$

$$S_2 = S \cap \{(u_{1i}, u_{2i}) : \sqrt{u_{1i}^2 + u_{2i}^2} > 0.98\},$$

where $S = \{(u_{1i}, u_{2i}) = (0.01, 0.01), (0.01, 0.03), \ldots, (0.99, 0.99)\}$. As discussed by Bouezmarni et al. (2013), this restricts the attention to the 25% extreme bottom left points and 25% extreme upper right points.

Table 4 reports the results for selected true copulas with $\tau = 0.25$ and $n = 100$. The selected copulas are known to be unbounded at one or both corners. We can see that the proposed estimator dominates all the alternatives, except perhaps the Bernstein polynomial estimator under the copula constraint.

Table 2: MSE (and MAD) over a 29-by-29 equi-spaced grid; Kendall's $\tau = 0.75$.

True Copula	DME^{CV}	NKE^{CV}	PESE^{CV}	EBCE	SMB^{AIC}	SMB^{BIC}	SMB^{AIC}_{CDC}	SMB^{BIC}_{CDC}	$\text{SE}^{AIC}_{d=1}$	$\text{SE}^{BIC}_{d=1}$	$\text{SE}^{CV}_{d=1}$	$\text{SE}^{AIC}_{d=2}$	$\text{SE}^{BIC}_{d=2}$
							$n = 100$						
Frank	0.391	2.143	0.257	0.378	0.909	1.094	0.918	1.11	0.578	0.602	0.514	0.25	0.67
	(0.343)	(0.794)	(0.336)	(0.332)	(0.742)	(0.827)	(0.749)	(0.834)	(0.642)	(0.68)	(0.606)	(0.345)	(0.632)
Clayton	3.259	5.053	3.467	2.104	4.26	4.506	4.276	4.527	3.698	3.864	3.431	3.459	3.568
	(0.663)	(1.113)	(0.516)	(0.448)	(0.856)	(0.938)	(0.866)	(0.949)	(0.608)	(0.692)	(0.606)	(0.615)	(0.763)
Gumbel	1.387	3.509	1.089	0.662	1.905	2.121	1.913	2.135	1.361	1.349	1.052	1.073	1.45
	(0.521)	(1.168)	(0.374)	(0.368)	(0.788)	(0.869)	(0.788)	(0.869)	(0.499)	(0.519)	(0.455)	(0.414)	(0.608)
Gaussian	0.772	2.789	0.455	0.297	1.134	1.34	1.138	1.358	0.632	0.724	0.304	0.374	0.743
	(0.447)	(1.051)	(0.307)	(0.294)	(0.749)	(0.837)	(0.749)	(0.838)	(0.417)	(0.743)	(0.287)	(0.346)	(0.628)
Time	1.3s	3.3s	649.7s	0.02s	1.4s	1.4s	0.8s	0.8s	1.3s	1.3s	200.4s	45.3s	45.4s
							$n = 150$						
Frank	0.373	1.538	0.251	0.372	0.758	1.092	0.772	1.01	0.229	0.386	0.213	0.224	0.486
	(0.338)	(0.776)	(0.327)	(0.326)	(0.669)	(0.825)	(0.679)	(0.814)	(0.306)	(0.43)	(0.25)	(0.313)	(0.498)
Clayton	3.166	4.389	3.443	2.01	4.079	4.44	4.103	4.464	3.128	3.765	2.573	3.032	3.525
	(0.651)	(1.033)	(0.496)	(0.417)	(0.8)	(0.921)	(0.807)	(0.929)	(0.481)	(0.638)	(0.426)	(0.481)	(0.652)
Gumbel	1.109	2.867	1.07	0.655	1.73	1.912	1.736	1.935	1.108	1.338	0.761	0.927	1.351
	(0.49)	(1.074)	(0.369)	(0.362)	(0.718)	(0.869)	(0.719)	(0.87)	(0.394)	(0.492)	(0.351)	(0.33)	(0.502)
Gaussian	0.559	1.379	0.435	0.27	0.952	1.339	0.97	1.352	0.369	0.596	0.225	0.248	0.506
	(0.433)	(0.754)	(0.303)	(0.287)	(0.671)	(0.827)	(0.673)	(0.827)	(0.298)	(0.429)	(0.26)	(0.244)	(0.408)
Time	2.2s	3.8s	641.7s	0.02s	1.6s	1.6s	0.8s	0.8s	1.3s	1.1s	211.0s	45.0s	45.6s

All values are averaged over 100 simulations. MAD is given in parentheses.

Table 3: MSE (and MAD) over a 29-by-29 equi-spaced grid; Kendall's $\tau = 0.25$.

True Copula	DME^{CV}	NKE^{CV}	PESE^{CV}	EBCE	SMB^{AIC}	SMB^{BIC}	SMB^{AIC}_{CDC}	SMB^{BIC}_{CDC}	$\text{SE}^{AIC}_{d=1}$	$\text{SE}^{BIC}_{d=1}$	$\text{SE}^{CV}_{d=1}$	$\text{SE}^{AIC}_{d=2}$	$\text{SE}^{BIC}_{d=2}$
						$n = 100$							
Frank	0.154	1.003	0.043	0.446	0.05	0.096	0.055	0.098	0.021	0.052	0.03	0.033	0.042
	(0.291)	(0.637)	(0.145)	(0.512)	(0.15)	(0.209)	(0.149)	(0.208)	(0.276)	(0.132)	(0.138)	(0.125)	(0.154)
Clayton	0.18	0.167	0.061	0.404	0.089	0.12	0.091	0.119	0.154	0.156	0.114	0.041	0.075
	(0.31)	(0.291)	(0.188)	(0.48)	(0.166)	(0.214)	(0.167)	(0.214)	(0.262)	(0.163)	(0.172)	(0.13)	(0.178)
Gumbel	0.138	1.883	0.085	0.471	0.04	0.077	0.041	0.076	0.139	0.127	0.098	0.041	0.05
	(0.24)	(0.719)	(0.168)	(0.534)	(0.105)	(0.155)	(0.109)	(0.155)	(0.256)	(0.241)	(0.21)	(0.153)	(0.163)
Gaussian	0.246	0.104	0.14	0.397	0.021	0.042	0.021	0.041	0.11	0.132	0.098	0.04	0.067
	(0.4)	(0.232)	(0.456)	(0.488)	(0.084)	(0.122)	(0.08)	(0.121)	(0.233)	(0.218)	(0.164)	(0.149)	(0.152)
Time	1.2s	3.1s	223.2s	0.01s	2.0s	2.1s	1.8s	1.8s	2.0s	2.1s	210.6s	69.4s	69.3s
						$n = 150$							
Frank	0.113	0.104	0.038	0.462	0.017	0.02	0.014	0.016	0.017	0.041	0.025	0.027	0.031
	(0.23)	(0.217)	(0.143)	(0.514)	(0.095)	(0.102)	(0.101)	(0.105)	(0.272)	(0.123)	(0.135)	(0.121)	(0.132)
Clayton	0.143	0.132	0.057	0.491	0.084	0.095	0.087	0.094	0.154	0.062	0.075	0.04	0.053
	(0.258)	(0.215)	(0.165)	(0.52)	(0.155)	(0.167)	(0.162)	(0.165)	(0.261)	(0.152)	(0.165)	(0.128)	(0.159)
Gumbel	0.131	0.215	0.06	0.43	0.041	0.07	0.04	0.069	0.137	0.043	0.052	0.023	0.029
	(0.231)	(0.327)	(0.149)	(0.511)	(0.11)	(0.154)	(0.104)	(0.153)	(0.251)	(0.145)	(0.15)	(0.115)	(0.122)
Gaussian	0.104	0.133	0.173	0.43	0.016	0.027	0.019	0.031	0.087	0.101	0.085	0.031	0.057
	(0.202)	(0.25)	(0.55)	(0.495)	(0.074)	(0.081)	(0.071)	(0.08)	(0.103)	(0.121)	(0.11)	(0.129)	(0.134)
Time	1.8s	3.8s	235.2s	0.02s	2.4s	2.3s	2.1s	2.2s	2.0s	1.9s	207.1s	66.2s	69.4s

All values are averaged over 100 simulations. MAD is given in parentheses.

Table 4: MSE (and MAE) near boundary; Kendall's $\tau = 0.25$, $n = 100$.

True Copula	DME^{CV}	NKE^{CV}	PESE^{CV}	EBCE	SMB^{AIC}_{CDC}	SMB^{BIC}_{CDC}	$\text{SE}^{AIC}_{d=2}$	$\text{SE}^{BIC}_{d=2}$
Clayton S_1	0.580	0.532	0.315	0.691	0.408	0.446	0.200	0.223
	(0.542)	(0.341)	(0.245)	(0.596)	(0.270)	(0.291)	(0.196)	(0.241)
Clayton S_2	0.187	0.209	0.033	0.778	0.009	0.027	0.096	0.117
	(0.363)	(0.365)	(0.133)	(0.662)	(0.074)	(0.117)	(0.212)	(0.224)
Gumbel S_1	0.251	0.234	0.071	0.690	0.049	0.069	0.069	0.79
	(0.422)	(0.374)	(0.212)	(0.619)	(0.138)	(0.173)	(0.189)	(0.210)
Gumbel S_2	0.628	0.480	0.324	0.758	0.412	0.459	0.179	0.198
	(0.426)	(0.356)	(0.255)	(0.611)	(0.340)	(0.374)	(0.223)	(0.255)
Gaussian S_1	0.218	0.243	0.105	0.682	0.087	0.112	0.086	0.109
	(0.356)	(0.338)	(0.200)	(0.673)	(0.183)	(0.192)	(0.199)	(0.221)
Gaussian S_2	0.272	0.266	0.074	0.662	0.043	0.146	0.042	0.046
	(0.354)	(0.366)	(0.178)	(0.647)	(0.124)	(0.214)	(0.154)	(0.161)

5 Applications to intergenerational BMI dependence and Gibson's Paradox

5.1 Intergenerational BMI dependence

In this section we investigate the intergenerational dependence of Body Mass Index (BMI) between children and parents. The dataset is part of the 2003 Community Tracking Study (CTS) Household Survey, which is the same data set used by Gao et al. (2015). We are interested in households with adult children (18-30) living with both parents. The sample consists of 691 female and 715 male adult children.

Table 5 reports summary statistics for the sample. It can be observed that the male children have higher average BMI than female and that the intergenerational dependence of BMI is stronger between female children and parents, and between mothers and children.

We report our copula estimates in Figures 3-8. The first column of the plots correspond to the spline estimator with tensor product B-splines of degree one. The number of grid points $k = 1, 2, 3, \ldots, 8, 9$; AIC is used for model selection and usually $k = 2$ or $k = 3$ is selected. The second column is for the spline estimator with quadratic B-splines. The third column is for the PESE, where we choose the penalty parameter from $\{2, 5, 10\}$. The last two columns are for the SMB with no copula property and under the copula density condition. In both columns four and five, the number of grid points $k = 1, 2, 3, \ldots, 8, 9$ is chosen using AIC.

In each column we present the interpolated copula density surface and its contour plot. In addition we

Table 5: BMI summary statistics (standard deviations in parentheses)

		Male	Female
Child	BMI	24.92	23.60
		(4.59)	(4.91)
Father	BMI	28.22	28.25
		(4.41)	(4.30)
Mother	BMI	26.79	27.00
		(5.30)	(5.55)
Correlation	Father	0.24	0.31
	Mother	0.29	0.36
Kendall's τ	Father	0.16	0.20
	Mother	0.18	0.23

report each of the two marginal densities implied by the estimator. If the third and fourth plots in each column show a horizontal line at one, this means the uniformity of the marginals is maintained. The extent to which the copula property is violated by PESE and SMB can be seen from the deviation of the marginal densities from uniform.

We note that SMB^{AIC}_{CDC} and $\text{SE}^{AIC}_{d=1}$ produce very similar fits, which is not surpising given that both impose the copula property and both are members of the same family of estimators. Interestingly, $\text{SE}^{AIC}_{d=2}$ is visibly different from other estimators and is more flexible. We also note that because we follow the literature and use ranks in all our estimations, rather than using estimates of parametric cdfs, the effect of using the constraints on SMB is not very prominent.

We follow Gao et al. (2015) and report the results by parent and child gender. For example, Figure 3 shows the results for fathers and sons. The four copula density estimators completely capture the dependence structures between generations. All four estimators clearly suggest a positive and asymmetric dependence structure, with strong dependence at the high end of the BMI distribution. In addition, the $\text{SE}^{AIC}_{d=2}$ seems to show stronger dependence at the high end while the other competitors seem to be over-smoothing. We observe similar results in Figures 4-8.

On the other hand, Figures 3-8 demonstrate that the dependence differs across children's and parents' gender, which is consistent with the results from the intergenerational health transfer literature. In addition, all four estimates in all the figures show stronger dependence at the higher end of BMI than at the lower end. The degree of asymmetry is especially pronounced in the tails and if one of the generations is female.

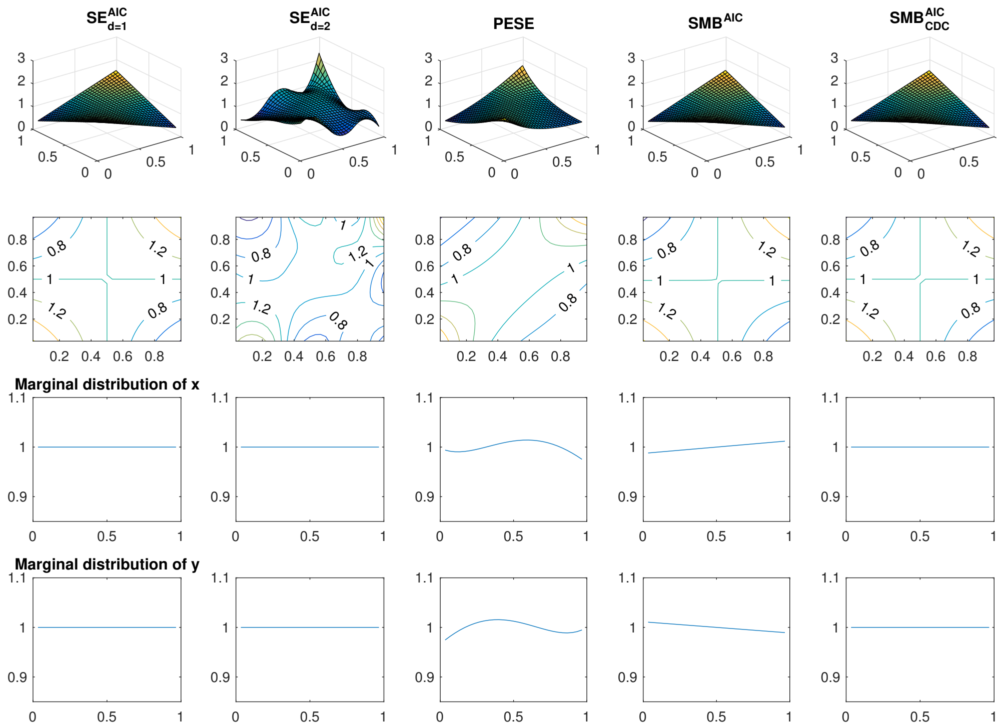


Figure 3: BMI copula density between son and dad

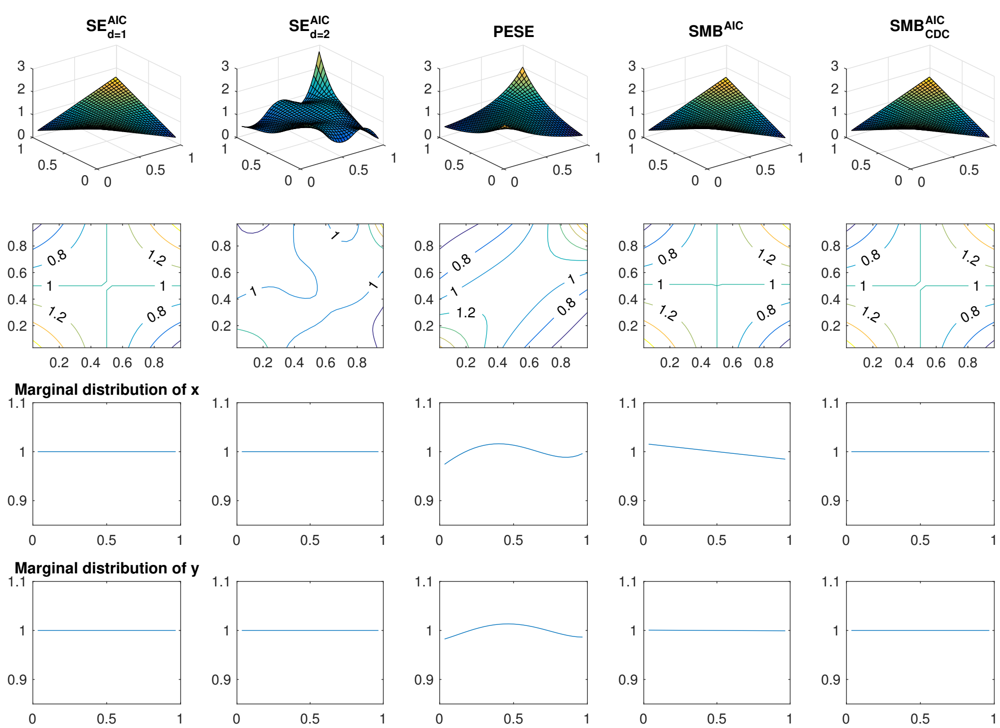


Figure 4: BMI copula density between son and mom

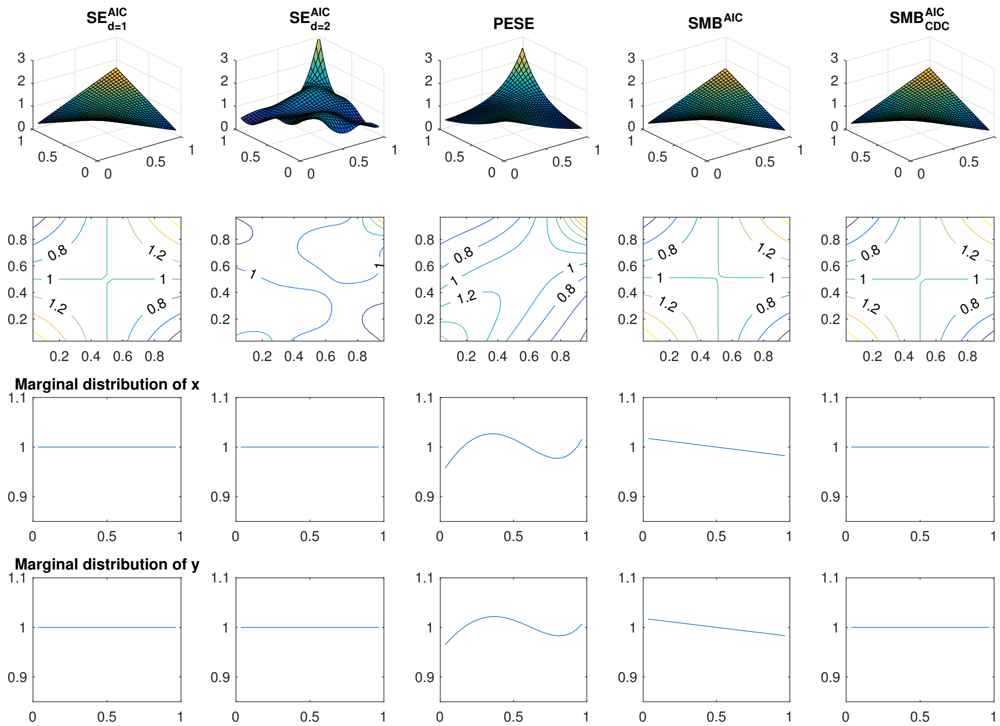


Figure 5: BMI copula density between daughter and dad

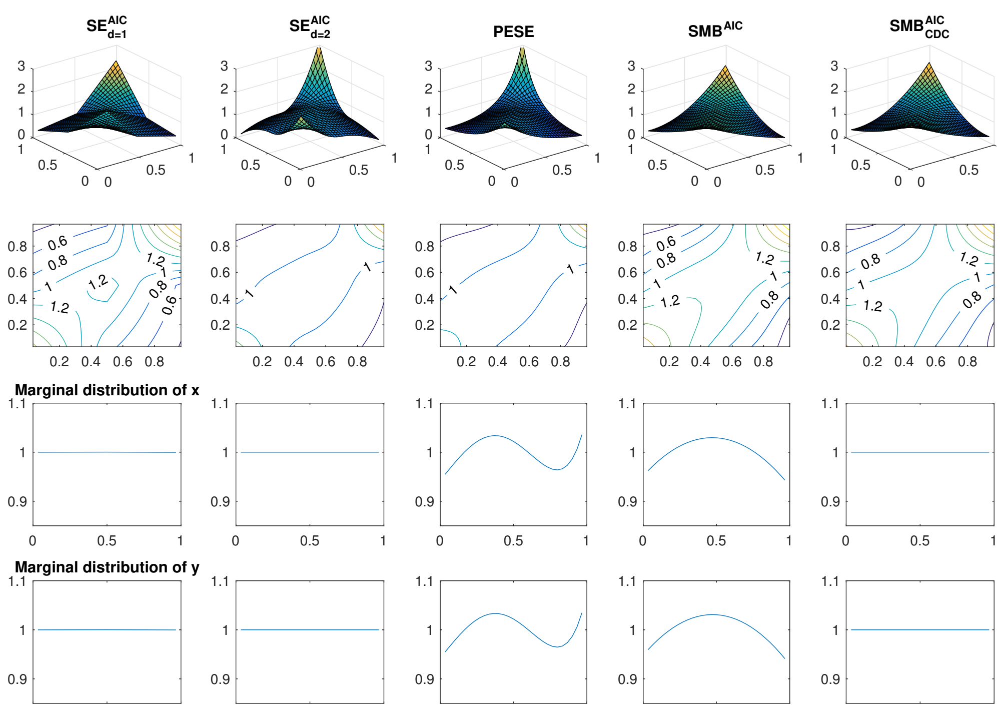


Figure 6: BMI copula density between daughter and mom

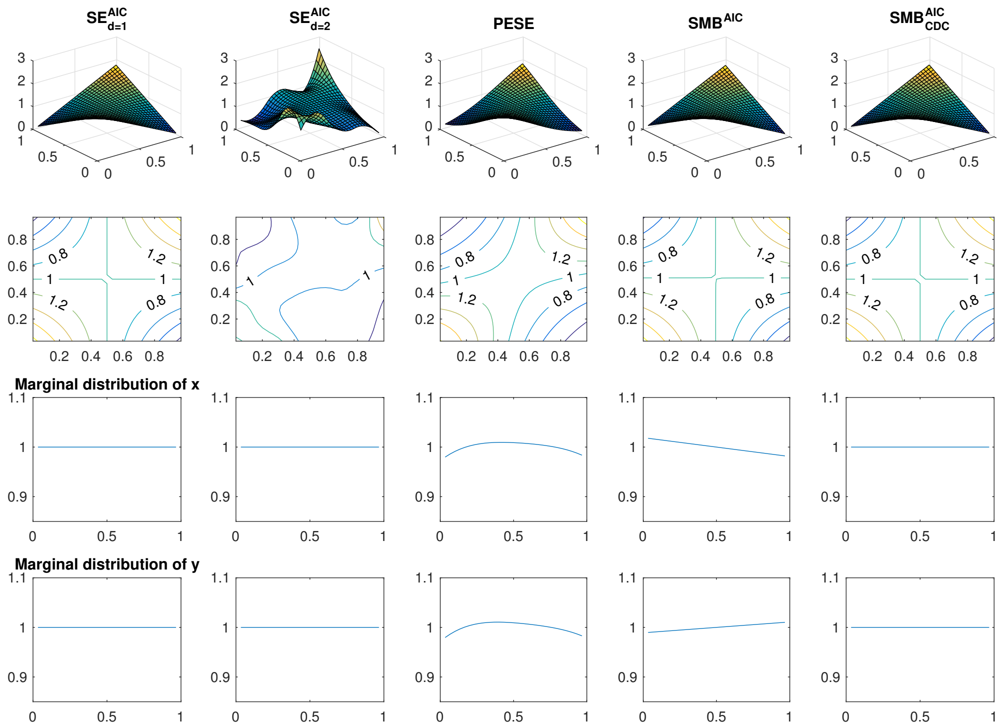


Figure 7: BMI copula density between son and mean (mom, dad)

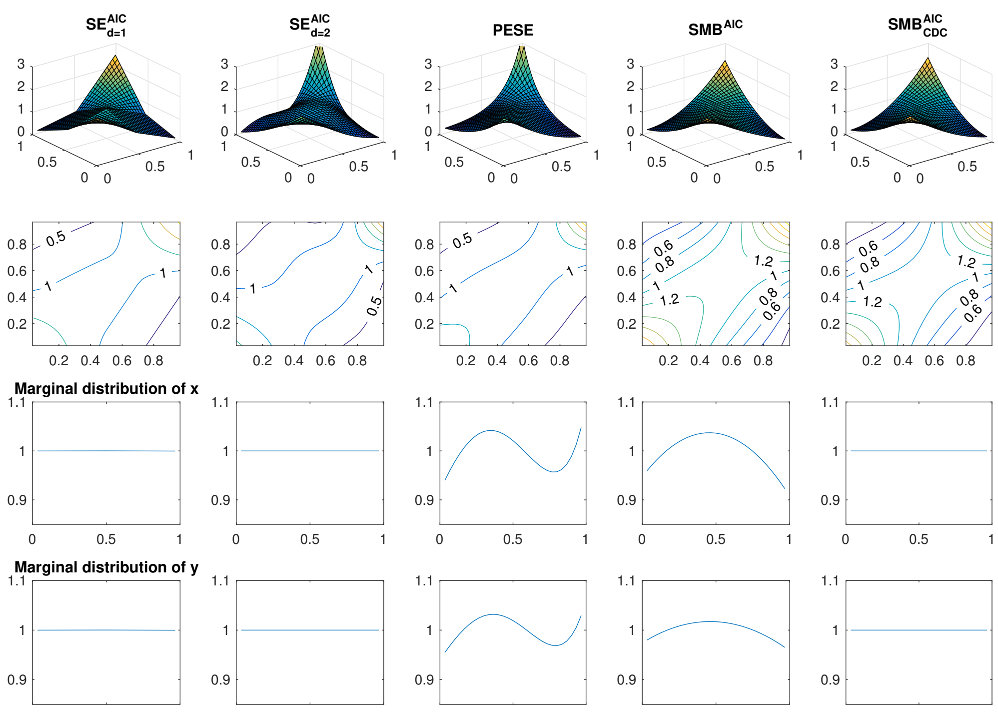


Figure 8: BMI copula density between daughter and mean (mom, dad)

5.2 Gibson's Paradox

Dependence is at the center of Gibson's paradox, which questions, whether prices represented by a general price index, and interest rates are directly or inversely related. Empirically they showed a positive correlation noted by Keynes as early as 1930s, while the quantitative theory of money suggested a negative correlation. Dowd (2008) argues that linear correlation coefficients are too restrictive and links prices and interest rates using parametric copulas. This section uses the same data set and provides four non-parametric estimators, $\mathrm{SE}^{AIC}_{d=1,2}$, PESE and SMB^{AIC}.

The data set consists of annual prices and interest rates in UK during 1821–1913. The price level is represented by the cost of living in the UK (see, e.g., Crafts and Mills, 1994, pp. 180–182). The interest rate is represented by a series combining the annual average yield on 3% consols for the period 1821–1849 (Homer, 1963, Table 19), and the annual consol yield series for the period 1850 –1914 (Klovland, 1994, pp. 184 –185).

We report summary statistics for the data set in Table 6 and plot the series in Figure 9. The patterns in Figure 9 strongly support a positive association between the average price level and interest rates.

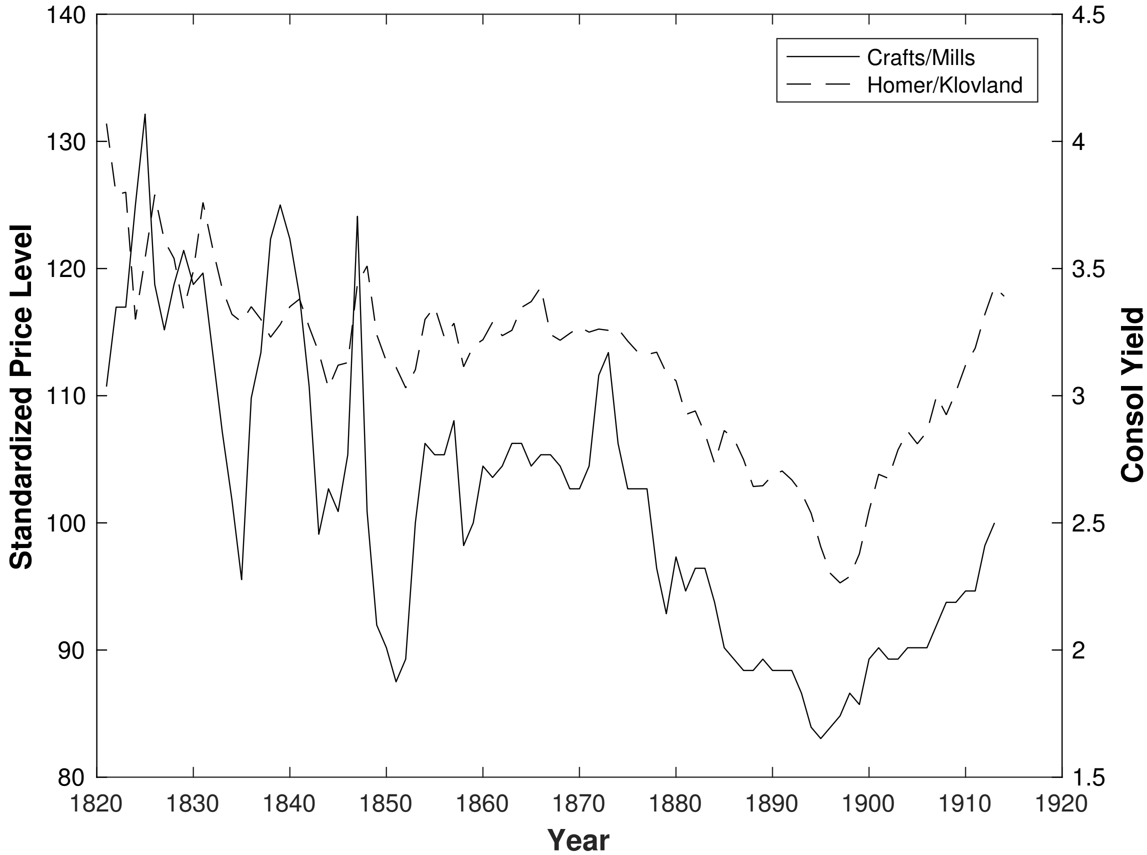
In Figure 10 we plot the copula density estimates for the two series. In Table 7 we present the estimates of Spearman's ρ for the series, based on the plug-in estimators using $\mathrm{SE}^{AIC}_{d=1,2}$, PESE, SMB^{AIC} and SMB^{AIC}_{CDC}[1]. All five copula density estimators show strong positive dependence between prices and interest rates, evident from the concentration of mass along the diagonal. It appears that tail dependence near the lower end is greater for the spline estimates than for the other estimates, and $\mathrm{SE}^{AIC}_{d=1,2}$ show fairly strong positive dependence in the center. This suggests that Gibson's paradox may be stronger than initially thought in the middle and lower parts of the joint distribution of prices and interest rates. The two SEs show higher estimates of Spearman's ρ than all the other estimators.

To conclude, summary statistics only capture the overall degree of dependence, while the estimated copula densities completely characterize the underlying dependence structure and, being genuine copula densities, provide reliable insights.

[1]Let X, Y be two random variables, and $c(\cdot,\cdot)$ be their copula density (assuming it exists). The plug-in estimator of Spearman's ρ for (X, Y) is given by

$$\hat{\rho} = 12 \iint_{[0,1]^2} uv\hat{c}(u,v)\, \mathrm{d}u\mathrm{d}v - 3,$$

where $\hat{c}$ is an estimator of c.. The integral is usually evaluated using sums over sample observations.



S

Figure 9: Plot of the price level and interest rate series

Table 6: Summary statistics

Series	Min	Max	Mean	Std	Skewness	Kurtosis
Crafts/Mills	83.04	132.14	101.41	11.62	0.48	2.41
Homer/Klovland	2.26	4.07	3.12	0.35	-0.32	3.19

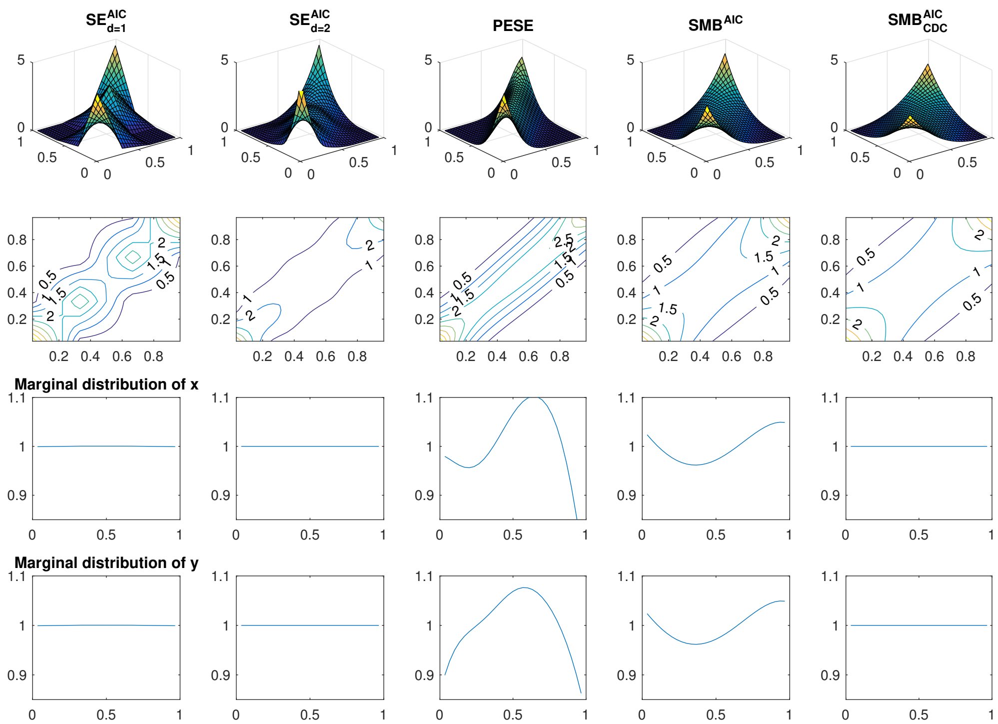


Figure 10: Copula density between Crafts/Mills and Homer/Klovland

Table 7: Estimates of Spearman's ρ for Crafts/Mills and Homer/Klovland

	Estimates of Spearman's ρ
Sample estimator	0.876
Plug-in estimator of Spearman's ρ based on	
$\text{SE}^{AIC}_{d=1}$	0.814
$\text{SE}^{AIC}_{d=2}$	0.785
PESE	0.781
SMB^{AIC}	0.753
SMB^{AIC}_{CDC}	0.600

6 Application to testing for independence

We now illustrate the use of our estimators in an independence test. Testing the null hypothesis of independence is equivalent to testing

$$\mathcal{H}_0 : c(\mathbf{u}) = 1, \quad \mathbf{u} \in [0,1]^2.$$

In order to test this null hypothesis, Belalia et al. (2017) proposed the following Cramér-von Mises-type test statistic

$$I_n(\mathbf{u}) = \int_{[0,1]^2} \left(c_{k,n}(\mathbf{u}) - 1\right)^2 \mathrm{d}\mathbf{u}, \tag{17}$$

where $c_{k,n}(\mathbf{u})$ is the empirical Bernstein copula density estimator.

Belalia et al. (2017, Theorem 3.1) show that, after appropriate standardization and scaling, the test statistic I_n converges under the null $\mathcal{H}_0$ to the standard normal distribution given a consistent estimator of the copula density. They also show that a simulation-based approach to calculating critical values (p-values) shows better behavior than using the asymptotic distribution.

Here we run a Monte Carlo experiment to investigate the performance of several versions of I_n based on different estimators of $c_{k,n}$. In particular, we study the power of the test statistics I_n based on two versions of SE and two versions of the Bernstein polynomial-based estimator, the Sieve MLE with Bernstein polynomials and the empirical Bernstein copula density estimator. The two Bernstein polynomial based estimators are denoted by I_n^B and I_n^{EB}, respectively.

The two spline estimators are different in that the copula property is imposed on one of them, while the other is only subject to the restriction that it is a density, not a copula density. The two statistics are denoted by I_n^S (copula property maintained) and I_n^{S*} (density property maintained).

For the estimators that use the smoothing parameters k and J, we use AIC for model selection. As for the empirical Bernstein copula estimator, we set $k = 10$, which generally gives better power in the simulations using this copula estimator (see, e.g., Belalia et al., 2017). In Appendix D, we also take various values of k and J to investigate the sensitivity of the power functions to the smoothing parameters. We do that for different sample sizes $n = \{100, 150\}$.

In order to calculate the critical values of these test statistics under the null for the nominal size of 5%, we generate $n = 100$ independent observations using the independent copula. We evaluate the empirical power of our tests by generating $n = 100$ observations using different copula functions under different values of Kendall's $\tau = \{0, 0.1, 0.25\}$. For Kendall's $\tau > 0.5$, all the tests provide very similar results and show high power so we do not report those results. The copulas we consider are Gaussian, Student, Clayton and Gumbel. We use Monte-Carlo approximations, based on 1,000 replications, to compute the critical values and the empirical power of all the tests.

We consider two scenarios for the marginal distributions. In the first one, we use the raw observations

$(u_i, v_i), i = 1, \ldots, n$. In the second scenario, we transform the raw observations into rank-based pseudo-observations, that is series of numbers from the set $\{\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n}\}$. The difference between the two scenarios is that the second scenario produces observations that are closer to being uniformly distributed than the first scenario. We report simulation results for I_n^S, I_n^{S*}, I_n^B and I_n^{EB} in Table 8 for the first scenario and in Table 9 for the second scenario.

Table 8: Empirical size and power of I_n^S, I_n^{S*}, I_n^B and I_n^{EB}; raw data.

	$\tau = 0$				$\tau = 0.10$				$\tau = 0.25$			
	I_n^S	I_n^{S*}	I_n^B	I_n^{EB}	I_n^S	I_n^{S*}	I_n^B	I_n^{EB}	I_n^S	I_n^{S*}	I_n^B	I_n^{EB}
Gaussian	0.051	0.049	0.052	0.050	0.200	0.062	0.107	0.127	0.942	0.173	0.716	0.632
Student	0.449	0.200	0.164	0.199	0.560	0.278	0.333	0.279	0.833	0.764	0.734	0.810
Clayton	0.049	0.050	0.051	0.048	0.186	0.073	0.110	0.150	0.887	0.342	0.723	0.685
Gumbel	0.048	0.049	0.050	0.049	0.187	0.071	0.119	0.105	0.890	0.302	0.658	0.714

Note: I_n^S is based on spline estimator with copula density constraints; I_n^{S*} is based on spline estimator only with joint density constraint; I_n^B is based on Sieve MLE with Bernstein polynomials with joint density constraint; I_n^{EB} is based on the empirical copula density.

Table 9: Empirical size and power of I_n^S, I_n^{S*}, I_n^B and I_n^{EB}; pseudo-observations.

	$\tau = 0$				$\tau = 0.10$				$\tau = 0.25$			
	I_n^S	I_n^{S*}	I_n^B	I_n^{EB}	I_n^S	I_n^{S*}	I_n^B	I_n^{EB}	I_n^S	I_n^{S*}	I_n^B	I_n^{EB}
Gaussian	0.046	0.048	0.055	0.060	0.226	0.156	0.225	0.165	0.932	0.841	0.932	0.867
Student	0.458	0.467	0.369	0.378	0.469	0.561	0.521	0.524	0.820	0.829	0.938	0.957
Clayton	0.046	0.051	0.065	0.057	0.221	0.156	0.178	0.250	0.921	0.866	0.882	0.926
Gumbel	0.052	0.050	0.054	0.062	0.217	0.153	0.221	0.189	0.896	0.863	0.812	0.887

Note: I_n^S is based on spline estimator with copula density constraints; I_n^{S*} is based on spline estimator only with joint density constraint; I_n^B is based on Sieve MLE with Bernstein polynomials with joint density constraint; I_n^{EB} is based on the empirical copula density.

By looking at Tables 8-9, we see that all tests generally control the size. Their power increases with the strength of dependence. I_n^S generally outperforms the rest of the tests by a large margin. We note that in

Table 8, I_n^S is the only test that is based on the estimator which is a true copula density function.

Table 9 shows the test power using pseudo-observations. We observe a significant improvement in the power of the test I_n^{EB} compared to the results in Table 8. This is because we are now using the histogram frequencies as parameters of the empirical Bernstein copula density estimator (EBDE), and if the histogram is using rank-based pseudo-observations with no ties, then the EBDE will satisfy the copula property, and it will do so exactly since k is a factor of the sample size.

Finally, we apply the independence test to the data from the two empirical applications of Section 5. For the BMI application, we compute the p-values for I_n^S by gender of child and parent. We use $\mathrm{SE}_{d=1}^{AIC}$ as the copula estimate and we decide on the smoothing parameter by ranging it between 1 and 20. Table 10 reports the p-values for these tests. For the Gibson Paradox application with only 93 pairs of observations, we use the same estimator but select the smoothing parameter in the range from 1 to 9. The p-value for that test is 0.011. So in both applications, we soundly reject independence.

Table 10: Testing for independence of children's and parents' BMI; p-value

	Male	Female
Father	0.033	0.019
Mother	0.026	0.020

7 Conclusion

We proposed a family of spline estimators which guarantee the uniform marginals property for the copula density estimator. The estimators are also strictly positive in the interior of the unit square and show excellent performance in terms of capturing the main features of the true data generating process. We illustrate this using Monte Carlo simulations.

The estimation procedure is a convex maximization problem with linear constraints, which is numerically easy to implement and costs less in terms of computation time compared to most other estimators of a similar complexity.

Our Monte Carlo simulations demonstrate the efficiency of the proposed estimators, both in terms of precision and in terms of time. We apply the proposed method to estimate the copula linking children's and parents' BMI and to estimate the copula linking prices and interest rates, and we investigate how the new estimators can be used in testing for independence.

A natural question is whether our results extend to higher dimensions. Unfortunately, we do not have much to say about this. Nonparametric estimators are plagued by the curse of dimensionality and our estimator is no exception. It is not difficult to generalize the tensor product spline surface in (6) to higher dimensions by including more terms in the product but this will inevitably affect convergence rates unless we can make use of dimension reduction techniques. There are not many such techniques for copulas primarily because a higher dimensional copula cannot in general be expressed as a simple function of lower dimensional copulas – this is known as the compatibility problem in the copula literature (see, e.g., Nelsen, 2006, pp. 105-107). One such technique is to use the vine decomposition (see, e.g., Joe, 1996) where a high-dimensional copula is expressed as a product of bivariate conditional copulas. In this case our two-dimensional estimator can be used as a building block of a copula with arbitrary dimension. We leave such extensions for future work.

Appendices

A Restrictions on the design points

We are going to show that the copula property puts heavy restrictions on the design points. Consider a vertical integral $\int_0^1 c(u,v)\mathrm{d}v$. Suppose this crosses triangulation lines $1,2,\ldots,M$ in order, where $M \in \mathbb{N}$. For $j \in 1,2,\ldots,M$, line j connects $(u_j^{(1)}, v_j^{(1)})$ and $(u_j^{(2)}, v_j^{(2)})$ with $v_1^{(1)} = v_1^{(2)} = 0$ and $v_M^{(1)} = v_M^{(2)} = 1$. The value of c at the point where we cross line j is

$$c_j = \frac{(u_j^{(2)} - u)c(u_j^{(1)}, v_j^{(1)}) + (u - u_j^{(1)})c(u_j^{(2)}, v_j^{(2)})}{(u_j^{(2)} - u_j^{(1)})},$$

and this occurs at a v value of

$$v_j = \frac{(u_j^{(2)} - u)v_j^{(1)} + (u - u_j^{(1)})v_j^{(2)}}{(u_j^{(2)} - u_j^{(1)})}.$$

The integral of c between line j and line $j+1$ is (using linearity)

$$\int_{v_j}^{v_{j+1}} \frac{(v_{j+1} - v)c_j + (v - v_j)c_{j+1}}{v_{j+1} - v_j}\mathrm{d}v = (v_{j+1} - v_j)\,\frac{(c_{j+1} + c_j)}{2}.$$

We need the total integral to be constant with changes in u. Note that

$$\frac{\partial c_j}{\partial u} = \frac{c(u_j^{(2)}, v_j^{(2)}) - c(u_j^{(1)}, v_j^{(1)})}{(u_j^{(2)} - u_j^{(1)})};$$

$$\frac{\partial v_j}{\partial u} = \frac{v_j^{(2)} - v_j^{(1)}}{(u_j^{(2)} - u_j^{(1)})}.$$

Hence

$$\sum_{j=1}^{M-1} \left(\left(\frac{\partial v_{j+1}}{\partial u} - \frac{\partial v_j}{\partial u} \right) (c_{j+1} + c_j) + (v_{j+1} - v_j) \left(\frac{\partial c_{j+1}}{\partial u} + \frac{\partial c_j}{\partial u} \right) \right) = 0.$$

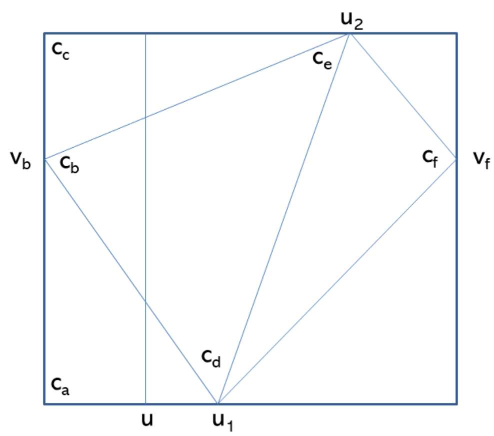


Figure 11

To examine the solution to the above system, we start with a simple case. Suppose that we have a situation as shown in Figure 11. To calculate the marginal integral at u, it crosses four triangulation lines. The first one connects $(0,0)$ and $(u_1,0)$. The second one connects $(0,v_b)$ and $(u_1,0)$. The third line connects $(0,v_b)$ and $(u_2,1)$, and the fourth line connects $(0,1)$ and $(u_2,1)$. Thus

$$c_1 = \frac{uc_d + (u_1-u)c_a}{u_1}, \frac{\partial c_1}{\partial u} = \frac{c_d - c_a}{u_1}, v_1 = 0$$
$$c_2 = \frac{uc_d + (u_1-u)c_b}{u_1}, \frac{\partial c_1}{\partial u} = \frac{c_d - c_b}{u_1}, v_2 = \frac{(u_1-u)v_b}{u_1}$$
$$c_3 = \frac{uc_e + (u_2-x)c_b}{u_2}, \frac{\partial c_3}{\partial u} = \frac{c_e - c_b}{u_2}, v_3 = \frac{u + (u_2-u)v_b}{u_2}$$
$$c_4 = \frac{uc_e + (u_2-u)c_c}{u_2}, \frac{\partial c_4}{\partial u} = \frac{c_e - c_c}{u_2}, v_4 = 1$$

and the integral is

$$\frac{v_2-v_1}{2}(c_2+c_1) + \frac{v_3-v_2}{2}(c_3+c_2) + \frac{v_4-v_3}{2}(c_4+c_3)$$

with derivative

$$\begin{aligned}
&\frac{(u_1-u)v_b}{u_1}\frac{2c_d-c_b-c_a}{u_1}-\frac{2uc_d+(u_1-u)(c_a+c_b)}{u_1}\frac{v_b}{u_1}\\
&+\left(\frac{u+(u_2-u)v_b}{u_2}-\frac{(u_1-u)v_b}{u_1}\right)\left(\frac{c_e-c_b}{u_2}+\frac{c_d-c_b}{u_1}\right)\\
&+\left(\frac{1-v_b}{u_2}+\frac{v_b}{u_1}\right)\left(\frac{uc_d+(u_1-u)c_b}{u_1}+\frac{uc_e+(u_2-u)c_b}{u_2}\right)\\
&+\left(1-\frac{u+(u_2-u)v_b}{u_2}\right)\frac{2c_e-c_b-c_c}{u_2}+\left(-\frac{1-v_b}{u_2}\right)\left(\frac{2uc_e+(u_2-u)(c_b+c_c)}{u_2}\right)\\
&=0.
\end{aligned}$$

Multiplying by $u_1^2u_2^2$, we get

$$\begin{aligned}
&2u_2^2v_b\left(uc_a+uc_b-2uc_d-u_1c_a-u_1c_b+u_1c_d\right)\\
&+2\left(u_1-u_1v_b+u_2v_b\right)\left(uu_1c_e-uu_1c_b-uu_2c_b+uu_2c_d+u_1u_2c_b\right)\\
&+2u_1^2\left(1-v_b\right)\left(uc_b-2uc_e+uc_c+u_2c_e-u_2c_b-u_2c_c\right)\\
&=0.
\end{aligned}$$

Since this is true for all u, we differentiate with respect to u and obtain

$$\begin{aligned}
u_2^2v_b\left(c_a+c_b-2c_d\right)+\left(u_1-u_1v_b+u_2v_b\right)\left(u_1c_e-u_1c_b-u_2c_b+u_2c_d\right)&\\
+u_1^2\left(1-v_b\right)\left(c_b-2c_e+c_c\right)&=0
\end{aligned}$$

which simplifies to

$$u_1^2(c_c-c_e)(1-v_b)+u_2^2v_b(c_a-c_d)+u_1u_2\left(c_d-c_b+c_ev_b-c_dv_b\right)=0. \tag{18}$$

Also setting $u=0$ gives

$$\begin{aligned}
2u_2^2v_b\left(-u_1c_a-u_1c_b+u_1c_d\right)+2\left(u_1-u_1v_b+u_2v_b\right)\left(u_1u_2c_b\right)&\\
+2u_1^2\left(1-v_b\right)\left(u_2c_e-u_2c_b-u_2c_c\right)&=0
\end{aligned}$$

which simplifies to

$$u_2v_b\left(c_d-c_a\right)+u_1\left(1-v_b\right)\left(c_e-c_c\right)=0. \tag{19}$$

We can then substitute this last equation in the term $u_2^2v_b(c_a-c_d)$ in (18) to get

$$\begin{aligned}
u_1^2(c_c-c_e)(1-v_b)+u_2u_1\left(1-v_b\right)\left(c_e-c_c\right)+u_1u_2\left(c_d-c_b+c_ev_b-c_dv_b\right)&=0\\
u_1(c_c-c_e)(1-v_b)+u_2\left(1-v_b\right)\left(c_e-c_c\right)+u_2\left(c_d-c_b+c_ev_b-c_dv_b\right)&=0.
\end{aligned}$$

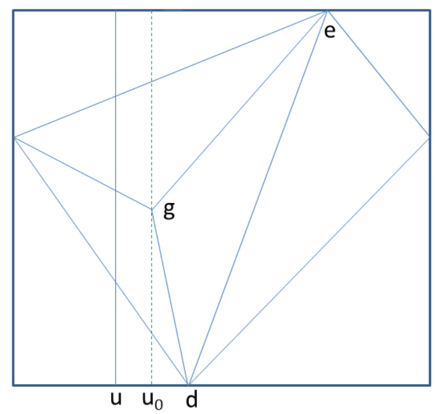


Figure 12

Dividing by u_1 and substituting for the term $u_1(c_c - c_e)(1 - v_b)$, we get

$$u_2 v_b (c_d - c_a) + u_2 (1 - v_b)(c_e - c_c) + u_2 (c_d - c_b + c_e v_b - c_d v_b) = 0$$

which simplifies to

$$c_e - c_b - c_c + c_d - c_a v_b + c_c v_b = 0. \tag{20}$$

We also need for the integral to be 1 at $u = 0$, so

$$v_b(c_a + c_b) + (1 - v_b)(c_b + c_c) = 2$$

giving

$$v_b(c_a - c_c) - 2 + c_b + c_c = 0. \tag{21}$$

Thus substituting in (20) gives

$$c_e + c_d = 2 \tag{22}$$

Thus we have three conditions (19), (21) and (22) on the design points that are implied by the copula property applied to the vertical integral crossing the triangle d-b-e.

However, if we add an internal point, then we may be unable to make this work. For example, with the points shown on Figure 12 there will be a change in the derivative with respect to u of the integral as u passes through u_0. This follows from the observation that the conditions implying that the marginal as u approaches u_0 from the left is constant, also imply that the marginal remains 1 were the surface to continue without the introduction of the triangle e-d-g. Hence introducing this triangulation cannot be consistent with constant marginals. Arguments of this sort imply that in many cases where there is an internal point, there must be a second point on the same horizontal line and another one on the same vertical line.

B Definitions and lemma for Proposition 2

Let $\boldsymbol{t} = (t_1 = \cdots = t_{d+1},\, t_{d+2}, \ldots, t_k,\, t_{k+1} = \cdots = t_{k+d+1})$ be a nondecreasing sequence of real numbers with $t_j < t_{j+d+1}$.

Definition 1 (de Boor, 1978) A B-spline $B_{j,d}$ of degree d over $\boldsymbol{t}$ is defined by the recurrence relation

$$B_{j,0}(u) := \begin{cases} 1 & \text{if } t_j \le u < t_{j+1} \\ 0 & \text{otherwise} \end{cases}$$

and

$$B_{j,r}(u) := \frac{u - t_j}{t_{j+r} - t_j} \cdot B_{j,r-1}(u) + \frac{t_{j+r+1} - u}{t_{j+r+1} - t_{j+1}} \cdot B_{j+1,r-1}(u)$$

for $1 \le r \le d$.

Definition 2 A knot vector $\boldsymbol{t} = (t_j)_{j=1}^{k+d+1}$ is said to be $(d+1)$-regular if $t_1 = ... = t_{d+1} < t_{d+2} < ... < t_k < t_{k+1} = ... = t_{k+d+1}$ and $k \ge d+1$.

A generic B-spline $B_{j,d}(x)$ has the following property on interval $[t_{d+1}, t_{k+1})$:

$$\sum_{j=1}^{k} B_{j,d}(x) \equiv 1. \tag{23}$$

Lemma (Bhatti and Brachen, 2006)

The integral over its support of a B-spline of degree d with $(d+1)$-regular knot vector $\boldsymbol{t} = (t_i)_{i=1}^{k+d+1}$ is given by

$$\begin{aligned} \int_{t_1}^{t_{k+d+1}} B_{j,d}(x)\mathrm{d}x &= \int_{t_j}^{t_{j+d+1}} B_{j,d}(x)\mathrm{d}x \\ &= \frac{t_{j+d+1} - t_j}{d+1}. \end{aligned}$$

C Cross-validation for PESE

Eq (8) of Gao et al. (2015, p.71) is

$$\mathcal{L}_{-} \approx \mathcal{L} - \frac{1}{(n-1)}\text{trace}(\Phi \hat{H}^{-1} \Phi') + \frac{1}{n(n-1)}(\iota' \Phi)\hat{H}^{-1}(\Phi' \iota). \tag{24}$$

Gao et al. (2015) skip the final steps in the derivation of Eq (8) but using their own formulas it can be shown that the correct expression is

$$\mathcal{L}_{-} \approx \mathcal{L} - \frac{1}{n(n-1)}\text{trace}(\Phi \hat{H}^{-1}\Phi') + \frac{1}{n^2(n-1)}(\iota'\Phi)\hat{H}^{-1}(\Phi'\iota).$$

To show this we use their notation and denote

$$\hat{\phi} = \frac{1}{n}\sum_{i=1}^{n}\phi(X_i, Y_i),$$
$$\hat{\phi}_{-i} = \frac{1}{n-1}\sum_{j\neq i}^{n}\phi(X_j, Y_j).$$

Their Eq (7) is

$$\mathcal{L}_{-} \approx \mathcal{L} - \frac{1}{n}\sum_{i=1}^{n}\phi'(X_i, Y_i)\hat{H}^{-1}(\hat{\phi} - \hat{\phi}_{-i})$$

To arrive at the correct expression, note that

$$\begin{aligned}
&\mathcal{L} - \frac{1}{n}\sum_{i=1}^{n}\phi'(X_i, Y_i)\hat{H}^{-1}(\hat{\phi} - \hat{\phi}_{-i}) \\
&= \mathcal{L} - \frac{1}{n}\sum_{i=1}^{n}\phi'(X_i, Y_i)\hat{H}^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\phi(X_i, Y_i) - \frac{1}{n-1}\sum_{j\neq i}^{n}\phi(X_j, Y_j)\right) \\
&= \mathcal{L} - \frac{1}{n}\sum_{i=1}^{n}\phi'(X_i, Y_i)\hat{H}^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\phi(X_i, Y_i) - \frac{1}{n-1}\left(\sum_{j=1}^{n}\phi(X_j, Y_j) - \phi(X_i, Y_i)\right)\right] \\
&= \mathcal{L} - \frac{1}{n}\sum_{i=1}^{n}\phi'(X_i, Y_i)\hat{H}^{-1}\left(\frac{1}{n-1}\phi(X_i, Y_i) - \frac{1}{n(n-1)}\sum_{i=1}^{n}\phi(X_i, Y_i)\right) \\
&= \mathcal{L} - \frac{1}{n(n-1)}\sum_{i=1}^{n}\phi'(X_i, Y_i)\hat{H}^{-1}\phi(X_i, Y_i) + \frac{1}{n^2(n-1)}\left(\sum_{i=1}^{n}\phi'(X_i, Y_i)\right)\hat{H}^{-1}\left(\sum_{i=1}^{n}\phi(X_i, Y_i)\right) \\
&= \mathcal{L} - \frac{1}{n(n-1)}\text{trace}(\Phi\hat{H}^{-1}\Phi') + \frac{1}{n^2(n-1)}(\iota'\Phi)\hat{H}^{-1}(\Phi'\iota),
\end{aligned}$$

where Φ is a $n \times M$ matrix with the ith row being $\big(\phi_1(X_i, Y_i), \ldots, \phi_M(X_i, Y_i)\big)$ and ι is a column vector of length n with each element being 1.

We further note that a similar cross-validation strategy is used by Gu and Wang (2003) and their formula agrees with ours.

D Sensitivity of test statistics to smoothing parameters

Table 11-12 report the empirical size and power for various values of grid size k, when the copula is estimated using pseudo-observations. We can see that the size/power results are consistent over all values of k. We can also see that I_n^S compares favorably with I_n^{S*} in most cases.

Table 11: Empirical size and power of test statistics I_n^S, I_n^{S*}, I_n^B and I_n^{EB} with bandwidth parameter $k \in \{1, 2, \ldots, 9\}$; marginals estimated by empirical density; sample size $n = 100$; number of replications = 1,000.

	Statistic I_n with Spline estimators							Statistic I_n^B and I_n^{EB}					
	$\tau = 0$		$\tau = 0.1$		$\tau = 0.25$			$\tau = 0$		$\tau = 0.1$		$\tau = 0.25$	
k	I_n^S	I_n^{S*}	I_n^S	I_n^{S*}	I_n^S	I_n^{S*}	J	I_n^B	I_n^{EB}	I_n^B	I_n^{EB}	I_n^B	I_n^{EB}
True Copula: Gaussian													
1	0.046	0.050	0.226	0.215	0.932	0.925	1	–		–		–	
2	0.051	0.046	0.164	0.125	0.834	0.808	2	0.055	0.206	0.196	0.152	0.921	0.460
3	0.068	0.042	0.109	0.091	0.759	0.688	3	0.054	0.055	0.208	0.187	0.886	0.803
4	0.047	0.042	0.094	0.082	0.654	0.606	4	0.062	0.037	0.159	0.180	0.884	0.893
5	0.042	0.058	0.086	0.071	0.594	0.506	5	0.042	0.050	0.125	0.213	0.854	0.891
6	0.051	0.044	0.073	0.070	0.522	0.434	6	0.046	0.049	0.125	0.173	0.824	0.881
7	0.041	0.045	0.074	0.061	0.430	0.352	7	0.059	0.060	0.123	0.189	0.813	0.891
8	0.042	0.053	0.072	0.069	0.429	0.347	8	0.039	0.044	0.118	0.185	0.765	0.886
9	0.048	0.058	0.070	0.053	0.363	0.321	9	0.041	0.034	0.105	0.168	0.740	0.850
True Copula: Clayton													
1	0.046	0.047	0.221	0.227	0.921	0.914	1	–		–		–	
2	0.054	0.040	0.189	0.168	0.908	0.872	2	0.038	0.205	0.196	0.138	0.904	0.460
3	0.045	0.049	0.148	0.141	0.857	0.828	3	0.068	0.049	0.230	0.146	0.920	0.791
4	0.043	0.045	0.127	0.120	0.809	0.751	4	0.052	0.032	0.178	0.212	0.913	0.853
5	0.048	0.058	0.112	0.110	0.755	0.693	5	0.047	0.052	0.207	0.180	0.912	0.887
6	0.049	0.047	0.104	0.125	0.716	0.670	6	0.062	0.051	0.132	0.230	0.883	0.921
7	0.051	0.064	0.102	0.091	0.635	0.590	7	0.054	0.061	0.147	0.214	0.905	0.907
8	0.064	0.052	0.078	0.090	0.564	0.549	8	0.031	0.050	0.166	0.226	0.859	0.916
9	0.058	0.056	0.072	0.078	0.533	0.489	9	0.039	0.069	0.126	0.228	0.846	0.929
True Copula: Gumbel													
1	0.052	0.051	0.217	0.201	0.896	0.888	1	–		–		–	
2	0.051	0.047	0.170	0.169	0.876	0.831	2	0.052	0.183	0.173	0.159	0.896	0.631
3	0.031	0.048	0.130	0.117	0.816	0.779	3	0.066	0.039	0.208	0.155	0.894	0.767
4	0.053	0.057	0.129	0.106	0.737	0.682	4	0.063	0.058	0.162	0.193	0.893	0.859
5	0.046	0.054	0.110	0.097	0.677	0.605	5	0.048	0.043	0.170	0.175	0.864	0.878
6	0.047	0.035	0.104	0.107	0.621	0.566	6	0.052	0.050	0.119	0.168	0.846	0.891
7	0.044	0.059	0.086	0.089	0.575	0.506	7	0.044	0.045	0.136	0.193	0.846	0.893
8	0.042	0.046	0.087	0.088	0.554	0.499	8	0.057	0.059	0.162	0.208	0.837	0.869
9	0.054	0.047	0.080	0.088	0.505	0.431	9	0.047	0.048	0.175	0.222	0.805	0.887
True Copula: Student													

Continued on next page

Table 11 – continued from previous page

	Statistic I_n with Spline estimators							Statistic I_n^B and I_n^{EB}					
	$\tau = 0$		$\tau = 0.1$		$\tau = 0.25$			$\tau = 0$		$\tau = 0.1$		$\tau = 0.25$	
k	I_n^S	I_n^{S*}	I_n^S	I_n^{S*}	I_n^S	I_n^{S*}	J	I_n^B	I_n^{EB}	I_n^B	I_n^{EB}	I_n^B	I_n^{EB}
1	0.042	0.040	0.092	0.107	0.661	0.657	1	–		–		–	
2	0.260	0.335	0.434	0.583	0.962	0.925	2	0.039	0.167	0.112	0.174	0.605	0.455
3	0.390	0.391	0.562	0.571	0.935	0.932	3	0.025	0.040	0.149	0.167	0.692	0.747
4	0.361	0.313	0.489	0.451	0.886	0.855	4	0.076	0.052	0.212	0.196	0.862	0.855
5	0.318	0.301	0.463	0.413	0.875	0.811	5	0.207	0.099	0.384	0.295	0.915	0.874
6	0.293	0.292	0.436	0.382	0.856	0.803	6	0.270	0.122	0.423	0.325	0.928	0.892
7	0.300	0.294	0.397	0.355	0.813	0.795	7	0.293	0.201	0.483	0.352	0.938	0.942
8	0.267	0.269	0.335	0.357	0.789	0.747	8	0.348	0.218	0.556	0.443	0.921	0.942
9	0.273	0.250	0.347	0.338	0.766	0.721	9	0.370	0.345	0.494	0.557	0.926	0.935

Note: I_n^S imposes copula density constraints; I_n^{S*} imposes only joint density constraint; I_n^B is based on Sieve MLE with Bernstein polynomials under joint density constraint; I_n^{EB} is based on the empirical copula density; blue entries indicate better size/power.

Table 12: Empirical size and power of test statistics I_n^S, I_n^{S*}, I_n^B and I_n^{EB} with bandwidth parameter $k \in \{1, 2, \ldots, 9\}$; marginals estimated by empirical density; sample size $n = 150$; number of replications = 1,000.

	Statistic I_n with Spline estimators							Statistic I_n^B and I_n^{EB}					
	$\tau = 0$		$\tau = 0.1$		$\tau = 0.25$			$\tau = 0$		$\tau = 0.1$		$\tau = 0.25$	
k	I_n^S	I_n^{S*}	I_n^S	I_n^{S*}	I_n^S	I_n^{S*}	J	I_n^B	I_n^{EB}	I_n^B	I_n^{EB}	I_n^B	I_n^{EB}
True Copula: Gaussian													
1	0.045	0.048	0.254	0.268	0.965	0.942	1	–		–		–	
2	0.051	0.045	0.256	0.187	0.891	0.848	2	0.050	0.181	0.205	0.192	0.945	0.510
3	0.054	0.046	0.161	0.132	0.789	0.715	3	0.051	0.052	0.225	0.198	0.921	0.823
4	0.048	0.048	0.144	0.111	0.724	0.641	4	0.051	0.043	0.177	0.206	0.884	0.893
5	0.045	0.052	0.092	0.089	0.644	0.526	5	0.047	0.051	0.141	0.252	0.854	0.912
6	0.049	0.048	0.087	0.088	0.564	0.465	6	0.046	0.049	0.125	0.173	0.824	0.881
7	0.045	0.046	0.101	0.075	0.499	0.380	7	0.050	0.052	0.141	0.205	0.841	0.901
8	0.041	0.051	0.095	0.076	0.447	0.372	8	0.045	0.050	0.159	0.207	0.815	0.901
9	0.050	0.052	0.091	0.067	0.420	0.356	9	0.048	0.042	0.145	0.189	0.782	0.912
True Copula: Clayton													
1	0.044	0.049	0.279	0.282	0.950	0.944	1	–		–		–	
2	0.049	0.046	0.222	0.196	0.939	0.907	2	0.045	0.189	0.251	0.182	0.919	0.521

Continued on next page

Table 12 – continued from previous page

	Statistic I_n with Spline estimators							Statistic I_n^B and I_n^{EB}					
	$\tau=0$		$\tau=0.1$		$\tau=0.25$			$\tau=0$		$\tau=0.1$		$\tau=0.25$	
k	I_n^S	I_n^{S*}	I_n^S	I_n^{S*}	I_n^S	I_n^{S*}	J	I_n^B	I_n^{EB}	I_n^B	I_n^{EB}	I_n^B	I_n^{EB}
3	0.044	0.051	0.168	0.154	0.891	0.885	3	0.055	0.051	0.262	0.183	0.956	0.810
4	0.042	0.047	0.143	0.131	0.866	0.768	4	0.049	0.049	0.226	0.246	0.945	0.901
5	0.046	0.053	0.162	0.148	0.783	0.776	5	0.045	0.051	0.227	0.212	0.945	0.919
6	0.050	0.051	0.157	0.145	0.808	0.720	6	0.055	0.049	0.171	0.263	0.883	0.921
7	0.050	0.048	0.155	0.122	0.681	0.629	7	0.048	0.051	0.181	0.237	0.905	0.907
8	0.058	0.047	0.107	0.115	0.680	0.661	8	0.048	0.058	0.197	0.246	0.873	0.936
9	0.052	0.049	0.103	0.132	0.593	0.514	9	0.044	0.057	0.165	0.248	0.861	0.944
True Copula: Gumbel													
1	0.051	0.050	0.257	0.226	0.933	0.917	1	–		–		–	
2	0.052	0.050	0.214	0.207	0.910	0.873	2	0.049	0.179	0.226	0.168	0.921	0.710
3	0.044	0.052	0.154	0.149	0.862	0.815	3	0.053	0.039	0.226	0.182	0.921	0.795
4	0.051	0.052	0.147	0.135	0.761	0.722	4	0.048	0.052	0.187	0.213	0.902	0.887
5	0.042	0.051	0.169	0.138	0.724	0.631	5	0.051	0.049	0.192	0.223	0.893	0.909
6	0.051	0.043	0.139	0.159	0.639	0.591	6	0.051	0.049	0.159	0.184	0.859	0.913
7	0.049	0.052	0.104	0.116	0.623	0.574	7	0.051	0.049	0.185	0.229	0.897	0.921
8	0.049	0.049	0.118	0.132	0.610	0.531	8	0.049	0.045	0.195	0.230	0.873	0.923
9	0.050	0.046	0.198	0.113	0.545	0.489	9	0.048	0.046	0.215	0.247	0.891	0.919
True Copula: Student													
1	0.067	0.076	0.142	0.164	0.735	0.767	1	–		–		–	
2	0.280	0.389	0.597	0.621	0.989	0.965	2	0.049	0.218	0.176	0.222	0.651	0.514
3	0.426	0.431	0.613	0.604	0.966	0.959	3	0.056	0.050	0.217	0.201	0.725	0.797
4	0.410	0.413	0.549	0.576	0.891	0.893	4	0.104	0.097	0.276	0.257	0.904	0.895
5	0.402	0.301	0.463	0.413	0.875	0.811	5	0.221	0.165	0.423	0.376	0.932	0.899
6	0.347	0.344	0.491	0.477	0.920	0.862	6	0.314	0.163	0.476	0.355	0.951	0.921
7	0.346	0.334	0.451	0.454	0.879	0.865	7	0.318	0.243	0.535	0.416	0.979	0.965
8	0.322	0.331	0.367	0.378	0.841	0.838	8	0.418	0.256	0.610	0.460	0.935	0.965
9	0.304	0.301	0.389	0.390	0.804	0.813	9	0.456	0.405	0.627	0.517	0.949	0.969

Note: I_n^S imposes copula density constraints; I_n^{S*} imposes only joint density constraint; I_n^B is based on Sieve MLE with Bernstein polynomials under joint density constraint; I_n^{EB} is based on the empirical copula density; blue entries indicate better size/power.

References

Aitchison, J. and S. D. Silvey (1958): "Maximum-Likelihood Estimation of Parameters Subject to Restraints," *The Annals of Mathematical Statistics*, 29, 813–828.

Belalia, M., T. Bouezmarni, F. C. Lemyre, and A. Taamouti (2017): "Testing independence based on Bernstein empirical copula and copula density," *Journal of Nonparametric Statistics*, 29, 346–380.

Bhatti, M. and P. Brachen (2006): "The calculation of integrals involving B-splines by means of recursion relations," *Applied Mathematics and Computation*, 172, 91–100.

Bouezmarni, T., A. El Ghouch, and A. Taamouti (2013): "Bernstein estimator for unbounded copula densities," *Statistics and Risk Modeling*, 30, 343–360.

Bouezmarni, T., J. V. Rombouts, and A. Taamouti (2010): "Asymptotic properties of the Bernstein density copula estimator for α-mixing data," *Journal of Multivariate Analysis*, 101, 1–10.

Bozic, M., J. Newton, C. S. Thraen, and B. W. Gould (2014): "Tails Curtailed: Accounting for Nonlinear Dependence in Pricing Margin Insurance for Dairy Farmers," *American Journal of Agricultural Economics*, 96, 1117–1135.

Burda, M. and A. Prokhorov (2014): "Copula based factorization in Bayesian multivariate infinite mixture models," *Journal of Multivariate Analysis*, 127, 200 – 213.

Chen, X. (2007): "Chapter 76 Large Sample Sieve Estimation of Semi-Nonparametric Models," vol. 6 of *Handbook of Econometrics*, 5549 – 5632.

Chui, C. and X. Wu (2009): "Exponential series estimation of empirical copulas with application to financial returns," in *Nonparametric Econometric Methods*, vol. 25 of *Advances in Econometrics*, 263–290.

Crafts, N. and T. C. Mills (1994): "Trends in Real Wages in Britain, 1750 –1913," *Explorations in Economic History*, 31, 176–194.

de Boor, C. (1978): *A practical guide to splines*, New York: Springer.

Deheuvels, P. (1979): "La Fonction de Dépendance Empirique et ses Propriétés. Un Test non Paramétrique D'independance," *Bullettin de l'académie Royal de Belgique, Classe des Sciences*, 65, 274–292.

Dowd, K. (2008): "Copulas in Macroeconomics," *Journal of International and Global Economic Studies*, 1, 1–26.

Durante, F., S. Saminger-Platz, and P. Sarkoci (2009): "Rectangular Patchwork for Bivariate Copulas and Tail Dependence," *Communications in Statistics - Theory and Methods*, 38, 2515–2527.

Erdely, A. (2016): "A note on linear B-spline copulas," *arXiv:1603.04933v2*.

Fan, Y. and A. J. Patton (2014): "Copulas in Econometrics," *Annual Review of Economics*, 6, 179–200.

Gao, Y., Y. Y. Zhang, and X. Wu (2015): "Penalized exponential series estimation of copula densities with an application to intergenerational dependence of body mass index," *Empirical Economics*, 48, 61–81.

Gijbels, I. and J. Mielniczuk (1990): "Estimating the density of a copula function," *Communications in Statistics - Theory and Methods*, 19, 445–464.

Gu, C. and J. Wang (2003): "Penalized Likelihood Density Estimation: Direct Cross-Validation and Scalable Approximation," *Statistica Sinica*, 13, 811–826.

Homer, S. (1963): *A History of Interest Rates*, New Brunswick, NJ: Rutgers University Press.

Hua, Y. Z. L. and J. Huang (2010): "A Spline-Based Semiparametric Maximum Likelihood Estimation Method for the Cox Model with Interval-Censored Data," *Scandinavian Journal of Statistics*, 37, 338–354.

Ibragimov, R. and A. Prokhorov (2017): *Topics in Dependence Modelling: Heavy Tails and Copulas in Economics and Finance*, World Scientific Publishing Company Pte Limited.

Joe, H. (1996): "Families of m-variate distributions with given margins and m (m-1)/2 bivariate dependence parameters," *in L. Rschendorf, B. Schweizer and M.D. Taylor (eds.), Distributions with Fixed Marginals and Related Topics, IMS Lecture Notes Monograph Series, Institute of Mathematical Statistics*, 120–141.

Kauerman, G., C. Schellhase, and D. Ruppert (2013): "Flexible Copula Density Estimation with Penalized Hierarchical B-splines," *Scandinavian Journal of Statistics*, 40, 684–705.

Klement, E. P., A. Kolesarova, R. Mesiar, and C. Sempi (2007): "Copulas Constructed from Horizontal Sections," *Communications in Statistics - Theory and Methods*, 36, 2901–2911.

Klovland, J. T. (1994): "Pitfalls in the Estimation of the Yield on British Consols, 1850-1914," *Journal of Economic History*, 54, 164–187.

Li, X., P. Mikusinski, and M. D. Taylor (1997): "Strong approximation of copulas," *Journal of Mathematical Analysis and Applications*, 225, 608–623.

Lu, M. (2010): "Spline-based sieve maximum likelihood estimation in the partly linear model under monotonicity constraints," *Journal of Multivariate Analysis*, 101, 2528 – 2542.

Nelsen, R. (2006): *An Introduction to Copulas*, Springer.

NELSEN, R. B., J. J. QUESADA-MOLINA, J. A. RODRGUEZ-LALLENA, AND M. UBEDA-FLORES (2008): "On the construction of copulas and quasi-copulas with given diagonal sections," *Insurance: Mathematics and Economics*, 42, 473 – 483.

OMELKA, M., I. GIJBELS, AND N. VERAVERBEKE (2009): "Improved kernel estimation of copulas: Weak convergence and goodness-of-fit testing," *The Annals of Statistics*, 37, 3023–3058.

PANCHENKO, V. AND A. PROKHOROV (2016): "Efficient estimation of parameters in marginals in semiparametric multivariate models," *Working Papers 2016-04, University of Sydney Business School, Discipline of Business Analytics.*

QU, L., Y. QIAN, AND H. XIE (2009): "Copula Density Estimation by Total Variation Penalized Likelihood," *Communications in Statistics - Simulation and Computation*, 38, 1891–1908.

QU, L. AND W. YIN (2012): "Copula density estimation by total variation penalized likelihood with linear equality constraints," *Computational Statistics and Data Analysis*, 56, 384–398.

SANCETTA, A. AND S. SATCHELL (2004): "The Bernstein Copula And Its Applications To Modelling And Approximations Of Multivariate Distributions," *Econometric Theory*, 20, 535–562.

SCHUSTER, E. (1985): "Incorporating support constraints into nonparametric estimators of densities," *Communications in Statistics - Theory and Methods*, 14, 1123–1136.

SEGERS, J., M. SIBUYA, AND H. TSUKAHARA (2017): "The empirical beta copula," *Journal of Multivariate Analysis*, 155, 35–51.

SHAO, Q. AND L. YANG (2017): "Oracally efficient estimation and consistent model selection for autoregressive moving average time series with trend," *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 79, 507–524.

SHEN, X., Y. ZHU, AND L. SONG (2008): "Linear B-spline copulas with applications to nonparametric estimation of copulas," *Computational Statistics and Data Analysis*, 52, 3806–3819.

SIBURG, K. F. AND P. A. STOIMENOV (2008): "A scalar product for copulas," *Journal of Mathematical Analysis and Applications*, 344, 429 – 439.

SKLAR, A. (1959): "Fonctions de répartition à n dimensions et leurs marges," *Publications de l'Institut de Statistique de l'Université de Paris*, 8, 229–231.

WANG, G. AND L. WANG (2015): "Spline estimation and variable selection for single-index prediction models with diverging number of index parameters," *Journal of Statistical Planning and Inference*, 162, 1–19.

Wang, L. and L. Yang (2009): "Spline estimation of single-index models," *Statistica Sinica*, 19, 765–783.

Xue, L. and H. Liang (2010): "Polynomial Spline Estimation for a Generalized Additive Coefficient Model," *Scandinavian Journal of Statistics*, 37, 26–46.

Zhang, Q., D. Li, and H. Wang (2013): "A note on tail dependence regression," *Journal of Multivariate Analysis*, 120, 163 – 172.