FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

# HABILITATION THESIS

Marek Omelka

# Nonparametric Estimation of Copulas, Conditional Copulas and Conditional Distribution Functions

Mathematics – Probability
and Mathematical Statistics

Prague, April 2017

# Contents

# Preface

The present habilitation thesis contributes to the theory of distribution function estimation with the emphasis on copulas. The thesis summarises the achievements of the following five papers:


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

**[2]** Gijbels, I., Veraverbeke, N., and Omelka, M. (2011). Conditional copulas, association measures and their applications, *Computational Statistics and Data Analysis*, 55, 1919–1932.

**[3]** Veraverbeke, N., Omelka, M., and Gijbels, I. (2011). Estimation of a conditional copula and association measures, *Scandinavian Journal of Statistics*, Vol. 38, 766–780.

**[4]** Gijbels, I., Omelka, M., and Veraverbeke, N. (2015). Estimation of a Copula when a Covariate Affects only Marginal Distributions, *Scandinavian Journal of Statistics*, Vol. 42, 1109–1126.

**[5]** Veraverbeke, N., Gijbels, I. and Omelka, M. (2014) Pre-adjusted nonparametric estimation of a conditional distribution function. *Journal of the Royal Statistical Society: Series B*, Vol. 76, 399–438.


These papers were written by Marek Omelka in collaboration with Irène Gijbels (KU Leuven) and Noël Veraverbeke (Hasselt University).

Paper **[1]** concerns kernel estimation of copulas. In this paper it is proposed to use a bandwidth shrinking factor in order to control the corner bias problem of the existing kernel estimators of copulas. It is also shown that this corner bias problem can be circumvented by a transformation approach. Details can be found in Chapter 2.

Paper **[2]** introduces nonparametric estimation of copulas and corresponding conditional measures of association based on copulas. The theoretical properties of the suggested estimators are investigated in paper **[3]**. The content of both papers is summarized in Chapter 3.

Paper **[4]** proposes a nonparametric estimator of a copula under the pairwise simplifying assumption and investigate its asymptotic properties. The derived theoretical results can be found in Chapter 4.

Finally, in paper **[5]** the authors combine nonparametric estimation together with the idea of a working model that does not have to hold exactly. They show that transforming the response observations so that they are 'less' dependent on the covariate usually results in an estimator of a conditional distribution function with considerably better bias properties than the standard nonparametric estimator. For details see Chapter 5.

The thesis contains the introduction to the problem and formulation of the main theoretical results proved in the papers. Some comments on further results and extensions are also included. Please note, that the notation in the thesis is unified and therefore it need not be the same as in the published versions of manuscripts which are attached to the thesis.

# Chapter 1

# Introduction

Dependence modelling (or at least measuring) is one of the most common tasks when analysing data. The tests of independence based on Pearson's and/or Spearman's correlation coefficient can be found almost in all textbooks of statistics. But researchers are often not satisfied by just rejecting the independence. They want to know how the dependence can be described. Additionally, in some applications it is already well understood that the variables are not independent (e.g. loss variables in insurance or daily maximum temperatures at different locations) and the task is to find a suitable model for the joint distribution.

## 1.1 Introducing a copula function

The copulas became very popular in the last twenty years in dependence modelling. The reason is that using copulas in applications enable to separate the modelling of marginals from modelling of dependence structure. The formal definition is as follows.

Consider a random vector $\boldsymbol{Y} = (Y_1, \dots, Y_d)^{\mathsf{T}}$ with joint cumulative distribution function $H$ and marginal distribution functions $F_1, \dots, F_d$. According to Sklar's theorem (see e.g. Nelsen, 2006) there exists a $d$-variate function $C$ such that

$$H(y_1, \dots, y_d) = C\big(F_1(y_1), \dots, F_d(y_d)\big) \ . \tag{1.1}$$

The function $C$ is called a copula. If the marginal distribution functions $F_1, \dots, F_d$ are continuous, then the function $C$ is unique and

$$C(u_1, \dots, u_d) = H\big(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)\big),$$

where, for $j = 1, \dots, d$, $F_j^{-1}(u) = \inf\big\{y : F_j(y) \geq u\big\}$, with $u \in [0, 1]$, is the quantile function of $F_j$. The copula $C$ 'couples' the joint distribution function $H$ to its univariate marginals, capturing as such the dependence structure between the components of $\boldsymbol{Y} = (Y_1, \dots, Y_d)^{\mathsf{T}}$.

The copula function can be also viewed as a distribution function of the transformed variables

$$\big(U_1, \dots, U_d\big)^{\mathsf{T}} = \big(F_1(Y_1), \dots, F_d(Y_d)\big)^{\mathsf{T}}.$$

For simplicity of the presentation, we will restrict to the case $d = 2$ in what follows.

In practice we usually do not know the marginal distributions $F_1$ and $F_2$ and so the variables $(U_1, U_2)^\mathsf{T}$ are not observable. Thus one needs to work with the vectors of pseudo-observations (see e.g. $(\widehat{U}_{1i}, \widehat{U}_{2i})^\mathsf{T}$ in (2.1)) that are not independent. This makes the theoretical investigation of estimators of copulas more challenging.

Methods for estimation of copulas usually depend on how much we are willing to assume about the joint distribution function $H$. In Chapter 2 we concentrate on nonparametric estimation of copulas. Having a nonparametric estimator is important of its known when one does not want to make any parametric assumption. But it is also important in parametric inference for model checking.

In Chapter 2 we introduce an empirical copula function and then we focus on contribution of Omelka et al. (2009) to kernel estimation of copula functions. It is well known that kernel (smooth) estimators of distribution functions can have better finite sample properties than the standard empirical estimators (see e.g. Reiss, 1981). The problem of a copula function is that its support is a unit square and thus one needs to tackle near borders bias problems of kernel estimators. Omelka et al. (2009) noticed that the standard approaches to this problem, i.e. using local linear kernels (as suggested in Chen and Huang, 2007) or mirror-reflection methods, do not work sufficiently well for many of the commonly used families of copulas. That is why Omelka et al. (2009) introduced 'a bandwidth shrinking function' $b$ (see Chapters 2.2.1 and 2.2.2). Further they showed that alternatively the problem can be circumvented by using a transformation estimator (see Chapter 2.2.3). Finally, they proved that all the suggested estimators have the same (first-order) asymptotic properties as an empirical copula function (see Theorem 2.2) but usually have better finite sample properties (see simulations in Omelka et al., 2009).

## 1.2 A conditional copula function

Suppose now that instead of just a random vector $(Y_1, Y_2)^\mathsf{T}$ (as in the previous section) now we observe a three-dimensional vector $(Y_1, Y_2, X)^\mathsf{T}$ and our main interest is in the relationship of $(Y_1, Y_2)^\mathsf{T}$ when $X$ is taken into account.

To adjust for the influence of the variable $X$, the most straightforward way is to use a partial correlation coefficient (either Pearson's or a rank based one) of $(Y_1, Y_2)^\mathsf{T}$ given $X$. But outside the trivariate normal distribution it might be difficult to construct an appropriate partial correlation coefficient. Further, using partial correlation coefficient may not answer all scientific questions. For instance it seems to be natural to compare the relationship of $(Y_1, Y_2)^\mathsf{T}$ for different values of $X$.

Let us illustrate this with an example. Suppose we have data on life expectancies at birth ('average lengths of lives') at different countries and the interest is in the relationship of the life expectancies of males $(Y_1)$ and females $(Y_2)$. Then a natural question is whether this relationship is different in poor and rich countries. Let us take e.g. gross domestic product (GDP) per capita $(X)$ as a proxy for the economic welfare of a country. Then, mathematically speaking, the question is about the relationship of $(Y_1, Y_2)^\mathsf{T}$ conditionally upon the given value of the covariate $X = x$ and whether this relationship depends on the value of $x$.

Patton (2006) introduced a very general concept of studying the dependence structure of $Y_1$ and $Y_2$ when $X$ is taken into consideration. To describe this, denote the joint and marginal distribution functions of $(Y_1, Y_2)^\mathsf{T}$, conditionally upon $X = x$, as

$$H_x(y_1, y_2) = \mathsf{P}(Y_1 \le y_1, Y_2 \le y_2 \,|\, X = x),$$

$$F_{1x}(y_1) = \mathsf{P}(Y_1 \le y_1 \,|\, X = x), \quad F_{2x}(y_2) = \mathsf{P}(Y_2 \le y_2 \,|\, X = x). \tag{1.2}$$

If $F_{1x}$ and $F_{2x}$ are continuous, then using Sklar's theorem (see e.g. Nelsen, 2006) analogously as in (1.1) there exists a unique copula $C_x$ such that

$$H_x(y_1, y_2) = C_x\big(F_{1x}(y_1), F_{2x}(y_2)\big). \tag{1.3}$$

From equation (1.3) we see that the conditional copula $C_x$ fully describes the conditional dependence structure of $(Y_1, Y_2)^\mathsf{T}$ given $X = x$ and it depends in a general way on the covariate value $x$.

In Chapter 3 we present the contribution of Gijbels et al. (2011) and Veraverbeke et al. (2011) to the nonparametric estimation of conditional copulas when $X$ is a univariate covariate. The main theoretical result is the weak convergence of the empirical copula process (see Theorems 3.1 and 3.2) which is proved in Veraverbeke et al. (2011). These results also show that it is useful to adjust the response observations ($Y_1$ and $Y_2$) for the effect of the covariate $X$ on their margins. This is also illustrated in the simulation study in Gijbels et al. (2011). To the best of our knowledge Gijbels et al. (2011) and Veraverbeke et al. (2011) are the first two papers where nonparametric estimators of conditional copulas were suggested and investigated.

The generalization when the covariate is a multivariate vector or even of a functional type can be found in Gijbels et al. (2012).

## 1.3 The conditional copula does not depend on the value of the covariate

Suppose we are in the situation of the previous section. Sometimes, it seems reasonable to assume that the covariate $X$ affects only the marginal distributions of $Y_1$ and $Y_2$, but it does not affect the conditional dependence structure so that the conditional copula $C_x$ does not depend on $x$. Denote this copula simply as $C$. Then the general model for the conditional joint distribution function $H_x$ given by (1.3) simplifies to

$$H_x(y_1, y_2) = C\big(F_{1x}(y_1), F_{2x}(y_2)\big). \tag{1.4}$$

This is also called *the simplified pair-copula construction* in the recent literature, see e.g. Hobæk Haff et al. (2010), Acar et al. (2012) and Stöber et al. (2013).

To the best of our knowledge Gijbels et al. (2015b) were the first who considered nonparametric estimation of the copula function $C$ in this setting. Their contribution is presented in Chapter 4. The main theoretical result is that if some information about the influence of the covariate on margins is available then one can often get the same asymptotic distribution of the estimator of $C$ as when the covariate $X$ is not present. See Theorems 4.4 a and 4.6. These generalise the results for empirical copulas.

## 1.4 Conditional distribution function

When estimating a conditional copula one needs to estimate conditional distribution functions $F_{1x}$ and $F_{2x}$ introduced in (1.2). If there is no further information about the effect of the covariate on the marginal conditional distributions then a general nonparametric estimator is needed. Simulations reveal that a good nonparametric estimator can improve the finite sample properties of conditional copulas.

To simplify the notation let $F_x$ stand for the conditional distribution function of $Y$ given $X = x$. Veraverbeke et al. (2014) suggested pre-adjusting original observations ($Y$) in order to improve in particular bias properties of a nonparameric estimator of $F_x$. The crucial point is that the model assumed for pre-adjusting does not have to hold 'exactly'. This contribution is presented in Chapter 5. Although the idea of pre-adjusting is not completely novel, Veraverbeke et al. (2014) were the first to investigate pre-adjusting in detail in the context of conditional distribution function estimation. Among others Veraverbeke et al. (2014) showed that when pre-adjusting non-parametrically through location and scale, then the asymptotic variance of the nonparametric estimator of $F_x$ stays the same while the asymptotic bias can be reduced substantially (see Theorem 5.1).

# Chapter 2

# Nonparametric copula estimation

In this chapter we introduce an empirical copula function. Then we describe the contribution of Omelka et al. (2009) to nonparametric kernel estimation of copula functions and formulate the main results proved in that paper.

For simplicity of the presentation we restrict to the case $d = 2$, and consider an independent and identically distributed sample $(Y_{11}, Y_{21})^{\mathsf{T}}, \ldots, (Y_{1n}, Y_{2n})^{\mathsf{T}}$ of a bivariate random vector $(Y_1, Y_2)^{\mathsf{T}}$ with joint distribution function $H$ and marginal distribution functions $F_1$ and $F_2$.

## 2.1 Empirical copula

Nonparametric estimation of copulas goes back to Deheuvels (1979) who proposed the following empirical copula estimator

$$(2.1)\quad C_n(u_1, u_2) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\{\widehat{U}_{1i} \le u_1, \widehat{U}_{2i} \le u_2\}, \quad \text{with} \quad \widehat{U}_{1i} = F_{1n}(Y_{1i}),\ \widehat{U}_{2i} = F_{2n}(Y_{2i}),$$

where $F_{1n}$ and $F_{2n}$ are the empirical cumulative distribution functions of the marginals, and where $\mathbb{I}\{A\}$ denotes the indicator of a set $A$. This estimator is asymptotically equivalent (up to a term $O(n^{-1})$) with the estimator based directly on Sklar's Theorem given by

$$(2.2)\qquad C_n(u_1, u_2) = H_n\big(F_{1n}^{-1}(u_1), F_{2n}^{-1}(u_2)\big),$$

with $H_n$ the empirical joint distribution function. Weak convergence studies of the empirical copula estimator can be found in Gänssler and Stute (1987), Fermanian et al. (2004), Tsukahara (2005) and Segers (2012). The limiting distribution is a two-dimensional pinned C-Brownian sheet on $[0, 1]^2$ which is described in Theorem 2.1.

## 2.2 Nonparametric kernel estimators of a copula

It is well known that kernel estimators of distribution functions are not only visually appealing but they can have a lower mean squared error. That is why people become soon interested if kernel estimation can improve also the finite sample properties of copulas estimators.

To the best of our knowledge, the first estimator of a smoothed version of the empirical copula can be found in Fermanian et al. (2004). Their proposal is a straightforward modification of (2.2). Let $k(t_1, t_2)$ be a given bivariate kernel density function and $b_n$ is a bandwidth sequence tending to zero with $n$. Then the kernel estimator of a copula function is defined as

$$\widehat{C}_n^{(\mathrm{SE})}(u_1, u_2) = \widehat{H}_n\big(\widehat{F}_{1n}^{-1}(u_1), \widehat{F}_{2n}^{-1}(u_2)\big), \tag{2.3}$$

where the quantities $\widehat{H}_n$, $\widehat{F}_{1n}$ and $\widehat{F}_{2n}$ are given by

$$\begin{gathered} \widehat{H}_n(y_1, y_2) = \frac{1}{n}\sum_{i=1}^{n} K_n(y_1 - Y_{1i}, y_2 - Y_{2i}), \\ \widehat{F}_{1n}(y_1) = \widehat{H}_n(y_1, +\infty), \ \ \widehat{F}_{2n}(y_2) = \widehat{H}_n(+\infty, y_2), \end{gathered}, \tag{2.4}$$

with

$$K_n(y_1, y_2) = K\big(\tfrac{y_1}{b_n}, \tfrac{y_2}{b_n}\big), \qquad K(y_1, y_2) = \int_{-\infty}^{y_1}\int_{-\infty}^{y_2} k(t_1, t_2)\, dt_1\, dt_2.$$

Fermanian et al. (2004) proved weak convergence of this estimator.

### 2.2.1 Local linear kernel estimator and its modification

A different approach to kernel estimation of copulas was introduced in Chen and Huang (2007). Their starting point is formula (2.1). In the first stage they estimate marginals by

$$\widehat{F}_{1n}(y_1) = \frac{1}{n}\sum_{i=1}^{n} K\big(\tfrac{y_1 - Y_{1i}}{b_{n1}}\big), \qquad \widehat{F}_{2n}(y_2) = \frac{1}{n}\sum_{i=1}^{n} K\big(\tfrac{y_2 - Y_{2i}}{b_{n2}}\big), \tag{2.5}$$

with $K$ the integral of a symmetric bounded kernel function $k$ supported on $[-1, 1]$. With the help of $\widehat{F}_{1n}$ and $\widehat{F}_{2n}$ they construct pseudo-observations $\widehat{U}_{1i} = \widehat{F}_{1n}(Y_{1i})$ and $\widehat{U}_{2i} = \widehat{F}_{2n}(Y_{2i})$. In the second stage $(\widehat{U}_{1i}, \widehat{U}_{2i})$ are used to estimate the copula function $C$. To prevent for boundary bias, Chen and Huang (2007) suggested using a local linear version of the kernel $k$ given by

$$k_{u,h}(y) = \frac{k(y)\{a_2(u,h) - a_1(u,h)\, y\}}{a_0(u,h)a_2(u,h) - a_1^2(u,h)}\, \mathbb{I}\{\tfrac{u-1}{h} < y < \tfrac{u}{h}\},$$

where

$$a_l(u,h) = \int_{\frac{u-1}{h}}^{\frac{u}{h}} t^l k(t)\, dt \qquad \text{for } l = 0, 1, 2.$$

Finally the Local Linear type estimator of the copula is given by

$$\widehat{C}_n^{(\mathrm{LL})}(u_1, u_2) = \frac{1}{n}\sum_{i=1}^{n} K_{u_1,h_n}\big(\tfrac{u_1 - \widehat{U}_{1i}}{h_n}\big) K_{u_2,h_n}\big(\tfrac{u_2 - \widehat{U}_{2i}}{h_n}\big), \tag{2.6}$$

where $K_{u,h}(y) = \int_{-\infty}^{y} k_{u,h}(t)\, dt$. Chen and Huang (2007) derived expressions for asymptotic bias, variance and mean squared error for this estimator and showed that a proper choice of the second stage smoothing constants $h = h_n$ may considerably decrease variance, as well as mean squared error of the copula estimate. Moreover their Monte Carlo experiments showed that the estimator $\widehat{C}_n^{(\mathrm{LL})}$ is quite insensitive to the choice of the constants $b_{1n}$ and $b_{2n}$ used for smoothing the marginals in the first stage. Variance considerations provided by the authors even showed that it is reasonable to take $b_{1n}$ and $b_{2n}$ as small as possible. Note that strong undersmoothing in the first stage, recommended in Chen and Huang (2007), results in using the pseudo-observations

$$(\widehat{U}_{1i}, \widehat{U}_{2i})^{\mathsf{T}} = \left(\tfrac{2\,n\,F_{1n}(Y_{1i})-1}{2\,n}, \tfrac{2\,n\,F_{2n}(Y_{2i})-1}{2\,n}\right)^{\mathsf{T}}, \tag{2.7}$$

which is asymptotically equivalent to the mostly-used pseudo-observations

$$(\widehat{U}_{1i}, \widehat{U}_{2i})^{\mathsf{T}} = \tfrac{n}{n+1}\big(F_{1n}(Y_{1i}), F_{2n}(Y_{2i})\big)^{\mathsf{T}}. \tag{2.8}$$

The theoretical inconvenience of the estimator (2.6) is that for many common families of copulas (e.g. Clayton, Gumbel, normal, Student) the bias of the estimator at some of the corners of the unit square is only of order $O(h_n)$. As the optimal bandwidth for distribution function estimation is of order $O(n^{-1/3})$, this violates the $n^{1/2}$-order weak convergence on the whole $[0,1]^2$.

The problem is caused by unboundedness of second order partial derivatives of many copula families. Although parametric models with unbounded densities are rather rare in ‘standard’ parametric models, copula families with unbounded densities are quite common. As a benchmark we can take the normal bivariate density which is usually supposed to be a well-behaved model. But the resulting normal copula density is unbounded.

To overcome this difficulty Omelka et al. (2009) proposed a method of shrinking the bandwidth when coming close to the borders of the unit square. The proposed method is based on the observation that when calculating the bias of the estimator (2.6) one has to deal with terms of the form $h^2\, C^{(1,1)}(u_1, u_2)$, $h^2\, C^{(1,2)}(u_1, u_2)$ and $h^2\, C^{(2,2)}(u_1, u_2)$, where $C^{(1,1)}(u_1, u_2)$, $C^{(1,2)}(u_1, u_2)$ and $C^{(2,2)}(u_1, u_2)$ are the second order partial derivatives of $C$, that is $C^{(1,1)}(u_1, u_2) = \partial^2 C(u_1, u_2)/\partial u_1^2$ and similarly for $C^{(1,2)}(u_1, u_2)$, $C^{(2,2)}(u_1, u_2)$. A closer inspection of the common copula families shows that

$$C^{(i,j)}(u_1, u_2) = O\left(\frac{1}{\sqrt{u_i(1-u_i)\,u_j(1-u_j)}}\right), \qquad i, j \in \{1, 2\}. \tag{2.9}$$

This is shown Omelka et al. (2009, Appendix D) for Clayton, Gumbel, normal and Student copulas. Thus in order to keep the bias bounded Omelka et al. (2009) suggested an improved ‘Shrinked’ version of (2.6) given by

$$\widehat{C}_n^{(\mathrm{LLS})}(u_1, u_2) = \frac{1}{n}\sum_{i=1}^{n} K_{u_1,h_n}\left(\tfrac{u_1-\widehat{U}_{1i}}{b(u_1)\,h_n}\right) K_{u_2,h_n}\left(\tfrac{u_2-\widehat{U}_{2i}}{b(u_2)\,h_n}\right), \tag{2.10}$$

with $b(w) = \min(\sqrt{w}, \sqrt{1-w})$. A straightforward adaptation of the result of Chen and Huang (2007) gives that for ‘the central part of the unit square’, i.e. for $(u_1/b(u_1), u_2/b(u_2)) \in [h_n, 1-h_n]^2$ (and no smoothing of the marginals in the first stage)

$$\mathsf{bias}\left\{\widehat{C}_n^{(\mathrm{LLS})}(u_1, u_2)\right\} = \tfrac{\sigma_K^2}{2}\, h_n^2 \left\{b^2(u_1)\, C^{(1,1)}(u_1, u_2) + b^2(u_2)\, C^{(2,2)}(u_1, u_2)\right\} + o(h_n^2),$$

$$
\begin{aligned}
&\operatorname{var}\Big\{\widehat{C}_n^{(\mathrm{LLS})}(u_1,u_2)\Big\}\\
&\quad=\tfrac{1}{n}\operatorname{var}\Big\{\mathbb{I}\{U_1\le u_1,U_2\le u_2\}-C^{(1)}(u_1,u_2)\mathbb{I}\{U_1\le u_1\}-C^{(2)}(u_1,u_2)\mathbb{I}\{U_2\le u_2\}\Big\}\\
&\qquad-\tfrac{h_n\,a_K}{n}\Big[b(u_1)\,C^{(1)}(u_1,u_2)\big(1-C^{(1)}(u_1,u_2)\big)+b(u_2)\,C^{(2)}(u_1,u_2)\big(1-C^{(2)}(u_1,u_2)\big)\Big]\\
&\qquad+o\big(\tfrac{h_n}{n}\big),
\end{aligned}
$$

with $\sigma_K^2 = \int_{-1}^{1} t^2 k(t)\,dt$, $a_K = 2\int_{-1}^{1} t\,k(t)\,K(t)\,dt$ and $b(\cdot)$ as defined in (2.10). Taking $b(w) = 1$ gives back the bias and variance expressions for $\widehat{C}_n^{(\mathrm{LL})}$ in Chen and Huang (2007) (in case of no smoothing at the first stage).

### 2.2.2 Mirror-reflection kernel estimator

Another version of a kernel estimator for the copula might be obtained by integration of the estimator of the density of the copula introduced and studied in Gijbels and Mielniczuk (1990). This estimator deals with the boundary problem by the technique known as mirror-reflection. If a multiplicative kernel $k(y_1, y_2) = k(y_1)\,k(y_2)$ is used, then the Mirror-Reflection estimate of the copula has a simple form

$$
\widehat{C}_n^{(\mathrm{MR})}(u_1,u_2)=\frac{1}{n}\sum_{i=1}^{n}\sum_{\ell=1}^{9}\left[K\Big(\tfrac{u_1-\widehat{U}_{1i}^{(\ell)}}{h_n}\Big)-K\Big(\tfrac{-\widehat{U}_{1i}^{(\ell)}}{h_n}\Big)\right]\left[K\Big(\tfrac{u_2-\widehat{U}_{2i}^{(\ell)}}{h_n}\Big)-K\Big(\tfrac{-\widehat{U}_{2i}^{(\ell)}}{h_n}\Big)\right], \tag{2.11}
$$

where $\{(\widehat{U}_{1i}^{(\ell)}, \widehat{U}_{2i}^{(\ell)}),\ i = 1,\dots,n,\ \ell = 1,\dots,9\} = \{(\pm\widehat{U}_{1i}, \pm\widehat{U}_{2i}),\ (\pm\widehat{U}_{1i}, 2-\widehat{U}_{2i}),\ (2-\widehat{U}_{1i}, \pm\widehat{U}_{2i}),\ (2-\widehat{U}_{1i}, 2-\widehat{U}_{2i}),\ i = 1,\dots,n\}$.

The mirror-type estimator (2.11) faces the same 'corner bias' problem as the local linear estimator (2.6). To prevent this problem Omelka et al. (2009) introduced 'shrinking' the bandwidth similarly as in (2.10) and proposed

$$
\widehat{C}_n^{(\mathrm{MRS})}(u_1,u_2)=\frac{1}{n}\sum_{i=1}^{n}\sum_{\ell=1}^{9}\left[K\Big(\tfrac{u_1-\widehat{U}_{1i}^{(\ell)}}{b(u_1)\,h_n}\Big)-K\Big(\tfrac{-\widehat{U}_{1i}^{(\ell)}}{b(u_1)\,h_n}\Big)\right]\left[K\Big(\tfrac{u_2-\widehat{U}_{2i}^{(\ell)}}{b(u_2)\,h_n}\Big)-K\Big(\tfrac{-\widehat{U}_{2i}^{(\ell)}}{b(u_2)\,h_n}\Big)\right].
$$

### 2.2.3 Transformation estimator

The unboundedness of the densities of many copula families brings us back to Sklar's theorem in (1.1) and to the estimator (2.3) proposed in Fermanian et al. (2004).

To control the bias of this estimator in order to achieve weak convergence, we need the boundedness of the second order partial derivatives of the original joint distribution $H$. As the bivariate normal benchmark example shows, this condition may be considerably weaker than the requirement of the bounded second order derivatives of the underlying copula $C$.

A possible methodological objection to the estimator $\widehat{C}_n^{(\mathrm{SE})}$, defined in (2.3), may be its dependence on the marginal distributions. This is confirmed by Monte-Carlo simulations which show that for a given copula the success of this estimator depends on the marginals crucially.

As the copula function is invariant to increasing transformations of the margins, it is possible to transform the original data to $Y_{1i}' = T_1(Y_{1i})$ and $Y_{2i}' = T_2(Y_{2i})$, where $T_1$ and $T_2$ are increasing functions, and then use $(Y'_{1i}, Y'_{2i})$ instead of the original observations $(Y_{1i}, Y_{2i})$ in the estimator $\widehat{C}_n^{(\mathrm{SE})}$. The aim of the transformation is to simplify the kernel estimation of the joint distribution. As the direct choice of functions $T_1, T_2$ is difficult, we propose the following procedure. Let us first construct the uniform pseudo-observations $\widehat{U}_{1i}^{(\mathrm{E})} = \frac{n}{n+1} F_{1n}(Y_{1i})$ and $\widehat{U}_{2i}^{(\mathrm{E})} = \frac{n}{n+1} F_{2n}(Y_{2i})$. Then for a given distribution function $\Phi$ put $\hat{S}_{1i} = \Phi^{-1}(\widehat{U}_{1i}^{(\mathrm{E})})$ and $\hat{S}_{2i} = \Phi^{-1}(\widehat{U}_{2i}^{(\mathrm{E})})$. Finally use these transformed pseudo-observations $(\hat{S}_{1i}, \hat{S}_{2i})$ instead of the original observations $(Y_{1i}, Y_{2i})$ in the estimator (2.4) of the joint distribution function. As we know the marginals to be given by the function $\Phi$, the suggested estimator has in the case of multiplicative kernel the following simple formula

$$\widehat{C}_n^{(\mathrm{T})}(u_1, u_2) = \frac{1}{n} \sum_{i=1}^{n} K\left(\frac{\Phi^{-1}(u_1) - \Phi^{-1}\big(\widehat{U}_{1i}^{(\mathrm{E})}\big)}{h_n}\right) K\left(\frac{\Phi^{-1}(u_2) - \Phi^{-1}\big(\widehat{U}_{2i}^{(\mathrm{E})}\big)}{h_n}\right).$$

The advantage of this estimator is that it is not affected by the marginal distributions. Further bias calculations show that if we choose $\Phi$, such that $\frac{\Phi'(x)^2}{\Phi(x)}$ is bounded, we take care of the 'corner bias problem' which is present if we try to estimate the joint distribution of pseudo-observations directly. The above condition is satisfied e.g. for $\Phi$ the normal cumulative distribution function.

### 2.2.4 Main results

The main theoretical contribution of Omelka et al. (2009) is the weak convergence of the kernel estimators $\widehat{C}_n^{(\mathrm{LL})}$, $\widehat{C}_n^{(\mathrm{LLS})}$, $\widehat{C}_n^{(\mathrm{MR})}$, $\widehat{C}_n^{(\mathrm{MRS})}$ and $\widehat{C}_n^{(\mathrm{T})}$.

For notational convenience, denote $\widehat{F}_{1n}$ and $\widehat{F}_{2n}$ the estimates of the marginals which are used to construct pseudo-observations, i.e. $\widehat{U}_{1i} = \widehat{F}_{1n}(Y_{1i})$ and $\widehat{U}_{2i} = \widehat{F}_{2n}(Y_{2i})$. For the weak convergence results these functions need to be asymptotically equivalent to the empirical cumulative distribution functions $F_{1n}$, $F_{2n}$, i.e.

$$\sup_{x\in\mathbb{R}} \big|\widehat{F}_{1n}(x) - F_{1n}(x)\big| = o_p\big(\tfrac{1}{\sqrt{n}}\big), \qquad \sup_{y\in\mathbb{R}} \big|\widehat{F}_{2n}(y) - F_{2n}(y)\big| = o_p\big(\tfrac{1}{\sqrt{n}}\big), \tag{2.12}$$

which further implies the standard weak convergence of the processes $\sqrt{n}\big(\widehat{F}_{jn} - F\big)$ $(j = 1, 2)$ to particular Brownian bridges. For technical reasons we will also suppose that the functions $\widehat{F}_{1n}$ and $\widehat{F}_{2n}$ are nondecreasing, which excludes higher order kernels (taking negative values) for the estimation of the marginals.

It is easy to see that (2.12) is satisfied if we define pseudo-observations as as in (2.7) or in a way given in (2.8).

If one decides for kernel smoothing of the marginals given in (2.5), then it is well known (see e.g. Lemma 7 Fermanian et al., 2004) that assumption (2.12) is met if there exists $\alpha > 0$ and a sequence $b_n$ such that for $j = 1, 2$, uniformly in $x$,

$$F_j(x + b_n) = F_j(x) + b_n\, f_j(x) + o(b_n^{1+\alpha}) \quad \text{with} \quad \sqrt{n}\, b_n^{1+\alpha} \to 0.$$

Let $\mathbb{C}_n^{(\mathrm{LL})}$, $\mathbb{C}_n^{(\mathrm{LLS})}$, $\mathbb{C}_n^{(\mathrm{MR})}$, $\mathbb{C}_n^{(\mathrm{MRS})}$, $\mathbb{C}_n^{(\mathrm{T})}$ be suitably normalized empirical copula processes on $[0, 1]^2$, i.e. for $(u_1, u_2) \in [0, 1]^2$

$$\mathbb{C}_n^{(\cdot)}(u_1, u_2) = \sqrt{n}\left[C_n^{(\cdot)}(u_1, u_2) - C(u_1, u_2)\right].$$

**Theorem 2.1.** *Suppose that $H$ has continuous marginal distribution functions and that the underlying copula function $C$ has bounded second order partial derivatives on $[0,1]^2$. If $h_n = O(n^{-1/3})$ and (2.12) is satisfied, then the (kernel) copula processes $\mathbb{C}_n^{(LL)}$, $\mathbb{C}_n^{(MR)}$ converge weakly to the Gaussian process $G_C$ in $\ell^\infty([0,1]^2)$ having representation*

$$G_C(u_1,u_2) = B_C(u_1,u_2) - C^{(1)}(u_1,u_2)\, B_C(u_1,1) - C^{(2)}(u_1,u_2)\, B_C(1,u_2), \tag{2.13}$$

*where $C^{(1)}$ and $C^{(2)}$ denote the first order partial derivatives of $C$, and $B_C$ is a two-dimensional pinned $C$-Brownian sheet on $[0,1]^2$, i.e. it is a centered Gaussian process with covariance function*

$$\mathsf{E}\left[B_C(u_1,u_2)B_C(u_1',u_2')\right] = C(u_1 \wedge u_1', u_2 \wedge u_2') - C(u_1,u_2)C(u_1',u_2').$$

While Theorem 2.1 requires boundedness of the second order partial derivatives of the copula $C$, the weak convergence result of Fermanian et al. (2004) for the estimator $\mathbb{C}_n^{(\mathrm{SE})}$ given by (2.3) requires boundedness of the second order derivatives of the original joint distribution function $H$. This may or may not be more stringent depending on the marginals. Unfortunately, Theorem 2.1 excludes many commonly-used families of copulas. In the next theorem the assumption of boundedness of the second order partial derivatives of $C$ is required only for each inner point of $[0,1]^2$. In Appendix D of Omelka et al. (2009) it is verified that this assumption is met or commonly-used copulas such as Clayton, Gumbel, normal and Student copulas. Further note this this also allows that the first order partial derivatives do not have to be continuous on $[0,1]^2$. In Omelka et al. (2009) it is proved that it sufficient to assume

$$C^{(1)}, C^{(2)} \text{ are continuous in } [0,1]^2 \backslash \big\{(0,0),(0,1),(1,0),(1,1)\big\}. \tag{2.14}$$

**Theorem 2.2.** *Suppose that $H$ has continuous marginal distribution functions. Further suppose that (2.14) holds and the second order partial derivatives are bounded in each point of $(0,1)^2$ and satisfies (2.9). If $h_n = O(n^{-1/3})$ and (2.12) is satisfied, then the (kernel) copula processes $\mathbb{C}_n^{(LLS)}$ and $\mathbb{C}_n^{(MRS)}$ converge weakly to the Gaussian process $G_C$ in $\ell^\infty([0,1]^2)$ given in Theorem 2.1.*

*Moreover, if the functions $\Phi'$ and $\frac{\Phi'(x)^2}{\Phi(x)}$ are bounded then the above statement holds also for the process $\mathbb{C}_n^{(T)}$.*

Segers (2012) proved that for the weak convergence of empirical copula the assumption (2.14) can be further weakened to

$$C^{(j)}, \text{ is continuous in } \big\{(u_1,u_2) \in [0,1]^2 : 0 < u_j < 1\big\}, \quad j = 1,2. \tag{2.15}$$

By checking the proof of Omelka et al. (2009) one can conclude that (2.15) would be sufficient also for Theorem 2.2. Note that assumption (2.15) is necessary so that the limiting process $G_C$ given by (2.13) exists and has continuous trajectories. To overcome this difficulty the empirical process has to be studied in a different space than the space of bounded functions equipped with the uniform norm, see Bücher et al. (2014).

Chapter 3

# Nonparametric conditional copula estimation (general case)

In this chapter empirical estimators of a conditional copula function suggested in Gijbels et al. (2011) are introduced. The main theoretical results proved in Veraverbeke et al. (2011) are formulated. Finally, conditional measures of association based on conditional copulas are shortly discussed.

## 3.1 Estimating the conditional copula

To estimate the conditional copula $C_x$ it is convenient to invert Sklar's theorem in (1.3) which enables to express $C_x$ as

$$C_x(u_1, u_2) = H_x\big(F_{1x}^{-1}(u_1), F_{2x}^{-1}(u_2)\big), \quad (u_1, u_2) \in [0,1]^2, \tag{3.1}$$

where $F_{1x}^{-1}(u) = \inf\{y : F_{1x}(y) \geq u\}$ is the conditional quantile function of $Y_1$ given $X = x$ and $F_{2x}^{-1}$ is the conditional quantile function of $Y_2$ given $X = x$.

Now suppose that we observe independent identically distributed three-dimensional vectors $(Y_{11}, Y_{21}, X_1)^\mathsf{T}, \ldots, (Y_{1n}, Y_{2n}, X_n)^\mathsf{T}$. Based on the sample of observations the empirical estimator for $H_x(y_1, y_2)$ is given by:

$$\widehat{H}_x(y_1, y_2) = \sum_{i=1}^{n} w_{ni}(x, h_n)\, \mathbb{I}\{Y_{1i} \leq y_1, Y_{2i} \leq y_2\},$$

where $\{w_{ni}(x, h_n)\}$ is a sequence of weights that smooth over the covariate space (see Section 3.2.2) and $h_n > 0$ is a bandwidth tending to zero as the sample size increases. Here $\mathbb{I}\{A\}$ denotes the indicator of an event $A$. Inspired by (3.1) Gijbels et al. (2011) suggested the following straightforward estimator of the conditional copula function

$$\begin{aligned}\widehat{C}_x(u_1, u_2) &= \widehat{H}_x\big(\widehat{F}_{1x}^{-1}(u_1), \widehat{F}_{2x}^{-1}(u_2)\big) \\ &= \sum_{i=1}^{n} w_{ni}(x, h_n)\, \mathbb{I}\big\{Y_{1i} \leq \widehat{F}_{1x}^{-1}(u_1), Y_{2i} \leq \widehat{F}_{2x}^{-1}(u_2)\big\},\end{aligned} \tag{3.2}$$

where $\widehat{F}_{1x}$ and $\widehat{F}_{2x}$ are corresponding marginal distribution functions of $\widehat{H}_x$.

Although the copula estimator $\widehat{C}_x$ given by (3.2) seems very natural, since it mimics the structure of the true copula $C_x$ given in (3.1), a closer inspection of the estimator points to some potential pitfalls of it. For instance suppose that $Y_1$ and $Y_2$ are conditionally independent given $X = z$, but that their marginal conditional distributions are stochastically increasing with $z$. Then, intuitively speaking, larger values of $Y_1$ will occur together with larger values of $Y_2$ purely because of the same trend in the covariate $z$ creating an artificial dependence. This results in a bias of the estimator of the conditional copula given by (3.2).

Note that the possible bias described in the previous paragraph is due to the fact that the covariate affects the marginal distributions of $Y_1$ and $Y_2$. Theoretically this bias could be eliminated by removing the effect of the covariates on the marginals. Recall that the copula function is invariant to increasing transformations. Thus if one knew $F_{1z}$, $F_{2z}$ (for each $z$ from the support of $X$) it would be advisable to base the estimator $\widehat{C}_x$ on the observations $\{(U_{1i}, U_{2i})^{\mathsf{T}}, i = 1, \ldots, n\}$ where

$$(U_{1i}, U_{2i})^{\mathsf{T}} = \big(F_{1X_i}(Y_{1i}), F_{2X_i}(Y_{2i})\big)^{\mathsf{T}}, \tag{3.3}$$

whose marginal distributions are uniform (for each $i = 1, \ldots, n$).

Unfortunately, we usually do not know the theoretical conditional marginal distribution functions $(F_{1z}, F_{2z})$, but we can estimate them in the same way as we estimate $F_{1x}$ and $F_{2x}$, that is

$$\begin{aligned} \widehat{F}_{1z}(y) &= \sum_{j=1}^{n} w_{nj}(z, g_{1n})\, \mathbb{I}\{Y_{1j} \le y\}, \\ \widehat{F}_{2z}(y) &= \sum_{j=1}^{n} w_{nj}(z, g_{2n})\, \mathbb{I}\{Y_{2j} \le y\}, \end{aligned} \tag{3.4}$$

where $g_1 = \{g_{1n}\} \searrow 0$ and $g_2 = \{g_{2n}\} \searrow 0$. Note that other estimators for the conditional distributions functions $F_{1z}$ and $F_{2z}$ can be used. See Chapter 5.

This leads to the following procedure. First, transform the original observations to reduce the effect of the covariate by

$$(\widetilde{U}_{1i}, \widetilde{U}_{2i})^{\mathsf{T}} = \big(\widehat{F}_{1X_i}(Y_{1i}), \widehat{F}_{2X_i}(Y_{2i})\big)^{\mathsf{T}}, \quad i = 1, \ldots, n. \tag{3.5}$$

Second, use the transformed observations $(\widetilde{U}_{1i}, \widetilde{U}_{2i})^{\mathsf{T}}$ in a similar way as the original observations, and construct

$$\widetilde{C}_x(u_1, u_2) = \widetilde{G}_x\Big(\widetilde{G}_{1x}^{-1}(u_1), \widetilde{G}_{2x}^{-1}(u_2)\Big), \tag{3.6}$$

where

$$\widetilde{G}_x(u_1, u_2) = \sum_{i=1}^{n} w_{ni}(x, h_n)\, \mathbb{I}\big\{\widetilde{U}_{1i} \le u_1, \widetilde{U}_{2i} \le u_2\big\},$$

and $\widetilde{G}_{1x}$ and $\widetilde{G}_{2x}$ are its corresponding marginals.

## 3.2 Theoretical results for conditional copula estimators

The aim of this section is to derive he asymptotic properties of the following (empirical copula) processes

$$\mathbb{C}^{(\mathrm{E})}_{xn}(u_1,u_2) = \sqrt{nh_n}\,\big(\widehat{C}_x(u_1,u_2) - C_x(u_1,u_2)\big), \qquad (0 \le u_1,u_2 \le 1), \tag{3.7}$$

$$\widetilde{\mathbb{C}}^{(\mathrm{E})}_{xn}(u_1,u_2) = \sqrt{nh_n}\,\big(\widetilde{C}_x(u_1,u_2) - C_x(u_1,u_2)\big), \qquad (0 \le u_1,u_2 \le 1). \tag{3.8}$$

All the theoretical results provided in this section are for a fixed but arbitrary value of $x$, and are uniform with respect to $u_1$ and $u_2$.

### 3.2.1 Regularity conditions

Let us denote $b_n = \max\{h_n, g_{1n}, g_{2n}\}$, $I_x^{(n)} = \{i : w_{ni}(x,b_n) \neq 0\}$ and

$$J_x^{(n)} = \Big[\min_{i\in I_x^{(n)}} X_i,\, \max_{i\in I_x^{(n)}} X_i\Big].$$

Let $a_n$ stand for a sequence of positive constants such that $(n\,a_n) \to \infty$ and $a_n = O(n^{-1/5})$. The following is a listing of assumptions on the system of weights $\{w_{ni};\ i = 1,\ldots,n\}$ in random design. The conditions for a fixed design, may be derived easily by replacing $X_i$ by $x_i$ and omitting the symbol $P$ in the subscript of $o_P$ and $O_P$ terms.

**(W1)** $\displaystyle\max_{1\le i\le n} |w_{ni}(x,h_n)| = o_P\Big(\tfrac{1}{\sqrt{n\,h_n}}\Big)$, **(W2)** $\displaystyle\sum_{i=1}^n w_{ni}(x,h_n) - 1 = o_P\Big(\tfrac{1}{\sqrt{n\,h_n}}\Big)$,

**(W3)** $\displaystyle\sum_{i=1}^n w_{ni}(x,h_n)(X_i-x) = O_P\Big(\tfrac{1}{\sqrt{n\,h_n}}\Big)$, **(W4)** $\displaystyle\sum_{i=1}^n w_{ni}(x,h_n)(X_i-x)^2 = O_P\Big(\tfrac{1}{\sqrt{n\,h_n}}\Big)$,

**(W5)** $\displaystyle\sum_{i=1}^n w_{ni}^2(x,h_n) = O_P\Big(\tfrac{1}{n\,h_n}\Big)$, **(W6)** $\displaystyle\Big(\max_{i\in I_x^{(n)}} X_i - \min_{i\in I_x^{(n)}} X_i\Big) = o_P(1)$,

**(W7)** $\displaystyle\sum_{i=1}^n |w_{ni}(x,h_n)| = O_P(1)$, **(W8)** $\displaystyle\sup_{z\in J_x^{(n)}}\Big|\sum_{i=1}^n w_{ni}(z,g_{jn}) - 1\Big| = o_P\big(g_{jn}^2\big)$,

**(W9)** $$\sup_{z\in J_x^{(n)}} \sum_{i=1}^n [w_{ni}(z,g_{jn})]^2 = O_P\Big(\frac{1}{n\,g_{jn}}\Big),$$

**(W10)** $$\sup_{z\in J_x^{(n)}} \sum_{i=1}^n [w'_{ni}(z,g_{jn})]^2 = O_P\Big(\frac{1}{n\,g_{jn}^3}\Big),$$

**(W11)** $$\exists_{C<\infty}\ \mathsf{P}\Big[\sup_{z\in J_x^{(n)}} \max_{1\le i\le n} |w_{ni}(z,h_n)\,\mathbb{I}\{|X_i - z| > C\,h_n\}| > 0\Big] = o(1),$$

**(W12)** $$\exists_{D_K<\infty}\,\forall_{a_n}\ \sup_{z\in J_x^{(n)}}\Big|\sum_{i=1}^n w_{ni}(z,a_n)(X_i - z) - a_n^2\,D_K\Big| = o_P\big(a_n^2\big),$$

**(W13)** $$\exists_{E_K<\infty}\,\forall_{a_n}\,\sup_{z\in J_x^{(n)}}\left|\sum_{i=1}^{n} w_{ni}(z,a_n)(X_i-z)^2 - a_n^2\,E_K\right| = o_P\left(a_n^2\right),$$

where $w'_{ni}(z,g_{jn})$ denotes the derivative with respect to $z$.

Conditions (**W7**)–(**W13**) make a finer control on the behaviour of the weights not only at the point $x$ but also in a (shrinking) neighbourhood of this point. This better control is needed to justify that the transformation (3.5) is 'painless'. Nevertheless, as argued in the next section, these conditions hold under usual regularity conditions on the distribution of the covariate $X$.

Further, we require the conditional copula $C_z$ and the conditional marginals $F_{1z}$ and $F_{2z}$ to satisfy:

**(R1)** The functions $\dot{H}_z\big(F_{1x}^{-1}(u_1), F_{2x}^{-1}(u_2)\big)$ and $\ddot{H}_z\big(F_{1x}^{-1}(u_1), F_{2x}^{-1}(u_2)\big)$ are uniformly continuous in $(z,u_1,u_2)$, where $z$ takes value in a neighbourhood of $x$.

**(R2)** The first order partial derivatives $C_x^{(1)}$, $C_x^{(2)}$ with respect to $u_1$ and $u_2$ respectively are continuous on $[0,1]^2 \setminus \{(0,0),(0,1),(1,0),(1,1)\}$

**($\tilde{\mathbf{R}}$1)** $\dot{C}_z(u_1,u_2) = \frac{\partial}{\partial z}C_z(u_1,u_2)$, $\ddot{C}_z(u_1,u_2) = \frac{\partial^2}{\partial z^2}C_z(u_1,u_2)$ exist and are continuous as functions of $(z,u_1,u_2)$, where $z$ takes value in a neighbourhood of $x$;

**($\tilde{\mathbf{R}}$2)** The functions $C_z^{(1)}(u_1,u_2)$ and $C_z^{(2)}(u_1,u_2)$ are uniformly continuous in $(z,u_1,u_2) \in U(x)\times\big([0,1]^2 \setminus \{(0,0),(0,1),(1,0),(1,1)\}\big)$, where $U(x)$ is a neighbourhood of the point $x$.

**($\tilde{\mathbf{R}}$3)** For $j=1,2$: $F_{jz}\big(F_{jz}^{-1}(u)\big)$, $\dot{F}_{jz}\big(F_{jz}^{-1}(u)\big)$, $\ddot{F}_{jz}\big(F_{jz}^{-1}(u)\big)$ are continuous as functions of $(z,u)$ for $z$ in a neighbourhood of $x$, where $\dot{F}_{jz}(y) = \frac{\partial}{\partial z}F_{jz}(y)$, $\ddot{F}_{jz}(y) = \frac{\partial^2}{\partial z^2}F_{jz}(y)$.

Similarly as noted at the end of Section 2.2.4 thanks to the results of Segers (2012) the set $\big([0,1]^2 \setminus \{(0,0),(0,1),(1,0),(1,1)\}\big)$ in assumptions **(R2)** and **($\tilde{\mathbf{R}}$2)** can be for $C_x^{(j)}$ and $C_z^{(j)}$ reduced to $\big\{(u_1,u_2)\in[0,1]^2 : 0 < u_j < 1\big\}$.

### 3.2.2 Some common choices of weights

As the list of conditions on the weights given in Section 3.2.1 might be rather discouraging in particular for readers who are less interested in technical details, we comment on several commonly used weight schemes.

Assume for concreteness that a kernel density function $k$ has support $[-1,1]$ and is symmetric and continuously differentiable. Further suppose that $h_n \sim n^{-1/5}$ and $g_{1n} \sim g_{2n} \sim n^{-1/5}$.

It can be shown that for *Nadaraya-Watson weights* (see Nadaraya, 1964; Watson, 1964), which are defined as

$$w_{ni}(x,h_n) = \frac{k\big(\frac{X_i-x}{h_n}\big)}{\sum_{j=1}^{n} k\big(\frac{X_j-x}{h_n}\big)}, \qquad i=1,\dots,n,$$

assumptions (**W1**)–(**W13**) hold, provided

**(F1)** $f_X = F_X'$ is continuous and positive at the point $x$,

**(F2)** $f_X' = F_X''$ is continuous in a neighbourhood of the point $x$,

where $F_X$ is the (marginal) distribution function of the covariate $X$.

Another system of weights, very commonly employed, is a *local linear* [LL] system of weights (see e.g. p. 20 of Fan and Gijbels, 1996), which is given by

$$w_{ni}(x, h_n) = \frac{\frac{1}{n\,h_n}\, k\big(\frac{X_i - x}{h_n}\big)\big(S_{n,2} - \frac{X_i - x}{h_n}\, S_{n,1}\big)}{S_{n,0}\, S_{n,2} - S_{n,1}^2}, \qquad i = 1, \ldots, n, \tag{3.9}$$

where

$$S_{n,j} = \frac{1}{n\,h_n} \sum_{i=1}^{n} \big(\tfrac{X_i - x}{h_n}\big)^j k\big(\tfrac{X_i - x}{h_n}\big), \qquad j = 0, 1, 2.$$

and $k$ is a kernel, a symmetric probability density function, with support $[-1, 1]$.

The nice thing about LL weights is that thanks to $\sum_{i=1}^n w_{ni}(x, h_n)(X_i - x) = 0$, it is sufficient to assume only **(F1)**.

In a fixed regular design case (see e.g. Müller, 1987), there exists an absolutely continuous distribution function $F_X$ (with associated density $f_X$) such that $x_i = F_X^{-1}\big(\frac{i}{n+1}\big)$. In this case the design points are ordered, that is $x_1 \leq x_2 \ldots \leq x_n$. In this setting *Gasser-Müller* [GM] weights (see Gasser and Müller, 1979) are quite popular. Consider fixed, but arbitrary values $x_0 < x_1$ and $x_{n+1} > x_n$. Then GM weights are defined as

$$w_{ni}(x, h_n) = \frac{1}{h_n} \int_{s_i}^{s_{i+1}} k\big(\tfrac{z - x}{h_n}\big)\, dz, \quad \text{where } s_i = (x_i + x_{i-1})/2, \quad i = 1, \ldots, n.$$

In a fixed regular design case, we conjecture that to verify (**W1**)–(**W13**) it is sufficient to assume that the design density satisfies **(F1)**.

### 3.2.3 The process $\mathbb{C}_{xn}^{(\mathrm{E})}$ given by (3.7)

Suppose

$$h_n = O(n^{-1/5}), \qquad n\,h_n \to \infty. \tag{3.10}$$

Note that (3.10) allows for $h_n \sim n^{-1/5}$, which is often the optimal rate for bandwidths in nonparametric problems.

**Theorem 3.1.** *Assume* (3.10), **(R1)**–**(R2)** *and* (**W1**)–(**W6**). *Then it holds uniformly in* $(u_1, u_2) \in [0, 1]^2$

$$\mathbb{C}_{xn}^{(\mathrm{E})}(u_1, u_2) = \sqrt{n\,h_n} \sum_{i=1}^{n} w_{ni}(x, h_n)\, \xi_i(u_1, u_2) + o_P(1),$$

*where*

$$\begin{aligned}\xi_i(u_1, u_2) &= \mathbb{I}\{Y_{1i} \leq F_{1x}^{-1}(u_1), Y_{2i} \leq F_{2x}^{-1}(u_2)\} - C_x(u_1, u_2) \\ &\quad - C_x^{(1)}(u_1, u_2)\, \big[\mathbb{I}\{Y_{1i} \leq F_{1x}^{-1}(u_1)\} - u_1\big] - C_x^{(2)}(u_1, u_2)\, \big[\mathbb{I}\{Y_{2i} \leq F_{2x}^{-1}(u_2)\} - u_2\big]\,.\end{aligned} \tag{3.11}$$

Define a process $Z_{xn} = \sqrt{n\,h_n}\sum_{i=1}^{n} w_{ni}(x,h_n)\,\xi_i$, where $\xi_i$'s are given in (3.11). As (**W5**) holds, typically there exists a finite positive constant $V$ such that

$$n\,h_n \sum_{i=1}^{n} w_{ni}^2(x,h_n) = V^2 + o_P(1).$$

Then for all $0 \le u_1, u_2, v_1, v_2 \le 1$

$$\mathsf{cov}\,\big(Z_{xn}(u_1,u_2), Z_{xn}(v_1,v_2)\big) \xrightarrow[n\to\infty]{} V^2\,\mathsf{cov}\,\big(\xi_x(u_1,u_2), \xi_x(v_1,v_2)\big), \tag{3.12}$$

where

$$\begin{aligned}\xi_x(u_1,u_2) &= \mathbb{I}\{F_{1x}(Y_{1x}) \le u_1, F_{2x}(Y_{2x}) \le u_2\} - C_x(u_1,u_2)\\ &\quad - C_x^{(1)}(u_1,u_2)\big[\mathbb{I}\{F_{1x}(Y_{1x}) \le u_1\} - u_1\big] - C_x^{(2)}(u_1,u_2)\big[\mathbb{I}\{F_{2x}(Y_{2x}) \le u_2\} - u_2\big].\end{aligned}$$

Thus with the help of (3.12) it is straightforward to verify the finite dimensional convergence of the process $\{Z_{xn}(u_1,u_2),\,(u_1,u_2)\in[0,1]^2\}$. As the asymptotic tightness of this process is (in a more general setting) verified in Step 1 of the proof of Theorem 3.1 given in Veraverbeke et al. (2011), one can deduce that $Z_{xn}$ converges weakly to a Gaussian process $Z_x$.

Further suppose that there exists $H$ such that $(n\,h_n^5) \to H^2$, with $H \ge 0$. Typically $h_n \sim n^{-1/5}$ so that $H > 0$. In that case, using Taylor expansion and assumption **(R1)** we can approximate the expectation of the limiting process $Z_x$ and find out that (uniformly in $(u_1,u_2)$)

$$\mathsf{E}\,Z_{xn}(u_1,u_2) = H\,\Big[D_K\,\dot{C}_x(u_1,u_2) + \tfrac{E_K}{2}\,B_x(u_1,u_2)\Big] + o(1),$$

with $D_K$ and $E_K$ being constants depending on the chosen system of weights $\{w_{ni}\}$ and on the type of the design (see (**W12**) and (**W13**) in Section 3.2.2) and

$$\begin{aligned}B_x(u_1,u_2) &= \ddot{H}_x\big(F_{1x}^{-1}(u_1), F_{2x}^{-1}(u_2)\big)\\ &\qquad - C_x^{(1)}(u_1,u_2)\,\ddot{F}_{1x}\big(F_{1x}^{-1}(u_1)\big) - C_x^{(2)}(u_1,u_2)\,\ddot{F}_{2x}\big(F_{2x}^{-1}(u_2)\big)\\ &= \ddot{C}_x(u_1,u_2) + 2\,\dot{C}_x^{(1)}(u_1,u_2)\,\dot{F}_{1x}\big(F_{1x}^{-1}(u_1)\big) + 2\,\dot{C}_x^{(2)}(u_1,u_2)\,\dot{F}_{2x}\big(F_{2x}^{-1}(u_2)\big)\\ &\qquad + C_x^{(1,1)}(u_1,u_2)\,\Big[\dot{F}_{1x}\big(F_{1x}^{-1}(u_1)\big)\Big]^2 + C_x^{(2,2)}(u_1,u_2)\,\Big[\dot{F}_{2x}\big(F_{2x}^{-1}(u_2)\big)\Big]^2\\ &\qquad + 2\,C_x^{(1,2)}(u_1,u_2)\,\dot{F}_{1x}\big(F_{1x}^{-1}(u_1)\big)\,\dot{F}_{2x}\big(F_{2x}^{-1}(u_2)\big),\end{aligned} \tag{3.13}$$

where a dot indicates a derivative with respect to the covariate $x$, e.g. $\dot{F}_z(u_1) = \frac{\partial}{\partial z}F_z(u_1)$, $\ddot{C}_z(u_1,u_2) = \frac{\partial^2}{\partial z^2}C_z(u_1,u_2)$; the symbol $^{(i)}$ indicates a derivative with respect to $u_i$, e.g. $C_x^{(i,j)}(u_1,u_2) = \frac{\partial^2 C_x(u_1,u_2)}{\partial u_i \partial u_j}$; and $\dot{C}_z^{(i)}(u_1,u_2) = \frac{\partial^2 C_z(u_1,u_2)}{\partial z\,\partial u_i}$, which is a mixture of the above notational rules.

**Corollary 3.1.** *If* (3.12), $(n\,h_n^5) \to H^2$, (**W12**), (**W13**) *and the assumptions of Theorem 3.1 hold, then the process* $\mathbb{C}_{xn}^{(\mathrm{E})}$ *converges in distribution to a Gaussian process* $Z_x$*, which can be written as*

$$Z_x(u_1,u_2) = V\,\Big\{W_x(u_1,u_2) - C_x^{(1)}(u_1,u_2)W_x(u_1,1) - C_x^{(2)}(u_1,u_2)W_x(1,u_2)\Big\} + R_x(u_1,u_2),$$

*where* $W_x$ *is a bivariate Brownian bridge on* $[0,1]^2$ *with covariance function*

$$\mathsf{E}\,[W_x(u_1,u_2)W_x(v_1,v_2)] = C_x(u_1 \wedge v_1, u_2 \wedge v_2) - C_x(u_1,u_2)C_x(v_1,v_2). \tag{3.14}$$

*and*

$$R_x(u_1,u_2) = H\,\Big[D_K\,\dot{C}_x(u_1,u_2) + \tfrac{E_K}{2}\,B_x(u_1,u_2)\Big]\,. \tag{3.15}$$

*Proof.* The proof follows from Theorem 3.1 and the reasoning given above. □

The constants $V$, $D_K$ and $E_K$ in general also depend on $x$, but for simplicity this is not made explicit in the notations.

It should be mentioned that to prove Corollary 3.1 it is only needed that assumptions (**W12**) and (**W13**) hold without supremum (for $z = x$) and for $a_n = h_n$.

### 3.2.4 The process $\widetilde{\mathbb{C}}_{xn}^{(\mathrm{E})}$ given by (3.8)

In the following we suppose that for $j = 1, 2$

$$\sqrt{n\,h_n}\,g_{jn}^2 = O(1), \quad \tfrac{h_n}{g_{jn}} = O(1), \quad n\,\min(h_n, g_{1n}, g_{2n}) \to \infty. \tag{3.16}$$

Note that (3.16) allows for the same rates of $h_n$ as in (3.10). Further, $h_n \sim n^{-1/5}$ implies that $g_{jn} \sim n^{-1/5}$ for $j = 1, 2$ as well.

**Theorem 3.2.** *Assume* (3.16), (**W1**)–(**W13**) *and* (**R̃1**)–(**R̃3**), *then uniformly in* $(u_1, u_2)$

$$\widetilde{\mathbb{C}}_{xn}^{(E)}(u_1,u_2) = \sqrt{n\,h_n}\,\sum_{i=1}^{n} w_{ni}(x,h_n)\,\tilde{\xi}_i(u_1,u_2) + o_P(1),$$

*where*

$$\begin{aligned}\tilde{\xi}_i(u_1,u_2) &= \mathbb{I}\{U_{1i} \le u_1, U_{2i} \le u_2\} - C_x(u_1,u_2)\\ &\quad - C_x^{(1)}(u_1,u_2)\,[\mathbb{I}\{U_{1i} \le u_1\} - u_1] - C_x^{(2)}(u_1,u_2)\,[\mathbb{I}\{U_{2i} \le u_2\} - u_2]\,,\end{aligned} \tag{3.17}$$

*and* $(U_{1i}, U_{2i})^{\mathsf{T}}$ *are given in* (3.3).

Similarly as in Section 3.2.3 we can state the following corollary.

**Corollary 3.2.** *If* (3.12), $(n\,h_n^5) \to H^2$ *and the assumptions of Theorem 3.2 hold, then the process* $\widetilde{\mathbb{C}}_{xn}^{(\mathrm{E})}$ *converges in distribution to a Gaussian process* $\widetilde{Z}_x$*, which can be written as*

$$\widetilde{Z}_x(u_1,u_2) = V\,\Big\{W_x(u_1,u_2) - C_x^{(1)}(u_1,u_2)W_x(u_1,1) - C_x^{(2)}(u_1,u_2)W_x(1,u_2)\Big\} + \widetilde{R}_x(u_1,u_2)$$

*where* $W_x$ *is a bivariate Brownian bridge on* $[0,1]^2$ *with covariance function* (3.14) *and*

$$\widetilde{R}_x(u_1,u_2) = H\,\Big[D_K\,\dot{C}_x(u_1,u_2) + \tfrac{E_K}{2}\,\widetilde{B}_x(u_1,u_2)\Big]\,, \tag{3.18}$$

*with* $\widetilde{B}_x(u_1,u_2) = \ddot{C}_x(u_1,u_2)$.

Thus comparing the limiting processes $Z_x$ and $\widetilde{Z}_x$ from Corollary 3.1 and 3.2 we see that the only difference is in the bias terms. This difference is a consequence of different random variables that are involved in the Bahadur representations of the processes $\sqrt{n\,h_n}\,(\widetilde{C}_x - C_x)$ and $\sqrt{n\,h_n}\,(\widehat{C}_x - C_x)$. The original observations $(Y_{1i}, Y_{2i})^{\mathsf{T}}$ in (3.11) are replaced by the

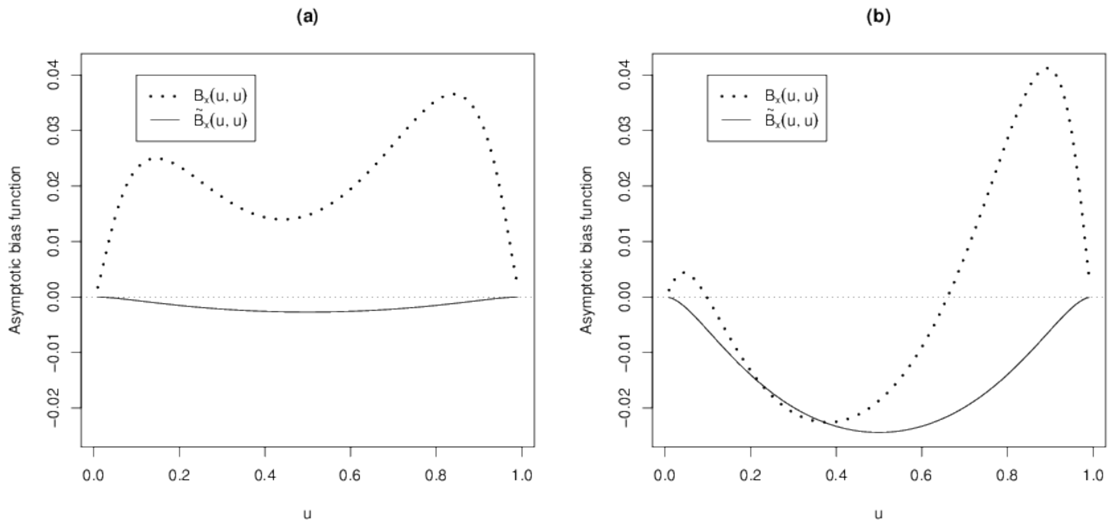


Figure 3.1: Diagonals of the functions $B_x$ and $\widetilde{B}_x$ for $\rho = 1$ (a) and $\rho = 7$ (b).

unobserved $(U_{1i}, U_{2i})^\mathsf{T}$ in (3.17). The key point is that the conditional marginal distributions of $(U_{1i}, U_{2i})^\mathsf{T}$ are uniform for each value of the covariate $X_i$ and thus do not depend on the values of the covariate, which results in a much simpler expression for the asymptotic bias given in (3.18).

To be fair it should be stressed that there is no guarantee that the asymptotic bias expression for the estimator $\widetilde{C}_x$ given by (3.18) is always closer to zero than that for $\widehat{C}_x$ given in expression (3.15). Suppose for simplicity that $D_K = 0$, which holds for example for a local linear system of weights (see Section 3.2.2). Then $B_x(u_1, u_2)$ of (3.15) may be closer to zero than $\widetilde{B}_x(u_1, u_2)$ if the additional terms in (3.13) turn out to be of opposite signs of the first term $\ddot{C}_x(u_1, u_2)$. For example, suppose that the covariate is standard normal distributed and we are interested in the point $X = 1$. The copula which joins the margins $(Y_1, Y_2)$ is taken to be a Frank copula with the parameter depending on the value of the covariate $X = z$ as $\theta(z) = 5 + \rho \sin\left(\frac{(z-1)\,\pi}{6}\right)$. Further, the margins are taken to be normal with unit variances and mean functions $\mu_1(z) = \mu_2(z) = \sin(z)$.

Consider two values of the parameter $\rho$. The case $\rho = 1$ represents a situation where the conditional dependence structure is only very mildly affected by the value of the covariate. The plot of the diagonals of the functions $B_x$ and $\widetilde{B}_x$ in Figure 3.1(a) clearly indicates that in terms of bias the estimator $\widetilde{C}_x$ is in this situation strongly preferable. This is further confirmed by calculating $L_2([0,1]^2)$-norms of the functions $B_x$ and $\widetilde{B}_x$, which equal 0.014 and 0.001 respectively.

When $\rho = 7$ the conditional dependence structure is strongly influenced by the covariate. Figure 3.1(b) shows that for this model it is not so easy to judge which estimator should be preferred. At some points $B_x$ is closer to zero and at other points it is the other way around. The $L_2([0,1]^2)$-norms of the functions $B_x$ and $\widetilde{B}_x$ now equal 0.011 and 0.012 indicating that the estimator $\widehat{C}_x$ might be slightly preferable if the interest is in estimation of the whole copula function and the mean integrated squared error is taken as the criterion for the quality of the estimate.

Our experience is that it is rather difficult to construct models where the estimator $\widehat{C}_x$

is (more than slightly) preferable to $\widetilde{C}_x$. In such models, both conditional marginals as well as the conditional dependence structure have to be strongly dependent on the value of the covariate. Further it must be the case that by a 'lucky coincidence' the additional terms in (3.13) help to reduce the effect of $\ddot{C}_x$. As this is difficult to predict, one stays on the safe side by using the estimator $\widetilde{C}_x$.

## 3.3 Some practical issues

### 3.3.1 Bandwidth selection

A crucial point of smoothing methods is the bandwidth selection. The proposed estimator $\widetilde{C}_x$ requires to choose three bandwidths – $g_{1n}$, $g_{2n}$ (the 'auxiliary' bandwidths to remove the effect of the covariate on the marginal distributions) and $h_n$ (the 'main' bandwidth to estimate the conditional copula). Generally speaking choosing bandwidth for conditional distribution estimation is more difficult than for instance in nonparametric regression as one cannot simply compare response and its estimate.

For choosing $h_n$ Gijbels et al. (2011) suggested an iterative method by adopting the idea of Gasser et al. (1991), which was further extended in Brockmann et al. (1993). Alternatively, one can consider the cross-validation procedure suggested in Gijbels et al. (2017b).

For choosing $g_{n1}$ and $g_{n2}$ one can either modify the above methods or try make use of the rules invented for nonparametric regression as for instance in Yu and Jones (1998).

Nevertheless, it should be said that the bandwidth choice is widely unexplored and all the methods mentioned above should be considered only as ad-hoc suggestions to be able to apply the suggested estimators to real data-sets.

### 3.3.2 Adjusting for the marginal effect of the covariate

Recall that the aim of the transformation (3.5) is to remove the effect of the covariate $X$ on the marginal distributions. For this reason nonparametric estimators of the conditional distribution functions are used. Of course, if we can assume a parametric or a semiparametric model for the influence of the covariate on the marginals, then it is advisable to use this model. Although it does not change asymptotic properties of the estimator, it may stabilize the finite sample properties.

Thus general strategy for conditional copula estimation in not very large samples may be as follows. First, check the scatterplots of the pairs $(X, Y_1)^{\mathsf{T}}$ and $(X, Y_2)^{\mathsf{T}}$. If there is no obvious pattern, then the estimator $\widehat{C}_x$ may be used. If this is not the case, we recommend to try to transform the variables $Y_1$ and $Y_2$ such that the influence of the covariate on the conditional marginal distributions is suppressed. This might be done in several ways. The transformation (3.5) is very general and in view of Theorem 3.2 it cannot be improved if we aim at eliminating the effect of the covariate on the marginals. The price we have to pay is that we have to specify two new bandwidths $g_{1n}$ and $g_{2n}$. Fortunately, the Monte Carlo simulation results of Gijbels et al. (2011) indicate that the rules for bandwidth selection in nonparametric regression may be employed or if $h_n$ is already fixed then using

$g_{1n} = g_{2n} = h_n$ for $\widetilde{C}_x$ usually results in an estimator which is at least as good as $\widehat{C}_x$.

Alternatively, in small samples we may try to stabilize the estimator by specifying a simple parametric model for the pairs $(X, Y_1)^\mathsf{T}$ and $(X, Y_2)^\mathsf{T}$. For instance, suppose simple linear regression models

$$Y_{1i} = \alpha_0 + \alpha_1 X_i + \varepsilon_{1i}, \quad Y_{2i} = \beta_0 + \beta_1 X_i + \varepsilon_{2i}, \tag{3.19}$$

where $\varepsilon_{1i}$ and $\varepsilon_{2i}$ are independent of $X_i$. Then it seems natural to replace the original observations $(Y_{1i}, Y_{2i})^\mathsf{T}$ with the estimated residuals from models (3.19). As the estimators of the unknown parameters converge at rate $n^{-1/2}$, the estimator based on the estimated residuals may be shown to be asymptotically equivalent with the one based on the unobserved residuals $(\varepsilon_{1i}, \varepsilon_{2i})^\mathsf{T}$ and thus the main effect of the covariate on the marginal distributions is usually removed.

A further step towards the general transformation (3.5) may be to assume nonparametric location-scale models

$$Y_{1i} = m_1(X_i) + \sigma_1(X_i)\,\varepsilon_{1i}, \quad Y_{2i} = m_2(X_i) + \sigma_1(X_i)\,\varepsilon_{2i}, \qquad i = 1, \dots, n,$$

where $\varepsilon_{1i}, \varepsilon_{2i}$ are independent of $X_i$ and $m_1(\cdot)$, $m_2(\cdot)$, $\sigma_1(\cdot)$, $\sigma_2(\cdot)$ are unknown, but sufficiently smooth functions. Let $\widehat{m}_1, \dots, \widehat{\sigma}_2$ be the corresponding nonparametric estimators. Then estimator $\widehat{C}_x$ can be based on the estimated residuals

$$\big(\widehat{\varepsilon}_{1i}, \widehat{\varepsilon}_{2i}\big)^\mathsf{T} = \Big(\tfrac{Y_{1i} - \widehat{m}_1(X_i)}{\widehat{\sigma}_1(X_i)}, \tfrac{Y_{2i} - \widehat{m}_2(X_i)}{\widehat{\sigma}_2(X_i)}\Big)^\mathsf{T}, \qquad i = 1, \dots, n.$$

### 3.3.3 Inference

Note that the limiting distributions of nonparametric estimators of conditional copulas suggested in Section 3.1 are rather involved.

Further, when choosing the optimal bandwidth then one cannot simply ignore the bias of the estimator as the (asymptotic) bias is typically of the same order as the (asymptotic) standard deviation of the estimator. Moreover, the bias involves second order derivatives that are very difficult to estimate. One possible way how handle this problem is to 'undersmooth', i.e. to choose a bandwidth that is of a smaller order than the optimal bandwidth (typically such that $n\,h_n^5 \to 0$ as $n \to \infty$). The price to pay is that a 'suboptimal' bandwidth which results in a higher (asymptotic) mean squared error of the estimator. Moreover, in applications it is almost impossible to specify the appropriate level of 'undersmoothing' that would guarantee that the bias can be safely ignored.

That is why as an alternative inference procedure Omelka et al. (2013) suggested a bootstrap approximation inspired by the resampling procedure introduced in Aerts et al. (1994).

## 3.4 Conditional measures of association

In many situations we would like to quantify the degree of dependence by only one number. In nonparametric statistics one uses a measure that does not depend on marginal distributions and thus can be expressed as a functional of a copula. With the help of conditional copulas it is straightforward to define conditional versions of these measures. We will illustrate this for Kendall's tau and Spearman's rho that are probably the most widely used nonparametric measures of associations.

### 3.4.1 Kendall's tau

For random variables $(Y_1, Y_2)^{\mathsf{T}}$ Kendall's tau is defined as

$$\tau = 2\,\mathsf{P}\big((Y_1 - Y_1')(Y_2 - Y_2') > 0\big) - 1,$$

where $(Y_1', Y_2')^{\mathsf{T}}$ is an independent copy of the random vector $(Y_1, Y_2)^{\mathsf{T}}$. It is well known (see e.g. Nelsen, 2006) that if $C$ is the copula for the vector $(Y_1, Y_2)^{\mathsf{T}}$, then $\tau$ may be expressed as

$$\tau = 4 \iint C(u_1, u_2)\, dC(u_1, u_2) - 1.$$

This leads immediately to an expression for the population version of the conditional Kendall's tau of $(Y_1, Y_2)^{\mathsf{T}}$ given $X = x$

$$\tau(x) = 4 \iint C_x(u_1, u_2)\, dC_x(u_1, u_2) - 1, \tag{3.20}$$

where $C_x$ is the appropriate conditional copula. The interpretation of the conditional Kendall's tau is

$$\tau(x) = 2\,\mathsf{P}\big((Y_1 - Y_1')(Y_2 - Y_2') > 0 \,|\, X = X' = x\big) - 1,$$

where $(Y_1', Y_2', X')^{\mathsf{T}}$ is an independent copy of the random vector $(Y_1, Y_2, X)^{\mathsf{T}}$.

The most straightforward way to estimate the conditional Kendall's tau is to replace the unknown quantity $C_x$ in (3.20) with the estimate $\widehat{C}_x$ to get

$$\widehat{\tau}_n^{I}(x) = 4 \iint \widehat{C}_x(u_1, u_2)\, d\widehat{C}_x(u_1, u_2) - 1. \tag{3.21}$$

Although expression (3.21) is convenient for exploring asymptotic properties of the estimator, in finite samples we have a slightly better experience with the formula

$$\widehat{\tau}_n(x) = \frac{4}{1 - W_n} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ni}(x, h_n)\, w_{nj}(x, h_n)\, \mathbb{I}\{Y_{1i} < Y_{1j}, Y_{2i} < Y_{2j}\} - 1, \tag{3.22}$$

where $W_n = \sum_{i=1}^{n} w_{ni}^2(x, h_n)$. Note that (3.22) mimics the formula for (unconditional) Kendall's tau estimation

$$\widehat{\tau}_n = \frac{4}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{I}\{Y_{1i} < Y_{1j}, Y_{2i} < Y_{2j}\} - 1.$$

Further, to improve the bias properties of the estimator of conditional Kendall's tau ($\widehat{\tau}_n(x)$) given by (3.22) it would be useful to replace the original observations $(Y_{1i}, Y_{2i})^{\mathsf{T}}$ with the observations already adjusted for the possible effect of the covariate on the margins as discussed in Section 3.3.2.

As pointed out by Veraverbeke et al. (2011) the asymptotic normality of $\widehat{\tau}_n(x)$ can be derived by Theorem 3.1 (or Theorem 3.2) from the Hadamard differentiability of Kendall's tau that is proved therein. On the other we conjecture that investigating directly the estimator $\widehat{\tau}_n(x)$ would yield asymptotic normality under less restrictive assumptions.

### 3.4.2 Spearman's rho

As the unconditional version of Spearman's rho may be expressed as

$$\rho = 12 \iint C(u_1, u_2)\, du_1\, du_2 - 3$$

the population conditional version is thus given by

$$\rho(x) = 12 \iint C_x(u_1, u_2)\, du_1\, du_2 - 3,$$

which may be estimated as

$$\widehat{\rho}_n(x) = 12 \iint \widehat{C}_x(u_1, u_2)\, du_1\, du_2 - 3 = 12 \sum_{i=1}^{n} w_{ni}(x, h_n)(1 - \widehat{U}_{1i})(1 - \widehat{U}_{2i}) - 3.$$

For interpretations of the population (unconditional) Spearman's rho see Nelsen (2006, Chapter 5.1.2).

## 3.5 Partial and average measures of association

Sometimes it might be desirable to summarize the dependence of $(Y_1, Y_2)^{\mathsf{T}}$ when $X$ is taken into account by one number. To do that one can use either partial measures of dependence or average conditional measures of dependence.

Gijbels et al. (2015a) suggested that *the partial measures* of association can be constructed analogously as the conditional measures of dependence but with the conditional copula replaced by the partial copula. This partial copula ($\bar{C}$) can be viewed as the copula of the random vector $\big(U_1, U_2\big)^{\mathsf{T}} = \big(F_{1X}(Y_1), F_{2X}(Y_2)\big)^{\mathsf{T}}$. In words, the partial copula is the copula of the response vector $(Y_1, Y_2)^{\mathsf{T}}$ adjusted for the effect of the covariate $X$ on the marginal distributions. Now for instance the partial Kendall's tau is defined as

$$\bar{\tau} = 4 \iint \bar{C}(u_1, u_2)\, d\bar{C}(u_1, u_2) - 1.$$

As an alternative to the partial measures one can use *average conditional measures*. These are simply defined as the expectations of the corresponding conditional measures. That is for instance average conditional Kendall's tau is defined as

$$\tau^A = \mathsf{E}\, \tau(X) = \int \tau(x)\, dF_X(x),$$

where $F_X$ is the distribution function of the covariate $X$. As discussed in Gijbels et al. (2015a) average conditional measures are generally different from partial measures but they coincide if the measure is a linear functional of a copula. Asymptotic properties of the estimators of these measures of association are investigated in Gijbels et al. (2015a).

# Chapter 4

# Nonparametric conditional copula estimation under simplifying assumptions

As in Chapter 3 suppose the observations are independent identically distributed three-dimensional vectors $(Y_{11}, Y_{21}, X_1)^\mathsf{T}, \ldots, (Y_{1n}, Y_{2n}, X_n)^\mathsf{T}$ from $(Y_1, Y_2, X)^\mathsf{T}$ with the joint distribution function $H$. But now the conditional copula $C_x$ does not depend on the value of $x$ and thus the conditional distribution function $H_x$ can be written as in (1.4).

Note, that if $Y_1$ and $Y_2$ were independent of $X$ then one could simply use the empirical copula estimator as introduced in (2.1). Thus the obvious idea is to replace the original observations $Y_{1i}$ and $Y_{2i}$ in (2.1) with observations that are already adjusted for the effect of the covariate $X$. Suppose for a moment that the conditional distribution functions $F_{1x}$ and $F_{2x}$ are known. Then one can construct the 'ideal' observations $(U_{1i}, U_{2i})^\mathsf{T}$ for estimating copula $C$ in (1.4) by (3.3). Thus an 'oracle' estimator of the copula function $C$ would be

$$C_n^{(or)}(u_1, u_2) = G_n\big(G_{1n}^{-1}(u_1), G_{2n}^{-1}(u_2)\big), \tag{4.1}$$

where

$$G_n(u_1, u_2) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\{U_{1i} \le u_1, U_{2i} \le u_2\},$$

and $G_{1n}$ and $G_{2n}$ are its corresponding marginals.

In case the conditional distribution functions $F_{1x}$ and $F_{2x}$ are unknown, we consider estimates for them, denoted by $\widehat{F}_{1x}$ and $\widehat{F}_{2x}$, and put

$$\big(\widetilde{U}_{1i}, \widetilde{U}_{2i}\big)^\mathsf{T} = \big(\widehat{F}_{1X_i}(Y_{1i}), \widehat{F}_{2X_i}(Y_{2i})\big)^\mathsf{T}, \quad i = 1, \ldots, n,$$

for the corresponding estimates of the unobserved $(U_{1i}, U_{2i})$ given in (3.3).

Mimicking formula (4.1) we construct the estimator of the copula function $C$ as

$$\widetilde{C}_n(u_1, u_2) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\big\{\widetilde{U}_{1i} \le \widetilde{G}_{1n}^{-1}(u_1), \widetilde{U}_{2i} \le \widetilde{G}_{2n}^{-1}(u_2)\big\}, \tag{4.2}$$

where

$$\widetilde{G}_{jn}(u) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\big\{\widetilde{U}_{ji} \le u\big\}, \quad j = 1, 2.$$

Note that in comparison to the estimator $\widetilde{C}_{xh}$ given by (3.6) there is no need to use the smoothing weights $w_{ni}(x, h_n)$ as the conditional copula $C_x$ does not depend on $x$ anymore. In what follows we present the asymptotic properties of the estimator $\widetilde{C}_n$ as derived in Gijbels et al. (2015b). Generally speaking, the results depend on what can be assumed about the dependence (on $X$) of the marginal distributions of $Y_1$ and $Y_2$. The more specific knowledge we have about the dependence of the marginals on $X$, the stronger the established result about $\widetilde{C}_n$ (under milder assumptions on $C$).

## 4.1 Consistency and weak convergence of the general estimator

The following theorem states that if both marginal distribution functions $F_{1x}$ and $F_{2x}$ allow uniformly consistent estimation at rate $r_n$ (a sequence of nonnegative numbers tending to zero with $n$), then the copula estimator $\widetilde{C}_n$ is also uniformly consistent with that rate (provided $r_n$ tends to zero not faster than $1/\sqrt{n}$).

**Theorem 4.1.** *Suppose that for $j = 1, 2$ (with $R_X$ the domain of $X$)*

$$\sup_{x \in R_X} \sup_{y \in \mathbb{R}} \big|\widehat{F}_{jx}(y) - F_{jx}(y)\big| = O_P(r_n), \quad \textit{as} \quad n \to \infty.$$

*Then*

$$\sup_{(u_1,u_2)\in[0,1]^2} \big|\widetilde{C}_n(u_1, u_2) - C(u_1, u_2)\big| = O_P\big(\max\{r_n, \tfrac{1}{\sqrt{n}}\}\big), \quad \textit{as} \quad n \to \infty. \tag{4.3}$$

To prove the weak convergence of the estimator $\widetilde{C}_n(u_1, u_2)$ (as a process on $[0, 1]^2$) one needs to assume more about the behaviour of $\widehat{F}_{jx}(y)$.

First, we need to control the (random) set of functions on $R_X \times \mathbb{R}^2$

$$\mathcal{F}_n = \Big\{(x, y_1, y_2) \mapsto \mathbb{I}\{y_1 \le \widehat{F}_{1x}^{-1}(u_1), y_2 \le \widehat{F}_{2x}^{-1}(u_2)\},\ u_1, u_2 \in [0, 1]\Big\}. \tag{4.4}$$

Second, we need a finer control of the rate of the process

$$Y_{jx}(u) = F_{jx}\Big(\widehat{F}_{jx}^{-1}\big(\widetilde{G}_{jn}^{-1}(u)\big)\Big) - u, \quad u \in [0, 1], \tag{4.5}$$

with a special attention to $u$ close to the endpoints of the unit interval. Roughly speaking, the faster the convergence of $Y_{jx}$ to zero, the lesser smoothness of the copula function $C$ is required. Two possible modes of this interplay are given in the following assumptions.

### General regularity assumptions

In what follows let $\epsilon_n = \epsilon/\sqrt{n}$ and $I_n = [\epsilon_n, 1 - \epsilon_n]$.

**(Yp)** For each $\epsilon > 0$, uniformly in $x \in R_X$ (the domain of $X$) and $u \in I_n$

$$Y_{jx}(u) = r(u)\, O_P\big(\tfrac{1}{\sqrt{n}}\big), \quad j = 1, 2,$$

where $r : [0, 1] \to \mathbb{R}$ is a bounded function such that

$$\lim_{u \to 0_+} r(u) = 0 \text{ and } \lim_{u \to 1_-} r(u) = 0.$$

**(Yn)** For each $\epsilon > 0$, uniformly in $x \in R_X$ and $u \in I_n$

$$Y_{jx}(u) = \sqrt{u(1-u)}\, o_P(n^{-1/4}), \quad j = 1, 2.$$

**(Cp)** The first-order partial derivatives $C^{(1)}$ and $C^{(2)}$ of the copula function $C$ exist and for $j = 1, 2$, $C^{(j)}$ is continuous on the set $\{(u_1, u_2) \in [0,1]^2 : 0 < u_j < 1\}$.

**(Cn)** The second-order partial derivatives o of the copula function $C$ satisfies (2.9).

As already noted at the end of Section 2.2.4 assumption **(Cp)** is the weakest assumption assumed so far to establish the weak convergence of the empirical copula process. The assumption **(Cn)** is slightly more restrictive, but it is satisfied for commonly-used copulas, see Omelka et al. (2009).

**Theorem 4.2.** *Suppose that there exists a set of functions $\mathcal{F}$ on $R_X \times [0,1]^2$ that is $P_H$-Donsker (with $P_H$ the probability measure associated with the joint distribution function $H$) and such that*

$$\lim_{n\to\infty} \mathsf{P}\left(\mathcal{F}_n \subset \mathcal{F}\right) = 1.$$

*Further suppose that either* **(Yp)** *and* **(Cp)** *or* **(Yn)** *and* **(Cn)** *hold. Then*

$$\sup_{(u_1,u_2)\in[0,1]^2} \left| \sqrt{n}\left(\widetilde{C}_n(u_1,u_2) - C_n^{(or)}(u_1,u_2)\right)\right| = o_P(1), \quad \text{as} \quad n \to \infty. \tag{4.6}$$

Theorem 4.2 together with the standard result on the weak convergence of the empirical copula process $\sqrt{n}\big(C_n^{(or)} - C\big)$ immediately implies the following corollary.

**Corollary 4.1.** *Suppose the assumptions of Theorem 4.2 are satisfied. Then the process*

$$\widetilde{\mathbb{C}}_n^{(E)}(u_1,u_2) = \sqrt{n}\left(\widetilde{C}_n(u_1,u_2) - C(u_1,u_2)\right), \qquad (0 \le u_1, u_2 \le 1),$$

*converges in the space $\ell^\infty\big([0,1]^2\big)$ to a centred Gaussian process $G_C$ described in Theorem 2.1.*

## 4.2 Location-scale models for $F_{1x}$ and $F_{2x}$

Often one is able to specify location-scale regression models for the influence of the covariate $X$ on the marginal distributions:

$$Y_1 = m_1(X) + \sigma_1(X)\,\varepsilon_1, \qquad Y_2 = m_2(X) + \sigma_2(X)\,\varepsilon_2, \tag{4.7}$$

where $m_1, m_2, \sigma_1, \sigma_2$ are functions (either fully or partially known) and $\varepsilon_1$ and $\varepsilon_2$ are independent of $X$ with unknown distribution functions $F_{1\varepsilon}$ and $F_{2\varepsilon}$.

Note that under this general location-scale model: for $j = 1, 2$

$$F_{jx}(y_j) = F_{j\varepsilon}\left(\tfrac{y_j - m_j(x)}{\sigma_j(x)}\right) \quad \text{and therefore} \quad U_{ji} = F_{jX_i}(Y_{ji}) = F_{j\varepsilon}\left(\tfrac{Y_{ji} - m_j(X_i)}{\sigma_j(X_i)}\right),$$

with $F_{j\varepsilon}(y) = \mathsf{P}(\varepsilon_j \le y)$.

Let $\widehat{m}_1$, $\widehat{m}_2$, $\widehat{\sigma}_1$ and $\widehat{\sigma}_2$ be estimates of respectively the functions $m_1$, $m_2$, $\sigma_1$ and $\sigma_2$. Then, in this general setting

$$\widehat{F}_{jx}(z) = \widehat{F}_{j\widehat{\varepsilon}}\big(\tfrac{z-\widehat{m}_j(x)}{\widehat{\sigma}_j(x)}\big),$$

where

$$\widehat{F}_{j\widehat{\varepsilon}}(z) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\{\widehat{\varepsilon}_{ji} \le z\}, \qquad \text{with} \qquad \widehat{\varepsilon}_{ji} = \tfrac{Y_{ji}-\widehat{m}_j(X_i)}{\widehat{\sigma}_j(X_i)}. \tag{4.8}$$

The estimator (4.2) then equals

$$\widetilde{C}_n(u_1,u_2) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\left\{\widehat{F}_{1\widehat{\varepsilon}}(\widehat{\varepsilon}_{1i}) \le u_1, \widehat{F}_{2\widehat{\varepsilon}}(\widehat{\varepsilon}_{2i}) \le u_2\right\}, \tag{4.9}$$

since $\big\{\widehat{F}_{j\widehat{\varepsilon}}(\widehat{\varepsilon}_{ji}) \le u_j\big\} = \{\widehat{F}_{j\widehat{\varepsilon}}(\widehat{\varepsilon}_{ji}) \le \widetilde{G}_{jn}^{-1}(u_j)\}$.

In this location-scale model setting, the (random) set of functions on $R_X \times \mathbb{R}^2$ (denoted by $\mathcal{F}_n$; see (4.4)), that enters in the mathematical derivations is

$$\mathcal{F}_n = \Big\{(x,y_1,y_2) \mapsto \mathbb{I}\{y_1 \le z_1\,\widehat{\sigma}_1(x) + \widehat{m}_1(x), y_2 \le z_2\,\widehat{\sigma}_2(x) + \widehat{m}_2(x)\},\ z_1,z_2 \in \mathbb{R}\Big\}.$$

and the process $Y_{jx}(u)$ (see (4.5)) that is involved is given by

$$Y_{jx}(u) = F_{jx}\big(\widehat{F}_{jx}^{-1}(u)\big) - u = F_{j\varepsilon}\left(\tfrac{\widehat{F}_{j\widehat{\varepsilon}}^{-1}(u)\widehat{\sigma}_j(x)+m_j(x)-\widehat{m}_j(x)}{\sigma_j(x)}\right) - u, \quad u \in [0,1].$$

See further Gijbels et al. (2015b, Appendix C).

In the sequel of this section we distinguish between parametric and nonparametric location-scale models (for both marginals). Clearly, a possibility could also be to assume a parametric location-scale model for one marginal component and a nonparametric one for the other marginal component. It then suffices to combine the results of both parts below. We do not elaborate on this further.

### 4.2.1 Parametric location-scale models for $F_{1x}$ and $F_{2x}$

Assume now that the following parametric location-scale models hold:

$$Y_1 = m_1(X,\boldsymbol{\theta}_1) + \sigma_1(X,\boldsymbol{\theta}_1)\,\varepsilon_1, \qquad Y_2 = m_2(X,\boldsymbol{\theta}_2) + \sigma_2(X,\boldsymbol{\theta}_2)\,\varepsilon_2, \tag{4.10}$$

where $m_1, m_2, \sigma_1, \sigma_2$ are (partially) known functions, up to unknown finite-dimensional parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. Finally $\varepsilon_1$ and $\varepsilon_2$ are as in (4.7).

Let $\widehat{\boldsymbol{\theta}}_1$ and $\widehat{\boldsymbol{\theta}}_2$ be estimates of the unknown parameters. Then, in this special setting:

$$\widehat{F}_{jx}(z) = \widehat{F}_{j\widehat{\varepsilon}}\Big(\tfrac{z-m_j(x,\widehat{\boldsymbol{\theta}}_j)}{\sigma_j(x,\widehat{\boldsymbol{\theta}}_j)}\Big),$$

where

$$\widehat{F}_{j\widehat{\varepsilon}}(z) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\{\widehat{\varepsilon}_{ji} \le z\}, \qquad \text{with} \qquad \widehat{\varepsilon}_{ji} = \tfrac{Y_{ji}-m_j(X_i,\widehat{\boldsymbol{\theta}}_j)}{\sigma_j(X_i,\widehat{\boldsymbol{\theta}}_j)}.$$

The estimator is then defined as in (4.9) and we denote this estimator as $\widetilde{C}_n^{(\mathrm{P})}$.

*Remark.* Note that if model (4.10) is not true then the estimator $\widetilde{C}_n^{(\mathrm{P})}$ does in general not estimate the copula $C$ of $(U_1, U_2)$. In order to characterize what is being estimated suppose that there exist $\boldsymbol{\theta}_1^*$ and $\boldsymbol{\theta}_2^*$ such that the estimates $\widehat{\boldsymbol{\theta}}_1$ and $\widehat{\boldsymbol{\theta}}_2$ converge in probability to the corresponding quantities. Further put

$$\mathbf{Y}^a = (Y_1^a, Y_2^a)^{\mathsf{T}} = \left(\tfrac{Y_1 - m_1(X,\boldsymbol{\theta}_1^*)}{\sigma_1(X,\boldsymbol{\theta}_1^*)}, \tfrac{Y_2 - m_2(X,\boldsymbol{\theta}_2^*)}{\sigma_2(X,\boldsymbol{\theta}_2^*)}\right)^{\mathsf{T}},$$

Then $\widetilde{C}_n^{(\mathrm{P})}$ estimates the following copula

$$C^{Y_1^a,Y_2^a}(u_1,u_2) = \int C\Big(F_{1x}^a\big(F_{Y_1^a}^{-1}(u_1)\big), F_{2x}^a\big(F_{Y_2^a}^{-1}(u_2)\big)\Big) f_X(x)\,dx, \tag{4.11}$$

where

$$F_{jx}^a(y) = \mathsf{P}\left(Y_j^a \leq y \,|\, X = x\right) = F_{jx}\big(y\,\sigma_j(x,\boldsymbol{\theta}_j^*) + m_j(x,\boldsymbol{\theta}_j^*)\big),$$

$F_{Y_j^a}(y) = \mathsf{E}_X F_{jX}^a(y)$, and $f_X$ be the marginal density of $X$. Thus in general, if model (4.10) does not hold then $\widetilde{C}_n^{(\mathrm{P})}$ is not a consistent estimator of $C$. On the other hand, if model (4.10) is not too far from the reality, then one can expect that $C^{Y_1^a,Y_2^a}$ is close to $C$. As the bias that results in estimating $C^{Y_1^a,Y_2^a}$ instead of $C$ can be, for small and moderate samples, negligible when compared to the variability of the estimator, it can be advantageous to use model (4.10) to adjust for the effect of the covariate on the marginal distribution even if this model does not hold. See also the simulation study in Gijbels et al. (2015b).

### Regularity assumptions

- **($\boldsymbol{\theta}$)** The estimates $\widehat{\boldsymbol{\theta}}_1$ and $\widehat{\boldsymbol{\theta}}_2$ are $\sqrt{n}$-consistent.
- **(F1p)** The functions $f_{j\varepsilon}(y)\,(1+y)$ are bounded, where $f_{j\varepsilon} = F_{j\varepsilon}'$.
- **(F2p)** The functions $f_{j\varepsilon}\big(F_{j\varepsilon}^{-1}(u)\big)$ and $f_{j\varepsilon}\big(F_{j\varepsilon}^{-1}(u)\big)\,F_{j\varepsilon}^{-1}(u)$ tend to zero when $u \to 0_+$ or $u \to 1_-$.
- **($\boldsymbol{m\sigma p}$)** The partial derivatives of the functions $m_j(x,\cdot)$ and $\sigma_j(x,\cdot)$ with respect to their second arguments are uniformly bounded as functions of both arguments at $(R_X, U(\boldsymbol{\theta}_j))$, where $R_X$ is the support of $X$ and $U(\boldsymbol{\theta}_j)$ is an open neighbourhood of $\boldsymbol{\theta}_j$. Moreover, $\inf_{x\in R_X,\mathbf{c}\in U(\boldsymbol{\theta}_j)} \sigma_j(x,\mathbf{c}) > 0$.

The following theorem establishes the rate of convergence of the estimator $\widetilde{C}_n^{(\mathrm{P})}$.

**Theorem 4.3.** *Suppose that assumptions* **($\boldsymbol{\theta}$)**, **(F1p)** *and* **($\boldsymbol{m\sigma p}$)** *are satisfied. Then the estimator* $\widetilde{C}_n^{(P)}$ *is* $\sqrt{n}$*-consistent, i.e.* (4.3) *holds with* $r_n = 1/\sqrt{n}$.

The next theorem establishes that under slightly stronger conditions (among which assuming **(F2p)** instead of **(F1p)**) one can even show that the estimator $\widetilde{C}_n^{(\mathrm{P})}$ is asymptotically equivalent to the oracle estimator $C_n^{(or)}$.

**Theorem 4.4.** *Statement* (4.6) *and the statement of Corollary 4.1 hold provided assumptions* **(Cp)**, **($\boldsymbol{\theta}$)**, **(F2p)** *and* **($\boldsymbol{m\sigma p}$)** *are satisfied.*

*Remark.* From the proof of Theorem 4.4 it follows that assumption **(F2p)** is needed to control the behaviour of the copula process $\widetilde{\mathbb{C}}_n^{(\mathrm{E})}$ close to the border of the unit square (more precisely close to the possible points of discontinuity of the first order partial derivatives of $C$). Note that **(F2p)** is not satisfied when the error distribution $\varepsilon_j$ follows a distribution whose density is not continuous (e.g. uniform, exponential distribution). For such distributions one can prove the weak convergence of the process $\widetilde{\mathbb{C}}_n^{(\mathrm{E})}$ on the set that excludes problematic points. More precisely if the functions $g_j(\delta) = F_{j\varepsilon}\big(y(1+\delta)+\delta\big)$ are Lipschitz for $j = 1, 2$ uniformly in $y$, then for every $\delta > 0$

$$\sup_{(u_1,u_2)\in[0,1]^2} \big|\widetilde{\mathbb{C}}_n^{(\mathrm{E})}(u_1,u_2)\big| = O_P(1), \qquad \widetilde{\mathbb{C}}_n^{(\mathrm{E})}\big|_{[\delta,1-\delta]^2} \xrightarrow[n\to\infty]{d} G_C\big|_{[\delta,1-\delta]^2},$$

where $\big|_{[\delta,1-\delta]^2}$ stands for the restriction of the process to $I_\delta = [\delta, 1-\delta]^2$. Such a result can often be sufficient to derive an asymptotic distribution of various functionals of copulas as e.g. Kendall's tau or Spearman's rho. For brevity we do not elaborate further on this.

### 4.2.2 Nonparametric location-scale models for $F_{1x}$ and $F_{2x}$

In nonparametric location-scale models (4.7) the functions $m_1$, $m_2$, $\sigma_1$ and $\sigma_2$ are fully unknown, and need to be estimated.

For simplicity of presentation we concentrate on local linear regression estimates of the functions $m(x)$ and $\sigma^2(x)$ defined as follows:

$$(4.12)\quad \widehat{m}_j(x) = \sum_{i=1}^n w_{ni}(x, g_{jn})\, Y_{ji}, \quad \widehat{\sigma}_j^{\,2}(x) = \sum_{i=1}^n w_{ni}(x, g_{jn})\, \big(Y_{ji} - \widehat{m}_j(X_i)\big)^2, \quad j = 1, 2,$$

where $g_{jn}$ is a bandwidth sequence going to zero. The weights $w_{ni}(x, g_{jn})$ in (4.12) are local linear weights given by (3.9).

The estimated residuals of model (4.7) are given by $\widehat{\varepsilon}_{ji}$ in (4.8), with $\widehat{m}_j$ and $\widehat{\sigma}_j$ as in (4.12). The copula function $C$ is then estimated by (4.9). Denote this estimator as $\widetilde{C}_n^{(\mathrm{NP})}$.

*Remark.* For simplicity of presentation the same bandwidth $g_{jn}$ is used in (4.12) for estimation of the location $m_j$ as well as estimation of the scale $\sigma_j$. Different bandwidths for location and scale estimation can be used provided that both satisfy the assumptions on the bandwidths given below.

#### Regularity assumptions

**(Bw)** For $j = 1, 2$: $g_{jn} = o(1)$ and $\frac{n\, g_{jn}}{\log n} \to \infty$.

**(Bwn)** For $j = 1, 2$ and for some $\delta > 0$: $\frac{n\, g_{jn}^{3+\delta}}{\log n} \to \infty$ and $n\, g_{jn}^5 \to 0$.

**(F1n)** The functions $F_{1\varepsilon}(y)$ and $F_{2\varepsilon}(y)$ are twice continuously differentiable and $\mathsf{E}\,\varepsilon_j^4 < \infty$.

**(F2n)** The functions $f_{j\varepsilon}\big(F_{j\varepsilon}^{-1}(u)\big)\big(1 + F_{j\varepsilon}^{-1}(u)\big)$ are of order $O(u^\delta(1-u)^\delta)$ for some $\delta \geq \frac{1}{4}$ and the functions $f'_{j\varepsilon}(y)(1+y)^2$ are bounded.

**(k)** The kernel $k$ is a symmetric and continuous function with support $[-1, 1]$.

**(kn)** The kernel $k$ is twice continuously differentiable, symmetric with support $[-1,1]$ and decreasing on $[0,1)$.

**(X)** The support $R_X$ of $X$ is a non-empty finite interval $(a,b)$ and $\inf_{x\in R_X} f_X(x) > 0$.

**(Xn)** The support $R_X$ of $X$ is a non-empty interval $(a,b)$, $\inf_{x\in R_X} f_X(x) > 0$ and $f_X(x)$ is twice continuously differentiable in $R_X$.

**($\boldsymbol{m\sigma}$1)** The second order derivatives of $m_j$ and $\sigma_j$ are bounded on the interior of $R_X$ and $\inf_{x\in R_X} \sigma_j(x) > 0$.

**($\boldsymbol{m\sigma}$2)** The functions $m_j$ and $\sigma_j$ are three-times continuously differentiable on the interior of $R_X$ and $\inf_{x\in R_X} \sigma_j(x) > 0$.

The following theorem states the rate of convergence of the estimator $\widetilde{C}_n^{(\mathrm{NP})}$.

**Theorem 4.5.** *Suppose that assumptions* **(Bwn)**, **(F1p)**, **(k)**, **($\boldsymbol{m\sigma}$1)** *and* **(Xn)** *are satisfied. Then the estimator $\widetilde{C}_n^{(NP)}$ satisfies* (4.3) *with*

$$r_n = \max\Big\{g_{1n}^2, g_{2n}^2, \sqrt{\tfrac{\log n}{n\, g_{1n}}}, \sqrt{\tfrac{\log n}{n\, g_{2n}}}\Big\}. \tag{4.13}$$

The next theorem establishes that under slightly stronger conditions one can improve the rate $r_n$ to $1/\sqrt{n}$ and moreover show that the estimator $\widetilde{C}_n^{(\mathrm{NP})}$ is even asymptotically equivalent to the oracle estimator $C_n^{(or)}$.

**Theorem 4.6.** *Statement* (4.6) *and the statement of Corollary 4.1 hold provided assumptions* **(Cn)**, **(Bwn)**, **(F1n)**, **(F2n)**, **(kn)**, **($\boldsymbol{m\sigma}$2)** *and* **(Xn)** *are satisfied.*

*Remark.* Making use of the Hadamard differentiability of the 'copula mapping' $\Phi : G \mapsto G(G_1^{-1}, G_2^{-1})$ proved in Theorem 2.4 of Bücher and Volgushev (2013) one can replace assumptions **(Cn)** and **(F2n)** by the following assumption on the joint distribution function $F_{\boldsymbol{\varepsilon}}$ of $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2)^{\mathsf{T}}$.

**($\mathbf{F}_\varepsilon$)** The second-order partial derivatives $F_\varepsilon^{(1,1)}$, $F_\varepsilon^{(1,2)}$ and $F_\varepsilon^{(2,2)}$ of the joint cumulative distribution function $F_\varepsilon(y_1,y_2) = \mathsf{P}(\varepsilon_1 \le y_1, \varepsilon_2 \le y_2)$, with $F_\varepsilon^{(j,k)}(y_1,y_2) = \frac{\partial^2 F_\varepsilon(y_1,y_2)}{\partial y_j \partial y_k}$, satisfy

$$\max_{j,k\in\{1,2\}} \sup_{y_1,y_2\in\mathbb{R}^2} \big|F_\varepsilon^{(j,k)}(y_1,y_2)(1+y_j)(1+y_k)\big| < \infty.$$

Further the innovation density $f_{j\varepsilon}$ $(j=1,2)$ satisfies

$$\lim_{u\to 0_+} \big(1+F_{j\varepsilon}^{-1}(u)\big) f_{j\varepsilon}\big(F_{j\varepsilon}^{-1}(u)\big) = 0 \quad \text{and} \quad \lim_{u\to 1_-} \big(1+F_{j\varepsilon}^{-1}(u)\big) f_{j\varepsilon}\big(F_{j\varepsilon}^{-1}(u)\big) = 0.$$

This is generally less restrictive assumption as it allows for an interplay between the marginal distributions $F_{1\varepsilon}$ and $F_{2\varepsilon}$ and the copula function $C$, i.e. the more it is assumed about $F_{1\varepsilon}$ and $F_{2\varepsilon}$, the less one can assume about $C$ and vice versa. The proof would go analogously as in Neumeyer et al. (2017).

*Remark.* Analogously as in the previous section, if model (4.7) does not hold, then the estimator $\widetilde{C}_n^{(\mathrm{NP})}$ does in general not estimate the copula $C$ but the copula $C^{Y_1^a,Y_2^a}$ given by (4.11), where $Y_j^a = \frac{Y_j - m_j(X)}{\sigma_j(X)}$ with $m_j(x) = \mathsf{E}\,[Y_j \,|\, X=x]$ and $\sigma_j^2(x) = \mathsf{var}\,[Y_j \,|\, X=x]$.

## 4.3 General nonparametric estimation of $F_{1x}$ and $F_{2x}$

Sometimes, one has no idea about the influence of $X$ on $Y_1$ and $Y_2$. Then one can construct general nonparametric estimators of $F_{1x}$ and $F_{2x}$ for instance as in (3.4).

### Regularity assumption

**(F)** The second derivatives, with respect to $x$, of the functions $F_{1x}(y)$ and $F_{2x}(y)$ are bounded.

The following theorem follows directly from Theorem 4.1 and Lemma 4 in Appendix D of Gijbels et al. (2015b).

**Theorem 4.7.** *Suppose that assumptions* **(Bw)**, **(F)**, **(k)** *and* **(X)** *are satisfied, then Theorem 4.1 holds with $r_n$ given in (4.13).*

It is worth noting that the best possible rate, from this theorem, is $r_n = \big(\frac{\log n}{n}\big)^{2/5}$ obtained by taking $g_{1n} = g_{2n} = O\Big(\big(\frac{\log n}{n}\big)^{1/5}\Big)$.

Note that in this general setting Gijbels et al. (2015b) were not able to establish the asymptotic equivalence (4.6) of the estimator $\widetilde{C}_n$ with the oracle estimator $C_n^{(or)}$. This was proved later by Portier and Segers (2015) where the authors consider smooth versions of the estimates of the conditional distribution functions $F_{jx}$.

## 4.4 Further extensions and discussion

The question of interest is what happens if the pairwise simplifying assumption (1.4) does not hold and the conditional copula function $C_x$ still depends on $x$. This is discussed Gijbels et al. (2015a). The authors of that paper show that (provided that the margins are properly adjusted for the effect of the covariate) the estimator $\widetilde{C}_n$ given by (4.2) estimates the partial copula which coincides with the average conditional copula and is given by

$$\bar{C}(u_1, u_2) = \int C_x(u_1, u_2)\, dF_X(x).$$

Further in Gijbels et al. (2015a) it is shown that if the margins follows a parametric or nonparametric location scale model as in Sections 4.2.1 and 4.2.2 then the estimator $\widetilde{C}_n$ is still $\sqrt{n}$-consistent of the partial copula $\bar{C}$. But in contrast to Theorems 4.4 and 4.6 the effect of using estimates $\widehat{F}_{jx}$ instead of the true $F_{jx}$ does not diminish and affects the limiting distribution.

As the pairwise simplifying assumption (1.4) is crucial in applications, it is of interest to have a test of the null hypothesis that the conditional copula does not depend on the covariate. A semiparametric approach can be found in Acar et al. (2013). Nevertheless, this approach needs that the copula family is well specified. To circumvent this problem Gijbels et al. (2017b) introduced a purely nonparametric test based on the conditional Kendall's tau (see Section 3.4.1). Later, Gijbels et al. (2017a) suggested a semiparametric test which works even if the copula family is misspecified.

Another question of interest is whether the asymptotic equivalence of the estimator $\widetilde{C}_n$ with the oracle estimator $C_n^{(or)}$ holds also in more complicated settings than just for independent and identically distributed random variables. A generalization to time series settings is considered in Neumeyer et al. (2017).

Chapter 5

# Nonparametric conditional distribution function estimation

Suppose we observe independent identically distributed vectors $(X_1, Y_1)^\mathsf{T}, \dots, (X_n, Y_n)^\mathsf{T}$ from $(X, Y)^\mathsf{T}$ with the cumulative distribution function $F(x, y)$. For various research questions it is necessary to estimate the conditional distribution function of $Y$ given $X = x$

$$F_x(y) = \mathsf{P}(Y \leq y \,|\, X = x), \quad y \in \mathbb{R}.$$

Note the estimator of the conditional distribution function is also necessary in conditional copula estimation as discussed in previous chapters.

When estimating $F_x$ a crucial information is whether one can make any assumption about the effect of the covariate $X$ on the response $Y$. In parametric approaches one assumes that $Y$ given $X = x$ follows a given distribution $F_x(y, \boldsymbol{\theta})$ known only up to the unknown parameter $\boldsymbol{\theta}$. A possible semiparametric approach would be to assume the location-scale model as in (4.10). Another alternative would be to use a nonparametric location-scale model

$$Y = m(X) + \sigma(X)\,\varepsilon, \tag{5.1}$$

where $m$ and $\sigma$ are unknown (smooth) functions and $\varepsilon$ is independent of $X$, with an unknown distribution. If an analyst is not willing to make any assumption about the relationship of $X$ and $Y$, then he/she is usually advised to use a fully nonparametric estimator of $F_x(y)$, for instance a weighted empirical distribution function

$$\widehat{F}_x(y) = \sum_{i=1}^{n} w_{ni}(x, h_n)\,\mathbb{I}\{Y_i \leq y\}, \tag{5.2}$$

where $\{w_{ni}(x, h_n)\}$ is a sequence of weights (e.g. Nadaraya-Watson weights, local linear weights, …) and $h_n > 0$ is a bandwidth sequence tending to zero as the sample size increases.

Often, it seems evident that a given model explains an important part of the variability of the observed data, but at the same time one is not willing to assume that the model completely describes the data generation process. In the following we propose a more general approach to improve the nonparametric estimator (5.2) of the conditional distribution function $F_x$, that is not relying on a specific model structure but is only inspired by it.

To illustrate the proposed estimation method consider a simple linear regression model

$$Y_i = \theta_0 + \theta_1 X_i + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{5.3}$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are identically independently distributed. Suppose that $\theta_1$ is large. Let $F_\varepsilon$ be the distribution function of the error term $\varepsilon$. Then the conditional distribution function of $Y$ given $X = x$ is

$$F_x(y) = F_\varepsilon(y - \theta_0 - \theta_1 x)$$

and the bias of the standard nonparametric estimator $\widehat{F}_x$ in (5.2) is for local linear weights proportional to

$$h_n^2\, \ddot{F}_x(y) = h_n^2\, F_\varepsilon''(y - \theta_0 - \theta_1 x)\, \theta_1^2,$$

where a dot indicates a derivative with respect to the covariate $x$, i.e. $\dot{F}_x(y) = \frac{\partial}{\partial x} F_x(y)$ and the symbol $'$ indicates a derivative with respect to $y$, i.e. $F_\varepsilon'(y) = \frac{\partial}{\partial y} F_\varepsilon(y)$. See Hall et al. (1999) for a study on methods of estimation of a conditional distribution function, and Van Keilegom and Veraverbeke (1997) for theoretical results on a conditional distribution function estimator in a fixed design (see the Appendix of that paper). See also Section 5.1.

As the variance of $\widehat{F}_x$ is proportional to $\frac{1}{n\,h_n}$, a large sample size is needed so that both variance and bias are sufficiently small and $\widehat{F}_x$ is a reasonable estimate of $F_x$. Of course, if one is sure that model (5.3) holds, then $F_x$ could be estimated simply by $\widehat{F}_\varepsilon\big(y - \widehat{\theta}_0 - \widehat{\theta}_1 x\big)$, where $\widehat{\theta}_0$, $\widehat{\theta}_1$ are consistent estimators of the parameters $\theta_0$, $\theta_1$ and $\widehat{F}_\varepsilon$ is the empirical distribution function of the residuals $\widehat{\varepsilon}_i = Y_i - \widehat{\theta}_0 - \widehat{\theta}_1 X_i$. Often it cannot be assumed that model (5.3) explains the effect of the covariate $X$ completely, but it is reasonable to assume that it captures an important part of this effect. This brings one to the idea of constructing an estimator of $F_x$ that is pre-adjusted by the linear model. This estimator is defined as

$$\widehat{F}_x^a(y) = \sum_{i=1}^{n} w_{ni}(x, h_n)\, \mathbb{I}\big\{\widehat{\varepsilon}_i \le y - \widehat{\theta}_0 - \widehat{\theta}_1 x\big\}.$$

As will be seen later this pre-adjusted nonparametric estimator of the conditional distribution function $F_x$ can work considerably better than the 'standard' nonparametric estimator $\widehat{F}_x$ given by (5.2), provided that the linear model (5.3) explains an important part of the effect of $X$ on $Y$.

The above considerations lead to the idea of pre-adjusting the responses for 'obvious' effects of the covariate. Pre-adjusted estimators can, in general, be constructed as follows. Let $G(x, y)$ be a function such that the distribution of the transformed observations $Y_i^a = G(X_i, Y_i)$ does not depend (or depends less) on $X_i$. Let $\widehat{G}_n$ be an estimate of this transformation. The pre-adjusted estimator, in general, is given by

$$\widehat{F}_x^a(y) = \sum_{i=1}^{n} w_{ni}(x, h_n)\, \mathbb{I}\big\{\widehat{G}_n(X_i, Y_i) \le \widehat{G}_n(x, y)\big\}. \tag{5.4}$$

In this paper we concentrate in particular on pre-adjusting by location and scale, inspired by the nonparametric model (5.1), that is $\widehat{G}_n(x, y) = \frac{y - \widehat{m}_n(x)}{\widehat{\sigma}_n(x)}$, where $\widehat{m}_n$ and $\widehat{\sigma}_n$ are for instance local linear estimators of the conditional mean $m$ and variance $\sigma^2$

$$m(x) = \mathsf{E}\,[Y \,|\, X = x], \qquad \sigma^2(x) = \mathsf{E}\,[Y^2 \,|\, X = x] - m^2(x). \tag{5.5}$$

We will show (both theoretically as well as via simulations) that the suggested estimator $\widehat{F}_x^a$ has better properties than the standard estimator $\widehat{F}_x$ given by (5.2) provided that the pre-adjusting can be supported by an underlying model. Of course other ways of standardizing the original observations are possible, e.g. using estimators for a median type of location quantity.

Note that the idea of pre-adjusting presented above is not entirely novel. Van Keilegom and Akritas (1999) used a nonparametric location-scale model to estimate $F_x$ in case of censoring, but their method assumes that model (5.1) describes the effect of $X$ on $Y$ completely. In our approach we only assume the existence of the first and second order conditional moments, without imposing a regression model. To the best of our knowledge the idea of simply using a standardization by nonparametric location and scale estimates has been little investigated in the statistical literature. In a working paper by Hansen (2004), the idea of pre-adjusting by means of a nonparametric location estimator (that is $\sigma(x) \equiv const.$) is used to estimate the conditional density function.

## 5.1 Estimating conditional distribution function by nonparametric pre-adjustment

In this section the suggested estimator is described in more detail and its asymptotic properties are discussed.

The key idea is to transform the observations $Y_i$ into $Y_i^a = G(X_i, Y_i)$ such that the distribution of the $Y_i^a$ depends considerably less on the $X_i$ than the original $Y_i$. Obviously there are various manners to try to achieve this goal, and for the ease of presentation we focus on the particular standardization

$$Y_i^a = \frac{Y_i - m(X_i)}{\sigma(X_i)}\ , \tag{5.6}$$

with $m(\cdot)$ and $\sigma^2(\cdot)$ the conditional mean and variance functions (see (5.5)), that are assumed to exist.

Let $\widehat{m}_n$ and $\widehat{\sigma}_n$ be estimates of the functions $m$ and $\sigma$ in (5.5) and define

$$\widehat{Y}_i^a = \frac{Y_i - \widehat{m}_n(X_i)}{\widehat{\sigma}_n(X_i)}, \qquad i = 1, \ldots, n. \tag{5.7}$$

This leads to the estimator

$$\widehat{F}_x^a(y) = \sum_{i=1}^n w_{ni}(x, h_n)\, \mathbb{I}\left\{ \frac{Y_i - \widehat{m}_n(X_i)}{\widehat{\sigma}_n(X_i)} \le \frac{y - \widehat{m}_n(x)}{\widehat{\sigma}_n(x)} \right\}. \tag{5.8}$$

For ease of presentation, we will concentrate on local linear regression estimates:

$$\widehat{m}_n(t) = \sum_{i=1}^n w_{ni}(t, g_{1n})\, Y_i, \qquad \widehat{\sigma}_n^2(t) = \sum_{i=1}^n w_{ni}(t, g_{2n})\, \big(Y_i - \widehat{m}_n(X_i)\big)^2, \tag{5.9}$$

where $g_{1n}$, $g_{2n}$ are bandwidth sequences going to zero.

The weights $w_{ni}(t, g_{1n})$, $w_{ni}(t, g_{2n})$ in (5.9), as well as the weights $w_{ni}(x, h_n)$ in (5.8), are local linear weights given by (3.9). Different kernel functions can be chosen in $w_{ni}(t, g_{1n})$, $w_{ni}(t, g_{2n})$ in (5.9) and in $w_{ni}(x, h_n)$ given in (3.9). Moreover other weighting schemes than the local linear weighting scheme can be considered.

Note that only existence of the first and second conditional moment functions ($m$ and $\sigma$) is required, as well as convergence in probability of the nonparametric estimates $\widehat{m}_n$ and $\widehat{\sigma}_n$ in a neighbourhood of the point of interest $x$ (see the assumptions below). So, the location-scale model (5.1) does not need to hold.

In the ideal situation that one knows the location and scale functions $m$ and $\sigma$, one would use the oracle estimator of $F_x$ based on the 'observations' $Y_i^a$ in (5.6):

$$\widetilde{F}_x^a(y) = \sum_{i=1}^{n} w_{ni}(x, h_n)\, \mathbb{I}\{Y_i^a \leq \tfrac{y-m(x)}{\sigma(x)}\}. \tag{5.10}$$

In other words, this is a kind of 'asymptotic version' of the estimator $\widehat{F}_x^a(y)$ that is based on (5.6) instead of on (5.7).

The main theoretical result in this paper (Theorem 1) tells us that the estimator $\widehat{F}_x^a(y)$ in (5.8) is close to the oracle estimator (5.10), in the sense that the supremum distance between the two estimators tends to zero fast. To prove this result we need to impose some regularity conditions.

A first set of conditions is

**(k')** The kernel $k$ is a continuously differentiable function with support $[-1, 1]$.

**(Bw')** $h_n = O(n^{-1/5})$, $\frac{h_n}{g_{1n}} \to 0$ and $\frac{h_n}{g_{2n}} \to 0$.

**(F1')** The density $f_X$ of $X$ is finite, positive and twice continuously differentiable in a neighbourhood of $x$.

**(F2')** There exists a neighbourhood $U_x$ of the point $x$ such that the conditional distribution function $F_t(y)$ viewed as a function of $(t, y)$ has uniformly continuous second order partial derivatives for $(t, y) \in U_x \times \mathbb{R}$. Furthermore,

$$\sup_{t\in U_x} \sup_{y} \left| y \dot{f}_t(y) \right| < \infty, \qquad \sup_{t\in U_x} \sup_{y} |y f_t(y)| < \infty, \qquad \sup_{t\in U_x} \sup_{y} \left| y^2 f_t'(y) \right| < \infty.$$

**($m\sigma$')** The functions $m$ and $\sigma$ are twice continuously differentiable in a neighbourhood $U_x$ of the point $x$ and $\inf_{t\in U_x} \sigma(t) > 0$.

*Remark.* Some common distributions whose support is not the whole real line $\mathbb{R}$ (e.g. uniform, exponential, ...) fail to meet assumption **(F2')** at the borders of their support. Let $A$ be a subset of $\mathbb{R}$ such that assumption **(F2')** is satisfied for $(t, y) \in U_x \times A$. Then Theorem 5.1 below still holds with $\sup_{y\in\mathbb{R}}$ being replaced with $\sup_{y\in A}$.

**Theorem 5.1.** *Assume that* **(k')**, **(Bw')**, **(F1')**, **(F2')** *and* **($m\sigma'$)** *are satisfied for the given point $x$. Then*

$$\sup_{y\in\mathbb{R}} \Big|\widehat{F}^a_x(y) - \widetilde{F}^a_x(y)\Big| = o_P\Big(\tfrac{1}{\sqrt{n\,h_n}}\Big). \tag{5.11}$$

Theorem 5.1 implies that the asymptotic properties of the estimator $\widehat{F}^a_x$ can be derived from the asymptotic properties of $\widetilde{F}^a_x$. Let $F^a_x$ stand for the conditional distribution function of $Y^a = \frac{Y-m(X)}{\sigma(X)}$ given $X = x$. With the help of Theorem 5.1 and similar calculations as in Fan and Gijbels (1996) (p. 66) one can derive the asymptotic representation for the pre-adjusted estimator in (5.8)

$$\widehat{F}^a_x(y) - F_x(y) = \sum_{i=1}^n w_{ni}(x, h_n)\, \Big[\mathbb{I}\Big\{Y^a_i \leq \tfrac{y-m(x)}{\sigma(x)}\Big\} - F^a_{X_i}\big(\tfrac{y-m(x)}{\sigma(x)}\big)\Big] + \tfrac{1}{2}\,h_n^2\,\mu_{2K}\ddot{F}^a_x\big(\tfrac{y-m(x)}{\sigma(x)}\big) + o_P\big(h_n^2\big) + o_P\Big(\tfrac{1}{\sqrt{n\,h_n}}\Big), \tag{5.12}$$

where $\mu_{2K} = \int u^2\,K(u)\,du$. Analogously, for the estimator in (5.2)

$$\widehat{F}_x(y) - F_x(y) = \sum_{i=1}^n w_{ni}(x, h_n)\, \big[\mathbb{I}\,\{Y_i \leq y\} - F_{X_i}(y)\big] + \tfrac{1}{2}\,h_n^2\,\mu_{2K}\ddot{F}_x(y) + o_P\big(h_n^2\big). \tag{5.13}$$

Now, using (5.12) and (5.13) one can derive the approximate asymptotic variances of the pre-adjusted and standard estimators

$$\mathsf{var}\,\big(\widehat{F}^a_x(y)\,|\,\mathbb{X}\big) = \tfrac{\|k\|}{n\,h_n f_X(x)}\,F^a_x\big(\tfrac{y-m(x)}{\sigma(x)}\big)\,\Big(1 - F^a_x\big(\tfrac{y-m(x)}{\sigma(x)}\big)\Big) + o_P\big(\tfrac{1}{n\,h_n}\big)\,, \tag{5.14}$$

$$\mathsf{var}\,\big(\widehat{F}_x(y)\,|\,\mathbb{X}\big) = \tfrac{\|k\|}{n\,h_n f_X(x)}\,F_x(y)\,(1 - F_x(y)) + o_P\big(\tfrac{1}{n\,h_n}\big)\,, \tag{5.15}$$

where $\|k\| = \int k^2(u)\,du$. Note that since

$$F^a_x(y) = \mathsf{P}\Big(\tfrac{Y-m(X)}{\sigma(X)} \leq y\,\Big|\,X = x\Big) = F_x\big(y\,\sigma(x) + m(x)\big),$$

one gets $F^a_x\big(\frac{y-m(x)}{\sigma(x)}\big) = F_x(y)$, which together with (5.14) and (5.15) implies that the approximate asymptotic variances of the standard estimator $\widehat{F}_x(y)$ and of the pre-adjusted estimator $\widehat{F}^a_x(y)$ are equal.

As pre-adjusting does not increase the asymptotic variance, the success of this method depends on the fact whether it reduces the bias. From representations (5.12) and (5.13) the approximate asymptotic bias is reduced if $\ddot{F}^a_x\big(\frac{y-m(x)}{\sigma(x)}\big)$ is closer to zero than $\ddot{F}_x(y)$. Note that if the location-scale model (5.1) holds, then $F^a_x(y) = \mathsf{P}(\varepsilon \leq y)$ which does not depend on $x$ anymore implying that $\ddot{F}^a_x(y) \equiv 0$. Thus, in case the nonparametric location-scale model holds, pre-adjusting is guaranteed to remove the leading term in the asymptotic bias.

What if the location-scale model (5.1) does not hold? Expressing $\ddot{F}^a_x\big(\frac{y-m(x)}{\sigma(x)}\big)$ in terms of the derivatives of $F_x$, $m(x)$ and $\sigma(x)$ yields

$$\ddot{F}^a_x\left(\tfrac{y-m(x)}{\sigma(x)}\right) = \ddot{F}_x(y) + 2\,\dot{f}_x(y)\,\left(\tfrac{(y-m(x))\,\sigma'(x)}{\sigma(x)} + m'(x)\right) + f'_x(y)\,\left(\tfrac{(y-m(x))\,\sigma'(x)}{\sigma(x)} + m'(x)\right)^2 + f_x(y)\,\left(\tfrac{(y-m(x))\,\sigma''(x)}{\sigma(x)} + m''(x)\right), \tag{5.16}$$

where $f_x = F_x'$ stands for the conditional density of $Y$ given $X = x$. The terms on the right-hand side of (5.16) should be compared to $\ddot{F}_x(y)$, the factor in the main term of the bias for the standard estimator (5.2). Since the terms on the right-hand side of (5.16) can be of opposite signs, no general comparison between the asymptotic bias of $\widehat{F}_x$ and $\widehat{F}_x^a$ is possible. From expression (5.16) it is seen that less or no advantage of a pre-adjusted estimator is to be expected when the conditional density $f_x(y)$ and/or its derivatives with respect to either $y$ or $x$ are high when compared to $\ddot{F}_x(y)$. See also the numerical study in Veraverbeke et al. (2014, Section 5).

*Remark.* If it is preferable to have a smoother estimator one can consider, in general, replacing the indicator in (5.4) by an integrated kernel function (a distribution function) $\mathbb{J}$. Together with pre-adjusting via nonparametric estimation of location and scale, the resulting estimator is then of the form

$$\widehat{F}_x^a(y) = \sum_{i=1}^n w_{ni}(x, h_n)\, \mathbb{J}_{b_n}\Big(\widehat{Y}_i^a - \tfrac{y - \widehat{m}_n(x)}{\widehat{\sigma}_n(x)}\Big), \quad \text{where} \quad \mathbb{J}_{b_n}(y) = \mathbb{J}(\tfrac{y}{b_n}),$$

with $b_n$ being a bandwidth sequence tending to zero. This double-kernel smoothing can be advantageous if one needs to plot the estimator or to use the estimator for quantile estimation (Yu and Jones, 1998).

## 5.2 Alternative methods of pre-adjusting

The nonparametric pre-adjusting method described in Section 2 is just one manner of pre-adjusting. There are many alternative methods that can be applied in the pre-adjusting step. In this section, we restrict to discuss some of these.

### 5.2.1 Alternative nonparametric pre-adjustments

Instead of pre-adjusting through nonparametric estimation of the conditional mean and variance as in (5.7), one can also perform a pre-adjustment by using other estimated quantities of conditional location and/or scale, such as using a conditional median, a conditional interquartile range, and so on.

Instead of aiming at reducing the influence of $X_i$ on the $Y_i$'s by standardizing the latter through a conditional mean and variance, an alternative road to go is to transform the $Y_i$'s as to make these observations more uniform-like (and thus not depending on the $X_i$). If the conditional distribution function of $Y$ given $X_i$, denoted by $F_{X_i}(y)$ would be given, then the transformed variable $F_{X_i}(Y_i)$ would have a uniform distribution. A method that follows this angle to tackle the problem, consists of estimating nonparametrically the function $F_{X_i}(y)$ and use this in the above described pre-adjustment step. Inspired by the work of Swanepoel and Van Graan (2005) a method of pre-adjusting would be to take $\widehat{Y}_i^a = \widehat{F}_{X_i g_n}(Y_i)$, where $\widehat{F}_{x g_n}(\cdot)$ is a pilot nonparametric estimator of the conditional distribution function given by (5.2) with $h_n$ replaced with a possible different bandwidth $g_n$. The resulting pre-adjusted estimator is then defined as

$$\widehat{F}_x^a(y) = \sum_{i=1}^n w_{ni}(x, h_n)\, \mathbb{I}\{\widehat{F}_{X_i g_n}(Y_i) \le \widehat{F}_{x g_n}(y)\}. \tag{5.17}$$

Note that the asymptotic version of the estimator is

$$\widetilde{F}_x^a(y) = \sum_{i=1}^n w_{ni}(x, h_n)\,\mathbb{I}\{F_{X_i}(Y_i) \le F_x(y)\}.$$

As $Y_i^a = F_{X_i}(Y_i)$ has a uniform distribution, $\ddot{F}_x^a(y) = 0$. Thus, this estimator aims at reducing the approximate asymptotic bias to the order of $o(h_n^2)$.

All previously discussed nonparametrically pre-adjusted estimators are simply examples of the key idea of transforming $Y_i$ into $G(X_i, Y_i)$, as to make the distribution of $Y_i$ to be less influenced by $X_i$. In the previously described methods we took $G(x, y) = (y - m(x))/\sigma(x)$, or $G(x, y) = F_x(y)$.

More generally, for an appropriate function $G(x, y)$, we would have the oracle estimator

$$\widetilde{F}_x^a(y) = \sum_{i=1}^n w_{ni}(x, h_n)\,\mathbb{I}\{G(X_i, Y_i) \le G(x, y)\}. \tag{5.18}$$

Replacing the unknown function $G$ by an estimator $\widehat{G}_n$ would then result into the general nonparametrically pre-adjusted estimator as in (5.4). In Section 5.2.3 we discuss the assumptions needed to guarantee result (5.11) for $\widehat{F}_x^a$ and $\widetilde{F}_x^a$ given by (5.4) and (5.18).

### 5.2.2 Alternative parametric pre-adjustments

In a general fashion, a parametric pre-adjustment would consist of transforming the observations $Y_i$ into $G(X_i, Y_i, \boldsymbol{\theta})$ where now $G(\cdot, \cdot, \boldsymbol{\theta})$ is a given parametric function, depending on a parameter vector $\boldsymbol{\theta}$. The unknown parameter vector is then replaced by an estimate $\widehat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}$, resulting into the general parametrically pre-adjusted $\widehat{Y}_i^a = G(X_i, Y_i, \widehat{\boldsymbol{\theta}}_n)$.

We next describe two specific examples of such parametrically pre-adjusted estimation approaches, that parallel the previously discussed nonparametric approaches. Consider parametric forms for the conditional mean and variance functions $m$ and $\sigma^2$ of Section 2, and denote them as $m(\cdot, \boldsymbol{\theta})$ and $\sigma^2(\cdot, \boldsymbol{\theta})$.

Let $\widehat{\boldsymbol{\theta}}_n$ be the estimator of the unknown parameter $\boldsymbol{\theta}$. The pre-adjusted observations $\widehat{Y}_i^a$ are now given by

$$\widehat{Y}_i^a = \frac{Y_i - m(X_i, \widehat{\boldsymbol{\theta}}_n)}{\sigma(X_i, \widehat{\boldsymbol{\theta}}_n)}, \qquad i = 1, \ldots, n.$$

*Remark.* A very simple pre-adjustment that can be often useful is given by:

$$\widehat{Y}_i^a = Y_i - \widehat{\theta}_0 - \widehat{\theta}_1 X_i - \widehat{\theta}_2 X_i^2 - \ldots - \widehat{\theta}_p X_i^p, \qquad i = 1, \ldots, n,$$

where $\widehat{\theta}_0, \widehat{\theta}_1, \ldots, \widehat{\theta}_p$ are the estimates of the unknown parameters in the conditional mean function $\mathsf{E}\,[Y|X] = \theta_0 + \theta_1 X + \ldots + \theta_p X^p$, obtained for instance by the least-squares method.

A second example is when the conditional distribution function of $Y$ given $X = x$, is of a parametric form, denoted by $F_x(y, \boldsymbol{\theta})$. The transformed observation $Y_i^a = F_{X_i}(Y_i, \boldsymbol{\theta})$ would then have a uniform distribution. Replacing the unknown parameter $\boldsymbol{\theta}$ by an estimate $\widehat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}$ leads to $\widehat{Y}_i^a = F_{X_i}(Y_i, \widehat{\boldsymbol{\theta}}_n)$ and the parametrically pre-adjusted estimator for $F_x$

$$\widehat{F}_x^a(y) = \sum_{i=1}^n w_{ni}(x, h_n)\,\mathbb{I}\{F_{X_i}(Y_i, \widehat{\boldsymbol{\theta}}_n) \le F_x(y, \widehat{\boldsymbol{\theta}}_n)\}. \tag{5.19}$$

Suppose that there exists a $\boldsymbol{\theta}$, such that $\widehat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}$ in probability. Then under appropriate regularity assumptions, asymptotic properties of the pre-adjusted estimator $\widehat{F}^a_x$ in (5.19) can be derived from the asymptotic properties of

$$\widetilde{F}^a_x(y) = \sum_{i=1}^{n} w_{ni}(x, h_n)\, \mathbb{I}\{F_{X_i}(Y_i, \boldsymbol{\theta}) \leq F_x(y, \boldsymbol{\theta})\}.$$

### 5.2.3 Further discussion

In this section we discuss whether for a more general transformation $Y_i^a = G(X_i, Y_i)$, the result of Theorem 5.1 continues to hold. Note that showing (5.11) would be quite straightforward provided that $\widehat{G}_n$ converges to $G$ at a faster rate than $(n\,h_n)^{-1/2}$ (see also Theorem 5.3 at the end of this section). But assuming such a fast rate of convergence of $\widehat{G}_n$ would either exclude nonparametric adjustment (that gives rate $O(n^{-2/5})$) or require that $h_n$ is of smaller order than $O(n^{-1/5})$ which is typically the optimal rate for $h_n$. The following technicalities are introduced in order to enable both the nonparametric adjustment as well as $h_n \sim n^{-1/5}$.

**(G1)** There exists a neighbourhood $U_x$ of the point $x$ and a sequence of deterministic functions $\{G_{(n)}(t, y)\}$ such that uniformly in $(t, y) \in U_x \times \mathbb{R}$

$$\widehat{G}_n(t, y) = G_{(n)}(t, y) + \big(1 + G_{(n)}(t, y)\big)\, o_p\big(\tfrac{1}{\sqrt{n\,h_n}}\big), \tag{5.20}$$

where $h_n$ is the bandwidth used in the construction of the weights $w_{ni}(x, h_n)$ in (5.4).

Note that assumption **(G1)** guarantees that there exists a sequence of deterministic functions $\{G_{(n)}(t, y)\}$ such that the difference $\widehat{G}_n(t, y) - G_{(n)}(t, y)$ converges to zero faster than $(n\,h_n)^{-1/2}$.

*Remark.* To illustrate assumption **(G1)**, consider $\widehat{G}_n(t, y) = \frac{y - \widehat{m}_n(t)}{\widehat{\sigma}_n(t)}$, where $\widehat{m}_n$ and $\widehat{\sigma}_n$ are given by (5.9).

Recall that $m(t) = \mathsf{E}\,[Y \mid X = t]$ and put

$$a_{nj}(t) = \int s^j k(s) f_X(t + s g_{1n})\, ds, \quad b_{nj}(t) = \int s^j k(s) f_X(t + s g_{1n})\, m(t + s g_{1n})\, ds. \tag{5.21}$$

Then for

$$m_{(n)}(t) = \frac{b_{n0}(t)\, a_{n2}(t) - b_{n1}(t)\, a_{n1}(t)}{a_{n0}(t)\, a_{n2}(t) - a_{n1}^2(t)}\,,$$

it holds that uniformly in $t \in U_x$:

$$\widehat{m}_n(t) - m_{(n)}(t) = O_P\big(\tfrac{1}{\sqrt{n\, g_{1n}}}\big). \tag{5.22}$$

Similarly (for details see Appendix A Veraverbeke et al., 2014), one can construct $\sigma_{(n)}(t)$ such that uniformly in $t \in U_x$:

$$\widehat{\sigma}_n(t) - \sigma_{(n)}(t) = O_P\big(\tfrac{1}{\sqrt{n\, g_{2n}}}\big). \tag{5.23}$$

Thus, by (5.22), (5.23) and assumption **(Bw')** one gets (5.20) with $G_{(n)}(t, y) = \frac{y - m_{(n)}(t)}{\sigma_{(n)}(t)}$.

*Remark.* If one considers a general nonparametric adjustment (5.17) with the transformation given by $\widehat{G}_n(t,y) = \widehat{F}_{tg_n}(y)$, then $G_{(n)}(t,y)$ can be taken as

$$G_{(n)}(t,y) = \frac{d_{n0}(t,y)\, a_{n2}(t) - d_{n1}(t,y)\, a_{n1}(t)}{a_{n0}(t)\, a_{n2}(t) - a_{n1}^2(t)},$$

with $a_{nj}(t)$ defined in (5.21) and $d_{nj}$ given by

$$d_{nj}(t,y) = \int s^j k(s) f_X(t + sg_n)\, F_{t+sg_n}(y)\, ds.$$

The next assumption guarantees that

$$\mathbb{I}\{G_{(n)}(x,Y) \leq G_{(n)}(x,y)\} = \mathbb{I}\{G(x,Y) \leq G(x,y)\} = \mathbb{I}\{Y \leq y\},$$

which is needed to recover the conditional distribution of $Y$ from the conditional distribution of the random variable $G_{(n)}(x,Y)$ or $G(x,Y)$.

**(G2)** Suppose that $G(t,y)$ as well as the functions $G_{(n)}(t,y)$ are for all sufficiently large $n$ increasing in $y$ for a fixed $t$. Let $H(t,u)$ ($H_{(n)}(t,u)$) be a function such that for a fixed $t$ the function $H(t,\cdot)$ ($H_{(n)}(t,\cdot)$) is an inverse function of $G(t,\cdot)$ ($G_{(n)}(t,\cdot)$). Suppose that for all sufficiently large $n$ it holds

$$H\big(x, G(x,y)\big) = y, \qquad H_{(n)}\big(x, G_{(n)}(x,y)\big) = y, \qquad \forall y \in \mathbb{R}. \tag{5.24}$$

The following assumptions are 'generic' replacements of the assumptions **(F2')** and ***(mσ')***. They guarantee that the functions $G_{(n)}(t,y)$ are sufficiently smooth and sufficiently close to the function $G(t,y)$. They can be either written in terms of the distribution of the original variables $Y$ or in terms of the transformed variables $G_{(n)}(x,Y)$. For brevity we formulate these assumptions only in terms of the transformed variables.

Let $F_{nt}^a(y)$ and $F_t^a(y)$ be the conditional distribution functions of the random variables $G_{(n)}(X_1,Y_1)$ and $G(X_1,Y_1)$ given $X_1 = t$. The corresponding densities are denoted as $f_{nt}^a(y)$ and $f_t^a(y)$.

**(F2g)** Suppose that

$$\lim_{n\to\infty} \sup_{(t,y)\in U_x\times\mathbb{R}} \big|F_{nt}^a(y) - F_t^a(y)\big| = o(1)$$

and

$$\lim_{n\to\infty} \sup_{(t,y)\in U_x\times\mathbb{R}} \big|y\, f_{nt}^a(y)\big| < \infty, \qquad \sup_{(t,y)\in U_x\times\mathbb{R}} \big|y\, f_t^a(y)\big| < \infty.$$

**(F3g)** There exists a neighbourhood $U_x$ of the point $x$ such that the function $\ddot{F}_t^a(G(x,y))$ exists and is uniformly continuous for $(t,y) \in U_x \times \mathbb{R}$.

**(F4g)** $\sup_{y\in\mathbb{R}} \left|\dot{F}_{nx}^a(y)\right| < \infty$ and

$$\sup_{(t,y)\in U_x\times\mathbb{R}} \left|\ddot{F}_{nt}^a\big(G_{(n)}(x,y)\big) - \ddot{F}_t^a\big(G(x,y)\big)\right| \to 0, \quad \text{as } n \to \infty.$$

**Theorem 5.2.** *Assume that* **(k')**, **(G1)**, **(G2)**, **(F1')**, **(F2g)**, **(F3g)** *and* **(F4g)** *are satisfied for the given point $x$ and $h_n = O(n^{-1/5})$. Then the result* (5.11) *of Theorem 5.1 holds, where $\widehat{F}_x^a$ is given by* (5.4) *and $\widetilde{F}_x^a$ by* (5.18).

The theorem is proved in Veraverbeke et al. (2014, Appendix B).

*Remark.* Note that the assumption of monotonicity of the function $G(t, y)$ in the second argument is not satisfied if one considers $G(t, y) = F_t(y)$ and the conditional distribution function $F_t(y)$ is not strictly increasing. Although this is distracting, it is only a technical problem. Generally, it seems natural to consider only the transformations $\widehat{G}_n(t, y)$ that are for fixed $t$ nondecreasing in $y$. Thus one can consider $G_{(n)}(t, y)$, $G(t, y)$ that are nondecreasing in $y$. Let $H_{(n)}(t, \cdot)$ and $H(t, \cdot)$ be a pseudo-inverse of $G_{(n)}(t, \cdot)$ and $G(t, \cdot)$ respectively. Further, let the assumption (5.24) hold in the points $y$ where $G(x, \cdot)$ is strictly increasing. Note that then it holds

$$F_x\big(H(x, G(x, y))\big) = F_x\big(H_{(n)}(x, G_{(n)}(x, y))\big) = F_x(y), \qquad y \in \mathbb{R},$$

and the proof of Theorem 5.2 works also for the situation when $G(t, y)$ is nondecreasing (but not strictly increasing) in $y$.

It is interesting to point out that the proof of Theorem 5.2 relies on a sufficiently fast convergence of $\widehat{G}_n$ to $G_{(n)}$ (see assumption **(G1)**), but it does not assume anything about the speed of convergence of $G_{(n)}$ to $G$. On the other hand smoothness assumptions formulated in **(F2g)**, **(F3g)** and **(F4g)** are required. Alternatively, one can considerably weaken the smoothness assumption about $F_t^a$ provided that already $\widehat{G}_n(x, y)$ converges to $G(x, y)$ at 'a sufficiently fast rate'. This is formalized by the following assumptions that are used in the subsequent theorem.

**(G1')** There exists a neighbourhood $U_x$ of the point $x$ and a function $G(t, y)$ such that uniformly in $(t, y) \in U_x \times \mathbb{R}$

$$\widehat{G}_n(t, y) = G(t, y) + \big(1 + G(t, y)\big)\, o_p\big(\tfrac{1}{\sqrt{n\, h_n}}\big), \tag{5.25}$$

where $h_n$ is the bandwidth used in the construction of the weights $w_{ni}(x, h_n)$ in (5.4).

**(G2')** Suppose that $G(t, y)$ are for all sufficiently large $n$ increasing in $y$ for a fixed $t$. Let $H(t, u)$ be a function such that for a fixed $t$ the function $H(t, \cdot)$ is an inverse function of $G(t, \cdot)$ Suppose that it holds

$$H\big(x, G(x, y)\big) = y, \qquad \forall y \in \mathbb{R}.$$

**(F2g')** It holds

$$\sup_{(t,y)\in U_x\times\mathbb{R}} \big|y\, f_t^a(y)\big| < \infty.$$

Note that assumption **(G1')** is useful for parametric pre-adjustment where the remainder term in (5.25) is typically of order $O_P(n^{-1/2})$.

The following theorem can be proved by mimicking Step 1 of the proof of Theorem 5.2 and replacing $G_{(n)}$, $H_{(n)}$ and $Y_{ni}^a$ with $G$, $H$ and $Y_i^a$.

**Theorem 5.3.** *Assume that* **(k')**, **(G1')**, **(G2')**, **(F1')** *and* **(F2g')** *are satisfied for the given point $x$ and $h_n = O(n^{-1/5})$. Then the result* (5.11) *of Theorem 5.1 holds.*

## 5.3 Discussion and further extensions

The idea of pre-adjusting can be applied in a straightforward way to the estimators of a conditional distribution function discussed in Hall et al. (1999). Pre-adjusting can also be applied together with the methods presented in Dette and Volgushev (2008) to have a monotone estimator of $F_x(y)$ when local linear weights are used.

As discussed in Veraverbeke et al. (2014, Section 4) the idea of pre-adjusting can be helpful in other estimation settings as nonparametric quantile estimation and conditional density estimation. Further, Veraverbeke et al. (2014) illustrate that pre-adjusting can be used also with censored data.

# Bibliography


Acar, E. F., Genest, C., and Nešlehová, J. (2012). Beyond simplified pair-copula constructions. *Journal of Multivariate Analysis*, **110**, 74–90. doi: 10.1016/j.jmva.2012.02.001.

Acar, E. F., Craiu, R. V., and Yao, F. (2013). Statistical testing of covariate effects in conditional copula models. *Electronic Journal of Statistics*, **7**, 2822–2850. doi: 10.1214/13-EJS866.

Aerts, M., Janssen, P., and Veraverbeke, N. (1994). Bootstrapping regression quantiles. *Journal of Nonparametric Statistics*, **4**, 1–20. doi: 10.1080/10485259408832597.

Brockmann, M., Gasser, T., and Hermann, E. (1993). Locally Adaptive Bandwidth Choice for Kernel Regression Estimators. *Journal of the American Statistical Association*, **88**, 1302–1309. doi: 10.1080/01621459.1993.10476411.

Bücher, A. and Volgushev, S. (2013). Empirical and sequential empirical copula processes under serial dependence. *Journal of Multivariate Analysis*, **119**, 61–70. doi: 10.1016/j.jmva.2013.04.003.

Bücher, A., Segers, J., and Volgushev, S. (2014). When uniform weak convergence fails: Empirical processes for dependence functions and residuals via epi-and hypographs. *Annals of Statistics*, **42**(4), 1598–1634. doi: 10.1214/14-AOS1237.

Chen, S. X. and Huang, T.-M. (2007). Nonparametric estimation of copula functions for dependence modelling. *Canadian Journal of Statistics*, **35**, 265–282. doi: 10.1002/cjs.5550350205.

Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés. *Acad. Roy. Belg. Bull. Cl. Sci.*, **65**, 274–292.

Dette, H. and Volgushev, S. (2008). Non-crossing non-parametric estimates of quantile curves. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, **70**(3), 609–627. doi: 10.1111/j.1467-9868.2008.00651.x.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman & Hall/CRC, London.

Fermanian, J.-D., Radulovič, D., and Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli*, **10**, 847–860. doi: 10.3150/bj/1099579158.

Gänssler, P. and Stute, W. (1987). *Seminar on Empirical Processes*. Birkhäuser, Basel.

Gasser, T. and Müller, H.-G. (1979). Kernel estimates of regression functions. In Gasser, T. and Rosenblatt, M., editors, *Lecture Notes in Mathematics 757*, pages 23–68. Springer, New York.

Gasser, T., Kneip, A., and Kohler, W. (1991). A Flexible and Fast Method for Automatic Smoothing. *Journal of the American Statistical Association*, **86**, 643–652. doi: 10.2307/2290393.

Gijbels, I., Omelka, M., and Veraverbeke, N. (2012). Multivariate and functional covariates and conditional copulas. *Electronic Journal of Statistics*, **6**, 1273–1306. doi: 10.1214/12-EJS712.

Gijbels, I. and Mielniczuk, J. (1990). Estimating the density of a copula function. *Communication in Statistics - Theory and Methods*, **19**, 445–464. doi: 10.1080/03610929008830212.

Gijbels, I., Veraverbeke, N., and Omelka, M. (2011). Conditional copulas, association measures and their application. *Computational Statistics & Data Analysis*, **55**, 1919–1932. doi: 10.1016/j.csda.2010.11.010.

Gijbels, I., Omelka, M., and Veraverbeke, N. (2015a). Partial and average copulas and association measures. *Electronic Journal of Statistics*, **9**(2), 2420–2474. doi: 10.1214/15-EJS1077.

Gijbels, I., Omelka, M., and Veraverbeke, N. (2015b). Estimation of a copula when a covariate affects only marginal distributions. *Scandinavian Journal of Statistics*, **42**(4), 1109–1126. doi: 10.1111/sjos.12154.

Gijbels, I., Omelka, M., Pešta, M., and Veraverbeke, N. (2017a). Score tests for covariate effects in conditional copulas. Submitted to JMVA. Under revision.

Gijbels, I., Omelka, M., and Veraverbeke, N. (2017b). Nonparametric testing for no covariate effects in conditional copulas. *Statistics*, **51**, 475–509. doi: 10.1080/02331888.2016.1258070.

Hall, P., Wolff, R. C. L., and Yao, Q. (1999). Methods for Estimating a Conditional Distribution Function. *Journal of the American Statistical Association*, **94**, 154–163. doi: 10.1080/01621459.1999.10473832.

Hansen, B. E. (2004). Nonparametric conditional density estimation. Technical report, University of Wisconsin.

Hobæk Haff, I., Aas, K., and Frigessi, A. (2010). On the simplified pair-copula construction – Simply useful or too simplistic? *Journal of Multivariate Analysis*, **101**(5), 1296–1310. doi: 10.1016/j.jmva.2009.12.001.

Müller, H.-G. (1987). Weighted local regression and kernel methods for nonparametric curve fitting. *Journal of the American Statistical Association*, **82**, 231–238. doi: 10.1080/01621459.1987.10478425.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and Its Application*, **9**, 141–142. doi: 10.1137/1109020.

Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer, New York. Second Edition.

Neumeyer, N., Omelka, M., and Hudecová, Š. (2017). A copula approach for dependence modeling in multivariate nonparametric time series. In preparation.

Omelka, M., Gijbels, I., and Veraverbeke, N. (2009). Improved kernel estimation of copulas: Weak convergence and goodness-of-fit testing. *Annals of Statistics*, **37**(5B), 3023–3058. ISSN 0090-5364. doi: 10.1214/08-AOS666.

Omelka, M., Veraverbeke, N., and Gijbels, I. (2013). Bootstrapping the conditional copula. *Journal of Statistical Planning and Inference*, **143**, 1–23. doi: 10.1016/j.jspi.2012.06.001.

Patton, J. A. (2006). Modeling asymmetric exchange rate dependence. *International Economic Review*, **47**(2), 527–556. doi: 10.1111/j.1468-2354.2006.00387.x.

Portier, F. and Segers, J. (2015). On the weak convergence of the empirical conditional copula under a simplifying assumption. *arXiv preprint arXiv:1511.06544*.

Reiss, R.-D. (1981). Nonparametric estimation of smooth distribution functions. *Scandinavian Journal of Statistics*, **8**, 116–119.

Segers, J. (2012). Weak convergence of empirical copula processes under nonrestrictive smoothness assumptions. *Bernoulli*, **18**, 764–782. doi: 10.3150/11-BEJ387.

Stöber, J., Joe, H., and Czado, C. (2013). Simplified pair copula constructions – Limitations and extensions. *Journal of Multivariate Analysis*, **119**, 101–118. doi: 10.1016/j.jmva.2013.04.014.

Swanepoel, J. W. H. and Van Graan, F. C. (2005). A New Kernel Distribution Function Estimator Based on a Non-parametric Transformation of the Data. *Scandinavian Journal of Statistics*, **32**(4), 551–562. doi: 10.1111/j.1467-9469.2005.00472.x.

Tsukahara, H. (2005). Semiparametric estimation in copula models. *Canadian Journal of Statistics*, **33**, 357–375. doi: 10.1002/cjs.5540330304.

Van Keilegom, I. and Akritas, M. G. (1999). Transfer of tail information in censored regression models. *Annals of Statistics*, **27**, 1745–1784. doi: 10.1214/aos/1017939150.

Van Keilegom, I. and Veraverbeke, N. (1997). Estimation and Bootstrap with Censored Data in Fixed Design Nonparametric Regression. *Annals of Institute of Statistical Mathematics*, **49**, 467–491. doi: 10.1023/A:1003166728321.

Veraverbeke, N., Omelka, M., and Gijbels, I. (2011). Estimation of a conditional copula and association measures. *Scandinavian Journal of Statistics*, **38**, 766–780. doi: 10.1111/j.1467-9469.2011.00744.x.

Veraverbeke, N., Gijbels, I., and Omelka, M. (2014). Pre-adjusted nonparametric estimation of a conditional distribution function. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, **76**, 399–438. doi: 10.1111/rssb.12041.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā A*, **26**, 359–372.

Yu, K. and Jones, M. C. (1998). Local linear quantile regression. *Journal of the American Statistical Association*, **93**, 228–237. doi: 10.1080/01621459.1998.10474104.

# Appendix: Attached papers

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. doi: 10.1214/08-AOS666.

GIJBELS, I., VERAVERBEKE, N., and OMELKA, M. (2011). Conditional copulas, association measures and their application. *Computational Statistics & Data Analysis*, **55**, 1919–1932. doi: 10.1016/j.csda.2010.11.010.

VERAVERBEKE, N., OMELKA, M., and GIJBELS, I. (2011). Estimation of a conditional copula and association measures. *Scandinavian Journal of Statistics*, **38**, 766–780. doi: 10.1111/j.1467-9469.2011.00744.x.

GIJBELS, I., OMELKA, M., and VERAVERBEKE, N. (2015). Estimation of a copula when a covariate affects only marginal distributions. *Scandinavian Journal of Statistics*, **42**, 1109–1126. doi: 10.1111/sjos.12154.

VERAVERBEKE, N., GIJBELS, I., and OMELKA, M. (2014). Pre-adjusted nonparametric estimation of a conditional distribution function. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, **76**, 399–438. doi: 10.1111/rssb.12041.