GENERALIZED ADDITIVE MODELING FOR MULTIVARIATE DISTRIBUTIONS

Vatter Thibault

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES

DÉPARTEMENT OPÉRATIONS

**GENERALIZED ADDITIVE MODELING
FOR MULTIVARIATE DISTRIBUTIONS**

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Etudes Commerciales
de l'Université de Lausanne

pour l'obtention du grade de
Docteur en Systèmes d'Information

par

Thibault Vatter

Directeur de thèse
Prof. Valérie Chavez-Demoulin

Jury

Prof. Hameri Ari-Pekka, Président
Prof. Eric Jondeau, expert interne
Prof. Christian Genest, expert externe
Prof. Claudia Czado, expert externe

LAUSANNE
2016

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES

DÉPARTEMENT OPÉRATIONS

**GENERALIZED ADDITIVE MODELING
FOR MULTIVARIATE DISTRIBUTIONS**

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Etudes Commerciales
de l'Université de Lausanne

pour l'obtention du grade de
Docteur en Systèmes d'Information

par

Thibault Vatter

Directeur de thèse
Prof. Valérie Chavez-Demoulin

Jury

Prof. Hameri Ari-Pekka, Président
Prof. Eric Jondeau, expert interne
Prof. Christian Genest, expert externe
Prof. Claudia Czado, expert externe

LAUSANNE
2016

IMPRIMATUR

Sans se prononcer sur les opinions de l'auteur, la Faculté des Hautes Études Commerciales de l'Université de Lausanne autorise l'impression de la thèse de Monsieur Thibault VATTER, Ingénieur diplômé EPFL, titulaire d'un B.Sc. en Physique et d'un M.Sc. en Ingénierie Physique avec mineure en Ingénierie Financière de l'Ecole Polytechnique Fédérale de Lausanne, en vue de l'obtention du grade de docteur en Systèmes d'Information.

La thèse est intitulée :

GENERALIZED ADDITIVE MODELING

FOR MULTIVARIATE DISTRIBUTIONS

Lausanne, le 10 mars 2016

Le doyen

Jean-Philippe Bonardi

Accredited by
Association
of MBAs

EFMD
EQUIS
ACCREDITED

THESIS COMMITTEE:

Valérie Chavez-Demoulin, Professor of Statistics, University of Lausanne

Eric Jondeau, Professor of Finance, University of Lausanne

Christian Genest, Professor of Statistics, McGill University

Claudia Czado, Professor of Statistics, Technische Universität München

University of Lausanne
Faculty of Business and Economics

Doctorate in Information Systems

I hereby certify that I have examined the doctoral thesis of

Thibault VATTER

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: _____ Date: ___13.04.2016_____

Prof. Valérie CHAVEZ-DEMOULIN
Thesis supervisor
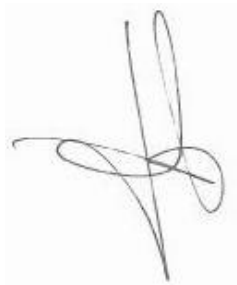
University of Lausanne

Faculty of Business and Economics

Doctorate in Information Systems

I hereby certify that I have examined the doctoral thesis of

Thibault VATTER

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: Date: 13.04.2016

Prof. Eric JONDEAU
Internal member of the doctoral committee

University of Lausanne
Faculty of Business and Economics

Doctorate in Information Systems

I hereby certify that I have examined the doctoral thesis of

Thibault VATTER

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: Date: 2016-03-14

Prof. Christian GENEST

University of Lausanne
Faculty of Business and Economics

Doctorate in Information Systems

I hereby certify that I have examined the doctoral thesis of

Thibault VATTER

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: _Claudia Czado_ Date: _15.3.2016_

Prof. Claudia CZADO
External member of the doctoral committee

I dedicate this thesis to the two most important women in my life: my mother, Chantal Vatter, and my fiancée, Fanny Pélissier. Their unconditional love and support have been an integral part of the completion of this PhD, and I am eternally grateful for everything they have done for me.

Acknowledgements

First and foremost, I wish to thank my advisor, Valérie Chavez-Demoulin, for her mentoring, trust and friendship. Without her, I would never have been able to (or maybe even wanted to) discover the wonderful field of statistics. I also would like to acknowledge my committee members for their time and expertise. In particular, I am grateful to Claudia Czado for her inputs related to dependence modeling during my two stays in the Department of Mathematics at the Technische Universität München.

I also wish to thank in no particular order:

- Damien Ackerer, a research colleague since our time at Swissquote and presently a PhD candidate in the Swiss Finance Institute at EPFL, for his friendship and numerous discussions in topics ranging from finance to nordic skiing and from statistics to belgian beers.

- Mikhail Zhelonkin, my office mate during most of my time in Lausanne, for his advices in statistics and being able to stand me in the office everyday,

- Thomas Nagler, a PhD candidate under Claudia Czado, for his conversations on copulas and collaboration on a research project that represent the third chapter of the present thesis.

- Hau-Tieng Wu, an assistant professor in the Mathematics department at the University of Toronto, for his patience in teaching me the basics of time-frequency analysis and lively discussions concerning our joint research project.

- Bin Yu, a professor in the Statistics Department of UC Berkeley, for welcoming me as a visiting scholar when I had just started my PhD, and her research group, as well as members from the Coleman Fung Risk Management Research Center, for many enriching debates.

- Fanny Pélissier, my fiancée and a PhD candidate in Biology at the University of Bergen, for putting up with me in the good as well as in the bad times, and her collaboration on interesting biostatistics problems.

- My family and friends from Geneva, Satellite, and many others, for their encouragements throughout this process.

Abstract

In this thesis, we develop tools to study the influence of predictors on multivariate distributions. We tackle the issue of conditional dependence modeling using generalized additive models, a natural extension of linear and generalized linear models allowing for smooth functions of the covariates. Compared to existing methods, the framework that we develop has two main advantages. First, it is completely flexible, in the sense that the dependence structure can vary with an arbitrary set of covariates in a parametric, nonparametric or semiparametric way. Second, it is both quick and numerically stable, which means that it is suitable for exploratory data analysis and stepwise model building. Starting from the bivariate case, we extend our framework to pair-copula constructions, and open new possibilities for further applied and methodological work. Our regression-like theory of the dependence, being built on conditional copulas and generalized additive models, is at the same time theoretically sound and practically useful.

Declaration

This thesis is a presentation of my original research work. Wherever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature. To avoid referencing the same work too often, note that most of the material from the second chapter of this thesis has been published by the author in Vatter and Chavez-Demoulin (2015). Similarly, the material from the third chapter has circulated as an unpublished manuscript in Vatter and Nagler (2015). To a lesser extent, some ideas have been influenced by two other manuscripts: Vatter et al. (2015) (published) and Ackerer and Vatter (2015) (unpublished).

<div align="right">

Thibault Vatter

March 2016

</div>

Table of contents

List of figures

List of tables

Chapter 1

Introduction

Because real-world statistical questions are seldom answered with univariate datasets, modeling the joint behavior of random variables has novel and promising applications, in line with topical questions and concerns from numerous fields: from social sciences like psychology, economics, or finance, to natural sciences like climatology, biology or robotics. Whether the goal is to describe the joint behavior of stocks in a portfolio to quantify and manage its risk, or of numerous geological and environmental characteristics to design a bridge or a damn, multivariate statistical tools are required.

Consider the following research questions:

- Cellular biologists study how predictor genes coordinate the expression of target genes. Can we further quantify how the association between the targets depends on the predictors?

- There are relationships between a population's life expectancy and the country's GDP, as well as between male and female life expectancy in a given country. Can we measure the effect of the GDP on the later while controlling for the former?

- The left/right and progressivism/conservatism political alignment of citizens in a given country are often related. Can we produce a geographical map of the strength of this effect, or further explain it using other demographic factors?

- Volatilities of intraday asset returns show periodicities due to the cyclical nature of market activity and macroeconomic news releases. Is this also true for their dependence structure?

To obtain a statistically sound answer, we need framework to model the joint distribution of random variables conditionally on exogenous predictors (covariates).

As an introducing example, consider the conditional correlation between the random variables Y_1 and Y_2 given X, which is defined by

$$corr\left(Y_1, Y_2 \mid X\right) = \frac{E\left[\{Y_1 - E\left(Y_1 \mid X\right)\}\{Y_2 - E\left(Y_2 \mid X\right)\} \mid X\right]}{\left(E\left[\{Y_1 - E\left(Y_1 \mid X\right)\}^2 \mid X\right] E\left[\{Y_2 - E\left(Y_2 \mid X\right)\}^2 \mid X\right]\right)^{1/2}}$$

For instance, when $Y_1 = A + BX$ and $Y_2 = C + DX$ where A, B, C, D are independently distributed random variables, then

$$corr\left(Y_1, Y_2 \mid X\right) = \frac{cov\left(A, C\right) + \{cov\left(A, D\right) + cov\left(B, C\right)\}X + cov\left(B, D\right)X^2}{\left[\begin{array}{c}\{var\left(A\right) + 2\,cov\left(A, B\right)X + var\left(B\right)X^2\} \\ \cdot \{var\left(C\right) + 2\,cov\left(C, D\right)X + var\left(D\right)X^2\}\end{array}\right]^{1/2}}.$$

Furthermore, if A, B, C, D have equal variance and correlation ρ, we have that

$$corr\left(Y_1, Y_2 \mid X\right) = \frac{\rho\left(1 + X\right)^2}{\left(1 + 2\,\rho X + X^2\right)}.$$

In this example, it is clear that the conditional correlation depends on the value of the conditioning variable explicitly. Hence, it is in general not equal to the so-called partial correlation. However, while the statistical literature abounds with models allowing the description of covariates effects on the behavior of a single random variable, letting the dependence structure itself be a function of exogenous predictors has only been recently explored. As often in statistics, it is useful to distinguish between distribution-free and parametric methods. In Gijbels et al. (2011), the authors suggest two kernel-based estimators of conditional copulas and corresponding conditional association measures. While useful as descriptive statistics, their estimators are limited to a single predictor, and they do not suggest a way of testing for covariates effects[1]. On the parametric side, Acar et al. (2011) consider a copula parameter that varies with a single covariate. The authors estimate their model using the concept of local likelihood, and they further suggest a testing framework in Acar et al. (2013). In Craiu and Sabeti (2012), the authors develop Bayesian inference tools for a bivariate copula, conditional on a single covariate, coupling mixed or continuous outcomes. It is extended to multiple covariates in the continuous case by Sabeti et al. (2014).

---

[1] Altough, at the time of writing, a paper of the same authors on the subject is in preparation.

In this thesis, we tackle the issue of conditional dependence modeling using generalized additive models (Hastie and Tibshirani 1986), a natural extension of linear and generalized linear models. Built on roughness penalty smoothing, a generalized additive model (GAM) is a flexible data analysis tool in a traditionally univariate context. Compared to existing methods, the framework that we develop benefits directly from the complete GAM toolbox, which has two main advantages. First, it is completely flexible, meaning that the dependence structure varies with an arbitrary set of covariates in a parametric, nonparametric or semiparametric way. Second, it is very fast and numerically stable, which means that it is suitable for exploratory data analysis and stepwise model building.

1.1 From Sklar to Patton

Before introducing copulas, we start by discussing how we can characterize the joint behavior of random variables. In this context, it is useful to distinguish between two closely related concepts, namely dependence and concordance. Loosely speaking, a dependence measure relates to any functional characteristic expressing how close the joint distribution is to the product of the margins. As for the concordance, it measures the degree of agreement between positive and negative comovements. For instance, consider two random variables X_1 and X_2. Provided that it exists, Pearson's correlation, namely

$$corr\left(X_1, X_2\right) = \frac{cov\left(X_1, X_2\right)}{\sqrt{var\left(X_1\right)}\sqrt{var\left(X_2\right)}},$$

is often used as a measure of concordance and its absolute value as a measure of dependence. It detects linear relationships between variables, but it has three drawbacks: it lacks robustness to outliers, it does not always exists (see Embrechts et al. 2002), and it depends on the marginal distribution of each random variable. While the first issue can be alleviated using robust methods, the second and third are arguably more fundamental. In essence, the last two drawbacks teach us that the moments of a distribution are not always appropriate to describe the joint behavior of the underlying random variables. For instance, assume that X_1 and X_2 both follow a Gaussian distribution, which means that $Y_1 = \exp\left(X_1\right)$ and $Y_2 = \exp\left(X_2\right)$ each have a log-normal distribution. Then if X_1 and X_2 have zero mean, the same variance σ^2,
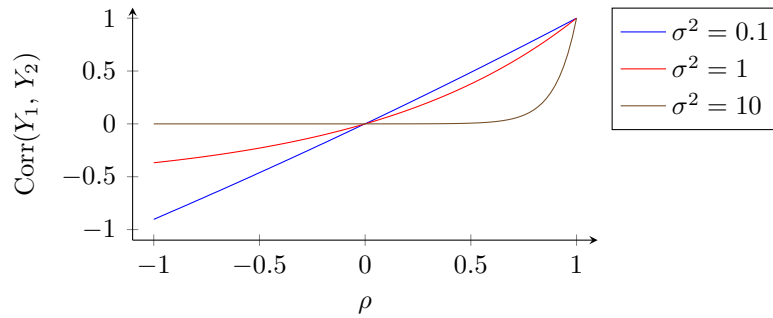
Fig. 1.1 **Correlation between log-normal random variables:** a function of the correlation between the underlying Gaussian random variables.

and $corr\,(X_1, X_2) = \rho$, we have that

$$corr\,(Y_1, Y_2) = \frac{\exp\left(\rho\sigma^2\right) - 1}{\exp\left(\sigma^2\right) - 1},$$

which is shown in Figure 1.1. Another issue related to the dependence on the margins is the range of attainable correlation. Let Y_1 and Y_2 be comonotonic (respectively countermonotonic) if

$$(Y_1, Y_2) \stackrel{d}{=} (T_1(X), T_2(X)),$$

where X is another random variable, T_1 and T_2 are increasing (respectively decreasing) transformations and $\stackrel{d}{=}$ the equality in distribution. Considering two log-normal random variables, comonotonicity obtains for instance when $Y_1 = \exp\left(X\right)$ and $Y_2 = \exp\left(\sigma X\right)$ with $X \sim N(0, 1)$, which means that the maximal attainable correlation is

$$\rho_{\max} = corr\,(Y_1, Y_2) = \frac{\exp(\sigma) - 1}{\sqrt{(\mathrm{e} - 1)\left\{\exp(\sigma^2) - 1\right\}}}.$$

Equivalently, countermonotonicity obtains when $Y_2 = \exp\left(-\sigma X\right)$, which means that the minimal attainable correlations is

$$\rho_{\min} = corr\,(Y_1, Y_2) = \frac{\exp(-\sigma) - 1}{\sqrt{(\mathrm{e} - 1)(\exp(\sigma^2) - 1)}}.$$

Both are shown in Figure 1.2, where one can observe that the complete $[-1, 1]$ interval cannot be attained.
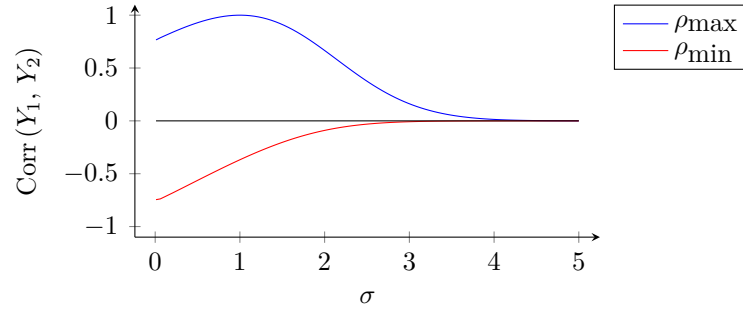
Fig. 1.2 **Correlation between comonotonic ($\rho_{\max}$) and countermonotonic ($\rho_{\min}$) log-normal random variables:** attainable correlation range, which does not span the complete $[-1, 1]$ interval.

Borrowing from Nelsen (1999), two desirable properties of a dependence (respectively concordance) measure are:

- invariance to monotone increasing transformations of the margins (respectively up to a sign change if one of the transformations is monotone decreasing);

- the existence and uniqueness of a minimum (respectively a zero), which is attained whenever the variables are independent.

For two random variables X_1 and X_2, Spearman's rho and Kendall's tau, denoted by $\rho(X_1, X_2)$ and $\tau(X_1, X_2)$ are rank correlation coefficients which satisfy the properties above for a concordance measure (or a dependence measure for their absolute value):

- $\rho(X_1, X_2)$ is simply the Pearson's correlation between the transformed variables $F_1(X_1)$ and $F_2(X_2)$, where F_1 and F_2 are the marginal distributions,

- and $\tau(X_1, X_2)$ is the difference between probability of concordance and discordance, namely

$$P\left\{\left(X_1 - \widetilde{X_1}\right)\left(X_2 - \widetilde{X_2}\right) > 0\right\} - P\left\{\left(X_1 - \widetilde{X_1}\right)\left(X_2 - \widetilde{X_2}\right) < 0\right\},$$

where $\widetilde{X_1}$ and $\widetilde{X_2}$ are independent copies following the same distribution than (X_1, X_2).

For instance, when $X_1 \sim \text{Exp}(1)$ and $X_2 = -\log(X_1)$, countermonotonicity obtains and $\rho(X_1, X_2) = \tau(X_1, X_2) = -1$, but $corr(X_1, X_2) = -\sqrt{6}/\pi$.

Note that copulas and dependence measures are closely related (see Joe 1997; Nelsen 1999 for textbook treatments). In fact, while the etymology of "copula" is bond or tie in latin, its main statistical raison d'etre is to glue together random variables

with arbitrary margins in a joint model. Mathematically, a copula is a multivariate distribution which is supported on the unit hypercube and has uniform margins. As for the main theoretical justification of copula-based modeling, it lies in the following theorem:

Theorem 1.1.1 (Sklar's Theorem 1959). *Let $X = (X_1, \ldots, X_d)$ be a random vector such that the marginal distribution of X_i is F_i. F is a joint distribution with margins F_i for $i \in \{1, \cdots, d\}$ if and only if there exists a copula C such that*

$$F(x_1, \ldots, x_d) = C\{F_1(x_1), \ldots, F_d(x_d)\} \tag{1.1}$$

for all $x \in \mathbb{R}^d$. Moreover, if the margins of F are continuous, then C is unique.

Note that, for all $u \in [0,1]^d$, the theorem implies that

$$\begin{aligned}
C(u_1, \ldots, u_d) &= F\left\{F_1^{-1}(u_1), \ldots, Y_d \le F_d^{-1}(u_d)\right\} \\
&= P\left\{X_1 \le F_1^{-1}(u_1), \ldots, X_d \le F_d^{-1}(u_N)\right\} \\
&= P\left\{F_1(X_1) \le u_d, \ldots, F_d(X_d) \le u_d\right\} \\
&= P\left(U_1 \le u_1, \ldots, U_d \le u_d\right),
\end{aligned}$$

where $U = (U_1, \cdots, U_d)$ is a random vector such that $U_i = F_i(X_i)$. In other words, the copula C is also the joint distribution of probability integral transforms. Furthermore, the first equality suggests a way of constructing the so-called implicit copulas from known multivariate distributions. For instance, if we consider $\Phi(\cdot; \Sigma)$ the multivariate Gaussian distribution with correlation matrix Σ and Φ the standard Normal distribution, then we obtain the Gaussian copula by writing

$$C(u_1, \ldots, u_d) = \Phi\left\{\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d); \Sigma\right\}.$$

Furthermore, with $t(\cdot; \Sigma, \nu)$ the multivariate Student's t distribution with correlation matrix Σ and degrees of freedom ν, and $t(\cdot; \nu)$ the Student's t distribution with degrees of freedom ν, then we obtain the Student's t copula by writing

$$C(u_1, \ldots, u_d) = t\left\{t^{-1}(u_1; \nu), \ldots, t^{-1}(u_d; \nu); \Sigma, \nu\right\}.$$

Alternatively, one can build the so-called explicit copulas by considering distribution supported in the unit hypercube directly. For instance, Archimedean copulas are built by considering a continuous and nonincreasing d-monotone (see McNeil and Nešlehová

Name	Generator ϕ	Inverse generator ϕ^{-1}	Parameter space Θ
Clayton	$(u^{-\theta}-1)/\theta$	$(1+\theta u)^{-1/\theta}$	$(0,\infty)$
Gumbel	$\{-\log(u)\}^{\theta}$	$\exp\left(-u^{1/\theta}\right)$	$[1,\infty)$
Frank	$-\log\left\{\frac{\exp(-\theta u)-1}{\exp(-\theta)-1}\right\}$	$-\frac{1}{\theta}\log\left[1+\exp\left(-t\right)\left\{\exp\left(-\theta\right)-1\right\}\right]$	$(-\infty,\infty)\setminus\{0\}$
Joe	$-\log\left\{1-(1-u)^{\theta}\right\}$	$1-\left\{1-\exp(-u)\right\}^{1/\theta}$	$[1,\infty)$
Independence	$-\log(u)$	$\exp(-u)$	

Table 1.1 **Archimedean copulas:** generator, inverse generator and parameter space.

2009) generator $\phi:[0,1]\times\Theta\to[0,\infty)$ such that $\phi(1;\theta)=0$ for all $\theta\in\Theta$, where Θ represents the parameter space. Using this generator, the copula is obtained by writing

$$C(u_1,\ldots,u_d)=\phi^{-1}\left\{\phi\left(u_1;\theta\right)+\cdots+\left(u_d;\theta\right);\theta\right\},$$

and we summarize the most popular in Table 1.1.

Note that the first of the two desirable properties of a dependence measure, namely invariance to monotone increasing transformations of the margins, essentially states that such a measure should depend on the copula only. For instance, if C is the copula of the two random variables X_1 and X_2, then their Kendall's tau is

$$\tau\left(X_1,X_2\right)=4\int_0^1\int_0^1 C\left(u_1,u_2\right)\,\mathrm{d}C\left(u_1,u_2\right)-1,$$

and their Spearman's rho can be written as

$$\begin{aligned}\rho\left(X_1,X_2\right)&=12\int_0^1\int_0^1 u_1\,u_2\,\mathrm{d}C\left(u_1,u_2\right)-3\\&=12\int_0^1\int_0^1 C\left(u_1,u_2\right)\,\mathrm{d}u_1\,\mathrm{d}u_2-3.\end{aligned}$$

Furthermore, in many cases, there is a simple mapping between the copula parameter and the dependence measure. For instance, if C is the Clayton copula, then we have that $\tau=\theta/(\theta+2)$, similarly, for the Gumbel copula, then $\tau=1-1/\theta$. As for the Gaussian copula (and all elliptical copulas such as the Student's t) with correlation parameter θ, then we have that $\tau=2\sin^{-1}\left(\theta\right)/\pi$.

While Sklar's theorem suggests a way of building copulas from known multivariate distributions, the opposite implication is arguably more useful in practice. Taking partial derivatives on both sides of (1.1) (provided that they exist), we have that

$$f(x_1,\ldots,x_d)=c\left\{F_1(x_1),\ldots,F_d(x_d)\right\}\prod_{i=1}^d f_i(x_i),\tag{1.2}$$
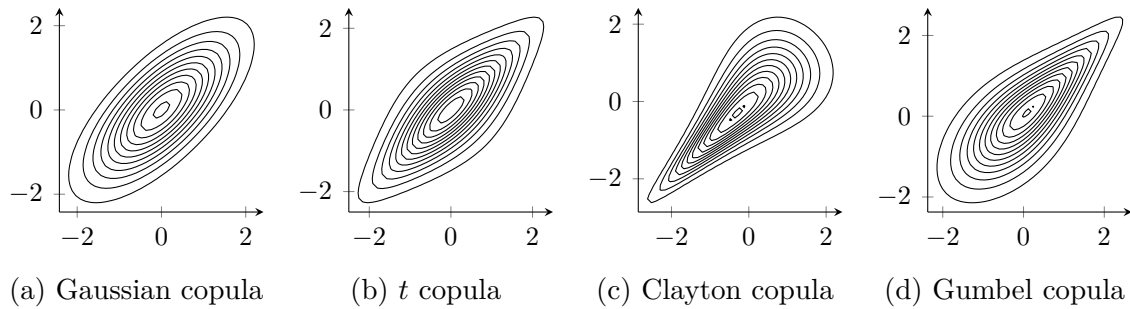
Fig. 1.3 **Gaussian margins:** density contours with different copulas.

(a) Gaussian copula (b) t copula (c) Clayton copula (d) Gumbel copula

(a) Gaussian copula (b) t copula (c) Clayton copula (d) Gumbel copula
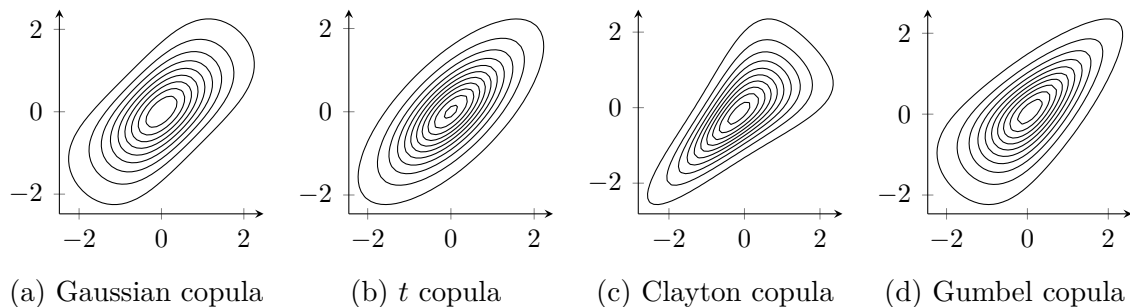
Fig. 1.4 **Student's t margins:** density contours with different copulas.

where f_i are the marginal densities and $c = \partial^d C/\partial u_1 \cdots \partial u_d$ is the so-called copula density. For instance, Figure 1.3 shows bivariate densities obtained using standard Normal margins and different copulas with a parameter chosen such that $\tau = 0.5$ (and four degrees of freedom for the Student's t). Alternatively, using Student's t margins with four degrees of freedom, we can build bivariate distributions with different margins but the same dependence structure, as we show in Figure 1.4.

Equation (1.2) also has important implications for inference. Because the right-hand side is a product, the joint log-likelihood can be written as a sum between the log-likelihood of each margin and the log-likelihood of the copula. This fact can be conveniently exploited in a two-step procedure

1. Estimate each of the margins separately to obtain $\widehat{F}_i$ for $i \in \{1, \ldots, d\}$.

2. Take the probability integral transform of the data using those margins to estimate the copula.

When parametric models are available for the margins, the asymptotics of such an approach are easily obtained. When this is not the case, Genest et al. (1995) investigate an alternative two-step approach, where the margins are nonparametrically estimated in the first step. The authors show that this semiparametric procedure is consistent

and asymptotically normal. While conditions for its semiparametric efficiency are given in Genest and Werker (2002), authors also argue that, except for the Gaussian copula, these requirements are seldom met. In Chen et al. (2006), a sieve maximum-likelihood estimator is shown to be semiparametrically efficient for both the copula parameter and unknown margins. The authors further argue that prior restrictions on the margins are easily incorporated, studying notably when some but not all are parametric.

Although the conditional correlation is useful as an introducing example, it suffers from the same issues discussed in the unconditional case. As such, to appropriately describe conditional dependence structures, measures that are invariant with respect to the monotonic transformations of the conditional margins are desirable. However, while copulas have been studied for more than fifty years, their formal extensions to conditional distributions have younger origins. To the best of our knowledge, it is Patton (2002) who first extended the standard theory by imposing a mutual conditioning algebra for each margin and the copula. Using the concept of conditional copula, namely a conditional distribution which is supported in the unit hypercube and has uniform conditional margins, Patton (2002)'s main theoretical contribution lies in the following theorem:

Theorem 1.1.2 (Patton's Theorem 2002)**.** *Let $Y = (Y_1, \ldots, Y_d)$ and $X = (X_1, \ldots, X_q)$ be two random vectors such that the conditional marginal distribution of $Y_i \mid X$ is $F_{Y_i|X}$ and F_X is the distribution of X. For all $x \in \mathbb{R}^q$, F is a joint conditional distribution with conditional margins $F_{Y_i|X}$ for $i \in \{1, \cdots, d\}$ if and only if there exists a conditional copula C such that*

$$F(y_1, \ldots, y_d \mid x) = C\left\{F_{Y_1|X}(y_1 \mid x), \ldots, F_{Y_d|X}(y_d \mid x) \mid x\right\} \qquad (1.3)$$

for all $y \in \mathbb{R}^d$. Moreover, if the conditional margins and F are continuous, then $C(\cdot \mid x)$ is unique.

Similarly as in the conditional case, C is the joint distribution of $U = (U_1, \cdots, U_d)$, where U is a random vector such that $U_i = F_{Y_i|X}(Y_i \mid X)$, namely the conditional probability integral transforms. Taking partial derivatives with respect to the conditioned vector on both sides of (1.3) and provided that they exist, we also have that

$$f(y_1, \ldots, y_d \mid x) = c\left\{F_{Y_1|X}(y_1 \mid x), \ldots, F_{Y_d|X}(y_d \mid x)\right\} \prod_{i=1}^{d} f_{Y_1|X}(y_1 \mid x), \qquad (1.4)$$

where $f_{Y_1|X}$ are the conditional margins and c the conditional copula density. As with (1.2), the product in the right-hand side of (1.4) means that a two-step procedure can be conveniently exploited.

1.2 The Epithelial to Mesenchymal Transition

Based on this theory, we can answer the first research question from the beginning of this chapter, in the context of breast cancer research. In this Section, we study how predictor genes coordinate the expression of target genes involved in the epithelial to mesenchymal transition (EMT).

Epithelial cells are essential components of every organ, including the mouth, lungs, stomach, liver, pancreas, prostate, mammary ducts, etc. They are organized in layers and are interconnected through cohesive interactions and to the surrounding environment through adhesive interactions. While mesenchymal cells are mobile and can migrate easily, normal epithelial cells require their surrounding neighborhood to stay alive and quiescent. In this context, EMT is a process by which epithelial cells lose their adhesive and cohesive properties and gain mesenchymal ones. When a tumor cell results from a normal epithelial tissue and undergoes EMT, it gains invasive properties by acquiring mesenchymal traits: it can enter the bloodstream, colonize other tissues and form secondary tumors and metastases.

To better understand EMT, we model the interplay of two transcription factors: SNAI1 and ZEB2, which primarily down-regulate the expression of cell-cohesion proteins and up-regulate migration proteins (see Lamouille et al. 2014 for a recent review). In this context, proteins such as HGF, PDGF-R, FOXC2 and FOXM1 (the predictors) activate complex signaling pathways in cells, leading to the expression of SNAI1 and ZEB2 (the targets). For a better understanding, we summarize the process in Figure 1.5.
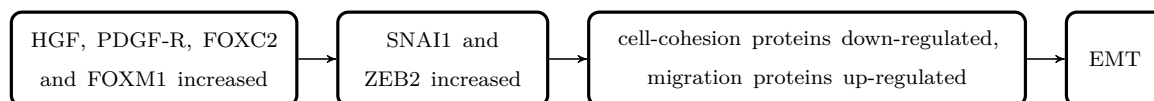
Fig. 1.5 **The epithelial to mesenchymal transition:** a schematic representation of the biological process.

Using Patton's theorem, we can quantify how the dependence structure between the targets varies as a function of the predictors. We perform the analysis in two steps. First, we fit a linear model on each individual target using all predictors. Second, we model the dependence structure between the univariate residuals by including the

effects of the predictors. To this end, we use a full RNA expression (standardized continuous variables) dataset consisting of 516 primary tumors (first studied in Network and Others 2012[2]).

In Table 1.2, we present the results of two univariate (linear) models, using HGF, PDGF-R, FOXC2 and FOXM1 as predictors for both SNAI1 and ZEB2. The effects of the predictors are all positive and highly significant, except FOXM1 on ZEB2.

$N = 522$	SNAI1			ZEB2		
	Estimate	Std. Error	p-value	Estimate	Std. Error	p-value
(Intercept)	9.59E-02	3.59E-02	7.85E-03	5.04E-02	2.61E-02	5.39E-02
HGF	1.36E-01	5.93E-02	2.21E-02	3.90E-01	4.30E-02	2.59E-18
PDGFR	3.64E-01	7.46E-02	1.42E-06	4.02E-01	5.41E-02	4.90E-13
FOXC2	2.54E-01	6.17E-02	4.57E-05	1.50E-01	4.48E-02	8.49E-04
FOXM1	2.68E-01	3.82E-02	7.07E-12	3.85E-02	2.77E-02	1.66E-01

Table 1.2 **EMT univariate results:** linear models with all predictors.

In order to describe the dependence structure, note that the correlation between the predictors equals 0.40 ($p < 10^{-16}$). Note that the p-value is computed using the test statistic $\sqrt{df}\,\widehat{r}/\sqrt{1 - \widehat{r^2}}$ which follows a Student's t-distribution with df degrees of freedom, where N is the sample of size, $\widehat{r}$ the usual estimator, and $df = N-2$. While the residual correlation decreases to 0.37 ($p < 10^{-16}$), it is still highly significant. However, the informativeness difference between the partial correlation and our approach is similar to the difference between a correlation coefficient and a complete regression analysis. Hence we fit different parametric copulas and we find that the Gaussian copula is the most appropriate, according to the AIC.

We then let Kendall's tau depend on a vector of predictors, say x, using

$$\tau(x) = g\left(x^\top \beta\right),$$

where β is a vector of parameters and $g(x) = (e^x - 1)/(e^x + 1)$ is a link to contrain $\tau(x)$ in $[-1, 1]$. In Table 1.3, we show the log-likelihood and AIC for five different models. Using likelihood-ratio statistics, we find that only HGF and PDGF-R have a significant effect ($p < 0.01$). Furthermore, when including both significant predictors, the AIC is the smallest. However, testing HGF+PDGF-R against HGF only does not provide significant evidence in favor of a more complex model ($p = 0.079$). Although this can be explained by a relatively high correlation between the two predictors, multicollinearity does not reduce the reliability of the model as a whole.

---

[2] tcga-data.nci.nih.gov/docs/publications/brca__2012: The dataset originally contains 522 primary tumors. Although it seldom happens, breast cancer also affects males and we remove those in the sample to avoid a potential confounding factor.

	HGF	PDGFRB	FOXC2	FOXM1	HGF+PDGFRB
logL	37.81	37.55	41.12	41.18	42.66
AIC	-71.59	-71.08	-78.22	-78.35	-79.28

Table 1.3 **EMT bivariate results:** log-likelihood and AIC values for different models for the Kendall's tau between SNAI1 and ZEB2.

In Figure 1.6, we show the results for the univariate models along with the dependence structure, in terms of linear correlation (which makes sense because we consider a linear model and a Gaussian copula). In each rightmost panel, we present a null model with constant correlation, the dependence as a function of the predictor and the 95% confidence intervals. As expected, confidence intervals become larger for values farther away from the mean predictor's expression (i.e., zero). Furthermore, the null model with constant correlation is always in the interval for non-significant predictors (FOXC2 and FOXM1).

The fact that complex signaling networks coordinate the expression and function of SNAI1 and ZEB2 and promote their interplay is not new (Peinado et al. 2007). However, our analysis quantifies the correlation increase between their expression, depending on the expression of HGF or PDGF-R. Our findings strongly support the conclusion that those predictors are inducers of EMT via activation of transcription factors SNAI1 and ZEB2.

1.3   Thesis Outline

In Chapter 2, we develop a generalized additive modeling framework for taking into account the effect of predictors on the dependence structure between two variables. We propose a maximum penalized log-likelihood estimator, derive its root-n-consistency and asymptotic normality, discuss details of the estimation procedure and the selection of the smoothing parameter. Finally, we present the results from a simulation study and apply the new methodology to a real dataset. In the application, we study the dependence structure of intraday asset returns. We show that the intraday dependence pattern, due to the cyclical nature of market activity, is shaped similarly to the conditional second moment.

In Chapter 3, we use generalized additive models to extend pair-copulas constructions (PCCs), namely flexible models that represent the distribution of a random vector as a cascade of bivariate copulas. We work specifically in two directions: the inclusion of exogenous variables and the relaxation of the so-called simplifying assumption. To model the effect of exogenous variables on a PCC, we let each pair-copula parameter
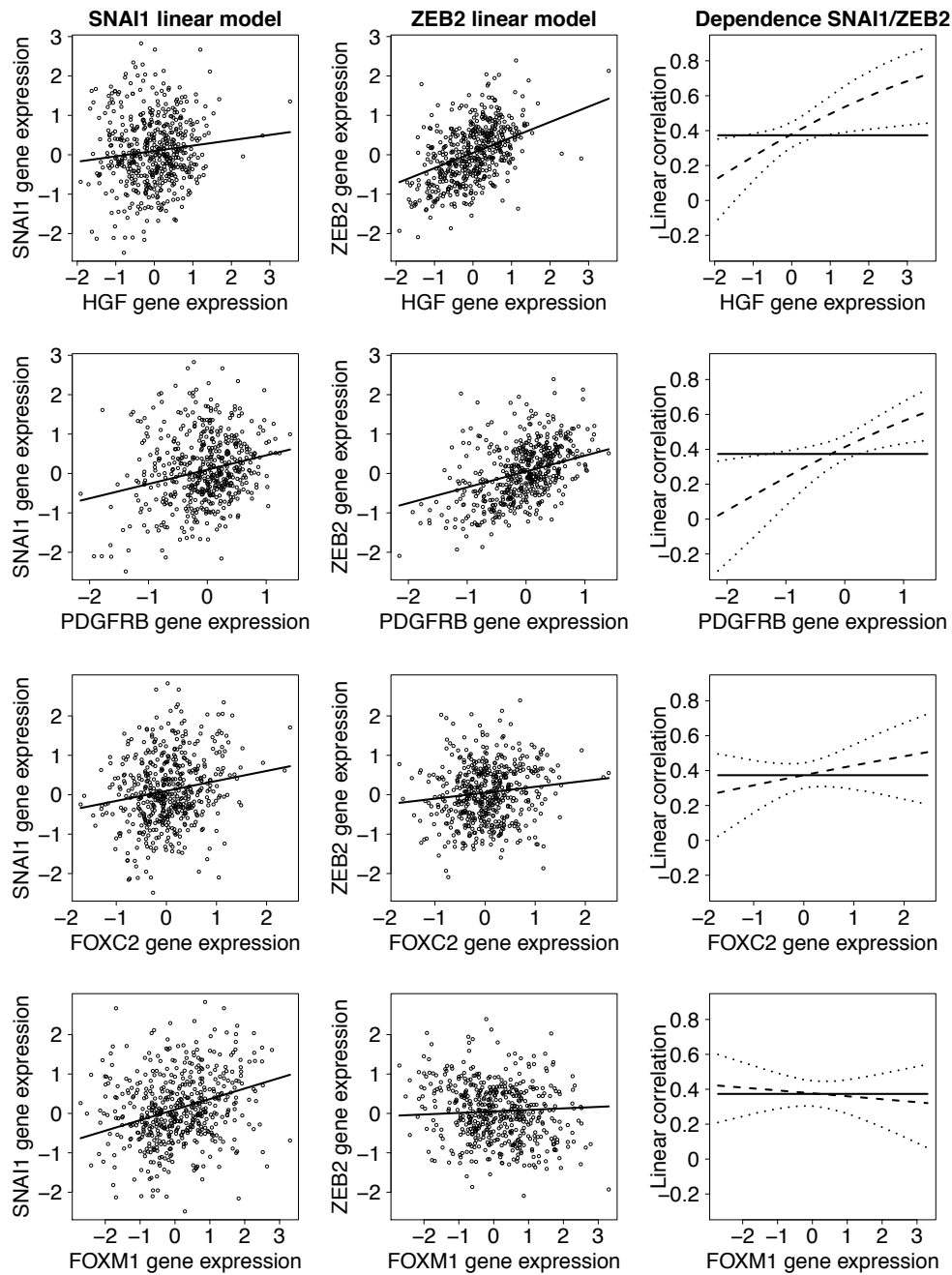
Fig. 1.6 **EMT results for HGF, PDGF-R, FOXC2 and FOXM1.** Left and middle panels: univariate marginal models; Right panel: dependence structure with constant dependence (plain line), the dependence as a function of the predictor (dashed line) and the 95 % confidence intervals (dotted lines). Note that, because we used a linear model and the Gaussian copula, the dependence structure is appropriately represented by Pearson's correlation.

depend directly on the exogenous variables in a parametric, semiparametric or nonparametric way. Furthermore, to relax the simplifying assumption, we let each pair-copula parameter depend on the set of conditioning variables implied by the PCC structure. We propose a sequential estimation method that we study with simulations. Finally, we apply our method to two data sets. First, we extend our analysis of the time-varying dependence structure between the intraday returns on foreign exchange rates to the three dimensional case. Second, we model the seven-dimensional distribution of an uranium exploration dataset.

In Chapter 4, we describe in details our implementation, which is freely available as the R (R Core Team 2013) package gamCopula at https://github.com/tvatter/gamCopula. In Chapter 5, we conclude and discuss directions of future research.

Chapter 2

Generalized Additive Models for Conditional Dependence Measures

In this Chapter, we develop the main theory of the thesis, that is our framework for taking into account the effect of predictors on the dependence structure between two variables. The structure of the Chapter is as follows: In Section 2.1, we provide a gentle introduction to generalized additive models and the roughness penalty approach. In Section 2.2, we develop the theoretical framework of generalized additive models for the dependence structure. We present the general model for the conditional dependence or concordance measure in Section 2.2.1. In Section 2.2.2, we state some asymptotic properties of the penalized log-likelihood estimator, namely its $\sqrt{n}$-consistency and asymptotic normality, assuming either known or unknown margins. In Section 2.2.3, we recast the penalized likelihood estimation as an iteratively reweighted generalized ridge regression. We close our theoretical considerations in Section 2.2.4, by discussing a measure of the penalized model's effective dimension and the selection of smoothing parameters. In Section 2.3, we present a simulation study and an application using a real dataset. We analyze the results of the simulation study in Section 2.3.1. We study the cross-sectional dynamics of intraday asset returns in Section 2.3.2. We conclude and suggest directions for further work in Section 2.4.

To set up the notations, we use uppercase (boldface) letters for scalar random variables (random vectors and matrices) and lowercase (boldface) letters for scalars (vectors and matrices). We differentiate scalar functions with identical names by their arguments, and similarly for boldfaced vectors and matrices functions. We use vectors columns, $\| \cdot \|^2$ for the Euclidean norm, subscripts for the elements of a given matrix or vector and superscripts for either independent copies of a random quantity or realized observations. We denote real intervals (or Cartesian products thereof) by double-struck

capital letters, except for the usual $\mathbb{N}, \mathbb{Z}, \mathbb{R}$, etc. For $k, l \in \mathbb{N}$ and $\mathbb{W} \subseteq \mathbb{R}^l$, we use $C^k(\mathbb{W})$ for the space of functions with k continuous (partial) derivatives on the interior of $\mathbb{W}$.

2.1   Generalized Additive Models

To model the distribution of Y given X, it is often useful to start by assuming that the conditional mean of the response variable is a linear function of the predictor, that is

$$Y = \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2 X + Z$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)^\top \in \mathbb{R}^2$ and Z is a zero-mean random variable. When facing a real dataset of n observations $\{x_j, y_j\}_{j=1}^n$, a common method to estimate the model's two parameters by minimizing the residual sum of squares. We define this so-called ordinary least-squares (OLS) estimator as

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^2}{\operatorname{argmin}} \sum_{j=1}^n (y_i - \beta_1 - \beta_2\, x_i)^2 = \underset{\boldsymbol{\beta} \in \mathbb{R}^2}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{x}\boldsymbol{\beta}\|^2 = \left(\mathbf{x}^\top \mathbf{x}\right)^{-1} \mathbf{x}^\top \mathbf{y},$$

where $\|\cdot\|$ denotes the L^2 norm, $\mathbf{y} = (y_1, \cdots, y_n)^\top$ and $\mathbf{x} = \left(\begin{smallmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{smallmatrix}\right)^\top$. The second
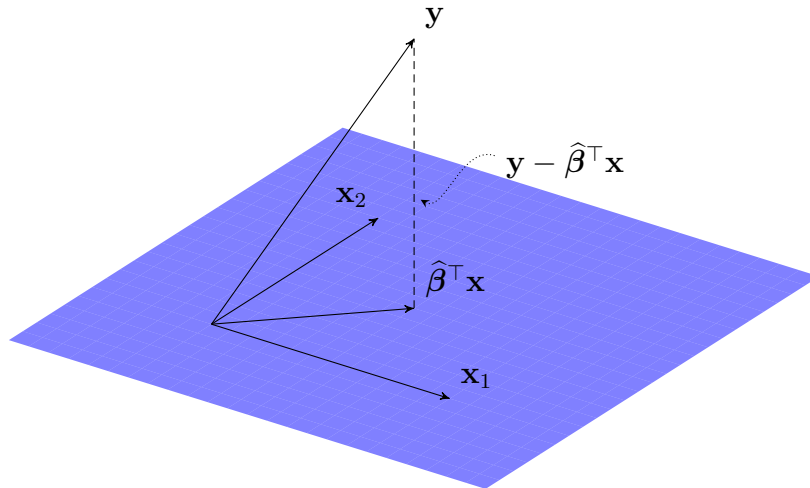
Fig. 2.1 **The ordinary least-squares:** a projection on the linear span of the predictors.

equality illustrates the fact that the OLS estimator can be viewed as a projection on the linear span of the predictors (see Figure 2.1 for a graphical example when $n = 3$).

While the linear relationship between the response and the predictor is a useful starting point, consider the data in Figure 2.2, which shows measurements of head

acceleration in a simulated motorcycle accident, used to test crash helmets (see Silverman 1985). As is often the case with real data, the linearity assumption is clearly not satisfied. In this case, a more complex model is required to capture the underlying features.
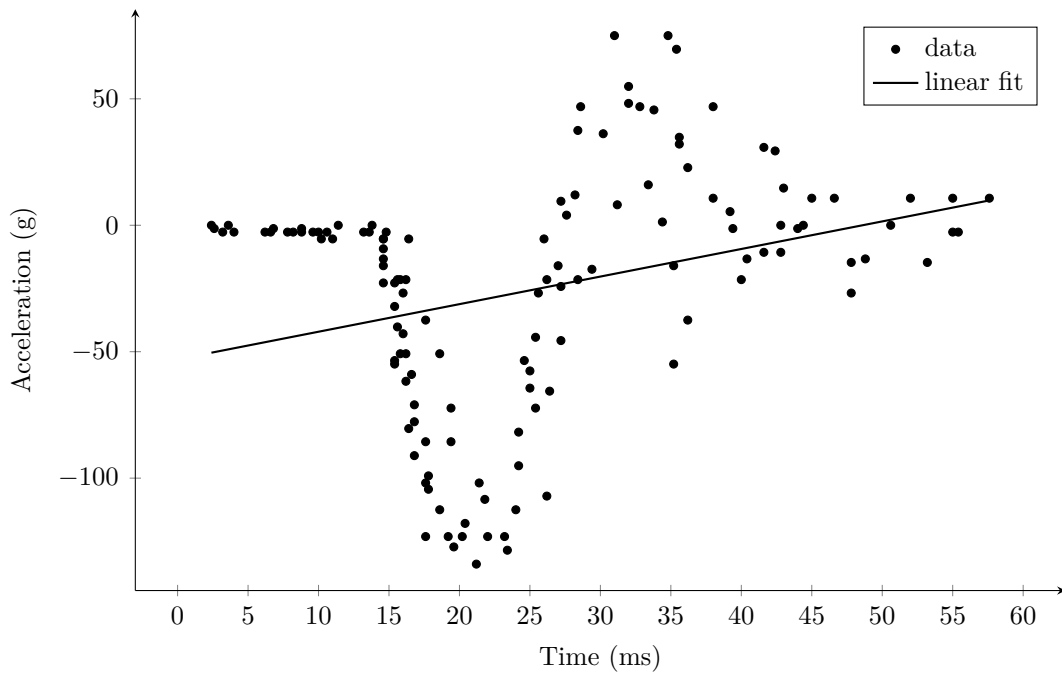
Fig. 2.2 **Data from a simulated motorcycle accident:** the linearity assumption is seldom verified in real datasets.

A natural extension of the linear framework is obtained by only assuming that the conditional mean of the response variable is a smooth function of the predictor, that is

$$Y = f(X) + Z, \tag{2.1}$$

where $f \in \mathcal{F} \subseteq \mathcal{C}^l$ for some $l \geq 1$ and Z is as before. Expending on the OLS ideas to estimate f, we start by choosing a basis of (or a close approximation of) $\mathcal{F}$ using the predictor(s). Then, by projecting the response variable on this appropriately chosen basis, a linear model obtains.

2.1.1   Basis Decompositions

Consider a collection of m basis functions $b_k \in \mathcal{F}$ for $k \in \{1, \cdots, m\}$, then we can write

$$f(X) = \sum_{k=1}^{m} \beta_k b_k(X), \tag{2.2}$$

where $\beta_k \in \mathbb{R}$ for $k \in \{1, \cdots, m\}$ are the parameters. Although the linearity assumption between the response and the predictor is relaxed, plugging (2.2) into (2.1) still defines a linear model, whose parameters can be obtained using the OLS estimator by rewritting the model matrix as

$$\mathbf{x} = \begin{pmatrix} b_1(x_1) & \cdots & b_m(x_1) \\ \vdots & \ddots & \vdots \\ b_1(x_n) & \cdots & b_m(x_n) \end{pmatrix}.$$

For instance, a polynomial basis can be constructed by using $b_k(x) = x^{k-1}$ for all $x \in \mathbb{R}$. While they are useful when considering the behavior of a function in the vicinity of a single point, they tend to perform badly when the whole domain is of interest. In this context, spline bases are usually preferred because of their appealing numerical properties[1]. Simply put, a spline is a piecewise polynomial that is constructed by dividing the domain of interest using a set of specified points, called knots. More specifically, a spline of order r has continuous derivatives up to order $r - 1$ on its domain.

As an example, let us consider cubic splines, namely piecewise cubic polynomial with continous first and second derivatives. Without loss of generality, we assume that the domain is $[0, 1]$, and that the knots are $0 = s_0 \leq s_1 \leq \cdots \leq s_{m-1} = 1$. Following Wood (2006), a basis is obtained by setting $b_1(x) = 1$, $b_2(x) = x$ and $b_{j+2}(x) = R(x, s_j)$ for $j \in \{1, \cdots, m-2\}$, where

$$R(x, z) = \left\{ (z - 1/2)^2 - 1/12 \right\} \left\{ (x - 1/2)^2 - 1/12 \right\} \tag{2.3}$$

$$- \left\{ (|x - z| - 1/2)^4 - 1/2 \, (|x - z| - 1/2)^2 + 7/240 \right\} / 24. \tag{2.4}$$

For instance, $\{b_1(x), \cdots, b_5(x)\}$ with $b_1(x) = 1$, $b_2(x) = x$, $b_3(x) = R(x, 1/6)$, $b_4(x) = R(x, 3/6)$ and $b_5(x) = R(x, 5/6)$ form a basis for cubic splines on $[0, 1]$ with three

---

[1]For instance, as an interpolant, the spline approximation error can often be made arbitrarily small using only low-degree polynomials, thus avoiding Runge's phenomenon (see de Boor 2001, Chapter 2 page 22, and Chapter 4 page 41).
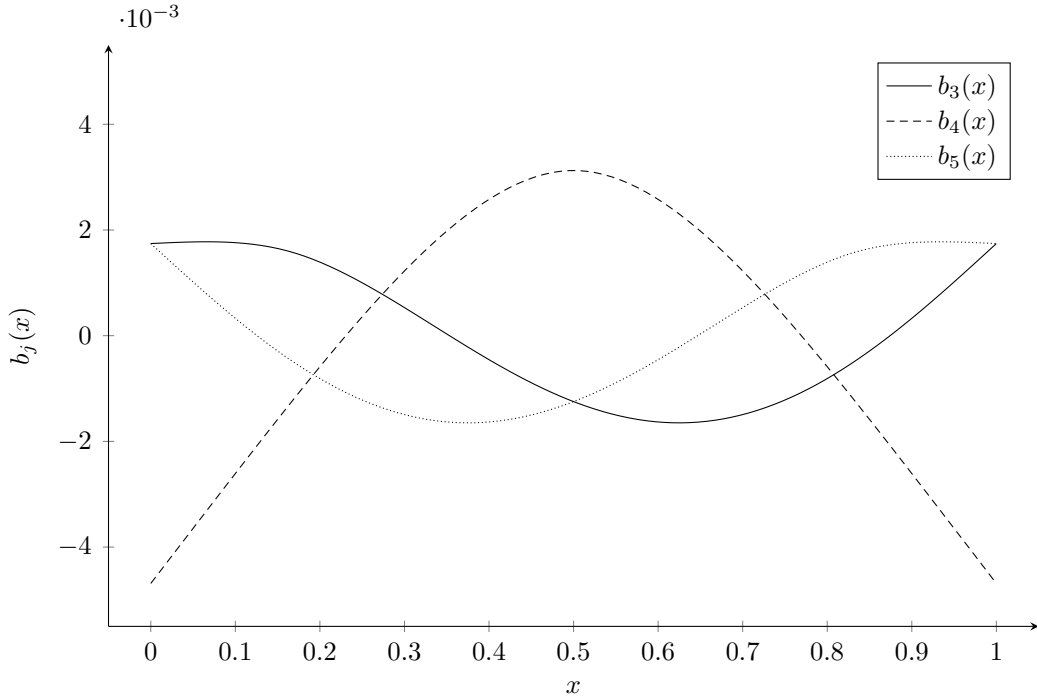
Fig. 2.3 **Cubic splines on** $[0,1]$ **with interior knots** $\{1/6, 3/6, 5/6\}$**:** $b_3(x) = R(x, 1/6)$, $b_4(x) = R(x, 3/6)$ and $b_5(x) = R(x, 5/6)$, where $R(x, z)$ is defined by (2.3).

interior knots $\{1/6, 3/6, 5/6\}$, and we show the last three basis functions in Figure 2.3. To fit such a model, we can simply replace row j of the model matrix by

$$(1, x_j, R(x_j, s_1), \cdots, R(x_j, s_{m-1})),$$

and use the OLS estimator. As an example, let us consider the data from Figure 2.2 again. In Figure 2.4, we show the resulting fits obtained using evenly spaced knots. As we observe, the basis choice has a strong influence on the quality of the fit: if the number of interior knots is too small, then some features are missed, whereas if it is too large, overfitting becomes more likely. Because of the linear setup, traditional model selection tools can readily be applied. However, since the knots are evenly spaced, the different models are not nested, which makes their comparison harder.

2.1.2 The Roughness Penalty Approach

To deal with this issue, an appealing alternative consists in setting a basis larger than necessary while controlling the smoothness by adding a wiggliness penalty. Note that the smoothness discussed here does not correspond to the classical definition from
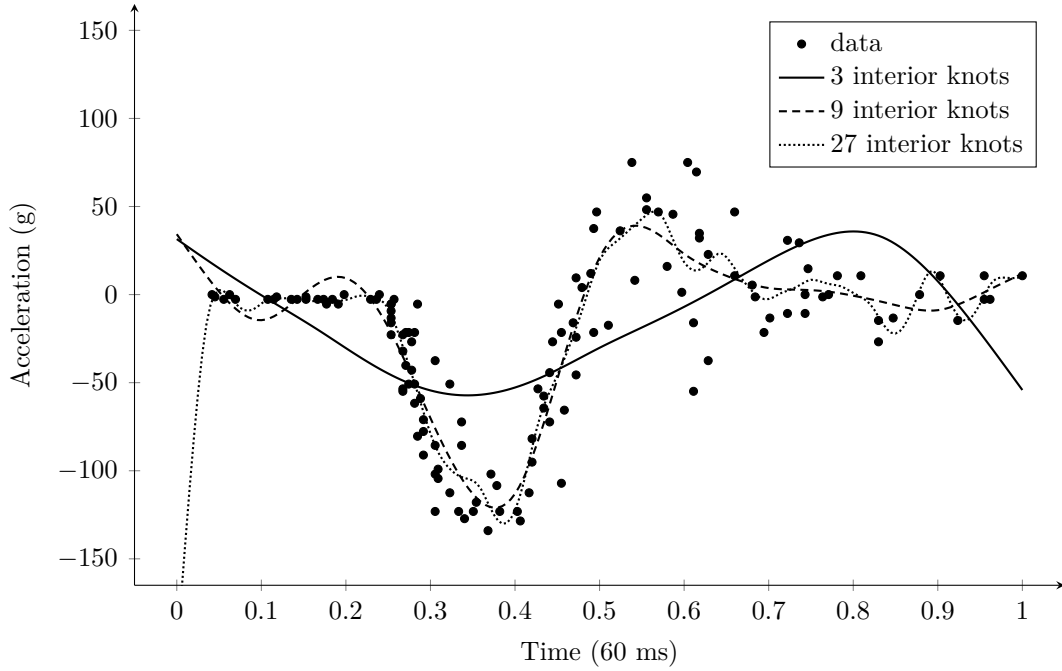
Fig. 2.4 **The simulated motorcycle accident:** cubic spline fits with different bases.

mathematical analysis[2]: it is a rather loosely defined concept that may depend on the problem at hand. In statistics, a function is often said smooth when belonging to $\mathcal{C}^2$, because discontinuities in the third derivative are difficult (if not impossible) to spot visually. However, as outlined in the example above, not all functions are equally smooth, and various measures of wigglinness can be defined as operators acting on a function's second derivative. Having a wigglinness of zero, a constant or linear function is then completely smooth in this sense. An especially useful penalty is the integrated squared second derivatives, that is

$$\int_{\mathcal{F}} f''(x)^2 \, dx.$$

While this is a global penalty and one may wish to penalize more locally the second derivative, the computational benefits of the quadratic form as a penalty are difficult to match. For instance, one can easily show that $\widehat{f}$, defined as

$$\widehat{f} = \operatorname*{argmin}_{f \in \mathcal{C}^2(\mathcal{F})} \|\mathbf{y} - \mathbf{x}\boldsymbol{\beta}\|^2 + \lambda \int_{\mathcal{F}} f''(x)^2 \, dx,$$

---

[2]In mathematical analysis, the space of smooth functions is $\mathcal{C}^\infty$, that is the space of infinitely differentiable functions.

with λ a smoothing parameter, is a natural cubic spline, that is a cubic spline with second derivatives equaling zero at the boundary knots (see e.g., Green and Silverman 2000). Furthermore, for a basis that possesses at least two derivatives integrable on $\mathcal{F}$, we have that

$$\int_{\mathcal{F}} f''(x)^2 \, \mathrm{d}x = \boldsymbol{\beta}^\top \mathbf{s}\boldsymbol{\beta},$$

where $\mathbf{s}$ depends only on the basis choice. With the cubic spline basis described above, then $\mathbf{s}_{i+2,\,j+2} = R(s_i, s_j)$ for $1 \leq i, j \leq m-2$ and zero otherwise. Finally, using this quadratic form, the penalized OLS estimator admits an analytic solution, since

$$\widehat{\boldsymbol{\beta}} = \operatorname*{argmin}_{\boldsymbol{\beta} \in \mathbb{R}^2} \|\mathbf{y} - \mathbf{x}\boldsymbol{\beta}\|^2 + \lambda \boldsymbol{\beta}^\top \mathbf{s}\boldsymbol{\beta} = \left(\mathbf{x}^\top \mathbf{x} + \lambda \mathbf{s}\right)^{-1} \mathbf{x}^\top \mathbf{y}.$$

Using the simulated motorcycle accident data again, we can observe the effect of the smoothing parameter on the penalized OLS estimator resulting from the basis with 27 interior knots: when the $\lambda \to 0$, we obtain the wiggly solution we obseved in Figure 2.4, and when $\lambda \to \infty$, the linear OLS estimator.
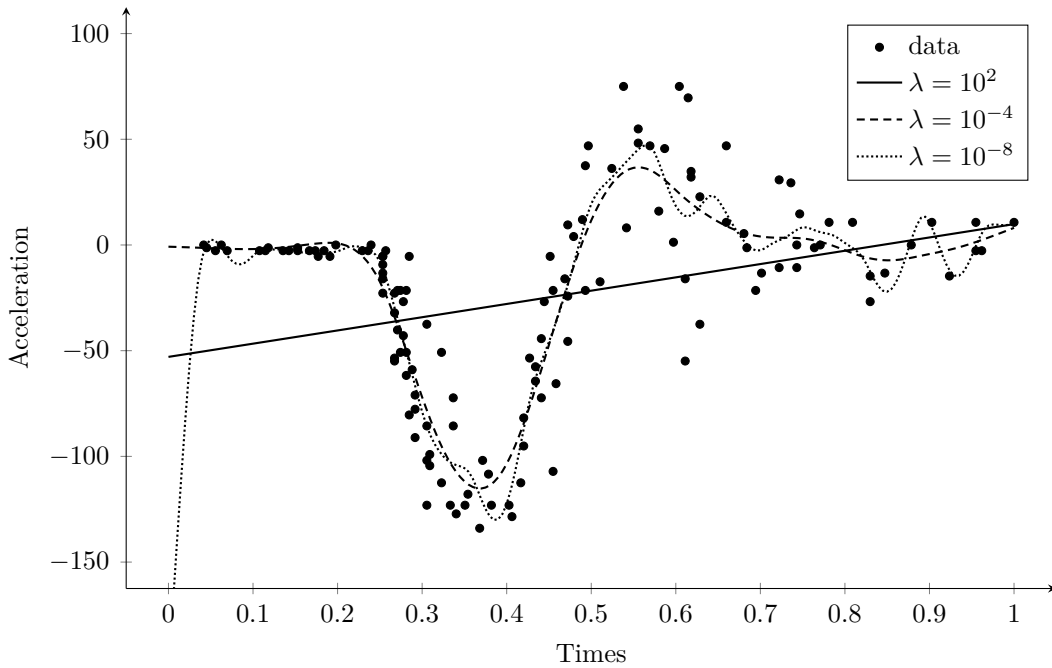
Fig. 2.5 **The simulated motorcycle accident:** penalized cubic spline fits with 27 interior knots and different smoothing parameters.

2.1.3 Smoothing Parameter Selection

As observed in the preceding example, the smoothing parameter has a big influence on the resulting curve estimate. If λ is too small (respectively large), then the data is undersmoothed (respectively oversmoothed). In order to choose the right amount of smoothing, it is convenient to define a measure of the model's penalized dimensionality. Let $\mathbf{h}(\lambda)$ be the so-called influence or hat matrix, defined such that $\hat{\mathbf{y}} = \mathbf{h}(\lambda)\mathbf{y}$, that is the matrix which yields the predicted response when premultiplying the observations:

$$\mathbf{h}(\lambda) = \mathbf{x}\left(\mathbf{x}^\top \mathbf{x} + \lambda \mathbf{s}\right)^{-1} \mathbf{x}^\top.$$

We can now define the equivalent (or effective) degrees of freedom (see e.g., Green and Silverman 2000; Hastie and Tibshirani 1990) as the hat matrix's trace, that is

$$EDF(\lambda) = \mathrm{Tr}\left\{\mathbf{x}\left(\mathbf{x}^\top \mathbf{x} + \lambda \mathbf{s}\right)^{-1} \mathbf{x}^\top\right\}.$$

It is clear that, when $\lambda = 0$, then the equivalent degrees of freedom is equal to the number of columns in the model matrix $\mathbf{x}$, that is the basis size m. Furthermore, when $\lambda \longrightarrow \infty$, then the equivalent degrees of freedom usually has 2 for limit. This holds whenever the linear model is in the span of the chosen basis, which is often the case when modeling $\mathcal{C}^2$ functions. For instance, this is the case for the cubic basis whose penalty matrix was described above.

Building on this measure of penalized dimensionality, a simple criterion to choose the smoothing parameter can now be devised. Similarly as in Craven and Wahba (1979), we define the generalized cross-validation (GCV) score as

$$GCV(\lambda) = \frac{n^{-1}\left\|\mathbf{y} - \mathbf{h}(\lambda)\mathbf{y}\right\|^2}{\left\{1 - EDF(\lambda)/n\right\}^2},$$

which is basically a trade-off between goodness of fit and penalized dimensionality. In Figure 2.6, we show the generalized cross-validation score as a function of the smoothing parameter, and we observe that it is minimized for $\lambda \approx 10^{-4}$ (recall this intermediate solution in Figure 2.5).
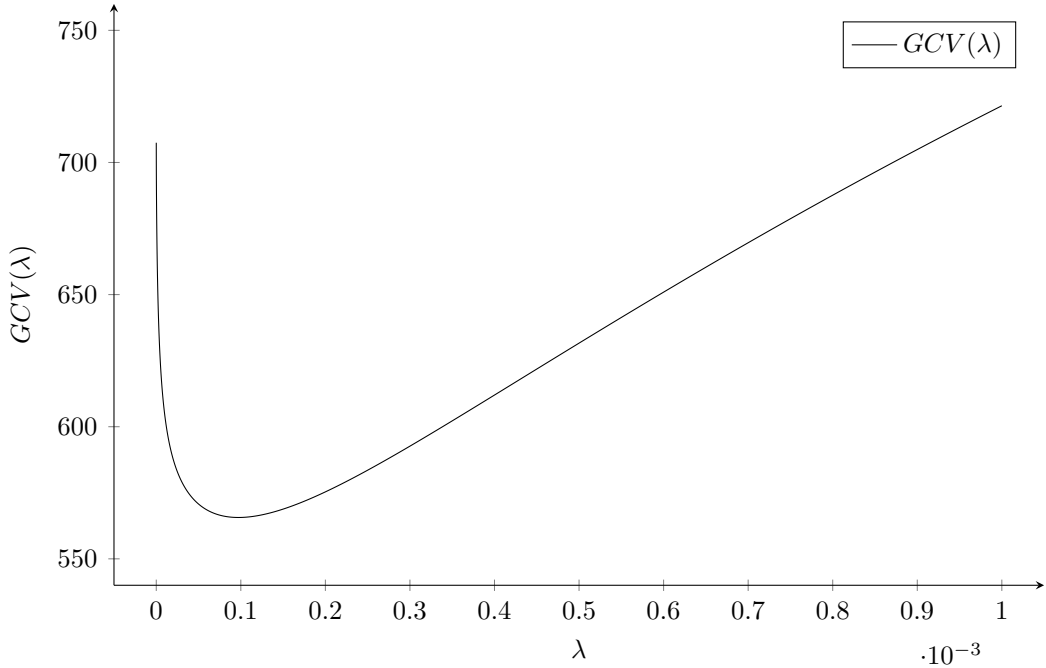
Fig. 2.6 **The simulated motorcycle accident:** minimizing the generalized cross-validation score.

2.2 Generalized Additive Models for Conditional Dependence Structures

In this section, we detail the approach to model the dependence structure between two random variables as a function of an arbitrary set of exogenous predictors (covariates).

Let $\mathbf{Y} \in \mathbb{Y} \subseteq \mathbb{R}^2$ be the random vector (responses) of interest, $\mathbf{X} \in \mathbb{X} \subseteq \mathbb{R}^q$ be a vector of q covariates (predictors). For $\mathbf{y} \in \mathbb{Y}$, $\mathbf{x} \in \mathbb{X}$ and $i \in \{1, 2\}$, we denote by $F_{\mathbf{Y}_i|\mathbf{X}}(\mathbf{y}_i \mid \mathbf{x}) = P(Y_i \leq y_i \mid \mathbf{X} = \mathbf{x})$ the conditional margins and by $\mathbf{U}$ the random vector of conditional probability integral transforms with $\mathbf{U}_i = F_{\mathbf{Y}_i|\mathbf{X}}(\mathbf{Y}_i \mid \mathbf{X})$. Assume further that all variables are continuous and have strictly positive densities. In what follows, we rely on Patton's theorem, which, in two dimensions, specializes to: for all $\mathbf{x} \in \mathbb{X}$, there exists a unique conditional copula $C(\cdot \mid \mathbf{x})$ which is the conditional distribution of $\mathbf{U} \mid \mathbf{X} = \mathbf{x}$. In other words, for $\mathbf{u} \in [0, 1]^2$, we have that

$$
\begin{aligned}
C(\mathbf{u} \mid \mathbf{x}) &= P(\mathbf{U} \leq \mathbf{u} \mid \mathbf{X} = \mathbf{x}) \\
&= F_{\mathbf{Y}|\mathbf{X}}\left\{ F_{\mathbf{Y}_1|\mathbf{X}}^{-1}(\mathbf{u}_1 \mid \mathbf{x}), F_{\mathbf{Y}_2|\mathbf{X}}^{-1}(\mathbf{u}_2 \mid \mathbf{x}) \mid \mathbf{x} \right\}.
\end{aligned}
\tag{2.5}
$$

Remark. In Patton (2002), the conditioning vector is the same for the two margins and copula. In time series, conditioning algebras are usually augmented with past observations. Therefore, the concept of conditional copulas can be rather restrictive (see Fermanian and Wegkamp 2012). However, conditional copulas and exogeneity of the predictors are sufficient to develop a regression-like theory for the dependence structure.

For $\mathbf{y} \in \mathbb{Y}$ and $\mathbf{x} \in \mathbb{X}$, the joint conditional density is

$$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x}) = \underbrace{\frac{\partial F_{\mathbf{Y}_1|\mathbf{X}}(\mathbf{y}_1 \mid \mathbf{x})}{\partial \mathbf{y}_1}}_{f_{\mathbf{Y}_1|\mathbf{X}}(\mathbf{y}_1 \mid \mathbf{x})} \underbrace{\frac{\partial F_{\mathbf{Y}_2|\mathbf{X}}(\mathbf{y}_2 \mid \mathbf{x})}{\partial \mathbf{y}_2}}_{f_{\mathbf{Y}_2|\mathbf{X}}(\mathbf{y}_2 \mid \mathbf{x})} \underbrace{\frac{\partial^2 C(\mathbf{u} \mid \mathbf{x})}{\partial \mathbf{u}}}_{c(\mathbf{u} \mid \mathbf{x})}, \qquad (2.6)$$

where $\mathbf{u}_i = F_{\mathbf{Y}_i|\mathbf{X}}(\mathbf{y}_i \mid \mathbf{x})$. Similarly as in the unconditional case, the right-hand-side of (2.6) is a product between conditional marginal and copula densities ($f_{\mathbf{Y}_1|\mathbf{X}}$, $f_{\mathbf{Y}_2|\mathbf{X}}$ and $c(\mathbf{u} \mid \mathbf{x})$).

In what follows, we focus on bivariate conditional copulas depending on the covariate only through a single parameter that we denote $\eta \in \mathbb{H}$. Let $c(\mathbf{u} \mid \mathbf{x}) = c\{\mathbf{u}; \eta(\mathbf{x})\}$ with the parameter being a function of the covariates, that is $\eta(\mathbf{x}) : \mathbb{X} \to \mathbb{H}$. In Section 2.2.1, we present the generalized additive model. We formulate the asymptotic properties assuming either known or estimated margins in Section 2.2.2. In Sections 2.2.3 and 2.2.4, we address the estimation of the model.

2.2.1   The Model

Consider an arbitrary dependence or concordance measure ψ between the components of $\mathbf{Y}$ that we want to condition on a realization of $\mathbf{X}$. We assume that

- ψ satisfies the two properties of Chapter 1 and takes values in a closed $\mathbb{P} \subset \mathbb{R}$;

- $\exists$ a mapping $\nu : \mathbb{P} \to \mathbb{H}$ such that $\eta = \nu(\psi)$ with $\nu \in C^{\infty}(\mathbb{P})$ strictly increasing.

For instance, Kendall's tau is a natural choice. In this case, the mappings for common copulas such as the Gaussian, Clayton and Gumbel are $\nu(\psi) = \sin\left(\frac{\pi}{2}\psi\right)$, $\nu(\psi) = 2\psi/(1-\psi)$ and $\nu(\psi) = 1/(1-\psi)$. For copula families with several parameters, all but one need to be treated as nuisances. The issue is discussed in Sections 2.3.1 and 2.4.

Remark. As it introduces the requirement of a one-to-one mapping, modeling a dependence or concordance measure instead of the copula parameter directly may seem an

unnecessary complication. In fact, this choice is dictated by the context of application and has two desirable properties. First, a dependence or concordance measure has a more natural interpretation than a copula parameter. Second, when the modeled distributional feature is the same, it is easier to compare various parametric families, for instance using information criteria. Nonetheless, the whole theory can be adapted in a straightforward fashion, using ν as the identity and modifying g accordingly (or vice-versa). In Section 2.3, we also illustrate models specified for the copula parameter.

For $\mathbf{x} \in \mathbb{X}$, a generalized additive model for the conditional measure can then be written as

$$\psi(\mathbf{x}; \boldsymbol{\theta}) = g\left\{ \mathbf{z}^\top \boldsymbol{\beta} + \sum_{k=1}^{K} h_k(\mathbf{t}_k) \right\}, \tag{2.7}$$

where

- $g : \mathbb{R} \to \mathbb{P}$ is a strictly increasing and $C^\infty(\mathbb{R})$ link expressing the relationship between the GAM and ψ (e.g., $g(x) = (e^x - 1)/(e^x + 1)$ when $\mathbb{P} = [-1, 1]$),

- $\mathbf{z} \in \mathbb{R}^p$ and $\mathbf{t} \in \mathbb{R}^K$ are subsets of $\mathbf{x}$ or products thereof to consider interactions,

- $\boldsymbol{\beta} \in \mathbb{R}^p$ is a vector of parameters,

- $h_k : \mathbb{T}_k \to \mathbb{R}$ are smooth functions supported on closed $\mathbb{T}_k \subset \mathbb{R}$ for all k and

- $\boldsymbol{\theta} \in \Theta$ is the vector of stacked parameters, containing both $\boldsymbol{\beta}$ and h_k for all k.

In this thesis, we assume that the smooth functions $h_k \in C^2(\mathbb{T}_k)$ admit a finite-dimensional basis-quadratic penalty representation (Green and Silverman 2000; Hastie and Tibshirani 1990; Wood 2006). A natural cubic spline (NCS) $h : \mathbb{T} \subset \mathbb{R} \to \mathbb{R}$ with fixed knots is a particular case. Suppose that the m fixed knots are such that $\inf \mathbb{T} = s_0 < s_1 < \cdots < s_m < s_{m+1} = \sup \mathbb{T}$. As an NCS is linear on the two extreme intervals $[s_0, s_1]$ and $[s_m, s_{m+1}]$ and twice continuously differentiable on its support $\mathbb{T}$, it has only m free parameters (say $\mathbf{h} \in \mathbb{R}^m$). Furthermore, there exists a unique $m \times m$ symmetric matrix $\mathbf{s}$ of rank $m - 2$ such that $\int_{\mathbb{T}} h''(t)^2 \, dt = \mathbf{h}^\top \mathbf{s} \mathbf{h}$. This matrix is fixed in the sense that it depends on the knots but not on $\mathbf{h}$. Apart from NCSs, many alternative C^2 smoothers admit this finite dimensional basis-quadratic penalty representation. For instance, tensor product splines (functions of multiple predictors) or cyclic cubic splines (to model periodic functions) are included in the GAM toolbox.

Remark. Additionally to potential multicollinearities arising from the linear part of (2.7), concurvity[3] may also play a role in the model's identifiability. To check for ill-conditioning in a matrix of linear predictors, the condition number is the standard measure. Similarly, it is possible to quantify concurvity with respect to the data and chosen bases.

Using such a representation for each h_k, we denote by $\mathbf{h}_k \in \mathbb{R}^{m_k}$ the m_k-dimensional parametrization and $\mathbf{h} = (\mathbf{h}_1^\top, \cdots, \mathbf{h}_K^\top)^\top$. Hence, the complete vector of parameters is $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \mathbf{h}^\top)^\top$, taking values in $\Theta \subseteq \mathbb{R}^d$ with $d = p + \sum_{k=1}^K m_k$. We also define the $m_k \times m_k$ matrices $\mathbf{s}_k$ such that

$$\int_{\mathbb{T}_k} h_k''(t)^2 \, dt = \mathbf{h}_k^\top \mathbf{s}_k \mathbf{h}_k.$$

Remark. Each h_k can be further constrained using a $l \times m_k$ matrix $\mathbf{c}$, such that an additional set of $l < m_k$ constraints is met whenever $\mathbf{ch}_k = 0$. For instance, a sensible identifiability requirement is that h_k integrates to zero over $\mathbb{T}_k$. Letting r be the nullity of $\mathbf{c}$, we represent an element of its null space by $\mathbf{nw}$, with $\mathbf{n}$ a $m_k \times r$ matrix and $\mathbf{w} \in \mathbb{R}^r$. Using $\mathbf{h}_k = \mathbf{nw}$ as a reparametrization of h_k, the constraints are automatically met.

For $\mathbf{x} \in \mathbb{X}$ and $\boldsymbol{\theta} \in \Theta$, the copula parameter is

$$\eta(\mathbf{x}; \boldsymbol{\theta}) = \nu \left\{ \psi(\mathbf{x}; \boldsymbol{\theta}) \right\}$$

and for $\mathbf{u} \in [0,1]^2$, we denote the log-likelihood function by

$$\ell_c(\mathbf{u}, \mathbf{x}; \boldsymbol{\theta}) = \log \left[c \left\{ \mathbf{u}; \eta(\mathbf{x}; \boldsymbol{\theta}) \right\} \right].$$

Considering a sample of n observations $\{\mathbf{u}^j, \mathbf{x}^j\}_{j=1}^n$, we estimate $\boldsymbol{\theta}$ by maximizing the penalized log-likelihood

$$\begin{aligned}
\ell_c(\boldsymbol{\theta}, \boldsymbol{\gamma}) &= \ell_c(\boldsymbol{\theta}) - \frac{1}{2} \sum_{k=1}^K \gamma_k \int_{\mathbb{T}_k} h_k''(\mathbf{t}_k)^2 d\mathbf{t}_k \\
&= \ell_c(\boldsymbol{\theta}) - \frac{1}{2} \sum_{k=1}^K \gamma_k \mathbf{h}_k^\top \mathbf{s}_k \mathbf{h}_k = \ell_c(\boldsymbol{\theta}) - \frac{1}{2} \boldsymbol{\theta}^\top \mathbf{p}(\boldsymbol{\gamma}) \boldsymbol{\theta},
\end{aligned} \tag{2.8}$$

---

[3]Concurvity is the nonparametric analogue of collinearity. When the model contains smooth components, we say that concurvity is present when there is $j \in \{1, \cdots, K\}$ such that $h_j(\mathbf{t}_j)$ is highly correlated with $\sum_{k \neq j} h_k(\mathbf{t}_k)$ (see e.g., Hastie and Tibshirani 1990, pages 118–123, where this definition is equivalent to their notion of approximate concurvity).

with $\ell_c(\boldsymbol{\theta}) = n^{-1} \sum_{j=1}^n \ell_c(\mathbf{u}^j, \mathbf{x}^j; \boldsymbol{\theta})$, $\boldsymbol{\gamma} \in (\mathbb{R}_+ \cup \{0\})^K$ and $\mathbf{p}(\boldsymbol{\gamma})$ is a $d \times d$ block diagonal matrix with $K + 1$ blocks; the first $p \times p$ is filled with zeros and the remaining K with $m_k \times m_k$ matrices are $\boldsymbol{\gamma}_k \mathbf{s}_k$. The integral terms are roughness penalties on each component and $\boldsymbol{\gamma}$ is a vector of smoothing parameters. We define the penalized maximum log-likelihood estimator as

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmax}} \; \ell_c(\boldsymbol{\theta}, \boldsymbol{\gamma}). \tag{2.9}$$

Remark. Fixing the number and location of the knots beforehand amounts to assuming that the true model can be represented using the finite-dimensional basis (Yu and Ruppert 2002). If this is not the case, consistency would require either additional assumptions on the smooth functions or an infinity of knots, or both. However, a finite-dimensional parameter space implies $\sqrt{n}$-consistency and asymptotic normality under standard regularity assumptions (see Section 2.2.2). In many practical cases, smoothers are used as approximations of the underlying functions of interest. In exploratory data analysis for instance, smoothers "let the data speak for themselves" using a minimal set of assumptions. When the true underlying functions do not admit a finite dimensional-quadratic basis representation, $\widehat{\boldsymbol{\theta}}$ is naturally interpreted as a projection. We refer to Yu and Ruppert (2002), Section 3 for a more detailed discussion.

2.2.2 Asymptotic Properties

In this section, we establish the required assumptions to ensure $\sqrt{n}$-consistency and asymptotic normality of the penalized maximum likelihood estimator. We start by assuming known margins and extend the results when they are parametrically estimated within a two-step procedure.

As noted in Joe (1997), Chapter 10, there are two approaches to derive asymptotic properties in the conditional copula context. With the first approach, one obtains $\sqrt{n}$-consistency and asymptotic normality using a Lindeberg-Feller type of condition, considering the covariates as fixed. With the second approach, similar results are derived by analyzing the joint distribution of the conditional probability integral transforms along with the covariates, which are treated as random variables. Because the finite sample estimators of the covariance matrices are identical, it is a matter of taste whether one should use one approach or the other. In what follows, we study $\mathbf{Z} = (\mathbf{U}^\top, \mathbf{X}^\top)^\top \in \mathbb{Z} = [0, 1]^2 \times \mathbb{X}$ directly, because it simplifies an analysis of the conditional two-step procedure similar to Joe (2005). For $\mathbf{z} \in \mathbb{Z}$, we write the joint

density as

$$f_{\mathbf{Z}}(\mathbf{z}) = c\left\{\mathbf{u}; \eta(\mathbf{x}; \boldsymbol{\theta})\right\} f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\omega}), \tag{2.10}$$

where $\boldsymbol{\omega}$ parametrizes the distribution of the covariates. Because (2.10) implies that $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$ are globally orthogonal (see Cox and Reid 1987), regularity conditions are only required for the conditional copula. However, for $\widehat{\boldsymbol{\theta}}$ to be asymptotically unbiased when the true $\boldsymbol{\theta}_0$ is in the interior of Θ, a penalty that vanishes with a $\sqrt{n}$-rate is also necessary.

Theorem 2.2.1. *Asymptotic properties assuming known margins*
If Assumptions 1 and 2 from Appendix A.1 hold, then $\widehat{\boldsymbol{\theta}}$ is $\sqrt{n}$-consistent and

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N\left\{0, \mathbf{i}(\boldsymbol{\theta}_0)^{-1}\right\},$$

where $\mathbf{i}(\boldsymbol{\theta}) = cov\left[\partial \log c\left\{\mathbf{U}; \eta(\mathbf{X}; \boldsymbol{\theta})/\right\} \partial \boldsymbol{\theta}\right]$. The detailed assumptions and proof are provided in Appendix A.1.

In many cases of practical importance, the margins are unknown. While full maximum likelihood estimation (MLE) is more efficient, the resulting numerical optimization problem may be very time-consuming or even infeasible. Because it is easier to implement and faster, two-step estimation is often preferred. In an unconditional and parametric context, this procedure is called inference functions for margins (IFM) in Xu (1996) and Joe (1997). In Xu (1996), simulations are further provided for discrete/categorical models, suggesting this procedure to be highly efficient. The theoretical results for the IFM are extended to the continuous case in Joe (2005), where it is also found that the two-stage efficiency loss is small. In what follows, we consider $\mathbf{W} = (\mathbf{Y}^{\top}, \mathbf{X}^{\top})^{\top} \in \mathbb{W} = \mathbb{Y} \times \mathbb{X}$ and use the second approach outlined in Joe (1997), Chapter 10 to extend Joe (2005) in the conditional context.

For the sake of simplicity, we assume that each conditional margin depends on the covariates only through a single scalar parameter that we denote $\alpha_i \in \mathbb{H}_i$. In other words, for $\mathbf{y} \in \mathbb{Y}$ and $\mathbf{x} \in \mathbb{X}$, we let $\mathbf{w} = (\mathbf{y}^{\top}, \mathbf{x}^{\top})^{\top}$ and $f_{\mathbf{Y}_i|\mathbf{X}}(\mathbf{y}_i \mid \mathbf{x}) = f_{\mathbf{Y}_i|\mathbf{X}}\left\{\mathbf{y}_i; \alpha_i(\mathbf{x})\right\}$ with the parameter being a function of the covariates, that is $\alpha_i(\mathbf{x}) : \mathbb{X} \to \mathbb{H}_i$. For each α_i, we assume a model similar to (2.7), containing a sum of K_i smooth functions admitting a finite-basis quadratic-penalty representation. We denote by

- $\boldsymbol{\alpha}_i \in \mathbb{A}_i \subseteq \mathbb{R}^{l_i}$ a l_i-dimensional parametrization,

- $\ell_i(\mathbf{y}_i, \mathbf{x}; \boldsymbol{\alpha}_i) = \log\left[f_{\mathbf{Y}_i|\mathbf{X}}\left\{\mathbf{y}_i; \boldsymbol{\alpha}_i(\mathbf{x})\right\}\right]$ the marginal log-likelihood,

- $\boldsymbol{\lambda}_i \in (\mathbb{R}_+ \cup \{0\})^{K_i}$ the vector of smoothing parameters,

- $\mathbf{p}_i(\boldsymbol{\lambda}_i)$ the compact representation of the quadratic penalty representation as in (2.8),

- $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^\top, \boldsymbol{\alpha}_2^\top)^\top \in \mathbb{A} = \mathbb{A}_1 \times \mathbb{A}_2$ and $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1^\top, \boldsymbol{\lambda}_2^\top)^\top$.

Remark. The dependency of the margins on the covariates through a single scalar parameter is by no means necessary, but yields clearer developments in what follows. For instance, this formulation includes univariate GAMs, with each margin being completely determined by $\boldsymbol{\alpha}_i$ and additional nuisance parameters specific to the chosen univariate distribution.

Considering a sample of n observations $\{\mathbf{y}^j, \mathbf{x}^j\}_{j=1}^n$, the parameters can be efficiently estimated by maximizing the penalized joint log-likelihood

$$\ell(\boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\gamma}) = \sum_{i=1}^2 \ell_i(\boldsymbol{\alpha}_i, \boldsymbol{\lambda}_i) + \ell_c(\boldsymbol{\theta}, \boldsymbol{\gamma}), \qquad (2.11)$$

with $\ell_i(\boldsymbol{\alpha}_i, \boldsymbol{\lambda}_i) = \ell_i(\boldsymbol{\alpha}_i) - \boldsymbol{\alpha}_i^\top \mathbf{p}_i(\boldsymbol{\lambda}_i)\boldsymbol{\alpha}_i/2$, $\ell_i(\boldsymbol{\alpha}_i) = n^{-1}\sum_{j=1}^n \ell_i(\mathbf{y}_i^j, \mathbf{x}^j; \boldsymbol{\alpha}_i)$ and $\ell_c(\boldsymbol{\theta}, \boldsymbol{\gamma})$ as in (2.8) using $\mathbf{u}_i = F_{\mathbf{Y}_i|\mathbf{X}}\{\mathbf{y}_i; \alpha_i(\mathbf{x})\}$. While $\ell_c(\boldsymbol{\theta}, \boldsymbol{\gamma})$ is a function of $\boldsymbol{\alpha}$ through the conditional margins, we hide this dependency for notational clarity. Because the joint maximization of (2.11) is seldom feasible, we define

$$\widehat{\boldsymbol{\alpha}} = \operatorname*{argmax}_{\boldsymbol{\alpha} \in \mathbb{A}} \sum_{i=1}^2 \ell_i(\boldsymbol{\alpha}_i, \boldsymbol{\lambda}_i).$$

Using $\widehat{\mathbf{u}}_i = F_{\mathbf{Y}_i|\mathbf{X}}\{\mathbf{y}_i; \widehat{\alpha}_i(\mathbf{x})\}$ in $\ell_c(\boldsymbol{\theta})$, the two-step estimator $\widehat{\boldsymbol{\theta}}$ is the same as in (2.9). Similarly as in Joe (2005), we define

$$\mathbf{g}(\mathbf{w}; \boldsymbol{\alpha}, \boldsymbol{\theta}) = \begin{pmatrix} \mathbf{g}_1(\mathbf{w}; \boldsymbol{\alpha}_1) \\ \mathbf{g}_2(\mathbf{w}; \boldsymbol{\alpha}_2) \\ \mathbf{g}_c(\mathbf{z}; \boldsymbol{\alpha}, \boldsymbol{\theta}) \end{pmatrix} \quad \text{with} \quad \begin{cases} \mathbf{g}_i(\mathbf{w}; \boldsymbol{\alpha}_i) = \partial \ell_i(\mathbf{y}_i, \mathbf{x}; \boldsymbol{\alpha}_i)/\partial \boldsymbol{\alpha}_i \\ \mathbf{g}_c(\mathbf{z}; \boldsymbol{\alpha}, \boldsymbol{\theta}) = \partial \ell_c(\mathbf{u}, \mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\theta})/\partial \boldsymbol{\theta} \end{cases}.$$

Hence, $(\widehat{\boldsymbol{\alpha}}^\top, \widehat{\boldsymbol{\theta}}^\top)^\top$, which solves

$$n^{-1}\sum_{j=1}^n \mathbf{g}(\mathbf{w}^j; \boldsymbol{\alpha}, \boldsymbol{\theta}) - \begin{pmatrix} \mathbf{p}_1(\boldsymbol{\lambda}_1)\boldsymbol{\alpha}_1 \\ \mathbf{p}_2(\boldsymbol{\lambda}_2)\boldsymbol{\alpha}_2 \\ \mathbf{p}(\boldsymbol{\gamma})\boldsymbol{\theta} \end{pmatrix} = \mathbf{0},$$

can be viewed as a $\sqrt{n}$-consistent and asymptotically normal Generalized Method of Moment (GMM) estimator with an identity weighting matrix (see Newey and McFadden 1994, Chapter 6). This is the case provided that both penalties vanish with a $\sqrt{n}$-rate, the usual regularity conditions hold for the joint conditional distribution and the model is jointly identifiable.

Theorem 2.2.2. *Asymptotic properties of the two-step procedure*
If Assumptions 1–5 from Appendix A.2 hold, then $\left(\widehat{\boldsymbol{\alpha}}^{\top}, \widehat{\boldsymbol{\theta}}^{\top}\right)^{\top}$ is $\sqrt{n}$-consistent with

$$\sqrt{n}\begin{pmatrix}\widehat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0 \\ \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\end{pmatrix} \xrightarrow{d} N\left\{0, \mathbf{j}(\boldsymbol{\alpha}_0, \boldsymbol{\theta}_0)^{-1}\mathbf{v}(\boldsymbol{\alpha}_0, \boldsymbol{\theta}_0)\mathbf{j}(\boldsymbol{\alpha}_0, \boldsymbol{\theta}_0)^{-\top}\right\},$$

where

$$\mathbf{j}(\boldsymbol{\alpha}, \boldsymbol{\theta}) = E\begin{pmatrix}\partial\mathbf{g}_1/\partial\boldsymbol{\alpha}_1^{\top} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \partial\mathbf{g}_2/\partial\boldsymbol{\alpha}_2^{\top} & \mathbf{0} \\ \partial\mathbf{g}_c/\partial\boldsymbol{\alpha}_1^{\top} & \partial\mathbf{g}_c/\partial\boldsymbol{\alpha}_2^{\top} & \partial\mathbf{g}_c/\partial\boldsymbol{\theta}^{\top}\end{pmatrix},$$

$$and \ \mathbf{v}(\boldsymbol{\alpha}, \boldsymbol{\theta}) = E\begin{pmatrix}\mathbf{g}_1\mathbf{g}_1^{\top} & \mathbf{g}_1\mathbf{g}_2^{\top} & \mathbf{0} \\ \mathbf{g}_2\mathbf{g}_1^{\top} & \mathbf{g}_2\mathbf{g}_2^{\top} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{g}_c\mathbf{g}_c^{\top}\end{pmatrix}.$$

Moreover, we have

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N\left\{0, \mathbf{i}(\boldsymbol{\alpha}_0, \boldsymbol{\theta}_0)^{-1}\boldsymbol{\omega}(\boldsymbol{\alpha}_0, \boldsymbol{\theta}_0)\mathbf{i}(\boldsymbol{\alpha}_0, \boldsymbol{\theta}_0)^{-\top}\right\},$$

where $\mathbf{i}(\boldsymbol{\alpha}, \boldsymbol{\theta}) = E\left(\partial\mathbf{g}_c/\partial\boldsymbol{\theta}^{\top}\right)$ and

$$\boldsymbol{\omega}(\boldsymbol{\alpha}, \boldsymbol{\theta}) = cov\left\{\mathbf{g}_c - \begin{pmatrix}\mathbf{j}(\boldsymbol{\alpha}, \boldsymbol{\theta})_{c1} & \mathbf{j}(\boldsymbol{\alpha}, \boldsymbol{\theta})_{c2}\end{pmatrix}\begin{pmatrix}\mathbf{j}(\boldsymbol{\alpha}, \boldsymbol{\theta})_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{j}(\boldsymbol{\alpha}, \boldsymbol{\theta})_{22}\end{pmatrix}^{-1}\begin{pmatrix}\mathbf{g}_1 \\ \mathbf{g}_2\end{pmatrix}\right\}.$$

The detailed assumptions and the proof are provided in Appendix A.2.

Provided that the various analytical expressions exist, the covariance matrices are consistently estimated using the sample average. Hence, the two theorems can be used to construct confidence intervals for the dependence measure or the copula parameter. Using the asymptotic distribution of $\widehat{\boldsymbol{\theta}}$, sampling a large number of parameter realizations is computationally straightforward. Then combining (2.7) and parameter realizations samples any derived quantity, which is much faster than bootstrap techniques.

The Wald, Lagrange multiplier, and likelihood-ratio tests apply directly to the situation where the margins are known. Viewing the two-step procedure as a GMM estimator allows for straightforward extensions of the usual large sample theory of hypothesis testing (see Newey and McFadden 1994, Chapter 9).

2.2.3   Penalized Maximum Likelihood Estimation As an Iteratively Reweighted Ridge Regression

In this section, we describe the estimation procedure, recasting the penalized maximum likelihood estimation into an iteratively reweighted ridge regression problem. The idea to use iteratively reweighted least squares for maximum likelihood estimation was first proposed in Green (1984). It was then extended to iteratively reweighted ridge regression for penalized maximum likelihood estimation in the context of exponential families in O'Sullivan et al. (1986). Finally, it appeared in the general setting that we use in Green (1987). This reformulation is particularly convenient, because algorithms solving the problem that appears at each iteration are implemented in standard software packages.

Let $\boldsymbol{\psi}(\boldsymbol{\theta})$ be the $n \times 1$ vector with $\boldsymbol{\psi}(\boldsymbol{\theta})_i = \psi(\mathbf{x}_i; \boldsymbol{\theta})$, $\mathbf{d}(\boldsymbol{\theta})$ the $n \times d$ matrix $\partial \boldsymbol{\psi}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$ and $\mathbf{q}(\boldsymbol{\theta})$ the $n \times 1$ vector $\partial \ell(\boldsymbol{\theta}, \boldsymbol{\gamma}) / \partial \boldsymbol{\psi}$. The penalized maximum log-likelihood estimator $\widehat{\boldsymbol{\theta}}$ satisfies d score equations

$$\partial \ell(\widehat{\boldsymbol{\theta}}, \boldsymbol{\gamma}) / \partial \boldsymbol{\theta} = \mathbf{d}(\widehat{\boldsymbol{\theta}})^{\top} \mathbf{q}(\widehat{\boldsymbol{\theta}}) - \mathbf{p}(\boldsymbol{\gamma}) \widehat{\boldsymbol{\theta}} = 0.$$

To obtain $\widehat{\boldsymbol{\theta}}$, $\boldsymbol{\theta}^{[l]}$, the lth estimate of $\boldsymbol{\theta}_0$, can be updated by Newton–Raphson:

$$\{\mathbf{l}(\boldsymbol{\theta}^{[l]}) + \mathbf{p}(\boldsymbol{\gamma})\}(\boldsymbol{\theta}^{[l+1]} - \boldsymbol{\theta}^{[l]}) = \mathbf{d}(\boldsymbol{\theta}^{[l]})^{\top} \mathbf{q}(\boldsymbol{\theta}^{[l]}) - \mathbf{p}(\boldsymbol{\gamma}) \boldsymbol{\theta}^{[l]}. \qquad (2.12)$$

Notice that

$$\mathbf{l}(\boldsymbol{\theta}) = -\partial^2 \ell(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \boldsymbol{\theta}^{\top} = \mathbf{d}^{\top}(\boldsymbol{\theta}) \mathbf{a}(\boldsymbol{\theta}) \mathbf{d}(\boldsymbol{\theta}) - \sum_{j=1}^{n} \mathbf{q}(\boldsymbol{\theta})_j \partial^2 \boldsymbol{\psi}(\boldsymbol{\theta})_j / \partial \boldsymbol{\theta} \boldsymbol{\theta}^{\top}, \qquad (2.13)$$

where $\mathbf{a}(\boldsymbol{\theta})$ is the $n \times n$ matrix with $-\partial^2 \ell(\boldsymbol{\theta}, \boldsymbol{\gamma}) / \partial \boldsymbol{\psi} \boldsymbol{\psi}^{\top}$.

Remark. When the margins are known, the observations are independent and $\mathbf{a}(\boldsymbol{\theta})$ is diagonal. Whereas this is seldom verified in practice, we make it a working assumption to carry out the estimation, similar as using weighted least squares instead of generalized least squares.

By standard arguments, when $\boldsymbol{\theta}^{[l]}$ is close to $\boldsymbol{\theta}_0$, then $E\left\{\mathbf{q}(\boldsymbol{\theta}^{[l]})\right\} \approx 0$. Using this approximation, we simplify the right-hand side of (2.13) by $\mathbf{l}(\boldsymbol{\theta}^{[l]}) \approx \mathbf{d}(\boldsymbol{\theta}^{[l]})^\top \mathbf{a}(\boldsymbol{\theta}^{[l]})\mathbf{d}(\boldsymbol{\theta}^{[l]})$. Plugging this expression in (2.12), we have

$$\left\{\mathbf{d}(\boldsymbol{\theta}^{[l]})^\top \mathbf{a}(\boldsymbol{\theta}^{[l]})\mathbf{d}(\boldsymbol{\theta}^{[l]}) + \mathbf{p}(\boldsymbol{\gamma})\right\}\left(\boldsymbol{\theta}^{[l+1]} - \boldsymbol{\theta}^{[l]}\right) = \mathbf{d}(\boldsymbol{\theta}^{[l]})^\top \mathbf{q}(\boldsymbol{\theta}^{[l]}) - \mathbf{p}(\boldsymbol{\gamma})\boldsymbol{\theta}^{[l]}. \quad (2.14)$$

Equivalently, (2.14) can be rewritten as

$$\left\{\mathbf{d}(\boldsymbol{\theta}^{[l]})^\top \mathbf{a}(\boldsymbol{\theta}^{[l]})\mathbf{d}(\boldsymbol{\theta}^{[l]}) + \mathbf{p}(\boldsymbol{\gamma})\right\}\boldsymbol{\theta}^{[l+1]} = \mathbf{d}(\boldsymbol{\theta}^{[l]})^\top \mathbf{a}(\boldsymbol{\theta}^{[l]})\mathbf{y}^{[l]}, \quad (2.15)$$

where $\mathbf{y}^{[l]} = \mathbf{a}(\boldsymbol{\theta}^{[l]})^{-1}\mathbf{q}(\boldsymbol{\theta}^{[l]}) + \mathbf{d}(\boldsymbol{\theta}^{[l]})\boldsymbol{\theta}^{[l]}$ are the "pseudodata" at the lth iteration. When $\mathbf{a}(\boldsymbol{\theta}^{[l]})$ is positive-definite, then (2.15) has the form of normal equations for a generalized ridge regression:

$$\boldsymbol{\theta}^{[l+1]} = \underset{\boldsymbol{\theta}\,\in\,\Theta}{\operatorname{argmin}}\left\{\|\mathbf{y}^{[l]} - \mathbf{d}(\boldsymbol{\theta}^{[l]})\boldsymbol{\theta}\|_{\mathbf{a}(\boldsymbol{\theta}^{[l]})}^2 + \|\boldsymbol{\theta}\|_{\mathbf{p}(\boldsymbol{\gamma})}^2\right\}, \quad (2.16)$$

where $\|\mathbf{x}\|_{\mathbf{W}}^2 = \mathbf{x}^\top \mathbf{w}\mathbf{x}$. In other words, $\boldsymbol{\theta}^{[l+1]}$ results from ridge regressing $\mathbf{y}^{[l]}$ on the columns of $\mathbf{d}(\boldsymbol{\theta}^{[l]})$ with weight matrix $\mathbf{a}(\boldsymbol{\theta}^{[l]})$ and penalty $\mathbf{p}(\boldsymbol{\gamma})$, and the procedure is iterated until convergence. Alternatively, Fisher's scoring technique is obtained by replacing $\mathbf{a}(\boldsymbol{\theta}^{[l]})$ by its expectation, that is the $n \times n$ matrix $E\left\{-\partial^2\ell(\boldsymbol{\theta}, \boldsymbol{\gamma})/\partial\boldsymbol{\psi}\boldsymbol{\psi}^\top\right\}$. This is especially useful whenever the observed information is not positive-definite.

Remark. As noted in Green (1987), convergence is not guaranteed or may be very slow for an arbitrary penalized maximum likelihood. In the framework of this thesis, this is not an issue: convergence for common parametric copulas occurs in a few steps even when the model is misspecified.

2.2.4   Smoothing Parameters Selection and Equivalent Degrees of Freedom

In Section 2.2.2, we mention that smoothers are useful for exploratory data analysis. For instance, when no theory guides the functional form of interest, there is seldom reason to assume a finite dimensional basis-quadratic representation. As a matter of fact, roughness penalty methods are often applied the other way: the dimension of the parameter space is arbitrarily augmented and its effective size controlled via the penalty.

Let $\mathbf{s}^{[l]}(\boldsymbol{\gamma})$ be the so-called influence or hat matrix at the lth iteration, defined such that $\mathbf{s}^{[l]}(\boldsymbol{\gamma})\mathbf{y}^{[l]} = \mathbf{d}(\boldsymbol{\theta}^{[l]})\boldsymbol{\theta}^{[l+1]}$, that is

$$\mathbf{s}^{[l]}(\boldsymbol{\gamma}) = \mathbf{d}(\boldsymbol{\theta}^{[l]}) \left\{ \mathbf{d}^{\top}(\boldsymbol{\theta}^{[l]})\mathbf{a}(\boldsymbol{\theta}^{[l]})\mathbf{d}(\boldsymbol{\theta}^{[l]}) + \mathbf{p}(\boldsymbol{\gamma}) \right\}^{-1} \mathbf{d}(\boldsymbol{\theta}^{[l]})^{\top}\mathbf{a}(\boldsymbol{\theta}^{[l]}).$$

Similarly as in Section 2.1.3, we can now define the effective or equivalent degrees of freedom at the lth iteration as

$$EDF^{[l]}(\boldsymbol{\gamma}) = \mathrm{tr}\left\{ \mathbf{s}^{[l]}(\boldsymbol{\gamma}) \right\}.$$

Remark. When the model contains only the parametric part (respectively when $\boldsymbol{\gamma} = 0$), the interpretation is straightforward: $EDF^{[l]}(\boldsymbol{\gamma})$ equals the dimension of the parameter space, that is p (respectively $d = p + \sum_{k=1}^{K} m_k$). More generally, we see that $0 \leq \mathbf{s}^{[l]}(\boldsymbol{\gamma})_{jj} \leq 1$ for $1 \leq j \leq d$. Hence the EDF of each smooth component h_k is the sum of the corresponding trace elements, smaller than or equal to m_k.

Following Section 2.1.3 again, we can now select $\boldsymbol{\gamma}$ by minimizing the GCV, that we define as

$$GCV^{[l]}(\boldsymbol{\gamma}) = \frac{n^{-1}\|\mathbf{y}^{[l]} - \mathbf{s}^{[l]}(\boldsymbol{\gamma})\mathbf{y}^{[l]}\|^2_{\mathbf{a}(\boldsymbol{\theta}^{[l]})}}{\left\{1 - n^{-1}EDF^{[l]}(\boldsymbol{\gamma})\right\}^2}.$$

Since the minimization is carried out for each generalized ridge iteration, computational stability and efficiency to calculate $GCV^{[l]}(\boldsymbol{\gamma})$ are of utmost importance (see Gu and Wahba 1991, Wood 2000 and Wood 2004 for details). In the remainder of this Chapter, the EDF of a model is the final $EDF^{[l]}(\boldsymbol{\gamma})$, obtained at convergence.

Remark. When GAMs are applied to the mean of an exponential family, the numerator at the optimum corresponds to the deviance (Hastie and Tibshirani 1990; Green and Silverman 2000, Section 5.4.3.). In the present context, it is only a first order approximation[4], namely a "linearized deviance" (a term coined in Green 1987).

---

[4]Assume that the notion of "saturated model" is well defined in the sense that the likelihood, as a function of ψ (freed from the dependence on $\boldsymbol{\theta}$), is uniquely maximized at $\widehat{\psi}$. At the optimum, the numerator is a first order approximation of $2\{\ell(\widehat{\psi}) - \ell(\widehat{\boldsymbol{\theta}}, \boldsymbol{\gamma})\}$.

2.3 Simulations and Application

2.3.1 Simulation Study

In order to assess the numerical stability of the procedure described in Sections 2.2.3 and 2.2.4, we proceed to a simulation study that we present in this section.

    We assume that the true underlying model is either a Gaussian, a Student's t with 4 degrees of freedom, a Clayton or a Gumbel copula with time-varying Kendall's tau. To encompass different cases of practical interest, we use three different deterministic functions

$$\left.\begin{array}{l} h_1(x) = a + b(x - x_0)^2 \\ h_2(x) = a + b\sin\left\{2\pi c\,(x - x_0)\right\} \\ h_3(x) = a + be^{-\frac{(x-x_0)^2}{2\sigma^2}} \end{array}\right\} \psi(\mathbf{x}) = g\left\{h_0 + \sum_{k=1}^{3} h_k(x_k)\right\}, \qquad (2.17)$$

with $a, b, x_0 \in \mathbb{R}$, $c, \sigma^2 \in \mathbb{R}_+$, $g(x) = (e^x - 1)/(e^x + 1)$.

A single deterministic covariate and unpenalized splines

We sample n observations in four steps:

1. Define a grid of times $t_1 < \cdots < t_n$ with linear spacing.

2. Compute Kendall's tau $\{\psi_k(t_j)\}_{j\in\{1,\ldots,n\}}^{k\in\{1,2,3\}}$ with $\psi_k(t) = g\{h_k(t)\}$.

3. Recover the copula parameter $\{\eta_k(t_j)\}_{j\in\{1,\ldots,n\}}^{k\in\{1,2,3\}}$ using the mapping $\eta(t) = \nu\{\psi(t)\}$.

4. Sample the corresponding bivariate pseudo-observation $\left\{\mathbf{u}_k^j\right\}_{j\in\{1,\ldots,n\}}^{k\in\{1,2,3\}}$.

    Using $n = 1000$, we set the various constants to values such that $\psi_1(t) \in [0.2, 0.80]$, $\psi_2(t) \in [0.2, 0.46]$ and $\psi_3(t) \in [0.2, 0.83]$ for all $t \in [0, 60]$. The data are represented in Figure 2.7 for the Gaussian copula.

    The left panels are related to the quadratic form in (2.17), the middle panels to the sinusoidal form and the right panels to the exponential form. In the top row, we represent the time-varying Kendall's tau for each deterministic function in (2.17). In the bottom row, we plot a corresponding sample of pseudo-observation: some linear dependence is visible but the time-varying feature is not. This illustrates why it is usually impossible to formulate a parametric hypothesis without prior knowledge of the data-generating process.
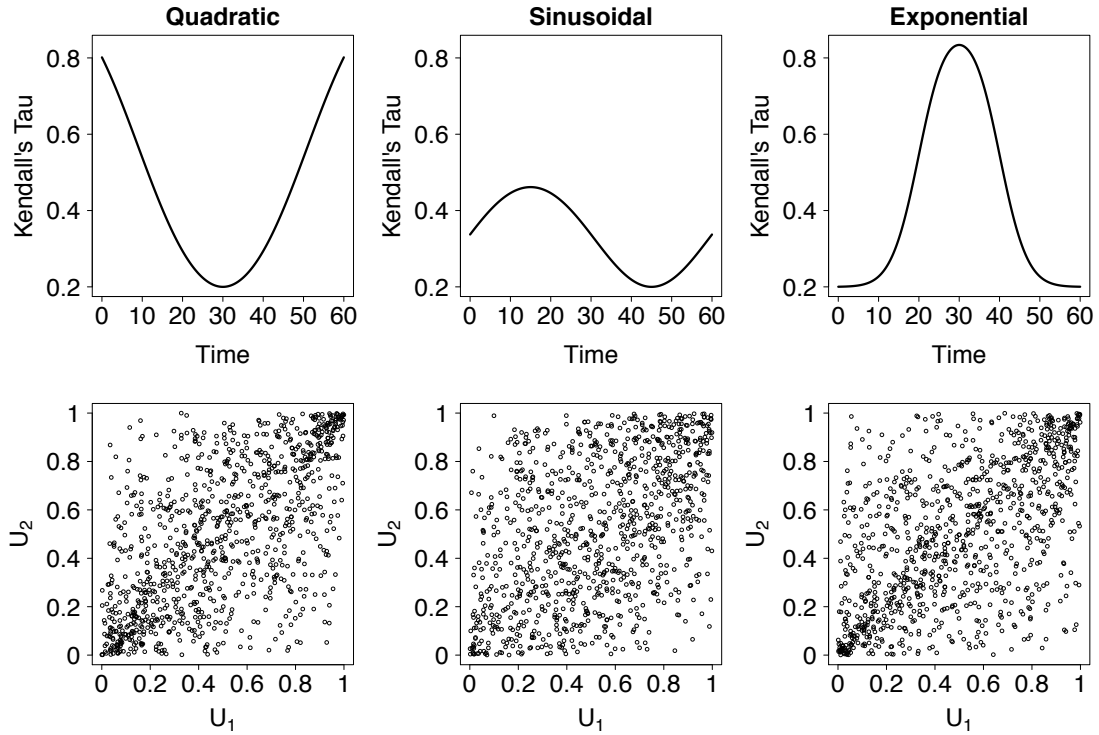
Fig. 2.7 **Gaussian copula with time-varying dependence:** Kendall's tau $\{\psi_k(t_j)\}_{j\in\{1,\dots,n\}}$ with $t_1 = 0$, $t_n = 60$ and $n = 1000$ for the quadratic ($k = 1$, left panel), sinusoidal ($k = 2$, middle panel) and exponential ($k = 3$, right panel) specification is in the top row. The corresponding pseudo-observations $\{(u_{kj,1}, u_{kj,2})\}_{j\in\{1,\dots,n\}}^{k\in\{1,2,3\}}$ are in the bottom row.

To approximate each underlying function $h_k(t)$, we use natural cubic splines (NCS) for $k \in \{1, 3\}$ and periodic cubic splines for $k = 2$ with homogeneously spaced knots (Valenzuela et al. 2013). Therefore, the models are not nested when changing the basis dimension and we use the Aikaike Information Criterion (AIC) to select $m_1 = 5$, $m_2 = 4$ and $m_3 = 7$ interior knots. We also estimate the quadratic specification without splines to benchmark the parametric and nonparametric estimation.

Remark. The three deterministic functions cannot be represented using cubic splines with a finite number of knots. Therefore we strictly study the misspecified case, where each function is projected on the corresponding spline basis.

Remark. While the Student's t copula has two parameters, Kendall's tau is a function of the correlation parameter only with mapping $\nu(\psi) = \sin\left(\frac{\pi}{2}\psi\right)$ like the Gaussian copula. We therefore treat the Student's t copula's degrees of freedom as a nuisance parameter that we update after each generalized ridge iteration. As the Student's t copula's Fisher information matrix is not diagonal, this procedure may affect the estimated

Kendall's tau or lead to non-convergence. An orthogonal reparametrization as in Cox and Reid (1987) is not conceivable because their suggested differential equation has no solution when the correlation parameter belongs to $(-1, 1)$.

We simulate 500 datasets of 1000 observations for each deterministic function. For each dataset, we repeat the (Fisher's scoring) estimation using either the parametric specification or the splines with m_1, m_2 and m_3 interior knots. We show the results in Figure 2.8 and Figure 2.9. In the top row for each copula, we observe that the mean estimate, $\widehat{E}\left\{\widehat{h}_k(t)\right\}$, is visually indistinguishable from the true curve, $h_k(t)$. The bootstrapped 95% confidence intervals (c.i.) are fairly narrow, except near the border for the exponential specification. In the middle row, the estimated bias is compared to the true model bias, which is simply the difference between $h_k(t)$ and its projection on a spline basis. Although numerical issues appear in some cases, the estimated biases closely follow their true counterpart. Finally, we show $\sqrt{n}$ times the variance in the bottom row for each copula. As is usual with many smoothers, the variance near the boundaries is the highest, except for cyclic cubic splines (using periodic boundary conditions).

In the Student's t copula case, the algorithm converges exactly as for the other copulas. Therefore, the non-orthogonality does not seem to be a practical issue here. Globally, the simulation study provides evidence of an accurate estimation procedure.

Three random covariates and penalized splines

We sample n observations in four steps:

1. Simulate 3-dimensional covariates $\{\mathbf{x}^j\}_{j\in\{1,\dots,n\}}$ with an equicorrelation Gaussian copula.

2. Compute Kendall's tau $\{\psi(\mathbf{x}^j)\}_{j\in\{1,\dots,n\}}$ using (2.17).

3. Recover the corresponding copula parameters.

4. Sample the corresponding bivariate pseudo-observation.

To ensure identifiability, we set the constant a such that $\int_0^1 h_k(t)dt = 0$ for each component $k \in \{1, 2, 3\}$. We then build a grid using $n \in \{200, 1000\}$ and $\rho \in \{0, 0.5, 0.9\}$ the correlation between the covariates . Furthermore, we also include the parametrization directly in term of the copula parameter. When using this "natural" parametrization, the mapping ν is the identity and the link g constraints the parameter space with $\eta(\mathbf{x}) = g\left\{h_0 + \sum_{k=1}^3 h_k(x_k)\right\}$.

We simulate 1000 datasets for the various combinations and repeat the estimation using either the Newton-Raphson (from (2.16)) or the Fisher's scoring method (replacing the observed information by its expectation). At each iteration of both algorithms, a minimization of the generalized cross-validation sum of squares is carried with $m_1 = 15$, $m_2 = 20$ and $m_3 = 25$ interior knots.

We present an example of such simulation in Figure 2.10, where the Clayton copula is parametrized using the Kendall's tau and estimated using Fisher's scoring method. Furthermore, the covariates are simulated using a small sample size ($n = 200$) and no correlation or with a large sample size ($n = 1000$) and a correlation of 0.9. We observe that the bias does not change with the sample size; whereas the width of the confidence intervals as well as the variance at the boundaries decrease.

To compare the results quantitatively, we use two criterion, namely the average integrated squared error (AISE)

$$\widehat{E}\left[\int_0^1 \left\{\widehat{h}_k(x) - h_k(x)\right\}^2 dx\right]$$

and the the average integrated absolute error (AIAE)

$$\widehat{E}\left\{\int_0^1 \mid \widehat{h}_k(x) - h_k(x) \mid dx\right\},$$

where the expectation is taken over all simulations for each component $k \in \{1, 2, 3\}$. The results presented in Table 2.1 and 2.2 can be summarized as follows:

- An intermediate value of the correlation ($\rho = 0.5$) does not seem to affect the estimation procedure, as the results are almost equal to the uncorrelated ones. However, for highly correlated covariates ($\rho = 0.9$), the estimation becomes more difficult, a fact consistently translated by higher values of the two criterions.

- As expected, only the average integrated squared error decreases with an increase in sample size. Furthermore, as previously seen in Figure 2.10, neither the bias nor the average integrated absolute error are affected.

- The simplified Newton-Raphson seems to perform slightly better than Fisher's scoring method. However, it is also slower, because the evaluation of the likelihood's second derivative is more demanding than its expectation, which is in closed form.

A remarkable feature when comparing the four pannels in Table 2.1 and 2.2 is the stability of the estimation across copula families. It indicates that the formulation of

the penalized maximum likelihood estimation as a generalized ridge regression problem (see Section 2.2.3) is appropriate for commonly used parametric copulas. Interestingly, the parametrization with Kendall's tau is numerically more stable than the natural one, but the results are qualitatively similar.

Remark. The number and position of the knots, which are spline/GAM related problems and outside the scope of this work, are deliberately not studied here.

2.3.2   The Foreign Exchange Market

In this section, we model the cross-sectional dynamics of asset returns, an important topic in finance and econometrics. More specifically, we study the foreign exchange (FX) market, which determines the relative value of currencies. This decentralized market has two main characteristics. First, it operates around the clock from Sunday 10pm to Friday 10pm UTC. Second, it is geographically dispersed because the main trading centers continuously accommodate various types of buyers and sellers.

In what follows, we use data graciously provided by Dukascopy Bank SA[5], an electronic broker holding a Securities Dealer License issued by the FINMA. It contains 15-minute spaced prices (i.e., 96 observations each day) for the EUR/USD and the USD/CHF from March 10, 2013 to November 1, 2013. Hence, in a total of 34 trading weeks, there are 16320 observations (170 days) excluding weekends. In Figure 2.11, we show the logarithm of the price $p_t = \log P_t$ and the log-return (or simply return) $r_t = p_t - p_{t-1}$.

Because intraday asset returns are heteroskedastic, we need to pre-filter the individual series before applying the methodology of this Chapter. A general procedure used in the high-frequency econometrics literature (see e.g., Andersen and Bollerslev 1997, 1998; Engle and Sokalska 2012; Vatter et al. 2015) consists of two steps:

1. filter the seasonality (i.e., intraday patterns due to the cyclical nature of market activity),

2. use a non-periodic model (e.g., from the GARCH family) to remove the residual heteroskedasticity.

For the first step, a popular model is the Fourier Flexible Form (FFF, see Gallant 1981), introduced in this context by Andersen and Bollerslev (1997, 1998). A simplified

---

[5]http://www.dukascopy.com/

Gaussian copula			$\rho = 0$	$\rho = 0.5$	$\rho = 0.9$
NR	$n = 200$	AISE	0.03 0.04 0.06	0.03 0.05 0.06	0.11 0.12 0.13
		AIAE	0.05 0.07 0.08	0.06 0.07 0.08	0.10 0.11 0.11
	$n = 1000$	AISE	0.01 0.01 0.01	0.01 0.01 0.01	0.02 0.02 0.02
		AIAE	0.05 0.07 0.08	0.06 0.07 0.08	0.10 0.11 0.11
FS	$n = 200$	AISE	0.03 0.05 0.06	0.04 0.06 0.07	0.11 0.13 0.13
		AIAE	0.13 0.16 0.18	0.14 0.17 0.19	0.23 0.26 0.27
	$n = 1000$	AISE	0.01 0.01 0.01	0.01 0.01 0.01	0.02 0.02 0.02
		AIAE	0.05 0.06 0.07	0.06 0.07 0.08	0.09 0.10 0.10

t copula			$\rho = 0$	$\rho = 0.5$	$\rho = 0.9$
NR	$n = 200$	AISE	0.03 0.06 0.07	0.04 0.06 0.07	0.15 0.16 0.18
		AIAE	0.06 0.08 0.10	0.07 0.09 0.10	0.12 0.13 0.14
	$n = 1000$	AISE	0.01 0.01 0.02	0.01 0.01 0.02	0.03 0.03 0.03
		AIAE	0.06 0.08 0.10	0.07 0.09 0.10	0.12 0.13 0.14
FS	$n = 200$	AISE	0.04 0.06 0.08	0.05 0.07 0.09	0.15 0.18 0.18
		AIAE	0.15 0.18 0.21	0.16 0.19 0.22	0.28 0.31 0.31
	$n = 1000$	AISE	0.01 0.01 0.01	0.01 0.01 0.01	0.02 0.03 0.03
		AIAE	0.06 0.08 0.09	0.07 0.08 0.09	0.11 0.12 0.12

Clayton copula			$\rho = 0$	$\rho = 0.5$	$\rho = 0.9$
NR	$n = 200$	AISE	0.03 0.05 0.06	0.04 0.06 0.06	0.13 0.12 0.15
		AIAE	0.07 0.09 0.10	0.07 0.09 0.10	0.12 0.12 0.14
	$n = 1000$	AISE	0.02 0.01 0.02	0.01 0.01 0.02	0.03 0.03 0.03
		AIAE	0.07 0.09 0.10	0.07 0.09 0.10	0.12 0.12 0.14
FS	$n = 200$	AISE	0.06 0.07 0.09	0.06 0.08 0.09	0.14 0.16 0.18
		AIAE	0.16 0.19 0.21	0.17 0.20 0.22	0.27 0.29 0.31
	$n = 1000$	AISE	0.02 0.01 0.01	0.01 0.01 0.02	0.02 0.02 0.03
		AIAE	0.07 0.08 0.09	0.07 0.08 0.10	0.11 0.11 0.12

Gumbel copula			$\rho = 0$	$\rho = 0.5$	$\rho = 0.9$
NR	$n = 200$	AISE	0.03 0.06 0.06	0.04 0.06 0.07	0.15 0.15 0.18
		AIAE	0.06 0.09 0.09	0.07 0.09 0.10	0.13 0.13 0.15
	$n = 1000$	AISE	0.01 0.01 0.01	0.01 0.01 0.02	0.03 0.03 0.04
		AIAE	0.06 0.09 0.09	0.07 0.09 0.10	0.13 0.13 0.15
FS	$n = 200$	AISE	0.06 0.08 0.10	0.06 0.09 0.10	0.18 0.20 0.22
		AIAE	0.17 0.20 0.23	0.18 0.21 0.23	0.30 0.32 0.34
	$n = 1000$	AISE	0.01 0.01 0.02	0.01 0.01 0.02	0.03 0.03 0.03
		AIAE	0.07 0.09 0.10	0.08 0.09 0.10	0.12 0.12 0.14

Table 2.1 **Simulation results for the parametrization with Kendall's tau:** NR/FS stands for Newton-Raphson/Fisher-Scoring, n for the sample size, AISE/AIAE for the average integrated squared/absolute error and ρ for the correlation between the covariates. In each cell, the three numbers correspond to the three components of the additive model.

version of their regression equation is

$$\log r_t^2 = s_t + \epsilon_t, \; s_t = s_0 + \sum_{l=1}^{L} \left\{ a_l \cos\left(2\pi l t/T\right) + b_l \sin\left(2\pi l t/T\right) \right\},$$

where $T = 96$. It contains two terms: a constant trend s_0 and a sum of cosines and sines with integer frequencies, designed to capture daily oscillations around the base level. Denoting by $\widehat{s}_t$ the ordinary least-squares estimate with $L = 5$, we refer to $\widehat{r}_t = r_t e^{-\widehat{s}_t/2}$ as the deseasonalized return.

In the top row of Figure 2.12, we show estimated autocorrelation functions for the absolute return, the deseasonalized absolute return and 95% confidence bands. We observe that the absolute return's estimated autocorrelation peaks every 96th observation, a feature captured by the FFF. In the bottom row of Figure 2.12, we relate the estimated empirical volatility in each time-interval to the FFF estimate. As observed in Andersen and Bollerslev (1997, 1998), we can connect the peaks of volatility to trading activity.

For the second step, we find that an additional GARCH(1,1) filter is sufficient to proceed to copula-based modeling. Even after pre-filtering, the residuals are highly dependent with a linear correlation and Kendall's tau of respectively -0.82 and -0.60. Furthermore, as usual with financial returns, both the residuals and their empirical distributions are highly (upper and lower) tail dependent (cf. McNeil et al. 2005, Example 5.59, p.235). This last observation explains why, when fitting different parametric families in what follows, the AIC selects the Student's t copula as the most appropriate.

To model a time-varying dependence between the two currency pairs, we use two predictors. First, the time-of-day (hereafter x_1), normalized to lie between zero and one with a period of 96, is designed to capture the quickly (and periodic) time-varying characteristics. Second, the absolute time (hereafter x_2), expressed in trading hours elapsed since the first observation, is designed to capture the slowly time-varying characteristics. Model (2.7) for the Kendall's tau is written as

$$\psi(\mathbf{x}) = g\left\{\mathbf{z}^{\top}(x_1)\boldsymbol{\beta} + h_1(x_1) + h_2(x_2)\right\},$$

where

- $\mathbf{z}(x_1) = \left\{1, \cos\left(2\pi x_1\right), \sin\left(2\pi x_1\right), \ldots, \cos\left(2\pi L x_1\right), \sin\left(2\pi L x_1\right)\right\}^{\top}$ is the vector of linear predictors,

- $\boldsymbol{\beta} = \left(s_0, a_1, b_1, \ldots, a_L, b_L\right)^{\top}$ is the vector of parameters,

- $\{h_1, h_2\}$ are built using (initially non-penalized) natural cubic splines.

Remark. The parametric part $\mathbf{z}^\top(x_1)\boldsymbol{\beta}$ presupposes the same intraday patterns for the dependence as for the individual volatilities.

In the top left panel of Figure 2.13, we show the AIC of models containing either $\mathbf{z}^\top(x_1)\widehat{\boldsymbol{\beta}}$ or $\widehat{h}_1(x_1)$ as a function of the number of parameters: $2L$ for the FFF (i.e., L cosines and L sines) and m interior knots for the splines. The panel indicates that 10 is the optimal number in either case and that the parametric model beats the splines. In the top right panel of Figure 2.13, we show $\widehat{\psi}(x_1) = g\left\{\mathbf{z}^\top(x_1)\widehat{\boldsymbol{\beta}}\right\}$ using the FFF with 95% confidence intervals (either from parametric bootstrap[6] samples or from the one-step asymptotics). Note that, because we model time-series and use a three-step procedure, the theory developed in Section 2.2.2 does not strictly apply. However, the confidence intervals resulting from the parametric bootstrap are only slightly wider than the one-step asymptotics ones. As expected, the estimated shape closely reflects that of the volatility presented in Figure 2.12. While this FFF and the splines are redundant, they provide an important sanity check when estimated separately. In fact, the shape estimated using the splines (not reported here) is very similar to the FFF.

In the top middle panel of Figure 2.13, we show the AIC of models containing both the linear function $\mathbf{z}^\top(x_1)\widehat{\boldsymbol{\beta}}$ with $L = 5$ and $\widehat{h}_2(x_2)$ with an increasing basis size. As the dots indicate, the AIC plateaus around 150 interior knots. Alternatively, we set the basis size at 200 and use generalized cross-validation to select 141.74 effective degrees of freedom. In the middle bottom panel of Figure 2.13, we show $\widehat{\psi}(x_2) = g\left\{\widehat{h}_2(x_2)\right\}$ with 95% confidence intervals. Unlike the parametric component of the model, the bootstrapped mean differs to some extent from the estimate, which is suggestive of oversmoothing. This also explains why the boostrapped confidence intervals are slightly smaller than the ones resulting from the one-step asymptotics. In the bottom panel of Figure 2.13, we show the combined effects of the time-of-day and the absolute time, that is $\widehat{\psi}(\mathbf{x}) = g\left\{\mathbf{z}^\top(x_1)\widehat{\boldsymbol{\beta}} + \widehat{h}_2(x_2)\right\}$, for the first four weeks of the sample.

In financial econometrics, a growing literature on multivariate GARCH-type models indicates a wide acceptance of time-varying dependence between asset returns. However, our decomposition into periodic and slowly time-varying components is a new empirical finding itself. While the joint time-series analysis of seasonality and longer-term trend patterns is common, a similar interplay in the dependence structure is natural.

---

[6]In this case, confidence intervals are obtained by simulating from the fitted model 1'000 samples of the same size than the underlying data.

2.4   Discussion

In this Chapter, we introduce a general approach to model the influence of covariates on the dependence structure between two variables. Based on an extension of copula theory to conditional distributions, our framework benefits directly from the flexibility of generalized additive models (GAM). As usual for roughness penalty-based methods, we propose a maximum penalized log-likelihood estimator and derive useful asymptotic properties. We discuss details of an estimation procedure and assess its performance via simulations. Finally, we model the dependence structure of intraday returns on two exchange rates. Interestingly, the intraday pattern, due to the cyclical nature of market activity, is shaped similarly to the univariate conditional second moments.

Although first restricted to the single-parameter and bivariate case, the approach has a natural higher-dimensional extension in the single-parameter Archimedean family, but multi-parameters copula families can also be handled in (at least) three distinct ways. First, for the Student's t copula, we update the degrees of freedom after each generalized ridge iteration. This is a viable option whenever the dependence or concordance measure of interest is a function of only a single copula parameter. However, this procedure may lead to numerical instability and convergence should be carefully checked. Second, when an orthogonal reparametrization as in Cox and Reid (1987) is analytically (or numerically) tractable, then the GAM can either be specified element by element or some parameters may be considered as nuisances. Third, to model the parameters vector as a whole directly, we could consider a multi-dimensional smoothing framework, such as vector generalized additive models (VGAM, Yee and Wild 1996). Note that, for implicit copulas, as the number of parameters usually increases quickly with the dimension, additional complications arise.

Relying on bivariate building blocks only, so-called pair-copula constructions (see Bedford and Cooke 2001, 2002 or Aas et al. 2009) represent an appealing alternative, and this approach is the subject of the next chapter.
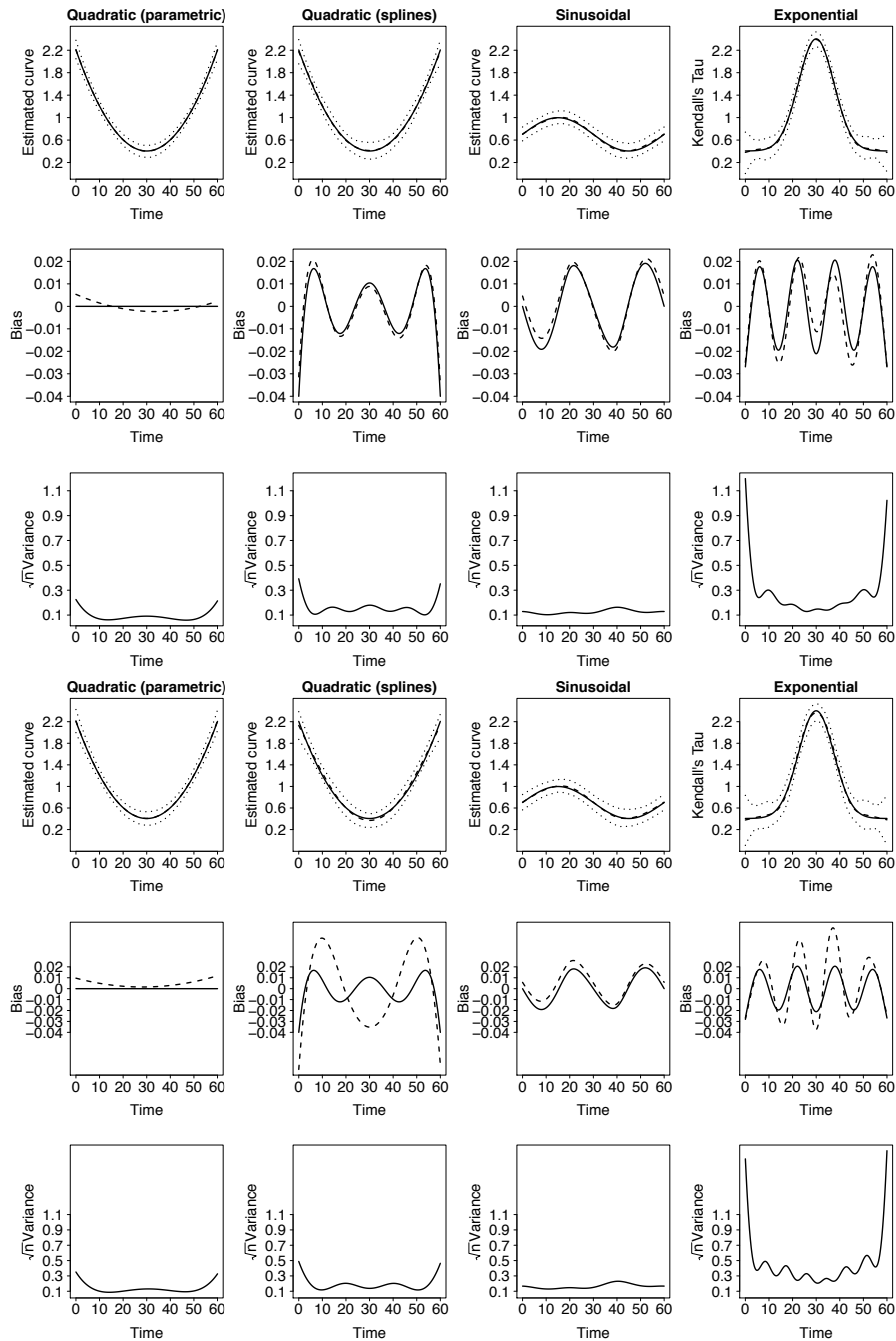
Fig. 2.8 **Simulation results for the Gaussian (top three rows) and Student's**
t **(bottom three rows) copulas:** the same row/column organization prevails for
both copulas. Top row: true curve (plain line), mean estimate (dashed line) and 95%
c.i. are the dotted lines with $k = 1$ estimated with the parametric model (left panel),
$k = 1$ estimated with splines (middle left panel), $k = 2$ (middle right panel) and $k = 3$
(right panel). Middle row: true model biases (plain line) and estimated model biases
(dashed lines). Bottom row: $\sqrt{n}$ times estimated variance.
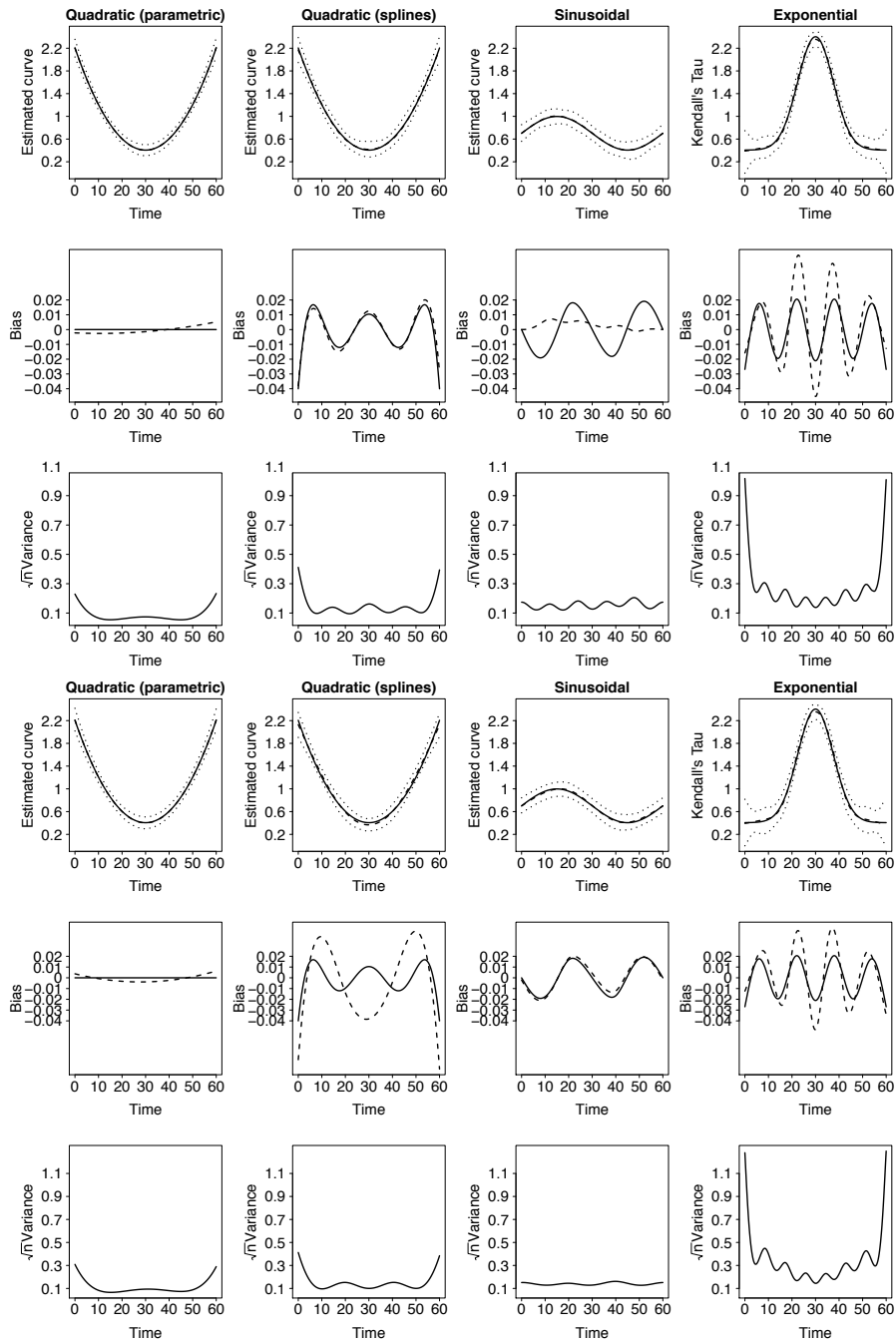
Fig. 2.9 **Simulation results for the Clayton (top three rows) and Gumbel (bottom three rows) copulas:** the same row/column organization prevails for both copulas. Top row: true curve (plain line), mean estimate (dashed line) and 95% c.i. are the dotted lines with $k = 1$ estimated with the parametric model (left panel), $k = 1$ estimated with splines (middle left panel), $k = 2$ (middle right panel) and $k = 3$ (right panel). Middle row: true model biases (plain line) and estimated model biases (dashed lines). Bottom row: $\sqrt{n}$ times estimated variance.
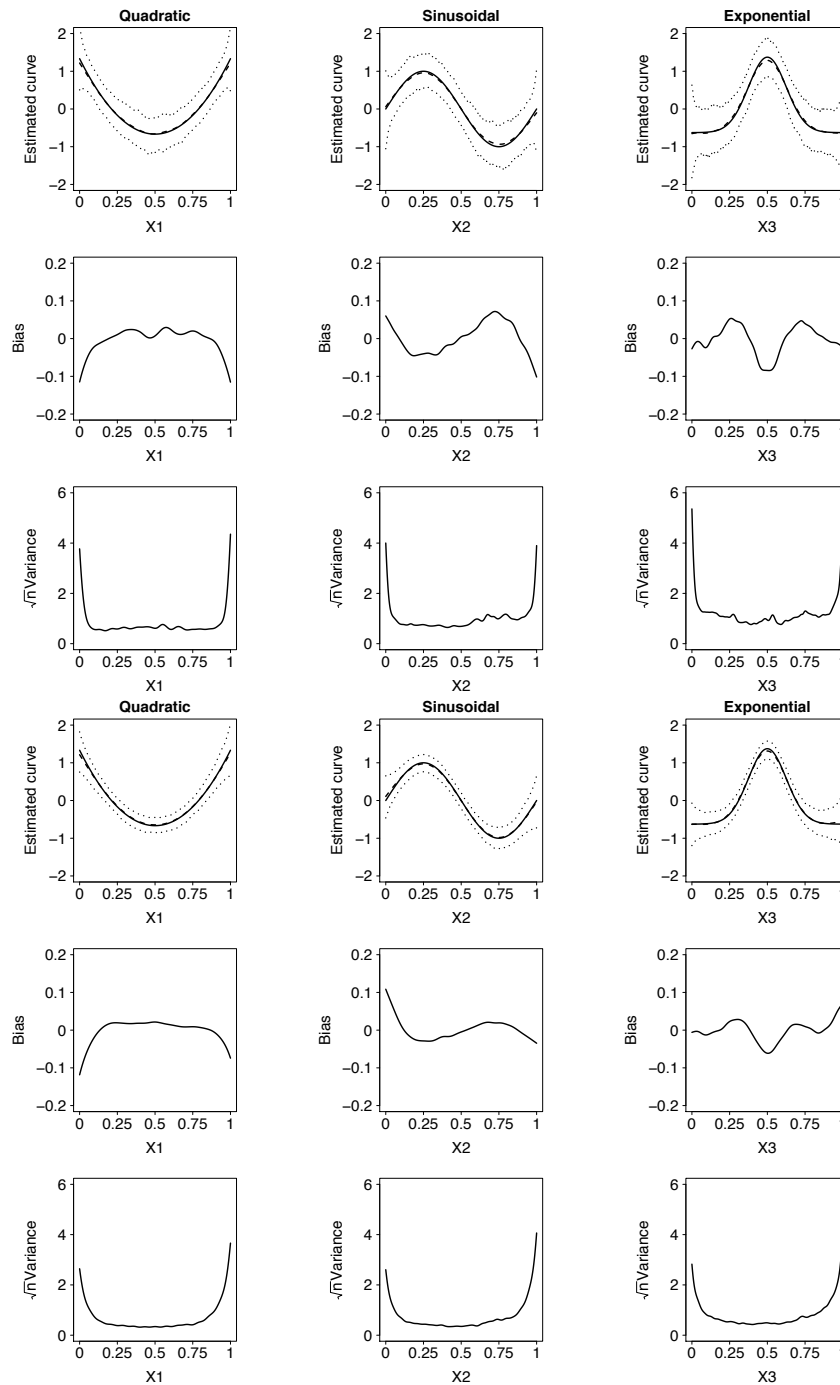
Fig. 2.10 **Clayton copula with three additive components between small sample size/no correlation (top three rows) and large sample size/correlation of 0.9 between the covariates (bottom three rows):** the same row/column organization prevails for both studies. Top row: true curve (plain line), mean estimate (dashed line) and 95% c.i. are the dotted lines with $k = 1$ (left panel), $k = 2$ (middle panel) and $k = 3$ (right panel). Middle row: estimated model biases. Bottom row: $\sqrt{n}$ times estimated variance.

Gaussian copula			$\rho = 0$	$\rho = 0.5$	$\rho = 0.9$
NR	$n = 200$	AISE	0.10 0.15 0.17	0.11 0.16 0.18	0.34 0.31 0.40
		AIAE	0.09 0.13 0.15	0.10 0.13 0.15	0.19 0.20 0.21
	$n = 1000$	AISE	0.02 0.03 0.04	0.02 0.03 0.04	0.07 0.07 0.08
		AIAE	0.09 0.13 0.15	0.10 0.13 0.15	0.19 0.20 0.21
FS	$n = 200$	AISE	0.09 0.15 0.18	0.12 0.17 0.19	0.35 0.38 0.40
		AIAE	0.22 0.28 0.31	0.24 0.29 0.32	0.43 0.45 0.47
	$n = 1000$	AISE	0.01 0.02 0.03	0.02 0.02 0.03	0.05 0.05 0.06
		AIAE	0.09 0.11 0.12	0.10 0.11 0.13	0.16 0.17 0.18

t copula			$\rho = 0$	$\rho = 0.5$	$\rho = 0.9$
NR	$n = 200$	AISE	0.12 0.17 0.21	0.15 0.19 0.23	0.41 0.37 0.50
		AIAE	0.11 0.15 0.18	0.13 0.16 0.19	0.22 0.23 0.25
	$n = 1000$	AISE	0.02 0.04 0.05	0.03 0.05 0.06	0.09 0.10 0.11
		AIAE	0.11 0.15 0.18	0.13 0.16 0.19	0.22 0.23 0.25
FS	$n = 200$	AISE	0.13 0.19 0.24	0.16 0.22 0.25	0.44 0.51 0.53
		AIAE	0.26 0.32 0.36	0.29 0.34 0.37	0.48 0.52 0.54
	$n = 1000$	AISE	0.02 0.03 0.03	0.02 0.03 0.04	0.06 0.07 0.08
		AIAE	0.10 0.13 0.14	0.11 0.13 0.15	0.19 0.20 0.21

Clayton copula			$\rho = 0$	$\rho = 0.5$	$\rho = 0.9$
NR	$n = 200$	AISE	0.13 0.18 0.21	0.15 0.17 0.22	0.37 0.25 0.42
		AIAE	0.11 0.16 0.16	0.12 0.16 0.17	0.24 0.20 0.26
	$n = 1000$	AISE	0.02 0.04 0.04	0.03 0.04 0.05	0.10 0.07 0.12
		AIAE	0.11 0.16 0.16	0.12 0.16 0.17	0.24 0.20 0.26
FS	$n = 200$	AISE	0.31 0.52 0.63	0.10 0.37 0.30	1.75 3.27 3.03
		AIAE	0.18 0.22 0.24	0.18 0.23 0.24	0.37 0.43 0.45
	$n = 1000$	AISE	0.01 0.01 0.01	0.01 0.01 0.02	0.03 0.03 0.03
		AIAE	0.07 0.08 0.09	0.07 0.08 0.09	0.12 0.13 0.14

Gumbel copula			$\rho = 0$	$\rho = 0.5$	$\rho = 0.9$
NR	$n = 200$	AISE	0.08 0.13 0.15	0.10 0.14 0.17	0.28 0.22 0.33
		AIAE	0.09 0.13 0.14	0.10 0.13 0.14	0.19 0.17 0.21
	$n = 1000$	AISE	0.02 0.03 0.04	0.02 0.03 0.03	0.07 0.05 0.08
		AIAE	0.09 0.13 0.14	0.10 0.13 0.14	0.19 0.17 0.21
FS	$n = 200$	AISE	0.04 0.06 0.07	0.04 0.06 0.07	1.50 2.36 1.17
		AIAE	0.14 0.18 0.19	0.15 0.18 0.20	0.32 0.36 0.35
	$n = 1000$	AISE	0.01 0.01 0.01	0.01 0.01 0.01	0.02 0.02 0.03
		AIAE	0.06 0.08 0.08	0.07 0.08 0.09	0.11 0.12 0.13

Table 2.2 **Simulation results for the natural parametrization:** NR/FS stands for Newton-Raphson/Fisher-Scoring, n for the sample size, AISE/AIAE for the average integrated squared/absolute error and ρ for the correlation between the covariates. In each cell, the three numbers correspond to the three components of the additive model.
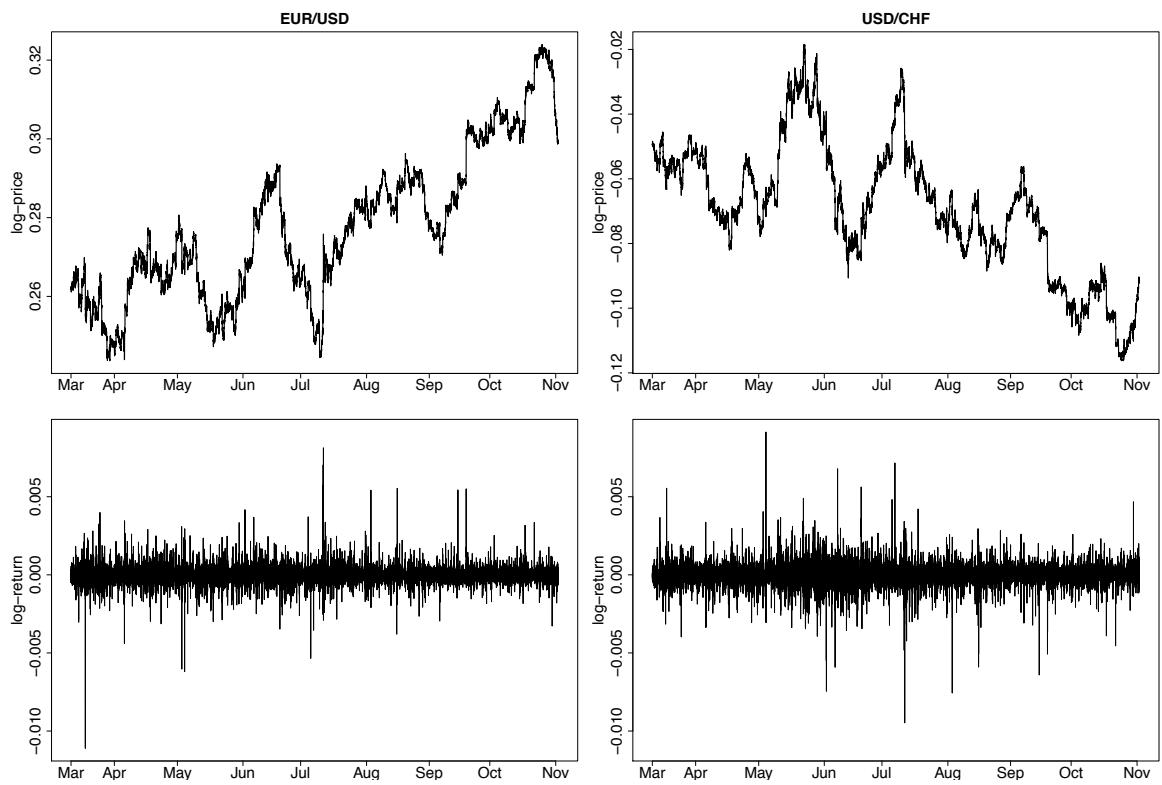
Fig. 2.11 **High-frequency FX data.** Top (respectively bottom) row: the log-price (respectively return) for the EUR/USD (left panel) and USD/CHF (right panel).
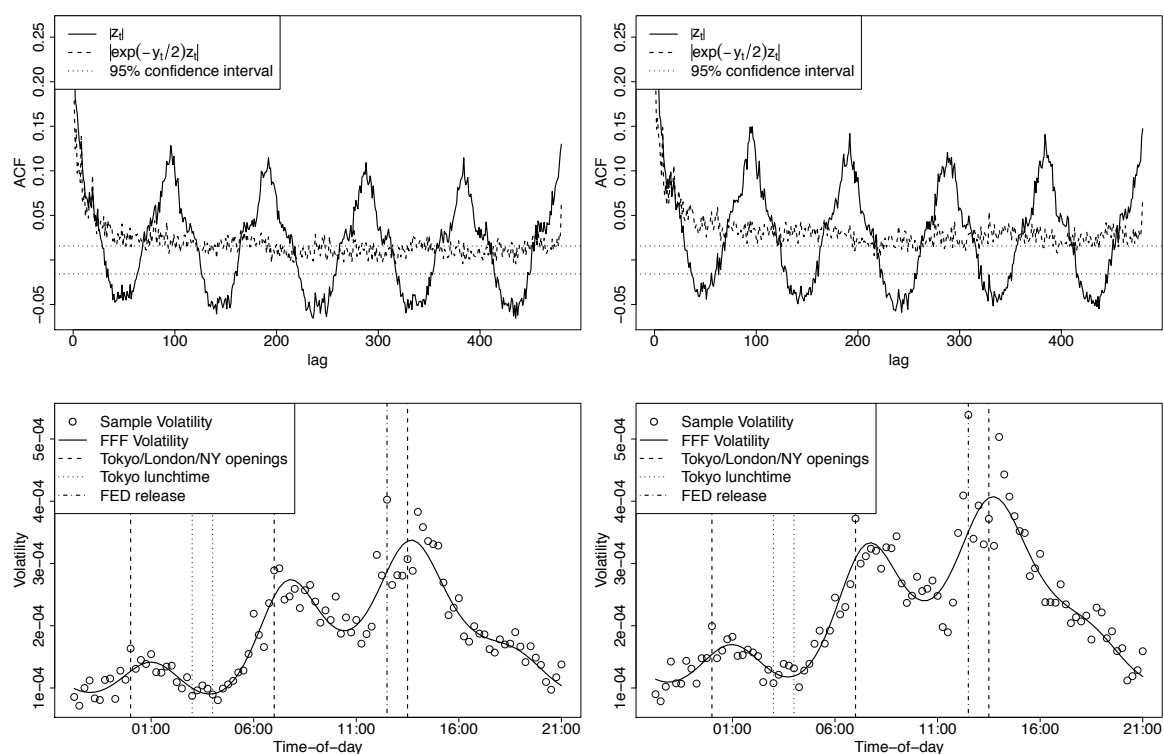
Fig. 2.12 **High-frequency FX ACF and volatility:** Top row: estimated autocorrelation function for the absolute return (plain lines), the deseasonalized absolute return (dashed lines) and 95% confidence bands (dotted lines). Bottom row: sample volatility (dots), FFF volatility (plain lines) and various events impacting the volatility level (dashed and dotted lines). The left panel corresponds to EUR/USD, and the right panel to USD/CHF.
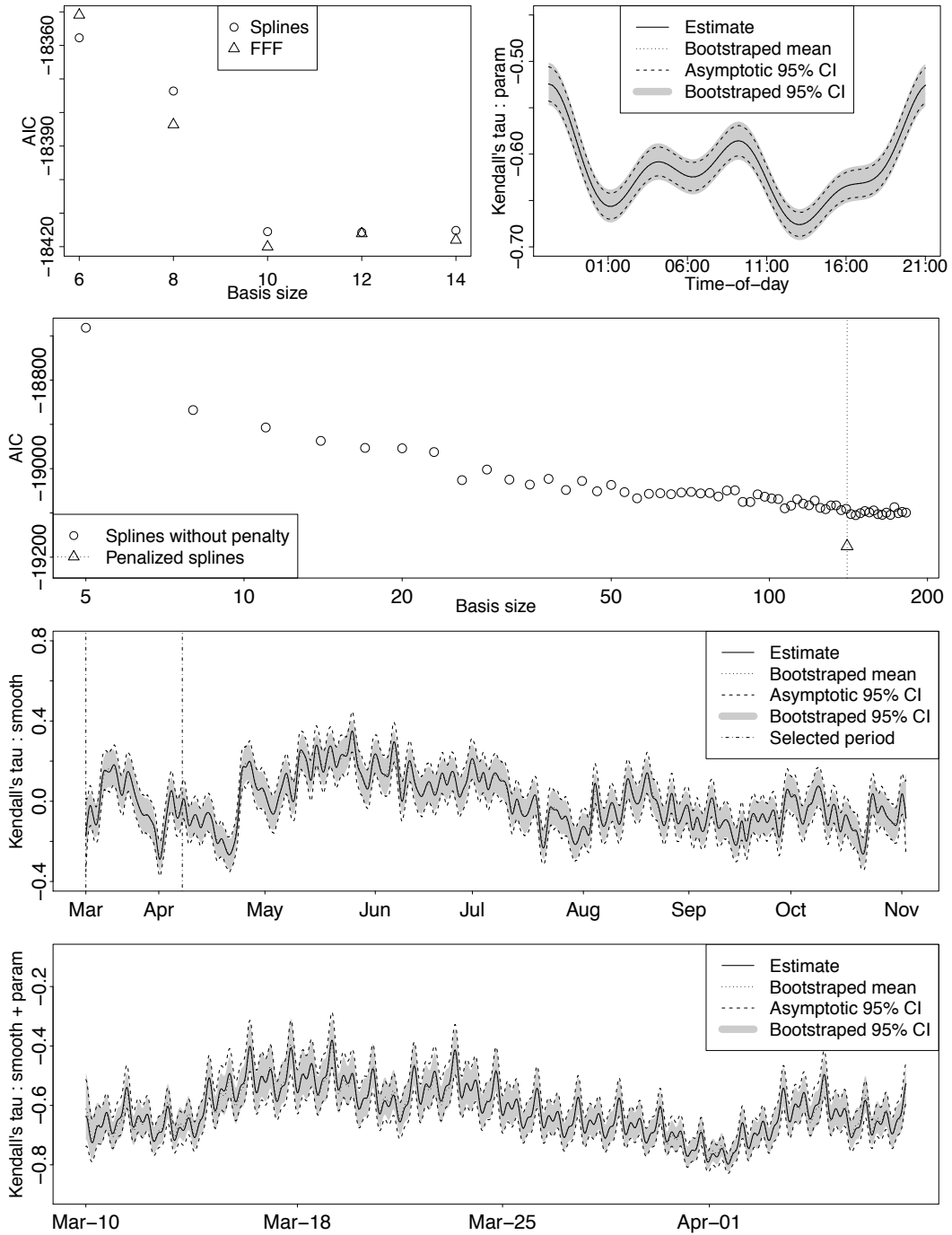
Fig. 2.13 **High-frequency FX results:** Top left: AIC and number of parameters in the $\mathbf{z}^{\top}(x_1)\boldsymbol{\beta}$ vs $h_1(x_1)$ (triangles vs dots) specification. Top right: $\widehat{\psi}(x_1)$ (plain and dotted lines) with 95% c.i. (dashed lines and gray area). Middle top: AIC and basis size for the splines vs penalized splines (dots vs triangle and dotted line) specification of $h_2(x_2)$. Middle bottom: $\widehat{\psi}(x_2)$ (plain and dotted lines), 95% c.i. (dashed lines and gray area) and zooming period for the next row (dashed and dotted lines). Bottom panel: $\widehat{\psi}(\mathbf{x})$ (plain and dotted lines) with 95% c.i. (dashed lines and gray area) on the zooming period.

Chapter 3

Generalized Additive Models for Pair-Copula Constructions

In this chapter, we extend the bivariate framework of Chapter 2 to conditional distributions with $d \geq 2$ dimensions. The structure of the Chapter is as follows: in Section 3.1, we first give an introduction to Pair-copula constructions (PCCs), which are flexible representations of the dependence underlying some multiviariate joint distribution. Using the framework of Chapter 2, we then describe two extensions: the modeling of simplified PCCs with exogenous covariates, as well as a first step toward non-simplified PCCs. In Section 3.2, we discuss model selection and estimation. The estimator's behavior is studied with simulations in Section 3.3. In Section 3.4, we present two applications of our method. First, we study the time-varying dependence structure between the intraday returns on three major foreign exchange rates using a PCC with exogenous covariates. Second, we model the seven-dimensional distribution of an uranium exploration dataset using a non-simplified PCC. We conclude with a discussion in Section 3.5.

3.1 Generalized Additive Models for Pair-Copula Constructions

3.1.1 Pair-Copula Constructions

This section starts with a brief introduction to pair-copula constructions (PCCs). For a more extensive treatment, we refer to Aas et al. (2009) and Czado (2010). Popularized in Aas et al. (2009), they became a hot-topic of multivariate analysis over the last

couple of years. The idea is to model the complete joint distribution of a d-dimensional random vector by considering pairs of (conditional) random variables. Let us consider a three dimensional example. The joint density $f_{1,2,3}(\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^3$, of a vector of continuous random variables $\boldsymbol{X} = (X_1, X_2, X_3)$ can be decomposed as

$$
\begin{aligned}
f_{1,2,3}(\boldsymbol{x}) =& f_1(x_1) \times f_2(x_2) \times f_3(x_3) \times c_{1,2}\left\{F_1(x_1), F_2(x_2)\right\} \times c_{2,3}\left\{F_2(x_2), F_3(x_3)\right\} \\
& \times c_{1,3;2}\left\{F_{1|2}(x_1 \mid x_2), F_{3|2}(x_3 \mid x_2); x_2\right\},
\end{aligned}
$$

where

- f_1, f_2, f_3 are the marginal densities,

- $c_{1,2}$ is the joint density of $(F_1(X_1), F_2(X_2))$,

- $c_{2,3}$ is the joint density of $(F_2(X_2), F_3(X_3))$,

- and $c_{1,3;2}$ is the joint density of $\left(F_{1|2}(X_1 \mid X_2), F_{3|2}(X_3 \mid X_2)\right)$.

The above decomposition can be generalized to arbitrary dimension d and leads to tractable and very flexible models.

Note that, in general, the conditional density $c_{1,3;2}$ is also a function of x_2. However, this effect is often ignored for the sake of tractability, in which case we speak about a simplified PCC. When this so-called simplifying assumption is made, then the complete joint distribution can be built using unconditional bivariate copulas. Discussions on the simplifying assumption can be found in Haff et al. (2010); Stöber et al. (2013). More recently, Spanhel and Kurz (2015) investigated theoretically what can go wrong when the simplifying assumption is not satisfied. They found that — besides missing out on the ignored effects — spurious dependence patterns can appear when the dimension exceeds three. This makes it difficult to interpret and estimate a simplified model.

Following the seminal work of Joe (1997) and Bedford and Cooke (2001, 2002), any copula density c can be decomposed into a product of $d(d-1)/2$ bivariate (conditional) copula densities. The decomposition is not unique, but all possible decompositions can be organized as graphical structure, called regular vine (R-vine) - a sequence of trees $T_m = (V_m, E_m), m = 1, \ldots, d-1$ satisfying the following conditions:

1. T_1 is a tree with nodes $V_1 = \{1, \ldots, d\}$ and edges E_1.

2. For $m \geq 2$, T_m is a tree with nodes $V_m = E_{m-1}$ and edges E_m.

3. (Proximity condition) Whenever two nodes in T_{m+1} are joined by an edge, the corresponding edges in T_m must share a common node.

The corresponding tree sequence is also called the structure of the PCC. A PCC identifies each edge e of the trees with a bivariate copula $c_{j_e,k_e;D_e}$ (a so-called pair-copula). The joint density of the PCC can then be written as the product of all pair-copula densities:

$$c(\boldsymbol{u}) = \prod_{m=1}^{d-1} \prod_{e \in E_m} c_{j_e,k_e;D_e} \left\{ C_{j_e|D_e}(u_{j_e}|\boldsymbol{u}_{D_e}), C_{k_e|D_e}(u_{k_e}|\boldsymbol{u}_{D_e}); \boldsymbol{u}_{D_e} \right\}, \qquad (3.1)$$

where $\boldsymbol{u}_{D_e} := (u_\ell)_{\ell \in D_e}$ is a subvector of $\boldsymbol{u} = (u_1, \ldots, u_d) \in [0,1]^d$ and $C_{j_e|D_e}$ is the conditional distribution of $U_{j_e}|\boldsymbol{U}_{D_e} = \boldsymbol{u}_{D_e}$. The set D_e is called conditioning set and the indices j_e, k_e form the conditioned set. Put differently, the pair-copula density $c_{j_e,k_e;D_e}$ describes the dependence between the two variables U_{j_e} and U_{k_e}, conditional on $\boldsymbol{U}_{D_e} = u_{D_e}$.

Example 1. *The density of a PCC corresponding to the tree sequence in Figure 3.1 is*

$$\begin{aligned}
c(u_1, \ldots, u_5) = {} & c_{1,2}(u_1, u_2) \times c_{1,3}(u_1, u_3) \times c_{3,4}(u_3, u_4) \times c_{3,5}(u_3, u_5) \\
& \times c_{2,3;1}(u_{2|1}, u_{3|1}; u_1) \times c_{1,4;3}(u_{1|3}, u_{4|3}; u_3) \times c_{1,5;3}(u_{1|3}, u_{5|3}; u_3) \\
& \times c_{2,4;1,3}(u_{2|1,3}, u_{4|1,3}; \boldsymbol{u}_{\{1,3\}}) \times c_{4,5;1,3}(u_{4|1,3}, u_{5|1,3}; \boldsymbol{u}_{\{1,3\}}) \\
& \times c_{2,5;1,3,4}(u_{2|1,3,4}, u_{5|1,3,4}; \boldsymbol{u}_{\{1,3,4\}}),
\end{aligned}$$

where we used the abbreviation $u_{j_e|D_e} := C_{j_e|D_e}(u_{j_e}|\boldsymbol{u}_{D_e})$.

Among others, Dissmann et al. (2013) suggest model selection and inference tools for R-vines. Brechmann et al. (2012) study the problem of exponentially increasing complexity in larger dimensions and suggest to either truncate or simplify R-vines. In Joe et al. (2010), the relationship between tail dependence of a vine copula and tail dependences and conditional tail dependences of lower-dimensional margins is investigated. In Nikoloulopoulos et al. (2012), vines are used to obtain stronger lower than upper tail dependence in the joint distribution of financial assets. In Min and Czado (2010, 2011), Markov chain Monte Carlo algorithms are developed to compute credible confidence intervals, detect (conditional) independence, perform model selection and estimation. In Panagiotelis et al. (2012), conditions under which any multivariate discrete distribution can be decomposed as a PCC are given. In Bauer et al. (2012), the vines framework is extended from undirected to directed acyclic graphs, allowing for new conditional independence assumptions.

An interesting extension of the PCC is to include the effect of exogenous variables. This is particularly useful when one wants to investigate the influence of a deterministic

(a) T_1

(b) T_2

(c) T_3

(d) T_4
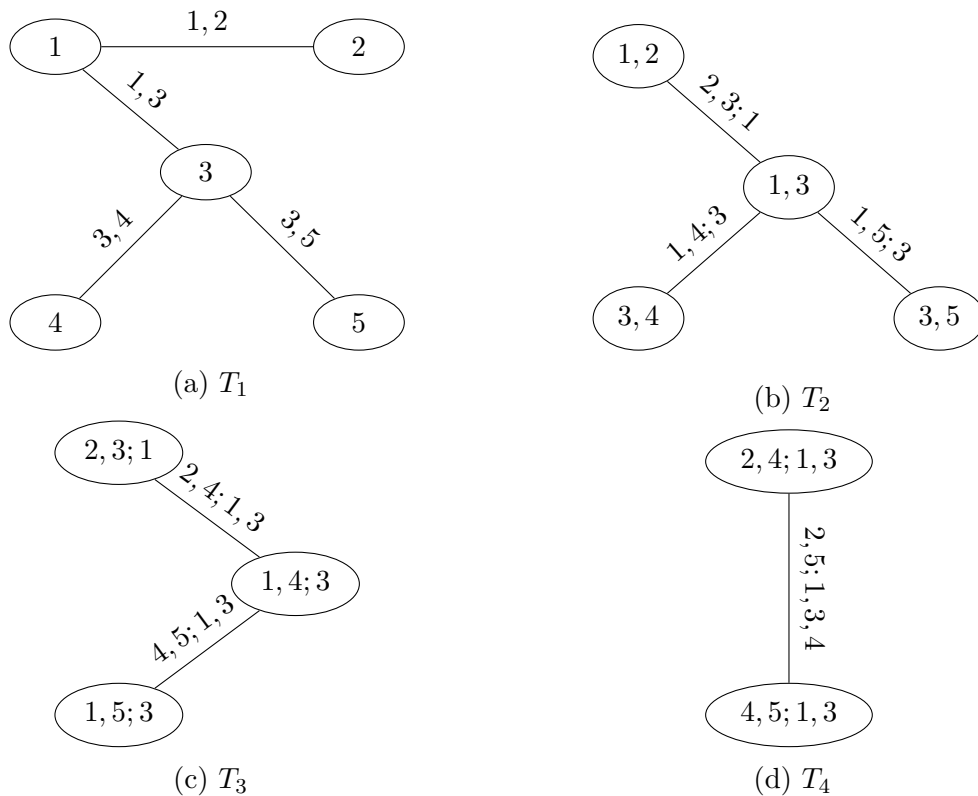
Fig. 3.1 **A regular vine tree sequence:** the four trees of a five dimensional PCC.

covariate (such as space or time) on a complex dependence structure. For instance, the joint spatio-temporal modeling of several hydrograph flood variables, such as the flood peak, the hydrograph volume and hydrograph duration, is necessary to design and manage risks for hydraulic structures like dams. Another example is the modeling of joint distribution of intraday returns on exchange rates, whose the dependence structure changes over time due to the cyclical nature of market activity. Even when the covariate is random, we are often only willing to study its effect on the joint distribution of a response vector of interest. In this case, it is usually unnecessary or inconvenient to model its stochastic behavior explicitly, and a regression-like theory for PCCs is required. In the hydrological example above, when a region under study characterized by large hydro-climatic heterogeneities, the inclusion of additional (potentially random) descriptors in the model is important; especially as the ultimate goal is the extrapolation (prediction) at ungauged sites. Similarly, scheduled economic news, such as the monthly release of the US unemployement rate, or the FOMC press conference, impact in a crucial way the dependence structure between intraday returns. Previous work in this direction include regime switching PCC (Stöber and Czado, 2014), and spatial PCC

models ([Erhardt et al.](), [2015a](),[b](); [Gräler](), [2014]()), where the individual parameters of the pair-copulas were modeled as linear functions of distances between different locations.

In the next Section, we discuss this extension, modeling explicitly the conditional copula density using the framework of Chapter [2](). We further generalize the model in Section [3.1.3]() by relaxing the simplifying assumption.

3.1.2 Pair-Copula Constructions with Exogenous Covariates

Recall that, in a simplified PCC, we ignore the influence of the covariate vector $\boldsymbol{u}_{De}$ on the pair-copula density $c_{j_e,k_e;D_e}$. In this case, the density of the PCC collapses to

$$c(\boldsymbol{u}) = \prod_{m=1}^{d-1} \prod_{e \in E_m} c_{j_e,k_e;D_e} \left\{ C_{j_e|D_e}(u_{j_e}|\boldsymbol{u}_{D_e}), C_{k_e|D_e}(u_{k_e}|\boldsymbol{u}_{D_e}) \right\}. \qquad (3.2)$$

Using the framework of Chapter [2](), we make each pair-copula depend on a vector of exogenous variables $\boldsymbol{w}$:

$$c(\boldsymbol{u}; \boldsymbol{w}) = \prod_{m=1}^{d-1} \prod_{e \in E_m} c_{j_e,k_e;D_e} \left\{ C_{j_e|D_e}(u_{j_e}|\boldsymbol{u}_{D_e}), C_{k_e|D_e}(u_{k_e}|\boldsymbol{u}_{D_e}); \psi_e(\boldsymbol{w}) \right\}, \qquad (3.3)$$

where the dependence parameter, which can be either a dependence measure such as Kendall's tau or a copula parameter directly, is expressed as

$$\psi_e(\boldsymbol{w}) = \psi_e(\boldsymbol{w}; \boldsymbol{\theta}_e) = g_e \left\{ \mathbf{z}_e^\top \boldsymbol{\beta}_e + \sum_{k=1}^{K_e} s_{e,k}(\mathbf{t}_{e,k}) \right\}.$$

As in Equation ([2.7]()), $\mathbf{z_e}$ and $\mathbf{t}_{e,k}$ are vectors of elements in $\boldsymbol{w}$ (or products thereof), and $g_e, \boldsymbol{\beta}_e, \boldsymbol{\theta}_e$ are defined similarly. Note that, in this Chapter, we use s instead of h to represent smooth components, because the h is usually reserved for the h-functions in the PCC context.

In this model, the covariate vector $\boldsymbol{w}$ is equal for all pair-copulas $c_{j_e,k_e;D_e}$. In particular, the number of covariates remains constant across the trees. For instance, consider the three dimensional PCC from Figure [3.2]() again, and assume that each copula parameter depends on a covariate x and time t. As such, the vector of covariates $\boldsymbol{w} = (x, t)$ is the same in T_1 and T_2. Supposing that we want the effect of time-variation to be nonparametric and the effect of the other covariate to be parametric, a model for

(a) T_1 (b) T_2

Fig. 3.2 **A simplified three-dimensional PCC with exogenous covariates:** the three conditional copula parameters for a PCC that depend on a covariate x parametrically and t nonparametrically.

the corresponding conditional pair-copulas can be written as

$$\psi_{1,2}(\boldsymbol{w}) = g_{1,2}\left\{x\beta_{1,2} + s_{1,2}(t)\right\},$$
$$\psi_{1,3}(\boldsymbol{w}) = g_{1,3}\left\{x\beta_{1,3} + s_{1,3}(t)\right\},$$
$$\psi_{2,3;1}(\boldsymbol{w}) = g_{2,3;1}\left\{x\beta_{2,3;1} + s_{2,3;1}(t)\right\}.$$

3.1.3   Non-simplified Pair-Copula Constructions

Recall that, in its most general form, a PCC consists of conditional copula densities: For a given edge e, the function $c_{j_e,k_e;D_e}$ is the copula density associated with the conditional random vector

$$(U_{j_e}, U_{k_e}) \,\big|\, \boldsymbol{U}_{D_e} = \boldsymbol{u}_{D_e}.$$

Note that $c_{j_e,k_e;D_e}(\cdot, \cdot; \boldsymbol{u}_{D_e})$ may be different for each value of $\boldsymbol{u}_{D_e}$. If we assume that the copula density is equal across all possible values of $\boldsymbol{u}_{D_e}$, the model is called "simplified". In contrast, our goal is to explicitly model the variation of the copula density across the values of $\boldsymbol{u}_{D_e}$.

Using the framework of Chapter 2 again, we model the density of a PCC as

$$c(\boldsymbol{u}) = \prod_{m=1}^{d-1} \prod_{e \in E_m} c_{j_e,k_e;D_e}\left\{C_{j_e|D_e}(u_{j_e}|\boldsymbol{u}_{D_e}), C_{k_e|D_e}(u_{k_e}|\boldsymbol{u}_{D_e}); \psi_e(\boldsymbol{u}_{D_e})\right\}, \qquad (3.4)$$

with

$$\psi_e(\boldsymbol{u}_{D_e}) = \psi_e(\boldsymbol{u}_{D_e}; \boldsymbol{\theta}_e) = g_e\left\{\mathbf{z}_e^\top \boldsymbol{\beta}_e + \sum_{k=1}^{K_e} s_{e,k}(\mathbf{t}_{e,k})\right\},$$

where $\mathbf{z_e}$ and $\mathbf{t}_{e,k}$ are vectors of elements in $\boldsymbol{u}_{D_e}$ (or products thereof), and $g_e, \boldsymbol{\beta}_e, s_{e,k}, \boldsymbol{\theta}_e$ are defined similar to (2.7).

(b) T_2

(a) T_1

(c) T_3
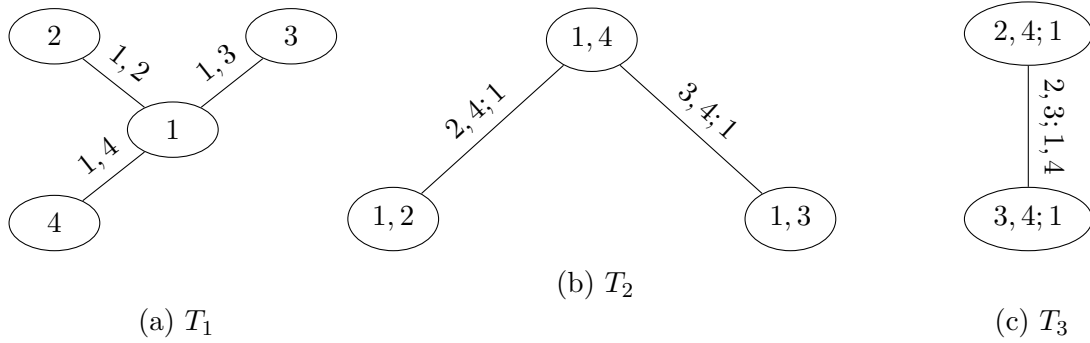
Fig. 3.3 **A non-simplified PCC:** the four trees of a four-dimensional regular vine.

It is noteworthy that the covariate vectors $\boldsymbol{u}_{D_e}$ usually differ between edges. In particular, the number of components in $\boldsymbol{u}_{D_e}$ grows with the tree level: for $e \in E_m$, the covariate vector $\boldsymbol{u}_{D_e}$ consists of $m - 1$ components. For instance, consider the four-dimensional PCC from Figure 3.3. In T_1, each edge has an empty conditioning set and the corresponding pair-copula is simplified. However, in T_2 and T_3, the conditioning sets are $\{1\}$ and $\{1, 4\}$. Hence, using only nonparametric smooth functions, a model for the corresponding conditional pair-copulas can be written as

$$\psi_{2,4;1}(u_1) = g_{2,4;1}\left\{s_{2,4;1}(u_1)\right\},$$

$$\psi_{3,4;1}(u_1) = g_{3,4;1}\left\{s_{3,4;1}(u_1)\right\},$$

$$\psi_{2,3;1,4}(\boldsymbol{u}_{2,3;1,4}) = g_{2,3;1,4}\left\{\sum_{j\in\{1,4\}} s_{2,3;1,4,j}(u_j)\right\}.$$

3.2   Sequential Estimation and Model Selection

We next discuss how the models proposed in Section 3.1 can be estimated. First, we give some heuristics for the selection of covariates and basis size for a single pair-copula when the parametric family is known. Second, we adapt the usual sequential estimation procedure to the class of PCC models proposed in Sections 3.1.2 and 3.1.3.

3.2.1   Model selection for a Single Pair-copula Family

At this step, the goal is to select a GAM model assuming a known pair-copula family. Generally speaking, there can be at most one unique smooth function $s_{e,k}$ for each

covariate. But usually there is no a priori knowledge of its shape. This gives rise to two questions:

1. Which of the covariates have an influence?

2. What is the appropriate basis size for each of the smooth functions?

Below, we give heuristically motivated answers to these questions.

To answer the **first question**, we first set the basis size for each component to five, which is at the same time big enough to detect obvious non-linear relations and small enough to be quickly estimated. Second, we use a variant of backward elimination, where we start with all the covariates, remove at each step the ones whose individual p-values are above a pre-specified level α, re-estimate the model and iterate until all remaining covariates are significantly non-zero. Note that, additionally to the usual issues related to step-wise selection methods, p-values for the smooth terms are necessarily approximate. As suggested in Marra and Wood (2012); Wood (2013a,b) in the exponential family context, we compute individual p-values using a Wald test[1].

As for the **second question**, once the set of covariates is selected, our answer is heuristic again. With GAMs, the choice of the basis size is usually not critical: it should be large enough so that the underlying features in the data can be well approximated, but small enough such as to maintain a good computational efficiency and low variance (although this last problem is partly addressed using the penalization).

To achieve a good trade-off, we start with a small basis size ($b_j = 5$ for the selected covariate j), fit the model, check whether the estimated equivalent degree of freedom corresponding to b_j is "close" to this upper limit[2], increase b_j if that is the case, and iterate until no further increase is required for any of the covariates. Additionally, at each step, we make sure that no individual basis size is greater than the sample size divided by thirty. While not theoretically justified, keeping a reasonably large ratio of number of observations to number of parameters is a rule of thumb that was found useful in this context[3].

---

[1]To compute the test statistics, we use a covariance matrix that results from the Bayesian interpretation of GAMs. While there is no optimality result for the power, Marra and Wood (2012), extending the analysis of Nychka 1988, motivated the use of this covariance matrix by showing that the resulting intervals have better frequentist performances (power and distribution under the null) than those computed using a strictly frequentist approximation.

[2]For cubic splines, the upper limit is in fact $b_j - 1$, because we require that the smooth integrates to zero for identifiability, which means that one degree of freedom is lost.

[3]Note that this restriction was not enforced in an earlier version of the code. However, it is clear that the variance of the estimates increases when the the ratio of number of observations to number of parameters decreases. Furthermore, because we noticed numerical instabilities in simulations when the ratio was around 30 and lower, we decided to use it as a bound in our algorithm.

We summarize this method in Algorithm 1, where we use $PMLE$, a function which computes the penalized maximum log-likelihood estimator $\widehat{\boldsymbol{\theta}}$ while selecting the vector of smoothing parameters $\boldsymbol{\gamma}$ automatically using GVC minimization. Its inputs are two $n \times 1$ vectors (say u_1 and u_2), a $n \times k$ matrix of k covariates, and a $k \times 1$ vector of basis sizes for the smooth components corresponding to each covariates. Its output is the fitted model.

3.2.2 Sequential Estimation of a Pair-Copula Construction

For estimation of PCC models it is common to follow a sequential estimation approach (see e.g. Aas et al., 2009; Hobæk Haff, 2013; Nagler and Czado, 2015), which we outline below. Assume that $\boldsymbol{u}^i = (u_1^i, \ldots, u_d^i), i = 1, \ldots, n$, are observations from a pair-copula construction and the vine structure is known. Then, the pair-copulas of the first tree, T_1, can be easily estimated using the method described in the previous subsection. This is not as straightforward for trees $T_m, m \geq 2$ since data from the densities $c_{j_e,d_e;D_e}$ are unobserved. However, we can sequentially construct pseudo-observations by an appropriate transformation of the data.

Define the h-functions (c.f. Aas et al., 2009) corresponding to a pair-copula density $c_{j_e,k_e;D_e}(u, v; \cdot)$ as

$$h_{j_e|k_e;D_e}(u|v; \cdot) = \int_0^u c_{j_e,k_e;D_e}(s, v; \cdot)ds, \quad h_{k_e|j_e;D_E}(u|v; \cdot) = \int_0^v c_{j_e,k_e;D_e}(u, s; \cdot)ds,$$

for all $(u, v) \in [0, 1]^2$. The dot in the third argument represents one of the GAM-formulations in Section 3.1. A crucial insight is the following: Assume we have (pseudo-)observations from the pair-copula density $c_{j_e,k_e;D_e}$, denoted as $(u_{j_e|D_e}^i, u_{k_e|D_e}^i), i = 1, \ldots, n$. Then we can construct pseudo-observations for the next tree by setting

$$u_{j_e|D_e \cup k_e}^i = h_{j_e|k_e;D_e}\left(u_{j_e|D_e}^i|u_{k_e|D_e}^i; \cdot\right), \quad u_{k_e|D_e \cup j_e}^i = h_{k_e|j_e;D_e}\left(u_{k_e|D_e}^i|u_{j_e|D_e}^i; \cdot\right), \quad (3.5)$$

for $i = 1, \ldots, n$.

As only the estimates of each pair-copula in T_l are required to compute pseudo-observations for tree T_{l+1}, we make use of the following sequential estimation procedure, starting with tree T_1:

1. For each edge in the tree:

    (a) Select covariates and estimate the GAM model for each copula family.

Algorithm 1 Model selection for a bivariate conditional copula

---

1: **Inputs:**
   $u_1, u_2, y_1, \ldots, y_k, \alpha$
2: **Initialize:**
   $b \leftarrow k$
   $basis_j \leftarrow 5, \; j = 1, \ldots, b$
   $cov \leftarrow$ a $n \times b$ matrix with columns $y_1, \ldots, y_b$
   $sel \leftarrow false$
3: **while** ANY($sel \neq true$) AND $b > 0$ **do**   ▷ Remove the "unsignificant" covariates
4:     $fitted \leftarrow$ PMLE($u_1, u_2, cov, basis$)
5:     Clear $basis$, PV, and sel
6:     $PV_j \leftarrow p$-value of $fitted$ corresponding to column j of cov, $j = 1, \ldots, b$
7:     $sel_j \leftarrow true, \; j = 1, \ldots, b$
8:     **for** $j = 1$ to b **do**
9:         **if** $PV_j \geq \alpha$ **then**
10:             $sel_j \leftarrow false$
11:             remove column j from cov
12:         **end if**
13:     **end for**
14:     $b \leftarrow \sum_{j=1}^{b} \mathbb{1}_{PV_j < \alpha}$
15:     $basis_j \leftarrow 5, \; j = 1, \ldots, b$
16: **end while**
17: $fitted \leftarrow$ PMLE($u_1, u_2, cov, basis$)
18: **if** $b \neq 0$ **then**   ▷ Select the basis size
19:     $EDF_j \leftarrow$ EDF of $fitted$ corresponding to column j of cov, $j = 1, \ldots, b$
20:     **if** ANY($EDF > (basis - 1)/2$) **then**
21:         **while** ANY($sel == true$) AND $ALL(basis < n/30)$ **do**
22:             Clear sel
23:             $sel_j \leftarrow false, \; j = 1, \ldots, b$
24:             **for** $j = 1$ to b **do**
25:                 **if** $EDF_j > (basis_j - 1)/2$ **then**
26:                     $sel_j \leftarrow true$
27:                     $basis_j \leftarrow 2 \cdot basis_j$
28:                 **end if**
29:             **end for**
30:             $fitted \leftarrow$ PMLE($u_1, u_2, cov, basis$)
31:             $EDF_j \leftarrow$ EDF of $fitted$ corresponding to column j of cov, $j = 1, \ldots, b$
32:         **end while**
33:     **end if**
34: **end if**
35: **return** $fitted$

(b) Use the AIC, BIC, or any suitable information criterion to choose a copula family.

(c) Use the estimated pair-copula density to construct pseudo-observations for the next tree via (3.5).

2. Go to next tree.

The fact that the tree sequence $T_1, T_2, \ldots,$ is a regular vine guarantees that at any step in this procedure, all required pseudo-observations are available.

3.3 Simulations

We evaluate the performance of the estimation and model selection procedures of Section 3.2 in a simulation study. In particular, we investigate the influence of copula family, type of smooth function, tree level, and sample size on the estimation accuracy.

3.3.1 Pair-Copula Constructions with Exogenous Covariates

In this section, we present simulation results to study the model described in Section 3.1.2. Before an analysis of the results, we outline how simulations were set up.

Setup

To encompass different cases of practical interest, we use the same three deterministic functions as in Chapter 2 normalized to ensure identifiability, namely

$$s_1(x) = -2/3 + 8(x - 0.5)^2 \tag{3.6}$$
$$s_2(x) = \sin(2\pi x) \tag{3.7}$$
$$s_3(x) = \sqrt{2\pi}/4 \left\{ \Phi_{1/2, 1/8}(0) - \Phi_{1/2, 1/8}(1) \right\} + \exp\left\{ -32(x - 1/2)^2 \right\}, \tag{3.8}$$

where $\Phi_{\mu, \sigma}$ denotes the Gaussian distribution function with mean μ and standard deviation σ. In order to make consistent estimation feasible, we represent each of the smooth functions using cubic splines interpolation based on 10 knots (equidistant on the unit interval). The Kendall's tau of each pair-copula is then specified as a function

$$\psi_e(\boldsymbol{x}) = g\left\{ s_{r_e(1)}(x_1) + s_{r_e(2)}(x_2) \right\}, \qquad g(x) = (e^x - 1)/(e^x + 1),$$

where, for each pair e, the indices $r_e(k), k = 1, 2$, are drawn with equal probability from the set $\{1, 2, 3\}$. This results in a Kendall's tau range of approximately $[-0.9, 0.9]$ for each pair. Additionally, we draw the copula family for each pair-copula with equal probability from the Gaussian, Student's t (with four degrees of freedom), Clayton and Gumbel families. Using independent covariates, each distributed uniformly on $[0, 1]$, the experiment is repeated $1\,000$ times for sample sizes $n = 500, 2\,000$.

Note that, since there are three pair-copulas in a 3-dimensional PCC, we sample six deterministic functions for each experiment. In order to also study the influence of the copula family, there is an impractically large number of possible combinations for each sample size ($3^6 \times 4^3 = 46\,656$), which explains why we opt for this randomization.

Furthermore, in this experiment, the Clayton and Gumbel families are extended to allow for $\tau < 0$ by using $90°$ and $270°$ rotations. For example, if $c^{Clay}(u_1, u_2; \eta)$ denotes the Clayton copula with parameter η, then $c^{Clay90}(u_1, u_2; \eta) = c^{Clay}(u_2, 1 - u_1; \eta)$ is its 90 degree (counter-clockwise) rotation and allows for negative dependence. We refer the reader to Chapter 4 for more details.

Results

We start with a detailed analysis of the case $n = 500$. Later, we compare the performance to the case $n = 2\,000$. Figure 3.4a shows the three deterministic functions s_1, s_2, s_3 (solid lines), median estimates (dashed lines), and 95% confidence bands (dotted). The top row corresponds to the optimal estimator (opt), where we assumed that the copula families and number of knots are known. The bottom row shows the results for the estimator that assumes no prior knowledge (except the structure of the vine) and uses the model selection algorithm proposed in Section 3.2.1 (sel) as well as the AIC to select the copula family. Both estimators capture the characteristics of each smooth function quite well. We observe some bias in areas of high curvature and increasing variability in the estimates towards the boundaries. These findings are in line with the results from Chapter 2. On a first glance, there seems to be little difference between the optimal estimator and the model selection algorithm.

In the top row of Figure 3.4b, the effect of the copula family on estimation accuracy is illustrated with boxplots of the integrated absolute error (IAE, see the simulation study of Chapter 2) between the true and estimated function. The results indicate that the family does have an impact on accuracy of both estimation methods. The Clayton copula seems to be a little problematic. This is most likely caused by the fact that the Clayton density explodes at the (0,0) corner very fast, leading to instabilities. In the two left plots of the bottom row the results are split by tree level. We see
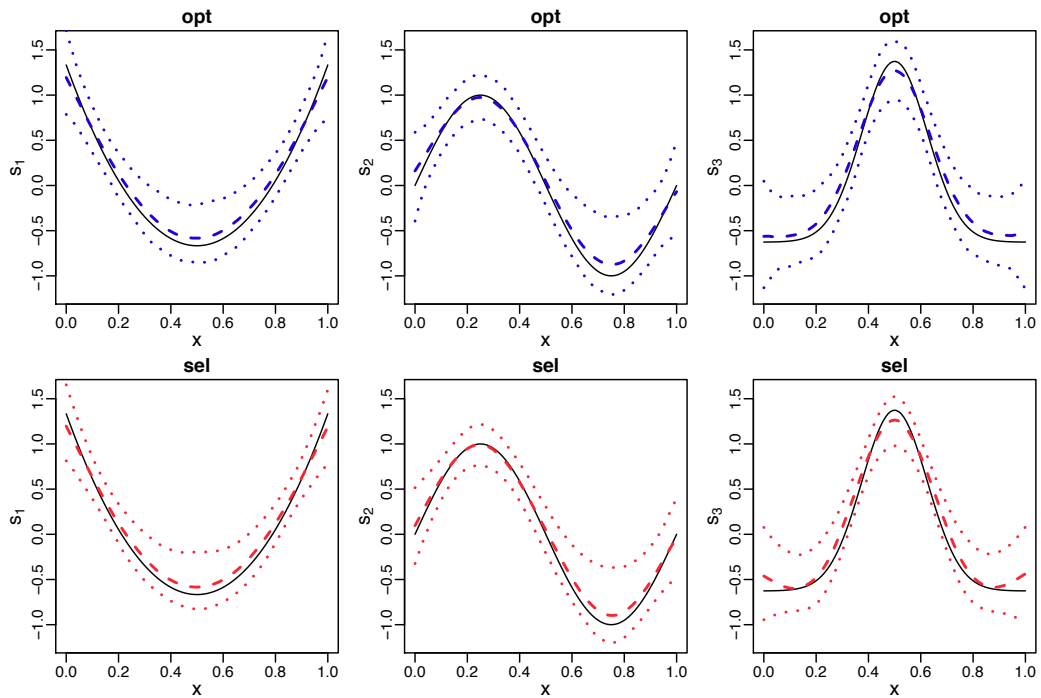
(a) Split by deterministic function s_1, s_2, s_3 ($n = 500$): true function (solid), median estimate (dashed), and 95% c.i. (dotted).
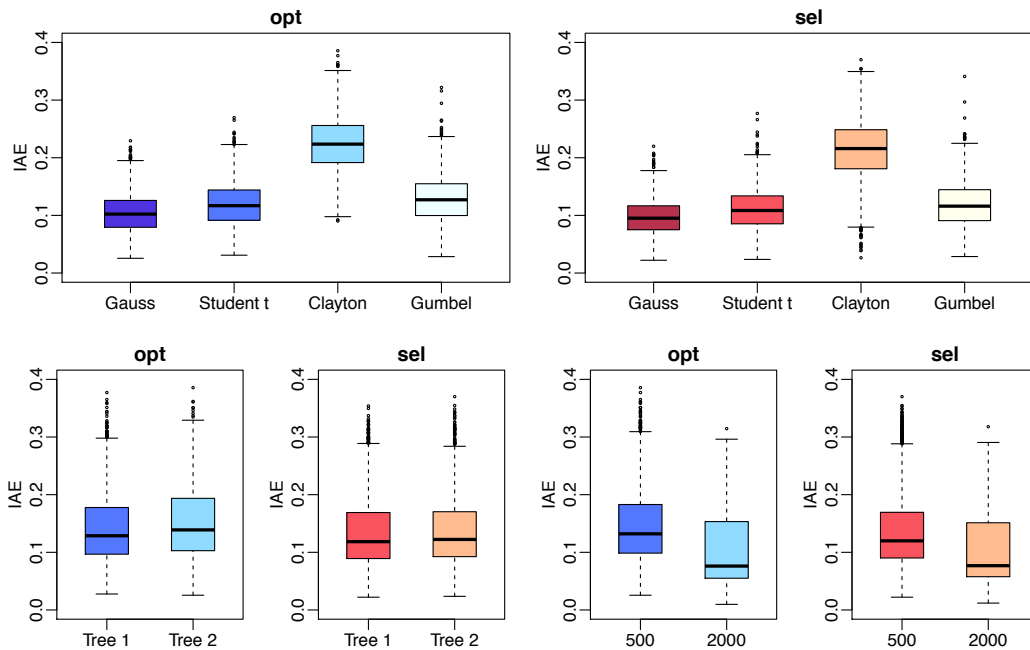
(b) Split by family and tree (both for $n = 500$) and sample size (bottom right).

Fig. 3.4 **PCC with exogenous covariates:** simulation results for the correctly specified estimator (opt) and model selection algorithm (sel).

	Gaussian	Student's t	Clayton	Gumbel	
Gaussian	24.2	0.0	0.1	0.4	24.8
Student's t	0.8	24.1	0.3	0.2	25.4
Clayton	0.0	0.0	24.9	0.0	24.9
Gumbel	0.0	0.1	0.0	24.7	24.8
	25.1	24.2	25.3	25.4	100.0

Table 3.1 **PCC with exogenous covariates:** contingency table of true family (rows) and family selected by the model selection algorithm (columns) in percent. All numbers are rounded to one digit.

that the estimation accuracy gets worse in the second tree. The main reason is that the sequential estimation approach (c.f., Section 3.2.2) relies on pseudo-observations. Estimation errors from the first tree propagate to the second tree, and so on. The remaining plots of the bottom row indicate that estimation accuracy improves notably when we increase the sample size from $n = 500$ to $n = 2\,000$.

Overall, we observe that the selection algorithm does a very good job in estimating the smooth functions s_1, s_2, s_3 — it even performs slightly better than the optimal estimator. This can be explained by the increased flexibility in choosing the number of knots: if the data doesn't convey enough information, a smaller number of knots can decrease the variance. However, this does not mean that the correct copula model was selected. This can be further investigated by comparing the true pair-copula families to the families selected by the algorithm. Table 3.1 shows a contingency table of true and selected families over all scenarios of this study. We see that, most of the time, the algorithm also manages to correctly identify the true family.

3.3.2 Nonsimplified Pair-Copula Constructions

The simulation setup for the nonsimplified case is similar to the above. But here, we specify the Kendall's tau of each pair-copula as

$$\psi_e(\boldsymbol{u}_{D_e}) = g\left\{\sum_{k \in D_e} s_{r_e(k)}(u_k)\right\}, \qquad g(x) = (e^x - 1)/(e^x + 1),$$

where, for each pair e, the indices $r_e(k)$ are drawn with equal probability from the set $\{1, 2, 3\}$. This results in a Kendall's tau range of approximately [-0.5, 0.5] in the second tree, and [-0.9, 0.9] in the third tree.

Figure 3.5 shows the results of the simulation study. They are very similar to the case with exogenous covariates and all conclusions from Section 3.3.1 transfer.
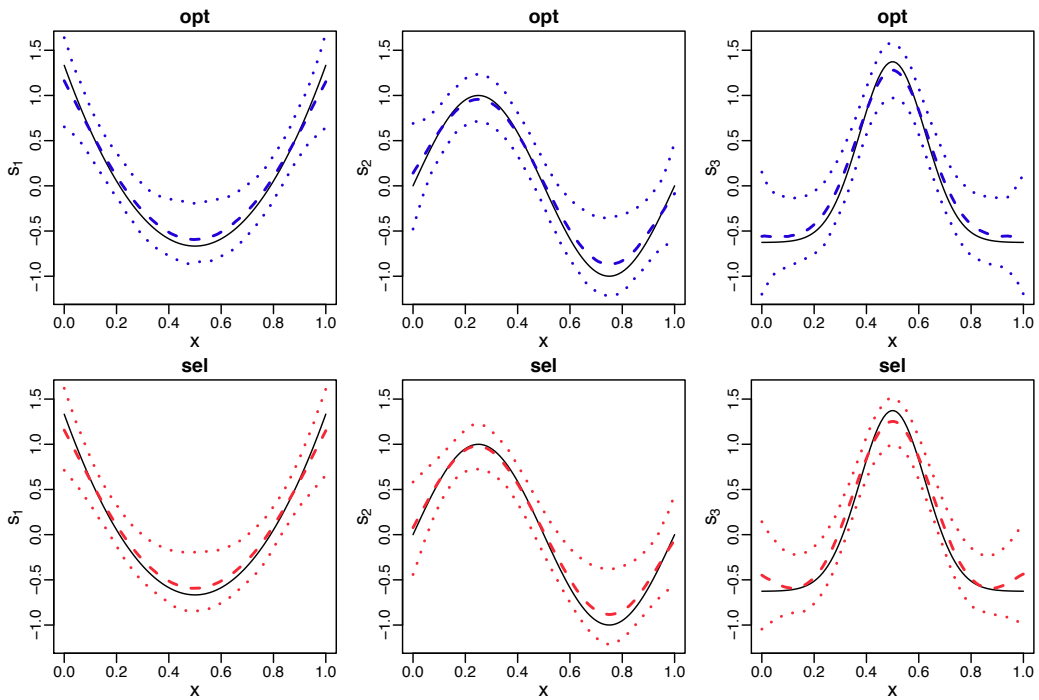
(a) Split by deterministic functions s_1, s_2, s_3 ($n = 500$): true function (solid), median estimate (dashed), and 95% c.i. (dotted).
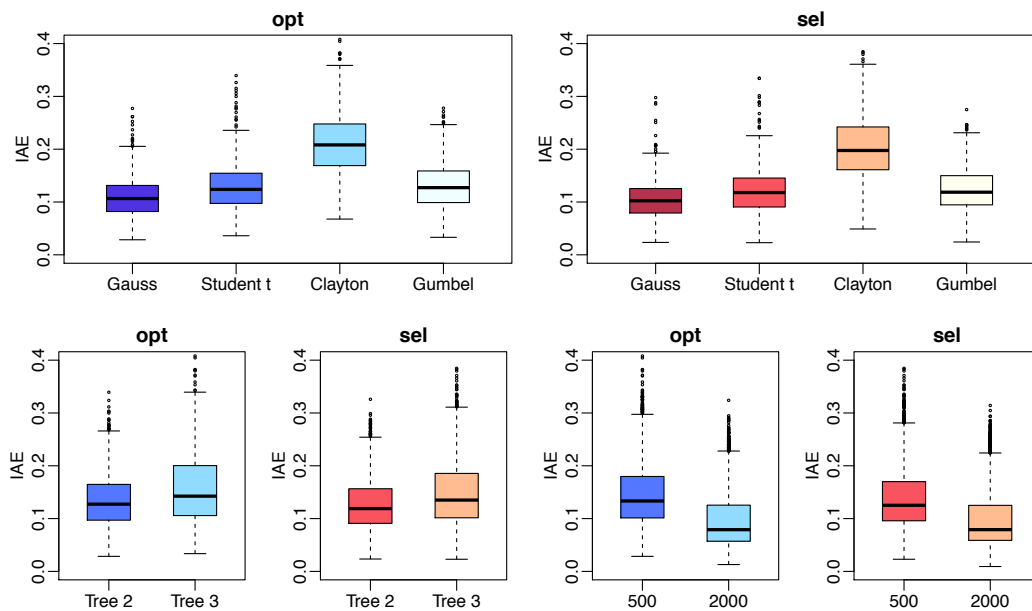
(b) Split by family. tree (both for $n = 500$) and sample size (bottom right).

Fig. 3.5 **Nonsimplified PCC:** simulation results for the correctly specified estimator (opt) and model selection algorithm (sel).

3.4   Applications

3.4.1   The Foreign Exchange Market Revisited

In our second application, we continue our study the cross-sectional dynamics of intraday asset return of Chapter 2. We extend the bivariate model to encompass three or more exchange rates using vines with exogenous covariates framework. We use the same data graciously provided by Dukascopy Bank SA[4], which contains 15-minute spaced returns (i.e., 96 observations each day) for the EURUSD, USDJPY and USDCHF from March 10, 2013 to November 1, 2013. Hence, in a total of 34 trading weeks, there are 16320 observations (170 days) excluding weekends.

As in Chapter 2, because intraday returns are heteroskedastic, we need to pre-filter the individual series before applying the methodology of this Chapter. However, contrasting with Chapter 2, we keep the decomposition of the volatility in two multiplicative components (a seasonal, often assumed deterministic, part and a stochastic part), but we model it directly within a one-step procedure. In fact, it is straightforward to achieve this within the GARCH-family, where $r_t = \sigma_t \, y_t$ for σ_t a function of $\{r_{t-1}, \sigma_{t-1}, r_{t-2}, \sigma_{t-2}, \cdots\}$ and y_t a white noise. Using sines and cosines to explain the intraday conditional variance, we write

$$\log \sigma_t^2 = \left[\omega + \sum_{k=1}^{K} \left\{ a_k \, \cos\left(2\pi kt/T\right) + b_k \, \sin\left(2\pi kt/T\right) \right\} \right]$$
$$+ \alpha \, \epsilon_{t-1} + \gamma \left(\mid \epsilon_{t-1} \mid - \mathbb{E} \mid \epsilon_{t-1} \mid \right) + \beta \log \sigma_{t-1}^2,$$

where $T = 96$. Note that this model is the EGARCH(1,1) from Nelson (1991), augmented with external regressors to take the seasonality into account. The sum of cosines and sines with integer frequencies, designed to capture daily oscillations around the base level, is similar to the Fourier Flexible Form (FFF, see Gallant 1981), introduced in this context by Andersen and Bollerslev (1997, 1998). Denoting by $\widehat{\sigma}_t$ the fitted volatility using the maximum log-likelihood estimator with $K = 5$, we refer to $\widehat{y}_t = r_t \, / \, \widehat{\sigma}_t$ as the residuals.

In Figure 3.6, we show the returns along with fitted conditional standard deviations for the first week of sample. From this picture, it is clear that the periodicity is stronger for the EURUSD and USDCHF than for the USDJPY. In Figure 3.7, the black (respectively red) curve represents the autocorrelation of the absolute value of the returns (respectively residuals), where we observe that our univariate models

---

[4]http://www.dukascopy.com/

EURUSD-USDCHF	EURUSD-USDJPY	USDCHF-USDJPY
-0.608	-0.171	0.252

Table 3.2 **Unconditional dependence between the foreign exchange rates:** the Kendall's tau computed using each univariate model's residuals.

appropriately capture the heteroskedasticity. In Figure 3.8, the black (respectively red) curve represents the empirical (respectively fitted) volatility per 15 minutes bin, where we recognize the usual modes at the opening time of the Tokyo, London and New-York markets.

From the residuals, we then compute observations on the copula scale by using the empirical cumulative distribution for each margins. Because the USDCHF residuals exhibit the higher dependence with the other two exchange rates (see Table 3.2), we choose this pair as the central node of the trivariate PCC. Furthermore, we use the same FFF regressors as in the conditional variance to model the periodic component of the dependence structure. We also add a smooth function of the time to model the evolution of the dependence over the sample period. Then, similarly as the example of Section 3.1.2, a model for the corresponding conditional pair-copulas can be written as

$$\psi_{1,2}(\boldsymbol{w}) = g_{1,2}\left\{\mathbf{x}(t)^\top \boldsymbol{\beta}_{1,2} + s_{1,2}(t)\right\},$$
$$\psi_{2,3}(\boldsymbol{w}) = g_{2,3}\left\{\mathbf{x}(t)^\top \boldsymbol{\beta}_{2,3} + s_{2,3}(t)\right\},$$
$$\psi_{2,3;2}(\boldsymbol{w}) = g_{2,3;2}\left\{\mathbf{x}(t)^\top \boldsymbol{\beta}_{2,3;2} + s_{2,3;2}(t)\right\},$$

where $\mathbf{x}(t) = (1, \cos\left(2\pi t/T\right), \ldots, \cos\left(2\pi 5t/T\right), \sin\left(2\pi t/T\right), \ldots, \sin\left(2\pi 5t/T\right))^\top$, $1 =$ EURUSD, $2 =$ USDCHF, and $3 =$ USDJPY. Before describing the results, two additional remarks concerning the model are:

- To facilitate its interpretation, we estimate the PCC directly using the Kendall's tau parametrization (and not the copula parameter) as in Chapter 2.

- We use the Student's t for all tree conditional pair-copulas, because it is favored by the AIC over the Gaussian or common Archimedean copulas, as is often the case when modeling financial data.

In order to capture the time-varying features, each smooth was estimated with 100 knots, but the generalized cross-validation minimization reduced the effective degrees of freedom to 80.24, 57.94 and 39.13. Furthermore, the average Kendall's tau, that is simply g(intercept), for the first two pair-copulas are -0.62 and 0.27, which is close to the unconditional values (see Table 3.2). As for the last conditional pair-copula, the
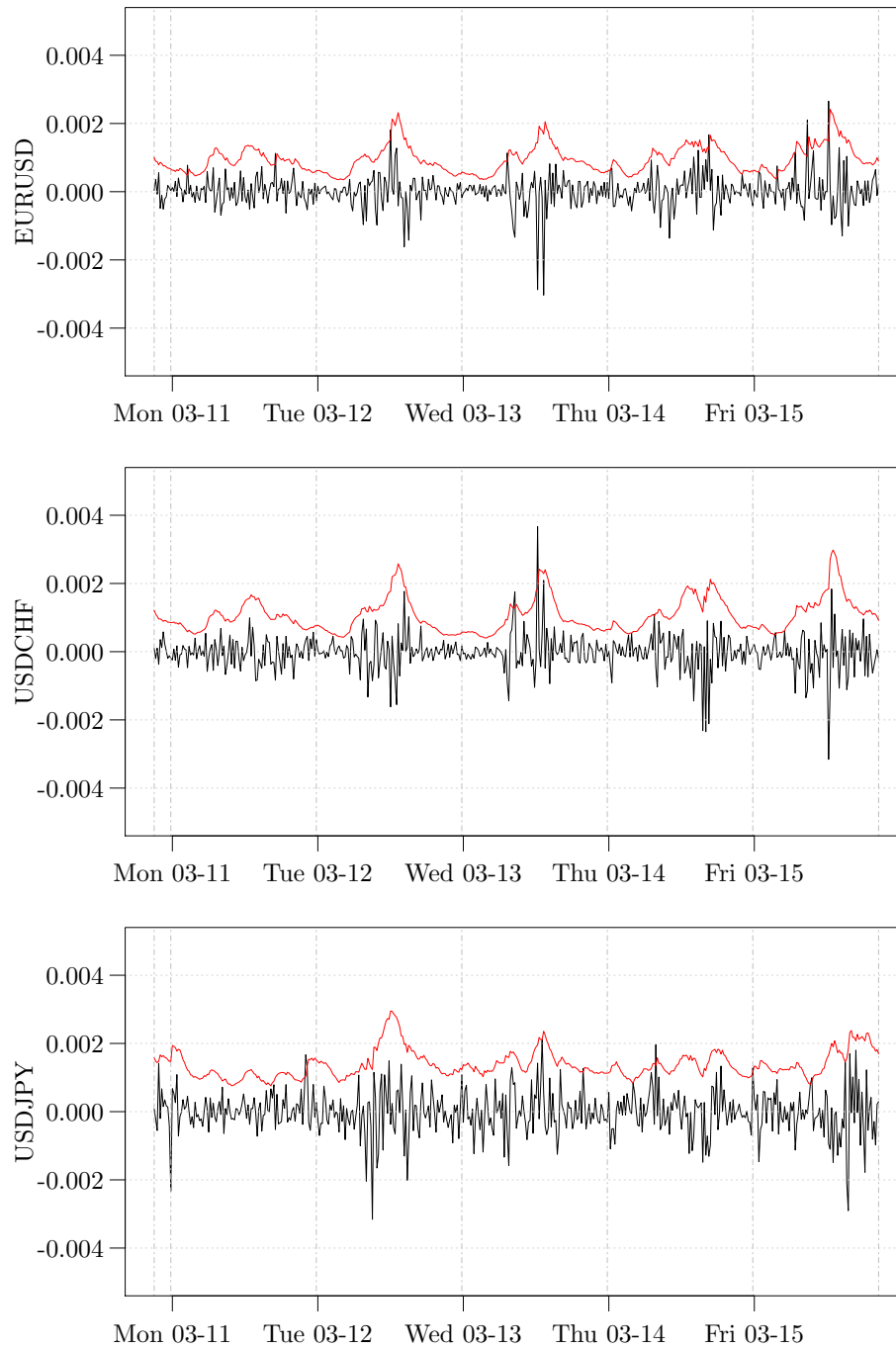
Fig. 3.6 **First week of the foreign exchange rates data:** in each row, the returns time-series is in black, and a two conditional standard deviations time-series is in red.
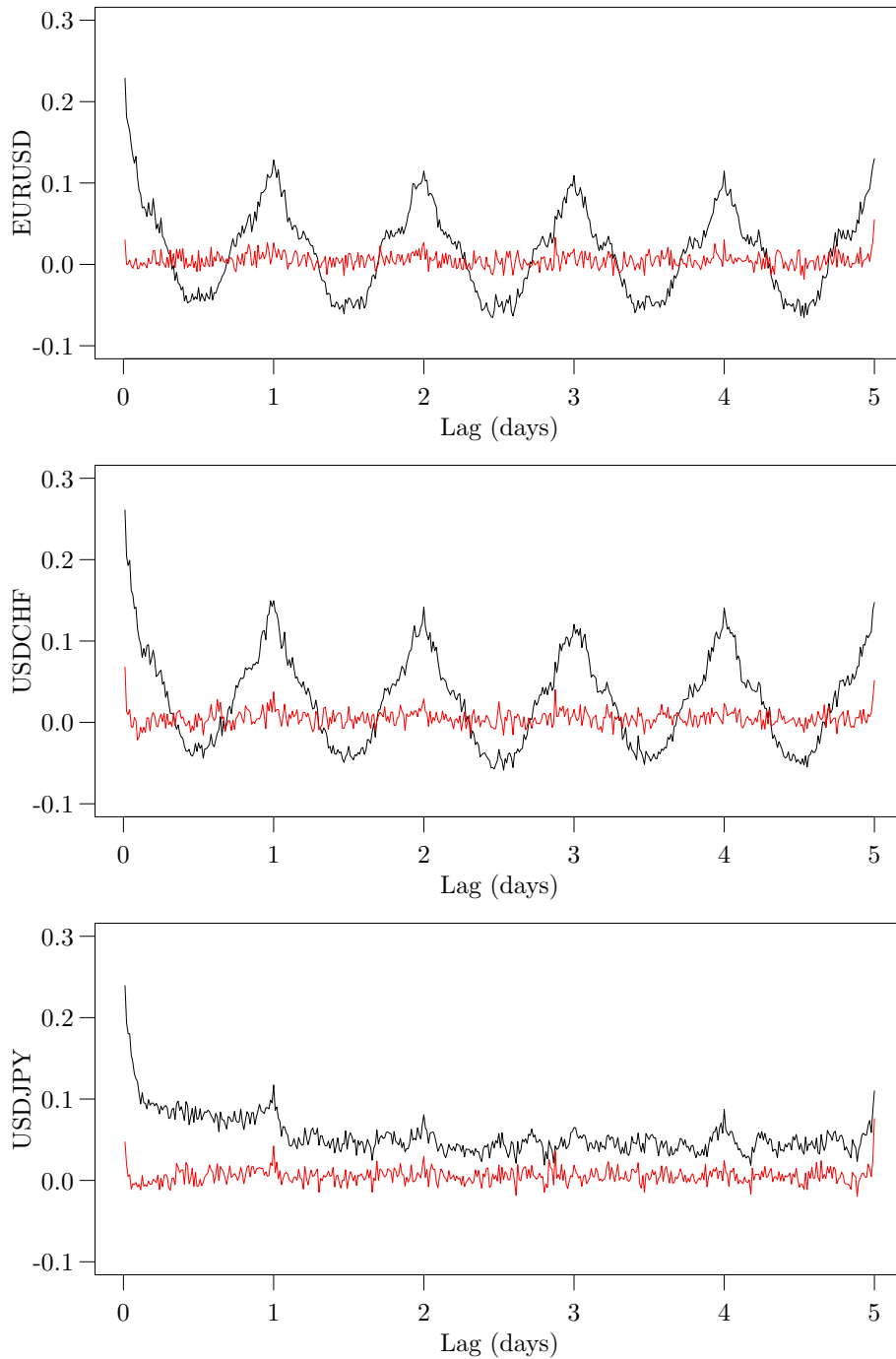
Fig. 3.7 **Autocorrelations of the foreign exchange rates:** in each row, the autocorrelation of the absolute value of the return is in black, and the deseasonalized residual is in red.
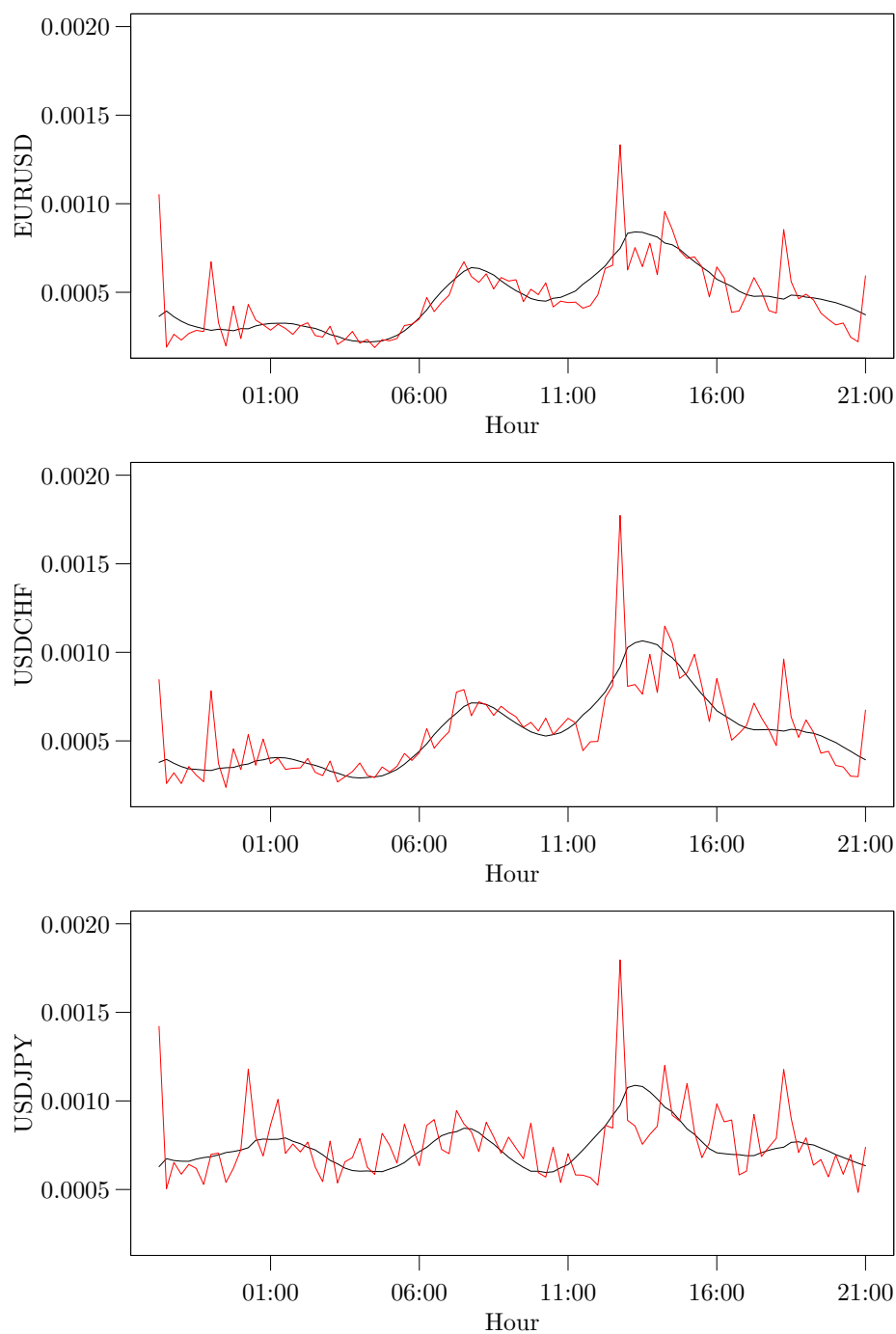
Fig. 3.8 **Periodic component of the foreign exchange rates:** in each row, the black (respectively red) curve represents the empirical (respectively fitted) volatility per 15 minutes bin.

average Kendall's tau is only 0.05, because most of the dependence is captured in the first tree.

In Figure 3.9, we show all smooth and periodic components (without the intercept). The confidence intervals are computed using a parametric bootstrap procedure, similarly as the one devised in the application of Chapter 2. Notice that the first two rows essentially mirror each other because the dependence for the EURUSD-USDCHF is negative while that of the USDCHF-USDJPY is positive. An interesting feature about their periodic components is that the dependence peaks at the opening of the Japanese market, as well as at the opening and closing of the US market, but it seems to be less impacted by the European markets. As for the third copula, its smooth component has less effective degrees of freedom (i.e., it is smoother) because some of the time-varying features are already captured in the first tree. However, unlike in the first tree, the main dependence peak is at the European opening, while that for the Japanese opening and the US opening and closing are slightly smaller.

Overall, when modeling intraday financial data using PCCs, our study suggests that a severe underestimation/overestimation of the dependence may result from failing to factor in its dynamic properties. Furthermore, while most of the time-varying features are captured in the first tree, special care should be taken for the second tree.

3.4.2   The Uranium Exploration Data

In second first application, we revisit the uranium exploration data initially used by Cook and Johnson (1981). The data represent 655 water samples collected in Colorado and contain measurements on the log-concentration (in parts per billion) of seven trace elements: uranium (U), lithium (Li), cobalt (Co), potassium (K), caesium (Cs), scandium (Sc), and titanium (Ti). More recently, Acar et al. (2012) fit a three-dimensional non-simplified PCC to a subset of the data. Their analysis indicates that the simplifying assumption does not hold.

We use the GAM-framework developed in this Chapter to fit a non-simplified PCC to the seven variables. First, we transform the original data to the copula scale by applying the empirical probability integral transform. The transformed data represents pseudo-observations from a seven-dimensional copula. Following Dissmann et al. (2013) for simplified PCCs, the vine tree structure is selected by sequentially maximizing the sum of absolute Kendall's tau, and the resulting first three trees are shown in Figure 3.10. For the selection of smooth components and pair-copula families we apply the approach described in Section 3.2.
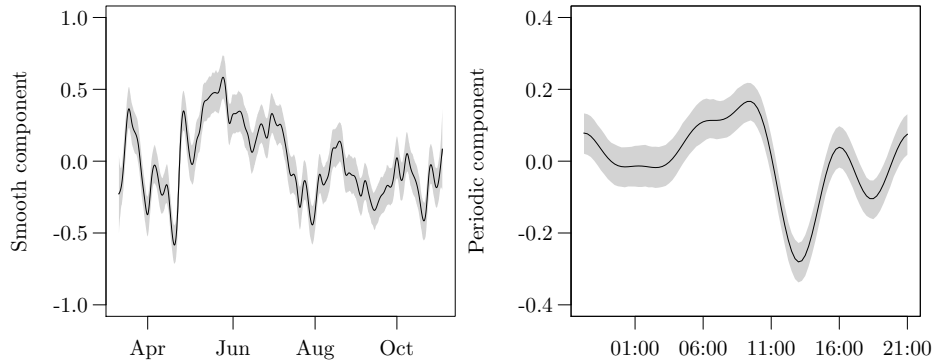
(a) The EURUSD-USDCHF copula.
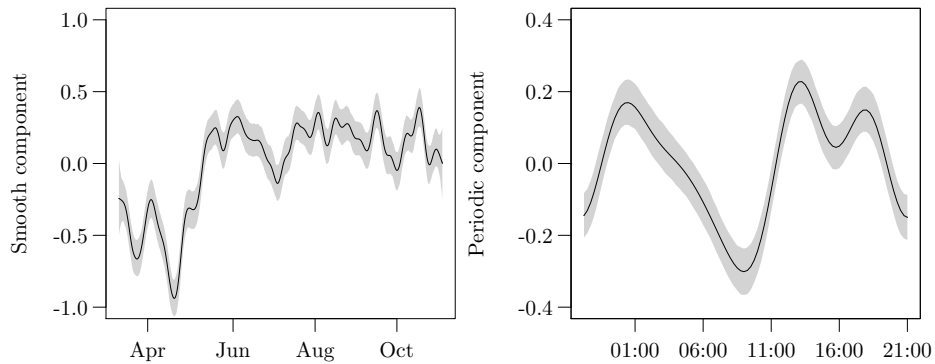
(b) The USDCHF-USDJPY copula.
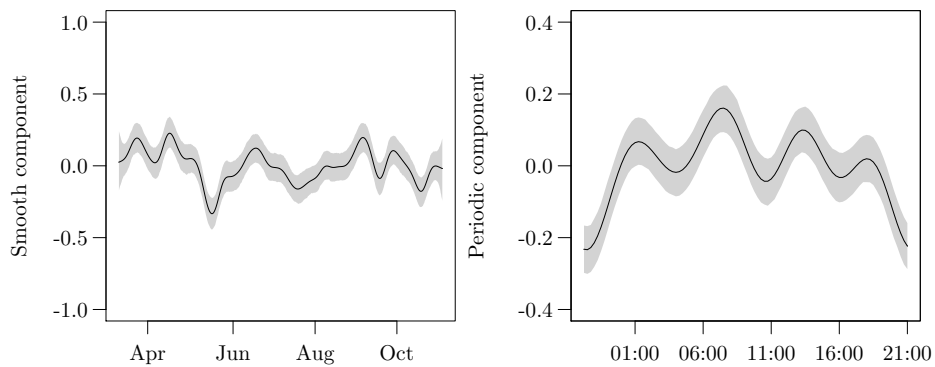
(c) The EURUSD–USDJPY;USDCHF copula.

Fig. 3.9 **Smooth and periodic components of the foreign exchange's PCC:** for the three non-simplified conditional copulas, $\widehat{s}(t)$ (plain line) with 95% c.i. (gray area) is represented in the left column, $\mathbf{x}(t)^\top \widehat{\boldsymbol{\beta}}(t)$ (plain line) with 95% c.i. (gray area) in the right column). Note that all smooths are centered around 0.
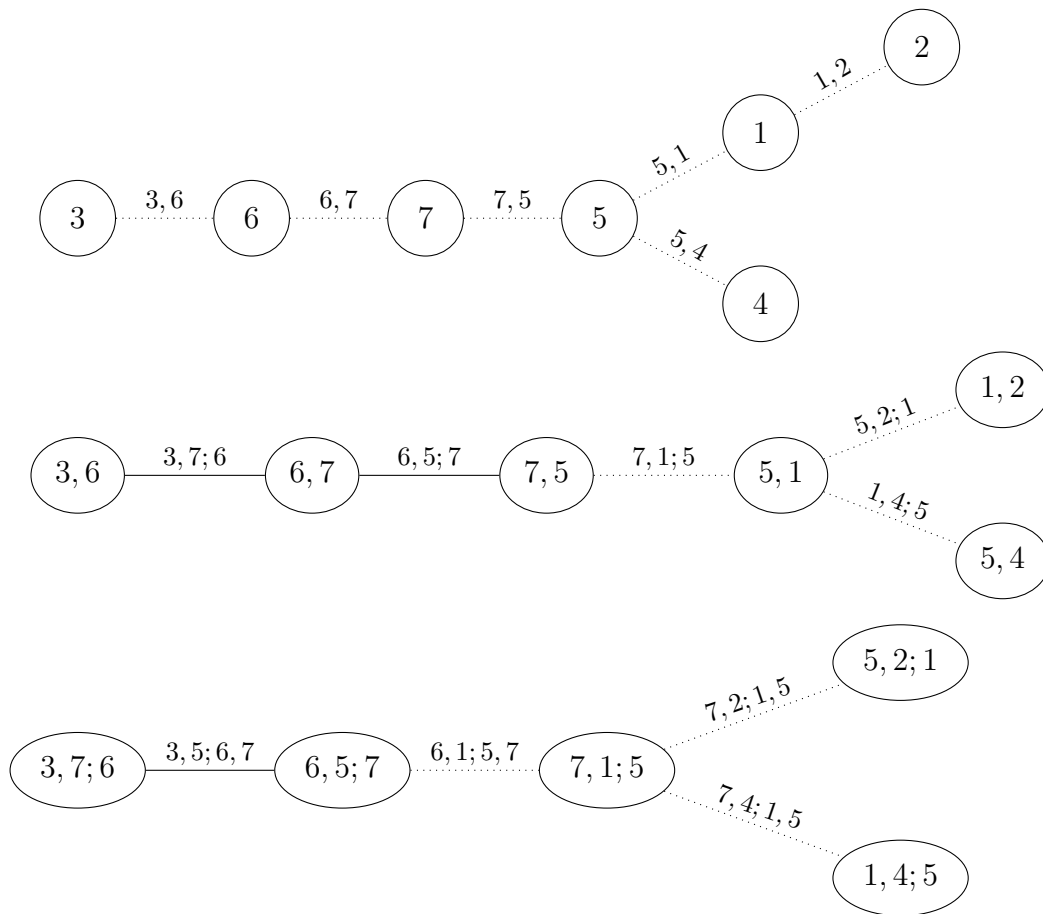
Fig. 3.10 **First three trees of the PCC selected for the uranium data:** dotted lines indicate simplified pair-copulas; solid lines non-simplified ones. The variables are encoded as follows: 1 – U, 2 – Li, 3 – Co, 4 – K, 5 – Cs, 6 – Sc, 7 – Ti.

In the first tree, the conditioning set is empty, so all pair-copulas are necessarily simplified (i.e., unconditional). In the second tree, only two out of five pair-copulas are modeled as non-simplified (indicated by a solid line in Figure 3.10) by the selection algorithm of Section 3.2.1. Then, in the third three, only one pair is modeled as non-simplified. Finally, in the lower trees (i.e, T_4, T_5, T_6), all pair-copulas are modeled as simplified.

For the non-simplified pair-copula, the resulting specification is

$$\psi_{3,7;6}(u_6) = g_{3,7;6}\left\{\beta_{3,7;6} + s_{3,7;6}(u_6)\right\},$$

$$\psi_{5,6;7}(u_7) = g_{5,6;7}\left\{\beta_{5,6;7} + s_{5,6;7}(u_7)\right\},$$

$$\psi_{3,5;6,7}(u_7) = g_{3,5;6,7}\left\{\beta_{3,5;6,7} + s_{3,5;6,7}(u_7)\right\},$$

where $\beta_{3,7;6}, \beta_{5,6;7}, \beta_{3,5;6,7} \in \mathbb{R}$ are intercepts. Note that the scandium (Sc) could have been part of $\psi_{3,5;6,7}$ but was not selected. The functions $s_{3,7;6}$, $s_{5,6;7}$ and $s_{3,5;6,7}$ are shown in Figure 3.11 (the confidence intervals are computed using parametric bootstrap once again). The pair-copula $c_{3,7;6}$, which captures the effect of scandium (Sc) on the dependence between cobalt (Co) and titanium (Ti) and has been investigated before in Acar et al. (2012). Similar to these authors, we find that the dependence declines for large values and has a local maximum for average values of the covariate (Sc). Concerning $s_{5,6;7}$, its estimated equivalent degree of freedom (< 1.5) suggests that a linear function of the covariate (Ti) is appropriate. Finally, since the confidence band is so large for the last non-simplified pair-copula, it is unclear whether the non-simplification really or whether the detected (linear) relationship is spurious.

Concerning all the other pair-copulas, we summarize the selected family and Kendall's tau (with 95% confidence intervals in parenthesis) in Table 3.3. By construction, most of the unconditional dependence is captured in the first tree, where the dependence is high for all selected pair-copulas. Then, the overall captured dependence quickly drops in the lower trees, and it is again unclear whether many of the estimated relationships are spurious or not.

Overall, an interesting effect is that most of the unconditional dependence is captured in the first tree, whereas most of the conditional dependence is captured in the second. This suggests that a more sophisticated model selection may be appropriate: for instance, after the first tree, we could minimize sequentially the sum of the p-values of a conditional independence test.
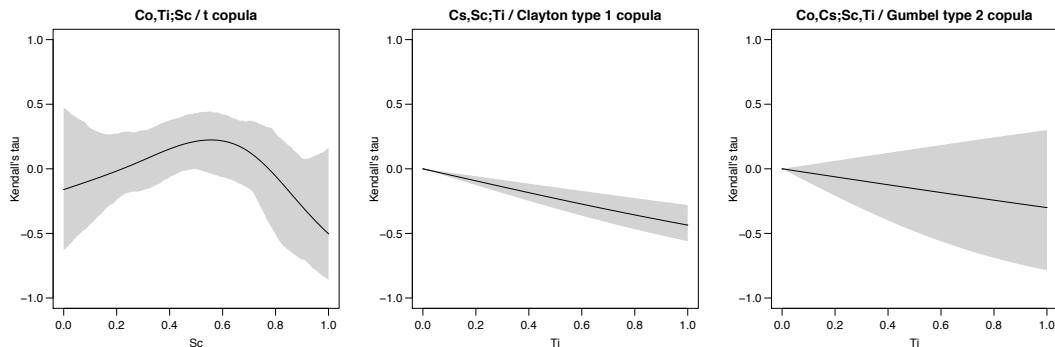
Fig. 3.11 **Smooth functions of the uranium data's PCC:** for the three selected non-simplified conditional copulas, $\widehat{g(s)}$ is the plain line and the 95% c.i. is the grey area.

3.5 Discussion

In this Chapter, we extend our bivariate frameworke to multivariate distribution, while introducing two extensions of the pair-copula constructions (PCCs). First, we take into account the effect of exogenous covariates on the dependence structure. Second, we relax the usual simplifying assumption, letting each copula parameter of the decomposition depend on its conditioning set. Both extensions are achieved using the flexibility of generalized additive models (GAMs), which allow for parametric, semiparametric or nonparametric specifications. Building on the maximum penalized log-likelihood estimator for bivariate copulas of Chapter 2, we propose a sequential estimation algorithm as well as a fully automatic model selection procedure. We evaluate both in a simulation study, and we find that

- the variance is higher in the lower trees, although the efficiency loss of the sequential estimation is small,

- and the performances of the selection and estimation are comparable to that of an oracle estimator (where the structure and copula families are known but not the smooth components).

Finally, we apply the new framework to two real datasets. First, we model the dependence structure of intraday returns on three exchange rates, and we observe that the bivariate results of Chapter 2 are extended directly in this higher-dimensional example. In other words, the data suggests that the dependence can be decomposed into a smooth and a periodic component. Furthermore, the periodic components for each conditional pair-copulas have peak that correspond to openings and/or closings

markets around the world. Finally, while most of the time-varying features are captured in the first tree, there is still a significant amount of periodicity left in the second tree. Second, we revisit the uranium exploration data of Cook and Johnson (1981), where we show empirically that the simplifying assumption. This example is interesting because it highlights that, for real datasets, the non-simplified PCC model can capture most of the unconditional dependence in the first tree, and the conditional dependence in the second.

We conclude this chapter with three observations representing potential directions for further work, although out of the scope of this thesis. First, while we present our framework in the general semiparametric form, the simulations are only concerned with the smooth components. The reason is that automatic model selection procedures for specifications containing parametric and nonparametric terms are seldom studied in the statistics literature. Second, as the dimension of the dataset under study grows, so does the conditioning set of the copulas in the lower trees. To alleviate this issue, we could use a dimensionality reduction method restricted to the unit hypercube. Third, while we used the algorithm of Dissmann et al. (2013) to select the vine structure, it may not be optimal for PCCs with exogenous covariates and/or non-simplified ones. For instance, depending on the application context, we may want to first select the pairs for which the covariates effects are stronger, or alternatively the pairs for which the simplifying assumption is violated the most.

	Li,U	Cs,K	Cs,U
Copula family	t	t	t
Kendall's tau	0.124 (0.072,0.175)	0.196 (0.15,0.245)	0.439 (0.405,0.475)

(a) Tree 1 (first three copulas)

	Co,Sc	Cs,Ti	Sc,Ti
Copula family	t	Gumbel type 2	t
Kendall's tau	0.527 (0.491,0.563)	0.315 (0.264,0.359)	0.429 (0.384,0.47)

(b) Tree 1 (last three copulas)

	Cs,Li;U	K,U;Cs	Ti,U;Cs
Copula family	Clayton type 2	Gaussian	t
Kendall's tau	0.049 (-0.003,0.094)	0.102 (0.05,0.152)	-0.1 (-0.15,-0.044)

(c) Tree 2

	K,Li;Cs,U	K,Ti;Cs,U	Sc,U;Cs,Ti
Copula family	Clayton type 2	t	t
Kendall's tau	0.083 (0.031,0.128)	-0.039 (-0.093,0.018)	-0.04 (-0.091,0.014)

(d) Tree 3

	Li,Ti;Cs,K,U	K,Sc;Cs,Ti,U	Co,U;Cs,Sc,Ti
Copula family	t	Gaussian	Clayton type 1
Kendall's tau	-0.013 (-0.062,0.038)	-0.194 (-0.243,-0.146)	-0.026 (-0.072,0)

(e) Tree 4

	Li,Sc;Cs,K,Ti,U	Co,K;Cs,Sc,Ti,U	Co,Li;Cs,K,Sc,Ti,U
Copula family	Clayton type 2	t	Clayton type 1
Kendall's tau	0.132 (0.092,0.172)	-0.036 (-0.088,0.012)	-0.053 (-0.091,-0.011)

(f) Tree 5 and 6

Table 3.3 **The simplified copulas of the uranium data's PCC:** families and Kendall's tau (with 95% c.i. in parenthesis).

Chapter 4

Code

The `R` (R Core Team 2013) functions for estimation and inference of the models described in Chapters 2 and 3 are collected in the package `gamCopula`, available at https://github.com/tvatter/gamCopula. Currently, `gamCopula` is heavily relying on the `mgcv` (see Wood 2006) and `VineCopula` (see Schepsmeier et al. 2015) packages, as it basically extends and mix both of them. In this Chapter, we describe practical details of our implementation, which revolves around two S4 classes, namely `gamBiCop` and `gamVine`.

4.1   The gamBiCop Class

`gamBiCop` is an S4 class used to store a generalized additive model either for the conditional Kendall's tau (as described in Chapter 2) or directly for the conditional copula parameter. Instances of this class can be created by using the function `gamBiCop()` or calls of the form `new("gamBiCop", ...)`. It contains four slots:

- `family` encodes the copula family.

- `model` is an S3 `gamObject` as returned by the `gam` function from the `mgcv` package.

- `par2` stores the second parameter for the Student's t copula.

- `tau` is `TRUE` for the Kendall's tau specification or `FALSE` for the copula parameter.

For the encoding of the copula family, we follow the convention of the package `VineCopula`. Currently, the Gaussian, Student's t and Frank are implemented using their standard parametrization (as well as the natural mapping with Kendall's tau). For the Clayton and Gumbel, as explained in Chapter 3, a modification is required to

allow for both positive and negative dependence. In order to span the complete $[-1, 1]$ interval for Kendall's tau, we consider the 90/180/270 degrees rotated versions of such copulas. For instance, if $c^{Clay}(u_1, u_2; \eta)$ denotes the Clayton copula with parameter η, then $c^{Clay90}(u_1, u_2; \eta) = c^{Clay}(u_2, 1 - u_1; \eta)$ is its 90 degrees (counter-clockwise) rotation and allows for negative dependence. Hence, the families of *double Archimedean* are obtained by mixing a copula of positive dependence (either the standard or survival, that is the 180 degrees rotated) and one of negative dependence (either the 90 or 270 degrees rotated). As such, additionally to the three families listed above, the `gamCopula` package contains

- the double Clayton type I (standard and rotated 90 degrees),

- the double Clayton type II (standard and rotated 270 degrees),

- the double Clayton type III (survival and rotated 90 degrees),

- the double Clayton type IV (survival and rotated 270 degrees),

- the double Gumbel type I (standard and rotated 90 degrees),

- the double Gumbel type II (standard and rotated 270 degrees),

- the double Gumbel type III (survival and rotated 90 degrees),

- and the double Gumbel type IV (survival and rotated 270 degrees).

There are three main functions related to the `gamBiCop` class, namely `gamBiCopEst`, `gamBiCopSim`, `gamBiCopPred`, respectively allowing to estimate a model, or simulate and predict from a fitted one.

For instance, let us assume a Gaussian copula conditional on a single covariate x with parameter

$$\theta(x) = g(1 + x),$$

where $g(x) = (e^x - 1)/(e^x + 1)$. We obtain 500 normally distributed observations of the covariate x using:

```
set.seed(0)
n <- 5e2
x <- rnorm(n)
```

For now without our package, we simulate from the model described above with:

```
g <- function(x) (exp(x)-1)/(exp(x)+1)
par <- g(1+x)
u <- t(sapply(par, function(par) BiCopSim(N = 1, family = 1, par)))
```

Note that `BiCopSim` is the bivariate copula simulation function from the `VineCopula` package, and we simulate 1 observation for the family 1 (i.e., the Gaussian) for each element of `par`. Before estimating the model, it is necessary to set-up a `dataframe` with the first two columns named `"u1"` and `"u2"`:

```
data <- data.frame(u, x)
names(data)[1:2] <- c("u1", "u2")
```

Finally, we can use `gamBiCopEst` to estimate the model:

```
cop <- gamBiCopEst(data, ~ x, family = 1 tau = FALSE)
```

Notice that the option `tau = FALSE` is used because the model is specified for the copula parameter. As for the second argument, it uses the standard way of specifying formulas in R (in this case a linear model with `x` as predictor). In fact, to obtain a smooth function of the covariate(s), the second argument of `gamBiCopEst` can also be a `gam` formula (which is described in detail in the help pages `?gam`, `?formula.gam` and `?gam.models` from the package `mgcv`).

The output from `gamBiCopEst` is a `list` containing various informations on the fit, as well the model fit (element `res` of the list), from which a summary can be obtained, in this case by calling `summary(cop[["res"]])`:

```
Gaussian copula with par(z) = (exp(z)-1)/(exp(z)+1) where
Formula:
 z ~ x


 Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
 (Intercept)  0.99429    0.07704   12.91   <2e-16 ***
 x            0.95659    0.07219   13.25   <2e-16 ***
 ---
 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



 R-sq.(adj) =  0.257   Deviance explained = 26.1%
 GCV = 0.93863  Scale est. = 0.93487   n = 500
```

Other methods such as `logLik`, `AIC`, `BIC`, `EDF` and `plot` are also available for this class. Furthermore, using `gamBiCopSim`, `gamBiCopPred`, `gamBiCopPDF` and `gamBiCopCDF`, it is then possible to simulate new observations, predict the calibration function, the parameter and/or Kendall's tau, compute the density or cumulative distribution from this fitted object. Note that those functions use a `gamBiCop` object as first argument and `newdata` as (optional) second argument, which, if provided, should contain all the variables required by the `gam` formula. Finally, the model selection algorithm described in Chapter 3 is implemented in the function `gamBiCopSel`.

4.1.1 A Complete Example

In this section, we give a complete example in a set-up similar to the simulation study of Chapter 2. We first load the required packages and set the seed for reproducibility:

```
require(copula)
require(mgcv)
set.seed(0)
```

Then, we set the simulation parameters (sample size, correlation between covariates, and Gaussian copula family):

```
n <- 5e2
rho <- 0.5
fam <- 1
```

We create a function factory for a copula parameter depending on three variables:

```
eta0 <- 1
calib.surf <- list(
  calib.quad <- function(t, Ti = 0, Tf = 1, b = 8) {
    Tm <- (Tf - Ti)/2
    a <- -(b/3) * (Tf^2 - 3 * Tf * Tm + 3 * Tm^2)
    return(a + b * (t - Tm)^2)},
  calib.sin <- function(t, Ti = 0, Tf = 1, b = 1, f = 1) {
    a <- b * (1 - 2 * Tf * pi/(f * Tf * pi +
                                 cos(2 * f * pi * (Tf - Ti))
                               - cos(2 * f * pi * Ti)))
    return((a + b)/2 + (b - a) * sin(2 * f * pi * (t - Ti))/2)},
  calib.exp <- function(t, Ti = 0, Tf = 1, b = 2, s = Tf/8) {
```

```
    Tm <- (Tf - Ti)/2
    a <- (b * s * sqrt(2 * pi)/Tf) * (pnorm(0, Tm, s) -
                                      pnorm(Tf, Tm, s))
    return(a + b * exp(-(t - Tm)^2/(2 * s^2)))})
```

We simulate a 3-dimensional matrix of covariates:

```
covariates.distr <- mvdc(normalCopula(rho, dim = 3),
                         c("unif"),
                         list(list(min = 0, max = 1)),
                         marginsIdentical = TRUE)
X <- rMvdc(n, covariates.distr)
```

Using the covariates, we simulate from the conditional model to obtain the copula
observations:

```
U <- condBiCopSim(fam, function(x1,x2,x3) {
  eta0+sum(mapply(function(f,x) f(x),
                  calib.surf,
                  c(x1,x2,x3)))},
  X[,1:3], par2 = 6, return.par = TRUE)
```

We merge the copula observations and covariates and give appropriate names to the
columns of the dataset:

```
data <- data.frame(U$data,X)
names(data) <- c(paste("u",1:2,sep=""),paste("x",1:3,sep=""))
```

We first fit a model fit with a basis size (arguably) too small and unpenalized cubic
spines:

```
pen <- FALSE
basis0 <- c(3, 4, 4)
formula <- ~s(x1, k = basis0[1], bs = "cr", fx = !pen) +
  s(x2, k = basis0[2], bs = "cr", fx = !pen) +
  s(x3, k = basis0[3], bs = "cr", fx = !pen)
system.time(fit0 <- gamBiCopEst(data, formula, fam))
```

Then we fit a model fit with a better basis size and penalized cubic splines (via GCV
minimization):

```
pen <- TRUE
basis1 <- c(3, 10, 10)
formula <- ~s(x1, k = basis1[1], bs = "cr", fx = !pen) +
  s(x2, k = basis1[2], bs = "cr", fx = !pen) +
  s(x3, k = basis1[3], bs = "cr", fx = !pen)
system.time(fit1 <- gamBiCopEst(data, formula, fam))
```

We extract the `gamBiCop` objects and show various methods:

```
res <- sapply(list(fit0,fit1), function(fit){fit$res})
metds <- list('logLik'=logLik,'AIC'=AIC,'BIC'=BIC,'EDF'=EDF)
lapply(res, function(x) sapply(metds, function(f) f(x)))
```

In this particular case, the last line returns:

```
[[1]]
[[1]]$logLik
'log Lik.' 267.9789 (df=9)

[[1]]$AIC
[1] -517.9578

[[1]]$BIC
[1] -480.0264

[[1]]$EDF
[1] 1 2 3 3


[[2]]
[[2]]$logLik
'log Lik.' 298.8658 (df=16.14058)

[[2]]$AIC
[1] -565.4505

[[2]]$BIC
[1] -497.4242
```

```
[[2]]$EDF
[1] 1.000000 1.999076 7.406505 5.735001
```

We can finally compare between the fitted and true model for each of the three smooth functions and basis choice, and display the results (see Figure 4.1):

```
fitted <- lapply(res, function(x)
  gamBiCopPred(x, data.frame(x1=u,x2=u,x3=u), type = "terms")$calib)

par(mfrow = c(1, 3), pty = "s")
yy <- c(-2,1.5)
for(k in 1:3){
  plot(u, true[[k]]$true, type = "l", ylim = yy,
       xlab = paste("Covariate",k), ylab = paste("Smooth",k),
       cex.axis = 1.5, cex.lab = 1.5, lwd = 2)
  lines(u, fitted[[1]][, k], col = "red", lwd = 2)
  lines(u, fitted[[2]][, k], col = "green", lwd = 2)
}
```
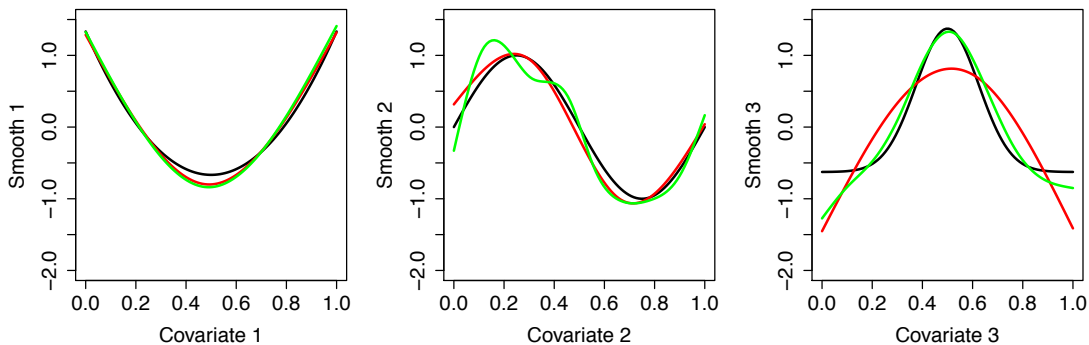

Fig. 4.1 **A complete example on `gamBiCopEst`:** in each panel, the true underlying function is in black, the fitted model with cubic splines and a fixed (small) basis is in red and the fitted model with cubic splines and a larger basis penalized via GCV minimizationgreen is in green.

4.2   The gamVine Class

`gamVine` is an S4 class used to store a conditional and potentially non-simplified pair-copula construction (as described in Chapter 3). Instances of this class can be

created by using the function `gamVine()` or calls of the form `new("gamVine", ...)`.
It contains four slots:

- `Matrix` is a lower triangular $d \times d$ matrix that defines the tree structure.

- `model` is a list containing $d \times (d-1)/2$ elements, with each element being either
  a list with three numeric items (`family`, `par` and `par2`) or an object of class
  `gamBiCop`.

- `names` is a vector of d names.

- `covariates` is an (optional) vector of names for the exogenous covariates.

Similarly as for the `gamBiCop` class, the `gamCopula` package deals with the `gamVine`
class via functions allowing to estimate a model or simulate from a fitted one. However,
the estimation part is arguably more complex because of the (often) required structure
selection (i.e., which pairs should be included). We build on the insights of the
`VineCopula` package and propose three estimation methods:

- `gamVineSeqEst`, like `RVineSeqEst`, assumes that both the structure and model
  for each pair are known. It takes a dataset and `gamVine` object as first and
  second inputs and uses the first input to estimate sequentially the model from
  the second input.

- `gamVineCopSelect`, like `RVineCopSelect`, assumes that only the structure is
  known. It takes a dataset and a `Matrix` object as first and second inputs and
  uses the first input to estimate sequentially the model from the second input.
  Because only the structure (i.e., which pairs should be included) is known, the
  function uses `gamBiCopSel` for the model selection of each pair-copula.

- `gamVineStructureSelect`, like `RVineStructureSelect`, assumes that nothing
  is known. It takes a dataset and uses it sequentially select and estimate the
  model. Similarly as `gamVineCopSelect`, it uses `gamBiCopSel` for the pair-copula
  selection. Furthermore, the structure is selected as in `RVineStructureSelect`,
  namely using maximum spanning trees with absolute values of pairwise Kendall's
  taus as weights (see Dissmann et al. 2013 for more details).

Once the model is estimated, one can simulate from a `gamVine` object by using
`gamVineSim`.

4.2.1 A Complete Example

In this section, we give a complete example in a set-up similar to the simulation study of Chapter 3. We first load the required packages and set the seed for reproducibility:

```
require(copula)
require(mgcv)
set.seed(0)
```

Then, we set the simulation parameters (sample size and all copula families allowed):

```
n <- 1e3
fam <- c(1:2,301:304,401:404)
```

Next, we define a 4-dimensional R-vine tree structure matrix

```
d <- 4
Matrix <- c(2,3,4,1,0,3,4,1,0,0,4,1,0,0,0,1)
Matrix <- matrix(Matrix,d,d)
nnames <- paste("X", 1:d, sep = "")
```

We create a function factory for a copula parameter depending on three variables:

```
eta0 <- 1
calib.surf <- list(
  calib.quad <- function(t, Ti = 0, Tf = 1, b = 8) {
    Tm <- (Tf - Ti)/2
    a <- -(b/3) * (Tf^2 - 3 * Tf * Tm + 3 * Tm^2)
    return(a + b * (t - Tm)^2)},
  calib.sin <- function(t, Ti = 0, Tf = 1, b = 1, f = 1) {
    a <- b * (1 - 2 * Tf * pi/(f * Tf * pi +
                                 cos(2 * f * pi * (Tf - Ti))
                               - cos(2 * f * pi * Ti)))
    return((a + b)/2 + (b - a) * sin(2 * f * pi * (t - Ti))/2)},
  calib.exp <- function(t, Ti = 0, Tf = 1, b = 2, s = Tf/8) {
    Tm <- (Tf - Ti)/2
    a <- (b * s * sqrt(2 * pi)/Tf) * (pnorm(0, Tm, s) -
                                        pnorm(Tf, Tm, s))
    return(a + b * exp(-(t - Tm)^2/(2 * s^2)))})
```

In other to create a `gamVine` object, we first define gam-vine model list:

```
count <- 1
model <- vector(mode = "list", length = d*(d-1)/2)
sel <- seq(d,d^2-d, by = d)
```

For the first tree, a family and its (uncondictional parameter) is then randomly selected:

```
for (i in 1:(d-1)) {
  # Select a copula family
  family <- sample(familyset, 1)
  model[[count]]$family <- family

  # Use the canonical link and a randomly generated parameter
  if (is.element(family,c(1,2))) {
    model[[count]]$par <- tanh(rnorm(1)/2)
    if (family == 2) {
      model[[count]]$par2 <- 2+exp(rnorm(1))
    }
  } else {
    if (is.element(family,c(401:404))) {
      rr <- rnorm(1)
      model[[count]]$par <- sign(rr)*(1+abs(rr))
    } else {
      model[[count]]$par <- rnorm(1)
    }
    model[[count]]$par2 <- 0
  }
  count <- count + 1
}
```

For trees 2 and 3, we select, for each edge, one copula family, and then for each conditioning variable, one function from the factory. Then, we approximate this function using a `gam` object from the `mgcv` package that we use as the basic building bloc of the model:

```
# A dummy dataset
data <- data.frame(u1 = runif(1e2), u2 = runif(1e2),
x = matrix(runif(1e2*d),1e2,d))
for(j in 2:(d-1)){
```

```r
for(i in 1:(d-j)){
  # Select a copula family
  family <- sample(familyset, 1)

  # Select the conditioning set and create a model formula
  cond <- nnames[sort(Matrix[(d-j+2):d,i])]
  tmpform <- paste("~",paste(paste("s(", cond, ", k=10, bs='cr')",
                                        sep = ""), collapse=" + "))
  l <- length(cond)
  temp <- sample(3, l, replace = TRUE)

  # Spline approximation of the true function
  m <- 1e2
  x <- matrix(seq(0,1,length.out=m), nrow = m, ncol = 1)
  if(l != 1){
    tmp.fct <- paste("function(x){eta0+",
                      paste(sapply(1:l, function(x)
                        paste("calib.surf[[",temp[x],"]](x[",x,"])",
                              sep="")), collapse="+"),"}",sep="")
    tmp.fct <- eval(parse(text = tmp.fct))
    x <- eval(parse(text = paste0("expand.grid(", paste0(rep("x",l),
      collapse = ","),")", collapse = "")))
    y <- apply(x,1,tmp.fct)
  }else{
    tmp.fct <- function(x) eta0+calib.surf[[temp]](x)
    colnames(x) <- cond
    y <- tmp.fct(x)
  }

  # Estimate the gam model
  form <- as.formula(paste0("y", tmpform))
  dd <- data.frame(y, x)
  names(dd) <- c("y", cond)
  b <- gam(form, data = dd)
  #plot(x[,1],(y-fitted(b))/y)
```

```
    # Create a dummy gamBiCop object
    tmp <- gamBiCopEst(data = data, formula = form,
      family = 1, n.iters = 1)$res

    # Update the copula family and the model coefficients
    attr(tmp, "model")$coefficients <- coefficients(b)
    attr(tmp, "model")$smooth <- b$smooth
    attr(tmp, "family") <- family
    if (family == 2) {
      attr(tmp, "par2") <- 2+exp(rnorm(1))
    }
    model[[count]] <- tmp
    count <- count+1
  }
}
```

Finally, we can create the `gamVine` object:

```
GVC <- gamVine(Matrix=Matrix,model = model,names=nnames)
print(GVC)
```

In this particular case, the last line returns:

```
GAM-Vine matrix:
     [,1] [,2] [,3] [,4]
[1,]    2    0    0    0
[2,]    3    3    0    0
[3,]    4    4    4    0
[4,]    1    1    1    1


 Where
1  <->  X1
2  <->  X2
3  <->  X3
4  <->  X4


 Tree 1:
X2,X1: Gumbel type 3 (survival and 90 degrees rotated)
```

```
X3,X1: Clayton type 4 (survival and 270 degrees rotated)
X4,X1: Gumbel type 3 (survival and 90 degrees rotated)

 Tree 2:
X2,X4|X1 : Gumbel type 4 (survival and 270 degrees rotated) copula
with tau(z) = (exp(z)-1)/(exp(z)+1) where
z ~ s(X1, k = 10, bs = "cr")
X3,X4|X1 : Gumbel type 4 (survival and 270 degrees rotated) copula
with tau(z) = (exp(z)-1)/(exp(z)+1) where
z ~ s(X1, k = 10, bs = "cr")

 Tree 3:
X2,X3|X4,X1 : Clayton type 4 (survival and 270 degrees rotated) copula
with tau(z) = (exp(z)-1)/(exp(z)+1) where
z ~ s(X1, k = 10, bs = "cr") + s(X4, k = 10, bs = "cr")
```

We can now simulate from the `gamVine` object:

```
sim <- gamVineSim(n, GVC)
```

We can also try various estimation methods:

```
fitGVC <- gamVineSeqEst(sim, GVC, verbose = TRUE)
fitGVC2 <- gamVineCopSelect(sim, Matrix, verbose = TRUE)
```

Here, by taking `GVC` as an input, `gamVineSeqEst` assumes that we now the model and simply re-estimate the parameters using the dataset. However, `gamVineCopSelect` takes only the R-vine matrix as an input, and performs automatically for each edge of the PCC the selection of the copula family, of the covariates, and of the basis size for each smooth component. Finally, we can plot the results (see Figure 4.2):

```
par(mfrow=c(3,4))
plot(GVC, ylim = c(-2.5,2.5))

plot(fitGVC, ylim = c(-2.5,2.5))

plot(fitGVC2, ylim = c(-2.5,2.5))
```
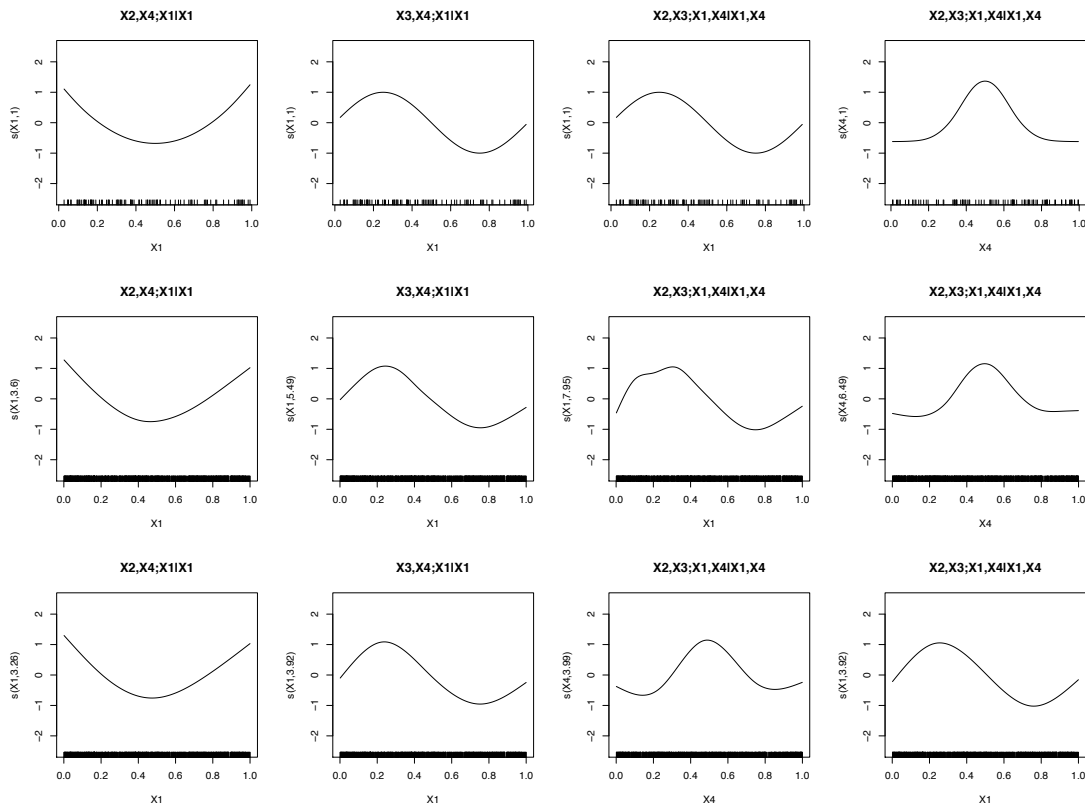
Fig. 4.2 **A complete example on the `gamVine` class:** in each row, the four smooth components are displayed. The first row corresponds to the true model, the second row to the output of `gamVineSeqEst`, and the third row to the output of `gamVineCopSelect`. In this case, notice that the ordering of the smooth functions in the third tree is different for `gamVineCopSelect` (the smooth function of X_4 appear before that of X_1). Furthermore, because the basis size is fixed by `GVC` in `gamVineSeqEst`, there is no a priori reason to believe that this size is optimal for this particular dataset. This explains why, in the third tree, the smooth function of the first covariate (i.e., $s(X_1, \cdot)$) estimated by `gamVineCopSelect` looks better in this particular case.

Chapter 5

Conclusion

In this thesis, we developed tools to model the influence of predictors on multivariate distributions. Our regression-like theory of the dependence, being built on conditional copulas and generalized additive models, is at the same time theoretically sound and practically useful. Both the bivariate dependence structures (Chapter 2) and pair-copula constructions (Chapter 3) open new possibilities for further applied and methodological work, of which we provide a non-exhaustive list in this chapter. Furthermore, the `R` implementation of all the methods suggested in this thesis (and more) being publicly available at https://github.com/tvatter/gamCopula, any researcher from either statistics or an applied domain can use it and/or improve on it. Note that, because Chapter 2 and 3 have their own concluding sections, we avoid repeating again what has been done, in order to focus on what lies ahead.

In the bivariate context, our univariate GAM framework comes short for multi-parameters copulas when several parameters are functions of the covariates. Note that, for common copulas, an orthogonal reparametrization à la Cox and Reid (1987) is seldom feasible analytically. As suggested in Chapter 2, a multi-dimensional smoothing framework, such as vector generalized additive models (VGAM, Yee and Wild 1996), represents a potential solution. After discussions and exchanges of code with the author of the `R` package `VGAM`, this idea is implemented for the Student's t-copula. However, a systematic study of this approach has yet to be devised. Furthermore, even when treating every parameter but one as nuisance, the numerical behavior of most multi-parameters bivariate families is unknown. This is why they are still missing from the `gamCopula` toolbox, whose only multi-parameters family is the Student's t.

Useful to the bivariate dependence structures, and necessary to (even moderate-dimensional) pair-copulas constructions, better automatic model selection methods for generalized additive models are desperately needed. In fact, tackling this issue,

which is not strictly related to conditional copulas, is more difficult than it appears when leaving the "linear world" (i.e., when allowing for smooth components). For instance, traditional methods usually fail to detect a quadratic function of a covariate. Furthermore, when non-linearity is assumed, then what basis size (or equivalently, what bandwidth in a local-likelihood or kernel context) should we choose to test for multiple covariates effects? And what if some of the additive components are linear and other are smooth? Given the complexity of this problem, it may be surprising that the crude algorithm suggested in Chapter 3 performs well in our simulations and applications. Because Marra and Wood (2011) is, to the best of our knowledge, the only publication on the subject, automatic model selection for GAMs is an important direction of further research to make them (even more) practically useful.

While diagnostics checks are available for GAMs, they have not been explored in the framework developed in this thesis. In the copula context, it is possible to use $\mathbf{y}^{[l]} - \mathbf{s}^{[l]}(\boldsymbol{\gamma})\mathbf{y}^{[l]}$ at convergence as "pseudo-residuals" (see Chapter 2 for the definition of the "pseudo-data" $\mathbf{y}^{[l]}$, and the hat-matrix $\mathbf{s}^{[l]}(\boldsymbol{\gamma})$). The pseudo-residuals can then be plotted against the linear predictor, or used in other diagnostic checks. Furthermore, the pseudo-data at convergence can also be plotted against $\mathbf{s}^{[l]}(\boldsymbol{\gamma})\mathbf{y}^{[l]}$, the "fitted response", to evaluate the quality of the fit. However, it is not clear how methods developed for distributions from the exponential family perform when applied to (conditional) copulas. Similarly, other GAM-related issues have not been considered in this thesis. First, the location of the knots was taken as equally spaced, but other settings (such as location based on quantiles of the covariate's distribution) could be useful. Second, we only used cubic splines, but higher-order (e.g., quartic) splines or other types (e.g., thin-plate splines) could also be considered.

Another related issue for PCCs is that the dimension of the conditioning set for each edge in the lower trees grows with the dimension of the dataset. Hence, for non-simplified constructions, this set, from which one has to select covariates, quickly becomes prohibitively large. As explained in the discussion of Chapter 3, this issue could be alleviated using dimensionality reduction restricted in the unit hypercube. However, when considering anything but the Euclidean space, there seems to be a lack of tailor-made approach. By this, we do not mean that principal component analysis, linear discriminant analysis or more modern methods such as t-distributed stochastic neighbor embedding cannot be applied to, for instance, categorical data or data restricted to any cartesian product of (open, or closed, or neither) intervals. Traditional methods are often applied with success to this kind of data, but they lack theoretical foundations. In other words, while heuristics may work, they are not

equivalent at exploiting the fundamental properties of the underlying space that is being "reduced". For instance, principal component analysis may sometimes be applied with success to data restricted to the unit hypercube, provided that the underlying copula is close to Gaussian. However, a sample reconstructed from a small number of components is in general not constrained to the unit hypercube.

Finally, we close this thesis with two potential applications that have been discussed with and sparked the interest of researchers from both climate science and financial econometrics. First, researchers from the Department of Civil Engineering of the Technical University of Madrid have a dataset of 60 annual maxima of hydrographs peak-flow and volume for 33 sites in the Ebro catchment. Their joint modeling is of particular interest to design and manage risks for hydraulic structures like dams, because the T-year flood at a given location is a function of both quantities. Furthermore, in this semi-arid region characterized by large hydro-climatic heterogeneities, the inclusion of additional descriptors (covariates) in the model is necessary; especially as the ultimate goal is the extrapolation (prediction) at ungauged sites. Note that modeling spatio-temporal maxima require the most recent tools of extreme value analysis. Furthermore, conditional copulas depending on the same set of covariates naturally link the two univariate models in a bivariate model for the joint distribution of spatio-temporal maxima. Second, while Andersen and Bollerslev (1997, 1998) first suggested to use the Fourier Flexible Form to capture daily periodic patterns, it is arguably their econometrics analysis of the impact of scheduled economic news on exchange rates volatility that promoted the two papers. While the authors lacked the methodology to go a step further, studying the impact of economic news on the dependence structure between intraday returns can be achieved using the framework developed in this thesis. Generally related to the usefulness of GAMs applied to financial data, it should be noted that the time-varying models developed in Chapters 2 and 3 are purely descriptive. While GAMs are useful data exploration tools, splines only allow for (often linear) extrapolation. Hence, in a time-series context, including a dynamic component (e.g., an autoregressive term) in the dependence structure would be required for prediction. Finally, an interesting alternative to the approach of Chapters 2 and 3 would be to first filter the linear component of the dependence structure through a multivariate GARCH, and then use a GAM to model the non-linear component.

References

Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182–198.

Acar, E. F., Craiu, R. V., and Yao, F. (2011). Dependence calibration in conditional copulas: a nonparametric approach. *Biometrics*, 67(2):445–453.

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

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

Ackerer, D. and Vatter, T. (2015). Valuation of multi-name credit derivatives. Unpublished manuscript.

Andersen, T. G. and Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, 4(2-3):115–158.

Andersen, T. G. and Bollerslev, T. (1998). Deutsche Mark - Dollar Volatility : Intraday Activity Patterns , Macroeconomic Announcements, and Longer Run Dependencies. *The Journal of Finance*, 53(1):219–265.

Bauer, A., Czado, C., and Klein, T. (2012). Pair-copula constructions for non-gaussian dag models. *Canadian Journal of Statistics*, 40(1):86–109.

Bedford, T. and Cooke, R. M. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, 32:245–268.

Bedford, T. and Cooke, R. M. (2002). Vines - a new graphical model for dependent random variables. *Annals of Statistics*, 30:1031–1068.

Brechmann, E. C., Czado, C., and Aas, K. (2012). Truncated regular vines in high dimensions with application to financial data. *Canadian Journal of Statistics*, 40(1):68–85.

Chen, X., Fan, Y., and Tsyrennikov, V. (2006). Efficient Estimation of Semiparametric Multivariate Copula Models. *Journal of the American Statistical Association*, 101(475):1228–1240.

Cook, R. D. and Johnson, M. E. (1981). A family of distributions for modelling non-elliptically symmetric multivariate data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 43(2):210–218.

Cox, D. R. and Reid, N. (1987). Parameter Orthogonality and Approximate Conditional Inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 49(1):1–39.

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

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik*, 31:377–403.

Czado, C. (2010). Pair-copula constructions of multivariate copulas. In Jaworski, P., Durante, F., Härdle, W. K., and Rychlik, T., editors, *Copula Theory and Its Applications*, Lecture Notes in Statistics, pages 93–109. Springer Berlin Heidelberg.

de Boor, C. (2001). *A Practical Guide to Splines - Revised Edition*. Springer-Verlag, New York.

Dissmann, J., Brechmann, E. C., Czado, C., and Kurowicka, D. (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics and Data Analysis*, 59:52–69.

Embrechts, P., McNeil, A., and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. *Risk management: value at risk and beyond*, pages 176–223.

Engle, R. F. and Sokalska, M. E. (2012). Forecasting intraday volatility in the US equity market. Multiplicative component GARCH. *Journal of Financial Econometrics*, 10(1):54–83.

Erhardt, T. M., Czado, C., and Schepsmeier, U. (2015a). R-vine models for spatial time series with an application to daily mean temperature. *Biometrics*, 71(2):323–332.

Erhardt, T. M., Czado, C., and Schepsmeier, U. (2015b). Spatial composite likelihood inference using local c-vines. *Journal of Multivariate Analysis*, 138:74–88. High-Dimensional Dependence and Copulas.

Fermanian, J.-D. and Wegkamp, M. H. (2012). Time-dependent copulas. *Journal of Multivariate Analysis*, 110:19–29.

Gallant, R. (1981). On the bias in flexible functional forms and an essentially unbiased form: The fourier flexible form. *Journal of Econometrics*, 15(2):211–245.

Genest, C., Ghoudi, K., and Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82(3):543–552.

Genest, C. and Werker, B. J. (2002). Conditions for the Asymptotic Semiparametric Efficiency of an Omnibus Estimator of Dependence Parameters in Copula Models. In Cuadras, C. M., Fortiana, J., and Rodríguez-Lallena, J. A., editors, *Distributions With Given Marginals and Statistical Modelling*. Springer Science & Business Media.

Gijbels, I., Veraverbeke, N., and Omelka, M. (2011). Conditional copulas, association measures and their applications. *Computational Statistics and Data Analysis*, 55(5):1919–1932.

Gräler, B. (2014). Modelling skewed spatial random fields through the spatial vine copula. *Spatial Statistics*, 10:87–102.

Green, P. J. (1984). Iteratively Reweighted Least Squares for Maximum Likelihood Estimation, and some Robust and Resistant Alternatives. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 46(2).

Green, P. J. (1987). Penalized likelihood for general semi-parametric regression models. *International Statistical Review/Revue Internationale de Statistique*, 55(3):245–259.

Green, P. J. and Silverman, B. W. (2000). *Nonparametric Regression and Generalized Linear Models – A roughness penalty approach*. Chapman and Hall/CRC.

Gu, C. and Wahba, G. (1991). Minimizing GCV/GML Scores with Multiple Smoothing Parameters via the Newton Method. *SIAM Journal on Scientific and Statistical Computing*, 12(2):383–398.

Haff, I., Aas, K., and Frigessi, A. (2010). On the simplified pair-copula construction - simply useful or too simplistic? *Journal of Multivariate Analysis*, 101(5):1296–1310.

Hastie, T. and Tibshirani, R. (1986). Generalized Additive Models. *Statistical Science*, 1(3):297–310.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. Chapman and Hall, London.

Hobæk Haff, I. (2013). Parameter estimation for pair-copula constructions. *Bernoulli*, 19(2):462–491.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman and Hall/CRC.

Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, 94(2):401–419.

Joe, H., Li, H., and Nikoloulopoulos, A. K. (2010). Tail dependence functions and vine copulas. *Journal of Multivariate Analysis*, 101(1):252–270.

Lamouille, S., Xu, J., and Derynck, R. (2014). Molechular mechanisms of epithelial-mesenchymal transition. *Nature Reviews Molecular Cell Biology*, 15(3):178–196.

Marra, G. and Wood, S. N. (2011). Practical variable selection for generalized additive models. *Computational Statistics & Data Analysis*, 55(7):2372 – 2387.

Marra, G. and Wood, S. N. (2012). Coverage properties of confidence intervals for generalized additive model components. *Scandinavian Journal of Statistics*, 39(1):53–74.

McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques, Tools*. Princeton University Press.

McNeil, A. J. and Nešlehová, J. (2009). Multivariate archimedean copulas, d-monotone functions and ℓ_1-norm symmetric distributions. *The Annals of Statistics*, pages 3059–3097.

Min, A. and Czado, C. (2010). Bayesian inference for multivariate copulas using pair-copula constructions. *Journal of Financial Econometrics*, 8(4):511–546.

Min, A. and Czado, C. (2011). Bayesian model selection for d-vine pair-copula constructions. *Canadian Journal of Statistics*, 39(2):239–258.

Nagler, T. and Czado, C. (2015). Evading the curse of dimensionality in multivariate kernel density estimation with simplified vines. *arXiv:1503.03305v2 [stat.ME]*.

Nelsen, R. B. (1999). *An Introduction to Copulas*. Springer.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, pages 347–370.

Network, C. G. A. and Others (2012). Comprehensive molecular portraits of human breast tumours. *Nature*, 490(7418):61–70.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In Engle, R. F. and McFadden, D. L., editors, *Handbook of Econometrics*, volume 4, pages 2111–2245. Elsevier North Holland.

Nikoloulopoulos, A. K., Joe, H., and Li, H. (2012). Vine copulas with asymmetric tail dependence and applications to financial return data. *Computational Statistics and Data Analysis*, 56(11):3659 – 3673.

Nychka, D. (1988). Bayesian confidence intervals for smoothing splines. *Journal of the American Statistical Association*, 83(404):1134–1143.

O'Sullivan, F., Yandell, B. S., and Raynor, W. J. (1986). Automatic Smoothing of Regression Functions in Generalized Linear Models. *Journal of the American Statistical Association*, 81(393):96–103.

Panagiotelis, A., Czado, C., and Joe, H. (2012). Pair copula constructions for multivariate discrete data. *Journal of the American Statistical Association*, 107(499):1063–1072.

Patton, A. J. (2002). *Applications of Copula Theory in Financial Econometrics*. PhD thesis, University of California, San Diego.

Peinado, H., Olmeda, D., and Cano, A. (2007). Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? *Nature Reviews Cancer*, 7(6):415–428.

R Core Team (2013). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Sabeti, A., Wei, M., and Craiu, R. V. (2014). Additive models for conditional copulas. *Stat*, 3(1):300–312.

Schepsmeier, U., Stoeber, J., Brechmann, E. C., Graeler, B., Nagler, T., and Erhardt, T. (2015). *VineCopula: Statistical Inference of Vine Copulas.* R package version 1.7.

Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 47(1):1–52.

Spanhel, F. and Kurz, M. S. (2015). Simplified vine copula models: Approximations based on the simplifying assumption. *arXiv:1510.06971 [stat.ME].*

Stöber, J. and Czado, C. (2014). Regime switches in the dependence structure of multidimensional financial data. *Computational Statistics & Data Analysis*, 76:672–686.

Stöber, J., Joe, H., and Czado, C. (2013). Simplified pair copula constructions — limitations and extensions. *Journal of Multivariate Analysis*, 119(0):101–118.

Valenzuela, O., Delgado-Marquez, B., and Pasadas, M. (2013). Evolutionary computation for optimal knots allocation in smoothing splines. *Applied Mathematical Modelling*, 37(8):5851–5863.

Vatter, T. and Chavez-Demoulin, V. (2015). Generalized additive models for conditional dependence structures. *Journal of Multivariate Analysis*, 141:147–167.

Vatter, T. and Nagler, T. (2015). Generalized additive models for pair-copula constructions. Unpublished manuscript.

Vatter, T., Wu, H.-T., Chavez-Demoulin, V., and Yu, B. (2015). Non-parametric estimation of intraday spot volatility: Disentangling instantaneous trend and seasonality. *Econometrics*, 3(4):864.

Wood, S. N. (2000). Modelling and smoothing parameter estimation with multiple quadratic penalties. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(2):413–428.

Wood, S. N. (2004). Stable and Efficient Multiple Smoothing Parameter Estimation for Generalized Additive Models. *Journal of the American Statistical Association*, 99(467):673–686.

Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R.* Chapman and Hall/CRC.

Wood, S. N. (2013a). On p-values for smooth components of an extended generalized additive model. *Biometrika*, 100(1):221–228.

Wood, S. N. (2013b). A simple test for random effects in regression models. *Biometrika*, 100(4):1005–1010.

Xingwei, T., Tao, H., and Hengjian, C. (2010). Hazard regression with penalized spline: The smoothing parameter choice and asymptotics. *Acta Mathematica Scientia*, 30(5):1759–1768.

Xu, J. J. (1996). *Statistical Modelling and Inference for Multivariate and Longitudinal Discrete Response Data*. PhD thesis, University of British Columbia.

Yee, T. W. and Wild, C. J. (1996). Vector generalized additive models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 58:481–493.

Yu, Y. and Ruppert, D. (2002). Penalized Spline Estimation for Partially Linear Single-Index Models. *Journal of the American Statistical Association*, 97(460):1042–1054.

Appendix A

Proof of the Theorems of Chapter 2

A.1 Proof of Theorem 2.2.1

In what follows, we define $c(\mathbf{z}; \boldsymbol{\theta}) = c\{\mathbf{u}; \eta(\mathbf{x}; \boldsymbol{\theta})\}$ as well as

$$\ell_c(\mathbf{z}; \boldsymbol{\theta}) = \log c(\mathbf{z}; \boldsymbol{\theta}), \ \ell_c(\boldsymbol{\theta}) = n^{-1} \sum_{j=1}^{n} \ell_c(\mathbf{z}^j; \boldsymbol{\theta}), \ \mathbf{g}_c(\mathbf{z}; \boldsymbol{\theta}) = \frac{\partial \ell_c(\mathbf{z}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \ \mathbf{g}_c(\boldsymbol{\theta}) = \frac{\partial \ell_c(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}},$$

$$\boldsymbol{\phi}_c(\mathbf{z}; \boldsymbol{\theta}) = \frac{\partial \mathbf{g}_c(\mathbf{z}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} = \frac{\partial^2 \ell_c(\mathbf{z}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \text{ and } \boldsymbol{\phi}_c(\boldsymbol{\theta}) = \frac{\partial \mathbf{g}_c(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} = \frac{\partial^2 \ell_c(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}.$$

The penalized log-likelihood is $\ell_c(\boldsymbol{\theta}, \boldsymbol{\gamma}) = \ell_c(\boldsymbol{\theta}) - \boldsymbol{\theta}^\top \mathbf{p}(\boldsymbol{\gamma}) \boldsymbol{\theta}/2$ with $\mathbf{p}(\boldsymbol{\gamma})$ as in (2.8).

Assumption 1. *Vanishing penalty:* $\boldsymbol{\gamma} = o(n^{-1/2})\mathbf{1}$.

Assumption 2. *Regularity conditions:*

1. *The model is identified in the sense that, for $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$ and $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$, then $c(\mathbf{z}; \boldsymbol{\theta}_1) \neq c(\mathbf{z}; \boldsymbol{\theta}_2)$. Furthermore, $E\left\{\sup_{\boldsymbol{\theta} \in \Theta} |\ell_c(\mathbf{Z}; \boldsymbol{\theta})|\right\} < \infty.$*

2. *The true vector of parameters $\boldsymbol{\theta}_0$ is in the interior of Θ, which is compact, and Θ_0 is an open neighborhood around $\boldsymbol{\theta}_0$.*

3. *For $\mathbf{z} \in \mathbb{Z}$, $c(\mathbf{z}; \boldsymbol{\theta}) \in C^3(\Theta_0)$.*

4. *$\int \sup_{\boldsymbol{\theta} \in \Theta_0} \|\mathbf{g}_c(\mathbf{z}; \boldsymbol{\theta})\| \, d\mathbf{z} < \infty$ and $\int \sup_{\boldsymbol{\theta} \in \Theta_0} \|\boldsymbol{\phi}_c(\mathbf{z}; \boldsymbol{\theta})\| \, d\mathbf{z} < \infty$*

5. *For $\boldsymbol{\theta} \in \Theta_0$, $\mathbf{i}(\boldsymbol{\theta}) = cov\{\mathbf{g}_c(\mathbf{Z}; \boldsymbol{\theta})\}$ exists and is positive-definite.*

6. *For each triplet $1 \leq q, r, s \leq d$, there exists a function $b : \mathbb{Z} \to \mathbb{R}$ such that, for $\boldsymbol{\theta} \in \Theta_0$ and $\mathbf{z} \in \mathbb{Z}$, $|\partial^3 \ell_c(\mathbf{z}; \boldsymbol{\theta})/\partial \boldsymbol{\theta}_{qrs}| \leq b(\mathbf{z})$, with $E\{b(\mathbf{Z})\} < \infty$.*

To prove Theorem 2.2.1, we require the following three lemmas.

Lemma 1. *Let $h(\mathbf{z}, \boldsymbol{\theta})$ be continuously differentiable, a.s. $d\mathbf{z}$, on $\boldsymbol{\theta} \in \Theta_0$. If $\int \sup_{\boldsymbol{\theta} \in \Theta_0} \|\partial h(\mathbf{z}; \boldsymbol{\theta})/\partial \boldsymbol{\theta}\| \, d\mathbf{z} < \infty$, then for $\boldsymbol{\theta} \in \Theta_0$,*

1. *$\int h(\mathbf{z}; \boldsymbol{\theta}) \, d\mathbf{z}$ is continuously differentiable, and*

2. *$\int \partial h(\mathbf{z}; \boldsymbol{\theta})/\partial \boldsymbol{\theta} \, d\mathbf{z} = \partial \int h(\mathbf{z}; \boldsymbol{\theta}) \, d\mathbf{z}/\partial \boldsymbol{\theta}$.*

Proof. See Newey and McFadden (1994), Lemma 3.6. $\qquad\square$

Lemma 2. *If Assumption 2 holds, then*

1. *$E\left\{\mathbf{g}_c(\mathbf{Z}; \boldsymbol{\theta}_0)\right\} = \mathbf{0}$, and*

2. *$E\left\{-\boldsymbol{\phi}_c(\mathbf{Z}; \boldsymbol{\theta}_0)\right\} = \mathbf{i}(\boldsymbol{\theta}_0)$.*

Proof. By the law of iterated expectations, $E\left\{\mathbf{g}_c(\mathbf{Z}; \boldsymbol{\theta}_0)\right\} = E\left[E\left\{\mathbf{g}_c(\mathbf{Z}; \boldsymbol{\theta}_0) \mid \mathbf{X} = \mathbf{x}\right\}\right]$ and $E\left\{-\boldsymbol{\phi}_c(\mathbf{Z}; \boldsymbol{\theta}_0)\right\} = E\left[E\left\{-\boldsymbol{\phi}_c(\mathbf{Z}; \boldsymbol{\theta}_0) \mid \mathbf{X} = \mathbf{x}\right\}\right]$ for all $\mathbf{x} \in \mathbb{X}$. Then by Assumption 2, (3) and (4) and Lemma 1,

$$
\begin{aligned}
E\left\{\mathbf{g}_c(\mathbf{Z}; \boldsymbol{\theta}_0) \mid \mathbf{X} = \mathbf{x}\right\} &= \int \mathbf{g}_c(\mathbf{z}; \boldsymbol{\theta}) c(\mathbf{z}; \boldsymbol{\theta}_0) \, d\mathbf{u} \\
&= \int \left.\frac{\partial c(\mathbf{z}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta} = \boldsymbol{\theta}_0} d\mathbf{u} \\
&= \left.\frac{\partial \int c(\mathbf{z}; \boldsymbol{\theta}) \, d\mathbf{u}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta} = \boldsymbol{\theta}_0} \\
&= \mathbf{0}
\end{aligned}
$$

proves the first part of Lemma 2 and

$$
\begin{aligned}
E\left\{\boldsymbol{\phi}_c(\mathbf{Z}; \boldsymbol{\theta}_0) \mid \mathbf{X} = \mathbf{x}\right\} &= \int \boldsymbol{\phi}_c(\mathbf{z}; \boldsymbol{\theta}_0) c(\mathbf{z}; \boldsymbol{\theta}_0) \, d\mathbf{u} \\
&= \int \left.\frac{\partial \mathbf{g}_c(\mathbf{z}; \boldsymbol{\theta}) c(\mathbf{z}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}\right|_{\boldsymbol{\theta} = \boldsymbol{\theta}_0} c(\mathbf{z}; \boldsymbol{\theta}_0) \, d\mathbf{u} \\
&\quad - \int \mathbf{g}_c(\mathbf{z}; \boldsymbol{\theta}_0) \mathbf{g}_c(\mathbf{z}; \boldsymbol{\theta}_0)^\top c(\mathbf{z}; \boldsymbol{\theta}_0) \, d\mathbf{u} \\
&= - \int \mathbf{g}_c(\mathbf{z}; \boldsymbol{\theta}_0) \mathbf{g}_c(\mathbf{z}; \boldsymbol{\theta}_0)^\top c(\mathbf{z}; \boldsymbol{\theta}_0) \, d\mathbf{u}
\end{aligned}
$$

the second. $\qquad\square$

Lemma 3. *Let $a \in \mathbb{R}_+$ and S_a be the surface of a sphere with radius $an^{-1/2}$ and center $\boldsymbol{\theta}_0$, that is*

$$S_a = \left\{ \boldsymbol{\theta} \in \Theta : \boldsymbol{\theta} = \boldsymbol{\theta}_0 + n^{-1/2}\mathbf{a}, \|\mathbf{a}\| = a \right\}.$$

For any $\epsilon > 0$, there exists a such that

$$P\left(\sup_{\boldsymbol{\theta} \in S_a} \ell_c(\boldsymbol{\theta}, \boldsymbol{\gamma}) < \ell_c(\boldsymbol{\theta}_0, \boldsymbol{\gamma}) \right) \geq 1 - \epsilon$$

when n is large enough.

Proof. Let $\boldsymbol{\theta} = \boldsymbol{\theta}_0 + n^{-1/2}\mathbf{a} \in S_a$. A Taylor expansion around $\boldsymbol{\theta}_0$ yields

$$n\ell_c(\boldsymbol{\theta}, \boldsymbol{\gamma}) - n\ell_c(\boldsymbol{\theta}_0, \boldsymbol{\gamma}) = n\ell_c(\boldsymbol{\theta}) - n\ell_c(\boldsymbol{\theta}_0) - \frac{n}{2}\left\{ \boldsymbol{\theta}^\top \mathbf{p}(\boldsymbol{\gamma})\boldsymbol{\theta} - \boldsymbol{\theta}_0^\top \mathbf{p}(\boldsymbol{\gamma})\boldsymbol{\theta}_0 \right\}$$

$$= \underbrace{n^{1/2}\mathbf{g}_c(\boldsymbol{\theta}_0)^\top \mathbf{a}}_{R_1(\mathbf{a})} + \underbrace{\frac{1}{2}\mathbf{a}^\top \boldsymbol{\phi}_c(\boldsymbol{\theta}_0)\mathbf{a}}_{R_2(\mathbf{a})} - \underbrace{\frac{1}{2}\mathbf{a}^\top \mathbf{p}(\boldsymbol{\gamma})\mathbf{a}}_{R_3(\mathbf{a})} - \underbrace{n^{1/2}\boldsymbol{\theta}_0^\top \mathbf{p}(\boldsymbol{\gamma})\mathbf{a}}_{R_4(\mathbf{a})}$$

$$+ \underbrace{\frac{n^{-1/2}}{6}\sum_q \sum_r \sum_s \mathbf{a}_q\mathbf{a}_r\mathbf{a}_s\, \partial^3 \ell_c(\boldsymbol{\theta})/\partial\boldsymbol{\theta}_{qrs}\Big|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}}}_{R_5(\mathbf{a})},$$

with $\widetilde{\boldsymbol{\theta}}$ in the interior of S_a. By Lemma 2, (1), Assumption 2, (4) and the CLT, $n^{1/2}\mathbf{g}_c(\boldsymbol{\theta}_0) \xrightarrow{d} N\{0, \mathbf{i}(\boldsymbol{\theta}_0)\}$ and $|R_1(\mathbf{a})| = O_p(1)a$. By Lemma 2, (2) and the law of large numbers, $\boldsymbol{\phi}_c(\boldsymbol{\theta}_0) \xrightarrow{p} -\mathbf{i}(\boldsymbol{\theta}_0)$. Thus, by the continuous mapping theorem, $R_2(\mathbf{a}) \xrightarrow{p} -\frac{1}{2}\mathbf{a}^\top \mathbf{i}(\boldsymbol{\theta}_0)\mathbf{a} \leq -\frac{1}{2}\lambda_{\min}a^2$ with $\lambda_{\min} > 0$ the smallest eigenvalue of $\mathbf{i}(\boldsymbol{\theta}_0)$. Assumption 1 directly implies $R_3(\mathbf{a}), R_4(\mathbf{a}) \xrightarrow{p} 0$. Finally, by Assumption 2, (6), $\left| \partial^3 \ell_c(\boldsymbol{\theta})/\partial\boldsymbol{\theta}_{qrs}\big|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} \right| \leq n^{-1}\sum_{j=1}^n b(\mathbf{z}^j) \xrightarrow{p} E\{b(\mathbf{Z})\} < \infty$. By this and Cauchy-Schwartz inequality, $|R_5(\mathbf{a})| \xrightarrow{p} 0$.

Collecting only the terms that do not vanish in probability, we obtain

$$n\ell_c(\boldsymbol{\theta}, \boldsymbol{\gamma}) - n\ell_c(\boldsymbol{\theta}_0, \boldsymbol{\gamma}) \leq Z = O_p(1)a - \frac{1}{2}\lambda_{\max}a^2 \tag{A.1}$$

when n is large enough. Because the choice of $\boldsymbol{\theta} \in S_a$ was arbitrary, (A.1) must also be satisfied for the supremum and we have

$$P\left\{ \sup_{\boldsymbol{\theta} \in S_a} \ell_c(\boldsymbol{\theta}, \boldsymbol{\gamma}) < \ell_c(\boldsymbol{\theta}_0, \boldsymbol{\gamma}) \right\} \geq P(Z < 0).$$

Because for all $\epsilon > 0$, there exists a such that $P\left(Z < 0\right) \geq 1 - \epsilon$, the proof is complete. $\square$

Remark. The proof of Lemma 3 is closely related to that of Xingwei et al. (2010) in the context of hazard regression, where the authors consider a unique spline component.

Lemma 1 is used to show that $\int c(\mathbf{z}; \boldsymbol{\theta}) \, d\mathbf{u}$ is twice continously differentiable and that the order of differentiation and integration can be interchanged. Lemma 2 is the proof that the score has zero mean and of the information matrix equality. By Lemma 3, the penalized log-likelihod has a local maximum $\widehat{\boldsymbol{\theta}}$ in the interior of a sphere centered on $\boldsymbol{\theta}_0$. Hence, we have that $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| = O_p(n^{-1/2})$.

Remark. By Assumption 2, (1) and Newey and McFadden (1994), Lemma 2.2, it is clear that $\boldsymbol{\theta}_0$ is the unique maximizer of $Q(\boldsymbol{\theta}) = E\left\{\ell_c(\mathbf{Z}, \boldsymbol{\theta})\right\}$. Furthermore, Assumption 1, Assumption 2, (1) and Newey and McFadden (1994), Lemma 2.4 also imply that $\ell_c(\boldsymbol{\theta}, \boldsymbol{\gamma})$ converges to $Q(\boldsymbol{\theta})$ uniformly in probability. However, using such arguments to prove consistency does not lead directly to the $\sqrt{n}$-rate.

To prove the asymptotic normality, we differentiate the log-likelihood (2.8) to obtain d score equations, that is

$$\mathbf{0} = \mathbf{g}_c(\widehat{\boldsymbol{\theta}}) - \mathbf{p}(\boldsymbol{\gamma})\widehat{\boldsymbol{\theta}}. \tag{A.2}$$

The second order Taylor expansion of (A.2) around $\boldsymbol{\theta}_0$ then reads

$$\mathbf{g}_c(\boldsymbol{\theta}_0) + \left\{\boldsymbol{\phi}_c(\boldsymbol{\theta}_0) - \mathbf{p}(\boldsymbol{\gamma})\right\}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) - \mathbf{p}(\boldsymbol{\gamma})\boldsymbol{\theta}_0 + \mathbf{r} = \mathbf{0}, \tag{A.3}$$

where $\mathbf{r}$ is a p-vector satisfying

$$\mathbf{r}_q = \frac{1}{2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top \left.\frac{\partial^2 \mathbf{g}_c(\boldsymbol{\theta})_q}{\partial \boldsymbol{\theta}\boldsymbol{\theta}^\top}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$$

with $\widetilde{\boldsymbol{\theta}}$ such that $\|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| \leq \|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|$. Multiplying the left-hand side of (A.3) by $\sqrt{n}$, we obtain the following system of random linear equations with d unknown:

$$\left\{\boldsymbol{\phi}_c(\boldsymbol{\theta}_0) - \mathbf{p}(\boldsymbol{\gamma}) + \mathbf{r}\right\}\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = n^{1/2}\left\{\mathbf{p}(\boldsymbol{\gamma})\boldsymbol{\theta}_0 - \mathbf{g}_c(\boldsymbol{\theta}_0)\right\},$$

where we redefine $\mathbf{r}$ as the $p \times p$ matrix satisfying

$$\mathbf{r}_{qr} = \frac{1}{2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top \left.\frac{\partial \boldsymbol{\phi}_c(\boldsymbol{\theta})_{qr}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}}.$$

By Assumption 1 (respectively Assumption 2, (6)), $\mathbf{p}(\boldsymbol{\gamma}) \overset{p}{\to} \mathbf{0}$ and $n^{1/2}\mathbf{p}(\boldsymbol{\gamma})\boldsymbol{\theta}_0 \overset{p}{\to} \mathbf{0}$ (respectively $\mathbf{r} \overset{p}{\to} \mathbf{0}$) is straightforward. As noted previously, $n^{1/2}\mathbf{g}_c(\boldsymbol{\theta}_0) \overset{d}{\to} N\{\mathbf{0}, \mathbf{i}(\boldsymbol{\theta}_0)\}$ and $\boldsymbol{\phi}_c(\boldsymbol{\theta}_0) \overset{p}{\to} -\mathbf{i}(\boldsymbol{\theta}_0)$. By Slutsky's theorem, we conclude that $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \overset{d}{\to} N\{\mathbf{0}, \mathbf{i}(\boldsymbol{\theta}_0)^{-1}\}$.

Remark. Using $\boldsymbol{\gamma} = o(1)\mathbf{1}$ and similar arguments, $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 - \mathbf{p}(\boldsymbol{\gamma})\boldsymbol{\theta}_0) \overset{d}{\to} N\{\mathbf{0}, \mathbf{i}(\boldsymbol{\theta}_0)^{-1}\}$.

A.2   Proof of Theorem 2.2.2

We use the same notations as in A.1, but we make explicit the dependency of the conditional copula on $\boldsymbol{\alpha}$ (e.g., $c(\mathbf{z}; \boldsymbol{\theta})$ becomes $c(\mathbf{z}; \boldsymbol{\alpha}, \boldsymbol{\theta})$). Recalling that $\mathbf{w} = (\mathbf{y}^\top, \mathbf{x}^\top)^\top$ and $\mathbf{z} = (\mathbf{u}^\top, \mathbf{x}^\top)^\top$ with $\mathbf{u}_i = F_{\mathbf{Y}_i|\mathbf{X}}\{\mathbf{y}_i; \alpha_i(\mathbf{x})\}$, we further define $f_{\mathbf{Y}_i|\mathbf{X}}(\mathbf{w}; \boldsymbol{\alpha}_i) = f_{\mathbf{Y}_i|\mathbf{X}}\{\mathbf{y}_i; \boldsymbol{\alpha}_i(\mathbf{x})\}$ and

$$
\mathbf{g}(\mathbf{w}; \boldsymbol{\alpha}, \boldsymbol{\theta}) = \begin{pmatrix} \mathbf{g}_1(\mathbf{w}; \boldsymbol{\alpha}_1) \\ \mathbf{g}_2(\mathbf{w}; \boldsymbol{\alpha}_2) \\ \mathbf{g}_c(\mathbf{z}; \boldsymbol{\alpha}, \boldsymbol{\theta}) \end{pmatrix} \text{ with } \begin{cases} \mathbf{g}_i(\mathbf{w}; \boldsymbol{\alpha}_i) = \partial \ell_i(\mathbf{w}; \boldsymbol{\alpha}_i)/\partial \boldsymbol{\alpha}_i \\ \mathbf{g}_c(\mathbf{z}; \boldsymbol{\alpha}, \boldsymbol{\theta}) = \partial \ell_c(\mathbf{z}; \boldsymbol{\alpha}, \boldsymbol{\theta})/\partial \boldsymbol{\theta} \end{cases},
$$

as well as $\boldsymbol{\phi}_{ci}(\mathbf{z}; \boldsymbol{\alpha}, \boldsymbol{\theta}) = \dfrac{\partial \mathbf{g}_c(\mathbf{z}; \boldsymbol{\alpha}, \boldsymbol{\theta})}{\partial \boldsymbol{\alpha}_i^\top} = \dfrac{\partial^2 \ell_c(\mathbf{z}; \boldsymbol{\alpha}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\alpha}_i^\top}$.

Assumption 3. *Regularity conditions for the margins:*

1. *Assumption 1 hold for $\boldsymbol{\lambda}$.*

2. *Assumption 2 hold for each margin.*

Assumption 4. *Joint identifiability: $E\{\mathbf{g}(\mathbf{W}; \boldsymbol{\alpha}, \boldsymbol{\theta})\} = \mathbf{0}$ if and only if $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$ and $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.*

Assumption 5. *Let $\mathbb{A}_0 \times \Theta_0$ be an open neighbourhood around $(\boldsymbol{\alpha}_0^\top, \boldsymbol{\theta}_0^\top)^\top$.*

1. *$\mathbf{g}_c(\mathbf{z}; \boldsymbol{\alpha}, \boldsymbol{\theta})$ is $C^1(\mathbb{A}_0 \times \Theta_0)$.*

2. *$E\left\{ \displaystyle\sup_{(\boldsymbol{\alpha}^\top, \boldsymbol{\theta}^\top)^\top \in \mathbb{A}_0 \times \Theta_0} \|\boldsymbol{\phi}_{ci}(\mathbf{z}; \boldsymbol{\alpha}, \boldsymbol{\theta})\| \right\} < \infty.$*

Lemma 4. $E\{\mathbf{g}_i(\mathbf{W}; \boldsymbol{\alpha}_{0i})\mathbf{g}_c(\mathbf{Z}; \boldsymbol{\alpha}_0, \boldsymbol{\theta}_0)\} = \mathbf{0}$

Proof. $E\{\mathbf{g}_i(\mathbf{W}; \boldsymbol{\alpha}_i)\mathbf{g}_c(\mathbf{Z}; \boldsymbol{\alpha}, \boldsymbol{\theta})\} = E[E\{\mathbf{g}_i(\mathbf{W}; \boldsymbol{\alpha}_i)\mathbf{g}_c(\mathbf{Z}; \boldsymbol{\alpha}, \boldsymbol{\theta}) \mid \mathbf{X} = \mathbf{x}\}]$ by the law of iterated expectations. $E\{\mathbf{g}_i(\mathbf{W}; \boldsymbol{\alpha}_{0i})\mathbf{g}_c(\mathbf{Z}; \boldsymbol{\alpha}_0, \boldsymbol{\theta}_0) \mid \mathbf{X} = \mathbf{x}\} = \mathbf{0}$ follows by Joe (2005), Appendix. $\qquad \square$

As noted in Section 2.2.2, $(\widehat{\boldsymbol{\alpha}}^\top, \widehat{\boldsymbol{\theta}}^\top)^\top$, which solves

$$n^{-1} \sum_{j=1}^{n} \mathbf{g}\left(\mathbf{w}^j; \boldsymbol{\alpha}, \boldsymbol{\theta}\right) - \begin{pmatrix} \mathbf{p}_1(\boldsymbol{\lambda}_1)\boldsymbol{\alpha}_1 \\ \mathbf{p}_2(\boldsymbol{\lambda}_2)\boldsymbol{\alpha}_2 \\ \mathbf{p}(\boldsymbol{\gamma})\boldsymbol{\theta} \end{pmatrix} = \mathbf{0},$$

can be viewed as a Generalized Method of Moment (GMM) estimator with an identity weighting matrix (see Newey and McFadden 1994, Chapter 6). Because $\partial \mathbf{g}\left(\mathbf{w}; \boldsymbol{\alpha}, \boldsymbol{\theta}\right) / \partial(\boldsymbol{\alpha}^\top, \boldsymbol{\theta}^\top)^\top$ is lower triangular, $\mathbf{j}(\boldsymbol{\alpha}_0, \boldsymbol{\theta}_0)^\top \mathbf{j}(\boldsymbol{\alpha}_0, \boldsymbol{\theta}_0)$ is nonsingular. Furthermore, Assumptions 1–5 directly imply the other conditions of Newey and McFadden (1994), Theorems 3.4 and 6.2. As in Joe (2005), the only non-trivial calculations concern the zeros of $v(\boldsymbol{\alpha}_0, \boldsymbol{\theta}_0)$, where

$$v(\boldsymbol{\alpha}, \boldsymbol{\theta}) = cov\left\{\mathbf{g}\left(\mathbf{W}; \boldsymbol{\alpha}, \boldsymbol{\theta}\right)\right\},$$

and Lemma 4 concludes the proof.