Model selection and goodness of fit tests for conditional copula models

Author:
Jacco Wielaard

Advisors:
Dr. A.F.F Derumigny
Prof. Dr. A.J Schmidt-Hieber

March 2021

Abstract

Conditional copulas, also known as copula models with covariates, are models that describe the dependence between several random variables of interest, conditionally to some known explanatory variables. It is often assumed that these conditional copulas belong to a given parametric family, with a (conditional) parameter depending on the explanatory variables. We propose several goodness-of-fit tests for the assumption of good specification of a parametric conditional copula model, without any constraint on the conditional margins. Two such tests that use different bootstrap resampling procedures are compared in a simulation study. Finally, these tests are applied to a dataset of financial returns.

Contents

1 Introduction

A copula is a distribution on the unit square with uniform margins. Copulas were designed in 1959 by Abe Sklar [1]. By using a copula to model the bivariate dependence, Sklar was able to join together two one-dimensional distribution functions to form a multivariate distribution function. He derived the term copula from the Latin noun which means “a link, tie, bond”. [2]

Copulas are popular in multivariate statistical applications as they allow one to easily model and estimate the distribution of random vectors by estimating marginals and copulas separately. These multivariate distributions are otherwise notoriously difficult to model and estimate.

Copulas were introduced to the field of financial risk management in 1999 [3]. Copulas are often applied to portfolio management, derivatives pricing and risk management. Copulas are also used to perform stress-tests and robustness checks that are especially important during crisis times where extreme events may occur (e.g., the financial crisis of 2007). In these stress-tests and robustness checks, copula models are used to estimate the probability distribution of losses on pools of loans or bonds. The incorrect usage of copula models is seen as one of the reasons for the financial crisis of 2007, see “The formula that killed wall street” [4] (also [5] for more discussion on this subject).

More recent is the use of copulas in climate models, see for example Nguyen et al. [6]. They used copula models to model the climate drivers to in turn measure their impact on wheat yield in Australia. Other examples of uses are in hydrology and biometric science. [7, 8]

Parametric models are the most standard choice for copulas, but it is often difficult to decide whether a given parametric model is appropriate or not. Therefore, lot of research has been done about goodness-of-fit tests for these copula models, see for example [9, 10, 11, 12, 13, 14]. The research about goodness-of-fit testing will be extended to conditional copulas in this thesis. Conditional copulas are an extension of copulas, where they not only couple a multivariate distribution to their one-dimensional marginal distribution functions, but do so in relation to other, explanatory variables. These explanatory variables are used to model the dependency between the explained variables.

1.1 Goal and Outline of the report

The main goal of this thesis is to research, implement and test goodness-of-fit (GOF) tests for conditional copula models. In Section 2, a short explanation of dependence measures, copulas, conditional copulas, bootstrapping and goodness of fit tests will be given. In Section 3 the mathematical theory behind copulas will be extended to conditional copulas, together with the proposed way to perform these GOF-tests. In Section 4 the methods used to make the computations discussed in Section 3 feasible will be explained, as well as the methods used to compute the results. The results themselves will be visible in Section 5. This includes the results of the simulations, as well as the application to real world stock data. In Section 6 the results will be discussed.

2 Basics

This section will give a brief description and explanation of some of the mathematics used in this thesis. In Section 2.1 the use of Kendall's Tau will be discussed. In Section 2.2 the basics of copulas, including parametric copulas will be explained. In Section 2.3 the theory of copulas will be expanded to conditional copulas. In Section 2.4 the nonparametric bootstrap procedure will be discussed. Finally in Section 2.5 goodness of fit testing will be treated.

2.1 Kendall's tau

Kendall's tau is a measure of dependence, just like Spearman's rho and the more well-known Pearson correlation. Where Pearson's correlation coefficient measures the linear dependence between two variables, both Kendall's tau and Spearman's rho measure the rank correlation. Since there is a need to linearly transform variables, and Pearson correlation is not invariant under linear transformation, using the Pearson correlation is not an option. Kendall's tau is used more often than Spearman's rho in the field of conditional copulas, which is the reason it is used in this thesis as well.

Kendall's tau is named after Maurice Kendall, who discussed the measure in 1938 [15]. Kendall's tau is a measure of rank correlation. A rank correlation measures the extent to which, as one variable increases, the other variable tends to increase, without requiring that increase to be represented by a linear relationship. Kendall's tau is Definition 1.

Definition 1 (Kendall's Tau)**.**
Let (X_1, Y_1) and (X_2, Y_2) be independent random vectors with the same distribution as (X, Y). Then Kendall's tau is de ned as

$$\begin{aligned} \tau_{1,2} &:= \mathbb{P}(\textit{concordant pair}) - \mathbb{P}(\textit{discordant pair}) \\ &= \mathbb{P}\big((X_1 - X_2)(Y_1 - Y_2) > 0\big) - \mathbb{P}\big((X_1 - X_2)(Y_1 - Y_2) < 0\big). \end{aligned}$$

A pair of points (x_1, y_1) and (x_2, y_2) is a concordant pair when both $x_1 < x_2$ and $y_1 < y_2$, or conversely, $x_1 > x_2$ and $y_1 > y_2$. A discordant pair is where neither of these is the case, that is $x_1 < x_2$ and $y_1 > y_2$ or $x_1 > x_2$ and $y_1 < y_2$. See Figure 1 for reference on concordant and discordant pairs.

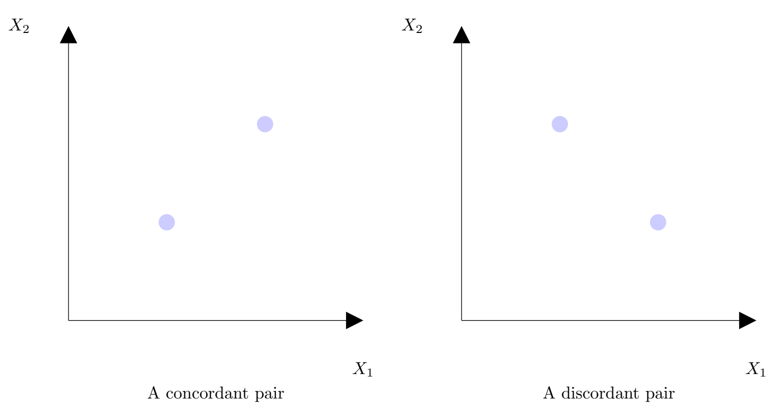


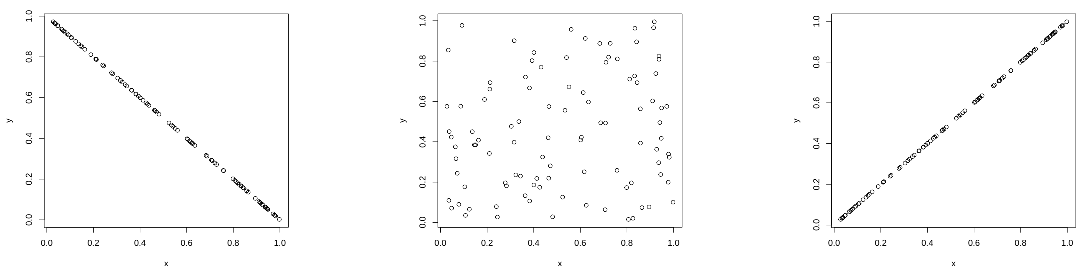
Figure 1: *Comparison between concordant and discordant pairs.*

Kendall's tau is usually estimated in the following way,

$$\hat{\tau} = \frac{(\text{Number of concordant pairs}) - (\text{Number of discordant pairs})}{\text{Number of pairs}}. \tag{1}$$

Where the number of pairs can easily be calculated by using the binomial coefficient of n and 2.

By design, Kendall's tau is always between −1 and 1, since both the number of concordant pairs and discordant pairs can never be larger than the number of pairs. A τ of −1 means perfect negative dependence, a τ of 0 means no rank correlation and a τ of 1 means perfect positive dependence. An illustration of these is also visible in Figure 2.



(a) *Complete negative dependence* **(b)** *Complete independence* **(c)** *Complete positive dependence*

Figure 2: *Dependence with Kendall's Tau*

2.2 Copulas

A copula is a distribution on the unit square with uniform margins. Copulas were invented in 1959 by Abe Sklar [1]. Sklar joined together two one-dimensional distribution functions to form a multivariate distribution function. His theorem, Theorem 2.2, is still the foundation for almost all copula related research. This theorem states that any multivariate joint distribution can be written in terms of univariate marginal distribution functions and a copula which describes the dependence structure between the variables.

Theorem 2.1 (Sklar, 1959)**.**
Let $F_{1,2}$ be a distribution function on R^2 with continuous margins F_1 and F_2. Then there exists a distribution C on $[0,1]^2$ with uniform margins, named the copula of X_1 and X_2 such that

$$\forall x_1, x_2 \in \mathsf{R}, \quad F_{1,2}(x_1, x_2) = C\big(F_1(x_1), F_2(x_2)\big),$$

and C is given by

$$\forall u_1, u_2 \in [0,1], \quad C(u_1, u_2) = F_{1,2}\big(F_1^{-1}(u_1), F_2^{-1}(u_2)\big).$$

where F_i^{-1} is the inverse of F_i, for $i = 1, 2$.

This also means that if we have two margins F_1 and F_2 and a copula C, it is possible to construct the joint distribution $F_{1,2}$. If instead we only have the joint distribution, it is possible to reclaim the marginals from this by integration in the following way,

$$F_1(x_1) = \int_{x_2=-\infty}^{x_2=\infty} F_{1,2}(x_1, x_2)dx_2,$$
$$F_2(x_2) = \int_{x_1=-\infty}^{x_1=\infty} F_{1,2}(x_1, x_2)dx_1.$$

All of the needed information to construct the copula is thus hidden in the joint distribution $F_{1,2}$. This makes us able to construct the copula from just the joint distribution using the second part of Sklar's theorem.

2.2.1 Parametric copulas

It is also possible to construct a copula based on the family it belongs to and a respective parameter. This means it is not necessary to construct the margins explicitly. The margins are implicitly obtained from the joint distribution of the corresponding family. An example of this is a copula belonging to the Gaussian family, also known as a Gaussian copula.

Definition 2 (Gaussian Copula).
Let Φ be the joint cdf of $\mathcal{N}(0, \Sigma_\theta)$, where

$$\Sigma_\theta := \begin{pmatrix} 1 & \theta \\ \theta & 1 \end{pmatrix}.$$

Let ϕ then be its marginal cdf. The Gaussian copula of parameter $\theta \in [-1, 1]$ is then defined as

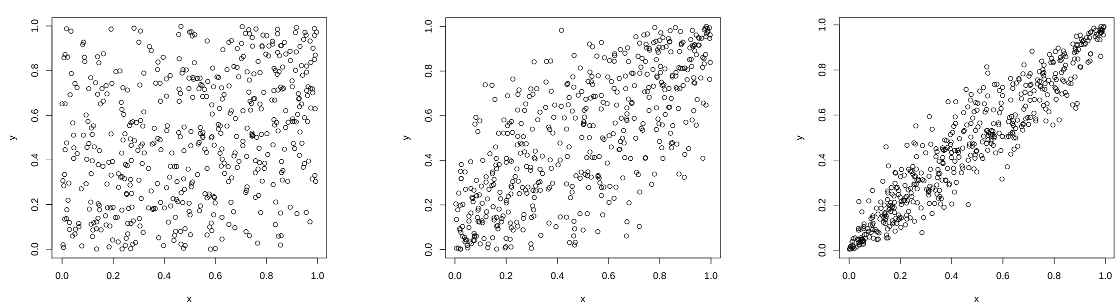
$$C_\theta^{Gaussian}(u_1, u_2) := \Phi_\theta(\phi^{-1}(u_1), \phi^{-1}(u_2)).$$

This means it is possible to construct a Gaussian copula based on only one parameter. The Kendall's tau of the Gaussian copula has a direct link with its parameter and therefore we reparametrize the copulas by their Kendall's tau equivalent. To use Kendall's tau as parameter for the Gaussian copula it is needed to transform it according to the following formula described by Hofert et al. [16, page 88].

$$\tau = \left(\frac{2}{\pi}\right) \arcsin\left(\frac{\theta}{2}\right) \tag{2}$$

The higher the absolute value of τ, the higher the dependence of the data. In Figure 3 different Gaussian copulas are simulated with different values of τ. In Figure 3a the dependence is barely noticeable. In Figure 3b the dependence is stronger, and almost no points exist in the bottom-right and top-left corners. When $\tau = 0.8$, in Figure 3c the dependence is already quite high, with all the points being close to the diagonal $y = x$.

In these figures the change of the shape of the copulas as τ increases is clearly visible. When τ is 0, there is no dependence and the independence copula is obtained, see also Figure 2b. When τ is 1, there is complete dependence as is visible in Figure 2c. For negative values of τ, the dependence between x and y would of course be reversed from what is visible in these figures.



(a) $\tau = 0.2$ **(b)** $\tau = 0.5$ **(c)** $\tau = 0.8$

Figure 3: *Comparison of how different values of τ influence the dependence of random variables following the Gaussian copula.*

Copulas that are described with the help of a parameter are also known as parametric copulas. If $\Theta = [-1, 1]$ is the space of possible parameters, the set of parametric copulas belonging to the Gaussian family can be described by $\{C_\theta^{Gaussian}(\mathbf{u}), \theta \in \Theta\}$. When not describing a specific family, this is denoted by $C_\theta(\mathbf{u})$.

Other commonly used copula families, besides the Gaussian family, are the Student, the Clayton and the Gumbel families. Each of these families has a distinct shape and will be discussed shortly in the following section.

First, the Student family. The Student family is a two parameter family, meaning it has two parameters that together describe the copula. It is defined in Definition 3. To reduce the complexity for this thesis however, only the Student copula with $\nu = 4$ degrees of freedom will be considered, which reduces the Student family to a one-parameter family.

Definition 3 (Student Copula)**.**
Let $(Y_1, Y_2) \sim \mathcal{N}(0, \Sigma_\theta)$, and let $\xi \sim \chi^2_\nu$ independently, then $\left(\frac{Y_1}{\sqrt{\xi/\nu}}, \frac{Y_2}{\sqrt{\xi/\nu}}\right)$ follows a multivariate Student distribution t_{ν,Σ_θ}. Its marginal distributions are (univariate) Student t_ν with ν degrees of freedom. The Student copula with correlation parameter $\theta \in [-1, 1]$ and $\nu > 2$ degrees of freedom is then

$$C^{Student}_{\theta,\nu}(u_1, u_2) := t_{\nu,\Sigma_\theta}\big(t^{-1}_\nu(u_1), t^{-1}_\nu(u_2)\big).$$

To reparametrize the Student family to use the Kendall's tau as parameter, the same transformation is applied as for the Gaussian copula, Equation 2.

Next the Clayton and the Gumbel copula. The Clayton and the Gumbel copula are both Archimedean copulas. Archimedean copulas can be described with the help of a generator function, g. The definition for these Archimedean copulas is Definition 4.

Definition 4 (Archimedean copulas)**.**
Let g be a continuous, strictly decreasing and convex function $[0, 1] \to [0, +\infty]$ such that $g(1) = 0$. Then

$$C^{Archimedean}_g(u_1, u_2) := g^{-1}\big(g(u_1) + g(u_2)\big)$$

is the Archimedean copula with generator g.

Equivalently, $C^{Archimedean}_g$ is the copula C such that

$$g\big(C(u_1, u_2)\big) = g(u_1) + g(u_2).$$

The Clayton copula and the Gumbel copula use different generator functions. The Clayton copula uses the generator function

$$g = (1 + t)^{-1/\theta}, \quad 0 < \theta < +\infty.$$

Which means that Clayton copula can be written explicitly as

$$C^{Clayton}_\theta(u_1, u_2) := \max\big(u_1^{-\theta} + u_2^{-\theta} - 1\,;\,0\big)^{-1/\theta}.$$

To reparametrize the Clayton copula to use Kendall's tau as parameter, the following transformation is needed,

$$\tau = \frac{\theta}{\theta + 2}.$$

The Gumbel copula uses the generator function

$$\big(-\log t\big)^\theta, \quad 1 < \theta < +\infty.$$

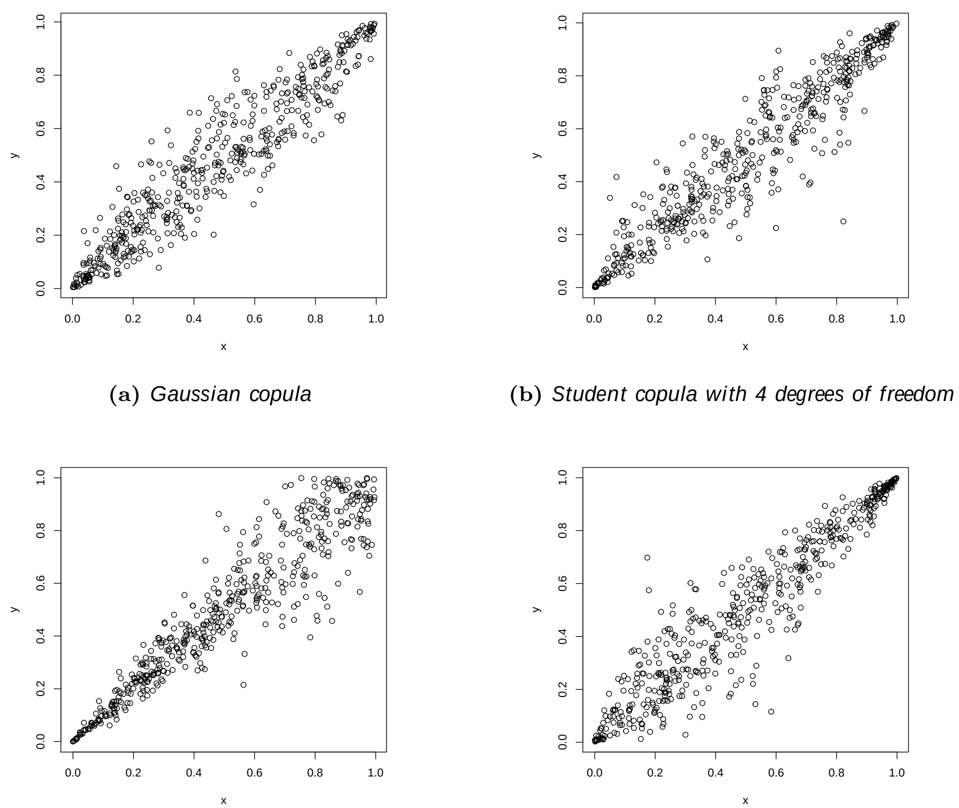
Which means that the Gumbel copula can be written explicitly as

$$C^{Gumbel}_\theta(u_1, u_2) := \exp\left[-\Big(\big(-\log u_1\big)^\theta + \big(-\log u_2\big)^\theta\Big)^{1/\theta}\right].$$

To use Kendall's tau as parameter for the Gumbel copula, the following transformation for τ is necessary,

$$\tau = \frac{\theta - 1}{\theta}.$$

The Gaussian copula, the Student copula with 4 degrees of freedom, the Clayton copula and the Gumbel copula are the main copulas that will be used in this thesis. In Figure 4 a sample of $N = 500$ points from each of these four copula families all simulated with a τ of 0.8 is displayed. This is done to give a general sense of the shape of each of these copulas. As is visible from these figures, the Gaussian and the Student copula look quite similar, although the Student copula has more outliers towards the top-left and bottom-right corner, which is characteristic for this family. The Clayton copula has a higher lower tail dependence than upper tail dependence, which means that the dependence for low values of x and y is higher than for high values of x and y. The Gumbel copula on the other hand has a higher upper tail dependence than lower tail dependence, which means that the dependence is higher for large values of x and y than for small values of x and y. Both the Clayton and the Gumbel copula are asymmetric in that regard.



(a) *Gaussian copula*

(b) *Student copula with 4 degrees of freedom*

(c) *Clayton copula*

(d) *Gumbel copula*

Figure 4: *Some samples of commonly used copula families displayed with a* τ *of* 0.8

2.3 Conditional Copulas

The previously described copulas are useful for modelling and certain financial calculations. If however there is more data available, which often is the case, it might be possible to make these models and calculations more accurate. A use for the extra data could be to use them as explanatory variables for the dependence. It would be interesting to see how a third variable $\mathbf{X}$ influences the dependence between Y_1 and Y_2. These types of models are called conditional copulas, see Figure 5 for a reference.

Let us use an often used example, see Gijbels et al. [17] and Veraverbeke et al. [18]. Suppose we have data on life expectancies at birth in 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 between (Y_1, Y_2) conditionally upon the given value of the covariate $X = x$ and whether this relationship changes with the values of x. This can be fully described by a function, which is the conditional copula.

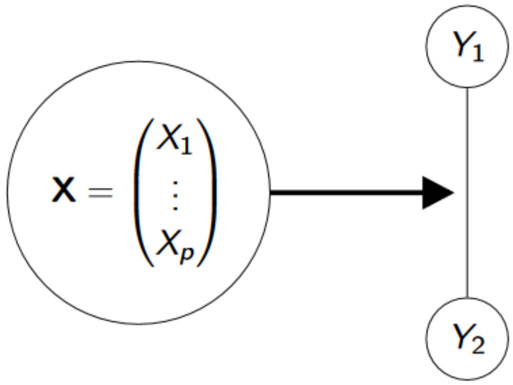


Figure 5: *The in uence of $\mathbf{X}$ on the conditional dependence between Y_1 and Y_2: the conditional copula $C_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(Y)$.*

We will now extend the theory of copulas to these conditional copulas. For $\mathbf{x} \in \mathsf{R}^p$ and $\mathbf{y} \in \mathsf{R}^d$, let us define the conditional distribution function of $\mathbf{Y}$ given $\mathbf{X} = \mathbf{x}$ by

$$F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{y}) := \mathbb{P}\left(\mathbf{Y} \leq \mathbf{y} \mid \mathbf{X} = \mathbf{x}\right).$$

Similarly, for $j = 1, \dots, d$, let $F_{Y_j|\mathbf{X}=\mathbf{x}}(y_j) := \mathbb{P}\left(Y_j \leq y_j \mid \mathbf{X} = \mathbf{x}\right)$. Finally we define the vector of marginal conditional distribution functions and the vector of marginal conditional quantiles by

$$\overrightarrow{F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{y}) := \left(F_{Y_1|\mathbf{X}=\mathbf{x}}(y_1), ..., F_{Y_d|\mathbf{X}=\mathbf{x}}(y_d)\right),$$
$$\overrightarrow{F^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{y}) := \left(F^{-1}_{Y_1|\mathbf{X}=\mathbf{x}}(y_1), ..., F^{-1}_{Y_d|\mathbf{X}=\mathbf{x}}(y_d)\right),$$

where for every $j = 1, \dots, d$, $F^{-1}_{Y_j|\mathbf{X}=\mathbf{x}}$ is the inverse of $F_{Y_j|\mathbf{X}=\mathbf{x}}$.

This means that it is possible to write the following conditional version of Sklar's theorem, introducing the concept of conditional copulas.

Theorem 2.2 (Conditional Sklar's theorem)**.**
Let $(\mathbf{X}, \mathbf{Y})$ be a random vector on R^{p+d} with continuous conditional marginals $F_{Y_j|\mathbf{X}=\mathbf{x}}$ for every $j = 1, \dots, d$ and every $\mathbf{x} \in \mathsf{R}^p$. Then for every $\mathbf{x} \in \mathsf{R}^d$ there exists a distribution $C_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}$ on $[0,1]^d$ with uniform margins conditionally to $\mathbf{X} = \mathbf{x}$, named the conditional copula of $\mathbf{Y}$ given $\mathbf{X} = \mathbf{x}$, such that

$$\forall \mathbf{y} \in \mathsf{R}^d, \quad F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{y}) = C_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\left(\overrightarrow{F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{y})\right),$$

and $C_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}$ is given by

$$\forall \mathbf{u} \in [0,1]^d, \quad C_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\left(\mathbf{u}\right) = F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\left(\overrightarrow{F^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right).$$

The theory of parametric copulas explained in Section 2.2 can be extended to conditional copulas. A conditional parameter $\theta(\mathbf{x})$ that changes based on the value(s) of $\mathbf{x}$ is used. The conditional parametric copula where the conditional dependence is modelled by the the function $\theta(x)$ is denoted by $C_{\theta(\mathbf{x})}(\mathbf{u})$. If Θ is the complete space of possible parameters, the set of conditional parametric copulas can be denoted by $\{C_{\theta(\mathbf{x})}(\mathbf{u}), \theta(\mathbf{x}) \in \Theta\}$. It is assumed that there exists a piecewise continuous mapping from the conditional variable(s) $\mathbf{x}$ to the parameter space Θ.

2.4 Nonparametric bootstrap

The nonparametric bootstrap is a resampling method in which random sampling with replacement is used to estimate statistical functionals of a distribution. The first bootstrap method was published by Bradly Efron [19]. The nonparametric bootstrap procedure will be explained by means of a short example.

We might want to estimate the average weight of a Dutch male. It would be almost impossible to measure each and every single male. We thus take a sample of 1000 Dutch males from the population and measure those. We then end up with 1000 measurements which we assume is a good enough representation of the entire population. This makes us able to estimate the mean of the sample, and by extension, the mean of the entire population. If we want to estimate a confidence interval for the average weight of a Dutch male however, we would need many of those samples to get this estimate, which defeats the purpose of only needing to take a small sample. This is where the nonparametric bootstrap method comes in. From the original sample we sample with replacement to get b bootstrap samples. If the original sample is a good representation of the total population, the bootstrap samples will be as well by assumption. This means that we do not have to get multiple samples from the original population, but instead can use the bootstrap method. See also Figure 6 for reference.

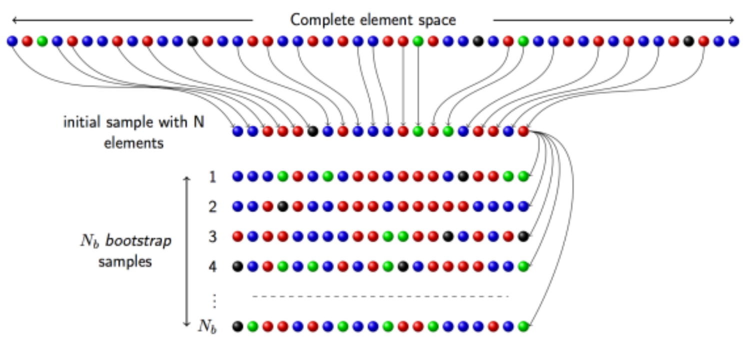


Figure 6: *Illustration of nonparametric bootstrap resampling. [20]*

A bootstrap technique will be used to calculate multiple bootstrapped test statistics. These so called bootstrapped test statistics will then be used to compute an approximate p-value for the test. This will be explained in Section 2.5.

2.5 Goodness of fit tests

The goodness of fit test is a statistical hypothesis test that tests how well sample data fits a distribution from a known population. An example of a more well known and commonly used goodness of fit test is the Shapiro-Wilk test to test the normality of a sample or distribution.

We want to test whether the data fits a specific copula family, or whether it does not fit that specific copula family. This means that the null hypothesis, $\mathcal{H}_0$, is that the data fits this certain copula family. The alternative hypothesis, $\mathcal{H}_1$, is that the data does not fit this specific copula family. For testing $\mathcal{H}_0$ on unconditional copulas, this comes down to testing

$$\mathcal{H}_0 : \quad \forall \mathbf{u} \in \mathsf{R}^d, \quad \exists \theta \in \Theta, \quad s.t. \quad C(\mathbf{u}) = C_\theta(\mathbf{u}),$$

against

$$\mathcal{H}_1 : \quad \exists \mathbf{u} \in \mathsf{R}^d, \quad \forall \theta \in \Theta, \quad s.t. \quad C(\mathbf{u}) \neq C_\theta(\mathbf{u}).$$

The alternative hypothesis states that the copula C is not any of the copulas $\{C_\theta, \, \theta \in \Theta\}$, i.e. $C \notin \{C_\theta, \, \theta \in \Theta\}$. For unconditional copulas different goodness of fit tests have been proposed and tested, see for example [12, 13]. One of the more straightforward approaches to do goodness of fit testing for copulas is measuring the distance between the estimated empirical copula and the copula under $\mathcal{H}_0$. The distance between these is T, the test statistic. A bootstrap procedure is then performed to calculate multiple bootstrapped test statistics, T^*, to compare to T. This goodness of fit test will be explained in more detail.

The distance is estimated between the non-parametric estimated copula of the original data, $\hat{C}$ and the parametric copula with the estimated parameter $\hat{\theta}$ under $\mathcal{H}_0$, $C_{\hat{\theta}}$. To compute the test statistic, T, the Cramer-von-Mises criterion, also known as the two-norm, will be used.

$$T = \left|\left|\hat{C}(\mathbf{u}) - C_{\hat{\theta}}(\mathbf{u})\right|\right|_2^2 = \int \left(\hat{C}(\mathbf{u}) - C_{\hat{\theta}}(\mathbf{u})\right)^2 d\mathbf{u}. \tag{3}$$

If the non-parametric estimated copula is close to the distribution of the tested family, this difference will be small. In the case that the estimated cumulative density is far from the distribution of the tested family, this difference will be large. This test statistic by itself does not tell us much however, since we do not yet know how small small is, or how large large is. The bootstrapped test statistics are needed as a comparison for the test statistic.

To obtain these bootstrapped test statistics, a bootstrap procedure is used to obtain a new sample of the original data. The distance between the estimated cumulative density of this bootstrap sample, $\hat{C}^*$ and the non-parametric estimated copula is then computed. Under usual regularity conditions, this distance should always be relatively small. The distance between the copula with the estimated parameter $\hat{\theta}$ and the copula with the estimated parameter from the boottrap sample, $\hat{\theta}^*$ is also computed. Again, since the bootstrap sample is expected to be close to the original sample, this difference should also be small. These two differences are added to obtain the bootstrapped test statistic, T^*, which is expected to be relatively small.

$$T^* = \left|\left|\hat{C}^*(\mathbf{u}) - \hat{C}(\mathbf{u}) + C_{\hat{\theta}}(\mathbf{u}) - C_{\hat{\theta}^*}(\mathbf{u})\right|\right|_2^2 \tag{4}$$

If the original sample is close to the tested distribution, both T and T^* should be approximately equally small. If however the original sample is not close to the tested distribution, T will be much larger than T^*. The approximate p-value is then the estimated probability that $\{T \leq T^*\}$ computed over several bootstrap resamplings. The amount of bootstrap resamplings in this case is K.

$$\widehat{\text{p-value}} = \frac{1}{K}\sum_{i=1}^{K} \mathbb{1}\left\{T \leq T^{*,i}\right\} \approx \mathbb{P}(T \leq T^*|T) = \text{p-value}$$

3 Theory

In this section more mathematical details will be explained about estimating the empirical copula, the bootstrap measures, and the test statistics for conditional copulas. The simplifications that are done and assumptions that are made will be discussed in Section 4.

First the test discussed in Section 2.5 needs to be extended. We want to test whether for all x, the tested copula belongs to a certain conditional parametric copula family, for example the Gaussian family. This means that the null hypothesis,

$$\mathcal{H}_0 : \forall \mathbf{x} \in \mathsf{R}^d, \quad \exists \theta(\mathbf{x}) \in \Theta, \quad s.t. \quad C_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\cdot|\mathbf{X}=\mathbf{x}) = C_{\theta(\mathbf{x})},$$

will be tested against the alternative hypothesis,

$$\mathcal{H}_1 : \exists \mathbf{x} \in \mathsf{R}^d, \quad \forall \theta \in \Theta, \quad s.t. \quad C_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\cdot|\mathbf{X}=\mathbf{x}) \neq C_{\theta}.$$

The alternative hypothesis means that the tested copula does not belong to the Gaussian family for any correlation parameter at the point x.

3.1 Estimation

The non-parametric estimated conditional copula, $\hat{C}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}$ can be defined as

$$\hat{C}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u}) := \hat{F}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\left(\overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})}\right),$$

where

$$\hat{F}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{y}) = \frac{1}{nh^p}\sum_{i=1}^{n} w_i(\mathbf{x},h)\mathbb{1}\left\{\mathbf{Y}_i \leq \mathbf{y}\right\},$$

and where $w_i(\mathbf{x}, h)$ are the weights belonging to the Epanechnikov kernel with bandwidth h at the point $\mathbf{x}$,

$$w_i(\mathbf{x},h) = K\left(\frac{\mathbf{X}_i - \mathbf{x}}{h}\right).$$

Under the assumption of the null hypothesis, $\mathcal{H}_0$, $\hat{\theta}(\mathbf{x})$ is estimated. The estimation of $\hat{\theta}(\mathbf{x})$ is done by using standard maximum likelihood in the following way,

$$\hat{\theta}(\mathbf{x}) := \arg\max_{\theta\in\Theta}\sum_{i=1}^{n} w_i(\mathbf{x},h)\log c_\theta\left(\overrightarrow{\hat{F}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{y})}\right),$$

where c_θ is the density of the copula C_θ. Using this $\hat{\theta}(\mathbf{x})$, the parametric estimated conditional copula $C_{\hat{\theta}(\mathbf{x})}$ is obtained, which is the copula with the estimated parameter $\hat{\theta}(\mathbf{x})$ at the point $\mathbf{x}$.

3.1.1 Test statistic

The test statistics are computed in the same way as in Equation 3. The test statistics are extended to use the conditional copula, instead of the unconditional copula. The test statistic is then computed in the following way,

$$T = ||\hat{C}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u}) - C_{\hat{\theta}(\mathbf{x})}(\mathbf{u})||_2^2 = \int \left(\hat{C}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u}) - C_{\hat{\theta}(\mathbf{x})}(\mathbf{u})\right)^2 d\mathbf{u}d\mathbf{x} \tag{5}$$

As mentioned before, this test statistic gives the baseline for how far the estimated empirical copula is from the copula under the assumption of $\mathcal{H}_0$.

3.2 Bootstrap

As described in Section 2.5 a bootstrap method will be used to resample the data to calculate multiple bootstrapped test statistics to then calculate an approximate p-value. The idea of bootstrapping conditional copulas came from Omelka et al. [21]. Two different bootstrap methods are proposed in this thesis, a nonparametric bootstrap, and a conditional parametric bootstrap. Both of these bootstrap methods will be explained in the following section. The results for both of the described bootstrap methods will be compared in Section 5.

3.2.1 Nonparametric bootstrap

Firstly the nonparametric bootstrap, this method was already discussed in Section 2.4. The extension to conditional copulas is rather straightforward. Instead of only resampling $Y_{i,1}$ and $Y_{i,2}$, the conditioning variable X_i is also included in the resampling. A similar estimation as described in Section 2.5 is performed to obtain $\hat{C}^*_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}$ and $C_{\hat{\theta}^*(x)}$. Afterwards it is possible to compute the bootstrapped test statistic. The full algorithm is described below.

1. From the original data, $\mathcal{D} = (Y_{i,1}, Y_{i,2}, X_i)_{i=1,\dots,n}$, sample with replacement to obtain the bootstrap sample $\mathcal{D}^* = (Y^*_{i,1}, Y^*_{i,2}, X^*_i)_{i=1,\dots,n}$.
2. Estimate $\hat{C}^*_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})$ nonparametrically using the dataset $\mathcal{D}^*$.
3. Estimate the parameter $\hat{\theta}^*(\mathbf{x})$ using the dataset $\mathcal{D}^*$ to obtain $C_{\hat{\theta}^*(x)}(\mathbf{u})$.
4. $T^* = ||\hat{C}^*_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u}) - \hat{C}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u}) + C_{\hat{\theta}(\mathbf{x})}(\mathbf{u}) - C_{\hat{\theta}^*(\mathbf{x})}(\mathbf{u})||_2^2$

Where the bootstrapped test statistic T^* is extended from Equation 4 to conditional copulas as was done for the test statistic in Equation 5. The p-value is thus the estimated probability that $\{T \leq T^*\}$ computed over several bootstrap resamplings. This means the p-value is estimated by

$$\text{p-value} = \frac{1}{K}\sum_{i=1}^{K} \mathbb{1}\left\{T \leq T^{*,i}\right\}.$$

This leads us to the following theorem about the consistency of the nonparametric bootstrap, for which the sketch of the proof is given in the appendix, Section 7. The proof relies on Bahadur representations of four empirical processes

1. $\sqrt{nh^p}\Big(\hat{C}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u}) - C_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})\Big)$,
2. $\sqrt{nh^p}\Big(\hat{C}^*_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u}) - \hat{C}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})\Big)$,
3. $\sqrt{nh^p}\Big(C_{\hat{\theta}(\mathbf{x})}(\mathbf{u}) - C_{\theta_0(\mathbf{x})}(\mathbf{u})\Big)$,
4. $\sqrt{nh^p}\Big(C_{\hat{\theta}^*(\mathbf{x})}(\mathbf{u}) - C_{\hat{\theta}(\mathbf{x})}(\mathbf{u})\Big)$.

The Bahadur representation of the first empirical process is due to [18]. We give a proof of the main arguments leading to Bahadur representations of the second and third empirical processes. We conjecture that similar arguments can be combined to prove the Bahadur representation of the fourth empirical process.

Theorem 3.1 (Consistency of the nonparametric bootstrap)**.**
Let $n, d, p > 0$. Let $(C_\theta, \theta \in \Theta)$ be a parametric copula family with $\Theta \subset \mathsf{R}^m$ for some $m > 0$. Assume that we observe an i.i.d. sample $(\mathbf{X}_i, \mathbf{Y}_i)_{i=1,\dots,n}$ such that for every $\mathbf{x} \in \mathsf{R}^p$, $C_{\mathbf{Y}|\mathbf{X}=\mathbf{x}} = C_{\theta_0(\mathbf{x})}$ for some value $\theta_0(\mathbf{x}) \in \Theta$. Let T_n be the test statistic computed using the sample $(\mathbf{X}_i, \mathbf{Y}_i)_{i=1,\dots,n}$ and T^*_n its bootstrapped test statistic. Then under appropriate regularity conditions,

$$\big(T_n, T^*_n\big) \overset{\text{law}}{\longrightarrow} \big(T_\infty, T^*_\infty\big),$$

as $n \to \infty$, where T_∞ and T^*_∞ are two independent random variables satisfying $T_\infty \overset{\text{law}}{=} T^*_\infty$.

As a particular case, the Gaussian copula family corresponds to the case where $m = 1$, $\Theta = [-1, 1]$ and C_θ is the Gaussian copula with parameter $\theta \in \Theta$.

3.2.2 Conditional parametric bootstrap

In Section 2.4 the nonparametric bootstrap was explained. A different way to approach the bootstrap is for conditional data. Instead of sampling with replacement from all the data, only the conditonal variable(s) are sampled with replacement. Then under the assumption of the null hypothesis new data is simulated for the other parameters. The full algorithm is described below.

1. For every $i = 1, \dots, n$
   (a) Sample with replacement to obtain $\mathbf{X}_i^{**}$ from the original $\mathbf{X}_i$.
   (b) Simulate $(Z_{i,1}^{**}, Z_{i,2}^{**}) \sim C_{\hat{\theta}(\mathbf{X}_i^{**})}$

   To obtain the complete bootstrap sample: $\mathcal{D}^{**} = (Z_{i,1}^{**}, Z_{i,2}^{**}, \mathbf{X}_i^{**})_i$.
2. Estimate $\hat{C}^{**}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})$ nonparametrically using the dataset $\mathcal{D}^{**}$.
3. Estimate the parameter $\hat{\theta}^{**}(\cdot)$ using the dataset $\mathcal{D}^{**}$.
4. $T^{**} = ||\hat{C}^{**}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u}) - C_{\hat{\theta}^{**}(\mathbf{x})}(\mathbf{u})||_2^2$

The bootstrapped sample of the conditioned variables is generated by sampling from the parametrically estimated conditional copula. Since a different bootstrap procedure is used than for the nonparametric bootstrap, the bootstrapped test statistic is also computed in a different way. Only the nonparametrically estimated conditional copula on the bootstrapped sample, $\hat{C}^{**}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})$, is compared to the conditional copula estimated under the parametric model on the bootstrapped sample, $C_{\hat{\theta}^{**}(\mathbf{x})}(\mathbf{u})$.

The p-value is still the estimated probability that $\{T \leq T^{**}\}$ computed over several bootstrap resamplings. It is thus estimated in the same way as is done for the nonparametric bootstrap,

$$\text{p-value} = \frac{1}{K}\sum_{i=1}^{K} \mathbb{1}\left\{T \leq T^{**,i}\right\}.$$

4 Method

Now that the theory of conditional copulas is discussed, there are some simplifications that need to be made to make the computations feasible. These will be discussed in the following section. Also the computational setting for the simulation study and real world data will be discussed in this section.

4.1 Estimations

For all of the following results, copulas with 2 explained variables and 1 explanatory variable will be considered. These variables will be called Y_1, Y_2 and X respectively. Adding more explanatory variables makes estimations of the conditional parameter much harder, which is not the focus of these results.

Since a bootstrap method is performed, not all of the same points that make up the original sample end up in the bootstrap sample. When using the conditional parametric bootstrap new points are even simulated. That means that for the computation of the test statistic, and the bootstrapped test statistics, fixed points need to be chosen to do the estimations. The standard practice is to design a grid and perform grid based estimation. For every point in this grid the computations are performed. Since everything with copulas happens on the unit interval, it is relatively straightforward to design a fixed grid. The grid that was chosen is $(0.1, 0.25, 0.4, 0.6, 0.75, 0.9)$. For the sake of simplicity this grid is used for the all variables, Y_1, Y_2 and X.

4.2 Level and Power

The level and the power of the test are an important measure for how good a certain test is when testing. To compute both of these measures for our test, it is important to know how they work.

The power of a hypothesis test is the probability of rejecting the null hypothesis when in fact it is false. This means that when the data does not follow for example a Gaussian copula, and it is tested whether the data belongs to a Gaussian copula, the test should reject the null hypothesis $\mathcal{H}_0$. To do this, data is simulated from a non-Gaussian copula, and the test is repeated 50 times. The power of the test is now the average probability of rejecting $\mathcal{H}_0$.

The level of a hypothesis test on the other hand is the probability of rejecting the null hypothesis when in fact it is true. This means that when the data does indeed conform to a Gaussian copula, and we test whether the data belongs to a Gaussian copula, the test should accept the null hypothesis $\mathcal{H}_0$. To do this, we simulate data from a Gaussian copula, and repeat the test 50 times once again. The level of the test is once again the average probability of rejecting $\mathcal{H}_0$.

This means that a hypothesis test that has a high power and a low level would be considered a "good" test. In practice this means that we want the difference between the level and the power of our test as large as possible.

4.3 Computational setting

The four main copula families mentioned in Section 2.2 will be compared. These are the Gaussian, the student with 4 degrees of freedom, the Clayton, and the Gumbel family. For both of the bootstrap methods, the non-parametric bootstrap and the conditional parametric bootstrap, $K = 100$ bootstrap replications are performed to obtain one approximate p-value. As mentioned before, this is repeated 50 times to obtain the level and the power in all of the simulations.

A τ to close to 1 means that the methods will not work. For $\tau = 1$ there exists only one copula with perfect positive dependence. Likewise, for $\tau = 0$ all copulas correspond to the independence copulas. For this reason the simulated values for τ in the simulation study will not be close to either 0 or 1. Data that has a perfect positive dependence, or perfect independence is very rare in practice so these problems do not occur when using real world data.

5 Results

In the previous sections we proposed two different bootstrap procedures to compute bootstrapped test statistics. In this section we will first perform a simulation study with these two different bootstrap procedures, the nonparametric bootstrap and the conditional parametric bootstrap. Afterwards we will apply the bootstrap methods to real world stock exchange data.

5.1 Simulation study

A simulation study is performed on the two different bootstrap procedures to gain more insight in their general performance. It is also important to see how the characteristics of the experimental design interact with each other and to see how they influence the performance. For example, it is known that the kernel bandwidth heavily influences the level and the power of tests like the ones performed here. A bandwidth that is too small means the test will not be able to accurately reject the null hypothesis when it is false. A low bandwidth value will thus usually result in a small power value. On the other hand, a large bandwidth value will result in almost always rejecting the null hypothesis, even when it is true, resulting in a very high level value. The 'best' kernel bandwidth in turn is heavily dependent on the sample size. [21]

5.1.1 Nonparametric bootstrap

A simulation study for the nonparametric bootstrap method will be performed first. The Gaussian copula ($\mathcal{H}_0$) is compared with the Clayton copula ($\mathcal{H}_1$). The data for this figure was simulated by using a sample size of $n = 1000$ points and $\tau = 0.05 + 0.95x$ was used with x uniform on $[0, 1]$. The level and the power as function of the kernel bandwidth are visible in Figure 7. The shaded areas in the figure are the 95% confidence bands for the level and the power. From Figure 7 it becomes clear that the ideal bandwidth would lie between about 0.6 and 0.8 for these characteristics, since the difference between the level and the power is the largest for these kernel bandwidth values.

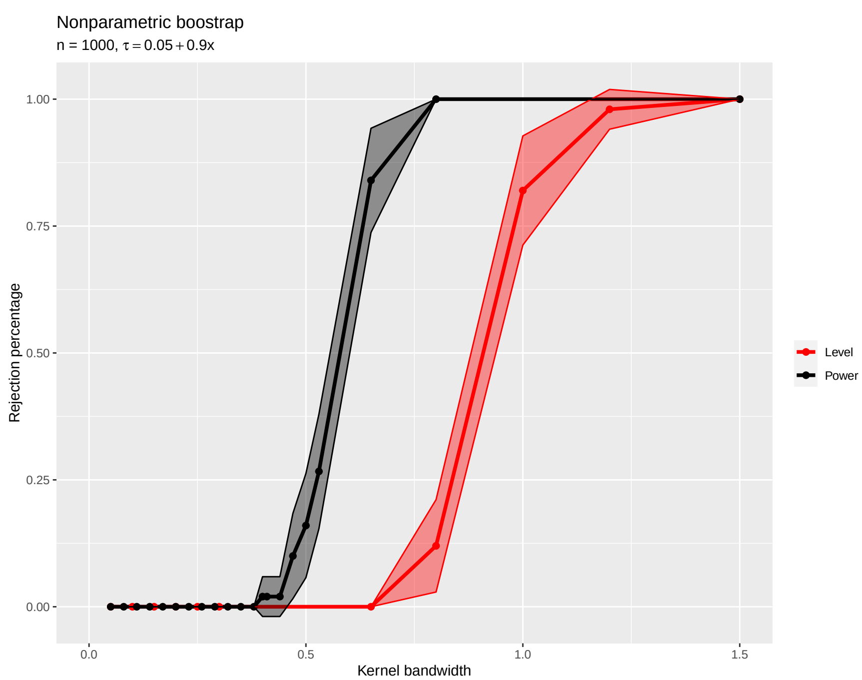


Figure 7: *The level and the power of the nonparametric bootstrap for $n = 1000$ and $\tau = 0.05 + 0.95x$ with x uniform on $[0, 1]$ when comparing the Gaussian copula ($\mathcal{H}_0$) with the Clayton copula ($\mathcal{H}_1$).*

Changing one or more of these characteristics however, such as the sample size, n, results in the result shown in Figure 8. In this figure, the sample size was decreased from $n = 1000$ to $n = 500$. It is observable that the ideal bandwidth now lies approximately between 1 and 1.5. The maximum difference between the level and the power also decreases. As mentioned before, the fact that as the sample size decreases, the ideal bandwidth value increases, is completely in line with the relevant literature.

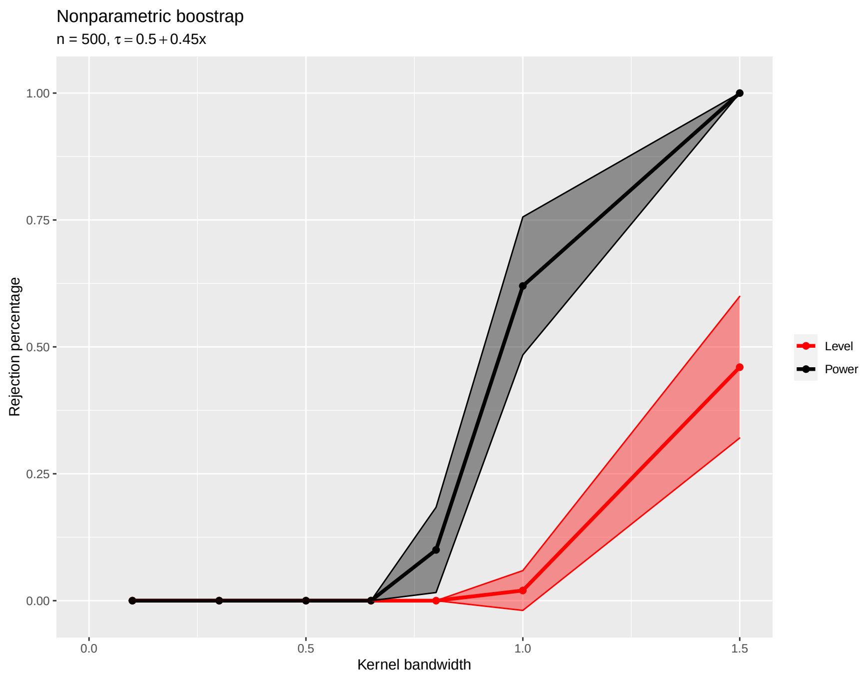


Figure 8: *The level and the power of the nonparametric bootstrap for $n = 500$ and $\tau = 0.05 + 0.95x$ with x uniform on $[0, 1]$, when comparing the Gaussian copula ($\mathcal{H}_0$) with the Clayton copula ($\mathcal{H}_1$).*

5.1.2 Conditional parametric bootstrap

We can do the same comparison we did for the nonparametric bootstrap for the conditional parametric bootstrap. The Gaussian copula ($\mathcal{H}_0$) and the Clayton copula ($\mathcal{H}_1$) are compared. Again we use $n = 1000$ points and $\tau = 0.05 + 0.9x$ with x uniform on $[0, 1]$. The results are displayed in Figure 9. From Figure 9 it becomes clear that the ideal bandwidth would lie between 0.3 and 0.5 for these parameters for this method, since the difference between the level and the power is the largest for these kernel bandwidth values. When comparing this to the nonparametric bootstrap in Figure 7 it stands out that a much lower kernel bandwidth value is desirable for the conditional parametric bootstrap.

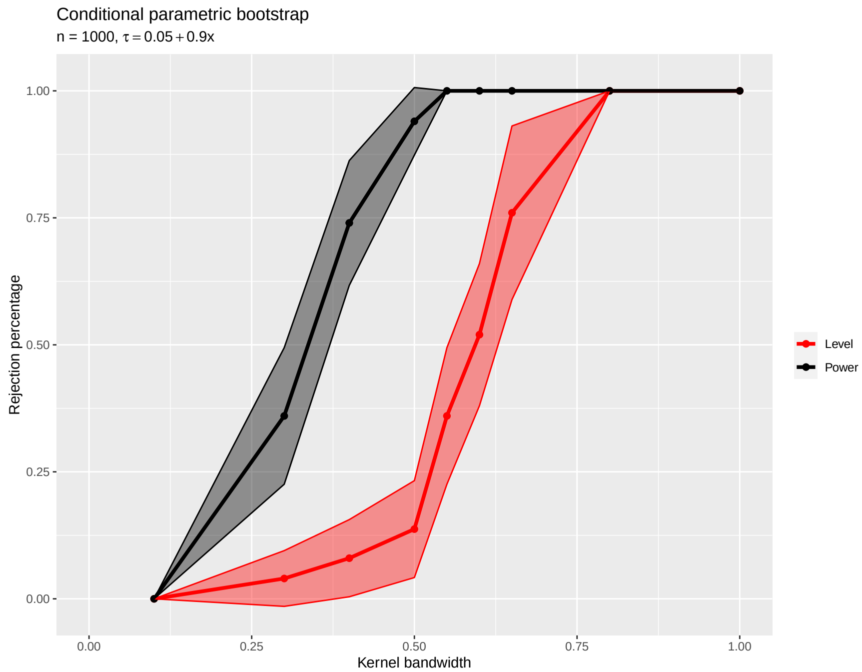


Figure 9: *The level and the power of the conditional parametric bootstrap for $n = 1000$ and $\tau = 0.05 + 0.95x$ with x uniform on $[0, 1]$, when comparing the Gaussian copula ($\mathcal{H}_0$) with the Clayton copula ($\mathcal{H}_1$).*

When decreasing the sample size from $n = 1000$ to $n = 500$, as was done for the nonparametric bootstrap, the results in Figure 10 is obtained. It is again observable that decreasing the sample size in this manner from 1000 to 500, the ideal bandwidth value increases. It now lies approximately between 0.55 and 0.65.

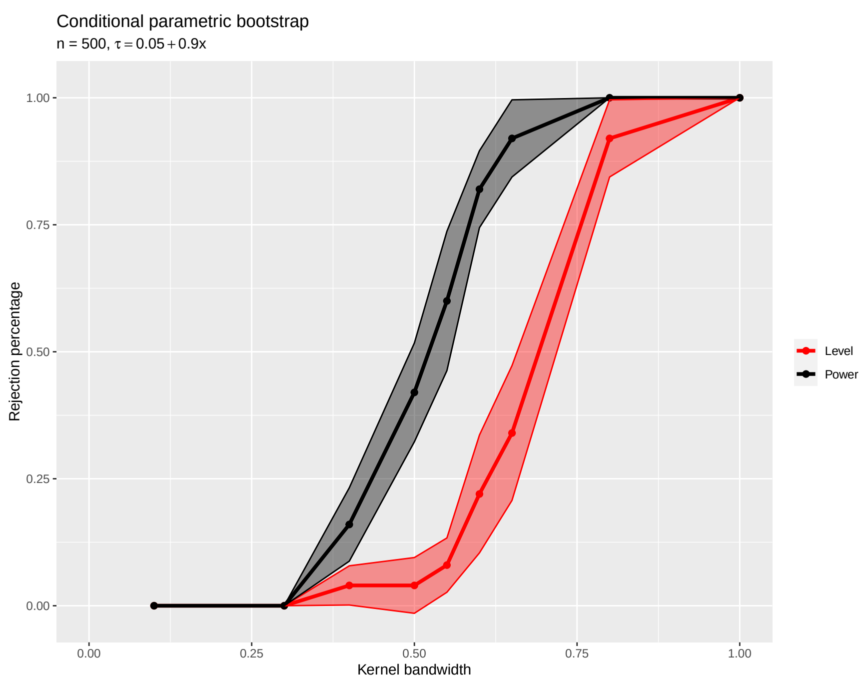


Figure 10: *The level and the power of the conditional parametric bootstrap for $n = 500$ and $\tau = 0.05 + 0.45x$ with x uniform on $[0, 1]$, when comparing the Gaussian copula ($\mathcal{H}_0$) with the Clayton copula ($\mathcal{H}_1$).*

In all of previous figures the τ that was used to simulate the original copula was uniformly distributed in the interval $(0.05; 0.95)$ by using $\tau = 0.05 + 0.9x$ with x uniform on $[0, 1]$. In the following figures a τ will be used that is higher. Using higher values for τ means that features specific to a certain copula family become more pronounced as was visible in Figure 3. This should make it easier for the test to distinguish different copula families. This is indeed visible in Figure 11, where values for τ were used that were uniformly distributed in the interval $(0.5; 0.95)$ using $\tau = 0.5 + 0.45x$ with x uniform on $[0, 1]$. When comparing this to Figure 9 the main difference is the much more steeply increasing power. This does not make a huge difference on the most ideal bandwidth values directly, this is still between about 0.3 and 0.5. It does however make a difference on the accuracy of the test, since the difference between the level and the power is larger for this range of bandwidth values.

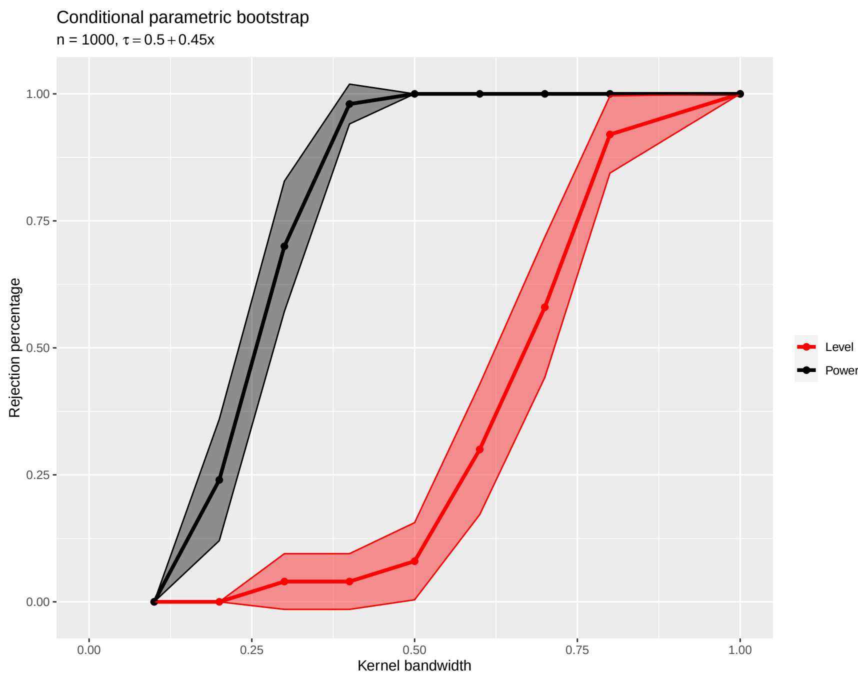


Figure 11: *The level and the power of the conditional parametric bootstrap for $n = 1000$ and $\tau = 0.5 + 0.95x$ with x uniform on $[0, 1]$, when comparing the Gaussian copula ($\mathcal{H}_0$) with the Clayton copula ($\mathcal{H}_1$).*

For these higher values of τ the sample size will also be decreased from $n = 1000$ to $n = 500$. The result for this is visible in Figure 12. Observe that when comparing this to Figure 10, not only the steepness of the power increases, the steepness of the level decreases too. The figure shows that the ideal bandwidth would be between 0.5 and 0.7, which is a much larger interval than for Figure 10, and the test should me more accurate for those values too.

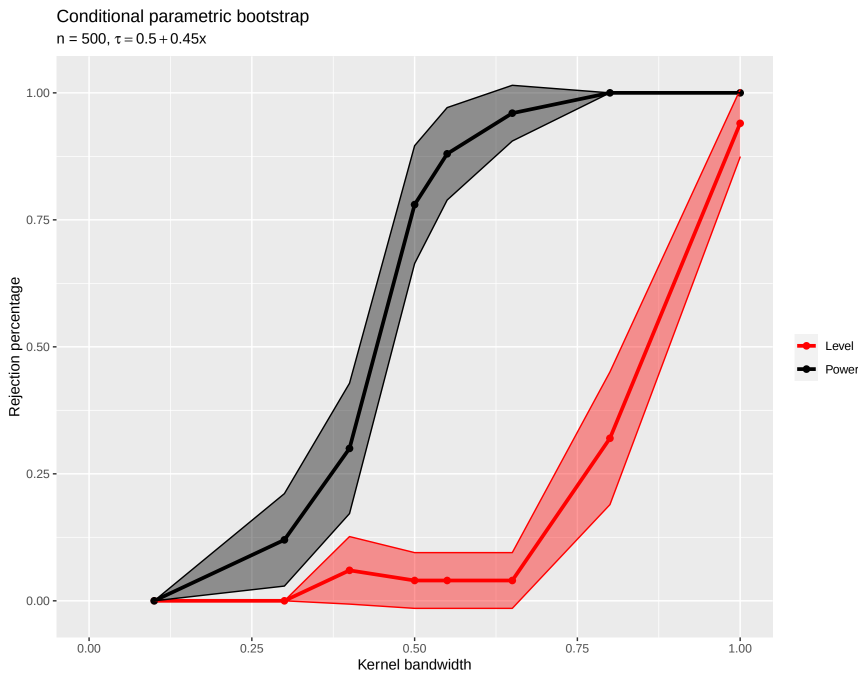


Figure 12: *The level and the power of the conditional parametric bootstrap for n = 500, and and $\tau = 0.5 + 0.45x$ with x uniform on $[0, 1]$, when comparing the Gaussian copula ($\mathcal{H}_0$) with the Clayton copula ($\mathcal{H}_1$).*

So far the Gaussian copula has always been the choice for $\mathcal{H}_0$. The data was either simulated from a Gaussian copula or a Clayton copula. Other choices for $\mathcal{H}_0$ and $\mathcal{H}_1$ should be considered too. This is done with the conditional parametric bootstrap, with $n = 1000$ points, and with the high values for $\tau = 0.5 + 0.45x$ with x uniform on $[0, 1]$. The results are visible in Figure 13. The Gaussian copula is the null hypothesis for each of the families. The other families are used as the alternative hypothesis one by one. The results when comparing the Gaussian copula to the Clayton copula are as already seen in Figure 11. The other families do not perform as well however. The Student copula with 4 degrees of freedom is too close to the Gaussian copula to easily distinguish for our test. The Gumbel copula is not nearly as difficult to distinguish for the test as the Student copula, but also not as easy as the Clayton copula. The similarity between the different copula families is amplified by the grid that we use, which is quite sparse, especially around the edges, where the largest differences between these copula families present themselves. The nonparametric bootstrap suffers from the same inability to distinguish between very similar copula families as the conditional parametric bootstrap.

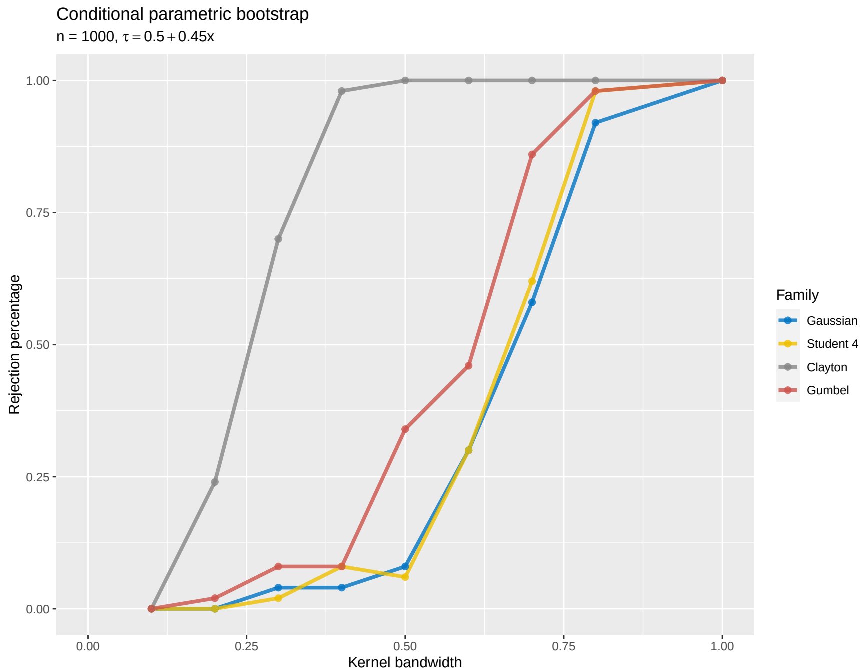


Figure 13: *Comparison of the di erent copula families, where $\mathcal{H}_0$ is the Gaussian copula*

5.1.3 Effect of the choice of the families on the level and power

Now that we have seen that not all families are as easily distinguished, it might be important to test all the different families against each other, instead of only testing against the Gaussian copula. This large scale comparison is visible in Table 1 for the nonparametric bootstrap and Table 2 for the conditional parametric bootstrap. All of the simulations for these tables were done with $n = 500$ points, and using $\tau = 0.5 + 0.45x$ with x uniform on $[0, 1]$. For the nonparametric bootstrap a bandwidth value of 1 was chosen, which was related from Figure 8. For the conditional parametric bootstrap a bandwidth value of 0.55 was chosen, derived from 12.

It becomes clear from those two tables that both the nonparametric bootstrap and the conditional parametric bootstrap suffer from the same problem of being unable to distinguish very similar families. It is important to note however that the results for these tables were obtained by using $n = 500$, which is a relatively small sample size, and as seen in the simulation study significantly reduces how well these tests perform.

Table 1: *Nonparametric bootstrap method comparison for n = 500 and a kernel bandwidth value of 1*

sim \ est	Gaussian	Student 4	Clayton	Gumbel
Gaussian	0	0.02	1	0
Student 4	0	0	0.98	0.02
Clayton	0.66	0.36	0.02	1
Gumbel	0.04	0.02	1	0

Table 2: *Conditional parametric bootstrap method comparison for n = 500 and a kernel bandwidth value of 0.55*

sim \ est	Gaussian	Student 4	Clayton	Gumbel
Gaussian	0.1	0.08	1	0.12
Student 4	0.05	0.08	0.92	0.04
Clayton	0.76	0.48	0.02	1
Gumbel	0.1	0	1	0.04

5.2 Power under local alternatives

Something else to consider is the power under local alternatives, sometimes known as the local power. It has been studied since the 1980s [22]. For conditional copulas the power under local alternatives considers how well a hypothesis test performs when not all of the data belongs to a certain copula. The data could for example be split between two different copula families.

For the simulation, data is generated in the following way. Let $z^* \in [0,1]$, then for every $z \in [0,1]$, if $z < z^*$, the data is simulated from the copula under $\mathcal{H}_0$, $C^{Gaussian}_{\theta(z)}$. If $z > z^*$, the data is simulated from a different copula $C^{Clayton}_{\theta(z)}$.

In the case that $z^* = 1$, there is no dilution happening, and the entire copula is simulated from a Gaussian copula. This means that the level of the test will be obtained. In the case that $z^* = 0$, the data is all simulated from a Clayton copula, there is nothing of the copula under $\mathcal{H}_0$, and the power of the test is obtained.

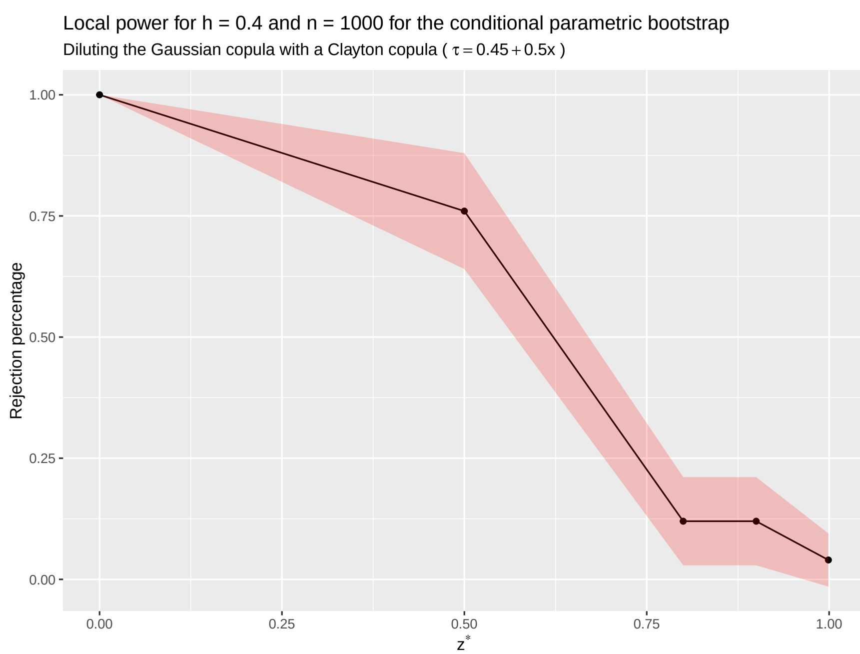


Figure 14: *Power under local alternatives*

Looking at the results in Figure 14, we can see that for these parameters ($h = 0.4$, $n = 1000$), the results for the level and the power are as expected, which is good. When only a little bit of dilution is happening, at $z^* = 0.8$ for example, the test still does not reject much, which is also to be expected, since most of the data is still coming from the Gaussian copula. When at $z^* = 0.5$, half of the simulated data is from a Clayton Copula, and half is from a Gaussian copula, about 60% of the tests are rejected, which is a good result.

5.3 Application to real world data

It is now time to look at real world stock data. For this, the data from about 3 years of stock trading have been used, consisting of 1000 data points from the French stock index (FCHI), the German stock index (GDAXI) and the Dutch stock index (AEX). The data that is used is the adjusted returns data after ARMA-GARCH filtering, also known as innovation data.

Since the sample size consists of 1000 points, it is possible to reference the simulation study, specifically Figure 9 to find a suitable kernel bandwidth value. From Figure 9 a kernel bandwidth value of 0.45 would be the most ideal bandwidth value. In Figure 15 the transformed FCHI and GDAXI data is displayed.

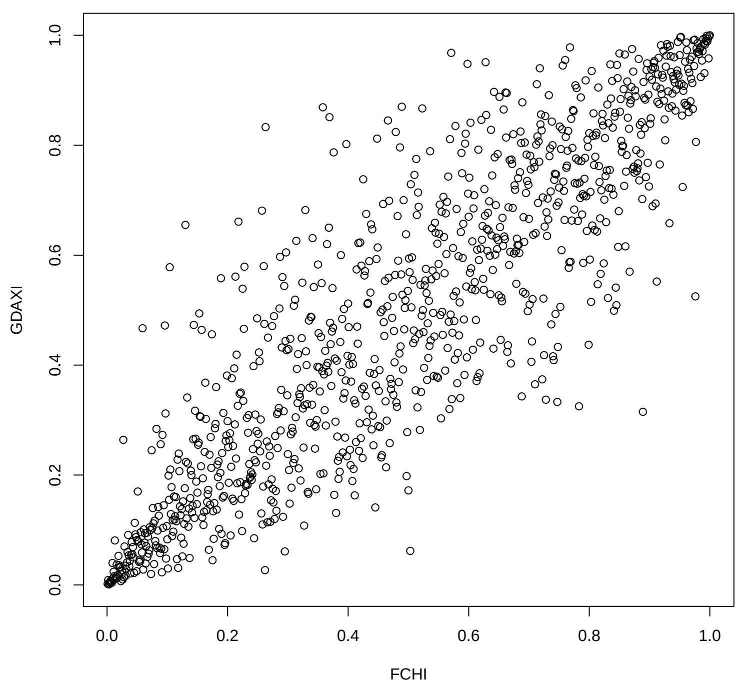
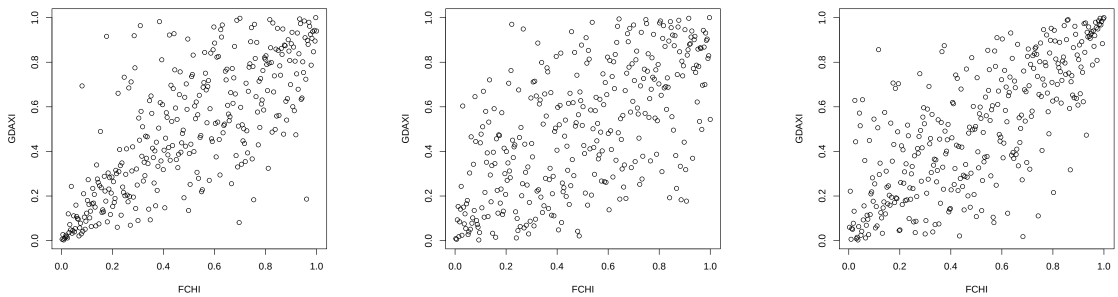


Figure 15: *FCHI and GDAXI innovations transformed to be uniform on* $[0, 1]$.

Since we are working with conditional copulas, we would like to estimate the distribution of the data conditional on the AEX. In Figure 16 the data is split in three different sets, where for each set the data is transformed to be uniform on [0,1] after the split.

- Figure 16a shows the data for when the transformed innovations of the AEX are less than 0.33. This set has a strong dependence on both tails, but there are still some points in the other corners, which would not be possible were this data from a Gaussian copula. Finally, this dataset is not very symmetric around the diagonal $y = x$, which is not possible for a Gaussian copula.
- Figure 16b shows the data when the transformed innovations of the AEX are between 0.33 and 0.67. This figure also features some tail dependence on both tails, but less strong than Figure 16a. The correlation is this figure looks to be very weak.
- Figure 16c shows the data when the transformed innovations of the AEX are more than 0.67. The dependence in the top-right corner of this data looks very strong, much stronger than in the bottom-left corner, which again does not correspond to a Gaussian copula.



(a) *AEX* $<$ *0.33* **(b)** *0.33* $\leq$ *AEX* $\leq$ *0.67* **(c)** *AEX* $>$ *0.67*

Figure 16: *The data for the FCHI and GDAXI when split in three parts according to the transformed innovations of the AEX.*

Figure 17 displays the estimated parameter $\hat{\theta}(x)$ of the conditional copula between the French and German stock innovations as a function of the Dutch stock market innovation x, computed using the kernel bandwidth $h = 0.45$. As mentioned in the above discussion about Figure 16, very negative values of the Dutch stock market innovation correspond to a higher dependence between the French and the German stock innovations. This is coherent with our observations of Figure 16a: the dependence increases during a crisis. When the Dutch stock market is stable (i.e. intermediate innovations), the dependence reaches its lowest value, as seen on Figure 16b. When Dutch stock market innovations are close to their highest values, the dependence increases, though not as much as in the opposite crisis-like situation, as Figure 16c shows.

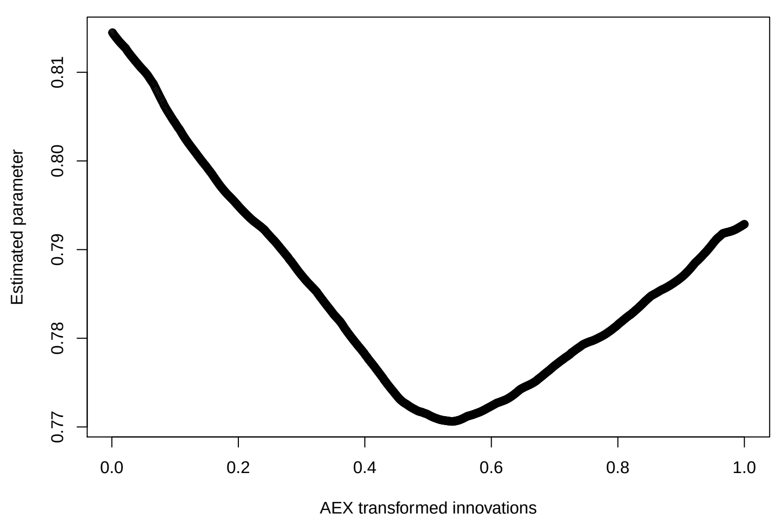


Figure 17: *The estimated parameter for a Gaussian copula for the FCHI and GDAXI conditional on the AEX with* $h = 0.45$

Performing a GOF-test on this data for the Gaussian copula, we obtain a p-value of 0 for both the nonparametric bootstrap method and the conditional parametric bootstrap method. It is thus very unlikely that the FCHI and GDAXI data is from a conditional Gaussian copula although the original data (Figure 15) looked somewhat like a bivariate Gaussian copula. This would mean that if a Gaussian copula was used for risk estimation on this stock portfolio, you would estimate your risk wrong.

6 Discussion and Conclusion

In this thesis, two different bootstrap methods were proposed, each with a different way to perform a goodness of fit test for conditional copula models. An extensive simulation study was done on the performance of both the nonparametric bootstrap and the conditional parametric bootstrap. Both of these methods performed well when distinguishing between the Clayton and the Gaussian copula. When comparing other families of copulas to each other, both of the methods had more difficulty correctly distinguishing them. A small test on the power under local alternatives gave promising results on the consistency of conditional parametric bootstrap.

The conditional parametric bootstrap was applied to real world stock data, where the data looked to be from a Gaussian copula, but the null hypothesis had to be rejected, since the conditional data did not correspond to a Gaussian copula.

We gave a proof of the main arguments leading to Bahadur representations of four empirical processes, from which we derive the consistency of the nonparametric bootstrap test procedure. The details of the proof are left for future work.

We close with some discussion and remarks. As explained in Section 4, a grid was chosen. This grid was designed to cover all the areas of the unit interval. It became apparent when comparing different families that the most important distinguishing features happen more towards the edges of the unit square. It would be beneficial to give more consideration to the grid. It would for example be helpful to add more points to the edges of the unit square, or add more points in general. This however would of course increase the computation time.

Extending the simulation study to more and larger values of sample sizes would increase the accuracy for estimation of different kernel bandwidth values. As seen in Figure 15, values for τ that are not uniformly distributed should be considered as well in an extension of the simulation study.

References

[1] Abe Sklar. "Fonctions de répartition à n dimensions et leurs marges". In: *Publications de l'Institut de statistique de l'Universite de Paris* 8 (1959), pp. 229–231.

[2] Nelson. *An Introduction to Copulas*. Ed. by P. Bickel, P. Diggle, S. Fienberg, U. Gather, I. Olkin, and S. Zeger. New York, NY: Springer, 2006. ISBN: 978-0387-28659-4.

[3] P. Embrechts, A. McNeil, and D. Straumann. "Correlation: Pitfalls and Alternatives". In: *RISK Magazine* May (1999), pp. 69–71.

[4] Donald MacKenzie and Taylor Spears. "'The formula that killed Wall Street': The Gaussian copula and modelling practices in investment banking". In: *Social Studies of Science* 44.3 (2014), pp. 393–417.

[5] Jean-David Fermanian. "In defence of the Gaussian copula". In: *CreditFlux* (2011), pp. 20–21.

[6] T. Nguyen-Huy et al. "Modeling the joint influence of multiple synoptic-scale, climate mode indices on Australian wheat yield using a vine copula-based approach". In: *European Journal of Agronomy* 98 (2018), pp. 65–81.

[7] Lu Chen and Shenglian Guo. *Copulas and Its Application in Hydrology and Water Resources*. New York, NY: Springer, 2019. ISBN: 978-981-13-0574-0.

[8] Satish Iyengar, P.K. Varshney, and Thyagaraju Damarla. "Biometric Authentication: A Copula Based Approach". In: *Multibiometrics for Human Identi cation* (Jan. 2011).

[9] Daniel Berg. "Copula goodness-of-fit testing: an overview and power comparison". In: *The European Journal of Finance* 15 (2009), pp. 675–701.

[10] Jean-David Fermanian. "Goodness of fit test for copulas". In: *Institut National De La Statistique Et Des Etudes Economiques* 34 (2003).

[11] Jean-David Fermanian. "Copulae in Mathematical and Quantitive Finance". In: *an Overview of the Goodness-of-Fit Test Problem for Copulas. Relativistic groups and analyticity*. Proceedings of the Workshop Held in Cracow (July 10–11, 2012). Ed. by Piotr Jaworski, Fabrizio Durante, and W.K. Härdle. Cracow: Springer, 2012, Chapter 4.

[12] Christian Genest, Bruno Rémillard, and David Beaudoin. "Goodness of fit test for copulas: A review and a power study". In: *Insurance: Mathematics and Economics* 44 (2009), pp. 199–213.

[13] Christian Genest and Bruno Rémillard. "Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models". In: *Annales de Institut Henri Poincare - Probabilites et Statistiques* 44 (2008), pp. 1096–1127.

[14] Olivier Scaillet. "Kernel-based goodness-of-fit tests for copulas with fixed smoothing parameters". In: *Journal of Multivariate analysis* 98 (2007), pp. 533–543.

[15] Maurice Kendall. "A New Measure of Rank Correlation". In: *Biometrika* 30 (1938), pp. 81–89.

[16] Marius Hofert et al. *Elements of Copula Modeling with R*. New York, NY: Springer, 2010. ISBN: 978-3-319-89634-2.

[17] Iréne Gijbels, Noël Veraverbeke, and Marek Omelka. "Conditional copulas, association measures and thier applications". In: *Computational Statistics and Data Analysis* 55 (2011), pp. 1919–1932.

[18] Noël Veraverbeke, Marek Omelka, and Iréne Gijbels. "Estimation of a Conditional Copula and Association Measures". In: *Scandinavian Journal of Statistics* 38.4 (2011), pp. 766–780.

[19] Bradly Efron. "Bootstrap methods: Another Look at the jackknife". In: *Annals of Statistics* 7 (1 1979), pp. 1–26.

[20] Yashu Seth. *Bootstrapping { A Powerful Resampling Method in Statistics*. 2017. URL: `https://yashuseth.blog/2017/12/02/bootstrapping-a-resampling-method-in-statistics/` (visited on 02/17/2020).

[21] Marek Omelka, Noël Veraverbeke, and Iréne Gijbels. "Boostrapping the Conditional Copula". In: *Journal of Statistical Planning and Inference* 143 (2013), pp. 1–23.

[22] Russell Davidson and James G. MacKinnon. "Implicit Alternatives and the Local Power of Test Statistics". In: *Econometrica* 55.6 (1987), pp. 1305–1329.

[23] A.D. van der Vaart and J.A. Wellner. *Weak convergence and Empirical processes*. New York, NY, 1996. ISBN: 978-1-4757-2547-6.

[24] Yanqin Fan, Qi Li, and Insik Min. "A Nonparametric Bootstrap Test of Conditional Distributions". In: *Econometric Theory* 22 (4 2006), pp. 587–613.

[25] Alexej Brauer. “Kernel Estimation of Conditional Copula Densities”. MA thesis. Technishe Iniversität München, 2016.

[26] Fentaw Abegaz, Irène Gijbels, and Noël Veraverbeke. “Semiparametric estimation of conditional copulas”. In: *Journal of multivariate analysis* 110 (2012), pp. 43–73.

[27] E.F. Acar, R.V. Craiu, and F. Yao. “Dependence Calibration in Conditional Copulas: A Nonparametric Approach”. In: *Biometrics* 67 (2011), pp. 445–453.

[28] Jean-David Fermanian and Olivier Lopez. “Single-index copulas”. In: *Biometrics* 165 (2018), pp. 27–55.

[29] Jin-Guan Lin, Kong-Sheng Zhang, and Yan-Yong Zhao. “Nonparametric estimation of multivariate multiparameter conditional copulas”. In: *Journal of the Korean Statistical Society* 46 (2017), pp. 126–136.

[30] Iréne Gijbels, Marek Omelka, and Noël Veraverbeke. “Estimation of a Copula when a Covariate Affects only Marginal Distributions”. In: *Scandinavian Journal of Statistics* 42 (2015), pp. 1109–1126.

[31] Marek Omelka, Sárka Hudecová, and Natalie Neumeyer. *Maximum pseudo-likelihood estimation based on estimated residuals in copula semiparametric models*. 2019. arXiv: 1903.04221 [math.ST].

[32] Alexis Derumigny and Jean-David Fermanian. “About tests of the “simplifying” assumption for conditional copulas”. In: *Salzburg Workshop on Dependence Models & Copulas* 5 (2017), pp. 154–197.

[33] L.J. Bain and M. Engelhardt. *Introduction to probability and mathematical statistics*. Pacific Grove, CA: Duxbury, 1992. ISBN: 0-534-92930-3.

[34] P. Billingsley. *Probability and Measure*. New York, NY: Wiley, 1995. ISBN: 0-471-00710-2.

[35] J.D. Fermanian, Dragan Radulovic, and Marten Wegkamp. “Weak convergence of empirical copula processes”. In: *Bernoulli* 10 (5 2004), pp. 847–860.

[36] Harry Joe. *Dependence Modeling with Copulas*. Ed. by F. Bunea, V. Isham, N. Keiding, T. Louis, R.L. Smith, and H. Tong. Boca Raton, FL: CRC Press, 2014. ISBN: 978-1-4665-8323-8.

[37] H. Tsukahara. “Semiparametric estimation in copula models”. In: *The Canadian Journal of Statistics* 33 (3 2005), pp. 357–375.

7 Proof

7.1 Bahadur representation of $\hat{C}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u}) - C_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})$

This Bahadur representation has been given in [18].

Lemma 5 (Theorem 1 in [18]). *Under some regularity conditions,*

$$\sqrt{nh^p}\left(\hat{C}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u}) - C_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})\right) = \frac{1}{\sqrt{nh^p}}\sum_{i=1}^{n} w_i(\mathbf{x},h)\xi_i(\mathbf{x},\mathbf{u}),$$

where

$$\xi_i(\mathbf{x},\mathbf{u}) := \mathbb{1}\left\{\mathbf{Y}_i \leq \overrightarrow{F^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\} - C_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u}) - \sum_{j=1}^{d}\frac{\partial C_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})}{\partial u_j} \times \left(\mathbb{1}\left\{Y_{i,j} \leq F^{-1}_{Y_j|\mathbf{X}=\mathbf{x}}(v)\right\} - u_j\right).$$

7.2 Bahadur representation of $\hat{C}^*_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u}) - \hat{C}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})$

We remark that the bootstrapped joint conditional distribution function and the bootstrapped marginal conditional distribution functions can be rewritten as

$$\hat{F}^*_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{y}) = \frac{1}{nh^p}\sum_{i=1}^{n} M_{n,i}w_i(\mathbf{x},h)\mathbb{1}\left\{\mathbf{Y} \leq \mathbf{y}\right\},$$
$$\hat{F}^*_{Y_i|\mathbf{X}=\mathbf{x}}(y_i) = \frac{1}{nh^p}\sum_{i=1}^{n} M_{n,i}w_i(\mathbf{x},h)\mathbb{1}\left\{Y_i \leq y_i\right\},$$

where $M_{n,i}$ is the number of times that X_i is "redrawn" from the original sample. These $M_{n,i}$ are not independent however. Our line of proof follows the construction detailed in [23, Chapter 3.6] where the dependence between the $M_{n,i}$ is removed by Poissonization. They also show that the Poissonized process and the ordinary bootstrap process are asymptotically equivalent in the case of the standard empirical process. The construction is the following: instead of a sample size n, take a random N_n number of replicates, where N_n follows a Poisson distribution with mean n and is independent of the original observations. Then $M_{N_n,1}, \cdots, M_{N_n,n}$ are i.i.d. Poisson variables with mean 1.

We defined the Poissonized versions of the bootstrapped joint conditional distribution function (respectively, bootstrapped marginal conditional distribution functions, bootstrapped conditional copula) by

$$\hat{F}^{*,Poiss}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{y}) = \frac{1}{nh^p}\sum_{i=1}^{n} M_{N_n,i}w_i(\mathbf{x},h)\mathbb{1}\left\{\mathbf{Y} \leq \mathbf{y}\right\},$$
$$\hat{F}^{*,Poiss}_{Y_i|X=x}(y_i) = \frac{1}{nh^p}\sum_{i=1}^{n} M_{N_n,i}w_i(\mathbf{x},h)\mathbb{1}\left\{Y_i \leq y_i\right\},$$
$$\hat{C}^{*,Poiss}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u}) := \hat{F}^{*,Poiss}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\left(\overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right).$$

We can now do the decomposition

$$\begin{aligned}&\sqrt{nh^p}\left(\hat{C}^*_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u}) - \hat{C}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})\right)\\ =&\sqrt{nh^p}\left(\hat{C}^*_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u}) - \hat{C}^{*,Poiss}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})\right) + \sqrt{nh^p}\left(\hat{C}^{*,Poiss}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u}) - \hat{C}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})\right) \qquad (6)\end{aligned}$$

Lemma 6. *Under some regularity conditions,*

$$E\left[\int_{\mathbf{x},\mathbf{u}} nh^p\left(\hat{C}^*_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u}) - \hat{C}^{*,Poiss}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})\right)^2 d\mathbf{x}d\mathbf{u}\right] \to 0,$$

as $n \to \infty$.

Lemma 7. *Under some regularity conditions,*

$$E\left[\int_{\mathbf{x},\mathbf{u}}\left(\sqrt{nh^p}\left(\hat{C}^{*,Poiss}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u}) - \hat{C}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})\right) - \frac{1}{\sqrt{nh^p}}\sum_{i=1}^n w_i(\mathbf{x},h)\xi_i^*(\mathbf{x},\mathbf{u})\right)^2\right] \to 0$$

as $n \to \infty$, *where*

$$\begin{aligned}\xi_i^*(\mathbf{x},\mathbf{u}) := &(M_{N_n,i}-1)\left(\mathbb{1}\left\{\mathbf{Y}_i \leq \overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\} - P\left(\mathbf{Y}_i \leq \overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right)\right)\\ &-\sum_{j=1}^d \frac{\partial C_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})}{\partial u_j} \times (M_{N_n,i}-1)\left(\mathbb{1}\left\{Y_{i,j} \leq \hat{F}^{-1}_{Y_j|\mathbf{X}=\mathbf{x}}(v)\right\} - u_j\right).\end{aligned}$$

Combining both of these lemmas, we get the following Bahadur representation

$$\sqrt{nh^p}\left(\hat{C}^*_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u}) - \hat{C}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})\right) = \sqrt{nh^p}\sum_{i=1}^n w_i(\mathbf{x},h)\xi_i^*(\mathbf{x},\mathbf{u}) + \varepsilon_n^*(\mathbf{x},\mathbf{u}), \tag{7}$$

where $\varepsilon_n^*(\mathbf{x},\mathbf{u}) := \sqrt{nh^p}\left(\hat{C}^*_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u}) - \hat{C}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})\right) - \sqrt{nh^p}\sum_{i=1}^n w_i(\mathbf{x},h)\xi_i^*(\mathbf{x},\mathbf{u})$ converges to 0 in L_2-norm.

We will treat these two parts seperately. First it will be shown that the first term (the Poissonization approximation) is asymptotically negligible as is done by Van der Vaart and Wellner [23]. Then it will be shown that we can rewrite the second term so that we can get the wanted Bahadur representation.

Proof of Lemma 6: justification for the Poissonization

We want to show that $\sqrt{nh^p}\left(\hat{C}^*_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u}) - \hat{C}^{*,Poiss}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})\right)$ is asymptotically negligible. For this, note that we can write the following

$$\begin{aligned}A_0 :=&\sqrt{nh^p}\left(\hat{C}^*_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u}) - \hat{C}^{*,Poiss}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})\right)\\ =&\sqrt{nh^p}\left(\hat{F}^*_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\left(\overrightarrow{\hat{F}^{*,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right) - \hat{F}^{*,Poiss}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\left(\overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right)\right)\\ =&\sqrt{nh^p}\left(\frac{1}{nh^p}\sum_{i=1}^n M_{n,i}w_i(\mathbf{x},h)\mathbb{1}\left\{\mathbf{Y}_i \leq \overrightarrow{\hat{F}^{*,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\} - \frac{1}{nh^p}\sum_{i=1}^n M_{N_n,i}w_i(\mathbf{x},h)\mathbb{1}\left\{\mathbf{Y}_i \leq \overrightarrow{\hat{F}^{*,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\}\right.\\ &\left.+\frac{1}{nh^p}\sum_{i=1}^n M_{N_n,i}w_i(\mathbf{x},h)\mathbb{1}\left\{\mathbf{Y}_i \leq \overrightarrow{\hat{F}^{*,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\} - \frac{1}{nh^p}\sum_{i=1}^n M_{N_n,i}w_i(\mathbf{x},h)\mathbb{1}\left\{\mathbf{Y}_i \leq \overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\}\right)\\ =&\frac{1}{\sqrt{nh^p}}\sum_{i=1}^n (M_{n,i}-M_{N_n,i})w_i(\mathbf{x},h)\mathbb{1}\left\{\mathbf{Y}_i \leq \overrightarrow{\hat{F}^{*,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\}\\ &+\frac{1}{\sqrt{nh^p}}\sum_{i=1}^n M_{N_n,i}w_i(\mathbf{x},h)\left(\mathbb{1}\left\{\mathbf{Y}_i \leq \overrightarrow{\hat{F}^{*,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\} - \mathbb{1}\left\{Y_i \leq \overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\}\right)\end{aligned} \tag{8}$$

Bound on the first term of Equation 8. We will follow the direct calculation of Van der Vaart and Wellner [23, p.348]. For any $\epsilon > 0$ there exists a sequence of integers l_n, with $l_n = O(\sqrt{n})$ such that $\mathrm{P}(|N_n - n| \geq l_n) \leq \epsilon$ for every n

$$\mathrm{P}\Big(\max_{1\leq i\leq n} |M_{n,i} - M_{N_n,i}| > 2\Big) \leq \epsilon + n\mathrm{P}\Big(bin(l_n, n^{-1}) \geq 2\Big) \to \epsilon$$

Thus for sufficiently large n, $|M_{N_n} - M_n|$ is 0, 1, or 2 with probability at least $1 - 2\epsilon$ for all i. We can then write

$$|M_{N_n,i} - M_{n,i}| = \sum_{j=1}^{\infty} \mathbb{1}\left\{|M_{N_n,i} - M_{n,i}| \geq j\right\}$$

Thus

$$\frac{1}{nh^p}\sum_{i=1}^{n}(M_{N_n,i} - M_{n,i})R = sign(N-n)\sum_{j=1}^{\infty}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} R$$

If F is a Glivenko-Cantelli class

$$\mathrm{P}\left(\left|\frac{1}{\sqrt{nh^p}}\sum_{i=1}^{n}\left(M_{n,i} - M_{N_n,i}\right)w_i(\mathbf{x},h)\mathbb{1}\left\{\mathbf{Y}_i \leq \overrightarrow{\hat{F}^{*,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\}\right|_F^* > \epsilon\right) \to 0$$

as $n \to \infty$, given almost all sequences $X_1, X_2, ...$, for every $\epsilon > 0$

Bound on the second term of Equation 8.

$$\phi(x,\mathbf{u}) := \frac{1}{\sqrt{nh^p}}\sum_{i=1}^{n} M_{N_n,i}w_i(\mathbf{x},h)\left(\mathbb{1}\left\{\mathbf{Y}_i \leq \overrightarrow{\hat{F}^{*,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\} - \mathbb{1}\left\{\mathbf{Y}_i \leq \overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\}\right)$$

Instead of showing that this term goes to 0, it is sufficient to show that the expectation of the square of this term goes to 0. (L2 convergence)

We have the decomposition (we can inverse the sums, integrals and expectations as we want by Fubini's theorem)

$$\begin{aligned} E\left[\left(\int \phi(\mathbf{x},\mathbf{u})d\mathbf{x}d\mathbf{u}\right)^2\right] &= E\left[\left(\int\sum_{i=1}^{n}\phi(i,\mathbf{x},\mathbf{u})d\mathbf{x}d\mathbf{u}\right)^2\right] \\ &= \sum_{i=1}^{n}\int E\Big[\phi(i,\mathbf{x},\mathbf{u})\phi(i,\mathbf{s},\mathbf{v})\Big]d\mathbf{x}d\mathbf{s}d\mathbf{u}d\mathbf{v} + \sum_{i\neq j}\int E\Big[\phi(i,\mathbf{x},\mathbf{u})\phi(j,\mathbf{s},\mathbf{v})\Big]d\mathbf{x}d\mathbf{s}d\mathbf{u}d\mathbf{v} \end{aligned} \tag{9}$$

Bounding the indicators by 1, the first term of Equation 9 becomes

$$\begin{aligned} &\left|\sum_{i=1}^{n}\int_{\mathbf{x},\mathbf{s},\mathbf{u},\mathbf{v}} E\Big[\phi(i,\mathbf{x},\mathbf{u})\phi(i,\mathbf{s},\mathbf{v})\Big]d\mathbf{x}d\mathbf{y}d\mathbf{u}d\mathbf{v}\right| \\ &\leq \sum_{i=1}^{n}\int_{\mathbf{x},\mathbf{s}} E\left|\frac{1}{\sqrt{nh^p}}(M_{N_n,i})^2 w_i(\mathbf{x},h)w_i(\mathbf{s},h)\right|d\mathbf{x}d\mathbf{y} \\ &\leq \frac{1}{h^p}E\big[(M_{N_n,1})^2\big]E\left|\int_{\mathbf{x},\mathbf{s}} w_1(\mathbf{x},h)w_1(\mathbf{s},h)d\mathbf{x}d\mathbf{y}\right|, \end{aligned}$$

where in the last line, we use the fact that the observations are identically distributed and that the $M_{N_n,i}$ are independent from the (X_i). We furthermore know that

$$E\big[(M_{N_n,i})^2\big] = Var\big[M_{N_n,i}\big] + E\big[M_{N_n,i}\big]^2 = 2$$

In practice, we use kernel-based weights, meaning that $w_1(\mathbf{x},h) = K\Big(\frac{X_1-\mathbf{x}}{h}\Big)$, so we can do in each integral the change of variable $\mathbf{z} = \frac{X_1-\mathbf{x}}{h}$, $\mathbf{t} = \frac{X_1-\mathbf{s}}{h}$, $d\mathbf{x} = hdz$, $d\mathbf{y} = hd\mathbf{t}$ and therefore we get

$$\begin{aligned}
&\frac{2}{h^p}E\bigg|\int_{\mathbf{x},\mathbf{s}} w_{1,n}(\mathbf{x},h)w_{1,n}(\mathbf{s},h)d\mathbf{x}d\mathbf{s}\bigg| \\
&= \frac{2}{h^p}E\bigg|\int_{\mathbf{z},\mathbf{t}} K(\mathbf{z})K(\mathbf{t})hd\mathbf{z}\ hd\mathbf{t}\bigg| \\
&= 2h^p\bigg(\int K\bigg)^2 \to 0.
\end{aligned}$$

Second term of Equation 9.
The second term of Equation 9 is more difficult to bound. We start by decomposing it in the following way.

$$\begin{aligned}
A_{0,1} &:= \frac{n(n-1)}{nh^p}\int_{\mathbf{x},\mathbf{s},\mathbf{u},\mathbf{v}} E\Bigg[w_i(\mathbf{x},h)w_j(\mathbf{s},h)\Bigg(\mathbb{1}\left\{\mathbf{Y}_i \le \overrightarrow{\hat{F}^{*,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\} - \mathbb{1}\left\{\mathbf{Y}_i \le \overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\}\Bigg) \\
&\qquad \times\Bigg(\mathbb{1}\left\{\mathbf{Y}_j \le \overrightarrow{\hat{F}^{*,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v})\right\} - \mathbb{1}\left\{\mathbf{Y}_j \le \overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v})\right\}\Bigg)\Bigg]d\mathbf{x}d\mathbf{s}d\mathbf{u}d\mathbf{v}, \\
&\le A_{0,1,1} + A_{0,1,2},
\end{aligned}$$

where

$$\begin{aligned}
A_{0,1,1} &:= \frac{n}{h^p}\int_{\mathbf{x},\mathbf{s},\mathbf{u},\mathbf{v}} E\Bigg[w_i(\mathbf{x},h)w_j(\mathbf{s},h)\Bigg(\mathbb{1}\left\{\mathbf{Y}_i \le \overrightarrow{\hat{F}^{*,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\} - \mathbb{1}\left\{\mathbf{Y}_i \le \overrightarrow{\hat{F}^{*,Poiss,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\}\Bigg) \\
&\qquad \times\Bigg(\mathbb{1}\left\{\mathbf{Y}_j \le \overrightarrow{\hat{F}^{*,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v})\right\} - \mathbb{1}\left\{\mathbf{Y}_j \le \overrightarrow{\hat{F}^{*,Poiss,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v})\right\}\Bigg)\Bigg]d\mathbf{x}d\mathbf{s}d\mathbf{u}d\mathbf{v}, \\
A_{0,1,2} &:= A_{0,1} - A_{0,1,1},
\end{aligned}$$

and

$$\overrightarrow{\hat{F}^{*,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u}) := \big(\hat{F}^{*,-1,-(i,j)}_{Y_1|\mathbf{X}=\mathbf{x}}(u_1),\cdots,\hat{F}^{*,-1,-(i,j)}_{Y_d|\mathbf{X}=\mathbf{x}}(u_d)\big),$$

and $\hat{F}^{*,-1,-(i,j)}_{Y_k|\mathbf{X}=\mathbf{x}}$ denotes the inverse of the estimated conditional distribution function of Y_k given $X = x$ based on the sample $\{1,2,\dots,n\}\backslash\{i,j\}$. The Poissonized version are defined similarly.

Bounding $A_{0,1,1}$.
We use the equality $E[X] = E[E[X|Z]]$.

$$\begin{aligned}
A_{0,1,1} &:= \frac{n}{h^p}\int_{\mathbf{x},\mathbf{s},\mathbf{u},\mathbf{v}} E\Bigg[w_i(\mathbf{x},h)w_j(\mathbf{s},h)\Bigg(\mathbb{1}\left\{\mathbf{Y}_i \le \overrightarrow{\hat{F}^{*,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\} - \mathbb{1}\left\{\mathbf{Y}_i \le \overrightarrow{\hat{F}^{*,Poiss,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\}\Bigg) \\
&\qquad \times\Bigg(\mathbb{1}\left\{\mathbf{Y}_j \le \overrightarrow{\hat{F}^{*,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v})\right\} - \mathbb{1}\left\{\mathbf{Y}_j \le \overrightarrow{\hat{F}^{*,Poiss,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v})\right\}\Bigg)\Bigg]d\mathbf{x}d\mathbf{s}d\mathbf{u}d\mathbf{v}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{n}{h^p}\int_{\mathbf{x},\mathbf{s},\mathbf{u},\mathbf{v}} E\Bigg[w_i(\mathbf{x},h)w_j(\mathbf{s},h)\Bigg(\mathbb{1}\left\{\mathbf{Y}_i \le \overrightarrow{\hat{F}^{*,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\}\mathbb{1}\left\{\mathbf{Y}_j \le \overrightarrow{\hat{F}^{*,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v})\right\}\\
&\qquad - \mathbb{1}\left\{\mathbf{Y}_i \le \overrightarrow{\hat{F}^{*,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\}\mathbb{1}\left\{\mathbf{Y}_j \le \overrightarrow{\hat{F}^{*,Poiss,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v})\right\}\\
&\qquad - \mathbb{1}\left\{\mathbf{Y}_i \le \overrightarrow{\hat{F}^{*,Poiss,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\}\mathbb{1}\left\{\mathbf{Y}_j \le \overrightarrow{\hat{F}^{*,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v})\right\}\\
&\qquad + \mathbb{1}\left\{\mathbf{Y}_i \le \overrightarrow{\hat{F}^{*,Poiss,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\}\mathbb{1}\left\{\mathbf{Y}_j \le \overrightarrow{\hat{F}^{*,Poiss,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v})\right\}\Bigg)\Bigg]d\mathbf{x}d\mathbf{s}d\mathbf{u}d\mathbf{v}\\
&= \frac{n}{h^p}\int_{\mathbf{x},\mathbf{s},\mathbf{u},\mathbf{v}} E\Bigg[w_i(\mathbf{x},h)w_j(\mathbf{s},h)\Bigg(\mathbb{P}\left[\mathbf{Y}_i \le \overrightarrow{\hat{F}^{*,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u}) \cap \mathbf{Y}_j \le \overrightarrow{\hat{F}^{*,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v})\middle| X_1,\dots X_n; Y_l, l\notin\{i,j\}\right]\\
&\qquad - \mathbb{P}\left[\mathbf{Y}_i \le \overrightarrow{\hat{F}^{*,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u}) \cap \mathbf{Y}_j \le \overrightarrow{\hat{F}^{*,Poiss,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v})\middle| X_1,\dots X_n; Y_l, l\notin\{i,j\}\right]\\
&\qquad - \mathbb{P}\left[\mathbf{Y}_i \le \overrightarrow{\hat{F}^{*,Poiss,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u}) \cap \mathbf{Y}_j \le \overrightarrow{\hat{F}^{*,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v})\middle| X_1,\dots X_n; Y_l, l\notin\{i,j\}\right]\\
&\qquad + \mathbb{P}\left[\mathbf{Y}_i \le \overrightarrow{\hat{F}^{*,Poiss,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u}) \cap \mathbf{Y}_j \le \overrightarrow{\hat{F}^{*,Poiss,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v})\middle| X_1,\dots X_n; Y_l, l\notin\{i,j\}\right]\Bigg)\Bigg]d\mathbf{x}d\mathbf{s}d\mathbf{u}d\mathbf{v}\\
&= \frac{n}{h^p}\int_{\mathbf{x},\mathbf{s},\mathbf{u},\mathbf{v}} E\Bigg[w_i(\mathbf{x},h)w_j(\mathbf{s},h)\Bigg(F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\overrightarrow{\hat{F}^{*,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u}))F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\overrightarrow{\hat{F}^{*,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v}))\\
&\qquad - F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\overrightarrow{\hat{F}^{*,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u}))F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\overrightarrow{\hat{F}^{*,Poiss,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v}))\\
&\qquad - F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\overrightarrow{\hat{F}^{*,Poiss,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u}))F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\overrightarrow{\hat{F}^{*,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v}))\\
&\qquad + F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\overrightarrow{\hat{F}^{*,Poiss,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u}))F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\overrightarrow{\hat{F}^{*,Poiss,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v}))\Bigg)\Bigg]d\mathbf{x}d\mathbf{s}d\mathbf{u}d\mathbf{v}.
\end{aligned}
$$

We then use the fact that $A_{0,1,1}$ does not depend on (i,j) and factor to get

$$
\begin{aligned}
A_{0,1,1} &= \frac{n}{h^p}\int_{\mathbf{x},\mathbf{s},\mathbf{u},\mathbf{v}} E\Bigg[w_1(\mathbf{x},h)w_2(\mathbf{s},h)\Bigg(F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\overrightarrow{\hat{F}^{*,-1,-(1,2)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})) - F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\overrightarrow{\hat{F}^{*,Poiss,-1,-(1,2)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u}))\Bigg)\\
&\qquad \times \Bigg(F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\overrightarrow{\hat{F}^{*,-1,-(1,2)}_{\mathbf{Y}|X=s}}(\mathbf{v})) - F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\overrightarrow{\hat{F}^{*,Poiss,-1,-(1,2)}_{\mathbf{Y}|X=s}}(\mathbf{v}))\Bigg)\Bigg]d\mathbf{x}d\mathbf{s}d\mathbf{u}d\mathbf{v}\\
&= \frac{n}{h^p}\int_{\mathbf{x},\mathbf{s},\mathbf{u},\mathbf{v}} E\Bigg[w_1(\mathbf{x},h)w_2(\mathbf{s},h)\nabla F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(Q(\mathbf{u}))\cdot\Bigg(\overrightarrow{\hat{F}^{*,-1,-(1,2)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u}) - \overrightarrow{\hat{F}^{*,Poiss,-1,-(1,2)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\Bigg)\\
&\qquad \times \nabla F_{\mathbf{Y}|X=s}(Q(\mathbf{v}))\cdot\Bigg(\overrightarrow{\hat{F}^{*,-1,-(1,2)}_{\mathbf{Y}|X=s}}(\mathbf{v}) - \overrightarrow{\hat{F}^{*,Poiss,-1,-(1,2)}_{\mathbf{Y}|X=s}}(\mathbf{v})\Bigg)\Bigg]d\mathbf{x}d\mathbf{s}d\mathbf{u}d\mathbf{v},
\end{aligned}
$$

by the mean value theorem in several variables.
Using first the inequality $x\cdot y \le |x|_\infty |y|_1$ and then Cauchy-Schwarz inequality, we get

$$
\begin{aligned}
|A_{0,1,1}| &\le \frac{n}{h^p}\int_{\mathbf{x},\mathbf{s},\mathbf{u},\mathbf{v}} E\Bigg[w_1(\mathbf{x},h)\left|\nabla F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(Q(\mathbf{u}))\right|_\infty \left|\overrightarrow{\hat{F}^{*,-1,-(1,2)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u}) - \overrightarrow{\hat{F}^{*,Poiss,-1,-(1,2)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right|_1\\
&\qquad \times w_2(\mathbf{s},h)\left|\nabla F_{\mathbf{Y}|X=s}(Q(\mathbf{v}))\right|_\infty \left|\overrightarrow{\hat{F}^{*,-1,-(1,2)}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v}) - \overrightarrow{\hat{F}^{*,Poiss,-1,-(1,2)}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v})\right|_1\Bigg]d\mathbf{x}d\mathbf{s}d\mathbf{u}d\mathbf{v}
\end{aligned}
$$

$$\lesssim \frac{n}{h^p}\int_{\mathbf{x},\mathbf{s}} E\Bigg[w_1(\mathbf{x},h)\Big(\sum_{k=1}^{d}\Big|\int_{u_k}(\hat{F}^{*,-1,-(1,2)}_{Y_k|\mathbf{X}=\mathbf{x}}(u_k)-\hat{F}^{*,Poiss,-1,-(1,2)}_{Y_k|\mathbf{X}=\mathbf{x}}(u_k))du_i\Big|\Big)$$
$$\times w_2(\mathbf{s},h)\Big(\sum_{l=1}^{d}\Big|\int_{v_l}(\hat{F}^{*,-1,-(1,2)}_{Y_l|X=s}(v_l)-\hat{F}^{*,Poiss,-1,-(1,2)}_{Y_l|X=s}(v_l))dv_l\Big|\Big)\Bigg]d\mathbf{x}d\mathbf{s}$$
$$\lesssim \frac{n}{h^p}\int_{x,s} E\Bigg[w_1(\mathbf{x},h)\Big(\sum_{k=1}^{d}\Big|\int_{y_k}(\hat{F}^{*,-(1,2)}_{Y_k|\mathbf{X}=\mathbf{x}}(y_k)-\hat{F}^{*,Poiss,-(1,2)}_{Y_k|\mathbf{X}=\mathbf{x}}(y_k))du_i\Big|\Big)$$
$$\times w_2(\mathbf{s},h)\Big(\sum_{l=1}^{d}\Big|\int_{z_i l}(\hat{F}^{*,-(1,2)}_{Y_l|X=s}(y_l)-\hat{F}^{*,Poiss,-(1,2)}_{Y_l|X=s}(y_l))dv_l\Big|\Big)\Bigg]d\mathbf{x}d\mathbf{s},$$

where in the last line, we use the fact that for two cdfs F_a and F_b, $\int_u |F_a^{-1}(u)-F_b^{-1}(u)|du = \int_y |F_a(y)-F_b(y)|du$. We can then compute these integrals explicitly.

$$|A_{0,1,1}| \lesssim \frac{n}{h^p}\sum_{k,l=1}^{d}\int_{\mathbf{x},\mathbf{s}} E\Bigg[w_{1,n}(\mathbf{x},h)\Big(\sum_{i_1=1}^{n-3}(Y_{(i_1+1),k}-Y_{(i_1),k})\frac{1}{nh^p}\Big|\sum_{j_1=1}^{i_1+1}w_{(j_1),n}M_{N_n,(j_1)}-M_{n,(j_1)}\Big|\Big)$$
$$\times w_{2,n}(\mathbf{s},h)\Big(\sum_{i_2=1}^{n-3}(Y_{(i_2+1),k}-Y_{(i_2),k})\frac{1}{nh^p}\Big|\sum_{j_2=1}^{i_2+1}w_{(j_2),n}M_{N_n,(j_2)}-M_{n,(j_2)}\Big|\Big)\Bigg]d\mathbf{x}d\mathbf{s}$$
$$\lesssim \frac{n}{h^p}\sum_{k,l=1}^{d}\int_{\mathbf{x},\mathbf{s}} E\Bigg[w_{1,n}(\mathbf{x},h)\Big(\sum_{i_1=1}^{n-3}(Y_{(i_1+1),k}-Y_{(i_1),k})\frac{1}{n}\Big|\sum_{j_1=1}^{i_1+1}M_{N_n,(j_1)}-M_{n,(j_1)}\Big|\Big)$$
$$\times w_{2,n}(\mathbf{s},h)\Big(\sum_{i_2=1}^{n-3}(Y_{(i_2+1),k}-Y_{(i_2),k})\frac{1}{n}\Big|\sum_{j_2=1}^{i_2+1}M_{N_n,(j_2)}-M_{n,(j_2)}\Big|\Big)\Bigg]d\mathbf{x}d\mathbf{s}$$
$$\lesssim \frac{n}{h^p}\sum_{k,l=1}^{d}\int_{\mathbf{x},\mathbf{s}} E\Bigg[w_{1,n}(\mathbf{x},h)\Big(\sum_{i_1=1}^{n-3}(Y_{(i_1+1),k}-Y_{(i_1),k})\frac{|N_n-n|}{n}\Big)w_{2,n}(\mathbf{s},h)\Big(\sum_{i_2=1}^{n-3}(Y_{(i_2+1),k}-Y_{(i_2),k})\frac{|N_n-n|}{n}\Big)\Bigg]d\mathbf{x}d\mathbf{s}$$
$$\lesssim \frac{1}{h^p}\sum_{k,l=1}^{d}\int_{\mathbf{x},\mathbf{s}} E\Big[w_{1,n}(\mathbf{x},h)w_{2,n}(\mathbf{s},h)\Big]d\mathbf{x}d\mathbf{s}$$
$$\lesssim \frac{1}{h^p}\int_{\mathbf{x},\mathbf{s}} E\Big[w_{1,n}(\mathbf{x},h)w_{2,n}(\mathbf{s},h)\Big]d\mathbf{x}d\mathbf{s}$$
$$\lesssim h^p.$$

Bounding $A_{0,1,2}$

$$A_{0,1,2} := \frac{n}{h^p}\int_{\mathbf{x},\mathbf{s},\mathbf{u},\mathbf{v}} E\Bigg[w_i(\mathbf{x},h)w_j(\mathbf{s},h)\Bigg(\Big(\mathbb{1}\Big\{\mathbf{Y}_i\le \overrightarrow{\hat{F}^{*,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\Big\}-\mathbb{1}\Big\{\mathbf{Y}_i\le \overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\Big\}\Big)$$
$$\times\Big(\mathbb{1}\Big\{\mathbf{Y}_j\le \overrightarrow{\hat{F}^{*,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v})\Big\}-\mathbb{1}\Big\{\mathbf{Y}_j\le \overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v})\Big\}\Big)$$
$$-\Big(\mathbb{1}\Big\{\mathbf{Y}_i\le \overrightarrow{\hat{F}^{*,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\Big\}-\mathbb{1}\Big\{\mathbf{Y}_i\le \overrightarrow{\hat{F}^{*,Poiss,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\Big\}\Big)$$
$$\times\Big(\mathbb{1}\Big\{\mathbf{Y}_j\le \overrightarrow{\hat{F}^{*,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v})\Big\}-\mathbb{1}\Big\{\mathbf{Y}_j\le \overrightarrow{\hat{F}^{*,Poiss,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v})\Big\}\Big)\Bigg)\Bigg]d\mathbf{x}d\mathbf{s}d\mathbf{u}d\mathbf{v}.$$

We first apply Fubini's theorem and use the linearity of the integral, so that

$$\begin{aligned} A_{0,1,2} = \frac{n}{h^p}\int_{\mathbf{x},\mathbf{s}} E\Bigg[w_i(\mathbf{x},h)w_j(\mathbf{s},h)\Bigg(&\Bigg(\int_{\mathbf{u}} \mathbb{1}\left\{\mathbf{Y}_i \le \overrightarrow{\hat{F}^{*,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\} d\mathbf{u} - \int_{\mathbf{u}} \mathbb{1}\left\{\mathbf{Y}_i \le \overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\} d\mathbf{u}\Bigg) \\ &\times \Bigg(\int_{\mathbf{v}} \mathbb{1}\left\{\mathbf{Y}_j \le \overrightarrow{\hat{F}^{*,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v})\right\} d\mathbf{v} - \int_{\mathbf{v}} \mathbb{1}\left\{\mathbf{Y}_j \le \overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v})\right\} d\mathbf{v}\Bigg) \\ &- \Bigg(\int_{\mathbf{u}} \mathbb{1}\left\{\mathbf{Y}_i \le \overrightarrow{\hat{F}^{*,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\} d\mathbf{u} - \int_{\mathbf{u}} \mathbb{1}\left\{\mathbf{Y}_i \le \overrightarrow{\hat{F}^{*,Poiss,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\} d\mathbf{u}\Bigg) \\ &\times \Bigg(\int_{\mathbf{v}} \mathbb{1}\left\{\mathbf{Y}_j \le \overrightarrow{\hat{F}^{*,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v})\right\} d\mathbf{v} - \int_{\mathbf{v}} \mathbb{1}\left\{\mathbf{Y}_j \le \overrightarrow{\hat{F}^{*,Poiss,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v})\right\} d\mathbf{v}\Bigg)\Bigg)\Bigg] d\mathbf{x}d\mathbf{s}. \end{aligned}$$

Note that, for any $\mathbf{Y}_i$ and any $\mathbf{x}$,

$$\begin{aligned} \int_{\mathbf{u}} \mathbb{1}\left\{\mathbf{Y}_i \le \overrightarrow{\hat{F}^{*,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\} d\mathbf{u} &= \int_{u_1,\ldots,u_d} \mathbb{1}\left\{Y_{i,1} \le \hat{F}^{*,-1}_{Y_1|\mathbf{X}=\mathbf{x}}(u_1),\ldots,Y_{i,d} \le \hat{F}^{*,-1}_{Y_d|\mathbf{X}=\mathbf{x}}(u_d)\right\} du_1\cdots du_d \\ &= \prod_{k=1}^{d}\int_{u_k} \mathbb{1}\left\{Y_{i,k} \le \hat{F}^{*,-1}_{Y_k|\mathbf{X}=\mathbf{x}}(u_k)\right\} du_k \\ &= \prod_{k=1}^{d}\left(\sum_{l_k=1}^{n} \mathbb{1}\left\{Y_{i,k} \le Y_{(l_k),k}\right\} \frac{1}{nh^p} w_{(l_k),n}(\mathbf{x},h) M_{n,(l_k)}\right) \\ &= \prod_{k=1}^{d}\left(\sum_{l_k=1}^{n} \mathbb{1}\left\{Y_{i,k} \le Y_{l_k,k}\right\} \frac{1}{nh^p} w_{l_k,n}(\mathbf{x},h) M_{n,l_k}\right) \\ &= \sum_{1\le l_1,\ldots,l_d\le n}\prod_{k=1}^{d} \mathbb{1}\left\{Y_{i,k} \le Y_{l_k,k}\right\} \frac{1}{nh^p} w_{l_k,n}(\mathbf{x},h) M_{n,l_k}. \end{aligned} \tag{10}$$

Therefore,

$$\begin{aligned} A_{0,1,2} = \frac{n}{h^p}\int_{\mathbf{x},\mathbf{s}} E\Bigg[w_i(\mathbf{x},h)w_j(\mathbf{s},h)\Bigg(&\sum_{1\le l_1,\ldots,l_d\le n}\prod_{k=1}^{d} \mathbb{1}\left\{Y_{i,k} \le Y_{l_k,k}\right\} \frac{1}{nh^p} w_{l_k,n}(\mathbf{x},h)\big(M_{n,l_k} - M_{N_n,l_k}\big) \\ &\times \sum_{1\le l'_1,\ldots,l'_d\le n}\prod_{k'=1}^{d} \mathbb{1}\left\{Y_{i,k'} \le Y_{l'_{k'},k'}\right\} \frac{1}{nh^p} w_{l'_{k'},n}(\mathbf{s},h)\big(M_{n,l_k} - M_{N_n,l'_{k'}}\big) \\ &- \sum_{1\le l_1,\ldots,l_d\le n,\ne i,j}\prod_{k=1}^{d} \mathbb{1}\left\{Y_{i,k} \le Y_{l_k,k}\right\} \frac{1}{nh^p} w_{l_k,n}(\mathbf{x},h)\big(M_{n,l_k} - M_{N_n,l_k}\big) \\ &\times \sum_{1\le l'_1,\ldots,l'_d\le n,\ne i,j}\prod_{k'=1}^{d} \mathbb{1}\left\{Y_{i,k'} \le Y_{l'_{k'},k'}\right\} \frac{1}{nh^p} w_{l'_{k'},n}(\mathbf{s},h)\big(M_{n,l_k} - M_{N_n,l'_{k'}}\big)\Bigg)\Bigg] d\mathbf{x}d\mathbf{s} \end{aligned}$$

$$= \frac{n}{h^p}\int_{\mathbf{x},\mathbf{s}} E\Bigg[w_i(\mathbf{x},h)w_j(\mathbf{s},h)$$

$$\times \sum_{\substack{1\le l_1,\dots,l_d\le n\\ 1\le l'_1,\dots,l'_d\le n\\ \text{at least one of them is } i \text{ or } j}} \prod_{k=1}^{d}\prod_{k'=1}^{d} \mathbf{1}\left\{Y_{i,k}\le Y_{l_k,k}\right\}\frac{1}{nh^p}w_{l_k,n}(\mathbf{x},h)\big(M_{n,l_k}-M_{N_n,l_k}\big)$$

$$\times\, \mathbf{1}\left\{Y_{i,k'}\le Y_{l'_{k'},k'}\right\}\frac{1}{nh^p}w_{l'_{k'},n}(\mathbf{s},h)\big(M_{n,l_k}-M_{N_n,l'_{k'}}\big)\Bigg]d\mathbf{x}d\mathbf{s}$$

$$= \sum_{q=0}^{2d-1} A_{0,2,q},$$

where for $q\in\{0,\dots,2d-1\}$,

$$A_{0,1,2,q} := \frac{n}{h^p}\sum_{\substack{1\le l_1,\dots,l_d\le n\\ 1\le l'_1,\dots,l'_d\le n\\ \text{at least one of them is } i \text{ or } j\\ Card(\{l_1,\dots,l_d,l'_1,\dots,l'_d\}\backslash\{i,j\})=q}} \int_{x,s} E\Bigg[w_i(\mathbf{x},h)w_j(s,h)\prod_{k=1}^{d}\prod_{k'=1}^{d}\mathbf{1}\left\{Y_{i,k}\le Y_{l_k,k}\right\}$$

$$\times \frac{1}{nh^p}w_{l_k,n}(x,h)\big(M_{n,l_k}-M_{N_n,l_k}\big)\mathbf{1}\left\{Y_{i,k'}\le Y_{l'_{k'},k'}\right\}\frac{1}{nh^p}w_{l'_{k'},n}(\mathbf{s},h)\big(M_{n,l_k}-M_{N_n,l'_{k'}}\big)\Bigg]d\mathbf{x}d\mathbf{s}$$

Thus

$$\big|A_{0,1,2,q}\big| \cdot \frac{n}{h^p}\sum_{\substack{1\le l_1,\dots,l_d\le n\\ 1\le l'_1,\dots,l'_d\le n\\ \text{at least one of them is } i \text{ or } j\\ Card(\{l_1,\dots,l_d,l'_1,\dots,l'_d\}\backslash\{i,j\})=q}} \int_{\mathbf{x},\mathbf{s}} E\Bigg[w_i(\mathbf{x},h)w_j(\mathbf{s},h)\prod_{k=1}^{d}\prod_{k'=1}^{d}\frac{1}{nh^p}w_{l_k,n}(\mathbf{x},h)\frac{1}{nh^p}w_{l'_{k'},n}(\mathbf{s},h)\Bigg]d\mathbf{x}d\mathbf{s}$$

In the sum, there are $O(n^q)$ terms, and there exists q indices $\widetilde{l}_1,\dots,\widetilde{l}_q$ such that all of these $\widetilde{l}_k$ are different from i and j and all different from each other.

$$\begin{aligned}
\big|A_{0,1,2,q}\big| &\cdot \frac{n}{h^p}n^q\frac{1}{n^{2d}h^{2pd}}\int_{\mathbf{x},\mathbf{s}} E\Big[w_i(\mathbf{x},h)w_j(\mathbf{s},h)\prod_{k=1}^{q}w_{\widetilde{l}_k,n}(\mathbf{x},h)\Big]d\mathbf{x}d\mathbf{s}\\
&\cdot \frac{1}{n^{2d-q-1}h^{(2d+1)p}}\int_{\mathbf{x},s} E\Bigg[w_i(\mathbf{x},h)w_j(\mathbf{s},h)\prod_{k=1}^{q}w_{\widetilde{l}_k,n}(\mathbf{x},h)\Bigg]d\mathbf{x}d\mathbf{s}\\
&\cdot \frac{1}{n^{2d-q-1}h^{(2d+1)p}}\int_{\mathbf{x},\mathbf{s}} E\Bigg[K\left(\frac{\mathbf{X}_i-\mathbf{x}}{h}\right)K\left(\frac{\mathbf{X}_j-\mathbf{s}}{h}\right)\prod_{k=1}^{q}K\left(\frac{\mathbf{X}_{\widetilde{l}_k}-x}{h}\right)\Bigg]d\mathbf{x}d\mathbf{s}\\
&\cdot \frac{1}{n^{2d-q-1}h^{(2d+1)p}}\int_{\mathbf{x},\mathbf{s},\mathbf{x}_i,\mathbf{x}_j,\mathbf{x}_{\widetilde{l}_1},\dots,\mathbf{x}_{\widetilde{l}_q}} K\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)K\left(\frac{\mathbf{x}_j-\mathbf{s}}{h}\right)\prod_{k=1}^{q}K\left(\frac{\mathbf{x}_{\widetilde{l}_k}-\mathbf{x}}{h}\right)\\
&\qquad\times f(\mathbf{x}_i)f(\mathbf{x}_j)f(\mathbf{x}_{\widetilde{l}_1})\cdots f(\mathbf{x}_{\widetilde{l}_q})d\mathbf{x}d\mathbf{s}d\mathbf{x}_i d\mathbf{x}_j d\mathbf{x}_{\widetilde{l}_1}\cdots d\mathbf{x}_{\widetilde{l}_q}.
\end{aligned}$$

After a change of variable, we get

$$\begin{aligned}
\big|A_{0,1,2,q}\big| &\cdot \frac{1}{n^{2d-q-1}h^{(2d+1)p}}h^{p(q+4)}\\
&= \frac{h^{2p}}{n^{2d-q-1}h^{(2d-q-1)p}} = \frac{h^{2p}}{(nh^p)^{2d-q-1}} = o(1),
\end{aligned}$$

since $2d - q - 1 \geq 0$, $h \to 0$ and $nh^p \to \infty$.

Proof of Lemma 7: the Bahadur representation of the Poissonized conditional copula

This leaves us with the Poissonized term, for which we can do the decomposition

$$\sqrt{nh^p}\left(\hat{C}^{*,Poiss}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u}) - \hat{C}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})\right) =: A_1 + A_2,$$

where

$$A_1 := \sqrt{nh^p}\left(\hat{F}^{*,Poiss}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\left(\overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right) - \hat{F}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\left(\overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right)\right)$$
$$A_2 := \sqrt{nh^p}\left(\hat{F}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\left(\overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right) - \hat{F}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\left(\overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right)\right).$$

Representation of A_1

We further decompose A_1 to separate the effect of the Poissonization and the bootstrap.

$$\begin{aligned} A_1 :=& \sqrt{nh^p}\left(\hat{F}^{*,Poiss}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\left(\overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right) - \hat{F}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\left(\overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right)\right) \\ =:& A_{1,1} + A_{1,2}, \end{aligned}$$

where

$$\begin{aligned} A_{1,1} :=& \sqrt{nh^p}\left(\hat{F}^{*,Poiss}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\left(\overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right) - \hat{F}^{*,Poiss}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\left(\overrightarrow{\hat{F}^{*,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right)\right. \\ &\left. - \hat{F}^{*}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\left(\overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right) + \hat{F}^{*}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\left(\overrightarrow{\hat{F}^{*,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right)\right) \\ A_{1,2} :=& \sqrt{nh^p}\left(\hat{F}^{*,Poiss}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\left(\overrightarrow{\hat{F}^{*,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right) - \hat{F}^{*}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\left(\overrightarrow{\hat{F}^{*,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right)\right). \end{aligned}$$

Proof that $A_{1,1} \to 0$ in L_2

$$\begin{aligned} A_{1,1} :=& \sqrt{nh^p}\left(\hat{F}^{*,Poiss}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\left(\overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right) - \hat{F}^{*,Poiss}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\left(\overrightarrow{\hat{F}^{*,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right)\right. \\ &\left. - \hat{F}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\left(\overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right) + \hat{F}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\left(\overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right)\right) \\ =& \sqrt{nh^p}\left(\frac{1}{nh^p}\sum_{i=1}^n M_{N_n,i} w_i(\mathbf{x},h)\mathbb{1}\left\{\mathbf{Y}_i \leq \overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\} - \frac{1}{nh^p}\sum_{i=1}^n M_{N_n,i} w_i(\mathbf{x},h)\mathbb{1}\left\{\mathbf{Y}_i \leq \overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\}\right. \\ &\left. - \frac{1}{nh^p}\sum_{i=1}^n w_i(\mathbf{x},h)\mathbb{1}\left\{\mathbf{Y}_i \leq \overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\} + \frac{1}{nh^p}\sum_{i=1}^n w_i(\mathbf{x},h)\mathbb{1}\left\{\mathbf{Y}_i \leq \overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\}\right) \\ =& \sqrt{nh^p}\left(\frac{1}{nh^p}\sum_{i=1}^n M_{N_n,i} w_i(\mathbf{x},h)\left(\mathbb{1}\left\{\mathbf{Y}_i \leq \overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\} - \mathbb{1}\left\{\mathbf{Y}_i \leq \overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\}\right)\right. \\ &\left. - \frac{1}{nh^p}\sum_{i=1}^n w_i(\mathbf{x},h)\left(\mathbb{1}\left\{\mathbf{Y}_i \leq \overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\} - \mathbb{1}\left\{\mathbf{Y}_i \leq \overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\}\right)\right) \\ =& \sqrt{nh^p}\left(\frac{1}{nh^p}\sum_{i=1}^n (M_{N_n,i} - 1) w_i(\mathbf{x},h)\left(\mathbb{1}\left\{\mathbf{Y}_i \leq \overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\} - \mathbb{1}\left\{\mathbf{Y}_i \leq \overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\right\}\right)\right). \end{aligned}$$

We will therefore prove that $\int_{\mathbf{x},\mathbf{u}} A_{1,1}(\mathbf{x},\mathbf{u})d\mathbf{x}d\mathbf{u}$ tends to 0. For this, we still use the same technique: showing that the expectation of the square tends to 0. We have the decomposition

$$\begin{aligned}
E\Big[\Big(\int A_{1,1}(\mathbf{x},\mathbf{u})d\mathbf{x}d\mathbf{u}\Big)^2\Big] &= E\Big[\Big(\int \sum_{i=1}^{n} a_{1,1}(i,\mathbf{x},\mathbf{u})d\mathbf{x}d\mathbf{u}\Big)^2\Big] \\
&= \sum_{i=1}^{n}\int E\Big[a_{1,1}(i,\mathbf{x},\mathbf{u})a_{1,1}(i,\mathbf{s},\mathbf{v})\Big]d\mathbf{x}d\mathbf{s}d\mathbf{u}d\mathbf{v} + \sum_{i\neq j}\int E\Big[a_{1,1}(i,\mathbf{x},\mathbf{u})a_{1,1}(j,\mathbf{s},\mathbf{v})\Big]d\mathbf{x}d\mathbf{s}d\mathbf{u}d\mathbf{v} \qquad (11)\\
&=: A_{1,1,1} + A_{1,1,2}
\end{aligned}$$

where

$$\begin{aligned}
A_{1,1,1} &:= \sum_{i=1}^{n}\int E\Big[a_{1,1}(i,\mathbf{x},\mathbf{u})a_{1,1}(i,\mathbf{s},\mathbf{v})\Big]d\mathbf{x}d\mathbf{s}d\mathbf{u}d\mathbf{v}\\
A_{1,1,2} &:= \sum_{i\neq j}\int E\Big[a_{1,1}(i,\mathbf{x},\mathbf{u})a_{1,1}(j,\mathbf{s},\mathbf{v})\Big]d\mathbf{x}d\mathbf{s}d\mathbf{u}d\mathbf{v}
\end{aligned}$$

(we can inverse the sums, integrals and expectations as needed by Fubini's theorem)

Bounding the indicators by 1, the first term of Equation 11 becomes

$$\begin{aligned}
\big|A_{1,1,1}\big| &= \Big|\sum_{i=1}^{n}\int_{\mathbf{x},\mathbf{s},\mathbf{u},\mathbf{v}} E\Big[a_{1,1}(i,\mathbf{x},\mathbf{u})a_{1,1}(i,\mathbf{s},\mathbf{v})\Big]d\mathbf{x}d\mathbf{s}d\mathbf{u}d\mathbf{v}\Big|\\
&\leq \sum_{i=1}^{n}\int_{\mathbf{x},\mathbf{s}} E\Big|\frac{1}{nh^p}(M_{N_n,i}-1)^2 w_i(\mathbf{x},h)w_i(\mathbf{s},h)\Big|d\mathbf{x}d\mathbf{s}\\
&\leq \frac{1}{h^p}E\big[(M_{N_n,1}-1)^2\big]E\Big|\int_{\mathbf{x},\mathbf{s}} w_1(\mathbf{x},h)w_1(\mathbf{s},h)d\mathbf{x}d\mathbf{s}\Big|,
\end{aligned}$$

where in the last line, we use the fact that the observations are identically distributed and that the $(M_{N_n,i}-1)$ are independent from the (X_i).

Using the fact that $M_{N_n,1}$ is a Poisson variable with mean 1 [23, p.346], and therefore variance 1, we conclude that

$$E\big[(M_{N_n,1}-1)^2\big] = E\big[(M_{N_n,1}-E[M_{N_n,1}])^2\big] = Var[M_{N_n,1}] = 1.$$

In practice, we use kernel-based weights, meaning that $w_1(\mathbf{x},h) = K\Big(\frac{\mathbf{X}_1-\mathbf{x}}{h}\Big)$, so we can do in each integral the change of variable $\mathbf{z} = \frac{X_1-\mathbf{x}}{h}$, $\mathbf{t} = \frac{X_1-\mathbf{s}}{h}$, $d\mathbf{x} = hd\mathbf{z}$, $d\mathbf{s} = hd\mathbf{t}$ and therefore we get

$$\begin{aligned}
&\frac{1}{h^p}E\Big|\int_{\mathbf{x},\mathbf{s}} w_1(\mathbf{x},h)w_1(\mathbf{s},h)d\mathbf{x}d\mathbf{s}\Big|\\
&= \frac{1}{h^p}E\Big|\int_{\mathbf{z},\mathbf{t}} K(z)K(t)hd\mathbf{z}hd\mathbf{t}\Big|\\
&= h\Big(\int K\Big)^2 \to 0.
\end{aligned}$$

For the second term of Equation 11, we will show that $A_{1,1,2} = o_{L_2}(1)$.

Together with the previous part, this means that

$$A_{1,1}\xrightarrow{L^2} 0.$$

We first decompose $A_{1,1,2}$ in two terms.

$$\begin{aligned}
A_{1,1,2} = \frac{1}{nh^p} \sum_{i \neq j} \int_{\mathbf{x},\mathbf{s},\mathbf{u},\mathbf{v}} E\Bigg[& (M_{N_n,i} - 1)(M_{N_n,j} - 1) w_i(\mathbf{x}, h) w_j(\mathbf{s}, h) \\
& \times \left(\mathbb{1}\left\{ \mathbf{Y}_i \leq \overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u}) \right\} - \mathbb{1}\left\{ \mathbf{Y}_i \leq \overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u}) \right\} \right) \\
& \times \left(\mathbb{1}\left\{ \mathbf{Y}_j \leq \overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v}) \right\} - \mathbb{1}\left\{ \mathbf{Y}_j \leq \overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{v}) \right\} \right) \Bigg] d\mathbf{x} d\mathbf{s} d\mathbf{u} d\mathbf{v} \\
=: A_{1,1,3} + A_{1,1,4},
\end{aligned}$$

where

$$\begin{aligned}
A_{1,1,3} = \frac{1}{nh^p} \sum_{i \neq j} \int_{\mathbf{x},\mathbf{s},\mathbf{u},\mathbf{v}} E\Bigg[& (M_{N_n,i} - 1)(M_{N_n,j} - 1) w_i(\mathbf{x}, h) w_j(\mathbf{s}, h) \\
& \left(\mathbb{1}\left\{ \mathbf{Y}_i \leq \overrightarrow{\hat{F}^{*,Poiss,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u}) \right\} - \mathbb{1}\left\{ \mathbf{Y}_i \leq \overrightarrow{\hat{F}^{-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u}) \right\} \right) \\
& \left(\mathbb{1}\left\{ \mathbf{Y}_j \leq \overrightarrow{\hat{F}^{*,Poiss,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v}) \right\} - \mathbb{1}\left\{ \mathbf{Y}_j \leq \overrightarrow{\hat{F}^{-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{v}) \right\} \right) \Bigg] d\mathbf{x} d\mathbf{s} d\mathbf{u} d\mathbf{v}
\end{aligned}$$

$$A_{1,1,4} := A_{1,1,2} - A_{1,1,3}.$$

Therefore,

$$\begin{aligned}
A_{1,1,3} = E[M_{N_n,i} - 1] E[M_{N_n,j} - 1] \times \frac{1}{nh^p} \sum_{i \neq j} \int_{\mathbf{x},\mathbf{s},\mathbf{u},\mathbf{v}} E\Bigg[& w_i(\mathbf{x}, h) w_j(\mathbf{s}, h) \\
& \left(\mathbb{1}\left\{ \mathbf{Y}_i \leq \overrightarrow{\hat{F}^{*,Poiss,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u}) \right\} - \mathbb{1}\left\{ \mathbf{Y}_i \leq \overrightarrow{\hat{F}^{-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u}) \right\} \right) \\
& \left(\mathbb{1}\left\{ \mathbf{Y}_j \leq \overrightarrow{\hat{F}^{*,Poiss,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v}) \right\} - \mathbb{1}\left\{ \mathbf{Y}_j \leq \overrightarrow{\hat{F}^{-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{v}) \right\} \right) \Bigg] d\mathbf{x} d\mathbf{s} d\mathbf{u} d\mathbf{v} \\
= 0,
\end{aligned}$$

because $E[M_{N_n,i} - 1] = 0$.

We now bound $A_{1,1,4}$.

$$\begin{aligned}
A_{1,1,4} = \frac{1}{nh^p} \sum_{i \neq j} \int_{\mathbf{x},\mathbf{s},\mathbf{u},\mathbf{v}} E\Bigg[& (M_{N_n,i} - 1)(M_{N_n,j} - 1) w_i(\mathbf{x}, h) w_j(\mathbf{s}, h) \\
& \Bigg(\left(\mathbb{1}\left\{ \mathbf{Y}_i \leq \overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u}) \right\} - \mathbb{1}\left\{ \mathbf{Y}_i \leq \overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u}) \right\} \right) \\
& \left(\mathbb{1}\left\{ \mathbf{Y}_j \leq \overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v}) \right\} - \mathbb{1}\left\{ \mathbf{Y}_j \leq \overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{v}) \right\} \right) \Bigg) \\
& - \Bigg(\left(\mathbb{1}\left\{ \mathbf{Y}_i \leq \overrightarrow{\hat{F}^{*,Poiss,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u}) \right\} - \mathbb{1}\left\{ \mathbf{Y}_i \leq \overrightarrow{\hat{F}^{-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u}) \right\} \right) \\
& \left(\mathbb{1}\left\{ \mathbf{Y}_j \leq \overrightarrow{\hat{F}^{*,Poiss,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v}) \right\} - \mathbb{1}\left\{ \mathbf{Y}_j \leq \overrightarrow{\hat{F}^{-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{v}) \right\} \right) \Bigg) \Bigg] d\mathbf{x} d\mathbf{s} d\mathbf{u} d\mathbf{v}
\end{aligned}$$

$$= \frac{1}{nh^p} \sum_{i \neq j} \int_{\mathbf{x},\mathbf{s}} E\Bigg[(M_{N_n,i}-1)(M_{N_n,j}-1) w_i(\mathbf{x},h) w_j(\mathbf{s},h)$$
$$\Bigg(\Bigg(\int_{\mathbf{u}} \mathbb{1}\left\{ \mathbf{Y}_i \leq \overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u}) \right\} d\mathbf{u} - \int_{\mathbf{u}} \mathbb{1}\left\{ \mathbf{Y}_i \leq \overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u}) \right\} d\mathbf{u} \Bigg)$$
$$\Bigg(\int_{\mathbf{v}} \mathbb{1}\left\{ \mathbf{Y}_j \leq \overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v}) \right\} d\mathbf{v} - \int_{\mathbf{v}} \mathbb{1}\left\{ \mathbf{Y}_j \leq \overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{v}) \right\} d\mathbf{v} \Bigg) \Bigg)$$
$$- \Bigg(\Bigg(\int_{\mathbf{u}} \mathbb{1}\left\{ \mathbf{Y}_i \leq \overrightarrow{\hat{F}^{*,Poiss,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u}) \right\} d\mathbf{u} - \int_{\mathbf{u}} \mathbb{1}\left\{ \mathbf{Y}_i \leq \overrightarrow{\hat{F}^{-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u}) \right\} d\mathbf{u} \Bigg)$$
$$\Bigg(\int_{\mathbf{v}} \mathbb{1}\left\{ \mathbf{Y}_j \leq \overrightarrow{\hat{F}^{*,Poiss,-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{s}}}(\mathbf{v}) \right\} d\mathbf{v} - \int_{\mathbf{v}} \mathbb{1}\left\{ \mathbf{Y}_j \leq \overrightarrow{\hat{F}^{-1,-(i,j)}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{v}) \right\} d\mathbf{v} \Bigg) \Bigg) \Bigg] d\mathbf{x} d\mathbf{s},$$

by linearity and Fubini's theorem. We now apply Equation 10 to decompose each of these integral terms. We therefore get:

$$A_{1,1,4} = \frac{1}{nh^p} \sum_{i \neq j} \int_{\mathbf{x},\mathbf{s}} E\Bigg[(M_{N_n,i}-1)(M_{N_n,j}-1) w_i(\mathbf{x},h) w_j(\mathbf{s},h)$$
$$\Bigg(\Bigg(\sum_{1 \leq l_1,\ldots,l_d \leq n} \prod_{k=1}^{d} \mathbb{1}\{Y_{i,k} \leq Y_{l_k,k}\} \frac{1}{nh^p} w_{l_k,n}(\mathbf{x},h) M_{N_n,l_k} - \sum_{1 \leq l_1,\ldots,l_d \leq n} \prod_{k=1}^{d} \mathbb{1}\{Y_{i,k} \leq Y_{l_k,k}\} \frac{1}{nh^p} w_{l_k,n}(\mathbf{x},h) \Bigg)$$
$$\times \Bigg(\sum_{1 \leq l'_1,\ldots,l'_d \leq n} \prod_{k=1}^{d} \mathbb{1}\left\{Y_{i,k} \leq Y_{l'_k,k}\right\} \frac{1}{nh^p} w_{l'_k,n}(\mathbf{x},h) M_{N_n,l'_k} - \sum_{1 \leq l'_1,\ldots,l'_d \leq n} \prod_{k=1}^{d} \mathbb{1}\left\{Y_{i,k} \leq Y_{l'_k,k}\right\} \frac{1}{nh^p} w_{l'_k,n}(\mathbf{x},h) \Bigg)$$
$$- \Bigg(\sum_{1 \leq l_1,\ldots,l_d \leq n, \neq i,j} \prod_{k=1}^{d} \mathbb{1}\{Y_{i,k} \leq Y_{l_k,k}\} \frac{1}{nh^p} w_{l_k,n}(\mathbf{x},h) M_{N_n,l_k} - \sum_{1 \leq l_1,\ldots,l_d \leq n, \neq i,j} \prod_{k=1}^{d} \mathbb{1}\{Y_{i,k} \leq Y_{l_k,k}\} \frac{1}{nh^p} w_{l_k,n}(\mathbf{x},h) \Bigg)$$
$$\times \Bigg(\sum_{1 \leq l'_1,\ldots,l'_d \leq n, \neq i,j} \prod_{k=1}^{d} \mathbb{1}\left\{Y_{i,k} \leq Y_{l'_k,k}\right\} \frac{1}{nh^p} w_{l'_k,n}(\mathbf{x},h) M_{N_n,l'_k} - \sum_{1 \leq l'_1,\ldots,l'_d \leq n, \neq i,j} \prod_{k=1}^{d} \mathbb{1}\left\{Y_{i,k} \leq Y_{l'_k,k}\right\} \frac{1}{nh^p} w_{l'_k,n}(\mathbf{x},h) \Bigg)$$
$$\Bigg) \Bigg] d\mathbf{x} d\mathbf{s},$$

By combining each pair of sums together, we get

$$A_{1,1,4} = \frac{1}{nh^p} \sum_{i \neq j} \int_{\mathbf{x},\mathbf{s}} E\Bigg[(M_{N_n,i}-1)(M_{N_n,j}-1) w_i(\mathbf{x},h) w_j(\mathbf{s},h)$$
$$\Bigg(\sum_{1 \leq l_1,\ldots,l_d \leq n} \prod_{k=1}^{d} \mathbb{1}\{Y_{i,k} \leq Y_{l_k,k}\} \frac{1}{nh^p} w_{l_k,n}(\mathbf{x},h) \Bigg(\prod_{k=1}^{d} M_{N_n,l_k} - 1 \Bigg)$$
$$\times \sum_{1 \leq l'_1,\ldots,l'_d \leq n} \prod_{k=1}^{d} \mathbb{1}\left\{Y_{i,k} \leq Y_{l'_k,k}\right\} \frac{1}{nh^p} w_{l'_k,n}(\mathbf{x},h) \Bigg(\prod_{k=1}^{d} M_{N_n,l'_k} - 1 \Bigg)$$
$$- \sum_{1 \leq l_1,\ldots,l_d \leq n, \neq i,j} \prod_{k=1}^{d} \mathbb{1}\{Y_{i,k} \leq Y_{l_k,k}\} \frac{1}{nh^p} w_{l_k,n}(\mathbf{x},h) \Bigg(\prod_{k=1}^{d} M_{N_n,l_k} - 1 \Bigg)$$
$$\times \sum_{1 \leq l'_1,\ldots,l'_d \leq n, \neq i,j} \prod_{k=1}^{d} \mathbb{1}\left\{Y_{i,k} \leq Y_{l'_k,k}\right\} \frac{1}{nh^p} w_{l'_k,n}(\mathbf{x},h) \Bigg(\prod_{k=1}^{d} M_{N_n,l'_k} - 1 \Bigg) \Bigg) \Bigg] d\mathbf{x} d\mathbf{s},$$

Therefore,

$$A_{1,1,4} = \frac{1}{nh^p} \sum_{i \neq j} \int_{\mathbf{x},\mathbf{s}} E\Bigg[(M_{N_n,i} - 1)(M_{N_n,j} - 1) w_i(\mathbf{x}, h) w_j(\mathbf{s}, h)$$

$$\times \sum_{\substack{1 \leq l_1, \ldots, l_d \leq n \\ 1 \leq l'_1, \ldots, l'_d \leq n \\ \text{at least one of them is } i \text{ or } j}} \prod_{k=1}^{d} \mathbb{1}\left\{ Y_{i,k} \leq Y_{l_k,k} \right\} \frac{1}{nh^p} w_{l_k,n}(\mathbf{x}, h) \prod_{k'=1}^{d} \mathbb{1}\left\{ Y_{i,k'} \leq Y_{l'_{k'},k'} \right\} \frac{1}{nh^p} w_{l'_{k'},n}(\mathbf{x}, h)$$

$$\times \left(\prod_{k=1}^{d} M_{N_n,l'_k} - 1 \right) \left(\prod_{k'=1}^{d} M_{N_n,l'_{k'}} - 1 \right) \Bigg] d\mathbf{x} d\mathbf{s}.$$

As before, we can separate the sum over sets of constant cardinals and each of them tends to 0. Finally, $A_{1,1,4} \to 0$. Thus $A_{1,1,2} \to 0$, and finally $A_{1,1} \to 0$ in L_2 as claimed.

Representation of $A_{1,2}$.
We now decompose $A_{1,2}$.

$$\begin{aligned}
A_{1,2} :=& \sqrt{nh^p} \left(\hat{F}^{*,Poiss}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}} \left(\overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})} \right) - \hat{F}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}} \left(\overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})} \right) \right) \\
=& \sqrt{nh^p} \left(\frac{1}{nh^p} \sum_{i=1}^{n} M_{N_n,i} w_i(\mathbf{x}, h) \mathbb{1}\left\{ \mathbf{Y}_i \leq \overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})} \right\} - \frac{1}{nh^p} \sum_{i=1}^{n} w_i(\mathbf{x}, h) \mathbb{1}\left\{ \mathbf{Y}_i \leq \overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})} \right\} \right) \\
=& \sqrt{nh^p} \left(\frac{1}{nh^p} \sum_{i=1}^{n} (M_{N_n,i} - 1) w_i(\mathbf{x}, h) \mathbb{1}\left\{ \mathbf{Y}_i \leq \overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})} \right\} \right) \\
=&: A_{1,2,1} + A_{1,2,2}
\end{aligned}$$

where this decomposition is like the one in Van der Vaart and Wellner [23, p.346].

$$A_{1,2,1} := \frac{1}{\sqrt{nh^p}} \sum_{i=1}^{n} (M_{N_n,i} - 1) w_i(\mathbf{x}, h) \left(\mathbb{1}\left\{ \mathbf{Y}_i \leq \overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})} \right\} - P\left(Y_i \leq \overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})} \right) \right)$$

and

$$\begin{aligned}
A_{1,2,2} :=& \frac{1}{\sqrt{nh^p}} \sum_{i=1}^{n} (M_{N_n,i} - 1) w_i(\mathbf{x}, h) P\left(\mathbf{Y}_i \leq \overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})} \right) \\
=& \frac{1}{\sqrt{nh^p}} \left(\sum_{i=1}^{n} M_{N_n,i} w_i(\mathbf{x}, h) P\left(\mathbf{Y}_i \leq \overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})} \right) - \sum_{i=1}^{n} w_i(\mathbf{x}, h) P\left(\mathbf{Y}_i \leq \overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})} \right) \right).
\end{aligned}$$

Similarly as in the theorem of Van der Vaart and Wellner [23, p.346], we conjecture that a similar result holds in our case.

Conjecture 8. *Under suitable regularity conditions,* $A_{1,2,2} \to 0$ *in* L_2*, as* $n \to \infty$.

Under Conjecture 8, we can write

$$A_{1,2} = \frac{1}{\sqrt{nh^p}} \sum_{i=1}^{n} (M_{N_n,i} - 1) w_i(\mathbf{x}, h) \left(\mathbb{1}\left\{ \mathbf{Y}_i \leq \overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})} \right\} - P\left(\mathbf{Y}_i \leq \overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})} \right) \right) + o_{L_2}(1).$$

Together with the previous section this means that

$$\begin{aligned}
A_1 &= A_{1,1} + A_{1,2} \\
&= \frac{1}{\sqrt{nh^p}} \sum_{i=1}^{n} (M_{N_n,i} - 1) w_i(\mathbf{x}, h) \left(\mathbb{1}\left\{ \mathbf{Y}_i \leq \overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})} \right\} - P\left(\mathbf{Y}_i \leq \overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})} \right) \right) + o_{L_2}(1).
\end{aligned}$$

Representation of A_2

Note that we can decompose the difference between $\hat{C}^{*,Poiss}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})$ and $\hat{C}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})$

$$\sqrt{nh^p}\Big(\hat{C}^{*,Poiss}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u}) - \hat{C}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})\Big) = A_1 + A_2 = A_3 + A_4,$$

where

$$\begin{aligned}
A_1(\mathbf{x},\mathbf{u}) &:= \sqrt{nh^p}\bigg(\hat{F}^{*,Poiss}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\Big(\overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\Big) - \hat{F}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\Big(\overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\Big)\bigg)\\
A_2(\mathbf{x},\mathbf{u}) &:= \sqrt{nh^p}\bigg(\hat{F}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\Big(\overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\Big) - \hat{F}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\Big(\overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\Big)\bigg)\\
A_3(\mathbf{x},\mathbf{u}) &:= \sqrt{nh^p}\bigg(\hat{F}^{*,Poiss}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\Big(\overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\Big) - \hat{F}^{*,Poiss}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\Big(\overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\Big)\bigg)\\
A_4(\mathbf{x},\mathbf{u}) &:= \sqrt{nh^p}\bigg(\hat{F}^{*,Poiss}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\Big(\overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\Big) - \hat{F}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\Big(\overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\Big)\bigg).
\end{aligned}$$

We have proved in the previous section that $A_1 = o(1)$. In the same way, we can prove that $A_3 = o(1)$.

Conjecture 9. *Under some regularity conditions,*

$$A_2(\mathbf{x},\mathbf{u}) = \frac{1}{\sqrt{nh^p}}\sum_{i=1}^{n}\sum_{j=1}^{d}\frac{\partial C_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})}{\partial u_j} \times w_i(\mathbf{x},h)(M_{N_n,i}-1)\Big(\mathbb{1}\Big\{Y_{i,j} \le \hat{F}^{-1}_{Y_j|\mathbf{X}=\mathbf{x}}(v)\Big\} - u_j\Big) + o_{L_2}(1).$$

Sketch of the proof: by definition,

$$\begin{aligned}
A_2(\mathbf{x},\mathbf{u}) &:= \sqrt{nh^p}\bigg(\hat{F}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\Big(\overrightarrow{\hat{F}^{*,Poiss,-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\Big) - \hat{F}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\Big(\overrightarrow{\hat{F}^{-1}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{u})\Big)\bigg)\\
&= \sqrt{nh^p}\bigg(\hat{C}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})\Big(\hat{F}_{Y_1|\mathbf{X}=\mathbf{x}}\Big(\hat{F}^{*,Poiss,-1}_{Y_1|\mathbf{X}=\mathbf{x}}(u_1)\Big), \cdots, \hat{F}_{Y_d|\mathbf{X}=\mathbf{x}}\Big(\hat{F}^{*,Poiss,-1}_{Y_d|\mathbf{X}=\mathbf{x}}(u_d)\Big)\Big)\\
&\quad - \hat{C}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})\Big(\hat{F}_{Y_1|\mathbf{X}=\mathbf{x}}\Big(\hat{F}^{-1}_{Y_1|\mathbf{X}=\mathbf{x}}(u_1)\Big), \cdots, \hat{F}_{Y_d|\mathbf{X}=\mathbf{x}}\Big(\hat{F}^{-1}_{Y_d|\mathbf{X}=\mathbf{x}}(u_d)\Big)\Big)\bigg)
\end{aligned}$$

By the mean value theorem, we have:

$$A_2(\mathbf{x},\mathbf{u}) \approx \sqrt{nh^p}\sum_{j=1}^{d}\frac{\partial C_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})}{\partial u_j} \times \bigg(\hat{F}_{Y_j|\mathbf{X}=\mathbf{x}}\Big(\hat{F}^{*,Poiss,-1}_{Y_j|\mathbf{X}=\mathbf{x}}(u_j)\Big) - \hat{F}_{Y_j|\mathbf{X}=\mathbf{x}}\Big(\hat{F}^{-1}_{Y_j|\mathbf{X}=\mathbf{x}}(u_j)\Big)\bigg) = A_{2,1} + A_{2,2},$$

where

$$\begin{aligned}
A_{2,1} &:= \sqrt{nh^p}\sum_{j=1}^{d}\frac{\partial C_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})}{\partial u_j} \times \bigg(\hat{F}^{*,Poiss}_{Y_j|\mathbf{X}=\mathbf{x}}\Big(\hat{F}^{-1}_{Y_j|\mathbf{X}=\mathbf{x}}(u_j)\Big) - \hat{F}_{Y_j|\mathbf{X}=\mathbf{x}}\Big(\hat{F}^{-1}_{Y_j|\mathbf{X}=\mathbf{x}}(u_j)\Big)\bigg)\\
A_{2,2} &:= \sqrt{nh^p}\sum_{j=1}^{d}\frac{\partial C_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})}{\partial u_j} \times \bigg(\hat{F}_{Y_j|\mathbf{X}=\mathbf{x}}\Big(\hat{F}^{*,Poiss,-1}_{Y_j|\mathbf{X}=\mathbf{x}}(u_j)\Big) - \hat{F}^{*,Poiss}_{Y_j|\mathbf{X}=\mathbf{x}}\Big(\hat{F}^{-1}_{Y_j|\mathbf{X}=\mathbf{x}}(u_j)\Big)\bigg)
\end{aligned}$$

Therefore, plugging in the definition of $\hat{F}^{*,Poiss}_{Y_j|\mathbf{X}=\mathbf{x}}$ and $\hat{F}_{Y_j|\mathbf{X}=\mathbf{x}}$, we get

$$\begin{aligned}
A_{2,1}(\mathbf{x},\mathbf{u}) &\approx \sqrt{nh^p}\sum_{j=1}^{d}\frac{\partial C_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})}{\partial u_j} \\
&\times\left(\frac{1}{nh^p}\sum_{i=1}^{n}M_{N_n,i}w_i(\mathbf{x},h)\mathbb{1}\left\{Y_{i,j}\le \hat{F}^{-1}_{Y_j|\mathbf{X}=\mathbf{x}}(u_j)\right\}-\frac{1}{nh^p}\sum_{i=1}^{n}w_i(\mathbf{x},h)\mathbb{1}\left\{Y_{i,j}\le \hat{F}^{-1}_{Y_j|\mathbf{X}=\mathbf{x}}(u_j)\right\}\right) \\
&= \sqrt{nh^p}\sum_{j=1}^{d}\frac{\partial C_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})}{\partial u_j}\times\frac{1}{nh^p}\sum_{i=1}^{n}(M_{N_n,i}-1)w_i(\mathbf{x},h)\mathbb{1}\left\{Y_{i,j}\le \hat{F}^{-1}_{Y_j|\mathbf{X}=\mathbf{x}}(u_j)\right\} \\
&= \frac{1}{\sqrt{nh^p}}\sum_{j=1}^{d}\frac{\partial C_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})}{\partial u_j}\times\sum_{i=1}^{n}(M_{N_n,i}-1)w_i(\mathbf{x},h)\mathbb{1}\left\{Y_{i,j}\le \hat{F}^{-1}_{Y_j|\mathbf{X}=\mathbf{x}}(u_j)\right\}.
\end{aligned}$$

Representation of $A_{2,2}$.
We denote by $A_{2,2,j}$ each term of the sum in $A_{2,2}$, that is

$$\begin{aligned}
A_{2,2,j} &:= \sqrt{nh^p}\hat{F}_{Y_j|\mathbf{X}=\mathbf{x}}\left(\hat{F}^{*,Poiss,-1}_{Y_j|\mathbf{X}=\mathbf{x}}(u_j)\right)-\hat{F}^{*,Poiss}_{Y_j|\mathbf{X}=\mathbf{x}}\left(\hat{F}^{-1}_{Y_j|\mathbf{X}=\mathbf{x}}(u_j)\right) \\
&= \frac{1}{nh^p}\sum_{i=1}^{n}w_i(\mathbf{x},h)\mathbb{1}\left\{Y_{i,j}\le \hat{F}^{*,Poiss,-1}_{Y_j|\mathbf{X}=\mathbf{x}}(u_j)\right\}-\sum_{i=1}^{n}M_{N_n,i}w_i(\mathbf{x},h)\mathbb{1}\left\{Y_{i,j}\le \hat{F}^{-1}_{Y_j|\mathbf{X}=\mathbf{x}}(u_j)\right\} \\
&= \frac{1}{nh^p}\sum_{i=1}^{n}w_i(\mathbf{x},h)\left(\mathbb{1}\left\{Y_{i,j}\le \hat{F}^{*,Poiss,-1}_{Y_j|\mathbf{X}=\mathbf{x}}(u_j)\right\}-M_{N_n,i}\mathbb{1}\left\{Y_{i,j}\le \hat{F}^{-1}_{Y_j|\mathbf{X}=\mathbf{x}}(u_j)\right\}\right).
\end{aligned}$$

Let us define

$$\begin{aligned}
A_{2,3,k}(\mathbf{x},\mathbf{u}) &:= -A_{2,2,k}-\sqrt{nh^p}\sum_{i=1}^{n}w_i(\mathbf{x},h)(M_{N_n,i}-1)u_k \\
&= \frac{1}{nh^p}\sum_{i=1}^{n}w_i(\mathbf{x},h)\left(M_{N_n,i}\mathbb{1}\left\{Y_{i,k}\le \hat{F}^{-1}_{Y_k|\mathbf{X}=\mathbf{x}}(u_k)\right\}-\mathbb{1}\left\{Y_{i,k}\le \hat{F}^{*,Poiss,-1}_{Y_k|\mathbf{X}=\mathbf{x}}(u_k)\right\}-(M_{N_n,i}-1)u_k\right).
\end{aligned}$$

We have the decomposition

$$E\left[\left(\int_{\mathbf{x},\mathbf{u}}A_{2,3,j}(\mathbf{x},\mathbf{u})\right)^2 d\mathbf{u}d\mathbf{x}\right]=A_{2,4,k}+A_{2,5,k},$$

where

$$\begin{aligned}
A_{2,4,k} &:= \frac{1}{nh^p}\sum_{i=1}^{n}\mathrm{E}\Bigg[\int_{u_k,v_k}w_i(\mathbf{x},h)^2 \\
&\times\left(M_{N_n,i}\left(\mathbb{1}\left\{Y_{i,k}\le \hat{F}^{-1}_{Y_k|\mathbf{X}=\mathbf{x}}(u_k)\right\}-\mathbb{1}\left\{Y_{i,k}\le \hat{F}^{*,Poiss,-1}_{Y_k|\mathbf{X}=\mathbf{x}}(u_k)\right\}\right)-(M_{N_n,i}-1)u_k\right) \\
&\times\left(M_{N_n,i}\left(\mathbb{1}\left\{Y_{i,k}\le \hat{F}^{-1}_{Y_k|\mathbf{X}=\mathbf{s}}(v_k)\right\}-\mathbb{1}\left\{Y_{i,k}\le \hat{F}^{*,Poiss,-1}_{Y_k|\mathbf{X}=\mathbf{s}}(v_k)\right\}\right)-(M_{N_n,i}-1)v_k\right)du_kdv_kd\mathbf{x}d\mathbf{s}\Bigg], \\
A_{2,5,k} &:= \frac{1}{nh^p}\mathrm{E}\Bigg[\sum_{i\neq j=1}^{n}\int_{u_k}w_i(\mathbf{x},h)w_j(\mathbf{x},h) \\
&\times\left(M_{N_n,i}\left(\mathbb{1}\left\{Y_{i,k}\le \hat{F}^{-1}_{Y_k|\mathbf{X}=\mathbf{x}}(u_k)\right\}-\mathbb{1}\left\{Y_{i,k}\le \hat{F}^{*,Poiss,-1}_{Y_k|\mathbf{X}=\mathbf{x}}(u_k)\right\}\right)-(M_{N_n,i}-1)u_k\right) \\
&\times\left(M_{N_n,j}\left(\mathbb{1}\left\{Y_{j,k}\le \hat{F}^{-1}_{Y_k|\mathbf{X}=\mathbf{s}}(v_k)\right\}-\mathbb{1}\left\{Y_{j,k}\le \hat{F}^{*,Poiss,-1}_{Y_k|\mathbf{X}=\mathbf{s}}(v_k)\right\}\right)-(M_{N_n,j}-1)v_k\right)du_kdv_kd\mathbf{x}d\mathbf{s}\Bigg]
\end{aligned}$$

As we did before, using change of variables, $A_{2,4,k}\to 0$ as $n\to\infty$.

Conjecture 10. *Under some regularity conditions,* $A_{2,5,k} \to 0$*.*

Therefore,

$$A_{2,2} = \sum_{j=1}^{p} A_{2,2,j} = -\sqrt{nh^p} \sum_{i=1}^{n} w_i(\mathbf{x}, h)(M_{N_n,i} - 1)u_k + o(1),$$

finishing the sketch of the proof of Conjecture 9.

7.3 Bahadur representation for $C_{\hat{\theta}(\mathbf{x})}(\mathbf{u}) - C_{\theta_0(\mathbf{x})}(\mathbf{u})$

Lemma 11. *Under some regularity conditions,*

$$C_{\hat{\theta}(\mathbf{x})}(\mathbf{u}) - C_{\theta_0(\mathbf{x})}(\mathbf{u}) = \frac{1}{nh^p} \sum_{i=1}^{n} w_i(\mathbf{x}, h)\tilde{\xi}_i(\mathbf{x}, \mathbf{u}) + \tilde{\varepsilon}_n(n, \mathbf{x}, \mathbf{u})), \tag{12}$$

for some $\tilde{\xi}_i(\mathbf{x}, \mathbf{u})$ *and* $\tilde{\varepsilon}_n(n, \mathbf{x}, \mathbf{u})$ *such that* $\tilde{\varepsilon}_n(n, \mathbf{x}, \mathbf{u}) = o(1)$.

To obtain this Bahadur representation, we will combine the Delta-method with the following lemma (under some regularity conditions on the parametric class of copulas).

Lemma 12. *Under some regularity conditions,*

$$\hat{\theta}(\mathbf{x}) - \theta_0(\mathbf{x}) = \frac{1}{nh^p} \sum_{i=1}^{n} \tilde{\xi}_i(\mathbf{x}) + \tilde{\varepsilon}_n(n, \mathbf{x}), \tag{13}$$

for some $\tilde{\xi}_i(\mathbf{x})$ *and* $\tilde{\varepsilon}_n(n, \mathbf{x})$ *such that* $\tilde{\varepsilon}_n(n, \mathbf{x}) = o(1)$.

We now give a sketch of the proof of Lemma 12.

By using the definition for the weighted maximum likelihood,

$$\hat{\theta}(\mathbf{x}) = \arg\max_{\theta \in \Theta} \sum_{i=1}^{n} w_i(\mathbf{x}, h) \log \Big(c_\theta\big(\overrightarrow{\hat{F}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{Y}_i)}\big)\Big) =: \arg\max_{\theta \in \Theta} \mathsf{M}_{n,\mathbf{x}}(\theta)$$

and

$$\theta_0(\mathbf{x}) = \arg\max_{\theta \in \Theta} E\Big[\log \Big(c_\theta\big(\overrightarrow{F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{Y})}\big)\Big) \Big| \mathbf{X} = \mathbf{x} \Big] = \arg\max_{\theta \in \Theta} \mathsf{M}_{\mathbf{x}}(\theta)$$

For this, we will apply the following result.

Theorem 7.1 (Theorem 3.2.16 in [23])**.** *Let* M_n *be stochastic processes indexed by an open subset* Θ *of Euclidean space and* $\mathsf{M} : \Theta \to \mathsf{R}$ *a deterministic function. Assume that* $\theta \to \mathsf{M}(\theta)$ *is twice continuously di erentiable at a point of maximum* θ_0 *with nonsingular second-derivative matrix* V*. Suppose that*

$$r_n(\mathsf{M}_n - \mathsf{M})(\tilde{\theta}_n) - r_n(\mathsf{M}_n - \mathsf{M})(\theta_0) = (\tilde{\theta}_n - \theta_0)^\top Z_n + o_P^*(||\tilde{\theta}_n - \theta_0|| + r_n||\tilde{\theta}_n - \theta_0||^2 + r_n^{-1}), \tag{14}$$

for every random sequence $\tilde{\theta}_n = \theta_0 + o_P^*(1)$ *and a uniformly tight sequence of random vectors* Z_n*. If the sequence* $\hat{\theta}_n$ *converges in outer probability to* θ_0 *and satis es* $\mathsf{M}_n(\hat{\theta}_n) \geq \sup_\theta \mathsf{M}_n(\theta) - o_P(r_n^{-2})$ *for every* n*, then*

$$r_n(\hat{\theta}_n - \theta_0) = -V^{-1} Z_n + o_P^*(1).$$

If it is known that the sequence $r_n(\hat{\theta}_n - \theta_0)$ *is uniformly tight, then the displayed condition needs to be veri ed for sequences* $\theta_n = \theta_0 + O_P^*(r_n^{-1})$ *only.*

Assume that the assumptions of Theorem 7.1 are satisfied for some fixed $\mathbf{x}$. Then we can apply it to get

$$\sqrt{nh^p}\big(\hat{\theta}(\mathbf{x}) - \theta(\mathbf{x})\big) \sim \sqrt{nh^p} V^{-1} \Bigg(\sum_{i=1}^{n} w_i(\mathbf{x}, h) \times \frac{\partial \log \Big(c_\theta\big(\overrightarrow{F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{Y}_i)}\big)\Big)}{\partial \theta}(\theta_0) - E\Bigg[\frac{\partial \log \Big(c_\theta\big(\overrightarrow{F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{Y})}\big)\Big)}{\partial \theta}(\theta_0(\mathbf{x})) \Bigg| \mathbf{X} = \mathbf{x} \Bigg] \Bigg). \tag{15}$$

Therefore,

$$\hat{\theta}(\mathbf{x}) - \theta(\mathbf{x}) = \frac{1}{nh^p}\sum_{i=1}^{n}\tilde{\xi}_i(\mathbf{x}) + \tilde{\varepsilon}_n(n,\mathbf{x}), \tag{16}$$

with

$$\tilde{\xi}_i(\mathbf{x}) := V^{-1}w_i(\mathbf{x},h) \times \frac{\partial \log\Big(c_\theta(\overrightarrow{F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{Y}_i))\Big)}{\partial\theta}(\theta_0) - E\Bigg[\frac{\partial \log\Big(c_\theta(\overrightarrow{F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{Y}))\Big)}{\partial\theta}(\theta_0(\mathbf{x}))\Bigg|\mathbf{X}=\mathbf{x}\Bigg].$$

This finishes the proof of Lemma 12.

There only remains to show that the assumptions of Theorem 7.1 are satisfied. We focus on showing that the condition given in Equation 14 is satisfied.

For the rest of this section, let $\mathbf{x} \in \mathsf{R}^p$ be fixed and $\tilde{\theta}_n(\mathbf{x})$ be any (random) sequence converging to $\theta_0(\mathbf{x})$. We can decompose the left-hand side of Equation 14 as

$$\begin{aligned}
&\sqrt{nh^p}\big(\mathsf{M}_{n,\mathbf{x}}(\tilde{\theta}_n(\mathbf{x})) - \mathsf{M}_x(\tilde{\theta}_n(\mathbf{x}))\big) - \sqrt{nh^p}\big(\mathsf{M}_{n,\mathbf{x}}(\theta_0(\mathbf{x})) - \mathsf{M}_{\mathbf{x}}(\theta_0(\mathbf{x}))\big)\\
&= \sqrt{nh^p}\Bigg(\sum_{i=1}^{n} w_i(\mathbf{x},h)\log\Big(c_{\tilde{\theta}_n(\mathbf{x})}(\overrightarrow{\hat{F}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{Y}_i)})\Big) - E\Bigg[\log\Big(c_{\tilde{\theta}_n(\mathbf{x})}(\overrightarrow{F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{Y}))\Big)\Bigg|\mathbf{X}=\mathbf{x}\Bigg]\Bigg)\\
&- \sqrt{nh^p}\Bigg(\sum_{i=1}^{n} w_i(\mathbf{x},h)\log\Big(c_{\theta_0(\mathbf{x})}(\overrightarrow{\hat{F}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{Y}_i)\Big) - E\Bigg[\log\Big(c_{\theta_0(\mathbf{x})}(\overrightarrow{F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{Y}))\Big)\Bigg|\mathbf{X}=\mathbf{x}\Bigg]\Bigg)\\
&=: B_1(\mathbf{x}) - B_2(\mathbf{x})
\end{aligned}$$

where

$$B_1(\mathbf{x}) = \sqrt{nh^p}\Bigg(\sum_{i=1}^{n} w_i(\mathbf{x},h)\log\Big(c_{\tilde{\theta}_n(\mathbf{x})}(\overrightarrow{\hat{F}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{Y}_i))\Big) - \sum_{i=1}^{n} w_i(\mathbf{x},h)\log\Big(c_{\theta_0(\mathbf{x})}(\overrightarrow{\hat{F}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{Y}_i))\Big)\Bigg)$$

and

$$B_2(\mathbf{x}) = \sqrt{nh^p}\Bigg(E\Bigg[\log\Big(c_{\tilde{\theta}_n(\mathbf{x})}(\overrightarrow{F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{Y}))\Big)\Bigg|\mathbf{X}=\mathbf{x}\Bigg]\Bigg) - E\Bigg[\log\Big(c_{\theta_0(\mathbf{x})}(\overrightarrow{F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{Y}))\Big)\Bigg|\mathbf{X}=\mathbf{x}\Bigg]\Bigg)$$

Representation of B_1
By a Taylor-Lagrange expansion, there exists $\check{\theta}_{j,k,1}(x) \in [\theta_0(\mathbf{x}), \tilde{\theta}_n(\mathbf{x})]$ such that we can decompose B_1 in a main term $B_{1,1}$ and a remainder term $B_{1,2}$.

$$\begin{aligned}
B_1(\mathbf{x}) :=& \sqrt{nh^p}\Bigg(\sum_{i=1}^{n} w_i(\mathbf{x},h)\log\Big(c_{\tilde{\theta}_n(\mathbf{x})}(\overrightarrow{\hat{F}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{Y}_i))\Big) - \sum_{i=1}^{n} w_i(\mathbf{x},h)\log\Big(c_{\theta_0(\mathbf{x})}(\overrightarrow{\hat{F}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{Y}_i))\Big)\Bigg)\\
=:& B_{1,1}(\mathbf{x}) + B_{1,2}(\mathbf{x})
\end{aligned}$$

where

$$B_{1,1}(\mathbf{x}) := \sqrt{nh^p}\sum_{i=1}^{n} w_i(\mathbf{x},h)(\tilde{\theta}_n(\mathbf{x}) - \theta_0(\mathbf{x}))^\top \times \frac{\partial\log\Big(c_\theta(\overrightarrow{\hat{F}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{Y}_i)))\Big)}{\partial\theta}(\theta_0),$$

$$B_{1,2}(x) = \sqrt{nh^p}\sum_{i=1}^{n} w_i(\mathbf{x},h)(\tilde{\theta}_n(\mathbf{x}) - \theta_0(\mathbf{x}))^2 \times \frac{1}{2}\sum_{j,k=1}^{p}\frac{\partial^2\log\Big(c_\theta(\overrightarrow{\hat{F}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{Y}_i)))\Big)}{\partial\theta_j\partial\theta_k}(\check{\theta}_{j,k,1}(\mathbf{x}))$$

Representation of B_2
Again by a Taylor-Lagrange expansion

$$\begin{aligned} B_2(\mathbf{x}) &:= \sqrt{nh^p}\Bigg(E\Bigg[\log\Big(c_{\tilde{\theta}_n(\mathbf{x})}(\overrightarrow{F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{Y}))\Big)\Bigg|\mathbf{X}=\mathbf{x}\Bigg]\Bigg) - E\Bigg[\log\Big(c_{\theta_0(\mathbf{x})}(\overrightarrow{F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{Y}))\Big)\Bigg|\mathbf{X}=\mathbf{x}\Bigg]\Bigg) \\ &= B_{2,1}(\mathbf{x}) + B_{2,2}(\mathbf{x}) \end{aligned}$$

where

$$B_{2,1}(\mathbf{x}) := \sqrt{nh^p}\Bigg(E\Bigg[(\tilde{\theta}_n(\mathbf{x}) - \theta_0(\mathbf{x}))' \frac{\partial \log\Big(c_\theta(\overrightarrow{F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{Y}))\Big)}{\partial\theta}(\theta_0(\mathbf{x}))\Bigg|\mathbf{X}=\mathbf{x}\Bigg]$$

$$B_{2,2}(\mathbf{x}) := \sqrt{nh^p}(\tilde{\theta}_n(\mathbf{x}) - \theta_0(\mathbf{x}))^2 \times \sum_{j,k=1}^{p} E\Bigg[\frac{\partial^2 \log\Big(c_\theta(\overrightarrow{F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{Y}))\Big)}{\partial\theta_j\partial\theta_k}(\check{\theta}_{j,k,2}(\mathbf{x}))\Bigg|\mathbf{X}=\mathbf{x}\Bigg]$$

Difference between the remainder terms
Difference between remainder of the sum and remainder of the expectation:

$$\begin{aligned} B_{3,2}(\mathbf{x}) :=& B_{1,2}(\mathbf{x}) - B_{2,2}(\mathbf{x}) \\ =& \sqrt{nh^p}(\tilde{\theta}_n(\mathbf{x}) - \theta_0(\mathbf{x}))^2 \times \frac{1}{2}\sum_{i=1}^{n} w_i(\mathbf{x},h) \sum_{j,k=1}^{p} \frac{\partial^2 \log\Big(c_\theta(\overrightarrow{\hat{F}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{Y}_i))\Big)}{\partial\theta_j\partial\theta_k}(\check{\theta}_{j,k,1}(\mathbf{x})) \\ & - \sum_{j,k=1}^{p} E\Bigg[\frac{\partial^2 \log\Big(c_\theta(\overrightarrow{F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{Y}))\Big)}{\partial\theta_j\partial\theta_k}(\check{\theta}_{j,k,2}(\mathbf{x}))\Bigg|\mathbf{X}=\mathbf{x}\Bigg] \end{aligned}$$

Conjecture 13. *Under some regularity conditions,* $B_{3,2} = o_P^*(||\tilde{\theta}_n - \theta_0|| + r_n||\tilde{\theta}_n - \theta_0||^2 + r_n^{-1})$.

Difference between the main terms

$$\begin{aligned} B_{3,1} :=& B_{1,1}(\mathbf{x}) - B_{2,1}(\mathbf{x}) \\ =& \sqrt{nh^p}\sum_{i=1}^{n} w_i(\mathbf{x},h)(\tilde{\theta}_n(\mathbf{x}) - \theta_0(\mathbf{x}))^\top \times \frac{\partial \log\Big(c_\theta(\overrightarrow{\hat{F}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{Y}_i))\Big)}{\partial\theta}(\theta_0) \\ & - \sqrt{nh^p}E\Bigg[(\tilde{\theta}_n(\mathbf{x}) - \theta_0(\mathbf{x}))^\top \frac{\partial \log\Big(c_\theta(\overrightarrow{F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{Y}))\Big)}{\partial\theta}(\theta_0(\mathbf{x}))\Bigg|\mathbf{X}=\mathbf{x}\Bigg] \\ =& \sqrt{nh^p}(\tilde{\theta}_n(\mathbf{x}) - \theta_0(\mathbf{x}))^\top\Bigg(\sum_{i=1}^{n} w_i(\mathbf{x},h) \times \frac{\partial \log\Big(c_\theta(\overrightarrow{\hat{F}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{Y}_i))\Big)}{\partial\theta}(\theta_0) \\ & - E\Bigg[\frac{\partial \log\Big(c_\theta(\overrightarrow{F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{Y}))\Big)}{\partial\theta}(\theta_0(\mathbf{x}))\Bigg|\mathbf{X}=\mathbf{x}\Bigg]\Bigg) \\ =:& \ B_{3,1,1} + B_{3,1,2} \end{aligned}$$

where

$$B_{3,1,1} = \sqrt{nh^p}(\tilde{\theta}_n(\mathbf{x}) - \theta_0(\mathbf{x}))^\top \left(\sum_{i=1}^n w_i(\mathbf{x}, h) \times \frac{\partial \log \left(c_\theta(\overrightarrow{F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{Y}_i)) \right)}{\partial \theta}(\theta_0) \right.$$
$$\left. - E\left[\frac{\partial \log \left(c_\theta(\overrightarrow{F_{Y|\mathbf{X}=\mathbf{x}}}(\mathbf{Y})) \right)}{\partial \theta}(\theta_0(\mathbf{x})) \middle| X = x \right] \right)$$
$$= (\tilde{\theta}_n(\mathbf{x}) - \theta_0(\mathbf{x}))^\top Z_n,$$
$$Z_n := \sqrt{nh^p}\left(\sum_{i=1}^n w_i(\mathbf{x}, h) \times \frac{\partial \log \left(c_\theta(\overrightarrow{F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{Y}_i)) \right)}{\partial \theta}(\theta_0) - E\left[\frac{\partial \log \left(c_\theta(\overrightarrow{F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{Y})) \right)}{\partial \theta}(\theta_0(\mathbf{x})) \middle| \mathbf{X} = \mathbf{x} \right] \right),$$
$$B_{3,1,2} = \sqrt{nh^p}(\tilde{\theta}_n(\mathbf{x}) - \theta_0(\mathbf{x}))^\top \sum_{i=1}^n w_i(\mathbf{x}, h) \times \left(\frac{\partial \log \left(c_\theta(\overrightarrow{\hat{F}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{Y}_i))) \right)}{\partial \theta}(\theta_0) - \frac{\partial \log \left(c_\theta(\overrightarrow{F_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{Y}_i)) \right)}{\partial \theta}(\theta_0) \right)$$

Conjecture 14. *Under some regularity conditions,* $B_{3,1,2} = o_P^*(||\tilde{\theta}_n - \theta_0|| + r_n||\tilde{\theta}_n - \theta_0||^2 + r_n^{-1})$.

In conclusion, we have

$$\begin{aligned}
&\sqrt{nh^p}\big(\mathsf{M}_{n,\mathbf{x}}(\tilde{\theta}_n(\mathbf{x})) - \mathsf{M}_x(\tilde{\theta}_n(\mathbf{x}))\big) - \sqrt{nh^p}\big(\mathsf{M}_{n,\mathbf{x}}(\theta_0(\mathbf{x})) - \mathsf{M}_{\mathbf{x}}(\theta_0(\mathbf{x}))\big) \\
&= B_1 - B_2 \\
&= B_{1,1} + B_{1,2} - B_{2,1} - B_{2,2} \\
&= B_{1,1} - B_{2,1} + o_P^*(||\tilde{\theta}_n - \theta_0|| + r_n||\tilde{\theta}_n - \theta_0||^2 + r_n^{-1}) \\
&= B_{3,1,1} + B_{3,1,2} + o_P^*(||\tilde{\theta}_n - \theta_0|| + r_n||\tilde{\theta}_n - \theta_0||^2 + r_n^{-1}) \\
&= B_{1,1} - B_{2,1} + o_P^*(||\tilde{\theta}_n - \theta_0|| + r_n||\tilde{\theta}_n - \theta_0||^2 + r_n^{-1}) \\
&= B_{3,1,1} + o_P^*(||\tilde{\theta}_n - \theta_0|| + r_n||\tilde{\theta}_n - \theta_0||^2 + r_n^{-1}) \\
&= (\tilde{\theta}_n(\mathbf{x}) - \theta_0(\mathbf{x}))^\top Z_n + o_P^*(||\tilde{\theta}_n - \theta_0|| + r_n||\tilde{\theta}_n - \theta_0||^2 + r_n^{-1}),
\end{aligned}$$

which matches the construction required by Equation 14.

7.4 Bahadur representation for $C_{\hat{\theta}^*(\mathbf{x})}(\mathbf{u}) - C_{\hat{\theta}(\mathbf{x})}(\mathbf{u})$

The proof for this section uses the method sketched in the previous section combined with the properties of the non-parametric bootstrap.

Lemma 15. *Under some regularity conditions, we have,*

$$C_{\hat{\theta}^*(\mathbf{x})}(\mathbf{u}) - C_{\hat{\theta}(\mathbf{x})}(\mathbf{u}) = \frac{1}{nh^p} \sum_{i=1}^{n} w_i(\mathbf{x}, h) \tilde{\xi}_i^*(\mathbf{x}, \mathbf{u}) + \tilde{\varepsilon}_n^*(n, \mathbf{x}, \mathbf{u})), \tag{17}$$

for some $\tilde{\xi}_i^*(\mathbf{x}, \mathbf{u})$ *and* $\tilde{\varepsilon}_n^*(n, \mathbf{x}, \mathbf{u})$ *such that* $\tilde{\varepsilon}_n^*(n, \mathbf{x}, \mathbf{u}) = o(1)$.

To obtain this Bahadur representation, we will combine the Delta-method with the following lemma (under some regularity conditions on the parametric class of copulas).

Lemma 16. *Under some regularity conditions,*

$$\hat{\theta}^*(\mathbf{x}) - \hat{\theta}(\mathbf{x}) = \frac{1}{nh^p} \sum_{i=1}^{n} \tilde{\xi}_i^*(\mathbf{x}) + \tilde{\varepsilon}_n^*(n, \mathbf{x}), \tag{18}$$

for some $\tilde{\xi}_i^*(\mathbf{x})$ *and* $\tilde{\varepsilon}_n^*(n, \mathbf{x})$ *such that* $\tilde{\varepsilon}_n^*(n, \mathbf{x}) = o(1)$.

We now give a sketch of the proof of Lemma 16. By using the definition for the weighted maximum likelihood,

$$\hat{\theta}^*(\mathbf{x}) = \arg\max_{\theta \in \Theta} \sum_{i=1}^{n} w_i(\mathbf{x}, h) \log\left(c_\theta\left(\overrightarrow{\hat{F}^*_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{Y}_i)\right)\right) =: \arg\max_{\theta \in \Theta} \mathsf{M}^*_{n,\mathbf{x}}(\theta)$$

and

$$\hat{\theta}(\mathbf{x}) = \arg\max_{\theta \in \Theta} \sum_{i=1}^{n} w_i(\mathbf{x}, h) \log\left(c_\theta\left(\overrightarrow{\hat{F}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}(\mathbf{Y}_i)\right)\right) =: \arg\max_{\theta \in \Theta} \mathsf{M}_{n,\mathbf{x}}(\theta)$$

For this, we will apply Theorem 7.1 as in the previous section.

7.5 End of the proof

As we assumed that the model is well-specified, so that $C_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u}) = C_{\theta(x)}(\mathbf{u})$ and we can use the results from Lemmas 5, 6 and 7 to get

$$\begin{aligned}
T_n =& nh^p \int \left(\hat{C}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u}) - C_{\hat{\theta}(x)}(\mathbf{u})\right)^2 W(d\mathbf{x}, d\mathbf{u}) \\
=& nh^p \int \left(\hat{C}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u}) - C_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u}) + C_{\theta(\mathbf{x})}(\mathbf{u}) - C_{\hat{\theta}(x)}(\mathbf{u})\right)^2 d\mathbf{x}d\mathbf{u} \\
=& \int \left(\sqrt{nh^p}\Big(\hat{C}_{Y|\mathbf{X}=\mathbf{x}}(\mathbf{u}) - C_{Y|\mathbf{X}=\mathbf{x}}(\mathbf{u})\Big) - \sqrt{nh^p}\Big(C_{\theta(\mathbf{x})}(\mathbf{u}) - C_{\hat{\theta}(\mathbf{x})}(\mathbf{u})\Big)\right)^2 W(dx, d\mathbf{u}) \\
=& \int \left(\frac{1}{\sqrt{nh^p}} \sum_{i=1}^n w_i(\mathbf{x}, h)\xi_i(\mathbf{x}, \mathbf{u}) + \varepsilon_n(\mathbf{x}, \mathbf{u}) - \frac{1}{\sqrt{nh^p}} \sum_{i=1}^n w_i(\mathbf{x}, h)\tilde{\xi}_i(\mathbf{x}, \mathbf{u}) - \tilde{\varepsilon}_n(\mathbf{x}, \mathbf{u})\right)^2 d\mathbf{x}d\mathbf{u} \\
=& \int \left(\frac{1}{\sqrt{nh^p}} \sum_{i=1}^n w_i(\mathbf{x}, h)\xi_i(\mathbf{x}, \mathbf{u}) - \frac{1}{\sqrt{nh^p}} \sum_{i=1}^n w_i(\mathbf{x}, h)\tilde{\xi}_i(\mathbf{x}, \mathbf{u})\right)^2 d\mathbf{x}d\mathbf{u} + o_P(1).
\end{aligned}$$

Similarly, we can combine Lemmas 12 and 16 to get

$$\begin{aligned}
T_n^* =& nh^p \int \left(\hat{C}^*_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u}) - \hat{C}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u}) + C_{\hat{\theta}(x)}(\mathbf{u}) - C_{\hat{\theta}(\mathbf{x})^*}(\mathbf{u})\right)^2 W(d\mathbf{x}, d\mathbf{u}) \\
=& \int \left(\sqrt{nh^p}\Big(\hat{C}^*_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u}) - \hat{C}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{u})\Big) - \sqrt{nh^p}\Big(C_{\hat{\theta}(x)}(\mathbf{u}) - C_{\hat{\theta}^*(\mathbf{x})}(\mathbf{u})\Big)\right)^2 d\mathbf{x}d\mathbf{u} \\
=& \int \left(\frac{1}{\sqrt{nh^p}} \sum_{i=1}^n w_i(\mathbf{x}, h)\xi^*_i(\mathbf{x}, \mathbf{u}) + \varepsilon^*_n(\mathbf{x}, \mathbf{u}) - \frac{1}{\sqrt{nh^p}} \sum_{i=1}^n w_i(\mathbf{x}, h)\tilde{\xi}^*_i(\mathbf{x}, \mathbf{u}) - \tilde{\varepsilon}^*_n(\mathbf{x}, \mathbf{u})\right)^2 d\mathbf{x}d\mathbf{u} \\
=& \int \left(\frac{1}{\sqrt{nh^p}} \sum_{i=1}^n w_i(\mathbf{x}, h)\xi^*_i(\mathbf{x}, \mathbf{u}) - \frac{1}{\sqrt{nh^p}} \sum_{i=1}^n w_i(\mathbf{x}, h)\tilde{\xi}^*_i(\mathbf{x}, \mathbf{u})\right)^2 d\mathbf{x}d\mathbf{u} + o_P(1).
\end{aligned}$$

The proof is completed by the following conjecture.

Conjecture 17. *Under some regularity conditions,*

$$\left(\frac{1}{\sqrt{nh^p}} \sum_{i=1}^n (w_i(\mathbf{x}, h)\xi_i - \tilde{\xi}_i)\,,\ \frac{1}{\sqrt{nh^p}} \sum_{i=1}^n (w_i(\mathbf{x}, h)\xi^*_i - \tilde{\xi}^*_i)\right) \xrightarrow{law} (\mathsf{G}, \mathsf{G}^*),$$

where $\mathsf{G}^* \overset{law}{=} \mathsf{G}$ *and* G^* *is independent of* G*.*

By the continuous mapping theorem, we can finally conclude that $(T_n, T_n^*) \to (T, T^*)$ where $T^* \overset{\text{law}}{=} T$ and T^* is independent of T.