# LASSO-type Regularization in
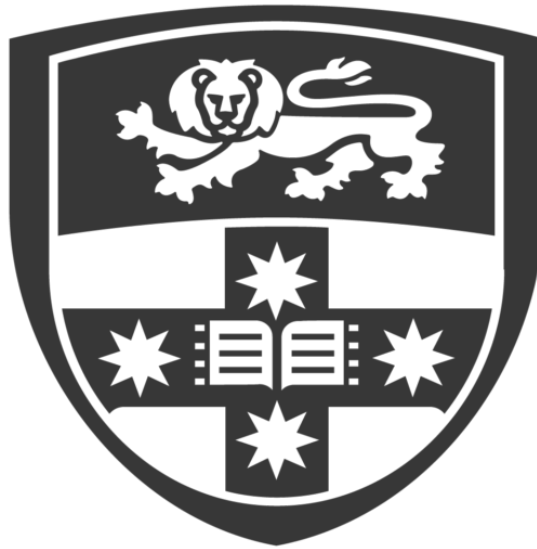
# Bernstein Copula and Its Application in Finance

Aixi Zhang

Business School

The University of Sydney

Supervised by

Professor Artem Prokhorov, Professor Peter Radchenko

A thesis submitted in fulfillment of the requirements for the degree of

Doctor of Philosophy

2023

# ABSTRACT

Copulas are receiving increasing attention due to their flexibility in modelling dependence between variables. Parametric copula models assume a specific form for the copula function, such as the Gaussian copula, and estimate the parameters from data. However, in many cases, the true underlying copula function may not be well-represented by the chosen parametric form, leading to inaccurate results. The data-driven nonparametric framework helps.

In this thesis, we propose a nonparametric Bernstein copula model to address this issue, which is flexible and capable of approximating any copulas arbitrarily well. Our proposed model is penalized with the reciprocal of the empirical Bernstein copula as the weight of adaptive LASSO. To overcome the ill-posed problem for nonparametric copula estimation, we utilize the Bernstein polynomial sieves as the sieve space and estimate it through the sieve maximum likelihood(SML). Extensive Monte Carlo simulations are carried out to evaluate the performance of the proposed model under different scenarios. We also perform an empirical analysis and successfully capture the financial contagion between four major markets.

Nonparametric density estimation provides a lot of flexibility in modelling complex data patterns and is particularly useful when the true distribution is unknown or difficult to model parametrically. However, in some cases, the data may follow a specific parametric distribution. For example, in non-life insurance, claim frequency and claim severity data are often modelled using specific parametric distributions such as the Poisson distribution and the Gamma distribution, respectively. Estimating their density based on their own parametric function is

accurate and efficient. Based on our nonparametric Bernstein copula model, we construct a semiparametric approach to improve the univariate density estimation. The copula model is employed to measure the joint density between the target variables and related ones to provide information in individual estimation. A three-step double selection method is developed to study the performance of the individual density estimation. We compare the univariate estimation method and copula-based methods using both cross-sectional data and time series one. The results show that the copula-based model can truly provide accuracy in estimating margins, but it is irrelevant to the linear dependency between margins.

According to the complexity of financial products significantly increasing, the copula is becoming a popular tool to study the interdependence in asset pricing. In the context of option pricing, it is often assumed that the copula function under the risk-neutral measure is from the same family as that under the physical measure. Sometimes, it can misspecify the true dependence structures. We conduct a nonparametric copula-based GJR-GARCH approach with GED innovation to evaluate the European bivariate rainbow options. Since the traditional local risk-neutral valuation relationship which connects the physical asset with the risk-neutral one is under the conditional normality, we modify it to accommodate the skewness and fat-tail properties of financial data with explicit transformation formula. Since the bivariate rainbow options are typically traded over the counter, price data is not available. Six pairs of price indices from both developed and emerging markets with different dependence structures are designed as the underlying assets to observe the copula's effect on option pricing.

# Acknowledgments

First, I would like to express my sincere gratitude to my main supervisor, Professor Artem Prokhorov, who guides me to the world of econometrics, takes me to the top conferences, helps me to get the scholarship, leads me to learn advanced quantitative courses, provides me informative and knowledgeable suggestions on research topics, answers all my concerns and questions anytime anywhere. I'm deeply impressed by his broad knowledge base, hard-working attitude, and motivational leadership. Without his enlightening guidance and consistent help, the thesis cannot be completed thoroughly. I would like to give my ultimate thank to my co-supervisor, Professor Peter Radchenko, who is warmhearted and has the magic to make complicated mathematics intuitively understandable.

I wish to show my appreciation to the University of Sydney Business School for their accepting me as a Ph.D. candidate and providing me with financial support, academic resources, and a comfortable office environment so that I can absorbedly concentrate on my research.

I would like to thank A/Prof Dmytro Matsypura, A/Prof Boris Choy, Dr Marcel Scharth, Dr Chao Wang for their detailed and valuable comments on my thesis proposal. I would like to thank Prof. Uri Gal and Prof. Barney Tan for helping me understand the philosophy and qualitative research. I would like to thank A/Prof Boris Choy, A/Prof Andrey Vasnev, and Prof. Richard Gerlach for providing teaching opportunities for me. I'd like to thank Prof. Junbin Gao and A/Prof. Minh-Ngoc Tran organized interesting and useful seminars every week during COVID and I'd like to thank Darae Jung, Colleen Liu, Jessica Ta, Ben

Breeds, and Mei Su who are always willing to help.

I would like to thank my colleagues and friends, Tina, Vivian, Mingyuan, Stella, Ethan, Olivia, Zhou, Bei, He, and Juan who have provided encouragement and happiness throughout my Ph.D. journey.

Pursuing a Ph.D. is a long road that I met too many people who deserve my thank and gratitude. Most importantly, I would highlight that without the support and external love from my parents, in-laws, husband, godparents, and grandma, it would be not possible for me to reach this milestone of my life.

# STATEMENT OF ORIGINAL AUTHORSHIP

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes. I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged

Aixi Zhang

January, 2023

# Contents

# List of Figures

# List of Tables

# CHAPTER 1

## Introduction

Dependence structures, describing the interdependence between random variables, are fundamental features of risk management. The recent global financial crisis of 2008, from the housing collapse to the downfall of a large scale of institutions, has had a ripple effect around the globe. The misspecification of the dependence structures between financial time series(Scheffer and Weiß, 2017) is the main reason for the unexpected contagion. The application of probability theory and statistical tools which capture the dependence structure have flourished in the financial and economic field(Torri et al., 2018). One of the most popular methods to capture a dependence structure is by using a correlation-based model. It is the canonical measure in the world of spherical and elliptical distributions. However, the empirical study in the finance field shows that the distributions of the real world seldom belong to this class (Embrechts et al., 2002).

To capture the dependence structure of the real world, an alternative method based on copulas garners particular attention in research. According to the literature, copula is well recognized as a flexible approach to modeling the joint behavior between random variables. It allows us to separately model the dependence structure and marginal behaviors by Sklar's theory (Sklar, 1959). Besides, the copula

function contains the information of both the degree and the structure of the dependence (Naifar, 2011). It has the ability to precisely capture the asymmetric and nonlinear dependence in financial data (Poon et al., 2004). The flexibility of copula estimation and the accuracy of its performance make it an attractive statistical tool in estimating dependency.

Estimation of copula function has been studied in several ways, including parametric (Malevergne and Sornette, 2003; Naifar, 2011; Genest et al., 1995; Chen et al., 2006) and nonparametric approaches (Deheuvels, 1979; Sancetta and Satchell, 2004). The parametric approach of estimating a copula model employs a parametric copula function, such as elliptical and Archimedean copula. It can be simply estimated through measuring a few key parameters of the specific models. The parametric method shows high efficiency and accuracy when the copula function is correctly specified. However, selecting a proper model for real data is nontrivial. And most of the parametric copula formulas exhibit a certain degree of symmetry or constraint by a fixed correlation structure, which is not always proper in modeling the dependence in the finance market. Nonparametric estimation of copulas and their densities does not assume a specific parametric form for the copula and the marginals, so it provides great flexibility and generality. This thesis relaxes all the restrictions on copula and margins and proposes a purely nonparametric model.

Nonparametric copula has been studied in several ways, such as kernel-based (Härdle, 1990), B-spline based(Shen et al., 2008; Eddie et al., 2018), infinite mixture model (Burda and Prokhorov, 2014). Noteworthily, the Bernstein copula is distinguishable among all the other nonparametric models. It has the property to represent any copula under regularity conditions (Sancetta and Satchell, 2004), which implies any continuous density function can be approximated by the density

coming from the Bernstein copula. Besides, during the process of converging to a specific dependence structure, the Bernstein copula shows a higher rate of consistency compared with other nonparametric models (Diers et al., 2012). Moreover, the Bernstein copula has an explicit formula that facilitates the data generation process in the Monte Carlo simulation study.

Nevertheless, the nonparametric multivariate model's estimation naturally leads to some major challenges: increasing computation burden, the risk of overfitting, and the curse of dimensionality. Regularization is a process of adding information to evade overfitting and reduce the number of parameters to be estimated (Bühlmann and Van De Geer, 2011). For example, Wahba (1990) and Ruppert et al. (2003) showed that a linear combination of basis functions in regression with coefficients that are regularized is an effective way to estimate a smooth unknown function. To tackle the problem of overfitting and computationally infeasibility, we propose the penalized Bernstein copula via the adaptive lasso penalty term. The adaptive LASSO inherits the excellent properties of LASSO and also reduces the bias of LASSO by using the weighted penalty approach (Zou, 2006). Besides, the weights of the adaptive lasso are set as reciprocals of the empirical copula so that the coefficient of the Bernstein copula can be shrunk to zeros if the corresponding grid cell contains few data points. In other words, during this process, a sparse model can be imposed. Sparsity enables faster and simpler processing of the data of interest since few coefficients reveal all the meaningful information.

The accurate estimation of copula function can provide more precision in the marginal distribution estimation. For the whole copula function estimation, semiparametric which set marginal distribution as nonparametric mode and copula as parametric is one of the most popular ones (Chen et al., 2006; Chen and Fan, 2006). Their research focuses on the dependence aspect. However, the individual

density measurement is still valuable in the application of microeconomics and actuarial studies. Thus, an inverse semiparametric model is studied.

Copula applications are presented in the concept of derivative pricing (Van den Goorbergh et al., 2005; Hull and White, 2006; Zhang and Guegan, 2008; Choe and Jang, 2011; Wang et al., 2015). Most of them are based on a parametric approach. The parametric copula model can be easily estimated and it can capture the dynamic information in time series data. But for the risk-neutral related asset pricing, the assumption: 'the copula model under physical and risk-neutral should belong to the same family' can be relaxed if a nonparametric approach is involved. Our nonparametric approach is studied in a bivariate option pricing context.

## 1.1 Thesis contribution

In this thesis, three significant contributions are stated:

1. Due to the complexity of financial data, the existing methods grounded on the parametric copula model are not competent for all possible cases. We address this problem by proposing a nonparametric Bernstein copula model, which is flexible and capable of approximating any copulas arbitrarily well. Our proposed model is penalized with the reciprocal of empirical Bernstein copula as the weight of adaptive LASSO. As the nonparametric model involves unknown parameters in infinite space, it can be computationally difficult to estimate based on a finite sample set. To overcome the ill-posed problem for nonparametric copula estimation, we utilize the Bernstein polynomial sieves as the sieve space and estimate it through sieve maximum likelihood. The estimation process reduces the nonparametric problem into a 'parametric' one. Besides, as the sample size increases, the optimal order of the empirical Bernstein copula also rises. It is prone to suffering from over-

fitting. When the Bernstein copula applies to a relatively higher dimension, the sample data tend to be much sparser. To improve prediction accuracy and efficiency, the adapted lasso regularization method is employed in our estimation. Extensive Monte Carlo simulations are carried out to assess the performance of the proposed model under different situations (the underlying data is Sparse/Less Sparse). As comparisons, our model outperforms the empirical Bernstein copula model(Sancetta and Satchell, 2004), penalized Bernstein copula model with LASSO, and penalized Bernstein copula with different weights of adaptive LASSO. We also perform an empirical analysis and successfully capture the financial contagion between four major markets.

2. Although the nonparametric estimation provides a lot of flexibility in density estimation, there are some kinds of data following specific parametric distribution, such as the claim frequency and claim severity of non-life insurance. Estimating their density based on their own parametric function is accurate and efficient when it is correctly specified. We construct a semiparametric approach of copula model to improve the univariate density estimation. The copula model is employed to measure the joint density between the target variables and related ones to provide information in individual estimation. It implies that if the unknown dependence structure between variables can be captured precisely, more information can be provided to margins. We develop a three-step double selection method and utilize the model we construct in the first contribution as the copula function to study the performance of the individual density estimation. We compare three kinds of methods, including 'Univariate' that estimates the individual density independently, 'Copula-based Empirical' method that estimates the univariate density with nonparametric Bernstein copula simultaneously through sieve

maximum likelihood estimation(Ivan et al., 2021), and 'Copula-based Penalized' method which estimates the univariate density after the selection of copula. Extensive simulation studies of the comparisons are provided based on settings of the different univariate density functions and various copula families, using cross-sectional data and time series data. In addition, the comparison in the context of Value at Risk for individual log return has been studied based on the data simulating from multivariate GARCH models. The forecast evaluation test of Value at Risk is applied to measure their performance.

3. According to the complexity of financial products significantly increasing, the copula is becoming a popular tool in asset pricing. We conduct a nonparametric copula-based GJR-GARCH approach with GED innovation to evaluate the European bivariate rainbow options. The GJR-GARCH model is utilized to capture the asymmetry and conditional volatility. The co-movement of residuals in the GARCH model is captured by the nonparametric model. Since the traditional local risk-neutral valuation relationship which connects the physical asset with the risk-neutral one is under the conditional normality, we relax it to accommodate the skewness and fat-tail properties of financial data. The generalized risk-neutral valuation relationship with explicit formula has been derived. The existing parametric copula models have been selected to compare under both static and dynamic scenarios for option pricing.

## 1.2  Thesis organization

The rest of the thesis is organized as followed. Chapter 6 reviews the literature about previous research on methods of nonparametric copula estimations and their application in financial markets. Section Chapter 3 summarizes some fundamental

theories about copula model and introduces the Bernstein copula with its definition, large sample properties, and the estimation method related to it. Chapter 4 explains the three steps of the double selection method and conducts various simulations to study its performance. Moreover, in Chapter 5, we apply the non-parametric Bernstein copula model in the real world and show its merits in options pricing.

# CHAPTER 2

## Literature Review of Nonparametric Copula Estimation

Copula represents a multivariate distribution as a function of univariate marginals. It measures the dependence structure separately from the marginal behaviors. The central Sklar theorem states that if $H$ is the multivariate distribution of a vector random variable $\mathbf{X} = [X_1, ..., X_d]'$ with marginal distributions $F_1, ...F_d$, there exist a n-dimensional copula: $C : [0, 1]^d \rightarrow [0, 1]$ such that

$$H(\mathbf{x}) = C(F_1(x_1), ..., F_n(x_d)), \quad \forall \mathbf{x} \in \mathbb{R}^d \tag{2.0.1}$$

and when $H$ is a continuous d-variate distribution function with univariate marginal distribution $F_1, ..., F_d$ and quantile functions $F_1^{-1}, ..., F_d^{-1}$, the copula function C is unique (Sklar, 1959):

$$C(\mathbf{u}) = H(F_1^{-1}(u_1), ..., F_d^{-1}(u_d)), \quad \mathbf{u} \in [0, 1]^d \tag{2.0.2}$$

A very powerful property of copula is its invariance concerning rank-preserving transformations of the random variables (Embrechts et al., 2002). The invariance theorem indicates that the full dependence between random variables can be captured by the copula with any shape of the marginal distributions.

Since the copula is introduced to develop the theory of probabilistic metric spaces, it stays in a purely mathematical structure for a long period. After Embrechts et al. (1999) was among the first to introduce it to finance literature, the copula model has become a widely used tool for modeling dependence for a large spectrum of financial problems: multi-asset pricing (Van den Goorbergh et al., 2005; Cherubini and Luciano, 2002), credit portfolio modeling (Glasserman et al., 2002), risk management (Kole et al., 2007), etc.

## 2.1   Estimation of copula functions

For any multivariate distribution $H$, the marginal distribution $F_i$ and the copula function $C$ can be extracted. Assume that the joint distribution $H$ is n-times differentiable, then taking the $d^{th}$ cross-partial derivative of equation 2.0.1:

$$h(\mathbf{x}) = \frac{\partial^d}{\partial x_1...\partial x_d} H(\mathbf{x})$$

$$= c(F_1(x_1), F_2(x_2), ..., F_d(x_d)) \prod_{i=1}^{d} f_i(x_i) \qquad (2.1.1)$$

where $f_i$'s are the marginal densities and c is the copula density function. Then, the following decomposition for the log-likelihood function $L = \ln[h(\mathbf{x})]$ of a random sample of matrix $\mathbf{x} = \{x_{1n}, ..., x_{dn}\}_{n=1}^{N}$ applies:

$$L = \sum_{n=1}^{N} \sum_{i=1}^{d} \ln f_i(x_{in}) + \sum_{n=1}^{N} \ln(c(F_1(x_{1n}), F_2(x_{2n}), ..., F_d(x_{dn}))) \qquad (2.1.2)$$

According to the function 2.1.2, the estimation of the joint density function involves several underlying parts: marginal cumulative distribution functions and a copula function. Considering the different settings on marginal distribution $f$ and copula function $c$, recent studies on copula estimations can be divided into three types: parametric, semi-parametric, and nonparametric estimation.

The parametric approach has been mostly studied, which assumes the copula function $c$ and marginal distribution $f$ belonging to a specific model. In this situation, the copula function can be simply estimated by measuring a few key parameters of the specific models by holding their assumptions. Because of its simplicity, this approach is widely applied in practice. Malevergne and Sornette (2003) indicates that some pairs of currencies and pairs of major stocks are compatible with the Gaussian copula. Naifar (2011) applies the Archimedean copula functions in exploring the dependence structure between stock market conditions and the default risk premium and successfully capture the asymmetric property. Once the parametric model is specified correctly, it shows high efficiency in estimation (Genest et al., 1995; Chen et al., 2006; Ibragimov and Prokhorov, 2017).

One step of modeling parametric copulas is estimating the parameter vector. Most studies concentrate on the fully parametric standard maximum likelihood estimator (MLE) by maximizing the log-likelihood function (Weiß, 2011; Kim et al., 2007):

$$(\hat{\beta}_1^{MLE}, ..., \hat{\beta}_d^{MLE}, \hat{\boldsymbol{\theta}}^{MLE}) = \arg \max_{\boldsymbol{\theta}, \beta_1,...,\beta_d} L(\beta_1, ..., \beta_d, \boldsymbol{\theta})$$

$$= \arg \max_{\boldsymbol{\theta}, \beta_1,...,\beta_d} (\sum_{n=1}^{N} \sum_{i=1}^{d} \ln f_i(x_{in}; \beta_i) + \sum_{n=1}^{N} \ln(c(F_1(x_{1n}; \beta_1), F_2(x_{2n}; \beta_2), ..., F_d(x_{dn}; \beta_d); \boldsymbol{\theta})))$$

$$(2.1.3)$$

over the unknown parameters $(\beta_1, \beta_2, ..., \beta_d, \theta)$ simultaneously. The ML-estimator performs relatively stable over the parameter interval of the respective parametric copula(Weiß, 2011). However, a multi-parameter numerical optimization in the multivariate model for MLE sometimes is infeasible or time-consuming (Joe, 2005), especially in the high-dimensional case.

Inference functions for margins (IFM) method is a method used for estimating the parameters of a statistical model based on the marginal distributions of the variables involved. As discussed by Joe and Xu (1996), it offers a straightforward approach at the expense of efficiency. It separates the whole estimation procedure into two steps: firstly the marginal density parameters $(\beta_1, ..., \beta_d)$ are estimated by performing univariate marginal distribution estimation

$$\left( \hat{\beta_1}^{IFM}, \hat{\beta_2}^{IFM}, ..., \hat{\beta_d}^{IFM} \right) = \arg \max_{\beta_1,...,\beta_d} \sum_{n=1}^{N} \sum_{i=1}^{d} \ln(f_i(x_{in}; \beta_i)) \qquad (2.1.4)$$

2.1.4 is the quasi-MLE (QMLE) estimator assuming independence between the marginals (independence over i and n). The QMLE remains consistent if there is dependence over i, but it is not efficient. Prokhorov and Schmidt (2009$b$) develop the improved QMLE (IQMLE) with stacked moment conditions. The optimal weights for individual moment functions are determined by the GMM. They show the efficiency gain is positive in comparison with the QMLE estimator if there is dependence over i and the marginal scores are not "equicorrelated".

Then, given $(\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_d)$ the parameter of copula $\theta$ will be evaluated.

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \sum_{n=1}^{N} \ln c(F_1(x_{1n}; \hat{\beta}_1), F_2(x_{2n}; \hat{\beta}_2), ..., F_d(x_{dn}; \hat{\beta}_d); \boldsymbol{\theta}) \qquad (2.1.5)$$

Under standard regularity conditions, the IFM estimator is consistent and asymptotically normally distributed, let $\hat{\boldsymbol{\theta}}^{IFM} = (\hat{\beta}_1^{IFM}, \hat{\beta}_2^{IFM}, ..., \hat{\beta}_d^{IFM}, \hat{\theta})$, $\boldsymbol{\theta}_0$ as the true parameters and $G(\theta)$ as the Godambe information matrix

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \to N(0, G^{-1}(\boldsymbol{\theta}_0)) \qquad (2.1.6)$$

Joe (2005) studies the IFM estimator in the case of extreme dependence, independence, and Fréchet bound with continuous margins. The results show that IFM has good efficiency (nearly as efficient as MLE) only when the margins are normally distributed and the copula function is also Gaussian. This equivalence does not hold in general. Prokhorov and Schmidt (2009a) indicate that the one-step method is equally or more efficient than the two-step method. However, both the full MLE and IFM are inconsistent and biased if the marginals or the copula function are misspecified. Besides, for a lot of dynamic models(e.g. stochastic volatility models), the corresponding margins cannot be written explicitly.

Considering the difficulty in precisely selecting margins, some researchers focus on the semi-parametric approach, which employs a parametric copula while leaving the marginals nonparametric. Similar to IFM procedures, the univariate margins $F_i$ are estimated non-parametrically in the first stage and then the parameters from copula are measured by using maximization of the contribution to the log-likelihood function (Genest et al., 1995):

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \sum_{i=1}^{n} \ln c(\hat{F}_1(x_{1i}), \hat{F}_2(x_{2i}), ..., \hat{F}_d(x_{di}); \boldsymbol{\theta}) \quad (2.1.7)$$

Genest et al. (1995) also shows that the resulting semi-parametric estimator $\hat{\theta}$ is consistent and asymptotically normally distributed under proper regularity conditions and $\hat{\theta}$ is fully efficient at independence under additional copula regularity assumptions which are satisfied by a large number of copula families (e.g. bivariate Gaussian, Eyraud-Farlie-Gumbel-Morgenstern(EFGM), Clayton, Frank). The main aim of semi-parametric estimation is to avoid possible misspecification of margins, but there still exists the risk of copula misspecification. The limitation of the parametric copula function mainly comes from its few types of distributional shapes (Kim et al., 2007). The robust copula is the incorrectly specified copula

with the expected value of copula scores zero. Prokhorov and Schmidt (2009*b*) states that robust parametric copulas exist, which means that the pseudo-MLE (PMLE) can provide consistent estimation under misspecification of parametric copula function. However, they also state that the robustness of an incorrect copula is determined by the nature of the model. A more flexible model is needed to study. A nonparametric copula estimation approach treats both the copula function and margins parameter-free, which provides the greatest flexibility and generality. In this thesis, a purely nonparametric framework for the copula model is proposed.

## 2.2 Nonparametric approaches for copula estimations

Nonparametric estimations of copula functions offer great generality. Different from estimating marginal distributions which are observable, the copula function is a hidden dependence structure. Proposing a suitable model for the copula function is nontrivial. The nonparametric estimator offers a sophisticated alternative in constructing initial information for revealing an underlying parametric model. The estimation of nonparametric copula can be dated back to Deheuvels (1979) proposes the copula estimator based on empirical distribution and constructs various nonparametric tests of independence. Let $(X_1, X_2)$ be the bivariate random vector with joint cumulative density function $H(x_1, x_2)$ and continuous marginal distribution $F_1(x_1)$ and $F_2(x_2)$ respectively. Based on the bivariate data series $(X_{11}, X_{21}), ..., (X_{1n}, X_{2n})$, the empirical distribution function

$$\mathbb{H}_N(x_1, x_2) = \frac{1}{N} \sum_{n=1}^{N} \mathbb{I}_{\{X_{1n} \leq x_1, X_{2n} \leq x_2\}}, \quad -\infty < x_1, x_2 < +\infty$$

where $\mathbb{I}_{\{A\}}$ denotes the indicator of a set A and the marginal distributions of $X_1$ and $X_2$ are

$$\mathbb{F}_{1N}(x_1) = \mathbb{H}_N(x_1, +\infty) = \frac{1}{N} \sum_{n=1}^{N} \mathbb{I}_{\{X_{1n} \leq x_1\}}$$

, and

$$\mathbb{F}_{2N}(x_2) = \mathbb{H}_N(x_2, +\infty) = \frac{1}{N} \sum_{n=1}^{N} \mathbb{I}_{\{X_{2n} \leq x_2\}}$$

Since the corresponding copula is defined as $\mathbb{C}_N(u_1, u_2) = \mathbb{H}_N(\mathbb{F}_{1N}^{-1}(u_1)), \mathbb{F}_{2N}^{-1}(u_2))$, the empirical copula function $\mathbb{C}_N$ can be

$$\mathbb{C}_N(u_1, u_2) = \frac{1}{N} \sum_{n=1}^{N} \mathbb{I}_{\{\mathbb{U}_{1n} \leq u_1, \mathbb{U}_{2n} \leq u_2\}}. \tag{2.2.1}$$

where $(\mathbb{U}_{1n}, \mathbb{U}_{2n}) = (\mathbb{F}_{1N}(X_{1n}), \mathbb{F}_{2N}(X_{2n}))$. The empirical estimator 2.2.1 is asymptotically equivalent to the estimator directly based on Sklar's theorem with a term $\mathcal{O}(n^{-1})$. Deheuvels (1980) studies the weak convergence of the empirical copula process in the case of independent marginal distributions for i.i.d random variables. Gaenssler and Stute (1987) considers the weak convergence of the empirical copula process in Skorokhod space $\mathcal{D}([0,1]^2)$ using ordinary empirical process techniques. Vaart and Wellner (1996) employ the functional delta method to show the weak convergence of the ordinary empirical copula process $\mathbb{Z}_n$ ($\mathbb{Z}_n(u_1, u_2) = \sqrt{n}(\mathbb{C}_n - C)(u_1, u_2)$) in $l^\infty((0,1)^2)$, for some $0 < u_1 < u_2 < 1$. Fermanian et al. (2004) conclude the empirical copula process $\mathbb{Z}_n$ converges weakly to the Gaussian process $\{\mathbb{G}_C(u_1, u_2), 0 \leq u_1, u_2 \leq 1\}$ in $\ell^\infty([0,1]^2)$ when its copula function has continuous partial derivatives. Nevertheless, the empirical copula model shows high discontinuity.

Some smoothed estimators are studied in a nonparametric framework. It has been extensively discussed in the kernel, spline, and mixture models.

### 2.2.1  Kernel-based estimator

In the world of statistics, many nonparametric estimation approaches are based on a kernel structure (Härdle, 1990). Several studies are proposed for the kernel-based copula function and its density(Omelka et al., 2009; Chen and Huang, 2007; Geenens et al., 2017). We will review different approaches to estimate the copula density function based on the kernel estimator. Assume $(U_{1n}, U_{2n})$ are i.i.d observations from a bivariate copula C and its corresponding density function is $c(u_1, u_2)$. The standard kernel-based estimator of the copula density using diagonal bandwidth can be expressed as (Wand and Jones, 1994)

$$\hat{c}_N(u_1, u_2) = \frac{1}{Nb_N^2} \sum_{n=1}^{N} K\left(\frac{u_1 - U_{1n}}{b_N}\right) K\left(\frac{u_2 - U_{2n}}{b_N}\right), \quad (u, v) \in [0, 1]^2, \qquad (2.2.2)$$

where $K : \mathbb{R}^2 \to \mathbb{R}$ is a kernel function which is symmetric and $\int K = 1$. $b_N > 0$ is the smoothing parameter. This estimator is not consistent on the boundaries of $[0, 1]^2$ in which $\hat{c}_N(u_1, u_2)$ has a jump. Such bias is significant near the boundaries, depending on the size of the bandwidth.

To cope with this boundary bias, several improved kernel-based copula models are constructed e.g. the method using mirror reflection, transformation, or boundary kernels. The mirror-reflection method in the bivariate case consists in reflecting each data point with respect to all edges and corners of the unit square. The kernel-based copula estimator is then built on the augmented data set(Gijbels and Mielniczuk, 1990).

$$\hat{c}_N^{MR}(u_1, u_2) = \frac{1}{Nb_N^2} \sum_{n=1}^{N} \sum_{k=1}^{9} K\left(\frac{u_1 - \tilde{U}_{1nk}}{b_N}\right) K\left(\frac{u_2 - \tilde{U}_{2nk}}{b_N}\right), \quad (u_1, u_2) \in [0, 1]^2,$$

where $\{(\tilde{U}_{1nk}, \tilde{U}_{2nk}) \quad n = 1, ..., N; \quad k = 1, ..., 9\} = \{(\pm U_{1n}, \pm U_{2n}), (\pm U_{1n}, 2 - U_{2n}), (2 - U_{1n}, \pm U_{2n}), (2 - U_{1n}, 2 - U_{2n}) \quad n = 1, ..., N\}$. Gijbels and Mielniczuk (1990) show strong consistency and asymptotic normality of this estimator. Nagler (2014) show that the asymptotic bias of $\hat{c}_N^{MR}$ is determined by the second order partial derivatives of the true copula density $c(u_1, u_2)$ after selecting proper kernel K and it corresponding bandwidth $b_N$. However, the second-order partial derivatives are not bounded near the corners of the unit square for many popular parametric copula families, which results in unbounded bias if the mirror-reflect estimator is applied (Omelka et al., 2009).

The transformation approach is first proposed by Marron and Ruppert (1994) to tackle the boundary bias of kernel density estimation. Charpentier et al. (2007) applied it in the estimation of copula density. The data is transformed first so that it has full support on $\mathbb{R}^2$ rather on the unit square, which is $\tilde{X}_1 = G^{-1}(U_1)$ and $\tilde{X}_2 = G^{-1}(U_2)$. $G$ is a continuous distribution function on $\mathbb{R}$ with differentiable strictly positive density $g$. Next, applying the standard kernel estimator to the transformed data can be described as

$$\hat{h}_N(x_1, x_2) = \frac{1}{Nb_N^2} \sum_{n=1}^{N} K\left(\frac{x_1 - \tilde{X}_{1n}}{b_N}, \frac{x_2 - \tilde{X}_{2n}}{b_N}\right).$$

Finally, based on the back transformation, the estimated copula density can be obtained from 2.1.1,

$$\hat{c}_N^T(u_1, u_2) = \frac{1}{Nb_N^2 g(G^{-1}(u_1))g(G^{-1}(u_2))}$$
$$\times \sum_{n=1}^{N} K\left(\frac{G^{-1}(u_1) - G^{-1}(U_{1n})}{b_N}, \frac{G^{-1}(u_2) - G^{-1}(U_{2n})}{b_N}\right), \; (u_1, u_2) \in [0, 1]^2.$$

Charpentier et al. (2007) also mention that this approach can be extended by considering different cumulative density function $G_{X_1}$ and $G_{X_2}$, different kernels

16

or different bandwidth for the marginals. In particular, Geenens et al. (2017) define the probit transformation of the marginals $(\Phi^{-1}(U_1), \Phi^{-1}(U_2))$ and show that combining with the local likelihood method, this kernel-type copula density estimator is easy to implement without boundary issues and able to manage the unbounded copula densities. They also show this estimator of copula density is uniformly consistent on any compact proper subset of $[0,1]^2$.

The boundary kernels approach is to use the matched kernels for the support of the target density. It includes a smoothing distortion near the boundary such that the kernels' shape and corresponding bandwidth can be adjusted. Beta kernels are one example of boundary kernels. Chen (1999) first introduces the beta kernel as a density estimator.

$$f_b(x) = \frac{1}{N} \sum_{n=1}^{N} K(X_n, \frac{x}{b} + 1, \frac{1-x}{b} + 1),$$

where $K(\cdot, \frac{x}{b} + 1, \frac{1-x}{b} + 1)$ stands for the density of the beta distribution with parameter $\frac{x}{b} + 1$ and $\frac{1-x}{b} + 1$,

$$K(x, \frac{x}{b} + 1, \frac{1-x}{b} + 1) = \frac{x^{\frac{x}{b}+1}(1-x)^{\frac{1-x}{b}+1}}{B(\frac{x}{b} + 1, \frac{1-x}{b} + 1)}.$$

$B(\cdot, \cdot)$ denotes the beta function and $b$ is the smoothing bandwidth. They also show the bias is of $\mathcal{O}(b)$ throughout $[0,1]$, which means that $f_b(x)$ is free of boundary bias. In addition, the asymptotic variances of the boundaries have comparably larger order $(N^{-1}b^{-1})$ than those of the interior $(N^{-1}b^{-1/2})$, but it has little effect on the mean integrated squared error(MISE) which is of order $\mathcal{O}(N^{-4/5})$. Bouezmarni and Rolin (2003) examine the asymptotics of the mean integrated absolute error (MIAE) for beta kernel estimator.Bouezmarni and Rolin (2003, 2007) prove that the beta kernel density estimator is consistent even the true density is un-

bounded on the boundaries. Bouezmarni and Rombouts (2010) extend the idea to the multivariate case. The multivariate kernel estimator is the product of univariate flexible and adapted kernels,

$$\hat{f}(x_1, .., x_d) = \frac{1}{N} \sum_{n=1}^{N} \prod_{s=1}^{d} K_B(b_s, X_{ns})(x_s),$$

where $b_s$ is the bandwidth parameter and $K_B$ is the beta kernel function $K_B(b, t)(x) = B(\frac{x}{b} + 1, \frac{1-x}{b} + 1)(t)$ and $B(\alpha, \beta)(t)$ is the beta density function with parameters $\alpha$ and $\beta$. When the multivariate density $f$ is uniform on $[0, 1]^d$, the nonparametric beta kernel estimator is unbiased for $f$ (Bouezmarni and Rombouts, 2010). Hirukawa et al. (2022) show the uniform consistency of the multivariate beta-kernel-based sample average estimator under weaker regularity conditions. The beta-kernel-based bivariate copula density function can be estimated by product beta kernels,

$$\hat{c}_N^B(u_1, u_2) = \frac{1}{Nb_N^2} \sum_{n=1}^{N} K(X_{1n}, \frac{u_1}{b_N} + 1, \frac{1-u_1}{b_N} + 1) \times K(X_{2n}, \frac{u_2}{b_N} + 1, \frac{1-u_2}{b_N} + 1),$$

$$(u_1, u_2) \in [0, 1]^2$$

The bias of $\hat{c}_N^B$ is of order $\mathcal{O}(b_N)$ and the variance of the estimator depends on the location, which is of order $\mathcal{O}(Nb_N^{\mathcal{K}})^{-1}$, $\mathcal{K} = 1$ in the interior of $[0, 1]^2$ and $\mathcal{K} = 3/2$ in borders (Charpentier et al., 2007; Nagler, 2014). Besides, $\hat{c}_N^B$ is asymptotically normal distributed,

$$\sqrt{Nb_N^{\mathcal{K}'}} \left[ \hat{c}_N^B(u_1, u_2) - c(u_1, u_2) \right] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma(u_1, u_2)^2)$$

where the value of $\mathcal{K}'$ determined by the location and $\sigma(u_1, u_2)^2$ is proportion to $c(u_1, u_2)$.

### 2.2.2 Spline estimator

A B-spline of order(m) greater than two is a polynomial-based smooth function on a given interval $[a, b]$. The places where the pieces overlap are knots. It is first proposed by Curry and Schoenberg (1966) and formally defined by De Boor et al. (1978). One property of B-splines is that any spline function on a given set of knots can be written as a linear combination of B-spline basis functions. Each basis function is a piecewise polynomial with degree $m - 1$ vanishing everywhere except on a small subinterval of $[a, b]$ and when the knots are distinct, the first $m - 2$ derivatives are continuous among each knot.

The B-spline basis $\phi_{1,m}, ..., \phi_{m+k,m}$ of order $m$ with knots $a < \tau_{m+1} < ... < \tau_{m+k} < b$ is constructed through the Cox-de Boor recursion formula(De Boor et al., 1978). Let $\tau_1 = ... = \tau_m = a$ and $\tau_{m+k} + 1 = ... = \tau_{2m+k} = b$. Then, we have for all $i \in \{1, ..., 2m + k - 1\}$, $j \in \{2, ..., m\}$, if $\phi_{i,1}$ is a indicator function of set $[\tau_i, \tau_{i+1})$ and $t \in [a, b]$,

$$\phi_{i,j}(t) = \frac{t - \tau_i}{\tau_{i+j-1}} \phi_{i,j-1}(t) + \frac{\tau_{i+j} - t}{\tau_{i+j} - \tau_{i+1}} \phi_{i+i,j-1}(t).$$

The first basis function $\phi_{1,m}$ is equal to 1 at point a and decreases monotonically to 0 as t goes from a to b. Homoplastically, the last basis function $\phi_{m+k,m}$ is equal to 0 at point a and increases monotonically to 1 as t goes from a to b. All the interior basis $\phi_{2,m}, ..., \phi_{m+k-1,m}$ is equal to 0 until it meets the knot after which, it will increase monotonically to a peak and fall back to 0, and it will remain 0 for the rest of interval.

Lambert (2007) applies the cubic B-spline to propose an estimator based on the Archimedean copula family. It uses a penalized smoothing B-spline to represent

the approximated ratio $\lambda(\cdot) = \frac{\phi(\cdot)}{\phi'(\cdot)}$ and then estimates the parameters via Markov chain Monte Carlo algorithm. Dimitrova et al. (2008) shows that although approximating the ratio estimator $\lambda(\cdot)$ is more convenient than using the spline estimator $\phi(\cdot)$ directly, it is complicated in simulation and inefficient when the data sample and parameter dimensions are large. Dimitrova et al. (2008) also states that if the copula estimation uses the spline directly of a fixed degree and set the coefficients and knots as unknown parameter $\theta$, the task of estimating copula density is a parametric approach with copula function:

$$C_\theta(u_1, u_2) = \phi_\theta^{-1}(\phi_\theta(u_1) + \phi_\theta(u_2)) \tag{2.2.3}$$

and with copula density function:

$$c_\theta(u_1, u_2) = \frac{\partial^2 C_\theta(u_1, u_2)}{\partial u_1 \partial u_2} = -\frac{\phi_\theta''(C_\theta(u_1, u_2))\phi_\theta'(u_1)\phi_\theta'(u_2)}{\phi_\theta'(C_\theta(u_1, u_2))^3} \tag{2.2.4}$$

Then a maximum likelihood approach could be applied to achieve asymptotic efficiency. But if using the spline estimator with free knots, the approximation approach is very difficult to implement numerically and is computationally infeasible since it leads to a multi-extrema-constrained non-linear optimization problem with the inversion of the spline generator.

A nonparametric approach of copula estimation using splines is proposed by Shen et al. (2008), which constructs a new class of copula model called linear B-spline copula. The new copula model overcomes the discontinuity problem of empirical copula and a comparably complicated copula can be uniformly approximated by the linear B-spline without losing essential properties. Kauermann et al. (2013) points out that considering the curse of dimensionality, this kind of nonparametric model is hardly feasible in high dimensions. Then, they use the B-splines to

model the copula density itself and replace the density with a linear combination of tensor products of univariate B-splines on [0,1]. To weaken the curse of dimensionality, they proposed reduced tensor products based on sparse grids. Anderson et al. (2020) illustrate that the uniformity condition of the marginal distributions is difficult to impose in nonparametric copula estimation. They construct a spline-based copula model with a triangular basis function, which enables certain constraints of conducting the uniformity to be imposed on the density surface.

### 2.2.3  Mixture model estimator

Considering the deficiency of representing true dependence using individual copula, the mixture copula model is introduced by Hu (2006) to take advantage of different shapes of copulas. They constructed the mixture copula model using the weighted average of several individual copulas and adjusting the weight to adapt different dependence estimations in the financial market:

$$C(u_1, u_2) = \sum_{i=1}^{n} \omega_i C_i(u_1, u_2), \quad u_2, u_2 \in [0, 1] \tag{2.2.5}$$

where $0 < \omega_i < 1$ and $\sum_{i=1}^{n} \omega_i = 1$. When the number of copulas $n$ goes to infinity, the mixture model change to a nonparametric framework.

Although the theory of an infinite mixture model is introduced by Ferguson (1973) and Antoniak (1974) in the 1970s, it becomes computationally feasible recently due to the development of the Markov Chain Monte Carlo(MCMC) method. The Bayesian infinite mixture model tends to be popular and increasing researchers start to study its properties and applications. Burda and Prokhorov (2014) propose a nonparametric univariate Gaussian mixture with a Bernstein polynomial copula as the link function under the Dirichlet process prior to constructing a stable dependence structure. Wu et al. (2015) construct systematic Bayesian non-

parametric procedures to measure any unconditional copula density function via an infinite mixture of Gaussian copula. Dalla Valle et al. (2018) extend the work of Wu et al. (2015) to estimate any conditional copula density under Dirichlet process priors. The infinite mixture models select an infinite number of parametric copulas as the mixture elements. However, the parametric copula is constrained by a specific shape, which means not all kinds of distribution shapes can be captured by even an infinite mixture model. Considering the computational feasibility, the research on mixture copula models can only choose a small number of parametric copulas as the base.

Bernstein copula can be considered a special case of the mixture estimator, which is based on the Bernstein polynomials. Bernstein polynomials have been proved to have the property of approximating any continuous function on the interval [0,1] by Lorentz (2013). The density is estimated by Bernstein polynomials introduced by Vitale (1975), and a modified version is applied by Gawronski and Stadtmüller (1981). Then, Tenbusch (1994) develops the estimation in the two-dimensional case and finds the mean squared error of the Bernstein polynomial is of the order $\mathcal{O}(N^{-2/3})$. Latter, Bouezmarni and Rolin (2007) show the consistency of Bernstein polynomial estimator with unbounded probability density functions.

The copula estimation employing Bernstein polynomial has first been studied by Li et al. (1997). They show that for any copula $C$:

$$\mathcal{B}_n(C)(u,v) = \sum_{i=1}^{n} \sum_{j=1}^{n} B_{i,n}(u) B_{j,n}(v) C(\frac{i}{n}, \frac{j}{n}) \qquad (2.2.6)$$

is also a copula and it converges to the copula $C$ with a strong sense. Sancetta and Satchell (2004) first define the Bernstein estimator for bounded copula densities with independent and identified distributed (i.i.d) data and study its consistency

22

in mean-square error. It defines the Bernstein copula function as:

$$\mathcal{C}(\mathbf{u}) = \sum_{j_1=0}^{J_1} \ldots \sum_{j_d=0}^{J_d} C(\frac{j_1}{J_1}, \ldots, \frac{j_d}{J_d}) \prod_{i=1}^{d} P_{j_i, J_i}(u_i) \tag{2.2.7}$$

with $\mathbf{u} = (u_1, \ldots, u_d) \in [0,1]^d$, $J_i$ is a smoothing parameter and $P_{j_i, J_i}(u_i)$ is the binomial distribution function. It has difficulty in applying the Bernstein copula in practice since the copula function $C$ is unknown. Sancetta and Satchell (2004) also provides a solution. They proposed a nonparametric Bernstein copula estimation using an empirical copula $C_n$ to substitute the copula function $C$. Based on the result of Sancetta and Satchell (2004), Janssen et al. (2012) investigate some asymptotic properties of the Bernstein copula estimator. In this research, they build the almost sure consistency and asymptotic normality of the Bernstein copula estimator.

The Bernstein copula has been widely applied in practice. In finance, Hurd et al. (2007) and Salmon et al. (2006) employ the Bernstein copula to capture the asymmetric dependence of foreign exchange rates, and then more accurate currency index options can be priced. Tavin (2015) applies the Bernstein copula to describe the absence of arbitrage opportunities in a market with multi-asset derivatives. Tavin (2018) constructs a set of dependence scenarios according to Bernstein copula by simulating random doubly stochastic matrices to compute hedging positions. In insurance, Diers et al. (2012) model the non-life insurance using Bernstein copula. It compares the Bernstein copula with another popular copula by fitting German flood, storm, and water claims data. It also shows the particular advantages of the Bernstein copula when the dependence structure is inhomogeneous, not highly correlated, and when the data used is sparse. Guo et al. (2017) measure the default risk of collateralized debt obligations based on the composite Bernstein copula(CBC), and it shows the merits of applying CBC

over other models through this empirical study.

Therefore, this thesis will use the Bernstein copula as the main model for several reasons. Firstly, according to the Weierstrass approximation theorem, the Bernstein copula density has the property to estimate any two-dimensional copula uniformly on $[0, 1]^2$. It implies that the density coming from the Bernstein copula can approximate any continuous bivariate density function (Dou et al., 2016). Secondly, the mathematical properties of the Bernstein copula are attractive, which will be detailed explained in the next section. Diers et al. (2012) indicates that when the Bernstein copula converges to an underlying dependence, it has a higher convergence rate than other nonparametric estimators. In addition, the Bernstein copula has an explicit formula, which facilitates the data generation process.

#### 2.2.3.1   Bernstein polynomial density estimation

The Bernstein polynomial used as a density estimator can be traced back to Vitale (1975) which introduces the estimation form as a linear combination of beta density with random coefficients. This form easily satisfies the non-negative and integrated-to-unity conditions required by a density function. Unlike the kernel estimator, it's also free from boundary bias. Tenbusch (1994) extended this idea to multi-dimensional densities estimation. Based on the approximation property of Bernstein polynomial, Petrone (1999) conduct a prior on the class of densities $(0, 1]$ with a full topology support. Ghosal et al. (2001) shows that the convergence rate of posterior distribution is nearly equal to the parametric rate $(\log n)/\sqrt{n}$ if the underlying distribution is the Bernstein density. Petrone and Wasserman (2002) study the posterior from a Bernstein prior and under the mild assumption, the posterior is weakly consistent for any distribution function $P_0$ on $[0, 1]$ with continuous and bounded Lebesgue density. They also show that under some stronger

assumptions on the Bernstein prior, the posterior is also Hellinger consistent.

## 2.2.4   The method of sieve

Since the nonparametric models involve unknown parameters in infinite-dimensional parameter spaces, it is computationally complicated to be estimated by finite samples. Moreover, optimizing a sample criterion over an infinite-dimensional space may lead to undesirable large sample properties, like inconsistency and/or slow rate of convergence (Chen, 2007). To remedy this issue, the method of sieves receives extensive attention.

The method of sieves was first introduced by Grenander (1981), which proves it to be a powerful technique in nonparametric estimation. Grenander (1981) states that the minimization problem can be a constraint in a subspace of the parameter space and then let the subspace grow with the sample size. It defines that the sequence of subspaces from which the estimator is obtained is called "sieve" and the resulting procedure is named as the method of sieves. An approximate sieve estimate, denote by $\hat{\theta}_n$, is defined as an approximate maximizer of $L_n(\theta)$ via $\Theta_n$, which is

$$L_n(\hat{\theta}_n) \geq \sup_{\theta \in \Theta_n} L_n(\theta) - \mathcal{O}(\epsilon_n^2) \qquad (2.2.8)$$

where $\epsilon_n \to 0$ as $n \to \infty$. The performance of the method of sieves relies on the sequence of sieves parameters. The sieves parameter should decrease to zero at a sufficiently slow rate, and this method shows consistency in the $L_1$ sense (Grenander, 1981).

The method of sieves was further developed by Geman and Hwang (1982) and Walter et al. (1984), which combines the method of sieves with maximum likelihood estimation. The maximum likelihood method has difficulty in applying to

the completely nonparametric estimation of a density function from an iid sample. Histogram is one famous example that uses the method of sieve to deal with non-parametric density estimation. Let $x_1, ..., x_n$ be an i.i.d sample from a continuous distribution with some unknown density function $\alpha_0(x)$. The maximum likelihood estimator for $\alpha_0$ is to optimize the likelihood function

$$\prod_{i=1}^{n} \alpha(x_i). \tag{2.2.9}$$

But the maximum of 2.2.9 cannot be reached within any of the natural parameter spaces for the nonparametric problem, which means that the traditional maximum likelihood method cannot be consistent in the nonparametric density estimation. A sieve which is a sequence of subsets of parameter space indexed by sample size can be defined as

$$S_{K_n} = \left\{ \alpha : \alpha \text{ is a probability density function which} \right.$$
$$\left. \text{is constant on} \left[ \frac{k-1}{K}, \frac{k}{K} \right), \; k = 0, \pm 1, \pm 2, ... \right\}$$

Let $S_{K_n}$ associated with maximum likelihood function 2.2.9, then it becomes

$$\max \prod_{i=1}^{n} \alpha(x_i) \quad \text{subject to} \quad \alpha \in S_{K_n}$$

and the solution to this function is

$$\hat{\alpha}(x) = \frac{K_n}{n} \# \left\{ x_i : \frac{k-1}{K_n} \leq x_i < \frac{k}{K_n} \right\} \quad \text{for } x \in \left[ \frac{k-1}{K_n}, \frac{k}{K_n} \right)$$

which is the histogram with bin width $1/K_n$. If the bin size $K_n$ grows slowly to infinity, the estimated $\alpha$ is consistent, $\int |\hat{\alpha}(x) - \alpha_0(x)| dx \to 0$ almost surely. Both Geman and Hwang (1982) and Walter et al. (1984) find that adding the method of sieves in nonparametric density estimation provides the possibility to

do nonparametric estimation using the maximum likelihood method. Wong and Severini (1991) and Wong (1992) construct the asymptotic efficiency of plug-in nonparametric maximum likelihood estimation of smooth functions. Shen et al. (1997) extends this method to the sieve MLE. Ghosal et al. (2001) measures the converge rate of sieve MLE for Bernstein polynomial densities' estimation. However, the study of nonparametric density estimation using sieve MLE is confined to the univariate case. Until Ivan et al. (2021), they apply the sieves method in multivariate density estimation. They employ the Bernstein polynomial sieves and show that the sieve MLE estimator improves over QMLE but does not suffer the drawbacks of the full MLE, which is almost as efficient as FMLE.

### 2.2.5   Model selection in density estimation

Regularization was first introduced to deal with the ill-posed inverse problem by Tikhonov (1943), from which an enormous amount of research in statistics studies various regularization methods in a wide spectrum of problems. One common situation that the regularization deal with is that the number of predictors exceeds the number of observations in a regression model. Consider a linear regression function

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{2.2.10}$$

where $\mathbf{Y}$ is the response variable with n observations, $\mathbf{X} = (X_1, ..., X_p)$ is an $n \times p$ predictor data matrix, $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^T$ is unknown regression coefficient and $\boldsymbol{\epsilon}$ is an n-vector independent and identically distributed random error with mean 0 and variance $\sigma^2$. If $p > n$, there are many solutions to the coefficients $\boldsymbol{\beta}$ and the model is not uniquely predictable. One regularization strategy to deal with this problem is to assume that the true number of the regression coefficients, which is the nonzero $\beta_i$, $i = 1, ..., p$, is small (Radchenko and James, 2010). Then, pushing any parameter weights connected to the negligible features to exactly zero can im-

prove estimation results. The sparse model, with many predictors set to zero, can be more interpretable and computationally feasible. More generally if the target function is sparse, enforcing the sparsity of the solution may be a way to avoid overfitting. Thus, a large amount of research explores an efficient and appropriate variable selection method through sparse-based regularization (Jenatton et al., 2011; Zhang et al., 2008).

$l_0$ regularization method comes up naturally in some classical model selections because of its accurate interpretation of best subset selection and attractive sampling properties (Barron et al., 1999). It is a measure to explain how complex is the regression function and how many variables are important. However, the best subset selection through $l_0$ regularization is as difficult as trying all possible subsets, especially when p is large. The most popular $l_1$ regularization method is LASSO (Tibshirani, 1996) since the regularization resulting from LASSO's $l_1$ penalty $\sum_{i=1}^{p} |\beta_i|$ leads to sparse solutions. The usual definition of sparseness for variable selection, as explained in Zhao and Yu (2006), is that a small number of regression coefficients $\beta_i$ are nonzero and all the nonzero coefficients are bounded away from zero uniformly at a certain rate. Besides, compared with the classical variable selection method, like the best subset selection, the LASSO has two advantages. Firstly, its selection process is based on continuous trajectories of regression coefficients as a function of the penalty, which ensures a more stable estimation. Besides, the LASSO is computationally feasible for high-dimension data (Osborne et al., 2000; Efron et al., 2004).

The merits of the LASSO estimator attract several researchers to study its consistency in the model selection process. For the low-dimensional case with a fixed number of predictors, Knight et al. (2000) show that the LASSO estimator is consistent when estimating the regression parameters $\beta_i$ and their limiting distri-

butions can have positive probability mass at 0 when the true value of the predictor is 0 under appropriate conditions. Then, Meinshausen et al. (2006) show that the LASSO is consistent even $p \to \infty$ at a rate faster than n, under a neighborhood stability condition on the design matrix and certain additional regularity conditions. Zhao and Yu (2006) organize the neighborhood stability condition as a strong irrepresentable condition in terms of the linear regression model. They prove that even when $p > \exp(n^a)$, $\exists \ a \in (0, 1)$, the LASSO is still consistent under that irrepresentable condition.

Although the LASSO has many excellent properties, it is a biased estimator and this bias does necessarily not go away as $n \to \infty$. There are numerous alternatives studied by recent researchers. It includes Smoothly Clipped Absolute Deviation(SCAD) (Fan and Li, 2001), the Elastic Net (Zou and Hastie, 2005), the adapted LASSO (Zou, 2006), the Dantzig selector (Candes et al., 2007), Variable inclusion and shrinkage algorithms(VISA) (Radchenko and James, 2008), DASSO (James et al., 2009). Among these methods, the adaptive LASSO inherits the outstanding properties and reduces the bias of LASSO by using a weighted penalty approach.

Inspired by the success of regularization for regression models, researchers start to study sparse regularization in nonparametric density estimation. It is known that the estimation of a nonparametric function with high dimension is feasible only if some regularity assumptions are satisfied (Devroye et al., 2013). Bunea et al. (2010) studies sparse density estimation via $l_1$ penalization (SPADES) using mixture models. They assume that the densities of true mixture components belong to a large known candidate density pool. Then, they select the components of mixture models by penalizing the mixing weights and show that SPADES can be effective only if the local distances of true mixture components are quite large

in comparison to their variances or covariance matrices.

In the multivariate case, the number of parameters in the nonparametric density model increases quadratically with the dimension, which poses challenges in high-dimensional applications. However, there are only a few studies in estimating the sparse density model. In the chapter 3, nonparametric density estimation is estimated through a nonparametric Bernstein copula with adapted LASSO regularization.

Both the sparse and less sparse scenarios will be investigated in the simulation study. In the sparse scenario, the majority of the true parameters are assigned a value of zero. In practice, this can be evaluated using a property testing algorithm, which aims to design an algorithm that determines whether the input data satisfies a predetermined property like sparsity(Barman et al., 2018). However, it is important to note that conducting a pre-test for sparsity is not within the scope of this thesis. Although our primary focus revolves around employing the lasso-type regularization method to handle sparse data and studying their properties, it does not imply that the model can only be utilized in the sparse scenario. The empirical Bernstein copula model has the capability to estimate any copula models, but it does not enforce sparsity. To address this limitation, our proposed model introduces a penalty term to encourage sparsity. In practice, if the underlying data is not sparse, the tuning parameter of our model will approach zero, effectively making the model equivalent to the empirical Bernstein copula model. As the level of sparsity increases, the tuning parameter also increases accordingly. Then the degree of sparsity can be inferred by observing the estimated tuning parameter. In a less sparse scenario, there is a possibility of falsely excluding certain parameters during the regularization process. Model selection mistakes made by lasso cause the distribution of this naive estimator to be biased and nonnnormal.

To address this issue, Belloni et al. (2014) develop a novel estimation and uniformly valid inference method for the treatment effect in this setting, called the "post-double-selection" method. Building on this, Belloni et al. (2016) present a general methodology to estimate an effect of interest using an instrument that guards against model selection errors, specifically applied to the case of a logistic binary choice model. Additionally, Chernozhukov et al. (2018) propose the double/debiased machine learning method to counteract the impact of regularization bias and overfitting. In the context of the Heckman selection model, Hirukawa et al. (2023) introduce the double lasso estimation method to correct errors arising from lasso. Beyond lasso-based methods, recent literature Ahmed et al. (2023) introduces the shrinkage method, which considers prior information about inactive parameters during coefficient estimation for active parameters. Research indicates that the shrinkage method outperforms penalized methods both analytically and numerically, but only when the number of inactive parameters is accurately specified.

## 2.3 Applications of the nonparametric copula in financial time series

The original motivation to use copula in finance is due to the emerging empirical evidence showing that the relationship between some crucial asset returns is non-normal. The financial asset returns exhibit relatively high concentration in market downturn compared with that in market upturn(Bae et al., 2003). This evidence has been widely studied for financial decision-making in risk management(Eling and Toplek, 2009; Ai et al., 2017), option pricing(Guegan and Zang, 2013; Van den Goorbergh et al., 2005), portfolio optimization(Boubaker and Sghaier, 2013; Ai et al., 2017) and the contagion between financial markets(Wen et al., 2012).

The implementation of the copula in time series data can be clarified into two distinct categories. One is the application to the multivariate time series and the focus for this category is to model the joint distribution of a random vector $S_t = [S_{1,t}, ..., S_{kt}]$ conditioning on a given information set $\mathcal{F}_{t-1}$. The information set $\mathcal{F}_{t-1}$ is the $\sigma-$field. The other one mainly focuses on considering the copula of a sequence of observations which is a univariate time series. It utilizes copula to study the joint behavior of $[S_t, S_{t+1}, ..., S_{t+n}]$ in a specific time period. This application connects with the Markov process and general nonlinear time series(Chen and Fan, 2006; Chen et al., 2009; Ibragimov, 2009). In this thesis, we concentrate on multivariate time series cases.

In multivariate time series applications, Rosenberg (2003) apply a fully nonparametric framework on the bivariate option pricing using kernel-based nonparametric copula. Patton (2006) extends the theory of copula for conditioning variables and combines Sklar's theorem to the time series case. For $t \in \{1, ..., T\}$, let

$$\boldsymbol{S}_t|\mathcal{F}_{t-1} \sim F(\cdot|\mathcal{F}_{t-1}), \qquad S_{i,t}|\mathcal{F}_{t-1} \sim F_i(\cdot|\mathcal{F}_{t-1})$$

Then,

$$F(\boldsymbol{s}|\mathcal{F}_{t-1}) = C(F_1(s_1|\mathcal{F}_{t-1}), ..., F_n(s_n|\mathcal{F}_{t-1})|\mathcal{F}_{t-1}) \qquad (2.3.1)$$

The attendant problem of applying Sklar's theorem to the conditional distribution is that the conditional set $\mathcal{F}_{t-1}$ should be the same for both the margins and copula function. The $F(\boldsymbol{s}|\mathcal{F}_{t-1})$ cannot be a valid joint distribution function if it is failed to use the same information set of each $F_i(\cdot|\mathcal{F}_{t-1})$(Patton, 2009). However, some information contained in the information set is not relevant to all variables, say $\mathcal{F}_{i,t-1}$ is the smallest subset of $\mathcal{F}_{t-1}$ such that $S_{i,t}|\mathcal{F}_{i,t-1} =^D S_{i,t}|\mathcal{F}_{t-1}$. Each marginal distribution is conditional on its own $\mathcal{F}_{i,t-1}$ and it will be different across

margins.

A preponderance of evidence from the econometrics literature indicates that the conditional volatility of financial time series changes through time (Andersen et al., 2006; Bauwens et al., 2006). Then, a natural question is whether the conditional dependence structure also significantly changes through time. Patton (2002) and Patton (2006) allow the time variation in the conditional copula by changing the dependence parameter to evolve through time according to a particular equation. They provide the evolution equation for both symmetrized Joe-Clayton copula and Gaussian copula based on an ARMA-type process. Guegan and Zang (2013) further studies this method and applies it to the generalized hyperbolic(GH)-GARCH process. The copula's parameters vary through the function

$$\theta_{l,t} = \theta_0 + \sum_{i=1}^{g} \eta_i \prod_{j=1}^{2} \varepsilon_{j,t-i} + \sum_{k=1}^{s} \gamma_k \theta_{l,t-k} \tag{2.3.2}$$

with $\eta$ and $\gamma$ are scalar model parameters and $(\varepsilon_{1,t}, \varepsilon_{2,t})$ are standardized innovations from GH-GARCH estimation. Van den Goorbergh et al. (2005) indicates that there exist the one-to-one relationship between the Kendall's nonparametric measure of association $\tau$ and the parameter $\theta$ of copula function. Then, the relationship is

$$\tau(\theta) = 4E(C_\theta(U_1, U_2)) - 1 \tag{2.3.3}$$

where $(U_1, U_2)$ is distributed as $C_\theta$ and the expectation is taken with respect to $U_1$ and $U_2$. Then, let the dynamic copula parameter be updated by the specific regression on Kendall's tau. However, this one-to-one relationship between Kendall's tau and copula function is only efficient and accurate for one parameter-based parametric copula.

# CHAPTER 3

## Adaptive LASSO Regularization with Nonparametric Bernstein Copula Estimation

## 3.1 Introduction

Copulas and copula densities have gained increasing attention for their flexibility in modeling multivariate distributions. However, in many cases, the dependence structure of the data cannot be fully captured by parametric copulas. The Bernstein copula, introduced by Sancetta and Satchell (2004), addresses this limitation by providing a nonparametric copula that can approximate any copula arbitrarily well. Its consistency and asymptotic properties have been systematically studied. The Bernstein copula exhibits high accuracy, especially in scenarios involving multiple risk sectors with inhomogeneous dependency, as shown by Diers et al. (2012).

Nevertheless, nonparametric estimation of Bernstein copula models involves dealing with unknown parameters in an infinite-dimensional parameter space, which is computationally challenging with finite samples and poses a non-well-posed optimization problem (Chen, 2007). To address this, Geman and Hwang (1982) introduced the method of sieves for nonparametric maximum likelihood estimation. The method of sieves optimizes a criterion function over a sequence of

parameter spaces with reduced complexity and finite dimensions. Building on this framework, Chen et al. (2006) proposed the sieve maximum likelihood estimation (SMLE) method for estimating copula parameters and unknown univariate marginals. The SMLE efficiently handles the estimation of these components within a unified framework. Furthermore, Ivan et al. (2021) extended the SMLE method to the multivariate case, specifically addressing situations where the copula density function is unknown. Their approach improves upon the Quasi-MLE method while avoiding the drawbacks associated with full maximum likelihood estimation. The SMLE strikes a balance by delivering comparable efficiency to full MLE while offering enhanced computational feasibility and robustness. In our proposed nonparametric model, we construct the sieve parameter space using Bernstein polynomials, enabling the estimation of the model in a parametric manner.

Since the Bernstein copula can be viewed as a special case of the mixture model, where the mixture components are beta densities. Determining the appropriate number of mixture components is a topic of intense research. Traditional methods for selecting the number of components rely on the likelihood function and information criteria. Leroux (1992) demonstrate the consistency of the maximum-penalized-likelihood estimator for estimating the mixing distribution and show that information criteria like AIC and BIC do not underestimate the true number of mixture components. Consistency of BIC in estimating the number of components for mixtures of normals is also discussed by Roeder and Wasserman (1997). Other approaches focus on measuring the discrepancy between the fitted mixture model and the nonparametric estimation of the population distribution, using methods such as the Kullback-Leibler distance (James et al., 2001; Lijoi et al., 2005), penalized minimum-distance approach (Chen and Kalbfleisch, 1996; Biau and Devroye, 2005), and minimum Hellinger distance method (Cutler and

Cordero-Brana, 1996; Woo and Sriram, 2006). Ray and Lindsay (2008) investigate model selection in multivariate mixture models through a global comparison tool called quadratic risk. However, these methods often involve computationally intensive model search algorithms, leading to high computational costs.

Another method is to use regularization strategies to improve the estimation results. Regularization, as a process of adding information to mitigate overfitting and reduce the number of parameters, has been widely used in nonparametric estimation (Bühlmann and Van De Geer, 2011). If the true underlying model exhibits sparsity, the idea of variable selection, commonly applied in regression, can enhance the performance of the fitted model. This approach also allows the number of mixture components (M) to exceed the sample size (n). One popular regularization technique is the LASSO (Least Absolute Shrinkage and Selection Operator), proposed by Tibshirani (1996). Other variable selection methods include SCAD (Smoothly Clipped Absolute Deviation)(Fan and Li, 2001), adaptive LASSO(Zou, 2006), VISA (Variable Selection via the Independence Screening Algorithm)(Radchenko and James, 2008), the Double Dantzig(James et al., 2009), and SPADE(Bunea et al., 2010). Some researchers have extended selection methods to high-dimensional cases (Banerjee et al., 2008), but many of these approaches are based on regression functions. In the context of estimating density functions, Bunea et al. (2010) introduced the possibility of estimating density functions through $l_1$ penalization (SPADES). They assume that the densities of true mixture components belong to a large known candidate density pool. By penalizing the mixing weights, SPADES effectively selects the components of the mixture models. They demonstrate that SPADES is effective when the local distances between true mixture components are substantially larger than their variances or covariance matrices. In this chapter, we demonstrate that the L1-regularization technique, commonly used in linear regression models, can also be successfully

applied to copula density estimation. L1-regularization introduces a penalty term based on the L1 norm of the estimated parameters, promoting sparsity and reducing the number of estimated parameters. By incorporating L1-regularization into the estimation procedure of the nonparametric Bernstein copula, we aim to improve the stability, robustness, and interpretability of the density estimates.

In this chapter, we present a nonparametric Bernstein copula approach for accurately measuring the underlying dependence structures among data. We employ the sieve maximum likelihood estimation (SMLE) technique to estimate the parameters within the sieve space. Additionally, we incorporate the adaptive LASSO method to introduce sparsity and enhance the predictability of the estimated model. The remainder of the chapter is structured as follows: Section 2 outlines the proposed model in a step-by-step manner. We describe the methodology and procedures involved in constructing the nonparametric Bernstein copula, estimating its parameters using SMLE, and incorporating adaptive LASSO for variable selection. Section 3 presents a comprehensive simulation study that demonstrates the properties of the Bernstein copula in approximating various types of copulas. We conduct Monte Carlo simulations under both sparse and less sparse scenarios to evaluate the performance of the proposed approach. In section 4, we provide empirical evidence to capture financial contagion between four markets. We apply the nonparametric Bernstein copula model to analyze and interpret the dependence patterns among these markets, shedding light on the presence and magnitude of contagion effects.

## 3.2 Problem formulation and methodology

One of the aims is to accurately estimate the target dependence structure. Compared with estimating the multivariate density function directly, the copula model

provides more flexibility which allows separately estimating the copula density function and different marginal density functions. For simplicity, consider the bivariate case $d = 2$ and it's not difficult to extend the following expression to the multivariate situation.

According to Sklar's theorem(Sklar, 1959), the bivariate density function $h$ of 2-dimension cumulative density function $F$ with univariate margin $u_1 = F_1(x_1), u_2 = F_2(x_2)$ and corresponding univariate density $f_1, f_2$ can be expressed as

$$h(u_1, u_2; \theta, \gamma) = c(u_1, u_2; \theta, \gamma) \prod_{i=1}^{2} f_i(x_i; \theta_i) \tag{3.2.1}$$

$c$ is the density function of the 2-dimensional copula. After taking the logarithm of 3.2.1:

$$L(u_1, u_2; \theta, \gamma) = L_c(\theta, \gamma) + L_m(\theta) \tag{3.2.2}$$

where the $L_c(\theta, \gamma) = \sum_{i=1}^{N} \log(c(u_{1i}, u_{2i}; \theta, \gamma)$ is loglikelihood of the copula function and $L_m(\theta) = \sum_{i=1}^{N} \log f_1(x_1; \theta_1) + \sum_{i=1}^{N} \log f_2(x_2; \theta_2)$ is the loglikelihood of marginal distribution functions. Assuming the marginal distributions are correctly specified, our problem is to maximize the log-likelihood of $L_c(\gamma, \hat{\theta})$. Since the copula function in this paper is a nonparametric one with $dim(\gamma) \to \infty$, we set an infinite-dimensional sieve space and penalty for regularization. The method of sieve solves the problem of estimating a nonparametric model over infinite dimensional space in maximum likelihood estimation(MLE) optimization. It constructs a subset of the parameter space to facilitate the estimation. The penalty part pushes more negligible parameters to zeros. Combining these two, the infinite-dimensional sieve space can be represented as followed

$$\Gamma_n = \{\gamma \in \Gamma : \gamma(\cdot) = \sum_{k_1=1}^{\infty} \sum_{k_2=1}^{\infty} a_{k_1,k_2} B_{k_1,k_2}(\cdot), \ pen(\gamma) \leq b_n\} \quad with \ b_n \to \infty \ slowly \tag{3.2.3}$$

where $B_{\mathbf{k}}(\cdot)$ is employed as Bernstein polynomials and $pen(\gamma)$ is the penalty part which is a function of $\gamma$. Since the loglikelihood of copula function $L_c(\gamma)$ is concave and the adaptive lasso penalty is convex, the sieve MLE estimation of $L_c(\gamma)$ becomes equivalent to the penalized MLE (Grenander, 1981):

$$\max_{\gamma \in \Gamma_n} L_{nc}(\gamma) \tag{3.2.4}$$

$$\text{subject to } pen(\gamma) \leq b_n \tag{3.2.5}$$

Then, the problem changes to a convex optimization problem.

### 3.2.1 The copula function and its properties

As mentioned in the last section, the copula function can be considered as a d-dimensional distribution with uniformly distributed marginals. $C : [0,1]^d \to [0,1]$ is a d-dimensional copula function if:

1. $C(u_1, u_2, ..., u_{m-1}, 0, u_{m+1}, ..., u_d) = 0$

2. $C(1, ..., 1, u, 1, ..., 1) = u$

3. $C(u_1, ..., u_d)$ is non-decreasing in each component, $u_i$. For $u_{j,1} \leq u_{j,2}$, $P(U_1 \in [u_{1,1}, u_{1,2}], ..., U_d \in [u_{d,1}, u_{d,2}])$ must be non-negative. This implies the rectangle inequality

$$\sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1+...+i_d} C(u_{1,i_1}, ..., u_{d,i_d}) \geq 0$$

Also, the reverse is true that if $C(\boldsymbol{u})$ is a copula function, it has the properties 1 to 3 immediately. From property 2, it is easy to confirm that $C(1, u_1, ..., u_{d-1})$ is a $(d-1)$-dimensional copula. Based on the idea of probability integral transformation(PIT), we have

**Proposition 3.2.1.** *let $X$ be a random variable on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with distribution function F. If F is continuous, then F(X) is uniformly distributed on [0,1]*

McNeil et al. (2015) summarize the quantile transform method which provides key concepts for further understanding of copulas based on PIT. Let $F$ denote the univariate distribution function and $\overleftarrow{F}$ represent the generalized inverse.

$$\overleftarrow{F}(x) := \inf\{v : F(v) \geq x\} \tag{3.2.6}$$

Then, we have the following well-known result:

**Proposition 3.2.2.** *If $U \sim U(0,1)$ and $F_X$ is the CDF funtion of random variable X, then*

$$P(\overleftarrow{F}(U) \leq x) = F_X(x) \tag{3.2.7}$$

*Considering the opposite direction, if $F_X$ is continuous, then*

$$F_X(X) \sim U(0,1) \tag{3.2.8}$$

Let the random variables $X_1,..., X_d$ with continuous distribution function $F_1,..., F_d$ correspondingly. Then, by proportion 3.2.2 we have $U_1 := F_1(X_1),..., U_d := F_d(X_d)$ are uniformly distributed on [0,1]. Therefore,

$$
\begin{aligned}
F(x_1, ..., x_d) &= \mathbb{P}(X_1 \leq x_1, ..., X_d \leq x_d) \\
&= \mathbb{P}(F_1(X_1) \leq F_1(x_1), ..., F_d(X_d) \leq F_d(x_d)) \\
&= \mathbb{P}(U_1 \leq F_1(x_1), ..., U_d \leq F_d(x_d)) \\
&= C(F_1(x_1), ..., F_d(x_d)) \tag{3.2.9}
\end{aligned}
$$

Here, $C$ is a copula function. This is one side of the famous Sklar's theorem.

**Theorem 3.2.1** (Sklar (1959))**.** *Let $(X_1, ..., X_d)$ be a multivariate random variable with joint distribution function $F$ and univariate marginal distributions $F_1, ..., F_d$. Then there exists a copula $C$, such that for all $\boldsymbol{x} \in \mathbb{R}^d$*

$$F(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d)) \tag{3.2.10}$$

*$C$ is uniquely determined on $Range(F_1) \times ... \times Range(F_d)$ and hence it is unique when $F_1, ..., F_d$ are continuous.*

Sklar's theorem provides the possibility that any copula can be combined with various univariate distribution functions to a d-dimensional multivariate distribution function by using equation 3.2.10. Particularly, copulas can deal with the situation where each margin is different and it provides a valid alternative way to estimate the classic multivariate distribution function, such as Gaussian, Pareto and etc. Another property of copula is its invariance under monotonic transformations.

**Proposition 3.2.3.** *Suppose the random variables $X_1, ..., X_d$ have copula function $C_{\boldsymbol{X}}$ and their marginal distributions are continuous. Let $T_i : \mathbb{R} \rightarrow \mathbb{R}$ represent the strictly increasing function for $i = 1, ..., d$:*

$$Y_1 := T_1(X_1), ..., Y_d := T_d(X_d) \tag{3.2.11}$$

*then $Y_1, ..., Y_d$ have the same copula function $C_{\boldsymbol{X}}$.*

Another crucial feature is derived independently by Fréchet–Hoeffding:

**Theorem 3.2.2** (The Fréchet–Hoeffding Bounds)**.** *Consider a copula function $C(\boldsymbol{u}) = C(u_1, ..., u_d)$, then*

$$\max\left\{1 - d + \sum_{i=1}^{d} u_i, 0\right\} \leq C(\boldsymbol{u}) \leq \min\{u_1, ..., u_d\} \tag{3.2.12}$$

The upper bound of the Fréchet–Hoeffding theorem is tight for all $d$ whereas the lower bound is tight only when $d = 2$. These two bounds represent two cases of extreme dependency: comonotonicity and countermonotonicity. The comonotonicity case constructs the upper Fréchet–Hoeffding bound and it stands for the extreme positive dependence.

### 3.2.2 Bernstein copula and its large sample theory

Bernstein polynomials as a linear combination of beta density are admitted to approximate any function $f(x)$ arbitrarily well in approximation theory.

**Definition 3.2.1.** *(Lorentz, 2013) Let $B(k, j, x) = \binom{k}{j} x^j (1 - x)^{k-j}$, $0 \leq x \leq 1$, $j = 1,...,k \in \mathbb{N}$. Then we have*

$$\int_0^1 kB(k - 1, j - 1, x)dx = 1$$

where $B(k, j, x)$ is the Bernstein basis polynomial. The first-order derivative of the basis Bernstein polynomial can be represented as the difference between the adjacent terms,

$$\frac{d}{dx}B(k, j, x) = k[B(k - 1, j - 1, x) - B(k - 1, j, x)] \quad \text{for j=0,...,k}$$

A Bernstein polynomial approximation to density function $f \in C_{[0,1]^k}$ is constructed through a linear operator $B_J^k$ to $f \in C_{[0,1]^k}$, such that

$$(B_J^k f)(x_1, x_2, ..., x_k) = \sum_{j_1=0}^{J_1} ... \sum_{j_k}^{J_k} f\left(\frac{j_1}{J_1}, ..., \frac{j_k}{J_k}\right) \prod_{s=1}^{k} P_{j_s, J_s}(x_s) \qquad (3.2.13)$$

and it can be written in the form of a Stieltjes integral in the variable t,

$$(B_J^k f)(x_1, x_2, ..., x_k) = \int_0^1 ... \int_0^1 f(t_1, ..., t_k) d_{t_1} K_{J_1}(x_1, t_1)...d_{t_k} K_{J_k}(x_k, t_k)$$

(3.2.14)

with the kernel

$$K_{J_s}(x_s, t_s) = \sum_{j_s \leq J_s t_s} \binom{J_s}{j_s} x_s^{j_s} (1 - x_s)^{J_s - j_s}, \quad 0 < t \leq 1$$ 

(3.2.15)

$$K_{J_s}(x_s, 0) = 0$$ 

(3.2.16)

which is constant in any interval $\frac{j_s}{J_s} \leq t \leq \frac{j_s+1}{J_s}$, $j_s = 0, 1, ..., J_s - 1$, and has the jump

$$\binom{J_s}{j_s} x_s^{j_s} (1 - x_s)^{J_s - j_s}$$

at the basic point of interpolation $t_s = j_s/J_s$ and $s = 1, ..., k$. Here, $C_{[0,1]^k}$ is the space of continuous bounded functions on $[0, 1]^k$.

**Theorem 3.2.3. (Uniform convergence)** *For a function $f(x_1, ..., x_k) \in C_{[0,1]^k}$, the relation*

$$\lim_{J \to \infty} (B_J^k f)(x_1, x_2, ..., x_k) = f(x_1, ..., x_k)$$

(3.2.17)

*holds at each point of continuity $x$ of $f$; and it converges uniformly on $[0, 1^k]$ if $f(x_1, ..., x_k)$ is continuous on this interval.*

**Theorem 3.2.4. (Error Bound for Lipschitz Conditions)** *if $f(x_1, ..., x_k) : [0, 1]^k \to \mathbb{R}$ is a continuous function satisfying Lipschitz condition*

$$\|f(\mathbf{x}) - f(\mathbf{y})\| < L\|\mathbf{x} - \mathbf{y}\|$$

*then the inequality holds,*

$$\|(B_J^k f)(x_1, x_2, ..., x_k) - f(x_1, ..., x_k)\| < \frac{L}{2} \left( \sum_{s=1}^{k} \frac{1}{J_s} \right)^{1/2}. \qquad (3.2.18)$$

Based on theorem 3.2.3, Sancetta and Satchell (2004) employs the multivariate Bernstein polynomial to construct the copula function. For simplicity, we state it in the bivariate situation and set the number of grids in each dimension $J_s$ to be the same number $J$. Let $\alpha(\frac{j_1}{J}, \frac{j_2}{J})$ be a real-valued constant indexed by $(j_1, j_2)$, such that $0 \le j_s \le J \in \mathbb{N}$. The bivariate mapping $C_B : [0, 1]^2 \to [0, 1]$, where

$$C_B(\mathbf{u}) = \sum_{j_1=0}^{J} \sum_{j_2=0}^{J} \alpha \left( \frac{j_1}{J}, \frac{j_2}{J} \right) P_{j_1, J}(u_1) P_{j_2, J}(u_2). \qquad (3.2.19)$$

And,

$$P_{j_s, J}(u_s) = \binom{J}{j_s} u_s^J (1 - u_s)^{J - j_s}, \quad s = 1, 2. \qquad (3.2.20)$$

where $\mathbf{u} = (u_1, u_2) \in [0, 1]^2$.

**Theorem 3.2.5.** *(Sancetta and Satchell, 2004) $C_B(\boldsymbol{u})$ is a bivariate Bernstein copula function if*

$$\sum_{l_1=0}^{1} \sum_{l_2=0}^{1} (-1)^{l_1+l_2} \alpha \left( \frac{j_1 + l_1}{J}, \frac{j_2 + l_2}{J} \right) \ge 0, \qquad (3.2.21)$$

$$\min \left( 0, \frac{j_1}{J} + \frac{j_2}{J} - 1 \right) \le \alpha \left( \frac{j_1}{J}, \frac{j_2}{J} \right) \le \min \left( \frac{j_1}{J}, \frac{j_2}{J} \right), 0 \le j_s \le J - 1, s = 1, 2.$$
$$(3.2.22)$$

*in particular,*

$$\lim_{j_s \to 0} \alpha \left( \frac{j_1}{J}, \frac{j_2}{J} \right) = 0, \quad and \quad \alpha \left( 1, \frac{j_s}{J} \right) = \frac{j_s}{J}, \forall s = 1, 2. \qquad (3.2.23)$$

The Berstein copula is evaluated at each vertex of the grid using the $\alpha$ coeffi-

cients, and the product of Bernstein polynomials is employed to smooth it. The Bernstein polynomials here can be considered as the weight function that adjusts the coefficient $\alpha$ through changing the corresponding J. Based on these properties of the Bernstein polynomials, the Bernstein copula can accommodate any given parametric copula function. Since the Bernstein copula is absolutely continuous, it has a copula density. After taking the derivative of equation 3.2.19 with respect to $u_1$ and $u_2$, the bivariate density function is as followed.

$$
c_B(\mathbf{u}) = \frac{\partial^2}{\partial u_1 \partial u_2} C_B(\mathbf{u}) = J^2 \sum_{j_1=0}^{J-1} \sum_{j_2=0}^{J-1} \gamma\left(\frac{j_1}{J}, \frac{j_2}{J}\right) \times P_{j_1,J-1}(u_1) \times P_{j_2,J-1}(u_2)
$$
(3.2.24)

Each marginal distribution $u_s$ of the Bernstein copula is split into J sections, which construct a grid-type structure in the two-dimensional hypercube. $\gamma(\frac{j_1}{J}, \frac{j_2}{J})$ is defined accordingly, which is directly linked to the two-dimensional rectangular inequality

$$
\gamma\left(\frac{j_1}{J}, \frac{j_2}{J}\right) = \Delta_{1,2}\alpha\left(\frac{j_1}{J}, \frac{j_2}{J}\right) = \sum_{l_1=0}^{1} \sum_{l_2=0}^{1} (-1)^{l_1+l_2+2} \alpha\left(\frac{j_1+l_1}{J}, \frac{j_2+l_2}{J}\right)
$$
$$
= \alpha\left(\frac{j_1}{J}, \frac{j_2}{J}\right) - \alpha\left(\frac{j_1+1}{J}, \frac{j_2}{J}\right) - \alpha\left(\frac{j_1}{J}, \frac{j_2+1}{J}\right) + \alpha\left(\frac{j_1+1}{J}, \frac{j_2+1}{J}\right)
$$

Under the condition 3.2.21, 3.2.22 and 3.2.23, $\sum_{j_1=0}^{J-1} \gamma\left(\frac{j_1}{J}, \frac{j_2}{J}\right) = \alpha\left(\frac{0}{J}, \frac{j_2}{J}\right) - \alpha\left(\frac{J}{J}, \frac{j_2}{J}\right) - \alpha\left(\frac{0}{J}, \frac{j_2+1}{J}\right) + \alpha\left(\frac{J}{J}, \frac{j_2+1}{J}\right) = 0 - \frac{j_2}{J} - 0 + \frac{j_2+1}{J} = \frac{1}{J}$ and also $\sum_{j_2=0}^{J-1} \gamma\left(\frac{j_1}{J}, \frac{j_2}{J}\right) = \frac{1}{J}$.

**Theorem 3.2.6.** *(Bernstein copula density function(Ivan et al., 2021))* $c_B(\mathbf{u})$ *is a Bernstein copula density function if the coefficient term $\gamma$ satisfies*

1. $\sum_{j_s=0}^{J-1} \gamma\left(\frac{j_1}{J}, \frac{j_2}{J}\right) = \frac{1}{J}, \; s = 1, 2$

2. $\sum_{j_1=0}^{J-1} \sum_{j_2=0}^{J-1} \gamma\left(\frac{j_1}{J}, \frac{j_2}{J}\right) = 1$

3. $0 \leq \gamma\left(\frac{j_1}{J}, \frac{j_2}{J}\right) < 1$

The error bound for the Bernstein polynomials can be easily applied to Bernstein copula, if we consider the coefficients of the Bernstein copula $\gamma$ have the Lipschitz property, such that(in bivariate case)

$$
\begin{aligned}
\|c_B(u_1, u_2) - c(u_1, u_2)\| &= \sup_{u_1, u_2 \in [0,1]} \left| \sum_{j_1=0}^{J} \sum_{j_2=0}^{J} \left[ \gamma \left( \frac{j_1}{J}, \frac{j_2}{J} \right) - c(u_1, u_2) \right] P_{j_1, J-1}(u_1) P_{j_2, J-1}(u_2) \right| \\
&\leq \sup_{u_1, u_2 \in [0,1]} \sum_{j_1=0}^{J} \sum_{j_2=0}^{J} \left[ \left| \frac{j_1}{J} - u_1 \right| + \left| \frac{j_2}{J} - u_2 \right| \right] P_{j_1, J-1}(u_1) P_{j_2, J-1}(u_2) \\
&\leq \frac{L_1}{2 J^{1/2}} + \frac{L_2}{2 J^{1/2}} = L J^{-1/2}
\end{aligned}
$$

Based on the Bernstein copula function, Sancetta and Satchell (2004) provide the empirical estimation procedure to make the Bernstein copula operational. The empirical Bernstein copula $\tilde{C}_B$ is considered that the copula coefficient $\alpha(\frac{j_1}{J}, \frac{j_2}{J}) = C_n(\frac{j_1}{J}, \frac{j_2}{J})$ and $C_n$ is the empirical copula:

$$
\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{\left\{ \cap_{s=1}^{2} [u_{s,i} \leq \frac{j_s}{J}] \right\}} \tag{3.2.25}
$$

and the order of the corresponding polynomial would be related to the smoothing capability of the estimator. After taking the first derivative of the empirical Bernstein copula function, the coefficient of the density function $\tilde{c}_B$ is equal to a bivariate histogram estimator. The coefficients are the fractions of data points in their grids.

$$
\tilde{c}_B = \sum_{j_1=0}^{J-1} \sum_{j_2=0}^{J-1} \Delta_{1,2} \left( \frac{J^2}{n} \sum_{i=1}^{n} \mathbb{I}_{\left\{ \cap_{s=1}^{2} [u_{s,i} \leq \frac{j_s}{J}] \right\}} \right) \times \prod_{s=1}^{2} \binom{J_1}{j_s} u_s^{j_s} (1 - u_s)^{J - j_s - 1}
$$

where $\Delta_{1,2}$ is the 2-dimensional differentiating operator and it is:

$$
\Delta_{1,2} \mathbb{I}_{\left\{ \cap_{s=1}^{2} [u_{s,i} \leq \frac{j_s}{J}] \right\}} = \sum_{l_1=0}^{1} \sum_{l_2=0}^{1} (-1)^{l_1 + l_2} \mathbb{I}_{\left\{ \cap_{s=1}^{2} [u_{s,i} \leq \frac{j_s}{J} + \frac{l_s}{J}] \right\}} \tag{3.2.26}
$$

46

The optimal empirical Bernstein copula is selected by maximizing the mean squared error $||\tilde{c}_B - c||_2^2$ where $||...||_2$ is the L-2 norm under the corresponding probability measure and $c$ represents the true copula density.

### 3.2.2.1 Large sample theories of Empirical Bernstein copula

The research on the asymptotic behavior of the Empirical Bernstein copula density estimator is mainly discussed by Sancetta and Satchell (2004), Bouezmarni et al. (2010), Janssen et al. (2012), Janssen et al. (2014) and Rose (2015). Sancetta and Satchell (2004) prove the consistency in mean-squared error. They state that the bias of the Bernstein copula estimator is $\mathcal{O}(J^{-1})$ which is of the same order as the bias for the histogram estimator. Bouezmarni et al. (2010) studies the Bernstein copula based on $\alpha-$mixing data. Janssen et al. (2012) derive the almost sure consistent rate and asymptotic normality properties of the empirical Bernstein copula function and it indicates if the copula function C has the first order partial derivatives $C_{u_1}$ and $C_{u_2}$ which are Lipschitz continuous of order $\alpha$ with $0 < \alpha < 1$ and $n/(J^{1+\alpha} \log \log n) \to c \geq 0$, then the $L_\infty$ distance of $\tilde{C}_B$ and the true copula $C$ is in order of $(n^{-1/2}(\log \log n)^{1/2}$. Janssen et al. (2014) further studies the asymptotic distributional behavior of the Bernstein copula density estimator. They show that the central limit theorem is valid for the empirical Bernstein copula density estimator without assuming the marginals to be known.

From the previous section, the number of coefficients that need to be estimated of the empirical Bernstein copula $\tilde{c}_B$ is determined by the order of $J$. The selection of J relies on the sample size n, and we have $J \to \infty$ if $n \to \infty$. To stress that the order J is based on n, we replace J with $J_n$. We first analyze the bias of the bivariate empirical Bernstein copula density estimator based on the second-order Taylor expansion(Bouezmarni et al., 2010). Assume that the copula density function $\tilde{c}_B$

has a continuous second-order derivative, then:

$$\mathbb{E}(\tilde{c}_B(u_1, u_2)) - c_B(u_1, u_2)$$

$$= J_n^2 \sum_{j_1=0}^{J_n-1} \sum_{j_2=0}^{J_n-1} \left\{ \int_{j_1/J_n}^{(j_1+1)/J_n} \int_{j_2/J_n}^{(j_2+1)/J_n} (c_B(x_1, x_2) - c(u_1, u_2)) dx_1 dx_2 \right\} \prod_{s=1}^{2} p_{j_s, J_n-1}(u_s)$$

$$= J_n^2 \sum_{j_1=0}^{J_n-1} \sum_{j_2=0}^{J_n-1} \left\{ \frac{\partial c(u_1, u_2)}{\partial x_1} \int_{j_1/J_n}^{(j_1+1)/J_n} \int_{j_2/J_n}^{(j_2+1)/J_n} (x_1 - u_1) dx_1 \right.$$

$$\left. + \frac{\partial c(u_1, u_2)}{\partial x_2} \int_{j_1/J_n}^{(j_1+1)/J_n} \int_{j_2/J_n}^{(j_2+1)/J_n} (x_2 - u_1) dx_2 \right\} \prod_{s=1}^{2} p_{j_s, J_n-1}(u_s)$$

$$+ \frac{J_n^2}{2} \sum_{j_1=0}^{J_n-1} \sum_{j_2=0}^{J_n-1} \left\{ \frac{\partial^2 c(u_1, u_2)}{\partial x_1 \partial x_2} \int_{j_1/J_n}^{(j_1+1)/J_n} \int_{j_2/J_n}^{(j_2+1)/J_n} (x_1 - u_1)(x_2 - u_2) dx_1 dx_2 \right\} \prod_{s=1}^{2} p_{j_s, J_n-1}(u_s)$$

$$+ \frac{J_n^2}{2} \sum_{j_1=0}^{J_n-1} \sum_{j_2=0}^{J_n-1} \left\{ \sum_{s=1}^{2} \frac{\partial^2 c(u_1, u_2)}{\partial x_s^2} \int_{j_1/J_n}^{(j_1+1)/J_n} \int_{j_2/J_n}^{(j_2+1)/J_n} (x_s - u_s)^2 dx \right\} \prod_{s=1}^{2} p_{j_s, J_n-1}(u_s) + o\left(\frac{1}{J_n}\right)$$

Based on

$$\frac{J_n^2}{2} \sum_{j_1=0}^{J_n-1} \sum_{j_2=0}^{J_n-1} \left\{ \frac{\partial^2 c(u_1, u_2)}{\partial x_1 \partial x_2} \int_{j_1/J_n}^{(j_1+1)/J_n} \int_{j_2/J_n}^{(j_2+1)/J_n} (x_1 - u_1)(x_2 - u_2) dx_1 dx_2 \right\} \prod_{s=1}^{2} p_{j_s, J_n-1}(u_s)$$

$$= o\left(\frac{1}{J_n^2}\right)$$

and the mean and variance of the Binomial distribution,

$$\mathbb{E}(\tilde{c}_B(u_1, u_2)) - c_B(u_1, u_2) =$$

$$\frac{1}{2} \sum_{s=1}^{2} \left\{ \frac{\partial c(u_1, u_2)}{\partial x_s} (1 - 2u_s) + \frac{\partial^2 c(u_1, u_2)}{\partial x_s^2} u_s(1 - u_s) \right\} \times \frac{1}{J_n} + o\left(\frac{1}{J_n}\right)$$

Bouezmarni et al. (2010) and Janssen et al. (2014) shows the variance of the Bernstein copula density estimator is

$$Var(\tilde{c}_B(u_1, u_2)) = \frac{J_n}{n} \times \frac{c(u_1, u_2)}{\prod_{s=1}^{2}(u_s(1 - u_s))^{1/2}} + o\left(\frac{J_n}{n}\right)$$

Since the $c_B(u_1, u_2)$ is the deterministic bias term, based on the assumption that c is continuous, we have

$$\int_{j_1/J_n}^{(j_1+1)/J_n} \int_{j_2/J_n}^{(j_2+1)/J_n} c(x_1, x_2)dx_1dx_2 = \frac{1}{J_n^2}c(\frac{j_1}{J_n - 1}, \frac{j_2}{J_n - 1}) + o\left(\frac{1}{J_n^3}\right)$$

uniformly in $j_1, j_2 \leq J_n - 1$. And

$$c_B(u_1, u_2) = J_n^2 \sum_{j_1=0}^{J_n-1} \sum_{j_2=0}^{J_n-1} \left\{ \int_{j_1/J_n}^{(j_1+1)/J_n} \int_{j_2/J_n}^{(j_2+1)/J_n} c(x_1, x_2)dx_1dx_2 \right\} \prod_{s=1}^{2} p_{j_s, J_n-1}(u_s)$$
$$= c(u, v) + o\left(\frac{1}{J_n}\right)$$

**Theorem 3.2.7.** *(Asymptotic normality (Janssen et al., 2014))Suppose that the copula density function c is twice differentiable and continuous on $[0, 1]^2$, $J_n$ depends on n and has the relationship $J_n = o\left(\frac{n^{1/2}}{(\log n)(\log \log n)^{1/2}}\right)$, then for $u_1, u_2 \in (0, 1)$, as $n \to \infty$*

$$\left(\frac{n}{J_n}\right)^{1/2} (\tilde{c}_B - c_B) \xrightarrow{d} \mathcal{N}\left(0, \frac{c(u_1, u_2)}{\prod_{s=1}^{2}(u_s(1 - u_s))^{1/2}}\right) \tag{3.2.27}$$

Then the asymptotic mean squared error of $\tilde{c}_B$ is

$$MSE(\tilde{c}_B) = Bias(\tilde{c}_B)^2 + Var(\tilde{c}_B)$$
$$\approx \frac{C_1}{J_n^2} + \frac{C_2 J_n}{n}$$

The optimal order $J_n$ of the polynomials follows by minimizing the $MSE(\tilde{c}_B)$ with respect to $J_n$ and set $\left(\frac{\partial}{\partial J_n}\right) MSE(\tilde{c}_B) = 0$. We have $J_o = n^{1/3}$. Combine with the assumption at the beginning, the optimal value of $J_o$ is close to including $n^\alpha$ and $1/3 < \alpha < 1/2$.

### 3.2.3 Sieve MLE

The infinite-dimensional unknown parameters in a nonparametric model are often considered as one of some function space having certain regularity like bounded second derivatives, monotone conditions, or concave requirements. Sieves provide accurate and reckonable ways to approximate any unknown functions. It can be conducted through linear spans of power series, Fourier series, splines, or many other basic functions to provide a numerical method dealing with the problems in economics and finance(Judd, 1998; Chen, 2007). The approximation space can usually be captured by a series of 'parameters', which means that a nonparametric estimation problem can be simplified to a parametric one when the method of sieves is implemented. Nonetheless, to study the properties of the estimator efficiently, as the sample size increases, it is necessary to require the number of parameters to grow slowly.

Let $c_o(u_1, u_2)$ be the true copula density and it belongs to $\mathcal{B}$. A space $\mathcal{B}_n$ is called a sieve space for $\mathcal{B}$ if for any copula density function $c \in \mathcal{B}$, there exists an element $\prod_n c \in \mathcal{B}_n$ such that $d(c, \prod_n c) \to 0$ as $n \to \infty$, where d is a metric on $\mathcal{B}_n$.

$$\mathcal{B}_n = \left\{ c_{J_n}(u_1, u_2) = J_n^2 \sum_{j_1=0}^{J_n-1} \sum_{j_2=0}^{J_n-1} \gamma_{j_1,j_2} p_{j_1}(u_1) p_{j_2}(u_2), \quad \int_0^1 \int_0^1 c_{J_n}(u_1, u_2) du_1 du_2 = 1 \right\},$$

$$J_n \to \infty, \frac{J_n}{n} \to 0.$$

where $p_{j_s}(u_s)$ is the beta density function and $\{\gamma_{j_1,j_2} : j_1 \geq 0, j_2 \geq 0\}$ is the collection of unknown sieve coefficients.

The sieved MLE is obtained by maximizing the likelihood function on a proper subset of the parameter space. Suppose $J_n$ is an increasing sequence and let

$\hat{c}_{J_n}(u_1, u_2) = J_n^2 \sum_{j_1=0}^{J_n-1} \sum_{j_2=0}^{J_n-1} \hat{\gamma}_{j_1,j_2} p_{j_1}(u_1) p_{j_2}(u_2) = B(:, J_n, \hat{\gamma}_{J_n})$ where $\hat{\gamma}_{J_n}$ is the maximum likelihood estimator for $\gamma_{J_n}$. Then, $\hat{c}_{J_n}(u_1, u_2)$ is called a sieve maximum likelihood estimator. Let $l(c, U_i)$ denote the contribution of i-th observation to the log-likelihood function.

$$l(c, U_i) = \log \{c(U_{1i}, U_{2i})\}$$

and,

$$\hat{c}_{J_n}(u_1, u_2) = \arg \max_{c \in \mathcal{B}_n} \sum_{i=1}^{n} l(c, U_i) \tag{3.2.28}$$

$\mathcal{F}_n = \cup_{r=1}^{J_n} \mathcal{B}_r$ be a sieve, which consists of the space of all Bernstein densities of the order $J_n$ or less, where $J_n$ is a sequence of integers tending to infinity. If the true copula density $c_0 = B(:, J_0, \gamma_0)$ for some $J_0$ and $\gamma_0 \in \Delta_J$, where

$$\Delta_J = \left\{ (x_{1,1}, x_{1,2}, ..., x_{J,J}) : 0 \leq x_{j_1,j_2} \leq 1, \sum_{j_1=0}^{J-1} \sum_{j_2=0}^{J-1} x_{j_1,j_2} = 1, \sum_{j_s=0}^{J-1} x_{j_s} = \frac{1}{J} \right.$$
$$\left. j_s = 0, 1, ..., J-1, s = 1, 2 \right\}$$

where $c_0(u_1, u_2)$ itself is a Bernstein copula density, then

$$d(\hat{c}(u_1, u_2), c_0(u_1, u_2)) = \mathcal{O}_P(J_n^{1/2} n^{-1/2} (\log n)^{1/2}). \tag{3.2.29}$$

In particular, when we set the $J_n$ to grow arbitrarily slowly, the convergence rate can be close to $n^{-1/2}(\log n)^{1/2}$. If the true copula density $f_0$ is not of the Bernstein type, but it is bounded away from 0 and has a bounded second derivative, then for the choice

$$c_1 n^{1/3}(\log n)^{-1/3} \leq J_n \leq c_2 n^{1/3}(\log n)^{1/3},$$

where $c_1$ and $c_2$ are constants, we have $d(\hat{c}(u_1, u_2), c_0(u_1, u_2)) = \mathcal{O}_P(n^{-1/3}(\log n)^{1/3})$ (see Ghosal et al. (2001); Wong and Shen (1995)).

### 3.2.4   Penalized maximum likelihood with adapted LASSO

The construction of the penalty term aims to enhance the prediction accuracy and interpretability of the model. In conjunction with the sieve method discussed in the previous section, the estimates of the nonparametric Bernstein copula density coefficients, denoted as $\hat{\gamma}$, are obtained by maximizing the penalized log-likelihood function:

$$\hat{\gamma} = \arg \max_{\gamma \in \Delta_J} l_\lambda(\gamma).$$

Here, $l_\lambda$ represents the penalized log-likelihood function defined as:

$$l_\lambda(\gamma) = l(\gamma) - N\lambda P(\gamma).$$

In this equation, $l(\gamma)$ corresponds to the log-likelihood function, $P(\cdot)$ is a penalty function, $N$ represents the sample size, and $\lambda$ is a nonnegative regularization parameter. The LASSO (Least Absolute Shrinkage and Selection Operator) regularization technique is a suitable approach for simultaneous estimation and variable selection(Tibshirani, 1996) and it produce sparse solutions for some $\lambda$, i.e., some of the coefficients of the nonparametric Bernstein copula density estimator $\gamma_{j_1 j_2}$ can be estimated as exactly zero. The LASSO penalty is set as

$$P(\gamma) = \|\gamma\|_1 = \sum_{\boldsymbol{j}} |\gamma_{\boldsymbol{j}}|$$

The continuous shrinkage reduces the estimation variance and improves the prediction accuracy of the statistical model due to the bias-variance trade-off. However, the traditional lasso, which penalizes the coefficients equally is not always consistent in variable selection (Zou, 2006; Zhao and Yu, 2006; Zhang, 2010). To address this limitation, the adaptive Lasso was proposed by Zou (2006) and its

the penalty term is expressed as:

$$P(\gamma) = \sum_{\boldsymbol{j}} w_{\boldsymbol{j}} |\gamma_{\boldsymbol{j}}|, \tag{3.2.30}$$

where the $\boldsymbol{w}$ is a known weights vector. In the regression model, Fan and Li (2001) introduced the concept of oracle procedures to describe coefficient estimators $\hat{\boldsymbol{\gamma}}(\delta)$ obtained from fitting procedures $\delta$, such as maximum likelihood estimation (MLE), ordinary least squares (OLS). If the following conditions can be satisfied, the procedure $\delta$ is considered an oracle procedure:

1. Identifies the right subset model, $\{j : \hat{\gamma}_j \neq 0\} = \mathcal{A}$

2. Has the optimal estimation rate, $\sqrt{n}(\hat{\boldsymbol{\gamma}}(\delta)_{\mathcal{A}} - \boldsymbol{\gamma}_{\mathcal{A}}^*) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\Sigma}^*)$, where $\boldsymbol{\Sigma}^*$ is the covariance matrix knowing the true subset model.

Here, $\mathcal{A} = \{j : \gamma_j^* \neq 0\}$ and also assume that $|\mathcal{A}| = p_0 < p$. p is the number of independent variables in the regression model. There is a consensus in the literature (Fan and Li, 2001; Fan and Peng, 2004) that a good fitting procedure should possess oracle properties, indicating its desirable performance in statistical estimation. These oracle properties ensure that the procedure achieves asymptotic consistency, correct variable selection, and efficient estimation. Additionally, it has been recognized that an optimal procedure should exhibit continuous shrinkage, allowing for effective regularization and improved estimation accuracy. When the weights are data-dependent and appropriately selected, the adaptive lasso 3.2.30 has the oracle property and the continuous shrinkage property for our proposed estimator is shown in the simulation study. Suppose that $\hat{\gamma}_e(SMLE)$ is the sieve maximum likelihood estimates in empirical nonparametric Bernstein copula model. It is obtained by maximizing $\frac{1}{N} \sum_{n=1}^{N} \log \sum_{j_1=0}^{J-1} \sum_{j_2=0}^{J-1} \gamma_{\mathbf{j}} \prod_{s=1}^{2} \beta(u_s; j_s, J - j_s)$. When there is a small number of data points concentrating in a specific square $\left[\frac{j_1}{J}, \frac{j_1+1}{J}\right] \times \left[\frac{j_2}{J}, \frac{j_2+1}{J}\right]$, the corresponding coefficient of the empirical Bernstein copula model,

denoted as $\hat{\gamma}_{e\ j_1 j_2}$, tends to be relatively small. We construct the weight sector as

$$\hat{w}_{j_1 j_2} = \frac{1}{|\hat{\gamma}_{e\ j_1 j_2}(SMLE)|} \quad j_s = 0, ..., J-1, \forall s = 1, 2. \tag{3.2.31}$$

As a result, the weight vector $w$ tends to be large, indicating a stronger penalty on the coefficient. This choice of the weights vector implies that coefficients associated with predictors with fewer data points will experience greater shrinkage. In other words, the weights vector assigns higher importance to squares that have a smaller number of data points, leading to increased regularization for those coefficients. Then, the adapted LASSO estimates in the nonparametric Bernstein copula model are given by

$$\hat{\gamma}(SMLE) = \arg\max_{\gamma} \left\{ \frac{1}{N} \sum_{n=1}^{N} \log \sum_{j_1=0}^{J-1} \sum_{j_2=0}^{J-1} \gamma_{j_1 j_2} \beta(u_1; j_1, J-j_1) \beta(u_2; j_2, J-j_1) \right.$$
$$\left. -\lambda \sum_{j_1=0}^{J-1} \sum_{j_2=0}^{J-1} \frac{\gamma_{j_1,j_2}}{\hat{w}_{j_1 j_2}} \right\}$$
$$\tag{3.2.32}$$

### 3.2.4.1 Other model to compare

We also compare our adaptive LASSO estimator with a different weight vector setting. Inspired by the empirical Bernstein copula model introduced by Sancetta and Satchell (2004), we adopt a weight setting based on the empirical copula. In the empirical Bernstein copula framework, the coefficients of the estimator are set to the empirical copula values. The empirical copula function is constructed by ranking the observations within each variable and calculating the joint probability based on the ranks. It is represented by the formula 3.2.25. Additionally, the density of the empirical copula can be obtained by counting the number of

observations in each grid. It can be expressed as follows:

$$\frac{1}{N} \sum_{i=1}^{N} \mathbb{I}_{\mathbf{u}_i \in H_{\mathbf{s}}},$$  (3.2.33)

Here, $\mathbf{u}_i = (u_{1i}, u_{2i})$, $H_{\mathbf{s}} = [\frac{j_1}{J}, \frac{j_1+1}{J}] \times [\frac{j_2}{J}, \frac{j_2+1}{J}]$, $\mathbb{I}_{...}$ is the indicator function, and $N$ represents the number of observations for each variable. The empirical copula, denoted as $c_B(\mathbf{u})$, can be viewed as a smooth copula histogram utilizing the beta density as the smoothing function.

Given a set of observations $(u_{1i}, u_{2i})$ for $i = 1, ..., N$, the estimates for the empirical copula coefficients are calculated as follows:

$$\hat{\alpha}_{j_1 j_2} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}_{\{\frac{j_1}{J} \leq u_{1i} \leq \frac{j_1+1}{J}, \frac{j_2}{J} \leq u_{2i} \leq \frac{j_2+1}{J}\}} \quad j_s = 0, ..., J-1, \forall s = 1, 2 \quad (3.2.34)$$

To construct the weight vector, we use the reciprocal of the empirical copula coefficients:

$$w_{j_1, j_2} = \frac{1}{\hat{\alpha}_{j_1, j_2}}.$$  (3.2.35)

Then, we obtain another nonparametric Bernstein copula estimator by incorporating the reciprocal of the empirical copula as a penalty term.

### 3.2.5 Goodness-of-fit test

In this chapter, our primary focus is on studying the estimation and properties of copula functions. While assuming the marginal distributions as known, it is crucial to ensure the correct specification of the marginal distributions in the empirical study(Tsukahara, 2005). To achieve this, we employ three statistical tests to assess the goodness-of-fit of the marginal distributions: the Kolmogorov-Smirnov (KS) test, Cramer-von Mises (CvM) test, and Anderson-Darling (AD) test, which are the common approach to comparing an empirical distribution with a theoretical

distribution is to measure the distance between these two distributions. All three tests are non-parametric, meaning they do not require any assumptions about the specific form of the underlying distribution. They can be applied to a wide range of distributions without relying on distributional assumptions.

The KS statistic, denoted as $\mathcal{T}_{KS}$, measures the maximum difference between the empirical cumulative distribution function (ECDF) $F_n(x)$ and the hypothesized cumulative distribution function $F(x)$.

$$\mathcal{T}_{KS} = \sup_S [|F_n(x) - F(x)|] \tag{3.2.36}$$

where $F_n(x) = n^{-1} \#\{1 \leq k \leq n : X_k \leq x\}$ $x \in \mathbb{R}$. It is particularly sensitive to deviations in the bulk of the distribution and less sensitive to tail deviations due to the lower variance of the ECDF (Malevergne and Sornette, 2003). Besides, the KS test treats all differences between ECDF and the hypothetical distribution equally, without any weighting. The CvM statistic, denoted as $\mathcal{T}_{CvM}$, is obtained by integrating the squared differences between $F_n(x)$ and $F(x)$, with a greater emphasis on highly concentrated areas.

$$\mathcal{T}_{CvM} = n \int_{-\infty}^{+\infty} (F_n(x) - F(x))^2 \, dF(x) \tag{3.2.37}$$

However, similar to the KS test, it is also less sensitive to tail deviations. The AD statistic, denoted as $\mathcal{T}_{AD}$, belongs to the Cramer-von Mises family and exhibits unbiasedness both at the center and tails of the distribution. It incorporates the ratio of the squared differences between $F_n(x)$ and $F(x)$ to the product of $F(x)$ and $1 - F(x)$(Anderson and Darling, 1952).

$$\mathcal{T}_{AD} = n \int_{-\infty}^{+\infty} \frac{(F_n(x) - F(x))^2}{F(x)(1 - F(x))} dF(x) \tag{3.2.38}$$

Each of these statistics possesses its own strengths and limitations. By employing all three tests, we aim to ensure the correct specification of the marginal distributions. The empirical evidence section will present the results of these tests, providing a comprehensive evaluation of the goodness-of-fit of the marginal distributions.

## 3.3   Monte Carlo simulation

In this section, the finite sample behaviors of the proposed model are investigated. It is known that the Bernstein copula has the property to represent all kinds of copulas. The first part shows the comparison between the empirical Bernstein copula, the proposed model (penalized Bernstein copula with empirical Bernstein copula penalty), and some popular parametric models. Here, we select four types of copula: Gaussian, Frank, Gumbel, and Clayton copula as the target copula. They are four different kinds of parametric copulas.

Then, we consider two cases for the dependence parameters to study whether the penalty term can accurately impose sparsity. First, we construct a sparse scenario in which most values parameters are set to exactly zero. For comparison, we set the second case as a less sparse scenario with more nonzero values. Four relative models are compared: empirical Bernstein copula, penalized Bernstein copula with LASSO penalty, penalized Bernstein copula with empirical copula penalty and penalized Bernstein copula with empirical Bernstein copula penalty (our proposed model). The performance is measured in terms of the Integrated Mean Square Error(IMSE) which is often used for capturing the closeness between the estimated and actual pdf in nonparametric estimation applications. The IMSE

has the format as 3.3.2 :

$$IMSE(\hat{c}) = \int E_{\hat{c}}[\hat{c}_\gamma(\mathbf{u}) - c(\mathbf{u})]^2 d\mathbf{u} \qquad (3.3.1)$$

$$= \int [[bias(\hat{c}_\gamma(\mathbf{u}))]^2 + var(\hat{c}_\gamma(\mathbf{u}))]d\mathbf{u} \qquad (3.3.2)$$

The above formula shows that IMSE can be decomposed into the bias and variance terms, which means that both the bias and variance affect the quantity of IMSE. The IMSE can be approximated using the following equation:

$$IMSE(\hat{c}) \approx \frac{1}{I}\frac{1}{N}\sum_{i=1}^{I}\sum_{j=1}^{N}(\hat{c}(\mathbf{u}_j;\boldsymbol{\gamma}_i) - c(\mathbf{u}_j))^2 \qquad (3.3.3)$$

where $\mathbf{u} = (u_1, u_2) \in \{(0.01, 0.01), (0.01, 0.02), ..., (0.99, 0.99)\}$. $\hat{c}_\gamma(\mathbf{u})$ is the estimated copula density, N is the number of grid knots which is $9801(99 \times 99)$, and $c(\mathbf{u})$ is the target density. Similar to previous studies such as the bivariate copula density estimation with total variation penalized method (Qu and Yin, 2012) and flexible copula density estimation with penalized Hierarchical B-splines (Kauermann et al., 2013), we have chosen to use 100 iterations in our research. This number of iterations has been found to be sufficient for obtaining meaningful results and drawing reliable conclusions for penalized bivariate copula density estimation. In sections 3.3.3 and 3.3.4, we present the results that demonstrate the desirable properties of the nonparametric Bernstein copula model with adaptive LASSO penalty. Specifically, we show that the model exhibits continuous shrinkage, effectively enforces sparsity as the regularization parameter increases, and outperforms the empirical Bernstein copula.

## 3.3.1 Random number generating from Bernstein copula

To explore the estimation quality of the proposed model, generating random numbers from the Bernstein copula is an important step. The most straightforward

method is to do an inversion. Denote the inverse function of $C$ is by $C^{-1}$, which can be defined even if the copula function $C$ is not invertible.

$$C^{-1}(u) = \inf\{x \in \mathbb{R} : C(x) \geq u\}. \qquad (3.3.4)$$

However, the method of inversion in simulating a random variable with arbitrary cdf is not available in closed form for many cumulative distribution functions. An alternative method is the acceptance/rejection algorithm. Suppose we aim to simulate a pair of random variables $(X_1, X_2)$ with copula density function $f$. $f$ is a complicated function, which is difficult to sample. It can apply the acceptance/rejection method if we are able to simulate the other pair of random variables $(Y_1, Y_2)$ whose probability density function $g$ satisfies $f(x) \leq Mg(x)$ (Casella et al., 2004)):

$$M = \sup_{x \in \chi_f} \frac{f(x)}{g(x)} < \infty. \qquad (3.3.5)$$

Here, $\chi_f$ is the support of $f$. Combining the reject sampling method with the characteristics of the Bernstein copula, the multivariate uniform distribution can be selected as the candidate density function $g$. In bivariate case, $g(x, y) = \frac{1}{area\ of\ S}, (x, y) \in S$, for $\mathbf{u} \in [0, 1]^2$, $g(\mathbf{u}) = 1$. Then, the boundary constant

$$M^* = \sup_{\mathbf{u} \in [0,1]^2} \frac{c_B(\mathbf{u})}{g(\mathbf{u})} = \sup_{\mathbf{u} \in [0,1]^2} c_B(\mathbf{u}). \qquad (3.3.6)$$

The probability of acceptance in each attempt is

$$P(U \leq f(Y_1, Y_2)/Cg(Y_1, Y_2)) = \int_0^1 \int_0^1 \frac{f(y_1, y_2)}{C} dy_1 dy_2 = \frac{1}{C}. \qquad (3.3.7)$$

Since the attempts are mutually independent, it follows the number of candidates generated until one is accepted. If the candidate multivariate distribution is close enough to the target one, the constant C will be close to 1. In extremely high-

dimension cases, the rejections will increase due to the curse of dimensionality. In other words, the ratio C will tend towards zero, then the method is infeasible in that situation. However, it is efficient enough to generate random numbers for a comparably lower dimensional case which is what we studied. The explicit simulation algorithm is as follows(Pfeifer et al., 2020; Liu and Prokhorov, 2016):

1. Generate $(u_1, u_2)$ from multivariate uniform distribution

2. Independently simulate $u_3$ from uniform distribution, $u_3 \sim Unif(0, 1)$

3. If $u_3 < \frac{c_B(u_1, u_2; \gamma)}{M^*}$, $u_1$ and $u_2$ are the pairs of random numbers from Bernstein copula. If not, turn to step 1 until the conditions are satisfied.

### 3.3.2   Property of representing parametric copulas

For the empirical Bernstein copula, dimension $J$ is the parameter that controls the smoothness. As $J \to \infty$, the empirical Bernstein copula interpolates all the points and tends to be an empirical copula. In this section, we need to select the tuning parameter $\lambda$ for the optimal penalized Bernstein copula model and $J$ for the empirical Bernstein copula model. Here, we set $J \in [2, 15]$, for each $J$, and repeat the procedure stated in section 3.3.1 to select the optimal tuning parameter for each J. Then, choose the dimension of coefficients $J$ with lowest IMSE.

The sample size for each simulation is set as n =1000. Generating random numbers from the elliptical copulas utilizes the inversion method. Figure 3.3.1 shows that even the IMSEs of estimating these three copulae are different, the trends are almost the same. As J goes up, the IMSEs are decreasing first and then increase. The line of penalized Bernstein copula(with empirical Bernstein copula penalty)'s IMSE is always under the line of empirical Bernstein copula's IMSE. Also, as the dimension of coefficients increases, the difference becomes larger. In other words,

the penalized model provides more accuracy compared with the empirical one for a higher dimension of coefficients. Besides, plot (a) in Figure 3.3.1 shows the IMSE for EBC with dimension J =10 is similar to the IMSE for PBC with dimension J=8. It means that for the same level of accuracy, our penalized model requires a comparably lower dimension of coefficients since the optimal order of $J_n$ is in a range.



(a) Gaussian        (b) Student's t        (c) Clayton
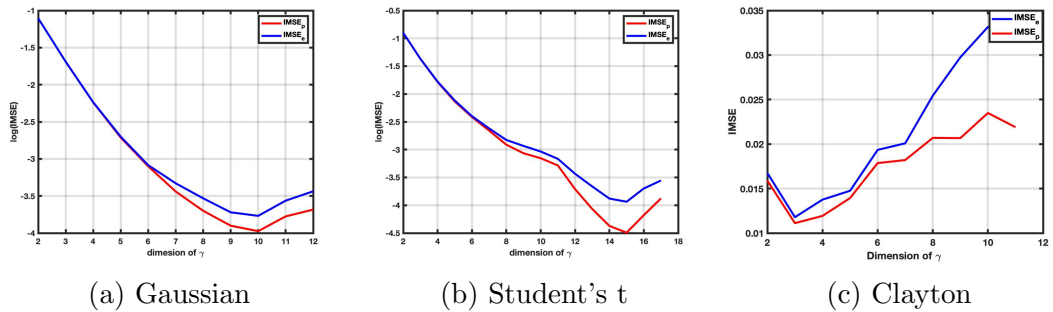
Figure 3.3.1: The relationship between IMSE and dimension of coefficients J. The red line shows the IMSE for penalized Bernstein copula model (with empirical Bernstein copula penalty) and the blue line represents the IMSE for the empirical Bernstein copula model. The true models are set as Gaussian copula($\rho = 0.5$), student's t copula ($\rho = 0.5, \nu = 10$) and Clayton copula($\theta = 0.25$) respectively. The optimal dimension of coefficients for Gaussian copula is 10, for student's t is 15, and for Clayton copula is 3.

The following table shows the simulation results for the optimal empirical Bernstein copula model (EBC), the optimal penalized Bernstein copula model with empirical Bernstein copula (PBC) as the penalty, Gaussian, student t and three copulas from Archimedean family. The IMSE and loglikelihood in the table are the means of all the iterations and the number of iteration is I = 10,000. The parameters of empirical Bernstein copula and penalized model are determined by the cross-validation method. All the parameters of the copula models are obtained from the training sample. The out-of-sample measurements are calculated using the testing sample. The training and testing samples have the size N = 100 individually.

| Copula Model | True copula model | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | Gaussian copula ($\rho = 0.3$) | | Frank copula ($\alpha = 5$) | | Clayton copula ($\theta = 2$) | | Gumbel copula ($\rho = 3$) | |
| | IMSE | Log-likelihood | IMSE | Log-likelihood | IMSE | Log-likelihood | IMSE | Log-likelihood |
| Gaussian | 0.0533 | 5.1433 | 0.1491 | 23.7488 | 0.2789 | 32.136 | 0.2083 | 66.8479 |
| student t | 0.1096 | 5.3923 | 0.1423 | 24.3554 | 0.2696 | 32.3104 | 0.294 | 69.2569 |
| Frank | 0.0761 | 4.7669 | 0.0977 | 27.0693 | 0.2775 | 31.6095 | 0.2343 | 63.1465 |
| Gumbel | 0.0856 | 4.2855 | 0.1925 | 5.6371 | 0.3848 | 23.2794 | 0.0599 | 72.8933 |
| Clayton | 0.1022 | 3.8988 | 0.2869 | 17.558 | 0.0562 | 43.7078 | 0.5006 | 42.7886 |
| EBC | 0.0679 | 5.0217 | 0.1230 | 24.1757 | 0.2682 | 32.5504 | 0.2039 | 68.2561 |
| PBC | 0.0632 | 5.1330 | 0.1008 | 25.3069 | 0.2636 | 32.6677 | 0.2030 | 68.2673 |

Table 3.3.1: The table shows the IMSE and out-of-sample average loglikelihood of the estimated models.

The Gaussian and Frank copulas are symmetric, except themselves, penalized Bernstein copula has the lowest IMSE. The shapes of Gumbel copula and Clayton are totally different from all the others. They are capturing the lower tail and upper tail dependency, so they have relatively worse statistics for estimating others. The simulation results verify that if the parametric model is correctly specified, it can work accurately and efficiently, otherwise, it might just provide the wrong estimation.

### 3.3.3 Sparse scenario

As discussed in section 3.3, the coefficient matrix of the Bernstein copula can be interpreted as the empirical copula. It represents the proportion of data points in each grid. The $\gamma_{j_1,j_2}$ in the following table stands for the probability of data points falling in the grid $[\frac{j_1-1}{J}, \frac{j_1}{J}] \times [\frac{j_2-1}{J}, \frac{j_2}{J}]$ and $j_1, j_2 = 1, ..., J$.

| | | | |
|---|---|---|---|
| $\gamma_{11}$ | $\gamma_{12}$ | $\cdots$ | $\gamma_{1J}$ |
| $\gamma_{21}$ | $\gamma_{22}$ | $\cdots$ | $\gamma_{2J}$ |
| $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ |
| $\gamma_{J1}$ | $\gamma_{J2}$ | $\cdots$ | $\gamma_{JJ}$ |

Table 3.3.2: The coefficient matrix of the Bernstein copula model in the bi-variate case with the number of grids J in each dimension.

In the sparse scenario, we set the majority parameters($\frac{J-1}{J}$) to be zero as the true model. The data series from the true copula model is generated through the method discussed in section 3.3.1. Here, the coefficient matrix $\boldsymbol{\gamma}$ is fixed to be an $8{\times}8$ matrix. In other words, the number of grids $J$ in each dimension is equal to 8. In the bivariate case, $J^2(=64)$ parameters are needed to be estimated. Based on the algorithm, two groups of samples are generated. One is for the sparse scenario. It sets the majority of the true coefficients is zero as followed, the ratio of zero components is 56/64=0.8750.

$$
\begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.125 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.125 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.125 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.125 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.125 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.125 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.125 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.125 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
$$

The above matrix indicates that it is a copula with high negative dependence. After generating the random number from this specific Bernstein copula, figure 3.3.2 shows the empirical copula scatter plot.
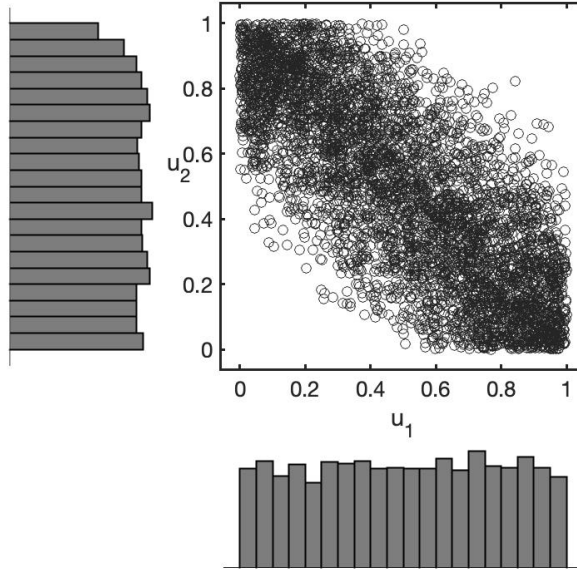
Figure 3.3.2: $u_1$ and $u_2$ are generated from the bi-variate Bernstein copula with the matrix above as the coefficients. The histograms show that both $u_1$ and $u_2$ have uniform marginal distributions. And the scatter plot exhibits a negative correlation between $u_1$ and $u_2$.

To select the optimal tuning parameter properly, the k-fold cross-validation method for penalized Bernstein copula density estimation is applied as follows:

1. Randomly partitions the original data sample into k sub-samples with roughly equal size. Of the k sub-samples, a single sub-sample is retained as the validation data for testing the model, and the remaining (k-1) sub-samples are used as training data.

2. Set $\lambda \in [\lambda_0, \lambda_1]$. Then, estimate $\hat{\gamma}_{\lambda_0}$ and $\hat{\gamma}_{\lambda_1}$ using the training data with $\lambda_0$ and $\lambda_1$ respectively.

3. Calculate the log-likelihood based on the following function:

$$\hat{L}_{\lambda_m} = \frac{k}{N} \sum_{n=1}^{N/k} \log \sum_{\boldsymbol{j}}^{\boldsymbol{J}} \hat{\boldsymbol{\gamma}}_{\boldsymbol{j}\lambda_m} \beta(u_{1n}, u_{2n}, \boldsymbol{j}) \ for \, m = 0, 1. and \, \boldsymbol{j} = [j_1, j_2] \quad (3.3.8)$$

using the validation data.

4. Repeat the step 2 and 3 for $K \times I$ times and get the average log-likelihood for the $\lambda_m$:

$$\overline{L}_{\lambda_m} = \frac{1}{I \times K} \sum_{i=1}^{I} \sum_{k=1}^{K} \hat{L}_{\lambda_m} \quad for \ \ m = 0, 1. \qquad (3.3.9)$$

The cross-validation log-likelihood is the average log-likelihood across all the observations in the testing sample.

5. If $\overline{L}_{\lambda_0} > \overline{L}_{\lambda_1}$ ,update the $\lambda_1 = \lambda_1 - \delta$.Otherwise, update the $\lambda_0 = \lambda_0 + \delta$. $\delta = (\lambda_1 - \lambda_0)/10$ is the increment.

6. Repeat step 2-5 until the accuracy reaches the setting.

7. The optimal tuning parameter $\lambda$ is chosen when the average log-likelihood becomes smallest.

The following plot shows that as the $\lambda$ increases, the log-likelihood increases first, and after reaching the largest value, it keeps as a constant. The trend of this curve is due to both the conditions on the Bernstein copula function, indicated in section 3.3.2, and the adapted lasso penalty. To keep the marginals uniform in the copula function, it requires that the sum of each row and each column in the coefficient matrix are the same. When the tuning parameter $\lambda$ is larger than some value, the penalty on the coefficient matrix pushes $(J \times (J - 1))$ parameters to zero. Only $J$ parameters are nonzero with the same value $1/J$. Here, it is a sparse scenario setting, and only J parameters are nonzero in the true model. Thus, the penalized Bernstein copula model is optimal when all the coefficients are penalized to the utmost.
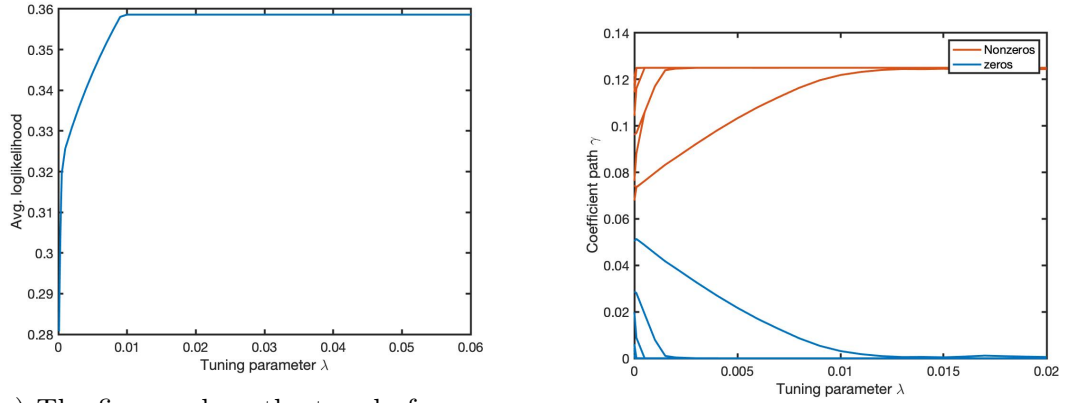
(a) The figures show the trend of average loglikelihood as the tuning parameter $\lambda$ increases. The results are based on the random sample with a sample size n = 5000. The Avg. loglikelihood is the average loglikelihood of the test sample in 5-fold cross-validation with iterations I = 100.

(b) The figures show the path of coefficient $\gamma_{j_1,j_2}$ as the tuning parameter $\lambda$ increases. The results are based on a random sample with sample size n = 5000, the number of coefficients in each dimension J = 8.

Figure 3.3.3: The loglikelihood curve and coefficient path for the sparse scenario.

The contour plots 3.3.4 and surface plots 3.3.5 are the best estimate for each model. Compared with the truth, the penalized Bernstein copula with empirical copula penalty(PBCEC) and the penalized Bernstein copula with LASSO penalty models are far more similar to the truth.
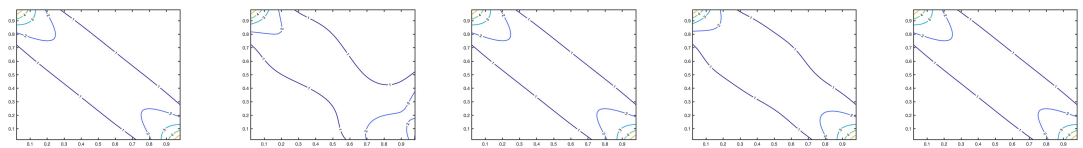


(a) True Copula     (b) EBC     (c) PBC LASSO     (d) PBC EC     (e) PBC EBC

Figure 3.3.4: The Contour plots for sparse scenario
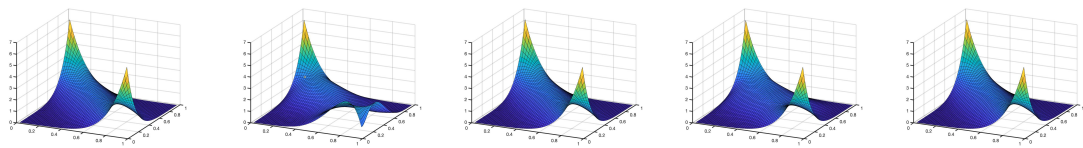


(a) True Copula     (b) EBC     (c) PBC LASSO     (d) PBC EC     (e) PBC EBC

Figure 3.3.5: The Surface plots for sparse scenario

66

### 3.3.4 Less sparse scenario

The other design is for the less sparse scenario. More non-zero elements are set in the coefficient matrix. It contains only 24 zero components and they are all in the corner. The ratio of zero components is $24/64 = 0.3750$. It still represents a negative dependence.

$$
\begin{bmatrix}
0 & 0 & 0 & 0 & 0.011 & 0.001 & 0.02 & 0.093 \\
0 & 0 & 0 & 0 & 0.001 & 0.015 & 0.088 & 0.021 \\
0 & 0 & 0.001 & 0.001 & 0.01 & 0.09 & 0.015 & 0.008 \\
0 & 0 & 0.001 & 0.024 & 0.08 & 0.015 & 0.002 & 0.003 \\
0.015 & 0.002 & 0.008 & 0.08 & 0.017 & 0.003 & 0 & 0 \\
0.028 & 0.005 & 0.07 & 0.015 & 0.006 & 0.001 & 0 & 0 \\
0.006 & 0.105 & 0.01 & 0.004 & 0 & 0 & 0 & 0 \\
0.076 & 0.013 & 0.035 & 0.001 & 0 & 0 & 0 & 0
\end{bmatrix}
$$
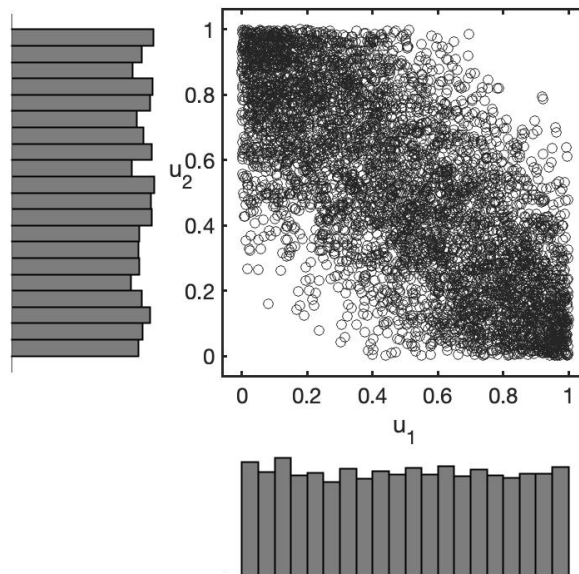


Figure 3.3.6: Scatter plot for random number generated by the given less sparse parameters of Bernstein copula

The estimated result of the empirical Bernstein copula shows that none of the coefficients is exactly zero, even though there are some not trivial values in the corner. After applying the cross-validation method to select the tuning parameter, the log-likelihood curve is shown in figure 3.3.7a. In the less sparse situation, the log-likelihood increases first, then decreases and finally stays constant. Comparing with the sparse scenario, the log-likelihood experiences a decreasing part, which is different from the previous curve. It is because the optimal model is not extremely sparse this time. When the tuning parameter is quite large, the proposed model will suffer from over-regularized and lose some important information from the trend.
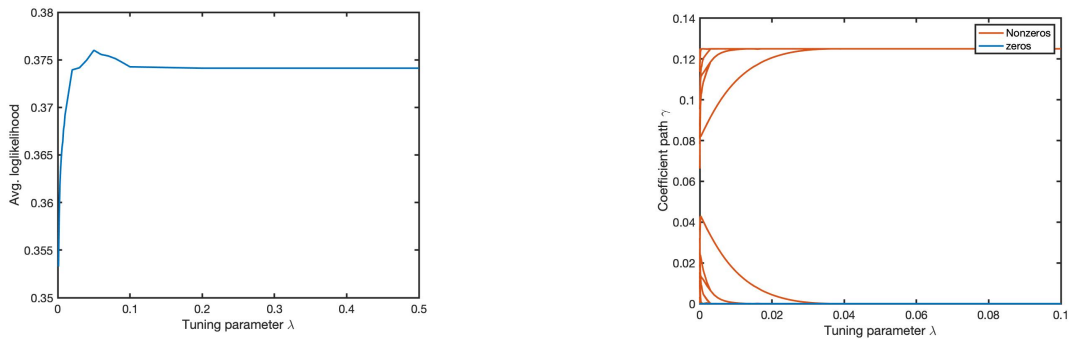


(a) The figures show the trend of average loglikelihood as the tuning parameter $\lambda$ increases. The results are based on the random sample with a sample size n = 5000. The Avg. loglikelihood is the average loglikelihood of the test sample in 5-fold cross-validation with iterations I = 100.

(b) The figures show the path of coefficient $\gamma_{j_1,j_2}$ as the tuning parameter $\lambda$ increases. The results are based on a random sample with sample size n = 5000, the number of coefficients in each dimension J = 8.

Figure 3.3.7: The loglikelihood curve and coefficient path for the less sparse scenario.

The table 3.3.3, 3.3.4, 3.3.5 and 3.3.6 show the model comparison between the Bernstein copula, penalized Bernstein copula with LASSO penalty, penalized Bernstein copula with empirical copula penalty and penalized Bernstein copula with empirical Bernstein copula penalty. The number of iterations is 100 with the 5-fold cross-validation method. As the sample size increases, the empirical

Bernstein copula's sparsity ratio reaches 0 first no matter in which scenario. In a small sample with a small order of polynomials, the IMSE and log likelihood of these four models are almost the same. As the order of polynomials increases, the LASSO penalized estimator fails to impose sparsity in the sparse scenario. All the results in the table are the average over iterations.



(a) True Copula   (b) EBC   (c) PBC LASSO   (d) PBC EC   (e) PBC EBC
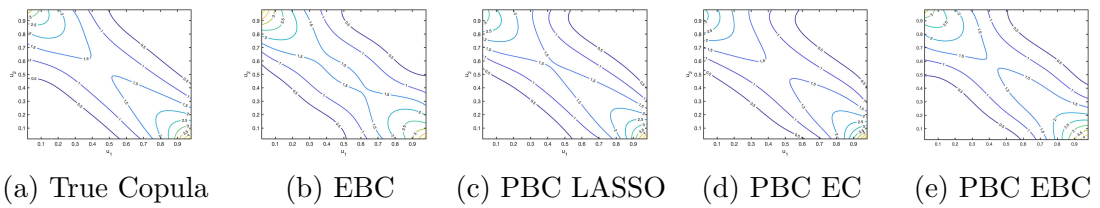
Figure 3.3.8: The Contour plots for less sparse scenario



(a) True Copula   (b) EBC   (c) PBC LASSO   (d) PBC EC   (e) PBC EBC
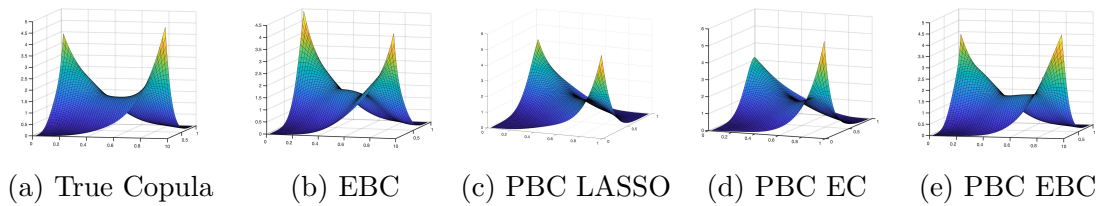
Figure 3.3.9: The Surface plots for less sparse scenario

**J = 4, n = 50, k = 5, I = 100**

| Model Type | Sparse Scenario | | | | | Less Sparse Scenario | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Tuning parameter | Avg. Log-likelihood | IMSE | Kullback-Leilur Distance | Sparsity (Ratio of Zeros) | Tuning parameter | Avg. Log-likelihood | IMSE | Kullback-Leilur Distance | Sparsity (Ratio of Zeros) |
| Empirical Bernstein copula | 0 | 0.2810 | 0.2889 | 0.1189 | 0.1290 | 0 | 0.3027 | 0.1901 | 0.048 | 0.1515 |
| Penalized Bernstein copula with LASSO penalty | 0.14 | 0.2810 | 0.2889 | 0.1189 | 0.1406 | 0.02 | 0.3027 | 0.1901 | 0.0479 | 0.1443 |
| Penalized Bernstein copula with Empirical copula penalty | 1e-18 | 0.2810 | 0.2889 | 0.1189 | 0.2926 | 2e-18 | 0.3027 | 0.1901 | 0.0479 | 0.152 |
| Penalized Bernstein copula with Empirical Bernstein copula penalty | 0.008 | 0.2834 | 0.2844 | 0.1130 | 0.6184 | 0.0321 | 0.3137 | 0.1738 | 0.0383 | 0.6346 |

**J = 8, n = 50, k = 5, I = 100**

| Model Type | Sparse Scenario | | | | | Less Sparse Scenario | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Tuning parameter | Avg. Log-likelihood | IMSE | Kullback-Leilur Distance | Sparsity (Ratio of Zeros) | Tuning parameter | Avg. Log-likelihood | IMSE | Kullback-Leilur Distance | Sparsity (Ratio of Zeros) |
| Empirical Bernstein copula | 0 | 0.3004 | 0.1663 | 0.0766 | 0.0098 | 0 | 0.2995 | 0.1652 | 0.0479 | 0.0109 |
| Penalized Bernstein copula with LASSO penalty | 0.01 | 0.3004 | 0.1663 | 0.0766 | 0.4519 | 1e-4 | 0.2995 | 0.1579 | 0.0471 | 0.0111 |
| Penalized Bernstein copula with Empirical copula penalty | 5e-18 | 0.3004 | 0.1663 | 0.0766 | 0.0072 | 5e-18 | 0.2996 | 0.1628 | 0.0479 | 0.6179 |
| Penalized Bernstein copula with Empirical Bernstein copula penalty | 0.009 | 0.3586 | 0.1402 | 0.0553 | 0.8502 | 0.02 | 0.3727 | 0.1604 | 0.0428 | 0.6897 |

**J = 16, n = 50, k = 5, I = 100**

| Model Type | Sparse Scenario | | | | | Less Sparse Scenario | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Tuning parameter | Avg. Log-likelihood | IMSE | Kullback-Leilur Distance | Sparsity (Ratio of Zeros) | Tuning parameter | Avg. Log-likelihood | IMSE | Kullback-Leilur Distance | Sparsity (Ratio of Zeros) |
| Empirical Bernstein copula | 0 | 0.2416 | 0.2874 | 0.1829 | 0.0123 | 0 | 0.2339 | 0.2918 | 0.1135 | 0.0197 |
| Penalized Bernstein copula with LASSO penalty | 0.03 | 0.2436 | 0.2863 | 0.1826 | 0.2536 | 7e-5 | 0.2342 | 0.2916 | 0.1132 | 0.0197 |
| Penalized Bernstein copula with Empirical copula penalty | 3e-5 | 0.2472 | 0.2835 | 0.1816 | 0.6447 | 6.00E-18 | 0.2339 | 0.2918 | 0.1135 | 0.0287 |
| Penalized Bernstein copula with Empirical Bernstein copula penalty | 0.02 | 0.4307 | 0.2824 | 0.1506 | 0.8163 | 9.6e-3 | 0.4369 | 0.2793 | 0.1003 | 0.8502 |

Table 3.3.3: The table shows the estimation results for four comparison models. The Estimation result is through a 5-fold cross-validation method with 100 iterations. The order of polynomials is set as 4,8 and 16, sample size n is 50.

**J = 4, n = 100, k = 5, I = 100**

| Model Type | Sparse Scenario | | | | | Less Sparse Scenario | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | Tuning parameter | Avg. Log-likelihood | IMSE | Kullback-Leibler Distance | Sparsity (Ratio of Zeros) | Tuning parameter | Avg. Log-likelihood | IMSE | Kullback-Leilur Distance | Sparsity (Ratio of Zeros) |
| Empirical Bernstein copula | 0 | 0.2805 | 0.2831 | 0.1115 | 0.1375 | 0 | 0.3073 | 0.1765 | 0.0391 | 0.1765 |
| Penalized Bernstein copula with LASSO penalty | 1e-7 | 0.2805 | 0.2831 | 0.1115 | 0.1367 | 0.002 | 0.3073 | 0.1765 | 0.0391 | 0.1955 |
| Penalized Bernstein copula with Empirical copula penalty | 2e-11 | 0.2805 | 0.2831 | 0.1115 | 0.1403 | 9e-11 | 0.3073 | 0.1764 | 0.0391 | 0.2219 |
| Penalized Bernstein copula with Empirical Bernstein copula penalty | 0.005 | 0.2811 | 0.2821 | 0.1102 | 0.6204 | 0.0543 | 0.3102 | 0.1703 | 0.0359 | 0.6660 |

**J = 8, n = 100, k = 5, I = 100**

| Model Type | Sparse Scenario | | | | | Less Sparse Scenario | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | Tuning parameter | Avg. Log-likelihood | IMSE | Kullback-Leibler Distance | Sparsity (Ratio of Zeros) | Tuning parameter | Avg. Log-likelihood | IMSE | Kullback-Leilur Distance | Sparsity (Ratio of Zeros) |
| Empirical Bernstein copula | 0 | 0.3068 | 0.1378 | 0.0453 | 0.0010 | 0 | 0.3124 | 0.1237 | 0.0259 | 0.0046 |
| Penalized Bernstein copula with LASSO penalty | 0.1 | 0.3068 | 0.1378 | 0.0453 | 0.0011 | 1.00E-07 | 0.3124 | 0.1237 | 0.0259 | 0.0046 |
| Penalized Bernstein copula with Empirical copula penalty | 7e-7 | 0.3086 | 0.1397 | 0.0415 | 0.4279 | 9e-7 | 0.3143 | 0.1199 | 0.0231 | 0.4248 |
| Penalized Bernstein copula with Empirical Bernstein copula penalty | 0.0509 | 0.3400 | 0.1303 | 0.0354 | 0.6421 | 0.0111 | 0.333 | 0.1221 | 0.0245 | 0.7229 |

**J = 16, n = 100, k = 5, I = 100**

| Model Type | Sparse Scenario | | | | | Less Sparse Scenario | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | Tuning parameter | Avg. Log-likelihood | IMSE | Kullback-Leibler Distance | Sparsity (Ratio of Zeros) | Tuning parameter | Avg. Log-likelihood | IMSE | Kullback-Leilur Distance | Sparsity (Ratio of Zeros) |
| Empirical Bernstein copula | 0 | 0.267 | 0.2358 | 0.0993 | 0.006 | 0 | 0.2721 | 0.2341 | 0.0679 | 0.0142 |
| Penalized Bernstein copula with LASSO penalty | 0.32 | 0.267 | 0.2358 | 0.0993 | 0.0061 | 2.00E-06 | 0.2721 | 0.2341 | 0.0679 | 0.0142 |
| Penalized Bernstein copula with Empirical copula penalty | 6.00E-06 | 0.2764 | 0.2217 | 0.0987 | 0.5436 | 5.00E-05 | 0.2723 | 0.234 | 0.0676 | 0.3879 |
| Penalized Bernstein copula with Empirical Bernstein copula penalty | 0.0631 | 0.2981 | 0.2103 | 0.0836 | 0.7387 | 0.0423 | 0.2935 | 0.2236 | 0.0628 | 0.8633 |

Table 3.3.4: The table shows the estimation results for four comparison models. The Estimation result is through the 5-fold cross-validation method with 100 iterations. The order of polynomials is set as 4, 8 and 16, sample size n is 100.

**J = 4, n = 512, k = 5, I = 100**

| Model Type | Sparse Scenario | | | | | Less Sparse Scenario | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Tuning parameter | Avg. Log-likelihood | IMSE | Kullback-Leibur Distance | Sparsity (Ratio of Zeros) | Tuning parameter | Avg. Log-likelihood | IMSE | Kullback-Leibur Distance | Sparsity (Ratio of Zeros) |
| Empirical Bernstein copula | 0 | 0.2852 | 0.2822 | 0.1102 | 0.1101 | 0 | 0.3101 | 0.1706 | 0.0360 | 0.2076 |
| Penalized Bernstein copula with LASSO penalty | 1e-4 | 0.2852 | 0.2822 | 0.1102 | 0.1098 | 1.3e-4 | 0.3101 | 0.1706 | 0.0360 | 0.1989 |
| Penalized Bernstein copula with Empirical copula penalty | 1e-12 | 0.2852 | 0.2822 | 0.1102 | 0.1904 | 1e-11 | 0.3101 | 0.1706 | 0.360 | 0.2054 |
| Penalized Bernstein copula with Empirical Bernstein copula penalty | 2.1e-4 | 0.2852 | 0.2821 | 0.1101 | 0.6146 | 1.5e-3 | 0.3102 | 0.1702 | 0.0359 | 0.6898 |

**J = 8, n = 512, k = 5, I = 100**

| Model Type | Sparse Scenario | | | | | Less Sparse Scenario | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Tuning parameter | Avg. Log-likelihood | IMSE | Kullback-Leibur Distance | Sparsity (Ratio of Zeros) | Tuning parameter | Avg. Log-likelihood | IMSE | Kullback-Leibur Distance | Sparsity (Ratio of Zeros) |
| Empirical Bernstein copula | 0 | 0.3265 | 0.0968 | 0.0188 | 0 | 0 | 0.3309 | 0.0667 | 0.0074 | 0 |
| Penalized Bernstein copula with LASSO penalty | 1.00E-04 | 0.3265 | 0.0968 | 0.0188 | 0 | 2e-6 | 0.3309 | 0.0667 | 0.0074 | 3.13e-5 |
| Penalized Bernstein copula with Empirical copula penalty | 9.00E-10 | 0.3265 | 0.0968 | 0.0187 | 0.1983 | 1.00E-09 | 0.331 | 0.0662 | 0.0071 | 0.1570 |
| Penalized Bernstein copula with Empirical Bernstein copula penalty | 1.1e-6 | 0.3267 | 0.0954 | 0.0167 | 0.1931 | 3e-6 | 0.3313 | 0.0647 | 0.0070 | 0.1554 |

**J = 16, n = 512, k = 5, I = 100**

| Model Type | Sparse Scenario | | | | | Less Sparse Scenario | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Tuning parameter | Avg. Log-likelihood | IMSE | Kullback-Leibur Distance | Sparsity (Ratio of Zeros) | Tuning parameter | Avg. Log-likelihood | IMSE | Kullback-Leibur Distance | Sparsity (Ratio of Zeros) |
| Empirical Bernstein copula | 0 | 0.3167 | 0.1576 | 0.0332 | 0 | 0 | 0.3167 | 0.1352 | 0.0139 | 0 |
| Penalized Bernstein copula with LASSO penalty | 1.00E-03 | 0.3167 | 0.1576 | 0.0332 | 0 | 5.00E-05 | 0.3167 | 0.1352 | 0.0139 | 0 0.0142 |
| Penalized Bernstein copula with Empirical copula penalty | 7.00E-10 | 0.3167 | 0.1576 | 0.0331 | 0.2035 | 1.00E-08 | 0.3168 | 0.1346 | 0.0138 | 0.1986 |
| Penalized Bernstein copula with Empirical Bernstein copula penalty | 0.01 | 0.3294 | 0.1329 | 0.0306 | 0.7833 | 8.00E-05 | 0.3185 | 0.1329 | 0.013 | 0.2031 |

Table 3.3.5: The table shows the estimation results for four comparison models. The Estimation result is through the 5-fold cross-validation method with 100 iterations. The order of polynomials is set as 4,8 and 16, sample size n is 512.

**J = 4, n = 1000, k = 5, I = 100**

| Model Type | Sparse Scenario | | | | | Less Sparse Scenario | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Tuning parameter | Avg. Log-likelihood | IMSE | Kullback-Leilur Distance | Sparsity (Ratio of Zeros) | Tuning parameter | Avg. Log-likelihood | IMSE | Kullback-Leilur Distance | Sparsity (Ratio of Zeros) |
| Empirical Bernstein copula | 0 | 0.2873 | 0.2822 | 0.11021847 | 0.0242 | 0 | 0.3103 | 0.1703 | 0.0358901 | 0.235 |
| Penalized Bernstein copula with LASSO penalty | 2E-04 | 0.2873 | 0.2822 | 0.11208521 | 0.0983 | 1.00E-03 | 0.3103 | 0.1703 | 0.035887 | 0.2374 |
| Penalized Bernstein copula with Empirical copula penalty | 9E-13 | 0.2873 | 0.2822 | 0.11020996 | 0.0965 | 2E-12 | 0.3103 | 0.1703 | 0.035885648 | 0.2255 |
| Penalized Bernstein copula with Empirical Bernstein copula penalty | 1E-12 | 0.2873 | 0.2822 | 0.11020986 | 0.0242 | 1.30E-03 | 0.3104 | 0.1702 | 0.035857709 | 0.7165 |

**J = 8, n = 1000, k = 5, I = 100**

| Model Type | Sparse Scenario | | | | | Less Sparse Scenario | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Tuning parameter | Avg. Log-likelihood | IMSE | Kullback-Leilur Distance | Sparsity (Ratio of Zeros) | Tuning parameter | Avg. Log-likelihood | IMSE | Kullback-Leilur Distance | Sparsity (Ratio of Zeros) |
| Empirical Bernstein copula | 0 | 0.3317 | 0.0821 | 0.0131 | 0 | 0 | 0.336 | 0.0534 | 0.0047 | 0 |
| Penalized Bernstein copula with LASSO penalty | 0.003 | 0.3317 | 0.0821 | 0.0131 | 0.0974 | 1.00E-04 | 0.336 | 0.0533 | 0.0047 | 0.0701 |
| Penalized Bernstein copula with Empirical copula penalty | 1.00E-10 | 0.3318 | 0.0821 | 0.013 | 0.1092 | 1.00E-09 | 0.336 | 0.0531 | 0.0047 | 0.1111 |
| Penalized Bernstein copula with Empirical Bernstein copula penalty | 1.00E-03 | 0.3345 | 0.0667 | 0.0105 | 0.6768 | 3.00E-06 | 0.3362 | 0.0521 | 0.0046 | 0.1096 |

**J = 16, n = 1000, k = 5, I = 100**

| Model Type | Sparse Scenario | | | | | Less Sparse Scenario | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Tuning parameter | Avg. Log-likelihood | IMSE | Kullback-Leilur Distance | Sparsity (Ratio of Zeros) | Tuning parameter | Avg. Log-likelihood | IMSE | Kullback-Leilur Distance | Sparsity (Ratio of Zeros) |
| Empirical Bernstein copula | 0 | 0.3258 | 0.1311 | 0.0217 | 0 | 0 | 0.3271 | 0.1083 | 0.0076 | 0 |
| Penalized Bernstein copula with LASSO penalty | 0.0024 | 0.3258 | 0.1311 | 0.0217 | 0.0083 | 1.00E-04 | 0.3271 | 0.1083 | 0.0076 | 0 |
| Penalized Bernstein copula with Empirical copula penalty | 1.00E-09 | 0.326 | 0.1311 | 0.0216 | 0.1764 | 2.00E-08 | 0.3271 | 0.1082 | 0.0076 | 0.1089 |
| Penalized Bernstein copula with Empirical Bernstein copula penalty | 0.025 | 0.3470 | 0.1206 | 0.0186 | 0.723 | 5.00E-05 | 0.3273 | 0.1079 | 0.0071 | 0.1232 |

Table 3.3.6: The table shows the estimation results for four comparison models. The Estimation result is through the 5-fold cross-validation method with 100 iterations. The order of polynomials is set as 4,8 and 16, sample size n is 1000.

73

## 3.4   Empirical evidence

In this section, we apply the proposed model in measuring the dependence structure of financial markets. From previous research on the financial crisis, the turmoil often occurred in one region and then quickly extended to a wide range of economics. Forbes and Rigobon (2002) defines this phenomenon as the financial contagion, which shows a significant increase in cross-market linkages after a shock to one country. Thus, precisely measuring the dependence between the international markets is a crucial task in risk management. We choose four representative markets: the United States, Europe, Hong Kong, and Japan which play important roles in the financial crisis after the 1980s.

Abbara and Zevallos (2014) indicates that the linkage between different economics is a sophisticated nonlinear dependency, so many researchers employ the copula-based model to measure the co-movements.

### 3.4.1   Data description

The data used to represent the four major markets is the monthly index returns of Morgan Stanley Capital International (MSCI). The index returns span 39 years from January 1980 to February 2019 for a total of 470 observations expressed in US dollars.

Descriptive Statistics for the United States(US), Europe(EU), Hong Kong(HK), and Singapore(SG) returns are presented in Table 3.4.1. The market of the United States shows the highest mean and median in index returns and the market of Hong Kong has the highest volatility. The unconditional distributions of Europe, Hong Kong, and the United States monthly returns show negative skewness and

expected express kurtosis. Besides, the results of the Jarque-Bera test indicate that neither of these equity returns is normally distributed under 1% significance level. In addition, the pair EU-US has the highest linear correlation which reaches 0.744. The EU-HK, EU-SG, and HK-US pairs show a similar moderate correlation of coefficients and Kendall's tau. The familiar volatility clustering effect with some extraordinarily large absolute value can be observed from the time series plots 5.4.2 of index returns for all these four markets.

|                        | EU       | HK       | SG        | US       |
|------------------------|----------|----------|-----------|----------|
| **Summary Statistics** |          |          |           |          |
| Mean(%)                | 0.511    | 0.628    | 0.415     | 0.679    |
| Median(%)              | 0.913    | 0.758    | 0.812     | 1.073    |
| Min(%)                 | -23.976  | -57.337  | -53.497   | -24.158  |
| Max(%)                 | 12.362   | 28.374   | 22.846    | 12.203   |
| Std. Dev.              | 0.051    | 0.083    | 0.074     | 0.043    |
| Skewness               | -0.790   | -1.088   | -1.238    | -0.893   |
| Kurtosis               | 5.064    | 9.355    | 10.434    | 6.128    |
| Jarque-Bera statistic  | 132.041  | 881.755  | 1199.700  | 253.639  |
| Jarque-Bera p-value    | 0.000    | 0.000    | 0.000     | 0.000    |

**Correlation matrix**

|    | HK            | SG            | US            |
|----|---------------|---------------|---------------|
| EU | 0.559 (0.376) | 0.575 (0.353) | 0.744 (0.500) |
| HK |               | 0.662 (0.431) | 0.498 (0.334) |
| SG |               |               | 0.601 (0.367) |

Table 3.4.1: The table shows the summary statistics of monthly index returns of Europe(EU), Hong Kong(HK), Singapore(SG), and the United States(US). Returns are defined as $r_t = \ln P_t / P_{t-1}$ and $P_t$ is the index of time t.
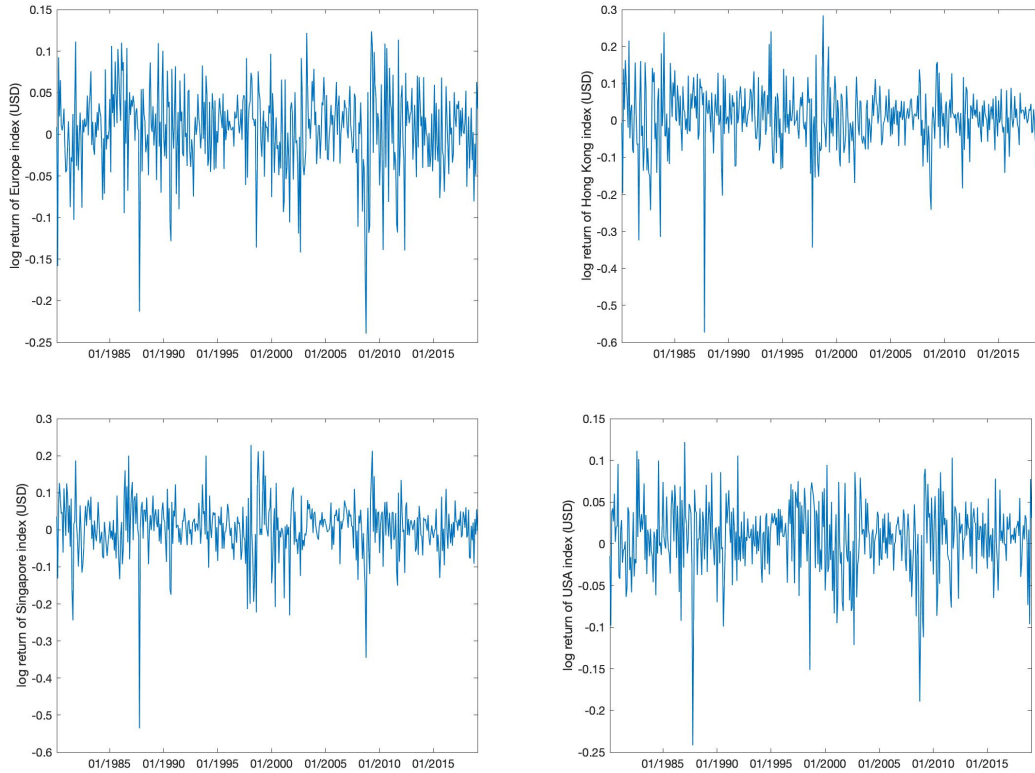
Figure 3.4.1: The time series plots for the log return of Europe(EU), Hong Kong(HK), Singapore(SG) and the United States(US) from Jan. 1980 to Feb. 2019.

### 3.4.2 The models of marginal distributions

An important step in accurately estimating the copula model is to set the margins properly. Considering the financial data presenting usually clustering volatility and conditional heteroscedasticity, we apply a GARCH(1,1) model for unconditional marginal distributions for index returns. Although various specifications of the GARCH model are implementable, Hansen and Lunde (2005) shows that selecting different orders of a GARCH model has trivial importance in improving forecasting accuracy. Also, another feature of the financial time series data is the leverage effect, which shows an asymmetric effect on the volatility changes in response to positive and negative shocks to the same extent. To capture this effect, the GJR-GARCH model proposed by Glosten et al. (1993) is taken into

consideration:

$$X_{it} = \mu_i + \epsilon_{it} \tag{3.4.1}$$

$$\epsilon_{it} = \sigma_{it}\eta_{it} \tag{3.4.2}$$

$$\sigma_{it}^2 = \alpha_i + \psi_i\epsilon_{i,t-1}^2 + \delta_i\sigma_{i,t-1}^2 + \phi_i\epsilon_{i,t-1}^2 I_{i,t-1} \tag{3.4.3}$$

where the $X_{it}$ stands for the return of the i-th market at time t and $I_{i,t-1} = 0$ if $\epsilon_{i,t-1} \geq 0$, $I_{i,t-1} = 1$ if $\epsilon_{i,t-1} < 0$. The term multiplying $\phi$ captures the asymmetry. Since the data series show conditional leptokurtic properties, the white noise process $\epsilon_{it}$ is assumed to follow the standardized t distribution. Let $\Omega_{it}$ denote $\{X_{it}, X_{i,t-1}, ...\}$:

$$\eta_{it}|\Omega_{it} \sim t_{\nu_i}(0,1) \quad i.i.d \tag{3.4.4}$$

The estimation results of selected models are shown in Table 3.4.2. The equity index of the United States(US) and Europe(EU) shows a significant asymmetry effect, so the GJR-based GARCH(1,1) models are selected for these two. The coefficients $\phi$ for both US and EU are positive, which indicates that the negative shocks provide more volatility than positive shocks with the same size in the subsequent period.

Correct specification of the marginal models determines the performance of the copula function. As the joint copula model is a function of marginal distributions, the misspecification of margins easily leads to the inaccurate result of copulas. Under this consideration, we provide several misspecification tests to confirm the empirical adequacy of the marginal models: the Kolmogorov–Smirnov test, the Cramer–von Mises test, and the Anderson–Darling tests.

| | GJR-GARCH(1,1) | | | | d.o.f | AIC | BIC |
|---|---|---|---|---|---|---|---|
| | $\alpha$ | $\psi$ | $\delta$ | $\phi$ | $\nu$ | | |
| | (s.e.) | (s.e.) | (s.e.) | (s.e.) | (s.e.) | | |
| US | 0.310E-3** | 0.710*** | | 0.313** | 8.295*** | -1661.7 | -1645.1 |
| | (0.000) | (0.102) | | (0.125) | (1.954) | | |
| EU | 0.472E-3** | 0.685*** | 0.058 | 0.151* | 7.681*** | -1506.1 | -1485.3 |
| | (0.000) | (0.122) | (0.063) | (0.089) | (2.515) | | |
| HK | 0.158E-3 | 0.892*** | 0.086*** | | 5.098*** | -1101.2 | -1084.6 |
| | (0.000) | (0.037) | (0.031) | | (0.980) | | |
| SG | 0.349E-3** | 0.762*** | 0.187*** | | 4.983*** | -1230.9 | -1214.3 |
| | (0.000) | (0.060) | (0.058) | | (0.993) | | |

Table 3.4.2: The table shows the ML estimates for the parameters of the marginal distribution model. The models are selected according to Akaike Information Criterion(AIC), Bayesian Information Criterion(BIC), and their statistical features.

The KS test is most sensitive when the estimated distributions differ in a global fashion near the center of the distribution. But if there are repeated deviations between the estimated distributions or the estimated distributions have the same mean values, then they cross each other multiple times and the maximum deviation between the distributions is reduced. The Cramer-von Mises (CvM) test measures the sum of squared deviations between the estimated distribution and target distribution and it treats this case well. But both the KS and CvM statistics are insensitive when the differences between the curves are most prominent near the beginning or end of the distributions. The Anderson-Darling (AD) test was developed in the 1950s as a weighted CvM test to overcome both of these problems. From the results shown in 3.4.3, all the marginal distribution models pass the KS test, CvM test, and AD test at 5% significance level, which means that the marginal models are well-specified.

| | KS | | CvM | | AD | |
|---|---|---|---|---|---|---|
| Margins | KS statistics | p-value | CM statistics | p-value | AD statistics | p-value |
| US | 0.037 | 0.533 | 0.160 | 0.365 | 1.246 | 0.251 |
| EU | 0.027 | 0.866 | 0.068 | 0.766 | 0.921 | 0.401 |
| HK | 0.024 | 0.937 | 0.053 | 0.860 | 0.534 | 0.713 |
| SG | 0.036 | 0.551 | 0.109 | 0.544 | 1.319 | 0.226 |

Table 3.4.3: The statistical results from Kolmogorov–Smirnov test, the Cramer–von Mises test, and the Anderson–Darling tests for the marginal distributions of the four markets: the United States(US), European(EU), Hong Kong(HK) and Singapore(SG).

### 3.4.3 The models for Copula function

Based on the previous analysis, our primary focus is to study the relationship between four pairs of markets: SG-HK, SG-US, US-EU, and EU-HK. These pairs exhibit high and similar correlation coefficients, as discussed in section 3.4.1. The scatter plots illustrating the relationships of these four pairs are presented in Figure 3.4.2, revealing distinct patterns among them. All four pairs exhibit asymmetry and a large number of outliers, which indicate violations of the assumption of an elliptical distribution. Our nonparametric copula approach provides a flexible way to capture this complex dependence structure. To assess the performance of our model, we divide the data into two samples: one for training and the other for testing purposes.

(a) SG-HK

(b) SG-US

(c) EU-US

(d) EU-HK
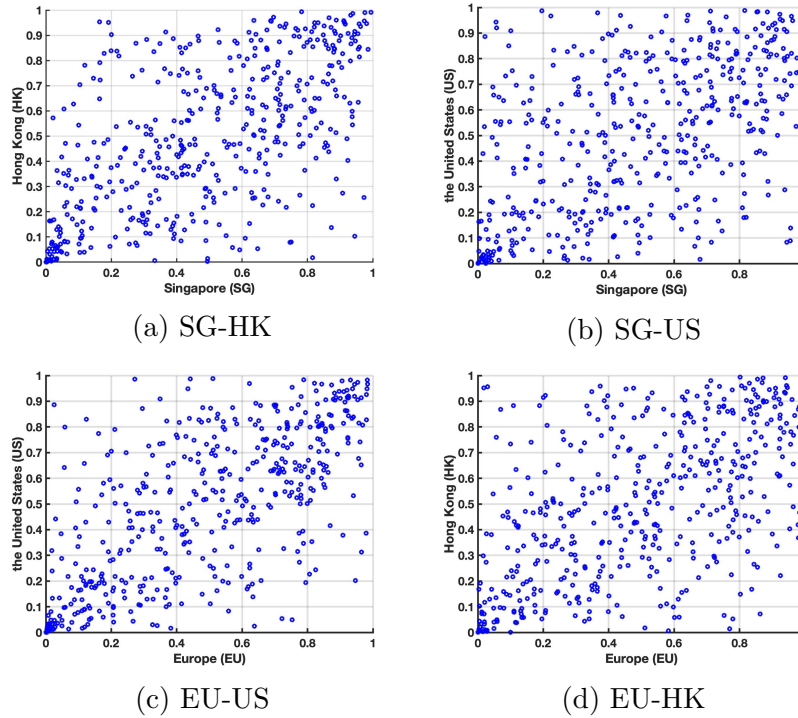
Figure 3.4.2: The scatter plot for these four pairs of markets.

The data points of most pairs are more concentrated in the left corner, which is similar to other financial time series data. It means that these four market equity indexes have the aggregation effect when the market is down-turn. Then we fit the data into the proposed model and do the selection of dimension of coefficients and the tuning parameter. The following figure shows that as the dimension of coefficients increases, the log-likelihood increases first and then decreases. For the pairs of SG-HK and EU-HK, the penalized Bernstein copula model with empirical Bernstein copula penalty(PBC) chooses a relatively smaller dimension of coefficients $J$ as the optimum and also has a higher log-likelihood compared with the empirical Bernstein copula model(EBC).

(a) SG-HK      (b) SG-US

(c) EU-US      (d) EU-HK
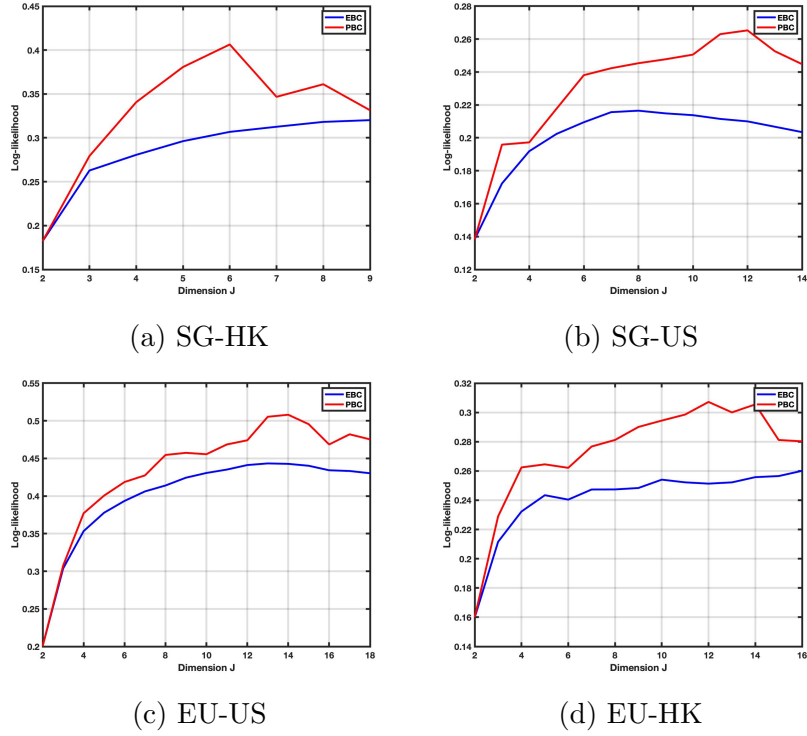
Figure 3.4.3: The process of selecting the optimal dimension of coefficient $\gamma$ and tuning parameter $\lambda$. The red line represents the loglikelihood of the empirical Bernstein copula(EBC) and the blue line stands for the loglikelihood of the penalized Bernstein copula with empirical Bernstein copula penalty(PBC). The loglikelihood of PBC is always larger than that of EBC.

From section 3.4.1, the pairs of SG-HK, SG-EU, and EU-HK have similar correlation coefficients and Kendall's tau. But they have totally different dependence structures which are shown in figure 3.4.4. The contour plots show that these four pairs of markets show various levels of asymmetry for the two tails.

| Pairs of markets | $J$ | $\lambda$ | Log-likelihood |
|---|---|---|---|
| SG-HK | 6 | 0.9560 | 0.4064 |
| SG-US | 12 | 0.8220 | 0.2653 |
| EU-US | 14 | 0.7720 | 0.5081 |
| EU-HK | 12 | 0.2050 | 0.3072 |

Table 3.4.4: This table represents the results of model selection, the J represents the dimension of coefficients, and $\lambda$ stands for the tuning parameter
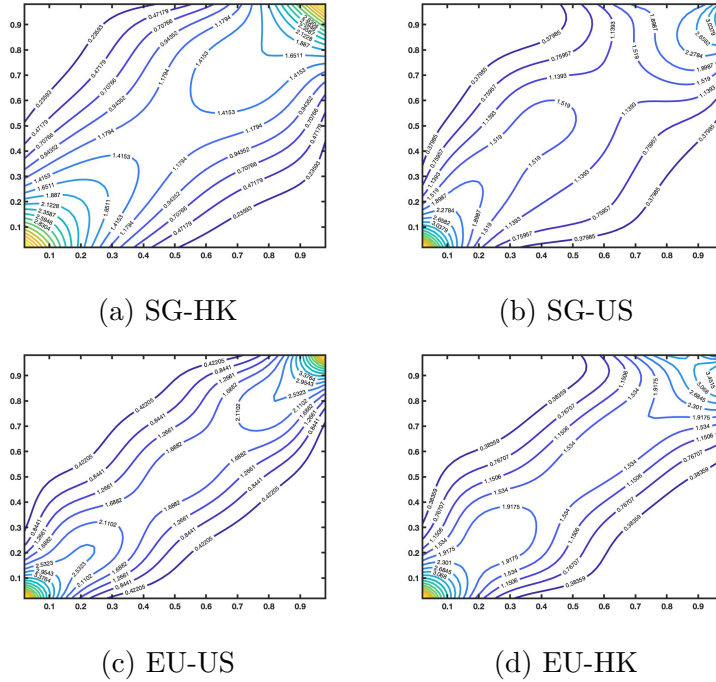
Figure 3.4.4: The contour plots of the estimated penalized Bernstein copula density function.

For the pair of SG-HK, the contour plot exhibits slightly greater dependence in the positive tail than in the negative one. In other words, the Singapore market and Hong Kong market are closely linked when there are some positive effects on the whole financial market. Turning to the SG-US and EU-HK pairs, they have similar asymmetric properties in tails. The contour plots (b) and (d) show greater dependence on the negative tail than on the positive one. And the EU-US pair has relatively symmetric tails.

## 3.5   Conclusion

In this chapter, we study a new approach in multivariate analysis based on the Bernstein copula estimation. We show that under some regularity, any copula can be approximated by the Bernstein copula both theoretically and practically. It allows us to take the advantage of the copula property when the multivariate

normality assumption is violated.

Considering the nonparametric approach has infinite-dimensional parameter space, we utilize the method of sieve to optimize the loglikelihood function over the Bernstein polynomial sieve. Since the approximating spaces can be characterized by a finite number of parameters, which is $J_n^2$ in our case, the nonparametric problem reduces to a parametric one. As the sample size increase, the order of polynomials $J_n$ grow slowly. To improve the prediction performance of the fitted model, we use the adapted LASSO to select the model. Inspired by the empirical Bernstein copula, the penalty function is set as the reciprocal of the empirical copula and empirical Bernstein copula respectively.

In the simulation study, we generate the random number of four different types of parametric copula and show that if the parametric model cannot be correctly selected, using the nonparametric Bernstein copula is the best choice. Also, we conduct two scenarios to study the capability of the model selection. Based on the simulation results, the proposed model can impose sparsity precisely and has the best out-of-sample log-likelihood in all the iterations compared to the empirical Bernstein copula and LASSO penalized model.

The empirical evidence verifies that the financial market equity indexes have the aggregation effect when the market is down-turned(Forbes and Rigobon, 2002). Besides, within the range of the optimal order of polynomials and under the same accuracy requirement, the penalized model can always have smaller $J_n$ compared with the empirical Bernstein.

# CHAPTER 4

## Double Selection of Marginal Parameters in Semiparametric Copula Models

## 4.1  Introduction

The semiparametric and nonparametric modeling approaches have gained increasing attention in both theoretical and applied econometrics. The parametric model is restrictive and sensitive to deviations from the parametric specification, whereas the semiparametric or nonparametric models are more flexible and robust. In the copula framework, many studies focus on the semiparametric setting with nonparametric marginal distribution and parametric copula function to avoid potential misspecification in margins (Chen and Fan, 2006; Tsukahara, 2005; Chen et al., 2006). Kim et al. (2007) shows that when the marginal distributions are misspecified, the behavior of the IFM (inference functions for margins) is unpredictable and likely to be inconsistent, resulting in large mean squared errors.

However, there is another line of research where the marginal distribution is known to follow a specific distribution (Ivan et al., 2021). In such cases, the copula function can be set to be nonparametric. The use of copulas to estimate marginal distributions offers several advantages. Firstly, copulas enable separate model-

ing of the marginal distributions and the dependence structure between variables. This flexibility proves particularly valuable when the marginal distributions are known or can be reasonably assumed to follow specific distributions. By estimating the copula function, we can accurately capture the dependence structure while preserving the characteristics of the marginal distributions. This approach provides more precise estimation compared to methods relying solely on empirical distributions.

Additionally, when the marginal distribution is known to have a parametric density function, estimating its density based on the correct parametric specification can be accurate and efficient. For example, in the case of non-life insurance claims, the claim frequency is often modeled using a Poisson distribution, and the claim severity follows a gamma distribution (Czado et al., 2012). Using a nonparametric copula model aims to provide more precise estimation in such scenarios. Similar settings can often be found in actuarial and microeconomic applications as well (Winkelmann, 2012; Amsler et al., 2014).

In this chapter, we present a three-step double selection method and utilize the model we constructed in Chapter 3 as the copula function to examine the performance of individual density estimation. The double selection procedure is introduced in section 4.2. To comprehensively evaluate the methods, extensive simulation studies are conducted. We compare three types of methods: 'Univariate' which independently estimates the individual density, 'Copula-based Empirical' method that simultaneously estimates the univariate density with a nonparametric Bernstein copula through sieve maximum likelihood estimation, and 'Copula-based Penalized' method which estimates the univariate density after the copula selection. We consider different univariate density functions and various copula families, using both cross-sectional data in section 4.3 and time series data in

section 4.4. Furthermore, we examine the comparison in the context of Value at Risk for individual log return, utilizing data simulated from multivariate GARCH models. To assess the performance, we employ a forecast evaluation test for Value at Risk.

## 4.2 Double selection methodology

In Chapter 3, we utilize the sieve MLE to estimate the coefficient of the nonparametric Bernstein copula density function and indicate that with correctly specifying the order of polynomial $J_n$, the metric between the sieve MLE estimator and the true estimator will reach zero as the sample size goes to infinity. In this chapter, a parametric marginal distribution will be taken into consideration. Without loss of generality, we focus on the bivariate case.

Let $(X_1, X_2) \sim H(x_1, x_2)$ , where $H$ is a cumulative function and $F_1, F_2$ is the absolutely continuous marginal CDF. By Sklar's theorem, there exists a unique copula $C(\cdot)$ such that

$$H(x_1, x_2) = C(F_1(x_1), F_2(x_2)) \quad \forall (x_1, x_2) \in \mathbb{R}^2 \qquad (4.2.1)$$

Assume the marginal distribution belongs to a specific parametric family and each margin has $\beta$ to be estimated. The double selection method is through the following procedure:

**Step 1: Obtain the coefficients $\hat{\gamma}_e$ of empirical Bernstein copula and the parameters $\hat{\beta}_{1,e}, \hat{\beta}_{2,e}$ from the marginals simultaneously.**

In the nonparametric empirical Bernstein copula framework, consider the estima-

tion of unknown parameters $\theta$, for this, let

$$\Theta = \{\theta = (\beta_1, \beta2, \Lambda)\} \in \mathcal{B} \otimes \mathcal{M}$$

denote the parameter space of $\theta$. Here, $\mathcal{B} = \{(\beta_1, \beta_2) \in R^{d_\beta}\}$. $\mathcal{M}$ is the collection of all bounded and continuous nondecreasing, nonnegative Bernstein copula functions. To estimate $\theta$, a natural approach is to maximize the log-likelihood function $l_n(\theta) = \log\{L_n(\theta)\}$. But, it's obvious to find that this is not an easy task since $l_n(\theta)$ involves both finite-dimensional and infinite-dimensional parameters. Followed by Chen et al. (2006), the sieve maximum likelihood estimation is employed. The sieve space can be defined as

$$\Theta_n = \{\theta_n = (\beta_1, \beta_2, \Lambda_n) \in \mathcal{B} \otimes \mathcal{M}_n\} \tag{4.2.2}$$

In the above,

$$\mathcal{M}_n = \left\{ \Lambda_n(u_1, u_2) = \sum_{j_1=0}^{J_n-1} \sum_{j_2=0}^{J_n-1} \gamma_{j_1,j_2} p_{j_1}(u_1) p_{j_2}(u_2) : \sum_{j_1=0}^{J_n-1} \sum_{j_2=0}^{J_n-1} \gamma_{j_1,j_2} = 1, \right.$$
$$\left. \sum_{j_s=0}^{J_n-1} \gamma_{j_1,j_2} = 1/J_n, \ s = 1, 2, \ 0 < \gamma_{j_1,j_2} < 1 \right\}$$

with $p_{j_s}(u_s)$ is a beta density defined in chapter 3.

$$l(\theta, Z_i) = \log\left\{ c(F_1(x_{1i}), F_2(x_{2i}); \theta) \prod_{j=1}^{2} f_j(x_{ji}) \right\}) \tag{4.2.3}$$

Then, it is natural to define the sieve maximum likelihood estimator $\hat{\theta}_n = (\hat{\beta}_{1n}, \hat{\beta}_{2n}, \hat{\Lambda}_n)$ of $\theta$ that maximizes the log-likelihood function $l_n(\theta)$ over $\Theta_n$.

$$\hat{\theta}_n = \arg\max_{\theta \in \Theta_n} \sum_{i=1}^{n} l(\theta, Z_i)$$

According to Ivan et al. (2021), under the non-singular and some convergence assumptions, the marginal coefficient $\beta$ based on sieve MLE is consistent and asymptotic efficient.

**Step 2: Estimate the $\hat{\gamma}_p$ based on the marginal coefficients $\hat{\beta}_{1e}$, $\hat{\beta}_{1e}$ from step 1.**

In this step, we use the same model as explained in chapter 3. Considering the marginal parameters and shapes as known value, $u_1 = \hat{F}_1(x_1)$ and $u_2 = \hat{F}_2(x_2)$. Still under the sieve space:

$$\mathcal{M}'_n = \left\{ \Lambda_n(u_1, u_2) = \sum_{j_1=0}^{J_n-1} \sum_{j_2=0}^{J_n-1} \gamma_{j_1,j_2} p_{j_1}(u_1) p_{j_2}(u_2) : \sum_{j_1=0}^{J_n-1} \sum_{j_2=0}^{J_n-1} \gamma_{j_1,j_2} = 1, \right.$$
$$\left. \sum_{j_s=0}^{J_n-1} \gamma_{j_1,j_2} = 1/J_n, \ s = 1, 2, \ 0 < \gamma_{j_1,j_2} < 1, pen(\Lambda) \leq b_n \right\}$$

$$\text{with } b_n \to \infty \text{ slowly.}$$

The penalty here is the adapted LASSO using the reciprocal of empirical Bernstein copula as the weight to penalize values with a small number of data points on its grids to zero. In other words, it selects the significant features that capture the dependence structure.

**Step 3: Remove the zeros of the coefficient matrix from step 2 $\hat{\gamma}$ outside the matrix, re-estimate the parameters of marginals.**

Based on step 2, using the penalization method, we select the optimal nonparametric copula model and remove the unrelated $\gamma_{j_1,j_2}$. Then, we revisit the nonpenalized Bernstein copula model, estimate $\gamma$ without zero elements and $\beta$ simultaneously through maximizing the loglikelihood of the copula function.

## 4.3 Simulation study

In this section, we perform the Monte Carlo simulations to compare the estimation accuracy of parametric marginal distribution between the univariate method and copula-based methods. We first generate the random number from the univariate density function to get independent and identically distributed(i.i.d) data, then fit the data using three different methods. Since the true parameters are given, the integrated mean squared error(IMSE) and Kullback Leibler distance are employed to measure the preciseness. Besides, correlated pairs are also generated based on Cholesky decomposition, which can secure the marginal distributions of the generated multivariate normal distribution are still normally distributed with known mean and variance. In addition, we simulate from multivariate GARCH models with Gaussian innovations to investigate the performances under the time series framework. In this section, the different dependencies between the two margins are specified.

### 4.3.1 Cross-sectional data

In this scenario, we generate random numbers from selected parametric models of marginal distribution including normal, gamma, and inverse Gaussian distribution. They are the most commonly applied distribution in finance and econometric world(Barndorff-Nielsen, 1997; Xie and Wu, 2017). The gamma distribution is a family of right-skewed continuous distributions which can accurately measure the asymmetry in asset pricing, such as options(Heston, 1993), non-life insurances(Gschlößl and Czado, 2007). The inversed Gaussian distribution is utilized to model the nonnegative skewed data and it has many similarities with standard normal distribution in estimation. The detailed setting for the parameters of margins can be found as follows.

1. Gaussian 1: $\mu_1 = 0.5, \sigma_1 = 3; \mu_2 = 3, \sigma_2 = 0, 5$.

2. Gaussian 2: $\mu_1 = 0.5, \sigma_1 = 10; \mu_2 = 5, \sigma_2 = 2$.

3. Gaussian 3: $\mu_1 = 1, \sigma_1 = 8; \mu_2 = 5, \sigma_2 = 2$.

4. Gamma 1: $a_1 = b_1 = a_2 = b_2 = 2$.

5. Gamma 2: $a_1 = b_1 = a_2 = b_2 = 5$.

6. Gamma 3: $a_1 = 0.5, a_2 = 1, a_3 = 0.5, a_4 = 1$

7. Inverse Gaussian 1: $a_1 = 1, \ b_1 = 10; \ a_2 = 1, \ b_2 = 10$.

8. Inverse Gaussian 1: $a_1 = 1, \ b_1 = 1; \ a_2 = 1, \ b_2 = 1$.

9. Inverse Gaussian 1: $a_1 = 2, \ b_1 = 4; \ a_2 = 2, \ b_2 = 4$.

Three models we aim to compare include the 'Univariate', 'Copula-based empirical', and 'Copula-based penalized'. The 'univariate' means the marginal distribution is obtained from IFM method and it is the optimal estimation based on an individual information, which is $\hat{\beta}_{ju} = \arg\max_{\beta_j} \sum_{i=1}^{n} \log f_j(x_{ji})$. The 'Copula-based empirical' is the $\hat{\beta}_e$ that is estimated simultaneously with the nonparametric Bernstein copula model through sieve MLE in step 1. The 'Copula-based penalized' is the $\hat{\beta}_p$ estimated after model selection. Based on the information of selection results, we simulate it with a sparse copula model through step 3.

The measure that we used for comparison includes the KL-distance, which is Kullback Leibler divergence. It is a kind of statistical distance to measure how one probability distribution is different from the other. For the distribution P and Q of a continuous random variable, the relative entropy from Q to P is

$$D_{KL}(P\|Q) = \int_{-\infty}^{\infty} p(x) \log\left(\frac{p(x)}{q(x)}\right) dx \qquad (4.3.1)$$

where P represents the true distribution of data and Q is an approximation of P. The KL-distance and integrated mean squared error (IMSE) are calculated based on the average of each iteration.

The simulation results can be found in table 4.3.1. Compare the univariate model with the copula-based model, the univariate one has a relatively larger distance from the true model. It means that using copula can truly impose preciseness in univariate density estimation for i.i.d data. The double selection method outperforms the other two. The 'copula based empirical' and the 'copula-based penalized' are all using the same order of polynomials here.

For the dependent scenario, we generated pairs of correlated random numbers based on the Cholesky decomposition. The data series that we simulated are bivariate normal distributed. The marginal distributions have mean 0 and standard deviation 1. Also, we employ different types of copula to study whether the improvement in evaluation correlates with the copula function. In other words, if the copula can capture the dependency more accurately, we are interested in whether it will affect the margins.

Here, we choose three different shapes of copulas. They are Gaussian copula, student t copula and Clayton copula. The Gaussian and t copulas are selected

| Marginal Type | | Gaussian1 | Gaussian2 | Gaussian3 | Gamma1 | Gamma2 | Gamma3 | Inv Gaussian1 | Inv Gaussian 2 | Inv Gaussian3 |
|---|---|---|---|---|---|---|---|---|---|---|
| Univariate | Margin 1 IMSE | 0.0031 | 0.0052 | 0.0000 | 0.0258 | 0.0000 | 1.4132 | 0.4110 | 0.3696 | 0.1434 |
| | Margin 1 KL | 2.4243 | 2.4882 | 0.0232 | 4.8259 | 5.2409 | 6.5516 | 6.6294 | 6.1558 | 7.0257 |
| | Margin 2 IMSE | 0.0048 | 0.0001 | 0.0001 | 0.0124 | 0.0000 | 1.5430 | 0.7948 | 0.8266 | 0.1152 |
| | Margin 2 KL | 0.6969 | 2.0541 | 2.0541 | 7.5469 | 3.3826 | 6.3874 | 5.5175 | 6.5886 | 5.8445 |
| Copula-Based Empirical | Margin 1 IMSE | 0.0001 | 0.0000 | 0.0000 | 0.0008 | 0.0000 | 0.0093 | 0.0055 | 0.0080 | 0.0023 |
| | Margin 1 KL | 0.0110 | 0.0109 | 0.0109 | 0.0130 | 0.0149 | 0.0112 | 0.0114 | 0.0107 | 0.0112 |
| | Margin 2 IMSE | 0.0000 | 0.0000 | 0.0000 | 0.0008 | 0.0000 | 0.0159 | 0.0074 | 0.0077 | 0.0022 |
| | Margin 2 KL | 0.0117 | 0.0098 | 0.0098 | 0.0125 | 0.0111 | 0.0105 | 0.0104 | 0.0102 | 0.0108 |
| Copula-Based Penalized | Margin 1 IMSE | 0.0000 | 0.0000 | 0.0000 | 0.0008 | 0.0000 | 0.0091 | 0.0058 | 0.0080 | 0.0023 |
| | Margin 1 KL | 0.0108 | 0.0103 | 0.0103 | 0.0128 | 0.0145 | 0.0110 | 0.0172 | 0.0107 | 0.0112 |
| | Margin 2 IMSE | 0.0000 | 0.0000 | 0.0000 | 0.0010 | 0.0000 | 0.0143 | 0.0072 | 0.0077 | 0.0022 |
| | Margin 2 KL | 0.0117 | 0.0098 | 0.0098 | 0.0125 | 0.0112 | 0.0101 | 0.0102 | 0.0102 | 0.0108 |

Table 4.3.1: Estimation results for i.i.d data generating from given marginal distribution with sample size n=100. Each path has been iterated for I =100 times to get the average error for each item. The 0.0000 means that the average distance between this estimated result and the true parameter is smaller than 0.00005.

due to their widespread application in finance(Cherubini et al., 2004; Demarta and McNeil, 2005). Choosing Clayton copula is because it works well if the correlation between the two random variables is strongest in the left tail of the joint distribution. For gaussian and Clayton $\rho$ is the only parameter to control the dependence. Since these three copula models are parametric, full MLE is applied to find the optimal parameters for both margins and copula simultaneously. We simulate $n = 100$ for 1,000 times for each dependence level $\rho = 0.1, 0.15, ..., 0.85, 0.9$. The following figure shows the correlated data that we generated with different $\rho$.



(a) Independent

(b) Correlation $\rho = 0.3$

(c) Correlation $\rho = 0.6$

(d) Correlation $\rho = 0.9$
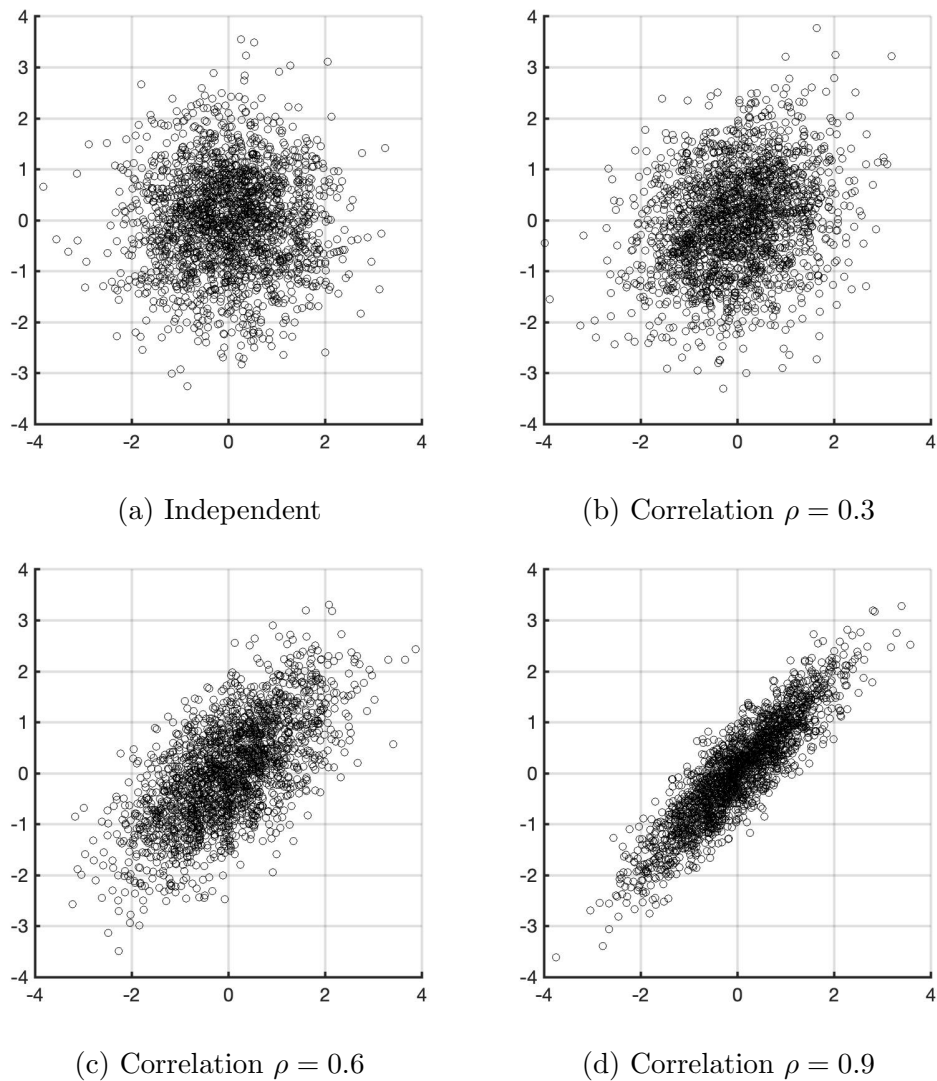
Figure 4.3.1: Pairs of correlated simulated from the bivariate normal distribution.

| | Gaussian Copula | | | | Student Copula | | | | Clayton Copula | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Margin 1 | | Margin 2 | | Margin 1 | | Margin 2 | | Margin 1 | | Margin 2 | |
| $\rho$ | IMSE | KL | IMSE | KL | IMSE | KL | IMSE | KL | IMSE | KL | IMSE | KL |
| 0.1 | 0.0007 | 0.0109 | 0.0007 | 0.0103 | 0.0006 | 0.0081 | 0.0007 | 0.0103 | 0.0006 | 0.0081 | 0.0007 | 0.0103 |
| 0.2 | 0.0007 | 0.0107 | 0.0007 | 0.0102 | 0.0006 | 0.0080 | 0.0007 | 0.0102 | 0.0006 | 0.0080 | 0.0007 | 0.0102 |
| 0.3 | 0.0007 | 0.0104 | 0.0007 | 0.0103 | 0.0006 | 0.0080 | 0.0007 | 0.0102 | 0.0006 | 0.0080 | 0.0007 | 0.0102 |
| 0.4 | 0.0007 | 0.0107 | 0.0007 | 0.0102 | 0.0006 | 0.0081 | 0.0007 | 0.0102 | 0.0006 | 0.0081 | 0.0007 | 0.0102 |
| 0.5 | 0.0007 | 0.0106 | 0.0007 | 0.0105 | 0.0006 | 0.0082 | 0.0007 | 0.0105 | 0.0006 | 0.0082 | 0.0007 | 0.0105 |
| 0.6 | 0.0007 | 0.0102 | 0.0007 | 0.0104 | 0.0006 | 0.0079 | 0.0007 | 0.0104 | 0.0006 | 0.0079 | 0.0007 | 0.0104 |
| 0.7 | 0.0007 | 0.0106 | 0.0008 | 0.0108 | 0.0006 | 0.0080 | 0.0008 | 0.0107 | 0.0006 | 0.0080 | 0.0008 | 0.0108 |
| 0.8 | 0.0007 | 0.0107 | 0.0007 | 0.0109 | 0.0006 | 0.0078 | 0.0007 | 0.0108 | 0.0006 | 0.0078 | 0.0007 | 0.0108 |
| 0.9 | 0.0007 | 0.0105 | 0.0007 | 0.0107 | 0.0006 | 0.0078 | 0.0007 | 0.0106 | 0.0006 | 0.0078 | 0.0007 | 0.0106 |

| | Univariate Estimation | | | | Empirical Bernstein Copula | | | | Penalized Bernstein Copula | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Margin 1 | | Margin 2 | | Margin 1 | | Margin 2 | | Margin 1 | | Margin 2 | |
| $\rho$ | IMSE | KL | IMSE | KL | IMSE | KL | IMSE | KL | IMSE | KL | IMSE | KL |
| 0.1 | 0.0007 | 0.0107 | 0.0007 | 0.0102 | 0.0008 | 0.0108 | 0.0007 | 0.0103 | 0.0007 | 0.0086 | 0.0006 | 0.0085 |
| 0.2 | 0.0007 | 0.0105 | 0.0007 | 0.0101 | 0.0007 | 0.0109 | 0.0007 | 0.0105 | 0.0007 | 0.0085 | 0.0006 | 0.0083 |
| 0.3 | 0.0007 | 0.0102 | 0.0007 | 0.0101 | 0.0007 | 0.0108 | 0.0007 | 0.0106 | 0.0006 | 0.0086 | 0.0006 | 0.0085 |
| 0.4 | 0.0007 | 0.0104 | 0.0007 | 0.0100 | 0.0008 | 0.0111 | 0.0007 | 0.0105 | 0.0007 | 0.0088 | 0.0006 | 0.0083 |
| 0.5 | 0.0007 | 0.0104 | 0.0007 | 0.0103 | 0.0007 | 0.0110 | 0.0008 | 0.0109 | 0.0007 | 0.0088 | 0.0007 | 0.0088 |
| 0.6 | 0.0007 | 0.0100 | 0.0007 | 0.0103 | 0.0007 | 0.0107 | 0.0008 | 0.0112 | 0.0007 | 0.0084 | 0.0007 | 0.0083 |
| 0.7 | 0.0007 | 0.0104 | 0.0008 | 0.0106 | 0.0008 | 0.0120 | 0.0009 | 0.0121 | 0.0008 | 0.0094 | 0.0008 | 0.0094 |
| 0.8 | 0.0007 | 0.0105 | 0.0007 | 0.0107 | 0.0007 | 0.0107 | 0.0007 | 0.0105 | 0.0007 | 0.0088 | 0.0007 | 0.0088 |
| 0.9 | 0.0006 | 0.0103 | 0.0007 | 0.0105 | 0.0006 | 0.0103 | 0.0008 | 0.0109 | 0.0006 | 0.0088 | 0.0007 | 0.0088 |

Table 4.3.2: Comparison between margins estimation based on different copula as the basis. The Basis copula includes Gaussian, student t, Clayton, Empirical Bernstein copula and Penalized Bernstein copula. Also, the Univariate is the one estimated individually. The sample size is n =100 and iteration I=1000 times.

The results in table 4.3.2 show the distance between the estimated marginal density function and the true density. Here, the true density is the standard normal distribution. KL and IMSE are the averages of 1000 times iteration. From the perspective of copula modeling, we can find that all these six models have similar fitness to the bivariate normal distribution. Only the penalized Bernstein copula model has a little bit of improvement in accuracy. Besides, there is no significant trend can be observed as the correlation coefficient $\rho$ increases. In order words, copula cannot provide more information due to their linear dependency. Even though the marginal distributions are highly correlated, the copula-based method provides similar estimation as the low dependency pairs.

## 4.4 Simulation study in the context of value-at-risk estimation

Investigating and understanding the time series dependency in the second-order moments of property returns is a crucial concept in financial econometrics. Lots of empirical evidence show the co-movement of financial volatilities over time between markets and assets(Calvet et al., 2006; Evans and McMillan, 2009; Hemche et al., 2016). Capturing this phenomenon using a multivariate way leads to more empirical models than working with separate univariate models. We aim to find if the dependency can be correctly specified, more information for the univariate models in estimation and prediction can be investigated.

The performance of the estimator can be compared through the valuation of the Value-at-Risk measure. Value at Risk (VaR) is defined as the maximum loss due to the market risk over a certain period under general conditions. In this chapter, we focus on the two-asset case. Let the log return series at a moment in time t,

be marked as $r_t$. The random variable of loss over the period $[t, t + 1]$ is marked as $L_{i+1} = -(r_{i+1} - r_i) = \Delta_r$. The $F_L$ is the cdf function of loss distribution and it holds that $F_L(x) = P(L \leq x)$. VaR at significance level $\alpha$ is actually the $\alpha-$quantile of the cdf function, i.e.

$$VaR_\alpha = \inf(x|F_L(x) \geq \alpha) \tag{4.4.1}$$

In this chapter, we study the VaR with $\alpha = 1\%$ and $5\%$. $1\%$ is the level that financial institutions must report which is required by Basel Accords and $5\%$ is the most popular used in financial industries. There are three fundamental methods to evaluate the VaR: (1)use a parametric probability distribution function; (2) apply the Monte Carlo simulation to estimate the quantile of the unknown distribution numerically; (3) obtain the quantile directly from historical observations. To study how the innovation term performs in the process of VaR estimation. We assumed that the financial return can be specified on the conditional variance, and the Value at Risk can be evaluated as follows(Santos et al., 2013):

$$VaR_t^\alpha = \mu_t + \sigma_t q_\alpha \tag{4.4.2}$$

where $\mu_t$ is the conditional mean and $\sigma_t$ is the standardized deviation at time t. $q_\alpha$ is the $\alpha-$quantile of the standardized innovation $z_t = (r_t - \mu_t)/\sigma_t$. Since the conditional distribution is often assumed to be normal or student's t distribution with degree of freedom $\nu$. After estimating the density of the standardized innovation $z_t$, the quantile can be obtained immediately. For the student t distribution, the quantile needs an adjustment due to the degree of freedom. If the standardized innovation is under t, $q_\alpha = \sqrt{\frac{\nu-2}{\nu}}\tilde{q}_\alpha$, where $\tilde{q}_\alpha$ is the $\alpha$ quantile of the estimated student t distribution.

For the estimation of conditional mean $\mu_t$ , we assume it to be a constant over time. Since the conditional mean shows the weak dynamic dependency of an asset return(Santos et al., 2013), this setting will not affect the results of Value at Risk estimation.

The comparison between univariate and copula-based methods is investigated through the accuracy of the estimation of the standardized innovation, then reflected in the Value at Risk formula 4.4.2.

## 4.4.1 Simulating from multivariate GARCH models

In this section, the Monte Carlo simulation is provided to simulate multivariate time series data from some famous models. Consider the log return as a vector stochastic process $\{r_t\}$ which is of dimension $2 \times 1$ for a bivariate case. It has been conditioned on the sigma field by $I_{t-1}$ which includes all the past information until time t-1. The conditional mean is assumed to be a constant in this chapter, and it usually specifies as a function of past through a vectorial autoregressive moving average (VARMA) representation for the process $r_t$(Spliid, 1983; Kascha, 2012). Then $r_t$ can be written as

$$r_t = \mu + \varepsilon_t$$

and,

$$\varepsilon_t = H_t^{1/2} z_t$$

where $H_t^{1/2}$ is a 2 by 2 positive definite matrix. The 2 by 1 random vector $z_t$ has the following two moments:

$$E(z_t) = 0, \quad Var(z_t) = I_2$$

where $I_2$ is the identity matrix of order 2. $H_t$ is the conditional variance matrix of $r_t$, since $Var(r_t|I_{t-1}) = Var_{t-1}(r_t) = Var_{t-1}(\varepsilon_t) = H_t^{1/2}Var_{t-1}(z_y)(H_t^{1/2})' = H_t$, and $H_t^{1/2}$ may obtain from the Cholesky decomposition. We simulate the multivariate GARCH model mainly relying on different specifications of the conditional covariance matrix $H_t$.

We employ two multivariate GARCH models to simulate log returns and generate 100 systems of N = 2 asset returns, each with a sample size of T = 1000 observations. Then, the first 500 observations are used to estimate the parameter of both underlying conditional variance models, copula models, and margins. The rest of them were used to assess the out-of-sample performance.

### 4.4.1.1 Baba-Engle-Kraft-Kroner (BEKK) model

The Bekk model is a simple extension to the popular GARCH model and it is a special case of the vector error correction model. The BEKK(1,1,K) model is defined as:

$$H_t = C^{*'}C^* + \sum_{k=1}^{K} A_k^{*'}\varepsilon_{t-1}\varepsilon_{t-1}'A_k^* + \sum_{k=1}^{K} G_k^{*'}H_{t-1}G_k^* \qquad (4.4.3)$$

where $C^*$, $A_k^*$ and $G_k^*$ are $N \times N$ matrices and $C^*$ is an upper triangular. The summation limit K controls the generality of the process. Let K =1, N=2, the model above contains 11 parameters and implies the following dynamics:

$$\begin{bmatrix} h_{11,t} & h_{12,t} \\ h_{21,t} & h_{22,t} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{1,2} \\ 0 & c_{22} \end{bmatrix}' \begin{bmatrix} c_{11} & c_{1,2} \\ 0 & c_{22} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{1,2} \\ a_{21} & a_{22} \end{bmatrix}' \begin{bmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{bmatrix}' \begin{bmatrix} a_{11} & a_{1,2} \\ a_{21} & a_{22} \end{bmatrix}$$

$$+ \begin{bmatrix} g_{11} & g_{1,2} \\ g_{21} & g_{22} \end{bmatrix}' \begin{bmatrix} h_{11,t} & h_{12,t} \\ h_{21,t} & h_{22,t} \end{bmatrix} \begin{bmatrix} g_{11} & g_{1,2} \\ g_{21} & g_{22} \end{bmatrix}$$

The drawback of the BeKK model is that it includes a large number of parameters to be estimated, even for moderate dimensions(Pedersen and Rahbek, 2014). From the perspective of simulation, the more parameters to set, the more difficult to generate hypothetical data series. The parameters we used are given by Franke et al. (2004), which provides the pattern of the exchange rate between two European currencies.

$$
C^* = \begin{bmatrix} 1.15 & 0.31 \\ 0 & 0.76 \end{bmatrix}, \quad A^* = \begin{bmatrix} 0.2820 & -0.05 \\ -0.057 & 0.2930 \end{bmatrix}, \quad G^* = \begin{bmatrix} 0.939 & 0.025 \\ 0.028 & 0.939. \end{bmatrix}
$$

### 4.4.1.2 Conditional Correlation Model

The CCC model is introduced by Bollerslev (1990). Compared with Bekk model, smaller number of parameters are needed to be provided in simulation. It is only $N(N-1)/2$. The CCC model is defined as:

$$
H_t = D_t R D_t = (\rho_{i,j} \sqrt{h_{ii,t} h_{jj,t}}) \tag{4.4.4}
$$

where

$$
D_t = diag(h_{11,t}^{1/2}, ..., h_{NN,t}^{1/2}) \tag{4.4.5}
$$

The $h_{ii,t}$ can be defined as any univariate GARCH model, and

$$
R = (\rho_{ij}) \tag{4.4.6}
$$

is a symmetric positive definite matrix with $\rho_{i,i} = 1, \forall i. R$ is the matrix containing the constant conditional correlation. The original CCC has a GARCH (1,1) specification for each conditional variance in $D_t$:

$$
h_{ii,t} = \omega_i + \alpha_i \varepsilon_{i,t-1}^2 + \beta_i h_{ii,t-1}, \quad i = 1, ..., N \tag{4.4.7}
$$

Besides, $H_t$ is positive definite if and only if all the N conditional variances are positive and R is positive definite. It's not difficult to find unconditional variances in univariate cases. However, it's quite complicated to calculate the unconditional covariance, since the relationship states in equation 4.4.4 are nonlinear.
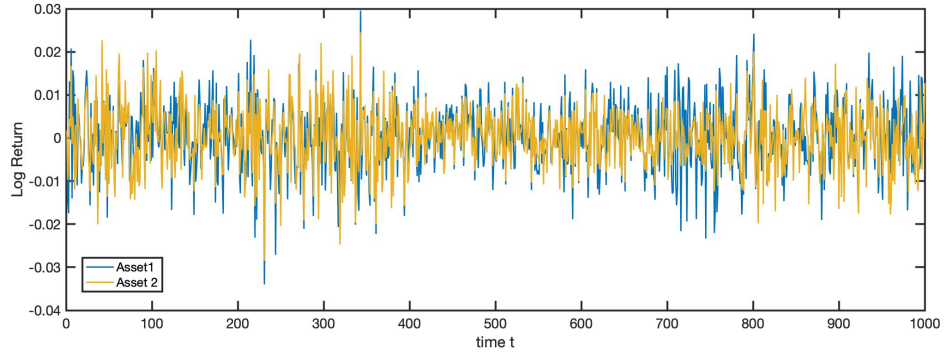


(a) The simulated data from BEKK GARCH model (T=1000).
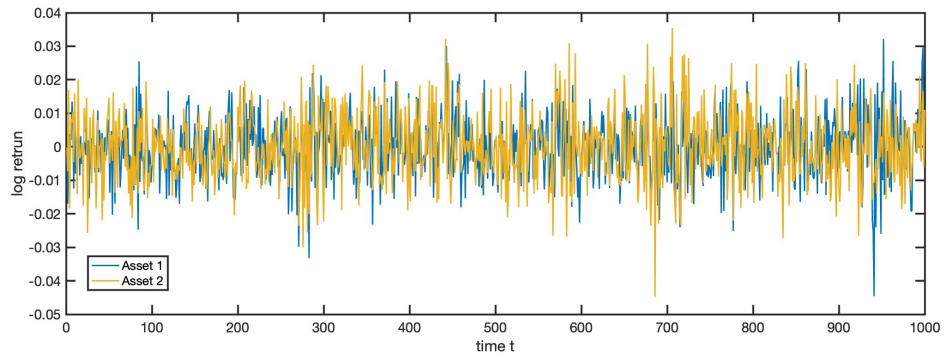


(b) The simulated data from CCC GARCH model (T=1000).

Figure 4.4.1: Simulated data from Bekk and CCC model. The figure shows one path out of 100, the conditional distribution for both these models is assumed to be normal.

## 4.4.2 Forcast evaluation of VaR models

Backtesting is the major technique to test the performance of Value at Risk. It includes the unconditional and conditional coverage and independence tests proposed by Kupiec et al. (1995) and Murphy (2012). In the content of log return,

the loss function-based backtests can be defined as

$$
I_{t+1}(\alpha) = \begin{cases} 1 & if \quad r_{t,t+1} \leq -VaR_t(\alpha) \\ 0 & if \quad r_{t,t+1} > -VaR_t(\alpha) \end{cases}
$$

where $r_{t,t+1}$ stands for the return or loss between the end of day t and t+1. Although the hit function above plays a crucial role in a variety of backtesting procedures, the hit function itself contains far less information. Based on the hit function. Christoffersen (1998) construct the independence test which is known as a Markov test and it tests the independent property. In other words, this test tries to find whether the probability of VaR violation is related to the result of the previous day. If we define $N_{i,j}$, with $i = 0, 1, j = 0, 1$, as the number of days it takes that state j violates on one date while state $i$ violate on the previous date and the following $2 \times 2$ contingency table including all the possible outcome:

|  | $T_{n-1} = 0$ | $T_{n-1} = 1$ |  |
|---|---|---|---|
| $T_n = 0$ | $N_{0,0}$ | $N_{1,0}$ | $N_{0,0} + N_{1,0}$ |
| $T_n = 1$ | $N_{0,1}$ | $N_{1,1}$ | $N_{0,1} + N_{1,1}$ |
|  | $N_{0,0} + N_{0,1}$ | $N_{1,0} + N_{1,1}$ | N |

Table 4.4.1: The transition matrix of the independence test.

Let $p_{01} = \frac{N_{0,1}}{N_{0,0}+N_{0,1}}$ denote the conditional probability of transition from state 0 to 1. Given the condition that there was no failure on period $t - 1$, $p_{01}$ is the probability of having a failure at time t. Let $p_{11} = \frac{N_{1,1}}{N_{1,0}+N_{1,1}}$ be the conditional probability of transition from state 1 to 1. Given the condition that there was a failure at time $t - 1$, $p_{11}$ is the probability of having a failure at time t. The unconditional probability of having a failure at time t is $p = \frac{N_{0,1}+N_{1,1}}{N}$. Then the likelihood ratio of this independence test is as followed and it is asymptotically

distributed as a chi-square distribution with the degree of freedom equal to 1.

$$LR_{ind} = -2\ln\left[(1-p)^{N_{0,0}+N_{0,1}} \times p^{N_{0,1}+N_{1,1}}\right] + 2\ln\left[(1-p_0)^{N_{0,0}}p_0^{N_{0,1}}(1-p_1)^{N_{1,0}}p_1^{N_{1,1}}\right]$$

(4.4.8)

The Kupiec's proportion of failures (PoF) test is utilized here to check whether the proportion of violation is consistent with the $\alpha \times 100\%$ of the whole sample. It is an extension of the standard /unconditional coverage test. The statistic

$$POF = 2\ln\left[\left(\frac{1-\hat{\alpha}}{1-\alpha}^{n-I(\alpha)}\right)\left(\frac{\hat{\alpha}}{\alpha}\right)^{I(\alpha)}\right]$$

(4.4.9)

where $\hat{\alpha}$ is the average of hit function on $I(\alpha)$ ($\hat{\alpha} = \frac{1}{n}I(\alpha)$)and $I((\alpha) = \sum_{t=1}^{n} I_t(\alpha)$. We also use the Conditional coverage mix test which we also call joint test. It is a combination of the independent statistics with Kupiec's PoF-test. It measures the correct failure rate but also the independence of violation. The test statistic is

$$LR = LR_{ind} + POF$$

(4.4.10)

The simulation results can be found in Tables 4.4.2 and 4.4.3. The VaR we tested in the table is calculated based on equation 4.4.2, which is dynamic but with same GARCH coefficients for each iteration. Parameters of the GARCH model and innovation density functions for the univariate model are calculated using the first 500 data points(training set). For the copula-based model, the first 500 data points in each iteration is divided into training set and validation set. The optimal copula-based empirical Bernstein model is selected when the number of grids in each dimension maximize the loglikelihood of the validation set. The optimal copula-based penalized Bernstein model is selected when the penalization part maximize the loglikelihood in validation set. The average coverage is the mean through 100 iterations and all the results of hypothesis testing in the table are the count of rejections in these 100 simulations. The out-of-sample test is based on

|  | Average Coverage(%) | Std.Dev. of coverage(%) | Independence test (No. of Rejection) | CC | PoF |
|---|---|---|---|---|---|
| **In sample** | | | | | |
| Univariate | | | | | |
| M1 | 5.0 | 0.37 | 1 | 1 | 0 |
| M2 | 4.8 | 0.48 | 3 | 0 | 1 |
| Copula-Based Empirical | | | | | |
| M1 | 5.2 | 0.43 | 2 | 1 | 0 |
| M2 | 5.1 | 0.36 | 2 | 0 | 1 |
| Copula-Based Penalized | | | | | |
| M1 | 4.9 | 0.52 | 2 | 1 | 0 |
| M2 | 4.9 | 0.38 | 2 | 0 | 1 |
|  | Average Coverage(%) | Std.Dev. of coverage(%) | Independence test (No. of Rejection) | CC | PoF |
| **Out of sample** | | | | | |
| Univariate | | | | | |
| M1 | 5.4 | 3.14 | 6 | 21 | 21 |
| M2 | 5.2 | 3.05 | 5 | 22 | 24 |
| Copula-Based Empirical | | | | | |
| M1 | 5.5 | 2.80 | 6 | 20 | 17 |
| M2 | 5.6 | 2.68 | 4 | 20 | 21 |
| Copula-Based Penalized | | | | | |
| M1 | 5.3 | 1.65 | 6 | 11 | 11 |
| M2 | 5.2 | 1.65 | 4 | 15 | 16 |

Table 4.4.2: Monte Carlo results of the backtesting analysis when the DGP from Bekk.

|  | Average Coverage(%) | Std.Dev. of coverage(%) | Independence test (No. of Rejection) | CC | PoF |
|---|---|---|---|---|---|
| **In sample** | | | | | |
| Univariate | | | | | |
| M1 | 5.0 | 0.45 | 2 | 0 | 1 |
| M2 | 5.0 | 0.48 | 2 | 0 | 1 |
| Copula-Based Empirical | | | | | |
| M1 | 5.0 | 0.37 | 2 | 0 | 1 |
| M2 | 5.0 | 0.41 | 2 | 0 | 0 |
| Copula-Based Penalized | | | | | |
| M1 | 5.0 | 0.33 | 2 | 0 | 1 |
| M2 | 5.0 | 0.41 | 2 | 0 | 0 |
|  | Average Coverage(%) | Std.Dev. of coverage(%) | Independence test (No. of Rejection) | CC | PoF |
| **Out of sample** | | | | | |
| Univariate | | | | | |
| M1 | 5.2 | 1.82 | 4 | 11 | 17 |
| M2 | 5.0 | 1.43 | 5 | 7 | 10 |
| Copula-Based Empirical | | | | | |
| M1 | 5.2 | 1.64 | 5 | 7 | 13 |
| M2 | 5.1 | 1.29 | 5 | 7 | 7 |
| Copula-Based Penalized | | | | | |
| M1 | 5.0 | 1.32 | 4 | 5 | 9 |
| M2 | 5.1 | 0.92 | 4 | 5 | 6 |

Table 4.4.3: Monte Carlo results of the backtesting analysis when the DGP from CCC.

the prediction value using estimated information from the training set. The $\alpha$ in

both DGP is set to be 5%.

The simulation results show that the estimation of innovation directly determines the level of Value-at-Risk. For each simulation, the estimated parameters of the standardized innovation are quite similar to each other. But the final Value at Risk has a comparably large discrepancy. For the in-sample estimation, these three methods have the similar average coverage and pass almost the same number of tests in backtesting. However, for the out-of-sample estimation, the copula-based methods provide more precise prediction, especially the penalized model. The penalized copula-based model has smaller standard deviation of coverage for both margin 1 and margin 2 and shows a significant decrease in the number of violations in the hypothesis testing.
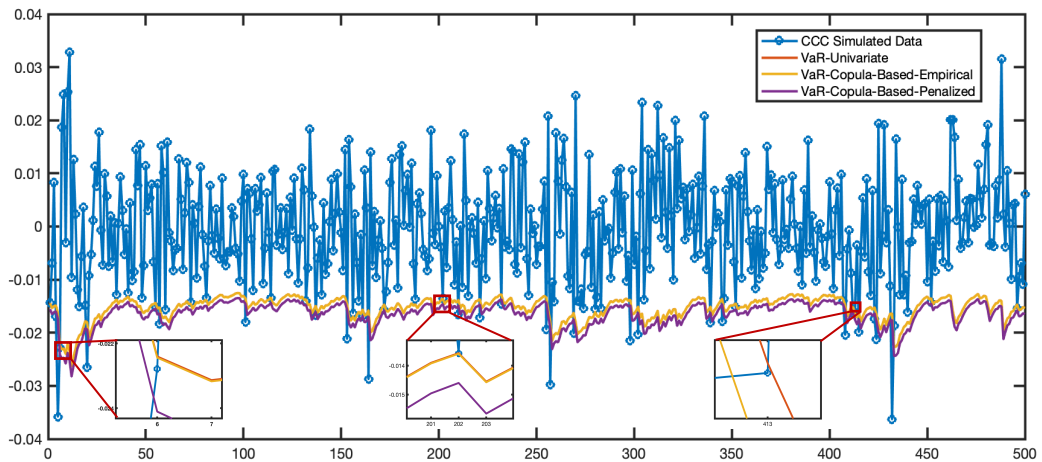


Figure 4.4.2: The estimated Value-at-Risks of marginal distribution 1 in an out-of-sample set.

## 4.5   Conclusion

In this chapter, we developed a three-step estimation method to improve the performance of the univariate density function. Compared with the univariate model, the copula-based model significantly provides more accuracy both in cross-

sectional data and time series data. The improvements are irrelevant to the linear dependency between random variables. Compared with the copula-based empirical model, the three-step method shows better prediction properties. In the context of VaR evaluation, some small changes in innovation estimation using penalized copula method can also lower the level of Value at Risk. Although the innovation distribution is pre-assumed, it is still valuable to study its properties and estimations.

# CHAPTER 5

## Pricing Rainbow Options Based on Dependence Structure Modeled by Penalized Bernstein Copula

## 5.1 Introduction

Rainbow (or multivariate) options are contracts linked to the performance of two or more underlying assets. They are usually calls (or puts) on the best (or worst) of several assets. One dominant aspect of evaluating rainbow options is the dependency between the underlying assets. Consider a bivariate call on min European option, it provides the holder with the right to purchase the minimum asset of the two for a pre-specified strike price. As the dispersion goes up(the two assets move in the opposite direction), the value of such an option decreases unequivocally. In general, the relationship between the underlying changes over time. The time-varying dependency involved methods can provide a more realistic and accurate valuation of rainbow options.

Option pricing has been studied extensively following the seminal work of Scholes and Black (1973) and Merton (1973). They provide the fundamental pro-

cedures for analyzing the value of contingent claim assets. Due to the increasing complexity of the options' structure, various research was conducted on modeling the multivariate option prices through traditional Black-Scholes model(Margrabe, 1978; Stulz, 1982). They estimate the inner dependency between assets by Pearson's correlation. However, unless the relationship between financial assets follows multivariate Gaussian distribution, the correlation coefficient is not a satisfactory choice for dependence modeling. Besides, the financial time series often exhibits stochastic volatility and jumps(Scaillet and Fermanian, 2002). Both features result in skewed and heavy-tailed distribution, which can be poorly captured by a lognormal density.

In this chapter, the relation between two assets in the bivariate option will be estimated through the nonparametric copula function. Using copula to price multivariate options is not a new topic. Rosenberg (1999) and Cherubini and Luciano (2002) firstly introduce copula in bivariate option pricing. Rosenberg (1999) use a Plackett copula to link the risk-neutral marginal distribution to obtain the bivariate risk-neutral density function. Rosenberg (2003) derive and implement a nonparametric, arbitrage-free estimator for marginal distribution and a kernel-based nonparametric copula to glue them together. The copula is applied statically. It works well in a relatively short period. While financial data often cover a reasonably long period, the economic environment may induce some changes in dependence structure. Zhang and Guegan (2008) indicates that the static Gaussian copula model gives higher option prices than the dynamic Gaussian copula model. So time-varying method is also considered in this chapter.

For pricing the contingent claims based on time series, Duan (1995) introduced the locally risk-neutral relationship(LRNVR) in the context that the asset return is assumed to follow the GARCH dynamic. It formulates that the one-period ahead

conditional variance is invariant with respect to the change to the risk-neutralized pricing measure. However, the LRNVR is based on conditional normality. Duan (1999) extend the LRNVR to generalized form which allows all kinds of innovations except student t. But it doesn't have an explicit formula for both log return and conditional volatility function, which increase the computational burden.

In our study, we adopt a nonparametric copula-based GJR-GARCH approach with generalized error distribution (GED) innovation to evaluate European bivariate rainbow options. The nonparametric model employed in our study captures the co-movement of residuals within the GARCH model. We recognize that the traditional local risk-neutral valuation relationship, which assumes conditional normality and connects the physical asset with the risk-neutral one, may not adequately account for the skewness and fat-tail properties observed in financial data. To address this limitation, we relax the assumption of conditional normality and derive a generalized risk-neutral valuation relationship with an explicit formula. To evaluate the performance of our approach, we compare it with existing parametric copula models in both static and dynamic scenarios for option pricing. The empirical results highlight the significance of selecting an appropriate copula for accurate option pricing across different scenarios, since the dependence structure changes under different measures(i.e. physical and risk-neutral measure). The remainder of this chapter is organized as follows: Section 5.2 provides background information on option valuation and explains the risk-neutral method based on the GJR-GARCH-GED model. In Section 5.3, we present the procedure for estimating the conditional copula. Section 5.4 presents the empirical study on bivariate option pricing, where we evaluate the performance of our nonparametric copula-based approach.

## 5.2 Preliminaries and related literature

### 5.2.1 Option valuation

The rainbow options include all the options whose payoff depends on more than one underlying risky asset. This chapter concentrates on European-type options with the bivariate call-on-max payoff. The technique can be sufficiently extended to other kinds of rainbow options as well. The payoff of a call-on-max option can be written as

$$\max(\max(S_{1,T}, S_{2,T}) - K, 0) \qquad (5.2.1)$$

where K is the strike price, T is the maturity, and $S_{i,t}$ is the value of the i-th asset (i=1,2) at time t. In the Black-Scholes model, it assumes that percentage changes in the asset price in a very short period are normally distributed. Assume that the expected return in a very short period is $\mu$ and the volatility of the underlying asset is $\sigma$. Then, the change of underlying asset $\Delta S$ in the asset $S$ in time $\Delta t$ follows:

$$\frac{\Delta S}{S} \sim \mathcal{N}(\mu \Delta t, \sigma^2 \Delta t)$$

Suppose $G = \ln S$, according to Ito's lemma

$$dG = \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma dz$$

In the simplest case, $\mu$ and $\sigma$ are constant, then this equation shows that $\ln S$ follows a generalized Wiener process. The log return of the underlying asset is therefore normally distributed between time 0 and some future time t.

$$\ln \frac{S_t}{S_0} \sim \mathcal{N}\left[\left(\mu - \frac{\sigma^2}{2}\right) t, \sigma^2 t\right]$$

Then, the individual asset price process in discrete time can be expressed as

$$S_t = S_0 \exp\left[\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t\right]$$

where $W_t$ stands for the Wiener process. Since observations are assumed to be equally spaced,

$$S_{t-1} = S_0 \exp\left[\left(\mu - \frac{\sigma^2}{2}\right)(t-1) + \sigma W_{t-1}\right]$$

which gives,

$$S_t = S_{t-1} \exp\left[\left(\mu - \frac{\sigma^2}{2}\right) + \sigma(W_t - W_{t-1})\right] \qquad (5.2.2)$$

For Wiener process, $W_{t+u} - W_t \sim \mathcal{N}(0, u)$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\Omega$ is a non-empty set, $\mathcal{F}$ is a $\sigma-$algebra of subsets of $\Omega$ and $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ is a probability measure on $\mathcal{F}$. Let

$$\epsilon_t | \mathcal{F}_{t-1} \overset{\mathbb{P}}{\sim} \mathcal{N}(0, 1)$$

which means $\epsilon_t$ is $\mathcal{F}_{t-1}$ measurable. $\mathcal{F}_t$ is the information set of all information through time t. 5.2.2 can be written as

$$S_t = S_{t-1} \exp\left(\mu - \frac{\sigma^2}{2} + \sigma\epsilon_t\right) \qquad (5.2.3)$$

Considering the financial data presenting usually clustering volatility and conditional heteroscedasticity, we apply a GARCH framework of Bollerslev (1986) to capture the dynamics of the volatility term of the underlying asset. Although various specifications of the GARCH model are implementable, Hansen and Lunde (2005) shows that selecting different orders of a GARCH model has trivial importance in improving forecasting accuracy. Also, another feature of the financial time series data is the leverage effect, which shows an asymmetric effect on the

volatility changes in response to positive and negative shocks to the same extent. To capture this effect, the GJR-GARCH model proposed by Glosten et al. (1993) is taken into consideration.

$$r_{i,t} = \mu_i + e_{i,t} \tag{5.2.4}$$

$$e_{i,t} = \sigma_{i,t}\varepsilon_{i,t}$$

$$\sigma_{i,t}^2 = \alpha_{i,0} + \alpha_{i,1}e_{i,t-1}^2 + \beta_i\sigma_{i,t-1}^2 + \eta_i e_{i,t-1}^2 I_{i,t-1} \tag{5.2.5}$$

where the $r_{i,t}$ stands for the return of the i-th market at time t and $I$ is the indicator function with:

$$I_{i,t-1} = \begin{cases} 1 & \text{if } \varepsilon_{i,t-1} < 0 \\ 0 & \text{if } \varepsilon_{i,t-1} \geq 0 \end{cases} \tag{5.2.6}$$

where the indicator term multiplying $\phi$ captures the asymmetry, $\alpha_{i,0} > 0$, $\alpha_{i,1} \geq 0$, $\beta_i \geq 0$ $\eta_i \geq 0$ secure that the conditional variance term is always positive and $\alpha_{i,1} + \beta_i + \frac{1}{2}\eta_i < 1$ is required for the convergence of the conditional variance. One common assumption of GARCH based option pricing model is that the return series of financial data is followed the conditional normality. However, the assumption is controversial. Bollerslev (1987) indicates that monthly S&P 500 Composite Index returns show conditional leptokurtosis and introduce the GARCH model with standardized student t innovation. Christoffersen et al. (2006) captures this conditional skewness through an inverse Gaussian distribution. Since the moment-generating function of a student t-distribution doesn't exist for all degrees of freedom, implementing t-distribution as the innovation of the GARCH model may lead to assuming ab unbounded expected asset return. To study the conditional fat-tail property, we employ the generalized error distribution (GED) as innovation Nelson (1991). Let $\mathcal{F}_{i,t}$ denote the $\sigma-$filed generated by all available information up to and including time t of i-th market and $f_\nu(\varepsilon_{i,t}|\mathcal{F}_{i,t-1})$ be the

conditional density function for $\varepsilon_{i,t}$.

$$\varepsilon_{i,t}|\mathcal{F}_{i,t-1} \sim G_\nu(\varepsilon_{i,t}) \text{ under measure } \mathbb{P}$$

$$f_\nu(\varepsilon_{i,t}|\mathcal{F}_{i,t-1}) = \frac{\nu \exp\left(-\frac{1}{2}\left|\frac{\varepsilon_{i,t}}{\theta(\nu)}\right|^\nu\right)}{2^{1+1/\nu}\theta(\nu)\Gamma(1/\nu)} \quad for \; 0 < \nu \leq \infty \qquad (5.2.7)$$

$$\text{where } \theta(\nu) = \left(\frac{2^{-\frac{2}{\nu}}\Gamma(1/\nu)}{\Gamma(3/\nu)}\right)^{1/2}$$

$\Gamma$ is the gamma function, the shape parameter $\nu$ determines the tail-fatness of the density function. 5.2.7 is a standard normal density function when $\nu = 2$ and it is the Laplace distribution when $\nu = 1$. As $\nu \to \infty$, the density converges pointwise to a uniform density. When $\nu > 2$, the density function has a thinner tail compared with the normal distribution. When $\nu < 2$, the fat-tail situation occurs. Duan (1999) indicates that even with $\nu = 1$, the expected asset return is still finite for $\sigma_{i,t} < 2$, and the utilization of GED innovation can analyze the fat-tail effect through the tradition of modeling the compound return continuously. Then, the conditional marginal distribution of $r_{i,t}$ can be defined as following:

$$P(r_{i,t} \leq r|\mathcal{F}_{i,t-1}) = P(\varepsilon_{i,t} \leq \frac{r - \mu_i}{\sqrt{\alpha_{i,0} + \alpha_{i,1}e_{i,t-1} + \beta_i\sigma_{i,t-1}^2 + \eta_i e_{i,t-1}^2 I_{i,t-1}}}|\mathcal{F}_{i,t-1})$$

## 5.2.2  Risk-neutralization with GJR-GARCH process

The risk-neutral valuation arises from one pivotal property of the Black-Scholes-Merton differential equation. It is that the equation does not include any variables affected by the risk preference of investors. In the risk-neutral world, the present value of any cash flow can be calculated by discounting its expected value at the risk-free rate. Camara (2003) derives a risk-neutral valuation(risk-neutral relationship, RNVR) equation to conduct a closed-form solution for European option pricing when the underlying asset has a transformed normal distribution

under a certain combination of preference and wealth distribution assumptions. Duan (1995) develops the locally risk-neutral valuation relationship(LRNVR) in the context that the asset return is assumed to follow the GARCH dynamic. The LRNVR formulates that the one-period ahead conditional variance is invariant concerning the change to the risk-neutralized pricing measure. However, the LRNVR is based on the conditional normality condition. To further relax the assumption, Duan (1999) conduct the generalized local risk-neutral valuation relationship (GLRNVR) to analyze the effect of conditional leptokurtosis and it shows that the LRNVR is still valid when conditional normality is restored by transformation.

We aim to find the bivariate risk-neutral return process from the objective marginals. Each marginal distribution can be transformed to risk neutralized one based on the risk-neutral valuation relationship. The previous section shows that the marginal process of individual asset return is assumed to be conditionally distributed under the physical measure $\mathbb{P}$ and the volatility term is a GARCH process with non-normal innovation. The conventional risk-neutral valuation relationship should be adjusted to take the heteroskedasticity of the return process into consideration. A risk-neutral probability measure $\mathbb{Q}$ is defined as follows:

**Definition 5.2.1.** *(Locally risk-neutralized pricing measure) Let $Y_t = ln\frac{U'(C_t)}{U'(C_{t-1})}$ be a process such that $Y_t|\mathcal{F}_{t-1}$ is normally distributed with constant mean and variance under the physical measure $\mathbb{P}$. Define the risk-neutralized measure $\mathbb{Q}$ as*

$$d\mathbb{Q} = exp\left((r - \rho)T + \sum_{s=1}^{T} Y_s\right) d\mathbb{P} \tag{5.2.8}$$

*then $\mathbb{Q}$ is a measure and equivalent to $\mathbb{P}$. Here, $C_t$ represents the aggregate consumption at time t, $U(C_t)$ is the strictly increasing utility function of the economic agent in the time separable and finitely additive exchange economy and $\rho$ is the*

*impatience factor.*

Duan (1999) indicates that the generalized risk neutralization has the invariance property which means that the underlying distribution for the transformed innovations keeps unchanged and the risk-neutral process only shifts the mean of the transformed innovation. Let $z_{i,t}$ be independent and identically distributed normal under the physical measure $\mathbb{P}$ and $z_{i,t}^*$ be independent and identically distributed normal under the risk-neutral measure $\mathbb{Q}$. Define the mean shift between the two measures by:

$$\lambda_{i,t} = z_{i,t}^* - z_{i,t} \tag{5.2.9}$$

For a GED-distributed innovation $\varepsilon_{i,t}$, the mapping can be written as

$$\lambda_{i,t} = z_{i,t}^* - \Phi^{-1}(G_\nu(\varepsilon_{i,t})),$$

where $\Phi^{-1}$ is the inverse cumulative distribution function (cdf) of the standard normal distribution function with respect to measuring $\mathbb{P}$. Then, $\Phi^{-1}(G_\nu(\varepsilon_{i,t}))$ is normally distributed with respect to measure $\mathbb{P}$. Combined with the generalized local risk-neutral valuation relationship(GLRNVR) theory by Duan (1999), the following

**Theorem 5.2.1.** ***Duan (1999)*** *A pricing measure $\mathbb{Q}$ is said to satisfy the generalized local risk-neutral valuation relationship (GLRNVR) if,*

1. *measure $\mathbb{Q}$ is mutually absolutely continuous with respect to the objective measure $\mathbb{P}$*

2. $E^{\mathbb{Q}}\left[\frac{S_{i,t}}{S_{i,t-1}} | \mathcal{F}_{i,t-1}\right] = \exp(r^f)$

3. *there exists a predictable process $\lambda_{i,t}$ such that*

$$z_{i,t}^* = \lambda_{i,t} + \Phi^{-1}(G_\nu(\varepsilon_{i,t}))|\mathcal{F}_{i,t-1} \sim \mathcal{N}(0,1) \text{ with respect to measure } \mathbb{Q}$$

The linear normal mapping in 5.2.9 as a nonlinear GED mapping given by

$$\varepsilon_{i,t} = G_\nu^{-1}(\Phi(z_{i,t}^* - \lambda_{i,t})).$$

It is necessary to solve for $\lambda_{i,t}$ to implement the risk-neutral model. The magnitude of mean shift $\lambda_{i,t}$ is determined by the second condition in theorem 5.2.1, which let the conditionally expected risk-neutral asset return in each period equal to the risk-free rate.

$$
\begin{aligned}
E^{\mathbb{Q}}\left[\frac{S_{i,t}}{S_{i,t-1}}|\mathcal{F}_{i,t-1}\right] &= E^{\mathbb{Q}}\left[\exp\{\mu_{i,t} + \sigma_{i,t}\varepsilon_{i,t}\}|\mathcal{F}_{i,t-1}\right] \\
&= E^{\mathbb{Q}}\left[\exp\{\mu_{i,t} + \sigma_{i,t}G_\nu^{-1}(\Phi(z_{i,t}^* - \lambda_{i,t}))\}|\mathcal{F}_{i,t-1}\right] \\
&= \exp(r^f)
\end{aligned}
$$

When $\nu = 2$, $G_\nu^{-1}(\Phi(z_{i,t}^* - \lambda_{i,t})) = z_{i,t}^* - \lambda_{i,t}$. It is not difficult to obtain an explicit solution in this case. When $\nu \neq 2$, the solution for $\lambda_{i,t}$ involves prohibitively cumbersome numerical calculations on each $t$. Here, we utilize the approximation method mentioned in Christoffersen et al. (2010). Since the normal and GED functions are both symmetric, and $G_\nu^{-1}(\Phi(0)) = 0$ for all the $\nu$. We use the linear approximation for

$$G_\nu^{-1}(\Phi(z)) \approx b_\nu z$$

here, $b_\nu$ can be found given the value of $\nu$ by fitting the linear model with a wide grid of z value.
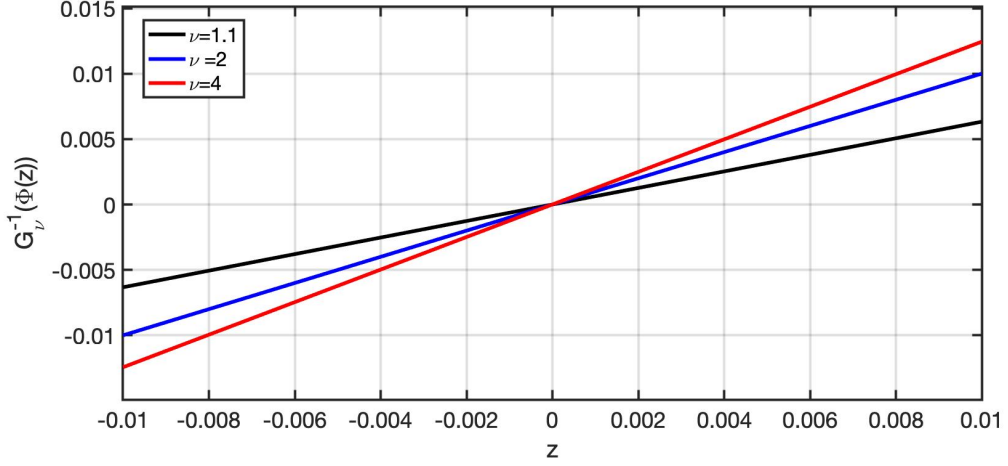
Figure 5.2.1: The scatter plot between $z$ and $G_\nu^{-1}(\Phi(z))$

With this approximation, and set $\mu_i = r_f + \theta\sigma_{i,t}$, here $\theta$ is the unit risk premium. $\theta = \frac{\mu - r_f}{\sigma}$, the $\sigma$ here is the long term standard deviation.

$$\exp(r_f) = E^{\mathbb{Q}}\left[\exp\{r_f + \theta\sigma_{i,t} + \sigma_{i,t}G_\nu^{-1}(\Phi(z_{i,t}^* - \lambda_{i,t}))\}|\mathcal{F}_{i,t-1}\right]$$

$$1 = \exp\{\theta\sigma_{i,t}\}E^{\mathbb{Q}}\left[\exp\{\sigma_{i,t}G_\nu^{-1}(\Phi(z_{i,t}^* - \lambda_{i,t}))\}|\mathcal{F}_{i,t-1}\right]$$

$$\approx \exp\{\theta\sigma_{i,t}\}E^{\mathbb{Q}}\left[\exp\{\sigma_{i,t}b_\nu(z_{i,t}^* - \lambda_{i,t})\}|\mathcal{F}_{i,t-1}\right]$$

then by the moment generating function for a normally distributed random variable we have,

$$1 = \exp\{\theta\sigma_{i,t}\}\exp\{-\lambda_{i,t}\sigma_{i,t}\}\exp\left\{\frac{1}{2}b_\nu^2\sigma_{i,t}^2\right\}$$

Taking logarithm on both side and solving for $\lambda_{i,t}$ yields

$$\lambda_{i,t} = \left(\frac{\theta}{b_\nu} + \frac{b_\nu}{2}\right)\sigma_{i,t}$$

The mapping between the physical GED and the risk-neutral normal shocks is

$$\varepsilon_{i,t} = G_\nu^{-1}\left(\Phi\left(z_{i,t}^* - \left(\frac{\theta}{b_\nu} + \frac{b_\nu}{2}\right)\sigma_{i,t}\right)\right) \tag{5.2.10}$$

117

Under some conditions, the generalized locally risk-neutral valuation relationship holds and the log-return process in 5.2.4 can be transformed to the one under the risk-neutral measure $\mathbb{Q}$:

**Theorem 5.2.2** (GJR-GARCH with conditional leptokurtosis). *Under the risk-neutral probability measure $\mathbb{Q}$, the model for one-period log return $r_{i,t}$ is*

$$r_{i,t} = r_f + \theta\sigma_{i,t}^2 + \sigma_{i,t}G_\nu^{-1}\left(\Phi\left(z_{i,t}^* - \lambda_{i,t}\right)\right)$$

$$\sigma_{i,t}^2 = \alpha_{i,0} + \alpha_{i,1}(G_\nu^{-1}(\Phi(z_{i,t-1}^* - \lambda_{i,t-1}))\sigma_{i,t-1})^2 + \beta_i\sigma_{i,t-1}^2 + \eta_i(G_\nu^{-1}(\Phi(z_{i,t-1}^* - \lambda_{i,t-1}))\sigma_{i,t-1})^2 I_{i,t-1}$$

$$\lambda_{i,t} = \left(\frac{\theta}{b_\nu} + \frac{b_\nu}{2}\right)\sigma_{i,t}$$

*where $z_{i,t}^*$ conditional on $\mathcal{F}_{i,t-1}$ is a $\mathbb{Q}-$standard normal random variable, and $G_\nu^{-1}[\cdot]$ represents the inverse GED cumulative density function with parameter $\nu$ and $\theta$ is the price of risk.*

It's obvious to find that when the underlying innovation is normally distributed $\nu = 2$, $b_\nu = 1$, the formula for the one-period log-return becomes $r_{i,t} = r_f - \frac{1}{2}\sigma_{i,t} + \sigma_{i,t}z_{i,t}^*$. Theorem 5.2.2 provides a relatively approximate efficient way from the physical model to the risk-neutral one. Based on this theorem, the following corollary can be defined for the terminal asset price.

**Corollary 5.2.1.** *When the generalized local risk-neutral valuation relationship (GLRNVR) holds, the i-th asset price for time $T$ can be expressed as*

$$S_{i,T} = S_{i,t}\exp\left\{(T-t)r_f + \sum_{s=t+1}^{T}\theta\sigma_{i,t}^2 + \sum_{s=t+1}^{T}\sigma_{i,t}G_\nu^{-1}(\Phi(z_{i,t-1}^* - \lambda_{i,t-1}))\right\}$$

$$(5.2.11)$$

*Proof.* From the theorem 5.2.2, we have

$$\ln\frac{S_t}{S_{t-1}} = r_f + \theta\sigma_{i,t}^2 + \sigma_{i,t}G_\nu^{-1}\left(\Phi\left(z_{i,t}^* - \lambda_{i,t}\right)\right)$$

for every $t \in \mathbb{R}$ under $\mathbb{Q}$. Thus,

$$\ln \frac{S_T}{S_t} = \sum_{T}^{s=t+1} \ln \frac{S_s}{S_{s-1}}$$

$$= \sum_{s=t+1}^{T} r_f + \theta \sigma_{i,s}^2 + \sigma_{i,s} G_\nu^{-1} \left( \Phi \left( z_{i,s}^* - \lambda_{i,s} \right) \right)$$

$$= r_f(T-t) + \theta \sum_{s=t+1}^{T} \sigma_{i,s}^2 + \sum_{s=t+1}^{T} \sigma_{i,s} G_\nu^{-1} \left( \Phi \left( z_{i,s}^* - \lambda_{i,s} \right) \right)$$

which means that

$$S_T = S_t \exp \left( r_f(T-t) + \theta \sum_{s=t+1}^{T} \sigma_{i,s}^2 + \sum_{s=t+1}^{T} \sigma_{i,s} G_\nu^{-1} \left( \Phi \left( z_{i,s}^* - \lambda_{i,s} \right) \right) \right)$$

by taking exponents on both sides of the equation. $\qquad\square$

The asset price which is discounted at the risk-free rate has the martingale property. The theory of contingent claim pricing' martingale properties has been introduced by Harrison and Kreps (1979) and further extend by Harrison and Pliska (1981). Since $\sigma_{i,t} z_{i,t^*}$ is conditionally normal with mean zero and variance $\sigma_{i,t}^2$ under the probability measure $\mathbb{Q}$, the discount asset price process $e^{-r_f t} S_{i,t}$ is a $\mathbb{Q}$-martingale.

Under the locally risk-neutral probability measure $\mathbb{Q}$, the rainbow (call-on-max) option with strike price K maturing at time T has the value at time t:

$$RO_t = e^{-(T-t)r_f} E^{\mathbb{Q}}[\max(\max(S_{1,T}, S_{2,T}) - K, 0)|\mathcal{F}_t]. \qquad (5.2.12)$$

Under the GJR-GARCH(1,1) specification, $\mathcal{F}_t = \{\mathcal{F}_{1,t}, \mathcal{F}_{2,t}\}$. It is the $\sigma$-field generated by $\{S_{i,t}, G_\nu^{-1} \left( \Phi \left( z_{i,t}^* - \lambda_{i,t} \right) \right), \sigma_{i,t}^2\}$.

## 5.3  Option pricing under GJR-GARCH process with penalized copula

To price the rainbow options dynamically, a time-varying copula approach is applied. The bivariate distribution of the log returns $(r_{1,t}, r_{2,t})$ is stated conditionally on the information set of all the information till time t-1, $\mathcal{F}_{t-1} = \sigma(r_{1,m}, r_{2,m})$ and $m \leq t - 1$. All conditional margins only depend on their own past. Based on Sklar's theorem of copula estimation, we estimated the marginal distributions of each log return and the copula function separately. The marginal distribution is modeled by the GJR-GARCH process with GED innovation to capture the asymmetry and potential heavy tails.

Considering the complex relationship between the assets during some special financial periods (like COVID-19), we use the penalized Bernstein copula to capture the inner dependence. The merit and main properties of this nonparametric copula have been thoroughly explained in Chapter 3. Since most of the data sets contain relatively long time periods, the difference in the macroeconomic environment may induce changes in the dependence structure. Consequently, we conduct a time-varying approach in moving windows based on the penalized Bernstein copula.

According to the formula 5.2.12, the value of the rainbow option price relies on the expectation of the payoff under risk-neutral measure $\mathbb{Q}$. Instead of converting the joint distribution into its risk-neutral counterpart as a whole, the physical margins explained in equation 5.2.5 are transformed into functions in theorem 5.2.2 respectively. Then, the joint distribution of assets should be under measure $\mathbb{Q}$, and the copula function is under $\mathbb{Q}$. Also, we assume that the optimal parametric

copula in the physical world and the risk-neutral world belong to the same family.

To estimate the rainbow option(call-on max), the dependency between $r_{1,t}$ and $r_{2,t}$ needs to be estimated. As the standardized innovation in the GJR-GARCH model is assumed to follow the generalized error distribution, the bivariate conditional distribution function is

$$H((r_{1,t}, r_{2,t})) = H((\varepsilon_{1,t}, \varepsilon_{2,t})) = C(G_{\nu_1}(\varepsilon_{1,t}), G_{\nu_2}(\varepsilon_{2,t}))$$

where $C$ is the copula function and $G_\nu$ is the cumulative density function of generalized error distribution with shape parameter $\nu$. Then, the corresponding conditional density function is

$$h((\varepsilon_{1,t}, \varepsilon_{2,t})) = c(G_{\nu_1}(\varepsilon_{1,t}), G_{\nu_2}(\varepsilon_{2,t})) \prod_{i=1}^{2} g_{\nu_i}(\varepsilon_{i,t}) \tag{5.3.1}$$

where the copula c is given by

$$c(u_1, u_2) = \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2}$$

with $(u_1, u_2) \in [0,1]^2$. Based on the IFM method, we estimate the marginal distribution parameter first,

$$\hat{\nu}_i = \arg\max_{\nu_i} \sum_{t=1}^{n} \log g_{\nu_i}(\varepsilon_{i,t})$$

and

$$\hat{\gamma}_t = \arg\max_{\gamma_t} \sum_{i=1}^{n} \log c(G_{\nu_1}(\varepsilon_{1,t}), G_{\nu_2}(\varepsilon_{2,t}); \gamma_t) - \lambda pen(\gamma)$$

From this estimation, the coefficients of dependence structure have been fitted for each window.

## 5.4 Empirical evidence

### 5.4.1 Analysis of data set

The analysis approach for rainbow(bivariate) option price using dynamic copula function with GJR-GARCH process to monitor the individual marginal distribution outlined in the previous section. This section shows the detailed analyzing procedure with the call-on-max option on the index from both developed markets and emerging markets. The sample includes 4 country indexes (United States, United Kingdom, Singapore, Germany ) from developed markets and 4 country indexes (Greece, China, India, Korea) from emerging markets. The data used to represent the eight representative markets is the daily index price of Morgan Stanley Capital International (MSCI). The country index from MSCI is designed to represent the performance of securities exhibiting the value/growth characteristics. The index returns span 2 years from January 2020 to December 2021 for a total of 522 observations expressed in US dollars. Figure5.4.1 shows how the market index prices change over this period. The market index prices of the United States, Singapore, the United Kingdom, and Germany experienced an enormous decrease in March 2020, when COVID-19 started to spread over the world. And they also show a similar downturn in November 2020. On the contrary, emerging markets do not display significant synchrony over the period.
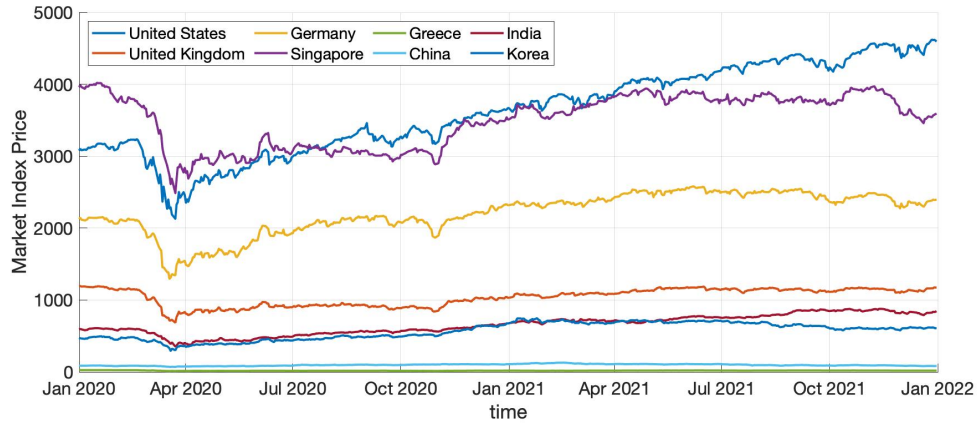
Figure 5.4.1: The market index price plot for all the markets and the index price span 2 years from January 2020 to December 2021.

Descriptive Statistics for the United States(US), United Kingdom(UK), Germany(DE), Singapore(SG), Greece(GR), China(CN), India(IN), and Korea(KR) returns are presented in Table 5.4.1. The market of the United States shows the highest mean and Indian shows the highest median in index returns. The market in Greece has the highest volatility. The unconditional distributions of all the markets' returns show negative skewness and expected express kurtosis.

The results of the Jarque-Bera test indicate that neither of these equity returns is normally distributed under 1% significance level. In addition, the pair DE-UK has the highest linear correlation based on both the correlation coefficient (0.8835) and Kendall's tau(0.6074). Four pairs ( US-SG, US-IN, US-KR, and GR-CN ) show very low correlation, while all the others have a similar moderate correlation(around 0.5 for Pearson correlation coefficients and around 0.3 for Kendall's tau). The least dependent pair is US-KR (0.2377 for correlation coefficients and 0.1102 for Kendall's tau). The familiar volatility clustering effect with some extraordinarily large absolute value can be observed from the time series plots 5.4.2 of index returns for all these four markets.

| | Developed Markets | | | | Emerging Markets Scenario | | | |
|---|---|---|---|---|---|---|---|---|
| | United States | United Kingdoms | Germany | Singapore | Greece | China | India | Korea |
| **Summary Statistics** | | | | | | | | |
| Mean(%) | 0.0768 | -0.0026 | 0.0232 | -0.0186 | -0.0611 | -0.0033 | 0.0680 | 0.0478 |
| Median(%) | 0.1100 | 0.1006 | 0.1048 | 0.0000 | 0.0354 | 0.0401 | 0.1204 | 0.0413 |
| Min(%) | -12.9221 | -14.2054 | -15.0943 | -7.9324 | -14.3226 | -6.0908 | -15.6226 | -10.9966 |
| Max(%) | 8.9831 | 10.9954 | 10.2431 | 7.3187 | 8.5469 | 4.9430 | 9.1706 | 10.1407 |
| Std. Dev. | 0.0163 | 0.0163 | 0.0163 | 0.0133 | 0.0216 | 0.0154 | 0.0166 | 0.0184 |
| Skewness | -1.1193 | -1.1600 | -1.5325 | -0.3721 | -1.1993 | -0.4018 | -2.1016 | -0.0248 |
| Kurtosis | 18.3438 | 19.0081 | 21.3669 | 10.2816 | 10.5341 | 4.4767 | 23.2309 | 9.3420 |
| Jarque-Bera statistic | 5229.6693 | 5690.7004 | 7541.5232 | 1165.2721 | 2486.8599 | 61.4752 | 9286.2788 | 874.8456 |
| Jarque-Bera p-value | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.00000 | 0.0000 |
| **Correlation matrix** | | | | | | | | |
| | United States | United Kingdom | Germany | Singapore | Greece | China | India | Korea |
| United States | 1.0000 | 0.6301 | 0.6220 | 0.3727 | 0.5447 | 0.4074 | 0.3769 | 0.2377 |
| | | (0.3070) | (0.3473) | (0.1598) | (0.2322) | (0.2238) | (0.1614) | (0.1102) |
| United Kingdom | | 1.0000 | 0.8835 | 0.6189 | 0.6560 | 0.4480 | 0.5747 | 0.4651 |
| | | | (0.6074) | (0.3574) | (0.3474) | (0.2732) | (0.3004) | (0.2449) |
| Germany | | | 1.0000 | 0.5948 | 0.6653 | 0.4544 | 0.5415 | 0.4474 |
| | | | | (0.3574) | (0.3769) | (0.2768) | (0.3160) | (0.2527) |
| Singapore | | | | 1.0000 | 0.4938 | 0.5066 | 0.6496 | 0.6947 |
| | | | | | (0.2613) | (0.3426) | (0.3442) | (0.4180) |
| Greece | | | | | 1.0000 | 0.3260 | 0.5087 | 0.3806 |
| | | | | | | (0.1643) | (0.2475) | (0.2090) |
| China | | | | | | 1.0000 | 0.4566 | 0.5253 |
| | | | | | | | (0.2662) | (0.3504) |
| India | | | | | | | 1.0000 | 0.5633 |
| | | | | | | | | (0.3342) |
| Korea | | | | | | | | 1.0000 |

Table 5.4.1: The table shows the summary statistics of the daily index return of four developed markets and four emerging markets. Returns are defined as $r_t = \ln S_t / S_{t-1}$ and $S_t$ is the index of time t. The correlation matrix shows the correlation coefficients and Kendall's taus in bracket
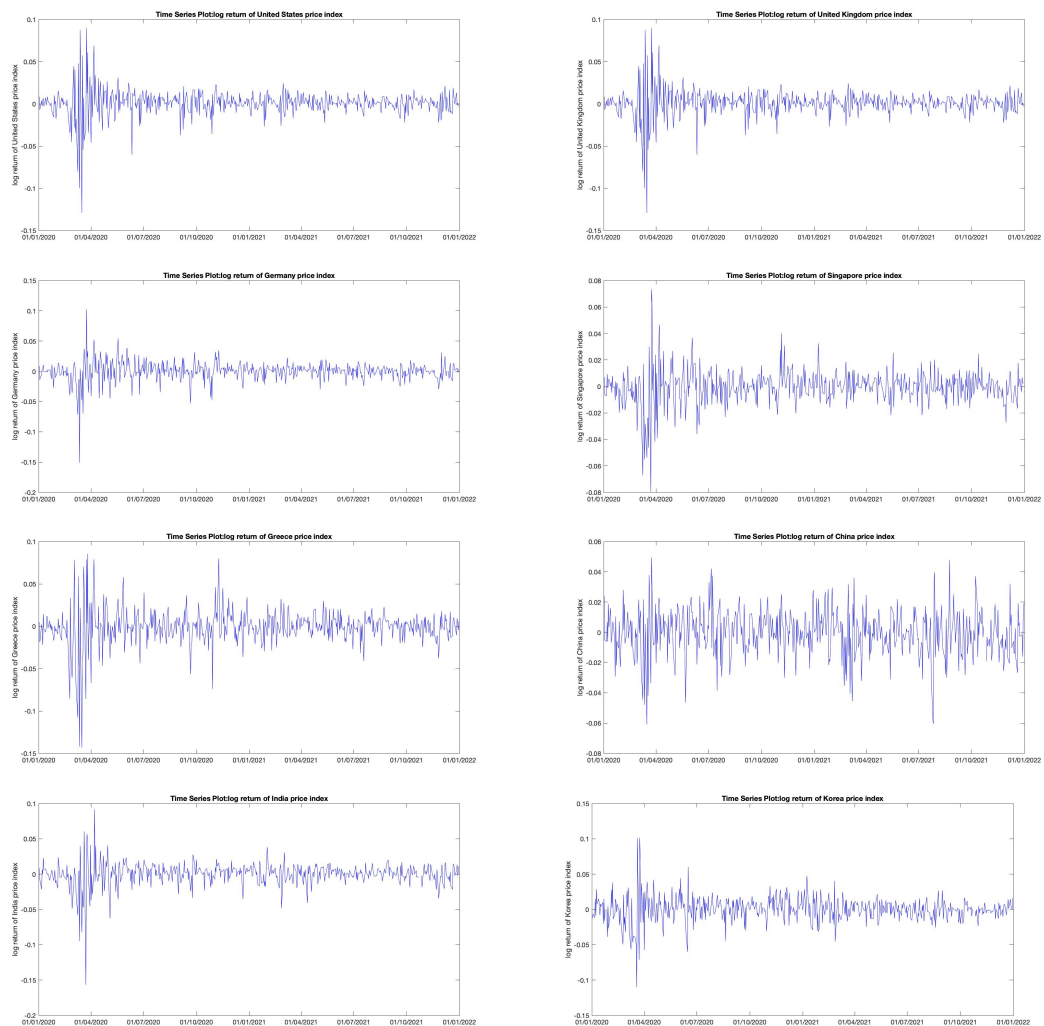
Figure 5.4.2: The time series plots for the log return of the United States, United Kingdom, Germany, Singapore, Greece, China, India, and Korea from Jan. 2020 to Dec. 2021.

## 5.4.2 The copula estimation method

We model the marginal distribution of returns in each individual market, which are assumed to be GED distributed. The estimation results of selected models are shown in Table 5.4.2. The GJR-GARCH(1,1) is employed to capture the conditional heteroscedasticity in returns. Since SG, CN and KR do not show significant asymmetry effects, so the GARCH(1,1) model is utilized for these three markets' returns. The asymmetry detected coefficient $\eta$ for US, UK, DE, GR, and

IN are all significant under 5% significance level and they are all positive. It indicates that the negative shocks provide more volatility than positive shocks of the same size in the subsequent period.

| | GARCH models | | | | Shape Parameter | Log-Likelihood | AIC | BIC |
|---|---|---|---|---|---|---|---|---|
| | $\alpha_{i,0}$ | $\alpha_{i,1}$ | $\beta_i$ | $\eta_i$ | $\nu$ | | | |
| | (s.e.) | (s.e.) | (s.e.) | (s.e.) | (s.e.) | | | |
| United | 9.6359E-06*** | 0.0945** | 0.6614*** | 0.4861*** | 1.2697*** | 1649.937 | -6.3024 | -6.2617 |
| States | (0.0000) | (0.0296) | (0.0423) | (0.1420) | (0.0954) | | | |
| United | 3.0661E-06 | 0.01661 | 0.925***6 | 0.0775* | 1.1008*** | 1554.13 | -5.9354 | -5.8946 |
| Kingdom | (0.0000) | (0.0698) | (0.0994) | (0.0361) | (0.0882) | | | |
| Singapore | 7.6492E-06*** | 0.1635*** | 0.7853*** | | 1.3118*** | 1639.762 | -6.2673 | -6.2347 |
| | (0.0000) | (0.0300) | (0.0340) | | (0.1067) | | | |
| Germany | 2.2623E-06** | 1.6833E-07 | 0.9269*** | 0.1142** | 1.0619*** | 1563.23 | -5.9702 | -5.9294 |
| | (0.0000) | (0.0048) | (0.0202) | (0.0354) | (0.0864) | | | |
| China | 3.6210E-05* | 0.1430** | 0.6987*** | | 1.5488** | 1467.784 | -5.6139 | -5.5731 |
| | (0.0000) | (0.0520) | (0.1050) | | (0.1055) | | | |
| Greece | 1.0180E-06*** | 0.0384* | 0.8652*** | 0.1201* | 1.1447*** | 1438.382 | -5.4919 | -5.4511 |
| | (0.0000) | (0.0197) | (0.0253) | (0.0490) | (0.0896) | | | |
| Korea | 1.6055E-05 | 0.15440** | 0.7878*** | | 1.6068*** | 1439.93 | -5.5012 | -5.4604 |
| | (0.0000) | (0.0501) | (0.0602) | | (0.2031) | | | |
| India | 6.6459E-06*** | 2.2120E-08 | 0.8829*** | 0.1464*** | 1.1896*** | 1576.977 | -6.0205 | -6.0045 |
| | (0.0000) | (0.0076) | (0.0174) | (0.0506) | (0.1195) | | | |

Table 5.4.2: The table shows the Quasi-Maximum likelihood estimates for the parameters of the marginal distribution model. The models are selected according to Akaike Information Criterion(AIC), Bayesian Information Criterion(BIC), and their statistical features.

In addition, the shape parameter $\nu$ for all the markets is significant and the values are between 1 and 2, which means the density function for these markets shows a fatter tail compared with the normal distribution. The shape parameter is far lower for Germany than for the other economies, implying the DE returns have the fattest tails during this time period.

Based on previous analysis, we mainly focus on studying the bivariate rainbow option constructed by the six pairs of the market: US-UK, UK-DE, DE-GC, SG-KR,

126

CN-KR, and IN-GC, which are representative: two pairs are within the developed markets, two pairs are across the developed and emerging market and two are within the emerging markets. And they all show the highest linear dependency in each category. We separate the data into two samples: one for training and the other for testing. Figure 5.4.3 draws the scatter plot of observed pairs of ranks for the estimated GJR-GARCH(some are GARCH). The data points of most pairs are more concentrated in the left corner, which is similar to other financial time series data. It means that these eight market equity indexes have the aggregation effect when the market is down-turn. The standard innovations($\varepsilon_{i,t} = e_{i,t}/\sigma_{i,t}$) of the six pairs of markets.

The Ljung-Box test for autocorrelation and the Kolmogorov-Smirrnov tests for density specification are also employed to secure the correct specification of margins. The test statistics and p values are displayed in table 5.4.3. All models pass the Ljung-Box test at 10% significance level and all the models pass the KS test at 1% significance level. It shows that the marginal models are well-specified.

| | Ljung box | | KS | |
|---|---|---|---|---|
| Margins | LB statistics | p-value | KS statistics | p-value |
| US | 7.4459 | 0.9949 | 0.0826 | 0.0565 |
| UK | 21.4173 | 0.3729 | 0.0488 | 0.5628 |
| DE | 24.6287 | 0.2160 | 0.0638 | 0.2380 |
| SG | 21.2947 | 0.3800 | 0.0312 | 0.9614 |
| GC | 7.7688 | 0.9933 | 0.0349 | 0.9071 |
| CN | 12.7357 | 0.8884 | 0.0414 | 0.7630 |
| IN | 17.2703 | 0.6354 | 0.0897 | 0.0298 |
| KR | 15.8054 | 0.7286 | 0.0559 | 0.3877 |

Table 5.4.3: The statistical results from the Ljung-Box test and Kolmogorov–Smirnov test for the marginal distributions of the eight markets.
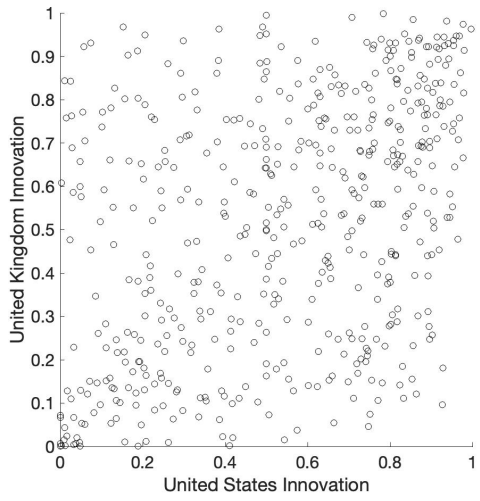
(a) US-UK
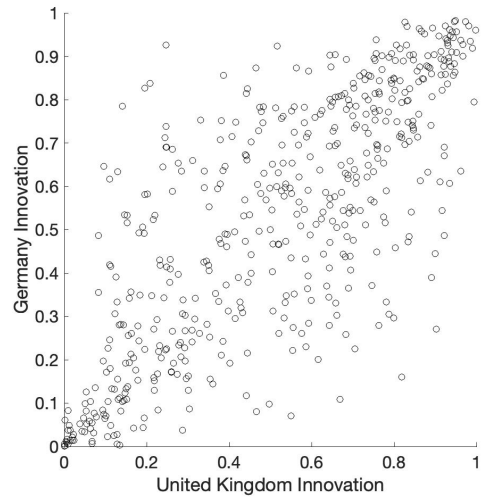
(b) UK-DE
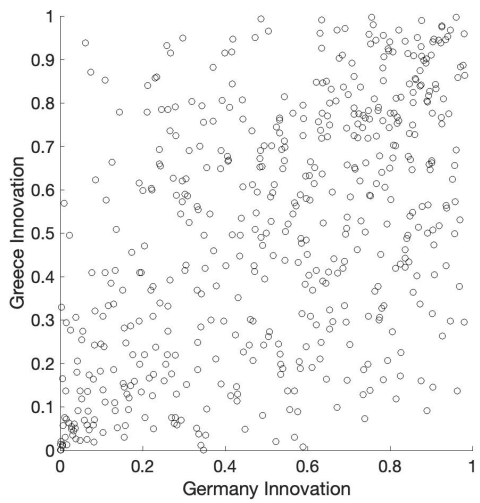
(c) DE-GC
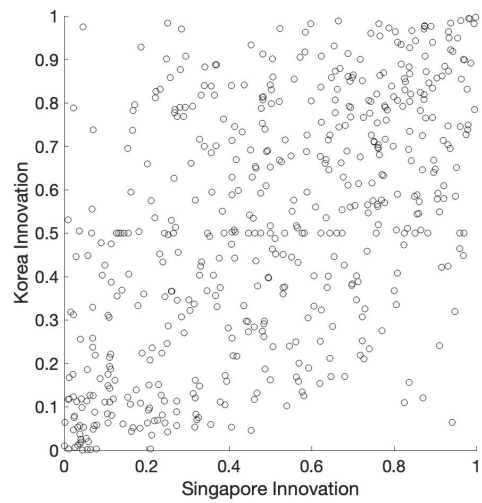
(d) SG-KR
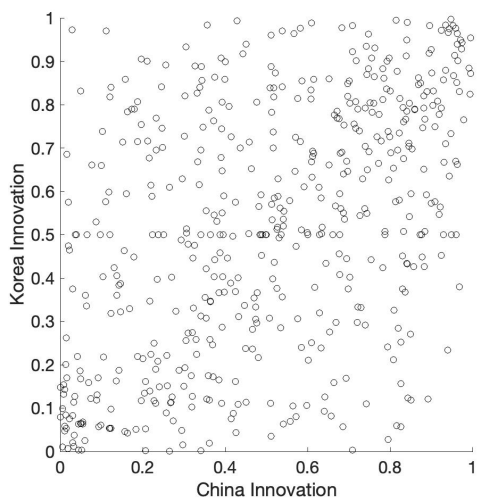
(e) CN-KR
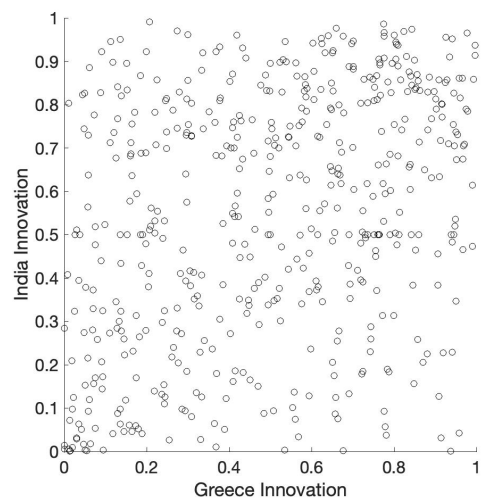
(f) GC-IN

Figure 5.4.3: Support set of the empirical copula of the standardized GJR-GARCH innovations(GED)

For the copula estimation, we select Gaussian, student t, and three Archimedean copulas: Frank, Gumbel, and Clayton as the parametric copula function to do the comparison. Table 5.4.4 shows the estimation result for the dependence structure of these assets on the whole period. All the copulas are fitted to the support set of the standardized innovations from the GJR-GARCH(1,1) and GARCH(1,1) models respectively. Among all the parametric families, student t copula shows the highest log-likelihood for all six pairs and it also has the smallest AIC.

| Copula Type | | US-UK | UK-DE | DE-GC | SG-KR | CN-KR | GC-IN |
|---|---|---|---|---|---|---|---|
| | Parameters | 0.4562 | 0.8057 | 0.5500 | 0.5871 | 0.4882 | 0.3712 |
| Gaussian | log-likelihood | 60.7028 | 273.3716 | 93.6497 | 109.9528 | 70.9808 | 37.8580 |
| | AIC | -119.4056 | -544.7432 | -185.2993 | -217.9056 | -139.9615 | -73.7160 |
| | Parameters | 0.4509 | 0.8059 | 0.5503 | 0.5883 | 0.4913 | 0.3683 |
| | d.o.f. | 7.5108 | 4.3702 | 7.8994 | 18.0453 | 18.8353 | 9.7282 |
| Student t | log-likelihood | 63.9732 | 282.4488 | 97.0611 | 111.1571 | 71.7572 | 41.2764 |
| | AIC | -123.9463 | -560.8976 | -190.1223 | -218.3143 | -139.5143 | -78.5528 |
| | Parameters | 2.8520 | 7.5922 | 3.7109 | 4.1337 | 3.3580 | 2.2406 |
| Frank | log-likelihood | 50.6505 | 236.1553 | 81.5889 | 100.0812 | 68.7534 | 33.2354 |
| | AIC | -99.3010 | -470.3106 | -161.1777 | -198.1624 | -135.5067 | -64.4708 |
| | Parameters | 1.4469 | 2.4307 | 1.5664 | 1.6019 | 1.4632 | 1.3017 |
| Gumbel | log-likelihood | 60.0655 | 264.3358 | 82.4954 | 97.9269 | 66.5909 | 30.9851 |
| | AIC | -118.1309 | -526.6716 | -162.9908 | -193.8538 | -131.1818 | -59.9703 |
| | Parameters | 0.5092 | 1.9658 | 0.8180 | 0.8677 | 0.5890 | 0.4351 |
| Clayton | log-likelihood | 49.5923 | 233.1102 | 89.3525 | 91.5353 | 52.9370 | 37.7950 |
| | AIC | -97.1846 | -464.2205 | -176.7050 | -181.0707 | -103.8739 | -73.5899 |

Table 5.4.4: Parametric copula fitting results. For the student t copula, the first parameter is correlation and the second is the degree of freedom.

Then, to observe whether the dependence structure has changed during this period, we utilize the moving window to show the variation of dependency. The whole sample is divided into 21 windows, and each window consists of 207 observations. The window is moved per 15 observations. Along with the window

movement, the best fitting copula on the corresponding sub-sample is selected through the log-likelihood based on the cross-validation method.

| Windows | US-UK | | UK-DE | | DE-GC | |
|---|---|---|---|---|---|---|
| i-th | Copula | Parameter | Copula | Parameter | Copula | Parameter |
| 1 | Gumbel | 1.7191 | t | 0.8943 1.7907 | Clayton | 1.2108 |
| 2 | Gumbel | 1.7264 | Gumbel | 3.3881 | Clayton | 1.0311 |
| 3 | Clayton | 0.8062 | t | 0.8893 2.2588 | Clayton | 0.9009 |
| 4 | Frank | 3.9029 | Gaussian | 0.8812 | t | 0.6047 12.5968 |
| 5 | Frank | 3.0690 | Gaussian | 0.8487 | Gaussian | 0.5498 |
| 6 | Frank | 2.8306 | Gaussian | 0.8366 | t | 0.5214 26.9187 |
| 7 | Frank | 2.3906 | Clayton | 2.0175 | Gaussian | 0.5123 |
| 8 | Frank | 2.1870 | Frank | 7.5906 | t | 0.4471 10.6564 |
| 9 | Frank | 2.2302 | t | 0.8144 9.4870 | Clayton | 0.5686 |
| 10 | Frank | 2.0673 | t | 0.7885 20.7487 | Clayton | 0.7189 |
| 11 | Clayton | 0.3007 | Frank | 7.5104 | Clayton | 0.8124 |
| 12 | Frank | 2.0866 | Gumbel | 2.1646 | Clayton | 0.9550 |
| 13 | Gumbel | 1.2828 | t | 0.7851 16.3412 | Clayton | 0.7764 |
| 14 | Gumbel | 1.2372 | Gaussian | 0.7463 | Gaussian | 0.5476 |
| 15 | Frank | 1.8268 | Gaussian | 0.7408 | Clayton | 0.6454 |
| 16 | Frank | 1.5573 | t | 0.7459 67.4079 | Gaussian | 0.4755 |
| 17 | Frank | 1.4206 | t | 0.7263 16.9855 | t | 0.4777 6.0106 |
| 18 | Frank | 1.3300 | t | 0.7105 34.0511 | Clayton | 0.6694 |
| 19 | Gumbel | 1.2569 | t | 0.7173 3.18E+06 | Clayton | 0.5226 |
| 20 | Gumbel | 1.2755 | Gumbel | 1.9825 | Clayton | 0.5316 |
| 21 | t | 0.3851 5.90361971 | Gumbel | 2.3130 | Clayton | 0.5582 |

Table 5.4.5: Dynamic copula analysis using a moving window for US-UK, UK-DE, DE-GC. For student t copula, the first parameter is correlation and the second is the degree of freedom.

| Windows | SG-KR | | CN-KR | | GC-IN | |
|---|---|---|---|---|---|---|
| i-th | Copula | Parameter | Copula | Parameter | Copula | Parameter |
| 1 | t | 0.7275 1.30E+07 | Frank | 6.4550 | Gumbel | 1.4547 |
| 2 | Clayton | 0.8255 | Clayton | 0.8458 | Frank | 2.9074 |
| 3 | Clayton | 0.8472 | Clayton | 0.9059 | t | 0.4132 11.6743 |
| 4 | t | 0.6845 1.19E+07 | Clayton | 0.8185 | Gaussian | 0.3594 |
| 5 | t | 0.6535 1.31E+07 | Clayton | 0.5715 | t | 0.2498 9.5732 |
| 6 | t | 0.6190 1.29E+07 | t | 0.4742 1.30E+07 | t | 0.2757 8.8084 |
| 7 | t | 0.6194 1.13E+07 | Gaussian | 0.4634 | t | 0.2431 15.9787 |
| 8 | Gaussian | 0.5866 | Gaussian | 0.4347 | t | 0.3029 24.0655 |
| 9 | Gumbel | 1.5323 | Frank | 2.4871 | Frank | 2.3331 |
| 10 | Gumbel | 1.4579 | Frank | 2.5566 | Gumbel | 1.3004 |
| 11 | Gumbel | 1.4842 | t | 0.3249 7.4457 | Frank | 3.0358 |
| 12 | Clayton | 0.8104 | t | 0.3896 8.7325 | Clayton | 0.4018 |
| 13 | Clayton | 0.9857 | Gaussian | 0.4182 | Clayton | 0.4559 |
| 14 | Clayton | 0.8530 | Gaussian | 0.4403 | Clayton | 0.3278 |
| 15 | Clayton | 0.8591 | Gumbel | 1.3277 | Clayton | 0.3308 |
| 16 | Clayton | 1.1275 | Gumbel | 1.3931 | Clayton | 0.3301 |
| 17 | t | 0.5324 3.1339 | Gumbel | 1.4047 | Clayton | 0.2765 |
| 18 | Clayton | 0.8898 | Clayton | 0.5505 | Clayton | 0.2123 |
| 19 | Clayton | 0.8453 | Gumbel | 1.4673 | Clayton | 0.2131 |
| 20 | Clayton | 0.8265 | Gumbel | 1.4049 | t | 0.1845 19.5530 |
| 21 | Clayton | 0.9500 | Gaussian | 0.3780 | t | 0.1714 11.3939 |

Table 5.4.6: Dynamic copula analysis using a moving window for SG-KR, CN-KR, GC-IN. For student t copula, the first parameter is correlation and the second is the degree of freedom,

The penalized Bernstein copula model is estimated through these 21 windows and the comparison of the optimal estimated parametric model and the optimal penalized Bernstein copula model is displayed in table 5.4.7. Except for some values for the pair of UK-DE, the penalized model has a comparably higher loglikelihood than the optimal parametric copula. From the moving windows table for

the parametric copula, all six pairs show significant variation in dependence structure. Most of the US-UK pair select Frank copula and its parameter show notable changes over time.

| Windows | US-UK | | UK-DE | | DE-GC | | SG-KR | | CN-KR | | GC-IN | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| i-th | Optimal | PBC | Optimal | PBC | Optimal | PBC | Optimal | PBC | Optimal | PBC | Optimal | PBC |
| 1 | 32.5478 | 36.5067 | 145.3476 | 127.8643 | 46.5033 | 42.1183 | 65.0055 | 62.20497927 | 44.3011 | 47.0583 | 16.4850 | 18.8291 |
| 2 | 35.7763 | 31.7301 | 147.6449 | 137.1256 | 41.3045 | 45.5013 | 42.9145 | 62.41770662 | 29.2580 | 44.8289 | 20.2155 | 22.4971 |
| 3 | 22.7418 | 29.7407 | 140.7814 | 128.3576 | 44.1350 | 46.5874 | 39.7258 | 58.33639014 | 27.0530 | 39.7376 | 23.2204 | 20.3431 |
| 4 | 25.9186 | 30.2971 | 125.7944 | 120.2853 | 44.6124 | 43.9356 | 56.2336 | 53.83839606 | 24.3992 | 34.1077 | 17.5725 | 11.9638 |
| 5 | 19.9636 | 22.6009 | 112.5944 | 109.4592 | 41.4143 | 39.4719 | 53.0214 | 49.30022042 | 18.9793 | 24.7103 | 13.4711 | 12.9844 |
| 6 | 18.8607 | 21.7372 | 111.8627 | 106.0765 | 39.4766 | 36.0276 | 51.6784 | 50.49322435 | 25.0712 | 25.5660 | 14.6564 | 15.4080 |
| 7 | 17.4364 | 20.0092 | 92.4307 | 101.9811 | 35.7151 | 42.1977 | 47.5108 | 51.42030112 | 23.3899 | 24.9883 | 12.5053 | 5.4970 |
| 8 | 13.8959 | 18.7645 | 88.4141 | 93.7755 | 30.1222 | 29.8549 | 41.4261 | 46.85127183 | 23.0015 | 24.0016 | 13.7376 | 12.8771 |
| 9 | 10.3441 | 17.8685 | 95.3231 | 87.9422 | 28.9459 | 38.1063 | 34.5597 | 36.99375302 | 21.2671 | 23.8471 | 13.7627 | 13.6315 |
| 10 | 8.3379 | 15.3141 | 89.1990 | 88.5704 | 31.0995 | 29.7431 | 32.1944 | 37.69008142 | 20.6133 | 24.8047 | 10.8243 | 11.3219 |
| 11 | 6.0078 | 15.9365 | 78.9725 | 81.3566 | 30.7172 | 32.1682 | 31.5105 | 33.72839263 | 17.3458 | 8.8731 | 14.9684 | 13.6227 |
| 12 | 10.9589 | 18.6314 | 77.0903 | 85.8963 | 28.6952 | 28.0154 | 28.1419 | 28.62487736 | 20.7127 | 25.3920 | 9.2026 | 8.3994 |
| 13 | 10.5132 | 19.6849 | 83.3633 | 83.5208 | 25.0835 | 23.5345 | 29.1997 | 29.32851357 | 19.3283 | 22.5462 | 9.2317 | 8.3320 |
| 14 | 11.2076 | 19.7248 | 86.2056 | 90.8473 | 22.3477 | 24.0221 | 25.8333 | 27.96557729 | 19.0241 | 21.5661 | 6.3845 | 7.6903 |
| 15 | 14.5426 | 21.6547 | 87.7718 | 92.2559 | 23.1805 | 27.1944 | 28.9648 | 30.53811013 | 17.7349 | 22.0549 | 6.7164 | 6.7602 |
| 16 | 11.6888 | 18.3250 | 83.6694 | 87.3478 | 23.4836 | 23.7506 | 31.9091 | 33.97936778 | 20.9046 | 22.9522 | 6.7662 | 10.1535 |
| 17 | 11.1356 | 17.0815 | 85.8010 | 91.4531 | 21.0377 | 20.6103 | 34.2462 | 27.53926506 | 22.3009 | 23.6730 | 6.6877 | 6.9340 |
| 18 | 11.0619 | 13.2579 | 84.8946 | 92.8686 | 23.3346 | 31.3235 | 28.2141 | 34.85636454 | 12.3194 | 19.5005 | 5.5444 | 8.7512 |
| 19 | 16.8519 | 16.8924 | 89.2023 | 94.6650 | 21.4499 | 25.0740 | 26.7344 | 27.43113769 | 23.2994 | 22.1408 | 5.0177 | 1.8664 |
| 20 | 15.4588 | 4.0366 | 81.4125 | 90.9421 | 20.7344 | 27.5979 | 26.0545 | 25.03409741 | 17.1899 | 17.1293 | 3.6991 | 2.9613 |
| 21 | 21.4077 | 28.2510 | 97.7245 | 95.0101 | 24.0061 | 23.5580 | 29.4505 | 32.08551197 | 15.6752 | 15.3363 | 6.2752 | 8.0974 |

Table 5.4.7: The loglikelihood of the optimal parametric copula and the optimal penalized Bernstein copula in the subsample.

## 5.4.3 Bivariate option pricing

In this section, the bivariate option will be evaluated by Monte Carlo simulation. The risk-free rate is set as the average of the 10-year treasure rate of the United States between the years 2020 and 2021, which is 2%. The unit risk premium is set as 5%. As the initial asset prices need to be close to let the option make sense, we assumed $S_{i,0}$ is normalized to unity(Zhang and Guegan, 2008). The maturity

is assumed to be 1 month (20 trading days) $T = 20$ and the strike price is set at a level between 0.9 to 1.1. The Monte Carlo simulation is based on the following procedures, the number of iterations $N = 10,000$:

1. Generate the uniform distributed random number $u_{1,t}, u_{2,t}$ from the chosen copula. If the chosen copula is from the parametric copula family, we use the inverse method. For the Bernstein copula estimator, we use the procedure states in section 3.3.1. In the static scenario, the random numbers generation process is utilizing the parameters estimated in Table 5.4.4. In the dynamic scenario, the random number is generated from the parameters estimated in each rolling window.

2. Set the initial conditional volatility is $\hat{\sigma}_{i,0}^2 = \hat{\alpha}_{i,0}/(1 - \hat{\alpha}_{i,1} - \hat{\beta}_i - 0.5\hat{\eta}_i)$. Based on the estimated parameter of the GJR-GARCH model and the function in theorem 5.2.2, using the generated $G_{\hat{\nu}_i}^{-1}(u_{i,t})$ to replace $\varepsilon_{i,t}$ in the formula and then each $\hat{\sigma}_{i,t}^2$ and $\hat{r}_{i,t}$ can be obtained through

$$\hat{\sigma}_{i,t}^2 = \hat{\alpha}_{i,0} + \hat{\alpha}_{i,1}(G_{\hat{\nu}}^{-1}(u_{i,t})\hat{\sigma}_{i,t-1})^2 + \hat{\beta}_i\hat{\sigma}_{i,t-1}^2 + \hat{\eta}_i(G_{\hat{\nu}}^{-1}(u_{i,t}))\hat{\sigma}_{i,t-1})^2 I_{i,t-1}$$

and,

$$\hat{r}_{i,t} = r_f + \theta\hat{\sigma}_{i,t}^2 + \hat{\sigma}_{i,t}G_{\hat{\nu}}^{-1}(u_{i,t})$$

3. Get $\hat{S}_{i,T}^j$ for each iteration j, j=1,....,N and i=1,2 by take the exponential of the sum of $r_{i,t}$ through t, which is

$$\hat{S}_{i,T}^j = S_{i,0} \exp\left(\sum_{t=1}^{T} \hat{r}_{i,t}\right)$$

4. The call on max option price can be found by the discounted average of the

max.

$$RO_{i,0} \approx e^{-T \times r_f} \times \frac{1}{N} \sum_{j=1}^{N} \left[ \max(\max(\hat{S}_{1,T}^j, \hat{S}_{2,T}^j) - K, 0) \right]$$

The subscript j refers to the j-th out of a total of the Monte Carlo simulated path.



(a) US-UK        (b) UK-DE        (c) DE-GC

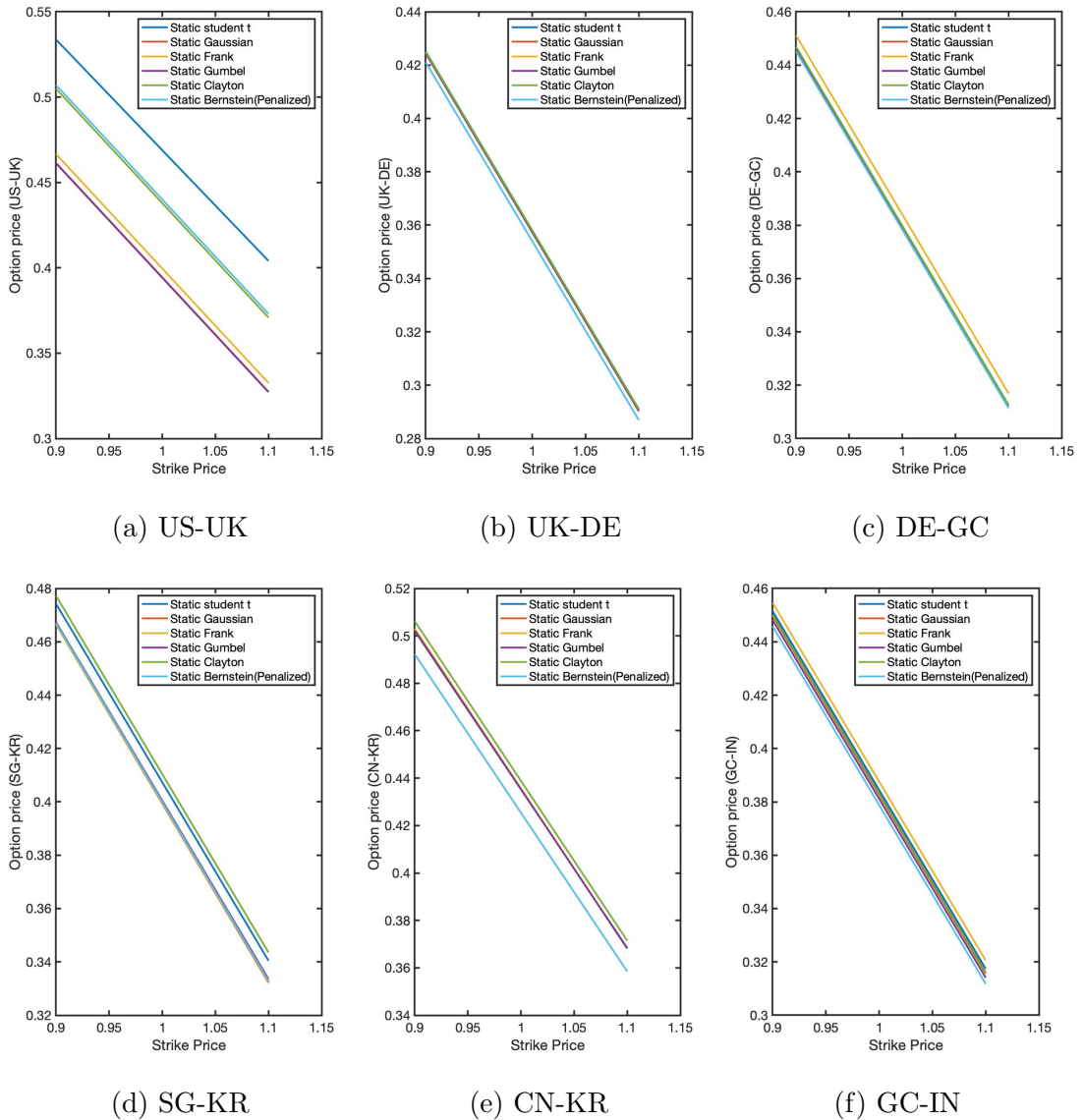(d) SG-KR        (e) CN-KR        (f) GC-IN

Figure 5.4.4: 1-month call-on-max European option prices in basis points as a function of the strike under static dependence.

The results for option pricing are illustrated here for comparison. The first pairs of figures show the call-on-max option prices as a function of the strike in the static

scenario. For high correlated pairs (UK-DE, DE-GC, SG-KR), the discrepancy between different models is not significant. For the pairs with smaller Kendall's tau, the nonparametric copula model has different option prices from the other parametric models. The Clayton copula with lowest loglikelihood and highest AIC value in the estimations of pair SG-KR and CN-KR provides higher option prices than the others. The Frank copula with lowest loglikelihood and highest AIC value in the estimations of pair DE-GC provides higher option prices than the others.
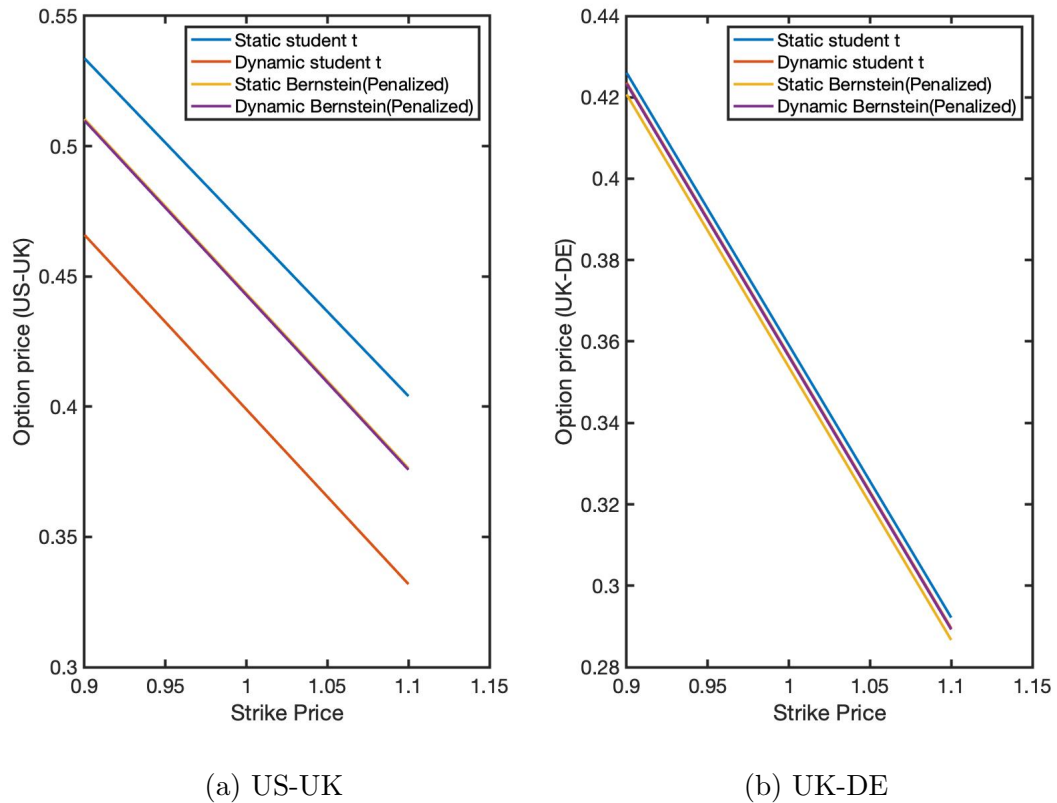


(a) US-UK $\qquad$ (b) UK-DE

Figure 5.4.5: 1-month call-on-max European option prices in basis points as a function of the strike under dynamic dependence.

Figure 5.4.5 shows the option prices in a dynamic scenario. We select one with the most concentrated (UK-DE )and one with the most discrepant (US-UK) option price lines in the static case. It shows consistent results with Zhang and Guegan (2008) that the dynamic student t copula model gives lower option prices than the

static one. For the nonparametric penalized Bernstein copula model, the dynamic one provides similar results to the static one.

## 5.5 Conclusion

In this chapter, a new approach for option price based on GJR-GARCH combined with penalized copula model has been introduced. The risk neutral process with non-normal innovation based on Duan (1999) has been derived with explicit formula which facilitate the Monte Carlo simulation. The introduce of GJR-GARCH-GED model on each asset allows most of the features from financial datasets to be captured, including conditional leptokurtosis, asymmetry and heavy tails. From the empirical results, the importance of selecting a proper copula is verified in both static and dynamic estimation, especially when the corresponding assets show low dependency.

# CHAPTER 6

## Limitation and Future Extensions

In this thesis, we propose a nonparametric Bernstein copula approach to accurately model the dependence structures among data. The sieve maximum likelihood estimation (SMLE) technique is employed to estimate the parameters within the Bernstein polynomial sieve space. To enhance prediction accuracy and introduce sparsity, we incorporate the adaptive LASSO method. Through extensive simulation studies, we demonstrate the superior performance of our model compared to parametric copula models, across both sparse and less sparse scenarios. As a future extension, we suggest exploring the application of penalized nonparametric Bernstein copula in higher dimensions, such as five dimensions. However, it should be noted that Diers et al. (2012) has shown that grid-type copula approaches outperform the Bernstein copula estimator in six-dimensional datasets. To address higher dimensions, additional techniques such as pair copula estimation can be employed. Moreover, studying the asymptotic behavior of the penalized maximum likelihood estimator with adapted LASSO can be a valuable direction, guided by the work of Zou and Hastie (2005).

While nonparametric estimation provides flexibility in density estimation, certain data types, like claim frequency and claim severity in non-life insurance, often adhere to specific parametric distributions. When accurately specified, estimating density based on the corresponding parametric model can yield accurate and

efficient results. In this study, we propose a three-step double selection procedure to enhance univariate density estimation. By conducting extensive simulations using both cross-sectional and time series data, we provide compelling evidence of the significant improvement achieved through our methodology. However, the theoretical foundations of the double selection procedure have not been explored in this research, presenting an opportunity for future investigation. Additionally, an interesting extension of our proposed methods is their application in higher dimensions, which warrants further exploration.

With the increasing complexity of financial products, the copula has emerged as a popular tool in asset pricing. We propose a nonparametric copula-based GJR-GARCH model with GED innovation for pricing European bivariate rainbow options. Our model captures residual co-movement and addresses limitations of the normality assumption. Empirical comparison with parametric copula models demonstrates the importance of selecting suitable copulas for accurate option pricing across static and dynamic scenarios under different measures. An extension to this work is to study the time variation in conditional correlation. Previous research has demonstrated that certain parametric copula models have a one-to-one relationship between their dependence parameters and Kendall's tau. Van den Goorbergh et al. (2005) specifically reveals that Kendall's tau can be expressed as a function of conditional variances over time. By estimating this function, the dependence parameter can be adjusted based on changes in Kendall's tau, offering a means to incorporate time-varying dependence in the model. Zhang and Guegan (2008) extend this relationship to t copula which has two dependence parameters. The relationship between Kendall's tau and Bernstein copula has been discussed by Sancetta and Satchell (2004), so the dynamic option pricing model based on nonparametric Bernstein copula can be further developed.

# Bibliography

Abbara, O. and Zevallos, M. (2014), 'Assessing stock market dependence and contagion', *Quantitative Finance* **14**(9), 1627–1641.

Ahmed, S. E., Ahmed, F. and Yüzbaşı, B. (2023), *Post-Shrinkage Strategies in Statistical and Machine Learning for High Dimensional Data*, CRC Press.

Ai, J., Brockett, P. L. and Wang, T. (2017), 'Optimal enterprise risk management and decision making with shared and dependent risks', *Journal of Risk and Insurance* **84**(4), 1127–1169.

Amsler, C., Prokhorov, A. and Schmidt, P. (2014), 'Using copulas to model time dependence in stochastic frontier models', *Econometric Reviews* **33**(5-6), 497–522.

Andersen, T. G., Bollerslev, T., Christoffersen, P. F. and Diebold, F. X. (2006), 'Volatility and correlation forecasting', *Handbook of economic forecasting* **1**, 777–878.

Anderson, E., Prokhorov, A. and Zhu, Y. (2020), 'A simple estimator of two-dimensional copulas, with applications 1', *Oxford Bulletin of Economics and Statistics* **82**(6), 1375–1412.

Anderson, T. W. and Darling, D. A. (1952), 'Asymptotic theory of certain" goodness of fit" criteria based on stochastic processes', *The annals of mathematical statistics* pp. 193–212.

Antoniak, C. E. (1974), 'Mixtures of dirichlet processes with applications to bayesian nonparametric problems', *The annals of statistics* pp. 1152–1174.

Bae, K.-H., Karolyi, G. A. and Stulz, R. M. (2003), 'A new approach to measuring financial contagion', *The Review of Financial Studies* **16**(3), 717–763.

Banerjee, O., Ghaoui, L. E. and d'Aspremont, A. (2008), 'Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data', *Journal of Machine learning research* **9**(Mar), 485–516.

Barman, S., Bhattacharyya, A. and Ghoshal, S. (2018), Testing sparsity over known and unknown bases, *in* 'International Conference on Machine Learning', PMLR, pp. 491–500.

Barndorff-Nielsen, O. E. (1997), 'Normal inverse gaussian distributions and stochastic volatility modelling', *Scandinavian Journal of statistics* **24**(1), 1–13.

Barron, A., Birgé, L. and Massart, P. (1999), 'Risk bounds for model selection via penalization', *Probability theory and related fields* **113**(3), 301–413.

Bauwens, L., Laurent, S. and Rombouts, J. V. (2006), 'Multivariate garch models: a survey', *Journal of applied econometrics* **21**(1), 79–109.

Belloni, A., Chernozhukov, V. and Hansen, C. (2014), 'Inference on treatment effects after selection among high-dimensional controls', *Review of Economic Studies* **81**(2), 608–650.

Belloni, A., Chernozhukov, V. and Wei, Y. (2016), 'Post-selection inference for generalized linear models with many controls', *Journal of Business & Economic Statistics* **34**(4), 606–619.

Biau, G. and Devroye, L. (2005), 'Density estimation by the penalized combinatorial method', *Journal of Multivariate Analysis* **94**(1), 196–208.

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of econometrics* **31**(3), 307–327.

Bollerslev, T. (1987), 'A conditionally heteroskedastic time series model for speculative prices and rates of return', *The review of economics and statistics* pp. 542–547.

Bollerslev, T. (1990), 'Modelling the coherence in short-run nominal exchange rates: a multivariate generalized arch model', *The review of economics and statistics* pp. 498–505.

Boubaker, H. and Sghaier, N. (2013), 'Portfolio optimization in the presence of dependent financial returns with long memory: A copula based approach', *Journal of Banking & Finance* **37**(2), 361–377.

Bouezmarni, T. and Rolin, J. (2007), 'Bernstein estimator for unbounded density function', *Nonparametric Statistics* **19**(3), 145–161.

Bouezmarni, T. and Rolin, J.-M. (2003), 'Consistency of the beta kernel density function estimator', *The Canadian Journal of Statistics/La Revue Canadienne de Statistique* pp. 89–98.

Bouezmarni, T. and Rombouts, J. V. (2010), 'Nonparametric density estimation for multivariate bounded data', *Journal of Statistical Planning and Inference* **140**(1), 139–152.

Bouezmarni, T., Rombouts, J. V. and Taamouti, A. (2010), 'Asymptotic properties of the bernstein density copula estimator for $\alpha$-mixing data', *Journal of Multivariate Analysis* **101**(1), 1–10.

Bühlmann, P. and Van De Geer, S. (2011), *Statistics for high-dimensional data: methods, theory and applications*, Springer Science & Business Media.

Bunea, F., Tsybakov, A. B., Wegkamp, M. H., Barbu, A. et al. (2010), 'Spades and mixture models', *The Annals of Statistics* **38**(4), 2525–2558.

Burda, M. and Prokhorov, A. (2014), 'Copula based factorization in bayesian multivariate infinite mixture models', *Journal of Multivariate Analysis* **127**, 200–213.

Calvet, L. E., Fisher, A. J. and Thompson, S. B. (2006), 'Volatility comovement: a multifrequency approach', *Journal of econometrics* **131**(1-2), 179–215.

Camara, A. (2003), 'A generalization of the brennan-rubinstein approach for the pricing of derivatives', *The Journal of Finance* **58**(2), 805–819.

Candes, E., Tao, T. et al. (2007), 'The dantzig selector: Statistical estimation when p is much larger than n', *The annals of Statistics* **35**(6), 2313–2351.

Casella, G., Robert, C. P., Wells, M. T. et al. (2004), Generalized accept-reject sampling schemes, *in* 'A Festschrift for Herman Rubin', Institute of Mathematical Statistics, pp. 342–347.

Charpentier, A., Fermanian, J.-D. and Scaillet, O. (2007), 'The estimation of copulas: Theory and practice', *Copulas: From theory to application in finance* pp. 35–64.

Chen, J. and Kalbfleisch, J. (1996), 'Penalized minimum-distance estimates in finite mixture models', *Canadian Journal of Statistics* **24**(2), 167–175.

Chen, S. X. (1999), 'Beta kernel estimators for density functions', *Computational Statistics & Data Analysis* **31**(2), 131–145.

Chen, S. X. and Huang, T.-M. (2007), 'Nonparametric estimation of copula functions for dependence modelling', *Canadian Journal of Statistics* **35**(2), 265–282.

Chen, X. (2007), 'Large sample sieve estimation of semi-nonparametric models', *Handbook of econometrics* **6**, 5549–5632.

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

Chen, X., Fan, Y. and Tsyrennikov, V. (2006), 'Efficient estimation of semiparametric multivariate copula models', *Journal of the American Statistical Association* **101**(475), 1228–1240.

Chen, X., Wu, W. B. and Yi, Y. (2009), 'Efficient estimation of copula-based semiparametric markov models', *The annals of statistics* **37**(6B), 4214–4253.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W. and Robins, J. (2018), 'Double/debiased machine learning for treatment and structural parameters'.

Cherubini, U. and Luciano, E. (2002), 'Bivariate option pricing with copulas', *Applied Mathematical Finance* **9**(2), 69–85.

Cherubini, U., Luciano, E. and Vecchiato, W. (2004), *Copula methods in finance*, John Wiley & Sons.

Choe, G. H. and Jang, H. J. (2011), 'Efficient algorithms for basket default swap pricing with multivariate archimedean copulas', *Insurance: Mathematics and Economics* **48**(2), 205–213.

Christoffersen, P., Dorion, C., Jacobs, K. and Wang, Y. (2010), 'Volatility components, affine restrictions, and nonnormal innovations', *Journal of Business & Economic Statistics* **28**(4), 483–502.

Christoffersen, P. F. (1998), 'Evaluating interval forecasts', *International economic review* pp. 841–862.

Christoffersen, P., Heston, S. and Jacobs, K. (2006), 'Option valuation with conditional skewness', *Journal of Econometrics* **131**(1-2), 253–284.

Curry, H. B. and Schoenberg, I. J. (1966), 'On pólya frequency functions iv: the fundamental spline functions and their limits', *Journal d'analyse mathématique* **17**(1), 71–107.

Cutler, A. and Cordero-Brana, O. I. (1996), 'Minimum hellinger distance estimation for finite mixture models', *Journal of the American Statistical Association* **91**(436), 1716–1723.

Czado, C., Kastenmeier, R., Brechmann, E. C. and Min, A. (2012), 'A mixed copula model for insurance claims and claim sizes', *Scandinavian Actuarial Journal* **2012**(4), 278–305.

Dalla Valle, L., Leisen, F. and Rossini, L. (2018), 'Bayesian non-parametric conditional copula estimation of twin data', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **67**(3), 523–548.

De Boor, C., De Boor, C., Mathématicien, E.-U., De Boor, C. and De Boor, C. (1978), *A practical guide to splines*, Vol. 27, springer-verlag New York.

Deheuvels, P. (1979), 'La fonction de dependence empirique et ses proprietes, un test non parametrique d'independance', *Bulletin de la classe des sciences, Academie Royale de Belgique, 5e serie* **65**, 274–292.

Deheuvels, P. (1980), Non parametric tests of independence, *in* 'Statistique non paramétrique asymptotique', Springer, pp. 95–107.

Demarta, S. and McNeil, A. J. (2005), 'The t copula and related copulas', *International statistical review* **73**(1), 111–129.

Devroye, L., Györfi, L. and Lugosi, G. (2013), *A probabilistic theory of pattern recognition*, Vol. 31, Springer Science & Business Media.

Diers, D., Eling, M. and Marek, S. D. (2012), 'Dependence modeling in non-life insurance using the bernstein copula', *Insurance: Mathematics and Economics* **50**(3), 430–436.

Dimitrova, D. S., Kaishev, V. K. and Penev, S. I. (2008), 'Ged spline estimation of multivariate archimedean copulas', *Computational Statistics & Data Analysis* **52**(7), 3570–3582.

Dou, X., Kuriki, S., Lin, G. D. and Richards, D. (2016), 'Em algorithms for estimating the bernstein copula', *Computational Statistics & Data Analysis* **93**, 228–245.

Duan, J.-C. (1995), 'The garch option pricing model', *Mathematical finance* **5**(1), 13–32.

Duan, J.-C. (1999), 'Conditionally fat-tailed distributions and the volatility smile in options', *Rotman School of Management, University of Toronto, Working Paper* .

Eddie, A., Artem, P. and Yajing, Z. (2018), 'A simple estimator of two-dimensional copulas, with applications', *Submitted to Oxford Bulletin of Economics and Statistics* .

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R. et al. (2004), 'Least angle regression', *The Annals of statistics* **32**(2), 407–499.

Eling, M. and Toplek, D. (2009), 'Modeling and management of nonlinear dependencies–copulas in dynamic financial analysis', *Journal of Risk and Insurance* **76**(3), 651–681.

Embrechts, P., McNeil, A. and Straumann, D. (2002), 'Correlation and dependence in risk management: properties and pitfalls', *Risk management: value at risk and beyond* **1**, 176–223.

Embrechts, P., Mcneil, E. and Straumann, D. (1999), Correlation: pitfalls and alternatives, *in* 'Risk Magazine', Citeseer.

Evans, T. and McMillan, D. G. (2009), 'Financial co-movement and correlation: evidence from 33 international stock market indices', *International Journal of Banking, Accounting and Finance* **1**(3), 215–241.

Fan, J. and Li, R. (2001), 'Variable selection via nonconcave penalized likelihood and its oracle properties', *Journal of the American statistical Association* **96**(456), 1348–1360.

Fan, J. and Peng, H. (2004), 'Nonconcave penalized likelihood with a diverging number of parameters'.

Ferguson, T. S. (1973), 'A bayesian analysis of some nonparametric problems', *The annals of statistics* pp. 209–230.

Fermanian, J.-D., Radulovic, D., Wegkamp, M. et al. (2004), 'Weak convergence of empirical copula processes', *Bernoulli* **10**(5), 847–860.

Forbes, K. J. and Rigobon, R. (2002), 'No contagion, only interdependence: measuring stock market comovements', *The journal of Finance* **57**(5), 2223–2261.

Franke, J., Härdle, W. K. and Hafner, C. M. (2004), *Statistics of financial markets*, Vol. 2, Springer.

Gaenssler, P. and Stute, W. (1987), *Seminar on empirical processes, DMV seminar*, Vol. 9(Birkhäuser Verlag, Basel).

Gawronski, W. and Stadtmüller, U. (1981), 'Smoothing histograms by means of lattice-and continuous distributions', *Metrika* **28**(1), 155–164.

Geenens, G., Charpentier, A., Paindaveine, D. et al. (2017), 'Probit transformation for nonparametric kernel estimation of the copula density', *Bernoulli* **23**(3), 1848–1873.

Geman, S. and Hwang, C.-R. (1982), 'Nonparametric maximum likelihood estimation by the method of sieves', *The Annals of Statistics* pp. 401–414.

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

Ghosal, S. et al. (2001), 'Convergence rates for density estimation with bernstein polynomials', *The Annals of Statistics* **29**(5), 1264–1280.

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

Glasserman, P., Heidelberger, P. and Shahabuddin, P. (2002), 'Portfolio value-at-risk with heavy-tailed risk factors', *Mathematical Finance* **12**(3), 239–269.

Glosten, L. R., Jagannathan, R. and Runkle, D. E. (1993), 'On the relation between the expected value and the volatility of the nominal excess return on stocks', *The journal of finance* **48**(5), 1779–1801.

Grenander, U. (1981), Abstract inference, Technical report.

Gschlößl, S. and Czado, C. (2007), 'Spatial modelling of claim frequency and claim size in non-life insurance', *Scandinavian Actuarial Journal* **2007**(3), 202–225.

Guegan, D. and Zang, J. (2013), Pricing bivariate option under garch-gh model with dynamic copula: application for chinese market, *in* 'Copulae and Multivariate Probability Distributions in Finance', Routledge, pp. 181–200.

Guo, N., Wang, F. and Yang, J. (2017), 'Remarks on composite bernstein copula and its application to credit risk analysis', *Insurance: Mathematics and Economics* **77**, 38–48.

Hansen, P. R. and Lunde, A. (2005), 'A forecast comparison of volatility models: does anything beat a garch (1, 1)?', *Journal of applied econometrics* **20**(7), 873–889.

Härdle, W. (1990), *Applied nonparametric regression*, number 19, Cambridge university press.

Harrison, J. M. and Kreps, D. M. (1979), 'Martingales and arbitrage in multiperiod securities markets', *Journal of Economic theory* **20**(3), 381–408.

Harrison, J. M. and Pliska, S. R. (1981), 'Martingales and stochastic integrals in the theory of continuous trading', *Stochastic processes and their applications* **11**(3), 215–260.

Hemche, O., Jawadi, F., Maliki, S. B. and Cheffou, A. I. (2016), 'On the study of contagion in the context of the subprime crisis: A dynamic conditional correlation–multivariate garch approach', *Economic Modelling* **52**, 292–299.

Heston, S. L. (1993), 'Invisible parameters in option prices', *The Journal of Finance* **48**(3), 933–947.

Hirukawa, M., Liu, D., Murtazashvili, I. and Prokhorov, A. (2023), 'Ds-heck: double-lasso estimation of heckman selection model', *Empirical Economics* pp. 1–29.

Hirukawa, M., Murtazashvili, I. and Prokhorov, A. (2022), 'Uniform convergence rates for nonparametric estimators smoothed by the beta kernel', *Scandinavian Journal of Statistics* **49**(3), 1353–1382.

Hu, L. (2006), 'Dependence patterns across financial markets: a mixed copula approach', *Applied financial economics* **16**(10), 717–729.

Hull, J. C. and White, A. D. (2006), 'Valuing credit derivatives using an implied copula approach', *The Journal of Derivatives* **14**(2), 8–28.

Hurd, M., Salmon, M. and Schleicher, C. (2007), 'Using copulas to construct bivariate foreign exchange distributions with an application to the sterling exchange rate index'.

Ibragimov, R. (2009), 'Copula-based characterizations for higher order markov processes', *Econometric Theory* **25**(3), 819–846.

Ibragimov, R. and Prokhorov, A. (2017), *Heavy tails and copulas: topics in dependence modelling in economics and finance*, World Scientific.

Ivan, M., Panchenko, V. and Prokhorov, A. (2021), 'Efficient estimation of parameters in marginals in semiparametric multivariate models', *RR, Journal of Computational and Graphical Statistics* .

James, G. M., Radchenko, P. and Lv, J. (2009), 'Dasso: connections between the dantzig selector and lasso', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71**(1), 127–142.

James, L. F., Priebe, C. E. and Marchette, D. J. (2001), 'Consistent estimation of mixture complexity', *Annals of Statistics* pp. 1281–1296.

Janssen, P., Swanepoel, J. and Veraverbeke, N. (2012), 'Large sample behavior of the bernstein copula estimator', *Journal of Statistical Planning and Inference* **142**(5), 1189–1197.

Janssen, P., Swanepoel, J. and Veraverbeke, N. (2014), 'A note on the asymptotic behavior of the bernstein estimator of the copula density', *Journal of Multivariate Analysis* **124**, 480–487.

Jenatton, R., Audibert, J.-Y. and Bach, F. (2011), 'Structured variable selection with sparsity-inducing norms', *Journal of Machine Learning Research* **12**(Oct), 2777–2824.

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

Joe, H. and Xu, J. J. (1996), 'The estimation method of inference functions for margins for multivariate models'.

Judd, K. L. (1998), *Numerical methods in economics*, MIT press.

Kascha, C. (2012), 'A comparison of estimation methods for vector autoregressive moving-average models', *Econometric Reviews* **31**(3), 297–324.

Kauermann, G., Schellhase, C. and Ruppert, D. (2013), 'Flexible copula density estimation with penalized hierarchical b-splines', *Scandinavian Journal of Statistics* **40**(4), 685–705.

Kim, G., Silvapulle, M. J. and Silvapulle, P. (2007), 'Comparison of semiparametric and parametric methods for estimating copulas', *Computational Statistics & Data Analysis* **51**(6), 2836–2850.

Knight, K., Fu, W. et al. (2000), 'Asymptotics for lasso-type estimators', *The Annals of statistics* **28**(5), 1356–1378.

Kole, E., Koedijk, K. and Verbeek, M. (2007), 'Selecting copulas for risk management', *Journal of Banking & Finance* **31**(8), 2405–2423.

Kupiec, P. H. et al. (1995), *Techniques for verifying the accuracy of risk measurement models*, Vol. 95, Division of Research and Statistics, Division of Monetary Affairs, Federal . . . .

Lambert, P. (2007), 'Archimedean copula estimation using bayesian splines smoothing techniques', *Computational Statistics & Data Analysis* **51**(12), 6307–6320.

Leroux, B. G. (1992), 'Consistent estimation of a mixing distribution', *The Annals of Statistics* pp. 1350–1360.

Li, X., Mikusiński, P., Sherwood, H. and Taylor, M. (1997), On approximation of copulas, *in* 'Distributions with given marginals and moment problems', Springer, pp. 107–116.

Lijoi, A., Prünster, I. and Walker, S. G. (2005), 'On consistency of nonparametric normal mixtures for bayesian density estimation', *Journal of the American Statistical Association* **100**(472), 1292–1296.

Liu, D. and Prokhorov, A. (2016), 'Sparse sieve mle'.

Lorentz, G. G. (2013), *Bernstein polynomials*, American Mathematical Soc.

Malevergne, Y. and Sornette, D. (2003), 'Testing the gaussian copula hypothesis for financial assets dependences', *Quantitative finance* **3**(4), 231.

Margrabe, W. (1978), 'The value of an option to exchange one asset for another', *The journal of finance* **33**(1), 177–186.

Marron, J. S. and Ruppert, D. (1994), 'Transformations to reduce boundary bias in kernel density estimation', *Journal of the Royal Statistical Society: Series B (Methodological)* **56**(4), 653–671.

McNeil, A. J., Frey, R. and Embrechts, P. (2015), *Quantitative Risk Management: Concepts, Techniques and Tools-revised edition*, Princeton university press.

Meinshausen, N., Bühlmann, P. et al. (2006), 'High-dimensional graphs and variable selection with the lasso', *The annals of statistics* **34**(3), 1436–1462.

Merton, R. C. (1973), 'Theory of rational option pricing', *The Bell Journal of economics and management science* pp. 141–183.

Murphy, K. P. (2012), *Machine learning: a probabilistic perspective*, MIT press.

Nagler, T. (2014), 'Kernel methods for vine copula estimation'.

Naifar, N. (2011), 'Modelling dependence structure with archimedean copulas and applications to the itraxx cds index', *Journal of Computational and Applied Mathematics* **235**(8), 2459–2466.

Nelson, D. B. (1991), 'Conditional heteroskedasticity in asset returns: A new approach', *Econometrica: Journal of the econometric society* pp. 347–370.

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

Osborne, M. R., Presnell, B. and Turlach, B. A. (2000), 'A new approach to variable selection in least squares problems', *IMA journal of numerical analysis* **20**(3), 389–403.

Patton, A. J. (2002), *Applications of copula theory in financial econometrics*, University of California, San Diego.

Patton, A. J. (2006), 'Modelling asymmetric exchange rate dependence', *International economic review* **47**(2), 527–556.

Patton, A. J. (2009), Copula–based models for financial time series, *in* 'Handbook of financial time series', Springer, pp. 767–785.

Pedersen, R. S. and Rahbek, A. (2014), 'Multivariate variance targeting in the bekk–garch model', *The Econometrics Journal* **17**(1), 24–55.

Petrone, S. (1999), 'Bayesian density estimation using bernstein polynomials', *Canadian Journal of Statistics* **27**(1), 105–126.

Petrone, S. and Wasserman, L. (2002), 'Consistency of bernstein polynomial posteriors', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**(1), 79–100.

Pfeifer, D., Strassburger, D. and Philipps, J. (2020), 'Modelling and simulation of dependence structures in nonlife insurance with bernstein copulas', *arXiv preprint arXiv:2010.15709* .

Poon, S.-H., Rockinger, M. and Tawn, J. (2004), 'Extreme value dependence in financial markets: Diagnostics, models, and financial implications', *The Review of Financial Studies* **17**(2), 581–610.

Prokhorov, A. and Schmidt, P. (2009*a*), 'Gmm redundancy results for general missing data problems', *Journal of Econometrics* **151**(1), 47–55.

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

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

Radchenko, P. and James, G. M. (2008), 'Variable inclusion and shrinkage algorithms', *Journal of the American Statistical Association* **103**(483), 1304–1315.

Radchenko, P. and James, G. M. (2010), 'Variable selection using adaptive nonlinear interaction structures in high dimensions', *Journal of the American Statistical Association* **105**(492), 1541–1553.

Ray, S. and Lindsay, B. G. (2008), 'Model selection in high dimensions: a quadratic-risk-based approach', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70**(1), 95–118.

Roeder, K. and Wasserman, L. (1997), 'Practical bayesian density estimation using mixtures of normals', *Journal of the American Statistical Association* **92**(439), 894–902.

Rose, D. (2015), Modeling and estimating multivariate dependence structures with the Bernstein copula, PhD thesis, lmu.

Rosenberg, J. V. (1999), 'Semiparametric pricing of multivariate contingent claims'.

Rosenberg, J. V. (2003), 'Non-parametric pricing of multivariate contingent claims', *The Journal of Derivatives* **10**(3), 9–26.

Ruppert, D., Wand, M. P. and Carroll, R. J. (2003), *Semiparametric regression*, number 12, Cambridge university press.

Salmon, M., Schleicher, C. et al. (2006), 'Pricing multivariate currency options with copulas', *Copulas: From Theory to Application in Finance, Risk Books, London* .

Sancetta, A. and Satchell, S. (2004), 'The bernstein copula and its applications to modeling and approximations of multivariate distributions', *Econometric theory* **20**(3), 535–562.

Santos, A. A., Nogales, F. J. and Ruiz, E. (2013), 'Comparing univariate and multivariate models to forecast portfolio value-at-risk', *Journal of financial econometrics* **11**(2), 400–441.

Scaillet, O. and Fermanian, J.-D. (2002), 'Nonparametric estimation of copulas for time series', *FAME Research paper* (57).

Scheffer, M. and Weiß, G. N. (2017), 'Smooth nonparametric bernstein vine copulas', *Quantitative Finance* **17**(1), 139–156.

Scholes, M. and Black, F. (1973), 'The pricing of options and corporate liabilities', *Journal of Political Economy* **81**(3), 637–654.

Shen, X., Zhu, Y. and Song, L. (2008), 'Linear b-spline copulas with applications to nonparametric estimation of copulas', *Computational Statistics & Data Analysis* **52**(7), 3806–3819.

Shen, X. et al. (1997), 'On methods of sieves and penalization', *The Annals of Statistics* **25**(6), 2555–2591.

Sklar, M. (1959), 'Fonctions de repartition an dimensions et leurs marges', *Publ. inst. statist. univ. Paris* **8**, 229–231.

Spliid, H. (1983), 'A fast estimation method for the vector autoregressive moving average model with exogenous variables', *Journal of the American Statistical Association* **78**(384), 843–849.

Stulz, R. (1982), 'Options on the minimum or the maximum of two risky assets: analysis and applications', *Journal of Financial Economics* **10**(2), 161–185.

Tavin, B. (2015), 'Detection of arbitrage in a market with multi-asset derivatives and known risk-neutral marginals', *Journal of Banking & Finance* **53**, 158–178.

Tavin, B. (2018), 'Measuring exposure to dependence risk with random bernstein copula scenarios', *European Journal of Operational Research* **270**(3), 873–888.

Tenbusch, A. (1994), 'Two-dimensional bernstein polynomial density estimators', *Metrika* **41**(1), 233–253.

Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 267–288.

Tikhonov, A. N. (1943), 'On the stability of inverse problem', *Doklady Akademii Nauk SSSR* **39**, 195–198.

Torri, G., Giacometti, R. and Paterlini, S. (2018), 'Robust and sparse banking network estimation', *European Journal of Operational Research* **270**(1), 51–65.

Tsukahara, H. (2005), 'Semiparametric estimation in copula models', *Canadian Journal of Statistics* **33**(3), 357–375.

Vaart, A. W. and Wellner, J. A. (1996), *Weak convergence and empirical processes: with applications to statistics*, Springer.

Van den Goorbergh, R. W., Genest, C. and Werker, B. J. (2005), 'Bivariate option pricing using dynamic copula models', *Insurance: Mathematics and Economics* **37**(1), 101–114.

Vitale, R. A. (1975), A bernstein polynomial approach to density function estimation, *in* 'Statistical inference and related topics', Elsevier, pp. 87–99.

Wahba, G. (1990), *Spline models for observational data*, SIAM.

Walter, G. G., Blum, J. R. et al. (1984), 'A simple solution to a nonparametric máximum likelihood estimation problem', *The Annals of Statistics* **12**(1), 372–379.

Wand, M. P. and Jones, M. C. (1994), *Kernel smoothing*, CRC press.

Wang, C.-W., Yang, S. S. and Huang, H.-C. (2015), 'Modeling multi-country mortality dependence and its application in pricing survivor index swaps—a dynamic copula approach', *Insurance: Mathematics and Economics* **63**, 30–39.

Weiß, G. (2011), 'Copula parameter estimation by maximum-likelihood and minimum-distance estimators: a simulation study', *Computational Statistics* **26**(1), 31–54.

Wen, X., Wei, Y. and Huang, D. (2012), 'Measuring contagion between energy market and stock market during financial crisis: A copula approach', *Energy economics* **34**(5), 1435–1446.

Winkelmann, R. (2012), 'Copula bivariate probit models: with an application to medical expenditures', *Health economics* **21**(12), 1444–1455.

Wong, W. H. (1992), 'On asymptotic efficiency in estimation theory', *Statistica Sinica* pp. 47–68.

Wong, W. H. and Severini, T. A. (1991), 'On maximum likelihood estimation in infinite dimensional parameter spaces', *The Annals of Statistics* pp. 603–632.

Wong, W. H. and Shen, X. (1995), 'Probability inequalities for likelihood ratios and convergence rates of sieve mles', *The Annals of Statistics* pp. 339–362.

Woo, M.-J. and Sriram, T. (2006), 'Robust estimation of mixture complexity', *Journal of the American Statistical Association* **101**(476), 1475–1486.

Wu, J., Wang, X. and Walker, S. G. (2015), 'Bayesian nonparametric estimation of a copula', *Journal of Statistical Computation and Simulation* **85**(1), 103–116.

Xie, H. and Wu, X. (2017), 'A conditional autoregressive range model with gamma distribution for financial volatility modelling', *Economic Modelling* **64**, 349–356.

Zhang, C.-H. (2010), 'Nearly unbiased variable selection under minimax concave penalty', *The Annuals of Statistics* pp. 894–942.

Zhang, C.-H., Huang, J. et al. (2008), 'The sparsity and bias of the lasso selection in high-dimensional linear regression', *The Annals of Statistics* **36**(4), 1567–1594.

Zhang, J. and Guegan, D. (2008), 'Pricing bivariate option under garch processes with time-varying copula', *Insurance: Mathematics and Economics* **42**(3), 1095–1103.

Zhao, P. and Yu, B. (2006), 'On model selection consistency of lasso', *Journal of Machine learning research* **7**(Nov), 2541–2563.

Zou, H. (2006), 'The adaptive lasso and its oracle properties', *Journal of the American Statistical Association* **101**(476), 1418–1429.

Zou, H. and Hastie, T. (2005), 'Regularization and variable selection via the elastic net', *Journal of the royal statistical society: series B (statistical methodology)* **67**(2), 301–320.

# APPENDIX A

---

# Some proofs of Bernstein polynomials

---

**Proof of theorem 3.2.3**

*Proof.*

$$\left|(B_J^k f)(x_1, x_2, ..., x_k) - f(x_1, ..., x_k)\right|$$

$$= \left|\sum_{j_1=0}^{J_1} ... \sum_{j_k=0}^{J_k} \left[f(\frac{j_1}{J_1}, ..., \frac{j_k}{J_k}) - f(x_1, ..., x_k)\right] p_{j_1}...p_{j_k}\right|$$

$$\leq \sum_{\left|\frac{j}{J_-}-x\right|<\delta} \left|f(\frac{j_1}{J_1}, ..., \frac{j_k}{J_k}) - f(x_1, ..., x_k)\right| p_{j_1}...p_{j_k} + \sum_{\left|\frac{j}{J}-x\right|\geq\delta} \left|f(\frac{j_1}{J_1}, ..., \frac{j_k}{J_k}) - f(x_1, ..., x_k)\right| p_{j_1}...p_{j_k}$$

Assume

$$T_s = \sum_{j_s=0}^{J_s}(j_s - J_s x_s)^2 p_{j_s} = \sum_{j_s=0}^{J_s}\{j_s(j_s - 1) - (2J_s x_s - 1)j_s + J_s^2 x_s^2\}p_{j_s}$$

Clearly, $\sum_{j_s=0}^{J_s} p_{j_s} = 1$, and we have

$$\sum_{j_s=0}^{J_s} j_s p_{j_s} = J_s x_s \sum_{\mu_s=0}^{J_s-1} \binom{J_s - 1}{\mu_s} x_s^\mu (1 - x_s^{J_s-1-\mu}) = J_s x_s$$

$$\sum_{j_s=0}^{J_s} j_s(j_s - 1)p_{j_s} = J_s(J_s - 1)x_s^2 \sum_{\mu_s=0}^{J_s-2} \binom{J_s - 2}{\mu_s} x_s^{\mu_s}(1 - x_s)^{J_s-2-\mu_s} = J_s(J_s - 1)x_s^2$$

157

Thus,

$$T_s = \left(J_s^2 x_s^2 - (2J_s x_s - 1)J_s x_s + J_s(J_s - 1)x_s^2\right) = J_s x_s(1 - x_s) \qquad (A.0.1)$$

Since $x_s(1 - x_s) \leq \frac{1}{4}$ on $[0, 1]$, we can obtain the inequality

$$\sum_{\left|\frac{j}{J} - x\right| \geq \delta} p_{j_1}...p_{j_k} \leq \frac{k}{\delta^2} \sum_{\left|\frac{j_s}{J_s} - x_s\right| \geq \delta} \left(\frac{j_s}{J_s} - x_s\right)^2 p_{j_s}$$

$$\leq \frac{1}{J_s^2 \delta^2} T_s = \frac{x_s(1 - x_s)}{J_s \delta^2} \leq \frac{k}{4J_s \delta^2}$$

Based on our assumptions, that $f(x_1, ..., x_k) \in C_{[0,1]^k}$, if $f$ is bounded, say $M :=$ $\max_{\mathbf{u} \in [0,1]^k|} f(u_1, ..., u_k)$, for a given $\epsilon > 0$, we can find a $\delta > 0$ such that $|x_s - x_s^*| <$ $\delta$ implies $|f(x_1, ..., x_k) - f(x_1^*, ..., x_k^*)| < \epsilon/2$. Then, we have

$$\left|(B_J^k f)(x_1, x_2, ..., x_k) - f(x_1, ..., x_k)\right| \qquad (A.0.2)$$

$$\leq \frac{\epsilon}{2} \sum_{\left|\frac{j}{J} - x\right| < \delta} p_{j_1}...p_{j_k} + \left|f(\frac{j_1}{J_1}, ..., \frac{j_k}{J_k}) - f(x_1, ..., x_k)\right| \sum_{\left|\frac{j}{J} - x\right| \geq \delta} p_{j_1}...p_{j_k} \qquad (A.0.3)$$

$$\leq \epsilon + 2M \times \frac{k}{4J_s \delta^2} \qquad (A.0.4)$$

and if $J$ is sufficient large, we have $\frac{Mk}{4J_s \delta^2} < \frac{\epsilon}{2}$ and $\left|(B_J^k f)(x_1, x_2, ..., x_k) - f(x_1, ..., x_k)\right| <$ $\epsilon$. If $f(x_1, ..., x_k$ is continuous in the whole interval $[0, 1]^k$ then A.0.4 holds with a $\delta$ independent of all $x_s$, so that $(B_J^k f)(x_1, x_2, ..., x_k) \to f(x_1, ..., x_k)$ uniformly. $\quad \square$

**Proof of theorem 3.2.4**

*Proof.*

$$(\|(B_J^k f)(x_1, x_2, ..., x_k) - f(x_1, ..., x_k)\|)^2$$

$$\leq \left( \sum_{j_1=0}^{J_1} ... \sum_{j_k=0}^{J_k} \left\| f(\frac{j_1}{J_1}, ..., \frac{j_k}{J_k}) - f(x_1, ..., x_k) \right\| p_{j_1}...p_{j_k} \right)^2$$

$$\leq \left( L \sum_{j_1=0}^{J_1} ... \sum_{j_k=0}^{J_k} \left\| \frac{\mathbf{j}}{\mathbf{J}} - \mathbf{x} \right\| p_{j_1}...p_{j_k} \right)^2$$

$$\leq L^2 \left( \sum_{j_1=0}^{J_1} ... \sum_{j_k=0}^{J_k} \left\| \frac{\mathbf{j}}{\mathbf{J}} - \mathbf{x} \right\| p_{j_1}...p_{j_k} \right) \left( \sum_{j_1=0}^{J_1} ... \sum_{j_k=0}^{J_k} p_{j_1}...p_{j_k} \right)$$

$$= L^2 \sum_{j_1=0}^{J_1} ... \sum_{j_k=0}^{J_k} \left( \sum_k^{s=1} \left( \frac{j_s}{J_s} - x_s \right)^2 \right) p_{j_1}...p_{j_k}$$

$$= L^2 \sum_k^{s=1} \frac{x_s(1-x_s)}{J_s} \leq L^2 \sum_k^{s=1} \frac{1}{4J_s}$$

$\square$

# APPENDIX B

---

## Parametric copula family

---

**Gaussian copula**

$$C(u_1, u_2; \theta) = \Phi_R(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \theta)$$

where $\Phi$ is the cumulative distribution function(cdf) of the standard normal distribution and $\Phi_R(u_1, u_2)$ is the standard bivariate normal distribution with the parameter of correlation $\theta \in (-1, 1)$.

**Student's t copula**

$$C(u_1, u_2; \theta_1, \theta_2) = \int_{-\infty}^{t_{\theta_1}^{-1}} \int_{-\infty}^{t_{\theta_2}^{-1}} \frac{1}{2\pi(1 - \theta_2^2)^{1/2}} \times \{1 + \frac{s^2 - 2\theta_2 st + t^2}{\nu(1 - \theta_2^2)}\}^{-(\theta_1+2)/2} ds dt$$

where $t_{\theta_1}^{-}1(u_1)$ is the inverse cdf of the standard univariate t-distribution with $\theta_1$ degree of freedom. So, the parameter $\theta_1$ controls the tails' heaviness. As $\theta_1 \to \infty$, the student's t copula behaves like Gaussian copula.

**Clayton copula**

$$C(u_1, u_2; \theta) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}, \ \theta \in (0, \infty)$$

when $\theta \to 0$, the margins tend to be independent. As $\theta \to \infty$, the copula attain Frechet upper bound. For Clayton copula, the lower bound cannot be attained.

It works well if the correlation between the two random variables is strongest in the left tail of joint distribution.

## Gumbel copula

The Gumbel copula is characterized by the upper tail dependence

$$C(u_1, u_2) = \exp\left(-\left[(-\ln(u_1))^\rho + (-\ln(u_2))^\rho\right]^{1/\rho}\right)$$

where $\rho \in [1, +\infty]$. When $\rho \to \infty$, the dependence structure between the two random variable is perfectly positive and when $\rho = 1$, the two random variable is independent as $C(u_1, u_2) = u_1 u_2$.

**Frank copula** The Frank copula is a symmetric Archimedean copula given by:

$$C(u_1, u_2) = -\frac{1}{\alpha} \ln\left(1 + \frac{(\exp(-\alpha u_1) - 1)(\exp(-\alpha u_2) - 1)}{\exp(-\alpha) - 1}\right)$$

where $\alpha \in \mathbb{R} \setminus \{0\}$.