NONPARAMETRIC ESTIMATION AND INference FOR THE COPULA PARAMETER IN CONDITIONAL COPULAS

by

Elif Fidan Acar

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Statistics
University of Toronto

Copyright © 2010 by Elif Fidan Acar
Abstract

The primary aim of this thesis is the elucidation of covariate effects on the dependence structure of random variables in bivariate or multivariate models. We develop a unified approach via a conditional copula model in which the copula is parametric and its parameter varies as the covariate. We propose a nonparametric procedure based on local likelihood to estimate the functional relationship between the copula parameter and the covariate, derive the asymptotic properties of the proposed estimator and outline the construction of pointwise confidence intervals. We also contribute a novel conditional copula selection method based on cross-validated prediction errors and a generalized likelihood ratio-type test to determine if the copula parameter varies significantly. We derive the asymptotic null distribution of the formal test. Using subsets of the Matched Multiple Birth and Framingham Heart Study datasets, we demonstrate the performance of these procedures via analyses of gestational age-specific twin birth weights and the impact of change in body mass index on the dependence between two consequent pulse pressures taken from the same subject.
Dedication

This thesis is dedicated to my mother. She sacrificed her education so that I could have my own.
Acknowledgements

It is impossible to recognize the multitude of people that have made this thesis possible. Nonetheless, some names must not go unmentioned.

Radu and Fang are fantastic advisors and deserve much of the credit for the direction this project has taken. Their contributions to my academic development are innumerable. Lei, whose work in statistical genetics influenced my own and who taught me much about the subject, deserves special thanks. Dr. Reid, a member of my thesis committee, requires special accolades for her invaluable advice. Dr. Yi of University of Waterloo, the external examiner of this thesis, also contributed insightful comments. Dr. Brenner, Dr. Evans, Dr. Virag and Dr. Quastel imparted important instruction during their courses. Indeed, the entire University of Toronto Department of Statistics is to be thanked and will be sorely missed after my departure. At the University of New Hampshire, Dr. Hadwin, Dr. Nordgren and Dr. Linder were important early influences. My colleagues and friends, Zeynep, Alex, Avideh, Lili, Lizhen, Shivon, Zi, Shelley, Gunho, Yan, Madeleine and innumerable others must not go unmentioned. I wish to offer a very special thanks to my mother, Ayse, and great-aunt, Piril, who initiated this journey, and to Michael, whose support is always present.
Contents

List of tables vii

List of figures ix

1 Introduction 1

1.1 Copulas 4

1.1.1 Basic definitions and properties 4

1.1.2 Examples of some copulas 7

1.1.3 Estimation in copula-based models 8

1.2 Conditional Copulas 10

1.3 Smoothing Methods 12

1.4 Thesis Outline 14

2 Motivation and Applications 15

2.1 Twin Birth Data 15

2.2 Framingham Heart Study 17

2.3 Summary 17

3 Dependence Calibration 21

3.1 Copulas versus Conditional Copulas 21

3.2 Calibration Model 24

3.3 Local Copula-Likelihood Estimation 25
3.4 Choice of Bandwidth .................................................. 27
3.5 Asymptotic Bias and Variance .................................. 28
3.6 Confidence Intervals ................................................. 32
3.7 Impact of Estimating Conditional Marginals .................... 37
3.8 Summary and Discussion .......................................... 38

4 Conditional Copula Selection & Parameter Inference .............. 39
  4.1 Conditional Copula Selection ................................... 39
    4.1.1 Proposed Method ........................................... 41
  4.2 Copula Parameter Inference .................................... 42
    4.2.1 Introduction .............................................. 42
    4.2.2 A generalized likelihood ratio test .................... 44
    4.2.3 Asymptotic null distribution ............................. 45
    4.2.4 Additional theoretical results .......................... 52
  4.3 Summary and Discussion ....................................... 56

5 Simulation Results and Data Applications .......................... 58
  5.1 Simulation Results .............................................. 58
    5.1.1 Clayton Copula ........................................ 60
    5.1.2 Frank Copula ........................................... 67
    5.1.3 Gumbel Copula .......................................... 71
    5.1.4 Generalized Likelihood Ratio Test .................... 75
  5.2 Data Examples .................................................. 78
    5.2.1 Twin Birth Data ...................................... 78
    5.2.2 Framingham Data ...................................... 83

6 Conclusion .......................................................... 90

Bibliography .......................................................... 94
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>The copula families used in simulation studies.</td>
<td>59</td>
</tr>
<tr>
<td>5.2</td>
<td>Clayton Copula. IMSE of the Kendall’s tau estimator.</td>
<td>63</td>
</tr>
<tr>
<td>5.3</td>
<td>Times of selection of the three families out of 100 Monte Carlo samples.</td>
<td>64</td>
</tr>
<tr>
<td>5.4</td>
<td>Frank Copula. IMSE of the Kendall’s tau estimator.</td>
<td>68</td>
</tr>
<tr>
<td>5.5</td>
<td>Times of selection of the three families out of 100 Monte Carlo samples.</td>
<td>68</td>
</tr>
<tr>
<td>5.6</td>
<td>Gumbel Copula. IMSE of the Kendall’s tau estimator.</td>
<td>72</td>
</tr>
<tr>
<td>5.7</td>
<td>Times of selection of the three families out of 100 Monte Carlo samples.</td>
<td>72</td>
</tr>
<tr>
<td>5.8</td>
<td>Linear Calibration Model. GLRT results for the linear null hypothesis.</td>
<td>76</td>
</tr>
<tr>
<td>5.9</td>
<td>Linear Calibration Model. Exact Binomial Tests.</td>
<td>76</td>
</tr>
<tr>
<td>5.10</td>
<td>Linear Calibration Model. GLRT results for the constant null hypothesis.</td>
<td>77</td>
</tr>
<tr>
<td>5.11</td>
<td>Nonlinear Calibration Model. GLRT results for the constant null hypothesis.</td>
<td>77</td>
</tr>
<tr>
<td>5.12</td>
<td>Nonlinear Calibration Model. GLRT results for the linear null hypothesis.</td>
<td>77</td>
</tr>
<tr>
<td>5.13</td>
<td>Twin birth data. Parametric conditional marginal models.</td>
<td>78</td>
</tr>
</tbody>
</table>
## List of Figures

1.1 Contour plots of some copulas with standard normal margins. . . . . . . 9

2.1 Twin birth data. Gestational age-specific intra-twin differences. . . . . 16

2.2 Twin birth data. Scatterplots of the twin birth weights at different gesta-
tional ages. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 18

2.3 Framingham data. Scatterplots of the log-pulse pressures at different co-
variate ranges. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 19

3.1 Simple example. Scatter and contour plots of the combined data. . . . 22

3.2 Simple example. Scatter and contour plots of each sample. . . . . . . . 23

5.1 Clayton Copula. An illustration of the local calibration function estimates
under the linear calibration model. . . . . . . . . . . . . . . . . . . . . . . . 61

5.2 Clayton Copula. An illustration of the local calibration function estimates
under the nonlinear calibration model. . . . . . . . . . . . . . . . . . . . . . 62

5.3 Clayton Copula. 90% confidence intervals for the Kendall’s tau under the
linear calibration model. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 65

5.4 Clayton Copula. 90% confidence intervals for the Kendall’s tau under the
nonlinear calibration model. . . . . . . . . . . . . . . . . . . . . . . . . . . . 66

5.5 Frank Copula. 90% confidence intervals for the Kendall’s tau under the
linear calibration model. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 69
5.6 Frank Copula. 90% confidence intervals for the Kendall’s tau under the nonlinear calibration model. ................................................................. 70

5.7 Gumbel Copula. 90% confidence intervals for the Kendall’s tau under the linear calibration model. ................................................................. 73

5.8 Gumbel Copula. 90% confidence intervals for the Kendall’s tau under the nonlinear calibration model. ................................................................. 74

5.9 Twin birth data. Histograms and scatterplots of the parametrically estimated conditional marginals. ................................................................. 80

5.10 Twin birth data. The Kendall’s tau estimates under three copula families using parametrically estimated conditional marginals. ...................... 81

5.11 Twin birth data. Histograms and scatterplot of the nonparametrically estimated conditional marginals. ................................................................. 82

5.12 Twin birth data. The Kendall’s tau estimates under three copula families using nonparametrically estimated conditional marginals. .............. 84

5.13 Twin birth data. Bootstrap Confidence Intervals under the Frank copula. 85

5.14 Framingham data. Histograms and scatterplots of the parametrically estimated conditional marginals. ................................................................. 87

5.15 Framingham data. The Kendall’s tau estimates under three copula families. 88

5.16 Framingham data. Constant versus variable bandwidth. ...................... 89
Chapter 1

Introduction

The statistics literature is well-equipped with graphical methods and inference techniques for assessing covariate effects on the mean or variance of a response variable. In the case of multivariate response data, however, a covariate may affect not only each response variable but also the dependence between them, rendering this task more complicated. Commonly used regression-based models, although well-suited for modeling conditional means, fail to elucidate complex dependence patterns. On the other hand, much less explored conditional covariance (correlation) models are used to represent linear dependencies. A more appropriate statistical strategy is needed when our primary interest is the effect of one or several covariates on dependence structures.

Multivariate dependence modeling is difficult even without taking into consideration the information added by covariates. One often needs to specify a complex joint distribution of random variables to have a complete view of the dependence structure. The challenge of constructing such multivariate distributions can be significantly reduced if we use a copula model to specify the dependence structure independently of the marginal models. Sklar’s theorem (1959) is central to the theoretical foundation of copulas. It holds that a continuous multivariate distribution is fully characterized by the marginal distributions and a uniquely determined copula, where the latter is a multivariate distri-
bution function having uniform \([0, 1]\) marginals.

In the last twenty years copulas have been used in a variety of applied work. We refer the reader to Embrechts et al. (2002), Cherubini et al. (2004) and Frees and Valdez (1998) for applications specific to finance and insurance. In survival analysis, Clayton (1978), Shih and Louis (1995), Wang and Wells (2000) and the monograph by Hougaard (2000) present copula techniques to model multivariate time-to-event data and competing risk problems.

As a result of their wide applicability, a large number of parametric families of copulas, typically indexed by a scalar parameter \(\theta\), have been proposed to represent various dependence patterns. While the copula family describes the functional form, within each family \(\theta\) controls the strength of dependence. A comprehensive introduction to copulas and their properties can be found in Nelsen (2006) and the connections between various copulas and dependence concepts are discussed in detail by Joe (1997). If a parametric form is assumed for the copula function, estimation can be achieved using maximum likelihood methods for a single copula parameter (Joe, 1997; Genest et al., 1995). Alternatively, estimation can be performed fully nonparametrically by using kernel estimators (Fermanian and Scaillet, 2003; Chen and Huang, 2007).

The importance of covariate adjustment was addressed by Clayton (1978) in the estimation of joint life table dependence. Joe (1997) suggested expressing model parameters as functions of covariates, Fine et al. (2001) and Peng and Fine (2006) included covariate adjustments in their analyses of semi-competing risk problems. However, a formal treatment incorporating Sklar’s theorem was unavailable until Patton’s advent of conditional copulas in 2002.

To our knowledge, so far conditional copulas have been applied only to financial time-series data using ARMA-type parametrization of dependence parameters (Patton, 2006; Jondeau and Rockinger, 2006; Bartram et al., 2007). This choice of parametric formulation has been undertaken without due justification of its particular appropriateness.
Indeed, Joe (1997, p. 312) makes a similar point:

A difficult modelling question may be whether dependence parameters should be functions of covariates. If so, what are natural functions to choose? ... For data analysis, one could split the covariate space into several clusters or subgroups, and then do a separate estimation of dependence parameters by clusters. If there are only binary (or categorical) predictor variables, then one could do estimation for each combination if the resulting (sub)sample sizes are large enough.

The main objective of this thesis is to answer the question posed by Joe: what are natural functions to choose? In our opinion, there is no simple (or even complete) answer to this question. Instead, we propose a natural estimation procedure that is flexible enough to adapt to a large number of functions via nonparametric smoothing. Smoothing methods for function estimation have been studied in depth for various problems. Here we use a local polynomial framework (see Fan and Gijbels, 1996, for a comprehensive review) for covariate-adjusted copula estimation via local likelihood-based models (Tibshirani and Hastie, 1987).

This goal includes, in addition to establishing an efficient and feasible estimation procedure, two subtasks. We must develop an applicable strategy of conditional copula selection. This is a challenging problem in general and has attracted considerable prior interest in the context of copula models but, to our knowledge, no methodology has been proposed for conditional copulas. Existing methods of copula selection include goodness-of-fit tests based on the empirical copula (Durrleman et al., 2000), on the Kendall process (Genest and Rivest, 1993; Wang and Wells, 2000) and on kernel density estimation (Scaillet, 2005; Craiu and Craiu, 2008). Our calibration procedure naturally leads to a novel conditional copula selection method based on cross-validated prediction errors. The proposed selection criterion makes possible comparisons across copula families due to its general applicability. Furthermore, we must develop a formal test of significance of covariate effects on dependence structures. This task is crucial
to complete the model specification for subsequent analyses. Nonparametric hypothesis testing has developed slowly compared to nonparametric estimation methods. Many goodness-of-fit tests have been constructed for nonparametric regression problems (see Hart, 1997, for a comprehensive list of references and an overview) but few are adaptable to other settings. We develop a generalized likelihood ratio (GLR) test in the same vein as Fan et al (2001), study its asymptotic distribution and use simulations to discuss its finite sample properties.

1.1 Copulas

1.1.1 Basic definitions and properties

A copula can be defined formally or less formally as follows.

**Definition 1** (Mathematical Definition). A bivariate copula is a function $C : [0,1]^2 \rightarrow [0,1]$ satisfying the following properties:

1. $C(u, 0) = C(0, u) = 0$ and $C(u, 1) = C(1, u) = u \ \forall u \in [0, 1]$,

2. $C(u_2, u_4) - C(u_1, u_4) - C(u_2, u_3) + C(u_1, u_3) \geq 0$ for all $u_1, u_2, u_3, u_4 \in [0, 1]$ such that $u_1 \leq u_2$ and $u_3 \leq u_4$.

It follows from the above properties that $C$ is non-decreasing in each component and continuous, and hence forms a valid bivariate distribution function on the unit square with uniform margins.

An alternative definition is widely used in statistical applications.

**Definition 2** (Statistical Definition). A bivariate copula is a joint distribution function of standard uniform random variables. That is,

$$C(u_1, u_2) = P(U_1 \leq u_1, U_2 \leq u_2),$$

(1.1)

where $U_i \sim Uniform(0, 1)$, $i = 1, 2$. 

The central result of copula theory is Sklar’s theorem (1959), to which copulas owe their name and popularity.

**Theorem 1 (Sklar’s Theorem for conditional distributions).** Let \( Y_1 \) and \( Y_2 \) be continuous random variables with joint distribution function \( H \) and marginal distributions \( F_1 \) and \( F_2 \). Then, there exists a unique copula \( C \) such that

\[
H(y_1, y_2) = C(F_1(y_1), F_2(y_2)), \quad \text{for all} \quad (y_1, y_2) \in \mathbb{R}^2. \tag{1.2}
\]

Conversely, if \( C \) is a copula and \( F_1 \) and \( F_2 \) are distribution functions, then \( H \), defined by (1.2), is a joint distribution function with margins \( F_1 \) and \( F_2 \).

A detailed proof of Sklar’s theorem can be found in Schweizer and Sklar (1983).

Sklar’s theorem has, among others, two important consequences. The information in a joint distribution \( H \) can be split into parts containing univariate characteristics, \( F_1(Y_1) \) and \( F_2(Y_2) \), and a part containing dependence information, \( C \). We can construct many bivariate distributions by linking any two univariate distributions, not necessarily of the same type, with any copula. Thus traditional multivariate distributions are no longer required to model complex dependence structures. Also, as can be easily demonstrated, copulas are invariant under strictly increasing transformations. Suppose we apply \( \phi_1 \) and \( \phi_2 \), both strictly increasing, to continuous variables \( Y_1 \) and \( Y_2 \). Denote by \( F_1, F_2, G_1 \) and \( G_2 \) the distributions of \( Y_1, Y_2, \phi_1(Y_1) \) and \( \phi_2(Y_1) \), respectively. Then, the copula of the transformed variables, \( C^* \), can be written as

\[
C^* (G_1(y_1), G_2(y_2)) = P[\phi_1(Y_1) \leq y_1, \phi_2(Y_2) \leq y_2] \\
= P[Y_1 \leq \phi_1^{-1}(y_1), Y_2 \leq \phi_2^{-1}(y_2)] \\
= C(F_1(\phi_1^{-1}(y_1)), F_2(\phi_2^{-1}(y_2))) \\
= C(G_1(y_1), G_2(y_2)).
\]

Under strictly monotone transformations, the copula of the transformed variables can be
obtained in a similar way. This result allows us to connect copulas to many scale-free
dependence measures.

Two such measures commonly used in practice are Kendall’s tau and Spearman’s rho.
For random variables $Y_1$ and $Y_2$ with marginal distributions $F_1$ and $F_2$, respectively, the
Spearman’s rho is the correlation of the transformed variables $F_1(Y_1)$ and $F_2(Y_2)$, while
the Kendall’s tau is defined as the difference between the probability of concordance
$P[(Y_1 - Y'_1)(Y_2 - Y'_2) > 0]$ and the probability of disconcordance $P[(Y_1 - Y'_1)(Y_2 - Y'_2) < 0]$ for the independent random vectors $(Y_1, Y_2)$ and $(Y'_1, Y'_2)$ having the same bivariate
distribution $H$. The population versions of Kendall’s tau and Spearman’s rho in terms
of copula functions are given below. Their derivations can be found in Nelsen (2006, pp.
159 and 167).

**Definition 3.** Kendall’s tau

$$\tau_C = 4 \int_0^1 \int_0^1 C(u_1, u_2) \, dC(u_1, u_2) - 1. \quad (1.3)$$

**Definition 4.** Spearman’s rho

$$\rho_C = 12 \int_0^1 \int_0^1 u_1 u_2 \, dC(u_1, u_2) - 3. \quad (1.4)$$

Tail dependence is another notion of dependence particularly useful in studying ex-
treme events.

**Definition 5.** For random variables $Y_1 \sim F_1$ and $Y_2 \sim F_2$ with copula $C$, tail dependence
indices are defined as

$$\lambda_L = \lim_{u \downarrow 0} P(Y_2 \leq F_2^{-1}(u) | Y_1 \leq F_1^{-1}(u)) = \lim_{u \downarrow 0} \frac{C(u, u)}{u},$$

$$\lambda_U = \lim_{u \uparrow 1} P(Y_2 > F_2^{-1}(u) | Y_1 > F_1^{-1}(u)) = \lim_{u \uparrow 1} \frac{1 - 2u + C(u, u)}{1 - u}. \quad (1.5)$$

$C$ is said to have lower (upper) tail dependence if $\lambda_L \in (0, 1]$ ($\lambda_U \in (0, 1]$), and no lower
(upper) tail dependence if $\lambda_L = 0$ ($\lambda_U = 0$).
1.1.2 Examples of some copulas

It is useful to use examples to investigate the dependence patterns of copulas.

Example 1. The Independence Copula is denoted by \( \Pi \), and has the form

\[
\Pi(u_1, u_2) = u_1 u_2.
\]

Example 2. The Minimum and Maximum Copulas, are denoted by \( W \) and \( M \), respectively.

\[
W(u_1, u_2) = \max(0, u_1 + u_2 - 1),
\]

\[
M(u_1, u_2) = \min(u_1, u_2).
\]

They are also known as Fréchet-Hoeffding lower and upper bounds of copulas, i.e.

\[
W(u_1, u_2) \leq C(u_1, u_2) \leq M(u_1, u_2).
\]

The following parametric copulas, all from the Archimedean class (see Chapter 4 in Nelsen, 2006, for details on Archimedean copulas), are widely used.

Example 3. The Clayton family has copula function

\[
C(u_1, u_2) = \left( u_1^{-\theta} + u_2^{-\theta} - 1 \right)^{-\frac{1}{\theta}}, \quad \theta \in (0, \infty).
\]

The Clayton copula exhibits lower tail dependence, with \( \lambda_L = 2^{1/\theta} \). Its Kendall’s tau is calculated as \( \tau = \frac{\theta}{\theta + 2} \). Its Spearman’s rho has no analytic expression.

Example 4. The Frank family has copula function

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

The Frank copula does not exhibit tail dependence, i.e. \( \lambda_L = \lambda_U = 0 \). Its Kendall’s tau is given by \( \tau = 1 + \frac{4}{\theta} \{ D_1(\theta) - 1 \} \), and its Spearman’s rho is computed as \( \rho = 1 - \frac{12}{\theta} \{ D_1(\theta) - D_2(\theta) \} \), where \( D \) denotes the Debye function,

\[
D_j(\theta) = \frac{j}{\theta^j} \int_{0}^{\theta} \frac{t^j}{e^t - 1} dt.
\]
Example 5. The Gumbel family has copula function

\[ C(u_1, u_2) = \exp \left[ -\left( -\ln u_1 \right)^\theta \left( -\ln u_2 \right)^\theta \right], \quad \theta \in [1, \infty), \quad (1.13) \]

The Gumbel copula exhibits upper tail dependence, with \( \lambda_U = 2 - 2^{1/\theta} \). Its Kendall’s tau is given by \( \tau = 1 - \frac{1}{\theta} \). There is no closed form expression for the Spearman’s rho of the Gumbel family.

Figure 1.1 gives the contour plots of bivariate densities, all with standard normal marginal distributions, defined by Clayton, Frank, and Gumbel copulas. The three values of the copula parameter under each family are chosen so that the corresponding Kendall’s tau values are approximately 0.1, 0.5, 0.8, representing weak, moderate and strong dependence, respectively.

1.1.3 Estimation in copula-based models

The estimation methods in copula based models are distinguished by whether parametric or nonparametric procedures are employed in specifying (i) the copula, and (ii) the marginal distributions. Since our main interest is the inference of the copula parameter, we will focus on the estimation methods in parametric copula models. For fully nonparametric treatment of copulas, see Fermanian and Scaillet (2003); Fermanian (2005); Chen and Huang (2007) and references therein.

The first model we consider is fully parametric, i.e. both copula and marginal distributions are specified by certain parametric forms. The likelihood inference of this model is achieved using the density equivalent (provided all densities exist) of the Sklar’s theorem,

\[ h(y_1, y_2|\theta, \alpha_1, \alpha_2) = \frac{\partial^2 H(y_1, y_2|\theta, \alpha_1, \alpha_2)}{\partial y_1 \partial y_2} = \frac{\partial^2 C(F_1(y_1|\alpha_1), F_2(y_2|\alpha_2)|\theta)}{\partial F_1(y_1|\alpha_1) \partial F_2(y_2|\alpha_2)} \frac{\partial F(y_1|\alpha_1)}{\partial y_1} \frac{\partial F(y_2|\alpha_2)}{\partial y_2} = c(F_1(y_1|\alpha_1), F_2(y_2|\alpha_2)|\theta) f(y_1|\alpha_1) f(y_2|\alpha_2). \quad (1.14) \]
Figure 1.1: Contour plots of some copulas with standard normal margins.
Estimation can then be performed by maximizing the log-likelihood function

\[ L(\theta, \alpha_1, \alpha_2) = \sum_{i=1}^{n} \ln \{c(F_1(y_{i1}|\alpha_1), F_2(y_{i2}|\alpha_2)|\theta)\} + \sum_{i=1}^{n} \ln \{f(y_{i1}|\alpha_1)\} + \sum_{i=1}^{n} \ln \{f(y_{i2}|\alpha_2)\} \]

over the unknown parameters, \( \alpha_1, \alpha_2 \) and \( \theta \), simultaneously. An alternative approach is to perform the estimation in two steps, separately maximizing the univariate likelihoods \( L_j(\alpha_j) \) over \( \alpha_j \), to get \( \hat{\alpha}_j, j = 1, 2 \), and then maximizing \( L_C(\theta, \hat{\alpha}_1, \hat{\alpha}_2) \) over \( \theta \) to obtain \( \hat{\theta} \). This two-step procedure, known as the method of inference functions for margins (IFM), introduced by Joe (1997), is less efficient in mean square error relative to the one-step estimator, but more attractive in practice as it significantly reduces the computational cost, especially in higher-dimensional problems.

A semiparametric copula model is usually formed by a parametric form of a copula function with unspecified marginal distributions. A common approach is to replace the marginal distributions in (1.2) by rescaled empirical counterparts, given by

\[ F_{jn}(y) = \frac{1}{n+1} \sum_{i=1}^{n} I(Y_{ji} \leq y), \quad j = 1, 2, \]  

where \( I \) denotes the indicator function. One can then estimate \( \theta \) by maximizing the pseudo log-likelihood

\[ L(\theta) = \sum_{i=1}^{n} \ln \{c(F_{1n}(y_{i1}), F_{2n}(y_{i2})|\theta)\}. \]  

The asymptotic properties of this semiparametric copula parameter estimator, including the asymptotic normality and consistency, can be found in Genest et al. (1995) and Shih and Louis (1995).

### 1.2 Conditional Copulas

**Definition 6.** The conditional copula of \((Y_1, Y_2)|X = x\), where \(Y_1|X = x \sim F_{1|X}(\cdot|x)\) and \(Y_2|X = x \sim F_{2|X}(\cdot|x)\), is the joint distribution function of \(U_1 \equiv F_{1|X}(Y_1|x)\) and \(U_2 \equiv F_{2|X}(Y_2|x)\) given \(X = x\).
Theorem 2 (Extension of Sklar’s Theorem for conditional distributions). Let
\( F_{1|X}(\cdot|x) \) be the conditional distribution of \( Y_1|X = x \), \( F_{2|X}(\cdot|x) \) be the conditional distribution of \( Y_2|X = x \), \( H_X(\cdot|x) \) be the joint distribution of \( (Y_1,Y_2)|X = x \), and \( X \) be the support of \( X \). Then, there exists a unique conditional copula \( C(\cdot|x) \), whenever \( F_{1|X}(\cdot|x) \)
and \( F_{2|X}(\cdot|x) \) are continuous in \( y_1 \) and \( y_2 \), for all \( x \in X \), such that
\[
H_X(y_1, y_2|x) = C(F_{1|X}(y_1|x), F_{2|X}(y_2|x)|x). \tag{1.17}
\]
Conversely, if \( F_{1|X}(\cdot|x) \) is the conditional distribution of \( Y_1|X = x \), \( F_{2|X}(\cdot|x) \) is the conditional distribution of \( Y_2|X = x \), and \( \{C(\cdot|x)\} \) is a family of conditional copulas measurable in \( x \), then the function \( H_X(\cdot|x) \) defined in (1.17) is a conditional bivariate distribution function with conditional marginal distributions \( F_{1|X}(\cdot|x) \) and \( F_{2|X}(\cdot|x) \).

See Patton (2002) for a generalized introduction to conditional copulas, and a proof of (1.17). Note that the marginal components and the copula must be conditioned on the same variable \( X \), otherwise, \( H_X(\cdot|x) \) will not be a valid bivariate distribution. The failure of this condition is illustrated by Patton (2002) in this example:

Example 6. Suppose we condition \( Y_1 \) on \( X_1 \), \( Y_2 \) on \( X_2 \), the copula on \( (X_1,X_2) \) and specify
\[
\tilde{H}_{(X_1,X_2)}(y_1, y_2|x_1, x_2) = C(F_{1|X_1}(y_1|x_1), F_{2|X_2}(y_2|x_2)|x_1, x_2).
\]
Then, \( \tilde{H}_{(X_1,X_2)}(y_1, \infty|x_1, x_2) = C(F_{1|X_1}(y_1|x_1), 1|x_1, x_2) = F_{1|X_1}(y_1|x_1) \), the conditional distribution of \( Y_1|X_1 \), which is the conditional marginal distribution of \( (Y_1, Y_2)|X_1 \). But, \( \tilde{H}_{(X_1,X_2)}(\infty, y_2|x_1, x_2) = C(1, F_{2|X_2}(y_2|x_2)|x_1, x_2) = F_{2|X_2}(y_2|x_2) \), the conditional distribution of \( Y_2|X_2 \), which is the conditional marginal distribution of \( (Y_1, Y_2)|X_2 \). Thus, the function \( \tilde{H}_{(X_1,X_2)} \) will not be the joint distribution function of \( (Y_1, Y_2|X_1,X_2) \) in general, unless \( F_{1|X_1}(y_1|x_1) = F_{1|X_1,X_2}(y_1|x_1, x_2) \) and \( F_{2|X_2}(y_2|x_2) = F_{2|X_1,X_2}(y_2|x_1, x_2) \).

Patton further argued that this special condition is not unrealistic, as one may find certain variables affecting one response variable but not the other. This consideration is
important especially in time-series contexts, as one usually models each marginal component conditional on its own lags but not on the other’s.

### 1.3 Smoothing Methods

Smoothing methods provide flexible ways to detect complex patterns in data without imposing strong functional assumptions. Most existing techniques are developed in a regression framework and later extended to more complicated settings. A good starting point is therefore the simple nonparametric regression problem,

\[ Y = \theta(X) + \varepsilon, \quad \varepsilon \sim (0, \sigma^2), \quad (1.18) \]

where \( \theta \) is an unknown smooth function.

The local polynomial regression provides a simple and intuitive approach to (1.18). Nonparametric estimation of \( \theta(x) \) is achieved by approximating

\[ \theta(X_i) \approx \sum_{j=0}^{p} \frac{\theta^{(j)}(x)}{j!} (X_i - x)^j \equiv \sum_{j=0}^{p} \beta_j (X_i - x)^j, \quad (1.19) \]

for data points \( X_i \) in a neighbourhood of \( x \), using a \( p^\text{th} \) order Taylor expansion and minimizing

\[ \sum_{i=1}^{n} \left( Y_i - \sum_{j=0}^{p} \beta_j (X_i - x)^j \right)^2 K \left( \frac{X_i - x}{h} \right). \quad (1.20) \]

Here, the kernel function \( K \) and the bandwidth parameter \( h \) determine the weighting scheme and the size of the local neighbourhood. Denoting the solution of (1.20) by \( \hat{\beta}_j, j = 0, 1, \ldots, p \), we obtain the estimators, \( \hat{\theta}^{(j)}(x) = j! \hat{\beta}_j, j = 0, 1, \ldots, p \).

The overall performance of the estimator \( \hat{\theta} \) is usually evaluated by its integrated mean square error (IMSE),

\[ \int_X E\left[ (\hat{\theta}(x) - \theta(x))^2 \right] dx = \text{IBIAS}^2(\hat{\theta}) + \text{IVAR}(\hat{\theta}), \quad (1.21) \]
where

\[
\text{IBIAS}^2(\hat{\theta}) = \int_X \left[ E\{\hat{\theta}(x)\} - \theta(x) \right]^2 \, dx,
\]
(1.22)

\[
\text{IVAR}(\hat{\theta}) = \int_X E\left( \{\hat{\theta}(x) - E\{\hat{\theta}(x)\}\}^2 \right) \, dx,
\]
(1.23)

are the integrated square bias and integrated variance, respectively.

Local polynomial approaches are but one of many avenues of study in this domain. Early work in this field includes Stone (1977) and Cleveland (1979). See Fan and Gijbels (1996) for a book-length treatment of local polynomial techniques specifically.

Local likelihood estimation is an important contribution to local polynomial approaches (Tibshirani and Hastie, 1987; Staniswalis, 1989) and paved the way for further studies of likelihood-based problems. The local likelihood idea can be summarized as follows. Suppose that the \(i\)th observation \((X_i, Y_i)\) has the contribution \(\ell(\theta(X_i), Y_i)\) to the conditional log-likelihood, where \(\theta\) is to be estimated. The conditional log-likelihood of \(n\) observations is given by

\[
\sum_{i=1}^{n} \ell(\theta(X_i), Y_i).
\]
(1.24)

The maximum likelihood estimation provides a solution to (1.24) when \(\theta\) is assumed to have a certain parametric form. The local likelihood removes this assumption. As in local regression setting, the unknown target function \(\theta(x)\) is approximated by (1.19). The local likelihood estimator \(\hat{\theta}(x)\) is then obtained by maximizing the kernel weighted local log-likelihood

\[
\sum_{i=1}^{n} \ell \left( \sum_{j=0}^{p} \beta_j (X_i - x)^j, Y_i \right) K \left( \frac{X_i - x}{h} \right),
\]
(1.25)

over \(\beta_j, j = 1, \ldots, p\). The estimators are given by \(\hat{\theta}^{(j)}(x) = j! \hat{\beta}_j, j = 0, 1, \ldots, p\).

The local likelihood approach has been extended to a large number of problems. Tibshirani and Hastie (1990) provided generalized additive models, Fan et al. (1995) generalized linear models and Kauermann and Tutz (1997), Cai et al. (2000) varying-
coefficient models. See Aerts and Claeskens (1997) for an important contribution to multiparameter likelihood models and Carroll et al. (1998) for local estimating equations.

1.4 Thesis Outline

Chapter 2 illustrates potential areas of application of conditional copulas outside of financial economics. Chapter 3 presents our proposed estimation procedure, bandwidth selection, pointwise confidence bands and the asymptotic properties of the nonparametric estimator. Chapter 4 presents a prediction-based conditional copula selection method and a generalized likelihood ratio test. Chapter 5 summarizes our simulation studies and data applications. Chapter 6 provides a concluding discussion.
Chapter 2

Motivation and Applications

There are few modeling techniques designed for conditional copulas and those which exist are usually limited to time-variation analyses of financial data. This chapter highlights the potential of conditional copulas in other applications, specifically the biological and medical. We consider the examples of twin birth weight analysis and an analysis of a subset of the Framingham Heart Study data.

2.1 Twin Birth Data

The twin birth data considered here is a subset of the Matched Multiple Birth Data Set (MMB) of the National Center for Health Statistics. Containing all twin, triple and quadruplet births in the United States between 1995 and 2000, the MMB allows research on several characteristics of multiple deliveries. The records include information on births, fetal deaths, birth weights of matched set members, birth order within each set, gestational age and maternal age.

We consider the twin births to mothers aged between 18 and 40 years of which both babies survived their first year. Of interest is the dependence between the birth weights of twins, denoted $BW_1$ and $BW_2$. The gestational age $GA$ is an important factor for fetal growth and is therefore chosen as an informative covariate. We consider a random
sample of 30 twin live births for each GA between 28 to 42 weeks.

Gestational age-specific intra-twin birth weight differences are usually examined in obstetrical studies to study fetal growth patterns in twin gestations (Kalish et al., 2005). A major drawback of this approach is that we lose information about intra-twin birth weight dependence when examining only the difference between twin weights. Since the birth weights of twins at different gestational ages have different ranges, the nominal differences yield limited and often misleading conclusions about growth patterns. For instance, in Figure 2.1, intra-twin birth weight differences at early gestations seem to have less variation, but this may be due to the small birth weights of pre-term twins.

Figure 2.1: Twin birth data. Gestational age-specific intra-twin differences.
In Chapter 6, we will study the inter-twin birth weight dependence as a function of gestational age using conditional copulas. Our initial investigation, shown in Figure 2.2, indicates a relatively stronger dependence between the birth weights (in grams) of the preterm (28–32 weeks) and post-term (38–42 weeks) twins compared to the twins delivered at term (33–37 weeks).

2.2 Framingham Heart Study

The Framingham Heart Study (FHS) is a longitudinal study conducted to identify factors associated with cardiovascular disease. Begun in 1948, the study details health records of the adult population of Framingham, MA over several decades. Using a subset of the data from 1956 to 1962, we consider 348 subjects who experienced a stroke during the rest of the follow-up period.

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, our interest is to investigate the dependence structure between the log-pulse pressures of the first two examination periods, \( \log(PP_1) \) and \( \log(PP_2) \), 1956 and 1962. We consider the change in body mass index (\( \Delta\text{BMI} \)) as the covariate.

For the initial investigation, we categorize the change in BMI, and focus on three broad cases, stable \((-0.7 \leq \Delta\text{BMI} \leq 0.7)\), increased \((\Delta\text{BMI} > 0.7)\) and decreased \((\Delta\text{BMI} < -0.7)\). Figure 2.3 indicates that when BMI increases the dependence between pulse pressures is weaker than in the other two cases.

2.3 Summary

In both data applications, it is difficult to specify how the strength of dependence varies with the covariate. Our initial investigation, in which we divided the data into subgroups according to the covariate range, as suggested by Joe (1997), may help with an
Figure 2.2: Twin birth data. Scatterplots of the twin birth weights for different ranges of gestational ages.
Figure 2.3: Framingham data. Scatterplots of the log-pulse pressures at different covariate ranges.
initial guess, but can be influenced by subjective choices. A more objective conclusion is obtained by using a nonparametric approach as an exploratory tool for detecting the underlying functional relationship between each copula parameter and the covariate.

As in the twin birth data, one can see different patterns of dependence depending on the range of the covariate value, suggesting that a copula model with a fixed parameter is not appropriate for our analysis.
Chapter 3

Dependence Calibration

This chapter contains our main proposal, referred to as dependence calibration (Acar et al., 2009). We briefly discuss the importance of the covariate adjustment in dependence modeling with a simple example. A conditional copula model from a parametric family of copulas is introduced by relaxing all functional assumptions on the copula parameter but that of adequate smoothness as a function of the covariate. A nonparametric estimation procedure based on local likelihood is used to estimate the unknown copula parameter function. The problem of choosing the bandwidth parameter is considered within the proposed framework. We examine asymptotic properties of the nonparametric estimator, discuss assessment of bias and variance and outline how to construct pointwise confidence intervals using these asymptotic results. A bootstrap confidence interval is also described to compensate for the impact of estimating conditional marginal distributions on the inference of the copula parameter.

3.1 Copulas versus Conditional Copulas

Let \( Y_1 \) and \( Y_2 \) be continuous variables whose statistical dependence is of interest and \( X \) a continuous covariate assumed to influence the strength/nature of the dependence between \( Y_1 \) and \( Y_2 \). If we ignore \( X \), there are several ways to study the relationship
Chapter 3. Dependence Calibration

between $Y_1$ and $Y_2$, from the Pearson’s correlation coefficient to copula analysis.

However, failing to account for the information brought in by $X$ can drastically alter our conclusions about dependence, and subsequent analyses. For instance, if of further interest is to predict $Y_2$ given $Y_1$, then $E(Y_2|Y_1)$ would not be equal to $E(Y_2|Y_1, X)$ unless $Y_2$ is independent of $X$ given $Y_1$.

To better illustrate this point, let us provide a simple example. The data on the left panel of Figure 3.1 is generated from the Clayton copula, with standard normal marginals. The copula parameter estimate under the Clayton copula is 1.09, which corresponds to 0.35 on Kendall’s tau scale. The right panel of Figure 3.1 displays the contour plot of the estimated bivariate density.

The data in fact consists of three different samples, each of size 200, generated from the Clayton copula with copula parameter values $\theta = 0.2, 1, \text{ and } 8$ (see Figure 3.2). Performing a separate analysis of the copula parameter, estimates are found to be $\hat{\theta} = 0.18, 1.10, \text{ and } 8.81$, with Kendall’s tau values of 0.08, 0.36, and 0.81, respectively. The contour plots of estimated bivariate densities are presented in Figure 3.2.

This example indicates that an unconditional copula model will yield the average
strength of dependence, whereas the conditional copula will capture the variation in the strength of dependence.

Note that, in practice, the covariate is observed but not the copula parameter. We need to know about their relationship in order to successfully account for the variation in the strength of dependence. In this simulation study, we consider the problem in the reverse sense, that is, the copula parameter is known but the covariate is not. In either case, we do not know the relationship between the copula parameter and the covariate, and in the following we provide a systematic approach to its inference.
3.2 Calibration Model

We consider the density equivalent of (1.17) with the conditional density $h_X(Y_1, Y_2| x; \theta, \alpha_1, \alpha_2)$ in which of main interest is the conditional copula parameter $\theta$, while the conditional marginal densities $f_{1|X}$ and $f_{2|X}$ are characterized by $\alpha_1$ and $\alpha_2$, respectively,

$$h_X(y_1, y_2| x; \theta, \alpha_1, \alpha_2) = f_{1|X}(y_1| x; \alpha_1) f_{2|X}(y_2| x; \alpha_2) c(u_1, u_2| x; \theta, \alpha_1, \alpha_2),$$

where $u_i = F_{i|X}(y_i| x; \alpha_i), i = 1, 2$ and $c(u_1, u_2| x; \theta, \alpha_1, \alpha_2)$ is the parametric conditional copula density with parameter $\theta$. We impose a minimal requirement that the parameters that govern the marginals are separate from the copula parameter. It is straightforward to demonstrate that such a requirement is not restrictive. In a regression setting, marginals may correspond to mean effects and the copula to covariance structure. Hence, the estimation can be performed in two stages, first for the marginal parameters and then for the copula. After replacing the estimates $\hat{F}_{1|X}(y_1| x)$ and $\hat{F}_{2|X}(y_2| x)$ in (1.17), we can estimate the functional form of the copula parameter.

Since our main focus is on the dependence structure, we assume that the conditional marginal distributions $F_{1|X}$ and $F_{2|X}$ are known and consider the following model:

$$(U_{1i}, U_{2i})|X_i \sim C\{u_{1i}, u_{2i}| \theta(x_i) \}, \quad (3.1)$$

where $\theta(x_i) = g^{-1}\{\eta(x_i)\}, \quad i = 1, \ldots, n$.

Here, $g^{-1}: \mathbb{R} \to \Theta$ is the known inverse link function, which ensures that the copula parameter has the correct range, and $\eta$ is the unknown calibration function to be estimated. The term calibration emphasizes that the level of dependence is adjusted for the covariate effect on the copula parameter. Analogous to generalized linear models, this model requires the choice of an appropriate link function to ensure that the estimate of $\theta$ is within the correct parameter range for the particular copula family under consideration. For instance, for the Clayton copula family $\theta \in (0, \infty)$, we use $g^{-1}(t) = \exp(t)$. If the link function is monotone then the choice is irrelevant since inference is invariant over monotone transformations of the copula parameter $\theta$. 
3.3 Local Copula-Likelihood Estimation

We begin with a classical polynomial functional form to motivate the proposed nonparametric strategy.

If the relationship between $\theta$ and $X$ falls into a pre-specified class of functions, for instance, polynomials up to degree $p$, we may estimate the calibration function $\eta(\cdot)$ via maximum likelihood estimation. More specifically, in this case we would write $\eta(X) = \sum_{j=0}^{p} \tilde{\beta}_j X^j$ and estimate $\tilde{\beta} = (\tilde{\beta}_0, \tilde{\beta}_1, \ldots, \tilde{\beta}_p)^T$ by maximizing

$$L(\tilde{\beta}) = \sum_{i=1}^{n} \ln c\{U_{1i}, U_{2i}| g^{-1}(\tilde{\beta}_0 + \tilde{\beta}_1 X_i + \cdots + \tilde{\beta}_p X_p^p)\}. \quad (3.2)$$

However, for most copula families, the function $\eta(\cdot)$ is not necessarily well-approximated by a preconceived polynomial model. Moreover, in contrast to classical regression, the form of the calibration function $\eta(\cdot)$ characterizing the underlying dependence structure is more difficult to discern by inspection. Therefore a nonparametric approach for estimating the target function is more needed here than in the classical regression context.

We adopt the local polynomial framework (Fan and Gijbels, 1996) within the local likelihood formulation (Tibshirani and Hastie, 1987) as follows. Assume $\eta$ has $(p + 1)^{th}$ continuous derivatives at an interior point $x$. For data points $X_i$ in the neighborhood of $x$, we approximate $\eta(X_i)$ by a Taylor expansion of order $p$,

$$\eta(X_i) \approx \eta(x) + \eta'(x)(X_i - x) + \cdots + \frac{\eta^{(p)}(x)}{p!}(X_i - x)^p \equiv x_{i,x}^T \beta,$$

where $x_{i,x} = (1, X_i - x, \ldots, (X_i - x)^p)^T$ and $\beta = (\beta_0, \beta_1, \ldots, \beta_p)^T$ with $\beta_\nu = \eta^{(\nu)}(x)/\nu!$. In our implementations we use the commonly adopted local linear fit, that is, $p = 1$, as in Fan and Gijbels (1996).

The contribution of each data point $(U_{1i}, U_{2i})|X_i$ in a neighborhood of $x$ to the local likelihood is given by $\ln c\{U_{1i}, U_{2i} | g^{-1}(x_{i,x}^T \beta)\}$. Let $L$ denote the local log-likelihood. Then, the weighted sum of contributions of each $(U_{1i}, U_{2i})|X_i$ forms the conditional local
log-likelihood
\[ L(\beta, x, p, h) = \sum_{i=1}^{n} \ln e \{ U_{1i}, U_{2i} | g^{-1}(x_{i,x}^{T} \beta) \} K_h(X_i - x), \]

(3.3)

where \( h \) is a bandwidth controlling the size of the local neighborhood and \( K_h(\cdot) = 1/h \) \( K(\cdot/h) \) with \( K \) a kernel function assigning weights to the data points in a local window. In our implementations we use the commonly adopted Epanechnikov kernel, \( K(z) = 3/4(1 - z^2)_+ \), where the subscript “+” denotes the positive part.

The local maximum likelihood estimator \( \hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p)^T \) is thus obtained by solving
\[ \nabla L(\beta, x) = \frac{\partial L(\beta, x, p, h)}{\partial \beta} = 0. \]

(3.4)

The numerical solution for (3.4) is found via the Newton-Raphson iteration
\[ \beta_{m+1} = \beta_m - \{\nabla^2 L(\beta_m, x)\}^{-1} \nabla L(\beta_m, x), \quad m = 0, 1, \ldots, \]

(3.5)

where
\[ \nabla L(\beta, x) = \frac{\partial L(\beta, x, p, h)}{\partial \beta} \]
\[ = \sum_{i} \frac{\partial}{\partial \beta} \ell \{ g^{-1}(x_{i,x}^{T} \beta), U_{1i}, U_{2i} \} K_h(X_i - x) \]
\[ = \sum_{i} \ell ' \{ g^{-1}(x_{i,x}^{T} \beta), U_{1i}, U_{2i} \} (g^{-1})' (x_{i,x}^{T} \beta) \ x_{i,x} K_h(X_i - x), \]

(3.6)

is the score vector, and
\[ \nabla^2 L(\beta, x) = \frac{\partial^2 L(\beta, x, p, h)}{\partial \beta^2} \]
\[ = \sum_{i} \frac{\partial}{\partial \beta} \left[ \ell ' \{ g^{-1}(x_{i,x}^{T} \beta), U_{1i}, U_{2i} \} (g^{-1})' (x_{i,x}^{T} \beta) \ x_{i,x} \right] K_h(X_i - x) \]
\[ = \sum_{i} \left[ \ell '^2 \{ g^{-1}(x_{i,x}^{T} \beta), U_{1i}, U_{2i} \} \ (g^{-1})' (x_{i,x}^{T} \beta)^2 \right. \]
\[ + \ell ' \{ g^{-1}(x_{i,x}^{T} \beta), U_{1i}, U_{2i} \} (g^{-1})'' (x_{i,x}^{T} \beta) \ x_{i,x} x_{i,x}^{T} K_h(X_i - x). \]

(3.7)
Chapter 3. Dependence Calibration

is the Hessian matrix.

We then obtain the estimator for the derivatives of the calibration function \( \eta^{(\nu)}(x), \ \nu = 0, \ldots, p \), where \( \hat{\eta}(x) = \hat{\beta}_0(x) \) is of particular interest. Finally, the copula parameter is estimated at covariate value \( x \) by applying the inverse link function

\[
\hat{\theta}(x) = g^{-1}\{\hat{\eta}(x)\}.
\]

3.4 Choice of Bandwidth

Choice of bandwidth parameter is an important issue in local estimation. Too small a bandwidth parameter will yield an estimator with a smaller bias but a greater variance, and the calibration function will be undersmoothed. Too large a bandwidth will produce less variance but a larger bias, and an oversmoothed calibration function.

Various methods for bandwidth selection exist, including cross-validation techniques and plug-in methods amongst others. Since our estimation procedure is based on the local copula likelihood, leave-one-out cross-validated local likelihood serves as a natural choice for the bandwidth selection.

Let \( \hat{\theta}_h(\cdot) \) denote the estimate of the copula parameter function depending on a bandwidth parameter \( h \). For each \( 1 \leq i \leq n \), we leave out the data point \( (U_{1i}, U_{2i}, X_i) \) and use the remaining data \( \{U_{1j}, U_{2j}, X_j, j \neq i\} \) to obtain \( \hat{\theta}_h^{(-i)}(X_i) \), the estimate of the copula parameter \( \theta \), at \( X_i \). The estimates obtained by leaving out the \( i \)th data point are then used to build the objective function depending on the bandwidth parameter

\[
B(h) = \sum_{i=1}^{n} \ln c\{U_{1i}, U_{2i} \mid \hat{\theta}_h^{(-i)}(X_i)\}.
\] (3.8)

The optimum bandwidth \( h^* \) is the one that maximizes (3.8).

In practice, when the underlying function is highly spiked and/or the covariate design highly nonuniform, we may use a nearest-neighbor type of variable bandwidth selection to choose a suitable proportion of data points contributing to each local estimate in a manner similar to (3.8).
3.5 Asymptotic Bias and Variance

Let \( f_X(\cdot) > 0 \) be the density function of the covariate \( X \). Denote the moments of \( K \) and \( K^2 \) by \( \mu_j = \int t^j K(t) dt \) and \( \nu_j = \int t^j K^2(t) dt \), respectively, and write the matrices
\[
S = (\mu_{j+\ell})_{0 \leq j, \ell \leq p}, \quad S^* = (\nu_{j+\ell})_{0 \leq j, \ell \leq p},
\]
the \((p + 1) \times (p + 1)\) vectors \( s_p = (\mu_{p+1}, \ldots, \mu_{2p+1})^T \), as well as the unit vector \( e_1 = (1, 0, \ldots, 0)^T \). For simplicity, we use \( \ell(\theta, U_1, U_2) = \ln c(U_1, U_2|\theta) \) for the log-copula density and denote its first and second derivatives with respect to \( \theta \) by \( \ell'(\theta, U_1, U_2) = \partial \ell(\theta, U_1, U_2)/\partial \theta \) and \( \ell''(\theta, U_1, U_2) = \partial^2 \ell(\theta, U_1, U_2)/\partial \theta^2 \).

For a fixed point \( x \) lying in the interior of the support of \( f_X \), we define \( \sigma^2(x) = -E\left( \ell''[g^{-1}\{\eta(x)\}, U_1, U_2] \mid X = x \right) \).

For our derivations, we require the following regularity assumptions, denoting \( N(x) \) the neighborhood of an interior point \( x \).

(A1) The functions \( \ell(\theta(x), U_1, U_2) \) and \( \ell''(\theta(x), U_1, U_2) \) exist and are continuous on \( N(x) \times (0, 1)^2 \), and can be bounded by integrable functions of \( u_1 \) and \( u_2 \) in \( N(x) \).

(A2) The functions \( f_X(\cdot), \eta^{(p+2)}, g''(\cdot) \text{ and } \sigma^2(\cdot) \) are continuous in \( N(x) \), and \( \sigma^2(x') \geq c \) for \( x' \in N(x) \) and some \( c > 0 \). Without loss of generality, the kernel density \( K(\cdot) \) has compact support \([-1, 1]\).

The assumption (A1) is to ensure that the copula density satisfies the first and second order Bartlett identities. Further discussion on condition (A1) for certain copula families can be found in Hu (1998) and Chen and Fan (2006). The mild regularity conditions in (A2) are commonly adopted in nonparametric regression.

Typically, an odd-order polynomial fit is preferred to an even-order fit in local polynomial modeling as the latter induces a higher asymptotic variance (see Fan and Gijbels, 1996, for details). Therefore, we consider only odd-order fits in the asymptotic expressions of conditional bias and variance. The following theorem summarizes our main results, denoting the collection of design variables \( \{X_1, \ldots, X_n\} \) by \( \mathbb{X} \).
Theorem 3. Assume that (A1) and (A2) hold, \( h \to 0 \) and \( nh \to \infty \) as \( n \to \infty \), for an odd-order local polynomial fit of degree \( p \). Then

\[
\begin{align*}
\text{Bias}(\hat{\eta}(x)|X) &= \mathbf{e}_1^T S^{-1} \mathbf{s}_p \frac{\eta^{(p+1)}(x)}{(p + 1)!} h^{p+1} + o_p(h^{p+1}), \\
\text{Var}(\hat{\eta}(x)|X) &= \frac{1}{nh f_X(x) [(g^{-1})'(\eta(x))]^2 \sigma^2(x)} \mathbf{e}_1^T S^{-1} S^* S^{-1} \mathbf{e}_1 + o_p \left( \frac{1}{nh} \right).
\end{align*}
\]

Proof of Theorem 3.

Let \( H = \text{diag}\{1, h, \ldots, h^p\} \), \( W = \text{diag}\{K_h(X_1 - x), \ldots, K_h(X_n - x)\} \) and

\[
X_x = \begin{pmatrix}
1 & X_1 - x & \cdots & (X_1 - x)^p \\
1 & X_2 - x & \cdots & (X_2 - x)^p \\
\vdots & \vdots & \ddots & \vdots \\
1 & X_n - x & \cdots & (X_n - x)^p
\end{pmatrix}.
\]

Define \((p+1) \times (p+1)\) matrices \( S_n = \sum_{i=1}^{n} \mathbf{x}_{i,x} \mathbf{x}_{i,x}^T K_h(X_i - x) \) and \( S_n^* = \sum_{i=1}^{n} \mathbf{x}_{i,x} \mathbf{x}_{i,x}^T K_h^2(X_i - x) \), with entries \( S_{n,j} = \sum_{i=1}^{n} (X_i - x)^j K_h(X_i - x) \) and \( S_{n,j}^* = \sum_{i=1}^{n} (X_i - x)^j K_h^2(X_i - x) \).

Our derivations rely on the Taylor expansion

\[
0 = \nabla \mathcal{L}(\hat{\beta}, x) \approx \nabla \mathcal{L}(\beta, x) + \nabla^2 \mathcal{L}(\beta, x) \{ \hat{\beta} - \beta \},
\]

where \( \beta = (\beta_0, \beta_1, \ldots, \beta_p)^T \) is the vector of true local parameters and \( \hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p)^T \) is the local likelihood estimator, yielding

\[
\hat{\beta} - \beta \approx - \{ \nabla^2 \mathcal{L}(\beta, x) \}^{-1} \nabla \mathcal{L}(\beta, x).
\]

Hence, the leading terms in the asymptotic conditional bias and variance can be written as

\[
\begin{align*}
E(\hat{\beta}|X) - \beta &\approx -E \{ \nabla^2 \mathcal{L}(\beta, x)|X \}^{-1} E \{ \nabla \mathcal{L}(\beta, x)|X \}, \quad (3.9) \\
\text{Var}(\hat{\beta}|X) &\approx E \{ \nabla^2 \mathcal{L}(\beta, x)|X \}^{-1} \text{Var} \{ \nabla \mathcal{L}(\beta, x)|X \} E \{ \nabla^2 \mathcal{L}(\beta, x)|X \}^{-1}. \quad (3.10)
\end{align*}
\]
In the following we approximate three terms appearing in (3.9) and (3.10). The expectation of the kernel weighted local score function is

\[ E \{ \nabla L(\beta, x) | X \} = \sum_{i=1}^{n} E[ \ell' \{ g^{-1}(x_{i,x}^T \beta), U_{1i}, U_{2i} \} | X] \ (g^{-1})'(x_{i,x}^T \beta) \ x_{i,x} \ K_h(X_i - x). \]

Since, for each \( i = 1, 2, \ldots, n \), we have \((U_{1i}, U_{2i}) | X_i \sim C\{u_1, u_2 | g^{-1}(\eta(X_i))\}\), from the first order Bartlett’s identity we have

\[ 0 = E[ \ell' \{ g^{-1}(\eta(X_i)), U_{1i}, U_{2i} \} | X] = E[ \ell' \{ g^{-1}(x_{i,x}^T \beta + r_i), U_{1i}, U_{2i} \} | X], \]

where \( r_i \) is the approximation error made by replacing the calibration function \( \eta(X_i) \) locally by a \( p^\text{th} \) degree polynomial and

\[ r_i = \sum_{j=p+1}^{\infty} \beta_j (X_i - x)^j = \beta_{p+1} (X_i - x)^{p+1} + o_p \{(X_i - x)^{p+1}\}. \tag{3.11} \]

If \( g^{-1} \) is continuous and twice differentiable, then

\[ \ell' \{ g^{-1}(\eta(X_i)), U_{1i}, U_{2i} \} \approx \ell' \{ g^{-1}(x_{i,x}^T \beta), U_{1i}, U_{2i} \} + \{r_i (g^{-1})'(x_{i,x}^T \beta)\} \ell'' \{ g^{-1}(x_{i,x}^T \beta), U_{1i}, U_{2i} \}, \]

and, after taking conditional expectations,

\[ E[\ell' \{ g^{-1}(x_{i,x}^T \beta), U_{1i}, U_{2i} \} | X] \approx - \{r_i (g^{-1})'(x_{i,x}^T \beta)\} E[\ell'' \{ g^{-1}(x_{i,x}^T \beta), U_{1i}, U_{2i} \} | X]. \]

Hence, we obtain

\[ E \{ \nabla L(\beta, x) | X \} \approx -E[ \ell'' \{ g^{-1}(\eta(x)), U_{1}, U_{2} \} | x] \ ((g^{-1})'(\eta(x)))^2 \ X_x^T W r, \]

where \( r = (r_1, r_2, \ldots, r_n)^T \).

Taking conditional expectation of the Hessian matrix in (3.7) yields

\[ E \{ \nabla^2 L(\beta, x) | X \} \approx E[ \ell'' \{ g^{-1}(\eta(x)), U_{1}, U_{2} \} | x] \ ((g^{-1})'(\eta(x)))^2 \ S_n. \]

The variance of the weighted local score function is

\[ Var \{ \nabla L(\beta, x) | X \} \approx \{(g^{-1})'(\eta(x)))^2 \ Var[ \ell' \{ g^{-1}(\eta(x)), U_{1}, U_{2} \} | x] \ S_n^*. \]
The second order Bartlett’s identity implies
\[
\sigma^2(x) = -E[\ell''(g^{-1}(\eta(x)), U_1, U_2)] = Var[\ell' \{g^{-1}(\eta(x)), U_1, U_2\}].
\]
Hence, we obtain
\[
E(\hat{\beta}|X) - \beta \approx S_{n}^{-1} X^T W r, \\
Var(\hat{\beta}|X) \approx \frac{1}{((g^{-1})'(\eta(x)))^2 \sigma^2(x)} S_{n}^{-1} S_{n}^{*} S_{n}^{-1}.
\]
Next, we need to find approximations for $S_n$, $S_n^*$ and $X^T W r$. The entries $S_{n,j} = \sum_{i=1}^{n} (X_i - x)^j K_h(X_i - x)$ of $S_n$ can be written as
\[
S_{n,j} = E S_{n,j} + O_p \left\{ \sqrt{Var(S_{n,j})} \right\}
= n h^j \int t^j K(t) f_X(x + ht) \, dt + O_p \left\{ \sqrt{n E[(X_1 - x)^{2j} K_h^2(X_1 - x)]} \right\}
= n h^j \left\{ f_X(x) \mu_j + o(1) + O_p \left( \frac{1}{\sqrt{nh}} \right) \right\}
= n h^j f_X(x) \mu_j \{1 + o_p(1)\},
\]
as $h \to 0$ and $nh \to \infty$. Thus, $S_n = n f_X(x) H S \{1 + o_p(1)\}$. Similarly, the entries $S_{n,j}^* = \sum_{i=1}^{n} (X_i - x)^j K_h^2(X_i - x)$ of $S_n^*$ can be approximated as
\[
S_{n,j}^* = E S_{n,j}^* + O_p \left\{ \sqrt{Var(S_{n,j}^*)} \right\}
= n h^{j-1} \int t^j K^2(t) f_X(x + ht) \, dt + O_p \left\{ \sqrt{n E[(X_1 - x)^{2j} K_h^4(X_1 - x)]} \right\}
= n h^{j-1} \left\{ f_X(x) \nu_j + o(1) + O_p \left( \frac{1}{\sqrt{nh}} \right) \right\}
= n h^{j-1} f_X(x) \nu_j \{1 + o_p(1)\},
\]
which yields $S_{n}^* = n h^{-1} f_X(x) H S^* H \{1 + o_p(1)\}$. Finally, the entries of $X^T W r$ can be written as
\[
k_j = \sum_{i=1}^{n} r_i (X_i - x)^j K_h(X_i - x)
= \sum_{i=1}^{n} \left\{ \sum_{m=p+1}^{\infty} \beta_m (X_i - x)^m \right\} (X_i - x)^j K_h(X_i - x)
= \beta_{p+1} S_{n,j+p+1} + o_p(n h^{j+p+1}).
\]
If we let $s_n = (S_{n,p+1}, S_{n,p+2}, \ldots, S_{n,2p+1})^T$, then we obtain $X_x^T W r = \beta_{p+1} s_n + o_p(nh^{p+1})$.

Hence, the bias and variance expressions become

$$E(\hat{\beta} | X) - \beta \approx H^{-1} S^{-1} s_p \beta_{p+1} h^{p+1} \{1 + o_p(1)\},$$

$$\text{Var}(\hat{\beta} | X) \approx \frac{1}{nh \ f_x(x) \ {(g^{-1})(\eta(x))}^2 \ \sigma^2(x)} H^{-1} S^{-1} S^* S^{-1} H^{-1} \{1 + o_p(1)\}.$$

The result of Theorem 3 is obtained by considering the first entry in the above expressions.

As a direct corollary of Theorem 3, we obtain the asymptotic conditional bias and variance of the copula parameter estimator, $\hat{\theta}(x) = g^{-1}\{\hat{\eta}(x)\}$.

**Corollary 1.** Assume that conditions of Theorem 3 holds, then

$$\text{Bias}(\hat{\theta}(x)|X) = e_1^T S^{-1} s_p \frac{\eta^{(p+1)}(x)}{g'\{\hat{\theta}(x)\} (p+1)!} h^{p+1} + o_p(h^{p+1}), \quad (3.12)$$

$$\text{Var}(\hat{\theta}(x)|X) = \frac{1}{nh \ f_x(x) \ \sigma^2(x)} e_1^T S^{-1} S^* S^{-1} e_1 + o_p\left(\frac{1}{nh}\right). \quad (3.13)$$

**Proof.** Since

$$\hat{\eta}(x) - \eta(x) = g(\hat{\theta}(x)) - g(\theta(x)) \approx g'(\theta(x)) \{\hat{\theta}(x) - \theta(x)\},$$

we reach the result of Corollary 1 by dividing the asymptotic bias and the variance of the local calibration estimator by $g'(\theta(x))$ and $\{g'(\theta(x))\}^2$, respectively. 

### 3.6 Confidence Intervals

One can use the result of Corollary 1 to approximate the bias and variance of the estimated copula parameter. The unknown quantity $\sigma^2(x)$ in the variance expression can be approximated by

$$\hat{\sigma}^2(x) = - \int_0^1 \int_0^1 \ell'' \{\hat{\theta}(x), U_1, U_2\} \ c\{U_1, U_2 \mid \hat{\theta}(x)\} \ dU_1 \ dU_2, \quad (3.14)$$
where $c(\cdot, \cdot)$ is the density corresponding to the underlying copula family. The approximate $100(1 - \alpha)\%$ pointwise confidence bands for the copula parameter are given by

$$\hat{\theta}(x) - \hat{b}(x) \pm z_{1-\alpha/2} \sqrt{V(x)}^{1/2}, \quad (3.15)$$

where $\hat{b}(x)$ and $\hat{V}(x)$ are the estimated bias and variance based on (3.12) and (3.13), and $z_{1-\alpha/2}$ is the $100(1 - \alpha/2)^{th}$ quantile of the standard normal distribution. In practice, estimating the bias can be difficult due to unknown higher-order derivatives (see Fan and Gijbels, 1996, for further discussion on bias correction). Alternatively, when variability plays a dominant role in (3.15), one may use a smaller bandwidth to reduce the bias to negligible levels (Fan and Zhang, 2000).

The asymptotic normality of the local copula parameter estimator can be shown under the following conditions. Without loss of generality, we may assume that the inverse link function $g^{-1}$ is the identity function and $p = 1$.

(B1) The density function $f_X(X) > 0$ of the covariate $X$ is Lipschitz continuous, and $X$ has a bounded support $\mathcal{X}$.

(B2) The kernel function $K(t)$ is symmetric, bounded and Lipschitz continuous.

(B3) The copula parameter function $\theta(x)$ has a continuous second derivative.

(B4) $E|\ell'(\theta(X), U_1, U_2)|X| < \infty$.

(B5) $E[\ell''\{\theta(X), U_1, U_2}\}|X]$ is Lipschitz continuous.

(B6) The function $\ell''(t, u_1, u_2) < 0$ for all $t \in \Theta$, and $u_1, u_2 \in (0, 1)$. For some integrable function $k$, and for $t_1$ and $t_2$ in a compact set,

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

In addition, let $\theta(x, X) = \theta(x) + \theta'(x)(X - x)$ and $z_x = (1, (X - x)/h)^T$. Then, for some constants $\xi > 2$ and $c_0 > 0$,

$$E \left\{ \sup_{x, \|a\| < c_0(nh)^{-1/2}} \left| \ell''(\theta(x, X) + a^T z_x, U_1, U_2) \right| \left| \frac{X - x}{h} \right|^{j-1} K \left( \frac{X - x}{h} \right) \right\}^\xi = O(1),$$
for \( j = 1, 2, 3 \).

As before, we denote the moments of \( K \) and \( K^2 \) by \( \mu_j \) and \( \nu_j \), respectively.

**Theorem 4.** Suppose the conditions B1-B6 hold and that \( h \to 0 \) and \( nh \to \infty \) as \( n \to \infty \). Then

\[
\left\{ \nu(x) \right\}^{-1/2} \left\{ \hat{\theta}(x) - \theta(x) - \frac{h^2}{2} \beta''(x) + o_p(h^2) \right\} \xrightarrow{\mathcal{L}} N(0,1),
\]

where \( \nu(x) = \frac{1}{nh} f_X(x) \sigma^2(x) \nu_0 \).

**Proof of Theorem 4.** Let \( \gamma_n = 1/\sqrt{nh} \) and define

\[
b = \gamma_n^{-1} (\beta_0 - \theta(x), h(\beta_1 - \theta'(x)))^T,
\]

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

\[
\beta_0 + \beta_1 (X_i - x) = \theta(x, X_i) + \gamma_n b^T z_{i,x},
\]

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

\[
\mathcal{L}(b, h) = \sum_{i=1}^n \left\{ \ell(\theta(x, X_i) + \gamma_n b^T z_{i,x}, U_{1i}, U_{2i}) K_h(X_i - x) \right\}.
\]

Note that \( \hat{b} = \gamma_n^{-1} (\hat{\beta}_0 - \theta(x), h(\hat{\beta}_1 - \theta'(x)))^T \) maximizes \( \mathcal{L}(b, h) \) since \( \hat{\beta} \) maximizes (3.3).

Equivalently, \( \hat{b} \) maximizes the following normalized function,

\[
\mathcal{L}^*(b) = \sum_{i=1}^n \left\{ \ell(\theta(x, X_i) + \gamma_n b^T z_{i,x}, U_{1i}, U_{2i}) - \ell(\theta(x, X_i), U_{1i}, U_{2i}) \right\} \left( \frac{X_i - x}{h} \right)
\]

\[
= h \gamma_n \sum_{i=1}^n \ell'(\theta(x, X_i), U_{1i}, U_{2i}) b^T z_{i,x} K_h(X_i - x)
\]

\[
+ h \gamma_n^2 \sum_{i=1}^n \ell''(\theta(x, X_i) + a_n^T z_{i,x}, U_{1i}, U_{2i}) (b^T z_{i,x})^2 K_h(X_i - x)
\]

\[
= b^T \left\{ \gamma_n \sum_{i=1}^n \ell'(\theta(x, X_i), U_{1i}, U_{2i}) z_{i,x} K \left( \frac{X_i - x}{h} \right) \right\}
\]

\[
+ \frac{1}{2} b^T \left\{ \frac{1}{n} \sum_{i=1}^n \ell''(\theta(x, X_i) + a_n^T z_{i,x}, U_{1i}, U_{2i}) z_{i,x} z_{i,x}^T K_h(X_i - x) \right\} b.
\]
In the following, we must show that

$$\frac{1}{n} \sum_{i=1}^{n} \ell''(\theta(x, X_i) + a_n^T z_{i,x}, U_{1i}, U_{2i}) z_{i,x} z_{i,x}^T K_h(X_i - x) = -\Delta + o_p(1),$$

where

$$\Delta = \sigma^2(x) f_X(x) \left( \begin{array}{cc} \mu_0 & \mu_1 \\ \mu_1 & \mu_2 \end{array} \right),$$

and $o_p(1)$ is uniform in $x \in \mathcal{X}$ and $||b|| < a_0$, for some fixed constant $a_0 > 0$. To show this, we need the following smoothness result.

Consider $A_n(x, a) = \ell''(\theta(x, X) + a^T z_x, U_1, U_2) z_x z_x^T K_h(X - x)$, with $||a|| < 1$. Then, under the conditions B1-B6, we can show that

$$|A_n(x_1, a_1) - A_n(x_2, a_2)| \leq |A_n(x_1, a_1) - A_n(x_1, a_2)| + |A_n(x_1, a_2) - A_n(x_2, a_2)|$$

$$\leq |\ell''(\theta(x_1, X) + a_1^T z_{x_1}, U_1, U_2) - \ell''(\theta(x_2, X) + a_2^T z_{x_2}, U_1, U_2)| ||z_{x_1} z_{x_1}^T K_h(X - x_1)||$$

$$+ |\ell''(\theta(x_1, X) + a_2^T z_{x_2}, U_1, U_2) - \ell''(\theta(x_2, X) + a_2^T z_{x_2}, U_1, U_2)| ||z_{x_2} z_{x_2}^T K_h(X - x_1)||$$

$$+ |\ell''(\theta(x_2, X) + a_2^T z_{x_2}, U_1, U_2)| ||z_{x_1} z_{x_1}^T K_h(X - x_1) - z_{x_2} z_{x_2}^T K_h(X - x_2)||$$

$$\leq \frac{1}{h^3} k(X, U_1, U_2)(||a_1 - a_2|| + |x_1 - x_2|)$$

for some integrable function $k(X, U_1, U_2)$. Using the triangle inequality, we can write

$$\left| \frac{1}{n} \sum_{i=1}^{n} \ell''(\theta(x, X_i) + a_n^T z_{i,x}, U_{1i}, U_{2i}) z_{i,x} z_{i,x}^T K_h(X_i - x) - (-\Delta) \right|$$

$$\leq \frac{1}{n} \sum_{i=1}^{n} \left| \ell''(\theta(x, X_i) + a_n^T z_{i,x}, U_{1i}, U_{2i}) - \ell''(\theta(x, X_i), U_{1i}, U_{2i}) \right| ||z_{i,x} z_{i,x}^T K_h(X_i - x)||$$

$$+ \sup_{\theta, x} \left[ \frac{1}{n} \sum_{i=1}^{n} \left| \ell''(\theta(x, X_i), U_{1i}, U_{2i}) - \ell''(\theta(X_i), U_{1i}, U_{2i}) \right| z_{i,x} z_{i,x}^T K_h(X_i - x) \right]$$

$$+ \sup_{\theta, x} \left[ \frac{1}{n} \sum_{i=1}^{n} \ell''(\theta(X_i), U_{1i}, U_{2i}) z_{i,x} z_{i,x}^T K_h(X_i - x) \right]$$

$$- E \left\{ \ell''(\theta(X), U_1, U_2) z_x z_x^T K_h(X - x)|x \right\}$$

$$+ E \left\{ \ell''(\theta(X), U_1, U_2) z_x z_x^T K_h(X - x)|x \right\} + \Delta \right]$$
for $\theta$ 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 $h_n(\xi-2)/\xi = O(1)$ and $||b|| < a_0$, for some fixed constant $a_0 > 0$. 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. Thus, we obtain

$$L^*(b) = b^T W_n(x) - \frac{1}{2} b^T \Delta b \left(1 + o_p(1)\right),$$

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

$$W_n(x) = \gamma_n \sum_{i=1}^n \ell'(\theta(x, X_i), U_{1i}, U_{2i}) z_{i,x} K((X_i - x)/h).$$

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

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

provided that $W_n$ is a stochastically bounded sequence of random vectors. The asymptotic normality of $\hat{b}$ follows from that of $W_n$.

To show asymptotic normality, we must first find the mean and covariance matrix of $W_n$. Using similar arguments to those in the proof of Theorem (3), we obtain

$$E(W_n) = \frac{h^2}{2\gamma_n} f_X(x) \sigma^2(x) \begin{pmatrix} \mu_2 \\ \mu_3 \end{pmatrix} \theta''(x)\{1 + o(1)\}$$

$$Var(W_n) = f_X(x) \sigma^2(x) \begin{pmatrix} \nu_0 & \nu_1 \\ \nu_1 & \nu_2 \end{pmatrix} \{1 + o(1)\} = \Lambda + o(1).$$

By the Cramér-Wold device, to obtain the asymptotic normality of $W_n$, it suffices to show that for any unit vector $a$,

$$\{a^T Var(W_n)a\}^{-1/2} (a^T W_n - a^T E(W_n)) \overset{D}{\to} N(0, 1),$$

which can be verified by checking the Lyapounov condition. Let $\psi_i = \ell'(\theta(x, X_i), U_{1i}, U_{2i}) a^T z_{i,x} K((X_i - x)/h)$. Then $a^T W_n = \gamma_n \sum_{i=1}^n \psi_i$. Using B2 and B4, we can show that

$$O \left(n \gamma_n^3 E |\ell'(\theta(x, X_1), U_{11}, U_{21}) a^T z_{1,x} K((X_1 - x)/h)|^3\right) = O(\gamma_n) = o(1).$$
as \( nh \to \infty \). Thus, the Lyapounov condition is satisfied and we obtain

\[
\hat{b} - \frac{n^{1/2} h^{5/2}}{2} \Delta^{-1} f_X(x) \sigma^2(x) \begin{pmatrix} \mu_2 \\ \mu_3 \end{pmatrix} \theta''(x) \{1 + o(1)\} \xrightarrow{\mathcal{L}} N(0, \Delta^{-1} \Lambda \Delta^{-1}).
\]

If \( K \) is symmetric, then \( \mu_1 = \mu_3 = 0 \) and

\[
\Delta^{-1} = \frac{1}{f_X(x) \sigma^2(x)} \begin{pmatrix} 1 & 0 \\ 0 & 1/\mu_2 \end{pmatrix}.
\]

Thus, we obtain

\[
\sqrt{nh} \left\{ \begin{pmatrix} \hat{\theta}(x) - \theta(x) \\ h \{\hat{\theta}'(x) - \theta'(x)\} \end{pmatrix} - \frac{h^2}{2} \begin{pmatrix} \mu_2 \\ \mu_3/\mu_2 \end{pmatrix} \theta''(x) \{1 + o(1)\} \right\} \xrightarrow{\mathcal{L}} N(0, \Delta^{-1} \Lambda \Delta^{-1}).
\]

The result directly follows from the first entry.

**3.7 Impact of Estimating Conditional Marginals**

In practice the estimation of the conditional marginal distributions may have an impact on the inference of the copula parameter. If the conditional marginal distributions can be adequately characterized by a parametric model, the \( \sqrt{n} \)-convergence rate is negligible when compared to the nonparametric convergence rate. The additional variability due to the estimation of the conditional marginals can thus be ignored.

In the case of the conditional marginal distributions estimated nonparametrically, the rate of convergence is of the same order as that of the copula estimator. It is thus difficult to analytically assess the two sources of uncertainty in a unified manner. A theoretical investigation of this scenario is of substantial interest, but requires a different framework. In practice, one viable way to incorporate the uncertainty from nonparametrically estimated marginals is to bootstrap the raw data and calculate the quantile-based bootstrap confidence bands. Thus in the absence of an adequate parametric marginal model, we suggest the use of the raw bootstrap approach. Nevertheless, the construction
of more computationally efficient and theoretically justified inference procedures is of great importance and constitutes a central topic of future efforts.

3.8 Summary and Discussion

We introduced a nonparametric procedure to estimate the functional relationship between the copula parameter and the covariate for a given parametric family of copulas. The *leave-one-out* cross-validated likelihood criterion was used to select the bandwidth parameter. We examined the asymptotic properties of the local copula parameter estimator and described how to construct pointwise confidence intervals when the conditional marginals are known or estimated.

Our approach is intuitive and flexible and allows us to directly study covariate effects on the strength of dependence. Although attention has focused on bivariate copulas, the above dependence calibration procedure can easily be extended to multivariate copulas as well.

The extension to include more than one covariate in the calibration function is, however, troubled by the curse of dimensionality. In this case, we recommend a careful selection of the variables prior to estimation of the calibration function. An adaptation of additive models, assuming an additive structure in the calibration functions associated with each covariate as per Tibshirani (1990), is an interesting future research question.
Chapter 4

Conditional Copula Selection and Copula Parameter Inference

This chapter consists of two parts. The first part concerns the choice of an appropriate family of copulas in a conditional copula model and introduces a novel conditional copula selection method based on cross-validated prediction errors. The focus of the second part is the nonparametric inference of the copula parameter of conditional copulas. We derive a generalized likelihood ratio test to determine whether an estimated copula parameter function significantly varies as a covariate and whether it can be adequately characterized by a certain parametric form (Acar et al., 2010).

4.1 Conditional Copula Selection

All inferential methods for copulas must be accompanied by a strategy to select among a number of copula families the one that best represents the data at hand. Several copula selection and goodness-of-fit tests have been proposed, but no single approach has yet emerged as a standard.

The best representation of the data can be sought by comparing the closeness of different families with a benchmark which is directly estimated from the data. The latter
can be an empirical copula (Durrleman et al., 2000), a nonparametric estimate of the copula density (Scaillet, 2005; Fermanian, 2005; Craiu and Craiu, 2008), the empirical distribution of a copula (Genest and Rivest, 1993), and different distance measures, such as L₂-norm, Kolmogorov and Kullbeck-Leibler, can be used to define closeness.

A further interest in evaluating the significance of a decision has directed attention towards goodness-of-fit procedures. The multivariate nature of the problem and the estimation of model components, i.e. the copula parameter and the marginal distributions, are the major complications in testing the goodness-of-fit of copulas. Besides the distance based goodness-of-fit tests, a pseudo-likelihood ratio test was proposed by Chen and Fan (2005) for semiparametric copula models. Genest et al. (2009) and Berg (2009) provide extensive surveys in this area.

A novel contribution to the copula selection literature was offered by Huard et al. (2006) in a Bayesian copula selection method. Treating the copula parameter as nuisance, this approach does not require parameter estimation prior to the copula selection, hence allows direct comparisons among copula families regardless of the parameter choice.

Unfortunately, there is no simple way to adapt these approaches to the problem of conditional copula selection. Choosing a benchmark model requires fully nonparametric estimation of the conditional copula or its density, which is possible using nonparametric kernel estimation, but can be tedious. A further complication that arises in adapting the goodness-of-fit tests is the variation in the conditional copula parameter, i.e. the data vectors are independent but not identically distributed. A Bayesian approach to conditional copula selection, on the other hand, is not recommended, since the variation in the copula parameter is of main interest in conditional copulas and should not be treated as nuisance.

In our selection procedure, we address both the functional form of the copula and the functional form of the copula parameter.
4.1.1 Proposed Method

Suppose we have a (finite) set of candidate families \( \mathcal{C} = \{C_q : q = 1, \ldots, Q\} \), from which we want to choose the one that best represents the data at hand. Each copula family is characterized by a copula parameter function, \( \theta_q(\cdot), q = 1, \ldots, Q \), which is estimated using the method proposed in Chapter 3. Since the estimation depends on the bandwidth parameter \( h \), we denote these estimates by \( \hat{\theta}_{h_q}(X), q = 1, \ldots, Q \). Hence, from each family there is a candidate model and we face the task of choosing the family whose representative best fits the data.

The general principle of cross-validated likelihood used in bandwidth selection does not apply to the selection of the copula family, as the scale of likelihoods varies across families. It is necessary to characterize the goodness-of-fit using different families on a comparable benchmark. Here we propose the cross-validated prediction of each response variable based on the other in a symmetric fashion. One can certainly modify the proposed criterion below if the variables are not of equal interest.

Note that, for each left-out sample point \((U_{1i}, U_{2i}, X_i)\), we have the estimate for the conditional copula parameter at the optimum bandwidth \( \hat{\theta}_{h_q}^{(-i)} \), which, in turn, leads to the best candidate model from the \( q \)th family, \( C_q\{U_{1i}, U_{2i} \mid \hat{\theta}_{h_q}^{(-i)}(X_i)\} \), with \( i = 1, \ldots, n \) and \( q = 1, \ldots, Q \).

We use the conditional expectation formula to measure the predictive ability of each candidate model. Within family \( C_q \), the cross-validated conditional prediction for \( U_{1i} \) is

\[
\hat{E}_q^{(-i)}(U_{1i} \mid U_{2i}, X_i) = \int_0^1 U_1 \ c_q\{U_1, U_{2i} \mid \hat{\theta}_{h_q}^{(-i)}(X_i)\} \ dU_1.
\]

Then, the cross-validated prediction errors (CVPE) is used to define the model selection criterion

\[
\text{CVPE}(C_q) = \sum_{i=1}^n \left[ \left\{ U_{1i} - \hat{E}_q^{(-i)}(U_{1i} \mid U_{2i}, X_i) \right\}^2 + \left\{ U_{2i} - \hat{E}_q^{(-i)}(U_{2i} \mid U_{1i}, X_i) \right\}^2 \right]. \quad (4.1)
\]

We select the copula family which yields the minimum CVPE value. Here we give a
heuristic justification of the model. If we denote $M_0$ the true copula family and $M$ the working copula family, then the first part in (4.1) normalized by $1/n$ is an approximation of $E_{M_0}[\{U_1 - E_M(U_1|U_2, X)\}^2|U_2, X]$. Then $E_{M_0}[\{U_1 - E_M(U_1|U_2, X)\}^2|U_2, X]$ is minimized when the model $M$ is correctly specified, i.e. $M = M_0$. This can be easily shown as follows

$$E_{M_0}[\{U_1 - E_M(U_1|U_2, X)\}^2|U_2, X]$$

$$= E_{M_0}[\{U_1 \pm E_{M_0}(U_1|U_2, X) - E_M(U_1|U_2, X)\}^2|U_2, X]$$

$$= E_{M_0}[\{U_1 - E_{M_0}(U_1|U_2, X)\}^2|U_2, X]$$

$$+ E_{M_0}[\{E_{M_0}(U_1|U_2, X) - E_M(U_1|U_2, X)\}^2|U_2, X]$$

$$+ 2E_{M_0}[\{U_1 - E_{M_0}(U_1|U_2, X)\}\{E_{M_0}(U_1|U_2, X) - E_M(U_1|U_2, X)\}|U_2, X]$$

$$> E_{M_0}[\{U_1 - E_{M_0}(U_1|U_2, X)\}^2|U_2, X].$$

A similar result holds for the second part in (4.1).

Note that, our conditional copula selection method is simple and intuitive. It directly follows from our bandwidth selection procedure without extra computational cost.

Here we assumed the marginal distributions were known. A similar approach can be taken when the marginals are estimated. However, for copula selection, we suggest considering the predictive accuracy on a uniform scale rather than the scale of the response variables, as the latter evaluates the performance of the joint model, both conditional marginals and the conditional copula.

### 4.2 Copula Parameter Inference

#### 4.2.1 Introduction

The nonparametric conditional copula parameter estimator serves two main purposes. The first is explanatory, and the second diagnostic. Its role as an explanatory tool to assess the copula parameter function has been discussed in previous chapters. We now
consider how it can be used in model diagnostics for parametrically specified copula parameters.

A typical diagnostic of interest is to determine whether a covariate has significant impact on the strength of dependence. We can visually inspect the variation in the nonparametric copula parameter estimates across the covariate range or, as outlined in Chapter 3, we can construct pointwise confidence intervals to check whether an estimated constant copula parameter falls within the confidence bands. Similar inspections can be performed for other parametric forms, e.g. linear in covariate. Although helpful, these approaches are not sufficient to reach a conclusion. Therefore, we need to take one further step towards completing the nonparametric inference for the conditional copula parameter.

Nonparametric inference has developed slowly in comparison to nonparametric estimation methods. By deriving the asymptotic theory for the maximal deviations of the kernel density estimator, Bickel and Rosenblatt (1973) spurred research on simultaneous confidence bands and the development of many tests based on discrepancy measures. In kernel regression, Härdle (1989) extended the Bickel-Rosenblatt result, Eubank and Speckman (1993) proposed bias-corrected confidence bands and Härdle and Marron (1991) provided bootstrap simultaneous error bars. There have been considerable efforts in testing the adequacy of a parametric model using nonparametric methods, known as smooth goodness-of-fit tests (see Hart, 1997, for a book-length reference). Härdle and Mammen (1993) considered weighted $L_2$ distances in their test and proposed a wild bootstrap procedure to obtain its critical values. Azzalini et al. (1989) proposed a pseudo-likelihood test for nonparametric regression problems. In fact, most of the smooth goodness-of-fit tests are developed for nonparametric regression problems, and only a few of them are adoptable to other settings.

A more general and widely applicable approach was proposed by Fan et al. (2001) in their generalized likelihood ratio tests (GLRT). The idea of generalized likelihood ratio
test is to extend the likelihood ratio tests for testing a parametric or a nonparametric null hypothesis versus a nonparametric alternative hypothesis. Since nonparametric maximum likelihood estimators are hard 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 linear estimator to specify the alternative model of a number of hypothesis testing problems, Fan et al. (2001) showed that the null distribution of the generalized likelihood ratio statistic follows asymptotically a $\chi^2$ distribution with 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 includes 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 now introduce a generalized likelihood ratio test for conditional copula models.

4.2.2 A generalized likelihood ratio test

Suppose we observe $\mathbf{(U_1, U_2, X_i)}$, $i = 1, \ldots, n$, with $U_{1i}$ and $U_{2i}$ uniformly distributed on the unit interval $(0, 1)$, from the conditional copula model (3.1). Of interest is whether or not the covariate $X$ has a significant impact on the strength of dependence between $Y_1$ and $Y_2$. This leads to the following simple hypothesis testing problem.

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta \neq \theta_0,$$

where $\theta_0 \in \mathbb{R}$. We construct a generalized likelihood ratio statistic using the local linear estimator, with $p = 1$, proposed in Chapter 3. Without loss of generality, we assume that the inverse link function $g^{-1}$ is the identity function.

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

$$\mathbb{L}_n(H_0) = \sum_{i=1}^{n} \ell(\theta_0, U_{1i}, U_{2i}),$$

(4.3)
where $\ell(t, u_1, u_2) = \ln c(u_1, u_2|t)$. Denoting $\hat{\theta}_h(x) = \hat{\theta}(x)$ the local linear estimator at a fixed point $x$ with the bandwidth parameter $h$, we can write the log-likelihood under the model (3.1) as

$$\mathbb{L}_n(H_1, h) = \sum_{i=1}^{n} \ell(\hat{\theta}(X_i), U_{1i}, U_{2i}).$$

(4.4)

The difference between the two log-likelihoods allows us to evaluate the evidence in the data in favor of or against the null model. We define the generalized likelihood ratio statistic as

$$\lambda_n(h) = \mathbb{L}_n(H_1, h) - \mathbb{L}_n(H_0).$$

(4.5)

Intuitively, large values of $\lambda_n(h)$ suggest the rejection of the null hypothesis.

One may follow a similar approach for testing composite null hypothesis of the form

$$H_0 : \theta \in \Theta_0 \quad \leftrightarrow \quad H_1 : \theta \notin \Theta_0,$$

(4.6)

where $\Theta_0$ is a set of predefined functions. For instance, for $H_0 : \theta = \alpha_0 + \alpha_1(X)$, we can use the maximum likelihood estimator of $(\alpha_0, \alpha_1)$ to obtain the log-likelihood under the null hypothesis (see Fan et al., 2001, for details).

Note that (4.5) defines a class of tests indexed by $h$. Hence, the choice of bandwidth parameter plays an important role in the formulation and performance of the generalized likelihood ratio test. Intuitively, a large (small) bandwidth parameter is more powerful for testing smooth (less smooth) alternatives. Bandwidth selection based on power maximization is, however, a challenging problem, which has not been completely solved yet. In practice, a natural choice is to use the bandwidth parameter employed in the estimation also in the nonparametric hypothesis testing. (Fan and Jiang, 2007).

### 4.2.3 Asymptotic null distribution

The asymptotic null distribution of the generalized likelihood ratio statistic relies on the following technical conditions.
(C1) The density function \( f(X) > 0 \) of the covariate \( X \) is Lipschitz continuous, and \( X \) has a bounded support \( \mathcal{X} \).

(C2) The kernel function \( K(t) \) is symmetric, bounded and Lipschitz continuous.

(C3) The copula parameter function \( \theta(x) \) has a continuous second derivative.

(C4) \( E \left| \ell'(\theta(x), u_1, u_2) \right|^4 < \infty \).

(C5) \( E(\ell''(\theta(x), u_1, u_2) | x) \) is Lipschitz continuous.

(C6) The function \( \ell''(t, u_1, u_2) < 0 \) 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,

\[
|\ell''(t_1, u_1, u_2) - \ell''(t_2, u_1, u_2)| < k(u_1, u_2)|t_1 - t_2|.
\]

In addition, for some constants \( \xi > 2 \) and \( c_0 > 0 \),

\[
E \left\{ \sup_{x, \|a\| < c_0 (nh)^{-1/2}} |\ell''(\theta(x, X) + a^T z_x, U_1, U_2)| \left| \frac{X - x}{h} \right|^{j-1} K \left( \frac{X - x}{h} \right) \right\}^\xi = O(1),
\]

for \( j = 1, 2, 3 \).

**Theorem 5.** Assume that (C1)-(C6) hold. Then, under the null hypothesis of (4.2), as \( h \to 0 \) and \( nh^{3/2} \to \infty \),

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

Furthermore, if \( \theta \) is linear or \( nh^{9/2} \to 0 \), then as \( nh \to \infty \),

\[
r_K \lambda_n(h) \xrightarrow{\text{asym}} \chi^2_{r_K \mu_n},
\]

where \( r_K = 2 \mu_n/\nu_n \),

\[
\mu_n = \frac{|\mathcal{X}|}{h} (K(0) - \frac{1}{2} \int K^2(t) dt),
\]

\[
\nu_n = \frac{2|\mathcal{X}|}{h} \int (K(t) - \frac{1}{2} K * K(t))^2 dt,
\]

\[
d_n = O_p(nh^4 + n^{3/2} h^2).
\]

Here \( K * K \) denotes the convolution of \( K \) and \( |\mathcal{X}| \) is the range of the covariate \( X \).
Proof. Let \( \gamma_n = 1/\sqrt{nh} \) and \( \theta(x, X_i) = \theta(x) + \theta'(x)(X_i - x) \). Define

\[
\alpha_n(x) = \frac{\gamma_n^2}{\sigma^2(x)f(x)} \sum_{i=1}^{n} \ell'(\theta(X_i), U_{1i}, U_{2i}) K\left(\frac{X_i - x}{h}\right),
\]

\[
R_n(x) = \frac{\gamma_n^2}{\sigma^2(x)f(x)} \sum_{i=1}^{n} \left[ \ell'(\theta(x, X_i), U_{1i}, U_{2i}) - \ell'\left(\theta(X_i), U_{1i}, U_{2i}\right) \right] K\left(\frac{X_i - x}{h}\right)
\]

and set

\[
R_{n1} = \sum_{k=1}^{n} \ell'(\theta(X_k), U_{1k}, U_{2k}) R_n(X_k),
\]

\[
R_{n2} = -\sum_{k=1}^{n} \ell''(\theta(X_k), U_{1k}, U_{2k}) \alpha_n(X_k) R_n(X_k),
\]

\[
R_{n3} = -\frac{1}{2} \sum_{k=1}^{n} \ell''(\theta(X_k), U_{1k}, U_{2k}) R_n^2(X_k).
\]

Also, define

\[
T_{n1} = \gamma_n^2 \sum_{i=1}^{n} \sum_{k=1}^{n} \frac{\ell'(\theta(X_k), U_{1k}, U_{2k})}{\sigma^2(X_k)f(X_k)} \ell'(\theta(X_i), U_{1i}, U_{2i}) K\left(\frac{X_i - X_k}{h}\right),
\]

\[
T_{n2} = \gamma_n^4 \sum_{i=1}^{n} \sum_{j=1}^{n} \ell'(\theta(X_i), U_{1i}, U_{2i}) \ell'(\theta(X_j), U_{1j}, U_{2j}) \times \left\{ \sum_{k=1}^{n} \frac{\ell''(\theta(X_k), U_{1k}, U_{2k})}{(\sigma^2(X_k)f(X_k))^2} K\left(\frac{X_i - X_k}{h}\right) K\left(\frac{X_j - X_k}{h}\right) \right\}.
\]

The following lemmas will be used in our derivations. Their proofs can be found under the additional technical results.

**Lemma 1.** Under conditions (C1) - (C6),

\[
\hat{\theta}(x) - \theta(x) = (\alpha_n(x) + R_n(x)) (1 + o_p(1)).
\]

**Lemma 2.** Under conditions (C1) - (C6), as \( h \to 0 \) and \( nh^{3/2} \to \infty \)

\[
T_{n1} = \frac{1}{h} K(0) \mathbb{E} f^{-1}(X) + \frac{1}{n} \sum_{k \neq i} \frac{\ell'(\theta(X_k), U_{1k}, U_{2k})}{\sigma^2(X_k)f(X_k)} \ell'(\theta(X_i), U_{1i}, U_{2i}) K_h(X_i - X_k) + o_p(h^{-1/2}),
\]

\[
T_{n2} = -\frac{1}{h} \mathbb{E} f^{-1}(X) \int K^2(t) dt - \frac{2}{nh} \sum_{i<j} \frac{\ell''(\theta(X_i), U_{1i}, U_{2i})}{\sigma^2(X_i)f(X_i)} \ell'(\theta(X_j), U_{1j}, U_{2j}) K * K((X_j - X_i)/h) + o_p(h^{-1/2}).
\]
We now start our derivations for the asymptotic null distribution of

\[ \lambda_n(h) = \sum_{k=1}^{n} \{ \ell(\hat{\theta}(X_k), U_{1k}, U_{2k}) - \ell(\theta_0, U_{1k}, U_{2k}) \}. \]

Approximation of \( \ell(\hat{\theta}(X_k), U_{1k}, U_{2k}) \) around \( \theta_0 = \theta(X_k) \) yields

\[ \lambda_n(h) \approx \sum_{k=1}^{n} \ell'(\theta(X_k), U_{1k}, U_{2k}) \{ \hat{\theta}(X_k) - \theta(X_k) \} + \frac{1}{2} \sum_{k=1}^{n} \ell''(\theta(X_k), U_{1k}, U_{2k}) \{ \hat{\theta}(X_k) - \theta(X_k) \}^2. \]

By Lemma 1, we have

\[
\begin{align*}
- \lambda_n(h) & = - \sum_{k=1}^{n} \ell'(\theta(X_k), U_{1k}, U_{2k}) \alpha_n - \frac{1}{2} \sum_{k=1}^{n} \sum_{k=1}^{n} \ell''(\theta(X_k), U_{1k}, U_{2k}) \alpha_n^2 \\
& \quad - \sum_{k=1}^{n} \ell'(\theta(X_k), U_{1k}, U_{2k}) R_n - \sum_{k=1}^{n} \ell''(\theta(X_k), U_{1k}, U_{2k}) \alpha_n R_n \\
& \quad - \frac{1}{2} \sum_{k=1}^{n} \ell''(\theta(X_k), U_{1k}, U_{2k}) R_n^2 + O_p\left(\frac{1}{nh^2}\right) \\
& = - \gamma_n^2 \sum_{k=1}^{n} \frac{\ell'(\theta(X_k), U_{1k}, U_{2k})}{\sigma^2(X_k)f(X_k)} \sum_{i=1}^{n} \ell'(\theta(X_i), U_{1i}, U_{2i}) K\left(\frac{X_i - X_k}{h}\right) \\
& \quad - \gamma_n^4 \sum_{k=1}^{n} \frac{\ell''(\theta(X_k), U_{1k}, U_{2k})}{(\sigma^2(X_k)f(X_k))^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \ell'(\theta(X_i), U_{1i}, U_{2i}) \ell'(\theta(X_j), U_{1j}, U_{2j}) \\
& \quad \times K\left(\frac{X_i - X_k}{h}\right) K\left(\frac{X_j - X_k}{h}\right) \\
& \quad - R_{n1} + R_{n2} + R_{n3} + O_p\left(\frac{1}{nh^2}\right).
\end{align*}
\]

Applying Lemma 2, we obtain

\[
\begin{align*}
- \lambda_n(h) & = - \frac{1}{h} \mathbb{E} f^{-1}(X) \left\{ K(0) - \frac{1}{2} \int K^2(t) dt \right\} \\
& \quad - \frac{1}{n} \sum_{i \neq j} \ell'(\theta(X_i), U_{1i}, U_{2i}) \ell'(\theta(X_j), U_{1j}, U_{2j}) K_{h}(X_j - X_i) \\
& \quad + \frac{1}{n} \sum_{i < k} \ell'(\theta(X_i), U_{1i}, U_{2i}) \ell'(\theta(X_j), U_{1j}, U_{2j}) K_{h} * K_{h}(X_j - X_i) \\
& \quad - R_{n1} + R_{n2} + R_{n3} + O_p\left(\frac{1}{nh^2}\right) + o_p(h^{-1/2}).
\end{align*}
\]
Calculating the leading terms of $R_{n1}$, $R_{n2}$ and $R_{n3}$, we have

\[
R_{n1} = \sum_{k=1}^{n} \frac{h^2}{2} \ell'(\theta(X_k), U_{1k}, U_{2k}) \theta''(X_k) \int t^2 K(t) dt (1 + o_p(1)) = O_p(n^{1/2}h^2)
\]

\[
-R_{n2} = \sum_{k=1}^{n} \frac{h^2}{4} \frac{\ell'(\theta(X_k), U_{1k}, U_{2k})}{\sigma^2(X_k)f(X_k)} \theta''(X_k) \omega_0 (1 + o_p(1)) = O_p(n^{1/2}h^2)
\]

\[
-R_{n3} = \frac{nh^4}{8} E\theta'(X)^2\sigma^2(X) \omega_0 (1 + o_p(1)) = O_p(nh^4)
\]

where $\omega_0 = \int \int t^2(s+t)K(t)K(s+t) ds dt$. Thus $R_{n3} - (R_{n1} - R_{n2}) = O_p(nh^4 + n^{1/2}h^2)$ and we directly obtain

\[
- \lambda_n(h) = -\mu_n + d_n - \frac{1}{2} h^{-1/2} W(n) + o_p(h^{-1/2})
\]

where

\[
W(n) = \frac{\sqrt{h}}{n} \sum_{i \neq j} \frac{1}{(\sigma^2(X_i)f(X_i))^2} \ell'(\theta(X_j), U_{1j}, U_{2j}) \ell'(\theta(X_i), U_{1i}, U_{2i}) \{2K_h(X_j - X_i) - K_h * K_h(X_j - X_i)\}.
\]

Next we show that

\[
W(n) \xrightarrow{\text{L}} N(0, \nu^*),
\]

where $\nu^* = 2 \|2K - K * K\|_2^2 Ef^{-1}(X)$ by letting

\[
W_{ij} = \frac{\sqrt{h}}{n} B_n(i, j) \ell'(\theta(X_i), U_{1i}, U_{2i}) \ell'(\theta(X_j), U_{1j}, U_{2j}),
\]

where

\[
B_n(i, j) = b_1(i, j) + b_2(i, j) - b_3(i, j) - b_4(i, j)
\]

and

\[
b_1(i, j) = 2K_h(X_j - X_i) \frac{1}{(\sigma^2(X_i)f(X_i))^2}, \quad b_2(i, j) = b_1(j, i),
\]

\[
b_3(i, j) = K_h * K_h(X_j - X_i) \frac{1}{(\sigma^2(X_i)f(X_i))^2}, \quad b_4(i, j) = b_3(j, i).
\]

Thus, we can write $W(n) = \sum_{i < j} W_{ij}$. The asymptotic normality of $W_n$ is obtained from Proposition 3.2 in de Jong (1987), also given here in the next section. If
1. $W(n)$ is clean,

2. $\text{Var}(W(n)) \to \nu^*$,

3. $G_I$ is of smaller order than $\text{Var}(W(n))^2$;

4. $G_{II}$ is of smaller order than $\text{Var}(W(n))^2$;

5. $G_{IV}$ is of smaller order than $\text{Var}(W(n))^2$,

where

$$
G_I = \sum_{1 \leq i < j \leq n} \mathbb{E}(W_{ij}^4),
$$

$$
G_{II} = \sum_{1 \leq i < j < k \leq n} \left\{ \mathbb{E}(W_{ij}^2W_{ik}^2) + \mathbb{E}(W_{ij}^2W_{jk}^2) + \mathbb{E}(W_{ki}^2W_{kj}^2) \right\},
$$

$$
G_{IV} = \sum_{1 \leq i < j < k < l \leq n} \left\{ \mathbb{E}(W_{ij}W_{ik}W_{jl}) + \mathbb{E}(W_{ij}W_{il}W_{jk}W_{kl}) + \mathbb{E}(W_{ik}W_{il}W_{jk}W_{jl}) \right\},
$$

then our results hold.

The first condition directly follows from the first Bartlett identity. For the variance of $W(n)$, note that

$$
\text{Var}(W(n)) = \sum_{i < j} E(W_{ij}^2).
$$

Thus, we calculate $E[\{B_n(i, j)\ell'(\theta(X_i), U_{1i}, U_{2i}) \ell'(\theta(X_j), U_{1j}, U_{2j})\}^2]$. To simplify our presentation, let $\ell'_i = \ell'(\theta(X_i), U_{1i}, U_{2i})$ and denote the $m$-fold convolution at $t$ by $K(t, m) = K * \cdots * K(t)$. Through direct calculations, we obtain

$$
E(b_i^2(i, j) \ell_i^2 \ell_j^2) = E \left[ \frac{4}{h^2} \frac{\ell_i^2 \ell_j^2}{\sigma^2(X_i)f(X_i)} K^2 \left( \frac{X_j - X_i}{h} \right) \right]
$$

$$
= \frac{4}{h^2} \int \frac{\sigma^2(X_1)}{\sigma^2(X_1)f(X_1)^2} \left\{ \int \sigma^2(X_2)K^2 \left( \frac{X_2 - X_1}{h} \right) f(X_2) dX_2 \right\} f(X_1) dX_1
$$

$$
= \frac{4}{h} \int \frac{f^{-2}(X_1)}{\sigma^2(X_1)} \int \sigma^2(X_1)K^2(t)df(X_1)dX_1 (1 + O(h))
$$

$$
= \frac{4}{h} K(0, 2)E f^{-1}(X)(1 + O(h))
$$
Similarly,

\[
E(b_2^2(i, j) \ell_i^2 \ell_j^2) = \frac{4}{h}K(0, 2)Ef^{-1}(X)(1 + O(h)),
\]

\[
E(b_2^2(i, j) \ell_i^2 \ell_j^2) = \frac{1}{h}K(0, 4)Ef^{-1}(X)(1 + O(h)),
\]

\[
E(b_3^2(i, j) \ell_i^2 \ell_j^2) = \frac{1}{h}K(0, 4)Ef^{-1}(X)(1 + O(h)),
\]

\[
E(b_1(i, j)b_2(i, j) \ell_i^2 \ell_j^2) = \frac{4}{h}K(0, 2)Ef^{-1}(X)(1 + O(h)),
\]

\[
E(b_1(i, j)b_3(i, j) \ell_i^2 \ell_j^2) = \frac{2}{h}K(0, 3)Ef^{-1}(X)(1 + O(h)),
\]

\[
E(b_1(i, j)b_4(i, j) \ell_i^2 \ell_j^2) = \frac{2}{h}K(0, 3)Ef^{-1}(X)(1 + O(h)),
\]

\[
E(b_2(i, j)b_3(i, j) \ell_i^2 \ell_j^2) = \frac{2}{h}K(0, 3)Ef^{-1}(X)(1 + O(h)),
\]

\[
E(b_2(i, j)b_4(i, j) \ell_i^2 \ell_j^2) = \frac{2}{h}K(0, 3)Ef^{-1}(X)(1 + O(h)),
\]

\[
E(b_3(i, j)b_4(i, j) \ell_i^2 \ell_j^2) = \frac{1}{h}K(0, 4)Ef^{-1}(X)(1 + O(h)).
\]

Thus,

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

The leading term of

\[
\frac{h}{n^2} \sum_{i < j} E[B_n(i, j)\ell_i^2 \ell_j^2]
\]

yields

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

For the third condition, note that

\[
E(b_1(1, 2)\ell_1 \ell_2)^4 = E(b_3(1, 2)\ell_1 \ell_2)^4 = O(h^{-3}).
\]

Hence, \(E(W_{12}^4) = \frac{h^2}{n^3}O(h^3)\), which implies \(G_I = O\left(\frac{1}{n^3h}\right) = o(1)\). Similarly, the fourth condition can be verified by noting that

\[
E(W_{12}^2W_{13}^2) = O(E(W_{12}^4)) = O\left(\frac{1}{n^4h}\right).
\]

Thus, \(G_{II} = O\left(\frac{1}{nh}\right) = o(1)\). For the last condition we need to check the order of
Chapter 4. Conditional Copula Selection & Parameter Inference

\[ E(W_{12}W_{23}W_{34}W_{41}). \] Calculations for few terms yield,

\[
\begin{align*}
E(b_1^2(1, 2)b_1^2(2, 3)b_1^2(3, 4)b_1^2(4, 1) & \ell_1^2 \ell_2^2 \ell_3^2 \ell_4^2) = O(h^{-1}) \\
E(b_1^2(1, 2)b_1^2(2, 3)b_1^2(3, 4)b_1^2(4, 1) & \ell_1^2 \ell_2^2 \ell_3^2 \ell_4^2) = O(h^{-1}) \\
E(b_1^2(1, 2)b_1^2(2, 3)b_1^2(3, 4)b_1^2(4, 1) & \ell_1^2 \ell_2^2 \ell_3^2 \ell_4^2) = O(h^{-1}) \\
E(b_1^2(1, 2)b_3^2(2, 3)b_3^2(3, 4)b_3^2(4, 1) & \ell_1^2 \ell_2^2 \ell_3^2 \ell_4^2) = O(h^{-1}) \\
E(b_3^2(1, 2)b_3^2(2, 3)b_3^2(3, 4)b_3^2(4, 1) & \ell_1^2 \ell_2^2 \ell_3^2 \ell_4^2) = O(h^{-1}). \\
\end{align*}
\]

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

\[ E(W_{12}W_{23}W_{34}W_{41}) = \frac{h^2}{n^4} O(h^{-1}) = O\left(\frac{h}{n^4}\right), \]

and \( G_{IV} = O(h) = o(1). \)

Hence, the first result directly follows, i.e.

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

where \( \nu_n = (4h)^{-1}\nu^* \). For the second result, note that the distribution \( N(a_n, 2a_n) \) is approximately same as the \( \chi^2 \) distribution with degrees of freedom \( a_n \), for a sequence \( a_n \to \infty \). Letting \( a_n = 2\mu_n/\nu_n \) and \( r_K = 2\mu_n/\nu_n \), we have

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

provided that \( nh^{9/2} \to 0 \).

\[ \square \]

4.2.4 Additional theoretical results

The following definition and proposition are from de Jong (1987). Note that, his notation is adapted to ours.

Let \( X_1, X_2, \ldots \) be independent variables, and \( w_{ijn}(\cdot, \cdot) \) Borel functions such that

\[ W(n) = \sum_{1 \leq i \leq n} \sum_{1 \leq j \leq n} w_{ijn}(X_i, X_j), \]
and

\[ W_{ij} = w_{ijn}(X_i, X_j) + w_{jin}(X_j, X_i). \]

The index \( n \) is suppressed in the notation \( W_{ij} \).

**Definition 7** (de Jong 1987, Definition 2.1). \( W_n \) is called clean if the conditional expectations of \( W_{ij} \) vanish:

\[ E[W_{ij}|X_i] = 0 \quad \text{a.s.} \quad \text{for all } i,j \leq n. \]

**Proposition 1** (de Jong 1987, Proposition 3.2). Let \( W(n) \) be clean with variance \( \nu_n^* \) and let \( G_I, G_{II} \) and \( G_{IV} \) be of lower order than \( \nu_n^{*2} \), then

\[ \nu_n^{*-1/2}W(n) \xrightarrow{L} N(0, 1), \quad n \to \infty, \]

where

\[
\begin{align*}
G_I &= \sum_{1 \leq i < j \leq n} E(W_{ij}^4), \\
G_{II} &= \sum_{1 \leq i < j < k \leq n} \{E(W_{ij}^2W_{ik}^2) + E(W_{ji}^2W_{jk}^2) + E(W_{ki}^2W_{kj}^2)\}, \\
G_{IV} &= \sum_{1 \leq i < j < k < l \leq n} \{E(W_{ij}W_{ik}W_{lj}W_{lk}) + E(W_{ij}W_{il}W_{kj}W_{lk}) + E(W_{ik}W_{il}W_{jk}W_{lj})\}.
\end{align*}
\]

**Proofs of lemmas**

The proofs of the lemmas used in our derivations are below.

**Proof of Lemma 1.** From (3.17), we have

\[ \gamma_n^{-1}\{\hat{\theta}(x) - \theta(x)\} = \frac{\gamma_n}{\sigma^2(x)f(x)} \sum_{i=1}^{n} \ell'((\theta(x, X_i), U_{1i}, U_{2i}))K\left(\frac{X_i - x}{h}\right) (1 + o_p(1)), \]

which directly yields

\[
\hat{\theta}(x) - \theta(x) = \left\{ \frac{\gamma_n^2}{\sigma^2(x)f(x)} \sum_{i=1}^{n} \{\ell'(\theta(x, X_i), U_{1i}, U_{2i}) - \ell'(\theta(X_i), U_{1i}, U_{2i})\}K\left(\frac{X_i - x}{h}\right) + \frac{\gamma_n^2}{\sigma^2(x)f(x)} \sum_{i=1}^{n} \ell'(\theta(X_i), U_{1i}, U_{2i}) K\left(\frac{X_i - x}{h}\right) \right\} (1 + o_p(1)).
\]
**Proof of Lemma 2.** Note that

\[ T_{n1} = \gamma_n^2 \sum_{k=1}^{n} \frac{1}{\sigma^2(X_k)f(X_k)} [\ell'(\theta(X_k), U_{1k}, U_{2k})]^2 K(0) \]

\[ + \gamma_n^2 \sum_{k \neq i} \frac{1}{\sigma^2(X_k)f(X_k)} \ell'(\theta(X_i), U_{1i}, U_{2i}) \ell'(\theta(X_k), U_{1k}, U_{2k}) K \left( \frac{X_i - X_k}{h} \right). \]

The approximation of the first term

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

yields the first result.

We can decompose \( T_{n2} \) into two parts: \( T_{n2} = T_{n21} + T_{n22} \), with

\[ T_{n21} = \frac{1}{(nh)^2} \sum_{i=1}^{n} [\ell'(\theta(X_i), U_{1i}, U_{2i})]^2 \sum_{k=1}^{n} \frac{\ell''(\theta(X_k), U_{1k}, U_{2k})}{(\sigma^2(X_k)f(X_k))^2} K^2 \left( \frac{X_i - X_k}{h} \right), \]

\[ T_{n22} = \frac{1}{n^2} \sum_{i \neq j} \ell'(\theta(X_i), U_{1i}, U_{2i}) \ell'(\theta(X_j), U_{1j}, U_{2j}) \left\{ \sum_{k=1}^{n} \frac{\ell''(\theta(X_k), U_{1k}, U_{2k})}{(\sigma^2(X_k)f(X_k))^2} K_h(X_i - X_k)K_h(X_j - X_k) \right\}. \]

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

\[ T_{n21} = \frac{1}{(nh)^2} \sum_{k=1}^{n} \ell'(\theta(X_k, U_{1k}, U_{2k}))^2 \ell''(\theta(X_k), U_{1k}, U_{2k}) \frac{\ell''(\theta(X_k), U_{1k}, U_{2k})}{(\sigma^2(X_k)f(X_k))^2} K^2(0) \]

\[ + \frac{1}{(nh)^2} \sum_{i \neq k} [\ell'(\theta(X_i), U_{1i}, U_{2i})]^2 \ell''(\theta(X_k), U_{1k}, U_{2k}) \frac{\ell''(\theta(X_k), U_{1k}, U_{2k})}{(\sigma^2(X_k)f(X_k))^2} K^2 \left( \frac{X_i - X_k}{h} \right). \]

The first sum can be shown to be

\[ \frac{1}{(nh)^2} \sum_{k=1}^{n} \sigma^2(X_k) \frac{\ell''(\theta(X_k), U_{1k}, U_{2k})}{(\sigma^2(X_k)f(X_k))^2} K^2(0) + o_p(h^{-1/2}) = O_p \left( \frac{1}{nh^2} \right) \]

Therefore, let

\[ V_n = \frac{2}{n(n-1)} \sum_{i<k} \{ \sigma^2(X_i) \frac{\ell''(\theta(X_i), U_{1i}, U_{2i})}{(\sigma^2(X_i)f(X_i))^2} + \sigma^2(X_k) \frac{\ell''(\theta(X_k), U_{1k}, U_{2k})}{(\sigma^2(X_k)f(X_k))^2} \} K_h^2(X_k - X_i), \]

and the second sum becomes

\[ \frac{1}{2} (V_n + o(1)) + O_p \left( \frac{1}{n^{3/2}h^2} \right) + o_p(h^{-1/2}). \]
The decomposition theorem for U-statistics (Hoeffding, 1948) allows us to show that the variance of $V_n$

$$\text{Var}(V_n) = O\left(\frac{1}{nh^2}\right).$$

Note that the leading term of $V_n$ is

$$-\frac{1}{h} \mathbb{E} f^{-1}(X) \int K^2(t) dt.$$

Hence, as $nh \to \infty$ and $h \to 0$, we obtain

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

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

$$T_{n221} = \frac{2}{n} \sum_{i<j} \ell'(\theta(X_i), U_{1i}, U_{2i}) \ell'(\theta(X_j), U_{1j}, U_{2j}) \frac{1}{n} \left\{ \sum_{k \neq i,j} \frac{\ell''(\theta(X_k), U_{1k}, U_{2k})}{(\sigma^2(X_k) f(X_k))^2} \right\} K_h(X_i - X_k) K_h(X_j - X_k),$$

$$T_{n222} = \frac{K(0)}{n^2 h} \sum_{i \neq j} \ell'(\theta(X_i), U_{1i}, U_{2i}) \ell'(\theta(X_j), U_{1j}, U_{2j})$$

$$\times \left\{ \frac{\ell''(\theta(X_i), U_{1i}, U_{2i})}{(\sigma^2(X_i) f(X_i))^2} + \frac{\ell''(\theta(X_j), U_{1j}, U_{2j})}{(\sigma^2(X_j) f(X_j))^2} \right\} K_h(X_i - X_j).$$

For $k \neq i, j$, define

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

It can be easily shown that

$$\text{Var}\left(\frac{1}{n} \sum_{k \neq i,j} Q_{ijk,h}\right) = O\left(\frac{1}{nh^2}\right).$$

Hence, $T_{n221}$ becomes

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

where

$$\mathbb{E}(Q_{ijk,h}|X_i, X_j) = -\frac{1}{h \sigma^2(X_i) f(X_i)} \int K(t) K \left(\frac{X_j - X_i}{h}\right) dt.$$
It is also easy shown

$$Var(T_{n222}) = O\left(\frac{1}{n^2 h^3}\right).$$

Hence,

$$T_{n222} = o_p(h^{-1/2}).$$

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

$$T_{n2} = -\frac{1}{h} \mathbb{E} f^{-1}(X) \int K^2(t)dt - \frac{2}{nh} \sum_{i<j} \frac{\ell'(\theta(X_i), U_{1i}, U_{2i})}{\sigma^2(X_i) f(X_i)} \ell'(\theta(X_j), U_{1j}, U_{2j})K \ast K((X_j - X_i)/h) + o_p(h^{-1/2}).$$

\[\square\]

### 4.3 Summary and Discussion

This chapter accomplished two important statistical tasks in conditional copulas: model selection and parameter inference.

The first section presented a novel conditional copula selection method, which naturally emerged from our bandwidth selection procedure. We compared the cross-validated predictive ability of candidate conditional copula models for choosing the copula family. Here our efforts were concentrated on conditional copula models with varying copula parameter. A prediction-based copula selection can easily be developed for unconditional copulas and it would be interesting to compare its performance with existing copula selection methods. We also would like to investigate different cross-validation strategies for copula selection.

In the second part, we focused on nonparametric inference of the conditional copula parameter, proposed a generalized likelihood ratio test and derived its asymptotic null distribution. Although, in practice, the bandwidth value chosen for estimation can be used in the generalized likelihood ratio test (Fan and Jiang, 2007), the optimal bandwidth in nonparametric hypothesis testing problems remains as an issue.
There is a wide literature on nonparametric inference based on Bickel-Rosenblatt approach (Bickel and Rosenblatt, 1973). Obtaining a Bickel-Rosenblatt type result for the local copula parameter estimator is a challenging problem, and requires finding the asymptotic distribution of the maximum or the squared normalized deviations of the local copula parameter estimate from its true value.

Another challenging question that arises in this chapter is that whether we can consider copula family selection and copula parameter inference together in a formal test. This brings a new concept smooth copula goodness-of-fit tests and will be an important direction in our future efforts.
Chapter 5

Simulation Results and Data Applications

This chapter contains the results of our simulation studies and data applications. First, we investigate the finite sample performances of the local copula parameter estimation, the copula selection method and the generalized likelihood ratio test in a set of simulation studies using the Clayton, Frank and Gumbel copulas. We then apply our proposed methodology to the two data examples introduced in Chapter 2.

5.1 Simulation Results

In our simulations, we consider the Clayton, Frank and Gumbel copulas. These choices cover a wide range of situations as the Clayton and Gumbel copulas are known to exhibit strong and weak lower tail dependence, respectively, while the Frank copula is symmetric and shows no tail dependence. Basic properties of these copulas are given in Chapter 1, and summarized in Table 5.1.

The inverse link functions are chosen as $g^{-1}(t) = \exp(t)$ for the Clayton copula, $g^{-1}(t) = t$ for the Frank copula and $g^{-1}(t) = \exp(t) + 1$ for the Gumbel copula, so that
Table 5.1: The copula families used in simulation studies

<table>
<thead>
<tr>
<th>Family</th>
<th>$C(u_1, u_2)$</th>
<th>$\theta \in$</th>
<th>$\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clayton</td>
<td>$(u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}}$</td>
<td>$(0, \infty)$</td>
<td>$\frac{\theta}{\theta + 2}$</td>
</tr>
<tr>
<td>Frank</td>
<td>$-\frac{1}{\theta} \ln \left{ 1 + \frac{(e^{-\theta u_1 - 1})(e^{-\theta u_2 - 1})}{e^{-\theta} - 1} \right}$</td>
<td>$(-\infty, \infty) \setminus {0}$</td>
<td>$1 + \frac{4}{\theta}[D_1(\theta) - 1]$</td>
</tr>
<tr>
<td>Gumbel</td>
<td>$\exp\left{ - \left[ (-\ln u)^\theta (-\ln v)^\theta \right]^{\frac{1}{\theta}} \right}$</td>
<td>$[1, \infty)$</td>
<td>$1 - \frac{1}{\theta}$</td>
</tr>
</tbody>
</table>

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

the resulting copula parameter estimates under these families will be in the correct range.

Under each copula family, we consider both a linear and a nonlinear calibration model. Our steps and objectives in the six experiments conducted can be summarized as follows.

i. To estimate the calibration function we perform the local linear estimation, as well as parametric estimation in which $\eta$ is assumed to be linear in $X$.

ii. We compare our proposed approach with the parametric estimation when the underlying calibration function is correctly specified and when it is misspecified.

iii. By performing the estimation also under the other two families, we investigate the impact of the copula (mis)specification.

iv. We evaluate the performance of our copula selection method in identifying the underlying family.

v. We test whether the calibration function is (a) constant, (b) linear and we determine the size of the generalized likelihood ratio test.

These steps are detailed below.
5.1.1 Clayton Copula

We generate the data \( \{ (U_{1i}, U_{2i}) : i = 1, 2, \ldots, n \} \) from the Clayton copula under each of the following models:

1. Linear calibration function: \( \eta(X) = 0.8X - 2 \)

\[
(U_1, U_2) \mid X \sim C(u_1, u_2 \mid \theta = \exp(0.8X - 2)) \quad \text{where } X \sim \text{Uniform}(2, 5),
\]

2. Nonlinear calibration function: \( \eta(X) = 2 - 0.3(X - 4)^2 \)

\[
(U_1, U_2) \mid X \sim C(u_1, u_2 \mid \theta = \exp(2 - 0.3(X - 4)^2)) \quad \text{where } X \sim \text{Uniform}(2, 5).
\]

We first generate the covariate values \( X_i \) from \( \text{Uniform}(2, 5) \). Then, for each \( i = 1, 2, \ldots, n \), we obtain the copula parameter, \( \theta_i = \exp(\eta(x_i)) \), imposed by the given calibration and link functions, and simulate the pairs \( (U_{1i}, U_{2i}) \mid X_i \) from the Clayton copula with the parameter \( \theta_i \). The true copula parameter varies from 0.67 to 7.39 in the linear calibration model, and from 2.22 to 7.39 in the nonlinear one. Under each model, we conduct experiments with sample sizes \( n = 200 \) and \( n = 500 \), each replicated \( m = 100 \) times.

We estimate the copula parameter under the Clayton, Frank and Gumbel families, using the local linear and parametric linear estimation, both with \( p = 1 \). For the bandwidth parameter, we consider 12 candidate values, ranging from 0.33 to 2.96, equally spaced on a logarithmic scale. The impact of the bandwidth parameter on the local calibration function estimates, under the Clayton copula \( (n = 200) \), is illustrated in Figure 5.1 and Figure 5.2 using typical examples from the linear and nonlinear calibration models, respectively. We use the leave-one-out cross-validated likelihood criterion (3.8) to select the optimum bandwidth. All results reported below are based on the local estimates at the chosen optimum bandwidth.

Since the accuracy of the calibration estimation is not directly comparable across different copula families, we convert the copula parameter estimates to a common scale
Figure 5.1: An illustration of the local calibration function estimates under the linear calibration model of the Clayton copula ($n = 200$): truth (solid line), local linear estimates (dashed line).

provided by the Kendall’s tau measure of association (1.3). For each of the three families, the connection between $\theta$ and Kendall’s tau is reported in Table 5.1.

Table 5.2 displays the Monte Carlo estimates of the integrated Mean Square Error (IMSE) along with its integrated square Bias (IBIAS$^2$) and integrated Variance (IVAR)
Figure 5.2: An illustration of the local calibration function estimates under the non-linear calibration model of the Clayton copula ($n = 200$): truth (solid line), local linear estimates (dashed line).

From Table 5.2, we see that the parametric estimation performs better under the linear calibration model when it uses the correct functional form. However, when there is no known parametric model for the calibration function, the proposed nonparametric
Table 5.2: Clayton Copula. Integrated Squared Bias, Integrated Variance and Integrated Mean Square Error (multiplied by 100) of the Kendall’s tau estimator. The last column shows the averages of the bandwidths $h^*$ selected using (3.8).

<table>
<thead>
<tr>
<th></th>
<th>Parametric estimation</th>
<th>Local estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>IBIAS$^2$</td>
</tr>
<tr>
<td><strong>Linear Calibration Model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clayton</td>
<td>200</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.018</td>
</tr>
<tr>
<td>Frank</td>
<td>200</td>
<td>0.490</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.479</td>
</tr>
<tr>
<td>Gumbel</td>
<td>200</td>
<td>3.739</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>3.499</td>
</tr>
<tr>
<td><strong>Nonlinear Calibration Model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clayton</td>
<td>200</td>
<td>0.414</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.423</td>
</tr>
<tr>
<td>Frank</td>
<td>200</td>
<td>0.324</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.357</td>
</tr>
<tr>
<td>Gumbel</td>
<td>200</td>
<td>4.914</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>4.862</td>
</tr>
</tbody>
</table>

Approach better captures the covariate effect on the copula parameter. The results also show that the performance of the estimation deteriorates as the properties of the used copula family significantly depart from the true one. Since the Frank copula, having
no tail dependence, is closer to the Clayton copula than is the Gumbel copula, it yields better results in all simulation scenarios when compared to Gumbel.

We also construct the approximate 90% pointwise confidence bands for the copula parameter under the correctly selected Clayton family, where the bias is much smaller than the variance as noticed in Table 5.2. We thus use half of the optimum bandwidth to assess $\sigma^2(x)$ in (3.14) while further reducing the bias to a negligible level (Fan and Zhang, 2000). To be consistent, for each Monte Carlo sample, the confidence intervals obtained for the copula parameter are converted to the Kendall’s tau scale. Figure 5.3 and Figure 5.4 display the confidence bands for the Kendall’s tau, averaged over 100 Monte Carlo samples ($n = 200$), under the linear and nonlinear calibration models, respectively. For comparison, we present the Monte Carlo based confidence bands obtained from 100 estimates of Kendall’s tau, which agree well with our proposal.

For the copula selection, we calculate the cross-validated prediction errors using the cross-validated parameter estimates at the optimum bandwidth. Our approach unifies the bandwidth selection and the copula selection in determining the model that best fits the data. We evaluate the performance of our copula selection method by counting the number of times the Clayton copula is selected. As indicated in Table 5.3, we successfully identify the true copula family, the Clayton copula, more than 90% of the time. Consistent with the results of Table 5.2, the Gumbel copula shows the worst performance in predictions and was not selected by any of the experiments.

Table 5.3: Times of selection of the three families out of 100 Monte Carlo samples.

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Clayton</th>
<th>Frank</th>
<th>Gumbel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear Calibration</td>
<td>200</td>
<td>91</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>99</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Nonlinear Calibration</td>
<td>200</td>
<td>97</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 5.3: Clayton Copula. 90% confidence intervals for the Kendall’s tau under the linear calibration model of the Clayton copula: truth (solid line), averaged local linear estimates (dashed line), approximate confidence intervals (dotted line), and Monte Carlo confidence intervals (dotdashed line).
Figure 5.4: Clayton Copula. 90% confidence intervals for the Kendall’s tau under the nonlinear calibration model of the Clayton copula: truth (solid line), averaged local linear estimates (dashed line), approximate confidence intervals (dotted line), and Monte Carlo confidence intervals (dotdashed line).
5.1.2 Frank Copula

This section presents the results for the data \(\{(U_{1i}, U_{2i} \mid X_i) : i = 1, 2, \ldots, n\}\) generated from the Frank copula under each of the following models:

1. Linear calibration function: \(\eta(X) = 25 - 4.2X\)

\[(U_1, U_2) \mid X \sim C(u_1, u_2 \mid \theta = 25 - 4.2X)\] where \(X \sim \text{Uniform}(2, 5),\)

2. Nonlinear calibration function: \(\eta(X) = 12 + 8\sin(0.4X^2)\)

\[(U_1, U_2) \mid X \sim C(u_1, u_2 \mid \theta = 12 + 8\sin(0.4X^2))\] where \(X \sim \text{Uniform}(2, 5).\)

The true copula parameter varies from 4 to 16.6 in the linear calibration model, and from 4 to 20 in the nonlinear one. Under each calibration model, we conduct \(m = 100\) experiments with sample size \(n = 200.\)

The local linear and parametric linear estimation, both with \(p = 1,\) are performed to estimate the copula parameter under the Clayton, Frank and Gumbel copulas, and the results are converted to the Kendall’s tau scale. Table 5.4 summarizes the performance of the Kendall’s tau estimates in terms of integrated mean square error, integrated square bias and integrated variance.

We construct approximate 90% pointwise confidence bands for the copula parameter under the correctly selected Frank family, using half of the optimum bandwidth. Figure 5.5 and Figure 5.6 displays the results in the Kendall’s tau scale, averaged over 100 Monte Carlo samples. We also present the Monte Carlo based confidence bands obtained from 100 estimates of the Kendall’s tau.

Table 5.5 reports the number of times each copula is selected by our copula selection method. Our copula selection method successfully identifies the Frank family.
Table 5.4: Frank Copula. Integrated Squared Bias, Integrated Variance and Integrated Mean Square Error of the Kendall’s tau estimator (multiplied by 100) and the averages of the selected bandwidths $h^*$.

<table>
<thead>
<tr>
<th></th>
<th>Linear Calibration Model</th>
<th>Nonlinear Calibration Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Parametric estimation</td>
<td>Local estimation</td>
</tr>
<tr>
<td></td>
<td>IBIAS$^2$, IVAR, IMSE</td>
<td>IBIAS$^2$, IVAR, IMSE, $h^*$</td>
</tr>
<tr>
<td>Clayton</td>
<td>8.611, 1.125, 9.736</td>
<td>8.334, 1.514, 9.848, 2.056</td>
</tr>
<tr>
<td></td>
<td>0.006, 0.316, 0.322</td>
<td>0.006, 0.444, 0.450, 2.206</td>
</tr>
<tr>
<td>Gumbel</td>
<td>1.887, 0.478, 2.365</td>
<td>1.793, 0.746, 2.539, 2.133</td>
</tr>
<tr>
<td></td>
<td>18.192, 3.470, 21.662</td>
<td>10.593, 2.833, 13.426, 0.822</td>
</tr>
<tr>
<td></td>
<td>6.162, 0.673, 6.835</td>
<td>0.127, 1.280, 1.407, 0.468</td>
</tr>
<tr>
<td></td>
<td>9.597, 1.183, 10.780</td>
<td>3.504, 2.067, 5.571, 0.569</td>
</tr>
</tbody>
</table>

Table 5.5: Times of selection of the three families out of 100 Monte Carlo samples.

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Clayton</th>
<th>Frank</th>
<th>Gumbel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear Calibration</td>
<td>200</td>
<td>0</td>
<td>95</td>
<td>5</td>
</tr>
<tr>
<td>Nonlinear Calibration</td>
<td>200</td>
<td>0</td>
<td>97</td>
<td>3</td>
</tr>
</tbody>
</table>
Figure 5.5: Frank Copula. 90% confidence intervals for the Kendall’s tau under the linear calibration model of the Frank copula: truth (solid line), averaged local linear estimates (dashed line), approximate confidence intervals (dotted line), and Monte Carlo confidence intervals (dotdashed line).
Figure 5.6: Frank Copula. 90% confidence intervals for the Kendall’s tau under the nonlinear calibration model of the Frank copula: truth (solid line), averaged local linear estimates (dashed line), approximate confidence intervals (dotted line), and Monte Carlo confidence intervals (dotdashed line).
5.1.3 Gumbel Copula

The data \{(U_{1i}, U_{2i} | X_i) : i = 1, 2, \ldots, n\} is generated from the Gumbel copula under each of the following models:

1. Linear calibration function: \( \eta(X) = -2.1 + 0.75 X \)

\( (U_1, U_2) | X \sim C(u_1, u_2 | \theta = \exp(-2.1 + 0.75 X) + 1) \) \quad \text{where} \quad X \sim \text{Uniform}(2, 5),

2. Nonlinear calibration function: \( \eta(X) = -1 + 2.5 \cos(X - 1)^2 \)

\( (U_1, U_2) | X \sim C(u_1, u_2 | \theta = \exp(-1 + 2.5 \cos(X - 1)^2) + 1) \) \quad \text{where} \quad X \sim \text{Uniform}(2, 5).\)

The true copula parameter varies from 1.55 to 6.21 in the linear calibration model, and from 1.37 to 5.46 in the nonlinear one. Under each calibration model, we conduct \( m = 100 \) experiments with sample size \( n = 200 \).

The local linear and parametric linear estimation, both with \( p = 1 \), are performed to estimate the copula parameter under the Clayton, Frank and Gumbel copulas, and the results are converted to the Kendall’s tau scale. Table 5.6 summarizes the performance of the Kendall’s tau estimates in terms of integrated mean square error, integrated square bias and integrated variance.

We construct approximate 90\% pointwise confidence bands for the copula parameter under the correctly selected Gumbel family, using half of the optimum bandwidth. Figure 5.7 and 5.8 display the results in the Kendall’s tau scale averaged over 100 Monte Carlo samples. We also present the Monte Carlo based confidence bands obtained from 100 estimates of the Kendall’s tau.

Table 5.7 reports the number of times each copula is selected by our copula selection method. The Gumbel family is selected more than 90\% of the time.
Table 5.6: Gumbel Copula. Integrated Squared Bias, Integrated Variance and Integrated Mean Square Error of the Kendall’s tau estimator (multiplied by 100) and the averages of the selected bandwidths $h^*$.  

<table>
<thead>
<tr>
<th></th>
<th>Linear Calibration Model</th>
<th>Nonlinear Calibration Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Parametric estimation</td>
<td>Local estimation</td>
</tr>
<tr>
<td></td>
<td>IBIAS$^2$</td>
<td>IVAR</td>
</tr>
<tr>
<td>Clayton</td>
<td>7.683</td>
<td>2.158</td>
</tr>
<tr>
<td>Frank</td>
<td>0.339</td>
<td>0.473</td>
</tr>
<tr>
<td>Gumbel</td>
<td>0</td>
<td>0.275</td>
</tr>
</tbody>
</table>

Table 5.7: Times of selection of the three families out of 100 Monte Carlo samples.  

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Clayton</th>
<th>Frank</th>
<th>Gumbel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear Calibration</td>
<td>200</td>
<td>0</td>
<td>6</td>
<td>94</td>
</tr>
<tr>
<td>Nonlinear Calibration</td>
<td>200</td>
<td>0</td>
<td>4</td>
<td>96</td>
</tr>
</tbody>
</table>
Figure 5.7: Gumbel Copula. 90% confidence intervals for the Kendall’s tau under the linear calibration model of the Gumbel copula: truth (solid line), averaged local linear estimates (dashed line), approximate confidence intervals (dotted line), and Monte Carlo confidence intervals (dotdashed line).
Figure 5.8: Gumbel Copula. 90% confidence intervals for the Kendall’s tau under the nonlinear calibration model of the Gumbel copula: truth (solid line), averaged local linear estimates (dashed line), approximate confidence intervals (dotted line), and Monte Carlo confidence intervals (dotdashed line).
5.1.4 Generalized Likelihood Ratio Test

We evaluate the performance of the generalized likelihood ratio test using the above simulated datasets. The tests are performed at the optimum bandwidth value selected by the leave-one-out cross-validated likelihood criterion (3.8). We obtained the asymptotic null distribution of the generalized likelihood ratio test using the tabulated values $r_K = 2.1153$, $c_K = 0.45$ and $\mu_n = r_K c_K |\mathcal{X}|/h^*$ for the Epanechnikov kernel. In our experiments, the range of the covariate $|\mathcal{X}| = 3$.

The hypotheses we test are whether the calibration function is (a) a constant, (b) a linear function of the covariate. We label these hypotheses the constant null hypothesis and the linear null hypothesis, respectively. We calculate the log-likelihood under each null hypothesis using the corresponding maximum likelihood estimates. To calculate the log-likelihood under the alternative, we perform the local linear estimation at each covariate value in the data, using the optimum bandwidth. We then obtain the generalized likelihood ratio statistic for each Monte Carlo sample by taking the difference between the alternative and null log-likelihoods. We apply the test at different significance levels using the asymptotic null distribution of the test statistic at the optimum bandwidth.

We first evaluate the performance of the generalized likelihood ratio test for the linear null hypothesis under the linear calibration model. Since the optimum bandwidth values are different for each Monte Carlo sample, we only consider the empirical significance levels. Table 5.8 displays these results. The results are consistent with the actual significance levels of the tests. Since we have 100 Monte Carlo samples, the results obtained under the Frank family should be reasonable. The exact binomial test is used to test the significance of the size of the generalized likelihood ratio tests. We display the $p$-values in Table 5.9. Except for a few values, $p$-values validate the use of the generalized likelihood ratio tests at the optimum bandwidth.

We also check the times of rejection of the null hypotheses under the alternative models, and present the results in Tables 5.10, 5.11, and 5.12. Overall rejection rates are
high. One interesting result is that under the nonlinear calibration model of the Clayton copula, times of rejection of the linear null hypothesis is lower, as is the variability of the strength of dependence (see Figure 5.4).

Table 5.8: Linear Calibration Model. GLRT results: Empirical significance levels under the linear null hypothesis

<table>
<thead>
<tr>
<th>Level of significance</th>
<th>0.25</th>
<th>0.20</th>
<th>0.15</th>
<th>0.10</th>
<th>0.05</th>
<th>0.025</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clayton 200</td>
<td>0.28</td>
<td>0.25</td>
<td>0.18</td>
<td>0.13</td>
<td>0.07</td>
<td>0.05</td>
<td>0.01</td>
</tr>
<tr>
<td>Frank 200</td>
<td>0.19</td>
<td>0.12</td>
<td>0.09</td>
<td>0.05</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Gumbel 200</td>
<td>0.25</td>
<td>0.22</td>
<td>0.16</td>
<td>0.12</td>
<td>0.09</td>
<td>0.06</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Table 5.9: Linear Calibration Model. Exact Binomial Tests for the size of the GLRT

<table>
<thead>
<tr>
<th>True probability</th>
<th>0.25</th>
<th>0.20</th>
<th>0.15</th>
<th>0.10</th>
<th>0.05</th>
<th>0.025</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clayton 200</td>
<td>0.489</td>
<td>0.212</td>
<td>0.400</td>
<td>0.315</td>
<td>0.352</td>
<td>0.106</td>
<td>1</td>
</tr>
<tr>
<td>Frank 200</td>
<td>0.203</td>
<td>0.045</td>
<td>0.121</td>
<td>0.130</td>
<td>0.01</td>
<td>0.186</td>
<td>0.630</td>
</tr>
<tr>
<td>Gumbel 200</td>
<td>1</td>
<td>0.617</td>
<td>0.779</td>
<td>0.503</td>
<td>0.100</td>
<td>0.039</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 5.10: Linear Calibration Model. GLRT results: Times of rejection of the constant null hypothesis out of 100 Monte Carlo samples

<table>
<thead>
<tr>
<th>Level of significance</th>
<th>0.25</th>
<th>0.20</th>
<th>0.15</th>
<th>0.10</th>
<th>0.05</th>
<th>0.025</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clayton</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Frank</td>
<td>200</td>
<td>100</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>97</td>
<td>95</td>
</tr>
<tr>
<td>Gumbel</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 5.11: Nonlinear Calibration Model. GLRT results: Times of rejection of the constant null hypothesis out of 100 Monte Carlo samples

<table>
<thead>
<tr>
<th>Level of significance</th>
<th>0.25</th>
<th>0.20</th>
<th>0.15</th>
<th>0.10</th>
<th>0.05</th>
<th>0.025</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clayton</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>98</td>
</tr>
<tr>
<td>Frank</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td>Gumbel</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 5.12: Nonlinear Calibration Model. GLRT results: Times of rejection of the linear null hypothesis out of 100 Monte Carlo samples

<table>
<thead>
<tr>
<th>Level of significance</th>
<th>0.25</th>
<th>0.20</th>
<th>0.15</th>
<th>0.10</th>
<th>0.05</th>
<th>0.025</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clayton</td>
<td>200</td>
<td>78</td>
<td>76</td>
<td>74</td>
<td>65</td>
<td>50</td>
<td>43</td>
</tr>
<tr>
<td>Frank</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>99</td>
<td>99</td>
<td>98</td>
</tr>
<tr>
<td>Gumbel</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
5.2 Data Examples

This section demonstrates the techniques proposed in this thesis using the two data examples introduced in Chapter 2.

5.2.1 Twin Birth Data

The first step is to fit the conditional marginal models both parametrically and nonparametrically. The results under each case are presented below.

Parametrically Estimated Conditional Marginals

The scatterplot and histograms of the birth weights, BW$_1$ and BW$_2$, for $n = 450$ twin pairs, are given in Figure 5.9(a), from which the marginals are seen to fit well by parametric cubic regression models with normal noise, shown in Figure 5.9(c) and 5.9(d), where the gestational age is denoted GA. Table 5.13 summarizes the (orthogonal) cubic regression models for each response variable.

<table>
<thead>
<tr>
<th></th>
<th>BW$_1$</th>
<th></th>
<th>BW$_2$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>2245.63</td>
<td>20.75</td>
<td>2214.12</td>
<td>20.62</td>
</tr>
<tr>
<td>poly(GA, degree = 3)$_1$</td>
<td>11612.92</td>
<td>440.16</td>
<td>10863.50</td>
<td>437.52</td>
</tr>
<tr>
<td>poly(GA, degree = 3)$_2$</td>
<td>-2955.23</td>
<td>440.16</td>
<td>-2508.71</td>
<td>437.52</td>
</tr>
<tr>
<td>poly(GA, degree = 3)$_3$</td>
<td>-1343.66</td>
<td>440.16</td>
<td>-1782.66</td>
<td>437.52</td>
</tr>
</tbody>
</table>

The resulting coefficients, all significant at the 1% level, substituted in

\[ U_{ij} = \Phi[(BW_{ij} - \hat{\mu}_{ij}(GA_j))/\hat{\sigma}_i], \quad j = 1, \ldots, 450, \quad i = 1, 2. \]  (5.1)
to transform the response variables to uniform scale, where $\hat{\mu}_{ij}$, depending on $\text{GA}_j$, are the fitted values of the cubic regression models, $\sigma_i^2$ the estimated residual errors, and $\Phi$ is the c.d.f. of $N(0,1)$. Figure 5.9(b) gives the scatterplot and histograms of the transformed random variables $U_j$, $j = 1, 2$, which support the parametric fit of marginal models considered here. We also check whether the transformed variables $U_1$ and $U_2$ are uniformly distributed via the Kolmogorov-Smirnov test and obtain $p$-values 0.424 and 0.226, respectively.

We then estimate the calibration function using the proposed nonparametric model with local linear fit ($p = 1$), under the Clayton, Frank and Gumbel families. For comparison, we also perform the parametric estimation with both a constant and a linear form. The results, converted to the Kendall’s tau scale, are given in Figure 5.10. In all cases, our approach yields a nonlinear pattern, with the highest strength of dependence in pre-term and post-term twins which cannot be detected by the parametric linear fit.

In the local linear estimation, the optimum bandwidths are chosen as 5.52, 4.04, and 5.52, for the Clayton, Frank and Gumbel copulas, respectively. We constructed approximate 90% confidence intervals under each family using the method introduced in Section 3.6. As seen in Figure 5.10, the parametric estimates are not within the confidence bands, suggesting that the linear model may not be adequate. Our copula selection method chooses the Frank family, having the minimum cross-validated prediction error (4.1) with value 47.05, followed by the Clayton copula with 47.36 and the Gumbel copula with 47.40.

Under the chosen Frank copula, we perform the generalized likelihood ratio test to check whether the gestational age has a significant effect on the strength of dependence. Using the optimum bandwidth, we obtain the degrees of freedom of the chi-square distribution as 2.87. The test statistic 29.90 yields a $p$-value less than $1.3 \times 10^{-6}$. Thus, we conclude that the variation in the strength of dependence between the twin birth weights at different gestational ages is highly significant. We also test the adequacy of a
Figure 5.9: Twin birth data. Histograms and scatterplots of (a) the twin birth weights, (b) the conditional marginal distributions. Scatterplots with fitted models: (c) gestational age and BW$_1$, (d) gestational age and BW$_2$. 
Figure 5.10: Twin birth data. The Kendall’s tau estimates under three copula families: global constant estimate (dashed line), global linear estimates (dotdash line), local linear estimates at the optimum bandwidth (longdash line), and an approximate 90% confidence interval (dotted lines).
parametric linear model for the copula parameter. The test statistic is obtained as 26.85, which yields a $p$-value less than $6 \times 10^{-6}$. Thus, we conclude that using a parametric linear form in the copula parameter will not adequately characterize the variation in the strength of dependence.

**Nonparametrically Estimated Conditional Marginals**

We use kernel based conditional distribution estimation (Hall et al., 1999; Li and Racine, 2008) to estimate the conditional marginals and then transform the responses to the uniform scale, as shown in Figure 5.11. We check whether the transformed variables $U_1$ and $U_2$ are uniformly distributed using the Kolmogorov-Smirnov test and obtain $p$-values 0.400 and 0.218, respectively.

![Figure 5.11: Twin birth data. Histograms and scatterplot of the nonparametric estimates of the conditional marginal distributions.](image)

In local estimation, the optimum bandwidths are chosen as 6.44, 4.72, and 8.79, for the Clayton, Frank and Gumbel copulas, respectively. We construct approximate 90%
confidence intervals under each family using the method introduced in Section 3.6. As seen in Figure 5.12, we obtain similar results when we use nonparametrically estimated conditional marginals.

**Bootstrap Confidence Intervals**

The confidence bands presented above are obtained using our asymptotic results (3.15). Here, we demonstrate the use of bootstrapping the raw data to construct confidence bands for Kendall’s tau.

For each bootstrap sample we use (a) parametric conditional distribution estimation, (b) kernel based conditional distribution estimation to estimate the marginals and then transform the responses to the uniform scale. Subsequently, we obtain the local linear estimate of the calibration function under the selected Frank copula model and construct the 90% quantile-based bootstrap interval based on 200 resamples. For comparison, we also obtain the asymptotic 90% confidence intervals using (3.15).

The results are shown in Figure 5.13. Consistent with our discussion in Section 3.7, the bootstrap bands are close to the asymptotic bands when conditional marginals are estimated parametrically, and wider when conditional marginals are estimated nonparametrically.

### 5.2.2 Framingham Data

The second data we analyze is the Framingham data.

The scatterplot and histograms of the log-pulse pressures are given in Figure 5.14(a), from which the marginals are seen to be well-modeled by linear regression models assuming normal errors. We consider $U_{ij} = \Phi((\log(PP_{ij})-\hat{\mu}_j(\Delta BMI_j))/\hat{\sigma}_i), \; j = 1, \ldots, 348, \; i = 1, 2,$ to transform the response variables to uniform scale, where $\hat{\mu}_j(\Delta BMI)$ are obtained from fitting linear models, $\hat{\sigma}_j$ the estimated standard errors in the linear regression models and $\Phi$ the c.d.f. of $N(0,1)$. Figure 5.14(b) gives the scatterplot and histograms of
Figure 5.12: Twin birth data. The Kendall’s tau estimates under three copula families: global constant estimate (dashed line), global linear estimates (dotdash line), local linear estimates at the optimum bandwidth (longdash line), and an approximate 90% confidence interval (dotted lines).
Figure 5.13: Twin birth data. The Kendall’s tau estimates under the Frank copula using (a) parametric conditional distribution estimation, (b) kernel based conditional distribution estimation: local linear estimates at the optimum bandwidth (longdashed line), an approximate 90% asymptotic confidence interval (dotted lines), a 90% bootstrap confidence interval (dotdashed line) based on 200 resamples.
the transformed random variables $U_j, j = 1, 2$. We check whether the transformed variables $U_1$ and $U_2$ are uniformly distributed via the Kolmogorov-Smirnov test and obtain $p$-values 0.928 and 0.947, respectively.

For comparison, the calibration function is estimated using both a parametric linear model and the proposed nonparametric model with local linear $(p = 1)$ specification under the Clayton, Frank and Gumbel families. In the local linear estimation, the optimum bandwidths are chosen as 9.47, 6.72, and 4.25, for the Clayton, Frank and Gumbel copulas, respectively. We constructed approximate 90% confidence intervals under each family using the method introduced in Section 3.6. Figure 5.15 displays the results converted to the Kendall’s tau scale.

A natural interest in modeling the conditional dependence of log-pulse pressures at different periods is to predict the latter from the previous measurement. Therefore, in conditional copula selection, we use the cross-validated prediction errors only for the second log-pulse pressure as the selection criterion which chooses the Frank copula family. Since, at a given change in BMI, observing two high pulse pressures together or two low pulse pressures together would be equally likely, our choice of Frank copula model seems practically sensible. The results obtained using the nonparametric approach under the Frank copula indicates that the dependence between two log-pulse pressures is stronger when BMI remains stable, as reflected by a steady health status. However, the nonlinear pattern does not seem to be significant as the parametric linear fit in fact falls within the approximate 90% pointwise confidence bands, suggesting that the linear model may already be adequate.

The generalized likelihood ratio test results suggest that both constant and linear models are adequate, with $p$-values 0.172 and 0.220, respectively. We therefore perform the likelihood ratio test to compare the constant model with the linear one, and obtain $p$-value 0.824. Thus we conclude that the change in body mass index in fact does not have any significant effect on the dependence between the two pulse pressures.
Figure 5.14: Framingham data. Histograms and scatterplots of (a) the log-pulse pressures at period 1 and period 2; (b) the marginal distributions of the transformed variables $(U_1, U_2)$. 
Figure 5.15: Framingham data. The Kendall’s tau estimates under three copula families: global constant estimate (dashed line), global linear estimates (dotdashed line), local linear estimates at the optimum bandwidth (longdashed line), and an approximate 90% pointwise confidence bands (dotted lines).
Variable Bandwidth

We also perform the estimation results using nearest-neighbour (variable) bandwidth (see Section 3.4), since the covariate $\Delta$ BMI does not have a uniform design. We obtain similar results as shown in Figure 5.16.

![Figure 5.16: Framingham data. The local linear estimates of the Kendall’s tau under three copula families using constant bandwidth (longdashed line) and variable bandwidth (dotdashed lines).](image)

Figure 5.16: Framingham data. The local linear estimates of the Kendall’s tau under three copula families using constant bandwidth (longdashed line) and variable bandwidth (dotdashed lines).
Chapter 6

Conclusion

We propose a conditional copula approach to model the strength and type of dependence between two responses. In order to capture the relationship between the strength of the dependence and a measured covariate, we allow the copula parameter to change according to a calibration function of the covariate. Statistical inference for the calibration function is obtained using local polynomial estimation.

The problem of modeling dependence between two or more random variables or processes when the strength of dependence varies according to the values of a measured covariate is discussed in the first chapter of this thesis. This chapter also provides some background information on copulas, conditional copulas and local polynomial approach.

The present problem has broad applications since one may observe (a) time-variation in the strength of dependence in multivariate data collected over time (b) a subject-specific covariate effect on the strength of dependence in multivariate response from several individuals (c) a pair-specific covariate effect on the strength of dependence in studies involving twins or paired units. Chapter 2 illustrates two such applications: twin birth weight analysis and an analysis of a subset of the Framingham Heart Study data.

In Chapter 3, we present the local copula parameter estimation, study the asymptotic properties of the estimator and describe how to construct pointwise confidence intervals
using these asymptotic results. A bootstrap procedure for confidence intervals is also outlined to accord the impact of estimating conditional marginal distributions on the inference of the copula parameter. The *leave-one-out cross-validation* likelihood criterion is used for choosing the bandwidth parameter.

The methodology proposed in Chapter 3 leads to a novel conditional copula selection procedure, in which we use prediction accuracy to select, via cross validation, among a number of copula families, the one that best approximates the data at hand. This conditional copula selection method is presented in the first part of Chapter 4.

A natural question that arises within the scope of this thesis is whether the copula parameter indeed significantly varies or may be adequately characterized by a certain predefined form. The second part of Chapter 4 answers this question via a generalized likelihood ratio test. We derive the asymptotic null distribution of the test and show that a Wilks-type result holds in conditional copulas.

Chapter 5 contains the results of our simulation studies and data applications. Our simulation studies convey that i) the nonparametric estimator of the calibration function is flexible enough to capture non-linear patterns, ii) the generalized likelihood ratio test and the copula selection procedure performs well in the cases studied herein. The analysis of twin birth data reveals a gestational age specific dependence pattern in the birth weights of twins that, to our knowledge, has not been detected before and may be of scientific interest. In the analysis of the Framingham data, we conclude that the change in the body mass index has no significant impact on the strength of dependence between the pulse pressures of the first two examination periods.

Although in this thesis we focus our attention on bivariate copulas, it is possible to extend our method to more general multivariate copulas. In addition, if more covariates are of potential interest for the conditional copula model, then we recommend a careful selection of the variables prior to estimation of the calibration function, as the estimation is troubled by the curse of dimensionality and computational cost increases significantly.
with each covariate added to the model.

Some future interests in our research are summarized below.

**Additive Models and Variable Selection** Additive models (Tibshirani and Hastie, 1990) can be adapted to conditional copulas to allow for more than one covariate in the calibration function. For instance, in the case of two covariates $X_1$ and $X_2$, we can formulate $\theta(X_1, X_2) = g^{-1}\{\eta_1(X_1) + \eta_2(X_2)\}$ and use a local copula-likelihood approach to estimate $\eta_1$ and $\eta_2$. Also, variable selection methods can be explored within additive calibration models via generalized likelihood ratio tests.

**Conditional Copula Selection** Conditional copula selection is a challenging problem. We want to further explore this direction in connection with cross-validated model selection strategies used in regression problems (Yang, 2007). This may allow us to examine the theoretical properties our conditional copula selection method. We are also interested in possible connections between goodness-of-fit tests for copulas and smooth goodness-of-fit tests in the nonparametric literature to develop smooth copula goodness-of-fit tests.

**Composite Likelihood** The difficulty in specifying the likelihood function of high dimensional data due to complex dependencies recently drew attention to composite likelihood methods. The objective of the composite likelihood approach is to combine lower dimensional likelihood objects (marginal, conditional or pairwise likelihoods) as alternatives to the usual joint likelihood. An important issue here is whether there exist higher dimensional multivariate distributions compatible with the lower dimensional (e.g.: bivariate) margins. We would like to investigate how one could exploit bivariate conditional copulas in composite likelihood and what improvements they would bring compared to pairwise or conditional composite likelihood methods. Further development includes adapting a local framework to composite likelihood for deriving nonparametric composite conditional copulas.
Competing and Semi-competing Risks  Copulas are widely used in competing and semi-competing risk problems. Although significant effort has been made to incorporate covariate effects on the marginal and joint survival functions, to best of my knowledge, conditional copulas have not been yet explored for these problems. We believe our approach would provide valuable insight into competing and semi-competing risk problems and deserves further investigation.

Mixture of Copulas  We are currently working on extending the framework of the work presented here by considering mixtures of conditional copulas in order to increase the spectrum of applications that can be tackled using our approach.
Bibliography


— (1990), Generalized Additive Models, Chapman & Hall.

