Dependence Modelling and Testing: Copula and Varying Coefficient Model with Missing Data

Yajing Zhu

A Thesis

in

The Department

of

Economics

Presented in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy (Economics) at

Concordia University

Montréal, Québec, Canada

May 2017

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Yajing Zhu

Entitled: Dependence Modelling and Testing: Copula and Varying Coefficient Model with Missing Data

and submitted in partial fulfillment of the requirements for the degree of

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Beverley Best — Chair

Taoufik Bouezmarni — External Examiner

Arusharka Sen — External to Program

Gordon Fisher — Examiner

Bryan Campbell — Examiner

Artem Prokhorov, Prosper Dovonon — Thesis Supervisor

Approved by

Chair of Department or Graduate Program Director

Dean of Faculty

Abstract

Dependence Modelling and Testing: Copula and Varying Coefficient Model with Missing Data

Yajing Zhu, Ph.D.

Concordia University, 2017

This thesis investigates three topics in theoretical econometrics: goodness-of-fit tests for copulas, copula density estimators which preserve the copula property, and bias-correction for the naive kernel local linear estimators in the two-sample varying coefficient model with missing data.

In the first topic a family of goodness-of-fit tests for copulas is proposed. The tests use generalizations of the information matrix equality of White (1982). The asymptotic distribution of the generalized tests is derived. In Monte Carlo simulations, the behavior of the new tests is compared with several Cramer-von Mises type tests and the desired properties of the new tests are confirmed in high dimensions. In the second topic, a semi-parametric copula density estimation procedure that guarantees that the estimator is a genuine copula density is outlined. A simulation-based study is constructed to examine the performance of the proposed copula density estimation method and compare it with the leading copula density estimators in the literature. The method is also applied to estimate copula densities in two empirical cases. The third topic shows that the naive kernel estimator using matching data is not consistent in the two-sample varying coefficient model with missing data. A bias-corrected consistent estimator is proposed and the asymptotic theory is discussed. A simulation study is conducted to support the theoretical results.

Acknowledgments

I would first like to thank my supervisor, Professor Artem Prokhorov for his meticulous and consistent help throughout my Ph.D. studies. His contributions to my research, both intellectual and financial, have been instrumental. Completing my thesis would not have been possible without his ingenuity, guidance and patience.

I would also like to extend my deepest gratitude toward Professor Gordon Fisher for his insightful input, and thoughtful and constructive recommendations he provided on my dissertation. His assistance with helping me compile my thesis and preparing for conference presentations has been invaluable.

Professor Prosper Dovonon has also been essential in helping me complete my thesis. His observant suggestions has been fundamental in helping me overcome any obstacles that arose throughout my research.

Lastly, I would like to thank my committee members Professor Bryan Campbell, Professor Arusharka Sen and Professor Taoufik Bouezmarni for their thorough examination of my thesis, for their comments, and for guiding me through the final stages of my Ph.D. studies.

Contribution of Authors

Chapter 2 is co-authored with Dr. Artem Prokhorov and Dr. Ulf Schepsmeier. Dr. Artem Prokhorov and I designed the Generalized Information Matrix Tests (GIMTs), and I implemented them and conducted the simulation study for copulas. Dr. Ulf Schepsmeier worked independently on cases with vine vopulas. Dr. Artem Prokhorov and Dr. Ulf Schepsmeier proved the asymptotic properties for the GIMTs.

Chapter 3 is co-authored with Dr. Artem Prokhorov and Dr. Eddie Anderson. Dr. Artem Prokhorov and Dr. Eddie Anderson proposed the idea of copula by triangulation and designed the estimation procedure with linear splines to estimate the copula density in their initial work. I generalized their estimation procedure to sieve estimators with higher degree splines and/or irregular grid. I conducted the simulation study for comparison of the newly proposed estimators with the leading competitors in the literature. I applied the new estimators of copula density estimation in two empirical cases.

Chapter 4 is co-authored with Dr. Xiaoqi He. Dr. Artem Prokhorov and Dr. Masayuki Hirukawa proposed the idea of bias-correction of the naive kernel estimators with matching data in two-sample varying coefficient model with missing data. I proposed the modified version of the above model and gave the identification condition. Dr. Xiaoqi He and I proved the asymptotic properties of the bias-corrected estimator collaboratively. I conducted the simulation study to show that the bias-corrected local linear estimator behaves better than the naive kernel local linear estimators in terms of mean squared error in several simulation settings.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis investigates three topics in theoretical econometrics: goodness-of-fit tests for copulas, copula density estimators which preserve the copula property, and bias-correction for the naive kernel local linear estimators in the two-sample varying coefficient model with missing data.

In Chapter 2 a family of goodness-of-fit tests for copulas is proposed. The tests use generalizations of the information matrix (IM) equality of White (1982) and so relate to the copula test proposed by Huang and Prokhorov (2014). The idea is that eigenspectrum-based statements of the IM equality reduce the degrees of freedom of the test's asymptotic distribution and lead to better size-power properties, even in high dimensions. The gains are especially pronounced for vine copulas, where additional benefits come from simplifications of score functions and the Hessian. The asymptotic distributions of the generalized tests are derived, accounting for the non-parametric estimation of the marginals, and apply a parametric bootstrap procedure, valid when asymptotic critical values are inaccurate. In Monte Carlo simulations, the behavior of the new tests are studied, comparing with several Cramer-von Mises type tests. The desired properties of the new tests in high dimensions are confirmed.

Chapter 3 focuses on simple arrangements for approximating copula densities with spline-type surfaces, while guaranteeing that our estimator is indeed a copula density. The difficulty of approximating copula densities with piecewise linear surface while guaranteeing the uniform marginal property is first explored. Next a straightforward method of applying the spline as basis functions for approximating copula densities is proposed. It is a semi-parametric copula density estimation procedure that guarantees that the estimator is indeed a copula density. The estimation procedure involves a maximum likelihood estimation of the coefficients of the splines. With simple linear constraints included in the maximization problem, we are solving a convex optimization problem which is easy to solve numerically. Our estimation procedure can be easily generalized onto an irregular grid on the unit square instead of a regular grid with equidistant knots, which implies a good localization property. Our estimator also can be easily generated to higher dimensions. We construct a simulation-based study to examine the performance of our copula density estimation method and compare it with the leading copula density estimators in the literature. This method is applied to estimate copula densities in two empirical cases.

Chapter 4 shows that the naive kernel estimator using matching data is not consistent in the two-sample varying coefficient model with missing data. A bias-corrected consistent estimator is proposed and the asymptotic theory is discussed. A simulation study is conducted to support the theoretical results.

Chapter 2

Generalized Information Matrix Tests for Copulas

2.1 Introduction

Consider a continuous random vector $\mathbf{X} = (X_1, \ldots, X_d)$ with a joint cumulative distribution function H and marginals $F_1, ..., F_d$. By Sklar's theorem, H has the following copula representation

$$H(x_1, ..., x_d) = C(F_1(x_1), \ldots, F_d(x_d)),$$

where C is a unique cumulative distribution function, whose marginals are uniform on $[0, 1]^d$. Copulas represent the dependence structure between elements of $\mathbf{X}$ and this allows one to model and estimate distributions of random vectors by estimating the marginals and the copula separately. In economics, finance and insurance, this ability is very important because it facilitates accurate pricing of risk (see, e.g., Zimmer, 2012). In such problems d is often quite high – tens or hundreds – and this has spurred a lot of interest to high dimensional copula modeling and testing in recent years (see, e.g., Patton, 2012).

In such high dimensions, classical multivariate parametric copulas such as the elliptical or Archimedean copulas are often insufficiently flexible in modeling different correlations or tail dependencies. On the other hand, they are very flexible and powerful in bivariate modeling. This advantage was used by Joe (1996) and later by Bedford and Cooke (2001, 2002) to construct multivariate densities using hierarchically bivariate copulas as building blocks. This process – known as a pair-copula construction (PCC, Aas et al., 2009) – results in a very flexible class of regular vine (R-vine) copula models, which can have a relatively large dimension, yet remain computationally tractable (see, e.g., Czado, 2010; Kurowicka and Cooke, 2006, for introductions to vine copulas).

A copula model for $\mathbf{X}$ arises when C is unknown but belongs to a parametric family $\mathcal{C}_0 = \{C_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \mathcal{O}\}$, where $\mathcal{O}$ is an open subset of $\mathbb{R}^p$ for some integer $p \geq 1$, and $\boldsymbol{\theta}$ denotes the copula parameter vector. There is a wide literature on the estimation of $\boldsymbol{\theta}$ under the assumption $\mathcal{H}_0 : C \in \mathcal{C}_0 = \{C_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \mathcal{O}\}$ given independent copies $\mathbf{X}_1 = (X_{11}, \ldots, X_{1d}), \ldots, \mathbf{X}_n = (X_{n1}, \ldots, X_{nd})$ of $\mathbf{X}$; see, e.g., Genest et al. (1995), Joe (2005), Prokhorov and Schmidt (2009). The complementary issue of testing

$$\mathcal{H}_0 : C \in \mathcal{C}_0 = \{C_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \mathcal{O}\} \quad \text{vs.} \quad \mathcal{H}_1 : C \notin \mathcal{C}_0 = \{C_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \mathcal{O}\}$$

is more recent – surveys of available tests can be found in Berg (2009) and Genest et al. (2009).

Currently, the main problem in testing is to develop operational "blanket" tests, powerful in high dimensions. This means we need tests which remain computationally feasible and powerful against a wide class of high-dimensional alternatives, rather than against specific low-dimensional families, and which do not require ad hoc choices, such as a bandwidth, a kernel, or a data categorization (see, e.g., Klugman and Parsa, 1999; Genest and Rivest, 1993; Junker and May, 2005; Fermanian, 2005; Scaillet, 2007; Kojadinovic and Yan, 2011). Genest et al. (2009) discuss five testing procedures that qualify as "blanket" tests. We will use some of them in our simulations.

Recently, Huang and Prokhorov (2014) proposed a "blanket" test based on the information matrix equality for copulas, and Schepsmeier (2016, 2015) extended that test to vine copulas. The point of this test is to compare the expected Hessian for $\boldsymbol{\theta}$ with the expected outer-product-of-the-gradient (OPG) form of the covariance matrix – under $\mathcal{H}_0$, their sum should be zero. This is the so called Bartlett identity and the test is called the Information Matrix Test (IMT) (see White, 1982). So in multi-parameter cases, the statistic is based on a random vector whose dimension – being equal to the number of distinct elements in the Hessian – grows as the square of the number of parameters.

Even though the statistic has a standard asymptotic distribution, simulations suggest that using analytical critical values leads to severe oversize distortions, especially when the dimension is high.

The tests we propose in this chapter are motivated by recent developments in information matrix equality testing by Golden et al. (2013). Specifically, we use alternative, eigenspectrum-based statements of the information matrix equality. This means we use functions of the eigenvalues of the two matrices, instead of the distinct elements of the matrices. This leads to a noticeable reduction in dimension of the random vector underlying the test statistic, which permits significant size and power improvements. The improvements are more pronounced for high dimensional dependence structures. Regular vine copulas are effective in this setting because of a further dimension reduction they permit. We argue that R-vines offer additional computational benefits for our tests. Compared to available alternatives, our tests applied to vine copula constructions remain operational and powerful in fairly high dimensions and seem to be the only tests allowing for copula specification testing in high dimensions.

The chapter is organized as follows. In Section 2.2, we introduce seven new goodness-of-fit tests for copulas and discuss their asymptotic properties. Section 2.3 describes the computational benefits that result from applying our tests to vine copulas. In Section 2.4 we use the new tests in a Monte Carlo study where we first study the new copula tests in terms of their size and power performance, and then examine the effect of dimensionality, sample size and dependence strength on size and power of these tests, as compared with three popular "blanket" tests that perform well in simulations. Section 2.5 presents the conclusions.

2.2 Generalized Information Matrix Test for Copulas

In the setting of general specification testing, Golden et al. (2013) introduced an extension to the original information equality test of White (1982), which they call the Generalized Information Matrix Test (GIMT). Unlike the original test, which is based on the negative expected Hessian and OPG, GIMT is based on functions of the eigenspectrum of the two matrices. In this section we develop a series of copula goodness-of-fit tests which draw on GIMT and we study their properties.

2.2.1 Generalized Tests and Hypothesis Functions

Let $\mathbf{X}_i = (X_{i1}, \ldots, X_{id})$, $i = 1, \ldots, n$, denote realizations of a random vector $\mathbf{X} = (X_1, \ldots, X_d) \in \mathbb{R}^d$. All tests we consider are based on a pseudo-sample $\mathbf{U}_1 = (U_{11}, \ldots, U_{1d})$, $\ldots$, $\mathbf{U}_n$ = $(U_{n1}, \ldots, U_{nd})$, where $\mathbf{U}_i = (U_{i1}, \ldots, U_{id}) = \left(\frac{R_{i1}}{n+1}, \ldots, \frac{R_{id}}{n+1}\right)$ are realizations of a random vector $\boldsymbol{\mathcal{U}} = (\mathcal{U}_1, \ldots, \mathcal{U}_d)$, and R_{ij} is the rank of X_{ij} amongst $X_{1j}, \ldots, X_{nj}$. The denominator $n+1$ is used instead of n to avoid numerical problems at the boundaries of $[0,1]^d$. Given a sample $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$, $\{\mathbf{U}_1, \ldots, \mathbf{U}_n\}$ can be viewed as a pseudo-sample from a copula C.

Note that $\mathbf{U}_1, \ldots, \mathbf{U}_n$ (and all functions thereof) depend on the sample $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ via the rank transformation but we do not reflect this in the notation (by using a hat or a subscript) in order to keep the notation under control.

Assume that the copula density $c_{\boldsymbol{\theta}}$ exists. Let $\mathbb{H}(\boldsymbol{\theta})$ denote the expected Hessian matrix of $\ln c_{\boldsymbol{\theta}}$ and let $\mathbb{C}(\boldsymbol{\theta})$ denote the expected outer product of the corresponding score function (OPG), i.e.,

$$\mathbb{H}(\boldsymbol{\theta}) := \mathbb{E}\nabla^2_{\boldsymbol{\theta}} \ln c_{\boldsymbol{\theta}}(\boldsymbol{\mathcal{U}}) \qquad \text{and} \qquad \mathbb{C}(\boldsymbol{\theta}) := \mathbb{E}\nabla_{\boldsymbol{\theta}} \ln c_{\boldsymbol{\theta}}(\boldsymbol{\mathcal{U}})\, \nabla'_{\boldsymbol{\theta}} \ln\, c_{\boldsymbol{\theta}}(\boldsymbol{\mathcal{U}}),$$

where "$\nabla_{\boldsymbol{\theta}}$" and "$\nabla^2_{\boldsymbol{\theta}}$" denote the first and second derivatives with respect to $\boldsymbol{\theta}$, respectively; the expectations are with respect to the true distribution H.

Let $\boldsymbol{\theta}_0$ denote the true value of $\boldsymbol{\theta}$, that is, $\boldsymbol{\theta}_0$ identifies the unique copula function C in Sklar's theorem. Assume $\mathbb{H}(\boldsymbol{\theta}_0)$ and $\mathbb{C}(\boldsymbol{\theta}_0)$ are in the interior of a compact set $S^{p\times p} \subseteq \mathbb{R}^{p\times p}$. For $i = 1, \ldots, n$, let

$$\mathbb{H}_i(\boldsymbol{\theta}) := \nabla^2_{\boldsymbol{\theta}} \ln c_{\boldsymbol{\theta}}(\mathbf{U}_i) \qquad \text{and} \qquad \mathbb{C}_i(\boldsymbol{\theta}) := \nabla_{\boldsymbol{\theta}} \ln c_{\boldsymbol{\theta}}(\mathbf{U}_i)\, \nabla'_{\boldsymbol{\theta}} \ln c_{\boldsymbol{\theta}}(\mathbf{U}_i).$$

For any $\boldsymbol{\theta} \in \mathcal{O}$, define the sample analogues of $\mathbb{H}(\boldsymbol{\theta})$ and $\mathbb{C}(\boldsymbol{\theta})$:

$$\bar{\mathbb{H}}(\boldsymbol{\theta}) := n^{-1} \sum_{i=1}^{n} \mathbb{H}_i(\boldsymbol{\theta}) \qquad \text{and} \qquad \bar{\mathbb{C}}(\boldsymbol{\theta}) := n^{-1} \sum_{i=1}^{n} \mathbb{C}_i(\boldsymbol{\theta}).$$

Then, given an estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}_0$, we can denote estimates of $\mathbb{H}(\boldsymbol{\theta}_0)$ and $\mathbb{C}(\boldsymbol{\theta}_0)$ by

$$\bar{\mathbb{H}}_n := \bar{\mathbb{H}}(\hat{\boldsymbol{\theta}}) \qquad \text{and} \qquad \bar{\mathbb{C}}_n := \bar{\mathbb{C}}(\hat{\boldsymbol{\theta}}),$$

where the subscript n denotes dependence on the estimator $\hat{\boldsymbol{\theta}}$.

The estimator we will use is known as the Canonical Maximum Likelihood Estimator (CMLE). It maximizes the copula-based likelihood evaluated at pseudo-observations and for this reason it is often called a maximum pseudo-likelihood estimator. The properties of CMLE are very well studied; for example, Proposition 2.1 of Genest et al. (1995) shows consistency and asymptotic normality of CMLE of $\boldsymbol{\theta}_0$.

Definition 1 (Hypothesis Function) *Let $s : S^{p\times p} \times S^{p\times p} \to \mathbb{R}^r$ be a continuous differentiable function in both of its matrix arguments. s is called a hypothesis function if for every $A, B \in S^{p\times p}$ it follows:*

$$\textit{If } A = -B \textit{ then } s(A, B) = \mathbf{0}_r,$$

where $\mathbf{0}_r$ is a zero vector of dimension r.

Here and in what follows we let $\mathbb{H}_0$ and $\mathbb{C}_0$ be the short-hand notation for the expected Hessian and OPG evaluated at the true value; that is, $\mathbb{H}_0 := \mathbb{H}(\boldsymbol{\theta}_0)$ and $\mathbb{C}_0 := \mathbb{C}(\boldsymbol{\theta}_0)$.

Definition 2 (GIMT) *A test statistic $\hat{s}_n := s(\bar{\mathbb{H}}_n, \bar{\mathbb{C}}_n)$ is a GIMT for copula $C_{\boldsymbol{\theta}}$ if it tests the null hypothesis:*

$$\mathcal{H}_0 : s(\mathbb{H}_0, \mathbb{C}_0) = \mathbf{0}_r.$$

Clearly, there are many choices for the hypothesis function $s(\cdot, \cdot)$. In particular, eigenspectrum functions such as the determinant $\det(\cdot)$ and the trace $tr(\cdot)$ can be used to construct $s(\cdot, \cdot)$. One of the main insights of Golden et al. (2013) is that different hypothesis functions permit misspecification testing in different directions. For example, a test comparing the determinants of $\mathbb{H}_0$ and $\mathbb{C}_0$ will detect small variations in eigenvalues of the two matrices, while a test comparing traces will focus on differences in the major principal components of the two matrices.

We consider the following choices:

(1) White Test $\mathcal{T}_n$: $vech(\mathbb{H}_0)+vech(\mathbb{C}_0) = \mathbf{0}_{p(p+1)/2}$, where $vech$ denotes vertical vectorization of the lower triangle of a square matrix.

(2) Determinant White Test $\mathcal{T}_n^{(D)}$: $det(\mathbb{H}_0 + \mathbb{C}_0) = 0$

(3) Trace White Test $\mathcal{T}_n^{(T)}$: $tr(\mathbb{H}_0 + \mathbb{C}_0) = 0$

(4) Information Ratio (IR) Test $\mathcal{Z}_n$: $tr(-\mathbb{H}_0^{-1}\mathbb{C}_0) - p = 0$

(5) Log Determinant IR Test $\mathcal{Z}_n^{(D)}$: $\log(det(-\mathbb{H}_0^{-1}\mathbb{C}_0)) = 0$

(6) Log Trace IMT Tr_n: $\log(tr(-\mathbb{H}_0)) - \log(tr(\mathbb{C}_0)) = 0$

(7) Log Generalized Akaike Information Criterion (GAIC) IMT $\mathcal{G}_n$: $\log[\frac{1}{p}(\mathbf{1}_p)'(\Lambda(-\mathbb{H}_0^{-1}) \odot \Lambda(\mathbb{C}_0))] = 0$, where $\odot$ denotes the Hadamard product, Λ denotes the eigenvalue function and $\mathbf{1}_p$ denotes a vector of p ones.

(8) Log Eigenspectrum IMT $\mathcal{P}_n$: $\log(\Lambda(-\mathbb{H}_0^{-1})) - \log(\Lambda(\mathbb{C}_0^{-1})) = \mathbf{0}_p$

(9) Eigenvalue Test $\mathcal{Q}_n$: $\Lambda(-\mathbb{H}_0^{-1}\mathbb{C}_0) - \mathbf{1}_p = \mathbf{0}_p$

The tests $\mathcal{T}_n$ and $\mathcal{Z}_n$ are the original White and IR tests (see, e.g., Huang and Prokhorov, 2014; Schepsmeier, 2016, 2015). The other tests are new. The *Trace White Test* $\mathcal{T}_n^{(T)}$ focuses on the sum of the eigenvalues of $\mathbb{H}_0 + \mathbb{C}_0$ and the *Determinant White Test* $\mathcal{T}_n^{(D)}$ focuses on the product of the eigenvalues of $\mathbb{H}_0 + \mathbb{C}_0$. The focused testing allows for directional power which we discuss later.

Two more tests are log-versions of the last two. The *(Log) Determinant IR Test* $\mathcal{Z}_n^{(D)}$ focuses on the determinant of the information matrix ratio, and the *Log Trace Test* Tr_n looks at whether the sum of the eigenvalues is the same for the negative Hessian and the OPG form. We use logarithms here as variance stabilizing transformations. In contrast to the White (or IR) version, the *Log Trace Test* does not use the eigenvalues of the sum (or the ratio) of $\mathbb{H}_0$ and $\mathbb{C}_0$, rather it looks at the eigenvalues of each matrix separately.

The *Log GAIC Test* $\mathcal{G}_n$ picks on the idea of the *IR Test* that the negative Hessian multiplied by the inverse of the OPG (or vice versa) equals the identity matrix. The new feature is that we focus on the average product of the Hessian-based eigenvalues and OPG-based eigenvalues. The last two tests are explicitly based on the full eigenspectrum. The *Eigenspectrum Test* $\mathcal{P}_n$ compares the eigenvalues of $\mathbb{H}_0$ and $\mathbb{C}_0$ separately, and the *Eigenvalue Test* $\mathcal{Q}_n$ uses the eigenvalues of the information matrix ratio.

In multivariate settings, the dimension of $\boldsymbol{\theta}$ often grows faster than the dimension of $\mathbf{X}$. For example, a d-variate t-copula has $O(d^2)$ parameters. The eigenspectrum-based hypothesis functions allow a reduction of the dimension of the test statistic (and thus the degrees of freedom of the test) from $p(p+1)/2$, where p is the number of copula parameters, to the number of values of the hypothesis function, r.

All these hypothesis functions represent equivalent equations under the null, yet the behavior of the tests varies widely. We first look at the asymptotic approximations of the behavior.

2.2.2 Asymptotic Results for a Generic Hypothesis Function

We start by looking at the asymptotic properties of the GIMT based on a generic hypothesis function. Since $\hat{s}_n$ is a function of CMLE these properties will mirror the properties of CMLE, which are known to be subject to certain regularity conditions. Therefore, the properties of the GIMT will be subject to the same regularity conditions. The regularity conditions are listed in many papers on semiparametric copula estimation (see, e.g., Genest et al., 1995; Shih and Louis, 1995; Hu, 1998; Tsukahara, 2005; Chen and Fan, 2006b,a). They include compactness of the parameter set, smoothness of the marginals, existence and continuity of the log-density derivatives up to the second order.

An additional assumption specific to our setting is the assumption of existence of the third-order derivatives of the log-density. Let $\nabla_{\boldsymbol{\theta}} s(\mathbb{H}_0, \mathbb{C}_0)$ denote the derivative matrix of the hypothesis function with respect to $\boldsymbol{\theta}$ evaluated at $\boldsymbol{\theta}_0$. We assume that $\nabla_{\boldsymbol{\theta}} s(\mathbb{H}_0, \mathbb{C}_0)$ has full row rank r.

The asymptotic distributions of the various test statistics we consider depend on the limiting properties of $\bar{\mathbb{H}}_n$ and $\bar{\mathbb{C}}_n$, and on the form of the hypothesis function. Let

$$\boldsymbol{d}_i(\boldsymbol{\theta}) := \begin{pmatrix} vech(\mathbb{H}_i(\boldsymbol{\theta})) \\ vech(\mathbb{C}_i(\boldsymbol{\theta})) \end{pmatrix} \in \mathbb{R}^{p(p+1)}$$

denote the lower triangle vectorizations of $\mathbb{H}_i(\boldsymbol{\theta})$ and $\mathbb{C}_i(\boldsymbol{\theta})$ and define the sample average $\bar{\boldsymbol{d}}(\boldsymbol{\theta}) := \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{d}_i(\boldsymbol{\theta})$. Clearly, the limiting behavior of $\hat{s}_n$ is determined by the behavior of $\bar{\boldsymbol{d}}(\hat{\boldsymbol{\theta}})$ and by the derivative of the various hypothesis functions with respect to $\mathbb{H}(\boldsymbol{\theta})$ and $\mathbb{C}(\boldsymbol{\theta})$.

Lemma 3 (Asymptotic Normality of $\sqrt{n}\hat{s}_n$) *Let $s : S^{p\times p} \times S^{p\times p} \to \mathbb{R}^r$ be a GIMT hypothesis function. Then, under $\mathcal{H}_0$,*

$$\sqrt{n}\hat{s}_n \xrightarrow{d} N(0, \Sigma_s(\boldsymbol{\theta}_0)),$$

where

$$\Sigma_s(\boldsymbol{\theta}_0) := S(\boldsymbol{\theta}_0)V(\boldsymbol{\theta}_0)S(\boldsymbol{\theta}_0)', \tag{1}$$

$$S(\boldsymbol{\theta}_0) := \left(\left.\frac{\partial s}{\partial vech(\mathbb{H}(\boldsymbol{\theta}))'}\right|_{\boldsymbol{\theta}_0}, \left.\frac{\partial s}{\partial vech(\mathbb{C}(\boldsymbol{\theta}))'}\right|_{\boldsymbol{\theta}_0} \right) \tag{2}$$

and $V(\boldsymbol{\theta}_0)$ is given in Eq.(25) of Appendix A.

Proof: see Appendix A for all proofs.

Lemma 3 essentially decomposes the two effects on the asymptotic distribution of $\hat{s}_n$. The common variance component $V(\boldsymbol{\theta}_0)$ is the variance of $\sqrt{n}\bar{\boldsymbol{d}}(\hat{\boldsymbol{\theta}})$ and the test-specific term $S(\boldsymbol{\theta}_0)$ captures the effect of using the different hypothesis functions.

The main difference between Lemma 3 and the specification tests of White (1982) and Golden et al. (2013) is in the form of $V(\boldsymbol{\theta}_0)$. The complication arises from the rank transformation which requires a non-trivial adjustment to the variance of $\hat{s}_n$, accounting for the estimation error (see Huang and Prokhorov, 2014). Therefore, the proof of Lemma 3 mimics that of Proposition 1 of Huang and Prokhorov (2014).

Let $\widehat{\Sigma}_s$ denote any consistent estimator of the asymptotic covariance matrix $\Sigma_s(\boldsymbol{\theta}_0)$. The following result is easy to show using Lemma 1 and consistency of $\widehat{\Sigma}_s$ so it is left without proof.

Theorem 4 *Under $\mathcal{H}_0$, the GIMT statistic for copulas*

$$\mathcal{W}_n := n\, \hat{s}_n' \widehat{\Sigma}_s^{-1} \hat{s}_n \tag{3}$$

is asymptotically χ^2_r distributed.

Note that the distribution has r degrees of freedom, where r is the number of components of the vector-valued hypothesis function. An improvement provided by the eigenspectrum-based GIMT is that for many tests $r = 1$.

Clearly, a consistent estimator $\widehat{\Sigma}_s$ would require a consistent estimation of $S(\boldsymbol{\theta}_0)$ and $V(\boldsymbol{\theta}_0)$. Given $\hat{\boldsymbol{\theta}}$, the task is to obtain consistent plug-in estimators of the derivatives and variance. Let $\widehat{S}$ and $\widehat{V}$ denote consistent estimators of $S(\boldsymbol{\theta}_0)$ and $V(\boldsymbol{\theta}_0)$, respectively. It follows that $\Sigma_s(\boldsymbol{\theta}_0)$ can be estimated as

$$\widehat{\Sigma}_s = \widehat{S}\widehat{V}\widehat{S}'.$$

How to obtain $\widehat{V}$ is discussed by Huang and Prokhorov (2014). This involves plugging $\hat{\boldsymbol{\theta}}$ in place of $\boldsymbol{\theta}_0$ in $V(\boldsymbol{\theta}_0)$ and replacing expectations in $V(\boldsymbol{\theta}_0)$ with sample averages.

In the propositions that follow we focus on the estimation of $S(\boldsymbol{\theta}_0)$.

2.2.3 Asymptotic Results for Specific Tests

We now specialize the result of Theorem 4 to the hypothesis functions we consider.

White Test for Copulas

In the case of the original White (1982) test, the asymptotic covariance matrix in Lemma 3 simplifies. Huang and Prokhorov (2014, Proposition 1) provide the asymptotic variance matrix for this case. It can be obtained by rearranging the building blocks used in the construction of the test statistic (elements of $\boldsymbol{d}_i(\boldsymbol{\theta})$), and by setting $\widehat{S} = \left[\boldsymbol{I}_{p(p+1)/2},\ \boldsymbol{I}_{p(p+1)/2}\right]$, where $\boldsymbol{I}_k$ is a $k \times k$ identity matrix.

Proposition 5 (Determinant White Test) *Define*

$$\widehat{S} = det(\bar{\mathbb{H}}_n + \bar{\mathbb{C}}_n) vech[(\bar{\mathbb{H}}_n + \bar{\mathbb{C}}_n)^{-1}]' \left[\boldsymbol{I}_{p(p+1)/2},\ \boldsymbol{I}_{p(p+1)/2}\right].$$

Then, under $\mathcal{H}_0$, the asymptotic distribution of the test statistic

$$\mathcal{T}_n^{(D)} := n \frac{[det(\bar{\mathbb{H}}_n + \bar{\mathbb{C}}_n)]^2}{\widehat{\Sigma}_s}$$

is χ_1^2.

Proposition 6 (Trace White Test) *Define*

$$\widehat{S} = \left[vech(\boldsymbol{I}_p)',\;\; vech(\boldsymbol{I}_p)'\right].$$

Then, under $\mathcal{H}_0$, the asymptotic distribution of the test statistic

$$\mathcal{T}_n^{(T)} := n\frac{tr(\bar{\mathbb{H}}_n + \bar{\mathbb{C}}_n)^2}{\widehat{\Sigma}_s}$$

is χ_1^2.

Note that $\hat{\Sigma}_s$ is scalar for these tests and the test statistics can be viewed as products of two standard normals where a square root of the numerator is scaled by a square root of $\hat{\Sigma}_s$. The two tests have one degree of freedom, rather than $p(p+1)/2$, but have important differences allowing for directional testing. Because larger eigenvalues have a larger effect on the determinant than on the corresponding trace, the *Trace White Test* will be less sensitive to changes in eigenvalues, especially small ones, and thus less powerful than the *Determinant White Test*.

Information Ratio Test for Copulas

As extensions of the original White test, Zhou et al. (2012) and Presnell and Boos (2004) consider using a ratio of the Hessian and OPG. Under correct specification, the matrix $-\mathbb{H}_0^{-1}\mathbb{C}_0$ is equal to a p-dimensional identity matrix. Two versions of this test for copulas are now proposed.

Proposition 7 (IR Test) *Define*

$$\widehat{S} = \left[vech\left(\bar{\mathbb{H}}_n^{-1}\bar{\mathbb{C}}_n\bar{\mathbb{H}}_n^{-1}\right)',\;\; vech\left(-\bar{\mathbb{H}}_n^{-1}\right)'\right].$$

Then, under $\mathcal{H}_0$, the asymptotic distribution of the test statistic

$$\mathcal{Z}_n := n\frac{\left[tr(-\bar{\mathbb{H}}_n^{-1}\bar{\mathbb{C}}_n) - p\right]^2}{\widehat{\Sigma}_s}$$

is χ^2_1.

Proposition 8 (Log Determinant IR Test) *Define*

$$\widehat{S} = \det(\bar{\mathbb{H}}_n^{-1}\bar{\mathbb{C}}_n)\left[vech\left(-\bar{\mathbb{C}}_n\bar{\mathbb{H}}_n^{-1}\bar{\mathbb{C}}_n\right)',\ \ vech\left(\bar{\mathbb{C}}_n^{-1}\right)'\right].$$

Then, under $\mathcal{H}_0$, the asymptotic distribution of the test statistic

$$\mathcal{Z}_n^{(D)} := n\frac{\left(\log(det(-\bar{\mathbb{H}}_n^{-1}\bar{\mathbb{C}}_n))\right)^2}{\widehat{\Sigma}_s}$$

is χ^2_1.

Log Trace Test for Copulas

Similar to the Log-Determinant IR Test we can construct a test using the log of traces of $-\mathbb{H}_0$ and $\mathbb{C}_0$, which should be identical under the null.

Proposition 9 (Log Trace Test) *Define*

$$\widehat{S} = \left[\frac{1}{tr(\bar{\mathbb{H}}_n)}vech(\mathbf{I}_p)', -\frac{1}{tr(\bar{\mathbb{C}}_n)}vech(\mathbf{I}_p)'\right].$$

Then, under $\mathcal{H}_0$, the asymptotic distribution of the test statistic

$$Tr_n := n\frac{\left[\log(tr(-\bar{\mathbb{H}}_n)) - \log(tr(\bar{\mathbb{C}}_n))\right]^2}{\widehat{\Sigma}_s}$$

is χ^2_1.

As mentioned earlier, trace-based tests pick up changes in larger eigenvalues easier than in smaller eigenvalues – a property that is desirable for some alternatives.

Log GAIC Test for Copulas

Define the Generalized Akaike Information Criterion as follows:

$$GAIC := -2\log\prod_{i=1}^{n} c(\mathbf{U}_i;\hat{\boldsymbol{\theta}}) + 2tr(-\bar{\mathbb{H}}_n^{-1}\bar{\mathbb{C}}_n).$$

It is well known (see, e.g., Takeuchi, 1976) that under model misspecification GAIC is an unbiased estimator of the expected value of $-2\log\prod_{i=1}^{n} c(\mathbf{U}_i;\hat{\boldsymbol{\theta}})$, where $\bar{\mathbb{H}}_n$ and $\bar{\mathbb{C}}_n$ come from a parametric likelihood. Under correct model specification $2tr(-\bar{\mathbb{H}}_n^{-1}\bar{\mathbb{C}}_n) \to 2p$, since $-\bar{\mathbb{H}}_n^{-1}\bar{\mathbb{C}}_n \to \mathbf{I}_p$ a.s., and so GAIC becomes AIC.

However, this definition ignores the fact that our likelihood has a non-parametric component and so would be valid in our setting only if $\bar{\mathbb{H}}_n$ and $\bar{\mathbb{C}}_n$ were based on observations from the copula rather than on the pseudo-observations. Gronneberg and Hjort (2014) provide a correction required to the conventional GAIC in order to account for the rank transformation used in CMLE. This link to GAIC motivates the name for the following form of the GIMT.

Let $\Lambda(A) = (\lambda_1,\ldots,\lambda_p)'$ denote the vector of sorted eigenvalues of $A \in \mathbb{R}^{p\times p}$. Further, let $\Lambda^{-1}(A) := 1/\Lambda(A)$ denote component-wise $\{1/\lambda_j\}_{j=1}^{p}$ and $\Lambda(A^{-1}) = \Lambda^{-1}(A)$. Then, under the null, $tr(-\mathbb{H}^{-1}\mathbb{C}) = (\mathbf{1}_p)'\left(\Lambda(-\mathbb{H}^{-1})\odot\Lambda(\mathbb{C})\right)$, where $\odot$ denotes the Hadamard product, i.e. component-wise multiplication. However, generally, eigenvalues of the product matrix are not equal to the product of eigenvalues of the components.

Proposition 10 (GAIC Test) *Define*

$$\widehat{S} = \frac{1}{\mathrm{tr}(\bar{\mathbb{H}}_n^{-1}\bar{\mathbb{C}}_n)}\left[vech\left(\bar{\mathbb{H}}_n^{-1}\bar{\mathbb{C}}_n\bar{\mathbb{H}}_n^{-1}\right)',\;\; vech\left(-\bar{\mathbb{H}}_n^{-1}\right)'\right].$$

Then, under $\mathcal{H}_0$, the asymptotic distribution of the test statistic

$$\mathcal{G}_n := n\frac{\left\{\log\left[\frac{1}{p}(1_p)'\left(\Lambda(-\bar{\mathbb{H}}_n^{-1})\odot\Lambda(\bar{\mathbb{C}}_n)\right)\right]\right\}^2}{\widehat{\Sigma}_s}$$

is χ^2_1.

In contrast to the *IR Test* the eigenvalues of the Hessian and the OPG are calculated separately.

Thus, similar to the *Log Determinant IR Test*, the *Log GAIC Test* is more sensitive to changes in the entire eigenspectrum than the *IR Test* (see Golden et al., 2013, for a more detailed discussion).

Eigenvalue Test for Copulas

The form of the *Log Eigenspectrum IMT* was initially proposed by Golden et al. (2013). The test has p degrees of freedom. So the reduction in the degrees-of-freedom from $p(p+1)/2$ is more noticeable for larger p, which would typically mean a higher dimensional copula.

In order to derive its asymptotic distribution we need additional notation. For a real symmetric matrix A, let $y_j(A)$ denote the normalized eigenvector corresponding to eigenvalue $\lambda_j(A)$, $j = 1, \ldots, p$. Let D denote the duplication matrix, i.e. such a matrix that $Dvech(A) = vec(A)$ (see, e.g. Magnus and Neudecker, 1999).

Proposition 11 (Log Eigenspectrum Test) *Define*

$$\widehat{S} = \begin{bmatrix} -\frac{1}{\lambda_1(\bar{\mathbb{H}}_n)}[y_1(\bar{\mathbb{H}}_n)' \otimes y_1(\bar{\mathbb{H}}_n)']D & \frac{1}{\lambda_1(\bar{\mathbb{C}}_n)}[y_1(\bar{\mathbb{C}}_n)' \otimes y_1(\bar{\mathbb{C}}_n)']D \\ \vdots & \vdots \\ -\frac{1}{\lambda_p(\bar{\mathbb{H}}_n)}[y_p(\bar{\mathbb{H}}_n)' \otimes y_p(\bar{\mathbb{H}}_n)']D & \frac{1}{\lambda_p(\bar{\mathbb{C}}_n)}[y_p(\bar{\mathbb{C}}_n)' \otimes y_p(\bar{\mathbb{C}}_n)']D \end{bmatrix}.$$

Then, under $\mathcal{H}_0$, the asymptotic distribution of the test statistic

$$\mathcal{P}_n := n\big[\log(\Lambda(-\bar{\mathbb{H}}_n^{-1})) - \log(\Lambda(\bar{\mathbb{C}}_n^{-1}))\big]'\widehat{\Sigma}_s^{-1}\big[\log(\Lambda(-\bar{\mathbb{H}}_n^{-1})) - \log(\Lambda(\bar{\mathbb{C}}_n^{-1}))\big]$$

is χ^2_p.

A similar approach uses the eigenspectrum of the information matrix ratio $\Lambda(-\mathbb{H}_0^{-1}\mathbb{C}_0)$. We will call this test the *Eigenvalue Test*.

Proposition 12 (Eigenvalue Test) *Define*

$$\widehat{S} = \begin{bmatrix} \frac{1}{\lambda_1(\bar{\mathbb{H}}_n)}[y_1(\bar{\mathbb{C}}_n)' \otimes y_1(\bar{\mathbb{C}}_n)']D & -\frac{\lambda_1(\bar{\mathbb{C}}_n)}{\lambda_1(\bar{\mathbb{H}}_n)^2}[y_1(\bar{\mathbb{H}}_n)' \otimes y_1(\bar{\mathbb{H}}_n)']D \\ \vdots & \vdots \\ \frac{1}{\lambda_p(\bar{\mathbb{H}}_n)}[y_p(\bar{\mathbb{C}}_n)' \otimes y_p(\bar{\mathbb{C}}_n)']D & -\frac{\lambda_p(\bar{\mathbb{C}}_n)}{\lambda_p(\bar{\mathbb{H}}_n)^2}[y_p(\bar{\mathbb{H}}_n)' \otimes y_p(\bar{\mathbb{H}}_n)']D \end{bmatrix}.$$

Then, under $\mathcal{H}_0$, the asymptotic distribution of the test statistic

$$\mathcal{Q}_n := n\big[\Lambda(-\bar{\mathbb{H}}_n^{-1}\bar{\mathbb{C}}_n) - \mathbf{1}_p\big]'\widehat{\Sigma}_s^{-1}\big[\Lambda(-\bar{\mathbb{H}}_n^{-1}\bar{\mathbb{C}}_n) - \mathbf{1}_p\big]$$

is χ^2_p.

2.2.4 On Applicability of Asymptotic Approximations

The asymptotic results in Propositions (1)-(8) have simple distributions and may seem very appealing. However, their implementation and validity is limited by several important considerations. One of the most important criticisms of the original White test is its slow convergence to the asymptotic distribution. For example, Schepsmeier (2016) shows that for a five-dimensional copula ($df = p(p+1)/2 = 55$), the number of observations needed to show acceptable size and power behavior using asymptotic critical values is at least 10,000; for an eight-dimensional copula ($df = 406$) that number is greater than 20,000. Unfortunately, the new tests inherit the same problem.

An important reason for the slow convergence to the asymptotic distribution is the complex form of $\Sigma_s(\boldsymbol{\theta}_0)$. Estimation of the asymptotic variance matrix of the hypothesis function involves numeral evaluation of d-dimensional integrals and numerical or analytical evaluation of copula derivatives of orders one to three. Such numerical evaluations are subject to approximation errors themselves and are rarely done in practice, especially in high dimensions. Instead, it is common to look at the bootstrap distribution of $\hat{s}_n$. Since the distribution depends on $\boldsymbol{\theta}_0$, one uses the parametric bootstrap.

One situation when using asymptotic critical values may be worthwhile is when the copula score simplifies. Vine copulas allow for such simplifications. Their structure eliminates the need for d-dimensional integration and they admit simpler derivatives. So in what follows we focus on vine copulas. For non-vine copulas, one can view the asymptotic results in Propositions (1)-(8) as justification for the parametric bootstrap using these hypothesis functions.

2.3 GIMTs for Vine Copulas

A regular vine (R-vine) copula is a nested set of bivariate copulas representing unconditional and conditional dependence between elements of the initial random vector (see, e.g., Joe, 1996; Bedford and Cooke, 2001, 2002). Any d-variate copula can be expressed as a product of such (conditional) bivariate copulas and there are many ways of writing this product. Graphically, R-vine copulas can be illustrated by a set of connected trees $\mathcal{V} = \{T_1, \ldots, T_{d-1}\}$, where each edge represents a bivariate conditional copula. The nodes illustrate the arguments of the associated copula. The edges of tree T_i form the nodes of tree T_{i+1}, $i \in \{1, \ldots, d-2\}$. The proximity condition of Bedford and Cooke (2001) then defines which possible edges are allowed between the nodes to form an R-vine. If we denote the set of bivariate copulas used in trees $\mathcal{V}$ by $\mathcal{B}(\mathcal{V})$ and the corresponding set of parameters by $\boldsymbol{\theta}(\mathcal{B}(\mathcal{V}))$, then we can specify an R-vine copula by $(\mathcal{V}, \mathcal{B}(\mathcal{V}), \boldsymbol{\theta}(\mathcal{B}(\mathcal{V})))$.

Let $U_1, \ldots, U_d$ denote a pseudo-sample as introduced in Section 2.2.1. The edges $j(e), k(e)|D(e)$ in E_i, for $1 \leq i \leq d-1$ correspond the set of bivariate copula densities $\mathcal{B} = \big\{c_{j(e),k(e)|D(e)}|e \in E_i,\ 1 \leq i \leq d-1\}$. The indices $j(e)$ and $k(e)$ form the *conditioned set* while $D(e)$ is called the *conditioning set*. Then a regular vine copula density is given by the product

$$c_{1,\ldots,d}(\boldsymbol{u}) = \prod_{i=1}^{d-1} \prod_{e \in E_i} c_{j(e),k(e);D(e)}(C_{j(e)|D(e)}(u_{j(e)}|\boldsymbol{u}_{D(e)}), C_{k(e)|D(e)}(u_{k(e)}|\boldsymbol{u}_{D(e)})). \tag{4}$$

The copula arguments $C_{j(e)|D(e)}(u_{j(e)}|\boldsymbol{u}_{D(e)})$ and $C_{k(e)|D(e)}(u_{k(e)}|\boldsymbol{u}_{D(e)})$ can be derived integral-free by the formula derived from the first derivative of the corresponding cdf with respect to the second copula argument. For details, see Eq.(2) in Schepsmeier (2016). An example of a 5-dimensional R-vine is given in Figure 2.1.

The canonical vine (C-vine) and the drawable vine (D-vine) are two special R-vines. The C-vine has in each tree a root node which is connected to all other nodes in this tree. In the D-vine each node is connected to two other nodes at most.

The copula parameter vector $\boldsymbol{\theta}(\mathcal{B}(\mathcal{V}))$ can be estimated either in a tree-by-tree approach called sequential estimation, or in a full maximum-likelihood estimation (MLE) procedure (Aas et al., 2009). The sequential procedure uses the hierarchical structure of R-vines and is quick – its results

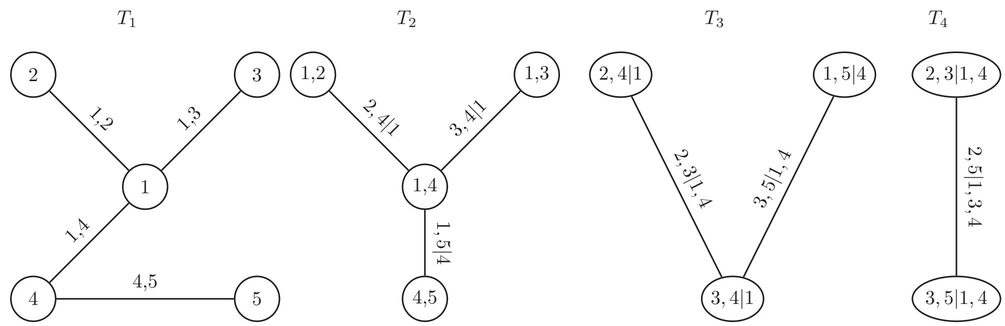


Figure 2.1: Tree structure of a 5-dimensional R-vine copula.

are often used as starting values for the MLE approach. Both are consistent estimators.

Vine copulas have gained popularity because of the benefits they offer when dimension d is high. First, they permit a decomposition of a d-variate copula with $O(d^2)$ or more parameters into $d(d-1)/2$ bivariate (one-parameter) copulas, which reduces computational burden. Second, they offer a natural way to impose conditional independence by dropping selected higher-order edges in $\mathcal{V}$. Finally, the integral free expressions for the conditional copulas offer an additional computational benefit.

Such a reduction of parameters using the conditional independence copula can be achieved in two ways. First, single conditional copulas can be assumed independent, especially if some pre-testing procedure confirms this (see, e.g., Genest and Favre, 2007). Further, by setting all pair-copula families above a certain tree order to the independence copula, the number of parameters can be reduced significantly. This involves no testing and is often done heuristically; Brechmann et al. (2012) call this approach *truncation*.

In our settings, vine copulas offer an additional advantage over conventional copulas. As an example, consider testing goodness-of-fit of a d-variate Eyraud-Farlie-Gumbel-Morgenstern (EFGM) copula. This copula has $p = 2^d - d - 1$ parameters so the number of degrees-of-freedom for the *White Test* is of order $O(2^{2d})$, while for the eigenspectrum-based tests that number is as low as one. Regardless of the GIMT, the calculation of the test statistic involves evaluating, analytically or numerically, the score function and the Hessian. If we use the asymptotic critical value we also need to evaluate the third derivative of the log-copula density and a d-variate integral. The score $\nabla_{\boldsymbol{\theta}} \ln c_{\boldsymbol{\theta}}$ is a vector-valued function with $2^d - d - 1$ elements, each a function of all $2^d - d - 1$ elements of $\boldsymbol{\theta}$. The Hessian is a $p \times p$ matrix-valued function, in which each component is a function of the entire vector $\boldsymbol{\theta}$. The third-order derivative is a $p^2 \times p$ matrix, with each element a function of p parameters. Now what changes if we replace that copula with a d-variate vine?

Consider the case of $d = 3$. Suppose we use the following R-vine representation

$$c_{123}(u_1, u_2, u_3; \boldsymbol{\theta}) = c_{12}(u_1, u_2; \theta_1)c_{23}(u_2, u_3; \theta_2)c_{13;2}(C_{1|2}(u_1|u_2; \theta_1), C_{3|2}(u_3|u_2; \theta_2); \theta_3),$$

where each bivariate copula is EFGM and $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$. Then, it is easy to see that $\nabla_{\boldsymbol{\theta}} \ln c_{\boldsymbol{\theta}}$ has the form

$$\begin{pmatrix} \nabla_{\theta_1} \ln c_{12} + \nabla_{\theta_1} \ln c_{13;2} \\ \nabla_{\theta_2} \ln c_{23} + \nabla_{\theta_2} \ln c_{13;2} \\ \nabla_{\theta_3} \ln c_{13;2} \end{pmatrix},$$

where each element is a score function for the corresponding element of $\boldsymbol{\theta}$ – a simpler function with fewer argument (see Stöber and Schepsmeier, 2013, for details). The term $\nabla_{\theta_1} \ln c_{13;2}$ is the only term that has all three parameters but if a sequential procedure is used, estimates of θ_1 and θ_2 come from previous steps and are treated as known so only θ_3 is effectively unknown in $c_{13;2}$. Regardless of the estimation method, only derivatives of bivariate copulas are needed, which are much simpler than in higher dimensions. Plus, d-dimensional integration needed for evaluation of $V(\boldsymbol{\theta}_0)$ is replaced with bivariate. Closed form expressions for the first two derivatives of several bivariate copulas are given in Schepsmeier and Stöber (2014, 2012). The Hessian will simplify accordingly – some cross derivatives will be zero (Stöber and Schepsmeier, 2013). The same is true for the third-order derivatives used to obtain $\widehat{\Sigma}_s$.

These are sizable simplifications when dealing with high dimensional copulas. The problem is that multivariate dependence requires sufficiently rich parametrization which affects the properties of the tests. For example, our simulations suggest that convergence to the asymptotic distribution of the new tests is never faster for non-vine copulas than for vine copulas. More generally, the properties of the goodness-of-fit tests including GIMTs deteriorate quickly and tests become infeasible for copulas with larger dimensions unless the copulas are vines. For example, we were unable to obtain stable simulation results for non-vine copulas for dimensions higher than 8 but had no difficulty doing so for vine-copulas.

For this reason, in the simulation study that follows we focus on vine copulas and on the bootstrap versions of these tests. To an extent, this makes comparisons with other tests fair as most available "blanket" tests use the parametric bootstrap.

2.4 Power Study

In this section we analyze the size and power properties of the new copula goodness-of-fit tests. We start by comparing performance of the various versions of GIMT for vine copulas. This is the case where we believe our tests are paticularly useful in high dimensions. Then, for classical (non-vine) copula specifications, we compare the best performing tests with "blanket" non-GIMT alternatives favored in an extensive simulation study by Genest et al. (2009). Genest et al. (2009) do not look at vine copulas so we return to the non-vine specification (and stay within low dimensions) for these comparisons.

2.4.1 Comparison Between GIMTs for Vine Copulas

Simulation Setup

We follow the simulation procedure of Schepsmeier (2016) and consider testing the null that the vine copula model is

$$M_0 = RV(\mathcal{V}_0, \mathcal{B}_0(\mathcal{V}_0), \boldsymbol{\theta}_0(\mathcal{B}_0(\mathcal{V}_0)))$$

against the alternative

$$M_1 = RV(\mathcal{V}_1, \mathcal{B}_1(\mathcal{V}_1), \boldsymbol{\theta}_1(\mathcal{B}_1(\mathcal{V}_1))), M_1 \neq M_0.$$

In each Monte Carlo simulation r, we generate n observations on $\boldsymbol{u}^r_{M_0} = (\boldsymbol{u}^{1r}_{M_0}, \dots, \boldsymbol{u}^{dr}_{M_0})$ from model M_0, estimate the vine copula parameters $\boldsymbol{\theta}_0(\mathcal{B}_0(\mathcal{V}_0))$ and $\boldsymbol{\theta}_1(\mathcal{B}_1(\mathcal{V}_1))$ and calculate the test statistic under the null, $t^r_n(M_0)$, and under the alternative, $t^r_n(M_1)$, for all the tests considered in Section 2. The number of simulations is $B = 5000$.

Then we obtain approximate p-values $\hat{p}_r$ for each test statistic as

$$\hat{p}_j := \hat{p}(t_j) := 1/B \sum_{r=1}^{B} \mathbf{1}\{t_r \geq t_j\}, j = 1, \dots, B$$

and the actual size $\hat{F}_{M_0}(\alpha)$ and (size-adjusted) power $\hat{F}_{M_1}(\alpha)$ using the formula

$$\hat{F}(\alpha) = \frac{1}{B} \sum_{r=1}^{B} \mathbf{1}\{\hat{p}_r \leq \alpha\}, \qquad \alpha \in (0, 1) \tag{5}$$

We use an R-vine copula with $d = 5$ and $d = 8$ as M_0. As M_1 we use (a) a multivariate Gaussian copula, which can also be represented as a vine, (b) a C-vine copula and (c) a D-vine copula. The details on the copulas under the null and alternatives, as well as on the method used for choosing the specific bivariate components, are provided in Appendix A.1. All calculations in this section were performed with **R** (R Development Core Team, 2013) and the R-package VineCopula of Schepsmeier et al. (2013).[1]

Simulation Results

We start by assessing the asymptotic approximation of the tests. Figures 2.2-2.3 show empirical distributions of the test statistics for two sample sizes, $n = 500$ and 1000. Several observations seem important here. First, overall we observe convergence to the asymptotic distribution even for the fairly high dimensional copulas we consider but asymptotics serve as a very poor approximator in all, except for a few, cases. Second, the sequential approach performs better than the MLE approach – an observation for which we do not have an explanation. Third, the sampling distributions of the Trace White and Determinant IR Tests – one-degree-of-freedom tests – are much closer to their asymptotic limits, regardless of the dimension, than tests with other functional forms and tests with greater degrees of freedom. Fourth, the Determinant White, Log Trace, and Eigenvalue Tests deteriorate quickly as dimension increases. The Trace White and Determinant IR Tests dominate other tests in terms of asymptotic approximation.

[1]The R code used in this section, as well as the Matlab codes used in the next section are available from the authors upon request.

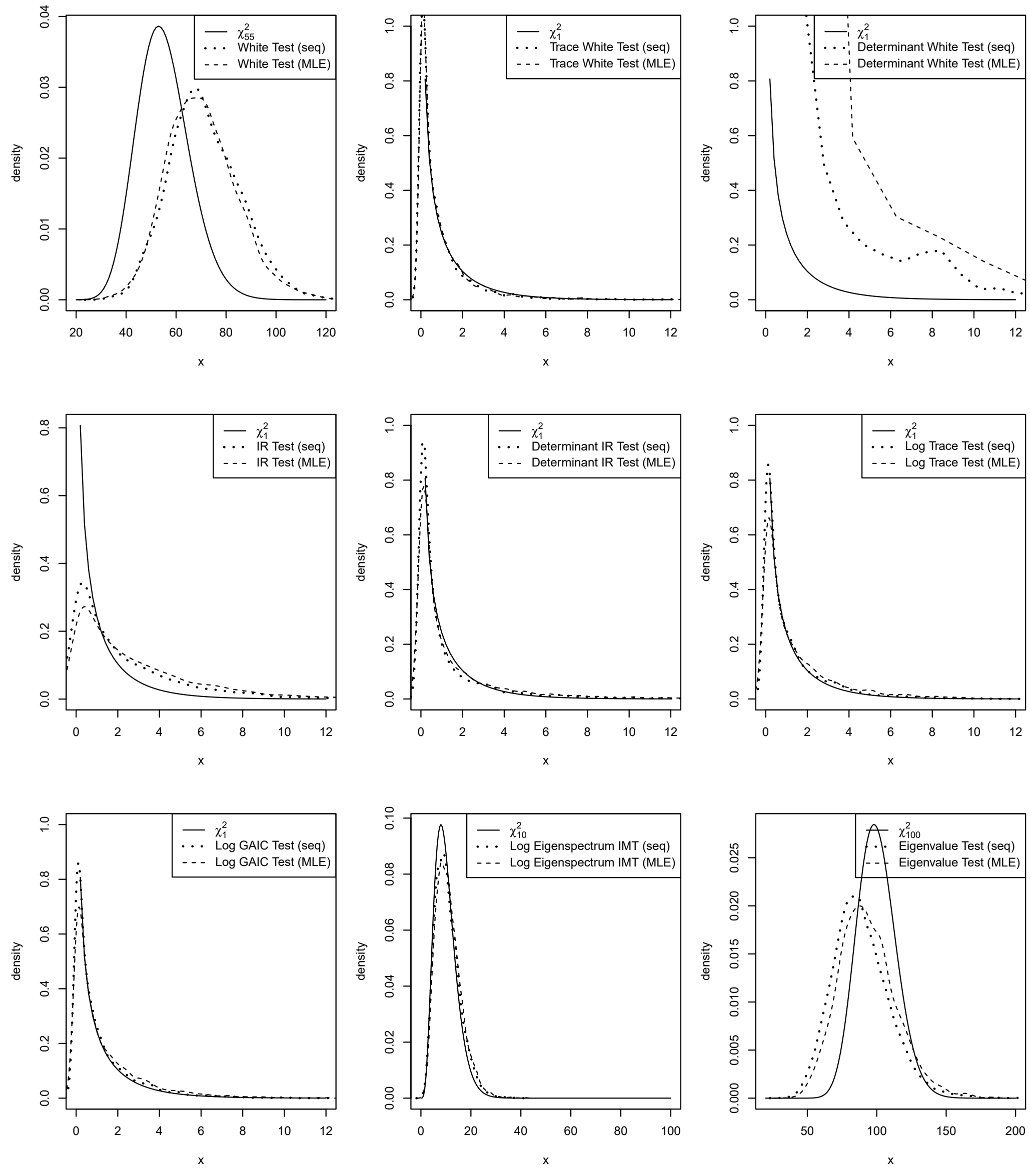


Figure 2.2: Empirical densities of GIMT for R-vine copulas: $d = 5, n = 500$

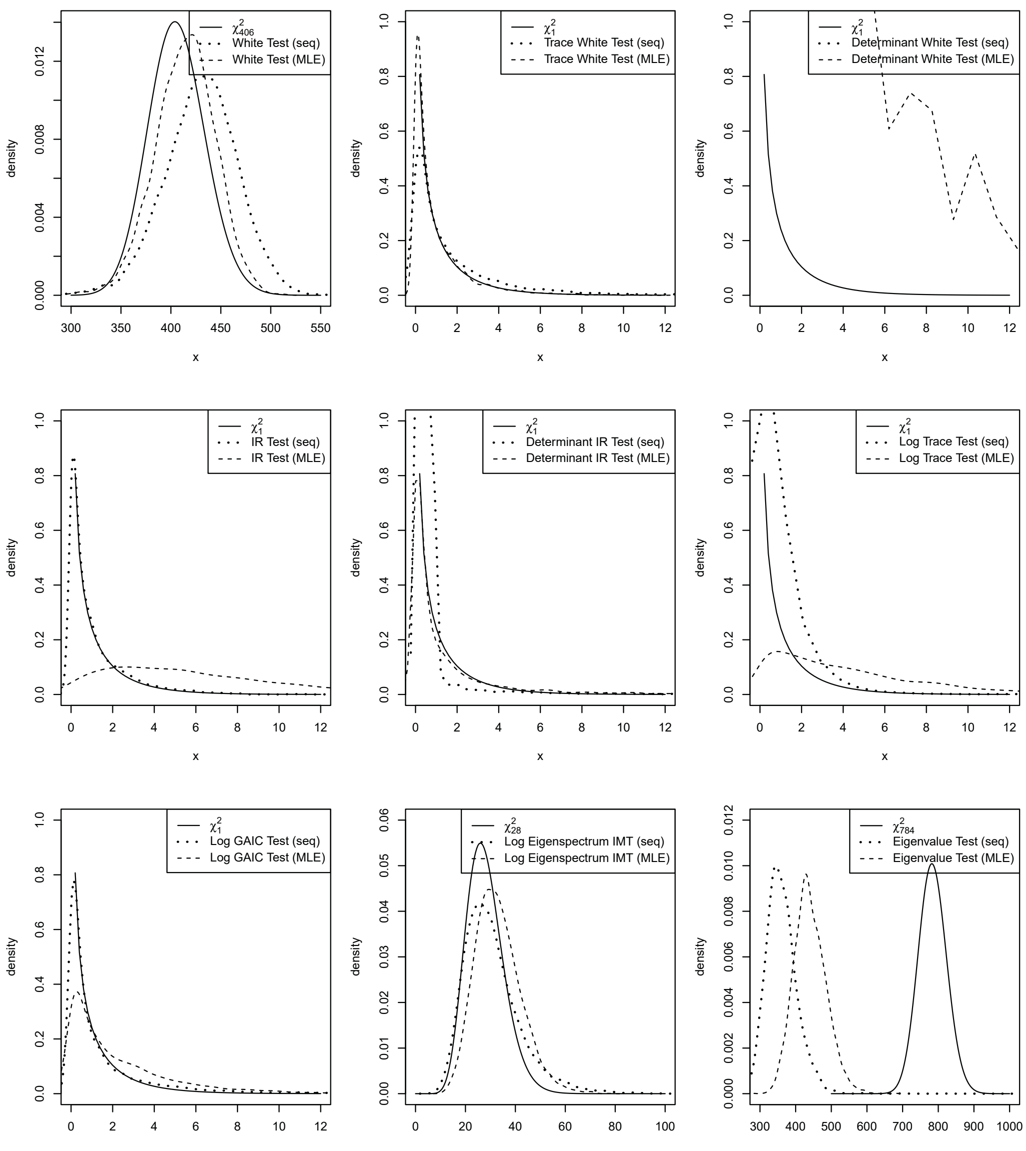


Figure 2.3: Empirical densities for GIMT for R-vine copulas: $d = 8, n = 1000$

Now we look at size-power behavior. Since some of the proposed tests face substantial numerical problems with the asymptotic variance estimation and many exhibit large deviations from the χ^2_r distribution in small samples, especially when dimension is high, we only investigate the bootstrap version of the tests. The parametric bootstrap version of the tests is quite common in the copula goodness of fit literature – for details of the parametric bootstrap procedure we refer the reader to Huang and Prokhorov (2014); Schepsmeier (2016). Figures 2.4-2.5 illustrate the estimated power of nine proposed tests. We consider three dimensions, $d = 5, 8$ and 16; and two versions, sequential (dotted lines) and MLE (solid lines). The two sample sizes we consider are $n = 500$ and 1000 for $d = 5$ and 8; and $n = 1000$ and 5000 for $d = 16$. Percentage of rejections of $\mathcal{H}_0$ is on the y-axis, while the truth (R-vine) and the alternatives are on the x-axis. Obviously, the power is equal to the actual size for the true model. A horizontal black dashed line indicates the 5% nominal size.

All proposed tests maintain their given size independently of the number of sample points, dimension or estimation method. For $d = 5$ we can observe increasing power as sample size increases for all tests except the Determinant White Test. If $d = 8$ the behavior of the tests, especially the MLE versions, is more erratic. The Determinant White Test seems to be the only test that continues to perform poorly in terms of power when sample size increases. Other tests show improvement in power for either the MLE or sequential version or both. Interestingly, the Trace White, Eigenvalue and IR Tests at times show very strong power in one of the two versions (MLE or sequential) and no power in the other. Overall, all tests except the Determinant White show power against each alternative, showing that they are consistent.

For $d = 16$ we report only sequential estimates as they were most time efficient. The Log Eigenspectrum, Eigenvalue, IR and Determinant IR tests show consistently good behavior in terms of power against the two alternatives. The power of the Determinant IR and Log Eigenspectrum Tests remains high independent of the dimension or the sample size.

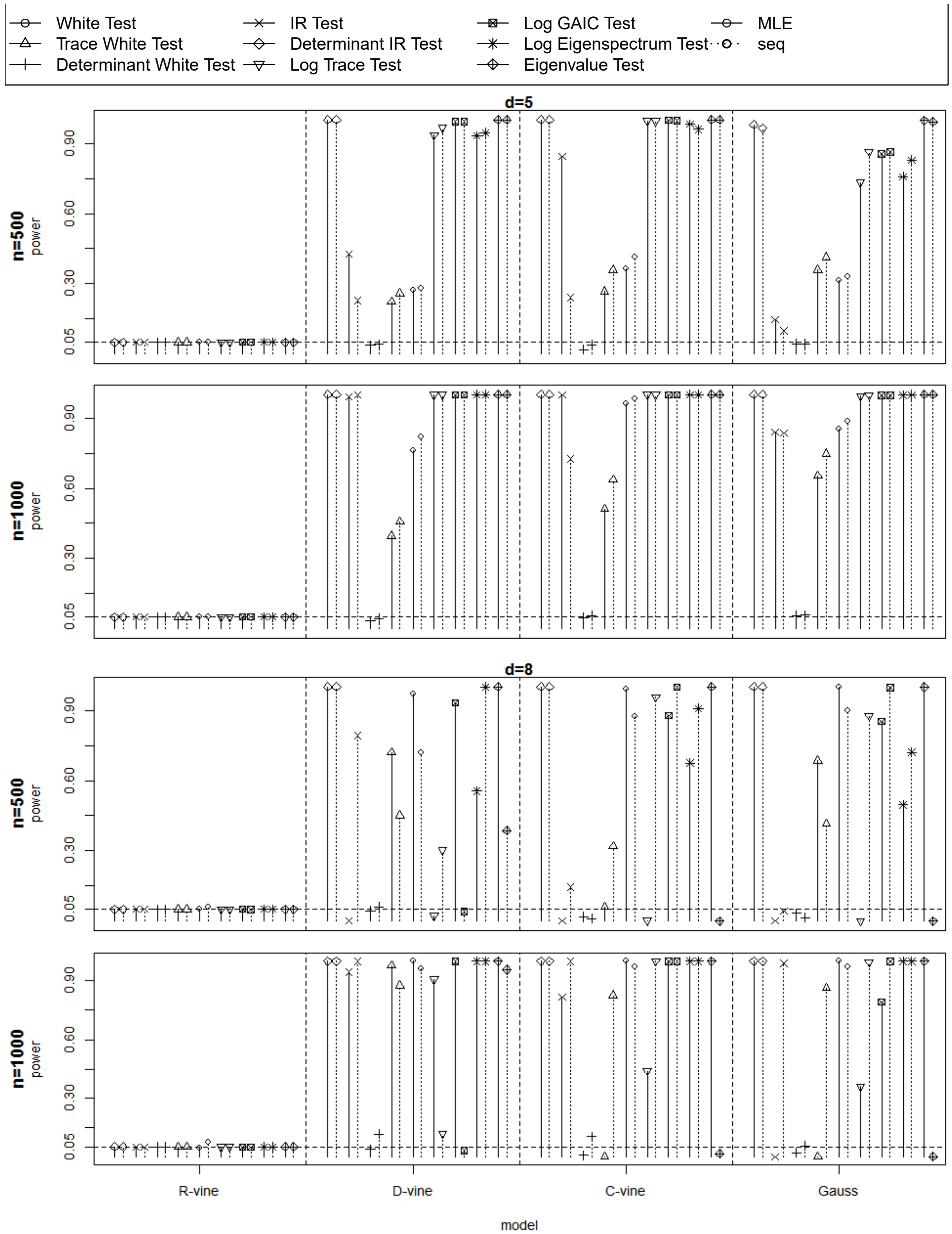


Figure 2.4: Size and power comparison for bootstrap versions of proposed tests in 5 and 8 dimensions with different sample sizes.

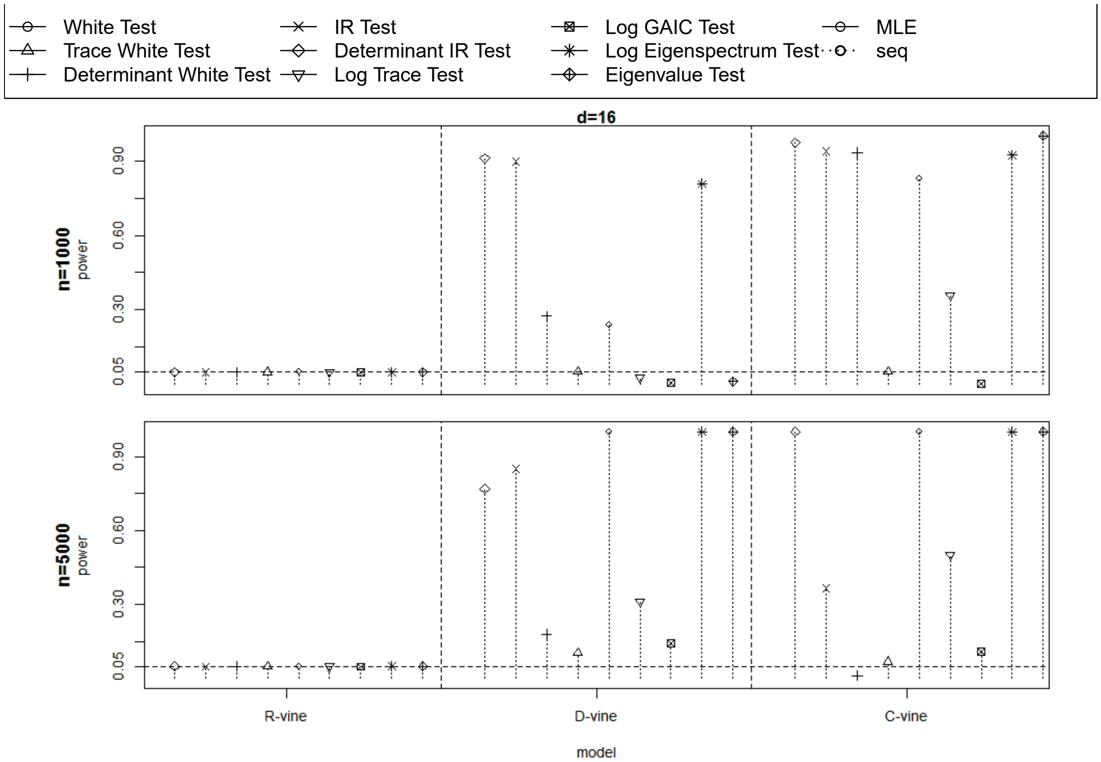


Figure 2.5: Size and power comparison for boostrap versions of proposed tests in 16 dimensions and different sample sizes (only sequential estimates are reported).

2.4.2 Comparison with Non-GIMT Tests

Simulation Setup

In this section we compare selected GIMTs for copulas with the original White test $\mathcal{T}_n$ and three "blanket" copula goodness-of-fit tests analyzed by Genest et al. (2009). Validity conditions for the parametric bootstrap method when testing for goodness-of-fit of families of copulas in semiparametric models are discussed in Genest and Rémillard (2008). The GIMTs we select are the Log GAIC Test $\mathcal{G}_n$ and the Eigenvalue Test $\mathcal{Q}_n$ – which showed acceptable size and power properties in the simulations of previous sections. The selected non-GIMTs are based on the empirical copula process and the Rosenblatt's and Kendall's transformation – which showed favorable size and power behavior in an extensive Monte Carlo study by Genest et al. (2009). We provide details on the three tests in Appendix A.3 and summarize them in Table 2.1. For vine copulas such comparisons are provided by Schepsmeier (2015), plus the simulations by Genest et al. (2009) do not include vine copulas so in this section we consider only classical (non-vine) copulas.

Table 2.1: Summary of non-GIMTs.

Empirical copula process	$\mathcal{S}_n$	$n\int_{[0,1]^d}(C_n(u)-C_{\hat{\boldsymbol{\theta}}}(u))^2\,dC_n(u)$
		$=\sum_{j=1}^n\{C_n(\mathbf{U}_j)-C_{\hat{\boldsymbol{\theta}}}(\mathbf{U}_j)\}^2$
Rosenblatt's transform	$\mathcal{S}_n^R$	$\{\mathbf{V}_j=\mathcal{R}_{C_{\hat{\boldsymbol{\theta}}}}(\mathbf{U}_j)\}_{j=1}^n$
		$\sum_{j=1}^n\{C_n(\mathbf{V}_j)-C_\perp(\mathbf{V}_j)\}^2$
Kendall's transform	$\mathcal{S}_n^K$	$C_{\boldsymbol{\theta}}(\mathbf{U})\sim K_{\boldsymbol{\theta}}$
		$n\int_{[0,1]}(K_n(v)-K_{\hat{\boldsymbol{\theta}}_n}(v))^2\,dK_{\hat{\boldsymbol{\theta}}}(v)$

Again, since the limiting approximation is poor and depends on an unknown parameter $\boldsymbol{\theta}$, we resort to parametric bootstrap to obtain valid p-values. We can use any consistent estimator of $\boldsymbol{\theta}_0$, e.g., the estimator based on Kendall's τ or the CMLE. In this section, we use the estimator based on Kendall's τ in all bivariate and multivariate cases except for tests involving the Outer Power Clayton and t-copula. For these two copulas, the true parameter vector $\boldsymbol{\theta}_0$ is estimated by CMLE. For details see Appendix A.2.

Simulation Results

We report selected size and power results in tables similar to those reported by Genest et al. (2009) and Huang and Prokhorov (2014). The point of the tables is to examine the effect of the sample size, degree of dependence and dimension on size and power of the seven tests. The nominal level is fixed at 5% as before.

We first report bivariate results for selected values of Kendall's τ. Gaussian, Frank, Clayton, Gumbel and Student-t copula families are considered both under the null hypothesis and under the alternative. When testing against the Student-t copula, we assume the degrees of freedom $\nu = 6$. For testing the first four one-parameter copula families, we obtain the estimate of the parameter by inverting the sample version of Kendall's τ. For testing the Student-t copula, the parameters are estimated by CMLE. The results are based on 1,000 random samples of size $n = 150$ and 500. Table 2.2 reports the size and power results for Kendall's $\tau = (0.5, 0.75)$. In each row we report the percentage of rejections of $\mathcal{H}_0$ associated with $\mathcal{S}_n$, $\mathcal{S}_n^R$, $\mathcal{S}_n^K$, $\mathcal{T}_n$ and $\mathcal{Q}_n$. As an example, Table 2.2 indicates that when testing the null of the Gaussian copula using $\mathcal{Q}_n$ and $n = 150$, we reject the null about 42% of the time when the true copula is Gumbel with Kendall's τ = 0.5. For all

Table 2.2: Percentage of rejections of $\mathcal{H}_0$ for $d = 2$.

Copula under $\mathcal{H}_0$	True copula	Kendall's $\tau = 0.50$ $n = 150$					Kendall's $\tau = 0.75$ $n = 150$					Kendall's $\tau = 0.50$ $n = 500$				
		$\mathcal{S}_n$	$\mathcal{S}_n^R$	$\mathcal{S}_n^K$	$\mathcal{T}_n$	$\mathcal{Q}_n$	$\mathcal{S}_n$	$\mathcal{S}_n^R$	$\mathcal{S}_n^K$	$\mathcal{T}_n$	$\mathcal{Q}_n$	$\mathcal{S}_n$	$\mathcal{S}_n^R$	$\mathcal{S}_n^K$	$\mathcal{T}_n$	$\mathcal{Q}_n$
Gaussian	Gaussian	*4.9*	*5.0*	*4.9*	*7.5*	***4.0***	*4.9*	*4.9*	***4.4***	*10.4*	*5.3*	*4.6*	*5.4*	*4.9*	*7.5*	***4.0***
	Frank	20.2	13.4	17.4	6.8	**36.0**	42.2	32.8	41.4	40.0	**86.6**	36.9	60.7	33.4	60.7	**66.5**
	Clayton	80.0	**90.8**	90.3	30.8	70.5	91.8	**99.9**	97.3	60.5	99.2	99.8	**100.0**	99.6	90.4	99.5
	Gumbel	38.3	**42.0**	16.1	15.4	**42.0**	38.5	55.5	17.9	23.7	**71.2**	65.3	18.9	62.9	62.3	**71.1**
	Student	7.9	30.7	6.4	10.5	**53.4**	8.0	17.4	5.8	10.6	**53.6**	16.7	41.7	6.8	63.4	**90.7**
Frank	Gaussian	19.9	8.9	**22.6**	14.6	2.0	**40.9**	18.4	40.2	8.0	3.6	**42.5**	35.1	32.7	20.6	15.2
	Frank	*4.8*	*4.8*	*4.8*	*9.4*	***4.0***	*4.7*	*5.0*	***4.5***	*11.0*	*5.3*	***4.2***	*6.4*	*4.7*	*7.1*	*4.8*
	Clayton	89.1	86.9	**98.6**	5.7	10.1	96.6	**99.7**	99.6	20.4	7.2	**100.0**	99.9	**100.0**	10.6	15.1
	Gumbel	**63.0**	44.1	28.3	5.3	12.0	**81.9**	59.9	53.2	8.7	2.2	**95.2**	47.5	85.8	13.3	14.9
	Student	36.4	**42.2**	35.7	11.4	17.6	**62.3**	60.9	58.5	32.3	40.8	**76.8**	36.7	64.5	41.6	46.2
Clayton	Gaussian	**93.7**	89.0	75.1	80.6	34.2	**99.8**	99.5	94.9	90.2	66.4	**100.0**	99.5	99.7	**100.0**	99.0
	Frank	**95.7**	94.4	89.5	90.2	70.0	99.1	**99.9**	97.0	91.6	99.8	**100.0**	99.4	99.9	99.2	99.9
	Clayton	*5.3*	*5.1*	***4.5***	*12.0*	*4.9*	*5.4*	*5.1*	*4.9*	*11.0*	***4.2***	*5.0*	*5.2*	***4.7***	*12.0*	*4.9*
	Gumbel	**99.9**	99.7	98.5	90.5	54.2	**99.9**	**99.9**	**99.9**	96.2	97.2	**100.0**	**100.0**	**100.0**	99.5	**100.0**
	Student	**84.2**	68.3	76.3	41.4	45.9	**99.8**	99.0	92.5	64.0	65.1	**99.9**	95.4	**99.9**	49.7	57.6
Gumbel	Gaussian	18.3	33.7	**37.7**	5.2	8.0	12.3	**60.7**	29.4	9.6	4.6	**74.1**	38.4	61.6	20.7	21.2
	Frank	39.8	**52.1**	42.4	29.3	37.6	51.7	83.8	61.6	76.4	**89.2**	95.5	47.8	85.1	89.3	**99.2**
	Clayton	99.6	99.7	**99.9**	75.5	78.8	99.9	99.9	99.9	90.4	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
	Gumbel	*4.6*	*4.5*	*4.6*	*10.0*	***4.4***	*4.5*	*5.2*	***4.4***	*10.9*	*5.0*	*5.2*	*5.5*	*5.0*	*7.2*	***4.4***
	Student	21.5	30.1	35.4	23.1	**36.8**	19.5	**54.4**	31.2	33.4	39.7	**78.4**	61.7	63.2	64.7	73.5
Student	Gaussian	4.9	4.2	5.0	**6.4**	4.6	4.9	4.6	5.0	**6.9**	5.2	4.6	5.4	5.1	**6.5**	4.9
	Frank	16.9	8.5	**18.1**	6.9	5.1	31.6	17.4	**32.4**	7.8	9.4	49.1	17.5	**53.2**	10.3	14.3
	Clayton	**65.4**	39.8	61.8	12.2	16.3	79.2	61.5	**81.2**	35.7	49.0	**99.9**	**99.9**	99.8	47.2	71.7
	Gumbel	25.3	10.4	**26.1**	5.6	4.5	42.2	19.7	**46.5**	8.9	4.9	**70.2**	52.4	68.4	10.5	9.6
	Student	***4.9***	*5.0*	*5.1*	*7.0*	***4.9***	***5.0***	*5.1*	*5.1*	*7.2*	*5.1*	***4.9***	*5.1*	*5.0*	*7.5*	*5.1*

Note: Italics indicate the test size, and bold entries indicate the best performing test.

tests, except $\mathcal{T}_n$, we bootstrap critical values. We use analytical values for $\mathcal{T}_n$ to show that the conventional version of IMT is badly oversized (more comparisons including bootstrap $\mathcal{T}_n$ can be found in Huang and Prokhorov (2014)).

The results indicate that all the tests, except perhaps $\mathcal{T}_n$, maintain the nominal size and generally have power against the alternatives. We note that in the bivariate case we use only one indicator in constructing $\mathcal{T}_n$ and so $\mathcal{Q}_n$ provides no dimension reduction. The analytical p-values used for $\mathcal{T}_n$ lead to noticeable oversize distortions, while $\mathcal{Q}_n$ retains size close to the nominal value and is often conservative compared with $\mathcal{S}_n$, $\mathcal{S}_n^R$, and $\mathcal{S}_n^K$. The table also shows that higher dependence or a larger sample size give higher power, which is true for all the tests we consider. The increase in power resulting from the sample size increase is an indication of $\mathcal{Q}_n$ being consistent.

Table 2.3 presents selected results for $d = 4$. Here we focus on $\mathcal{S}_n$, $\mathcal{T}_n$ and $\mathcal{Q}_n$ but report two versions of $\mathcal{T}_n$, one based on the bootstrapped critical values ($\mathcal{T}_n^b$) and the other based on the analytical asymptotic critical values ($\mathcal{T}_n^a$) – this high dimensional comparison was not considered by Huang and Prokhorov (2014). We do not include $\mathcal{S}_n^R$ and $\mathcal{S}_n^K$ because their behavior appears similar to that of $\mathcal{S}_n$. Under the null, we have three one-parameter Archimedean copulas, the Gaussian and the t-copula, each with six distinct parameters in the correlation matrix and the Outer Power Clayton copula with two parameters. The alternatives are six four-dimensional copula families.

Several observations are unique to the multivariate simulations because they involve more than one parameter and more than two marginals. To simulate from the Outer Power Clayton copula, which has two parameters, we set $(\beta, \theta) = (4/3, 1)$, which corresponds to Kendall's τ equal 0.5. For the Gaussian copula, after estimating the pairwise Kendall's τs, we invert them to obtain the corresponding elements of the correlation matrix. For the Archimedean copulas, we follow Berg (2009) and obtain the dependence parameter by inverting the average of six pairwise Kendall's τs. For the Outer Power Clayton and Student-t copula, we can only estimate the parameters by CMLE. Details on simulating from and estimation of the Outer Power Clayton copula can be found in Hofert et al. (2012). For a given value of τ and each combination of copulas under the null and alternative, the results reported are based on 1,000 random samples of size $n = 150$. Each of these samples is then used to test goodness-of-fit. Table 2.3 reports size and power for (the average of) Kendall's τ equal 0.5. (We do not report results for other values of n and τ in order to save space.)

Table 2.3: Percentage of rejections of $\mathcal{H}_0$ for $d = 4$, $n = 150$, and Kendall's $\tau = 0.50$.

Copula under $\mathcal{H}_0$	True copula	Test based on			
		$\mathcal{S}_n$	$\mathcal{T}_n^a$	$\mathcal{T}_n^b$	$\mathcal{Q}_n$
Gaussian	Gaussian	*5.0*	***4.9***	*5.0*	***4.9***
	Frank	15.4	4.7	6.5	**56.1**
	Clayton	**88.5**	14.4	10.2	72.5
	Gumbel	52.1	12.1	13.6	**75.5**
	Student	11.3	14.6	7.0	**90.4**
	Outer Power Clayton	60.2	13.9	11.4	**72.4**
Frank	Gaussian	43.4	16.3	19.6	**47.8**
	Frank	***4.2***	*7.3*	*5.3*	*4.9*
	Clayton	**97.0**	14.5	7.1	27.3
	Gumbel	**67.3**	7.0	4.5	25.6
	Student	56.7	77.3	50.5	**80.9**
	Outer Power Clayton	**77.6**	8.2	13.1	42.7
Clayton	Gaussian	92.2	**99.4**	42.6	98.8
	Frank	94.1	**99.9**	38.1	**99.9**
	Clayton	*5.1*	*10.3*	***4.2***	*4.7*
	Gumbel	99.3	**99.9**	55.4	99.8
	Student	96.7	**98.5**	50.8	96.9
	Outer Power Clayton	70.3	50.6	12.5	**75.8**
Gumbel	Gaussian	76.3	49.8	20.2	**83.4**
	Frank	60.1	33.8	16.9	**76.1**
	Clayton	99.4	99.6	82.6	**99.9**
	Gumbel	***5.0***	*6.5*	*5.2*	*5.1*
	Student	77.5	79.0	30.3	**93.2**
	Outer Power Clayton	**89.7**	50.9	22.3	78.5
Outer Power Clayton	Gaussian	**62.8**	14.6	6.7	18.4
	Frank	**60.1**	20.2	9.1	45.1
	Clayton	9.4	8.9	9.0	**11.1**
	Gumbel	**25.4**	13.5	8.1	20.9
	Student	19.5	8.4	7.9	**75.7**
	Outer Power Clayton	*5.3*	*7.7*	*5.0*	***4.8***
Student	Gaussian	5.2	**6.8**	5.1	4.9
	Frank	12.3	10.7	8.3	**16.2**
	Clayton	**86.5**	24.2	20.7	41.5
	Gumbel	**45.1**	6.2	5.4	6.9
	Student	*5.1*	*7.2*	***5.0***	*5.1*
	Outer Power Clayton	**27.5**	22.6	10.1	18.3

Note: Italics indicate the test size, and bold entries indicate the best performing test.

Table 2.4: Percentage of rejections of $\mathcal{H}_0$ for $d = 5$, $n = 150$, and Kendall's $\tau = 0.50$.

Copula under $\mathcal{H}_0$	True copula	Test based on			
		$\mathcal{S}_n$	$\mathcal{Q}_n$	$\mathcal{T}_n^b$	$\mathcal{G}_n$
Gaussian	Gaussian	*5.1*	***4.8***	*5.0*	*5.0*
	Frank	15.2	**63.4**	7.1	50.6
	Clayton	**93.8**	76.9	17.7	71.2
	Gumbel	52.3	**74.6**	12.4	62.5
	Student	9.1	**92.6**	7.6	90.1
	Outer Power Clayton	61.7	**74.7**	13.5	57.5
Frank	Gaussian	60.4	**61.4**	21.3	51.7
	Frank	*5.0*	***4.9***	*5.1*	***4.9***
	Clayton	**98.3**	34.6	8.3	30.5
	Gumbel	**69.7**	20.1	4.1	19.2
	Student	**64.2**	51.8	60.4	56.4
	Outer Power Clayton	75.4	77.3	13.9	**80.1**
Clayton	Gaussian	91.4	98.1	50.4	**92.0**
	Frank	89.9	99.2	38.9	**99.4**
	Clayton	***4.9***	***4.9***	*5.0*	***4.9***
	Gumbel	97.5	**99.9**	59.5	99.8
	Student	97.1	98.1	55.4	**98.9**
	Outer Power Clayton	72.6	**74.1**	17.6	64.3
Gumbel	Gaussian	81.0	**86.5**	24.9	85.4
	Frank	67.5	77.4	20.7	**82.0**
	Clayton	99.3	**99.9**	83.4	**99.9**
	Gumbel	*5.1*	***5.0***	*5.1*	*5.1*
	Student	74.2	**90.4**	40.2	76.5
	Outer Power Clayton	**91.1**	80.5	30.5	62.1
Outer Power Clayton	Gaussian	**60.2**	17.3	8.2	12.8
	Frank	**60.6**	51.6	17.4	41.3
	Clayton	7.5	11.3	10.2	**15.9**
	Gumbel	**26.7**	21.7	13.1	17.8
	Student	5.2	**76.4**	10.4	63.7
	Outer Power Clayton	*5.3*	*5.0*	***4.9***	*5.0*
Student	Gaussian	5.1	4.9	**5.3**	5.0
	Frank	15.9	21.4	12.4	**24.5**
	Clayton	**89.0**	49.3	24.6	43.2
	Gumbel	**54.4**	8.8	6.9	8.6
	Student	*5.0*	*5.0*	***4.8***	*5.2*
	Outer Power Clayton	**38.3**	31.5	17.6	34.9

Note: Italics indicate the test size, and bold entries indicate the best performing test.

Table 2.5: Percentage of rejections of $\mathcal{H}_0$ for $d = 8$, $n = 150$, and Kendall's $\tau = 0.50$.

Copula under $\mathcal{H}_0$	True copula	Test based on			
		$\mathcal{S}_n$	$\mathcal{Q}_n$	$\mathcal{T}_n^b$	$\mathcal{G}_n$
Gaussian	Gaussian	*5.0*	***4.8***	*5.0*	*5.0*
	Frank	25.6	**86.3**	22.5	81.5
	Clayton	**98.7**	91.2	29.6	93.8
	Gumbel	75.5	87.2	36.1	**90.5**
	Student	12.2	**99.9**	18.9	**99.9**
	Outer Power Clayton	75.4	**95.6**	39.2	82.7
Frank	Gaussian	**97.8**	87.9	32.3	82.2
	Frank	***4.9***	***4.9***	*5.0*	***4.9***
	Clayton	**99.5**	60.2	19.4	42.2
	Gumbel	**85.6**	32.4	9.8	29.3
	Student	**99.5**	79.8	64.4	82.3
	Outer Power Clayton	91.4	93.7	42.3	**96.7**
Clayton	Gaussian	99.7	**99.9**	75.4	**99.9**
	Frank	97.9	**100.0**	62.2	99.9
	Clayton	***4.9***	***4.9***	*5.0*	*5.0*
	Gumbel	**99.9**	**99.9**	82.3	**99.9**
	Student	**99.9**	**99.9**	65.2	**99.9**
	Outer Power Clayton	81.1	**95.8**	34.6	81.6
Gumbel	Gaussian	**99.5**	98.9	42.1	97.5
	Frank	63.4	81.9	40.3	**85.1**
	Clayton	100.0	**99.9**	99.0	**99.9**
	Gumbel	*5.2*	***5.0***	*5.1*	*5.1*
	Student	**99.5**	**99.5**	54.2	90.1
	Outer Power Clayton	**99.9**	**99.9**	42.2	82.1
Outer Power Clayton	Gaussian	**67.6**	38.2	33.4	20.7
	Frank	**71.4**	54.1	16.2	42.9
	Clayton	14.2	12.5	11.7	**16.6**
	Gumbel	**45.3**	28.4	32.3	35.8
	Student	18.6	**97.6**	52.4	67.9
	Outer Power Clayton	***5.0***	*5.1*	*5.3*	***5.0***
Student	Gaussian	5.0	4.9	**5.2**	5.0
	Frank	21.7	32.8	20.7	**33.7**
	Clayton	**96.4**	69.3	31.4	64.5
	Gumbel	**72.5**	14.7	9.6	15.2
	Student	*5.1*	*5.0*	***4.9***	*5.1*
	Outer Power Clayton	**69.7**	54.3	33.6	57.2

Note: Italics indicate the test size, and bold entries indicate the best performing test.

The key observation from Table 2.3 is that $\mathcal{Q}_n$ dominates both versions of $\mathcal{T}_n$ in terms of power. We attribute this to the dimension reduction permitted by $\mathcal{Q}_n$. The table also shows that our test maintains a nominal size of 5% in the multivariate cases. Overall, the behavior of $\mathcal{Q}_n$ is as good, if not better than, that of $\mathcal{S}_n$. A remarkable case of the better performance of $\mathcal{Q}_n$ is the tests involving the Student-t alternative, where $\mathcal{S}_n$ does worse, regardless of the copula under the null.

An interesting observation is how the power of $\mathcal{Q}_n$ changes between Table 2.2 and Table 2.3. Consider, for example, the test of the null of the Frank copula. Regardless of the alternative, $\mathcal{Q}_n$ performs poorly in the bivariate case. However, with the increased dimension the behavior of $\mathcal{Q}_n$ improves substantially. This is especially pronounced in comparison with $\mathcal{T}_n$, whose power remains particularly low against the Archimedean alternatives. At the same time, for the Student-t and Gaussian alternatives, the performance of $\mathcal{Q}_n$ stands out even compared with $\mathcal{S}_n$.

Table 2.4 and Table 2.5 present selected results for $d = 5$ and $d = 8$, respectively. Here we focus on $\mathcal{S}_d$, $\mathcal{Q}_n$, $\mathcal{T}_n$ and $\mathcal{G}_n$. We use $\mathcal{T}_n$ (bootstrap) as a benchmark. The Log GAIC Test $\mathcal{G}_n$ is another GIMT that performed well in Section 2.4.1 – we use it to illustrate further the dimension reduction permitted by GIMTs. In Tables 2.4 and 2.5, under the null we have three one-parameter Achimedean copulas, the Outer Power Clayton copula with two parameters, the Gaussian copula with $\frac{d(d-1)}{2}$ distinct parameters in the correlation matrix and the Student-t copula with $\frac{d(d-1)}{2} + 1$ distinct parameters. The alternatives are Frank, Clayton, Gumbel, Outer Power Clayton, Gaussian, and t copulas. Samples in every scenario are simulated from a copula with Kendall's τ equal to 0.5. The parameter estimation here is done by CMLE, rather than by conversion of Kendall's τ used for $d = 4$ in Table 2.4. The explicit expressions of the score functions of the selected Archimedean copulas can be found in Hofert et al. (2012).

The results in Tables 2.4-2.5 show that, as expected, $\mathcal{Q}_n$, $\mathcal{G}_n$ and $\mathcal{T}_n$ all maintain the nominal size and show power. More interestingly, the power of the three GIMT tests increases as the dimension increases. In particular, $\mathcal{Q}_n$ and $\mathcal{G}_n$ behave similarly under all null hypotheses and both show significant increases in power in almost all scenarios as the dimension grows. This may be due to the fact that, for regular copula, the Kendall's τs between each pair of the elements in random vector are set to be the same, therefore more information about the true parameter can be obtained as the dimension increases. Therefore the power increase is largely the result of our specific design in this simulation study – same Kendall's τs across pairs. We also see that $\mathcal{Q}_n$ and $\mathcal{G}_n$ dominate $\mathcal{T}_n$ in all scenarios. Note that for the Frank, Clayton, and Gumbel copulas, both Hessian and OPG matrices degenerate to scalars; therefore there is no dimension reduction in $\mathcal{Q}_n$ and $\mathcal{G}_n$ compared to $\mathcal{T}_n$. Yet, we observe that $\mathcal{Q}_n$ and $\mathcal{G}_n$ are more powerful than $\mathcal{T}_n$, which may be due to the fact that the eigenvalues of $-\mathbb{H}^{-1}\mathbb{C}$ are more sensitive to changes in $\mathbb{H}$ and $\mathbb{C}$ than the eigenvalues of $\mathbb{H}+\mathbb{C}$. When testing multi-parameter copulas, e.g., multivariate Gaussian, due to the additional dimension reduction, $\mathcal{Q}_n$ and $\mathcal{G}_n$ perform much better than $\mathcal{T}_n$.

2.5 Conclusion

We consider a battery of tests resulting from eigenspectrum-based versions of the information matrix equality applied to copulas. The benefit of this generalization is due to a reduction in degrees of freedom of the tests and to the focused hypothesis function used to construct them. For example, in testing goodness of fit of high-dimensional multi-parameter copulas we manage to reduce the information matrix based test statistic to an asymptotically χ^2 with one degree of freedom. Moreover, we can focus on the effect of larger or smaller eigenvalues by using specific functions of the eigenspectrum such as det or $trace$. However, only a few of the proposed tests can be well approximated by their asymptotic distributions in realistic sample sizes so we have also looked at the boostrap version of the tests.

The main argument of the chapter is that the bootstrap versions of GIMTs dominate other available tests of copula goodness of fit when copulas are high-dimensional and multi-parameter. We use this argument to motivate the use of GIMTs on vine copulas, where additional simplifications result from the functional form of the Hessian and the score.

Chapter 3

Copula by Triangulation

3.1 Introduction

Copulas are a broadly used tool for modelling dependence which recently found many applications in economics, finance and risk management. A key feature of copulas is that they have uniform margins which amounts to each marginal integral of a copula density being equal to one. If an estimator does not satisfy this restriction, copula based quantities such as the tail dependence coefficient are badly biased and can take infeasible values.

While nonparametric copulas offer great flexibility and serve as a robust means of estimating dependence, available estimators suffer from several major drawbacks. Conventional kernel density estimation methods exhibit a severe boundary problem (see, e.g., Gijbels and Mielniczuk, 1990; Omelka et al., 2009). Most of newer nonparametric copula estimators such as the Bernstein-Kantorovich polynomial and exponential series estimators do not impose the uniform marginal property in finite samples (see, e.g. Sancetta and Satchell, 2004; Gao et al., 2015). The few exceptions that do, are computationally inefficient and have not been shown to be consistent (see, e.g., Qu and Yin, 2012). Finally, very few methods are easily adaptable to cases when dependence is sparse, that is, when some parts of the copula domain are populated by vastly fewer observations than others.

This chapter proposes a new class of copula density estimators obtained by triangulation over a possibly sparse grid. Approximate copula densities with spline-type surfaces while ensuring that our estimator is indeed a copula density is proposed. The difficulty of approximating copula densities – bivariate for simplicity – with piecewise linear surfaces while guaranteeing the uniform marginal property is first explored. The difficulty is that such estimation procedure involves mixed integer optimization which is hard to work with. Next a straightforward method applying a specific spline basis function is proposed, which reduces this problem to a convex non-parametric maximum likelihood estimation, subject to linear equality constraints – an easy problem to handle in most available software packages.

The estimator is generalized to higher degree of spline and irregular grids on the unit square. The latter contribution is important because it provides a natural but overlooked way of imposing denser dependence at the corners and along the diagonal and sparser elsewhere. That is to say, our estimator has a natural localization property.

The new estimator is compared to the empirical beta copula density estimator, Bernstein-Kantorovich polynomial, exponential series, data-mirror and naive kernel estimators. This covers the most serious competitors in nonparametric copula density estimation. The effect of strength of dependence on performance is examined in the simulation study. In addition, computational time is also considered. As an application, new insights into several well-studied econometric data sets characterized by high tail dependence is also provided.

3.2 The Estimation of Copula Densities

First recall Sklar's representation for multivariate distributions. Let H be a d-dimensional distribution function with one-dimensional marginals $F_1, \dots, F_d$; then there exists a function C: $[0,1]^d \to [0,1]$ such that

$$H(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d));$$

here $C(\mathbf{u}) = F(F_1^{-1}(u_1), \dots, F_d{-}1(u_d))$ is referred to as the d-copula. If each marginal is continuous, C is unique. The copula density $c(\mathbf{u})$ is defined as $\frac{\partial^d}{\partial_1 \cdots \partial_d} C(\mathbf{u})$.

Many nonparametric estimation procedures for the density of a copula density function have already been proposed in the literature. The basic ones rely on symmetric kernels. Unfortunately, these techniques are not consistent on the boundaries of $[0,1]^d$ and suffer from boundary bias. Some techniques have been introduced to get better estimation on the boundaries, e.g., mirror image modification. The series estimators are also commonly used for copula density estimation because they are flexible in describing complicated relationships among variables and have the advantage of smoothness. Most of the copula density estimators in the literature are not genuine copulas because they do not satisfy the key copula property – uniform marginals. If an estimator does not satisfy this restriction, copula based quantities such as the Spearman's ρ coefficient and the upper tail dependence are badly biased and can take invalid values. For details, see Appendix B.1.

In the following, some popular copula density estimators are listed. To ease the notation without a lack of generality, we will restrict ourselves to the bivariate case.

3.2.1 The Naive Kernel Estimator

Kernel density estimators are popular choices for multivariate density estimation. Gijbels and Mielniczuk (1990) estimate a bivariate copula using smoothing kernel methods. Here we consider the simplest case. The copula density at a point (x, y) is:

$$c(x, y) = \frac{1}{4h^2} \lim_{h \to 0} \mathcal{P}(|X - x| \leq h, |Y - y| \leq h).$$

This can be estimated by not taking the limit and then replacing the probability with the relative frequency in this small region:

$$\hat{c}_h(x, y) = \frac{1}{4h^2 n} \#(i\,;\ |X_i - x| \leq h, |Y_i - y| \leq h).$$

As one can imagine, it exhibits the well-known boundary bias problem of the kernel methods.

3.2.2 The Data-Mirror Estimator

The problem with the naive kernel estimator and other regular kernel estimators is that the copula densities are underestimated at the boundaries. Several techniques have been introduced to obtain better estimation on the boundaries. One of them is based on the data-mirror modification (see, e.g., Schuster, 1985), where artificial data are obtained using symmetric transformations with respect to boundaries. To be specific, in bivariate case for example, instead of using only the observed data (X_i, Y_i), additional observations including the images of (X_i, Y_i) with respect to all edges and corners of the unit square are considered; i.e., the $(\pm X_i, \pm Y_i)$, the $(\pm X_i, 2 - Y_i)$, the $(2 - X_i, \pm Y_i)$

and the $(2 - X_i, 2 - Y_i)$. The estimator is then given by

$$\begin{aligned}
\hat{c}_h(x,y) =& \frac{1}{nh^2} \sum_{i=1}^{n} \left\{ k\left(\frac{x - X_i}{h}\right) k\left(\frac{y - Y_i}{h}\right) + k\left(\frac{x + X_i}{h}\right) k\left(\frac{y - Y_i}{h}\right) \right. \\
& + k\left(\frac{x - X_i}{h}\right) k\left(\frac{y + Y_i}{h}\right) + k\left(\frac{x + X_i}{h}\right) k\left(\frac{y + Y_i}{h}\right) \\
& + k\left(\frac{x - X_i}{h}\right) k\left(\frac{y + Y_i - 2}{h}\right) + k\left(\frac{x + X_i}{h}\right) k\left(\frac{y + Y_i - 2}{h}\right) \\
& + k\left(\frac{x + X_i - 2}{h}\right) k\left(\frac{y - Y_i}{h}\right) + k\left(\frac{x + X_i - 2}{h}\right) k\left(\frac{y + Y_i}{h}\right) \\
& \left. + k\left(\frac{x + X_i - 2}{h}\right) k\left(\frac{y + Y_i - 2}{h}\right) \right\},
\end{aligned}$$

where $k(\cdot)$ can be any symmetric kernel with support [-1, 1]. Note that although underestimation is corrected on the boundaries, the convergence rate of the bias will be of $\mathcal{O}(h)$ on the boundaries, which is larger than the usual rate $\mathcal{O}(h^2)$ obtained in the interior. This method can be generalized to cases of higher dimensions.

3.2.3 The Penalized Exponential Series Estimator

Gao et al. (2015) propose a penalized exponential series estimator (ESE) for copula density estimation. Unlike series density estimators, the penalized exponential series estimator always generates positive density estimates. The idea is that we can approximate the log copula density function by a linear combination of basis functions and penalize the roughness to balance between goodness-of-fit and parsimony, which leads to penalized maximum-likelihood estimation (MLE). Akaike information criterion (AIC), Bayesian information criterion (BIC) and the cross-validation method can be applied for model selection. Note that it is rather expensive to implement the leave-one-out cross validation for multivariate ESE with a large number of basis functions. To make the penalized ESE practical in the multivariate case, Gao et al. (2015) propose an approximate cross-validated log likelihood which requires calculating ESE based on the full sample only once.

Let $\phi_k(x,y), k = 1 \ldots K$ be a series of linearly independent basis functions defined on the unit square. We approximate $c(x, y)$ by

$$\hat{c}(x, y) = \frac{exp(g(x, y))}{\int exp(g(x, y))dxdy},$$

where $g(x, y) = a'\phi(x, y)$ with $a = (a_1, \ldots, a_K)'$ and $\phi(x, y) = (\phi_1(x, y), \ldots, \phi_K(x, y))'$. The penalized MLE objective function is given by

$$\mathcal{Q} = \frac{1}{n}\sum_{i=1}^{n} a'\phi(X_i, Y_i) - \ln \int exp(g(x, y))dxdy - \frac{\lambda}{2}a'Wa,$$

where W is a positive definite weight matrix for the roughness penalty and λ is the smoothing parameter. If one applies the leave-one-out cross-validation using the above penalized MLE objective function to choose λ, it is numerically impractical. To overcome this problem, the cross validated log likelihood approximation is given below

$$\mathcal{L}_{-} \approx \mathcal{L} - \frac{1}{n(n-1)}\text{trace}(\Phi\hat{H}^{-1}\Phi') + \frac{1}{n^2(n-1)}(\iota'\Phi)\hat{H}^{-1}(\Phi'\iota),$$

where H denotes the Hessian matrix of $\mathcal{Q}$ and $\mathcal{L}$ denotes the quasi-likelihood function

$$\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n} a'\phi(X_i, Y_i) - \ln \int exp(g(x, y))dxdy,$$

and Φ is a $n \times K$ matrix with the ith row being $\big(\phi_1(X_i, Y_i), \ldots, \phi_K(X_i, Y_i)\big)$ and ι is an $n \times 1$ vector with every element equal to unity.

3.2.4 The Sieve MLE Based on the Bernstein Polynomials

The Bernstein copula estimator was first studied by Sancetta and Satchell (2004) for independent and identically distributed data and then by Bouezmarni et al. (2010, 2013) for dependent data and for unbounded density copula functions. Here the Sieve MLE based on the Bernstein polynomials is considered. For a given point $\mathbf{u} = (u_1, u_2)$ in the unit square $(0, 1)^2$, the Bernstein copula density at $\mathbf{u}$ is given by

$$c_J(u_1, u_2; \boldsymbol{\omega}) = J_N^2 \sum_{v_1=0}^{J-1} \sum_{v_2=0}^{J-1} \omega_{(v_1,v_2)} \prod_{l=1}^{2} \binom{J-1}{v_l} u_l^{v_l} (1-u_l)^{J-v_l-1}, \tag{6}$$

where J_N is an integer that plays the role of a bandwidth parameter and $\boldsymbol{\omega} = \left\{\omega_{(v_1,v_2)}\right\}_{v_1,v_2=0,\ldots,J-1}$ denotes the coefficient of the Bernstein polynomials indexed by $v = (v_1, v_2)$. The estimation problem given sample set $(X_i, Y_i)_{i=1}^N$ is in fact a parametric likelihood maximization problem: for a given $J \in \mathbb{N}$

$$\begin{aligned} &\arg\max_{\boldsymbol{\omega}} \sum_{i=1}^{N} \log c_J(X_i, Y_i; \boldsymbol{\omega}), \\ &\text{s.t.} \quad \omega_{(v_1,v_2)} > 0; \;\; v_1, v_2 = 0, \ldots, J-1 \\ &\qquad \sum_{v_1,v_2} \omega_{(v_1,v_2)} = 1; \end{aligned} \tag{7}$$

note that $\sum\limits_{v_1,v_2} \omega_{(v_1,v_2)} = 1$ guarantees that $c(\,\cdot\,,\cdot\,; \boldsymbol{\omega})$ is indeed a density function. The sieve MLE with Bernstein polynomials can be easily generalized to higher dimensions by applying the following multivariate Bernstein copula density at $\mathbf{u} = (u_1, \cdots, u_d)$

$$c_{J_N}(\mathbf{u}) = J_N^d \sum_{v_1=0}^{J_N-1} \cdots \sum_{v_d=0}^{J_N-1} \omega_v \prod_{l=1}^{d} \binom{J_N-1}{v_l} u_l^{v_l} (1-u_l)^{J_N-v_l-1}. \tag{8}$$

For initial values of the coefficient, we can let ω_v be the multivariate empirical density estimator, i.e., $\omega_v = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}(U_i \in H_v)$, where $U = (F_1(X_1), \ldots, F_d(X_d))$, $\mathbb{I}(\cdot)$ is the indicator function and

$$H_v = \left[\frac{v_1}{J_N}, \frac{v_1+1}{J_N}\right] \times \cdots \times \left[\frac{v_d}{J_N}, \frac{v_d+1}{J_N}\right].$$

c_{J_N} with the above ω_v is related to the empirical Bernstein estimator proposed in Sancetta and Satchell (2004).

3.2.5 The Empirical Beta Copula

Segers et al. (2017) propose to use the empirical beta copula to estimate the copula nonparametrically. It is shown in their paper that the empirical beta copula is a genuine copula. The estimation procedure is described below.

Let $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,d}), i \in \{1, \ldots, n\}$ be independent and identically distributed random vectors, and assume that the cumulative distribution function, F, of $\boldsymbol{X}_i$ is continuous. For $i \in \{1, \ldots, n\}$and $j \in \{1, \ldots, d\}$, let $R_{i,j}^{(n)}$ be the rank of $X_{i,j}$ among $X_{1,j}, \ldots, X_{n,j}$; namely,

$$R_{i,j}^{(n)} = \sum_{k=1}^{n} \mathbb{I}\{X_{k,j} \leq X_{i,j}\}.$$

The empirical beta copula is defined by

$$\mathbb{C}_n^{\beta}(\boldsymbol{u}) = \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{d} F_{n,R_{i,j}^{(n)}}(u_j),\ \boldsymbol{u} = (u_1, \ldots, u_d) \in [0,1]^d,$$

where, for $u \in [0,1]$ and $r \in \{1, \ldots, n\}$, $F_{n,r}(u)$ is the beta distribution $\mathcal{B}(r, n+1-r)$.

The copula density estimation is the main interest of this paper instead of copula estimation. The copula density derived from the empirical beta copula above is:

$$\mathbb{c}_n^{\beta}(\boldsymbol{u}) = \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{d} f_{n,R_{i,j}^{(n)}}(u_j),\ \boldsymbol{u} = (u_1, \ldots, u_d) \in [0,1]^d,$$

where, for $u \in [0,1]$ and $r \in \{1, \ldots, n\}$, $f_{n,r}(u)$ is the density function of the beta distribution, $\frac{n!}{(r-1)!(n-r)!} x^{r-1}(1-x)^{n-r}$. Note that the condition for $\mathbb{c}_n^{\beta}$ to be a copula density is that we need to break the ties at random with the ranks $R_{i,j}^{(n)}$.

The empirical beta estimator does not require any smoothing-parameter selection. It is shown to perform very well in terms of mean squared error compared with other empirical estimators as shown in the simulation study in Segers et al. (2017).

3.3 The Idea of the New Estimator

Interest centres on simple arrangements for approximating copula densities with spline type surfaces. For ease of exposition, we will confine ourselves to a bivariate model for now. The multivariate case is considered in the Appendix B. The main purpose is to develop an adaptive estimator that preserves, even in finite samples, the copula property that

$$\int_0^1 c(x,y)dx = 1, y \in (0,1) \tag{9}$$

$$\int_0^1 c(x,y)dy = 1, x \in (0,1), \tag{10}$$

where $c(x,y)$ is a copula density. The most basic approach will be to take a piecewise linear surface on the unit square. Suppose that $f(x,y)$ is determined at points $p \in P = \{p_1, \ldots, p_n\}$, where this set includes the corners of the unit square. Then the piecewise linear points surface is defined elsewhere through a triangulation.

In fact, the uniform marginal properties 9 and 10 are quite restricted on the set of the design points P in the way that for each internal design point (u,v), there must be two more design points in the set P to guarantee that Eq.(9) and Eq.(10) hold, and one of the points have the same $x-$coordinate u and the other point should have the same $y-$coordinate v. More details are discussed in Appendix B.2. Therefore the simplest method to construct set P would be to use a set of grid points.

For now, assume that we have a regular grid on the unit square with grid points

$$\Gamma = \{(i/N, j/N)\}_{i,j \in \{0,1,\ldots,N\}}.$$

The most basic approach would be to take a piecewise linear surface on $[0,1] \times [0,1]$ with knots at Γ and then to define all points interior to a grid through a triangulation. However, the grid points Γ will not uniquely define the triangulation.

To see this, consider any grid cell i with corners denoted by $\{x_i^{(1)}, x_i^{(2)}, x_i^{(3)}, x_i^{(4)}\}$. We can fix different non-overlapping triangles where each T_i is the convex hull of three of the corners. For example, we can have lower left triangle and upper right triangle to form a cover of the cell, or have lower right triangle and upper left triangle to form a cover.

In a more general form, we can write a basis expansion of $c(x,y)$ as follows

$$c(x,y) = \sum_{i=0}^{N}\sum_{j=0}^{N} f_{ij} b_{ij}(x,y),$$

where $\{f_{ij}\}_{i,j\in\{0,1,\ldots,N\}}$ are copula density estimates at the grid points and $b_{ij}(x,y), i,j \in \{0,1,\ldots,N\}$ are certain basis functions. The basis functions are meant to smooth the density estimate in between the grid points and so it has the following properties:

$$\begin{aligned} &b_{ij}(i/N, j/N) = 1, \\ &\qquad b_{ij}(x,y) = 0 \quad \text{if } \left|x - \frac{i}{N}\right| \geq 1/N, \text{ or } \left|y - \frac{j}{N}\right| \geq 1/N. \end{aligned}$$

For reasons that will become clear shortly, the choice of b_{ij} we propose is

$$b_{ij}(x,y) = g(x - i/N)g(y - j/N),$$

where g is the triangular function $g(z) = (1 - N\,|z|)_+$.

The basis is a B-spline and the choice of the basis function is ideal for our purposes because it guarantees the copula properties (9) and (10) provided that they hold on the grid knots $x = i/N$ and $y = i/N$. This is because for $0 \leq \alpha, \beta \leq 1$ it can be established that

$$f((\alpha + i)/N, (\beta + j)/N) = (1-\alpha) f(i/N, (\beta+j)/N) + \alpha f((i+1)/N, (\beta+j)/N)$$

Thus

$$\int_0^1 f((\alpha+i)/N, y)dy = (1-\alpha)\int_0^1 f(i/N, y)dy + \alpha \int_0^1 f((i+1)/N, y)dy,$$

and with a similar argument

$$\int_0^1 f(x, (\beta+j)/N)dx = (1-\beta)\int_0^1 f(x, j/N)dx + \beta\int_0^1 f(x,(j+1)/N)dx.$$

The assumption that the uniform marginal properties hold at knots is innocuous because we do not observe the copula values at the knots but need to estimate them. Triangulation permits a nonparametric estimation of these values as well as other values of the copula density based on observations within the grid cells. The estimates obey the restrictions (9) and (10) by construction. More precisely, given a sample (x_k, y_k), $k = 1, 2, ...K$, we first calculate the estimates of $(F_1(x_k), F_2(y_k))$, for example, the pseudo-sample (u_k, v_k), $k = 1, 2, ...K$, where $(u_k, v_k) = \left(\frac{R_{k1}}{K+1}, \frac{R_{k2}}{K+1}\right)$, R_{k1} is the rank of x_k amongst $x_1, \ldots, x_K$, and R_{k2} is the rank of y_k amongst $y_1, \ldots, y_K$. Then we search for values f_{ij} $i, j \in \{0, 1, ...N\}$ to maximize

$$\sum_{k=1}^{K} \log f(u_k, v_k),$$

subject to the constraints

$$f_{i0} + 2\sum_{j=1}^{N-1} f_{ij} + f_{iN} = 2N, \;\; i = 0, 1, ..., N,$$
$$f_{0j} + 2\sum_{i=1}^{N-1} f_{ij} + f_{Nj} = 2N, \;\; j = 0, 1, ..., N,$$
$$0 \leq f_{ij} \leq 1.$$

The above procedure with linear B-splines is a convex non-parametric maximum likelihood estimation, subject to linear equality constraints - an easy problem that most available software packages can handle. The above procedure with linear B-splines is only an example of a family of desired estimators. In fact, this procedure can be extended to series estimators using B-splines with sparse grids and/or of higher degrees, without losing its two main benefits: preserving uniform marginal property and remaining easy to handle. For details see Appendix B.3. These spline estimators are consistent under mild regularity conditions. For details, see Theorem 3.1 in Chen (2007). These spline estimators as well as the Bernstein copula density estimator, unlike other series estimators, always produce non-negative density estimates. A limitation of the spline estimators and the Bernstein copula density estimators is that they cannot be used to model extreme tail behaviour defined in terms of the coefficient of tail dependence. Nevertheless, it can capture increasing dependence as we move to the tails, which can be seen in the empirical study section.

3.4 Simulation Study

To investigate the finite sample performance of the proposed spline estimators and compare them with the leading competitors described in Section 3.2 , a series of Monte Carlo simulations is conducted. The number of observations is set to be $K = 100$. Four common bivariate copulas are considered in the study: Gaussian, Frank, Clayton and Gumbel. One of the practical problems is the choice of the grid number N which serves as the "bandwidth" in the sieve estimators including the Bernstein estimator and the spline estimators. AIC, BIC and cross-validation (CV) for model selection are used. However, the theoretical implications of using these techniques are not explored. We compare the performance of the spline estimators (SE) in various cases to the data mirror estimator (DME), the naive kernel estimator (NKE), the penalized exponential estimator ($PESE$), the sieve MLE with Bernstein polynomials (SMB) and the empirical beta copula density estimator ($EBCE$).

Given that the sample size $K = 100$, we let the grid parameter N for the spline estimators range from 1 to $9 - d$ to make sure that the sample size is greater than the number of parameters to estimate. Note that d denotes the degree of the B-splines and only $d = 1$ and $d = 2$ are considered in this section. The number of estimators to be estimated for each N are $(N + d)^2$. The grid parameter J_N for the sieve MLE with Bernstein polynomials ranges from 1 to 9, to make sure that the sample size is greater than the number of parameters to estimate.

For $PESE$, the truncated power series we used is given by

$$\phi(x) = [1, x, x^2, \ldots, x^r, (x - x_1^*)_+^r, \ldots, (x - x_k^*)_+^r],$$

where $(x)_+ = \max(0, x)$ and $x_1^*, \ldots, x_k^*$ are the knots of the spline basis functions. This truncated power series performs relatively well in Gao et al. (2015). Set $r = 2, k = 2$ with $x_1^* = 1/3$ and $x_2^* = 2/3$. The tensor product contains a total of 24 basis functions, which implies that 24 parameters are to be estimated for each smoothing parameter. In the simulation, we pick among three values of the smoothing parameters: two, five, 10. For DM and NK estimators we use CV for bandwidth selection.

Table 3.1 - 3.2 contain the simulation results of estimating four bivariate copulas with Kendall's $\tau = 0.5$. Every entry is based on averages of 100 repetitions of the mean squared error (MSE) or mean squared deviation (MSD) of the estimated densities, evaluated on a 29-by-29 equally spaced grid on the unit square $(0,1)^2$.

The result in Table 3.1 - 3.2 suggests that all spline estimators perform better in terms of MSE and MAD compared to the kernel estimators, and all spline estimators perform better in terms of time cost compared to $PESE$. $SE^{AIC}_{d=1}$ dominates SMB^{AIC} both in accuracy and time cost. $SE^{AIC}_{d=2}$ dominates $PESE$ both in accuracy and time cost. $SE^{AIC}_{d=2}$ performs better that SMB^{AIC} in terms of MSE while having similar time cost.

Figure 3.1- 3.2 visualize all the copula density estimators in the case where the sample is generated from Gaussian copula, which give us a rough idea how these estimators perform . Figure 3.1 plots the copula density estimates on a 29-by-29 equally spaced grid on the unit square, based on a single simulation, while Figure 3.2 is based on the averages over 10 simulations. It seems that the spline estimators have a better performance of capturing the tail behavior, compared to other estimators except the $EBCE$. The $EBCE$ is extremely undersmoothed.

Table 3.1: MSE and MAD between the estimated densities and the true copula densities, evaluated on a 29-by-29 equally spaced grid on the unit square.

True Copula	Estimators: DME^{CV}	NKE^{CV}	$SE^{CV}_{d=1}$	$PESE^{CV}$	$EBCE$
Frank	0.3231(0.4266)	0.6410(0.6150)	0.1345(0.1538)	0.0724(0.1583)	0.3835(0.4419)
Clayton	0.3082(0.3365)	0.7132(0.5723)	0.2466(0.2346)	0.3303(0.2819)	0.4315(0.4259)
Gumbel	0.3021(0.3583)	0.9218(1.1454)	0.1271(0.1345)	0.1143(0.1728)	0.4282(0.4365)
Gaussian	0.5802(0.3975)	0.6275(0.7266)	0.1069(0.1045)	0.1057(0.1143)	0.3785(0.4330)
Time	5.710s	10.687s	779.864s	1629.233s	0.448s

Note: Applied to samples of size K = 100 with dependence parameter Kendall's $\tau = 0.5$.
All values are averaged over 100 simulations. MAD is given in parentheses.

Table 3.2: MSE and MAD between the estimated densities and the true copula densities, evaluated on a 29-by-29 equally spaced grid on the unit square.

True Copula	Estimators: $SE^{BIC}_{d=1}$	$SE^{BIC}_{d=2}$	$SE^{AIC}_{d=1}$	$SE^{AIC}_{d=2}$	SMB^{AIC}
Frank	0.2418(0.4167)	0.0878(0.1603)	0.0317(0.1376)	0.0163(0.1074)	0.0915(0.1609)
Clayton	0.4458(0.3026)	0.2308(0.1823)	0.3743(0.2414)	0.2341(0.1857)	0.3846(0.2722)
Gumbel	0.1757(0.1876)	0.1088(0.1660)	0.2005(0.2018)	0.1106(0.1542)	0.1562(0.2441)
Gaussian	0.2385(0.3721)	0.0844(0.1803)	0.0725(0.1665)	0.0612(0.1463)	0.0748(0.1015)
Time	6.675s	181.368s	1.447s	32.600s	15.12s

Note: Applied to samples of size K = 100 with dependence parameter Kendall's $\tau = 0.5$.
All values are averaged over 100 simulations. MAD is given in parentheses.
Settings are the same as in Table 3.1.

Table 3.3 - 3.4 contain the simulation results of estimating four bivariate copulas with Kendall's $\tau = 0.75$. Every entry is based on averages of 100 repetitions of MSE and MAD of the estimated densities, evaluated on a 29-by-29 equally spaced grid on the unit square $(0,1)^2$. Similar results as in Table 3.1 - 3.2 are observed in this case.

Table 3.3: MSE and MAD between the estimated densities and the true copula densities, evaluated on a 29-by-29 equally spaced grid on the unit square.

True Copula	Estimators: DME^{CV}	NKE^{CV}	$SE^{CV}_{d=1}$	$PESE^{CV}$
Frank	4.7941(1.1099)	4.8297(1.2.75)	0.2356(0.3598)	0.2395(0.3350)
Clayton	3.1369(0.5244)	3.1744(0.5556)	3.7136(0.7665)	3.4402(0.5230)
Gumbel	1.1890(0.3021)	1.1374(0.3272)	1.0815(0.1754)	1.0756(0.3753)
Gaussian	3.5897(0.9874)	4.1440(1.1629)	0.3043(0.2871)	0.4511(0.3051)
Time	0.82s	0.417s	820.235s	410.498s

Note: Applied to samples of size K = 100 with dependence parameter Kendall's $\tau = 0.75$.
All values are averaged over 100 simulations. MAD is given in parentheses.

Table 3.4: MSE and MAD between the estimated densities and the true copula densities, evaluated on a 29-by-29 equally spaced grid on the unit square.

True Copula	Estimators: $SE^{BIC}_{d=1}$	$SE^{BIC}_{d=2}$	$SE^{AIC}_{d=1}$	$SE^{AIC}_{d=2}$	SMB^{AIC}
Frank	0.2056(0.3245)	0.2475(0.3864)	0.1804(2665)	0.2235(0.3266)	0.6489(0.6104)
Clayton	3.1643(0.4972)	3.5368(0.8361)	3.2883(0.4886)	3.4915(0.6068)	3.7263(0.6743)
Gumbel	1.235(0.3597)	1.1524(0.3684)	1.0624(0.3696)	0.9776(0.3503)	1.2824(0.5147)
Gaussian	0.6329(0.4175)	0.2918(0.2990)	0.5920(0.4238)	0.2861(0.2781)	0.7088(0.5512)
Time	6.953s	193.487s	7.262s	210.167s	205.682s

Note: Applied to samples of size K = 100 with dependence parameter Kendall's $\tau = 0.75$.

All values are averaged over 100 simulations. MAD is given in parentheses.

Settings are the same as in Table 3.3.

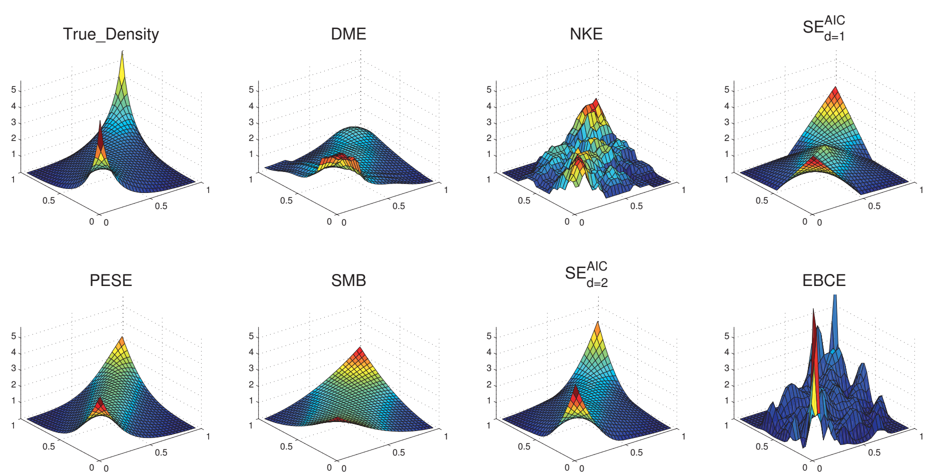


Figure 3.1: Plot of copula density estimates of a 29-by-29 equally spaced grid on the unit square, a single simulation. Gaussian copula with Kendall's $\tau = 0.5$.

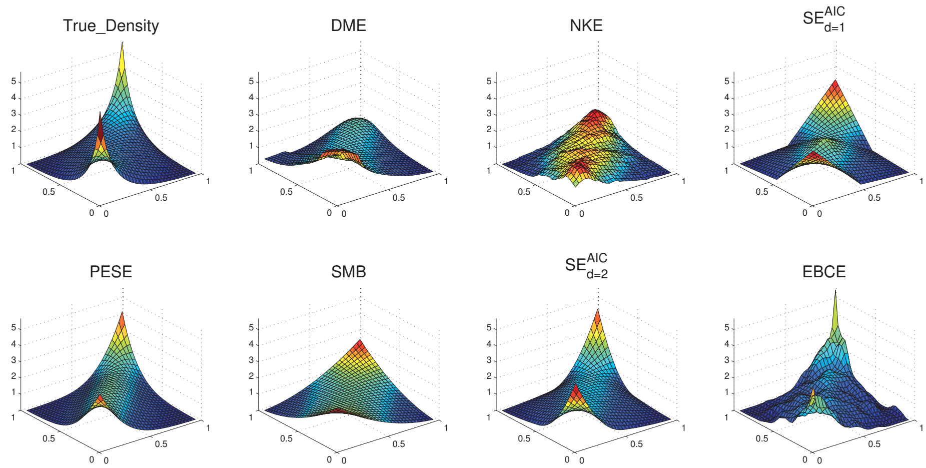


Figure 3.2: Plot of copula density estimates of a 29-by-29 equally spaced grid on the unit square, averaged over 10 simulations . Gaussian copula with Kendall's $\tau = 0.5$.

3.5 Applications to Intergenerational BMI Dependence and Gibson's Paradox

3.5.1 Application to Intergenerational BMI Dependence

In this section we investigate the intergenerational dependence of Body Mass Index (BMI) between children and parents. The dataset is part of the 2003 Community Tracking Study (CTS) Household Survey, which is the same dataset used by Gao et al. (2015). The interest is in households with adult children (18-30) living with both parents. 691 female and 715 male adult children are sampled. Table 3.5 reports some BMI summary statistics for the sample. It can be observed that the male children have higher average BMI and the intergenerational dependence of BMI is stronger between female children and parents, and between mother and children.

Figure 3.3 - Figure 3.8 report the estimated copula densities.

The first estimator is the spline estimator with linear B-splines. The number of grid parameters is $N = 1, 2, 3, \ldots, 8, 9$. AIC is used for model selection. Usually $N = 2$ or $N = 3$ is chosen.

The second estimator is the spline estimator with quadratic B-splines. The third estimator is the $PESE$ estimator. The penalty parameter is chosen from $\{2, 5, 10\}$. The fourth estimator is the

Table 3.5: BMI summary statistics (standard deviations in parentheses)

		Male	Female
Child	BMI	24.9242	23.6055
		(4.5904)	(4.9106)
Father	BMI	28.2215	28.2599
		(4.4115)	(4.3035)
Mother	BMI	26.7987	27.0045
		(5.3084)	(5.5559)
Correlation	Father	0.2491	0.3146
	Mother	0.2910	0.3668
Kendall's τ	Father	0.1633	0.2097
	Mother	0.1880	0.2328

SMB. The number of grid parameters is $N = 1, 2, 3, \ldots, 8, 9$. AIC is used for model selection.

Figure 3.3 shows the result of copula density estimates between son and father. The four copula density estimators completely capture the dependence structures between generations. All four estimators clearly suggest a positive and asymmetric dependence structure, with strong dependence at the high end of the BMI distribution. In addition, the $SE_{d=2}^{AIC}$ seems to show stronger dependence at the high end while the other competitors seem over-smoothing. We observe similar results in Figure 3.4 - Figure 3.8. On the other hand, Figure 3.3 - Figure 3.8 show that the dependence relationship differs across children's and parents' gender, which is consistent with the results from summary statistics. In addition, all four estimates in all the figures show stronger dependence at the higher end of the BMI than at the lower end. The degree of asymmetry in terms of difference in dependence at two ends is greater when a female is involved.

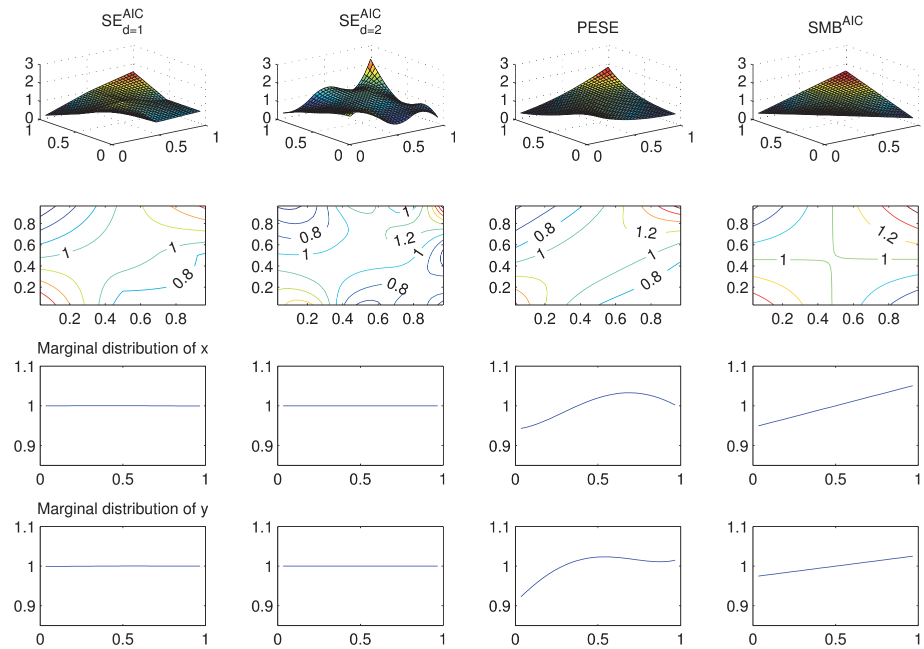


Figure 3.3: BMI copula density between son and dad

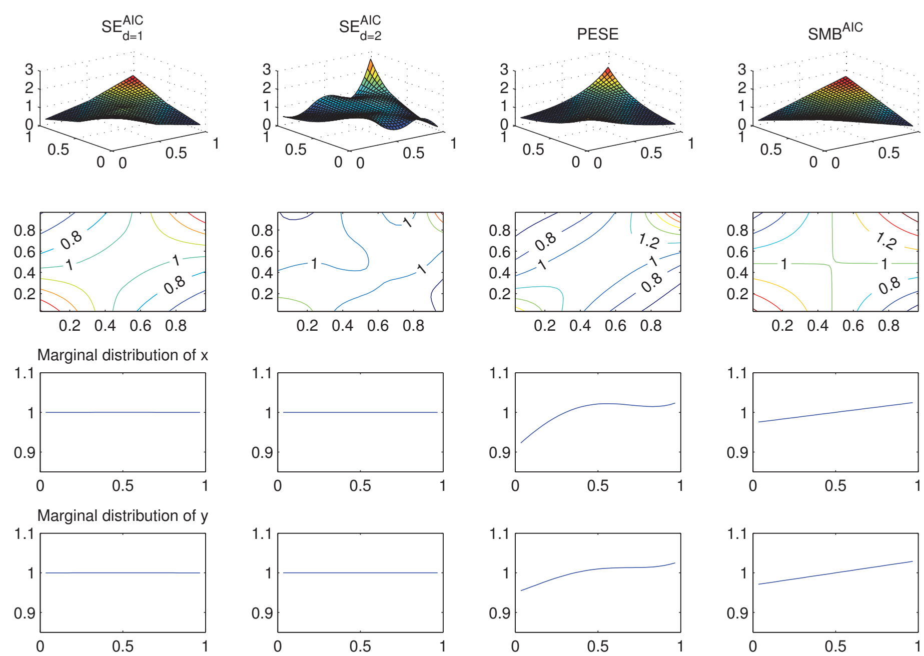


Figure 3.4: BMI copula density between son and mom

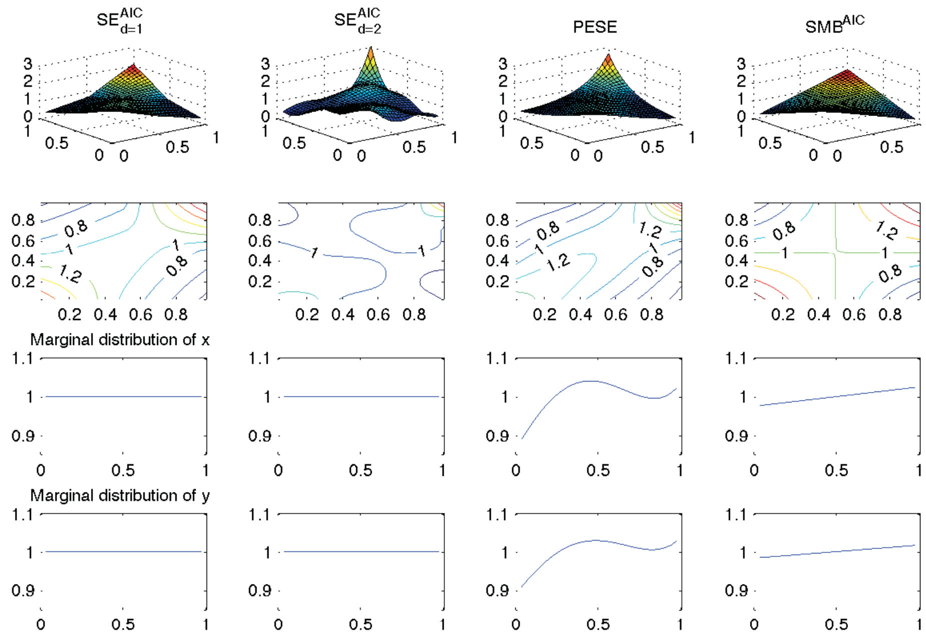


Figure 3.5: BMI copula density between daughter and dad

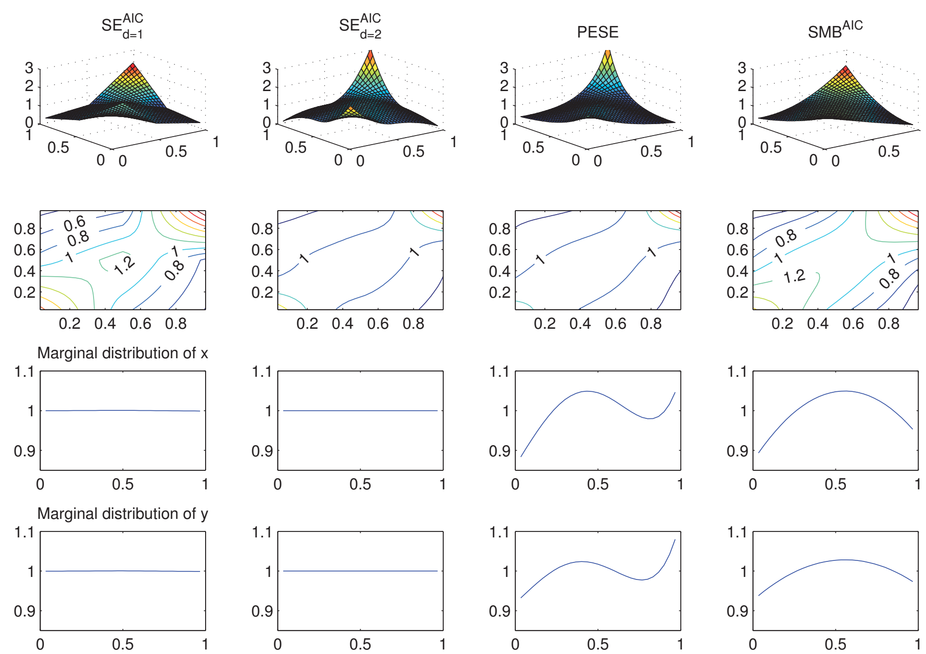


Figure 3.6: BMI copula density between daughter and mom

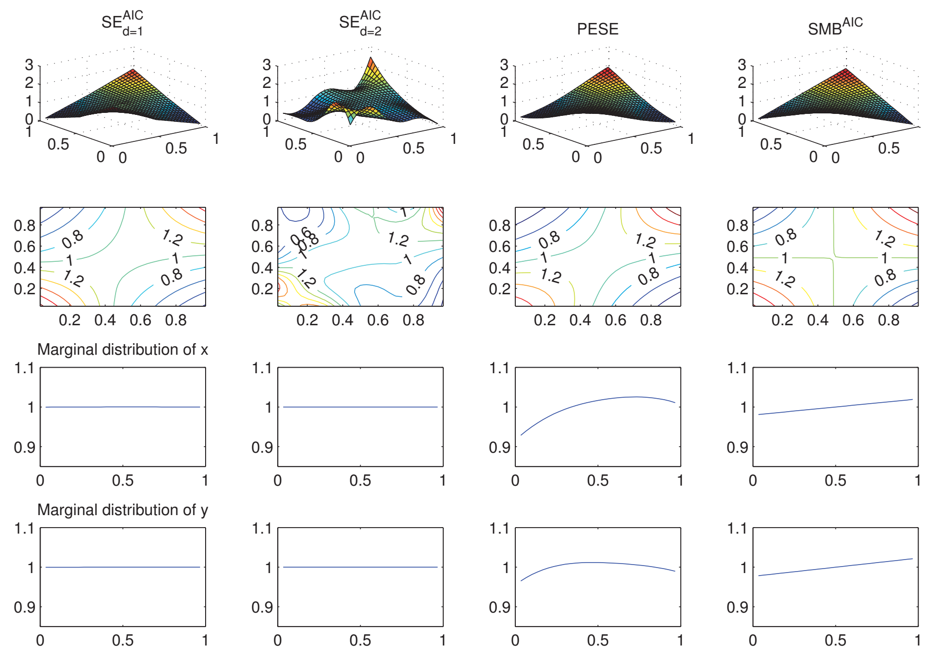


Figure 3.7: BMI copula density between son and mean (mom, dad)

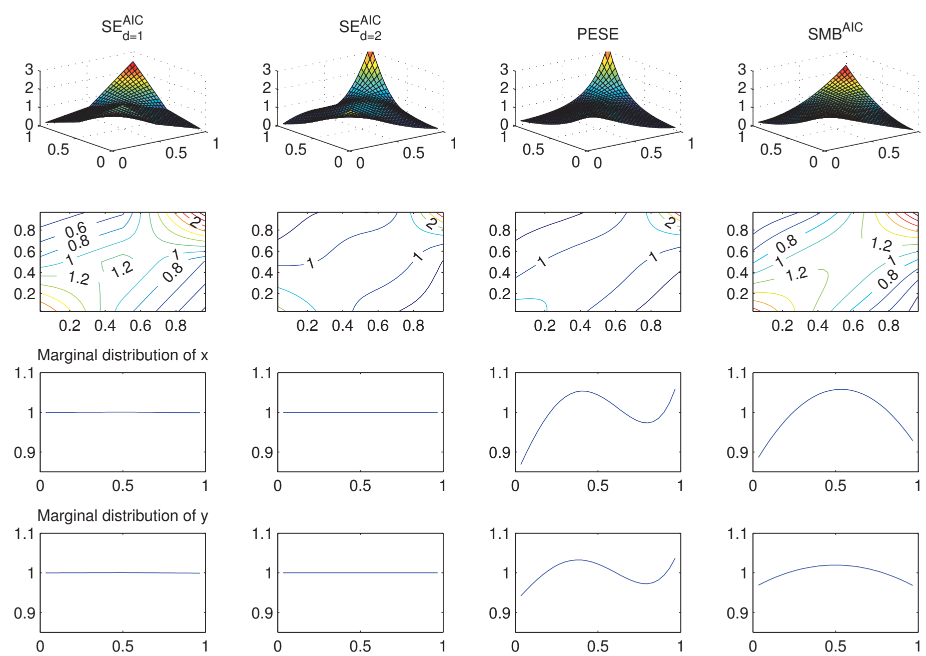


Figure 3.8: BMI copula density between daughter and mean (mom, dad)

3.5.2 Application to the Gibson's Paradox

The dependence relationship is the highlight of the Gibson's paradox. Dowd (2008) studies the prices and interest rates using parametric copulas. In the following, the same dataset is studied using four estimators $SE^{AIC}_{d=1,2}$, $PESE$ and SMB^{AIC}.

The dataset consists of prices and interest rates of UK during 1821–1913. The price level was represented by the UK cost of living (namely, Crafts and Mills, 1994, pp. 180 –182). The interest rate was represented by a series combining the annual average yield on 3% consols for the period 1821–1849 (Homer, 1963, Table 19), and the annual consol yield series for the period 1850 –1914 (Klovland, 1994, pp. 184 –185).

The positive sample Kendall's τ in Table 3.6, as well as the positive slope in Figure 3.11 strongly suggests that on average the price level series and the interest rate series have a positive association.

Figure 3.12 plots the copula density estimates of the Crafts/Mills price level series and the Homer/Klovland interest rate series. Table 3.7 presents the estimates of the Spearman's ρ of those two series, based on the plug-in estimators using $SE^{AIC}_{d=1,2}$, $PESE$ and SMB^{AIC}. All four copula density estimators show strong positive relations between price and interests along the diagonal. It seems that the tail dependence near the lower end is greater in the spline estimates than in the other estimates, and $SE^{AIC}_{d=1}$ shows fairly strong positive dependence in the centre. Those properties are consistent with the scatterplot for the combination of the Crafts-Mills price level and the Homer-Klovland consol yield in Figure 3.11, which suggest that the dependence is relatively strong in the centre area and at the lower end. However, the two SEs have relatively smaller estimates of the Spearman's ρ.

To conclude, summary statistics only capture the overall degree of dependence, while the copula densities completely summarize the underlying dependence structure between variables.

Table 3.6: Summary statistics

Series	Min	Max	Mean	Std	Skewness	Kurtosis
Crafts/Mills	83.0357	132.1429	101.4113	11.6216	0.4826	2.4141
Homer/Klovland	2.2640	4.0700	3.1245	0.3546	-0.3208	3.1868
Sample Kendall's τ	0.708					

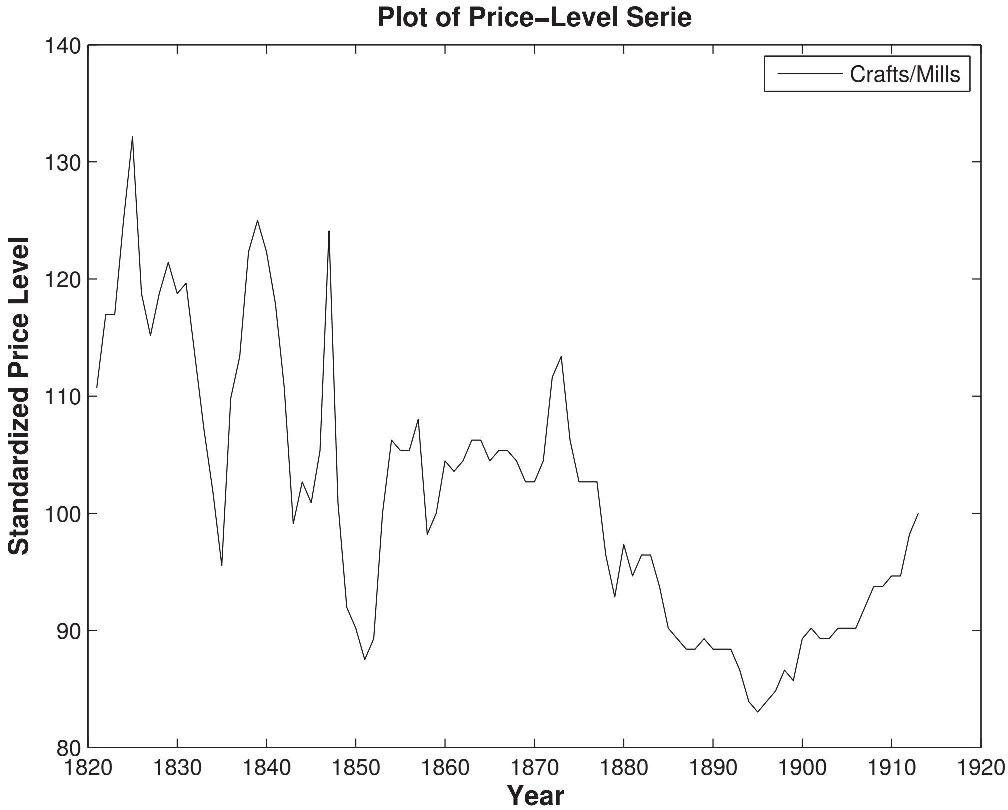


Figure 3.9: Plot of price level serie

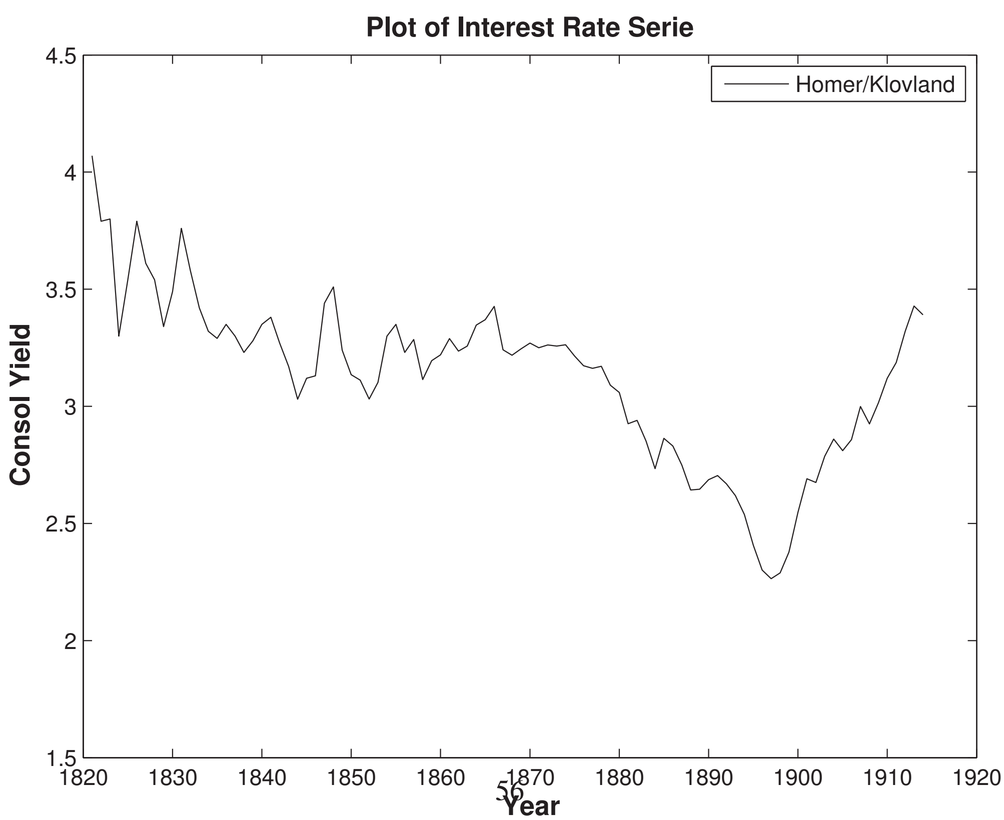


Figure 3.10: Plot of interest rate serie

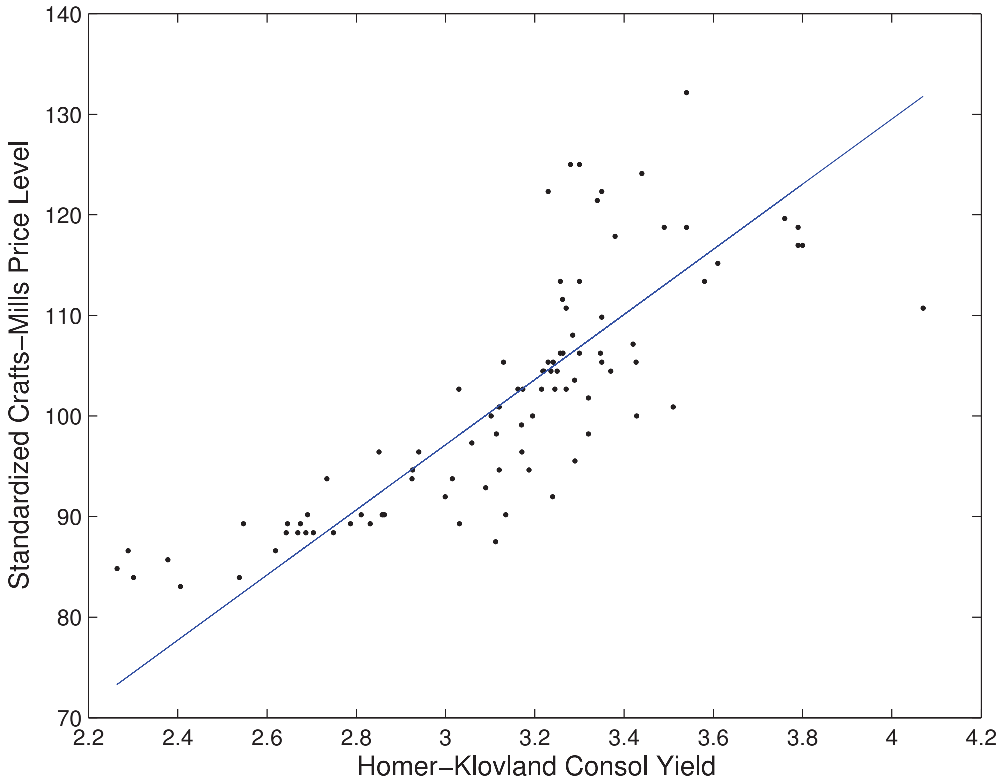


Figure 3.11: Scatterplot for the combination of the Crafts-Mills price level and the Homer-Klovland consol yield.

Table 3.7: Estimates of the Spearman's ρ of Crafts/Mills and Homer/Klovland

	Estimates of the Spearman's ρ
Sample estimator	0.876
Plug-in estimator of the Spearman's ρ based on	
$SE_{d=1}$	0.813
$SE_{d=2}$	0.785
$PESE$	0.871
SMB	0.825

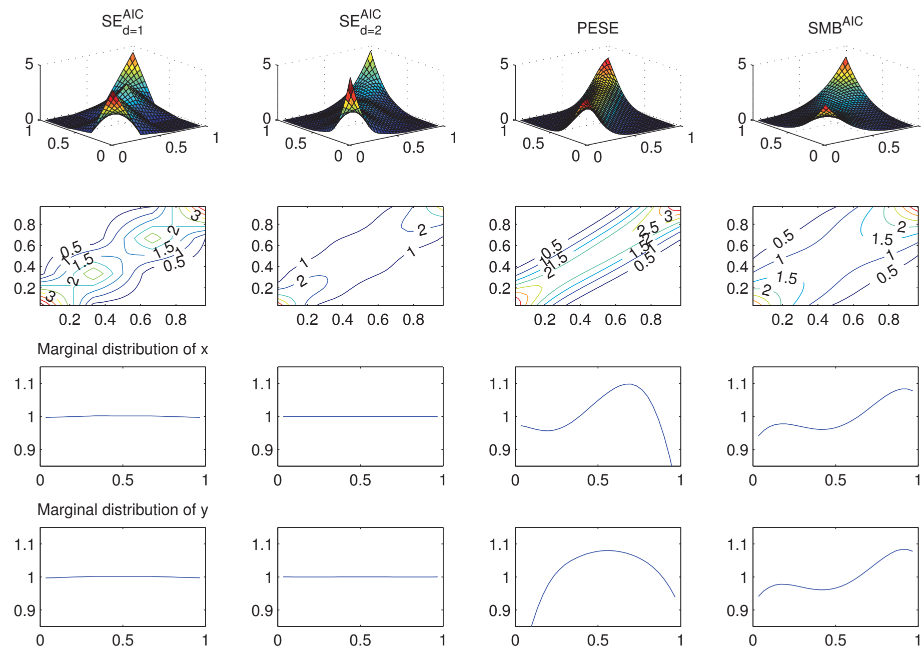


Figure 3.12: Copula density between Crafts/Mills and Homer/Klovland

3.6 Conclusion

We propose a family of spline estimators which guarantee the uniform marginal property for copula density. The family of spline estimators is strictly positive in the interior and behaves well in terms of capturing the behaviour while moving to the tails compared to the leading competitors in the literature. The estimation procedure is a convex maximization problem with linear constraints, which is numerically easy to implement and costs less time in computation compared to other sieve estimators. Our Monte Carlo simulations demonstrate the efficiency of the proposed estimators. We apply the proposed method to estimate the copula densities between children's and parents' BMI. The proposed estimators show similar results as the penalized exponential series estimators and the Bernstein estimator that the dependence relationship is generally asymmetric and stronger when BMI is high. We also apply the proposed method to examine the Gibson's paradox.

Chapter 4

Varying Coefficient Model with Missing Data

4.1 Introduction

The purpose of this chapter is to investigate the estimation of the varying coefficients model that involves matching estimators. The varying coefficient model is:

$$Y = X \cdot \beta(Z) + u. \tag{11}$$

Unlike a linear parametric regression model with constant coefficients, we study a case of nonlinear parametric regression models, of which coefficients vary with the models' specifications. For example, the marginal propensity to consume would be different between generations, and the rate of return to schooling would be different for individuals with different work experience. When we investigate issues between generations, we sometimes face the problem that the data (Y, X, Z) cannot be obtained from a single dataset and we have to combine information from two or more samples drawn from the same population. For example, there is a great deal of literature that studies intergenerational income mobility. Let Y be a son's income, and X_1 be control variables of the son's characteristics, such as education and/or years of work experience. Let X_2 be the father's or family income at the time of the son's childhood and Z be the father's education. It is likely that X_2 and Y cannot be observed in the same sample. Usually we can only observe (Y, X_1, Z) in one sample and (X_2, Z) in another sample, where Z represents some common variables (not necessarily common observations). Combining different datasets is quite common when a complete dataset is not available. Arellano and Meghir (1992) estimate female labor supply using two sets of survey data, the UK Labour Force Survey (LFS) and the Family Expenditure Survey (FES). The LFS contains the information about labour supply and the information on job-search activity, while the FES contains the information on the wage rate, other income and consumption. Here labour supply is the dependent variable and job-search activity is one of the explanatory variables. The common variables are education, age of husband and regional labor-market conditions. In Arellano and Meghir (1992), the common variables are excluded from the supply equation and are only used for the imputation of wage and other income in the combined dataset.

There is an abundance of literature studying how to identify and estimate the joint density of (Y, X, Z) based on data combination. See the literature review by Ridder and Moffitt (2007). However, as noted in Ridder and Moffitt (2007), what can be recovered from the combined data is largely dependent on the nature of the available samples and the additional assumptions that we have to make. For example, when the population moment conditions are additively separable into two samples and the available samples are rich enough that we can construct the required moment conditions from the two samples, then a two-sample generalized method of moments (GMM) estimation is enough and no additional assumptions are needed (see Angrist and Krueger, 1995; Murtazashvili et al., 2015).

However, when the two-sample GMM is not feasible due to the nature of the available samples, we need more assumptions. Suppose that Y and X are only available in two different datasets. Usually, either the assumption of conditional independence between Y and X, given the common variables Z, or exclusion conditions must be added for full inference. Obviously, the conditional independence assumption is not very attractive when we are interested in the estimation of $\mathbb{E}(Y \mid X, Z)$. On the other hand, exclusion conditions are similar to the instrumental variable estimation (IVE). We need to find variables that are excluded from the regression of interest, but are highly correlated with the missing data we want to impute into the combined data. This approach is also known as two-sample IVE. Instead of making more assumptions or requiring rich samples, we can also simply combine the missing data samples by applying a matching method, and investigate the properties of the estimators using matching. This matching method has been rigorously studied in the average treatment effects models (see Abadie and Imbens, 2006, 2011).

We consider a general case where (Y, X_1, Z) and (X_2, Z) are collected from two samples with one sample size potentially greater than the other but of the same order. Then we can apply a nearest-neighbour matching method to match these two samples based on the covariates Z. The first intuition, as in the literature on the average treatment effect, is to approximate the missing X_2 in (Y, X_1, Z) by reasonable X_2 in the larger sample determined through nearest-neighbour matching over the corresponding Z. Our investigation shows that the simple local linear estimator based on matching is inconsistent, due to the "matching discrepancy" termed in Abadie and Imbens (2006). Moreover, it is shown that the rate of convergence of the simple local linear estimator is dominated by the error of the matching discrepancy, which in turn depends on the number of matching variables. In particular, the simple local linear estimator reaches the parametric convergence rate only if the matching is conducted over one variable, instead of a high dimensional Z. In addition to the above results, we also discuss possible bias-corrected estimators.

The rest of this chapter is organized as follows. Section 4.2 reveals the inconsistency of simple local linear estimation of the regression model (12), using matched samples. Section 4.3 proposes bias-corrected estimators and examines their convergence properties. Section 4.4 examines the performance of the bias correction in finite samples using Monte Carlo simulations. The conclusions are summarized in Section 4.5. Some proofs are given in the Appendix.

4.2 Two-Sample Matching Estimator

4.2.1 Setting and Notation

Consider the following varying coefficient model

$$\begin{aligned} Y_i &= X_i^T\beta(Z_i) + u_i \\ &= X_{1i}^T\beta_1(Z_i) + X_{2i}^T\beta_2(Z_i) + u_i, \\ &\mathbb{E}(u_i \mid X_i, Z_i) = 0 \\ &\mathbb{E}({u_i}^2 \mid X_i, Z_i) = \sigma^2 \quad a.s., i \in \{1, ...n\} \end{aligned} \tag{12}$$

where the dependent variable Y_i is a scalar random variable, and $\beta(\cdot) = (\beta_1(\cdot)^T, \beta_2(\cdot)^T)^T$ is a $d_1 + d_2$ dimensional vector of unknown functions. Let $X_i = (X_{1i}^T, X_{2i}^T)^T \in \mathcal{R}^{d_1+d_2}$, where X_{1i} and X_{2i} denote d_1 and d_2 dimensional vectors of exogenous regressors, respectively. The set of $Z_i \in \mathcal{R}^1$ are continuous covariates with compact support.

Suppose that we observe two independent random samples from the same population, namely, $\{Y_i, X_{1i}, Z_{1i}\}_{i=1}^n$ and $\{X_{2j}, Z_{2j}\}_{j=1}^m$, and we construct a matching dataset of n observations $\{(Y_i, X_{1i}, X_{2j(i)}, Z_{1i}, Z_{2j(i)})\}_{i=1}^n$, where $Z_{2j(i)}$ denotes the nearest match to Z_{1i} and $X_{2j(i)}$ is the observation paired with $Z_{2j(i)}$ in the sample $\{X_{2j}, Z_{2j}\}_{j=1}^m$. Define

$$j(i) := \text{argmin}_{j\in\{1,\dots,m\}}|Z_{2j} - Z_{1i}|. \tag{13}$$

In other words, we will match the missing values X_{2i} in $\{X_{1i}, X_{2i}\}_{i=1}^n$, with $X_{2j(i)}$ in $\{X_{2j}, Z_{2j}\}_{j=1}^m$, where $j(i)$ is the index of the element that is the nearest match for Element i in terms of the matching variables Z[1].

Define $C(j)$, the number of times that Element j in the sample $\{X_{2j}, Z_{2j}\}_{j=1}^m$ is used as a match to Element i in the sample $\{Y_i, X_{1i}, Z_{1i}\}_{i=1}^n$,

$$C(j) = \sum_{i=1}^n \mathbb{1}(j = j(i)), \quad j \in \{1, ..., m\}$$

[1]As discussed in Abadie and Imbens (2006), we can also apply the nearest kth match method such that we first find k nearest matches to each i and then average the k matches to impute into the combined sample set. This will improve the performance of our corrected estimator using a single match. Further proof will be explored in future work.

where $\mathbb{1}(\cdot)$ is the indicator function, equal to one if $j = j(i)$ is true and zero otherwise.

We also define $\mathbb{A}(j)$ as the subset of the indices i, $i \in \{1, ..., n\}$, such that j is used as a match to each observation indexed from $\mathbb{A}(j)$; for instance, if $i \in \mathbb{A}(j)$, then $j = j(i)$. Clearly, the number of the elements in the set $\mathbb{A}(j)$ is $C(j)$.

4.2.2 Identification of the Two-Sample Estimator

If we can observe a complete sample, the moment condition is $\mathbb{E}(X_i u_i \mid Z_i) = 0_{d_1+d_2}$, or $\begin{pmatrix} \mathbb{E}(X_{1i} u_i \mid Z_i) \\ \mathbb{E}(X_{2i} u_i \mid Z_i) \end{pmatrix} = 0_{d_1+d_2}$. If we now have two samples of missing data $\{Y_i, X_{1i}, Z_{1i}\}_{i=1}^n$ and $\{X_{2j}, Z_{2j}\}_{j=1}^m$ as well as the constructed sample $\{(Y_i, X_{1i}, X_{2j(i)}, Z_{1i}, Z_{2j(i)})\}_{i=1}^n$, the previous moment condition cannot be used directly to identify parameters. Instead, the applicable moment condition in this case would be $\begin{pmatrix} \mathbb{E}(X_{1i} u_i \mid Z_i) \\ \mathbb{E}(X_{2j(i)} u_i \mid Z^{1,2}) \end{pmatrix} = 0_{d_1+d_2}$, where $Z^{1,2}$ contains all Z from two samples. Therefore with the above moment condition, a straightforward calculation yields

$$\begin{aligned} &\begin{pmatrix} \mathbb{E}(X_{1i} y_i \mid Z_i) \\ \mathbb{E}(X_{2j(i)} y_i \mid Z^{1,2}) \end{pmatrix} \\ &= \begin{pmatrix} \mathbb{E}(X_{1i} X_{1i}^T \mid Z_i), & \mathbb{E}(X_{1i} X_{2i}^T \mid Z_i) \\ \mathbb{E}(X_{2i} \mid Z_i)\mathbb{E}(X_{1i}^T \mid Z_i), & \mathbb{E}(X_{2i} \mid Z_i)\mathbb{E}(X_{2i}^T \mid Z_i) \end{pmatrix} \begin{pmatrix} \beta_1(Z_i) \\ \beta_2(Z_i) \end{pmatrix}. \end{aligned} \tag{14}$$

To further simplify Eq.(14), let $\Omega(z) = \mathbb{E}(XX' \mid Z = z)$ be positive definite for each z and uniformly continuous in z, and let $\Omega(z)_{(ij)}$ be the (i, j)th block element of matrix $\Omega(z)$. Further denote the conditional expectations of X_1 and X_2, given Z, as

$$g_1(Z) = \mathbb{E}(X_1 \mid Z), \quad g_2(Z) = \mathbb{E}(X_2 \mid Z).$$

Define v_1 and v_2:

$$v_1 = X_1 - g_1(Z), \quad v_2 = X_2 - g_2(Z).$$

Let $g(Z) = (g_1(Z)^T, g_2(Z)^T)^T$ and $v = (v_1^T, v_2^T)^T$. Denote the conditional variance of v as $\mathbb{E}(vv^T|Z) = \Sigma$, the conditional variance of v_1 as $\mathbb{E}(v_1 v_1^T|Z) = \Sigma_{11}$, the conditional variance of v_2 as $\mathbb{E}(v_2 v_2^T \mid Z) = \Sigma_{22}$, and the conditional cross-covariances are $\mathbb{E}(v_1 v_2^T | Z) = \Sigma_{12}$ and $\mathbb{E}(v_2 v_1^T | Z) = \Sigma_{21}$.

Recall that the model is identifiable if the matrix

$$Q = \begin{pmatrix} \mathbb{E}(X_{1i} X_{1i}^T \mid Z_i) & \mathbb{E}(X_{1i} X_{2i}^T \mid Z_i) \\ \mathbb{E}(X_{2i} \mid Z_i)\mathbb{E}(X_{1i}^T \mid Z_i) & \mathbb{E}(X_{2i} \mid Z_i)\mathbb{E}(X_{2i}^T \mid Z_i) \end{pmatrix} \tag{15}$$

from Eq.(14) is invertible.

Note that if we assume $\Sigma_{12} = 0_{d_1 \times d_2}$, which means that X_1 and X_2 are uncorrelated conditional on Z, then the matrix in (15) is equivalent to

$$\begin{pmatrix} \Omega(z)_{(11)} & \Omega(z)_{(12)} \\ \Omega(z)_{(21)} & \Omega(z)_{(22)} - \Sigma_{22} \end{pmatrix}.$$

In this case, Q is positive-semidefinite matrix. If Q is invertible, its leading principal minors should be all positive for Q to be positive definite. Therefore both $\mathbb{E}(X_{1i} X_{1i}^T \mid Z_i)$ and $\mathbb{E}(X_{2i} \mid Z_i)\mathbb{E}(X_{2i}^T \mid Z_i)$ should be invertible, which means X_{2i}'s dimension can only be $d_2 = 1$ and X cannot include the intercept.

To avoid the above situation, instead of assuming $\Sigma_{12} = 0_{d_1 \times d_2}$, we can estimate Σ_{12} directly. Assume that the conditional moment of each element of X_2 is a linear function of X_1 and Z, $X_{2i}^k = X_{1i}^T \beta_x^k + Z_i \beta_z^k + \epsilon_i, \forall k \in 1, \ldots, d_2$, then β_x^k and β_z^k are identifiable if

$$M^k = \begin{pmatrix} \mathbb{E}(X_{1i})\mathbb{E}(X_{1i}^T), & \mathbb{E}(X_{1i})\mathbb{E}(Z_i) \\ \mathbb{E}(X_{1i} Z_i), & \mathbb{E}(Z_i^2) \end{pmatrix}$$

is invertible. See Hirukawa and Prokhorov (2016) for details on linear regression models using matched data.

In the following sections, we will confine ourselves to the simpler case where $\Sigma_{12} = 0_{d_1 \times d_2}$ is assumed.

4.2.3 Two-Sample Naive Local Linear Estimator

Suppose that $\{Y_i, X_{1i}, X_{2j(i)}, Z_{1i}, Z_{2j(i)}\}_{i=1}^n$ is a matched sample from two samples $\{Y_i, X_{1i}, Z_{1i}\}_{i=1}^n$ and $\{X_{2j}, Z_{2j}\}_{j=1}^m$. Denote $X_{j(i)} = (X_{1i}^T, X_{2j(i)})^T$ as the matching pair for $X_i = (X_{1i}^T, X_{2i})^T$, $i \in \{1, ..., n\}$. We can approximate $\beta(Z_{1i})$ in a small neighbourhood of z by a linear function:

$$\beta(Z_{1i}) \approx \theta_0 + \frac{1}{h}(Z_{1i} - z)\theta_1,$$

where $\theta_0 = \beta(z)$, and $\theta_1 = h\beta'(z)$. Then we consider a naive local linear estimator as minimizers with respect to (θ_0, θ_1) of the following weighted local least-squares problem:

$$\sum_{i=1}^n \left(Y_i - X_{j(i)}^T\theta_0 - \frac{Z_{2j(i)} - z}{h} X_{j(i)}^T\theta_1\right)^2 K_h\left(Z_{2j(i)} - z\right), \tag{16}$$

where $K(\cdot)$ is a kernel function, h is a bandwidth and $K_h(\cdot) = \frac{K(\cdot/h)}{h}$.

The local linear estimator $\hat{\beta}(z)$ is given by the solution for θ_0 to the problem of minimizing (16). And the solution admits the following expression:

$$\left(\hat{\beta}(z)^T, h\hat{\beta}'(z)^T\right)^T = \left(\left(D^{X_m}\right)^T W D^{X_m}\right)^{-1} \left(D^{X_m}\right)^T WY, \tag{17}$$

where

$$D^{X_m} = \begin{pmatrix} X_{j(1)}^T & X_{j(1)}^T \frac{Z_{2j(1)}-z}{h} \\ ... & ... \\ X_{j(n)}^T & X_{j(n)}^T \frac{Z_{2j(n)}-z}{h} \end{pmatrix},$$

$$W = diag\left(K_h\left(Z_{2j(1)} - z\right), ..., K_h\left(Z_{2j(n)} - z\right)\right),$$

and

$$Y = (Y_1, ..., Y_n)^T.$$

Let $\Theta(z) = \left(\beta(z)^T, \beta'(z)^T\right)^T$, and define its estimator to be

$$\hat{\Theta}(z) = H^{-1}\left((D^{X_m})^T W D^{X_m}\right)^{-1}\left(D^{X_m}\right)^T WY,$$

where $H = diag(1, ..., 1, h, ..., h)$ is a $2(d_1 + 1) \times 2(d_1 + 1)$ matrix with the first $d_1 + 1$ diagonal elements being 1 and the remaining diagonal elements h. We can write the estimator of $\beta(z)$ as

$$\hat{\beta}(z)^T = eH^{-1}((D^{X_m})^T W D^{X_m})^{-1}(D^{X_m})^T WY,$$

where e is a $1 \times 2(d_1 + 1)$ matrix with the first $d_1 + 1$ elements being 1 and the rest of the diagonal elements 0. We call $\hat{\beta}(z)$ the two-sample naive local linear estimator for varying coefficient models. For the following analysis, we modify the expression for the naive local linear estimator in a concise way,

$$H\hat{\Theta}(z) = D_n(z)^{-1} N_n(z),$$

where $D_n(z) = \begin{pmatrix} D_{n,0} & D_{n,1} \\ D_{n,1} & D_{n,2} \end{pmatrix}$ and $N_n(z) = \begin{pmatrix} N_{n,0} \\ N_{n,1} \end{pmatrix}$,

$$\begin{aligned}
D_{n,0}(z) &= \frac{1}{n}\sum_{i=1}^{n} K_h(Z_{2j(i)} - z) X_{j(i)} X_{j(i)}^T, \\
D_{n,1}(z) &= \frac{1}{nh}\sum_{i=1}^{n} K_h(Z_{2j(i)} - z)(Z_{2j(i)} - z) X_{j(i)} X_{j(i)}^T, \\
D_{n,2}(z) &= \frac{1}{nh^2}\sum_{i=1}^{n} K_h(Z_{2j(i)} - z)(Z_{2j(i)} - z)^2 X_{j(i)} X_{j(i)}^T, \\
N_{n,0}(z) &= \frac{1}{n}\sum_{i=1}^{n} K_h(Z_{2j(i)} - z) X_{j(i)} Y_i, \\
N_{n,1}(z) &= \frac{1}{nh}\sum_{i=1}^{n} K_h(Z_{2j(i)} - z)(Z_{2j(i)} - z) X_{j(i)} Y_i.
\end{aligned}$$

4.2.4 Large Sample Properties of the Two-Sample Naive Local Linear Estimator

Assumption 1 *$\{Y_i, X_{1i}, Z_{1i}\}_{i=1}^n$ and $\{X_{2j}, Z_{2j}\}_{j=1}^m$ are two independent samples from the same population $\{Y, X_1, X_2, Z\}$ with missing data.*

Assumption 2 *(1) The density function $f(\cdot)$ of Z is bounded, and has continuous second derivatives on a compact set.*

(2) *The matrix $f(z)\Omega(z)$ is invertible, and so is the matrix*

$$f(z)\begin{pmatrix}\Omega(z)_{(11)}, & \Omega(z)_{(12)} \\ \Omega(z)_{(21)}, & \Omega(z)_{(22)}-\Sigma_{22}\end{pmatrix}$$

over the domain of z.

(3) *$\beta_j(\cdot)$, with $j \in \{1, ..., d_1 + d_2\}$ has continuous second derivatives at each point z in the support of Z.*

Assumption 3 *The kernel function $K(\cdot)$ is symmetric and a bounded second-order kernel function with compact support. $K(\cdot)$ is Lipschitz continuous. The bandwidth h satisfies $nh \to \infty$ and $h \to 0$ as $n \to \infty$, $nh^8 \to 0$ and $nh^2/(\log n)^2 \to \infty$ as $n \to \infty$.*

Assumption 4 (1) *Functions $g_1(\cdot)$ and $g_1(\cdot)$ have continuous second derivatives at each point z on the support of Z,*

(2) *the fourth moment of the conditional distribution of Y given $Z = z$ exists and is bounded uniformly in z,*

(3) *σ^2 is bounded away from zero,*

(4) *$\frac{m}{n} \to \kappa \in (0, \infty)$ as $n,\ m \to \infty$ jointly.*

Assumption 1 specifies a two-sample setup. As both sample sizes $n \to \infty$ and $m \to \infty$, we cannot apply the law of large numbers or the central limit theorem to the combined sample constructed by nearest matching, because the i.i.d. property of the combined sample is destroyed by replacing the fixed index i with a random index $j(i)$. Assumptions 2−3 are standard assumptions for the consistency and asymptotic normality of the local linear estimators of the varying coefficient models. Assumption 4 adds additional assumptions needed to re-establish the consistency and asymptotic normality results for our two-sample nearest matching estimators.

We first show that the denominator in our two-sample matching estimator is consistent for its expectation. We then show that the numerator is also consistent but the resulting two-sample matching estimator is biased. Then we will prove that, without the conditional bias term, the matching estimator is $N^{1/2}$ consistent and asymptotically normal. In the following analysis, we denote

$$\mu_j = \int u^j K(u) du,$$

and

$$\nu_j = \int u^j K^2(u) du, \;\; j = 1, 2, 3.$$

We use $\bigotimes$ to denote the kronecker product.

Lemma 13 ***Convergence of the Denominator***. *Under Assumptions 1–4,*

$$D_n(z) \xrightarrow{p} f(z)\Omega(z) \bigotimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix},$$

Proof. See Appendix C.1. ■

Lemma 14 ***Convergence of the Numerator***. *Under Assumptions 1–4,*

$$N_n(z) - \frac{1}{n}(D^{X_m})^T W D^{X_m} \begin{pmatrix} \beta(z) \\ h\beta'(z) \end{pmatrix} \xrightarrow{p}$$
$$f(z) \begin{pmatrix} 0 & \Sigma_{12} \\ 0 & -\Sigma_{22} \end{pmatrix} \beta(z) \bigotimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} + f(z) \begin{pmatrix} 0 & \Sigma_{12} \\ 0 & -\Sigma_{22} \end{pmatrix} h\beta'(z) \bigotimes \begin{pmatrix} 0 \\ \mu_2 \end{pmatrix}.$$

Proof. See Appendix C.2. ■

Theorem 15 ***Inconsistency of the Naive Estimator***. *Suppose that Assumptions 1–4 hold. Then*

$$\hat{\beta}(z) - \beta(z) = bias(z)\beta(z) + O_p\left(h^2 + \frac{1}{\sqrt{nh}}\right), \tag{18}$$

where

$$bias(z) = \Omega(z)^{-1} \begin{pmatrix} 0 & \Sigma_{12} \\ 0 & -\Sigma_{22} \end{pmatrix}.$$

Proof. Theorem 15 holds by combining the results from Lemma 13 and 14. ■

Theorem 16 ***Asymptotic Normality for the Naive Estimator***. *Suppose that Assumptions 1–4 hold. Then*

$$\sqrt{nh}V(z)^{-1/2}D(z)\left(H(\hat{\Theta}(z)-\Theta(z))-bias(z)\bigotimes\begin{pmatrix}\beta(z)\\ h\beta'(z)\end{pmatrix}\right)\xrightarrow{d}\mathcal{N}(0,I),$$

where $D(z)=f(z)\Omega(z)\bigotimes\begin{pmatrix}1 & 0\\ 0 & \mu_2\end{pmatrix}$ *and*

$$V(z)=\sum_{j=1}^{m}\sigma^2(C(j)K_h(Z_{2j}-z))^2\Omega(Z_{2j})\bigotimes\begin{pmatrix}1 & (Z_{2j}-z)/h\\ (Z_{2j}-z)/h & ((Z_{2j}-z)/h)^2\end{pmatrix}.$$

Proof. From the argument in Lemma 14,

$$\sqrt{nh}\left(H(\hat{\Theta}(z)-\Theta(z))-bias(z)\bigotimes\begin{pmatrix}\beta(z)\\ h\beta'(z)\end{pmatrix}\right)=\sqrt{nh}D_n^{-1}(z)I_4,$$

where $I_4=\begin{pmatrix}\frac{1}{n}\sum_{i=1}^{n}X_{j(i)}K_h(Z_{2j(i)}-z)u_i\\ \frac{1}{n}\sum_{i=1}^{n}X_{j(i)}K_h(Z_{2j(i)}-z)u_i(Z_{2j(i)}-z)/h\end{pmatrix}$. By Lemma 13,

$$D_n(z)\xrightarrow{p}f(z)\Omega(z)\bigotimes\begin{pmatrix}1 & 0\\ 0 & \mu_2\end{pmatrix}.$$

Firstly, we will show I_4 is a random vector made up of sums of conditionally independent random variables, given all of Z from the two samples. Denote $I_4=(I_{4,0}^T,I_{4,1}^T)^T$, where

$$I_{4,0}=\frac{1}{n}\sum_{i=1}^{n}X_{j(i)}K_h(Z_{2j(i)}-z)u_i,$$

and

$$I_{4,1}=\frac{1}{n}\sum_{i=1}^{n}X_{j(i)}K_h(Z_{2j(i)}-z)u_i(Z_{2j(i)}-z)/h.$$

Since $\mathbb{A}(j)$ indicates the subset of the index i, $i\in\{1,...,n\}$, such that j is used as a match to each observation indexed by i, $i \in \{1, ..., n\}$, we can rewrite $I_{4,0}$ and $I_{4,1}$ as

$$\begin{aligned} I_{4,0} &= \frac{1}{n}\sum_{j=1}^{m} I_{4,0}^{j} \\ &= \frac{1}{n}\sum_{j=1}^{m}\sum_{i\in\mathbb{A}(j)} X_{(i,j)}K_h(Z_{2j}-z)u_i, \end{aligned}$$

and

$$\begin{aligned} I_{4,1} &= \frac{1}{n}\sum_{j=1}^{m} I_{4,1}^{j} \\ &= \frac{1}{n}\sum_{j=1}^{m}\sum_{i\in\mathbb{A}(j)} X_{(i,j)}K_h(Z_{2j}-z)u_i(Z_{2j}-z)/h, \end{aligned}$$

where $X_{(i,j)} = (X_{1i}^T, X_{2j}^T)^T$.

Recall that $Z^{1,2}$ represents all of the Z from the two samples. Then conditional on $Z^{1,2}$, the individual terms in $I_{4,0}^j = \sum_{i\in\mathbb{A}(j)} X_{(i,j)}K_h(Z_{2j}-z)u_i$ are independent with zero means and non-identically distributed. To see this, first note that the number of elements in the index set $\mathbb{A}(j)$ is $C(j)$, which is the number of times Z_{2j} is used as a match, $j \in \{1, ..., m\}$. Conditional on $Z^{1,2}$, $C(j)$ is nonstochastic. Therefore conditional on $Z^{1,2}$, the sum in $I_{4,0}$ are made up of all $i.i.d.$ terms for all j, $j \in \{1, ..., m\}$. As a result, $I_{4,0}^j$ are independent for all j, $j \in \{1, ..., m\}$. The conditional variance of $I_{4,0}^j$ is $(C(j)K_h(Z_{2j}-z))^2\sigma^2\Omega(Z_{2j}) + o(1)$.

Likewise, conditional on all the $Z^{1,2}$, the individual terms in $I_{4,1}^j$ are also independent with zero means and non-identically distributed. The conditional variance of $I_{4,1}^j$ is $(C(j)K_h(Z_{2j}-z)(Z_{2j}-z)/h)^2\sigma^2\Omega(Z_{2j}) + o(1)$.

Next, we will use the Cramèr-Wold device and the Lindeberg-Feller central limit theorem to derive the asymptotic distribution of $\sqrt{n}I_4$. Denote the dimension of $X = (X_1, X_2)$ as $p = d_1 + d_2$.

For any $2p \times 1$ nonzero vector $\tau = (\tau_1, ..., \tau_{2p})^T$, we have

$$\sqrt{nh}\tau^T I_4 = \sqrt{\frac{m}{n}} \frac{\sqrt{h}}{\sqrt{m}} \sum_{j=1}^{m} \left\{ \sum_{k=1}^{p} \sum_{i \in \mathbb{A}(j)} \tau_k X_{(i,j)k} K_h(Z_{2j} - z) u_i \right.$$
$$\left. + \sum_{k=1}^{p} \sum_{i \in \mathbb{A}(j)} \tau_{k+p} X_{(i,j)k} K_h(Z_{2j} - z) \left((Z_{2j} - z)/h \right) u_i \right\}.$$

Similarly to the argument for the conditional independence of $I_{4,0}^j$ and $I_{4,1}^j$ given $Z^{1,2}$,

$$\sqrt{h} I_4^j = \sqrt{h} \left\{ \sum_{k=1}^{p} \sum_{i \in \mathbb{A}(j)} \tau_k X_{(i,j)k} K_h(Z_{2j} - z) u_i \right.$$
$$\left. + \sum_{k=1}^{p} \sum_{i \in \mathbb{A}(j)} \tau_{k+p} X_{(i,j)k} K_h(Z_{2j} - z) \left((Z_{2j} - z)/h \right) u_i \right\}$$

are independent and we can apply the Lindeberg-Feller central limit theorem if the Lindeberg-Feller conditions are satisfied. Denote the variance of $\sum_{j=1}^{m} I_4^j$ is V_τ.

For given $Z^{1,2}$, the Lindeberg-Feller condition requires that

$$\frac{1}{mV_\tau} \sum_{j=1}^{m} \mathbb{E} \left[(I_4^j)^2 \mathbb{1}\{|I_4^j| \geq \eta \sqrt{mV_\tau}\} \mid Z^{1,2} \right] \to 0$$

for all $\eta > 0$. By applying Hölder's and Markov's inequalities, we have

$$\frac{1}{mV_\tau} \sum_{j=1}^{m} E \left[(I_4^j)^2 \mathbb{1}\{|I_4^j| \geq \eta \sqrt{mV_\tau}\} \mid Z^{1,2} \right]$$
$$\leq \frac{1}{mV_\tau} \sum_{j=1}^{m} \left(\mathbb{E} \left[(I_4^j)^4 \mid Z^{1,2} \right] \right)^{\frac{1}{2}} \frac{\mathbb{E} \left[(I_4^j)^2 \mid Z^{1,2} \right]}{\eta^2 m V_\tau}$$
$$\leq \frac{1}{mV_\tau} \sum_{j=1}^{m} \left(C(j)^4 K_h(Z_{2j} - z)^4 \mathbb{E}[u_j^4 \mid Z^{1,2}] \Psi(\tau, Z_{2j}) \right)^{\frac{1}{2}}$$
$$\frac{C(j)^2 K_h(Z_{2j} - z)^2 \sigma^2 \Phi(\tau, Z_{2j})}{\eta^2 m V_\tau}$$
$$\leq \frac{\bar{c}^{\frac{1}{2}}}{\eta^2 \sigma^2} \frac{1}{m} \left(\frac{1}{m} \sum_{j=1}^{m} C(j)^4 \right),$$

where $\bar{c} = sup_z \mathbb{E}[u_j^4 \mid Z = z] < \infty$, both $\Psi(\tau, Z_{2j})$ and $\Phi(\tau, Z_{2j})$ are functions comprising deterministic coefficients. Because $\mathbb{E}(C(j)^4)$ is uniformly bounded by Lemma 3 in Abadie and Imbens (2006), by Markov's inequality, the last term is bounded in probability. Hence, the Lindeberg-Feller condition is satisfied for almost all z. As a result,

$$\sqrt{nh} V(z)^{-1/2} I_4 \xrightarrow{d} \mathcal{N}(0, \mathrm{I}),$$

where

$$V(z) = \sum_{j=1}^{m} \sigma^2 (C(j) K_h(Z_{2j} - z))^2 \Omega(Z_{2j}) \bigotimes \begin{pmatrix} 1 & (Z_{2j} - z)/h \\ (Z_{2j} - z)/h & ((Z_{2j} - z)/h)^2 \end{pmatrix} + o(1).$$

The boundedness of $V(z)$ is proved in Appendix C.4. Then

$$\sqrt{nh} V(z)^{-1/2} D(z) \left(H(\hat{\Theta}(z) - \Theta(z)) - bias(z) \bigotimes \begin{pmatrix} \beta(z) \\ h\beta'(z) \end{pmatrix} \right) \xrightarrow{d} \mathcal{N}(0, I),$$

where $D(z) = f(z)\Omega(z) \bigotimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix}$. ■

4.3 Bias Correction and a Consistent Estimator

In this section we analyze the asymptotic properties of the bias-corrected matching estimator. From Theorem 15, the bias term in the two-sample matching estimator is given by

$$bias(z) \bigotimes \begin{pmatrix} \beta(z) \\ h\beta'(z) \end{pmatrix} = \Omega(z)^{-1} \begin{pmatrix} 0 & \Sigma_{12} \\ 0 & -\Sigma_{22} \end{pmatrix} \bigotimes \begin{pmatrix} \beta(z) \\ \mu_2 h\beta'(z) \end{pmatrix}.$$

This bias comes from the fact that the denominator of the two-sample matching estimator, $D_n(z)$, converges to its expectation $D(z)$, while the numerator $N_n(z)$ converges to

$$\left(D(z) + f(z) \begin{pmatrix} 0 & \Sigma_{12} \\ 0 & -\Sigma_{22} \end{pmatrix} \bigotimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix} \right) \begin{pmatrix} \beta(z) \\ h\beta'(z) \end{pmatrix}.$$

Therefore, we can replace the denominator $D_n(z)$ with a consistent estimator of

$$D_n(z) + f(z) \begin{pmatrix} 0 & \Sigma_{12} \\ 0 & -\Sigma_{22} \end{pmatrix} \bigotimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix},$$

and leave the numerator $N_n(z)$ unchanged to eliminate the bias in the two-sample matching estimator.

In order to establish the asymptotic properties of the bias-corrected estimator, we need to estimate Σ_{12} and Σ_{22} consistently. For Σ_{22}, we consider the difference-based variance estimator discussed in Rice (1984). In particular,

$$\hat{\Sigma}_{22} = \frac{1}{2(m-1)} \sum_{j=2}^{m} (X_{2(j)} - X_{2(j-1)})(X_{2(j)} - X_{2(j-1)})^T,$$

where $X_{2(j)}$ and $Z_{2(j)}$ are from the ordered sample $\{(X_{2(j)}, Z_{2(j)})\}_{j=1}^{m}$ based on $Z_{2(1)} \leq \cdots \leq Z_{2(m)}$.

However, the estimator of Σ_{12} is more complicated, because Σ_{12} reflects the population correlation between X_1 and X_2, while X_1 and X_2 are not available in a single sample . There is rich literature about this kind of two-sample-combination problem and solution to recover the population joint density of X_1 and X_2. These assume either that X_1 and X_2 are conditionally independent or that more exclusive variables are needed, which are excluded from the regression of Y on X_1 and X_2, but highly correlated to both X_1 and X_2. As a result, here we assume $\Sigma_{12} = 0$, and we can also consistently estimate the density function $f(z)$ by any nonparametric method.

Define bias-corrected two-sample matching estimator as

$$H\hat{\Theta}^{bcll}(z) = \left(D_n(z) + \hat{f}(z) \begin{pmatrix} 0 & 0 \\ 0 & -\hat{\Sigma}_{22} \end{pmatrix} \bigotimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix} \right)^{-1} N_n(z).$$

Then we can estimate $\beta(z)$ consistently using

$$\hat{\beta}^{bcll}(z) = eH^{-1}\left(D_n(z) + \hat{f}(z)\begin{pmatrix} 0 & 0 \\ 0 & -\hat{\Sigma}_{22} \end{pmatrix} \bigotimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix}\right)^{-1} N_n(z), \tag{19}$$

where e is a $1 \times 2(d_1 + d_2)$ matrix with the first $(d_1 + d_2)$ elements being 1 and the remaining diagonal elements 0.

Theorem 17 ***Asymptotic Normality for the Bias-corrected Matching Estimator***. *Suppose that Assumptions 1–4 hold, and assume $\Sigma_{12} = 0$, then*

$$\sqrt{nh}V(z)^{-1/2}D(z)\left(H(\hat{\Theta}^{bcll}(z) - \Theta(z))\right) \xrightarrow{d} \mathcal{N}(0, I),$$

where $D(z) = f(z)\Omega(z) \bigotimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix}$ *and*

$$V(z) = \sum_{j=1}^{m} \sigma^2(C(j)K_h(Z_{2j} - z))^2\Omega(Z_{2j}) \bigotimes \begin{pmatrix} 1 & (Z_{2j} - z)/h \\ (Z_{2j} - z)/h & ((Z_{2j} - z)/h)^2 \end{pmatrix}.$$

Proof. Firstly, we will show that

$$\sqrt{n}\left(H\hat{\Theta}^{bcll}(z) - H\hat{\Theta}(z) + bias(z) \bigotimes \begin{pmatrix} \beta(z) \\ h\beta'(z) \end{pmatrix}\right) \xrightarrow{p} 0,$$

where $bias(z) = \Omega(z)^{-1} \begin{pmatrix} 0 & 0 \\ 0 & -\Sigma_{22} \end{pmatrix}$. Note

$$
\begin{aligned}
& H\hat{\Theta}^{bcll}(z) - H\hat{\Theta}(z) \\
&= \left(\left(D_n(z) + \hat{f}(z) \begin{pmatrix} 0 & 0 \\ 0 & -\hat{\Sigma}_{22} \end{pmatrix} \bigotimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix} \right)^{-1} - (D_n(z))^{-1} \right) N_n(z) \\
&= (D_n(z))^{-1} \left(\hat{f}(z) \begin{pmatrix} 0 & 0 \\ 0 & -\hat{\Sigma}_{22} \end{pmatrix} \bigotimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix} \right) \\
&\qquad \left(D_n(z) + \hat{f}(z) \begin{pmatrix} 0 & 0 \\ 0 & -\hat{\Sigma}_{22} \end{pmatrix} \bigotimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix} \right)^{-1} N_n(z) \\
&= (D_n(z))^{-1} \left(\hat{f}(z) \begin{pmatrix} 0 & 0 \\ 0 & -\hat{\Sigma}_{22} \end{pmatrix} \bigotimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix} \right) H\hat{\Theta}^{bcll}(z).
\end{aligned}
$$

The consistency of $H\hat{\Theta}^{bcll}(z)$ can be established in line with the proof of Theorem 15. Therefore we have $H\hat{\Theta}^{bcll}(z) \xrightarrow{p} H\Theta$. As a result,

$$
H\hat{\Theta}^{bcll}(z) - H\hat{\Theta}(z) \xrightarrow{p} \Omega(z)^{-1} \begin{pmatrix} 0 & 0 \\ 0 & -\Sigma_{22} \end{pmatrix} \bigotimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix} H\Theta.
$$

Then the asymptotic normality results will be applied by Theorem 16. ■ The result of Theorem 17 suggests that the bias-corrected matching estimator has the same asymptotic variance as the naive local linear matching estimator does.

4.4 Monte Carlo Simulations

To evaluate the naive local linear and the bias-corrected local linear estimator, we consider the following data generating process (DGP):

$$
Y = \beta_0(Z)X_{10} + \beta_1(Z)X_{11} + \beta_2(Z)X_2 + U, \tag{20}
$$

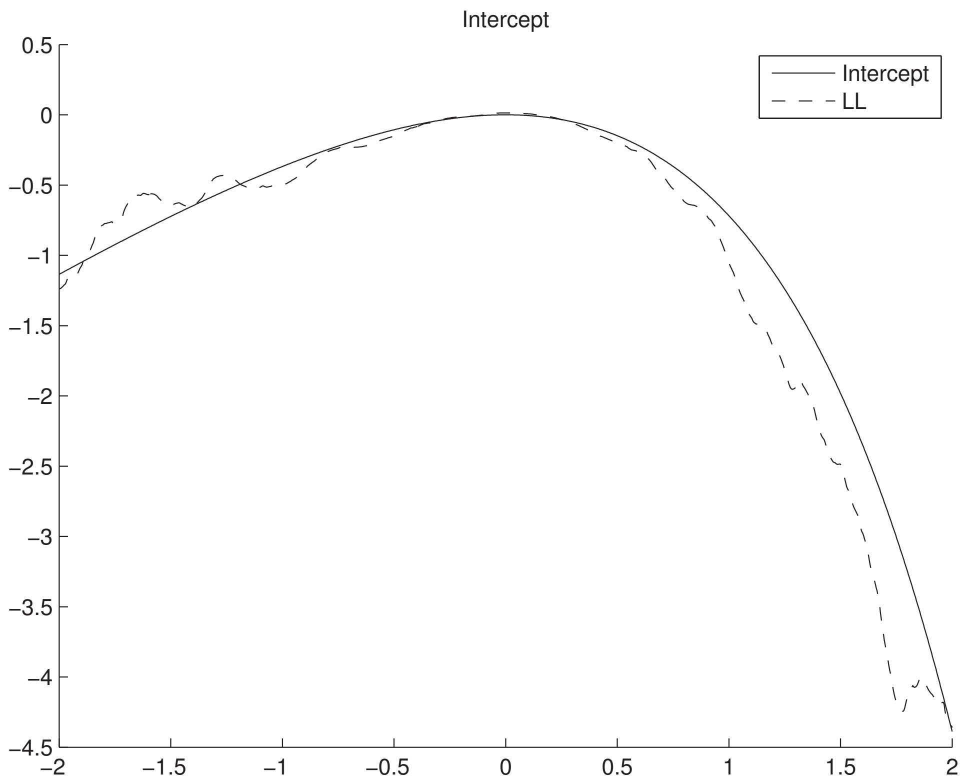


Figure 4.1: Local linear estimator in one complete sample case: $\beta_0(Z)$

where $\beta_0(Z) = (1 - e^Z + Z)$ and $\beta_1(Z) = (0.5 + 0.5Z)$. We consider two functional forms of $\beta_2(\cdot)$. We assume $\beta_2^{(1)}(z) = 0.5 + 0.5Z + 0.25Z^3$ for DGP (1) and $\beta_2^{(2)}(z) = 1 + e^z$ for DGP (2). We set $Z \sim N(0, 1)$ truncated at ± 2, $X_{10} = 2Z + \xi_1$, $X_{11} = 0.5X_{10} + \xi_2$, $X_2 = 3Z + \xi_3$, $(\xi_1, \xi_2, , \xi_3)' \sim N(0, \mathbf{I}_3)$, and $U \perp (\xi_1, \xi_2, \xi_3)'$.

The optimal bandwidth h of all the estimators are determined by the cross validation (CV) method throughout this section.

4.4.1 The case with One Complete Sample

In this section we show that the naive local linear estimator performs well when we have one complete sample generated from DPG (1). The sample size $n = 3000$. Figure 4.1 - Figure 4.3 are based on 500 replications. In fact, the performance of the estimator is based on one estimation, including the order of bias and variance, is similar with that based on 500 replications.

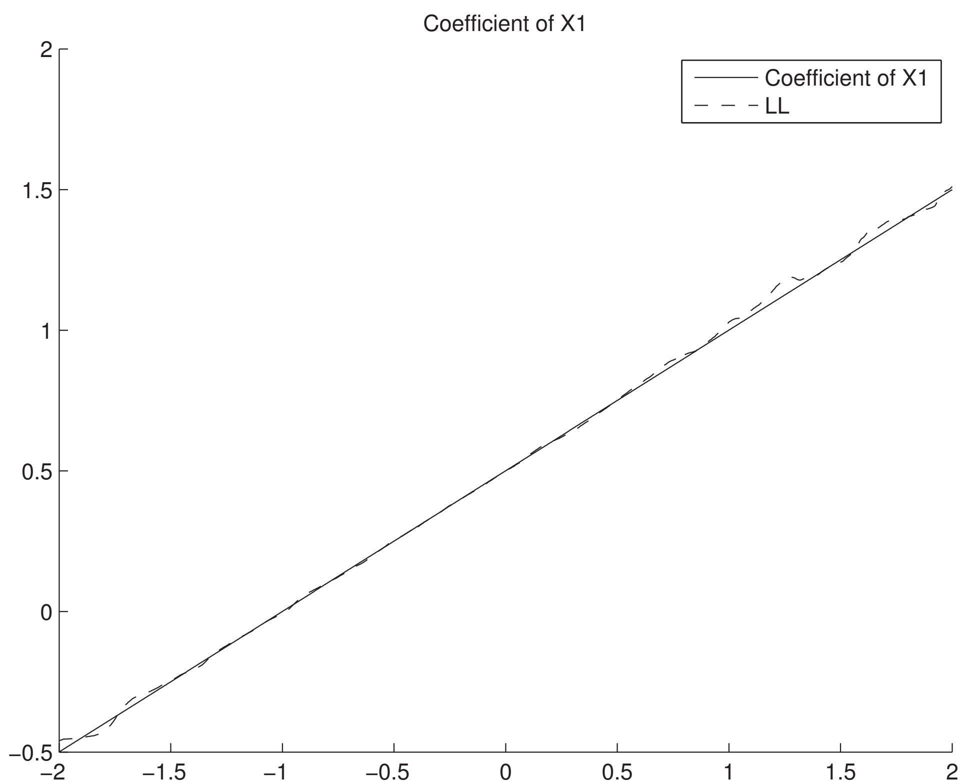


Figure 4.2: Local linear estimator in one complete sample case: $\beta_1(Z)$

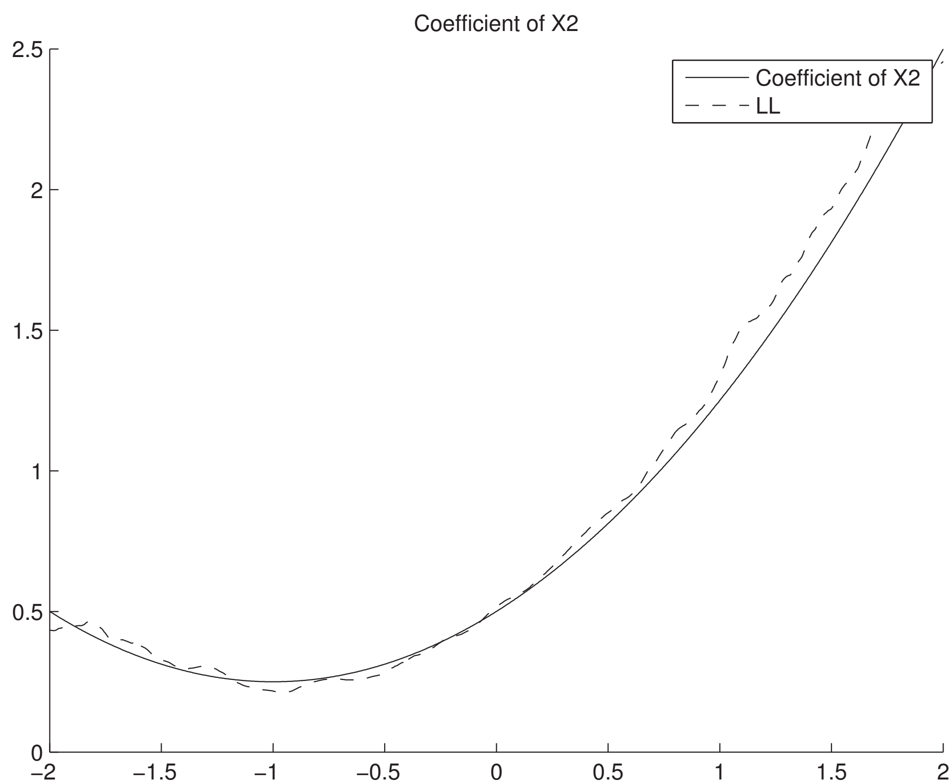


Figure 4.3: Local linear estimator in one complete sample case: $\beta_2(Z)$

4.4.2 Case with Two Missing-Data Samples

In this section we first study the behaviour of the naive local linear estimator (LL) and the bias-corrected local linear estimator (BCLL) when we have two i.i.d samples with missing data generated from DPG (1), $\{Y_i, X_{1i}, Z_{1i}\}_{i=1}^{n}$, $\{X_{2j}, Z_{2j}\}_{j=1}^{m}$, where $n = 3000$, $m = 4000$. Figure 4.4 - Figure 4.6 are based on 500 replications. In fact, the performance of the LL and BCLL estimators in one estimation, including the order of bias and variance, are similar to the performance based on 500 replications, respectively. Figure 4.4, 4.5 and 4.6 show the performance of the LL and BCLL estimators estimating of the intercept, the coefficient function for X_1 and the coefficient function for X_2, respectively. In these figures, the solid lines represent the true function parameters, while the dashed lines and the dotted lines represent the LL and BCLL respectively. It can be easily seen from Figure 4.4 that, the BCLL estimator has less average bias than the LL estimator in most parts of the support of Z despite the boundary effect. In addition, BCLL identifies the right shape of the intercept function while LL does not. Figure 4.6 shows similar properties of two estimators as in Figure 4.4. For Figure 4.5, the two estimators both identify the true shape of the coefficient function while BCLL has less average bias.

To evaluate the finite sample performance of the LL and BCLL estimator of the functional coefficient, we calculate both the mean absolute deviation (MAD) and mean squared error (MSE) for each estimate evaluated at 100 evenly-spaced points between the support of Z, which is $[-2, 2]$. In this case we consider both DGP (1) and DGP(2) with sample size $n = 300$ and $m = 400$.

Table 4.1 reports the results where the MSEs and MADs are averages over 500 replications for each functional coefficient. As expected, the bias-corrected local linear (BCLL) estimators perform better than the local linear (LL) estimators both in DGP(1) and DGP(2), which are specified differently only for the coefficient of X_2, $\beta_2(Z)$, in Eq.(20). However, all the estimators of the coefficients $\beta_0(Z)$, $\beta_1(Z)$ and $\beta_2(Z)$ are very sensitive to this change in the DGP. This reflects the fact that even with the additional assumption that $\Sigma_{12} = 0$, the BCLL estimation does not just correct the coefficient β_2; all the estimators of coefficients are affected because of the inverse of the additive bias-corrected term in the denominator of the BCLL estimator.

Table 4.1: Finite sample comparison of the local linear (LL) estimator and the bias-corrected local linear (BCLL) estimator

DGP	Estimators	$\beta_1(Z)$		$\beta_2(Z)$		$\beta_3(Z)$	
		MSE	MAD	MSE	MAD	MSE	MAD
$DGP(1)$	LL	1.1625	1.3886	0.7254	0.6272	0.8537	0.6072
	BCLL	0.2233	0.3040	0.2784	0.2297	0.3082	0.2237
$DGP(2)$	LL	1.3908	1.0438	1.4761	1.1572	1.3490	0.9765
	BCLL	0.8634	0.7173	0.1902	0.3199	0.3856	0.5050

Notes: samples are generated from DGP (1) and DGP (2) respectively. The pair of sample sizes are $(n = 300, m = 400)$, which is the same for both DGP (1) and DPG (2). MSEs and MADs are averages over 500 replications.

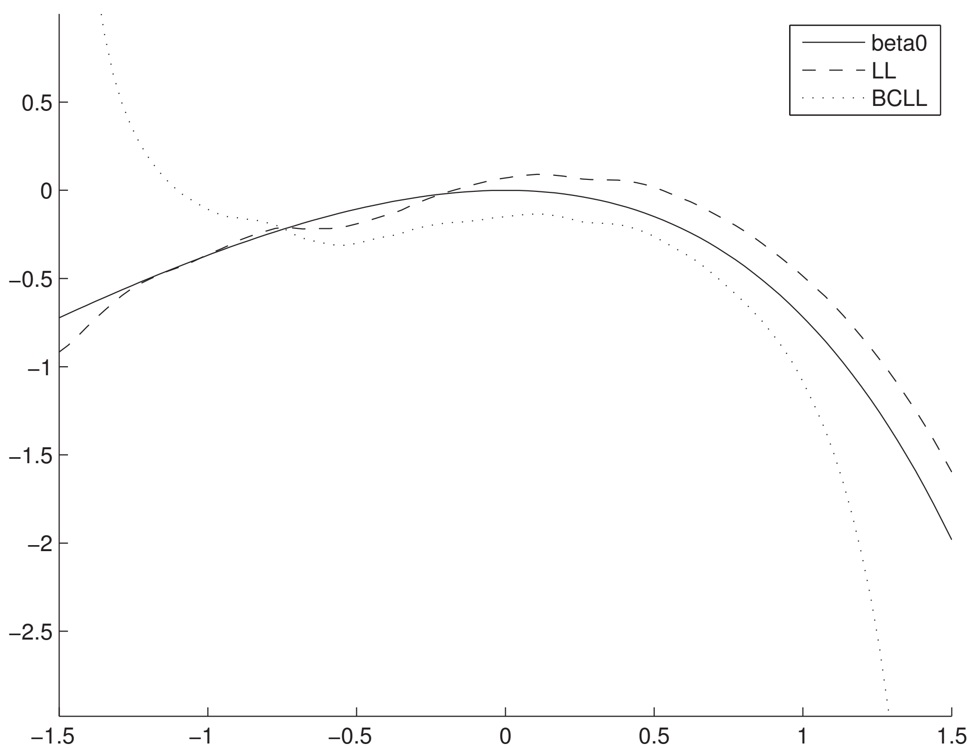


Figure 4.4: Estimators in the two-sample case: $\beta_0(Z)$

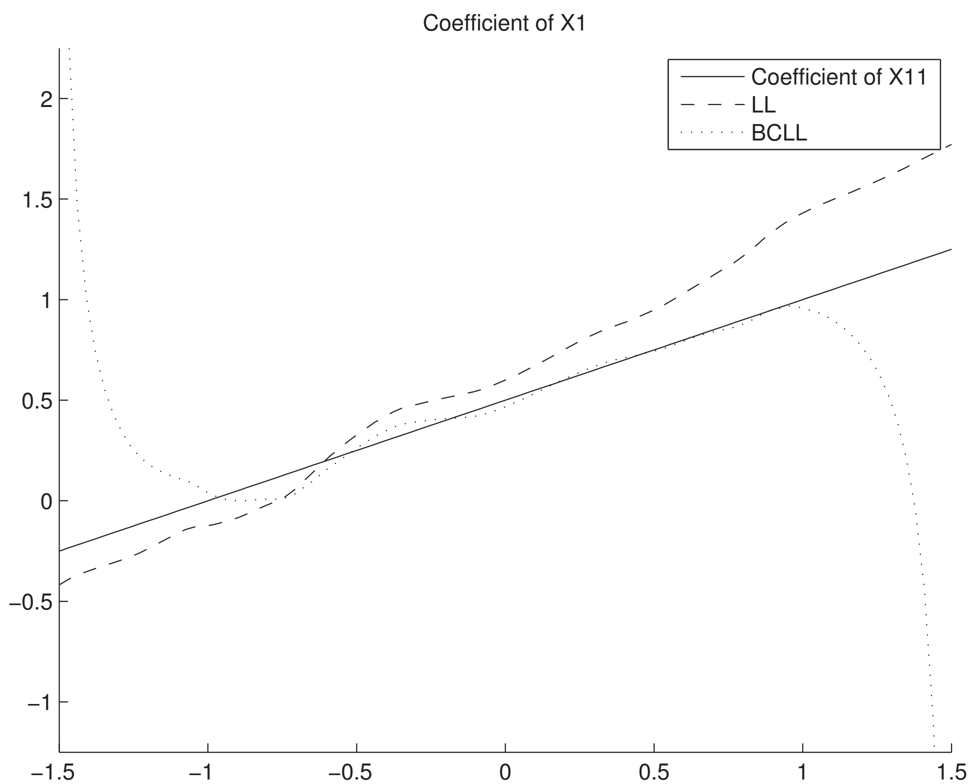


Figure 4.5: Estimators in the two-sample case: $\beta_1(Z)$

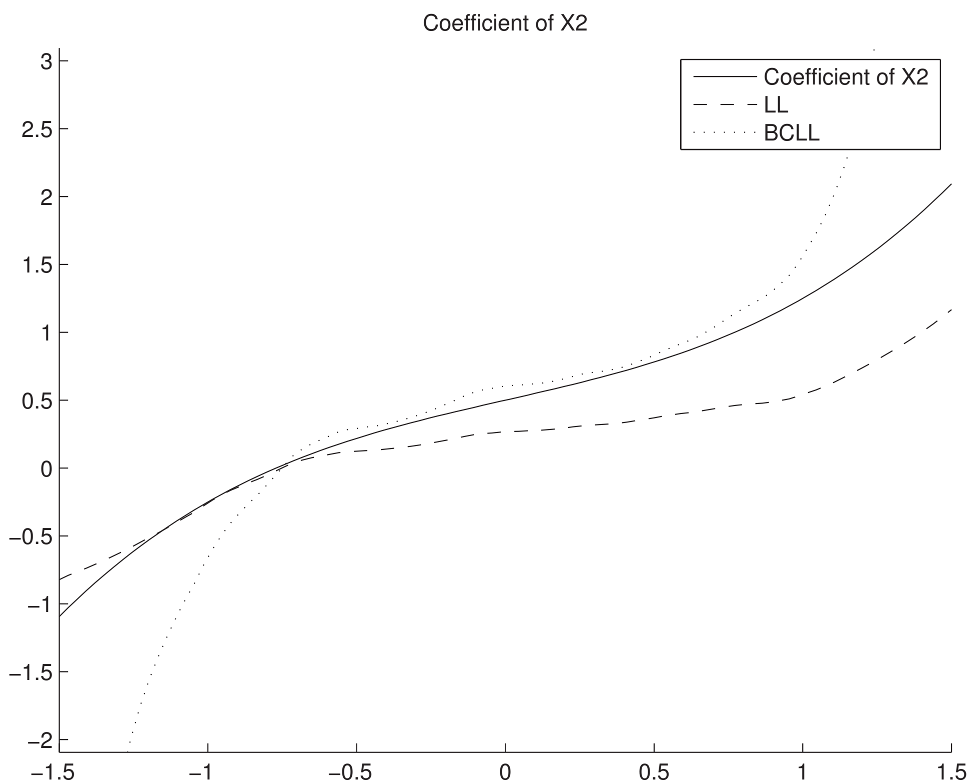


Figure 4.6: Estimators in the two-sample case: coefficient of X_2, $\beta_2(Z)$

4.5 Conclusion

In this chapter, we show the inconsistency of the simple local linear estimation of the two-sample varying coefficient model with missing data, using matched samples. The bias-corrected estimator is proposed and it is proven to be consistent and asymptotically normally distributed. According to the simulation study, it shows better performance in terms of mean squared error than the non-corrected version in finite samples.

Chapter 5

Conclusions

In Chapter 2, we consider a battery of tests resulting from eigenspectrum-based versions of the information matrix equality applied to copulas. The benefit of this generalization is due to a reduction in degrees of freedom of the tests and to the focused hypothesis function used to construct them. For example, in testing the goodness of fit of high-dimensional multi-parameter copulas we manage to reduce the information matrix based test statistic to an asymptotically χ^2 with one degree of freedom. Moreover, we can focus on the effect of larger or smaller eigenvalues by using specific functions of the eigenspectrum such as det or $trace$. However, only a few of the proposed tests can be well approximated by their asymptotic distributions in realistic sample sizes, so we have also looked at the boostrap version of the tests. The main argument of this chapter is that the bootstrap versions of GIMTs dominate other available tests of copula goodness of fit when copulas are high-dimensional and multi-parameter. We use this argument to motivate the use of GIMTs on vine copulas, where additional simplifications result from the functional form of the Hessian and the score.

In Chapter 3, we propose a family of spline estimators which guarantee the uniform marginal property for copula density. The family of spline estimators is strictly positive in the interior and behaves well in terms of capturing the behaviour while moving to the tails compared to the leading competitors in the literature. The estimation procedure is a convex maximization problem with linear constraints, which is numerically easy to implement and has less or similar computational burdens in terms of time cost compared to other sieve estimators. Our Monte Carlo simulations demonstrate the efficiency of the proposed estimators. We apply the proposed method to estimate the copula densities between children's and parents' BMI. The proposed estimators show similar results as the penalized exponential series estimators and Bernstein estimator that the dependence relationship is generally asymmetric and stronger when BMI is high. We also apply the proposed method to examine the Gibson's paradox. The family of spline estimators is strictly positive in the interior and behaves well in terms of capturing the behaviour while moving to the tails compared to the leading competitors in the literature.

In Chapter 4, we show the inconsistency of simple local linear estimation of the two-sample varying coefficient model with missing data, using matched samples. The bias-corrected estimator is proposed and it proves to be consistent and asymptotically normally distributed. The simulation study shows that in finite samples it has better performance in terms of mean squared error than the non-corrected version.

Bibliography

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

Abadie, A. and G. Imbens (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica 74*, 235–267.

Abadie, A. and G. Imbens (2011). Bias-corrected matching estimators for average treatment effects. *Journal of Business & Economic Statistics 29*, 1–11.

Angrist, J. D. and A. B. Krueger (1995). Split-sample instrumental variables estimates of the return to schooling. *Journal of Business & Economic Statistics 13*(2), 225–235.

Arellano, M. and C. Meghir (1992). Female labour supply and on-the-job search: An empirical model estimated using complementary data sets. *The Review of Economic Studies 59*, 537–559.

Bedford, T. and R. M. Cooke (2001, August). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence 32*(1), 245–268.

Bedford, T. and R. M. Cooke (2002, August). Vines: A new graphical model for dependent random variables. *The Annals of Statistics 30*(4), 1031–1068.

Berg, D. (2009, December). Copula goodness-of-fit testing: An overview and power comparison. *The European Journal of Finance 15*(7-8), 675–701.

Bhatti, M. and P. Brachen (2006). The calculation of integrals involving B-splines by means of recursion relations. *Applied Mathematics and Computation 172*, 91–100.

Bouezmarni, T., A. El Ghouch, and A. Taamouti (2013). Bernstein estimator for unbounded copula densities. *Statistics and Risk Modeling 30*(4), 343–360.

Bouezmarni, T., J. V. Rombouts, and A. Taamouti (2010). Asymptotic properties of the Bernstein density copula estimator for α-mixing data. *Journal of Multivariate Analysis 101*(1), 1–10.

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

Brechmann, E. C. and U. Schepsmeier (2013, January). Modeling dependence with C- and D-Vine copulas: The R package CDVine. *Journal of Statistical Software 52*(3), 1–27.

Chen, X. (2007). Chapter 76 large sample sieve estimation of semi-nonparametric models. Volume 6, Part B of *Handbook of Econometrics*, pp. 5549–5632. Elsevier.

Chen, X. and Y. Fan (2006a). Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *Journal of Econometrics 135*(1–2), 125–154.

Chen, X. and Y. Fan (2006b). Estimation of copula-based semiparametric time series models. *Journal of Econometrics 130*(2), 307–335.

Crafts, N. and T. C. Mills (1994). Trends in real wages in Britain, 1750 - 1913. *Explorations in Economic History 31*, 176–194.

Czado, C. (2010). Pair-copula constructions of multivariate copulas. In *Copula Theory and Its Applications. Lecture Notes in Statistics*, Volume 198, Berlin, Heidelberg, pp. 93–109. Springer Berlin Heidelberg.

Dowd, K. (2008, June). Copulas in macroeconomics. *Journal of International and global Economic Studies 1*(1), 1–26.

Fermanian, J.-D. (2005, July). Goodness-of-fit tests for copulas. *Journal of Multivariate Analysis 95*, 119–152.

Gao, Y., Y. Y. Zhang, and X. Wu (2015, February). Penalized exponential series estimation of copula densities with an application to intergenerational dependence of body mass index. *Empirical Economics 48*(1), 61–81.

Genest, C. and A. Favre (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering 12*, 347–368.

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

Genest, C., J.-F. Quessy, and B. Remillard (2006). Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scandinavian Journal of Statistics 33*, 337–366.

Genest, C. and B. Rémillard (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Annales de l'Institut Henri Poincare - Probabilités et Statistiques 44*(6), 1096–1127.

Genest, C., B. Rémillard, and D. Beaudoin (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics 44*, 199–213.

Genest, C. and L.-P. Rivest (1993). Statistical inference procedures for bivariate archimedean copulas. *Journal of the American Statistical Association 88*, 1034–1043.

Gijbels, I. and J. Mielniczuk (1990). Estimating the density of a copula function. *Communications in Statistics - Theory and Methods 19*(2), 445–464.

Golden, R., S. Henley, H. White, and T. M. Kashner (2013). New directions in information matrix testing: Eigenspectrum tests. In X. Chen and N. R. Swanson (Eds.), *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis*, pp. 145–177. Springer New York.

Gronneberg, S. and N. L. Hjort (2014, September). The copula information criteria. *Scandinavian Journal of Statistics 41*(2), 436– 459.

Hirukawa, M. and A. Prokhorov (2016, October). Consistent estimation of linear regression models using matched data. *Revised and resubmitted to Journal of Econometrics*.

Hofert, M., M. Machler, and A. J. McNeil (2012, September). Likelihood inference for archimedean copulas in high dimensions under known margins. *Journal of Multivariate Analysis 110*, 133–150.

Homer, S. (1963). *A History of Interest Rates*. New Brunswick, NJ: Rutgers University Press.

Hu, H.-L. (1998). Large sample theory of pseudo-maximum likelihood estimates in semiparametric models. *Ph.D. dissertation, University of Washington*.

Huang, W. and A. Prokhorov (2014, March). A goodness-of-fit test for copulas. *Econometric Reviews 98*, 533–543.

Joe, H. (1996). Families of m-variate distributions with given margins and m(m-1)/2 bivariate dependence parameters. *Lecture Notes-Monograph Series 28*, 120–141.

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

Junker, M. and A. May (2005, November). Measurement of aggregate risk with copulas. *The Econometrics Journal 8*, 428–454.

Klovland, J. T. (1994). Pitfalls in the estimation of the yield on British consols, 1850-1914. *Journal of Economic History 54*, 164–187.

Klugman, S. and R. Parsa (1999, March). Fitting bivariate loss distributions with copulas. *Insurance: Mathematics and Economics 24*(1), 139–148.

Kojadinovic, I. and J. Yan (2011). A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems. *Statistics and Computing 21*(1), 17–30.

Kollo, T. and D. von Rosen (2006). *Advanced Multivariate Statistics with Matrices. Mathematics and Its Applications*. Springer.

Kurowicka, D. and R. M. Cooke (2006). *Uncertainty Analysis with High Dimensional Dependence Modelling*. John Wiley & Sons Ltd, Chichester.

Magnus, J. (1985). On differentiating eigenvalues and eigenvectors. *Econometric Theory 1*, 179–191.

Magnus, J. and H. Neudecker (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons Ltd, Chichester.

Murtazashvili, I., D. Liu, and A. Prokhorov (2015). Two-sample nonparametric estimation of intergenerational income mobility in the United States and Sweden. *Canadian Journal of Economics 48*(5), 1733–1761.

Omelka, M., I. Gijbels, and N. Veraverbeke (2009, October). Improved kernel estimation of copulas: Weak convergence and goodness-of-fit testing. *The Annals of Statistics 37*(5B), 3023–3058.

Patton, A. J. (2012, September). A review of copula models for economic time series. *Journal of Multivariate Analysis 110*, 4–18.

Presnell, B. and D. D. Boos (2004). The IOS test for model misspecification. *Journal of the American Statistical Association 99*(465), 216–227.

Prokhorov, A. and P. Schmidt (2009, November). Likelihood-based estimation in a panel setting: Robustness, redundancy and validity of copulas. *Journal of Econometrics 153*(1), 93–104.

Qu, L. and W. Yin (2012). Copula density estimation by total variation penalized likelihood with linear equality constraints. *Computational Statistics and Data Analysis 56*(2), 384–398.

R Development Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Rice, J. (1984). Bandwidth choice for nonparametric regression. *Annals of Statistics 12*(4), 1215–1230.

Ridder, G. and R. Moffitt (2007). The econometrics of data combination. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics* (1 ed.), Volume 6B, Chapter 75, pp. 5469–5547. Elsevier.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics 23*(3), 470–472.

Sancetta, A. and S. Satchell (2004). The Bernstein copula and its applications to modelling and approximations of multivariate distributions. *Econometric Theory 20*(03), 535–562.

Scaillet, O. (2007, March). Kernel based goodness-of-fit tests for copulas with fixed smoothing parameters. *Journal of Multivariate Analysis 98*, 533–543.

Schepsmeier, U. (2015). Efficient goodness-of-fit tests in multi-dimensional vine copula models. *Journal of Multivariate Analysis 138*, 35–52.

Schepsmeier, U. (2016). A goodness-of-fit test for regular vine copula models. *Econometric Reviews*, 1–22.

Schepsmeier, U. and J. Stöber (2012). Web supplement: Derivatives and Fisher information of bivariate copulas. Technical report, TU München.

Schepsmeier, U. and J. Stöber (2014). Derivatives and fisher information of bivariate copulas. *Statistical Papers 55*(2), 525–542.

Schepsmeier, U., J. Stoeber, and E. C. Brechmann (2013). *VineCopula: statistical inference of vine copulas*. R package version 1.2.

Schuster, E. (1985). Incorporating support constraints into nonparametric estimators of densities. *Communications in Statistics - Theory and Methods 14*(5), 1123–1136.

Segers, J., M. Sibuya, and H. Tsukahara (2017). The empirical beta copula. *Journal of Multivariate Analysis 155*, 35–51.

Shih, J. H. and T. A. Louis (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics 51*(4), 1384–1399.

Stöber, J. and U. Schepsmeier (2013). Estimating standard errors in regular vine copula models. *Computational Statistics 28*(6), 2679–2707.

Takeuchi, K. (1976). Distribution of information statistics and a criterion of model fitting for adequacy of models. *Mathematical Sciences 153*, 12–18.

Tsukahara, H. (2005). Semiparametric estimation in copula models. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique 33*(3), 357–375.

White, H. (1982, January). Maximum likelihood estimation of misspecified models. *Econometrica 50*(1), 1–25.

Zhou, Q. M., P. X.-K. Song, and M. E. Thompson (2012). Information ratio test for model misspecification in quasi-likelihood inference. *Journal of the American Statistical Association 107*(497), 205–213.

Zimmer, D. (2012). The role of copulas in the housing crisis. *Review of Economics and Statistics 94*(2), 607–620.

Appendix A

Proofs in Chapter 2

Proof of Lemma 1: The proof is based on combining the results of Golden et al. (2013) and Huang and Prokhorov (2014). It also relates to the work of Presnell and Boos (2004) on information ratio test. We start with $d = 2$ for simplicity and later give the formulas for any d. We will need some additional notation.

Recall

$$\boldsymbol{d}_i(\boldsymbol{\theta}) := \begin{pmatrix} vech(\mathbb{H}_i(\boldsymbol{\theta})) \\ vech(\mathbb{C}_i(\boldsymbol{\theta})) \end{pmatrix} \in \mathbb{R}^{p(p+1)}. \tag{21}$$

Under the assumption that the derivatives and expectation exist, let $D_{\boldsymbol{\theta}} := \mathbb{E}\nabla_{\boldsymbol{\theta}}\boldsymbol{d}_i(\boldsymbol{\theta}) \in \mathbb{R}^{p(p+1)\times p}$ denote the expected Jacobian matrix of the random vector $\boldsymbol{d}_i(\boldsymbol{\theta})$. Note that we can estimate $\mathbb{E}\boldsymbol{d}_i(\boldsymbol{\theta}_0)$ by $\bar{\boldsymbol{d}}(\hat{\boldsymbol{\theta}})$.

Let $F_{ij} := F_j(X_{ij})$, $j = 1, 2$, $i = 1, \dots, n$, denote the marginal cdf of X_j evaluated at point X_{ij} and let $\hat{F}_{ij} := \hat{F}_j(X_{ij})$, $j = 1, 2$, $i = 1, \dots, n$, denote the empirical cdf of X_j evaluated at point X_{ij}. Then, Eq.(21) can be written as follows:

$$\boldsymbol{d}_i(\boldsymbol{\theta}) = \left\{ vech[\nabla^2_\theta \ln c(\hat{F}_{i1}, \hat{F}_{i2}; \boldsymbol{\theta})]', vech[\nabla_\theta \ln c(\hat{F}_{i1}, \hat{F}_{i2}; \boldsymbol{\theta})\nabla'_\theta \ln c(\hat{F}_{i1}, \hat{F}_{i2}; \boldsymbol{\theta})]' \right\}'.$$

Define the sample equivalent of $D_{\boldsymbol{\theta}}$ as follows

$$\bar{D}_{\boldsymbol{\theta}} = n^{-1} \sum_{i=1}^{n} \nabla_\theta \boldsymbol{d}_i(\boldsymbol{\theta}).$$

The asymptotic normality proof for $\sqrt{n}\bar{d}(\hat{\boldsymbol{\theta}})$ is provided by White (1982) for generic multivariate distributions and can be easily transfered to the case of copulas with known margins. To extend the proof to empirical margins we first expand $\sqrt{n}\bar{\boldsymbol{d}}(\hat{\boldsymbol{\theta}})$ with respect to $\boldsymbol{\theta}$ at $\boldsymbol{\theta}_0$:

$$\sqrt{n}\bar{\boldsymbol{d}}(\hat{\boldsymbol{\theta}}) = \sqrt{n}\bar{\boldsymbol{d}}(\boldsymbol{\theta}_0) + D_{\boldsymbol{\theta}_o}\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + o_p(1). \tag{22}$$

The remainder term in this expansion is controlled by assumptions on continuity of copula derivatives such as the conditions used in Theorem 1 of Tsukahara (2005) or Proposition 2.1 of Genest et al. (1995). We do not list these conditions explicitly for space considerations.

Chen and Fan (2006a) show that the second term in the right-hand side of Eq.(22) is normally distributed, i.e.,

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \to N(0, \mathbb{H}_0^{-1} G \mathbb{H}_0^{-1}),$$

where

$$\begin{aligned} G &= \lim_{n\to\infty} Var(\sqrt{n}A_n^*), \\ A_n^* &= \frac{1}{n}\sum_{i=1}^{n}(\nabla_\theta \ln c(F_{i1}, F_{i2}; \boldsymbol{\theta}_0) + W_1(F_{i1}) + W_2(F_{i2})). \end{aligned}$$

Here the terms $W_1(F_{i1})$ and $W_2(F_{i2})$ are the adjustments needed to account for the empirical distributions used in place of the true distributions. These terms are calculated as follows:

$$\begin{aligned} W_1(F_{i1}) &= \int_0^1\int_0^1 [\mathbf{1}\{F_{i1} \le u\} - u]\nabla^2_{\theta,u} \ln c(u, v; \boldsymbol{\theta}_0)\, c(u, v; \boldsymbol{\theta}_0)\mathrm{d}u\mathrm{d}v, \\ W_2(F_{i2}) &= \int_0^1\int_0^1 [\mathbf{1}\{F_{i2} \le v\} - v]\nabla^2_{\theta,v} \ln c(u, v; \boldsymbol{\theta}_0)\, c(u, v; \boldsymbol{\theta}_0)\mathrm{d}u\mathrm{d}v. \end{aligned}$$

So, rewriting the consistency result from Chen and Fan (2006a) we have

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -\mathbb{H}_0^{-1}\sqrt{n}A_n^* + o_p(1).$$

The explicit conditions for this result to hold are Conditions A1 through A6 of Chen and Fan (2006a, p. 319).

Second, let $\nabla_j \boldsymbol{d}_i(\boldsymbol{\theta}_0), j = 1, 2$, denote $\frac{\partial \boldsymbol{d}_i(\boldsymbol{\theta}_0)}{\partial U_{ij}}|_{U_{ij}=F_{ij}}$ and expand $\sqrt{n}\bar{\boldsymbol{d}}(\boldsymbol{\theta}_0)$ with respect to U_1 and U_2 around the point (F_{i1}, F_{i2}):

$$\begin{aligned}\sqrt{n}\bar{\boldsymbol{d}}(\boldsymbol{\theta}_0) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \boldsymbol{d}_i(\boldsymbol{\theta}_0)|_{U_{ij}=F_{ij}} + \frac{1}{n}\sum_{i=1}^{n} \nabla_1 \boldsymbol{d}_i(\boldsymbol{\theta}_0)\sqrt{n}(\hat{F}_{i1} - F_{i1}) \\ &+ \frac{1}{n}\sum_{i=1}^{n} \nabla_2 \boldsymbol{d}_i(\boldsymbol{\theta}_0)\sqrt{n}(\hat{F}_{i2} - F_{i2}) + o_p(1). \qquad (23)\end{aligned}$$

In order to control the behavior of the remainder term in the expansion it is standard to use assumptions on existence and boundedness of copula derivatives such as assumptions A5-A6 of Chen and Fan (2006a).

Now let ∇_u denote the derivative w.r.t. u and let $\nabla_{\boldsymbol{\theta}}$ denote the vertical derivative vector w.r.t. $\boldsymbol{\theta}$. Then, following Chen and Fan (2006a), we can write

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n} \nabla_1 \boldsymbol{d}_i(\boldsymbol{\theta}_0)\sqrt{n}(\hat{F}_{i1} - F_{i1}) \\ &\simeq \int_0^1 \int_0^1 \nabla_u \{vech[\nabla_\theta^2 \ln c(u, v; \boldsymbol{\theta}_0)]', vech[\nabla_\theta \ln c(u, v; \boldsymbol{\theta}_0)\nabla_\theta' \ln c(u, v; \boldsymbol{\theta}_0)]'\}' \\ &\qquad \sqrt{n}(\hat{F}_1(F_1^{-1}(u)) - u)c(u, v; \boldsymbol{\theta}_0)\mathrm{d}u\mathrm{d}v \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^1 \int_0^1 [\mathbf{1}\{F_{i1} \leq u\} - u] \\ &\qquad \nabla_u \{vech[\nabla_\theta^2 \ln c(u, v; \boldsymbol{\theta}_0)]', vech[\nabla_\theta \ln c(u, v; \boldsymbol{\theta}_0)\nabla_\theta' \ln c(u, v; \boldsymbol{\theta}_0)]'\}' c(u, v; \boldsymbol{\theta}_0)\mathrm{d}u\mathrm{d}v.\end{aligned}$$

Denote

$$\begin{aligned}M_1(F_{i1}) = &\int_0^1 \int_0^1 [\mathbf{1}\{F_{i1} \leq u\} - u] \\ &\nabla_u \{vech[\nabla_\theta^2 \ln c(u, v; \boldsymbol{\theta}_0)]', vech[\nabla_\theta \ln c(u, v; \boldsymbol{\theta}_0)\nabla_\theta' \ln c(u, v; \boldsymbol{\theta}_0)]'\}' c(u, v; \boldsymbol{\theta}_0)\mathrm{d}u\mathrm{d}v,\end{aligned}$$

then

$$\frac{1}{n}\sum_{i=1}^{n} \nabla_1 \boldsymbol{d}_i(\boldsymbol{\theta}_0)\sqrt{n}(\hat{F}_{i1} - F_{i1}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} M_1(F_{i1}).$$

Similarly, denote

$$M_2(F_{i2}) = \int_0^1 \int_0^1 [\mathbf{1}\{F_{i2} \le v\} - v]$$
$$\nabla_v \{vech[\nabla_\theta^2 \ln c(u,v;\boldsymbol{\theta}_0)]', vech[\nabla_\theta \ln c(u,v;\boldsymbol{\theta}_0)\nabla'_\theta \ln c(u,v;\boldsymbol{\theta}_0)]'\}' c(u,v;\boldsymbol{\theta}_0)\mathrm{d}u\mathrm{d}v,$$

then

$$\frac{1}{n}\sum_{i=1}^n \nabla_2 \boldsymbol{d}_i(\boldsymbol{\theta}_0)\sqrt{n}(\hat{F}_{i2} - F_{i2}) = \frac{1}{\sqrt{n}}\sum_{t=i}^n M_2(F_{i2}).$$

Therefore, Eq.(23) can be rewritten as

$$\sqrt{n}\bar{\boldsymbol{d}}(\boldsymbol{\theta}_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{d}_i(\boldsymbol{\theta}_0)|_{U_{ij}=F_{ij}} + \sqrt{n}B_n^* + o_p(1), \tag{24}$$

where

$$B_n^* = \frac{1}{n}\sum_{i=1}^n [M_1(F_{i1}) + M_2(F_{i2})].$$

Finally, combining the expansions (22) and (24) gives

$$\sqrt{n}\bar{\boldsymbol{d}}(\hat{\boldsymbol{\theta}}) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{d}_i(\boldsymbol{\theta}_0)|_{U_{ij}=F_{ji}} + \sqrt{n}B_n^* - D_{\boldsymbol{\theta}_0}\mathbb{H}_0^{-1}\sqrt{n}A_n^* + o_p(1).$$

So $\bar{\boldsymbol{d}}(\hat{\boldsymbol{\theta}})$ converges in distribution to a multivariate normal with variance matrix $V(\boldsymbol{\theta}_0)$:

$$\sqrt{n}\bar{\boldsymbol{d}}(\hat{\boldsymbol{\theta}}) \to N(0, V(\boldsymbol{\theta}_0)),$$

where

$$\begin{aligned}
V(\boldsymbol{\theta}_0) = \mathbb{E}\,\{&\boldsymbol{d}_i(\boldsymbol{\theta}_0) + M_1(F_{i1}) + M_2(F_{i2}) \\
&- D_{\boldsymbol{\theta}_0}\mathbb{H}_0^{-1}\left[\nabla_\theta \ln c(F_{i1}, F_{i2};\boldsymbol{\theta}_0) + W_1(F_{i1}) + W_2(F_{i2})\right]\big\} \\
&\times \{\boldsymbol{d}_i(\boldsymbol{\theta}_0) + M_1(F_{i1}) + M_2(F_{i2}) \\
&- D_{\boldsymbol{\theta}_0}\mathbb{H}_0^{-1}\left[\nabla_\theta \ln c(F_{i1}, F_{i2};\boldsymbol{\theta}_0) + W_1(F_{i1}) + W_2(F_{i2}))\right]\big\}'.
\end{aligned}$$

Extension to $d \geq 2$ is straightforward. Now

$$\boldsymbol{d}_i(\boldsymbol{\theta}) = \begin{pmatrix} vech(\nabla^2_\theta \ln c(\hat{F}_{i1}, \hat{F}_{i2}, \ldots, \hat{F}_{id}; \boldsymbol{\theta})) \\ vech(\nabla_\theta \ln c(\hat{F}_{i1}, \hat{F}_{i2}, \ldots, \hat{F}_{id}; \boldsymbol{\theta}) \nabla'_\theta \ln c(\hat{F}_{i1}, \hat{F}_{i2}, \ldots, \hat{F}_{id}; \boldsymbol{\theta})) \end{pmatrix}$$

and the asymptotic variance matrix becomes

$$\begin{aligned} V(\boldsymbol{\theta}_0) = \mathbb{E} &\left\{ \boldsymbol{d}_i(\theta_0) - D_{\boldsymbol{\theta}_0} \mathbb{H}_0^{-1} \left[\nabla_\theta \ln c(F_{i1}, F_{i2}, \ldots, F_{id}; \boldsymbol{\theta}_0) + \sum_{j=1}^d W_j(F_{ij}) \right] + \sum_{j=1}^d M_j(F_{ij}) \right\} \\ &\times \left\{ \boldsymbol{d}_i(\boldsymbol{\theta}_0) - \nabla D_{\boldsymbol{\theta}_0} \mathbb{H}_0^{-1} \left[\nabla_\theta \ln c(F_{i1}, F_{i2}, \ldots, F_{id}; \boldsymbol{\theta}_0) + \sum_{j=1}^d W_j(F_{ij}) \right] + \sum_{j=1}^d M_j(F_{ij}) \right\}', \end{aligned} \tag{25}$$

where, for $j = 1, 2, \ldots, d$,

$$\begin{aligned} W_j(F_{ij}) = \int_0^1 \int_0^1 \cdots \int_0^1 [\mathbf{1}\{F_{ij} \leq u_n\} - u_j] \nabla^2_{\boldsymbol{\theta}, u_j} \ln c(u_1, u_2, \ldots, u_d; \boldsymbol{\theta}_0) \\ c(u_1, u_2, \ldots, u_d; \boldsymbol{\theta}_0) \mathrm{d}u_1 \mathrm{d}u_2 \cdots \mathrm{d}u_d, \end{aligned}$$

and

$$\begin{aligned} M_j(F_{ij}) = \int_0^1 \int_0^1 \cdots \int_0^1 [\mathbf{1}\{F_{ij} \leq u_j\} - u_j] \nabla_{u_j} vech[\nabla^2_\theta \ln c(u_1, u_2, \ldots, u_d; \boldsymbol{\theta}_0) \\ + \nabla_\theta \ln c(u_1, u_2, \ldots, u_d; \boldsymbol{\theta}_0) \nabla'_\theta \ln c(u_1, u_2, \ldots, u_d; \boldsymbol{\theta}_0)] \\ c(u_1, u_2, \ldots, u_d; \boldsymbol{\theta}_0) \mathrm{d}u_1 \mathrm{d}u_2 \cdots \mathrm{d}u_d. \end{aligned}$$

Now, since $\hat{s}_n$ is a function of $\bar{\boldsymbol{d}}(\hat{\boldsymbol{\theta}})$, its asymptotic distribution follows trivially using the delta method:

$$\sqrt{n} \hat{s}_n \xrightarrow{d} N(0, \Sigma_s(\boldsymbol{\theta}_0)),$$

where

$$\Sigma_s(\boldsymbol{\theta}_0) := S(\boldsymbol{\theta}_0) V(\boldsymbol{\theta}_0) S(\boldsymbol{\theta}_0)'.$$

Lemma A1: For any real-valued square matrices A and B, let the elements of $B \in \mathbb{R}^{r\times r}$ be functions of $A \in \mathbb{R}^{p\times p}$. Let the matrix $\frac{dB}{dA} \in \mathbb{R}^{p^2\times r^2}$ be called matrix derivative of B by A if

$$\frac{dB}{dA} = \frac{\partial}{\partial vec(A)} vec(B)',$$

where vec denotes the vectorization operator. Let D denote the transition matrix, i.e. such a matrix that for, any A, $vech(A) = Dvec(A)$ and $D^{+}vech(A) = vec(A)$, where D^{+} is the Moore-Penrose inverse of D. Then, the following results hold Kollo and von Rosen (see, e.g., 2006):

$$\begin{aligned}
\frac{dA}{dA} &= \mathbf{I}_{p^2} \\
\frac{dC'A}{dA} &= \mathbf{I}_p \otimes C, \text{ where C is a matrix of proper size with constant elements} \\
\frac{d(C'B)}{dA} &= \frac{dB}{dA}(\mathbf{I} \otimes C) \\
\frac{dBC}{dA} &= \frac{dB}{dA}(C \otimes \mathbf{I}) \\
\frac{dA^{-1}}{dA} &= -A^{-1} \otimes (A')^{-1} \\
\frac{d\mathrm{tr}(B)}{dA} &= \frac{dB}{dA} vec(\mathbf{I}_r) \\
\frac{d\mathrm{tr}(C'A)}{dA} &= vec(C), \text{ where C is a matrix of proper size with constant elements} \\
\frac{d\det(A)}{dA} &= \det(A) vec(A^{-1})' \\
\frac{dA(B(C))}{dC} &= \frac{dB}{dC}\frac{dA}{dB}
\end{aligned}$$

Lemma A2: Let λ denote an eigenvalue of a symmetric matrix A and let y denote the corresponding normalized eigenvector, i.e. the solution of the equation system $Ay = \lambda y$, such that $y'y = 1$. Let D denote the duplication matrix. Then, the following result holds Magnus (see 1985):

$$\frac{\partial \lambda}{\partial vech(A)} = [y' \otimes y']D$$

Proof of Proposition 1: First use Lemma A1 on determinant differentiation, as well as properties of vec and $vech$ operators, to obtain

$$S(\boldsymbol{\theta}_0) = det(\mathbb{H}(\boldsymbol{\theta}_0) + \mathbb{C}(\boldsymbol{\theta}_0))vech((\mathbb{H}(\boldsymbol{\theta}_0) + \mathbb{C}(\boldsymbol{\theta}_0))^{-1})' \left[\mathbf{I}_{p(p+1)/2}, \mathbf{I}_{p(p+1)/2}\right]$$

Now use $\hat{\boldsymbol{\theta}}$, which is consistent for $\boldsymbol{\theta}_0$, and the sample equivalents $\bar{\mathbb{H}}_n$ and $\bar{\mathbb{C}}_n$, which are consistent for $\mathbb{H}_0$ and $\mathbb{C}_0$, to obtain the consistent estimator $\widehat{S}$ given in the proposition.

The asymptotic distribution of $\mathcal{T}_n^{(D)}$ then follows from Theorem 1.

Proof of Proposition 2: First use Lemma A1 on trace differentiation to obtain the form of $S(\boldsymbol{\theta}_0)$, then the result follows trivially from Theorem 1.

Proof of Proposition 3: First use Lemma A1 on trace and inverse differentiation as well as the fact that $[C' \otimes A]vec(B) = vec(ABC)$, to obtain

$$S(\boldsymbol{\theta}_0) = \left(vech\left(\mathbb{H}(\boldsymbol{\theta}_0)^{-1}\mathbb{C}(\boldsymbol{\theta}_0)\mathbb{H}(\boldsymbol{\theta}_0)^{-1}\right)',\ \ vech\left(-\mathbb{H}(\boldsymbol{\theta}_0)^{-1}\right)'\right)$$

then replace the population values with consistent estimates as before, and apply Theorem 1 to obtain the result.

Proof of Proposition 4: Similar to previous propositions, using Lemma A1 on determinant differentiation to obtain

$$S(\boldsymbol{\theta}_0) = \det(\mathbb{H}(\boldsymbol{\theta}_0)^{-1}\mathbb{C}(\boldsymbol{\theta}_0))\left(vech\left(-\mathbb{C}(\boldsymbol{\theta}_0)^{-1}\mathbb{H}(\boldsymbol{\theta}_0)^{-1}\mathbb{C}(\boldsymbol{\theta}_0)\right)',\ \ vech\left(\mathbb{C}(\boldsymbol{\theta}_0)^{-1}\right)'\right).$$

Proof of Proposition 5: Similar to previous propositions, using Lemma A1 on trace differentiation to obtain

$$S(\boldsymbol{\theta}_0) = \left(-\frac{1}{\mathrm{tr}(-\mathbb{H}(\boldsymbol{\theta}_0))}vec\left(\mathbf{I}_p\right)',\ \ \frac{1}{\mathrm{tr}(\mathbb{C}(\boldsymbol{\theta}_0))}vec\left(\mathbf{I}_p\right)'\right).$$

Proof of Proposition 6: Under the null, this is a log version of the IR test, so

$$S(\boldsymbol{\theta}_0) = \frac{1}{\mathrm{tr}(\mathbb{H}(\boldsymbol{\theta}_0)^{-1}\mathbb{C}(\boldsymbol{\theta}_0))}\left(vech\left(\mathbb{H}(\boldsymbol{\theta}_0)^{-1}\mathbb{C}(\boldsymbol{\theta}_0)\mathbb{H}(\boldsymbol{\theta}_0)^{-1}\right)',\ \ vech\left(-\mathbb{H}(\boldsymbol{\theta}_0)^{-1}\right)'\right)$$

The rest of the proof is the same as in previous propositions.

Proof of Proposition 7: Similar to above, using Lemma A2 to obtain

$$S(\boldsymbol{\theta}_0) = \begin{bmatrix} -\frac{1}{\lambda_1(\mathbb{H}(\boldsymbol{\theta}_0))}[y_1(\mathbb{H}(\boldsymbol{\theta}_0))' \otimes y_1(\mathbb{H}(\boldsymbol{\theta}_0))']D & \frac{1}{\lambda_1(\mathbb{C}(\boldsymbol{\theta}_0)))}[y_1(\mathbb{C}(\boldsymbol{\theta}_0))' \otimes y_1(\mathbb{C}(\boldsymbol{\theta}_0))']D \\ \vdots & \vdots \\ -\frac{1}{\lambda_p(\mathbb{H}(\boldsymbol{\theta}_0))}[y_p(\mathbb{H}(\boldsymbol{\theta}_0))' \otimes y_p(\mathbb{H}(\boldsymbol{\theta}_0))']D & \frac{1}{\lambda_p(\mathbb{C}(\boldsymbol{\theta}_0)))}[y_p(\mathbb{C}(\boldsymbol{\theta}_0))' \otimes y_p(\mathbb{C}(\boldsymbol{\theta}_0))']D \end{bmatrix}.$$

Proof of Proposition 8: Similar to above, using Lemma A2 to obtain

$$S(\boldsymbol{\theta}_0) = \begin{bmatrix} \frac{1}{\lambda_1(\mathbb{H}(\boldsymbol{\theta}_0))}[y_1(\mathbb{C}(\boldsymbol{\theta}_0))' \otimes y_1(\mathbb{C}(\boldsymbol{\theta}_0))']D & -\frac{\lambda_1(\mathbb{C}(\boldsymbol{\theta}_0))}{\lambda_1(\mathbb{H}(\boldsymbol{\theta}_0))^2}[y_1(\mathbb{H}(\boldsymbol{\theta}_0))' \otimes y_1(\mathbb{H}(\boldsymbol{\theta}_0))']D \\ \vdots & \vdots \\ \frac{1}{\lambda_p(\mathbb{H}(\boldsymbol{\theta}_0))}[y_p(\mathbb{C}(\boldsymbol{\theta}_0))' \otimes y_p(\mathbb{C}(\boldsymbol{\theta}_0))']D & -\frac{\lambda_p(\mathbb{C}(\boldsymbol{\theta}_0))}{\lambda_p(\mathbb{H}(\boldsymbol{\theta}_0))^2}[y_p(\mathbb{H}(\boldsymbol{\theta}_0))' \otimes y_p(\mathbb{H}(\boldsymbol{\theta}_0))']D \end{bmatrix}.$$

A.1 Vines Used in Simulations

In Section 2.4.1 we used the following vine copula for our simulation study. Table A.1 for $d = 5$ and Table A.2 for $d = 8$ give details about the vine copula decomposition (structure) $\mathcal{V}$, their selected pair-copula families $\mathcal{B}$ and Kendall's τ for the vine copula under the null hypothesis. For the C-vine and D-vine, $\mathcal{V}$ as well as $\mathcal{B}$ are selected by the algorithms provided in the VineCopula package (Schepsmeier et al., 2013). $\hat{\tau}$ denotes the estimated Kendall's τ in the pre-run step of the simulation procedure of Schepsmeier (2016). Note that the vine copula density is written in a short hand notation omitting the pair-copula arguments. The notation of the pair-copula families follows Brechmann and Schepsmeier (2013).

For the C- and D-vine the calculation of the vine copula density (4) simplifies. For the five-dimensional example used in the simulation study, (4) can be expressed as

$$c_{12345} = c_{1,2} \cdot c_{2,3} \cdot c_{2,4} \cdot c_{2,5} \cdot c_{1,3;2} \cdot c_{1,4;2} \cdot c_{1,5;2} \cdot c_{3,4;1,2} \cdot c_{4,5;1,2} \cdot c_{3,5;1,2,4}$$

$$c_{12345} = c_{1,2} \cdot c_{1,5} \cdot c_{4,5} \cdot c_{3,4} \cdot c_{2,5;1} \cdot c_{1,4;5} \cdot c_{3,5;4} \cdot c_{2,4;1,5} \cdot c_{1,3;4,5} \cdot c_{2,3;1,4,5}$$

Similar representations used for $d = 8$ and 16 as well as a similar table for $d = 16$ are available from the authors upon request.

	R-vine			C-vine			D-vine		
T	$\mathcal{V}_R^5$	$\mathcal{B}_R^5(\mathcal{V}_R^5)$	τ	$\mathcal{V}_C^5$	$\mathcal{B}_C^5(\mathcal{V}_C^5)$	$\hat{\tau}$	$\mathcal{V}_D^5$	$\mathcal{B}_D^5(\mathcal{V}_D^5)$	$\hat{\tau}$
1	$c_{1,2}$	N	0.71	$c_{1,2}$	N	0.71	$c_{1,2}$	N	0.71
	$c_{1,3}$	N	0.33	$c_{2,3}$	N	0.51	$c_{1,5}$	F	0.70
	$c_{1,4}$	C	0.71	$c_{2,4}$	$G180$	0.70	$c_{4,5}$	G	0.75
	$c_{4,5}$	G	0.74	$c_{2,5}$	F	0.73	$c_{3,4}$	G	0.48
2	$c_{2,4;1}$	G	0.38	$c_{1,3;2}$	$G90$	-0.33	$c_{2,5;1}$	N	0.37
	$c_{3,4;1}$	G	0.47	$c_{1,4;2}$	$G180$	0.29	$c_{1,4;5}$	$G180$	0.22
	$c_{1,5;4}$	G	0.33	$c_{1,5;2}$	$G180$	0.25	$c_{3,5;4}$	C	0.15
3	$c_{2,3;1,4}$	C	0.35	$c_{3,4;1,2}$	N	0.27	$c_{2,4;1,5}$	F	0.18
	$c_{3,5;1,4}$	C	0.31	$c_{3,5;1,2}$	N	0.25	$c_{1,3;4,5}$	F	-0.26
4	$c_{2,5;1,3,4}$	N	0.13	$c_{4,5;1,2,3}$	G	0.20	$c_{2,3;1,4,5}$	$G180$	0.31

Table A.1: Chosen vine copula structures, copula families and Kendall's τ values for the R-vine copula model and the C- and D-vine alternatives in the five-dimensional case (N:=Normal, C:=Clayton, G:=Gumbel, F:=Frank, J:=Joe; 90, 180, 270:= degrees of rotation).

A.2 Outer Power Clayton Copula

The Outer Power Clayton copula is defined as follows:

$$C(u) = \psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d)),$$

where $\psi(t) = \widetilde{\psi}(t^{1/\beta})$ for some $\beta \in [1, \infty)$ and $\widetilde{\psi}(t)$ is the Clayton copula generator $\widetilde{\psi}(t) = (1+t)^{-1/\theta}$ for some $\theta \in (0, \infty)$. The inversion of Kendall's τ is not feasible here because $\tau = \tau(\theta, \beta) = 1 - \frac{2}{\beta(\theta+2)}$ and so (β, θ) are not identifiable individually. Our simulations using the CMLE instead of the inversion of Kendall's τ for other copulas (not reported here) suggest that the CMLE leads to a substantial power improvement of some GIMT, e.g., of $\mathcal{Q}_n$. We do not have an explanation for this phenomenon and so only report the least favorable results. The power reported in Section 2.4.2 for tests that do not involve the Outer Power Clayton copula is therefore conservative.

A.3 Non-GIMTs for Copulas

Here we provide details on the non-GIMTs used in Section 2.4.2. We start with a few definitions.

	R-vine			C-vine			D-vine		
T	$\mathcal{V}_R^8$	$\mathcal{B}_R^8(\mathcal{V}_R^8)$	τ	$\mathcal{V}_C^8$	$\mathcal{B}_C^8(\mathcal{V}_C^8)$	$\hat{\tau}$	$\mathcal{V}_D^8$	$\mathcal{B}_D^8(\mathcal{V}_D^8)$	$\hat{\tau}$
1	$c_{1,2}$	J	0.41	$c_{1,8}$	F	0.59	$c_{1,4}$	N	0.61
	$c_{1,4}$	N	0.59	$c_{2,8}$	F	0.51	$c_{4,5}$	G180	0.71
	$c_{1,5}$	N	0.59	$c_{3,8}$	N	0.55	$c_{5,8}$	F	0.60
	$c_{1,6}$	F	0.23	$c_{4,8}$	G180	0.59	$c_{7,8}$	G	0.65
	$c_{3,6}$	F	0.19	$c_{5,8}$	F	0.60	$c_{3,7}$	G180	0.41
	$c_{4,7}$	C	0.44	$c_{6,8}$	F	0.27	$c_{2,3}$	G	0.52
	$c_{7,8}$	G	0.64	$c_{7,8}$	G	0.65	$c_{2,6}$	J180	0.57
2	$c_{2,6;1}$	C	0.58	$c_{1,2;8}$	J	0.10	$c_{1,5;4}$	C	0.22
	$c_{1,3;6}$	G	0.44	$c_{2,3;8}$	J	0.29	$c_{4,8;5}$	C	0.22
	$c_{4,6;1}$	F	0.11	$c_{2,4;8}$	G	0.24	$c_{5,7;8}$	J90	-0.05
	$c_{4,5;1}$	C	0.53	$c_{2,5;8}$	G	0.29	$c_{3,8;7}$	G	0.41
	$c_{1,7;4}$	C	0.29	$c_{2,6;8}$	J180	0.52	$c_{2,7;3}$	J	0.10
	$c_{4,8;7}$	N	0.53	$c_{2,7;8}$	N	-0.17	$c_{3,6;2}$	G270	-0.48
3	$c_{5,6;1,4}$	N	0.19	$c_{1,4;2,8}$	N	0.28	$c_{1,8;4,5}$	N	0.20
	$c_{6,7;1,4}$	F	0.03	$c_{3,4;2,8}$	N	0.22	$c_{4,7;5,8}$	N	-0.13
	$c_{1,8;4,7}$	G	0.22	$c_{4,5;2,8}$	G180	0.41	$c_{3,5;7,8}$	G	0.18
	$c_{3,4;1,6}$	N	0.41	$c_{4,6;2,8}$	G270	-0.20	$c_{2,8;3,7}$	G	0.25
	$c_{2,3,1,6}$	G	0.68	$c_{4,7;2,8}$	I	0	$c_{6,7;2,3}$	C	0.08
4	$c_{6,8;1,4,7}$	C	0.17	$c_{1,6;2,4,8}$	J180	0.09	$c_{6,8;2,3,7}$	C	0.05
	$c_{5,7;1,4,6}$	N	0.09	$c_{3,6;2,4,8}$	N	-0.33	$c_{2,5;3,7,8}$	G	0.19
	$c_{3,5;1,4,6}$	F	0.21	$c_{5,6;2,4,8}$	F	-0.04	$c_{3,4;5,7,8}$	C180	0.09
	$c_{2,4;1,3,6}$	G	0.57	$c_{6,7;2,4,8}$	I	0	$c_{1,7;4,5,8}$	J180	0.06
5	$c_{2,5;1,3,4,6}$	J	0.25	$c_{1,5;2,4,6,8}$	C	0.23	$c_{5,6;2,3,7,8}$	C90	-0.04
	$c_{3,7;1,4,5,6}$	G	0.17	$c_{3,5;2,4,6,8}$	F	0.10	$c_{2,4;3,5,7,8}$	C90	-0.02
	$c_{5,8;1,4,6,7}$	F	0.02	$c_{5,7;2,4,6,8}$	F	0.05	$c_{1,3;4,5,7,8}$	G90	-0.09
6	$c_{2,7;1,3,4,5,6}$	G	0.31	$c_{1,3;2,4,5,6,8}$	F	0.07	$c_{4,6;2,3,5,7,8}$	C90	-0.14
	$c_{3,8;1,4,5,6,7}$	C	0.20	$c_{3,7;2,4,5,6,8}$	I	0	$c_{1,2;3,4,5,7,8}$	G90	-0.13
7	$c_{2,8;1,3,4,5,6,7}$	F	0.03	$c_{1,7;2,3,4,5,6,8}$	I	0	$c_{1,6;2,3,4,5,7,8}$	G180	0.24

Table A.2: Chosen vine copula structures, copula families and Kendall's τ values for R-vine copula model and the C- and D-vine alternatives in the eight-dimensional case (I:=indep., N:=Normal, C:=Clayton, G:=Gumbel, F:=Frank, J:=Joe; 90, 180, 270:= degrees of rotation).

Given a multivariate distribution, the Rosenblatt's transformation (Rosenblatt, 1952) yields a set of independent uniforms on $[0,1]$ from possibly dependent realizations obtained using that multivariate distribution. The Rosenblatt's transform can be specialized to copulas as follows:

Definition 18 *Rosenblatt's probability integral transformation (PIT) of a copula C is the mapping $\mathcal{R} : (0,1)^d \to (0,1)^d$ which to every $\boldsymbol{u} = (u_1, \ldots, u_d) \in (0,1)^d$ assigns a vector $\mathcal{R}(u) = (e_1, \ldots, e_d)$ with $e_1 = u_1$ and, for $i \in \{2, \ldots, d\}$,*

$$e_i = \frac{\partial^{i-1} C(u_1, \ldots, u_i, 1, \ldots, 1)}{\partial u_1 \cdots \partial u_{i-1}} \Big/ \frac{\partial^{i-1} C(u_1, \ldots, u_{i-1}, 1, \ldots, 1)}{\partial u_1 \cdots \partial u_{i-1}}. \tag{26}$$

As noted by Genest et al. (2009), the initial random vector $\mathbf{U}$ has distribution C, denoted $\mathbf{U} \sim C$, if and only if the distribution of the Rosenblatt's transform $\mathcal{R}(\mathbf{U})$ is the d-variate independence copula defined as $C_\perp(e_1, \ldots, e_d) = \prod_{j=1}^d e_j$. Thus $\mathcal{H}_0 : \mathbf{U} \sim C \in \mathcal{C}_0$ is equivalent to $\mathcal{H}_0^* : \mathcal{R}_{\boldsymbol{\theta}}(\mathbf{U}) \sim C_\perp$.

The PIT algorithm for R-vine copulas is given in the Appendix of Schepsmeier (2015). It makes use of the hierarchical structure of the R-vine, which simplifies the calculation of (26).

Definition 19 *Kendall's transformation is the mapping $\boldsymbol{X} \mapsto V = C(U_1, \ldots, U_d)$, where $U_i = F_i(X_i)$ for $i = 1, \ldots, d$ and C denotes the joint distribution of $\boldsymbol{U} = (U_1, \ldots, U_d)$.*

Let K denote the (univariate) distribution function of Kendall's transform V and let K_n denote the empirical analogue of K defined by

$$K_n(v) = \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}\{V_j \leq v\},\ v \in [0,1], \tag{27}$$

where $\mathbf{1}\{\cdot\}$ is the indicator function. Then, under standard regularity conditions, K_n is a consistent estimator of K. Also, under $\mathcal{H}_0$, the vector $\mathbf{U} = (U_1, \ldots, U_d)$ is distributed as $C_{\boldsymbol{\theta}}$ for some $\boldsymbol{\theta} \in \mathcal{O}$, and hence Kendall's transformation $C_{\boldsymbol{\theta}}(\mathbf{U})$ has distribution $K_{\boldsymbol{\theta}}$.

Note that K is not available for all parametric copula families in closed form, especially not for vine copulas. Thus Genest et al. (2009) use a bootstrap procedure to approximate K in such cases.

We now describe the non-GIMTs used in the simulation study.

A.3.1 Empirical Copula Process Test

This test is based on the empirical copula defined as follows:

$$C_n(\mathbf{u}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{U_{i1} \leq u_1, \ldots, U_{id} \leq u_d\}. \tag{28}$$

It is a well-known result that, under regularity conditions, C_n is a consistent estimator of the true underlying copula C, whether or not $\mathcal{H}_0$ is true. Note that $C_n(\mathbf{u})$ is different from $K_n(v)$, which is a univariate empirical distribution function.

A natural goodness-of-fit test would be based on a "distance" between C_n and an estimated copula C_{θ_n} obtained under $\mathcal{H}_0$. In this paper, $\hat{\boldsymbol{\theta}} = \Gamma_n(\mathbf{U}_1, \ldots, \mathbf{U}_n)$ stands for an estimator of $\boldsymbol{\theta}$ obtained using the pseudo-observations.

Thus the test relies on the empirical copula process (ECP) $\sqrt{n}(C_n - C_{\hat{\boldsymbol{\theta}}})$. In particular, it has the following rank-based Cramér-von Mises form:

$$\mathcal{S}_n = \int_{[0,1]^d} (C_n - C_{\hat{\boldsymbol{\theta}}})^2 \, dC_n(u) = \sum_{j=1}^{n} \{C_n(\mathbf{U}_j) - C_{\hat{\boldsymbol{\theta}}}(\mathbf{U}_j)\}^2, \tag{29}$$

where large values of $\mathcal{S}_n$ would lead to a rejection of $\mathcal{H}_0$. Genest et al. (2009) demonstrate that the test is consistent, that is, that if $C \notin \mathcal{C}_0$ then $\mathcal{H}_0$ is rejected with probability one as $n \to \infty$.

In the vine copula case we have to perform a double bootstrap procedure to obtain p-values since $C_{\hat{\boldsymbol{\theta}}_n}$ is not available in closed form.

A.3.2 Rosenblatt's Transformation Test

As an alternative to $\mathcal{S}_n$, Genest and Rémillard (2008) proposed using $\{\mathbf{V}_j = \mathcal{R}_{C_{\hat{\boldsymbol{\theta}}}}(\mathbf{U}_j)\}_{j=1}^{n}$ instead of $\mathbf{U}_j$, where $\mathcal{R}_{C_{\hat{\boldsymbol{\theta}}}}$ represents Rosenblatt's transformation with respect to the copula $C_{\hat{\boldsymbol{\theta}}_n} \in \mathcal{C}_0$ and $\hat{\boldsymbol{\theta}}$ is a consistent estimator of the true value $\boldsymbol{\theta}_0$, under $\mathcal{H}_0 : C \in \mathcal{C}_0 = \{C_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \mathcal{O}\}$.

The idea is then to compare $C_n(\mathbf{V}_j)$ with the independence copula $C_{\perp}(\mathbf{V}_j)$ and the corresponding Cramér-von Mises type statistic can be written as follows:

$$\mathcal{S}_n^R = \sum_{j=1}^{n} \{C_n(\mathbf{V}_j) - C_{\perp}(\mathbf{V}_j)\}^2. \tag{30}$$

In the vine copula context Schepsmeier (2015) called this GOF test ECP2 test addressing its close relation to the ECP.

A.3.3 Kendall's Transformation Test

Since under $\mathcal{H}_0$, the Kendall's transformation $C_{\boldsymbol{\theta}}(\mathbf{U})$ has distribution $K_{\boldsymbol{\theta}}$, the distance between K_n and a parametric estimator $K_{\hat{\boldsymbol{\theta}}}$ of K is another natural testing criterion. We are testing the null $\mathcal{H}_0^{**} : K \in \mathcal{K}_0 = \{K_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \mathcal{O}\}$ using the empirical process $\mathbb{K} = \sqrt{n}(K_n - K_{\hat{\boldsymbol{\theta}}})$. The specific statistic considered by Genest et al. (2006) is the following rank-based analogue of the Cramér-von Mises statistic

$$S_n^K = \int_0^1 \mathbb{K}_n(v)^2 dK_{\hat{\boldsymbol{\theta}}}(v)$$

Appendix B

Proofs in Chapter 3

B.1 The Plug-in Estimators of the Spearman's ρ and the Upper Tail Dependence Based on the Copula Density Estimators

B.1.1 The Spearman's ρ

In this section we are going to show that, in the bivariate case, if the copula density estimator is a density function but does not satisfy the uniform marginal property of copula density, then the plug-in estimator of the Spearman's ρ may take an invalid value.

Let $U, V \sim U(0,1)$ with joint distribution function C and its corresponding copula density c. Then the Spearman's ρ for (U, V) is given by

$$\begin{aligned}\rho_s(U,V) &= 12 \iint_{[0,1]^2} uv \,\mathrm{d}C(u,v) - 3 \\ &= 12 \iint_{[0,1]^2} uvc(u,v) \,\mathrm{d}u\mathrm{d}v - 3.\end{aligned}$$

We sometimes will use $\rho_s(c)$ to emphasis the dependence on c.

To estimate ρ_s, we only need the density estimator $c_n(\cdot,\cdot)$ and apply the plug-in estimator

$$\hat{\rho_s} = \rho_s(c_n) = 12 \iint_{[0,1]^2} uvc_n(u,v) \,\mathrm{d}u\mathrm{d}v - 3. \tag{31}$$

Now we are going to show that if $c_n(u,v)$ does not satisfy the uniform marginal property and is just a bivariate density function, then $\hat{\rho_s}$ (or $\rho_s(c_n)$) may not fall into [-1, 1].

Assume $c_n(\cdot,\cdot)$ is a copula density estimator which is a density function but not a copula density, e.g., the penalized exponential series estimator or the sieve MLE with Bernstein polynomials. Assume $F(\cdot,\cdot)$ is the corresponding bivariate distribution function, and F_u, F_v are the marginals, respectively. C^* is the corresponding copula of F.

Now consider $\hat{\rho_s} = \rho_s(c_n)$.

$$\begin{aligned}
\hat{\rho_s} &= \rho_s(c_n) \\
&= 12 \iint_{[0,1]^2} uvc_n(u,v)\,\mathrm{d}u\mathrm{d}v - 3 \\
&= 12 \iint_{[0,1]^2} uv\,\mathrm{d}F(u,v) - 3 \\
&= 12\ \mathrm{E}[UV] - 3, \qquad (32)
\end{aligned}$$

where U, V are random variables with joint distribution function $F(\cdot,\cdot)$, copula C^*, and marginals F_u, F_v respectively. Therefore,

$$\begin{aligned}
\hat{\rho_s} &= \rho_s(c_n) \\
&= 12\ \mathrm{E}[UV] - 3 \\
&= 12\left(\mathrm{Cov}(U,V) + \mathrm{E}[U]\,\mathrm{E}[V]\right) - 3 \\
&= 12\ \mathrm{Cov}(U,V) + (12\ \mathrm{E}[U]\,\mathrm{E}[V] - 3) \\
&= \frac{\mathrm{Cov}[U,V]}{\sqrt{\mathrm{Var}[U]}\sqrt{\mathrm{Var}[V]}}\left(\sqrt{\frac{\mathrm{Var}[U]}{\frac{1}{12}}}\sqrt{\frac{\mathrm{Var}[V]}{\frac{1}{12}}}\right) + (12\ \mathrm{E}[U]\,\mathrm{E}[V] - 3) \\
&= A_1 \cdot \mathrm{Corr}[U,V] + A_0, \qquad (33)
\end{aligned}$$

where $A_1 = \sqrt{\frac{\mathrm{Var}[U]}{\frac{1}{12}}}\sqrt{\frac{\mathrm{Var}[V]}{\frac{1}{12}}}$, and $A_0 = 12\,\mathrm{E}[U]\,\mathrm{E}[V] - 3$.

When $U, V \sim U(0,1)$, i.e., c_n is a copula density, $\hat{\rho_s} = \mathrm{Corr}[U,V] \in [-1,1]$. This is because, in this case, $A_1 = 1, A_0 = 0, \mathrm{E}[U] = \mathrm{E}[V] = \frac{1}{2}$, and $\mathrm{Var}[U] = \mathrm{Var}[V] = \frac{1}{12}$.

When $U, V \nsim U(0,1)$, i.e., c_n is just a density function but not a copula density, we will examine the values of A_1 and A_0 in the following. First

$$\begin{aligned} \mathrm{E}[U] &= \int_0^1 u \mathrm{d}F_u(u) \\ &= uF_u(u) \,\big|_0^1 - \int_0^1 F_u(u)\mathrm{d}u \\ &= 1 - \int_0^1 F_u(u)\mathrm{d}u \\ &= 1 - M_u^{(0)}, \end{aligned} \tag{34}$$

where $M_u^{(i)} = \int_0^1 u^i F_u(u)\mathrm{d}u$ for $i = 0, 1, \ldots$. It is obvious that $\int_0^1 F_u(u)\mathrm{d}u \in (0,1)$, therefore $\mathrm{E}[U] \in (0,1)$, and $A_0 \in (-3, 9)$.

Now consider A_1. Note that

$$\begin{aligned} \mathrm{Var}[U] &= \int_0^1 u^2 \,\mathrm{d}F_u(u) - (1 - M_u^{(0)})^2 \\ &= u^2 F_u(u) \,\big|_0^1 - 2\int_0^1 uF_u(u)\mathrm{d}u - (1 - M_u^{(0)})^2 \\ &= 1 - 2M_u^{(1)} - (1 - M_u^{(0)})^2 \\ &= -2M_u^{(1)} + 2M_u^{(0)} - M_u^{(0)^2}. \end{aligned}$$

When $U \nsim U(0,1)$, $\mathrm{Var}[U]$ is an increasing function with respect to $M_u^{(0)}$ on $[0,1]$. When $M_u^{(0)}$ goes to 0, $\mathrm{Var}[U]$ goes to 0, because $M_u^{(1)} \leq \int_0^1 u\mathrm{d}u = \frac{1}{2}$. Therefore A_1 goes to 0 while A_0 goes to 9, which means that when $\mathrm{E}[U]$ and $\mathrm{E}[V]$ are much greater than $\frac{1}{2}$ and close to 1, we will have a $\hat{\rho}_s$ much greater than 1.

To conclude, when c_n is not a copula density, it is possible to have $\rho_s(c_n) \notin [-1, 1]$.

B.1.2 The Upper Tail Dependence

In this section we are going to show that, in the bivariate case, if the copula density estimator is a density function but not a copula density, then the plug-in estimator of the upper tail dependence may take an invalid value.

Let X and Y be two random variables. The upper tail dependence λ_U is defined as

$$\lambda_U = \lim_{u \to 1^-} 2 - \frac{1 - C(u,u)}{1-u}, \tag{35}$$

where C is the true copula of (X, Y). We can estimate λ_U by a plug-in estimator $\hat{\lambda}_U$ replacing C with a copula estimator $\hat{C}$.

If we have a copula density estimator $\hat{c}$ for X and Y, which is a density function but not a copula density, then the copula estimator $\hat{C}$ based on $\hat{c}$ is not a copula but only a joint distribution function on $[0,1]^2$.

Denote $\hat{C}(\cdot,\cdot)$ by $F(\cdot,\cdot)$. Let F_1, F_2 and C^* be the two marginals and the copula associated with $F(\cdot,\cdot)$, respectively. Also let f_1 and f_2 be the density functions associated with F_1 and F_2, respectively. The support of F_1, F_2, f_1, and f_2 is $[0,1]$. f_1 and f_2 are not always equal to one on $[0,1]$ since $\hat{C}$ or F is not a copula function.

Now consider the plug-in estimator $\hat{\lambda}_U$ of λ_U defined in Eq.(35) by replacing C with $\hat{C}$.

$$\begin{aligned}
\hat{\lambda}_U &= 2 - \lim_{v \to 1^-} \frac{1 - \hat{C}(v,v)}{1-v} \\
&= 2 - \lim_{v \to 1^-} \frac{1 - F(v,v)}{1-v} \\
&= 2 - \lim_{v \to 1^-} \frac{1 - C^*(F_1(v), F_2(v))}{1-v} \\
&= 2 - \lim_{v \to 1^-} \left\{ \left. \frac{\partial C^*(x_1,x_2)}{\partial x_1} \right|_{(F_1(v),\, F_2(v))} \cdot f_1(v) + \left. \frac{\partial C^*(x_1,x_2)}{\partial x_2} \right|_{(F_1(v),\, F_2(v))} \cdot f_2(v) \right\} \\
&= 2 - \lim_{v \to 1^-} \left\{ \left. \frac{\partial C^*(x_1,x_2)}{\partial x_1} \right|_{(F_1(v),\, F_2(v))} \right\} \cdot f_1(1) - \left\{ \left. \frac{\partial C^*(x_1,x_2)}{\partial x_2} \right|_{(F_1(v),\, F_2(v))} \right\} \cdot f_2(1).
\end{aligned} \tag{36}$$

It can be easily shown that

$$\begin{aligned}
\left. \frac{\partial C^*(x_1,x_2)}{\partial x_1} \right|_{(a,\, b)} &= \Pr(Y \le b | X = a) \in [0,1], \\
\left. \frac{\partial C^*(x_1,x_2)}{\partial x_2} \right|_{(a,\, b)} &= \Pr(X \le a | Y = b) \in [0,1],
\end{aligned}$$

where (X, Y) is a random vector with uniform marginal and copula C^*, and $a, b \in [0, 1]$. Therefore,

$$\lim_{v \to 1^-} \left. \frac{\partial C^*(x_1, x_2)}{\partial x_1} \right|_{(F_1(v),\ F_2(v))} \in [0, 1],$$
$$\lim_{v \to 1^-} \left. \frac{\partial C^*(x_1, x_2)}{\partial x_2} \right|_{(F_1(v),\ F_2(v))} \in [0, 1].$$

If $\hat{C}$ is just a joint cumulative distribution function but not a copula, then $f_1(1)$ and $f_2(2)$ could be any non-negative numbers. It is possible that the estimate of the upper tail dependence $\hat{\lambda}_U$ is negative when $f_1(1)$ and $f_2(2)$ are large enough in Eq.(36).

B.2 Restrictions on the Design Points

In this section we are going to show that the copula property puts heavy restrictions on the design points. Consider a vertical integral $\int_0^1 f(x, y)\mathrm{d}y$. Suppose this crosses triangulation lines $1, 2, \ldots, M$ in order, where $M \in \mathbb{N}$. For $j \in 1, 2, \ldots, M$, line j connects $(x_j^{(1)}, y_j^{(1)})$ and $(x_j^{(2)}, y_j^{(2)})$ with $y_1^{(1)} = y_1^{(2)} = 0$ and $y_M^{(1)} = y_M^{(2)} = 1$. The value of f at the point where we cross line j is

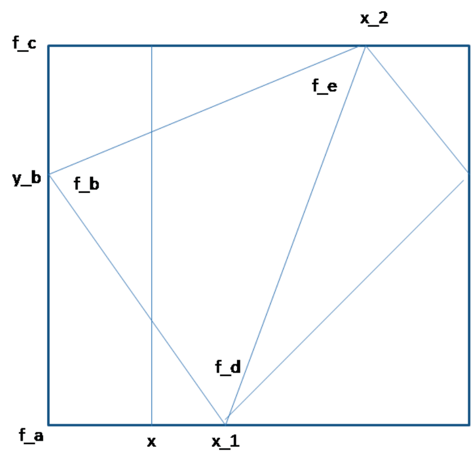
$$f_j = \frac{(x_j^{(2)} - x) f(x_j^{(1)}, y_j^{(1)}) + (x - x_j^{(1)}) f(x_j^{(2)}, y_j^{(2)})}{(x_j^{(2)} - x_j^{(1)})},$$

and this occurs at a y value of

$$y_j = \frac{(x_j^{(2)} - x) y_j^{(1)} + (x - x_j^{(1)}) y_j^{(2)}}{(x_j^{(2)} - x_j^{(1)})}.$$

The integral of f between line j and line $j + 1$ is

$$\begin{aligned}
&\int_{y_j}^{y_{j+1}} \frac{(y_{j+1} - y) f_j + (y - y_j) f_{j+1}}{y_{j+1} - y_j} \mathrm{d}y \\
&= \frac{1}{y_{j+1} - y_j} \left[(y_{j+1} y - \frac{y^2}{2}) f_j + (\frac{y^2}{2} - y_j y) f_{j+1} \right]_{y_j}^{y_{j+1}} \\
&= y_{j+1} f_j - y_j f_{j+1} + (\frac{y_{j+1} + y_j}{2})(f_{j+1} - f_j) \\
&= \frac{y_{j+1} - y_j}{2}(f_{j+1} + f_j).
\end{aligned}$$



We need the total integral to be constant with changes in x. Note that

$$\frac{\partial f_j}{\partial x} = \frac{f(x_j^{(2)}, y_j^{(2)}) - f(x_j^{(1)}, y_j^{(1)})}{(x_j^{(2)} - x_j^{(1)})};$$
$$\frac{\partial y_j}{\partial x} = \frac{y_j^{(2)} - y_j^{(1)}}{(x_j^{(2)} - x_j^{(1)})}.$$

Hence

$$\sum_{j=1}^{M-1} \left(\left(\frac{\partial y_{j+1}}{\partial x} - \frac{\partial y_j}{\partial x} \right) (f_{j+1} + f_j) + (y_{j+1} - y_j) \left(\frac{\partial f_{j+1}}{\partial x} - \frac{\partial f_j}{\partial x} \right) \right) = 0.$$

To examine the solution to the above system, we start with a simple case. Suppose that we have a situation as shown below. To calculate the marginal integral at x, it crosses four triangulation lines. The first one connects $(0, 0)$ and $(x_1, 0)$. The second one connects $(0, y_b)$ and $(x_1, 0)$. The third line connects $(0, y_b)$ and $(x_2, 1)$, and the fourth line connects $(0, 1)$ and $(x_2, 1)$. The value of f at the point where we cross line j is f_j, then

$$
\begin{aligned}
f_1 &= \frac{xf_d + (x_1 - x)f_a}{x_1}, \frac{\partial f_1}{\partial x} = \frac{f_d - f_a}{x_1}, y_1 = 0 \\
f_2 &= \frac{xf_d + (x_1 - x)f_b}{x_1}, \frac{\partial f_1}{\partial x} = \frac{f_d - f_b}{x_1}, y_2 = \frac{(x_1 - x)y_b}{x_1} \\
f_3 &= \frac{xf_e + (x_2 - x)f_b}{x_2}, \frac{\partial f_3}{\partial x} = \frac{f_e - f_b}{x_2}, y_3 = \frac{x + (x_2 - x)y_b}{x_2} \\
f_4 &= \frac{xf_e + (x_2 - x)f_c}{x_2}, \frac{\partial f_4}{\partial x} = \frac{f_e - f_c}{x_2}, y_4 = 1
\end{aligned}
$$

and the integral is

$$
= \frac{y_2 - y_1}{2}(f_2 + f_1) + \frac{y_3 - y_2}{2}(f_3 + f_2) + \frac{y_4 - y_3}{2}(f_4 + f_3)
$$

with derivative

$$
\begin{aligned}
&\frac{(x_1 - x)y_b}{x_1}\frac{2f_d - f_b - f_a}{x_1} - \frac{2xf_d + (x_1 - x)(f_a + f_b)}{x_1}\frac{y_b}{x_1} \\
&+ (\frac{x + (x_2 - x)y_b}{x_2} - \frac{(x_1 - x)y_b}{x_1})(\frac{f_e - f_b}{x_2} + \frac{f_d - f_b}{x_1}) \\
&+ \left(\frac{1 - y_b}{x_2} + \frac{y_b}{x_1}\right)\left(\frac{xf_d + (x_1 - x)f_b}{x_1} + \frac{xf_e + (x_2 - x)f_b}{x_2}\right) \\
&+ \left(1 - \frac{x + (x_2 - x)y_b}{x_2}\right)\frac{2f_e - f_b - f_c}{x_2} + \left(-\frac{1 - y_b}{x_2}\right)\left(\frac{2xf_e + (x_2 - x)(f_b + f_c)}{x_2}\right) \\
&= 0.
\end{aligned}
$$

Multiplying by $x_1^2 x_2^2$, we get

$$
\begin{aligned}
&x_2^2(x_1 - x)y_b(2f_d - f_b - f_a) - x_2^2\left(2xf_d + (x_1 - x)(f_a + f_b)\right)y_b \\
&+ \left(x_1\left(x + (x_2 - x)y_b\right) - x_2(x_1 - x)y_b\right)\left(x_1(f_e - f_b) + x_2\left(f_d - f_b\right)\right) \\
&+ \left(x_1(1 - y_b) + x_2 y_b\right)\left(x_2\left(xf_d + (x_1 - x)f_b\right) + x_1\left(xf_e + (x_2 - x)f_b\right)\right) \\
&+ x_1^2\left(x_2 - (x + (x_2 - x)y_b)\right)\left(2f_e - f_b - f_c\right) - x_1^2\left(1 - y_b\right)\left(2xf_e + (x_2 - x)(f_b + f_c)\right) \\
&= 0
\end{aligned}
$$

thus

$$\begin{aligned}
&2x_2^2y_b\left(xf_a+xf_b-2xf_d-x_1f_a-x_1f_b+x_1f_d\right)\\
&+2\left(x_1-x_1y_b+x_2y_b\right)\left(xx_1f_e-xx_1f_b-xx_2f_b+xx_2f_d+x_1x_2f_b\right)\\
&+2x_1^2\left(1-y_b\right)\left(xf_b-2xf_e+xf_c+x_2f_e-x_2f_b-x_2f_c\right)\\
&=0.
\end{aligned}$$

Since this is true for all x, we have

$$\begin{aligned}
x_2^2y_b\left(f_a+f_b-2f_d\right)+\left(x_1-x_1y_b+x_2y_b\right)\left(x_1f_e-x_1f_b-x_2f_b+x_2f_d\right)&\\
+x_1^2\left(1-y_b\right)\left(f_b-2f_e+f_c\right)&=0\\
x_1^2(f_c-f_e)(1-y_b)+x_2^2y_b(f_a-f_d)+x_1x_2\left(f_d-f_b+f_ey_b-f_dy_b\right)&=0
\end{aligned}$$

and

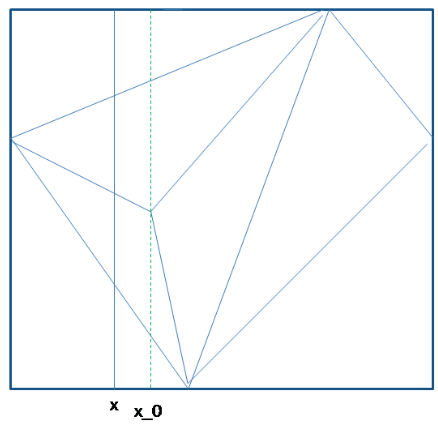
$$\begin{aligned}
2x_2^2y_b\left(-x_1f_a-x_1f_b+x_1f_d\right)+2\left(x_1-x_1y_b+x_2y_b\right)\left(x_1x_2f_b\right)&\\
+2x_1^2\left(1-y_b\right)\left(x_2f_e-x_2f_b-x_2f_c\right)&=0\\
x_2y_b\left(f_d-f_a-f_b\right)+\left(x_1(1-y_b)+x_2y_b\right)\left(f_b\right)+x_1\left(1-y_b\right)\left(f_e-f_b-f_c\right)&=0\\
x_2y_b\left(f_d-f_a\right)+x_1\left(1-y_b\right)\left(f_e-f_c\right)&=0.
\end{aligned}$$

We can then substitute this last equation in the above to get

$$\begin{aligned}
x_1^2(f_c-f_e)(1-y_b)+x_2x_1\left(1-y_b\right)\left(f_e-f_c\right)+x_1x_2\left(f_d-f_b+f_ey_b-f_dy_b\right)&=0\\
x_1(f_c-f_e)(1-y_b)+x_2\left(1-y_b\right)\left(f_e-f_c\right)+x_2\left(f_d-f_b+f_ey_b-f_dy_b\right)&=0.
\end{aligned}$$

Substituting again, we get

$$\begin{aligned}
x_2y_b\left(f_d-f_a\right)+x_2\left(1-y_b\right)\left(f_e-f_c\right)+x_2\left(f_d-f_b+f_ey_b-f_dy_b\right)&=0\\
y_b\left(f_d-f_a\right)+\left(1-y_b\right)\left(f_e-f_c\right)+\left(f_d-f_b+f_ey_b-f_dy_b\right)&=0\\
f_e-f_b-f_c+f_d-f_ay_b+f_cy_b&=0.
\end{aligned}$$



We also need for the integral to be 1 so

$$\begin{aligned} y_b(f_a + f_b) + (1 - y_b)(f_b + f_c) = 2 \\ y_b = \frac{2 - f_b - f_c}{f_a - f_c} \end{aligned}$$

and if $f_a = f_c$ then $f_b = 2 - f_a$. Thus

$$\begin{aligned} f_e - f_b - f_c + f_d - (2 - f_b - f_c) = 0 \\ f_e + f_d = 2 \end{aligned}$$

The three conditions that we need are:

$$\begin{aligned} f_e + f_d = 2 \\ y_b = \frac{2 - f_b - f_c}{f_a - f_c} \\ x_2 y_b \left(f_d - f_a\right) = x_1 \left(1 - y_b\right) \left(f_c - f_e\right). \end{aligned}$$

However, if we add an internal point, then we may be unable to make this work.

In conclusion, the uniform marginal property is quite restrictive on the set of the design points. And a simple set of design points that would satisfy the uniform marginal property would be an equidistant grid on the unit square.

For example, with the points shown above there will be a change in the derivative with respect to x of the integral as x passes through x_0. So for any internal point, there must be a second point on the same horizontal line and another one on the same vertical line.

We need to think about the same question with respect to the previous example. Is there a problem as x moves past x_1 if $x_1 \neq x_2$? This can certainly be avoided if the f values on the boundary are linear through the point in question. If not, then

$$g(x) = \int_0^\delta f(x,y)dy$$

cannot be linear in a range $(x_1 - \varepsilon, x_1 + \varepsilon)$ and yet

$$1 - g(x) = \int_\delta^1 f(x,y)dy$$

will be linear, thus giving a contradiction.

So this adds so much more conditions to the example in the case where $x_1 \neq x_2$ and $y_b \neq y_f$ (the unlabelled point on the opposite boundary) that in the end this will rule out any solution with $x_1 \neq x_2$ and $y_b \neq y_f$.

B.3 Generalization to General B-Splines

We are going to show that the originally proposed copula density estimation method using simple linear B-splines can be generalized to a similar method using B-splines of higher degrees while still preserving the uniform marginal property.

We are going to show that the tensor product spline surface generated by two sets of univariate B-splines with arbitrary degrees can preserve the uniform marginal property of the bivariate copula under mild assumptions on the knot vectors $\boldsymbol{t}$ and $\boldsymbol{s}$, and on coefficients $\{f_{ij}\}$.

Some lemmas that are needed for the proof of Proposition 22 are listed below.

Lemma 20 *A knot vector is said to be "$d+1$" regular if* $t_1 = ... = t_{d+1} < t_{d+2} < ... < t_n <$

$t_{n+1} = ... = t_{n+d+1}$ and $n \geq d+1$. On interval $[t_{d+1}, t_{n+1})$,

$$\sum_{j=1}^{n} B_{j,d}(x) \equiv 1. \tag{37}$$

Lemma 21 (Bhatti and Brachen (2006)) *The integral of a B-spline on its support is given by*

$$\begin{aligned}\int_{t_1}^{t_{n+d+1}} B_{j,d}(x)\mathrm{d}x &= \int_{t_j}^{t_{j+d+1}} B_{j,d}(x)\mathrm{d}x \\ &= \frac{t_{j+d+1} - t_j}{d+1}.\end{aligned}$$

Proposition 22 *Assume that we have a spline space $\mathcal{S}_1$ of degree d_t with a "$d_t + 1$ regular " knot vector $\boldsymbol{t} = (t_i)_{i=1}^{n_t+d_t+1}$ and $t_1 = 0, t_{n_t+d_t+1} = 1$, and another spline space $\mathcal{S}_2$ of degree d_s with a "$d_s + 1$ regular " knot vector $\boldsymbol{s} = (s_j)_{j=1}^{n_s+d_s+1}$ and $s_1 = 0, s_{n_s+d_s+1} = 1$. Let $\{B_{i,d_t,\boldsymbol{t}}\}_{i=1}^{n_t}$ and $\{B_{j,d_s,\boldsymbol{s}}\}_{j=1}^{n_s}$ denote the B-splines in $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively. We claim that*

$$f(x,y) = \sum_{i=1}^{N}\sum_{j=1}^{N} f_{ij}\ B_{i,d_t,\boldsymbol{t}}(x)\ B_{j,d_s,\boldsymbol{s}}(y) \tag{38}$$

with constraints

$$\begin{aligned}&\sum_{j=1}^{n_s} \frac{s_{j+d_s+1} - s_j}{d_s+1} f_{ij} = 1,\ i = 1, 2, \ldots, n_t, \\ &\sum_{i=1}^{n_t} \frac{t_{i+d_t+1} - t_i}{d_t+1} f_{ij} = 1,\ j = 1, 2, \ldots, n_s\end{aligned} \tag{39}$$

can be used to approximate the copula density while preserving the uniform marginal property.

Proof. We are going to prove that $\forall x \in [0, 1)$, $\int_0^1 f(x, y)\mathrm{d}y = 1$.

$$\begin{aligned}
\int_0^1 f(x,y)\mathrm{d}y &= \int_0^1 \sum_{i=1}^{n_t}\sum_{j=1}^{n_s} f_{ij}\ B_{i,d_t,\boldsymbol{t}}(x)\ B_{j,d_s,\boldsymbol{s}}(y)\mathrm{d}y \\
&= \int_0^1 \sum_{i=1}^{n_t}\ B_{i,d_t,\boldsymbol{t}}(x)\sum_{j=1}^{n_s} f_{ij}\ B_{j,d_s,\boldsymbol{s}}(y)\mathrm{d}y \\
&= \sum_{i=1}^{n_t}\ B_{i,d_t,\boldsymbol{t}}(x)\sum_{j=1}^{n_s} f_{ij}\int_0^1 B_{j,d_s,\boldsymbol{s}}(y)\mathrm{d}y \\
&= \sum_{i=1}^{n_t}\ B_{i,d_t,\boldsymbol{t}}(x)\sum_{j=1}^{n_s} f_{ij}\int_{s_1}^{s_{n_s+d_s+1}} B_{j,d_s,\boldsymbol{s}}(y)\mathrm{d}y \\
&= \sum_{i=1}^{n_t}\ B_{i,d_t,\boldsymbol{t}}(x)\sum_{j=1}^{n_s} f_{ij}\frac{s_{j+d_s+1}-s_j}{d_s+1} \\
&= \sum_{i=1}^{n_t}\ B_{i,d_t,\boldsymbol{t}}(x) \\
&= 1.
\end{aligned}$$

The last two equations are due to Eq.(39) in the assumptions and Eq.(37), respectively. ■

Appendix C

Proofs in Chapter 4

C.1 Convergence of the Denominator

Proof of Lemma 13: Recall that

$$D_{n,0}(z) = \frac{1}{n}\sum_{i=1}^{n} K_h(Z_{2j(i)} - z)\begin{pmatrix} X_{1i}X_{1i}^T & X_{1i}X_{2j(i)} \\ X_{2j(i)}X_{1i}^T & X_{2j(i)}^2 \end{pmatrix}.$$

Let $D_{n,0_{(j_1,j_2)}}(z)$ denote the (j_1, j_2)th block matrix element of $D_{n,0}(z)$. Then we will prove that $D_{n,0_{(22)}}(z) \xrightarrow{p} f(z)\Omega_{(22)}(z)$ by showing that $D_{n,0_{(22)}}(z)$ converges to $f(z)\Omega_{(22)}(z)$ in mean-square. The convergence of the rest of $D_{n,0_{(j_1,j_2)}}(z)$ can be shown in a similar way.

Before we proceed, we first outline some properties of $Z_{2j(i)}, i \in \{1, ..., n\}$. Let $f_{j(i)}(\cdot), i \in \{1, ..., n\}$ denote the density of $Z_{2j(i)}, i \in \{1, ..., n\}$, which is also the distribution of Z_{2j} conditional on it being a nearest match to Z_{1i}, $j \in \{1, ...m\}$. Because the density of Z_{2j} is $f(z)$, then

$$\begin{aligned} f_{j(i)}(z) &= \sum_{j=1}^{m} Pr(j(i) = j \mid Z_{2j} = z)f(z) \\ &= f(z) \cdot \sum_{j=1}^{m}(\frac{1}{m} + o(\frac{1}{m})) \\ &= f(z)(1 + o(1)), \end{aligned}$$

where the second equality comes from the conditional probability result derived in Abadie and Imbens (2006)'s Additional Proofs on Page 5. The above implies that the density of any $Z_{2j(i)}$ differs from another density of $Z_{2j(k)}$ and the population density by $o(1)$, $i, k \in \{1, ..., n\}, k \neq i$. Also, $C(j)$ is defined as the number of times that Element j in the sample $\{X_{2j}, Z_{2j}\}_{j=1}^{m}$ is used as a match to Element i in the sample $\{Y_i, X_{1i}, Z_{1i}\}_{i=1}^{n}$,

$$C(j) = \sum_{i=1}^{n} \mathbb{1}(j = j(i)), \quad j \in \{1, ...m\}$$

where $\mathbb{1}(\cdot)$ is the indicator function, equal to one if $j = j(i)$ is true and zero otherwise. Then $\mathbb{E}(C(j) \mid Z_{2j}) = \frac{n}{m}(1 + o(1))$ is given in Abadie and Imbens (2006)'s Additional Proofs on Page 11.

Let $Z^{1,2}$ denote all the Z from the two samples. Note that, the conditional expectation of $D_{n,0_{(22)}}(z)$ is

$$\begin{aligned}\mathbb{E}\left(D_{n,0_{(22)}}(z) \mid Z^{1,2}\right) &= \frac{1}{n}\sum_{i=1}^{n} K_h(Z_{2j(i)} - z)\Omega_{(22)}(Z_{2j(i)}) \\ &= \frac{1}{n}\sum_{j=1}^{m} C(j)K_h(Z_{2j} - z)\Omega_{(22)}(Z_{2j}).\end{aligned}$$

And the unconditional expectation is

$$\begin{aligned}\mathbb{E}\left(D_{n,0_{(22)}}(z)\right) &= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} K_h(Z_{2j(i)} - z)\Omega_{(22)}(Z_{2j(i)})\right) \\ &= \mathbb{E}\left(\frac{1}{n}\sum_{j=1}^{m} C(j)K_h(Z_{2j} - z)\Omega_{(22)}(Z_{2j})\right) \\ &= \frac{1}{n}\sum_{j=1}^{m} \mathbb{E}\left(\mathbb{E}(C(j) \mid Z_{2j})K_h(Z_{2j} - z)\Omega_{(22)}(Z_{2j})\right) \\ &= \frac{m}{n}\mathbb{E}\left(\frac{n}{m}(1 + o(1))K_h(Z_{2j} - z)\Omega_{(22)}(Z_{2j})\right) \\ &= f(z)\Omega_{(22)}(z)(1 + o(1) + O(h^2)).\end{aligned}$$

Next, we are going to show the convergence of the variance of $D_{n,0_{(22)}}(z)$. Apply the law of total variance,

$$
\begin{aligned}
&Var\left(D_{n,0_{(22)}}(z)\right) \\
&= \mathbb{E}\left(Var\left(D_{n,0_{(22)}}(z) \mid Z^{1,2}\right)\right) + Var\left(\mathbb{E}\left(D_{n,0_{(22)}}(z) \mid Z^{1,2}\right)\right).
\end{aligned}
$$

First, we consider $Var\left(\mathbb{E}\left(D_{n,0_{(22)}}(z) \mid Z^{1,2}\right)\right)$. Since $\Omega(z) = E(XX' \mid z)$ is a matrix, we transform it into a vector for its convergence. Let $\Omega_{(22)}(z)$ be the $(2,2)$th block matrix element of $\Omega(z)$. Denote

$$
\begin{aligned}
A_j &= vec\left(K_h(Z_{2j}-z)\Omega_{(22)}(Z_{2j}) - f(z)\Omega_{(22)}(z)\right), \\
A_{j(i)} &= vec\left(K_h(Z_{2j(i)}-z)\Omega_{(22)}(Z_{2j(i)}) - f(z)\Omega_{(22)}(z)\right).
\end{aligned}
$$

Then $Var\left(\mathbb{E}\left(D_{n,0_{(22)}}(z) \mid Z^{1,2}\right)\right)$ can be rewritten as

$$
\begin{aligned}
& Var\left(\mathbb{E}\left(D_{n,0_{(22)}}(z) \mid Z^{1,2}\right)\right) \\
& = \mathbb{E}\left((\frac{1}{n}\sum_{i=1}^{n} A_{j(i)})(\frac{1}{n}\sum_{i=1}^{n} A_{j(i)})^T \right) \\
& = \frac{1}{n^2}\mathbb{E}\left(\sum_{i=1}^{n}\sum_{t=1}^{n} A_{j(i)}A_{j(t)}^T \right) \\
& = \frac{1}{n^2}\mathbb{E}\left(\sum_{i=1}^{n} \sum_{\substack{j(t)=j(i) \\ t\in\{1,\ldots,n\}}} A_{j(i)}A_{j(t)}^T \right) + \frac{1}{n^2}\mathbb{E}\left(\sum_{i=1}^{n} \sum_{\substack{j(t)\neq j(i) \\ t\in\{1,\ldots,n\}}} A_{j(i)}A_{j(t)}^T \right) \\
& = \frac{1}{n^2}\mathbb{E}\left(\sum_{i=1}^{n} A_{j(i)}C(j(i))A_{j(i)}^T \right) + \frac{1}{n^2}\sum_{i=1}^{n} \sum_{\substack{j(t)\neq j(i) \\ t\in\{1,\ldots,n\}}} \mathbb{E}\left(A_{j(i)}\right)\mathbb{E}\left(A_{j(t)}^T\right) \\
& \leq \frac{1}{n^2}\mathbb{E}\left(\sum_{i=1}^{n} (\max_{j=1,\ldots,m} C(j))A_{j(i)}A_{j(i)}^T \right) + \frac{n^2-n}{n^2}(o(1)^2 + O(h^2)^2) \\
& = \frac{1}{n^2}\mathbb{E}\left(\sum_{j=1}^{m} (\max_{j=1,\ldots,m} C(j))C(j)A_jA_j^T \right) + O(h^4) + o(1)^2 \\
& = \frac{1}{\sqrt{n}}\left(\frac{m}{n}\right)^{\frac{3}{2}}\mathbb{E}\left((\frac{1}{\sqrt{m}} \max_{j=1,\ldots,m} C(j))C(j)A_jA_j^T \right) + O(h^4) + o(1)^2 \\
& = \frac{1}{\sqrt{n}}\left(\frac{m}{n}\right)^{\frac{3}{2}}\mathbb{E}\left(\frac{1}{\sqrt{m}} (\max_{j=1,\ldots,m} C(j))^2 A_jA_j^T \right) + O(h^4) + o(1)^2 \\
& \leq \frac{1}{\sqrt{n}}\left(\frac{m}{n}\right)^{\frac{3}{2}}\mathbb{E}\left(\frac{1}{\sqrt{m}} (\max_{j=1,\ldots,m} C(j))^2 \right)\bar{\mu}_2 + O(h^4) + o(1)^2,
\end{aligned}
$$

where $\bar{\mu}_2 = \sup_{Z_j} \|A_jA_j^T\|$. $\bar{\mu}_2$ is finite due to Assumption 2 and Assumption 3, which implies that Ω and K_h are continuous on the finite support of Z, and in turn implies that K_h and Ω satisfy the Lipschitz condition. Next, we are going to prove that $\mathbb{E}\left(\frac{1}{\sqrt{m}}(\max_{j=1,\ldots,m} C(j))^2\right)$ is bounded uniformly in m, which implies that $Var\left(\mathbb{E}\left(D_{n,0_{(22)}}(z) \mid Z^{1,2}\right)\right)$ converges to zero matrix as n, m go to infinity of the same order.

Following the proofs on Page 23 in Abadie and Imbens (2006)'s Additional Proofs, by Bonferroni's inequality:

$$
\begin{aligned}
\mathbb{E}\left((\frac{1}{\sqrt{m}}(\max_{j=1,\dots,m} C(j))^2)^2 \right) &= \frac{1}{m}\mathbb{E}\left(\max_{j=1,\dots,m} C(j)^4 \right) \\
&= \frac{1}{m}\sum_{N=0}^{\infty} \text{Pr}(\max_{j=1,\dots,m} C(j)^4 > N) \\
&\leq \frac{1}{m}\sum_{N=0}^{\infty} m\text{Pr}(C(j)^4 > N) \\
&\leq \mathbb{E}\left(C(j)^4\right).
\end{aligned} \tag{40}
$$

The first equation holds because $C(j)$ is a positive integer for all $j = 1, \dots, m$. According to Lemma 3 in Abadie and Imbens (2006) , given Assumption 1 and part 1 of Assumption 2, $\mathbb{E}\left(C(j)^q\right)$ is bounded uniformly in m for all $q > 0$. Since

$$
\mathbb{E}\left(\frac{1}{\sqrt{m}}(\max_{j=1,\dots,m} C(j))^2 \right)^2 \leq \mathbb{E}\left((\frac{1}{\sqrt{m}}(\max_{j=1,\dots,m} C(j))^2)^2 \right), \tag{41}
$$

which implies the boundedness of $\mathbb{E}\left(\frac{1}{\sqrt{m}}(\max_{j=1,\dots,m} C(j))^2 \right)$, and in turn implies that

$$
Var\left(\mathbb{E}\left(D_{n,0_{(22)}}(z) \mid Z^{1,2} \right) \right)
$$

converges to zero matrix as n, m go to infinity of the same order. By the same method, we can easily show that under certain finite moment assumptions, the expectation of $Var\left(D_{n,0_{(22)}}(z) \mid Z^{1,2} \right)$ goes to zero as sample sizes go to infinity.

To conclude, we have showed that $Var\left(D_{n,0_{(22)}}(z) \right)$ goes to zero as sample sizes go to infinity and therefore $D_{n,0_{(22)}}(z)$ converges to $f(z)\Omega_{(22)}(z)$ in mean-square. Similarly, we can show that all the block matrix elements of $D_{n,0}(z)$ converge to the respective block matrix elements of $f(z)\Omega(z)$ in mean-square. The convergence of $D_n(z)$ to $f(z)\Omega(z) \bigotimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix}$ is proved in the same line of arguments.

C.2 Convergence of the Numerator

Proof of Lemma 14: We approximate the expression of $Y = X_i^T\beta(Z_{1i}) + u_i \simeq X_i^T\beta(Z_{2j(i)}) + u_i$ by a Taylor expansion in the neighbourhood of $|Z_{1i} - z| < h$ and $|Z_{2j(i)} - z| < h$.

$$X_i^T\beta(Z_{1i}) = X_i^T\beta(z) + (Z_{2j(i)} - z)X_i^T\beta'(z) + \frac{h^2}{2}\left(\frac{Z_{2j(i)} - z}{h}\right)^2 \beta(z)'' + o(h^2) \;\; a.s.,$$

where $\beta'(z)$ and $\beta(z)''$ are the vectors consisting of the first and the second derivatives of the function $\beta(z)$.

Then we substitute the approximate expression of Y into the expression of $N_n(z)$.

$$\begin{aligned} N_n(z) &= \frac{1}{n}(D^{X_m})^T WY \\ &= \frac{1}{n}(D^{X_m})^T W \left(D^X \begin{pmatrix} \beta(z) \\ h\beta'(z) \end{pmatrix} + \frac{h^2}{2} A_z X^T \beta''(z) + u + o(h^2) \right), \end{aligned}$$

where

$$D^{X_m} = \begin{pmatrix} X_{j(1)}^T & X_{j(1)}^T \frac{Z_{2j(1)} - z}{h} \\ ... & ... \\ X_{j(n)}^T & X_{j(n)}^T \frac{Z_{2j(n)} - z}{h} \end{pmatrix},$$

$$W = diag\left(K_h(Z_{2j(1)} - z), ..., K_h(Z_{2j(n)} - z)\right),$$

$$D^X = \begin{pmatrix} X_1^T & X_1^T \frac{Z_{2j(1)} - z}{h} \\ ... & ... \\ X_n^T & X_n^T \frac{Z_{2j(n)} - z}{h} \end{pmatrix},$$

and

$$A_z = diag\left(\left(\frac{Z_{2j(1)} - z}{h}\right)^2, ..., \left(\frac{Z_{2j(n)} - z}{h}\right)^2\right).$$

Recall that $X_{j(i)} = (X_{1i}^T, X_{2j(i)})^T$ is the matching pair for $X_i = (X_{1i}^T, X_{2i})^T$, $i \in \{1, ...n\}$.

Next, we decompose $N_n(z) - \frac{1}{n}(D^{X_m})^T W D^{X_m} \begin{pmatrix} \beta(z) \\ h\beta'(z) \end{pmatrix} = I_1 + I_2 + I_3 + I_4$, where

$$I_1 = \frac{1}{n}(D^{X_m})^T W \left(D^X \begin{pmatrix} \beta(z) \\ h\beta'(z) \end{pmatrix} - \mathbb{E}(D^X \mid Z^{1,2}) \begin{pmatrix} \beta(z) \\ h\beta'(z) \end{pmatrix} \right),$$

$$I_2 = \frac{1}{n}(D^{X_m})^T W \left(\mathbb{E}(D^X \mid Z^{1,2}) \begin{pmatrix} \beta(z) \\ h\beta'(z) \end{pmatrix} - D^{X_m} \begin{pmatrix} \beta(z) \\ h\beta'(z) \end{pmatrix} \right),$$

$$I_3 = \frac{1}{n}(D^{X_m})^T W \frac{h^2}{2} A_z X^T \beta''(z),$$

$$I_4 = \frac{1}{n}(D^{X_m})^T W u.$$

Firstly, we consider I_1. Since

$$D^X - \mathbb{E}(D^X \mid Z^{1,2}) = \begin{pmatrix} (X_1 - g(Z_{11}))^T & (X_1 - g(Z_{11}))^T \frac{(Z_{2j(1)}-z)}{h} \\ ... & ... \\ (X_n - g(Z_{1n}))^T & (X_n - g(Z_{1n}))^T \frac{(Z_{2j(n)}-z)}{h} \end{pmatrix},$$

then

$$I_1 = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n X_{j(i)} \left(X_i - g\left(Z_{1i}\right)\right)^T K_h\left(Z_{2j(i)} - z\right) \\ \frac{1}{nh}\sum_{i=1}^n X_{j(i)} \left(X_i - g\left(Z_{1i}\right)\right)^T \left(Z_{2j(i)} - z\right) K_h\left(Z_{2j(i)} - z\right) \end{pmatrix} \beta(z)$$
$$+ \begin{pmatrix} \frac{1}{nh}\sum_{i=1}^n X_{j(i)} \left(X_i - g\left(Z_{1i}\right)\right)^T \left(Z_{2j(i)} - z\right) K_h\left(Z_{2j(i)} - z\right) \\ \frac{1}{nh^2}\sum_{i=1}^n X_{j(i)} \left(X_i - g(Z_{1i})\right)^T \left(Z_{2j(i)} - z\right)^2 K_h\left(Z_{2j(i)} - z\right) \end{pmatrix} h\beta'(z).$$

We will show that

$$I_1 \xrightarrow{p} f(z) \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ 0 & 0 \end{pmatrix} \beta(z) \bigotimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} + f(z) \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ 0 & 0 \end{pmatrix} h\beta'(z) \bigotimes \begin{pmatrix} 0 \\ \mu_2 \end{pmatrix}.$$

Denote

$$I_{10} = \frac{1}{n}\sum_{i=1}^{n} X_{j(i)}(X_i - g(Z_{1i}))^T K_h(Z_{2j(i)} - z),$$

$$I_{11} = \frac{1}{nh}\sum_{i=1}^{n} X_{j(i)}(X_i - g(Z_{1i}))^T (Z_{2j(i)} - z) K_h(Z_{2j(i)} - z),$$

and

$$I_{12} = \frac{1}{nh}\sum_{i=1}^{n} X_{j(i)}(X_i - g(Z_{1i}))^T (Z_{2j(i)} - z)^2 K_h(Z_{2j(i)} - z).$$

Next, we will show that $I_{10} \xrightarrow{p} f(z)\begin{pmatrix} \Sigma_{11}(z) & \Sigma_{12}(z) \\ 0 & 0 \end{pmatrix}$. Suppose that $j(i) = j$, then X_{2j} is the match to X_{2i}. Denote $X_{(i,j)} = (X_{1i}^T, X_{2j}^T)^T$ as the match to $X_i = (X_{1i}^T, X_{2i})$, when $j(i) = j$.

$$\begin{aligned}
\mathbb{E}(I_{10} \mid Z^{1,2}) &= \frac{1}{n}\sum_{j=1}^{m} \mathbb{E}\left(\sum_{i\in\mathbb{A}(j)} X_{(i,j)}(X_i - g(Z_{1i}))^T K_h(Z_{2j} - z) \mid Z^{1,2}\right) \\
&= \frac{1}{n}\sum_{j=1}^{m} K_h(Z_{2j} - z)\mathbb{E}\left(\sum_{i\in\mathbb{A}(j)} X_{(i,j)}(X_i - g(Z_{1i}))^T \mid Z^{1,2}\right) \\
&= \frac{1}{n}\sum_{j=1}^{m} K_h(Z_{2j} - z)\sum_{i\in\mathbb{A}(j)} \mathbb{E}\left(\begin{pmatrix} X_{1i} \\ X_{2j} \end{pmatrix}(v_{1i}, v_{2i})^T \mid Z^{1,2}\right) \\
&= \frac{1}{n}\sum_{j=1}^{m} K_h(Z_{2j} - z)\sum_{i\in\mathbb{A}(j)} \begin{pmatrix} \Sigma_{11}(Z_{1i}) & \Sigma_{12}(Z_{1i}) \\ 0 & 0 \end{pmatrix} \\
&= \frac{1}{n}\sum_{j=1}^{m} K_h(Z_{2j} - z)C(j)\begin{pmatrix} \Sigma_{11}(Z_{2j}) & \Sigma_{12}(Z_{2j}) \\ 0 & 0 \end{pmatrix}(1 + o(1)).
\end{aligned}$$

The next to the last equality is from the fact that $\{X_{1i}\}_{i=1}^n$ and $\{X_{2j}\}_{j=1}^m$ are two independent samples from the same population. Since $\mathbb{A}(j)$ is defined as the subset of the index i, $i \in \{1, ..., n\}$, such that j is used as a match to each indexed observation, then the number of elements in the set $\mathbb{A}(j)$ is $C(j)$, $j \in \{1, ...m\}$. Also $C(j)$ is nonstochastic conditional on $Z^{1,2}$. As we discussed before, for a pair of match, $Z_{2j(i)}$ and Z_{1i}, their densities only differ by $o(1)$. So for all $i \in \mathbb{A}(j)$, the densities of Z_{1i} and Z_{2j} differ by $o(1)$ as well. Therefore, the last equality holds.

To compute the unconditional expectation of I_{10}, we apply the result of $\mathbb{E}(C(j)|Z^{1,2}) = \frac{n}{m}(1+o(1))$ from Abadie and Imbens (2006). Then

$$\mathbb{E}(I_{10}) = f(z)\begin{pmatrix} \Sigma_{11}(z) & \Sigma_{12}(z) \\ 0 & 0 \end{pmatrix}.$$

The convergence of the variance $Var(I_{10})$ is in the Appendix C.3. As a result, we have the convergence result of I_{10}. Similarly, we can show the convergence of I_{11} and I_{12}. And then the convergence of I_1 is straightforward.

Next, we consider the convergence of I_2. Since

$$\mathbb{E}(D^X \mid Z^{1,2}) - D^{X_m} = \begin{pmatrix} (g(Z_{11}) - X_{j(1)})^T & (g(Z_{11}) - X_{j(1)})^T \frac{(Z_{1j(1)}-z)}{h} \\ \dots & \dots \\ (g(Z_{1n}) - X_{j(n)})^T & (g(Z_{1n}) - X_{j(n)})^T \frac{(Z_{nj(n)}-z)}{h} \end{pmatrix},$$

then

$$\begin{aligned} I_2 = & \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n X_{j(i)}(g(Z_{1i}) - X_{j(i)})^T K_h(Z_{2j(i)} - z) \\ \frac{1}{nh}\sum_{i=1}^n X_{j(i)}(g(Z_{1i}) - X_{j(i)})^T (Z_{1i} - z) K_h(Z_{2j(i)} - z) \end{pmatrix} \beta(z) \\ & + \begin{pmatrix} \frac{1}{nh}\sum_{i=1}^n X_{j(i)}(g(Z_{1i}) - X_{j(i)})^T (Z_{1i} - z) K_h(Z_{2j(i)} - z) \\ \frac{1}{nh^2}\sum_{i=1}^n X_{j(i)}(g(Z_{1i}) - X_{j(i)})^T (Z_{1i} - z)^2 K_h(Z_{2j(i)} - z) \end{pmatrix} h\beta'(z). \end{aligned}$$

Following a similar argument, we can show that

$$\begin{aligned} I_2 \xrightarrow{p} & f(z)\begin{pmatrix} -\Sigma_{11} & 0 \\ 0 & -\Sigma_{22} \end{pmatrix} \beta(z) \bigotimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ & + f(z)\begin{pmatrix} -\Sigma_{11} & 0 \\ 0 & -\Sigma_{22} \end{pmatrix} h\beta'(z) \bigotimes \begin{pmatrix} 0 \\ \mu_2 \end{pmatrix}. \end{aligned}$$

Finally, we can compute in a similar way to prove that $I_3 = o(h^2)$ and $I_4 = o_p(1)$.

C.3 Convergence of $Var(I_{10})$

Let $vec(\cdot)$ denote the vectorization transformation of matrix and $VI_{10} = vec(I_{10})$. Consider the variance decomposition,

$$Var\left(VI_{10}\right) = \mathbb{E}\left(Var(VI_{10} \mid Z^{1,2})\right) + Var\left(\mathbb{E}(VI_{10} \mid Z^{1,2})\right).$$

In the following, we will prove that the convergence of both $\mathbb{E}\left(Var\left(VI_{10} \mid Z^{1,2}\right)\right)$ and $Var\left(\mathbb{E}\left(VI_{10} \mid Z^{1,2}\right)\right)$ as sample sizes go to infinity.

First, consider $\mathbb{E}\left(Var\left(VI_{10} \mid Z^{1,2}\right)\right)$. Let

$$B_{(i,j)} = K_h(Z_{j(i)} - z) \cdot vec\left(X_{j(i)}(X_i - g(Z_i))^T - \begin{pmatrix} \Sigma_{11}(Z_i) & \Sigma_{12}(Z_i) \\ 0 & 0 \end{pmatrix}\right),$$
$$B_{(i,i)} = K_h(Z_i - z) \cdot vec\left(X_i(X_i - g(Z_i))^T - \Sigma\right).$$

Then

$$\begin{aligned}
&\mathbb{E}\left(Var\left(VI_{10} \mid Z^{1,2}\right)\right) \\
&= \mathbb{E}\left(\left(VI_{10} - \mathbb{E}\left(VI_{10} \mid Z^{1,2}\right)\right)\left(VI_{10} - \mathbb{E}\left(VI_{10} \mid Z^{1,2}\right)\right)^T\right) \\
&= \mathbb{E}\left(\left(\frac{1}{n}\sum_{i=1}^{n} B_{(i,j)}\right)\left(\frac{1}{n}\sum_{i=1}^{n} B_{(i,j)}\right)^T\right) \\
&\leq \frac{1}{\sqrt{n}}\left(\frac{m}{n}\right)^{\frac{3}{2}} \mathbb{E}\left(\frac{1}{\sqrt{m}}(\max_{j=1,\ldots,m} C(j))^2 B_{(i,i)} B_{(i,i)}^T\right) + O(h^4) + o(1)^2.
\end{aligned}$$

Following the same assumptions and argument in proof of Lemma 13, we have that

$$\mathbb{E}\left(\frac{1}{\sqrt{m}}(\max_{j=1,\ldots,m} C(j))^2 B_{(i,i)} B_{(i,i)}^T\right)$$

is bounded uniformly in all m and therefore $\mathbb{E}\left(Var\left(VI_{10} \mid Z^{1,2}\right)\right)$ converges to zero matrix as sample size n, m go to infinity of the same order.

Before we proceed, we introduce a useful lemma from Abadie and Imbens (2006) about the distribution of the *matching discrepancy*. Suppose that we have a random sample $Z_1, \ldots, Z_N$, with density f over bounded support Z. Now consider the closest match to a $z \in$ Z in the sample. Let $j_1 = \text{argmin}_{j=1,\ldots,N}\|Z_j - z\|$ and let $U_{j_1} = Z_{j_1} - z$ be the matching discrepancy.

Lemma 23 ***Matching Discrepancy - Asymptotic Properties:*** *Suppose that f is differentiable in a neighbourhood of z. Then $U_{j_1} = O_p(N^{-1})$. Moreover, the first two moments of U_{j_1} are $O(N^{-\frac{1}{2}})$.*

Now, consider $Var\left(\mathbb{E}\left(VI_{10} \mid Z^{1,2}\right)\right)$. Define

$$A_{i,j(i)} = K_h(Z_{j(i)} - z) \cdot vec\begin{pmatrix} \Sigma_{11}(Z_i) & \Sigma_{12}(Z_i) \\ 0 & 0 \end{pmatrix} - f_z(z) \cdot vec\begin{pmatrix} \Sigma_{11}(z) & \Sigma_{12}(z) \\ 0 & 0 \end{pmatrix},$$

$$A_{i,i} = K_h(Z_i - z) \cdot vec\begin{pmatrix} \Sigma_{11}(Z_i) & \Sigma_{12}(Z_i) \\ 0 & 0 \end{pmatrix} - f_z(z) \cdot vec\begin{pmatrix} \Sigma_{11}(z) & \Sigma_{12}(z) \\ 0 & 0 \end{pmatrix}.$$

By Lemma 23 and Lipschitz assumption on K_h, $\|A_{i,j(i)} - A_{i,i}\| = O_p(N^{-1})$. Also note that by simple calculation, we have

$$\mathbb{E}\left(K_h(Z_i - z)\begin{pmatrix} \Sigma_{11}(Z_i) & \Sigma_{12}(Z_i) \\ 0 & 0 \end{pmatrix}\right) = f_z(z)\begin{pmatrix} \Sigma_{11}(z) & \Sigma_{12}(z) \\ 0 & 0 \end{pmatrix} + O(h^2),$$

which implies that $\mathbb{E}(A_{i,i}) = O(h^2)$. Therefore,

$$\begin{aligned}
&Var\left(\mathbb{E}(VI_{10} \mid Z^{1,2})\right) \\
&= \mathbb{E}\left((\frac{1}{n}\sum_{i=1}^{n} A_{i,j(i)})(\frac{1}{n}\sum_{i=1}^{n} A_{i,j(i)})^T\right) \\
&= \frac{1}{\sqrt{n}}\left(\frac{m}{n}\right)^{\frac{3}{2}} \mathbb{E}\left(\frac{1}{\sqrt{m}}(\max_{j=1,\ldots,m} C(j))^2 A_{(i,i)} A_{(i,i)}^T\right) + O(h^4) + o(1)^2.
\end{aligned}$$

Note that under the Lipschitz assumption on K_h and Σ, $Var\left(\mathbb{E}(VI_{10} \mid Z^{1,2})\right)$ goes to zero matrix as sample sizes go to infinity. Therefore we have demonstrated that I_{10} converges in probability to $f(z)\begin{pmatrix} \Sigma_{11}(z) & \Sigma_{12}(z) \\ 0 & 0 \end{pmatrix}$.

C.4 Boundedness of $V(z)$

The following shows that the expectation of $V(z)$ is finite. The result follows Lemma 3 in Abadie and Imbens (2006).

$$V(z) = \sum_{j=1}^{m} \sigma^2 (C(j) K_h(Z_j - z))^2 \Omega(Z_j) \bigotimes \begin{pmatrix} 1 & (Z_j - z)/h \\ (Z_j - z)/h & ((Z_j - z)/h)^2 \end{pmatrix}.$$

From Lemma 3 in Abadie and Imbens (2006), $C(j)$ is $O(1)$, $j \in \{1, ..., m\}$. By Assumptions 3–4, the kernel functions we consider are Lipschitz continuous on a compact set, and $\nu_j = \int u^j K^2(u) du, j = 1, 2, 3$ are also bounded, therefore $V(z)$ is also bounded.