Electronic Journal of Statistics
Vol. 16 (2022) 2550–2606
ISSN: 1935-7524
https://doi.org/10.1214/22-EJS2007

Smooth bootstrapping of copula functionals*

Maximilian Coblenz

Department of Services and Consulting, Ludwigshafen University of Business and Society, Germany
e-mail: maximilian.coblenz@hwg-lu.de

Oliver Grothe

Institute of Operations Research, Karlsruhe Institute of Technology, Karlsruhe, Germany
e-mail: oliver.grothe@kit.edu

Klaus Herrmann

Département de mathématiques, Université de Sherbrooke, Sherbrooke, Canada
e-mail: klaus.herrmann@usherbrooke.ca

and

Marius Hofert

Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Canada
e-mail: marius.hofert@uwaterloo.ca

Abstract: The smooth bootstrap for estimating copula functionals in small samples is investigated. It can be used both to gauge the distribution of the estimator in question and to augment the data. Issues arising from kernel density and distribution estimation in the copula domain are addressed, such as how to avoid the bounded domain, which bandwidth matrix to choose, and how the smoothing can be carried out. Furthermore, we investigate how the smooth bootstrap impacts the underlying dependence structure or the functionals in question and under which conditions it does not. We provide specific examples and simulations that highlight advantages and caveats of the approach.

Keywords and phrases: Smooth bootstrap, kernel distribution estimation, bandwidth selection, kernel smoothing, bandwidth matrix, dependence distortion, data augmentation.

Received October 2021.

arXiv: 2110.03397

*The third author gratefully acknowledges financial support of the Karlsruhe Institute of Technology (KIT) where part of this research was carried out. The third author would like to thank NSERC for financial support for this work through Discovery Grant RGPIN-2020-05784. The fourth author acknowledges support from NSERC through grants RGPIN-2020-04897 and RGPAS-2020-00093.

1. Introduction and notation

Contrary to resampling from the observed values, resampling in the smooth bootstrap introduced in Efron (1982) is done from a smoothed version of the empirical distribution function. We consider the multivariate smooth bootstrap for functionals T defined on the set (or possibly only a subset) of copulas which represent the possible dependence structures at hand. Examples for such functionals include measures of association such as Kendall's tau or Spearman's rho, the upper and lower tail dependence coefficients, or level sets that are used to quantify the risk inherent in joint events; see, for example, Salvadori et al. (2016), and Coblenz, Dyckerhoff and Grothe (2018) where in the latter the smooth bootstrap has recently been used to estimate level sets of copulas.

Our investigation is motivated by the question of how much the smoothing aspect of smooth bootstrap influences the underlying dependence structure in a multivariate framework. The whole procedure can also be seen as a form of data augmentation – a topic which recently has drawn considerable attention in the machine learning community, see, e.g., Shorten and Khoshgoftaar (2019), Wong et al. (2016), and Taylor and Nitschke (2018). In general the extent of this *dependence distortion* introduced by the smooth bootstrap may depend on the functional, the smoothing kernel or the sample size.

The smooth bootstrap used in this paper is different from jittering as, e.g., investigated in Nagler (2018). Usually, jittering is used for making discrete variables continuous and is not a resampling scheme in the classical sense. In addition to that, jittering is carried out in the data space, whereas we employ the smooth bootstrap in a transformed data space and focus specifically on the dependence structure, cf. Figure 1.

Furthermore, we want to point out that there are links to other empirical versions of the copula such as the empirical beta copula, see Segers, Sibuya and Tsukahara (2017), and the checkerboard copula. In contrast to the empirical copula, a sample from the latter two is not a simple resampling of the initial observations, but produces previously unseen observations. However, the sampling schemes are different from the resampling scheme employed in this paper. While the empirical beta copula smoothes based on the distribution of rank orders, the smoothing considered here relies on adding a smoothing kernel from a given family. Further, the 'smoothing' implied by drawing from the checkerboard copula is related to latin hypercube sampling, see Packham and Schmidt (2010), and thus also different in basic principle from the kernel based smoothing analyzed here.

We contribute to the existing literature in the following ways. Firstly, we provide theoretical details on the smooth bootstrap for copula functionals. We focus on elliptical distributions at the population level where the bandwidth matrices are obtained by the commonly used sphering approach. In this framework we show that the dependence distortion of the underlying elliptical copula is solely due to a distortion of the associated characteristic generator, whereas the associated parameter matrix remains unaffected. Our investigation reveals surprising cases where kernel smoothing has no impact on the underlying de-

pendence structure and we identify and characterize the responsible mechanism in detail. We also discuss that this fact holds true for certain functionals of copulas which remain unaffected by kernel smoothing. Concerning the rate of convergence in our framework, we connect the absolute value difference between the characteristic functions with the regular variation property of the characteristic generator of the smoothing kernel and we discuss related examples and practical implications. While most motivating examples are bivariate in nature, our discussion is valid in arbitrary dimensions d making the results viable for high-dimensional settings and data science applications in general.

Secondly, when considering a fixed sample we provide practical details concerning the evaluation of the copula of the smoothed distribution estimator, where our approach is not limited to product kernels when elliptical smoothing kernels are used. Furthermore, in Algorithm 4.1 we discuss a sampling strategy for constructing smooth bootstrap samples. The resulting smooth bootstrap sample can then be used to gauge the variability of a functional $T_n(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ based on a sample of size n. Furthermore, as in Coblenz, Dyckerhoff and Grothe (2018), the suggested algorithm can be used for data augmentation, for example, to obtain smoothed results in order to facilitate the numerical computation of T or to circumvent problems with ties in the original or an alternative (ordinary) bootstrap sample. Data augmentation in this sense is particularly valuable for small sample sizes.

Thirdly, we generalise the bandwidth selection procedure of Bowman, Hall and Prvan (1998) for kernel distribution estimation to the multivariate case. While bandwidth selection is a crucial part of kernel distribution estimation it is generally discussed for diagonal bandwidth matrices, see, for example, Liu and Yang (2008). We introduce a cross-validation-based bandwidth matrix selection procedure that is not limited to diagonal bandwidth matrices or product kernels to overcome the aforementioned limitations and can select non-diagonal bandwidth matrices in a data driven and optimal way.

Lastly we illustrate the smooth bootstrap for copula functionals with a number of theoretical examples and simulation studies.

The remainder of this paper is structured as follows: In the next section we introduce our notation and provide necessary preliminaries on copulas and multivariate kernel smoothing. Additionally, we give an overview of the smooth bootstrap setting and the estimators involved. Section 3 comprises our theoretical considerations. We start in 3.1 by reviewing basic facts about the connection of multivariate kernel smoothing to convolution and thus sums of independent random vectors. Section 3.2 introduces dependence distortion at the population level for elliptical distributions and smoothing kernels. Aside from deriving conditions under which the dependence distortion can be quantified precisely, we also discuss examples of functionals that do not exhibit dependence distortion. In Section 3.3 we link our results to regular variation properties of the characteristic generator associated with the elliptical kernel distribution. We discuss convergence rates of the distorted characteristic function towards the undistorted characteristic function and hence dependence structure in terms of regular variation of the radial distribution associated with the smoothing kernel.

Section 3.4 identifies special classes of distributions for which the dependence structure is unaffected by smoothing at the population level. To complete the picture we also present examples for which the dependence distortion effects are present and can be worked out in detail. In Section 4 we review the multivariate smooth bootstrap for a given sample. We introduce an algorithm to obtain a smooth bootstrap sample from the copula. Furthermore, we not only introduce and discuss a method to select appropriate non-diagonal bandwidth matrices via cross-validation but also cover some asymptotic considerations of our approach. A simulation study to show the impact of dependence distortion for popular functionals is conducted in Section 5. Final conclusions are given in Section 6. Appendix A and Appendix B provide details about characteristic functions and elliptical random vectors, respectively.

2. Notation, preliminaries, and setting

In this section we provide necessary preliminaries and introduce our setting. Also, we introduce the notation as needed. Much of the attention of this article is devoted to the impact of kernel smoothing and bootstrapping on the dependence structure between random variables which we represent by copulas. Copulas are one way to model the dependence between components of random vectors. While we only review the notions relevant for our exposition, textbook introductions to dependence modeling via copulas can be found, e.g., in Nelsen (2006); Mai and Scherer (2012); Joe (2015); Durante and Sempi (2016); Hofert et al. (2018).

A function $\boldsymbol{C}\colon [0,1]d \to [0,1]$ is called a *d-copula* (or *copula*), if $\boldsymbol{C}$ is the distribution function of a d-dimensional random vector $\boldsymbol{U} = (U_1, \dots, U_d)$ with standard uniform univariate marginals, that is, $\mathbb{P}[U_k \leq u_k] = u_k$ for all $k \in \{1, \dots, d\}$ and $u_k \in [0,1]$. The importance of copulas stems from Sklar's Theorem, see Sklar (1959), which states that any d-dimensional distribution function $\boldsymbol{F}$ with margins $F_1, \dots, F_d$ can be decomposed as (where $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$)

$$\boldsymbol{F}(x_1, \dots, x_d) = \boldsymbol{C}(F_1(x_1), \dots, F_d(x_d)), \quad \boldsymbol{x} \in \overline{\mathbb{R}}^d, \tag{1}$$

where $\boldsymbol{C}$ is a copula. If $F_1, \dots, F_d$ are all continuous, then $\boldsymbol{C}$ is unique. Otherwise, $\boldsymbol{C}$ is uniquely determined on $\times_{j=1}^{d} \operatorname{range} F_j$, where $\operatorname{range} F_j = F_j(\overline{\mathbb{R}})$ for $j = 1, \dots, d$. Conversely, if $\boldsymbol{C}$ is a d-copula and $F_1, \dots, F_d$ are distribution functions, then the function $\boldsymbol{F}$ defined by (1) is a d-dimensional distribution function with margins $F_1, \dots, F_d$.

Throughout we generally consider a d-dimensional random vector $\boldsymbol{X}$ defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. In our presentation we draw clear distinctions between (i) the (theoretical) underlying copula $\boldsymbol{C}_{\boldsymbol{X}}$ related to the joint distribution $\boldsymbol{F}_{\boldsymbol{X}}$ of $\boldsymbol{X}$ via Sklar's theorem, (ii) the (theoretical) copula $\widehat{\boldsymbol{C}}_n^{\mathbb{E}}$ constructed by convolving the underlying joint distribution $\boldsymbol{F}_{\boldsymbol{X}}$ with the kernel $\mathbf{K}$ with bandwidth matrix H_n, and (iii) the smoothed copula $\widehat{\boldsymbol{C}}_n^{|\mathbb{X}}$ based on a sample $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ of $\boldsymbol{X}$ that is used to generate a smooth bootstrap sample $\{\boldsymbol{u}_1^*, \dots, \boldsymbol{u}_m^*\}$ of size m based on Algorithm 4.1. Figure 1 shows the schematic

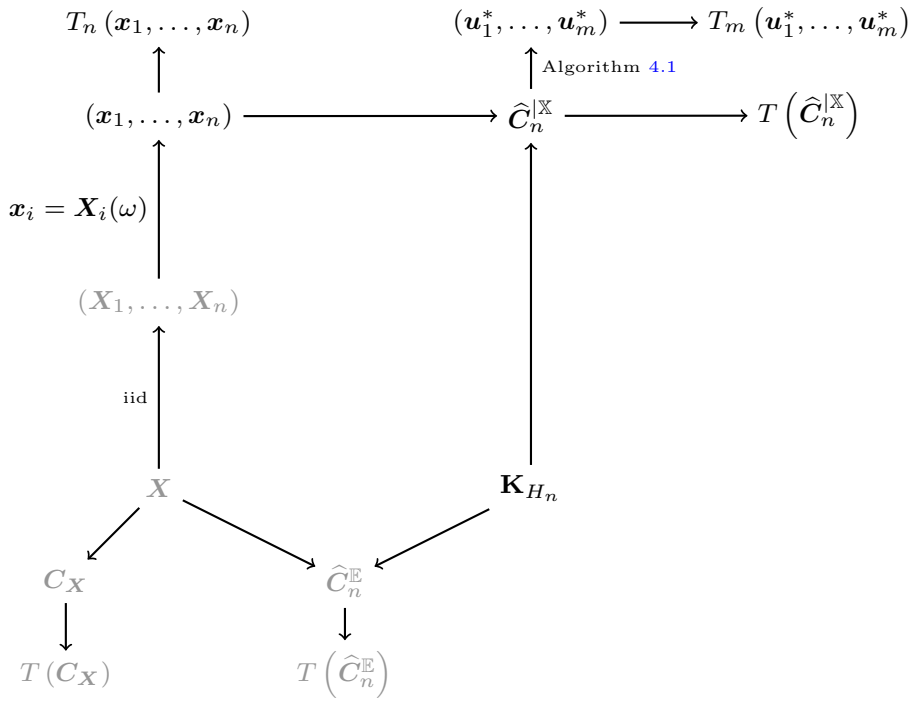

FIG 1. *Schematic relationships between the objects under consideration. Objects in gray are at the (inaccessible) population level, while objects in black are tangible to the statistician.* $\boldsymbol{C}_{\boldsymbol{X}}$ *is the theoretical copula,* $\widehat{\boldsymbol{C}}_n^{\mathbb{E}}$ *the theoretical copula constructed by convolving the underlying joint distribution* $\boldsymbol{F}_{\boldsymbol{X}}$ *with the kernel* $\mathbf{K}_{H_n}$*, and* $\widehat{\boldsymbol{C}}_n^{|\mathbb{X}}$ *the smoothed copula.* T *and* T_n *denote the population version and estimated version of the considered functional, respectively.*

relationship between the considered objects; gray entries represent the inaccessible objects at the population level, black entries are observable or under the direct control of the statistician.

To understand the ideas behind the smooth bootstrap later on, we need to briefly review multivariate kernel density estimation. We follow the notation of Wand and Jones (1995); other references are Devroye and Gyorfi (1985); Silverman (1986); Härdle et al. (2012) and Scott (2015). We denote the d-dimensional identity matrix by $\boldsymbol{I}_d$. Vectors are generally understood as column vectors, and $^\top$ is used to denote the transpose when required.

The key idea in kernel density estimation is to smooth out the empirical mass function by means of a kernel which is defined as follows.

Definition 2.1 (Kernel)**.** *A function* $\mathbf{k}$ *is called a* d-dimensional kernel*, if*

1. $\mathbf{k}$ *is the density function of an absolutely continuous* d*-dimensional random vector* $\boldsymbol{Y}$*, i.e.,* $\mathbf{k}(\boldsymbol{x}) \geq 0$ *for all* $\boldsymbol{x} \in \mathbb{R}^d$*,* $\mathbf{k}$ *is integrable and* $\int_{\mathbb{R}^d} \mathbf{k}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = 1$*;*

2. $\mathbf{k}$ *has zero mean, i.e.,* $\mathbb{E}[\boldsymbol{Y}] = \int_{\mathbb{R}^d} \boldsymbol{x}\mathbf{k}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \mathbf{0}$*; and*
3. $\mathbf{k}$ *has uncorrelated components with finite and identical marginal second moments, i.e.,* $\operatorname{cov}[\boldsymbol{Y}] = \int_{\mathbb{R}^d} \boldsymbol{x}\boldsymbol{x}^\top \mathbf{k}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \mu_{2,\mathbf{k}}\boldsymbol{I}_d$*, where the marginal second moment* $\mu_{2,\mathbf{k}} = \int_{\mathbb{R}^d} x_j^2 \mathbf{k}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$ *is equal for each* j*.*

More general definitions of kernels (such as higher order kernels) are available in the literature, but are not considered here.

For a given kernel $\mathbf{k}$ we define a rescaled version via

$$\mathbf{k}_H(\boldsymbol{x}) = \frac{1}{\sqrt{\det(H)}}\mathbf{k}\left(H^{-\frac{1}{2}}\boldsymbol{x}\right), \quad \boldsymbol{x} \in \mathbb{R}^d, \tag{2}$$

where H is a symmetric positive definite (spd) matrix called the *bandwidth matrix*. Accordingly, $\mathbf{K}$ and $\mathbf{K}_H$ denote the corresponding distribution functions of the kernel and rescaled kernel. Note that if $\boldsymbol{Y} \sim \mathbf{K}$, then the density of $\boldsymbol{Y}_H = H^{1/2}\boldsymbol{Y}$ is $\mathbf{k}_H$. Based on the properties of $\mathbf{k}$ it is then straightforward to see that $\mathbf{k}_H$ has zero mean, i.e., $\mathbb{E}[\boldsymbol{Y}_H] = \mathbf{0}$, and covariance matrix $\operatorname{cov}[\boldsymbol{Y}_H] = \mu_{2,\mathbf{k}}H$.

For a random sample $\mathbb{X} = (\boldsymbol{X}_i)_{i=1}^n$ from a d-dimensional random vector $\boldsymbol{X}$, the kernel density estimate of the underlying d-dimensional density $\boldsymbol{f}_{\boldsymbol{X}}$ is defined by

$$\widehat{\boldsymbol{f}}_n(\boldsymbol{x}) = \widehat{\boldsymbol{f}}_n(\boldsymbol{x}; \mathbb{X}) = \frac{1}{n}\sum_{i=1}^n \mathbf{k}_{H_n}(\boldsymbol{x} - \boldsymbol{X}_i), \tag{3}$$

where the bandwidth matrix H_n typically only depends on the sample size n. However, it is possible to consider a local bandwidth matrix $H_n(\boldsymbol{x})$ that (possibly) changes with the evaluation point $\boldsymbol{x} \in \mathbb{R}^d$. In a practical setting, H_n can also depend on $\mathbb{X}$, for example, when considering a scaled version of the empirical covariance matrix; we discuss the selection of H_n for our problem setting in Section 4.2. The estimate $\widehat{\boldsymbol{F}}_n$ of the corresponding joint distribution function $\boldsymbol{F}_{\boldsymbol{X}}$ is thus given by

$$\begin{aligned}\widehat{\boldsymbol{F}}_n(\boldsymbol{x}) = \widehat{\boldsymbol{F}}_n(\boldsymbol{x}; \mathbb{X}) &= \int_{(-\boldsymbol{\infty}, \boldsymbol{x}]} \widehat{\boldsymbol{f}}_n(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} = \frac{1}{n}\sum_{i=1}^n \int_{(-\boldsymbol{\infty}, \boldsymbol{x}]} \mathbf{k}_{H_n}(\boldsymbol{y} - \boldsymbol{X}_i)\,\mathrm{d}\boldsymbol{y} \\ &= \frac{1}{n}\sum_{i=1}^n \mathbf{K}_{H_n}(\boldsymbol{x} - \boldsymbol{X}_i),\end{aligned}$$

where $(-\boldsymbol{\infty}, \boldsymbol{x}] = \times_{j=1}^d (-\infty, x_j]$. Note that by Sklar's Theorem, see Sklar (1959), there is a unique (random) copula associated to $\widehat{\boldsymbol{F}}_n$ which is given by

$$\widehat{\boldsymbol{C}}_n(\boldsymbol{u}) = \widehat{\boldsymbol{F}}_n\left(\widehat{F}_{n1}^{-1}(u_1), \ldots, \widehat{F}_{nd}^{-1}(u_d)\right), \tag{4}$$

where $\widehat{F}_{n1}^{-1}, \ldots, \widehat{F}_{nd}^{-1}$ are the marginal quantile functions associated to $\widehat{\boldsymbol{F}}_n$.

While (2) provides a direct relationship to evaluate $\mathbf{k}_H$ in terms of $\mathbf{k}$, there is in general no such relationship between $\mathbf{K}_H$ and $\mathbf{K}$. For certain kernel distributions $\mathbf{K}$, such as the multivariate normal distribution or, more generally,

elliptical distributions, the corresponding rescaled distribution function $\mathbf{K}_H$ can be identified and is (at least numerically) accessible. Special cases, such as diagonal bandwidth matrices and product kernels, allow in general for a direct evaluation of $\mathbf{K}_H$ in terms of $\mathbf{K}$. In the case of product kernels and a strictly stationary and geometrically strongly mixing data generating process, Liu and Yang (2008) derive pointwise first and second order asymptotics of $\widehat{\boldsymbol{F}}_n$.

The kernel density estimate in (3) can also be understood from the viewpoint of convolutions. The *convolution* $\boldsymbol{f} * \boldsymbol{g}$, see Rudin 1991, p. 170, of two d-dimensional functions $\boldsymbol{f}, \boldsymbol{g} \in L^1(\mathbb{R}^d)$ at $\boldsymbol{x} \in \mathbb{R}^d$ is defined by

$$(\boldsymbol{f} * \boldsymbol{g})(\boldsymbol{x}) = \int_{\mathbb{R}^d} \boldsymbol{f}(\boldsymbol{x} - \boldsymbol{y}) \boldsymbol{g}(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y}. \tag{5}$$

If $\boldsymbol{f}$ and $\boldsymbol{g}$ are density functions, the convolution also has a probabilistic interpretation. When two independent random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ have densities $\boldsymbol{f}$ and $\boldsymbol{g}$, the density $\boldsymbol{f}_{\boldsymbol{X}+\boldsymbol{Y}}$ of $\boldsymbol{X} + \boldsymbol{Y}$ is given by $\boldsymbol{f}_{\boldsymbol{X}+\boldsymbol{Y}} = \boldsymbol{f} * \boldsymbol{g}$. To also cover probability mass functions in $\boldsymbol{f} * \boldsymbol{g}$ one can extend the convolution formula in (5) in terms of a Lebesgue-Stieltjes integral to $(\boldsymbol{f} * \boldsymbol{g})(\boldsymbol{x}) = \int_{\mathbb{R}^d} \boldsymbol{f}(\boldsymbol{x} - \boldsymbol{y}) \, \mathrm{d}\boldsymbol{G}(\boldsymbol{y})$ when $\boldsymbol{g}$ is a probability mass function. The kernel density estimator in (3) can then be seen as the convolution of the (rescaled) kernel density $\mathbf{k}_{H_n}$ with the point masses $\sum_{i=1}^n \frac{1}{n} \delta_{\boldsymbol{X}_i}$ belonging to the empirical distribution function; here $\delta_{\boldsymbol{a}}$ denotes a point mass at $\boldsymbol{a} \in \mathbb{R}^d$, i.e., $\delta_{\boldsymbol{a}}(\boldsymbol{x}) = 1$ if $\boldsymbol{x} = \boldsymbol{a}$ and zero otherwise. For specific observations $\{\boldsymbol{x}_i\}_{i=1}^n$, so realizations $\boldsymbol{x}_i = \boldsymbol{X}_i(\omega)$ for some $\omega \in \Omega$ (or, equivalently, given $\mathbb{X}$), the kernel density estimate is a mixture density where the ith mixing density has mean $\boldsymbol{x}_i$ and is given by $\mathbf{k}_{H_n}(\boldsymbol{x} - \boldsymbol{x}_i)$ while the mixing weights are $\frac{1}{n}$.

The bias and variance of $\widehat{\boldsymbol{f}}_n$ can be derived under additional assumptions on $\mathbf{k}$, H_n and $\boldsymbol{f}_{\boldsymbol{X}}$; see Wand and Jones 1995, Chapter 4.3. To state the results, we denote by $\operatorname{tr}(H) = \sum_{j=1}^d H_{jj}$ the trace of a $d \times d$ matrix H. Furthermore, for a twice continuously differentiable function $f \colon \mathbb{R}^d \to \mathbb{R}$, the Hessian matrix of second order partial derivatives at $\boldsymbol{x} \in \mathbb{R}^d$ is denoted by $\mathcal{H}_{\boldsymbol{x}}(f) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{x}) \right)_{i,j=1,\ldots,d}$. To simplify limits we utilize the Landau symbols o and O. Specifically, we assume the following conditions to hold:

1. $\|\mathbf{k}\|_2^2 = \int_{\mathbb{R}^d} \mathbf{k}^2(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} < \infty$;
2. each entry of the Hesse matrix $\mathcal{H}_{\boldsymbol{x}}(\boldsymbol{f}_{\boldsymbol{X}})$ is piecewise continuous and square integrable;
3. $(H_n)_{n \geq 1}$ is a sequence of bandwidth matrices such that $1/(n\sqrt{\det(H_n)})$ and all entries of H_n approach zero as $n \to \infty$;
4. the ratio of the largest and smallest eigenvalue of H_n is bounded for all n; and
5. $\mathbf{k}$ is a bounded and compactly supported d-dimensional kernel.

Under these conditions the bias and variance of the kernel density estimate $\widehat{\boldsymbol{f}}_n$ can be computed as a function of the sample size n and the function $\boldsymbol{f}_{\boldsymbol{X}}$ via

$$\mathbb{E}\left[\widehat{\boldsymbol{f}}_n(\boldsymbol{x})\right] - \boldsymbol{f}_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{2} \mu_{2,\mathbf{k}} \operatorname{tr}\left(H_n \mathcal{H}_{\boldsymbol{x}}(\boldsymbol{f}_{\boldsymbol{X}})\right) + o(\operatorname{tr}(H_n))$$

$$\approx \frac{1}{2}\mu_{2,\mathbf{k}} \operatorname{tr}\left(H_n \mathcal{H}_{\boldsymbol{x}}(\boldsymbol{f}_{\boldsymbol{X}})\right),$$

$$\operatorname{var}\left[\widehat{\boldsymbol{f}}_n(\boldsymbol{x})\right] = \frac{1}{n\sqrt{\det(H_n)}} \|\mathbf{k}\|_2^2 \, \boldsymbol{f}_{\boldsymbol{X}}(\boldsymbol{x}) + o\left(\frac{1}{n\sqrt{\det(H_n)}}\right)$$

$$\approx \frac{1}{n\sqrt{\det(H_n)}} \|\mathbf{k}\|_2^2 \, \boldsymbol{f}_{\boldsymbol{X}}(\boldsymbol{x}).$$

Furthermore, under the above conditions the estimate is consistent at any fixed point $\boldsymbol{x}$, that is, $\widehat{\boldsymbol{f}}_n(\boldsymbol{x}) \xrightarrow{\mathbb{P}} \boldsymbol{f}_{\boldsymbol{X}}(\boldsymbol{x})$ for $n \to \infty$.

In the following section we discuss kernel smoothing and the dependence distortion it (possibly) introduces. This will be important for the subsequent analysis of the smooth bootstrap.

3. Kernel smoothing and dependence distortion

This section investigates the effects of kernel smoothing on the dependence structure on the population level. We start by considering a fixed sample size.

3.1. Population version of kernel smoothing for a fixed sample size

To prepare the presentation of the smooth bootstrap later on we now discuss the dependence distortion introduced by kernel density estimation for a fixed sample size n. This links to the bootstrap in that the number of sample points n is fixed, but the number of (bootstrap) samples B is effectively unlimited. If we *had* access to B independent samples $(\mathbb{X}_b)_{b=1}^B$, where $\mathbb{X}_b = \{\boldsymbol{X}_{b1}, \ldots, \boldsymbol{X}_{bn}\}$ is a collection of iid random vectors with common density $\boldsymbol{f}_{\boldsymbol{X}}$, we could indeed average over our samples and obtain from the strong law of large numbers that

$$\frac{1}{B}\sum_{b=1}^{B} \widehat{\boldsymbol{f}}_n(\boldsymbol{x}; \mathbb{X}_b) \xrightarrow{\text{a.s.}} \mathbb{E}\left[\widehat{\boldsymbol{f}}_n(\boldsymbol{x})\right], \quad (B \to \infty).$$

When considering the smooth bootstrap in Section 4, bootstrap samples $\mathbb{X}_b^*$ will replace the unavailable samples $\mathbb{X}_b$. When comparing (as a function of $\boldsymbol{x}$) the expected density and distribution function estimators

$$\widehat{\boldsymbol{f}}_n^{\mathbb{E}}(\boldsymbol{x}) = \mathbb{E}\left[\widehat{\boldsymbol{f}}_n(\boldsymbol{x})\right] = \int_{\mathbb{R}^d} \mathbf{k}_{H_n}(\boldsymbol{x} - \boldsymbol{y}) \boldsymbol{f}_{\boldsymbol{X}}(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y}, \tag{6}$$

$$\widehat{\boldsymbol{F}}_n^{\mathbb{E}}(\boldsymbol{x}) = \mathbb{E}\left[\widehat{\boldsymbol{F}}_n(\boldsymbol{x})\right] = \int_{\mathbb{R}^d} \mathbf{K}_{H_n}(\boldsymbol{x} - \boldsymbol{y}) \boldsymbol{f}_{\boldsymbol{X}}(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y} \tag{7}$$

with the convolution formula (5), we see that $\widehat{\boldsymbol{f}}_n^{\mathbb{E}}$ coincides with the density of $\boldsymbol{Z} = \boldsymbol{X} + \boldsymbol{Y}_{H_n}$, where $\boldsymbol{X}$ is distributed with density $\boldsymbol{f}_{\boldsymbol{X}}$ and $\boldsymbol{Y}_{H_n}$ with density $\mathbf{k}_{H_n}$, and $\boldsymbol{X}$ and $\boldsymbol{Y}_{H_n}$ are independent. The mean of $\boldsymbol{Z}$ is therefore $\mathbb{E}[\boldsymbol{X}]$ and hence undistorted compared to $\boldsymbol{f}_{\boldsymbol{X}}$. However, the covariance matrix

is $\operatorname{cov}[\boldsymbol{Z}] = \operatorname{cov}[\boldsymbol{X}] + \operatorname{cov}[\boldsymbol{Y}_{H_n}] = \operatorname{cov}[\boldsymbol{X}] + \mu_{2,\mathbf{k}} H_n$ by independence of $\boldsymbol{X}$ and $\boldsymbol{Y}_{H_n}$ and the properties of $\mathbf{k}_{H_n}$, and thus in general not equal to $\operatorname{cov}[\boldsymbol{X}]$. The independence between $\boldsymbol{X}$ and $\boldsymbol{Y}_{H_n}$ also allows us to compute the characteristic function of $\widehat{\boldsymbol{f}}_n^{\mathbb{E}}$ as $\phi_{\widehat{\boldsymbol{f}}_n^{\mathbb{E}}}(\boldsymbol{t}) = \phi_{\boldsymbol{X}}(\boldsymbol{t})\phi_{\mathbf{k}}\left(H_n^{1/2}\boldsymbol{t}\right)$, see Appendix A where we provide further details on properties of characteristic functions.

According to Sklar's Theorem there is a unique copula associated to $\widehat{\boldsymbol{F}}_n^{\mathbb{E}}$ which is given by

$$\widehat{\boldsymbol{C}}_n^{\mathbb{E}}(\boldsymbol{u}) = \widehat{\boldsymbol{F}}_n^{\mathbb{E}}\left(\widehat{F}_{n1}^{-1}(u_1), \ldots, \widehat{F}_{nd}^{-1}(u_d)\right),$$

where $\widehat{F}_{n1}^{-1}, \ldots, \widehat{F}_{nd}^{-1}$ are the marginal quantile functions associated to $\widehat{\boldsymbol{F}}_n^{\mathbb{E}}$. It is important to notice here that $\widehat{\boldsymbol{C}}_n^{\mathbb{E}}$, just like $\widehat{\boldsymbol{f}}_n^{\mathbb{E}}$ and $\widehat{\boldsymbol{F}}_n^{\mathbb{E}}$, does not depend on the sample or the sample size n directly. However, the bandwidth matrix H_n typically depends on the sample size (and possibly the sample) which establishes an indirect connection and justifies the subscript n in the notation. Further properties of $\widehat{\boldsymbol{C}}_n^{\mathbb{E}}$ and the differences between $\widehat{\boldsymbol{C}}_n^{\mathbb{E}}$ and $\boldsymbol{C}_{\boldsymbol{X}}$ in the case of elliptical distributions are the subject of the following section.

3.2. *Population version of dependence distortion for elliptical random vectors and elliptical smoothing kernels*

Combining the observations from Section 3.1 with the properties of elliptical random vectors gives a first idea on how smoothing distorts the underlying dependence structure for the statistically important class of elliptical distributions. An elliptical random vector can be described by three components: a location vector $\boldsymbol{\mu}$, a dispersion matrix $\boldsymbol{\Sigma}$ and a characteristic generator ψ, which is a real-valued function such that $\phi(\boldsymbol{t}) = \exp\left(\mathrm{i}\boldsymbol{t}^\top\boldsymbol{\mu}\right)\psi\left(\boldsymbol{t}^\top\boldsymbol{\Sigma}\boldsymbol{t}\right)$ where ϕ denotes the multivariate characteristic function. Appendix B provides further details on elliptical random vectors. In the notation of McNeil, Frey and Embrechts 2015, Chapter 6, consider $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi_{\boldsymbol{X}})$ and $\boldsymbol{Y} \sim E_d(\mathbf{0}, \boldsymbol{I}_d, \psi_{\boldsymbol{Y}})$ (related to the kernel density $\mathbf{k}$) to be independent and elliptically distributed, where we directly have $\mu_{2,\mathbf{k}} = -2\psi'_{\boldsymbol{Y}}(0)$. In general, see Theorem B.2, we have for an elliptical random vector $\boldsymbol{Z} \sim E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi_{\boldsymbol{Z}})$ with an associated radial part $R \geq 0$ with finite second moment that $\mathbb{E}[\boldsymbol{Z}] = \boldsymbol{\mu}$ and $\operatorname{cov}[\boldsymbol{Z}] = \frac{\mathbb{E}[R^2]}{\operatorname{rank}(\boldsymbol{\Sigma})}\boldsymbol{\Sigma} = -2\psi'_{\boldsymbol{Z}}(0)\boldsymbol{\Sigma}$. However, the associated correlation matrix is independent of $\psi_{\boldsymbol{Z}}$ and given as $\operatorname{corr}[\boldsymbol{Z}] = \operatorname{diag}\left(1/\sqrt{\boldsymbol{\Sigma}_{11}}, \ldots, 1/\sqrt{\boldsymbol{\Sigma}_{dd}}\right)\boldsymbol{\Sigma}\operatorname{diag}\left(1/\sqrt{\boldsymbol{\Sigma}_{11}}, \ldots, 1/\sqrt{\boldsymbol{\Sigma}_{dd}}\right)$ so that $\operatorname{corr}[Z_i, Z_j] = \boldsymbol{\Sigma}_{ij}/\sqrt{\boldsymbol{\Sigma}_{ii}\boldsymbol{\Sigma}_{jj}}$.

Furthermore, if $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent, the distribution of $\boldsymbol{Z} = \boldsymbol{X} + \boldsymbol{Y}_{H_n}$ with density $\widehat{\boldsymbol{f}}_n^{\mathbb{E}}$ is elliptical if $H_n = c_n\boldsymbol{\Sigma}$ for some $c_n > 0$; see McNeil, Frey and Embrechts 2015, Remark 6.26. In this case $\boldsymbol{Y}_{H_n} \sim E_d(\mathbf{0}, c_n\boldsymbol{\Sigma}, \psi_{\boldsymbol{Y}})$. In Corollary B.1 we derive general conditions that imply $\boldsymbol{Z} \sim E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi_{\boldsymbol{Z}})$ in the given context, where $\psi_{\boldsymbol{Z}}(u) = \psi_{\boldsymbol{X}}(u)\psi_{\boldsymbol{Y}}(c_n u)$. This shows that if $H_n = c_n\boldsymbol{\Sigma}$, then, on average (i.e., when considering the expected density estimate $\widehat{\boldsymbol{f}}_n^{\mathbb{E}}$), the distortion introduced by kernel smoothing affects the rescaling of the characteristic generator of $\boldsymbol{X}$ by a factor of $\psi_{\boldsymbol{Y}}(c_n u)$ for every $u \geq 0$.

While the specific choice $H_n = c_n\boldsymbol{\Sigma}$ seems limiting at first, it corresponds to the population version of the *sphering* approach commonly used in multivariate kernel density estimation; see Wand and Jones 1995, Chapter 4.6 and references therein. In this case the bandwidth matrix is determined by a one-dimensional parameter h by setting $H = h\widehat{\boldsymbol{S}}_n$. This links to our theoretical discussion since $\widehat{\boldsymbol{S}} \xrightarrow{\mathbb{P}} \operatorname{cov}[\boldsymbol{X}]$ and hence $H \approx -2h\psi'(0)\boldsymbol{\Sigma}$ for elliptical random vectors. As we will see, in the context of elliptical random vectors, sphering will allow us to derive theoretical results without being limited to product kernels or diagonal bandwidth matrices. We return to a general discussion of bandwidth matrices in Section 4.2.

Furthermore, since $\boldsymbol{Z} \sim E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi_{\boldsymbol{Z}})$, the covariance matrix of $\boldsymbol{Z}$ is given by

$$\operatorname{cov}[\boldsymbol{Z}] = -2\psi'_{\boldsymbol{Z}}(0)\boldsymbol{\Sigma} = \left(1 + c_n \frac{\psi'_{\boldsymbol{Y}}(0)}{\psi'_{\boldsymbol{X}}(0)}\right) \operatorname{cov}[\boldsymbol{X}], \tag{8}$$

where $\operatorname{cov}[\boldsymbol{X}] = -2\psi'_{\boldsymbol{X}}(0)\boldsymbol{\Sigma}$. As a consequence of $\operatorname{cov}[\boldsymbol{Z}]$ being a re-scaling of $\operatorname{cov}[\boldsymbol{X}]$, the correlation matrix of $\boldsymbol{Z}$ is thus

$$\operatorname{corr}[\boldsymbol{Z}] = \operatorname{corr}[\boldsymbol{X}]$$

and hence, although the covariance matrix is distorted in general, the correlation matrix is not. Note that the jth margin of $\boldsymbol{X}$ is $X_j \sim E_1(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_{jj}, \psi_{\boldsymbol{X}})$, see (34), so the margins of $\boldsymbol{Z}$ are $Z_j \sim E_1(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_{jj}, \psi_{\boldsymbol{Z}})$, i.e., their characteristic generators have also been altered. While we still have that $\mathbb{E}[Z_j] = \boldsymbol{\mu}_j$, the variance changes to $\operatorname{var}[Z_j] = -2\psi'_{\boldsymbol{Z}}(0)\boldsymbol{\Sigma}_{jj} = \left(1 + c_n \frac{\psi'_{\boldsymbol{Y}}(0)}{\psi'_{\boldsymbol{X}}(0)}\right) \operatorname{var}[X_j]$ and so do the marginal quantile functions when going from $\boldsymbol{X}$ to $\boldsymbol{Z}$.

In summary, the correlation structure remains unchanged when going from $\boldsymbol{X}$ to $\boldsymbol{Z}$, but the altered characteristic generator affects the marginal distributions. Differences in the resulting elliptical copulas, see Definition B.3, are hence not due to the respective correlation matrices (which enter the copulas as parameters), but are solely due to differences of the characteristic generators, i.e., the distributions of the underlying radial parts.

Contrasting the copula of $\boldsymbol{Z}$ given by $\widehat{\boldsymbol{C}}_n^{\mathbb{E}} = \boldsymbol{C}_{\operatorname{corr}[\boldsymbol{Z}],\psi_{\boldsymbol{Z}}}$ with the original copula of interest $\boldsymbol{C}_{\boldsymbol{X}} = \boldsymbol{C}_{\operatorname{corr}[\boldsymbol{X}],\psi_{\boldsymbol{X}}}$, $\operatorname{corr}[\boldsymbol{Z}] = \operatorname{corr}[\boldsymbol{X}]$ implies that the only difference between $\widehat{\boldsymbol{C}}_n^{\mathbb{E}}$ associated to the average kernel density estimate based on samples of size n and the copula $\boldsymbol{C}_{\boldsymbol{X}}$ underlying the data generating process is due to the difference in the characteristic generators $\psi_{\boldsymbol{Z}}$ versus $\psi_{\boldsymbol{X}}$. At this point it is worth noticing that $\lim_{c_n \to 0} \psi_{\boldsymbol{Y}}(c_n u) = 1$ for all $u \geq 0$ due to uniform continuity of characteristic functions, see Sasvári 2013, Theorem 1.1.2, and thus for $c_n \to 0$ we have that $\psi_{\boldsymbol{Z}}(u) \to \psi_{\boldsymbol{X}}(u)$ for all points $u \geq 0$. Generally this implies, even in the absence of estimation of $\psi_{\boldsymbol{X}}$ and $\operatorname{corr}[\boldsymbol{X}]$, that the average estimated density $\widehat{\boldsymbol{f}}_n^{\mathbb{E}}$ has a different copula $\widehat{\boldsymbol{C}}_n^{\mathbb{E}}$ than the original sample $\boldsymbol{C}_{\boldsymbol{X}}$. Interestingly this is not always the case and we discuss conditions for such exceptions and examples in Section 3.4. Furthermore, when $\widehat{\boldsymbol{C}}_n^{\mathbb{E}}$ is used to estimate functionals of $\boldsymbol{C}_{\boldsymbol{X}}$, the distortion introduced may or may not affect the

result depending on the functional under consideration. We illustrate this point by considering a number of popular copula functionals and their properties in the case of elliptical distributions in the following paragraphs.

In the case of absolutely continuous meta-elliptical distributions, Kendall's tau does not depend on the characteristic generator, but solely on the entries of the dispersion matrix $\boldsymbol{\Sigma}$; see Lindskog, McNeil and Schmock (2003). As shown in Schmid and Schmidt 2007, Proposition 8, the same is true for Blomqvist's beta which equals Kendall's tau for elliptical models. On the population level in the elliptical class these measures of association hence coincide surprisingly for the underlying random vector $\boldsymbol{X}$ and the smooth version $\boldsymbol{Z}$ as long as $H_n = c_n\boldsymbol{\Sigma}$.

However, Spearman's rho may depend on the density generator of the elliptical density; see Abdous, Genest and Rémillard (2005) and Hult and Lindskog (2002) for an example showing that Spearman's rho is not invariant among (meta) elliptical models with a common correlation structure. For absolutely continuous multivariate elliptical random vectors the corresponding density generator provides a simple description of the multivariate density in terms of a univariate function, see Theorem B.5 for the details. The density generator in turn depends on the associated characteristic generator ψ as the characteristic function fully describes the joint distribution. The same is true for the tail dependence coefficient of elliptical distributions for which Schmidt (2002) shows that it depends on the regular variation property of the density generator. As expected, smoothing thus leads to different values of these functionals for the underlying random vector $\boldsymbol{X}$ and the smooth version $\boldsymbol{Z}$ even if $H_n = c_n\boldsymbol{\Sigma}$.

For spherical distributions it is possible to give a precise condition when a test statistic will be invariant under changes of the underlying spherical distribution. This complements the earlier discussion on the invariance of certain functionals of elliptical copulas and it potentially opens an alternative route of investigation.

Theorem 3.1 (Fang, Kotz and Ng 1990, Theorem 2.22)**.** *The distribution of a statistic $T(\boldsymbol{X})$ remains unchanged as long as $\boldsymbol{X} \sim E_d(\mathbf{0}, \boldsymbol{I}_d, \psi_{\boldsymbol{X}})$ with $\mathbb{P}[\boldsymbol{X} = \mathbf{0}] = 0$, provided that*

$$T(\alpha\boldsymbol{X}) \overset{d}{=} T(\boldsymbol{X})$$

for each $\alpha > 0$. In this case $T(\boldsymbol{X}) \overset{d}{=} T(\boldsymbol{Y})$ where $\boldsymbol{Y} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_d)$.

Next, we investigate the convergence of the characteristic generator and the characteristic function.

3.3. Convergence rates of $\psi_{\boldsymbol{Z}}(u) \to \psi_{\boldsymbol{X}}(u)$ and $\phi_{\boldsymbol{Z}}(t) \to \phi_{\boldsymbol{X}}(t)$

In this section, we investigate the difference between the characteristic generator $\psi_{\boldsymbol{Z}}(u) = \psi_{\boldsymbol{X}}(u)\psi_{\boldsymbol{Y}}(c_n u)$ and the original $\psi_{\boldsymbol{X}}$ and the impact on the difference between the respective characteristic functions. The difference between $\psi_{\boldsymbol{Z}}$ and $\psi_{\boldsymbol{X}}$ only depends on the convergence rate $\psi_{\boldsymbol{Y}}(c_n u) \to 1$ as $c_n \to 0$, since

$$|\psi_{\boldsymbol{X}}(u) - \psi_{\boldsymbol{Z}}(u)| = |\psi_{\boldsymbol{X}}(u)|\,(1 - \psi_{\boldsymbol{Y}}(c_n u)) \leq 1 - \psi_{\boldsymbol{Y}}(c_n u).$$

The equality as well as the inequality hold since $|\psi(z)| \leq 1$ for all $z \in [0, \infty)$ and all characteristic generators ψ. Also the relative error hence conveniently takes the form

$$\left| \frac{\psi_{\boldsymbol{X}}(u) - \psi_{\boldsymbol{Z}}(u)}{\psi_{\boldsymbol{X}}(u)} \right| = |1 - \psi_{\boldsymbol{Y}}(c_n u)| = 1 - \psi_{\boldsymbol{Y}}(c_n u). \tag{9}$$

For certain generators $\psi_{\boldsymbol{Y}}$ this rate may be slower than for others. For example, in case of a Gaussian kernel with $\psi_{\boldsymbol{Y}}(u) = \exp(-u/2)$ or a Laplace kernel with $\psi_{\boldsymbol{Y}}(u) = (1 + u/2)^{-1}$ we have in both cases

$$1 - \psi_{\boldsymbol{Y}}(c_n u) = c_n \frac{u}{2} + O\left(c_n^2\right), \quad c_n \to 0.$$

For the Cauchy distribution with $\psi_{\boldsymbol{Y}}(u) = \exp\left(-\sqrt{u}\right)$ we have

$$1 - \psi_{\boldsymbol{Y}}(c_n u) = \sqrt{c_n}\sqrt{u} + O\left(c_n\right), \quad c_n \to 0.$$

The convergence rate of the relative error in (9), that simultaneously acts as an upper bound for the absolute error, is hence slower in the latter case. For a given characteristic generator $\psi_{\boldsymbol{Y}}$ the behaviour of $1 - \psi_{\boldsymbol{Y}}(y)$ for $y \to 0$ is analyzed in Bingham (1972). It is connected to the behavior of the corresponding radial distribution as follows. Denote by F_R the distribution function of the radial part corresponding to the spherical distribution $E_d(\mathbf{0}, \boldsymbol{I}_d, \psi_{\boldsymbol{Y}})$. For $0 < \alpha < 2$ we then have that

$$1 - \psi_{\boldsymbol{Y}}(y) \sim y^{\alpha/2} L\left(1/\sqrt{y}\right), \qquad (y \to 0), \tag{10}$$

for a function L varying slowly at infinity, if and only if

$$1 - F_R(y) \sim \frac{L(y)}{y^{\alpha}} \frac{2^{\alpha}\Gamma\left((d+\alpha)/2\right)}{\Gamma\left(d/2\right)\Gamma\left(1-\alpha/2\right)}, \qquad (y \to \infty); \tag{11}$$

see Bingham (1972), where one can also find a discussion of the cases $\alpha = 0$ and $\alpha \geq 2$. It is important to note that we adapted the result in Bingham (1972) to our convention concerning the characteristic generators, whereas Bingham (1972) works with comparable functions $\widetilde{\psi}_{\boldsymbol{Y}}$ such that $\widetilde{\psi}_{\boldsymbol{Y}}\left(\|\boldsymbol{t}\|_2\right) = \mathbb{E}\left[\exp\left(\mathrm{i}\boldsymbol{t}^\top \boldsymbol{Y}\right)\right] = \phi_{\boldsymbol{Y}}(\boldsymbol{t})$. However, neither convention impacts the role of the radial distribution F_R. The convergence rate of the relative error in (9) representing the discrepancy between the dependence structures is therefore directly linked to the asymptotic behavior of the survival function $1 - F_R$ of the radial distribution of $\boldsymbol{Y}$ representing the kernel.

We can connect the previous discussion to the absolute difference between the characteristic functions of the initial random vector $\boldsymbol{X}$ and the smoothed version $\boldsymbol{Z}$ as follows. With $\left|e^{i\boldsymbol{\mu}^\top \boldsymbol{t}}\right| = 1$ we have

$$\begin{aligned} |\phi_{\boldsymbol{X}}(\boldsymbol{t}) - \phi_{\boldsymbol{Z}}(\boldsymbol{t})| &= \left| e^{i\boldsymbol{\mu}^\top \boldsymbol{t}} \psi_{\boldsymbol{X}}\left(\boldsymbol{t}^\top \boldsymbol{\Sigma} \boldsymbol{t}\right) - e^{i\boldsymbol{\mu}^\top \boldsymbol{t}} \psi_{\boldsymbol{Z}}\left(\boldsymbol{t}^\top \boldsymbol{\Sigma} \boldsymbol{t}\right) \right| \\ &= \left| \psi_{\boldsymbol{X}}\left(\boldsymbol{t}^\top \boldsymbol{\Sigma} \boldsymbol{t}\right) - \psi_{\boldsymbol{Z}}\left(\boldsymbol{t}^\top \boldsymbol{\Sigma} \boldsymbol{t}\right) \right| \leq 1 - \psi_{\boldsymbol{Y}}\left(c_n \boldsymbol{t}^\top \boldsymbol{\Sigma} \boldsymbol{t}\right), \end{aligned}$$

implying pointwise convergence for $c_n \to 0$. We also have a bound on the relative error given by

$$\left|\frac{\phi_{\boldsymbol{X}}(\boldsymbol{t}) - \phi_{\boldsymbol{Z}}(\boldsymbol{t})}{\phi_{\boldsymbol{X}}(\boldsymbol{t})}\right| \leq 1 - \psi_{\boldsymbol{Y}}\left(c_n \boldsymbol{t}^\top \boldsymbol{\Sigma} \boldsymbol{t}\right).$$

The absolute difference $|\phi_{\boldsymbol{X}}(\boldsymbol{t}) - \phi_{\boldsymbol{Z}}(\boldsymbol{t})|$ can be used to construct an upper bound on the uniform distance $\|\boldsymbol{F}_{\boldsymbol{X}} - \boldsymbol{F}_{\boldsymbol{Z}}\|_\infty$ via the smoothing inequality and its higher dimensional analogues, see Theorem A.3 in Appendix A. For ease of presentation we proceed with a univariate example.

Example 3.1. *For simplicity, we discuss this approach in the one-dimensional case where a Laplace kernel with $\psi_Y(u) = (1 + u/2)^{-1}$ is used to smooth a one dimensional elliptical random variable $X \sim E_1(\mu, \sigma^2, \psi_X)$. Combining the fact that*

$$|\phi_X(t) - \phi_Z(t)| \leq 1 - \psi_Y\left(c_n \sigma^2 t^2\right) = \frac{c_n \sigma^2 t^2/2}{c_n \sigma^2 t^2/2 + 1}$$

with the smoothing inequality we have for $T > 0$ that

$$\begin{aligned}
\|F_X - F_Z\|_\infty &\leq \frac{1}{\pi}\int_{-T}^{T}\left|\frac{\phi_X(t) - \phi_Z(t)}{t}\right| \mathrm{d}t + \frac{24}{\pi T}\sup_{x\in\mathbb{R}}|f_Z(x)| \\
&\leq \frac{2}{\pi}\int_0^T \frac{1 - \psi_Y(c_n\sigma^2 t^2)}{t}\,\mathrm{d}t + \frac{24}{\pi T}\sup_{x\in\mathbb{R}}|f_Z(x)| \\
&= \frac{1}{\pi}\int_0^T \frac{c_n\sigma^2 t}{c_n\sigma^2 t^2/2 + 1}\,\mathrm{d}t + \frac{24}{\pi T}\sup_{x\in\mathbb{R}}|f_Z(x)| \\
&= \frac{\log(c_n\sigma^2 T^2/2 + 1)}{\pi} + \frac{24}{\pi T}\sup_{x\in\mathbb{R}}|f_Z(x)| \\
&\leq \frac{\log(c_n\sigma^2 T^2/2 + 1)}{\pi} + \frac{M}{\pi T}, \qquad (12)
\end{aligned}$$

where M is such that $24\sup_{x\in\mathbb{R}}|f_Z(x)| \leq M$. Given that f_Z is the convolution of the kernel density f_Y with f_X, a simple upper bound independent of c_n is given by $M = 24\sup_{x\in\mathbb{R}}|f_X(x)|$.

Optimizing the bound in (12) with respect to T we obtain the first order condition $c_n\sigma^2 T^3 - M c_n \sigma^2 T^2/2 - M = 0$ with the unique (real) solution

$$\begin{aligned}
T^* &= \frac{1}{6}\left(\frac{c_n\sigma^2 M^2}{a^{1/3}} + \frac{a^{1/3}}{c_n\sigma^2} + M\right), \quad \text{where} \\
a &= c_n^3\sigma^6 M^3 + 108 c_n^2\sigma^4 M + 6\sqrt{6}\sqrt{c_n^5\sigma^{10}M^4 + 54 c_n^4\sigma^8 M^2}.
\end{aligned}$$

Given that T^* is the unique stationary point and $\lim_{T\to 0}\frac{\log(c_n\sigma^2T^2/2+1)}{\pi} + \frac{M}{\pi T} = \lim_{T\to\infty}\frac{\log(c_n\sigma^2T^2/2+1)}{\pi} + \frac{M}{\pi T} = \infty$ it is clear that T^* leads to the minimal upper bound. Considering $T_n = T^*(c_n)$ as a function of c_n we have $T_n = O\left(c_n^{-1/3}\right)$

for $c_n \to 0$ *and hence* $c_n T_n^2 = O\left(c_n^{1/3}\right)$ *which leads to*

$$\|F_X - F_Z\|_\infty = \left\|F_X - F_{X+Y_{c_n\sigma^2}}\right\|_\infty = O\left(c_n^{1/3}\right).$$

It is important to note that the rate $O\left(c_n^{1/3}\right)$ *is universal in the sense that it only depends on the kernel distribution* Y *and holds simultaneously for all absolutely continuous univariate elliptical random variables* $X \sim E_1(\mu, \sigma^2, \psi_X)$ *with a bounded density. The dependence on* X *is in fact only visible in the chosen bound* M*.*

In the next section, we provide cases for which the dependence structure remains unaffected by smoothing at the population level.

3.4. Exceptions to dependence distortions at the population level

Section 3.2 has established how smoothing impacts the dependence structure for elliptical models and kernels at the population level. In this section we give conditions under which these effects surprisingly do not impact the dependence structure. As can be expected, these examples are rather artificial but serve the purpose of establishing a comprehensive view of dependence distortion in kernel density estimation. To complete the picture we also discuss in Section 3.4.4 and 3.4.5 examples for which the effects of dependence distortion are present and can be worked out in detail.

3.4.1. Multivariate normal distribution

We start by considering the Gaussian random vector $\boldsymbol{X}$, assume that the smoothing random vector $\boldsymbol{Y}$ is also Gaussian and that the bandwidth matrix $H_n = c_n\boldsymbol{\Sigma}$ is a rescaled version of the dispersion matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{X}$. In this case, see Fang, Kotz and Ng 1990, Example 2.3, page 28, the characteristic generators of $\boldsymbol{X}$ and $\boldsymbol{Y}$ are given by $\psi_{\boldsymbol{X}}(u) = \psi_{\boldsymbol{Y}}(u) = \exp(-u/2)$, and hence we can compute the characteristic generator of the expected density $\widehat{\boldsymbol{f}}_n^{\mathbb{E}}$ of the smoothed random vector $\boldsymbol{Z} = \boldsymbol{X} + \boldsymbol{Y}$ since

$$\psi_{\boldsymbol{Z}}(u) = \psi_{\boldsymbol{X}}(u)\psi_{\boldsymbol{Y}}(c_n u) = e^{-u/2}e^{-c_n u/2} = e^{-(1+c_n)u/2} = \psi_{\boldsymbol{X}}((1+c_n)u).$$

This allows us to represent $\boldsymbol{Z}$ in two different ways, namely as $\boldsymbol{Z} \sim E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi_{\boldsymbol{Z}})$ as in the previous section, and as $\boldsymbol{Z} \sim E_d(\boldsymbol{\mu}, (1+c_n)\boldsymbol{\Sigma}, \psi_{\boldsymbol{X}})$ by invoking a reparameterization; see Remark B.1. Recalling from (8) that $\operatorname{corr}[\boldsymbol{Z}] = \operatorname{corr}[\boldsymbol{X}]$, the second parameterization in fact yields that

$$\widehat{\boldsymbol{C}}_n^{\mathbb{E}} = \boldsymbol{C}_{\operatorname{corr}[\boldsymbol{Z}],\psi_{\boldsymbol{Z}}} = \boldsymbol{C}_{\operatorname{corr}[\boldsymbol{X}],\psi_{\boldsymbol{X}}} = \boldsymbol{C}_{\boldsymbol{X}}.$$

This is due to the specific properties of the characteristic generator that allows us to shift the effects of multiplying the two characteristic generators $\psi_{\boldsymbol{X}}$ and

$\psi_{\boldsymbol{Y}}$ into a rescaling of the dispersion matrix. This rescaling in turn gets lost when considering the associated correlation matrix. The effect of reverting $\psi_{\boldsymbol{Z}}$ back to the standard form $\psi_{\boldsymbol{X}}$, however, is persisting.

The key property we have used to derive this result is that for all $u \geq 0$ and $\beta \geq 0$ we have

$$\psi_{\boldsymbol{X}}(u)\psi_{\boldsymbol{Y}}(\beta u) = \psi_{\boldsymbol{X}}(\gamma u) \tag{13}$$

for some $\gamma > 0$. The functional equation in (13) is reminiscent of a characterization of the exponential function stating that, for non-zero continuous functions, the property

$$\phi(x)\phi(y) = \phi(x+y) \tag{14}$$

uniquely characterizes the exponential function; see, for example, Rudin 1976, Exercise 6 of Chapter 8.

3.4.2. Multivariate elliptical stable distributions

Since the requirement in (13) is less strict than (14), it is possible to find solutions other than the normal distribution discussed above. One class of distributions that allows one to shift multiplicative scalars as in (13) is the class of multivariate elliptical stable distributions; see Nolan (2013) for an overview (note that this reference calls this class of distributions multivariate elliptically contoured stable distributions). Multivariate elliptical stable distributions are at the intersection of stable and elliptical distributions.

Definition 3.1 (Multivariate elliptical stable distributions)**.** *For $0 < \alpha \leq 2$ a d-dimensional random vector $\boldsymbol{X}$ has a* multivariate elliptical stable distribution *if its characteristic function takes the form*

$$\phi_{\boldsymbol{X}}(\boldsymbol{t}) = \exp\left(\mathrm{i}\boldsymbol{\mu}^\top \boldsymbol{t} - \left(\boldsymbol{t}^\top \boldsymbol{\Sigma} \boldsymbol{t}\right)^{\alpha/2}\right),$$

where $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$ is symmetric positive definite and $\boldsymbol{\mu} \in \mathbb{R}^d$.

The characteristic generator of a multivariate elliptical stable distribution is given by $\psi(u) = \exp\left(-u^{\alpha/2}\right)$. For $\beta \geq 0$, we thus have that

$$\psi(u)\psi(\beta u) = \exp(-(1+\beta^{\alpha/2})u^{\alpha/2}) = \psi(\gamma u)$$

with $\gamma = (1+\beta^{\alpha/2})^{2/\alpha}$. If the original random vector and the smoothing kernel both belong to the class of multivariate elliptical stable distributions with the same shape parameter α and dispersion matrices $\boldsymbol{\Sigma}$ and $c_n\boldsymbol{\Sigma}$, it is thus possible to shift the scaling factor c_n into the dispersion matrix analogously to the case of the multivariate normal distribution. This in turn leaves the underlying dependence structure undistorted, i.e., $\widehat{\boldsymbol{C}}_n^{\mathbb{E}} = \boldsymbol{C}_{\boldsymbol{X}}$. As a consequence, functionals that only depend on the copula are unaffected when smoothing.

One application of this result are the values of the upper and lower tail dependence coefficients which are derived in Schmidt (2002) for elliptical random vectors. For multivariate elliptical stable distributions we have for $0 < \alpha \leq 2$ that

$$1 - \psi(u) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{u^{k\alpha/2}}{k!} \sim u^{\alpha/2}, \qquad (u \to 0),$$

and thus by the results of Bingham (1972) in (10) and (11) that the distribution function of the associated radial part is regularly varying with index $-\alpha$ for $0 < \alpha < 2$. Following Theorem 5.2 and (5.2) of Schmidt (2002), this implies a non-zero upper and lower tail dependence coefficient for multivariate elliptical stable distributions, except for the Gaussian case ($\alpha = 2$) where there is no tail dependence. However, since the tail dependence coefficients are functionals of the copula, our calculations show that the tail dependence will not be distorted if the kernel and bandwidth matrix are chosen appropriately.

3.4.3. Multivariate stable distributions

To complement the discussion in Section 3.4.2 we briefly touch upon the general case of multivariate stable distributions. We follow Samorodnitsky and Taqqu 1994, Chapter 2, with the following definition.

Definition 3.2 (Multivariate stable distribution)**.** *A random vector $\boldsymbol{X}$ is said to be* stable *in $\mathbb{R}^d$ if for any positive numbers $b_1 > 0$ and $b_2 > 0$ there exists a vector $\boldsymbol{d} \in \mathbb{R}^d$ such that*

$$b_1 \boldsymbol{X}_1 + b_2 \boldsymbol{X}_2 \stackrel{d}{=} (b_1^\alpha + b_2^\alpha)^{1/\alpha} \boldsymbol{X} + \boldsymbol{d}, \tag{15}$$

where $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent copies of $\boldsymbol{X}$, and $0 < \alpha \leq 2$ does not depend on b_1 and b_2. A stable random vector is called strictly stable*, if (15) holds with $\boldsymbol{d} = \boldsymbol{0}$ for any $b_1 > 0$ and $b_2 > 0$.*

It is straightforward to link Definition 3.2 to the framework of multivariate kernel smoothing by choosing a smoothing kernel $\mathbf{K} = \boldsymbol{F}_{\boldsymbol{X}}$. For $\boldsymbol{Y} \sim \mathbf{K}$ we then have for $h > 0$ that

$$\boldsymbol{Z} = \boldsymbol{X} + h\boldsymbol{Y} \sim \boldsymbol{F}_{\boldsymbol{X}} \left((1 + h^\alpha)^{-1/\alpha} (\boldsymbol{x} - \boldsymbol{d}) \right).$$

While it is straightforward to show that we have under the current assumptions $\widehat{\boldsymbol{C}}_n^{\mathbb{E}} = \boldsymbol{C}_{\boldsymbol{X}}$, the result is not of much practical use. On the one hand, if the underlying distribution function $\boldsymbol{F}_{\boldsymbol{X}}$ is known and hence can be used as the kernel distribution no estimation is necessary. On the other hand, if a smoothing kernel in the class of multivariate stable distributions is chosen, the chances that the underlying data generating process follows the same distribution are (without additional knowledge) slim.

From a theoretical point of view it is, however, noteworthy that multivariate stable distributions are presumably the largest class of distributions for which no dependence distortion occurs if the smoothing kernel is chosen appropriately.

3.4.4. Multivariate Student t distribution

For the Student t family with $\nu > 0$ degrees of freedom, the characteristic generator is derived in Sutradhar (1986). We utilize the form derived in Joarder and Ali (1996) and Song, Park and Kim (2014), given by

$$\psi_\nu(x) = \frac{K_{\nu/2}\left(\sqrt{\nu x}\right)\left(\sqrt{\nu x}\right)^{\nu/2}}{\Gamma\left(\nu/2\right)2^{\nu/2-1}},$$

where K_α denotes the modified Bessel function of the second kind (Joarder and Ali (1996) refer to it as the Mcdonald function, Song, Park and Kim (2014) and Kotz, Kozubowski and Podgórski (2001) refer to it as the modified Bessel function of the third kind) which can be represented as

$$K_\alpha(t) = \left(\frac{2}{t}\right)^\alpha \frac{\Gamma\left(\alpha + 1/2\right)}{\sqrt{\pi}} \int_0^\infty \left(1+u^2\right)^{-\left(\alpha+\frac{1}{2}\right)} \cos(tu)\,\mathrm{d}u \tag{16}$$

for $t > 0$ and $\alpha > -1/2$; see, for example, NIST, Equation 10.32.11. For parameters of the form $\alpha = r + 0.5$ with $r \in \{0, 1, 2, \ldots\}$ we have the explicit formula

$$K_\alpha(t) = \sqrt{\frac{\pi}{2t}} e^{-t} \sum_{k=0}^{r} \frac{(r+k)!}{(r-k)!k!} (2t)^{-k}, \tag{17}$$

see Kotz, Kozubowski and Podgórski 2001, Equation (A.0.10). For the special case $\alpha = 1/2$ we consequently have

$$K_{\frac{1}{2}}(t) = \sqrt{\frac{\pi}{2t}} e^{-t},$$

see also NIST, Equation 10.39.2. For the special case of the multivariate Cauchy distribution we have $\nu = 1$ and therefore obtain the characteristic generator as

$$\psi_1(u) = \frac{K_{\frac{1}{2}}\left(\sqrt{u}\right)\left(\sqrt{u}\right)^{1/2}}{\Gamma\left(\frac{1}{2}\right)2^{-\frac{1}{2}}} = \sqrt{\frac{2}{\pi}} u^{\frac{1}{4}} \sqrt{\frac{\pi}{2\sqrt{u}}} e^{-\sqrt{u}} = e^{-\sqrt{u}}.$$

If the smoothing kernel is also a multivariate Cauchy distribution we consequently have for $\beta \geq 0$ that,

$$\psi_1(u)\psi_1(\beta u) = e^{-\sqrt{u}} e^{-\sqrt{\beta u}} = e^{-\sqrt{u\left(1+\sqrt{\beta}\right)^2}} = \psi_1\left(u\left(1+\sqrt{\beta}\right)^2\right).$$

This is as expected since the Cauchy distribution is an elliptical stable distribution with index $\alpha = 1$.

For general $\nu \neq 1$ the modified Bessel function of the second kind does not reduce to the exponential function. A similar rescaling is therefore in general not possible for the Student t distribution. This implies that for $\nu \neq 1$ the

dependence structure will be distorted even when the smoothing kernel is chosen to match the distribution of the original data. For a concrete example we set $\nu = 3$ and obtain via (17) that

$$\psi_3(u) = e^{-\sqrt{3u}}\left(1 + \sqrt{3u}\right).$$

Consequently we have

$$\begin{aligned}\psi_3(u)\psi_3(\beta u) &= e^{-\sqrt{3u\left(1+\sqrt{\beta}\right)^2}}\left(1 + \sqrt{3u\left(1+\sqrt{\beta}\right)^2} + 3u\sqrt{\beta}\right)\\ &= \psi_3(\gamma u)e^{-\sqrt{3\gamma u}}3u\sqrt{\beta},\end{aligned}$$

where $\gamma = \left(1+\sqrt{\beta}\right)^2$. Hence we can identify the term $e^{-\sqrt{3\gamma u}}3u\sqrt{\beta}$ on the right as being responsible for the dependence distortion.

3.4.5. Elliptical distributions not elliptical stable

As a last example, we consider an elliptical distribution which is not elliptical stable. We focus on the multivariate Laplace distribution, in which case (13) will not hold. For this model the effect of the dependence distortion can be worked out explicitly. A general introduction to the multivariate Laplace distribution can be found in Kotz, Kozubowski and Podgórski (2001).

Denote by $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi_L)$ a d-dimensional random vector with characteristic generator

$$\psi_L(u) = \frac{1}{1+u/2}.$$

For such a random vector $\boldsymbol{X}$ the density generator is given by

$$g_L(t) = \frac{2}{(2\pi)^{d/2}}\left(\frac{t}{2}\right)^{(2-d)/4} K_{(2-d)/2}\left(\sqrt{2t}\right),$$

see Kotz, Kozubowski and Podgórski 2001, Equation (5.2.2), and $\boldsymbol{X}$ follows a multivariate Laplace distribution; see Kotz, Kozubowski and Podgórski 2001, Equations (5.2.1) and (5.2.2). The corresponding radial distribution R_L of $\boldsymbol{X}$ has the density

$$f_{R_L}(x) = \frac{2x^{d/2}K_{d/2-1}\left(x\sqrt{2}\right)}{\left(\sqrt{2}\right)^{d/2-1}\Gamma(d/2)}, \quad x > 0;$$

see Kotz, Kozubowski and Podgórski 2001, Proposition 6.3.1. In this case, the product of the characteristic generators related to kernel smoothing for $\beta \geq 0$ is given by

$$\psi_{\boldsymbol{Z}}(u) = \psi_L(u)\psi_L(\beta u) = \frac{1}{1+u/2}\frac{1}{1+\beta u/2} = \frac{1}{1+u(1+\beta)/2+\beta u^2/4},$$

where the term $\beta u^2/4$ in the denominator prevents a simplification as in the Gauss and Cauchy (and general multivariate elliptical stable) cases discussed before. In this example it is therefore clearly not possible to convert the effects of dependence distortion into a rescaling of the dispersion matrix. When trying to identify the radial distribution connected to the generator $\psi_{\boldsymbol{Z}}$ resulting from smoothing with a Laplace kernel (with an appropriate bandwidth matrix) we can use partial fraction decomposition for $\beta \neq 1$ to get

$$\psi_{\boldsymbol{Z}}(u) = \frac{2}{1-\beta}\left(\frac{1}{u+2} - \frac{\beta}{\beta u+2}\right) = \frac{1}{1-\beta}\psi_L(u) - \frac{\beta}{1-\beta}\psi_L(\beta u).$$

To identify the radial distribution connected to $\psi_{\boldsymbol{Z}}$ we can draw on the connection between characteristic generators and radial distributions; see Fang, Kotz and Ng 1990, Section 2.1. Specifically, for any spherical distribution $E_d(\mathbf{0}, \boldsymbol{I}_d, \psi)$ there exists a distribution function F_R of an a.s. positive random variable R (the radial part) such that

$$\psi(u) = \int_0^\infty \Omega_d\left(ut^2\right)\, \mathrm{d}F_R(t),$$

where Ω_d is the characteristic generator of a random vector $\boldsymbol{S}$ uniformly distributed on the unit sphere $\{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\|_2 = 1\}$, i.e., $\phi_{\boldsymbol{S}}(\boldsymbol{t}) = \mathbb{E}\left[e^{\mathrm{i}\boldsymbol{t}^\top \boldsymbol{S}}\right] = \Omega_d\left(\|\boldsymbol{t}\|_2^2\right)$, see Fang, Kotz and Ng 1990, Theorem 2.2, page 29. Concerning $\psi_{\boldsymbol{Z}}$, we consequently have that

$$\begin{aligned}
\psi_L(u) &= \int_0^\infty \Omega_d\left(ut^2\right) f_{R_L}(t)\,\mathrm{d}t,\\
\psi_L(\beta u) &= \int_0^\infty \Omega_d\left(\beta u t^2\right) f_{R_L}(t)\,\mathrm{d}t = \int_0^\infty \Omega_d\left(ut^2\right) f_{R_L}\left(t/\sqrt{\beta}\right)/\sqrt{\beta}\,\mathrm{d}t,
\end{aligned}$$

where the last equality is obtained from a substitution. Using linearity and the integral representation of ψ_L, we now have that

$$\begin{aligned}
\psi_{\boldsymbol{Z}}(u) &= \frac{1}{1-\beta}\psi_L(u) - \frac{\beta}{1-\beta}\psi_L(\beta u)\\
&= \frac{1}{1-\beta}\int_0^\infty \Omega_d\left(ut^2\right) f_{R_L}(t)\,\mathrm{d}t\\
&\quad - \frac{\beta}{1-\beta}\int_0^\infty \Omega_d\left(ut^2\right) f_{R_L}\left(t/\sqrt{\beta}\right)/\sqrt{\beta}\,\mathrm{d}t\\
&= \int_0^\infty \Omega_d\left(ut^2\right) f_{R_{\boldsymbol{Z}}}(t)\,\mathrm{d}t,
\end{aligned}$$

where the radial density $f_{R_{\boldsymbol{Z}}}$ connected to $\psi_{\boldsymbol{Z}}$ can be identified as

$$f_{R_{\boldsymbol{Z}}}(x) = \frac{1}{1-\beta}\left(f_{R_L}(x) - \sqrt{\beta} f_{R_L}\left(x/\sqrt{\beta}\right)\right), \quad x > 0. \tag{18}$$

When smoothing the considered multivariate Laplace distribution with a matching multivariate Laplace kernel and an appropriate bandwidth matrix $H = \beta\boldsymbol{\Sigma}$, the smoothing thus affects the radial distribution as shown in (18). The difference between $f_{R_{\boldsymbol{Z}}}$ and the original radial density f_{R_L} distorts the joint distribution and hence the implied copula.

This concludes our theoretical discussion and investigation of dependence structure distortion at the population level. In the next section, we link this to the smooth bootstrap.

4. Smooth bootstrap

Section 3 focuses on dependence distortion at the population level for kernel based estimators. In this section, we connect our previous results to bootstrapping. While we discuss asymptotic theory where appropriate, we mainly focus on the smooth bootstrap as a computational tool. A general introduction to the theory of the non-parametric bootstrap can be found in Hall 1992, Chapter 1, and Shao and Tu (1995).

The smooth bootstrap is motivated by, and closely related to, kernel density estimation as discussed in Section 2. To introduce the smooth bootstrap we denote by n the original sample size, while B denotes the overall number of smooth bootstrap replications. Each bootstrap sample is of size m. While $m = n$ is a typical choice, it is possible to use $m \gg n$ in a data augmentation situation. Furthermore, τ_b, $b \in \{1, \dots, B\}$, denotes a random vector uniformly distributed on $\times_{i=1}^{m}\{1, \dots, n\}$ with components τ_{bi} for $i \in \{1, \dots, m\}$ independent of the sample $\mathbb{X} = \{\boldsymbol{X}_i\}_{i=1}^{n}$. In this case the individual components of τ_b are clearly independent and uniformly distributed on $\{1, \dots, n\}$. A bootstrap sample $\mathbb{X}_b^* = (\boldsymbol{X}_{bi}^*)_{i=1}^{m}$, $b \in \{1, \dots, B\}$, is now generated via $\boldsymbol{X}_{bi}^* = \boldsymbol{X}_{\tau_{bi}}$. The draws $\tau_1, \dots, \tau_B$ are assumed to be independent which thus carries over to the (non-parametric) bootstrap samples $\{\mathbb{X}_b^*\}_{b=1}^{B}$.

The smooth bootstrap sample $\mathbb{S}_b^* = \{\boldsymbol{Z}_{bi}^*\}_{i=1}^{m}$ is obtained by setting $\boldsymbol{Z}_{bi}^* = \boldsymbol{X}_{bi}^* + \boldsymbol{Y}_{bi}$, where $\boldsymbol{Y}_{bi}$ is distributed according to a smoothing kernel $\mathbf{K}_{H_n}$ independently of $\{\boldsymbol{X}_i\}_{i=1}^{n}$ and $\{\tau_b\}_{b=1}^{B}$. This again leads to independence among all the components of $\{\mathbb{S}_b^*\}_{b=1}^{B}$. The smooth bootstrap sampling scheme can also be interpreted from the kernel smoothing perspective discussed in Sections 2 and 3. Considering the unconditional distribution of $\boldsymbol{Z}_{bi}^*$ it can readily be seen that $\boldsymbol{Z}_{bi}^* \overset{d}{=} \boldsymbol{X} + \boldsymbol{Y}$, where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent random vectors with densities $\boldsymbol{f}_{\boldsymbol{X}}$ and $\mathbf{k}_{H_n}$. The (unconditional) density of $\boldsymbol{Z}_{bi}^*$ is therefore given by $\widehat{\boldsymbol{f}}_n^{\mathbb{E}}$ defined in (6), the expected kernel density estimate, and the observations in Sections 3.1–3.4 apply accordingly.

Conditionally on $\mathbb{X}$, the smooth bootstrap is equivalent to sampling from the mixture density obtained from the kernel density estimation for a given dataset. We denote this mixture density by

$$\widehat{\boldsymbol{f}}_n^{|\mathbb{X}}(\boldsymbol{x}) = \widehat{\boldsymbol{f}}_n(\boldsymbol{x} \,|\, \mathbb{X}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{k}_{H_n}(\boldsymbol{x} - \boldsymbol{x}_i), \tag{19}$$

where $\boldsymbol{x}_i = \boldsymbol{X}_i(\omega)$, $i \in \{1, \dots, n\}$, for some fixed $\omega \in \Omega$. Along the same lines we denote the joint distribution function conditional on $\mathbb{X}$ by

$$\widehat{\boldsymbol{F}}_n^{|\mathbb{X}}(\boldsymbol{x}) = \widehat{\boldsymbol{F}}_n(\boldsymbol{x} \,|\, \mathbb{X}) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{K}_{H_n}(\boldsymbol{x} - \boldsymbol{x}_i). \tag{20}$$

In general, sampling from a mixture density proceeds in two steps. First, one of n possible groups is selected with equal probability n^{-1}. The second step consist of drawing a random vector from the corresponding mixing density, which in our case is represented by a random vector that follows $\mathbf{K}_{H_n}$ centered at the randomly selected $\boldsymbol{x}_i$ representing the ith group. Given $\mathbb{X}$, we thus equivalently have that $\boldsymbol{Z}_{bi}^* = \boldsymbol{x}_{bi}^* + \boldsymbol{Y}_{bi}$ in the smooth bootstrap, where $\boldsymbol{x}_{bi}^*$ is selected uniformly from $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ according to τ_b.

In the following section, we apply the smooth bootstrap to copula functionals and provide an algorithm that draws a smoothed sample from given observations.

4.1. The smooth bootstrap for copula functionals

In Section 3.2 we have discussed how smoothing impacts the joint distribution and dependence structure of the expected kernel density estimate $\widehat{\boldsymbol{f}}_n^{\mathbb{E}}$ with a special focus on elliptical distributions and kernels. In this section we discuss properties of $\widehat{\boldsymbol{f}}_n^{|\mathbb{X}}$ with a focus on the implied dependence structure.

If the marginal quantile functions $\widehat{F}_{n1}^{-1}, \dots, \widehat{F}_{nd}^{-1}$ associated to the joint distribution function $\widehat{\boldsymbol{F}}_n^{|\mathbb{X}}$ implied by the density $\widehat{\boldsymbol{f}}_n^{|\mathbb{X}}$ are known, Sklar's Theorem can be used to extract the corresponding copula (conditional on the data $\mathbb{X}$) via

$$\widehat{\boldsymbol{C}}_n^{|\mathbb{X}}(\boldsymbol{u}) = \widehat{\boldsymbol{F}}_n^{|\mathbb{X}}\left(\widehat{F}_{n1}^{-1}(u_1), \dots, \widehat{F}_{nd}^{-1}(u_d)\right). \tag{21}$$

However, recovering the marginal distribution and quantile functions from $\widehat{\boldsymbol{f}}_n^{|\mathbb{X}}$ defined in (19) is not an easy task in general. In the following discussion we will thus limit ourselves to kernels that are elliptical. Aside from elliptical kernels, it is straightforward to extract the marginal distributions in the case of product kernels. There, the marginal distributions are given by standard univariate kernel distribution estimates where the respective bandwidths are selected individually by virtue of a diagonal bandwidth matrix $H_n = \operatorname{diag}(h_{n1}, \dots, h_{nd})$.

If the kernel $\mathbf{k}$ is the density of an elliptical random vector $\boldsymbol{Y} \sim E_d(\mathbf{0}, \boldsymbol{I}_d, \psi_{\boldsymbol{Y}})$, the rescaled and shifted kernel $\mathbf{k}_{H_n}(\boldsymbol{x} - \boldsymbol{x}_i)$, $\boldsymbol{x} \in \mathbb{R}^d$, can be identified as the density of $\boldsymbol{Z}_i = \boldsymbol{x}_i + H_n^{1/2}\boldsymbol{Y} \sim E_d(\boldsymbol{x}_i, H_n, \psi_{\boldsymbol{Y}})$; see Theorem B.3. Considering (34), the jth marginal distribution of $\boldsymbol{Z}_i$ is hence given by $Z_{ij} \sim E_1(\boldsymbol{x}_{ij}, H_{njj}, \psi_{\boldsymbol{Y}})$. Denoting by $F_{\psi_{\boldsymbol{Y}}}$ the common univariate marginal distribution function of $\boldsymbol{Y}$, the distribution function of Z_{ij}, $j \in \{1, \dots, d\}$, is then given in terms of a location-scale model taking the form

$$F_{Z_{ij}}(z) = \mathbb{P}\left[Z_{ij} \leq z\right] = F_{\psi_{\boldsymbol{Y}}}\left(\frac{z - \boldsymbol{x}_{ij}}{\sqrt{H_{njj}}}\right), \quad z \in \mathbb{R}.$$

In this specific setup, the marginal distribution functions of $\widehat{\boldsymbol{F}}_n^{|\mathbb{X}}$ are given by mixture distribution functions of location-scale models of $F_{\psi_{\boldsymbol{Y}}}$. For the jth margin we have that

$$\widehat{F}_{nj}(x) = \frac{1}{n}\sum_{i=1}^{n} F_{\psi_{\boldsymbol{Y}}}\left(\frac{x-\boldsymbol{x}_{ij}}{\sqrt{H_{njj}}}\right), \quad x \in \mathbb{R}. \tag{22}$$

The marginal distributions of $\widehat{\boldsymbol{F}}_n^{|\mathbb{X}}$ are thus given by univariate kernel distribution estimates with bandwidth $h_{nj} = \sqrt{H_{njj}}$ and kernel density $f_{\psi_{\boldsymbol{Y}}} = F'_{\psi_{\boldsymbol{Y}}}$.

To make use of Sklar's Theorem, the corresponding quantile functions $\widehat{F}_{nj}^{-1}$ need to be available. For $p \in (0,1)$ the quantile function $\widehat{F}_{nj}^{-1}(p)$ is (here) defined via the inverse relationship

$$\widehat{F}_{nj}(x) = p. \tag{23}$$

The probabilistic behaviour of univariate quantile functions for $n \to \infty$ in this case is studied in Nadaraya (1964a,b) and Azzalini (1981). Since in our case the marginal distributions are given in closed form in (22), the corresponding quantile functions $\widehat{F}_{nj}^{-1}$ can be computed for a fixed n via numerical inversion. This leads to a tractable numerical evaluation of $\widehat{\boldsymbol{C}}_n^{|\mathbb{X}}$ for an argument $\boldsymbol{u}$ in the case of elliptical kernels.

Concerning the numerical inversion, the value of a sample quantile can serve as a starting point for numerical algorithms as suggested in Azzalini (1981). When a starting interval instead of a starting point is required for the numerical search, the value of $\widehat{F}_{nj}$ at the first and last order statistic (in component j) can be used to obtain a first estimate of the relevant search region.

In Azzalini (1981), an optimal bandwidth for deriving a quantile via the implicit definition in (23) is given. In this case the asymptotic mean square optimal bandwidth is proportional to $n^{-\frac{1}{3}}$. However, the optimal bandwidth for smoothing the joint density (or distribution) will in general depend on the dimension d. This makes it necessary to compromise either on the marginal or joint distributional level when selecting the bandwidth. When different bandwidths are chosen for the marginal smoothing and quantile computation via (23), and for the smoothing of the joint distribution via (20), the resulting combination in (21) is not a proper copula since the resulting margins are not adapted to the joint distribution. It is important to point out that in an asymptotic $n \to \infty$ consideration different bandwidth choices for the margins and joint distribution might not pose any problems as long as the usual conditions are obeyed. In a setting with a fixed n, it is, however, not possible to mix different bandwidths and obtain a proper copula, even if they might be optimal when considered individually.

If an elliptical kernel (or any other multivariate kernel with accessible marginal distributions) is used to construct the kernel density estimate, it is possible to simulate from the implied conditional copula in (21). This is done by combining

the simulation of mixture distributions discussed at the end of the previous section with the marginal distributions given in (22). This leads to the following algorithm to draw a random sample from $\widehat{\boldsymbol{C}}_n^{|\mathbb{X}}$.

Algorithm 4.1 (Smooth bootstrap sample from $\widehat{\boldsymbol{C}}_n^{|\mathbb{X}}$)**.** *Denote by* $\{\boldsymbol{x}_i\}_{i=1}^n$ *a given set of observations and assume that a sensible bandwidth matrix* H_n *has been determined (see, e.g., the discussion in Section 4.2). To draw a pseudo-random sample* $(\boldsymbol{u}_\ell^*)_{\ell=1}^m$ *from* $\widehat{\boldsymbol{C}}_n^{|\mathbb{X}}$ *of size* m *repeat the following steps for* $\ell \in \{1, \dots, m\}$*:*

1. *Draw a pseudo-random* $\boldsymbol{y}_\ell$ *distributed according to the kernel density* $\mathbf{k}$*.*
2. *Draw an index* i *uniformly from* $\{1, \dots, n\}$ *and set* $\boldsymbol{z}_\ell^* = \boldsymbol{x}_i + H_n^{1/2}\boldsymbol{y}_\ell$*.*
3. *Return* $\boldsymbol{u}_\ell^* = \left(\widehat{F}_{n1}(z_{\ell 1}^*), \dots, \widehat{F}_{nd}(z_{\ell d}^*)\right)$*, where* $\widehat{F}_{nj}$ *is the* j*th marginal distribution of* $\widehat{\boldsymbol{F}}_n^{|\mathbb{X}}$*.*

Remark 4.1. *Algorithm 4.1 is based on a direct application of the kernel density estimate. However, in certain situations a straightforward application of the kernel density estimation might not be possible. The most common situation of the latter type is when the multivariate joint distribution is only supported on a compact set, leading to a boundary bias in the estimation. Specifically relevant to our investigation are the cases when the original sample* $\mathbb{X}$ *consists (i) of observations of a copula* $\boldsymbol{C}$*, or (ii) of copula pseudo-observations obtained by applying the marginal empirical distribution functions to the component samples. To avoid boundary issues in such cases it is possible to transform the observations from* $[0,1]^d$ *to* $\mathbb{R}^d$ *by an appropriate (bijective) marginal transformation. Although any continuous marginal distribution function can be used for this transformation, it is most common to use the standard normal distribution function, see, for example, (Joe, 2015, Section 5.10.3). Transformation re-transformation approaches in the context of non-parametric copula estimation have also been studied in Omelka, Gijbels and Veraverbeke (2009) and Geenens, Charpentier and Paindaveine (2018). In the new domain the smoothing can then be carried out according to Algorithm 4.1. Since the copula is invariant under strictly increasing marginal transformations, the resulting sample is still representative of the underlying copula.*

The smooth bootstrap outlined in Algorithm 4.1 can now be used in two situations when dealing with a copula functional T and its empirical version T_n defined for samples $\mathbb{X}$ (of arbitrary size n). First, the smooth bootstrap can be used to gauge the distribution (and other characteristics) of T_n for a fixed sample size n. While the original sample $\mathbb{X}$ only allows for one realization of T_n, one can use the smooth bootstrap to gauge for example the distribution of T_n. Via Algorithm 4.1 one can draw B smooth bootstrap samples $(\mathbb{S}_b^*)_{b=1}^B$ of size n. This leads to B smooth bootstrap observations $(T_n^b)_{b=1}^B$ of T_n, which are based on $\widehat{\boldsymbol{F}}_n^{|\mathbb{X}}$ with underlying copula $\widehat{\boldsymbol{C}}_n^{|\mathbb{X}}$. While the number of bootstrap samples B is under the control of the statistician, it is crucial to verify whether the resulting bootstrap distribution is (asymptotically for $n \to \infty$) representative

for the distribution of T_n at the population level. A general discussion of this issue for the smooth bootstrap can be found in Shao and Tu 1995, Chapter 3.5.

Second, the bootstrap can also be used as a method of data augmentation. This method can come into play when an approximation of $T(\boldsymbol{C_X})$ is constructed by replacing the unknown copula $\boldsymbol{C_X}$ with either the empirical copula or the smooth version $\widehat{\boldsymbol{C}}_n^{|\mathbb{X}}$. This can either be necessary when the computations based on the empirical copula, i.e., the original sample $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$, are too coarse to be useful, or to facilitate the (numerical) approximation of $T\left(\widehat{\boldsymbol{C}}_n^{|\mathbb{X}}\right)$. For certain functionals T, such as level sets, or Kendall's tau and Spearman's rho which are given as multivariate integrals, the computation of $T\left(\widehat{\boldsymbol{C}}_n^{|\mathbb{X}}\right)$ might pose (numerical) challenges even if $\widehat{\boldsymbol{C}}_n^{|\mathbb{X}}$ is in principle known and can be evaluated via (21). From a practical point of view it can then be easier to use an approximation

$$\lim_{m\to\infty} T_m\left(\boldsymbol{u}_1^*, \ldots, \boldsymbol{u}_m^*\right) = T\left(\widehat{\boldsymbol{C}}_n^{|\mathbb{X}}\right),$$

if a suitable sample version T_m is available. To make sense of the limit we need to formally define a distance between $T_m\left(\boldsymbol{u}_1^*, \ldots, \boldsymbol{u}_m^*\right)$ and $T\left(\widehat{\boldsymbol{C}}_n^{|\mathbb{X}}\right)$ and we circle back to this issue in Section 4.3.

In both cases, instead of creating B samples of size n to assess the distribution of T_n, only one smooth bootstrap sample $\mathbb{S}_m^* = (\boldsymbol{u}_i^*)_{i=1}^m$ of size $m \gg n$ is created. Contrary to the approximation based on the original sample $T_n\left(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\right)$, Algorithm 4.1 allows us to sample an arbitrary number m of pseudo-observations. If the functional is well behaved, the resulting $T_m^* = T\left(\mathbb{S}_m^*\right)$ is then a close approximation to $T\left(\widehat{\boldsymbol{C}}_n^{|\mathbb{X}}\right)$.

An open question is which bandwidth matrix to use in Algorithm 4.1. This is addressed in the subsequent section.

4.2. Cross-validation bandwidth selection

A crucial part of the suggested procedure is the selection of the bandwidth matrix H. In the univariate case the asymptotic mean integrated squared error (AMISE) optimal bandwidth is of order $O\left(n^{-1/5}\right)$ in the case of density estimation. However, as shown in Azzalini (1981) the AMISE optimal bandwidth for distribution estimation is of order $O\left(n^{-1/3}\right)$. In the multivariate case we can thus not expect that bandwidth selection techniques designed for density estimation will work well when estimating distribution functions.

While a variety of bandwidth selection methods are available in the case of multivariate kernel density estimation, this is not the case when estimating distribution functions. When restricting oneself to product kernels, and hence diagonal bandwidth matrices in our setting, a plug-in estimator can be found in Liu and Yang (2008). Given that product kernels are too restrictive in our setup we instead turn to cross-validation for bandwidth selection.

In the univariate case Bowman, Hall and Prvan (1998) adapt cross-validation for kernel distribution estimators by introducing the objective function

$$\begin{aligned}\mathrm{CV}_n^1(h) &= \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}}\left(\mathbb{1}_{(-\infty,x]}(X_i) - \widehat{F}_{-i}(x)\right)^2 \mathrm{d}x \\ &= \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}}\left(\mathbb{1}_{[X_i,\infty)}(x) - \widehat{F}_{-i}(x)\right)^2 \mathrm{d}x,\end{aligned}$$

where $\widehat{F}_{-i}$ is the (in this case one-dimensional) leave-one-out kernel distribution estimator. Minimization of $\mathrm{CV}_n^1(h)$ with respect to h then leads to a sensible optimal bandwidth as argued in Bowman, Hall and Prvan (1998). While the integral is finite for compactly supported kernel functions, kernels with support on $\mathbb{R}$ can be used if they decay fast enough which can be seen when decomposing the integral as

$$\int_{\mathbb{R}}\left(\mathbb{1}_{[X_i,\infty)}(x) - \widehat{F}_{-i}(x)\right)^2 \mathrm{d}x = \int_{-\infty}^{X_i}\left(\widehat{F}_{-i}(x)\right)^2 \mathrm{d}x + \int_{X_i}^{\infty}\left(1 - \widehat{F}_{-i}(x)\right)^2 \mathrm{d}x.$$

However, in the multivariate case a direct generalization of CV_n^1 is only valid for compactly supported kernels. When supported over $\mathbb{R}^d$, the respective integrals will generally not converge. To solve this issue we introduce a weight function $w\colon \mathbb{R}^d \to [0,\infty)$ and define a weighted multivariate version of CV_n^1 as

$$\mathrm{CV}_n^d(H;w) = \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}^d}\left(\mathbb{1}_{(-\boldsymbol{\infty},\boldsymbol{x}]}(\boldsymbol{X}_i) - \widehat{\boldsymbol{F}}_{-i}(\boldsymbol{x})\right)^2 w(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}, \tag{24}$$

where $(-\boldsymbol{\infty},\boldsymbol{x}] = \times_{i=1}^{d}(-\infty,x_i]$ for $\boldsymbol{x} = (x_1,\ldots,x_d) \in \mathbb{R}^d$. Here $\widehat{\boldsymbol{F}}_{-i}$ denotes the leave-one-out kernel distribution estimate when disregarding the ith observation, i.e.,

$$\widehat{\boldsymbol{F}}_{-i}(\boldsymbol{x};\mathbb{X}) = \widehat{\boldsymbol{F}}_{-i}(\boldsymbol{x}) = \frac{1}{n-1}\sum_{\substack{j=1\\ j\neq i}}^{n}\mathbf{K}_H(\boldsymbol{x}-\boldsymbol{X}_j).$$

As a measure of performance we consider the weighted mean integrated squared error (WISE)

$$\mathrm{WISE}_n^d(H;w) = \mathbb{E}\left[\int_{\mathbb{R}^d}\left(\widehat{\boldsymbol{F}}_n(\boldsymbol{x}) - \boldsymbol{F}_{\boldsymbol{X}}(\boldsymbol{x})\right)^2 w(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right] \tag{25}$$

as a function of the bandwidth matrix H. A WISE optimal bandwidth matrix is any matrix that minimizes (25).

In the univariate case with $w \equiv 1$, Bowman, Hall and Prvan (1998) show that $\mathbb{E}\left[\mathrm{CV}_n^1(h)\right] = \mathrm{WISE}_{n-1}^1(h)$ up to a constant shift term that is independent of h which justifies minimizing $\mathrm{CV}_n^1(h)$ to find a sensible bandwidth. In the multivariate case we derive the following generalization concerning the objective function in (24).

Theorem 4.1. *If* $\int_{\mathbb{R}^d} w(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} < \infty$ *then*

$$\mathbb{E}\left[\mathrm{CV}_n^d(H;w)\right] = \mathrm{WISE}_{n-1}^d(H;w) + D_{\boldsymbol{X}}(w), \tag{26}$$

where $D_{\boldsymbol{X}}(w)$ *is independent of* H *and* n *and given by*
$D_{\boldsymbol{X}}(w) = \mathbb{E}\left[\int_{\mathbb{R}^d}\left(\mathbb{1}_{(-\boldsymbol{\infty},\boldsymbol{x}]}(\boldsymbol{X}) - \boldsymbol{F}_{\boldsymbol{X}}(\boldsymbol{x})\right)^2 w(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right]$.

Proof. We first note that the existence of the involved integrals is guaranteed by integrability of the weight function. Due to the iid setting we have that $\mathbb{E}\left[\mathbb{1}_{(-\boldsymbol{\infty},\boldsymbol{x}]}(\boldsymbol{X}_i)\widehat{\boldsymbol{F}}_{-i}(\boldsymbol{x})\right] = \boldsymbol{F}_{\boldsymbol{X}}(\boldsymbol{x})\mathbb{E}\left[\widehat{\boldsymbol{F}}_{-i}(\boldsymbol{x})\right]$ and therefore

$$\begin{aligned}
&\mathbb{E}\left[\mathrm{CV}_n^d(H)\right] - D_{\boldsymbol{X}}(w)\\
&= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}^d}\left(\mathbb{1}_{(-\boldsymbol{\infty},\boldsymbol{x}]}(\boldsymbol{X}_i) - \widehat{\boldsymbol{F}}_{-i}(\boldsymbol{x})\right)^2 w(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right]\\
&\quad - \mathbb{E}\left[\int_{\mathbb{R}^d}\left(\mathbb{1}_{(-\boldsymbol{\infty},\boldsymbol{x}]}(\boldsymbol{X}) - \boldsymbol{F}_{\boldsymbol{X}}(\boldsymbol{x})\right)^2 w(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right]\\
&= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}^d}\left(\mathbb{1}_{(-\boldsymbol{\infty},\boldsymbol{x}]}(\boldsymbol{X}_i) - \widehat{\boldsymbol{F}}_{-i}(\boldsymbol{x})\right)^2 w(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right]\\
&\quad - \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}^d}\left(\mathbb{1}_{(-\boldsymbol{\infty},\boldsymbol{x}]}(\boldsymbol{X}_i) - \boldsymbol{F}_{\boldsymbol{X}}(\boldsymbol{x})\right)^2 w(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right]\\
&= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}^d}\left(\mathbb{1}_{(-\boldsymbol{\infty},\boldsymbol{x}]}(\boldsymbol{X}_i)^2 + \widehat{\boldsymbol{F}}_{-i}(\boldsymbol{x})^2 - 2\mathbb{1}_{(-\boldsymbol{\infty},\boldsymbol{x}]}(\boldsymbol{X}_i)\widehat{\boldsymbol{F}}_{-i}(\boldsymbol{x})\right)\right.\\
&\quad \left. - \mathbb{E}\left[\mathbb{1}_{(-\boldsymbol{\infty},\boldsymbol{x}]}(\boldsymbol{X}_i)^2 - \boldsymbol{F}_{\boldsymbol{X}}(\boldsymbol{x})^2 + 2\mathbb{1}_{(-\boldsymbol{\infty},\boldsymbol{x}]}(\boldsymbol{X}_i)\boldsymbol{F}_{\boldsymbol{X}}(\boldsymbol{x})\right)w(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right]\\
&= \frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}^d}\left(\mathbb{E}\left[\widehat{\boldsymbol{F}}_{-i}(\boldsymbol{x})^2\right] - 2\mathbb{E}\left[\mathbb{1}_{(-\boldsymbol{\infty},\boldsymbol{x}]}(\boldsymbol{X}_i)\widehat{\boldsymbol{F}}_{-i}(\boldsymbol{x})\right] - \boldsymbol{F}_{\boldsymbol{X}}(\boldsymbol{x})^2\right.\\
&\qquad \left. + 2\mathbb{E}\left[\mathbb{1}_{(-\boldsymbol{\infty},\boldsymbol{x}]}(\boldsymbol{X}_i)\right]\boldsymbol{F}_{\boldsymbol{X}}(\boldsymbol{x})\right)w(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\\
&= \frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}^d}\left(\mathbb{E}\left[\widehat{\boldsymbol{F}}_{-i}(\boldsymbol{x})^2\right] - 2\boldsymbol{F}_{\boldsymbol{X}}(\boldsymbol{x})\mathbb{E}\left[\widehat{\boldsymbol{F}}_{-i}(\boldsymbol{x})\right] - \boldsymbol{F}_{\boldsymbol{X}}(\boldsymbol{x})^2\right.\\
&\qquad \left. + 2\boldsymbol{F}_{\boldsymbol{X}}(\boldsymbol{x})^2\right)w(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\\
&= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\int_{\mathbb{R}^d}\left(\widehat{\boldsymbol{F}}_{-i}(\boldsymbol{x}) - \boldsymbol{F}_{\boldsymbol{X}}(\boldsymbol{x})\right)^2 w(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right]\\
&= \mathbb{E}\left[\int_{\mathbb{R}^d}\left(\widehat{\boldsymbol{F}}_{n-1}(\boldsymbol{x}) - \boldsymbol{F}_{\boldsymbol{X}}(\boldsymbol{x})\right)^2 w(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right] = \mathrm{WISE}_{n-1}^d(H;w)
\end{aligned}$$

from which the claim follows by rearranging terms. The step from the third to last equality to the last line is justified by the identical distribution of the $\boldsymbol{X}_i$, leading to n times the same expectation. □

Clearly, if the kernel is compactly supported we can set $w \equiv 1$ and obtain a direct generalization of the result in Bowman, Hall and Prvan (1998) along the same lines. Although the constant $D_{\boldsymbol{X}}$ is typically unknown in a realistic setting, $D_{\boldsymbol{X}}$ does not depend on H and is hence irrelevant for the minimization. This justifies minimizing $\mathrm{CV}_n^d(H; w)$ with respect to H to obtain an approximation to the WISE optimal bandwidth.

In the absence of specific preferences it seems natural that the weight function w decays evenly in all directions from a central point. While any measure of centrality can in principle be used as a central point, we will be using the (sample) mean. A possible example of an appropriately shifted weight function is hence given by $w(\boldsymbol{x}) = \exp\left(-\left\|\boldsymbol{x} - \mathbb{E}\left[\boldsymbol{X}\right]\right\|_2^2\right)$.

In terms of selecting the bandwidth matrix H we fall back to the sphering approach introduced in Section 3.2, see also Wand and Jones 1995, Chapter 4.6 and references therein. Instead of optimizing over all $d(d+1)/2$ entries in H we instead compute the empirical covariance matrix $\widehat{\boldsymbol{S}}_n$ and introduce a one-dimensional optimization parameter h by setting $H = h\widehat{\boldsymbol{S}}_n$. Not only does this approach avoid the otherwise high dimensional optimization, but it also fits to the theoretical discussion in Section 3.2 since $\widehat{\boldsymbol{S}} \xrightarrow{\mathbb{P}} \operatorname{cov}\left[\boldsymbol{X}\right]$ and hence $H \approx -2h\psi'(0)\boldsymbol{\Sigma}$ for elliptical random vectors.

Finally, depending on the kernel $\widehat{\boldsymbol{F}}_{-i}$, the evaluation of the integrals in (24) is not possible in closed form. In this case multivariate numerical integration can be used to compute the integrals. The previously introduced weight function $w(\boldsymbol{x}) = \exp\left(-\left\|\boldsymbol{x} - \mathbb{E}\left[\boldsymbol{X}\right]\right\|_2^2\right)$ fits especially well with Gauss-Hermite quadrature and is utilized in our numerical examples where $d = 2$. Multivariate Gauss-Hermite quadrature that is compatible with the chosen weight function can be accomplished via a tensor grid, or, for higher dimensions, by the more efficient sparse grid integration. Sparse grid integration introduced by Smolyak (1963) efficiently combines univariate quadrature rules into multivariate ones; see Gerstner and Griebel (1998) for an overview.

Figure 2 visualizes (26) by showing the approximation of $\mathrm{WISE}_n^2(H; w)$ by the expectation of $\mathrm{CV}_n^2(H; w)$ in the case of a bivariate normal $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with parameters

$$\boldsymbol{\mu} = \begin{pmatrix} -1.0 \\ 1.0 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1.0 & 1.05 \\ 1.05 & 1.96 \end{pmatrix},$$

yielding a correlation coefficient of 0.75 between X_1 and X_2. In our example we also use a standard bivariate normal distribution $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_2)$ for the kernel $\mathbf{K}$. Due to the lack of closed form solutions we approximate $\mathrm{WISE}_n^2(H; w)$, $D_{\boldsymbol{X}}(w)$, as well as $\mathbb{E}\left[\mathrm{CV}_n^2(H; w)\right]$ numerically. All necessary integrals inside the respective expectations are computed via a bivariate Gauss-Hermite tensor product rule with 25 points in each dimension, totalling to 625 evaluation points. The weight

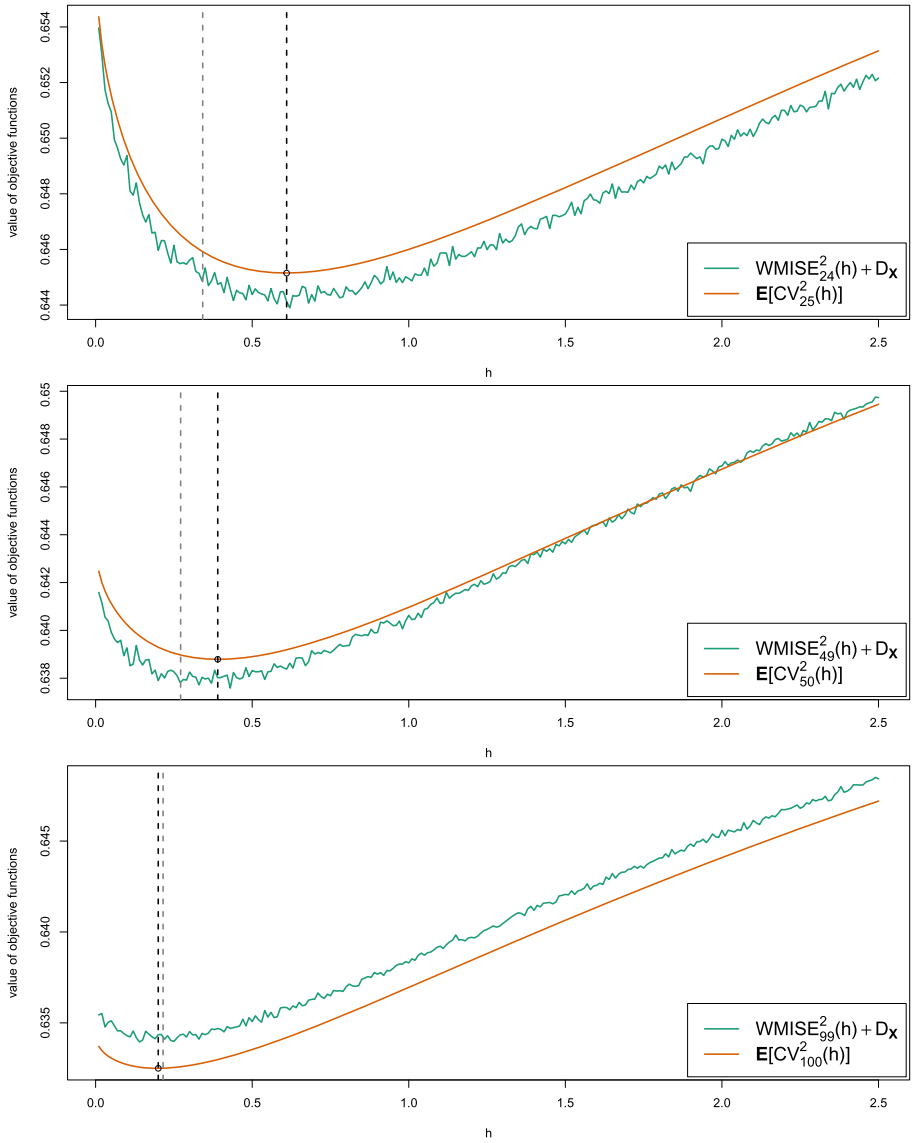

FIG 2. *Numerical approximation of* $\mathbb{E}\left[\mathrm{CV}_n^2(H;w)\right]$ *(orange) and* $\mathrm{WISE}_{n-1}^2(H;w) + D_{\boldsymbol{X}}$ *(green) for* $n = 25$ *(top),* $n = 50$ *(middle) and* $n = 100$ *(bottom) with* $H = h\boldsymbol{\Sigma}$ *for* $h \in \{0.01, 0.02, \ldots, 2.5\}$. *The minimum of* $\mathbb{E}\left[\mathrm{CV}_n^2(H;w)\right]$ *is indicated by a circle and the dashed vertical black line in all cases. The dashed vertical gray line indicates the choice of* h *following Silverman's rule of thumb.*

function is always centered at $\boldsymbol{\mu}$, i.e., $w(\boldsymbol{x}) = \exp\left(-\|\boldsymbol{x} - \boldsymbol{\mu}\|_2^2\right)$. Given that $D_{\boldsymbol{X}}(w)$ is independent of H and n we compute it only once based on 10 000 in-

TABLE 1
Simulation settings for the example in Section 4.2.

n	Sample size used for $\mathbb{E}\left[\mathrm{CV}_n^2(H;w)\right]$	Sample size used for $\mathrm{WISE}_{n-1}^2(H;w)$	Gauss-Hermite nodes used
25	300	5 000	25^2
50	60	2 500	25^2
100	20	1 000	25^2

dependent samples from $\boldsymbol{X}$ and find $D_{\boldsymbol{X}}(w) = 0.629732$. For a given bandwidth matrix H we compute $\mathrm{WISE}_n^2(H;w)$ for three sample sizes $n \in \{24, 49, 99\}$. For $n = 24$ we approximate the outer expectation by the mean over 5 000 independent samples of $\boldsymbol{X}$, for $n = 49$ we use 2 500 independent samples and for $n = 99$ the approximation is based on 1 000 independent samples. The approximation of $\mathbb{E}\left[\mathrm{CV}_n^2(H;w)\right]$ for the same bandwidth matrix H is thus based on samples of size $n \in \{25, 50, 100\}$. For $n = 25$ the value of $\mathbb{E}\left[\mathrm{CV}_n^2(H;w)\right]$ is approximated by 300 independent samples of size 25. In the case of $n = 50$ we use 60 independent samples and for $n = 100$ we use 20 independent samples from $\boldsymbol{X}$. For $\mathrm{WISE}_n^2(H;w)$ and $\mathbb{E}\left[\mathrm{CV}_n^2(H;w)\right]$, the bandwidth matrices are chosen as $H = h\boldsymbol{\Sigma}$ with $h \in \{0.01, 0.02, \ldots, 2.5\}$. The resulting approximation for $\mathbb{E}\left[\mathrm{CV}_n^2(H;w)\right]$ as a function of h is shown in orange for $n = 25$ (top), $n = 50$ (middle) and $n = 100$ (bottom). The green line shows the approximation to $\mathrm{WISE}_{24}^2(H;w) + D_{\boldsymbol{X}}(w)$ (top), $\mathrm{WISE}_{49}^2(H;w) + D_{\boldsymbol{X}}(w)$ (middle) and $\mathrm{WISE}_{99}^2(H;w) + D_{\boldsymbol{X}}(w)$ (bottom). The minimum of $\mathbb{E}\left[\mathrm{CV}_n^2(H;w)\right]$ is indicated by a circle and the dashed vertical black line in all cases. The settings are summarized in Table 1.

From Figure 2 we see that the approximation to $\mathrm{WISE}_n^2(H;w)$ seems to be more erratic than for $\mathbb{E}\left[\mathrm{CV}_n^2(H;w)\right]$. While the functions should perfectly match according to Theorem 4.1, the remaining differences can be attributed to the limited sample sizes. In line with intuition, the bandwidth parameter h is decreasing with sample size. We also see that the population version $\mathbb{E}\left[\mathrm{CV}_n^2(H;w)\right]$, i.e., the average over the sample versions $\mathrm{CV}_n^2(H;w)$, is a smooth function with a unique minimum. However, as known from cross-validation in other contexts, this is not necessarily the case for a given sample $\mathbb{X} = \{\boldsymbol{X}_i\}_{i=1}^n$. In Figure 3 we show two out of the 300 curves of $\mathrm{CV}_{25}^2(H;w)$ that are used in the computation of $\mathbb{E}\left[\mathrm{CV}_{25}^2(H;w)\right]$. The two underlying samples are denoted by $\mathbb{X}_1 = \{\boldsymbol{X}_i\}_{i=1}^{25}$ and $\mathbb{X}_2 = \{\boldsymbol{X}_i\}_{i=1}^{25}$. While the orange curve generated from $\mathbb{X}_2$ has a shape that is conducive to optimization, the black curve generated from $\mathbb{X}_1$ is monotonically decreasing over the considered range. When considering $\mathbb{E}\left[\mathrm{CV}_{25}^2(H;w)\right]$ all 300 curves get averaged which finally yields a reasonable target for minimization, but the individual curves might not be good optimization targets.

In our numerical experiments we find that this issue is more pronounced for small values of n. To address this issue in a practical situation it is possible to use a bootstrap approach to generate artificial samples that can then be averaged. Taking for example the sample $\mathbb{X}_1$ that generated the black curve in

Figure 3, we generate 25 new samples of size $n = 25$ by resampling from $\mathbb{X}_1$ with replacement. Based on these 25 new bootstrap samples we then compute an approximation to $\mathbb{E}\left[\mathrm{CV}_{25}^2(H; w)\right]$ by averaging. The resulting curve is shown in Figure 4 and shows a preferable shape compared to the initial black curve in Figure 3. Although the optimal bandwidth h taken from Figure 4 does not match the population version shown in Figure 2, it is important to recall that the starting sample $\mathbb{X}_1$ was problematic from the point of view of bandwidth selection by cross-validation to begin with due to its decreasing shape. In this sense bootstrapping helped to obtain a reasonable bandwidth under difficult conditions.

While our bandwidth selection approach is specifically tailored towards estimating multivariate distribution functions, we compare it to the popular rule of thumb bandwidth selection for multivariate kernel density estimation. In the multivariate case, Silverman's rule of thumb, see, e.g., Wand and Jones 1995, Chapter 4, is given by setting

$$h(d, n) = \left(\frac{4}{n(d+2)}\right)^{2/(d+4)}, \tag{27}$$

and then using sphering as before to arrive at $H_n = h(d, n)\widehat{\boldsymbol{S}}_n$. While this choice can be justified when estimating multivariate normal densities, it is not theoretically justified for estimating multivariate distribution functions, even in cases like our example setup where all involved distributions are multivariate normal. It is, however, computationally fast and easy to implement.

The dashed vertical gray line in Figure 2 indicates the choice of h when following Silverman's rule of thumb. As indicated in Figure 2, this bandwidth choice leads to undersmoothing for $n = 25$ and $n = 50$. On the contrary, for $n = 100$ the rule of thumb bandwidth is slightly larger than the WISE_n^2 optimal bandwidth. It is not surprising that the rule of thumb bandwidth differs from the WISE_n^2 optimal bandwidth. The difference can on the one hand be attributed to different objectives, distribution versus density estimation, that both methods are trying to accomplish. On the other hand, the weight function w is not part of the rule of thumb bandwidth selection procedure while it is explicitly necessary for the WISE_n^2 procedure.

4.3. Consistency

In this section, we outline the difference between $T(\boldsymbol{C}_{\boldsymbol{X}})$, $T\left(\widehat{\boldsymbol{C}}_n^{\mathbb{E}}\right)$ and $T\left(\widehat{\boldsymbol{C}}_n\right)$ from an asymptotic perspective, where $\widehat{\boldsymbol{C}}_n$ is the random version of $\widehat{\boldsymbol{C}}_n^{|\mathbb{X}}$ defined in (4). While a full development of the asymptotic theory is beyond the scope of this paper, we highlight the differences between the population and sample versions of the functionals and their relationship, creating a link to data augmentation as discussed in Section 4.1. We then choose Spearman's rho as an example to highlight the challenges at hand when developing asymptotic theory for the smooth bootstrap.

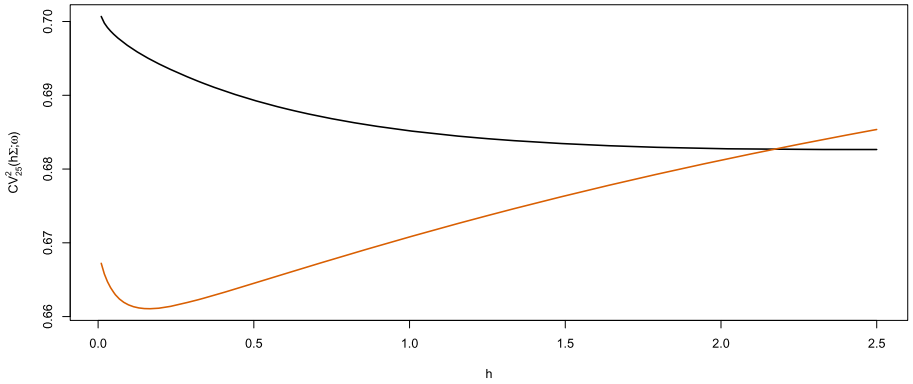

FIG 3. $\mathrm{CV}^2_{25}(h\boldsymbol{\Sigma}; w)$ *for two independent samples* $\mathbb{X}_1 = \{\boldsymbol{X}_i\}_{i=1}^{25}$ *(black) and* $\mathbb{X}_2 = \{\boldsymbol{X}_i\}_{i=1}^{25}$ *(orange) of* $\boldsymbol{X}$ *over a grid* $h \in \{0.01, 0.02, \ldots, 2.5\}$.

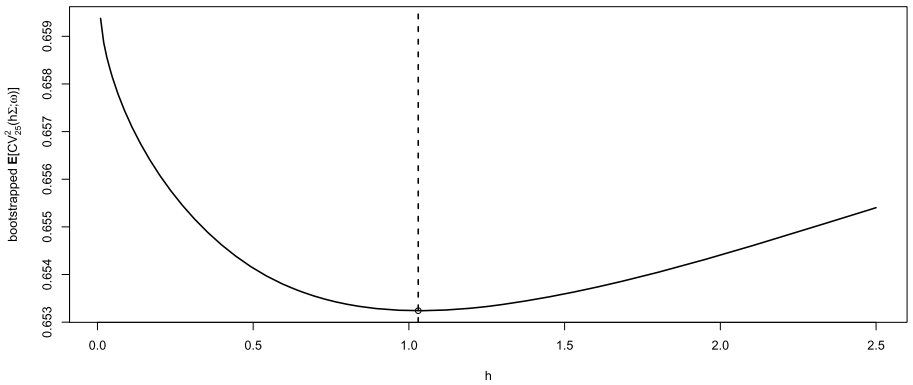

FIG 4. *Bootstrapped version of* $\mathbb{E}\left[\mathrm{CV}^2_{25}(h\boldsymbol{\Sigma}; w)\right]$ *based on the sample* $\mathbb{X}_1$ *over a grid* $h \in \{0.01, 0.02, \ldots, 2.5\}$. *The approximation is based on* 25 *independent bootstrap samples drawn from* $\mathbb{X}_1$. *The minimum is indicated by a circle and the dashed vertical line.*

In order to make the dependence of estimates on the respective bandwidth matrices clear we use the bandwidth matrix as an argument in the following. For example, $\widehat{\boldsymbol{C}}_n(H_n)$ denotes the estimate $\widehat{\boldsymbol{C}}_n$ based on the bandwidth matrix H_n. In principle, the quantity of interest is the (random) approximation error

$$E_{\mathbb{X}} = \mathrm{d}\left(T\left(\boldsymbol{C}_{\boldsymbol{X}}\right), T\left(\widehat{\boldsymbol{C}}_n\left(h_n \widehat{\boldsymbol{S}}_n\right)\right)\right),$$

where $\mathrm{d}(\cdot,\cdot)$ is an appropriate distance function (metric) chosen with regard to the functional T. Using the triangle inequality for $\mathrm{d}(\cdot,\cdot)$ we can incorporate our results concerning $\widehat{\boldsymbol{C}}_n^{\mathbb{E}}$ into this discussion via

$$\begin{aligned} E_{\mathbb{X}} &\leq \mathrm{d}\left(T\left(\boldsymbol{C}_{\boldsymbol{X}}\right), T\left(\widehat{\boldsymbol{C}}_n^{\mathbb{E}}\left(h_n^* \boldsymbol{\Sigma}\right)\right)\right) + \mathrm{d}\left(T\left(\widehat{\boldsymbol{C}}_n^{\mathbb{E}}\left(h_n^* \boldsymbol{\Sigma}\right)\right), T\left(\widehat{\boldsymbol{C}}_n\left(h_n \widehat{\boldsymbol{S}}_n\right)\right)\right) \\ &= E_1 + E_2, \end{aligned} \tag{28}$$

where h_n^* is the non-random, WISE_n optimal bandwidth factor and h_n is a data-driven and hence random choice for the bandwidth factor. It is important to observe that the first error E_1 is non-random. In the case of elliptical distributions and smoothing kernels, E_1 is a deterministic function of only the functional T and the bandwidth matrix $h_n^*\boldsymbol{\Sigma}$, which introduces the differences of the characteristic generators $\psi_{\boldsymbol{X}}$ and $\psi_{\boldsymbol{Z}}$ as addressed in Section 3. From our previous investigation we know that $E_1 = 0$ if either (i) the elliptical distributions for the data generating process and the kernel share the same dispersion matrix up to a scale factor and the functional T does not depend on the characteristic generator, cf. Section 3.2, or if (ii) the circumstances discussed in Section 3.4 are met. If $E_1 > 0$ in the elliptical setting of Section 3, the rate of convergence of $\psi_{\boldsymbol{X}} \to \psi_{\boldsymbol{Z}}$ discussed in Section 3.3 is linked to the convergence rate $\mathrm{d}\left(T\left(\boldsymbol{C}_{\boldsymbol{X}}\right), T\left(\widehat{\boldsymbol{C}}_n^{\mathbb{E}}\left(h_n^*\boldsymbol{\Sigma}\right)\right)\right) \to 0$.

As an example, we consider Spearman's rho ρ_{S} of a bivariate random vector $\boldsymbol{X}$ with an associated copula $\boldsymbol{C}_{\boldsymbol{X}}$ which is given by

$$\rho_{\mathrm{S}}\left(\boldsymbol{C}_{\boldsymbol{X}}\right) = 12 \int_0^1 \int_0^1 \boldsymbol{C}_{\boldsymbol{X}}(u, v)\, \mathrm{d}u\, \mathrm{d}v - 3; \tag{29}$$

see Nelsen 2006, Chapter 5. The corresponding sample version $T_n\left(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\right)$ is given by the Pearson correlation coefficient of the ranks of the first and second components of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$. As discussed in Section 3.2, similar alternatives given by Kendall's tau or Blomqvist's beta lead to $E_1 = 0$ in the considered setup. Given that Spearman's rho takes values in $[-1, 1]$, a suitable metric $\mathrm{d}(\cdot, \cdot)$ is in this case given by the absolute value $\mathrm{d}(x, y) = |x - y|$. For the error $E_{\mathbb{X}}$ this leads to

$$\begin{aligned}
E_{\mathbb{X}} &= \mathrm{d}\left(T\left(\boldsymbol{C}_{\boldsymbol{X}}\right), T\left(\widehat{\boldsymbol{C}}_n\left(h_n\widehat{\boldsymbol{S}}_n\right)\right)\right) = \left|\rho_{\mathrm{S}}\left(\boldsymbol{C}_{\boldsymbol{X}}\right) - \rho_{\mathrm{S}}\left(\widehat{\boldsymbol{C}}_n\left(h_n\widehat{\boldsymbol{S}}_n\right)\right)\right| \\
&= 12 \left| \int_0^1 \int_0^1 \boldsymbol{C}_{\boldsymbol{X}}(u, v) - \widehat{\boldsymbol{C}}_n\left(h_n\widehat{\boldsymbol{S}}_n\right)(u, v)\, \mathrm{d}u\, \mathrm{d}v \right| \\
&\leq 12 \int_0^1 \int_0^1 \left| \boldsymbol{C}_{\boldsymbol{X}}(u, v) - \widehat{\boldsymbol{C}}_n\left(h_n\widehat{\boldsymbol{S}}_n\right)(u, v) \right| \mathrm{d}u\, \mathrm{d}v \\
&\leq 12\, \mathrm{d}_{\mathrm{KS}}\left(\boldsymbol{C}_{\boldsymbol{X}}, \widehat{\boldsymbol{C}}_n\left(h_n\widehat{\boldsymbol{S}}_n\right)\right);
\end{aligned}$$

see (32) for the definition of the Kolmogorov-Smirnov distance d_{KS}. We can follow (28) to include $\widehat{\boldsymbol{C}}_n^{\mathbb{E}}\left(h_n^*\boldsymbol{\Sigma}\right)$ to obtain

$$\begin{aligned}
\mathrm{d}_{\mathrm{KS}}\left(\boldsymbol{C}_{\boldsymbol{X}}, \widehat{\boldsymbol{C}}_n\left(h_n\widehat{\boldsymbol{S}}_n\right)\right) \leq\ & \mathrm{d}_{\mathrm{KS}}\left(\boldsymbol{C}_{\boldsymbol{X}}, \widehat{\boldsymbol{C}}_n^{\mathbb{E}}\left(h_n^*\boldsymbol{\Sigma}\right)\right) \\
& + \mathrm{d}_{\mathrm{KS}}\left(\widehat{\boldsymbol{C}}_n^{\mathbb{E}}\left(h_n^*\boldsymbol{\Sigma}\right), \widehat{\boldsymbol{C}}_n\left(h_n\widehat{\boldsymbol{S}}_n\right)\right).
\end{aligned}$$

In principle the bivariate Kolmogorov-Smirnov distance between distributions can be bounded in terms of the characteristic functions, see Sadikova (1966) and Heuberger and Kropf (2018) for which we give the details in Theorem A.3, and we followed this approach at the end of Section 3.3. However, in this case,

a direct application is too restrictive due to the need for bounded derivatives, a condition that is not met for a number of popular copula families. As an alternative we utilize the invariance of the Kolmogorov-Smirnov distance under strictly increasing transforms. For continuous distribution functions $F_1, \ldots, F_d$ supported on $\mathbb{R}$ we have

$$\begin{aligned}
&\mathrm{d_{KS}}\left(\boldsymbol{C}_{\boldsymbol{X}}, \widehat{\boldsymbol{C}}_n\left(h_n \widehat{\boldsymbol{S}}_n\right)\right) \\
&= \sup_{\boldsymbol{u}\in(0,1)^d} \left|\boldsymbol{C}_{\boldsymbol{X}}(\boldsymbol{u}) - \widehat{\boldsymbol{C}}_n\left(h_n \widehat{\boldsymbol{S}}_n\right)(\boldsymbol{u})\right| \\
&= \sup_{\boldsymbol{x}\in\mathbb{R}^d} \left|\boldsymbol{C}_{\boldsymbol{X}}(F_1(x_1), \ldots, F_d(x_d)) - \widehat{\boldsymbol{C}}_n\left(h_n \widehat{\boldsymbol{S}}_n\right)(F_1(x_1), \ldots, F_d(x_d))\right| \\
&= \mathrm{d_{KS}}\left(\boldsymbol{F}_{\boldsymbol{X}}, \widetilde{\boldsymbol{F}}_n\right),
\end{aligned}$$

where $\boldsymbol{F}_{\boldsymbol{X}}$ and $\widetilde{\boldsymbol{F}}_n$ are the joint distribution functions with the respective copulas and identical margins $F_1, \ldots, F_d$. While showing that $\mathrm{d_{KS}}\left(\boldsymbol{F}_{\boldsymbol{X}}, \widetilde{\boldsymbol{F}}_n\right) \xrightarrow{\text{a.s.}} 0$, or $\mathrm{d_{KS}}\left(\boldsymbol{F}_{\boldsymbol{X}}, \widetilde{\boldsymbol{F}}_n\right) \xrightarrow{\mathbb{P}} 0$, is beyond the scope of this paper we outline one possible approach: Devroye and Wagner (1979) give conditions under which the measure associated to the kernel density estimate converges to the unknown measure of the true underlying density in total variation distance. For two measures μ and ν on $\mathbb{R}^d$ the total variation distance is defined as $\mathrm{d_{TV}}(\mu, \nu) = \sup_{B\in\mathscr{F}} |\mu(B) - \nu(B)|$, where $\mathscr{F}$ is the Borel sigma algebra on $\mathbb{R}^d$. In place of the measures μ and ν we will also use the distribution or density functions associated to them. From the definitions we immediately have $\mathrm{d_{KS}}(\boldsymbol{F}, \boldsymbol{G}) \leq \mathrm{d_{TV}}(\boldsymbol{F}, \boldsymbol{G})$ for any two distributions functions $\boldsymbol{F}$ and $\boldsymbol{G}$. To show that $\mathrm{d_{TV}}\left(\widehat{\boldsymbol{f}}_n, \boldsymbol{f}_{\boldsymbol{X}}\right) \xrightarrow{\text{a.s.}} 0$, and $\mathrm{d_{TV}}\left(\widehat{\boldsymbol{f}}_n, \boldsymbol{f}_{\boldsymbol{X}}\right) \xrightarrow{\mathbb{P}} 0$, Devroye and Wagner (1979) however rely, amongst other assumptions, on a diagonal bandwidth matrix in the definition of $\widehat{\boldsymbol{f}}_n$. We leave the adaptation to elliptical kernels with non-diagonal bandwidth matrices in our setting for further research.

As exemplified by our treatment of Spearman's rho, a detailed analysis of the asymptotic behaviour depends on the functional under consideration. If, for example, the functional T is a level set of the underlying copula, the absolute value is not a suitable metric since we need to quantify the distance between two sets or their respective boundaries. A suitable metric in this context is given by the Hausdorff distance between the sets enclosed by the contour lines. We discuss this approach in our simulation studies in the next section.

5. Simulation study

In this section, we illustrate Algorithm 4.1 with several examples for two different functionals T, namely copula level curves and copula based dependence measures. In our examples, we compare $T\left(\widehat{\boldsymbol{C}}_n^{|\mathbb{X}}\right)$ to $T(\boldsymbol{C})$, where we use the

smooth bootstrap to approximate $T\left(\widehat{\boldsymbol{C}}_n^{|\mathbb{X}}\right)$ via data augmentation, as discussed in Section 4.1. In our simulations we use elliptical smoothing kernels but we do not restrict ourselves to elliptical data generating processes to highlight that the approach is not limited to the specific situation discussed in Section 3.2.

5.1. Copula level curves and copula diagonals

Inspired by the application in Coblenz, Dyckerhoff and Grothe (2018) we first focus on bootstrapping level curves for copulas. For a copula $\boldsymbol{C}$ we define the sublevel set at level $t \in (0,1)$ as

$$\mathbb{L}_t(\boldsymbol{C}) = \{\boldsymbol{u} \in [0,1]^d : \boldsymbol{C}(\boldsymbol{u}) \leq t\}.$$

Sublevel sets of copulas have an interpretation as multivariate quantiles, see Salvadori et al. (2016), and are important for applications, e.g., in finance and hydrology. To assess whether the sublevel sets of two copulas are close we measure their distance in terms of the Hausdorff distance. For two subsets A and B of a metric space (M, d) the Hausdorff distance $\mathrm{d_H}(\cdot,\cdot)$ is defined as

$$\mathrm{d_H}(A,B) = \max\left(\sup_{x\in A}\inf_{y\in B}\mathrm{d}(x,y), \sup_{y\in B}\inf_{x\in A}\mathrm{d}(x,y)\right),$$

where in our simulations we use the standard Euclidean distance $\mathrm{d}(\boldsymbol{x},\boldsymbol{y}) = \|\boldsymbol{x}-\boldsymbol{y}\|_2$ over the unit cube.

In our simulation we draw pseudorandom numbers $\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_n\}$ from a fixed bivariate copula $\boldsymbol{C}$ and record the distance $\mathrm{d_H}\left(\mathbb{L}_t(\boldsymbol{C}), \widehat{\mathbb{L}}_t\right)$ for different values of t over a number of independent simulations. Here $\widehat{\mathbb{L}}_t$ is the estimated sublevel set based either on the original observations $\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_n\}$, $\boldsymbol{u}_i \in [0,1]^2$ for $i \in \{1,\ldots,n\}$, only, or on the augmented sample $\{\boldsymbol{u}_1^*,\ldots,\boldsymbol{u}_m^*\}$ produced by the smooth bootstrap. Sufficient conditions for the convergence $\mathrm{d_H}\left(\mathbb{L}_t(\boldsymbol{C}), \widehat{\mathbb{L}}_t\right) \xrightarrow{\text{a.s.}} 0$ are discussed in Coblenz, Dyckerhoff and Grothe (2018).

Concerning the bandwidth matrix in the smooth bootstrap we use the empirical variance-covariance matrix based on a transformation of $\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_n\}$ with the standard normal quantile function together with Silverman's rule of thumb smoothing parameter given in (27). We utilize this choice throughout the section due to the computational efficiency of the method in the simulation re-runs. To transform the smooth bootstrap sample back into $[0,1]^2$ we use the marginal distribution functions of the associated kernel distribution mixture as outlined in Algorithm 4.1. Finally, the multivariate normal density is used as smoothing kernel.

For a given sample, either $\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_n\}$ or $\{\boldsymbol{u}_1^*,\ldots,\boldsymbol{u}_m^*\}$, the boundary of the sublevel set is estimated by computing the contour lines of the associated empirical copula at level t, resulting in a piecewise linear approximation to $\mathbb{L}_t(\boldsymbol{C})$; to this end we apply R's contourLines() to the empirical copula constructed via empCopula() of the R package `copula` of Hofert et al. (2022). For a given level t

it is, from a theoretical perspective, also clear that the copula contour lines ultimately need to pass through $(t, 1)$ and $(1, t)$. Due to the inherent randomness in the samples the estimated contour lines do not necessarily fulfill this constraint. For points (u_i, v_i) on the estimated contour line it is however possible to modify the results accordingly by replacing all values $u_i < t$ with t and all values $v_i < t$ with t. This modification ensures the validity of the boundary conditions and is utilised in our computations. As a result of the algorithm we obtain (by adding the points $(0, 0)$, $(1, 0)$ and $(0, 1)$) the vertices of a polygon that approximates $\mathbb{L}_t(\boldsymbol{C})$, where the number of vertices depends on the chosen number of grid points used to discretize the x-axes between t and 1. Concerning the sublevel sets of the true underlying copula $\mathbb{L}_t(\boldsymbol{C})$ it is also necessary to discretize the boundary of $\mathbb{L}_t(\boldsymbol{C})$ into a list of vertices over a sufficiently fine approximation grid. Concerning the numerical computation of $\mathrm{d_H}(\cdot, \cdot)$ we implement the algorithm outlined in Taha and Hanbury (2015). The algorithm efficiently computes the Hausdorff distance between two polygons which then applies directly to the estimated boundary of $\widehat{\mathbb{L}}_t$ and the discretized boundary of $\mathbb{L}_t(\boldsymbol{C})$.

For the simulation setup we consider the following data. We choose the underlying true copula in the Archimedean class of copulas, see Nelsen (2006) for an overview. This allows to obtain a closed-form expression for the boundary of $\mathbb{L}_t(\boldsymbol{C})$ and hence allows to accurately discretize the boundary of the true underlying copula $\mathbb{L}_t(\boldsymbol{C})$. Specifically, we simulate from a Clayton copula $\boldsymbol{C}_\theta$ where we fix the model parameter θ in such a way that Kendall's tau takes specific values, $\tau(\boldsymbol{C}_\theta) \in \{-0.9, -0.8, -0.7, \ldots, 0.9\}$. We also consider $t \in \{0.1, 0.2, \ldots, 0.9\}$ and $n \in \{25, 50, 100\}$. For the smooth bootstrap we generally set the sample size of the resulting (augmented) sample to $m = 5\,000$. The difference between n and m will (generally) lead to different discretization step sizes for the respective boundary approximations. We repeat each simulation independently $M = 10\,000$ times.

The resulting distances for $n = 25$ are presented in the boxplots in Figure 5. The results for $n = 50$ can be found in Figure 6, while Figure 7 shows the results for $n = 100$. The figures clearly highlight the benefit of using the smooth bootstrap in this situation. Without smoothing, the estimated contour lines based on the original data samples are too coarse. On the one hand this makes them unusable in practice, see the discussion in Coblenz, Dyckerhoff and Grothe (2018), on the other hand this leads to a significant distance from the theoretical target. Using an augmented data sample constructed by a smooth bootstrap procedure on the other hand leads to an estimated curve that bears more resemblance with a contour line compared to the step-function like estimation result of the standard estimation. Not only is the resulting curve more suitable for practical applications, but also it is closer to the theoretical target. A visual representation of this situation can be found in Figure 8, where we depict in black the theoretical contour line of a Clayton copula at level $t = 0.3$ with parameter $\theta = 2$ and an associated Kendall's tau of $\tau = 1/2$. For a sample of size $n = 25$ we also give the estimated contour lines based on the original sample (green) and the estimated contour lines based on the smooth bootstrap (orange). As before, we use Silverman's rule of thumb, see (27), in combination

with the empirical variance-covariance matrix to construct the bandwidth matrix for the bivariate Gaussian kernel. The empirical variance-covariance matrix is computed based on the data once they have been transformed into $\mathbb{R}^2$ via the standard normal quantile function. As discussed, the smooth bootstrap contour lines provide a better approximation to the theoretical target by smoothing out the sharp kinks of the direct estimate which is clearly visible in Figure 8.

Similar improvements are visible in the estimation of the copula diagonal $\delta(u) = \boldsymbol{C}(u, \ldots, u)$ which is presented in Figure 9 for a twelve dimensional Clayton copula with parameter $\theta = 5$. In small samples, $n = 10$ in the example, the estimation based on the empirical copula is too coarse to be useful in a practical situation. When using the smooth bootstrap to generate additional observations the empirical copula diagonal based on the enlarged sample, $m = 10\,000$ in the example, is less similar to a step function and closer to the unknown copula diagonal. As for the contour lines, the smooth bootstrap is advantageous when the target functional is a curve where the evaluation $\delta(u)$ needs to exhibit a smooth behavior for nearby values of u.

5.2. Copula based dependence measures

As an alternative to copula contour lines we now consider the estimation of copula based dependence measures where we focus on Spearman's rho and Kendall's tau. In our simulations we estimate Spearman's rho and Kendall's tau based on samples of size $n \in \{5, 10, 20, 25, 50, 75, 100\}$ where we consider Clayton, Student-t, Gumbel, Joe and Gaussian copulas. In these copula families both dependence measures can be computed in closed form which allows us to compare our estimates to the true underlying values. Based on the original observations we use the smooth bootstrap to generate $m = 10\,000$ observations based on a Gaussian kernel where we use Silverman's rule of thumb to establish our bandwidth matrix. Finally, we repeat the simulations 2 000 times. For the Clayton copula with parameter $\theta = 4$ this leads to the boxplots in Figure 10 (top) and the mean squared error curves in Figure 10 (bottom) for Spearman's rho, and the boxplots in Figure 11 (top) and the mean squared error curves in Figure 11 (bottom) for Kendall's tau. The figures show that for small sample sizes the smooth bootstrap leads to an improved estimation in terms of the mean squared error for both, Spearman's rho and Kendall's tau. For sample sizes larger than 75 the advantage of the smooth bootstrap disappears. The figures in Appendix C show that the same conclusions hold for our numerical experiments using the Student-t copula with $\rho = 0.9$ and $\nu = 4$ degrees of freedom (see Figure 12 and 13), the Gumbel copula with parameter $\theta = 4$ (see Figure 14 and 15), the Joe copula with parameter $\theta = 4$ (see Figure 16 and 17) and the Gaussian copula with $\rho = 0.9$ (see Figure 18 and 19). Overall these findings seem to indicate that the smooth bootstrap improves the estimation of Spearman's rho and Kendall's tau for small sample sizes. In the next section we conclude our results.

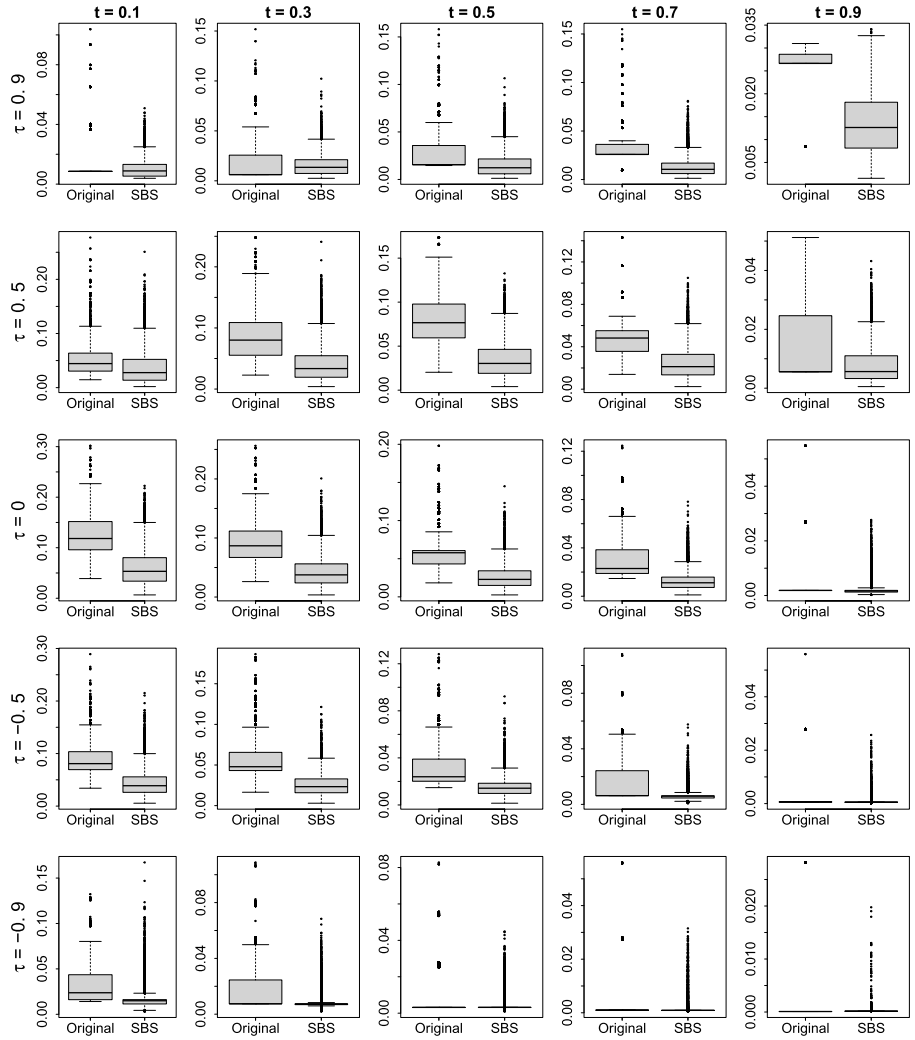

FIG 5. *Hausdorff distance simulation results between the true and estimated level sets for a Clayton copula* $\boldsymbol{C}_\theta$ *with* θ *such that* $\tau(\boldsymbol{C}_\theta) \in \{-0.9, -0.5, 0, 0.5, 0.9\}$*, for levels* $t \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$*. Original sample size* $n = 25$*, augmented smooth bootstrap sample size* $m = 5\,000$*. Each boxplot is based on* $M = 10\,000$ *independent reruns.*

6. Conclusion

We investigate the distortion of the underlying dependence structure that arises as a side effect of the smooth bootstrap. In the framework of elliptical distributions and elliptical smoothing kernels with a sphering type bandwidth matrix we provide the exact mechanism that leads to the distortion of the resulting elliptical copula. Even though sphering is at first glance a strong restriction on the

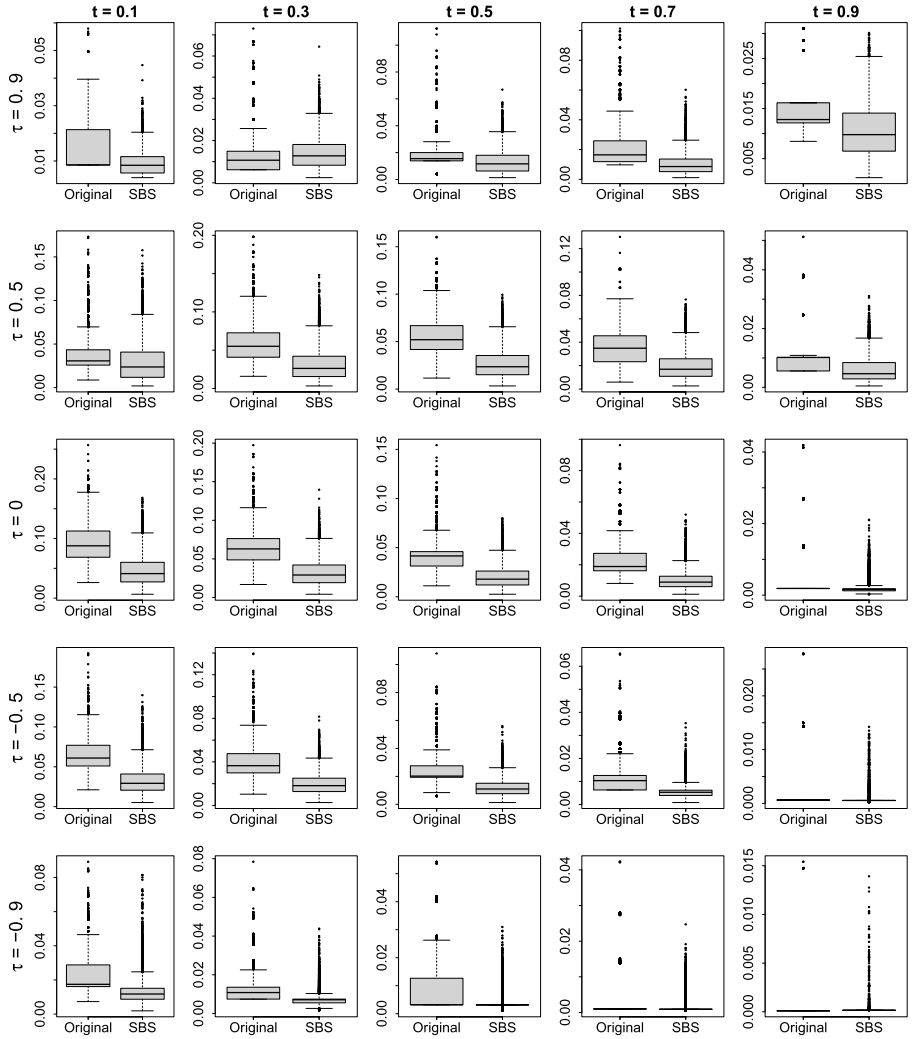

FIG 6. *Hausdorff distance simulation results between the true and estimated level sets for a Clayton copula* $\boldsymbol{C}_\theta$ *with* θ *such that* $\tau(\boldsymbol{C}_\theta) \in \{-0.9, -0.5, 0, 0.5, 0.9\}$*, for levels* $t \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$*. Original sample size* $n = 50$*, augmented smooth bootstrap sample size* $m = 5\,000$*. Each boxplot is based on* $M = 10\,000$ *independent reruns.*

choice of a possible bandwidth matrix, it allows us to bypass otherwise necessary restrictions such as product kernels or diagonal bandwidth matrices. While in our results the parameter matrix of the elliptical copula remains unchanged, the associated characteristic generator is distorted by a multiplicative factor related to the smoothing kernel. We connect our investigation to the previous result of Bingham (1972) to show that in general the pointwise convergence rate between the original and smoothed characteristic function is linked to the regular vari-

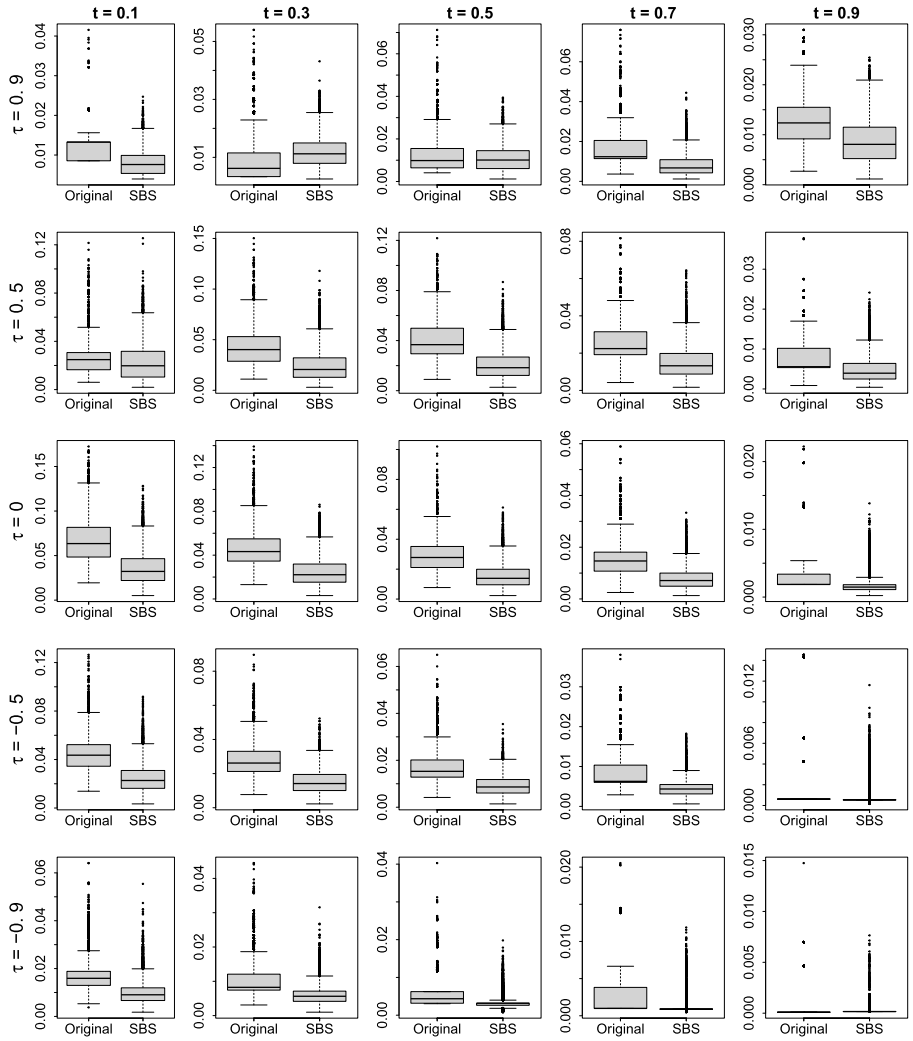

FIG 7. *Hausdorff distance simulation results between the true and estimated level sets for a Clayton copula* $\boldsymbol{C}_\theta$ *with* θ *such that* $\tau(\boldsymbol{C}_\theta) \in \{-0.9, -0.5, 0, 0.5, 0.9\}$*, for levels* $t \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$*. Original sample size* $n = 100$*, augmented smooth bootstrap sample size* $m = 5\,000$*. Each boxplot is based on* $M = 10\,000$ *independent reruns.*

ation of the characteristic generator of the smoohting kernel. Surprisingly we however also uncover situations where the underlying elliptical copula remains completely unaffected by the smooth bootstrap on the population level. To complement this finding we discuss examples where the dependence distortion can be worked out in detail. Furthermore, given that the parameter matrix remains unchanged, the dependence distortion introduced by kernel smoothing does not have any impact on a functional of the copula if the functional does not depend

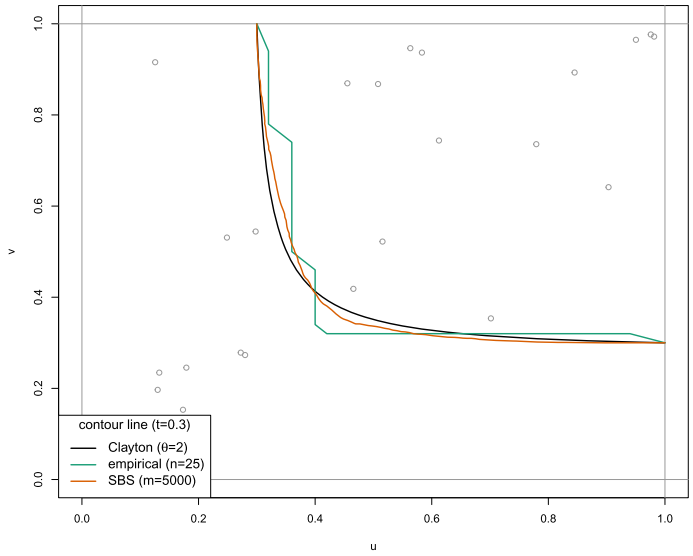

FIG 8. *Theoretical (black) and estimated contour lines for a Clayton copula with parameter* $\theta = 2$ *at level* $t = 0.3$. *Direct estimation (green) is based on* $n = 25$ *sample points. For the smooth bootstrap (orange)* $m = 5\,000$ *pseudo-observations are generated based on the initial sample.*

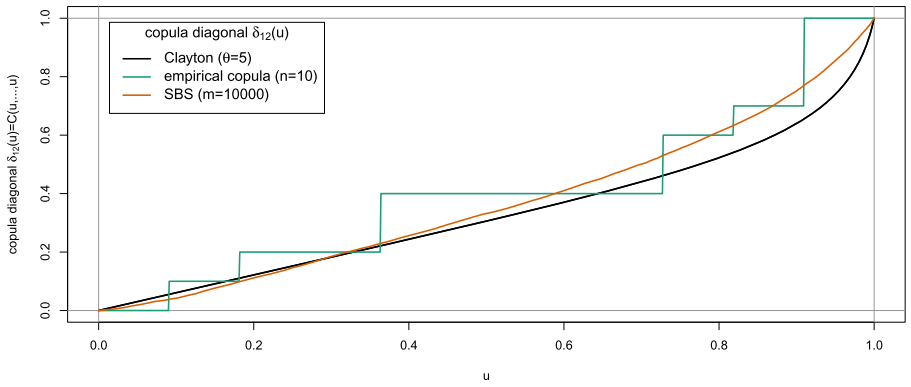

FIG 9. *Theoretical (black) and estimated copula diagonal for a twelve dimensional Clayton copula with parameter* $\theta = 5$. *Direct estimation (green) via the empirical copula is based on* $n = 10$ *sample points. For the smooth bootstrap (orange)* $m = 10\,000$ *pseudo-observations are generated based on the initial sample which are then used to estimate the diagonal via the empirical copula.*

on the characteristic generator of the underlying elliptical copula. Examples of such functionals are Kendall's tau and Blomqvist's beta. However, in practical applications, even if it is known that the data generating process is elliptical, the estimation of $\boldsymbol{\Sigma}$ may still impact our results and the estimation uncertainty

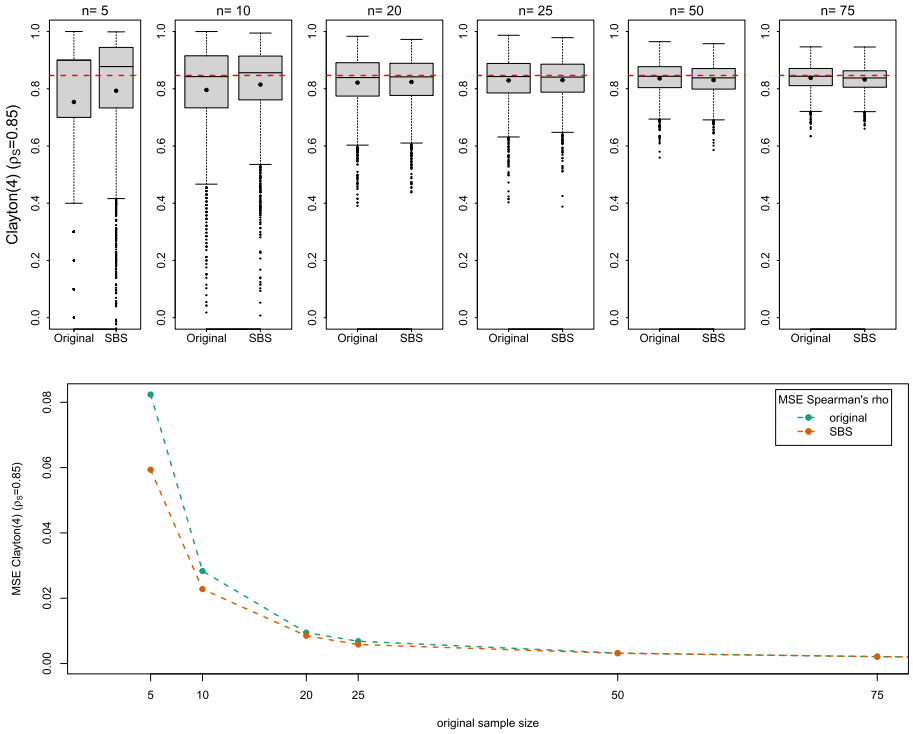

FIG 10. *Top: Estimated Spearman's rho for the bivariate Clayton(4) copula. The results are based on an original sample size of* $n \in \{5, 10, 20, 25, 50, 75\}$*, while the augmented smooth bootstrap sample size is* $m = 10\,000$*. Each boxplot is based on* $M = 2\,000$ *independent reruns. The red dashed line indicates the theoretical value of* ρ_S *while black dots indicate the means. Bottom: Mean squared error for estimation of Spearman's rho for the bivariate Clayton(4) copula. The results are based on an original sample size of* $n \in \{5, 10, 20, 25, 50, 75\}$*, while the augmented smooth bootstrap sample size is* $m = 10\,000$*.*

connected to $\widehat{\boldsymbol{\Sigma}}$ has to be taken into account.

From a practical perspective, we outline how the smooth bootstrap can be utilized to generate observations from the smoothed copula. Thus, it serves as a data augmentation scheme. As well as stating an algorithm which can be used in this situation, we discuss details and options concerning the marginal transforms and application scenarios. As a necessary part of the algorithm we generalise the univariate bandwidth selection procedure of Bowman, Hall and Prvan (1998) to the multivariate case. This bandwidth selection procedure is not limited to diagonal bandwidth matrices and allows us to select an optimal full bandwidth matrix in terms of a weighted mean integrated squared error criterion.

In our simulation studies we utilize the smooth bootstrap to improve the estimation of copula contour lines in the bivariate and the copula diagonal in the multivariate case. While our theoretical investigation is limited to elliptical

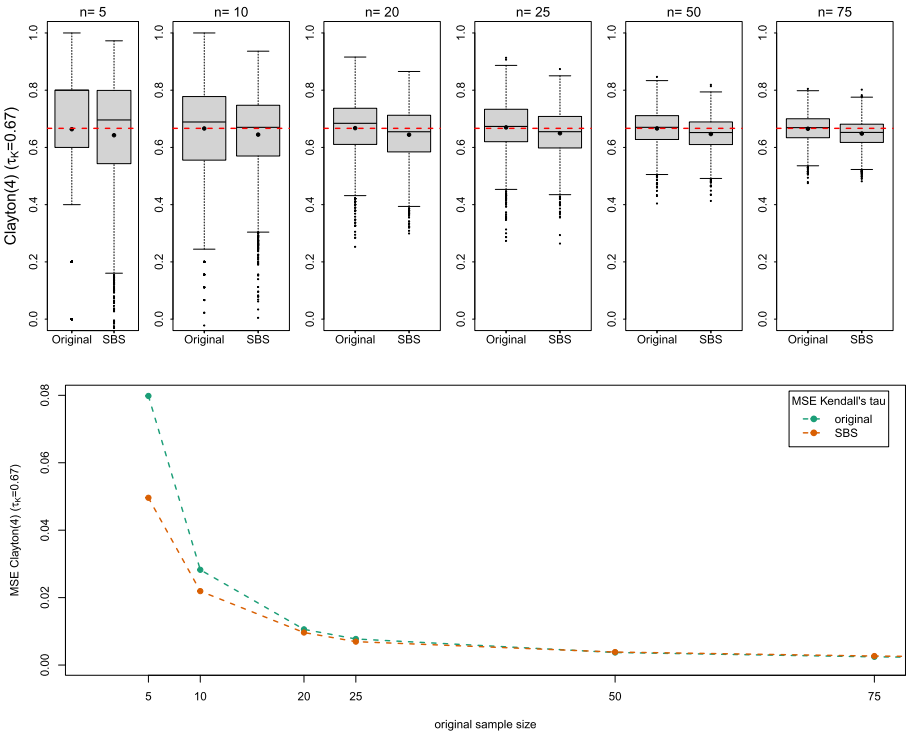

FIG 11. *Top: Estimated Kendall's tau for the bivariate Clayton(4) copula. The results are based on an original sample size of* $n \in \{5, 10, 20, 25, 50, 75\}$*, while the augmented smooth bootstrap sample size is* $m = 10\,000$*. Each boxplot is based on* $M = 2\,000$ *independent reruns. The red dashed line indicates the theoretical value of* τ_K *while black dots indicate the means. Bottom: Mean squared error for estimation of Kendall's tau for the bivariate Clayton(4) copula. The results are based on an original sample size of* $n \in \{5, 10, 20, 25, 50, 75\}$*, while the augmented smooth bootstrap sample size is* $m = 10\,000$*.*

distributions, our simulations show a vast improvement of the approximation measured in terms of the Hausdorff distance even though the utilised Clayton copula is not in the class of elliptical copulas. In a second set of simulations we consider copula based measures of dependence where we focus on Spearman's rho and Kendall's tau. Our results show that the smooth bootstrap improves the estimation in small samples for a number of copulas in the elliptical and Archimedean class.

Based on our simulation results the smooth bootstrap can lead to an improved estimation of copula functionals for small sample sizes. The procedure is especially advantageous if the target functional is smooth, such as level curves or the copula diagonal in the considered examples, but the estimator based on the initial (small) sample is too coarse. Here an application of the smooth bootstrap leads to a virtually unlimited number of observations which, in the considered applications, leads to a dramatic improvement of the estimation.

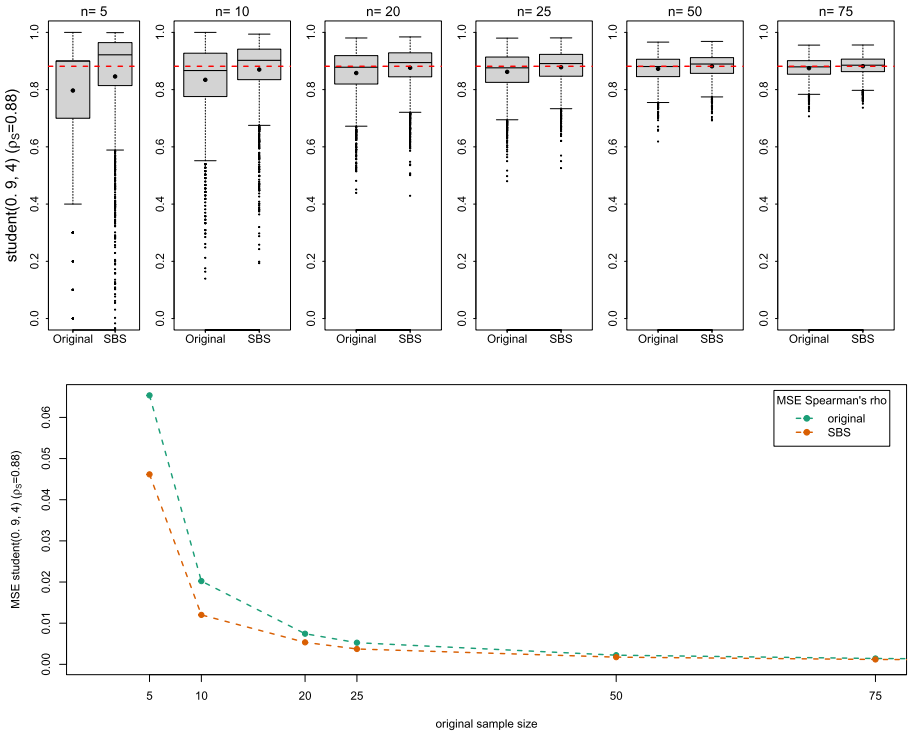

FIG 12. *Top: Estimated Spearman's rho for the bivariate Student-t copula with* $\rho = 0.9$ *and* $\nu = 4$ *degrees of freedom. The results are based on an original sample size of* $n \in \{5, 10, 20, 25, 50, 75\}$*, while the augmented smooth bootstrap sample size is* $m = 10\,000$*. Each boxplot is based on* $M = 2\,000$ *independent reruns. The red dashed line indicates the theoretical value of* ρ_S *while black dots indicate the means.*
Bottom: Mean squared error for estimation of Spearman's rho. The results are based on an original sample size of $n \in \{5, 10, 20, 25, 50, 75\}$*, while the augmented smooth bootstrap sample size is* $m = 10\,000$.

Appendix A: Properties of characteristic functions

Here, we review properties of multivariate characteristic functions.

Definition A.1 (d-dimensional Characteristic function)**.** *The characteristic function of* d*-dimensional random vector* $\boldsymbol{X}$ *is defined as*

$$\phi_{\boldsymbol{X}} \colon \mathbb{R}^d \to \mathbb{C}, \quad \boldsymbol{t} \mapsto \phi_{\boldsymbol{X}}(\boldsymbol{t}) = \mathbb{E}\left[e^{\mathrm{i}\boldsymbol{t}^\top \boldsymbol{X}}\right],$$

where $\mathrm{i} \in \mathbb{C}$ *is the imaginary unit with* $\mathrm{i}^2 = -1$.

In an abuse of notation we may also write $\phi_{\boldsymbol{F}}$ and $\phi_{\boldsymbol{f}}$ when $\boldsymbol{X}$ has distribution $\boldsymbol{F}$ and density $\boldsymbol{f}$. Among the properties of characteristic functions the following two theorems will be useful in the context of kernel smoothing. A comprehensive

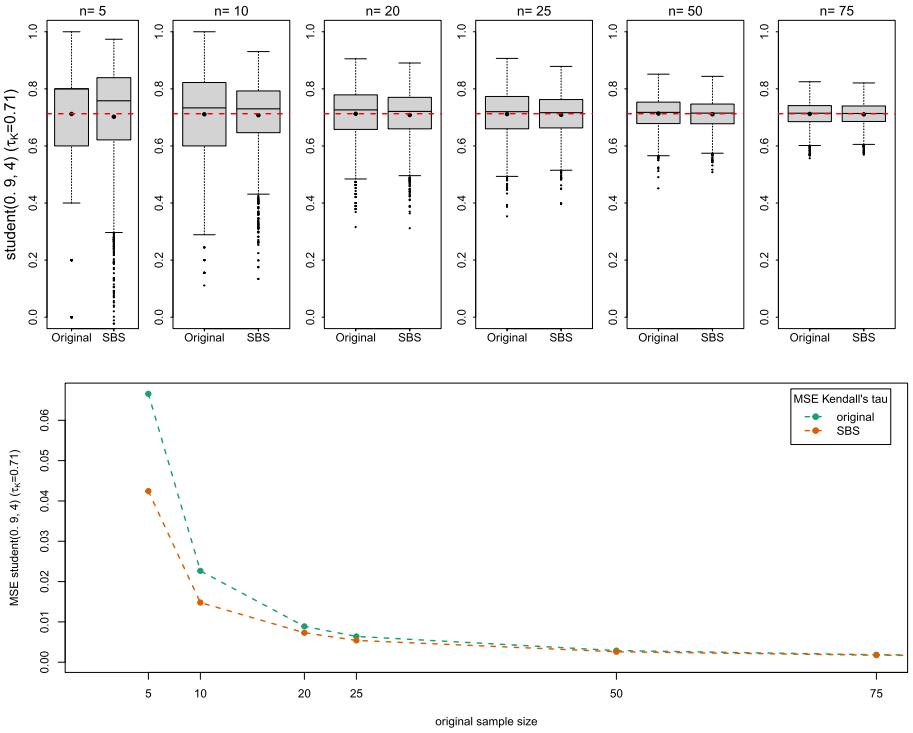

FIG 13. *Top: Estimated Kendall's tau for the bivariate Student-t copula with* $\rho = 0.9$ *and* $\nu = 4$ *degrees of freedom. The results are based on an original sample size of* $n \in \{5, 10, 20, 25, 50, 75\}$*, while the augmented smooth bootstrap sample size is* $m = 10\,000$*. Each boxplot is based on* $M = 2\,000$ *independent reruns. The red dashed line indicates the theoretical value of* τ_K *while black dots indicate the means.*
Bottom: Mean squared error for estimation of Kendall's tau. The results are based on an original sample size of $n \in \{5, 10, 20, 25, 50, 75\}$*, while the augmented smooth bootstrap sample size is* $m = 10\,000$*.*

treatment of multivariate characteristic functions can be found, e.g., in Sasvári (2013).

Theorem A.1 (Characteristic function of convolutions, Sasvári 2013, Theorem 1.1.3)**.** *If* $\boldsymbol{X}$ *and* $\boldsymbol{Y}$ *are independent* d*-dimensional random vectors, then the characteristic function of their sum is* $\phi_{\boldsymbol{X}+\boldsymbol{Y}} = \phi_{\boldsymbol{X}} \cdot \phi_{\boldsymbol{Y}}$*.*

Theorem A.2 (Characteristic function of affine transformations, Sasvári 2013, Theorem 1.1.7)**.** *Let* $\boldsymbol{X}$ *be a* d*-dimensional random vector. Then the equation*

$$\phi_{A\boldsymbol{X}+\boldsymbol{b}}(\boldsymbol{t}) = e^{\mathrm{i}\boldsymbol{t}^\top \boldsymbol{b}} \cdot \phi_{\boldsymbol{X}}\left(A^\top \boldsymbol{t}\right), \boldsymbol{t} \in \mathbb{R}^n$$

holds for every linear mapping $A\colon \mathbb{R}^d \to \mathbb{R}^n$ *and* $\boldsymbol{b} \in \mathbb{R}^n$*.*

Due to the assumed independence of the underlying random vector and the smoothing kernel we can directly compute the characteristic function for our

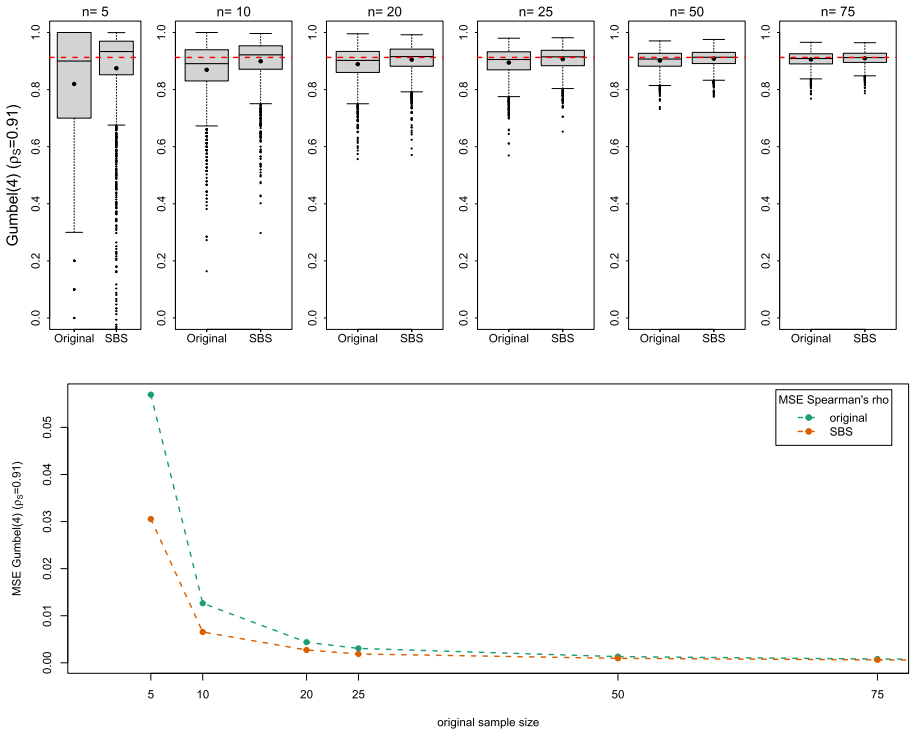

FIG 14. *Top: Estimated Spearman's rho for the bivariate Gumbel copula with parameter* $\theta = 4$. *The results are based on an original sample size of* $n \in \{5, 10, 20, 25, 50, 75\}$, *while the augmented smooth bootstrap sample size is* $m = 10\,000$. *Each boxplot is based on* $M = 2\,000$ *independent reruns. The red dashed line indicates the theoretical value of* ρ_S *while black dots indicate the means.*
Bottom: Mean squared error for estimation of Spearman's rho. The results are based on an original sample size of $n \in \{5, 10, 20, 25, 50, 75\}$, *while the augmented smooth bootstrap sample size is* $m = 10\,000$.

main objects under consideration. Concerning the expected density function estimator, we immediately obtain that $\phi_{\widehat{\boldsymbol{f}}_n^{\mathbb{E}}}(\boldsymbol{t}) = \phi_{\boldsymbol{X}}(\boldsymbol{t})\phi_{\boldsymbol{Y}_{H_n}}(\boldsymbol{t})$ due to the independence of $\boldsymbol{X}$ and $\boldsymbol{Y}_{H_n}$; see Theorem A.1. The characteristic function of $\boldsymbol{Y}_{H_n}$ is obtained in terms of the characteristic function of $\boldsymbol{Y}$, or equivalently $\mathbf{k}$, as $\phi_{\boldsymbol{Y}_{H_n}}(\boldsymbol{t}) = \phi_{\mathbf{k}}\left(H_n^{1/2}\boldsymbol{t}\right)$, see Theorem A.2, which overall leads to

$$\phi_{\widehat{\boldsymbol{f}}_n^{\mathbb{E}}}(\boldsymbol{t}) = \phi_{\boldsymbol{X}}(\boldsymbol{t})\phi_{\mathbf{k}}\left(H_n^{1/2}\boldsymbol{t}\right). \tag{30}$$

Via Theorem A.2 also the characteristic function of the conditional density estimate $\widehat{\boldsymbol{f}}_n^{|\mathbb{X}}$ can be computed as

$$\phi_{\widehat{\boldsymbol{f}}_n^{|\mathbb{X}}}(\boldsymbol{t}) = \phi_{\mathbf{k}}\left(\left(H_n^{1/2}\right)^{\top}\boldsymbol{t}\right)\frac{1}{n}\sum_{i=1}^{n} e^{\mathrm{i}\boldsymbol{t}^{\top}\boldsymbol{x}_i}, \tag{31}$$

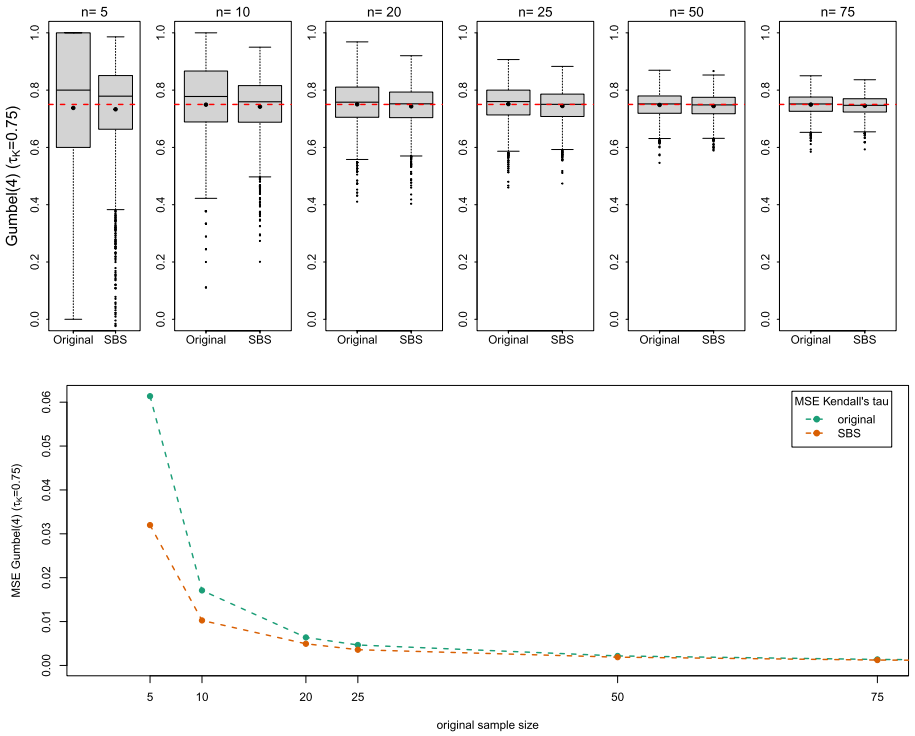

FIG 15. *Top: Estimated Kendall's tau for the bivariate Gumbel copula with parameter* $\theta = 4$. *The results are based on an original sample size of* $n \in \{5, 10, 20, 25, 50, 75\}$, *while the augmented smooth bootstrap sample size is* $m = 10\,000$. *Each boxplot is based on* $M = 2\,000$ *independent reruns. The red dashed line indicates the theoretical value of* τ_K *while black dots indicate the means.*
Bottom: Mean squared error for estimation of Kendall's tau. The results are based on an original sample size of $n \in \{5, 10, 20, 25, 50, 75\}$, *while the augmented smooth bootstrap sample size is* $m = 10\,000$.

from which we can recover the characteristic function of the expected kernel density estimate given in (30) when taking the expectation with respect to the underlying random vector $\boldsymbol{X}$.

Characteristic functions will also play an important role in determining the distance between two distribution functions. The Kolmogorov-Smirnov distance between two d-dimensional distribution functions is denoted by

$$\mathrm{d_{KS}}\left(\boldsymbol{F}, \boldsymbol{G}\right) = \sup_{\boldsymbol{x} \in \mathbb{R}^d} \left|\boldsymbol{F}(\boldsymbol{x}) - \boldsymbol{G}(\boldsymbol{x})\right|. \tag{32}$$

The distance $\mathrm{d_{KS}}$ can be bounded by the average scaled difference of the associated characteristic functions. In the univariate case this smoothing inequality is linked to the Berry–Esseen theorem and for $T > 0$ takes the form

$$\mathrm{d_{KS}}\left(F, G\right) \leq \frac{1}{\pi} \int_{-T}^{T} \left| \frac{\phi_X(t) - \phi_Y(t)}{t} \right| \mathrm{d}t + \frac{24}{\pi T} \sup_{x \in \mathbb{R}} \left|G'(x)\right|,$$

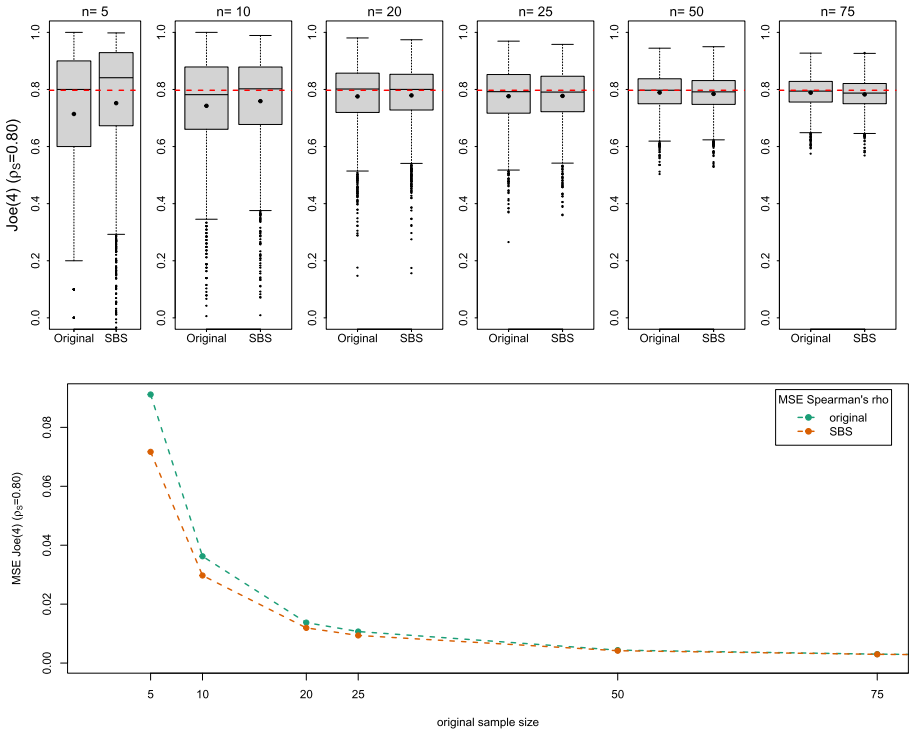

FIG 16. *Top: Estimated Spearman's rho for the bivariate Joe copula with parameter* $\theta = 4$. *The results are based on an original sample size of* $n \in \{5, 10, 20, 25, 50, 75\}$, *while the augmented smooth bootstrap sample size is* $m = 10\,000$. *Each boxplot is based on* $M = 2\,000$ *independent reruns. The red dashed line indicates the theoretical value of* ρ_S *while black dots indicate the means.*
Bottom: Mean squared error for estimation of Spearman's rho. The results are based on an original sample size of $n \in \{5, 10, 20, 25, 50, 75\}$, *while the augmented smooth bootstrap sample size is* $m = 10\,000$.

see, for example, Feller 1971, Lemma 2, page 538. In the multivariate case extensions are available. To keep the notation to a minimum we present the bivariate case where we use the presentation of Heuberger and Kropf (2018). Here the cut-off interval $[-T, T]$ is generalized to a ball. Alternatively, a representation in terms of a cube $[-T, T]^2$ is also possible; see Sadikova (1966).

Theorem A.3 (Smoothing inequality; Heuberger and Kropf 2018, Theorem 2)**.** *Denote by* $\boldsymbol{X} \sim \boldsymbol{F_X}$ *and* $\boldsymbol{Y} \sim \boldsymbol{F_Y}$ *two 2-dimensional random vectors. Assume that* $\boldsymbol{F_Y}$ *is differentiable. Let* $T > 0$ *be fixed, then*

$$\begin{aligned} &\mathrm{d_{KS}}\left(\boldsymbol{F_X}, \boldsymbol{F_Y}\right) \\ &\leq \frac{2}{(2\pi)^2} \int_{\|\boldsymbol{t}\|_2 \leq T} \left| \frac{\phi_{\boldsymbol{X}}(\boldsymbol{t}) - \phi_{X_1}(t_1)\phi_{X_2}(t_2) - \phi_{\boldsymbol{Y}}(\boldsymbol{t}) + \phi_{Y_1}(t_1)\phi_{Y_2}(t_2)}{\prod_{i=1}^{2} t_i} \right| \mathrm{d}\boldsymbol{t} \end{aligned}$$

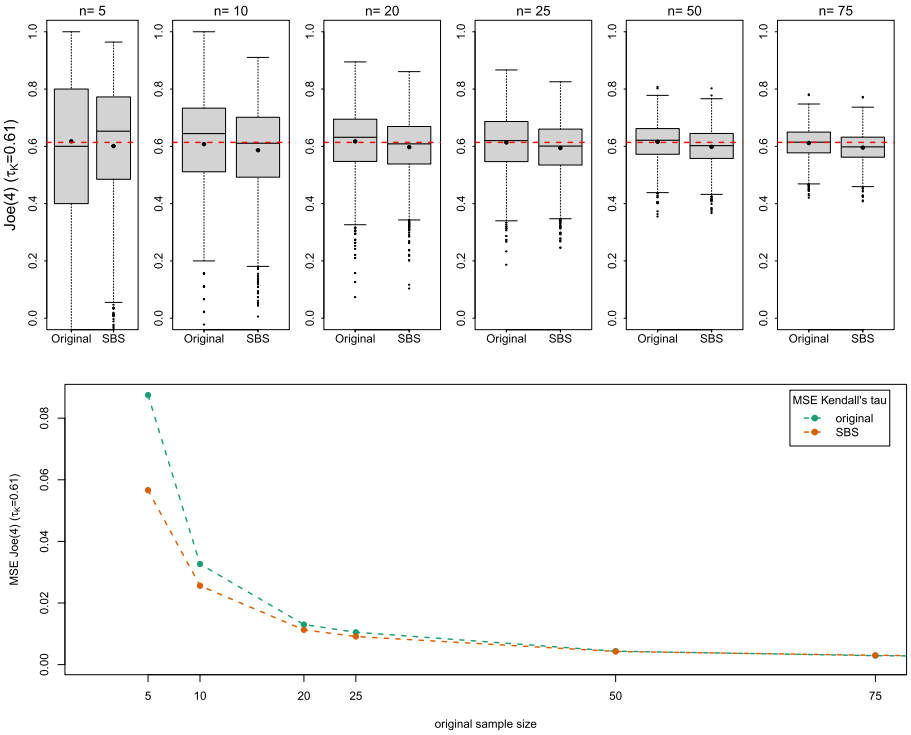

FIG 17. *Top: Estimated Kendall's tau for the bivariate Joe copula with parameter* $\theta = 4$. *The results are based on an original sample size of* $n \in \{5, 10, 20, 25, 50, 75\}$, *while the augmented smooth bootstrap sample size is* $m = 10\,000$. *Each boxplot is based on* $M = 2\,000$ *independent reruns. The red dashed line indicates the theoretical value of* τ_K *while black dots indicate the means.*
Bottom: Mean squared error for estimation of Kendall's tau. The results are based on an original sample size of $n \in \{5, 10, 20, 25, 50, 75\}$, *while the augmented smooth bootstrap sample size is* $m = 10\,000$.

$$
\begin{aligned}
&+ \frac{2}{\pi} \int_{|t_1| \le T} \left| \frac{\phi_{X_1}(t_1) - \phi_{Y_1}(t_1)}{t_1} \right| \mathrm{d}t_1 + \frac{2}{\pi} \int_{|t_2| \le T} \left| \frac{\phi_{X_2}(t_2) - \phi_{Y_2}(t_2)}{t_2} \right| \mathrm{d}t_2 \\
&+ 2 \frac{\left\| \frac{\partial \boldsymbol{F_Y}}{\partial y_1} \right\|_\infty + \left\| \frac{\partial \boldsymbol{F_Y}}{\partial y_2} \right\|_\infty}{T} \left(\frac{12}{\pi} + \sqrt[3]{\frac{32}{\pi \left(1 - \left(\frac{3}{4}\right)^{1/2}\right)}} \right) \\
&+ \frac{4 \left\| f_{Y_1} \right\|_\infty}{T} \left(\frac{12}{\pi} + \sqrt[3]{\frac{128}{\pi}} \right) + \frac{4 \left\| f_{Y_2} \right\|_\infty}{T} \left(\frac{12}{\pi} + \sqrt[3]{\frac{128}{\pi}} \right).
\end{aligned}
$$

Proof. This is the bivariate case of Heuberger and Kropf 2018, Theorem 2. □

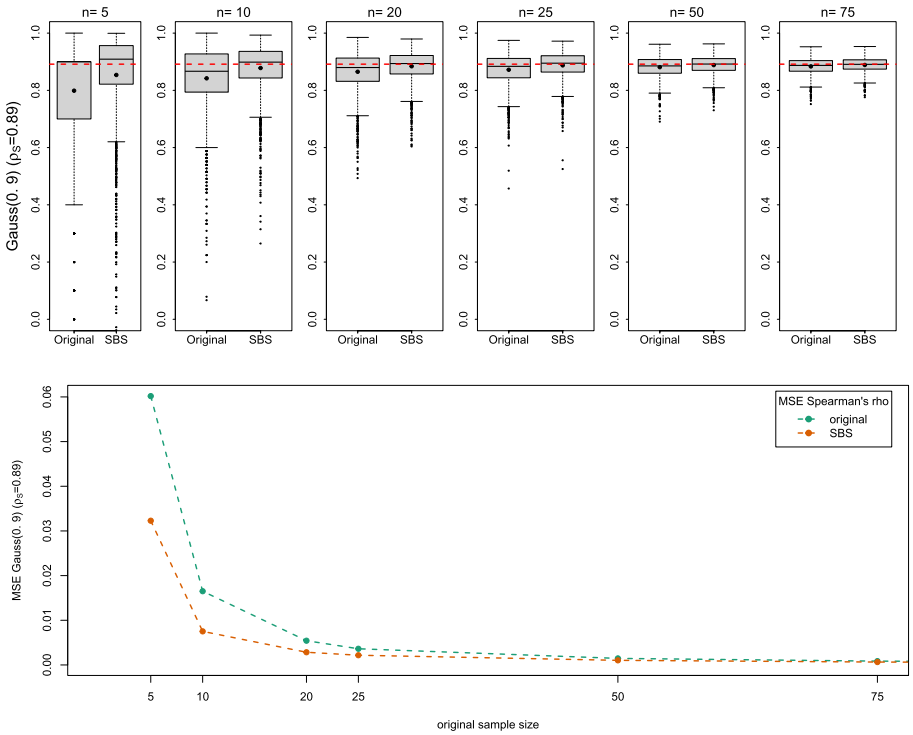

FIG 18. *Top: Estimated Spearman's rho for the bivariate Gaussian copula with* $\rho = 0.9$ *The results are based on an original sample size of* $n \in \{5, 10, 20, 25, 50, 75\}$*, while the augmented smooth bootstrap sample size is* $m = 10\,000$*. Each boxplot is based on* $M = 2\,000$ *independent reruns. The red dashed line indicates the theoretical value of* ρ_S *while black dots indicate the means.*
Bottom: Mean squared error for estimation of Spearman's rho. The results are based on an original sample size of $n \in \{5, 10, 20, 25, 50, 75\}$*, while the augmented smooth bootstrap sample size is* $m = 10\,000$.

Appendix B: Elliptically distributed random vectors and copulas

For our considerations the class of elliptical distributions will play a central role. Textbook introductions can be found in Fang, Kotz and Ng (1990) and McNeil, Frey and Embrechts (2015). Before going to elliptical distributions we first introduce spherical distributions as a necessary stepping stone.

Definition B.1 (Spherical distribution; McNeil, Frey and Embrechts 2015, Definition 3.18)**.** *A d-dimensional random vector* $\boldsymbol{Y} = (Y_1, \ldots, Y_d)^\top$ *has a spherical distribution if, for every orthogonal matrix* $U \in \mathbb{R}^{d \times d}$*,* $UU^\top = U^\top U = \boldsymbol{I}_d$*, we have* $U\boldsymbol{X} \stackrel{d}{=} \boldsymbol{X}$.

Spherical distributions can equivalently be characterized via their characteristic functions or by randomly scaling a uniform distribution on the unit sphere.

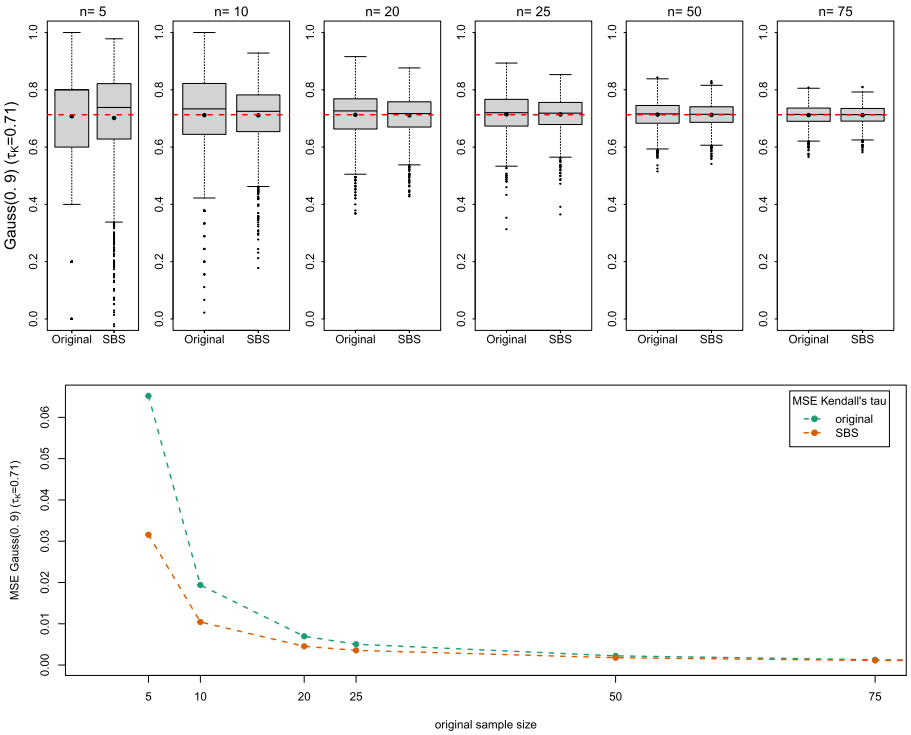

FIG 19. *Top: Estimated Kendall's tau for the bivariate Gaussian copula with* $\rho = 0.9$ *The results are based on an original sample size of* $n \in \{5, 10, 20, 25, 50, 75\}$*, while the augmented smooth bootstrap sample size is* $m = 10\,000$*. Each boxplot is based on* $M = 2\,000$ *independent reruns. The red dashed line indicates the theoretical value of* τ_K *while black dots indicate the means.*
Bottom: Mean squared error for estimation of Kendall's tau. The results are based on an original sample size of $n \in \{5, 10, 20, 25, 50, 75\}$*, while the augmented smooth bootstrap sample size is* $m = 10\,000$*.*

Theorem B.1 (Equivalent characterization of spherical distributions; McNeil, Frey and Embrechts 2015, Theorem 3.19 and Theorem 3.22)**.** *Denote by* $\boldsymbol{Y}$ *a* d*-dimensional random vector. Then the following are equivalent:*

1. $\boldsymbol{Y}$ *has a spherical distribution,*
2. *there exists a real valued function* $\psi\colon [0,\infty) \to [-1,1]$ *with* $\psi(0) = 1$ *such that the characteristic function of* $\boldsymbol{Y}$ *is given by* $\phi_{\boldsymbol{Y}}(\boldsymbol{t}) = \psi\left(\boldsymbol{t}^\top \boldsymbol{t}\right)$*,*
3. $\boldsymbol{Y}$ *has the stochastic representation* $\boldsymbol{Y} \overset{d}{=} R\boldsymbol{S}$*, where* $\boldsymbol{S}$ *is uniformly distributed on the unit sphere* $\{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\|_2 = 1\}$ *and* R *is an almost surly non-negative random variable independent of* $\boldsymbol{S}$*.*

The random variable R *specific to* $\boldsymbol{Y}$ *is called the radial distribution of* $\boldsymbol{Y}$*. The function* ψ *is called the characteristic generator of* $\boldsymbol{Y}$*. We denote the distribution of* $\boldsymbol{Y}$ *by* $S_d(\psi)$*. The subclass of spherical distributions for which* $\mathbb{P}[R = 0] = 0$

TABLE 2
Characteristic generator ψ and value of its derivative at 0 for popular spherical and elliptical models.

Distribution	$\psi(u)$	$\lim_{u\to 0+}\psi'(u)$
Gauss	$\exp(-u/2)$	$-1/2$
Laplace	$(1+u/2)^{-1}$	$-1/2$
Student t $(\nu=1)$	$\exp\left(-\sqrt{u}\right)$	$-\infty$
Student t $(\nu=2)$	$K_1\left(\sqrt{2x}\right)\sqrt{2x}$	$-\infty$
Student t $(\nu=4)$	$K_2\left(\sqrt{4x}\right)2x$	-1
Student t $(\nu>2)$	$\frac{K_{\nu/2}\left(\sqrt{\nu x}\right)\left(\sqrt{\nu x}\right)^{\nu/2}}{\Gamma(\nu/2)2^{\nu/2-1}}$	$-\nu/(2\nu-4)$

is denoted by $S_d^+(\psi)$.

In Table 2 we collect the characteristic generators and the value of their derivative at 0 for popular spherical and elliptical models. Based on Definition B.1 we can now go on to define elliptical distributions.

Definition B.2 (Elliptical distribution; McNeil, Frey and Embrechts 2015, Definition 3.26)**.** *A d-dimensional random vector $\boldsymbol{X}$ has an elliptical distribution if*

$$\boldsymbol{X} \stackrel{d}{=} \boldsymbol{\mu} + A\boldsymbol{Y}, \tag{33}$$

where $\boldsymbol{Y} \sim S_d(\psi)$ and $A \in \mathbb{R}^{d\times k}$ and $\boldsymbol{\mu} \in \mathbb{R}^d$ are a matrix and vector of constants, respectively. The distribution of $\boldsymbol{X}$ is denoted by $E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$ where $\boldsymbol{\Sigma} = AA^\top$ is called the dispersion matrix. The radial distribution of $\boldsymbol{X}$ is the radial distribution R associated to $\boldsymbol{Y}$ in (33).

Given the stochastic representation of spherical random vectors an elliptical random vector naturally has the representation

$$\boldsymbol{X} \stackrel{d}{=} \boldsymbol{\mu} + RA\boldsymbol{S}$$

where R is the associated radial distribution, A is a non-random matrix and $\boldsymbol{S}$ is uniformly distributed on the unit sphere. Based on the characteristic function of spherical random vectors, the characteristic function of an elliptical random vector $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$ is given by

$$\phi_{\boldsymbol{X}}(\boldsymbol{t}) = e^{\mathrm{i}\boldsymbol{t}^\top \boldsymbol{\mu}}\psi\left(\boldsymbol{t}^\top \Sigma \boldsymbol{t}\right).$$

If the radial distribution of an elliptical random vector has a finite second moment, the mean vector and covariance matrix admit the following convenient expressions.

Theorem B.2 (Moments of elliptical random vectors; Fang, Kotz and Ng 1990, Theorem 2.17, p. 43)**.** *If $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$ and $\mathbb{E}\left[R^2\right] < \infty$, then $\mathbb{E}[\boldsymbol{X}] = \boldsymbol{\mu}$ and $\operatorname{cov}[\boldsymbol{X}] = \frac{\mathbb{E}[R^2]}{\operatorname{rank}(\boldsymbol{\Sigma})}\boldsymbol{\Sigma} = -2\psi'(0)\boldsymbol{\Sigma}$.*

Remark B.1 (Re-parameterization of elliptical random vectors)**.** *The parameterization of an elliptical random vector is non-unique since the dispersion matrix and the characteristic generator can be rescaled. From the characteristic function representation it is clear that the pairs* $(\boldsymbol{\Sigma}, \psi)$ *and* $(c\boldsymbol{\Sigma}, \psi(\cdot/c))$ *lead to the same distribution for every* $c > 0$*. If* $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$ *we can consequently set* $\widetilde{\boldsymbol{\Sigma}} = -2\psi'(0)\boldsymbol{\Sigma}$ *and* $\widetilde{\psi}(u) = \psi(u/(-2\psi'(0)))$ *to obtain a parameterization* $\boldsymbol{X} \sim E_d\left(\boldsymbol{\mu}, \widetilde{\boldsymbol{\Sigma}}, \widetilde{\psi}\right)$ *such that* $\operatorname{cov}[\boldsymbol{X}] = \widetilde{\boldsymbol{\Sigma}}$*.*

Elliptically distributed random vectors are especially well behaved when considering affine transformations and sums as shown in the following theorems.

Theorem B.3 (Linear combinations of elliptical random vectors; Fang, Kotz and Ng 1990, Theorem 2.16, p. 43)**.** *Denote by* $\boldsymbol{B}$ *a* $k \times d$ *matrix and by* $\boldsymbol{b} \in \mathbb{R}^k$ *a vector. If* $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$ *with* $\operatorname{rank}(\boldsymbol{\Sigma}) = k$ *then* $\boldsymbol{B}\boldsymbol{X} + \boldsymbol{b} \sim E_k(\boldsymbol{B}\boldsymbol{\mu} + \boldsymbol{b}, \boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}^\top, \psi)$*.*

As an application of Theorem B.3 we can obtain the marginal distributions of an elliptical random vector by considering $X_j = e_j^\top \boldsymbol{X}$, where $e_j = (0, \ldots, 1, \ldots, 0)^\top$ denotes the jth unit vector. We then obtain

$$X_j \sim E_1(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_{jj}, \psi). \tag{34}$$

Summation of elliptical random vectors produces again an elliptical random vector under certain conditions. Our first theorem in this direction is due to Hult and Lindskog (2002) and relaxes the commonly requested independence assumption by allowing for non-independent radial distributions.

Theorem B.4 (Hult and Lindskog 2002, Theorem 4.1)**.** *Let* R_1 *and* R_2 *be non-negative random variables and let* $\boldsymbol{X}_1 = \boldsymbol{\mu}_1 + R_1 A_1 \boldsymbol{S}_1 \sim E_d(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}, \psi_1)$ *and* $\boldsymbol{X}_2 = \boldsymbol{\mu}_2 + R_2 A_2 \boldsymbol{S}_2 \sim E_d(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}, \psi_2)$*, where the random vectors* (R_1, R_2)*,* $\boldsymbol{Z}_1$ *and* $\boldsymbol{Z}_2$ *are mutually independent. Then* $\boldsymbol{X}_1 + \boldsymbol{X}_2 \sim E_d(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2, \boldsymbol{\Sigma}, \psi_3)$*. Moreover, if* R_1 *and* R_2 *are independent, then* $\psi_3(u) = \psi_1(u)\psi_2(u)$*.*

In our context a slight generalization of Theorem B.4 for non-equal dispersion matrices is needed. This is essentially a combination of Theorem B.4 and Lindskog, McNeil and Schmock 2003, Lemma 1.

Corollary B.1. *Let* R_1 *and* R_2 *be non-negative random variables and let* $\boldsymbol{X}_1 = \boldsymbol{\mu}_1 + R_1 A_1 \boldsymbol{S}_1 \sim E_d(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}, \psi_1)$ *and* $\boldsymbol{X}_2 = \boldsymbol{\mu}_2 + R_2 A_2 \boldsymbol{S}_2 \sim E_d(\boldsymbol{\mu}_2, c\boldsymbol{\Sigma}, \psi_2)$*, where the random vectors* (R_1, R_2)*,* $\boldsymbol{Z}_1$ *and* $\boldsymbol{Z}_2$ *are mutually independent and* $c > 0$*. Then* $\boldsymbol{X}_1 + \boldsymbol{X}_2 \sim E_d(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2, \boldsymbol{\Sigma}, \psi_c)$ *for some characteristic generator* ψ_c*. Moreover, if* R_1 *and* R_2 *are independent, then* $\psi_c(u) = \psi_1(u)\psi_2(cu)$*.*

Proof. Denote by ϕ_1 and ϕ_2 the conditional characteristic functions

$$\phi_1(\boldsymbol{t}) = \mathbb{E}\left[e^{\mathrm{i}(\boldsymbol{\mu}_1 + R_1 A_1 \boldsymbol{S}_1)^\top \boldsymbol{t}} \middle| R_1 = r_1\right] = e^{\mathrm{i}\boldsymbol{t}^\top \boldsymbol{\mu}_1} \psi_1\left(r_1^2 \boldsymbol{t}^\top \boldsymbol{\Sigma} \boldsymbol{t}\right),$$
$$\phi_2(\boldsymbol{t}) = \mathbb{E}\left[e^{\mathrm{i}(\boldsymbol{\mu}_2 + R_2 A_2 \boldsymbol{S}_2)^\top \boldsymbol{t}} \middle| R_1 = r_1\right] = e^{\mathrm{i}\boldsymbol{t}^\top \boldsymbol{\mu}_2} \psi_2^{r_1}\left(\boldsymbol{t}^\top c\boldsymbol{\Sigma} \boldsymbol{t}\right),$$

where ψ_1 is the characteristic generator of $A_1\boldsymbol{S}_1$ and $\psi_2^{r_1}$ is the characteristic generator of $R_2A_2\boldsymbol{S}_2$ given $R_1 = r_1$. Following the same steps as in the proof of Hult and Lindskog 2002, Theorem 4.1, we obtain

$$\phi_{\boldsymbol{X}_1+\boldsymbol{X}_2}(\boldsymbol{t}) = e^{\mathrm{i}\boldsymbol{t}^\top(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2)}\mathbb{E}\left[\psi_1\left(R_1^2\boldsymbol{t}^\top\boldsymbol{\Sigma}\boldsymbol{t}\right)\psi_2^{R_1}\left(c\boldsymbol{t}^\top\boldsymbol{\Sigma}\boldsymbol{t}\right)\right],$$

showing that $\boldsymbol{X}_1 + \boldsymbol{X}_2 \sim E_d(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2, \boldsymbol{\Sigma}, \psi_c)$ with characteristic generator

$$\psi_c(u) = \int_0^\infty \psi_1\left(r_1^2u\right)\psi_2^{r_1}\left(cu\right)\,\mathrm{d}F_{R_1}(r_1),$$

where F_{R_1} is the marginal distribution of R_1. If R_1 and R_2 are independent we directly get $\psi_c(u) = \psi_1(u)\psi_2(cu)$ without conditioning on $R_1 = r_1$. □

If a spherical or elliptical random vector is absolutely continuous with respect to the Lebesgue measure, the density takes a particular form.

Theorem B.5 (Density generator; McNeil, Frey and Embrechts 2015, Equation (3.46))**.** *If a spherical random vector $\boldsymbol{Y} \sim S_d(\psi)$ is absolutely continuous with respect to the Lebesgue measure, the density $f_{\boldsymbol{Y}}$ for $\boldsymbol{y} \in \mathbb{R}^d$ takes the form $f_{\boldsymbol{Y}}(\boldsymbol{y}) = g_{\boldsymbol{Y}}(\boldsymbol{y}^\top\boldsymbol{y})$, where $g_{\boldsymbol{Y}}$ is a positive function $g_{\boldsymbol{Y}}\colon [0,\infty) \to [0,\infty), t \mapsto g_{\boldsymbol{Y}}(t)$. The function $g_{\boldsymbol{Y}}$ is called the* density generator *of $\boldsymbol{Y}$.*

If an elliptical random vector $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$ is absolutely continuous with respect to the Lebesgue measure, the density $f_{\boldsymbol{X}}$ for $\boldsymbol{x} \in \mathbb{R}^d$ takes the form $f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{\sqrt{\det(\boldsymbol{\Sigma})}}g_{\boldsymbol{Y}}\left((\boldsymbol{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)$, where $g_{\boldsymbol{Y}}$ is the density generator of the spherical distribution $\boldsymbol{Y}$ associated to $\boldsymbol{X}$.

Based on Sklar's theorem every elliptically distributed random vector gives rise to an associated copula. Due to the fact that the mean and variances are only marginal attributes they do not play a role when focusing on the dependence structure inherent to a given elliptical random vector. Hence it is sufficient to neglect location parameters and to only consider correlation matrices in the following definition. For a one-dimensional distribution function F we denote its quantile function by F^{-1}.

Definition B.3 (Elliptical copula)**.** *Denote by $\boldsymbol{R}$ a correlation matrix and by ψ a characteristic generator. The* elliptical copula *$\boldsymbol{C}_{\boldsymbol{R},\psi}$ is given by the copula associated to the d-dimensional random vector $\boldsymbol{X} \sim E_d(\boldsymbol{0}, \boldsymbol{R}, \psi)$ by virtue of Sklar's theorem*

$$\boldsymbol{C}_{\boldsymbol{R},\psi}(u_1,\ldots,u_d) = \mathbb{P}\left[X_1 \le F^{-1}(u_1),\ldots,X_d \le F^{-1}(u_d)\right],$$

where $F = E_1(0,1,\psi)$ is the marginal distribution common to $X_1,\ldots,X_d$.

Remark B.2. *Given that copulas are invariant under strictly increasing transformations, different elliptical random vectors (and their parameterizations) can give rise to the same elliptical copula. Specifically, the copula $\boldsymbol{C}_{\boldsymbol{X}}$ of an elliptical random vector $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$ is the same as the elliptical copula $\boldsymbol{C}_{\mathrm{corr}[\boldsymbol{X}],\psi}$. This can readily be seen by considering a random vector $\boldsymbol{Y}$ with*

elliptical copula $\boldsymbol{C}_{\mathrm{corr}[\boldsymbol{X}],\psi}$*, i.e.,* $\boldsymbol{Y} \sim E_d(\mathbf{0}, \mathrm{corr}\,[\boldsymbol{X}], \psi)$ *and defining* $\boldsymbol{D} = \mathrm{diag}(\mathrm{sd}[X_1], \ldots, \mathrm{sd}[X_d])/\sqrt{-2\psi'(0)}$*. This yields* $\boldsymbol{DY} + \boldsymbol{\mu} \sim E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$ *and hence* $\boldsymbol{DY} + \boldsymbol{\mu} \overset{d}{=} \boldsymbol{X}$*. We therefore have* $\boldsymbol{C}_{\boldsymbol{X}} = \boldsymbol{C}_{\mathrm{corr}[\boldsymbol{X}],\psi}$*, since the transformation applied to* $\boldsymbol{Y}$ *is strictly increasing in every component.*

Considering for example a multivariate normal random vector $\boldsymbol{X} \sim E_d(\mathbf{0}, c\boldsymbol{R}, \psi)$*, where* $\boldsymbol{R}$ *is a correlation matrix and* $c > 0$*, we have a characteristic generator* $\psi(u) = \exp(-u/2)$ *with* $-2\psi'(0) = 1$*. Adjusting the margins by* $\widetilde{D}\boldsymbol{X}$*, where* $\widetilde{D} = \mathrm{diag}(1/\sqrt{c}, \ldots, 1/\sqrt{c})$*, corrects the correlation structure to* $\mathrm{corr}\left[\widetilde{D}\boldsymbol{X}\right] = \boldsymbol{R}$*. This shows that the random vectors* $\boldsymbol{X}$ *and* $\widetilde{D}\boldsymbol{X}$ *share the same Gaussian copula. Alternatively, the copula can also be obtained by the multivariate probability integral transform which in this case takes the form* $\left(\Phi\left(\frac{X_1}{\sqrt{c}}\right), \ldots, \Phi\left(\frac{X_d}{\sqrt{c}}\right)\right)$.

Appendix C: Numerical results

Figures 12 to 19 contain the numerical results concerning the estimation of Spearman's rho and Kendall's tau using the smooth bootstrap discussed in Section 5.2.

Acknowledgments

We thank an anonymous reviewer for helpful comments which improved the quality of the paper.

References

Abdous, B., Genest, C. and Rémillard, B. (2005). Dependence Properties of Meta-Elliptical Distributions. In *Statistical Modeling and Analysis for Complex Data Problems* (P. Duchesne and B. Rémillard, eds.) 1–15. Springer. MR2189528

Azzalini, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika* **68** 326–328. MR0614972

Bingham, N. H. (1972). A Tauberian Theorem for Integral Transforms of Hankel Type. *Journal of the London Mathematical Society* **2** 493–503. MR0435748

Bowman, A., Hall, P. and Prvan, T. (1998). Bandwidth selection for the smoothing of distribution functions. *Biometrika* **4** 799–808. MR1666695

Coblenz, M., Dyckerhoff, R. and Grothe, O. (2018). Nonparametric Estimation of Multivariate Quantiles. *Environmetrics* **29** 1–23. MR3775342

Devroye, L. and Gyorfi, L. (1985). *Nonparametric Density Estimation: The* L_1 *View*. Wiley. MR0780746

Devroye, L. and Wagner, T. J. (1979). The L_1 Convergence of Kernel Density Estimates. *Ann. Statist.* **7** 1136–1139. MR0536515

Durante, F. and Sempi, C. (2016). *Principles of copula theory*. Chapman & Hall/CRC. MR3443023

Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*. SIAM. MR0659849

Fang, K. T., Kotz, S. and Ng, K. W. (1990). *Symmetric multivariate and related distributions*. Chapman & Hall. MR1071174

Feller, W. (1971). *An introduction to probability theory and its applications*. Wiley. MR1570945

Geenens, G., Charpentier, A. and Paindaveine, D. (2018). Probit transformation for nonparametric kernel estimation of the copula density. *Bernoulli* **23** 1848–1873. MR3624880

Gerstner, T. and Griebel, M. (1998). Numerical integration using sparse grids. *Numerical Algorithms* **18** 209–232. MR1669959

Hall, P. (1992). *The bootstrap and Edgeworth expansion*. Springer. MR1145237

Härdle, W. K., Müller, M., Sperlich, S. and Werwatz, A. (2012). *Nonparametric and Semiparametric Models*. Springer. MR2061786

Heuberger, C. and Kropf, S. (2018). Higher dimensional quasi-power theorem and Berry–Esseen inequality. *Monatshefte für Mathematik* **187** 293–314. MR3850313

Hofert, M., Kojadinovic, I., Mächler, M. and Yan, J. (2018). *Elements of copula modeling with R*. Springer. MR3887600

Hofert, M., Kojadinovic, I., Mächler, M. and Yan, J. (2022). copula: Multivariate Dependence with Copulas R package version 1.0.1.

Hult, H. and Lindskog, F. (2002). Multivariate extremes, aggregation and dependence in elliptical distributions. *Adv. in Appl. Probab.* **34** 587–608. MR1929599

Joarder, A. and Ali, M. (1996). On the Characteristic Function of the Multivariate t-Distribution. *Pak. J. Statist.* **12** 55–62. MR1409301

Joe, H. (2015). *Dependence modeling with copulas*. CRC Press. MR3328438

Kotz, A., Kozubowski, T. and Podgórski, K. (2001). *The Laplace Distribution and Generalizations. A Revisit with Applications to Communications, Economics, Engineering, and Fincance*. Springer Science+Business Media, New York. MR1935481

Lindskog, F., McNeil, A. and Schmock, U. (2003). Kendall's Tau for Elliptical Distributions. In *Credit Risk: Measurement, Evaluation and Management* (G. Bol, G. Nakhaeizadeh, S. Rachev, T. Ridder and K. H. Vollmer, eds.) 149–156. Physica-Verlag HD. MR1675308

Liu, R. and Yang, L. (2008). Kernel estimation of multivariate cumulative distribution function. *Journal of Nonparametric Statistics* **4** 661–677. MR2467701

Mai, J. F. and Scherer, M. (2012). *Simulating Copulas: Stochastic Models, Sampling Algorithms, and Applications*. World Scientific Pub Co Inc. MR3729417

McNeil, A., Frey, R. and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*, 2nd ed. Princeton University. MR3445371

Nadaraya, E. A. (1964a). Some new estimates for distribution functions. *The-*

ory of Probability & its Applications **9** 497–500. MR0166862

Nadaraya, E. A. (1964b). Nekotorye novye ocenki funkcij raspredelenija. *Teor. Veroyatnost. i Primenen* **9** 550–554.

Nagler, T. (2018). Asymptotic analysis of the jittering kernel density estimator. *Mathematical Methods of Statistics* **27** 32–46. MR3800980

Nelsen, R. (2006). *An introduction to copulas*, 2nd ed. Springer. MR2197664

NIST NIST Digital Library of Mathematical Functions. http://dlmf.nist.gov/, Release 1.0.18 of 2018-03-27. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller and B. V. Saunders, eds.

Nolan, J. P. (2013). Multivariate elliptically contoured stable distributions: theory and estimation. *Computational Statistics* **28** 2067–2089. MR3107292

Omelka, M., Gijbels, I. and Veraverbeke, N. (2009). Improved kernel estimation of copulas: Weak convergence and goodness-of-fit testing. *Annals of Statistics* **37** 3023–3058. MR2541454

Packham, N. and Schmidt, W. M. (2010). Latin hypercube sampling with dependence and applications in finance. *Journal of Computational Finance* **13** 81–111. MR2608991

Rudin, W. (1976). *Principles of Mathematical Analysis*, 3rd ed. McGraw-Hill, Inc. MR0385023

Rudin, W. (1991). *Functional Analysis*, 2nd ed. McGraw-Hill, Inc., Singapore. MR1157815

Sadikova, S. M. (1966). On two-dimensional analogues of an inequality of Esseen and their application to the central limit theorem. *Theory of Probability and its Applications* **XI** 325–335. MR0207016

Salvadori, G., Durante, F., De Michele, C., Bernardi, M. and Petrella, L. (2016). A multivariate copula-based framework for dealing with hazard scenarios and failure probabilities. *Water Resources Research* **52** 3701–3721.

Samorodnitsky, G. and Taqqu, M. S. (1994). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman & Hall/CRC, New York. MR1280932

Sasvári, Z. (2013). *Multivariate Characteristic and Correlation Functions*. De Gruyter. MR3059796

Schmid, F. and Schmidt, R. (2007). Nonparametric inference on multivariate versions of Blomqvist's beta and related measures of tail dependence. *Metrika* **66** 323–354. MR2336484

Schmidt, R. (2002). Tail dependence for elliptically contoured distributions. *Mathematical Methods of Operations Research* **55**. MR1919580

Scott, D. W. (2015). *Multivariate Density Estimation: Theory, Practice, and Visualization*, 2nd ed. Wiley. MR3329609

Segers, J., Sibuya, M. and Tsukahara, H. (2017). The empirical beta copula. *Journal of Multivariate Analysis* **155** 35–51. MR3607880

Shao, J. and Tu, D. (1995). *The jackknife and bootstrap*. Springer Science & Business Media. MR1351010

Shorten, C. and Khoshgoftaar, T. M. (2019). A survey on Image Data

Augmentation for Deep Learning. *Journal of Big Data* **6** 241–267.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall. MR0848134

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* **8** 229–231. MR0125600

Smolyak, S. (1963). Quadrature and interpolation formulas for tensor products of certain classes of functions. *Soviet Mathematics (Translation of Doklady Akademii Nauk SSSR)* **4** 240–243.

Song, D.-K., Park, H.-J. and Kim, H.-M. (2014). A Note on the Characteristic Function of Multivariate t Distribution. *Communications for Statistical Applications and Methods* **21** 81–91.

Sutradhar, B. C. (1986). On the Characteristic Function of Multivariate Student t-Distribution. *The Canadian Journal of Statistics* **4** 329–337. MR0876759

Taha, A. and Hanbury, A. (2015). An efficient algorithm for calculating the exact Hausdorff distance. *IEEE transactions on pattern analysis and machine intelligence* **37** 2153–2163.

Taylor, L. and Nitschke, G. (2018). Improving Deep Learning with Generic Data Augmentation. In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)* 1542-1547.

Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing*. Chapman & Hall. MR1319818

Wong, S. C., Gatt, A., Stamatescu, V. and McDonnell, M. D. (2016). Understanding Data Augmentation for Classification: When to Warp? In *2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA)* 1-6. MR3793154