Electronic Journal of Statistics
Vol. 7 (2013) 2822–2850
ISSN: 1935-7524
DOI: 10.1214/13-EJS866

Statistical testing of covariate effects in conditional copula models

Elif F. Acar

Department of Statistics
University of Manitoba
Winnipeg, Manitoba, R3T 2N2, Canada
e-mail: elif.acar@umanitoba.ca

Radu V. Craiu

Department of Statistics
University of Toronto
Toronto, Ontario, M5S 3G3, Canada
e-mail: craiu@utstat.utoronto.ca

and

Fang Yao

Department of Statistics
University of Toronto
Toronto, Ontario, M5S 3G3, Canada
e-mail: fyao@utstat.utoronto.ca

Abstract: In conditional copula models, the copula parameter is deterministically linked to a covariate via the calibration function. The latter is of central interest for inference and is usually estimated nonparametrically. However, in many applications it is scientifically important to test whether the calibration function is constant or not. Moreover, a correct model of a constant relationship results in significant gains of statistical efficiency. We develop methodology for testing a parametric formulation of the calibration function against a general alternative and propose a generalized likelihood ratio-type test that enables conditional copula model diagnostics. We derive the asymptotic null distribution of the proposed test and study its finite sample performance using simulations. The method is applied to two data examples.

AMS 2000 subject classifications: Primary 62H20; secondary 62G10.
Keywords and phrases: Constant copula, covariate effects, dynamic copula, local likelihood, model diagnostics, nonparametric inference.

Received August 2012.

1. Introduction

Copulas are an important tool for modeling dependence. The recent development of conditional copulas by Patton (2006) widely expands the range of possible applications, as it allows covariate adjustment in copula structures and thus enables their use in regression settings. Specifically, if X is a covariate that affects the dependence between the continuous random variables Y_1 and Y_2, then the conditional joint distribution H_x of Y_1 and Y_2 given $X = x$ can be written as $H_x(y_1, y_2 \mid x) = C_x\{F_{1|x}(y_1 \mid x), F_{2|x}(y_2 \mid x) \mid x\}$, where $F_{i|x}$ is the conditional marginal distribution of Y_i given $X = x$ for $i = 1, 2$ and C_x is the conditional copula, i.e. the joint distribution of $U_1 \equiv F_{1x}(Y_1 \mid x)$ and $U_2 \equiv F_{2x}(Y_2 \mid x)$ given $X = x$.

When the conditional dependence structure is in the inferential focus, one needs to specify a functional connection between the covariate X and the copula C_x as in Jondeau and Rockinger (2006); Patton (2006); Bartram et al. (2007); Rodriguez (2007). Facing the problem of specifying a model for this functional relationship, one cannot usually rely on graphical or empirical pointers and must therefore use flexible models (e.g. semi- or non-parametric) that can potentially capture a large variety of patterns. There is now a rich body of work on modelling and estimating conditional copula models, e.g. Gijbels et al. (2011); Abegaz et al. (2012); Veraverbeke et al. (2011); Craiu and Sabeti (2012). In the context of a parametric copula family, Acar et al. (2011) have studied a nonparametric estimator of the calibration function $\eta(X)$ in

$$(U_1, U_2) \mid X = x \sim C_x\{u_1, u_2 \mid \theta(x) = g^{-1}(\eta(x))\}, \tag{1}$$

where $g : \Theta \to \mathbf{R}$ is a known link function that allows unrestricted estimation for η. In this model, the copula family, i.e. the form of dependence between uniform random variables U_1 and U_2 remains the same for each value of the covariate $X = x$, but the strength of the dependence between U_1 and U_2, measured by the copula parameter θ, is allowed to vary with X according to a smooth function η. Hence, in order to assess the covariate effect on the strength of dependence, one needs to infer the functional form of $\eta(X)$.

It is known that if a parametric model for $\eta(X)$ is suitable, then fitting a nonparametric model leads to an unnecessary loss of efficiency. For instance, in Table 1 in Acar et al. (2011) this loss is illustrated in the case of an underlying linear calibration function. It is thus of great practical importance to construct rigorous hypothesis tests for the specification of calibration functions.

Our development focuses on the hypotheses of the form H_0: "$\eta(\cdot)$ is linear in X" versus H_1: "$\eta(\cdot)$ is not linear in X" under the conditional copula model in (1). This class of hypotheses includes the important special case of $\mathrm{H}_0^{(c)}$: "$\eta(\cdot)$ is constant" versus $\mathrm{H}_1^{(c)}$: "$\eta(\cdot)$ is not constant". In most applications we have encountered, the constant calibration hypothesis is most relevant scientifically. In comparison, the precise specification of a non-constant calibration function (linear, quadratic, cubic) is of relatively much smaller importance. Therefore, the paper focuses on the constant/nonconstant dichotomy and uses the linear/nonlinear one to illustrate the possible generalizations of the testing procedure proposed here. From a statistical viewpoint, it is important to establish the validity of a constant calibration because then one can rely on inferential methods developed for the classical copula model. Canonical approaches based on likelihood ratio tests are possible when the calibration function is specified parametrically (Jondeau and Rockinger, 2006). Within a Bayesian approach in which regression splines are used to model η, Craiu and Sabeti (2012) suggest that novel criteria for testing $\mathrm{H}_0^{(c)}$ are needed.

Hypotheses like H_0 cannot be tested using the canonical likelihood ratio test (LRT) because estimation under the alternative hypothesis is performed non-parametrically. Exploration of the asymptotic distribution of the ratio test falls within the scope of the generalized likelihood ratio test (GLRT) developed by Fan et al. (2001) for testing a parametric null hypothesis versus a nonparametric alternative hypothesis. Since nonparametric maximum likelihood estimators are difficult to obtain and may not even exist, Fan et al. (2001) suggested using any reasonable nonparametric estimator under the alternative model. In particular, using a local polynomial estimator to specify the alternative model of a number of hypothesis testing problems, Fan et al. (2001) showed that the null distribution of the GLRT statistic follows asymptotically a chi-square distribution with the number of degrees of freedom independent of the nuisance parameters. This result, referred to as *Wilks phenomenon*, holds for Gaussian white-noise model (Fan et al., 2001), varying-coefficient models, which include the regression model as a special case (Fan et al., 2001), spectral density (Fan and Zhang, 2004), additive models (Fan and Jiang, 2005) and single-index models (Zhang et al., 2010).

We expand the GLRT-based approach to testing the calibration function in conditional copula models. The test procedure employs the nonparametric estimator proposed by Acar et al. (2011) when evaluating the local likelihood under the alternative hypothesis. The major contribution of this work is the construction of a rigorous framework for such GLRTs in the conditional copula context. It is worth mentioning that the proposal can easily accommodate the test for an arbitrary parametric form specified under the null hypothesis. The description of the test, the derivation of its asymptotic null distribution and the discussion of practical implementation are included in Section 2. The finite sample performance of the test is illustrated using simulations and two data examples in Section 3 and 4, respectively. The paper ends with concluding remarks.

2. Generalized likelihood ratio test for copula functions

The construction of the GLRT is detailed under the assumption that the conditional marginal distributions $U_1 \equiv F_{1x}(Y_1 \mid x)$ and $U_2 \equiv F_{2x}(Y_2 \mid x)$ are known. A discussion of the impact of estimating the conditional marginal distributions is provided at the end of this section.

Suppose that $\{(U_{11}, U_{21}, X_1), \dots, (U_{1n}, U_{2n}, X_n)\}$ is a random sample from the conditional copula model (1). The null hypothesis of interest restricts the space of calibration functions to a subspace $\mathfrak{f}$ that is fully specified parametrically. Without loss of generality, we assume $\mathfrak{f} = \mathfrak{f}_L = \{\eta(\cdot) : \exists\ a_0, a_1 \in \mathbb{R}$ such that $\eta(X) = a_0 + a_1 X, \quad \forall X \in \mathcal{X}\}$ is the set of all linear functions on $\mathcal{X}$. Then we are interested in testing

$$\mathrm{H}_0 : \eta(\cdot) \in \mathfrak{f}_L \qquad \text{versus} \qquad \mathrm{H}_1 : \eta(\cdot) \notin \mathfrak{f}_L. \tag{2}$$

In what follows, we assume that the density c_x of C_x exists and for simplicity we use the notation $\ell(t, u_1, u_2) = \ln c_x\{u_1, u_2; g^{-1}(t)\}$. Furthermore, the first and second partial derivatives of ℓ with respect to t are assumed to exist and are denoted by $\ell_j(t, u_1, u_2) = \partial^j \ell(t, u_1, u_2)/\partial t^j$, for $j = 1, 2$.

2.1. Proposed GLRT for the conditional copula model

A natural way to approach (2) is through the likelihood ratio of the *restricted* (i.e., conditional copula with a linear calibration function) and the *full* (i.e., conditional copula with an arbitrary calibration function) models, or equivalently, through the difference

$$\sup_{\eta(\cdot)\notin\mathfrak{f}_L} \{\mathbb{L}_n(\mathrm{H}_1)\} - \sup_{\eta(\cdot)\in\mathfrak{f}_\mathrm{L}} \{\mathbb{L}_\mathrm{n}(\mathrm{H}_0)\},$$

where

$$\begin{aligned} \mathbb{L}_n(\mathrm{H}_0) &= \sum_{i=1}^n \ell(a_0 + a_1 X_i, U_{1i}, U_{2i}), \\ \mathbb{L}_n(\mathrm{H}_1) &= \sum_{i=1}^n \ell(\eta(X_i), U_{1i}, U_{2i}). \end{aligned}$$

The supremum of the log-likelihood function under the null hypothesis is given by

$$\mathbb{L}_n(\mathrm{H}_0, \tilde{\eta}) = \sum_{i=1}^n \ell(\tilde{\eta}(X_i), U_{1i}, U_{2i}).$$

where $\tilde{\eta}(X) = \tilde{a}_0 + \tilde{a}_1 X$, with $\tilde{\boldsymbol{a}} = (\tilde{a}_0, \tilde{a}_1)$ denoting the maximum likelihood estimator of the parameter $\boldsymbol{a} = (a_0, a_1)$.

Under the alternative, the general unknown form of $\eta(\cdot)$ adds significant complexity to the calculation of the supremum. We use the nonparametric estimator of $\eta(\cdot)$ proposed by Acar et al. (2011) to define the log-likelihood under the full model. Specifically, for each observation X_i in a neighbourhood of an interior point x, we approximate $\eta(X_i)$ linearly by

$$\eta(X_i) \;\approx\; \eta(x) + \eta'(x)(X_i - x) \;\equiv\; \beta_0 + \beta_1(X_i - x),$$

provided that $\eta(x)$ is twice continuously differentiable. Estimates of $\boldsymbol{\beta} = (\beta_0, \beta_1)$ and of $\eta(x) = \beta_0$, are then obtained by maximizing a kernel-weighted local likelihood function

$$\mathcal{L}(\boldsymbol{\beta}, x) = \sum_{i=1}^n \ell\{\beta_0 + \beta_1(X_i - x), U_{1i}, U_{2i}\}\, K_h(X_i - x), \tag{3}$$

where $h > 0$ is a bandwidth parameter controlling the size of the neighbourhood around x, K is a symmetric kernel density function and $K_h(\cdot) = K(\cdot/h)/h$

weighs the contribution of each data point based on their proximity to x. Similarly, if one uses a pth order local polynomial estimator, the local linear approximation in (3) will be replaced by $\sum_{\ell=0}^{p} \beta_\ell (X_i - x)^\ell$ and the resulting estimator is given by $\hat{\eta}_h(x) = \hat{\beta}_0$. The bandwidth h is chosen to maximize the cross-validation criterion

$$\mathcal{B}(h) = \sum_{i=1}^{n} \ln c(U_{1i}, U_{2i} \mid \hat{\theta}_h^{(-i)}(X_i)),$$

where $\hat{\theta}_h^{(-i)}(X_i)$ is the estimate of the copula parameter θ at X_i when the ith sample (U_{1i}, U_{2i}) is left out.

Then we evaluate the log-likelihood function under the alternative hypothesis of (2) as

$$\mathbb{L}_n(\mathrm{H}_1, \hat{\eta}_h) = \sum_{i=1}^{n} \ell\{\hat{\eta}_h(X_i), U_{1i}, U_{2i}\}.$$

The difference between the two log-likelihoods allows us to evaluate the evidence in the data in favor of (or against) the null model. Hence, the generalized likelihood ratio statistic is given by

$$\lambda_n(h) = \mathbb{L}_n(\mathrm{H}_1, \hat{\eta}_h) - \mathbb{L}_n(\mathrm{H}_0, \tilde{\eta}). \tag{4}$$

While large values of $\lambda_n(h)$ suggest the rejection of the null hypothesis, we need to determine the rejection region for the test. In order to inform the decision in finite samples we investigate the asymptotic distribution of the GLRT statistic under the null hypothesis.

2.2. Asymptotic distributions of proposed GLRT statistic

To facilitate our presentation we introduce the following notation. Let $f(x) > 0$ be the density function of X with support $\mathcal{X}$ and denote by $|\mathcal{X}|$ the range of the covariate X. Also, denote by $K * K$ the convolution of the kernel K and define

$$\begin{aligned}
\mu_n &= \frac{|\mathcal{X}|}{h}\left(K(0) - \frac{1}{2}\int K^2(t)\mathrm{d}t\right) = \frac{|\mathcal{X}|}{h} c_K, \\
\nu_n &= \frac{2|\mathcal{X}|}{h}\int (K(t) - \frac{1}{2}K * K(t))^2 \mathrm{d}t, \\
c_K &= K(0) - \frac{1}{2}\int K^2(t)\mathrm{d}t.
\end{aligned}$$

The following result states that the GLRT statistic follows asymptotically a normal or equivalently a chi-square distribution in the case of negligible bias, where the mean and variance are related to the quantities μ_n and ν_n, respectively. The technical conditions and proofs are deferred to Appendix I.

Theorem 1. *Assume that the conditions* (A1)–(A7) *in Appendix I hold and the GLRT statistic* $\lambda_n(h)$ *is constructed from* (4) *with a local linear estimator.*

Then, as $h \to 0$ *and* $nh^{3/2} \to \infty$,

$$\nu_n^{-1/2}(\lambda_n(h) - \mu_n + d_n) \xrightarrow{\mathcal{L}} N(0,1), \tag{5}$$

where $d_n = O_p(nh^4 + n^{1/2}\, h^2)$.

Furthermore, if η *is linear or* $nh^{9/2} \to 0$*, then, as* $nh^{3/2} \to \infty$,

$$r_K \lambda_n(h) \overset{\text{asym}}{\sim} \chi^2_{r_K\ \mu_n}, \tag{6}$$

where $r_K = 2\ \mu_n/\nu_n$.

2.3. Practical implications and further aspects

In order to use the asymptotic result proven in Theorem 1 in a practical setting, one needs to choose values for the bandwidth parameter and the order of the polynomial fitting. We give below guidelines for these choices and discuss other aspects relevant to the implementation of the GLRT.

Choice of bandwidth parameter. It should be noted that when η is linear, the asymptotic bias d_n becomes exactly zero, as shown in (8) in the Appendix I, and thus the condition $nh^{9/2} \to 0$ is not needed (the optimal bandwidth for estimation is of the order $n^{-1/5}$, see Acar et al., 2011). More importantly, this facilitates the calculation of the GLRT statistic $\lambda_n(h)$ in practice, since one can use directly the bandwidth used for estimation, chosen by the leave-one-out cross-validated likelihood (Acar et al., 2011). Our simulation study in Section 3 provides empirical support for this suggestion.

Order of local polynomial fitting and testing polynomial functions. The asymptotic results in Theorem 1 can be easily extended to the case where $\lambda_n(h)$ is based on a pth order local polynomial estimator, by substituting the kernel function K with its equivalent kernel K^* in c_K and r_K (see Fan and Gijbels, 1996, page 64, for the expression of K^*) induced by the local polynomial fitting (Fan et al., 2001). The asymptotic chi-square distribution (6) continues to hold if either η is a polynomial of degree p or $nh^{(4p+5)/2} \to 0$, as the asymptotic bias $d_n = O_p(nh^{2p+2} + n^{1/2}h^{p+1})$. The practical implication of such an extension is that, if the interest is to test a null hypothesis of a polynomial form $\eta(x) = \sum_{\ell=0}^{p} \beta_\ell x^\ell$, it is recommended to calculate $\lambda_n(h)$ using the local polynomial estimator with the corresponding degree p. This avoids the possible necessity of undersmoothing in order to have the asymptotic bias negligible.

Testing constancy of the copula parameter. As pointed out earlier, the hypothesis of η being constant is a special case of the linearity constraint and leads to the classical copula model (i.e., no covariate adjustment is required). If this hypothesis is of interest, using a local constant estimator, i.e., $p = 0$, to calculate $\lambda_n(h)$ may be more appealing (as confirmed by the simulations in Section 3) than using a local linear estimator. The latter tends to overfit even with large bandwidth when H_0 indeed holds, thus resulting in an inflated type I error.

Testing independence. While the test is quite general and can be used, in principle, to test any preset copula parameter value, extra caution is needed when testing independence. Under the null hypothesis of independence, the copula parameter is at the boundary of the parameter space (e.g., $\theta = 0$ for the Frank and Clayton families, and $\theta = 1$ for the Gumbel family), and the asymptotic result in Theorem 1 does not hold. For the canonical likelihood ratio test, the asymptotic null distribution is known to be a mixture of chi-squares when some parameters lie on boundary of the parameter space (Self and Liang, 1987). A similar result is expected to hold for the GLRT as the two tests are fairly similar in nature (see the next paragraph). However, an investigation in this direction may not be too much of practical value, considering that there are a number of independence tests in the copula literature (see, for instance, Genest and Rémillard, 2004).

Relationship to the canonical likelihood ratio test. One can conclude from Theorem 1 that the GLRT is fairly similar to the classical likelihood ratio test. The tabulated value of the scaling constant r_K is close to 2 for commonly used kernels. For instance, $r_K = 2.115$ for the commonly used Epanechnikov kernel $K(u) = 0.75(1 - u^2)\mathbf{1}_{\{|u|\leq 1\}}$. The degrees of freedom (df) $r_K\ c_K\ |\mathcal{X}|/h$ of the asymptotic null distribution of the GLRT tends to infinity when $h \to 0$, due to the nonparametric nature of the alternative hypothesis. One can interpret the quantity $|\mathcal{X}|/h$ as the number of nonintersecting intervals on $\mathcal{X}$, and thus $r_K\ c_K\ |\mathcal{X}|/h$ approximates the effective number of parameters in the nonparametric estimation. For the Epanechnikov kernel with $c_K = 0.45$, the degrees of freedom is given by $0.968\ |\mathcal{X}|/h$.

Impact of estimating conditional marginal distributions. An important aspect in (conditional) copula model implementation is the estimation of unknown (conditional) marginal distributions. Although the proposed GLRT procedure was presented assuming that the conditional marginal distributions are known, Theorem 1 provides a basis for more general approach where estimation is performed jointly or using a two-step method (i.e., parameters for the marginal distributions are estimated first and only subsequently the inference for the conditional copula parameters is performed).

In a conditional copula model with parametric conditional marginal distributions, one can easily accommodate joint estimation under the null hypothesis since the model for the calibration function is then parametric. On the other hand, joint analysis under the alternative requires iterative estimation of parametric conditional marginal distributions and nonparametric calibration function model. Although, such iterative procedures have been long studied for nonparametric regression problems, for instance in partially linear models, the problem of joint estimation has not been addressed yet in the conditional copula setting.

As shown in Appendix I, the asymptotic distribution of the GLR statistic $\lambda_n(h)$ is governed by the nonparametric part $\lambda_{1n}(h)$ since the parametric part λ_{2n}, which corresponds to the canonical likelihood ratio statistic, vanishes compared to $\lambda_{1n}(h)$. Hence, even one employs joint maximum likelihood estimation or a less efficient inference for margins approach (Joe, 2005) under the null hypothesis, the result in Theorem 1 will not change. Generally speaking, the proposed test remains valid as long as the conditional marginal distributions are estimated with the usual parametric rate $\sqrt{n}$, both under the null and under the alternative (see Remark 1 at the end of Appendix I). A two-step approach can be safely employed in the testing procedure provided that the conditional marginal distributions are estimated parametrically. This observation, in fact, motivated the use of parametric models to specify the conditional marginal distributions in the data examples of Section 4. Note that an alternative two-step approach is to use kernel-based smoothing methods as in Abegaz et al. (2012) in the first stage. However, this case requires a careful treatment as both the null and the alternative models will have nonparararametric convergence rates due to nonparametric specification of the conditional marginal distributions.

Choice of copula family. The proposed GLRT approach assumes that the true copula family is used in the test procedure. If the copula is misspecified, the asymptotic result in Theorem 1 would not hold. This phenomenon is similar to the departures from the chi-squared asymptotic limit exhibited by the canonical likelihood ratio tests under misspecified models. Furthermore, copula misspecification can lead to serious bias in the estimation results. Nonetheless, in our simulations we have observed a good performance of the cross-validated prediction error criterion of Acar et al. (2011) in choosing the true copula family.

3. Simulation study

We conduct simulations to evaluate the finite sample performance of the proposed test for the linear hypothesis given in (2). We consider four simulation scenarios corresponding to four calibration functions,

$$\begin{aligned}
\mathbf{M}_0^{(F)}: \quad \eta_0(X) &= 8,\\
\mathbf{M}_1^{(F)}: \quad \eta_1(X) &= 25 - 4.2\,X,\\
\mathbf{M}_2^{(F)}: \quad \eta_2(X) &= 1 + 2.5(3 - X)^2,\\
\mathbf{M}_3^{(F)}: \quad \eta_3(X) &= 12 + 8\sin(0.4\,X^2).
\end{aligned}$$

The copula used belongs to the Frank family and has the form

$$C(u_1, u_2|\theta) = -\frac{1}{\theta}\ln\left\{1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1}\right\}, \quad \theta \in (-\infty, \infty) \setminus \{0\}.$$

Since the range of θ is $\mathbb{R}\setminus\{0\}$ for the Frank copula, an identity link is used, i.e., $\theta_k(X) = \eta_k(X)$ for $k = 0, 1, 2, 3$. Similar findings, summarized in Appendix II, were obtained in simulations produced using the Clayton and Gumbel copulas in the true generating models.

As mentioned earlier, the copula family describes the overall shape of the dependence. For the Frank family, this shape is symmetric and shows no tail dependence as shown in the left panel of Figure 1. The four calibration models

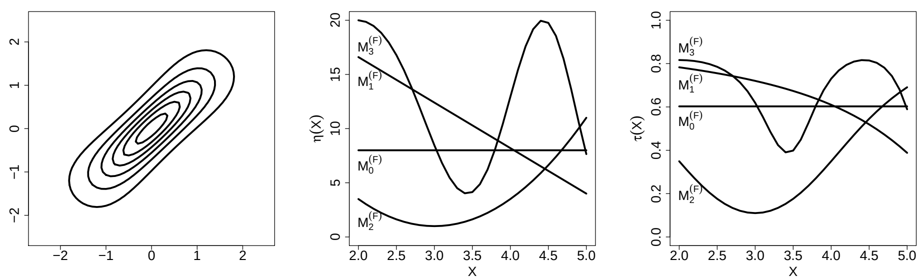

Fig 1. *Contour plot of the density of Frank copula under $\mathbf{M}_0^{(F)}$, illustrated with standard normal marginal distributions (left panel), and graphical summary of the four calibration models (middle panel) and the corresponding Kendall's tau values (right panel).*

are displayed in the middle panel of Figure 1 over the covariate range $(2,5)$. For a scale-free interpretation of the strength of dependence, we also provide a graphical summary of the variation in Kendall's tau for each η (see the right panel of Figure 1) obtained using the conversion

$$\tau = 1 + \frac{4}{\theta}\{D_1(\theta) - 1\},$$

where $D_1(\theta) = \frac{1}{\theta}\int_0^\theta \frac{t}{e^t-1}dt$ is the Debye function.

Our Monte Carlo experiment consists of 200 replicated samples of sizes $n = 50, 100, 200$ and 500 generated from each calibration model $\mathbf{M}_k^{(F)}$, $k = 0,1,2,3$. Specifically, under model $\mathbf{M}_k^{(F)}$, we first simulate the covariate values X_i from Uniform$(2,5)$. Then, for each $i = 1,\ldots,n$, we calculate the copula parameter $\theta_k(X_i) = \eta_k(X_i)$ induced by the calibration model $\mathbf{M}_k^{(F)}$, and simulate the uniform pairs (U_{1i}, U_{2i}) conditional on X_i from the Frank family with copula parameter $\theta_k(X_i)$. Throughout the simulations we have used the Epanechnikov kernel. For each Monte Carlo sample, the leave-one-out cross-validated likelihood method of Acar et al. (2011) is employed to select, out of 12 pilot values ranging from 0.33 to 2.96 and equally spaced in logarithmic scale, the optimum bandwidth h for the local polynomial estimation of the calibration function.

We have followed the suggestion made in Section 2 and have calculated the nonparametric estimator for η using a local polynomial of the same degree as specified by the null hypothesis. For instance, in Table 1, when testing $H_0 : \eta = c$, we consider a local constant estimator (with $p = 0$) for η under the alternative model. Subsequently, the GLRT statistic $\lambda_n(h)$ is computed using the same bandwidth h that is used for estimation. We also assume that in practice one would *first* test for constant calibration function and, conditional on rejection, would test for linear calibration. For this reason, in Table 1 we do not report the results of testing $H_0 : \eta(x) = a_0 + a_1 x$ when the generating model is $\mathbf{M}_0^{(F)}$.

TABLE 1
Demonstration of the proposed GLRT for testing the linear/constant null hypothesis H_0 at $\alpha = 0.10, 0.05$ and 0.01, respectively, under the Frank copula. Shown are the rejection frequencies assessed from 200 Monte Carlo replicates. The sample sizes are $n = 50, 100, 200$ and 500; the generating calibration models are shown in the "True Model" column. Those entries in the table reflecting the power of the testing procedure are shown in bold face

True Model		Null Model					
		$H_0 : \eta(x) = a_0 + a_1 x$			$H_0 : \eta = c$		
	n	.10	.05	.01	.10	.05	.01
$\mathbf{M}_0^{(F)}$	50	—	—	—	.075	.025	.015
	100	—	—	—	.110	.055	.010
	200	—	—	—	.105	.040	.020
	500	—	—	—	.110	.045	.005
$\mathbf{M}_1^{(F)}$	50	.060	.020	.005	**.695**	**.520**	**.255**
	100	.100	.060	.010	**.910**	**.870**	**.760**
	200	.100	.055	.005	**.995**	**.990**	**.955**
	500	.085	.055	.010	**1.00**	**1.00**	**1.00**
$\mathbf{M}_2^{(F)}$	50	**.425**	**.285**	**.085**	**.640**	**.515**	**.255**
	100	**.650**	**.510**	**.245**	**.940**	**.860**	**.620**
	200	**.895**	**.790**	**.560**	**1.00**	**1.00**	**.975**
	500	**1.00**	**1.00**	**.980**	**1.00**	**1.00**	**1.00**
$\mathbf{M}_3^{(F)}$	50	**.735**	**.635**	**.435**	**.755**	**.645**	**.385**
	100	**.865**	**.840**	**.780**	**.965**	**.945**	**.870**
	200	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
	500	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**

One can notice from Table 1 that the rejection rates under the null are very close to the target values of the type I error probabilities $\alpha \in \{0.1, 0.05, 0.01\}$, for both linear and constant nulls (models $\mathbf{M}_0^{(F)}$ and $\mathbf{M}_1^{(F)}$), except for the case with $n = 50$, which is slightly conservative. Since nonparametric methods often require relatively large samples, the latter observation confirms that caution is recommended when working with small samples. Overall, our approach leads to high power in detecting departures from the null, as one can see from the results generated under models $\mathbf{M}_1^{(F)}$, $\mathbf{M}_2^{(F)}$ and $\mathbf{M}_3^{(F)}$. For clearer visualization, the entries in the table that correspond to power are shown in bold face. As expected, the rejection rates increase with the sample size and the *nonlinearity* of the underlying calibration function (see the middle or right panel of Figure 1 for a visual comparison). For instance, when testing the linear calibration hypothesis, we observe lower rejection rates for $\mathbf{M}_2^{(F)}$, which is quadratic, than for $\mathbf{M}_3^{(F)}$, which exhibits more variation.

4. Data application

In this section, we apply the GLRT to the two data examples studied in Acar et al. (2011). Our aim is to check whether a constant copula model or a conditional copula model with a linear calibration function fits these examples reasonably well, i.e. whether the nonparametric calibration estimates of Acar et al. (2011) are in fact necessary.

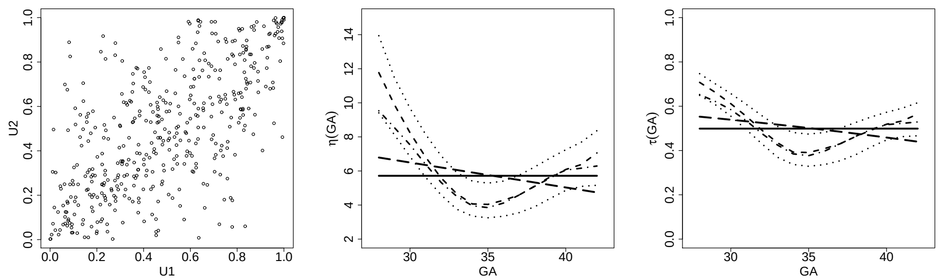

Fig 2. *Scatterplot of transformed birth weights (left panel), and the plots of calibration function estimates (middle panel) and the corresponding Kendall's tau values (right panel) under the Frank copula: maximum likelihood estimate of the constant calibration function (solid line), maximum likelihood estimate of the linear calibration function (long-dashed line), local constant estimates (dot-dashed line), local linear estimates (dashed line),* 90% *pointwise confidence intervals for the local linear estimates (dotted lines).*

4.1. Twin birth data

This data set contains information on 450 pairs of twins from the Matched Multiple Birth Data Set (MMB) of the National Center for Health Statistics. We consider the birth weights (BW_1, BW_2) of the first- and second-born twins who survived their first year, and whose mothers were between 18 and 40 years old. The gestational age GA is an important factor for fetal growth and is therefore included in the analysis as a covariate. The question of interest here is whether the dependence between the weights BW_1 and BW_2 varies with GA. Subsequent interrogations regarding the *exact* parametric forms of η are much less relevant. We transform the data on the uniform scale, as shown in the left panel of Figure 2, after fitting parametric marginal regression models. We use the Frank family of copulas to model the dependence structure, as this was the family chosen by Acar et al. (2011) according to their cross-validated prediction error criterion. The middle panel of Figure 2 shows the maximum likelihood estimates obtained under the constant calibration assumption (solid line), linear calibration assumption (long-dash line), the nonparametric estimates with $p = 0$ (dot-dashed line), $p = 1$ (dashed line) and 90% pointwise confidence intervals for the local linear estimates (dotted lines), obtained as in Acar et al. (2011). These results are also displayed in terms of Kendall's tau in the right panel of Figure 2.

As seen in Figure 2, the maximum likelihood estimates under constant and linear calibration assumptions are not within the pointwise confidence intervals of the local linear estimates, suggesting that these simple parametric formulations may not be appropriate. This empirical observation is confirmed by the GLRT tests, which yielded p-values smaller than 10^{-5} for both tests (test statistics are 13.58 on 3.92 df and 12.95 on 3.36 df for the constant and linear hypothesis, respectively).

Thus, we conclude that the variation in the strength of dependence between the twin birth weights at different gestational ages, as represented by the non-

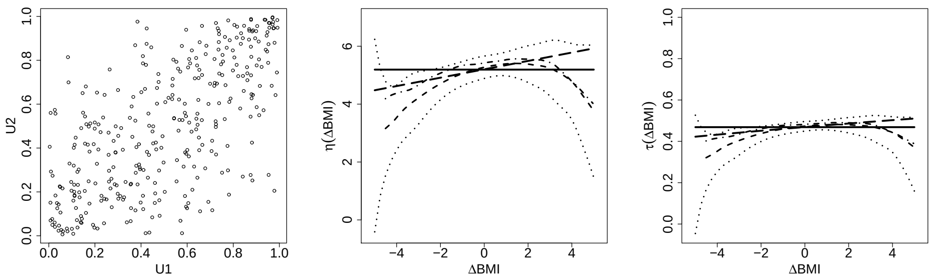

FIG 3. *Scatterplot of the transformed log-pulse pressures (left panel), and the plots of calibration function estimates (middle panel) and the corresponding Kendall's tau values (right panel) under the Frank copula: maximum likelihood estimate of the constant calibration function (solid line), maximum likelihood estimate of the linear calibration function (long-dashed line), local constant estimates (dot-dashed line), local linear estimates (dashed line),* 90% *pointwise confidence intervals for the local linear estimates (dotted lines).*

parametric estimates in the right panel of Figure 2 is statistically significant. The estimated nonlinear pattern indicates a relatively stronger dependence between the birth weights of the preterm (28–32 weeks) and post-term (38–42 weeks) twins compared to the twins delivered at term (33–37 weeks). While the factors affecting the twin fetal growth patterns are not fully known, the relatively stronger intra-twin dependence for preterm twins may be due to the fact that fat accumulation only begins in the third trimester of gestation (around 28 weeks). Hence, during the weeks 28–32, twins are expected to be still very *similar* in their growth. The increase in the strength of dependence after week 38, on the other hand, is more puzzling, but perhaps can be explained by intrauterine growth restriction factors.

4.2. Framingham heart study data

This data set comes from the Framingham Heart Study (FHS) and contains the log-pulse pressures of 348 subjects at the first two examination periods (1956 and 1962), denoted by $\log(PP_1)$ and $\log(PP_2)$, respectively, as well as the change in body mass index ΔBMI between these periods. Pulse pressure, defined as the difference between systolic and diastolic blood pressure, reflects arterial stiffness and is associated with an increased risk in stroke incidence. For the 348 subjects, who experienced a stroke during the rest of follow-up period, of interest is to investigate the dependence structure between $\log(PP_1)$ and $\log(PP_2)$ conditional on ΔBMI. The left panel of Figure 3 displays the conditional marginal distributions of the log-pulse pressures given ΔBMI, which are obtained parametrically as in Acar et al. (2011).

In conditional copula selection, Acar et al. (2011) used the cross-validated prediction errors only for the second log-pulse pressure as the selection criterion, and chose the Frank copula family. Proceeding with their choice, we obtained the calibration function estimates using the maximum likelihood estimation with constant and linear calibration forms and the nonparametric estimation with $p = 0$ and $p = 1$. The results are shown in the middle panel of Figure 3, and converted to the Kendall's tau scale in the right panel.

Based on the Figure 3 we suspect that a constant copula model may be appropriate, even though it is not fully contained in the pointwise confidence intervals. Note that the latter provides only partial guidance, and simultaneous confidence intervals are needed for a decisive visual conclusion. To decide whether the fitted constant copula model is appropriate, we perform the GLRT using the local constant estimates at the bandwidth value $h = 3.45$. This bandwidth choice leads to 2.66 df of the chi-square distribution. The difference between the log-likelihoods of the alternative and null conditional copula models is 0.91 and consequently the p-value is 0.514. Thus, we conclude that the change in body mass index does not have any significant effect on the strength of dependence between the two log-pulse pressures.

5. Conclusion

Adjusting statistical dependence for covariates via conditional copulas is an active area of research where model fitting and validation are currently in early development. This paper takes a first step towards establishing conditional copula model diagnostics by presenting a formal test of hypothesis for the calibration function. Inspired by the generalized likelihood ratio idea of Fan et al. (2001), the proposed test uses the local likelihood estimator of Acar et al. (2011) to specify the model under the alternative when testing a parametric calibration function hypothesis. The asymptotic null distribution of the test statistic, shown to be a chi-squared distribution with the number of degrees of freedom determined by the estimation-optimal bandwidth, is used to determine the rejection region in finite samples. Simulations suggest that the method has high power of detecting departures from the null model and yields the targeted type I error probability.

The GLRT procedure presented here can be easily adapted to test an arbitrary parametric calibration function. Furthermore, the approach can be extended to employ other nonparametric estimators, such as smoothing splines, although with additional effort of deriving the asymptotic null distribution. Nevertheless, the asymptotic null distribution may not always be appropriate for determining the rejection region in finite samples. While conditional bootstrap is usually used to assess the null distribution of the GLRT in regression-based problems, defining a similar bootstrap procedure in the conditional copula setting is not straightforward and requires further study.

Although the focus of the paper is on bivariate copulas, the GLRT approach is potentially applicable to multivariate copulas. However, complications arise in the latter case since the number of copula parameters is likely to increase and simultaneous testing is therefore necessary. This represents an interesting direction for future research.

One current restriction of the test is that it requires the covariate to be univariate. This restriction is mainly due to the lack of a nonparametric estimation procedure that can accommodate multiple covariates. The latter is subject of ongoing research together with a covariate selection method based on the GLRT framework.

Appendix I: Regularity conditions and technical proofs

The asymptotic distribution of the GLRT statistic relies on the following technical conditions. The conditions (A1)–(A3) are standard in nonparametric estimation and the conditions (A4)–(A7) are required to regularize the conditional copula density.

(A1) The density function $f(X) > 0$ of the covariate X is Lipschitz continuous, and X has a bounded support $\mathcal{X}$.
(A2) The kernel function $K(t)$ is a symmetric probability density function that is bounded and Lipschitz continuous.
(A3) The functions η and g^{-1} have $(p+1)$th continuous derivatives, where $p = 1$ when a local linear estimator is used for $\lambda_n(h)$.
(A4) The functions $\ell_1\{\eta(x), u_1, u_2\}$ and $\ell_2\{\eta(x), u_1, u_2\}$ exist and are continuous on $\mathcal{X} \times (0,1)^2$, and can be bounded by integrable functions of u_1 and u_2.
(A5) $\mathrm{E}\big|\{\ell_1(\eta(\mathrm{x}), \mathrm{u}_1, \mathrm{u}_2) \mid \mathrm{x}\}\big|^4 < \infty$.
(A6) $\mathrm{E}\{\ell_2(\eta(\mathrm{x}), \mathrm{u}_1, \mathrm{u}_2) \mid \mathrm{x}\}$ is Lipschitz continuous.
(A7) The function $\ell_2(t, u_1, u_2) < 0$ for all $t \in \mathbb{R}$, and $u_1, u_2 \in (0,1)$. For some integrable function k, and for t_1 and t_2 in a compact set,

$$|\ell_2(t_1, u_1, u_2) - \ell_2(t_2, u_1, u_2)| < k(u_1, u_2)|t_1 - t_2|.$$

In addition, for some constants $\xi > 2$ and $k_0 > 0$, $j = 1, 2, 3$,

$$E\Big\{\sup_{x, ||\boldsymbol{m}|| < k_0/\sqrt{nh}} |\ell_2(\bar{\eta}(x, X) + \boldsymbol{m}^T \boldsymbol{z}_x, U_1, U_2)| \times \Big|\frac{X - x}{h}\Big|^{j-1} K\Big(\frac{X - x}{h}\Big)\Big\}^{\xi} = O(1),$$

where $\bar{\eta}(x, X) = \eta(x) + \eta'(x)(X - x)$.

Before proving Theorem 1, we shall introduce additional notation. Let $\gamma_n = 1/\sqrt{nh}$ and define

$$\alpha_n(x) = \frac{\gamma_n^2}{\sigma^2(x) f(x)} \sum_{i=1}^{n} \ell_1(\eta(X_i), U_{1i}, U_{2i})\ K((X_i - x)/h),$$

$$R_n(x) = \frac{\gamma_n^2}{\sigma^2(x) f(x)} \sum_{i=1}^{n} \Big\{\ell_1(\bar{\eta}(x, X_i), U_{1i}, U_{2i}) - \ell_1(\eta(X_i), U_{1i}, U_{2i})\Big\} \times K((X_i - x)/h),$$

where $\sigma^2(x) = -\mathrm{E}\big[\ \ell_2\ \{\eta(\mathrm{x}), \mathrm{U}_1, \mathrm{U}_2\} \mid \mathrm{X} = \mathrm{x}\big]$ denotes the Fisher Information for $\eta(x)$ at any $x \in \mathcal{X}$.

Recall that $\bar{\eta}(x, X_i) = \eta(x) + \eta'(x)(X_i - x)$, define

$$\begin{aligned}
R_{n1} &= \sum_{k=1}^{n} \ell_1(\eta(X_k), U_{1k}, U_{2k})\ R_n(X_k),\\
R_{n2} &= -\sum_{k=1}^{n} \ell_2(\eta(X_k), U_{1k}, U_{2k})\ \alpha_n(X_k)\ R_n(X_k),\\
R_{n3} &= -\frac{1}{2}\sum_{k=1}^{n} \ell_2(\eta(X_k), U_{1k}, U_{2k})\ R_n^2(X_k).
\end{aligned}$$

and set

$$\begin{aligned}
T_{n1} &= \gamma_n^2 \sum_{i=1}^{n}\sum_{k=1}^{n} \frac{\ell_1(\eta(X_k), U_{1k}, U_{2k})}{\sigma^2(X_k) f(X_k)}\ \ell_1(\eta(X_i), U_{1i}, U_{2i})\ K((X_i - x)/h),\\
T_{n2} &= \gamma_n^4 \sum_{i=1}^{n}\sum_{j=1}^{n}\ \ell_1(\eta(X_i), U_{1i}, U_{2i})\ \ell_1(\eta(X_j), U_{1j}, U_{2j})\\
&\qquad \times \left\{ \sum_{k=1}^{n} \frac{\ell_2(\eta(X_k), U_{1k}, U_{2k})}{(\sigma^2(X_k) f(X_k))^2}\ K((X_i - x)/h) K((X_i - x)/h) \right\}.
\end{aligned}$$

Lemma 1–3 are used in our derivations, and their proofs are given at the end of this appendix.

Lemma 1. *Under conditions* (A1)–(A7)*,*

$$\hat{\eta}_h(x) - \eta(x) = \{\alpha_n(x) + R_n(x)\}\ (1 + o_p(1)).$$

Lemma 2. *Under conditions* (A1)–(A7)*, as* $h \to 0$ *and* $nh^{3/2} \to \infty$

$$\begin{aligned}
T_{n1} = \frac{1}{h} K(0) \mathrm{E}[\mathrm{f}^{-1}(\mathrm{X})] + \frac{1}{\mathrm{n}} \sum_{\mathrm{k} \neq \mathrm{i}} \frac{\ell_1(\eta(\mathrm{X_k}), \mathrm{U_{1k}}, \mathrm{U_{2k}})}{\sigma^2(\mathrm{X_k}) \mathrm{f}(\mathrm{X_k})} \ell_1(\eta(\mathrm{X_i}), \mathrm{U_{1i}}, \mathrm{U_{2i}})&\\
K_h\,(X_i - X_k) +\ o_p(h^{-1/2}),&
\end{aligned}$$

$$\begin{aligned}
T_{n2} = -\frac{1}{h} \mathrm{E}[\mathrm{f}^{-1}(\mathrm{X})] \int \mathrm{K}^2(\mathrm{t}) \mathrm{dt}\ -\ \frac{2}{\mathrm{nh}} \sum_{\mathrm{i<j}} \frac{\ell_1(\eta(\mathrm{X_i}), \mathrm{U_{1i}}, \mathrm{U_{2i}})}{\sigma^2(\mathrm{X_i}) \mathrm{f}(\mathrm{X_i})}&\\
\times\, \ell_1(\eta(X_j), U_{1j}, U_{2j}) K * K((X_j - X_i)/h) + o_p(h^{-1/2}).&
\end{aligned}$$

To introduce Lemma 3, we first restate a proposition in de Jong (1987), where the notation is adapted to ours. Let $X_1, X_2, \ldots$ be independent variables, and

$w_{ijn}(\cdot,\cdot)$ Borel functions such that $W(n) = \sum_{1\le i\le n}\sum_{1\le j\le n} w_{ijn}(X_i, X_j)$, and $W_{ij} = w_{ijn}(X_i, X_j) + w_{jin}(X_j, X_i)$, where the index n is suppressed in W_{ij}. Following de Jong (1987, Definition 2.1), W_n is called clean if the conditional expectations of W_{ij} vanish: $E[W_{ij}|X_i] = 0$ $a.s.$ for all $i, j \le n$.

Proposition 3.2 (de Jong, 1987) *Let $W(n)$ be clean with variance ν_n^*, if G_I, G_{II} and G_{IV} be of lower order than ν_n^{*2}, then*

$$\nu_n^{*-1/2} W(n) \xrightarrow{\mathcal{L}} N(0,1), \qquad n \to \infty,$$

where

$$\begin{aligned} G_I &= \sum_{1\le i<j\le n} \mathbb{E}(W_{ij}^4), \quad G_{II} = \sum_{1\le i<j<k\le n} \{\mathbb{E}(W_{ij}^2 W_{ik}^2) + \mathbb{E}(W_{ji}^2 W_{jk}^2) + \mathbb{E}(W_{ki}^2 W_{kj}^2)\},\\ G_{IV} &= \sum_{1\le i<j<k<l\le n} \{\mathbb{E}(W_{ij}W_{ik}W_{lj}W_{lk}) + \mathbb{E}(W_{ij}W_{il}W_{kj}W_{kl}) + \mathbb{E}(W_{ik}W_{il}W_{jk}W_{jl})\}. \end{aligned}$$

We now define the following U-statistic,

$$\begin{aligned} W(n) &= \frac{\sqrt{h}}{n} \sum_{i\ne j} \frac{1}{(\sigma^2(X_i) f(X_i))^2} \ell_1(\eta(X_j), U_{1j}, U_{2j})\ \ell_1(\eta(X_i), U_{1i}, U_{2i}) \\ &\qquad \{2K_h(X_j - X_i) - K_h * K_h(X_j - X_i)\}. \end{aligned} \tag{7}$$

Lemma 3. *Under conditions (A1)–(A7), W_n defined in (7) is clean and, as $h \to 0$ and $nh^{3/2} \to \infty$, $W(n) \xrightarrow{\mathcal{L}} N(0, \nu^*)$, where $\nu^* = 2\ ||2K - K*K||_2^2\ E[f^{-1}(X)]$.*

Proof of Theorem 1. To provide a general framework, we use $\eta(X_k)$ and $\tilde{\eta}(X_k)$ to denote the true value under the null hypothesis and its maximum likelihood estimator, respectively. Then, the GLRT statistic can be written as

$$\begin{aligned} \lambda_n(h) &= \sum_{k=1}^n [\ell(\hat{\eta}_h(X_k), U_{1k}, U_{2k}) - \ell(\eta(X_k), U_{1k}, U_{2k}) \\ &\qquad - \{\ell(\tilde{\eta}(X_k), U_{1k}, U_{2k}) - \ell(\eta(X_k), U_{1k}, U_{2k})\}] \\ &\equiv \lambda_{1n}(h) - \lambda_{2n}. \end{aligned}$$

Here λ_{2n} corresponds to the canonical likelihood ratio statistic and it is $\lambda_{1n}(h)$ that governs the asymptotic distribution of $\lambda_n(h)$.

To derive the asymptotic distribution of $\lambda_{1n}(h)$, first approximate $\ell(\hat{\eta}_h(X_k), U_{1k}, U_{2k})$ around $\eta(X_k)$

$$\begin{aligned} \lambda_{1n}(h) &\approx \sum_{k=1}^n \ell_1(\eta(X_k), U_{1k}, U_{2k})\ \{\hat{\eta}_h(X_k) - \eta(X_k)\} \\ &\qquad + \frac{1}{2} \sum_{k=1}^n \ell_2(\eta(X_k), U_{1k}, U_{2k})\ \{\hat{\eta}_h(X_k) - \eta(X_k)\}^2. \end{aligned}$$

Applying Lemma 1 and 2 yields

$$\begin{aligned}
-\lambda_{1n}(h) &= -h^{-1}\mathrm{E}[\mathrm{f}^{-1}(\mathrm{X})]\left\{\mathrm{K}(0)-\int \mathrm{K}^2(\mathrm{t})\mathrm{dt}/2\right\} \\
&\quad - n^{-1}\sum_{i\neq j}\frac{\ell_1(\eta(X_i),U_{1i},U_{2i})}{(\sigma^2(X_i)f(X_i))^2}\,\ell_1(\eta(X_j),U_{1j},U_{2j})K_h\left(X_j-X_i\right) \\
&\quad + n^{-1}\sum_{i<k}\frac{\ell_1(\eta(X_i),U_{1i},U_{2i})}{(\sigma^2(X_i)f(X_i))^2}\,\ell_1(\eta(X_j),U_{1j},U_{2j})K_h * K_h(X_j-X_i) \\
&\quad - R_{n1}+R_{n2}+R_{n3}+O_p\left(n^{-1}h^{-2}\right)+o_p(h^{-1/2}).
\end{aligned}$$

By calculating of the leading terms R_{n1}, R_{n2} and R_{n3}, one can show that

$$\begin{aligned}
R_{n1} &= \sum_{k=1}^{n}\frac{h^2}{2}\,\ell_1(\eta(X_k),U_{1k},U_{2k})\eta''(X_k)\int t^2\;K(t)dt(1+o_p(1)) \\
&= O_p(n^{1/2}h^2) \\
-R_{n2} &= \sum_{k=1}^{n}\frac{h^2}{4}\;\frac{\ell_1(\eta(X_k),U_{1k},U_{2k})}{\sigma^2(X_k)f(X_k)}\eta''(X_k)\;\omega_0(1+o_p(1)) = O_p(n^{1/2}h^2) \\
-R_{n3} &= \frac{nh^4}{8}E\eta''(X)^2\sigma^2(X)\;\omega_0(1+o_p(1)) = O_p(nh^4)
\end{aligned}$$

where $\omega_0=\int\int t^2(s+t)^2K(t)K(s+t)\;ds\;dt$. Thus,

$$R_{n3}-(R_{n1}-R_{n2})=O_p(nh^4+n^{1/2}h^2).$$

This results in

$$-\lambda_{1n}(h)=-\mu_n+d_n-h^{-1/2}\;W(n)/2+o_p(h^{-1/2})$$

where W_n is as defined in (7). Applying Lemma 3, we arrive at $W(n)\xrightarrow{\mathcal{L}} N(0,\nu^*)$, where $\nu^*=2\;||2K-K*K||_2^2\;E[f^{-1}(X)]$. Hence,

$$\nu_n^{-1/2}(\lambda_{1n}(h)-\mu_n+d_n)\xrightarrow{\mathcal{L}}N(0,1),$$

where $\nu_n=(4h)^{-1}\nu^*$. For the asymptotic null distribution of $\lambda_n(h)$, this result can be re-written as

$$\nu_n^{-1/2}\{(\lambda_{1n}(h)-\lambda_{2n})-\mu_n+d_n+\lambda_{2n}\}\xrightarrow{\mathcal{L}}N(0,1).$$

Since $\lambda_{2n}=O_p(1)$, it vanishes compared to $\lambda_{1n}(h)=O_p(h^{-1})$ and we obtain

$$\nu_n^{-1/2}(\lambda_n(h)-\mu_n+d_n)\xrightarrow{\mathcal{L}}N(0,1).$$

For the second result, note that the distribution $N(a_n,2a_n)$ is approximately same as the chi-square distribution with degrees of freedom a_n, for a sequence $a_n\to\infty$. Letting $a_n=2\mu_n^2/\nu_n$ and $r_K=2\mu_n/\nu_n$, we have

$$(2a_n)^{-1/2}(r_K\lambda_n(h)-a_n)\xrightarrow{\mathcal{L}}N(0,1),$$

provided that d_n vanishes. □

Proof of Lemma 1. Define

$$\boldsymbol{b} = \gamma_n^{-1}(\beta_0 - \eta(x), h(\beta_1 - \eta'(x)))^T,$$

so that each component has the same rate of convergence. Then, we have

$$\beta_0 + \beta_1(X_i - x) = \bar{\eta}(x, X_i) + \gamma_n \boldsymbol{b}^T \boldsymbol{z}_{i,x},$$

where $\boldsymbol{z}_{i,x} = (1, (X_i - x)/h)^T$. The local log-likelihood function can be re-written in terms of $\boldsymbol{b}$,

$$\mathcal{L}(\boldsymbol{b}) = \sum_{i=1}^{n} \ell(\bar{\eta}(x, X_i) + \gamma_n \boldsymbol{b}^T \boldsymbol{z}_{i,x}, U_{1i}, U_{2i}) K_h(X_i - x).$$

Note that $\hat{\boldsymbol{b}} = \gamma_n^{-1}(\hat{\beta}_0 - \eta(x), h(\hat{\beta}_1 - \eta'(x)))^T$ maximizes $\mathcal{L}(\boldsymbol{b})$. It also maximizes following normalized function,

$$\mathcal{L}^*(\boldsymbol{b}) = \sum_{i=1}^{n} \Big\{ \ell(\bar{\eta}(x, X_i) + \gamma_n \boldsymbol{b}^T \boldsymbol{z}_{i,x}, U_{1i}, U_{2i}) - \ell(\bar{\eta}(x, X_i), U_{1i}, U_{2i}) \Big\} K((X_i - x)/h),$$

which can be written as

$$\begin{aligned}
\mathcal{L}^*(\boldsymbol{b}) &= h\gamma_n \sum_{i=1}^{n} \ell_1(\bar{\eta}(x, X_i), U_{1i}, U_{2i})\ \boldsymbol{b}^T \boldsymbol{z}_{i,x} K_h(X_i - x) \\
&\quad + h\frac{\gamma_n^2}{2} \sum_{i=1}^{n} \ell_2(\bar{\eta}(x, X_i) + \boldsymbol{m_n}^T \boldsymbol{z}_{i,x}, U_{1i}, U_{2i})\ (\boldsymbol{b}^T \boldsymbol{z}_{i,x})^2\ K_h(X_i - x) \\
&= \boldsymbol{b}^T \Big\{ \gamma_n \sum_{i=1}^{n} \ell_1(\bar{\eta}(x, X_i), U_{1i}, U_{2i}) \boldsymbol{z}_{i,x} K((X_i - x)/h) \Big\} \\
&\quad + 2^{-1} \boldsymbol{b}^T \Big\{ \frac{1}{n} \sum_{i=1}^{n} \ell_2(\bar{\eta}(x, X_i) + \boldsymbol{m_n}^T \boldsymbol{z}_{i,x}, U_{1i}, U_{2i})\ \boldsymbol{z}_{i,x} \boldsymbol{z}_{i,x}^T\ K_h(X_i - x) \Big\} \boldsymbol{b}.
\end{aligned}$$

In the following, we will show that

$$n^{-1} \sum_{i=1}^{n} \ell_2(\bar{\eta}(x, X_i) + \boldsymbol{m_n}^T \boldsymbol{z}_{i,x}, U_{1i}, U_{2i})\ \boldsymbol{z}_{i,x} \boldsymbol{z}_{i,x}^T\ K_h(X_i - x) = -\Delta + o_p(1),$$

where $\Delta = \sigma^2(x) f_X(x) \left(\begin{smallmatrix} \mu_0; & \mu_1 \\ \mu_1; & \mu_2 \end{smallmatrix} \right)$, with $\mu_i = \int t^i K(t) dt$, and $o_p(1)$ is uniform in $x \in \mathcal{X}$ and $||\boldsymbol{b}|| < m_0$, for some fixed constant $m_0 > 0$. To show this, we need the following smoothness result. Let $A_n(x, \boldsymbol{m}) = \ell_2(\bar{\eta}(x, X) + \boldsymbol{m}^T \boldsymbol{z}_x, U_1, U_2)\ \boldsymbol{z}_x \boldsymbol{z}_x^T\ K_h(X - x)$, with $||\boldsymbol{m}|| < 1$. Then, under the conditions (A1)–(A6), we can show that

$$|A_n(x_1, \boldsymbol{m_1}) - A_n(x_2, \boldsymbol{m_2})|\ \le h^{-3}\ k(X, U_1, U_2)(||\boldsymbol{m_1} - \boldsymbol{m_2}|| + |x_1 - x_2|)$$

for some integrable function $k(X, U_1, U_2)$. Thus, using the triangle inequality,

$$\left|\frac{1}{n}\sum_{i=1}^{n}\ell_2(\bar{\eta}(x,X_i)+\boldsymbol{m_n}^T\boldsymbol{z}_{i,x},U_{1i},U_{2i})\ \boldsymbol{z}_{i,x}\boldsymbol{z}_{i,x}^T\ K_h(X_i-x)-(-\Delta)\right|$$

$$\leq \quad \frac{1}{n}\sum_{i=1}^{n}\Big|\{\ell_2(\bar{\eta}(x,X_i)+\boldsymbol{m_n}^T\boldsymbol{z}_{i,x},U_{1i},U_{2i})-\ell_2(\bar{\eta}(x,X_i),U_{1i},U_{2i})\}$$

$$\times\ \boldsymbol{z}_{i,x}\boldsymbol{z}_{i,x}^T K_h(X_i-x)\Big| + \sup_{\eta,x}\left[\frac{1}{n}\Big|\sum_{i=1}^{n}\{\ell_2(\bar{\eta}(x,X_i),U_{1i},U_{2i})-\ell_2(\eta(X_i),\right.$$

$$\left.U_{1i},U_{2i})\}\times \boldsymbol{z}_{i,x}\boldsymbol{z}_{i,x}^T\ K_h(X_i-x)\Big|\right] + \sup_{\eta,x}\left[\Big|\frac{1}{n}\sum_{i=1}^{n}\ell_2(\eta(X_i),U_{1i},U_{2i})\right.$$

$$\times\ \boldsymbol{z}_{i,x}\boldsymbol{z}_{i,x}^T\ K_h(X_i-x)-E\{\ell_2(\eta(X),U_1,U_2)\ \boldsymbol{z}_x\boldsymbol{z}_x^T\ K_h(X-x)|x\}\Big|$$

$$\left.+\Big|E\{\ell_2(\eta(X),U_1,U_2)\ \boldsymbol{z}_x\boldsymbol{z}_x^T\ K_h(X-x)|x\}+\Delta\Big|\right],$$

for η in a compact set and $x \in \mathcal{X}$. The first sum goes to zero by the previous argument and the Dominated Convergence theorem. Similarly, the second sum converges to zero provided that $hn^{(\xi-2)/\xi} = O(1)$ and $||\boldsymbol{b}|| < m_0$, for some fixed constant $m_0 > 0$. The first part in the last term goes to zero with probability one by the uniform weak law of large numbers and the second part vanishes by direct calculation. We thus obtain

$$\mathcal{L}^*(\boldsymbol{b}) \quad = \quad \boldsymbol{b}^T\ W_n(x) - 2^{-1}\boldsymbol{b}^T\Delta\boldsymbol{b}\ (1+o_p(1)),$$

uniformly for $x \in \mathcal{X}$, where

$$W_n(x) = \gamma_n\sum_{i=1}^{n}\ell_1(\bar{\eta}(x,X_i),U_{1i},U_{2i})\ \boldsymbol{z}_{i,x}K((X_i-x)/h).$$

Using the quadratic approximation lemma (Fan and Gijbels, 1996, p. 210),

$$\hat{\boldsymbol{b}} = \Delta^{-1}\ W_n(u) + o_p(1),$$

provided that W_n is a stochastically bounded sequence of random vectors. The first entry of $\hat{\boldsymbol{b}}$ directly yields the result, i.e.

$$\gamma_n^{-1}\{\hat{\eta}_h(x)-\eta(x)\} = \frac{\gamma_n}{\sigma^2(x)f(x)}\left[\sum_{i=1}^{n}\ell_1(\eta(X_i),U_{1i},U_{2i})K((X_i-x)/h)\right.$$

$$\left.+\sum_{i=1}^{n}\Big\{\ell_1(\bar{\eta}(x,X_i),U_{1i},U_{2i})-\ell_1(\eta(X_i),U_{1i},U_{2i})\Big\}K((X_i-x)/h)\right](1+o_p(1)).$$

Note that, when η is linear, then the second sum directly becomes zero as for each $i = 1, \ldots, n$

$$\bar{\eta}(x,X_i) = a_0 + a_1x + a_1(X_i - x) = \eta(X_i). \tag{8}$$

This is clearly also the case when η is constant. □

Proof of Lemma 2. Note that

$$T_{n1} = \gamma_n^2 \sum_{k=1}^{n} \frac{1}{\sigma^2(X_k)f(X_k)}[\ell_1(\eta(X_k), U_{1k}, U_{2k})]^2 \; K(0)$$
$$+ \gamma_n^2 \sum_{k \neq i} \frac{1}{\sigma^2(X_k)f(X_k)} \ell_1(\eta(X_i), U_{1i}, U_{2i}) \ell_1(\eta(X_k), U_{1k}, U_{2k}) \; K((X_i - X_k)/h).$$

The approximation of the first term

$$\gamma_n^2 \sum_{k=1}^{n} \frac{[\ell_1(\eta(X_k), U_{1k}, U_{2k})]^2}{\sigma^2(X_k)f(X_k)} \; K(0) = h^{-1}K(0)\,\mathbb{E}\, f^{-1}(X) + o_p(h^{-1/2})$$

yields the first result. We can decompose $T_{n2} = T_{n21} + T_{n22}$, where

$$T_{n21} = \frac{1}{(nh)^2} \sum_{i=1}^{n} [\ell_1(\eta(X_i), U_{1i}, U_{2i})]^2 \sum_{k=1}^{n} \frac{\ell_2(\eta(X_k), U_{1k}, U_{2k})}{(\sigma^2(X_k)f(X_k))^2} K^2((X_i - X_k)/h),$$

$$T_{n22} = \frac{1}{n^2} \sum_{i \neq j} \ell_1(\eta(X_i), U_{1i}, U_{2i}) \; \ell_1(\eta(X_j), U_{1j}, U_{2j}) \Big\{ \sum_{k=1}^{n} \frac{\ell_2(\eta(X_k), U_{1k}, U_{2k})}{(\sigma^2(X_k)f(X_k))^2}$$
$$K_h(X_i - X_k) K_h(X_j - X_k) \Big\}.$$

We deal with T_{n21} and T_{n22} separately. For T_{n21}, note that

$$T_{n21} = \frac{1}{(nh)^2} \sum_{k=1}^{n} \ell_1(\eta(X_k), U_{1k}, U_{2k})]^2 \frac{\ell_2(\eta(X_k), U_{1k}, U_{2k})}{(\sigma^2(X_k)f(X_k))^2} \; K^2(0)$$
$$+ \frac{1}{(nh)^2} \sum_{i \neq k} [\ell_1(\eta(X_i), U_{1i}, U_{2i})]^2 \frac{\ell_2(\eta(X_k), U_{1k}, U_{2k})}{(\sigma^2(X_k)f(X_k))^2} \; K^2((X_i - X_k)/h).$$

The first sum can be shown to be

$$\frac{1}{(nh)^2} \sum_{k=1}^{n} \sigma^2(X_k) \frac{\ell_2(\eta(X_k), U_{1k}, U_{2k})}{(\sigma^2(X_k)f(X_k))^2} K^2(0) + o_p(h^{-1/2}) = O_p(n^{-1}h^{-2}).$$

Therefore, let

$$V_n = \frac{2}{n(n-1)} \sum_{i<k} \{ \sigma^2(X_i) \frac{\ell_2(\eta(X_k), U_{1k}, U_{2k})}{(\sigma^2(X_k)f(X_k))^2}$$
$$+ \sigma^2(X_k) \frac{\ell_2(\eta(X_i), U_{1i}, U_{2i})}{(\sigma^2(X_i)f(X_i))^2} \} K_h^2\big(X_k - X_i\big),$$

and the second sum becomes $(V_n + o(1))/2 + O_p\left(n^{-3/2}h^{-2}\right) + o_p(h^{-1/2})$. The decomposition theorem for U-statistics (Hoeffding, 1948) allows us to show that $Var(V_n) = O(n^{-1}h^{-2})$ as follows. First note that the leading term of V_n is $-h^{-1}\,\mathbb{E}\, f^{-1}(X) \int K^2(t)dt$. Hence, as $nh \to \infty$ and $h \to 0$, we obtain

$$T_{n21} = -h^{-1}\,\mathbb{E}\, f^{-1}(X) \int K^2(t)dt + o_p(h^{-1/2}).$$

Similarly, we can decompose $T_{n22} = T_{n221} + T_{n222}$ with

$$T_{n221} = \frac{2}{n}\sum_{i<j} \ell_1(\eta(X_i), U_{1i}, U_{2i})\ \ell_1(\eta(X_j), U_{1j}, U_{2j})\ \frac{1}{n}\Big\{ \sum_{k\neq i,j} \frac{\ell_2(\eta(X_k), U_{1k}, U_{2k})}{(\sigma^2(X_k)f(X_k))^2} K_h(X_i - X_k)K_h(X_j - X_k)\Big\},$$

$$T_{n222} = \frac{K(0)}{n^2h}\sum_{i\neq j} \ell_1(\eta(X_i), U_{1i}, U_{2i})\ \ell_1(\eta(X_j), U_{1j}, U_{2j}) \times \Big\{ \frac{\ell_2(\eta(X_i), U_{1i}, U_{2i})}{(\sigma^2(X_i)f(X_i))^2} + \frac{\ell_2(\eta(X_j), U_{1j}, U_{2j})}{(\sigma^2(X_j)f(X_j))^2}\Big\} K_h(X_i - X_j).$$

For $k \neq i, j$, define

$$Q_{ijk,h} = \frac{\ell_2(\eta(X_k), U_{1k}, U_{2k})}{(\sigma^2(X_k)f(X_k))^2} K_h(X_k - X_i)K_h(X_k - X_j).$$

It can be easily shown that $Var(n^{-1}\sum_{k\neq i,j} Q_{ijk,h}) = O(n^{-1}h^{-2})$. Then,

$$T_{n221} = 2n^{-2}(n-2)\sum_{i<j} \ell_1(\eta(X_i), U_{1i}, U_{2i})\ \ell_1(\eta(X_j), U_{1j}, U_{2j}) \times\ \mathbb{E}(Q_{ijk,h}|X_i, X_j) + o_p(h^{-1/2}),$$

where

$$\mathbb{E}(Q_{ijk,h}|X_i, X_j) = -\ \{h\ \sigma^2(X_i)f(X_i)\}^{-1}\int K(t)\ K((X_j - X_i)/h)dt.$$

It is also easy to show $Var(T_{n222}) = O\left(n^{-2}h^{-3}\right)$, implying $T_{n222} = o_p(h^{-1/2})$.

Combining T_{n21}, T_{n221} and T_{n222} yields

$$T_{n2} = -\ \frac{1}{h}\,\mathbb{E}\,f^{-1}(X)\int K^2(t)dt - \frac{2}{nh}\sum_{i<j} \frac{\ell_1(\eta(X_i), U_{1i}, U_{2i})}{\sigma^2(X_i)f(X_i)} \ell_1(\eta(X_j), U_{1j}, U_{2j}) \times K * K((X_j - X_i)/h) + o_p(h^{-1/2}).$$

□

Proof of Lemma 3. Recall that

$$W(n) \quad = \quad n^{-1}h^{1/2}\sum_{i\neq j} \{\sigma^2(X_i)f(X_i)\}^{-2}\ell_1(\eta(X_j), U_{1j}, U_{2j})\ \ell_1(\eta(X_i), U_{1i}, U_{2i}) \{2K_h(X_j - X_i) - K_h * K_h(X_j - X_i)\}.$$

We shall show that W_n satisfies conditions in Proposition 3.2. Let

$$W_{ij} = n^{-1}h^{1/2}B_n(i,j)\ell_1(\eta(X_i), U_{1i}, U_{2i})\ \ell_1(\eta(X_j), U_{1j}, U_{2j}),$$

where

$$B_n(i,j) = b_1(i,j) + b_2(i,j) - b_3(i,j) - b_4(i,j),$$

and

$$b_1(i,j) = 2K_h(X_j - X_i)\{\sigma^2(X_i)f(X_i)\}^{-2}, \qquad b_2(i,j) = b_1(j,i),$$
$$b_3(i,j) = K_h * K_h(X_j - X_i)\{\sigma^2(X_i)f(X_i)\}^{-2}, \qquad b_4(i,j) = b_3(j,i).$$

Thus we can write $W(n) = \sum_{i<j} W_{ij}$, and $W(n)$ is clean directly follows from the first Bartlett identity. For the variance of $W(n)$, note that $Var(W(n)) = \sum_{i<j} E(W_{ij}^2)$. Thus we calculate $E[\{B_n(i,j)\ell_1(\theta(X_i), U_{1i}, U_{2i})\ \ell_1(\theta(X_j), U_{1j}, U_{2j})\}^2]$. To simplify our presentation, let $\ell_{1i} = \ell_1(\theta(X_i), U_{1i}, U_{2i})$ and denote the m-fold convolution at t by $K(t,m) = K * \cdots * K(t)$. Through direct calculations, we obtain

$$\begin{aligned}
E(b_1^2(i,j)\ \ell_{1i}^2\ \ell_{1j}^2) &= E\left[\frac{4}{h^2}\frac{\ell_{1i}^2\ \ell_{1j}^2}{\{\sigma^2(X_i)f(X_i)\}^2}K^2\left(\frac{X_j - X_i}{h}\right)\right] \\
&= \frac{4}{h^2}\int \frac{\sigma^2(X_1)}{\{\sigma^2(X_1)f(X_1)\}^2}\left\{\int \sigma^2(X_2)K^2\left(\frac{X_2 - X_1}{h}\right)f(X_2)dX_2\right\}f(X_1)dX_1 \\
&= \frac{4}{h}\int \frac{f^{-2}(X_1)}{\sigma^2(X_1)}\int \sigma^2(X_1)f(X_1)K^2(t)dtf(X_1)dX_1(1 + O(h)) \\
&= \frac{4}{h}K(0,2)Ef^{-1}(X)(1 + O(h)).
\end{aligned}$$

Similarly,

$$\begin{aligned}
E(b_2^2(i,j)\ \ell_{1i}^2\ \ell_{1j}^2) &= 4h^{-1}K(0,2)Ef^{-1}(X)(1 + O(h)), \\
E(b_3^2(i,j)\ \ell_{1i}^2\ \ell_{1j}^2) &= h^{-1}K(0,4)Ef^{-1}(X)(1 + O(h)), \\
E(b_4^2(i,j)\ \ell_{1i}^2\ \ell_{1j}^2) &= h^{-1}K(0,4)Ef^{-1}(X)(1 + O(h)), \\
E(b_1(i,j)b_2(i,j)\ \ell_{1i}^2\ \ell_{1j}^2) &= 4h^{-1}K(0,2)Ef^{-1}(X)(1 + O(h)), \\
E(b_1(i,j)b_3(i,j)\ \ell_{1i}^2\ \ell_{1j}^2) &= 2h^{-1}K(0,3)Ef^{-1}(X)(1 + O(h)), \\
E(b_1(i,j)b_4(i,j)\ \ell_{1i}^2\ \ell_{1j}^2) &= 2h^{-1}K(0,3)Ef^{-1}(X)(1 + O(h)), \\
E(b_2(i,j)b_3(i,j)\ \ell_{1i}^2\ \ell_{1j}^2) &= 2h^{-1}K(0,3)Ef^{-1}(X)(1 + O(h)), \\
E(b_2(i,j)b_4(i,j)\ \ell_{1i}^2\ \ell_{1j}^2) &= 2h^{-1}K(0,3)Ef^{-1}(X)(1 + O(h)), \\
E(b_3(i,j)b_4(i,j)\ \ell_{1i}^2\ \ell_{1j}^2) &= h^{-1}K(0,4)Ef^{-1}(X)(1 + O(h)).
\end{aligned}$$

Thus,

$$E[B_n(i,j)\ell_{1i}^2\ \ell_{1j}^2] = h^{-1}\{16K(0,2) - 16K(0,3) + 4K(0,4)\}Ef^{-1}(X)(1 + O(h)).$$

The leading term of $n^{-2}h\sum_{i<j} E[\{B_n(i,j)\ell_{1i}^2\ \ell_{1j}^2]$ yields

$$\nu^* = 2\{4K(0,2) - 4K(0,3) + K(0,4)\}Ef^{-1}(X) = 2\ ||2K - K * K||_2^2\ Ef^{-1}(X).$$

For the condition on G_I, note that $E(b_1(1,2)\ell_{11}\ell_{12})^4 = E(b_3(1,2)\ell_{11}\ell_{12})^4 = O(h^{-3})$. Then $E(W_{12}^4) = n^{-4}h^2O(h^3)$, which implies $G_I = O(n^{-2}h^{-1}) = o(1)$. Similarly, the condition on G_{II} can be verified by noting that $E(W_{12}^2 W_{13}^2) = O(E(W_{12}^4)) = O(n^{-4}h^{-1})$. Thus, $G_{II} = O(n^{-1}h^{-1}) = o(1)$. For the last condition we need to check the order of $E(W_{12}W_{23}W_{34}W_{41})$. Calculations for few terms yield,

$$\begin{aligned}
E(b_1^2(1,2)b_1^2(2,3)b_1^2(3,4)b_1^2(4,1)\ \ell_1'^2\ \ell_2'^2\ \ell_3'^2\ \ell_4'^2) &= O(h^{-1})\\
E(b_1^2(1,2)b_1^2(2,3)b_1^2(3,4)b_3^2(4,1)\ \ell_1'^2\ \ell_2'^2\ \ell_3'^2\ \ell_4'^2) &= O(h^{-1})\\
E(b_1^2(1,2)b_1^2(2,3)b_3^2(3,4)b_3^2(4,1)\ \ell_1'^2\ \ell_2'^2\ \ell_3'^2\ \ell_4'^2) &= O(h^{-1})\\
E(b_1^2(1,2)b_3^2(2,3)b_3^2(3,4)b_3^2(4,1)\ \ell_1'^2\ \ell_2'^2\ \ell_3'^2\ \ell_4'^2) &= O(h^{-1})\\
E(b_3^2(1,2)b_3^2(2,3)b_3^2(3,4)b_3^2(4,1)\ \ell_1'^2\ \ell_2'^2\ \ell_3'^2\ \ell_4'^2) &= O(h^{-1}).
\end{aligned}$$

Since terms with other combinations will be of the same order, we conclude that

$$E(W_{12}W_{23}W_{34}W_{41}) = n^{-4}h^2O(h^{-1}) = O(n^{-4}h),$$

and $G_{IV} = O(h) = o(1)$. This completes the proof. □

Remark 1. In the case where conditional marginal distributions are estimated, say by $\widehat{U}_1$ and $\widehat{U}_2$, both under the null and under the alternative, the GLR statistic takes the form

$$\lambda_n(h) = \sum_{k=1}^{n}[\ell(\hat{\eta}_h(X_k),\widehat{U}_{1k},\widehat{U}_{2k}) - \ell(\tilde{\eta}(X_k),\widehat{U}_{1k},\widehat{U}_{2k})],$$

which can be re-written as

$$\begin{aligned}
\lambda_n(h) &= \sum_{k=1}^{n}[\ell(\hat{\eta}_h(X_k),\widehat{U}_{1k},\widehat{U}_{2k}) - \ell(\eta(X_k),U_{1k},U_{2k})\\
&\qquad -\{\ell(\tilde{\eta}(X_k),\widehat{U}_{1k},\widehat{U}_{2k}) - \ell(\eta(X_k),U_{1k},U_{2k})\}].\\
&\equiv \lambda_{1n}(h) - \lambda_{2n}.
\end{aligned}$$

In this case, specific to the estimation method used in $\widehat{U}_1$ and $\widehat{U}_2$, one has to revise Lemma 1 and Theorem 1. Nevertheless, denoting the partial derivatives by

$$\ell_{rsq}(t,u_1,u_2) = \frac{\partial^{r+s+q}\ell(t,u_1,u_2)}{\partial t^r\ \partial u_1^s\ \partial u_2^q},$$

for arbitrary integers r, s, q, we can provide a fairly general argument on the asymptotic behaviour of $\lambda_n(h)$ using three-dimensional Taylor approximations,

$$\lambda_{1n}(h) \approx \sum_{k=1}^{n}\ell_{100}(\eta(X_k),U_{1k},U_{2k})\ \{\hat{\eta}_h(X_k) - \eta(X_k)\}$$

$$
\begin{aligned}
&+ \sum_{k=1}^{n} \ell_{010}(\eta(X_k), U_{1k}, U_{2k}) \, \{\widehat{U}_{1k} - U_{1k}\} \\
&+ \sum_{k=1}^{n} \ell_{001}(\eta(X_k), U_{1k}, U_{2k}) \, \{\widehat{U}_{2k} - U_{2k}\} \\
&+ \frac{1}{2} \sum_{k=1}^{n} \ell_{200}(\eta(X_k), U_{1k}, U_{2k}) \, \{\hat{\eta}_h(X_k) - \eta(X_k)\}^2 \\
&+ \frac{1}{2} \sum_{k=1}^{n} \ell_{020}(\eta(X_k), U_{1k}, U_{2k}) \, \{\widehat{U}_{1k} - U_{1k}\}^2 \\
&+ \frac{1}{2} \sum_{k=1}^{n} \ell_{002}(\eta(X_k), U_{1k}, U_{2k}) \, \{\widehat{U}_{2k} - U_{2k}\}^2 \\
&+ \sum_{k=1}^{n} \ell_{110}(\eta(X_k), U_{1k}, U_{2k}) \, \{\hat{\eta}_h(X_k) - \eta(X_k)\}\{\widehat{U}_{1k} - U_{1k}\} \\
&+ \sum_{k=1}^{n} \ell_{101}(\eta(X_k), U_{1k}, U_{2k}) \, \{\hat{\eta}_h(X_k) - \eta(X_k)\}\{\widehat{U}_{2k} - U_{2k}\} \\
&+ \sum_{k=1}^{n} \ell_{011}(\eta(X_k), U_{1k}, U_{2k}) \, \{\widehat{U}_{1k} - U_{1k}\}\{\widehat{U}_{2k} - U_{2k}\},
\end{aligned}
$$

and

$$
\begin{aligned}
\lambda_{2n} \approx & \sum_{k=1}^{n} \ell_{100}(\eta(X_k), U_{1k}, U_{2k}) \, \{\tilde{\eta}(X_k) - \eta(X_k)\} \\
&+ \sum_{k=1}^{n} \ell_{010}(\eta(X_k), U_{1k}, U_{2k}) \, \{\widehat{U}_{1k} - U_{1k}\} \\
&+ \sum_{k=1}^{n} \ell_{001}(\eta(X_k), U_{1k}, U_{2k}) \, \{\widehat{U}_{2k} - U_{2k}\} \\
&+ \frac{1}{2} \sum_{k=1}^{n} \ell_{200}(\eta(X_k), U_{1k}, U_{2k}) \, \{\tilde{\eta}(X_k) - \eta(X_k)\}^2 \\
&+ \frac{1}{2} \sum_{k=1}^{n} \ell_{020}(\eta(X_k), U_{1k}, U_{2k}) \, \{\widehat{U}_{1k} - U_{1k}\}^2 \\
&+ \frac{1}{2} \sum_{k=1}^{n} \ell_{002}(\eta(X_k), U_{1k}, U_{2k}) \, \{\widehat{U}_{2k} - U_{2k}\}^2 \\
&+ \sum_{k=1}^{n} \ell_{110}(\eta(X_k), U_{1k}, U_{2k}) \, \{\tilde{\eta}(X_k) - \eta(X_k)\}\{\widehat{U}_{1k} - U_{1k}\} \\
&+ \sum_{k=1}^{n} \ell_{101}(\eta(X_k), U_{1k}, U_{2k}) \, \{\tilde{\eta}(X_k) - \eta(X_k)\}\{\widehat{U}_{2k} - U_{2k}\}
\end{aligned}
$$

$$+ \sum_{k=1}^{n} \ell_{011}(\eta(X_k), U_{1k}, U_{2k}) \{\widehat{U}_{1k} - U_{1k}\}\{\widehat{U}_{2k} - U_{2k}\}.$$

After directly cancelling out the common terms, i.e. the ones only involving the differences $\{\widehat{U}_{ik} - U_{ik}\}$, $i = 1, 2$, we obtain

$$\begin{aligned} \lambda_n(h) \approx & \sum_{k=1}^{n} \big[\ell_{100}(\eta(X_k), U_{1k}, U_{2k}) + \ell_{110}(\eta(X_k), U_{1k}, U_{2k})\{\widehat{U}_{1k} - U_{1k}\} \\ & \qquad + \ell_{101}(\eta(X_k), U_{1k}, U_{2k}) \{\widehat{U}_{2k} - U_{2k}\}\big]\{\hat{\eta}_h(X_k) - \eta(X_k)\} \\ & + \frac{1}{2} \sum_{k=1}^{n} \ell_{200}(\eta(X_k), U_{1k}, U_{2k}) \{\hat{\eta}_h(X_k) - \eta(X_k)\}^2 \\ & - \sum_{k=1}^{n} \big[\ell_{100}(\eta(X_k), U_{1k}, U_{2k}) + \ell_{110}(\eta(X_k), U_{1k}, U_{2k})\{\widehat{U}_{1k} - U_{1k}\} \\ & \qquad + \ell_{101}(\eta(X_k), U_{1k}, U_{2k}) \{\widehat{U}_{2k} - U_{2k}\}\big]\{\tilde{\eta}(X_k) - \eta(X_k)\} \\ & - \frac{1}{2} \sum_{k=1}^{n} \ell_{200}(\eta(X_k), U_{1k}, U_{2k}) \{\tilde{\eta}(X_k) - \eta(X_k)\}^2 \end{aligned}$$

If the conditional marginal distributions are estimated parametric rates, for instance, as in Section 4, then the second and the third terms in the first sum, and the last two sums will vanish. Hence, the result in Theorem 1 will hold. However, if the conditional marginal distributions are estimated with nonparametric rates (Abegaz et al., 2012), then the terms involving $\widehat{U}_{ik}$, $i = 1, 2$ are expected to alter the asymptotic distribution of the GLRT. The latter case requires further study.

Appendix II: Additional Simulation Results

We investigate the finite sample performance of the proposed GLRT in simulations also using the Clayton and Gumbel families. Together with the Frank family, these copulas cover wide range of dependence patterns. The Clayton family has the copula function

$$C(u_1, u_2) = \left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{-\frac{1}{\theta}}, \qquad \theta \in (0, \infty),$$

and exhibits lower tail dependence; while the Gumbel copula has the form

$$C(u_1, u_2) = \exp\left[-\{(-\ln u_1)^{\theta}(-\ln u_2)^{\theta}\}^{\frac{1}{\theta}}\right], \quad \theta \in [1, \infty),$$

and exhibits upper tail dependence (see the top left and right panels of Figure 4). Considering their restricted copula parameter range, the inverse link functions are chosen as $g^{-1}(t) = \exp(t)$ for the Clayton copula, and $g^{-1}(t) = \exp(t) + 1$ for the Gumbel copula.

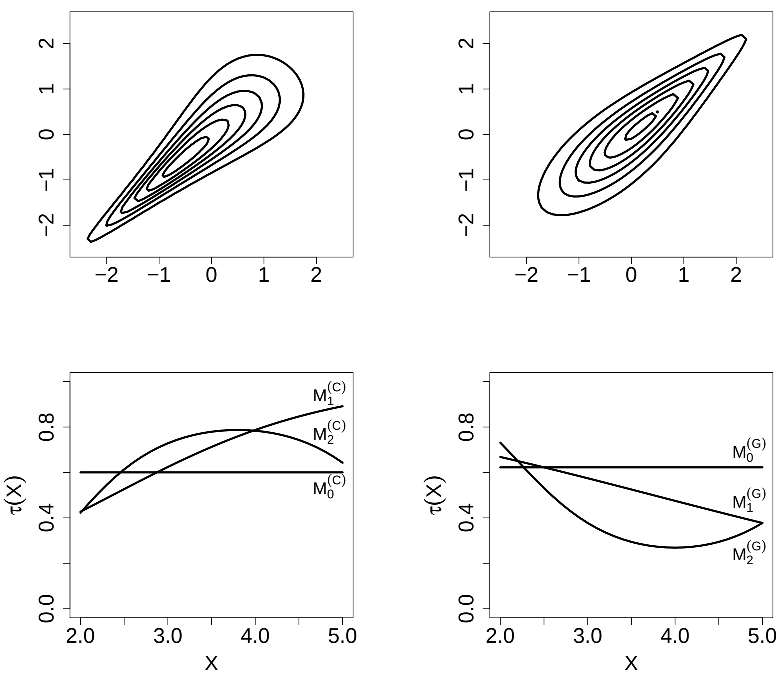

FIG 4. *Contour plots of the densities of the Clayton (top left panel) and Gumbel copulas (top right panel) under* $\mathbf{M}_0^{(C)}$ *and* $\mathbf{M}_0^{(G)}$*, respectively, illustrated with standard normal marginal distributions; and graphical summaries of the calibration models under the Clayton (bottom left panel) and Gumbel copulas (bottom right panel) in the Kendall's tau scale.*

In these set of simulations, we focus on the following constant, linear and quadratic calibration models, indexed by $0, 1$ and 2, respectively. We also indicate the first letter of the data generating copula as superscript. The three calibration models for the Clayton family are

$$\begin{aligned}
\mathbf{M}_0^{(\mathrm{C})}: \quad \eta_0(X) &= 1.1, \\
\mathbf{M}_1^{(\mathrm{C})}: \quad \eta_1(X) &= -1.2 + 0.8\,X, \\
\mathbf{M}_2^{(\mathrm{C})}: \quad \eta_2(X) &= 2 - 0.5\,(X - 3.8)^2,
\end{aligned}$$

and for the Gumbel family, we consider

$$\begin{aligned}
\mathbf{M}_0^{(\mathrm{G})}: \quad \eta_0(X) &= 0.5, \\
\mathbf{M}_1^{(\mathrm{G})}: \quad \eta_1(X) &= 1.5 - 0.4\,X, \\
\mathbf{M}_2^{(\mathrm{G})}: \quad \eta_2(X) &= -1 + 0.5(X - 4)^2.
\end{aligned}$$

Figure 4 displays the variations in the strength of dependence for these calibration models, summarized separately for each copula family in the Kendall's tau scale using the conversions $\tau = \theta/(\theta+2)$ for the Clayton copula, and $\tau = 1-1/\theta$ for the Gumbel copula.

We consider sample sizes of $n = 100, 200$ and 500, and generate 200 replicated samples following the same steps as in Section 3. First, we simulate the covariate values X_i from Uniform $(2, 5)$. Then, for each $i = 1, 2, \ldots, n$, we obtain the copula parameter, θ_i imposed by the given calibration and link functions, and finally simulate the pairs $(U_{1i}, U_{2i}) \mid X_i$ from the underlying family with the parameter θ_i. The results for testing the linear and constant null hypotheses are obtained using the local linear and local constant estimates, respectively, at the optimum bandwidth values chosen according to the leave-one-out cross-validated likelihood method among the same 12 pilot bandwidth values considered in Section 3. As can be seen in Table 2, the empirical rejection rates under the null hypotheses roughly attain the nominal type I error rates $\alpha \in \{0.1, 0.05, 0.01\}$ for both the Clayton (models $\mathbf{M}_0^{(C)}$ and $\mathbf{M}_1^{(C)}$) and Gumbel families (models $\mathbf{M}_0^{(G)}$ and $\mathbf{M}_1^{(G)}$). Consistent with the results in Section 3, the empirical power in detecting departures from the null depends heavily on the underlying calibration model. For instance, in both Clayton and Gumbel families, the quadratic models $\mathbf{M}_2^{(C)}$ and $\mathbf{M}_2^{(G)}$ show modest departures from linearity (see bottom panels of Figure 4), therefore moderate rejection rates were observed in the cases with smaller sample size.

Table 2

Demonstration of the proposed GLRT for testing the linear/constant null hypothesis H_0 at $\alpha = 0.10, 0.05$ and 0.01, respectively, under the Clayton and Gumbel copulas. Shown are the rejection frequencies assessed from 200 Monte Carlo replicates. The sample sizes are $n = 100$, 200 and 500, where the generating calibration models are shown in the "True Model" column. Those entries in the table reflecting the power of the testing procedure are shown in bold face

True Model		Null Model					
		$H_0: \eta(x) = a_0 + a_1 x$			$H_0: \eta = c$		
	n	.10	.05	.01	.10	.05	.01
$\mathbf{M}_0^{(C)}$	100	—	—	—	.105	.055	.010
	200	—	—	—	.110	.040	.000
	500	—	—	—	.085	.040	.005
$\mathbf{M}_1^{(C)}$	100	.075	.040	.000	**1.00**	**1.00**	**1.00**
	200	.130	.060	.010	**1.00**	**1.00**	**1.00**
	500	.060	.035	.000	**1.00**	**1.00**	**1.00**
$\mathbf{M}_2^{(C)}$	100	**.765**	**.665**	**.435**	**.915**	**.855**	**.670**
	200	**.975**	**.930**	**.820**	**.995**	**.995**	**.960**
	500	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
$\mathbf{M}_0^{(G)}$	100	—	—	—	.090	.045	.005
	200	—	—	—	.120	.045	.000
	500	—	—	—	.140	.070	.015
$\mathbf{M}_1^{(G)}$	100	.100	.065	.015	**.520**	**.395**	**.165**
	200	.090	.020	.010	**.615**	**.515**	**.355**
	500	.110	.035	.005	**1.00**	**.990**	**.975**
$\mathbf{M}_2^{(G)}$	100	**.330**	**.200**	**.050**	**.775**	**.635**	**.355**
	200	**.585**	**.410**	**.210**	**.970**	**.960**	**.805**
	500	**.945**	**.870**	**.685**	**1.00**	**1.00**	**0.995**

Acknowlegements

The authors would like to thank the editor, the associate editor and the two referees for their careful review and valuable comments. E. F. Acar, R. V. Craiu and F. Yao were partially supported by the Discovery Grants and Discovery Accelerator Supplements from Natural Sciences and Engineering Research Council of Canada (NSERC).

References

Abegaz, F., Gijbels, I., and Veraverbeke, N. (2012). Semiparametric estimation of conditional copulas. *J. Multivariate Anal.*, 110:43–73. MR2927509

Acar, E. F., Craiu, R. V., and Yao, F. (2011). Dependence calibration in conditional copulas: A nonparametric approach. *Biometrics*, 67:445–453. MR2829013

Bartram, S., Taylor, S., and Wang, Y. (2007). The euro and european financial market dependence. *Journal of Banking and Finance*, 31:1461–1481.

Craiu, R. V. and Sabeti, A. (2012). In mixed company: Bayesian inference for bivariate conditional copula models with discrete and continuous outcomes. *J. Multivariate Anal.*, 110:106–120. MR2927512

de Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probability Theory and Related Fields*, 75:261–277. MR0885466

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*, volume 66. Chapman & Hall, London, 1st edition. MR1383587

Fan, J. and Jiang, J. (2005). Nonparametric inferences for additive models. *Journal of American Statistical Association*, 100(471):890–907. MR2201017

Fan, J., Zhang, C., and Zhang, J. (2001). Generalized likelihood ratio statistics and wilks phenomenon. *Annals of Statistics*, 29(1):153–193. MR1833962

Fan, J. and Zhang, W. (2004). Generalized likelihood ratio tests for spectral density. *Biometrika*, 91(1):195–209. MR2050469

Genest, C. and Rémillard, B. (2004). Tests of independence and randomness based on the empirical copula process. *Test*, 13:335–370. MR2154005

Gijbels, I., Veraverbeke, N., and Omelka, M. (2011). Conditional copulas, association measures and their application. *Comput. Stat. Data An.*, 55:1919–1932. MR2765054

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics*, 19(3):293–325. MR0026294

Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *J. Multivariate Anal.*, 94:401–419. MR2167922

Jondeau, E. and Rockinger, M. (2006). The copula-garch model of conditional dependencies: An international stock market application. *Journal of International Money and Finance*, 25:827–853.

Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. *Internat. Econom. Rev.*, 47:527–556. MR2216591

Rodriguez, J. C. (2007). Measuring financial contagion: A copula approach. *J. of Empirical Finance*, 14(3):401–423.

Self, S. G. and Liang, K. (1987). Asymptotic properlies of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Amer. Statist. Assoc.*, 82(398):605–610. MR0898365

Veraverbeke, N., Omelka, M., and Gijbels, I. (2011). Estimation of a conditional copula and association measures. *Scandinavian Journal of Statistics*, early view. MR2859749

Zhang, R., Huang, Z., and Lv, Y. (2010). Statistical inference for the index parameter in single-index models. *Journal of Multivariate Analysis*, 101(4):1026–1041. MR2584917