A Nonparametric Bayesian Estimator of Copula Density with an Application to Option Pricing

Qiaoyu Wang* Ximing Wu†

Abstract

We propose a nonparametric Bayesian copula density estimator based on the Logistic Gaussian Process density estimation method. A Gaussian process prior with flexible mean and covariance functions are placed on the latent functions. To avoid typical boundary issues in copula density estimation, a transformation approach is adopted such that the latent process is defined on unrestricted support and subsequently back-transformed to obtain the posterior copula density. We also develop a sampler to facilitate sampling from the posterior copula distribution with ease and efficiency. Monte Carlo simulations demonstrate good overall and tail performance of the proposed estimator and the efficiency of the posterior sampler. We apply the proposed method to pricing options based on a flexible time-varying copula model of the underlying asset returns.

Keywords: copula; nonparametric Bayesian estimation; Gaussian process; option pricing

JEL codes: C11; C14; G13

*Department of Economics, Texas A&M University; email: wqy680426@tamu.edu

†Department of Agricultural Economics, Texas A&M University, College Station, TX 77843; email: xwu@tamu.edu

1 Introduction

The celebrated Sklar's theorem (Sklar, 1959) states that a d-dimensional distribution can be generally expressed as $F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d))$, where F is the joint distribution and C is the copula distribution whose arguments are the marginal distributions $F_1, \ldots, F_d$. This construction separates the marginal distributions from their dependence, which is completely captured by the copula. There has been a rapid growth in the exploration and application of copula methods; see, among others, Joe (1997) and Nelsen (2007) for book-length treatments and Charpentier et al. (2007), Patton (2009), Fan and Patton (2014), and Fermanian (2017) for overviews of recent developments in economics and finance.

Both parametric and nonparametric methods have been used in copula estimation. Most parametric copulas are parameterized by a few parameters and thus can be restricted. Nonparametric methods seek a flexible approximation based on methods such as kernel or spline estimation. These estimators, however, encounter the challenge of boundary biases since the copula density is defined on a unit hypercube. Methods such as reflection kernel, boundary adaptive kernel and domain transformation have been proposed to tackle this difficulty; see Geenens et al. (2017) and Wen and Wu (2020) and references therein.

This study contributes to this literature by proposing a novel nonparametric Bayesian copula density estimator. Leonard (1978) designed a flexible density estimator based on the Logistic transformation: $g(s) = \frac{\exp(f(s))}{\int \exp(f(s'))ds'}$. The key advantage of this approach is that one can freely model the latent function f; the subsequent Logistic transformation then yields a proper density g. We combine this approach and the domain transformation method in copula modeling to develop a copula density estimator. Define the Probit transformation $s_i = \Phi^{-1}(F_i(x_i)), i = 1, \ldots, d$ and let $\boldsymbol{s} = (s_1, \ldots, s_d)^T$. This method transforms variables from the unit hypercube $\mathcal{I}^d$ to $\mathcal{R}^d$. We then model the density of $\boldsymbol{s}$ on $\mathcal{R}^d$ and back-transform it to obtain a copula density estimate that is free of boundary bias. Specifically, we adopt a Bayesian approach, employing the Logistic transformation and placing a Gaussian Process prior on the latent function:

$$f(\boldsymbol{s}) \sim \mathcal{GP}(m(\boldsymbol{s}), K(\boldsymbol{s}, \boldsymbol{s}')),$$

where m and K are the mean and covariance functions that describe this process. The Gaussian Process is a versatile method of probabilistic learning (Rasmussen and Williams, 2006). The mean function reflects the global features of the latent function. We consider both polynomial and piecewise polynomial mean functions. The expressive covariance process, parameterized by an infinitely differentiable kernel and hyperdistributions of tuning parameters, accommodates local features. Model configuration is guided by the Bayesian Deviance Information Criterion. Since the posterior distribution is intractable, we use the Laplace approximation augmented with importance sampling for inference. We further propose a posterior sampler that permits us to sample from the posterior copula distribution with ease and efficiency.

The proposed copula density estimator enjoys several advantages. First, the Gaussian processes are appealing priors for the latent function due to their ability to specify the smoothness and differentiability properties of a stochastic path through the prior covariance structure. Upon exponentiation and back-transformation, the resulting LGP copula density can be viewed as a product of a baseline density parametrized by $m(\cdot)$ and an exponentialized zero mean Gaussian process parametrized by $K(\cdot,\cdot)$. This is in spirit close to the nonparametric density with a parametric start by Hjort and Glad (1995), in spite of the Bayesian approach taken in this study. Second, the transformation kernel density suffers from likely erratic tail behaviors. To tackle this difficulty, Geenens et al. (2017) consider local likelihood estimation and Wen and Wu (2020) employ a smooth tapering device. The LGP estimator, thanks to the intrinsic averaging over a continuum of distributions, effectively smoothes the copula density over its entire support and behaves well in the tails. Third, all aspects of model specification, including the mean and covariance functions of the Gaussian process and the hyperdistributions of tuning parameters are governed by a Bayesian information criterion. Lastly, the proposed sampler provides a cost-effective way to sample from the posterior distribution, greatly enhancing the usefulness of the LGP copula density estimator.

Our Monte Carlo simulations demonstrate the merit of the LGP copula density estimator. It generally outperforms its competitors, sometimes by substantial margins in terms of the overall goodness-of-fit and tail estimations. The efficiency of the posterior sampler is also demonstrated. We apply the proposed method to the analysis of time varying dependence of financial asset returns and option pricing. The posterior sampler proves to be particularly useful for simulation-based option pricing.

This rest of the paper is organized as follows. Section 2 briefly reviews the literature on copula and the general procedure for Logistic Gaussian Process density estimation. Section 3 presents the formulation and inference of the LGP copula density estimator and its posterior sampler. Sections 4 and 5 present Monte Carlo simulations and the empirical application to financial data analysis and option pricing. The last section concludes. Some technical details are provided in the Appendix.

2 Preliminaries

We provide in this section a brief review on copula, the subject of our investigation and the proposed method — a Gaussian Process logistic density estimator.

2.1 Copula

To ease exposition, we focus on bivariate distributions in this study. Let $\boldsymbol{x} = (x_1, x_2)^T$ be a bivariate random vector with a continuous cumulative distribution function (CDF) F. According to Sklar (1959), there exists a unique copula function $C(\cdot, \cdot) : [0,1]^2 \to [0,1]$, such that

$$F(\boldsymbol{x}) = C(F_1(x_1), F_2(x_2)), \ \ \boldsymbol{x} \in \mathcal{R}^2,$$

where F_1 and F_2 are the marginal distributions of x_1 and x_2. Sklar's theorem de-couples the marginal distributions of $\boldsymbol{x}$ from their dependence, which is completely captured by the copula. This decomposition facilitates separate modeling of marginal distributions and the dependence of random variables.

Let $\boldsymbol{u} = (u_1, u_2)^T = (F_1(x_1), F_2(x_2))^T$ be a random vector defined on the unit square $\mathcal{I}^2 = [0,1]^2$. Suppose both F_1 and F_2 are continuously differentiable, thus the copula is a cumulative distribution function with standard uniform margins. In addition, if C is absolutely continuous, it admits a probability density function (PDF), given by the derivative of C with respect to each of its arguments:

$$c(\boldsymbol{u}) = \frac{\partial^2 C}{\partial u_1 \partial u_2}(u_1, u_2), \quad \boldsymbol{u} \in \mathcal{I}^2.$$

Generally copula densities are more interpretable than copula distributions and therefore often the subject of statistical investigations.

Commonly used parametric copulas include the Gaussian, Student-t, Clayton, Frank, and Gumbel copulas. These copulas are often parameterized by one or two coefficients and may lack flexibility. For instance, the popular Gaussian copula is symmetric and has zero tail dependence. Consequently, it is inadequate for modeling financial data that tend to move in tandem under extreme market conditions, especially during market downturns. Except for the family of elliptical copulas (e.g. Gaussian and t copulas), most copulas are defined for bivariate distributions and generalizations to higher dimensions can be challenging.

There exists a number of nonparametric copula estimators in the literature. The classical kernel estimation of copula densities suffers from boundary biases as copula density is defined on a bounded support. This adversity is further exacerbated by that copula density can be unbounded at the corners. To mitigate these difficulties, boundary-bias correcting or adaptive methods such as mirror reflection (Gijbels and Mielniczuk (1990)) or beta-density-kernel (Charpentier et al. (2007) have been proposed. Bayesian methods have been considered for copula estimation as well; see, e.g., Smith (2013) and Ning and Shephard (2018) and references therein.

Transformation-kernel density estimator provides an apt solution to the boundary bias issue. Let $s_1 = \Phi^{-1}(u_1)$ and $s_2 = \Phi^{-1}(u_2)$ be the probit transformation; the induced joint density of $\boldsymbol{s} = (s_1, s_2)$ is given by

$$g(\boldsymbol{s}) = c(\Phi(s_1), \Phi(s_2))\phi(s_1)\phi(s_2), \quad \boldsymbol{s} \in \mathcal{R}^2,$$

where Φ and ϕ are the standard normal CDF and PDF. The copula density, upon back-transformation, is then given by

$$c(\boldsymbol{u}) = \frac{g(\Phi^{-1}(u_1), \Phi^{-1}(u_2))}{\phi(\Phi^{-1}(u_1))\phi(\Phi^{-1}(u_2))}, \quad \boldsymbol{u} \in \mathcal{I}^2. \tag{1}$$

Notice that the back-transformation entails a multiplicative factor $\left\{\phi(\Phi^{-1}(u_1))\phi(\Phi^{-1}(u_2))\right\}^{-1}$, which tends to infinity as u_1 and/or u_2 approaches zero or one. The resultant estimate is free of boundary bias and not bounded at the boundaries, a desirable property for an estimator of copula densities. This approach, however, suffers from a different boundary difficulty — slight variations in the tails of g might be magnified substantially due to the multiplicative factor, resulting in rugged irregularities of the copula density near the boundaries, especially at the corners of the unit square.

Geenens et al. (2017) proposed a local likelihood estimator that regularizes the tails and Wen and Wu (2020) developed a refinement that smoothly tapers the multiplicative factor in the tails. The former requires location specific bandwidths; the latter, while retaining global bandwidths, entails a couple of extra tuning parameters for the tapering device.

2.2 Logistic Gaussian Process Density Estimator

In this study, we adopt the domain transformation approach and develop a nonparametric Bayesian copula density estimator through a Logistic transformation. Our estimator is motivated by the nonparametric density estimator with a Logistic Gaussian Process (LGP) prior; see e.g. Leonard (1978), Daniel (1986), Lenk (1988) and Lenk (1991).

Let $g(\cdot) : \mathcal{X}_g \to \mathcal{R}^+$ be a density function of interest for a d-dimensional random vector $\boldsymbol{x}$. Leonard (1978) proposed a nonparametric estimator based on the following Logistic transform

$$g(\boldsymbol{x}) = \frac{\exp(f(\boldsymbol{x}))}{\int \exp(f(\boldsymbol{x}'))d\boldsymbol{x}'}, \tag{2}$$

where $f(\cdot) : \mathcal{X}_g \to \mathcal{R}$ is an unconstrained latent function. The exponentiation of f and the normalizing denominator ensure that $g(\cdot)$ is a *bona fide* density, which is positive and integrates to one on its support. A Gaussian process prior is placed on the latent function:

$$f(\boldsymbol{x}) \sim \mathcal{GP}(m(\boldsymbol{x}), K(\boldsymbol{x}, \boldsymbol{x}')),$$

where $m(\boldsymbol{x})$ is the mean function and the kernel function $K(\boldsymbol{x}, \boldsymbol{x}')$ prescribes the covariance between a pair of inputs $\boldsymbol{x}$ and $\boldsymbol{x}'$. Throughout, we use $\mathcal{GP}$ to denote an infinite dimensional Gaussian Process and $\mathcal{N}$ for a finite dimensional Gaussian distribution.

The gist of Gaussian process modeling is the covariance, which encodes the modeler's knowledge about the latent process such as smoothness, stationarity and periodicity. A commonly used covariance function is the squared exponential (SE) covariance function

$$K(\boldsymbol{x}, \boldsymbol{x}') = \sigma^2 \exp(-\frac{1}{2}\sum_{k=1}^{d} l_k^{-2}(x_k - x_k')^2), \tag{3}$$

where $\boldsymbol{x} = (x_1, \ldots, x_d)'$. Given this covariance function, a Gaussian process is governed by its

hyperparameters $\boldsymbol{\theta} = (l_1, l_2 \ldots, l_d, \sigma^2)$: the length scale l_k's control the smoothness of its path and σ^2 determines its overall variation. The hyperparameters are to be learned from data. For a full Bayesian treatment, prior (hyper) distributions for these hyperparameters are used and their posterior distribution is inferred using Bayes' rule. The posterior distribution for the latent function $f(\boldsymbol{x})$ is then derived through marginalization over the posterior distribution of hyperparameters. Finally the posterior distribution of $g(\boldsymbol{x})$ is obtained via the exponentiation and proper normalization of the posterior latent function $f(\boldsymbol{x})$.

The posterior distributions of Gaussian process models generally are not tractable, except for Gaussian observation models. Researchers often resort to deterministic approximations (such as the Laplace approximation, variational approximation or expectation propagation) or MCMC methods for inference. A practical bottleneck of Gaussian process models is its high computation cost. For a sample with n observations, this method requires inverting the covariance matrix, an $O(n^3)$ operation. The actual computation cost can be substantially higher since repeated matrix inversions are needed in the search of optimal hyperparameters or MCMC iterations. There has been a growing literature on methods to improve the scalability of Gaussian process models. The interested reader is referred to Rasmussen and Williams (2006) for a lucid exposition of theories and implementations of Gaussian process models. Theoretical properties of LGP estimators have been investigated by Tokdar and Ghosh (2007). They establish the conditions under which the posterior is consistent for estimating densities supported on a closed bounded interval in $\mathcal{R}^d$. Their results were extended by Tokdar (2007) to distributions with unbounded support. The convergence rate of LGP estimators are established by van der Vaart et al. (2009).

3 LGP Copula Density Estimator

Although it provides a viable remedy to several difficulties encountered in the nonparametric estimation of copula density, the transformation approach introduces to the final estimate a multiplicative factor that can be volatile in the tails. In this study, we consider an alternative Bayesian estimator that eschews the intricate regularization or tapering discussed in the preceding section.

3.1 Formulation

Upon the probit transformation $s_1 = \Phi^{-1}(F_1(x_1))$ and $s_2 = \Phi^{-1}(F_2(x_2))$, we consider a LGP density estimator for $\boldsymbol{s} = (s_1, s_2)^T \in \mathcal{R}^2$:

$$g(\boldsymbol{s}) = \frac{\exp(f(\boldsymbol{s}))}{\int \exp(f(\boldsymbol{s}'))d\boldsymbol{s}'},$$

and place a Gaussian process prior on the latent function $f : \mathcal{R}^2 \to \mathcal{R}$. To facilitate capturing the global feature of f, we augment a zero mean Gaussian process $f_0(\boldsymbol{s}) \sim \mathcal{GP}(0, K(\boldsymbol{s}, \boldsymbol{s}'))$ with a polynomial mean function $\boldsymbol{h}(\boldsymbol{s})^T\boldsymbol{\beta}$ where

$$\boldsymbol{h}(\boldsymbol{s}) = (s_1, s_2, s_1^2, s_2^2, s_1 s_2)^T \tag{4}$$

and a normal prior $\mathcal{N}(\boldsymbol{0}, \boldsymbol{B})$ is placed on the coefficients $\boldsymbol{\beta}$. It follows that

$$f(\boldsymbol{s}) \sim \mathcal{GP}(0, K(\boldsymbol{s}, \boldsymbol{s}') + \boldsymbol{h}^T(\boldsymbol{s})\boldsymbol{B}\boldsymbol{h}(\boldsymbol{s}')).$$

Remark 1. *Notice that given the above prior distribution f, $g(\boldsymbol{s}) \propto \exp(\boldsymbol{h}^T(\boldsymbol{s})\boldsymbol{\beta})\exp(f_0(\boldsymbol{s}))$. The first factor coincides with the bivariate Gaussian distribution that is capable of reflecting the global feature of the underlying distribution, while the second factor is a flexible Gaussian process that captures deviations from the global baseline. This is in spirit similar to the nonparametric density estimator with a parametric start by Hjort and Glad (1995). Upon back-transformation, the resulting copula density estimator can be viewed as a nonparametric copula estimator with a Gaussian copula start.*

To further improve the flexibility of the baseline density, we also consider the following alternative piecewise polynomial basis functions

$$\boldsymbol{h}(\boldsymbol{s}) = (s_1, s_2, s_{1-}^2, s_{1+}^2, s_{2-}^2, s_{2+}^2, s_{1-}s_{2-}, s_{1+}s_{2+}, s_{1-}s_{2+}, s_{1+}s_{2-})^T, \tag{5}$$

where $x_- = \min(x, 0)$ and $x_+ = \max(x, 0)$. The resulting baseline density resembles the multivariate split-Gaussian density of Geweke (1989). We use the Deviance Information Criterion (DIC) to guide the specification of the basis functions. The DIC is a measure of predictive accuracy in

Bayesian models. Given a sample $\boldsymbol{y}$ and hyperparameters $\boldsymbol{\eta}$ with $\hat{\boldsymbol{\eta}}$ being its posterior mean, it is defined as DIC $= -2\log p(\boldsymbol{y}|\hat{\boldsymbol{\eta}}) + 2p_{\text{DIC}}$. The first term captures the goodness-of-fit and the second term is a penalty of model complexity, given by $p_{\text{DIC}} = 2\big(\log p(\boldsymbol{y}|\hat{\boldsymbol{\eta}}) - E_{\text{post}}(\log p(\boldsymbol{y}|\hat{\boldsymbol{\eta}})\big)$. The expectation in the second term of p_{DIC} is taken with respect to the posterior distribution and can be calculated through simulations (Gelman et al., 2013).

As the true marginal distribution functions are usually unknown in practice, it is customary in copula estimations to use the so-called pseudo-observations instead. Given the observations $\{X_{i1}, X_{i2}\}_{i=1}^n$, we start to estimate the copula density by calculating the pseudo-observations through

$$\hat{U}_{i1} = \frac{n}{n+1}\hat{F}_{1,n}(X_{i1}) \quad \text{and} \quad \hat{U}_{i2} = \frac{n}{n+1}\hat{F}_{2,n}(X_{i2}),$$

where $\hat{F}_{1,n}(x_1) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{X_1 \le x_1\}}$ is the empirical CDF of X_1 and $\hat{F}_{2,n}$ is similarly defined.

This approach avoids parametric assumptions on the marginal distributions and the rank statistics of the margins preserve sufficient information required for copula estimation. The influence of using the pseudo-observations instead of true ones is asymptotically negligible; see discussions in Genest et al. (1995), Geenens et al. (2017) and Wen and Wu (2020).

The inference of the LGP estimator is challenging since both the posterior distribution of g and its normalization factor are intractable. Following Riihimäki et al. (2014), we discretize the data to simplify the inference. We first calculate the probit transforms $(\Phi^{-1}(\hat{U}_{i1}), \Phi^{-1}(\hat{U}_{i2}))$ and then partition the rectangular empirical support spanned by the transformed data into an $m \times m$ equal-sized grid. Denote the number of observations in the j-th cell of the grid by Y_j (such that $\sum_{j=1}^{m^2} Y_j = n$). With a slight abuse of notation, let $\boldsymbol{S} = (\boldsymbol{S}_1, \ldots, \boldsymbol{S}_{m^2})^T$ with $\boldsymbol{S}_j$ being the center of the j-th cell and f_j be the latent value associated with $\boldsymbol{S}_j$. The probability of an observation falling into the j-th cell is then estimated by

$$\pi_j = \exp(f_j) / \sum_{i=1}^{m^2} \exp(f_i).$$

Next let $\boldsymbol{Y} = (Y_1, \ldots, Y_{m^2})^T$ and $\boldsymbol{f} = (f_1, \ldots, f_{m^2})^T$. The conditional distribution of the observed counts $\boldsymbol{Y}$ given latent values $\boldsymbol{f}$ is $p(\boldsymbol{Y}|\boldsymbol{f}) = \prod_{j=1}^{m^2} \pi_j^{Y_j}$. It follows that the log-likelihood

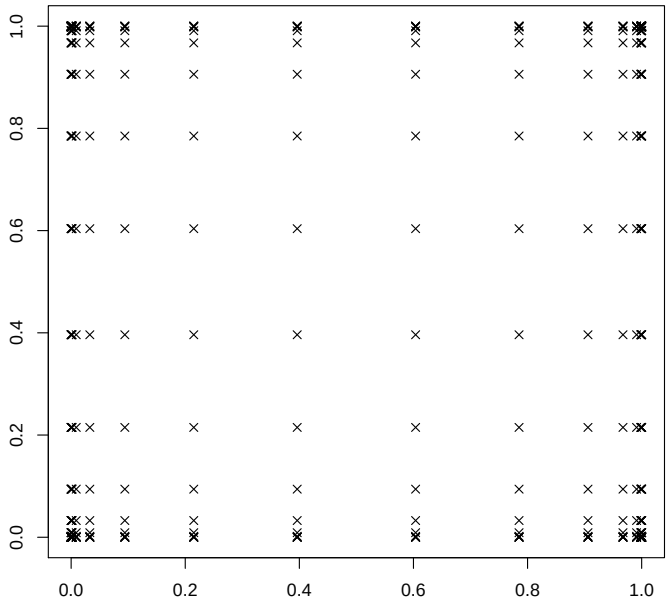

Figure 1: A 20×20 equally-spaced grid $\boldsymbol{s} \in [-5, 5]^2$ transformed to $\boldsymbol{u} \in [0, 1]^2$ via $\boldsymbol{u} = \Phi(\boldsymbol{s})$

function is given by

$$\log p(\boldsymbol{Y}|\boldsymbol{f}) = \boldsymbol{Y}^T \boldsymbol{f} - n \log(\sum_{j=1}^{m^2} \exp(f_j)). \tag{6}$$

With a Gaussian process prior, the conditional distribution of $\boldsymbol{f}$ is then

$$p(\boldsymbol{f}|\boldsymbol{S}, \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{f}|\boldsymbol{0}, \boldsymbol{A}),$$

where $\boldsymbol{A} = \boldsymbol{K} + \boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^T$, with $\boldsymbol{H}$ and $\boldsymbol{K}$ being the basis functions $\boldsymbol{h}$ and covariance function K evaluated at $\boldsymbol{S}$ respectively. Specifically, $\boldsymbol{H}$ is an $m^2 \times j$ matrix and $\boldsymbol{B}$ is $j \times j$ with $j = 5$ or 10 under basis functions (4) or (5) respectively.

Remark 2. *Since the mass of a copula density tends to concentrate near the boundaries of the unit square, it is desirable to use a finer grid near the boundaries when discretizing the data. We notice that the probit transformation achieves this automatically: upon back-transformation via* $\boldsymbol{u} = \Phi(\boldsymbol{s})$*, an equally-spaced grid on* $\mathcal{R}^2$ *becomes increasingly denser towards the boundaries of the unit square, especially near the corners. See Figure 1 for an illustration.*

3.2 Approximate Inference

The posterior distribution of the latent values is given by

$$p(\boldsymbol{f}|\boldsymbol{Y},\boldsymbol{S},\boldsymbol{\theta}) = \frac{p(\boldsymbol{Y}|\boldsymbol{f})p(\boldsymbol{f}|\boldsymbol{S},\boldsymbol{\theta})}{p(\boldsymbol{Y}|\boldsymbol{S},\boldsymbol{\theta})}, \tag{7}$$

where $p(\boldsymbol{Y}|\boldsymbol{S},\boldsymbol{\theta}) = \int p(\boldsymbol{Y}|\boldsymbol{f})p(\boldsymbol{f}|\boldsymbol{S},\boldsymbol{\theta})d\boldsymbol{f}$. Since the log likelihood function (6) is non-Gaussian, the posterior distribution is also non-Gaussian and intractable. We use the Laplace approximation for inference. Ignoring the denominator which is free of $\boldsymbol{f}$, we define

$$\Psi(\boldsymbol{f}) = \log p(\boldsymbol{Y}|\boldsymbol{f}) + \log p(\boldsymbol{f}|\boldsymbol{S},\boldsymbol{\theta}) = \log p(\boldsymbol{Y}|\boldsymbol{f}) - \frac{1}{2}\boldsymbol{f}^T\boldsymbol{A}^{-1}\boldsymbol{f} - \frac{1}{2}\log|\boldsymbol{A}| - \frac{m^2}{2}\log 2\pi. \tag{8}$$

The Laplace approximation of $p(\boldsymbol{f}|\boldsymbol{Y},\boldsymbol{S},\boldsymbol{\theta})$ is based on the second order Taylor expansion of $\Psi(\boldsymbol{f})$ around its posterior mode $\hat{\boldsymbol{f}} = \arg\max_{\boldsymbol{f}} \Psi(\boldsymbol{f})$. It is given by

$$q(\boldsymbol{f}|\boldsymbol{Y},\boldsymbol{S},\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{f}|\hat{\boldsymbol{f}},(\boldsymbol{A}^{-1}+\boldsymbol{W})^{-1}), \tag{9}$$

where $\boldsymbol{W} = -\nabla\nabla \log p(\boldsymbol{Y}|\boldsymbol{f})|_{\boldsymbol{f}=\hat{\boldsymbol{f}}}$ with $W_{ij} = n(-\pi_i\pi_j)$ and $W_{jj} = n(\pi_j - \pi_j^2)$.

The marginal likelihood can be similarly approximated using a second order Taylor expansion:

$$\begin{aligned} p(\boldsymbol{Y}|\boldsymbol{S},\boldsymbol{\theta}) &= \int p(\boldsymbol{Y}|\boldsymbol{f})p(\boldsymbol{f}|\boldsymbol{S},\boldsymbol{\theta})d\boldsymbol{f} = \int \exp(\Psi(\boldsymbol{f}))d\boldsymbol{f} \\ &\approx \int \exp(\Psi(\hat{\boldsymbol{f}}))\exp\left(-\frac{1}{2}(\boldsymbol{f}-\hat{\boldsymbol{f}})^T(\boldsymbol{A}^{-1}+\boldsymbol{W})(\boldsymbol{f}-\hat{\boldsymbol{f}})\right)d\boldsymbol{f}. \end{aligned}$$

It follows that the Laplace approximation of $\log p(\boldsymbol{Y}|\boldsymbol{S},\boldsymbol{\theta})$ takes the form

$$\log q(\boldsymbol{Y}|\boldsymbol{S},\boldsymbol{\theta}) = \log p(\boldsymbol{Y}|\hat{\boldsymbol{f}}) - \frac{1}{2}\hat{\boldsymbol{f}}^T\boldsymbol{A}^{-1}\hat{\boldsymbol{f}} - \frac{1}{2}\log|\boldsymbol{I}+\boldsymbol{W}^{1/2}\boldsymbol{A}\boldsymbol{W}^{1/2}|. \tag{10}$$

Derivations of the above Laplace approximation and marginal likelihood are given in Appendix A and B.

Thus far we have treated the hyperparameters $\boldsymbol{\theta}$ as given. Given a prior distribution $p(\boldsymbol{\theta})$, the

posterior distribution of $\boldsymbol{\theta}$ takes the form

$$p(\boldsymbol{\theta}|\boldsymbol{Y},\boldsymbol{S}) = \frac{p(\boldsymbol{Y}|\boldsymbol{S},\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{Y}|\boldsymbol{S})}$$

where $p(\boldsymbol{Y}|\boldsymbol{S}) = \int p(\boldsymbol{Y}|\boldsymbol{S},\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$. This posterior distribution can be learned via several approaches. One possibility is the approach of maximum a posterior (MAP), which seeks the posterior mode

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \log p(\boldsymbol{Y}|\boldsymbol{S},\boldsymbol{\theta}) + \log p(\boldsymbol{\theta}). \tag{11}$$

Under Laplace approximation, the approximate posterior $q(\boldsymbol{Y}|\boldsymbol{S},\boldsymbol{\theta})$ given in (10) is used in the place of $p(\boldsymbol{Y}|\boldsymbol{S},\boldsymbol{\theta})$. Next the predictive mean of $\exp(f^*)$ associated with a test location $\boldsymbol{s}^* = (s_1^*, s_2^*)$ is given by

$$E_q[\exp(f^*)|\boldsymbol{Y},\boldsymbol{S},\hat{\boldsymbol{\theta}}] = \int \exp(\boldsymbol{A}^*\nabla \log p(\boldsymbol{Y}|\boldsymbol{f}))q(\boldsymbol{f}|\boldsymbol{Y},\boldsymbol{S},\hat{\boldsymbol{\theta}})d\boldsymbol{f},$$

where $\boldsymbol{A}^* = K(\boldsymbol{s}^*,\boldsymbol{S}) + \boldsymbol{h}(\boldsymbol{s}^*)\boldsymbol{B}\boldsymbol{H}^T$.

Since the Laplace approximation may not adequately describe the non-Gaussian posterior distribution, we further employ importance sampling in the approximate inference. Suppose that $\tilde{q}(\boldsymbol{f})$ is an approximating distribution, which can be sampled with ease, to the Laplace approximate $q(\boldsymbol{f}|\boldsymbol{Y},\boldsymbol{S},\hat{\boldsymbol{\theta}})$; for instance, the multivariate split Gaussian distribution proposed by Geweke (1989). Let $\boldsymbol{f}_s, s = 1,\ldots,S$, be an i.i.d. sample from $\tilde{q}(\boldsymbol{f})$. We estimate the posterior mean of $\exp(f^*)$ by

$$\bar{\mu}^* = \frac{1}{S}\sum_{s=1}^{S} \frac{\exp\left(\boldsymbol{A}^*\nabla \log p(\boldsymbol{Y}|\boldsymbol{f}_s)\right)}{\tilde{q}(\boldsymbol{f}_s)/q(\boldsymbol{f}_s|\boldsymbol{S},\hat{\boldsymbol{\theta}})}.$$

It follows that the posterior mean of the density $g(\boldsymbol{s}^*|\boldsymbol{Y},\boldsymbol{S},\hat{\boldsymbol{\theta}})$ is calculated as

$$\bar{g}(\boldsymbol{s}^*) = \frac{\bar{\mu}^*}{\Delta\sum_{j=1}^{m^2}\bar{\mu}_j}, \tag{12}$$

where $\bar{\mu}_j$ is the posterior mean of $\exp(f_j)$ at training location $\boldsymbol{S}_j$ and Δ is the size of each grid cell. Finally upon back-transformation, the predictive mean of copula density at $\boldsymbol{u}^* = (u_1^*, u_2^*) =$

$(\Phi(s_1^*), \Phi(s_2^*))$ is given by

$$\bar{c}(u_1^*, u_2^* | \boldsymbol{Y}, \boldsymbol{S}, \hat{\boldsymbol{\theta}}) = \frac{\bar{g}(\Phi^{-1}(u_1^*), \Phi^{-1}(u_2^*))}{\phi(\Phi^{-1}(u_1^*))\phi(\Phi^{-1}(u_2^*))}.$$

3.3 Posterior sampler

It is often of interest to generate random samples from a copula distribution since simulation-based estimations have been widely used in finance and risk management for tasks such as option pricing, multivariate Value at Risk, and expected shortfall. These tasks typically involve calculating quantities in the form of $E_c[t(\boldsymbol{u})] = \int t(\boldsymbol{u})c(\boldsymbol{u})d\boldsymbol{u}$, where t is a function of interest defined on the support of a copula density. Oftentimes this integral does not admit a simple analytical form and Monte Carlo integration is used. Since the posterior distribution of the proposed LGP estimator is intractable, we propose an importance sampling approach that samples indirectly from an approximating distribution $\tilde{C}$ to facilitate the calculation of $E_c[t(\boldsymbol{u})]$.

Given an approximating density $\tilde{c}$, we can write

$$E_c[t(\boldsymbol{u})] = \frac{\int \frac{t(\boldsymbol{u})c(\boldsymbol{u})}{\tilde{c}(\boldsymbol{u})}\tilde{c}(\boldsymbol{u})d\boldsymbol{u}}{\int \frac{c(\boldsymbol{u})}{\tilde{c}(\boldsymbol{u})}\tilde{c}(\boldsymbol{u})d\boldsymbol{u}}.$$

Suppose $\{\boldsymbol{U}_s\}_{s=1}^S$ is an i.i.d. sample from $\tilde{C}$, which can be easily sampled from. We estimate $E_c[t(\boldsymbol{u})]$ using

$$\hat{t} = \frac{\sum_{s=1}^S w_s t(\boldsymbol{U}_s)}{\sum_{s=1}^S w_s},$$

where $w_s = c(\boldsymbol{U}_s)/\tilde{c}(\boldsymbol{U}_s)$. See e.g. Geweke (1996) and Liu (2008) for the methods of importance sampling and Monte Carlo integration. Denote by $\tilde{g}$ the transformed density of copula $\tilde{c}$ on $\mathcal{R}^2$. The identity (1), which links c to g, suggests

$$\frac{c(u_1, u_2)}{\tilde{c}(u_1, u_2)} = \frac{g(s_1, s_2)}{\tilde{g}(s_1, s_2)}.$$

It follows that the importance sampling can be equivalently conducted in terms of g or c. We base our importance sampling on g since it contains a parametric component coinciding with the bivariate Gaussian distribution, a natural candidate for the approximating distribution $\tilde{g}$.

The efficiency of importance sampling generally depends on how closely the approximating

density resembles the target density. One commonly used indicator of sampling efficiency is the effective sample size:

$$S_{\text{eff}} = \frac{1}{\sum_{s=1}^{S}(w_s^*)^2}, \tag{13}$$

where $w_s^* = w_s / \sum_{s'=1}^{S} w_{s'}$. The closer S_{eff} is to S, the higher the sampling efficiency. Liu (2008) reviewed methods of efficient importance sampling. The method of adaptive importance sampling seeks an approximating distribution that minimizes the variance of sampling weights. This is equivalent to maximizing the effective sample size and theoretically desirable. It is, however, known to be numerically unstable.

A less 'greedy' but more stable alternative is to seek an approximating density that minimizes some distance measure to the target density. In particular, we consider a bivariate approximating density $\tilde{g}(\boldsymbol{s}) = \mathcal{N}(\boldsymbol{s}|\tilde{\mu}, \tilde{\Sigma})$ that minimizes the integrated squared error (ISE) $\int(\bar{g}(\boldsymbol{s}) - \tilde{g}(\boldsymbol{s}))^2 d\boldsymbol{s}$, where $\bar{g}$ is given in (12). The coefficients of $\tilde{g}$ are obtained by minimizing the sample ISE:

$$(\tilde{\mu}, \tilde{\Sigma}) = \arg\min_{\mu,\Sigma} \sum_{j=1}^{m^2} (\bar{g}(\boldsymbol{S}_j) - \mathcal{N}(\boldsymbol{S}_j|\mu, \Sigma))^2, \tag{14}$$

subject to the constraint that Σ is positive definite. The detail of this procedure is given in Appendix C.

Lastly let $\{\tilde{S}_s\}_{s=1}^{S}$ be an i.i.d. sample from the approximating distribution $\tilde{g}$, the Monte Carlo estimate of $E_c[t(\boldsymbol{u})]$ is then calculated as

$$\hat{t} = \frac{\sum_{s=1}^{S} w_s t(\Phi(\tilde{S}_s))}{\sum_{s=1}^{S} w_s},$$

where $w_s = \bar{g}(\tilde{S}_s)/\tilde{g}(\tilde{S}_s)$. Numerical simulations reported in the next section demonstrate the good performance of this posterior sampler.

4 Simulations

We conduct Monte Carlo simulations to examine the finite sample performance of the LGP copula density estimator and the efficiency of the proposed sampling method. We consider two configurations of LGP estimators: $\hat{c}_1$ with the quadratic mean basis functions (4) and $\hat{c}_2$ with the

more flexible piecewise mean basis functions (5). We use the DIC to select a preferred model, which is denoted by $\hat{c}_3$. We assume a mean zero normal prior distribution for the coefficients of the mean basis functions (4) and (5) with covariance $\boldsymbol{B} = 10^2\boldsymbol{I}_5$ and $\boldsymbol{B} = 10^2\boldsymbol{I}_{10}$ respectively. Following Riihimäki et al. (2014), we use the half Student-t_4 distribution as the prior for $\boldsymbol{\theta} = (l_1, l_2, \sigma^2)$, the hyperparameters of the covariance function. The variances of these t_4-priors are set respectively at 10 for l_1, l_2 and 1000 for σ^2. These weakly informative prior distributions are rather generic. We have experimented with alternative prior distributions and found that our simulation results are not sensitive to the configuration of hyper prior distributions. The same holds for alternative covariance functions rather than the squared exponential covariance function used in our simulations.

To facilitate comparisons with existing studies, we follow the experiment design of Wen and Wu (2020) and consider two groups of copulas. The first group includes four commonly used parametric copulas that differ considerably from the Gaussian copula: (A) the Student t-copula with 5 degrees of freedom, with parameters $\rho = 0.454$ and $\rho = 0.809$; (B) the Frank copula with parameters $\theta = 2.92$ and $\theta = 7.93$; (C) the Gumbel copula with parameters $\theta = 10/7$ and $\theta = 2.5$; (D) the Clayton copula with parameters $\theta = 6/7$ and $\theta = 3$. The second group includes four Gaussian or near Gaussian copulas: (E) the Gaussian copula with parameters $\rho = 0.454$ and $\rho = 0.809$; (F) mixture of 85% Gaussian copula and 15% Clayton copula with two pairs of parameters ($\rho = 0.454,\ \theta = 6/7$) and ($\rho = 0.809,\ \theta = 3$); (G) the Student t-copula with 15 degrees of freedom, with parameters $\rho = 0.454$ and $\rho = 0.809$; (H) mixture of 85% student t-copula with 15 degrees of freedom and 15% Clayton copula with two pairs of parameters ($\rho = 0.454,\ \theta = 6/7$) and ($\rho = 0.809,\ \theta = 3$). The coefficients for each copula are selected to allow a low dependence with Kendall's $\tau = 0.3$ or a high dependence with $\tau = 0.6$. For the LGP estimation on the transformed data, we consider two sample sizes $n = 500$ and $n = 2000$ with a 30×30 and 50×50 equal-sized grids on $\mathcal{R}^2$ respectively. Each experiment is repeated 1000 times.

We evaluate the global performance by the integrated squared error (ISE) on a 99×99 equally spaced grid within the unit square. Wen and Wu (2020) conducted a detailed investigation of many nonparametric copula density estimators. To save space, we shall only compare our estimators to the best overall estimator reported in that study, an improved transform-kernel estimator denoted by $\tilde{c}_m$. Table 1 reports the simulation results. The LGP estimators $\hat{c}_1, \hat{c}_2$ and $\hat{c}_3$ outperform $\tilde{c}_m$ with only a few exceptions. Their better performance is evident across copula specifications, levels

of dependence and sample sizes. Among them, $\hat{c}_1$ with quadratic mean functions performs better for the Gaussian and near Gaussian copulas in Group 2. This is expected as its mean function corresponds to the Gaussian copula. In contrast, $\hat{c}_2$ — thanks to its more flexible piecewise mean function — fares better for the non-Gaussian copulas in Group 1. In practice, the choice between these two configurations can be difficult since the type of underlying copula is generally unknown. It is reassuring to observe that $\hat{c}_3$, the preferred estimator between these two according to the DIC, clearly provides the best overall performance. It is also worthy to note that according to Wen and Wu (2020), $\tilde{c}_m$ obtains a faster (than the standard) convergence rate for Gaussian copulas and is particularly advantageous for the estimation of Gaussian and near-Gaussian copulas. Nonetheless, our experiments suggest that the LGP estimators offer further and oftentimes substantial improvements for these copulas.

Table 1: Average ISE of estimated copula densities

	n=500				n=2000			
Copulas	$\hat{c}_1$	$\hat{c}_2$	$\hat{c}_3$	$\tilde{c}_m$	$\hat{c}_1$	$\hat{c}_2$	$\hat{c}_3$	$\tilde{c}_m$
Group 1								
A1	0.0261	**0.0169**	**0.0202**	0.0208	**0.0067**	0.0087	**0.0063**	0.0102
A2	0.0672	**0.0500**	**0.0547**	0.0608	**0.0233**	0.0260	**0.0228**	0.0278
B1	0.0167	**0.0158**	0.0165	**0.0120**	0.0075	**0.0052**	**0.0053**	0.0064
B2	0.0426	0.0428	**0.0393**	**0.0383**	**0.0091**	**0.0099**	**0.0091**	0.0195
C1	0.0294	**0.0207**	**0.0267**	0.0332	0.0158	**0.0153**	**0.0148**	0.0174
C2	0.1427	**0.1225**	**0.1215**	0.1441	**0.0594**	0.0822	**0.0608**	0.0717
D1	0.0249	**0.0207**	**0.0227**	0.0481	**0.0101**	0.0151	**0.0110**	0.0238
D2	**0.2720**	0.3618	**0.2934**	0.3531	**0.1265**	0.1639	**0.1269**	0.1596
Group 2								
E1	**0.0026**	0.0073	**0.0028**	0.0083	**0.0006**	0.0018	**0.0007**	0.0034
E2	**0.0035**	0.0095	**0.0048**	0.0255	**0.0009**	0.0026	**0.0013**	0.0105
F1	**0.0043**	0.0077	**0.0046**	0.0098	**0.0021**	**0.0022**	**0.0022**	0.0042
F2	0.0259	**0.0217**	**0.0251**	0.0421	0.0196	**0.0155**	**0.0167**	0.0219
G1	**0.0050**	0.0084	**0.0058**	0.0104	**0.0033**	**0.0030**	**0.0033**	0.0045
G2	**0.0099**	0.0136	**0.0114**	0.0327	0.0075	**0.0068**	**0.0073**	0.0138
H1	**0.0075**	0.0093	**0.0082**	0.0124	0.0055	**0.0038**	**0.0050**	0.0057
H2	0.0408	**0.0329**	**0.0378**	0.0504	0.0333	**0.0267**	0.0285	**0.0260**

NOTE: A1 and A2 refer to copula specification A with Kendall's τ being 0.3 and 0.6 respectively; other copulas are similarly denoted. The boldface font with and without an underline indicate the smallest and the second smallest ISEs among the four estimators in each experiment.

We next examine the tail performance of the copula estimators, which is of critical importance in risk management. We evaluate the tail performance by restricting the calculation of ISE to the

four corners of the unit square $\mathcal{I}^2$ with a 'tail' grid $\{0.0005, 0.0015, 0.0025, \cdots, 0.0485, 0.0495\} \cup \{0.9505, 0.9515, 0.9525, \cdots, 0.9985, 0.9995\}$. The results for $n = 2000$ are reported in Table 2. Evidently, the LGP estimators outperform their competitor in the tails across the board.

Table 2: Average ISE in the tails

Copulas	$\hat{c}_1$	$\hat{c}_2$	$\hat{c}_3$	$\tilde{c}_m$	Copulas	$\hat{c}_1$	$\hat{c}_2$	$\hat{c}_3$	$\tilde{c}_m$
	Group 1					Group 2			
A1	7.02	**5.58**	6.85	**3.80**	E1	**0.10**	0.23	**0.11**	0.62
A2	13.24	**12.94**	**12.99**	13.76	E2	**0.31**	0.71	**0.36**	6.12
B1	0.84	**0.34**	**0.41**	0.68	F1	**0.55**	**0.48**	**0.55**	1.08
B2	**1.80**	1.98	**1.82**	2.66	F2	5.14	**4.66**	**4.89**	10.67
C1	**5.06**	7.28	**6.41**	8.52	G1	0.72	**0.43**	**0.64**	1.14
C2	**28.65**	39.42	**29.90**	36.48	G2	1.98	**1.35**	**1.80**	7.39
D1	**4.56**	7.18	**5.08**	9.98	H1	1.57	**0.98**	**1.33**	1.68
D2	**55.94**	69.50	**55.91**	83.18	H2	9.20	**8.40**	**8.77**	12.54

NOTE: A1 and A2 refer to copula specification A with Kendall's τ being 0.3 and 0.6 respectively; other copulas are similarly denoted. The boldface font with and without an underline indicate the smallest and the second smallest ISEs among the four estimators in each experiment.

Lastly we investigate the efficiency of the proposed posterior sampler based on the LGP copula density estimator. For the sake of brevity, we focus on the posterior copula distribution estimated by $\hat{c}_1$ with sample sizes $n = 2000$. In particular, $S = 1000$ samples are drawn from the approximating distribution described in the previous section and the effective sample size S_{eff} is calculated according to (13). Each experiment is repeated 500 times. Summary statistics of the relative effective sample size, S_{eff}/S, are reported in Table 3. It is evident that the proposed sampling method provides a practically efficient algorithm to sample from the posterior copula distribution. As expected, the sampling efficiency is particularly high for the Gaussian and near-Gaussian copulas (i.e. Group 2) thanks to the Gaussian approximating distribution used in the posterior sampler.

Table 3: Summary statistics of S_{eff}/S

Copulas	Median	Mean	Copulas	Median	Mean
	Group 1			Group 2	
A1	0.9770	0.9518	E1	0.9999	0.9999
A2	0.9803	0.9532	E2	0.9999	0.9999
B1	0.9868	0.9813	F1	0.9998	0.9991
B2	0.8764	0.8211	F2	0.9998	0.9953
C1	0.9739	0.9695	G1	0.9999	0.9997
C2	0.9264	0.9029	G2	0.9999	0.9999
D1	0.9280	0.9235	H1	0.9997	0.9984
D2	0.6713	0.5994	H2	0.9997	0.9936

NOTE: A1 and A2 refer to copula specification A with Kendall's τ being 0.3 and 0.6 respectively; other copulas are similarly denoted.

5 Empirical Application

Copula has found a wide range of applications in time series and financial econometrics studies; see e.g. Fermanian and Scaillet (2003), Chen and Fan (2006a), Chen and Fan (2006b), Lee and Long (2009), Creal and Tsay (2015) and Chen et al. (2020). In this section, we illustrate the usefulness of the proposed copula density estimator with an application to option pricing.

Following Chiou and Tsay (2008), we analyze the daily log returns of Taiwan weighted stock index (TAIEX) and the New York Stock Exchange composite price index (NYSE) from January 1, 2017 to December 31, 2019. The data are obtained from Datastream and consists of 754 observations for NYSE and 737 for TAIEX. In total, there are 709 overlapped daily log returns because of different holidays for the two stock markets. Some descriptive statistics for the two return series are given in Table 4.

Table 4: Summary statistics of the daily log returns (in percentages)

	Min	1Q	Median	3Q	Max	Mean	S.D.	Skewness	Kurtosis
NYSE	−3.99	−0.26	0.06	0.42	3.95	0.03	0.72	−0.77	4.84
TAIEX	−6.52	−0.32	0.07	0.48	2.86	0.03	0.76	−1.35	10.58

5.1 Copula Estimation

Similarly to Chiou and Tsay (2008), we use the GARCH$(1,1)$ model to fit the marginal process of each asset return and use copula to model their dependence. The marginal process for return

$X_{i,t}, t = 1, \cdots, T, i = 1, 2$, is given by

$$\begin{aligned} X_{i,t} &= \mu_i + a_{i,t},\ a_{i,t} = \sqrt{h_{i,t}}\varepsilon_{i,t},\ \varepsilon_{i,t} \sim N(0,1) \\ h_{i,t} &= \kappa_i + \alpha_i a_{i,t-1}^2 + \beta_i h_{i,t-1}, \end{aligned} \tag{15}$$

where $h_{i,t}$ is the GARCH conditional volatility. The estimated parameters are displayed in Table 5. All parameters are statistically significant at the 5% confidence level. We next apply the proposed LGP estimator to estimate the copula density of the standardized residuals. Surface and contour plots of the estimated copula are reported in Figure 2. The results suggest that the two series are positively dependent and the dependence is stronger during market downturns.

Table 5: Estimated parameters and standard error (in parentheses) for the marginal GARCH models

	$\mu_i \times 10^4$	$\kappa_i \times 10^6$	α_i	β_i
NYSE	7.7204	2.9124	0.2362	0.7238
	(1.9543)	(0.9578)	(0.0347)	(0.0348)
TAIEX	7.9323	4.7148	0.1504	0.7790
	(2.5373)	(1.7170)	(0.0158)	(0.0298)

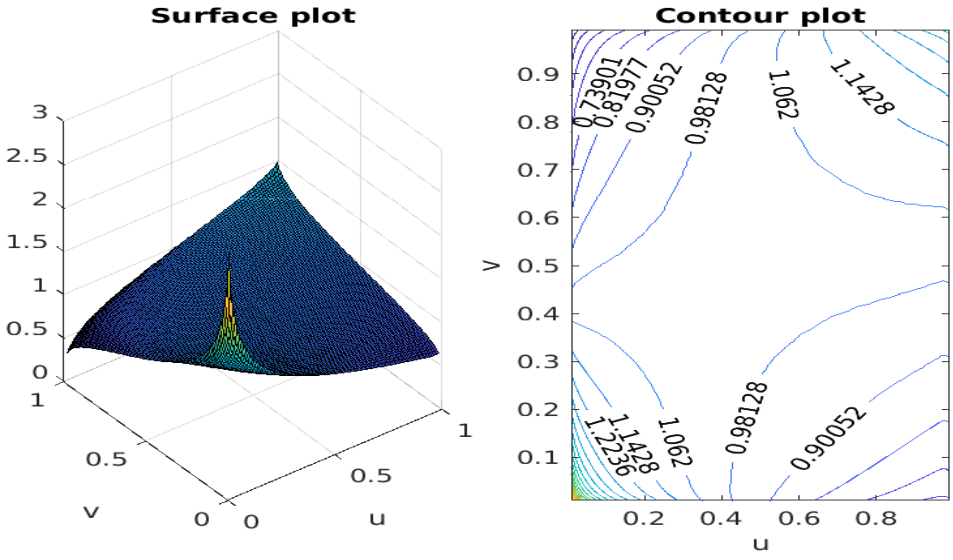

Figure 2: Surface and contour plots for estimated copula density

Since the degree of dependence between these two returns may vary over time, we calculate

Pearson's correlation coefficient, Spearman's rho, and Kendall's tau between them based on a rolling window of 250 days. Specifically using the sampler discussed in Section 3.3, we sample from the posterior copula distribution for each rolling window and then calculate the three measures based on the these samples. The results are reported in Figure 3. All three dependence measures exhibit a common pattern: a positive dependence increased between January 2018 and July 2019 and then gradually declined during the second half of 2019.

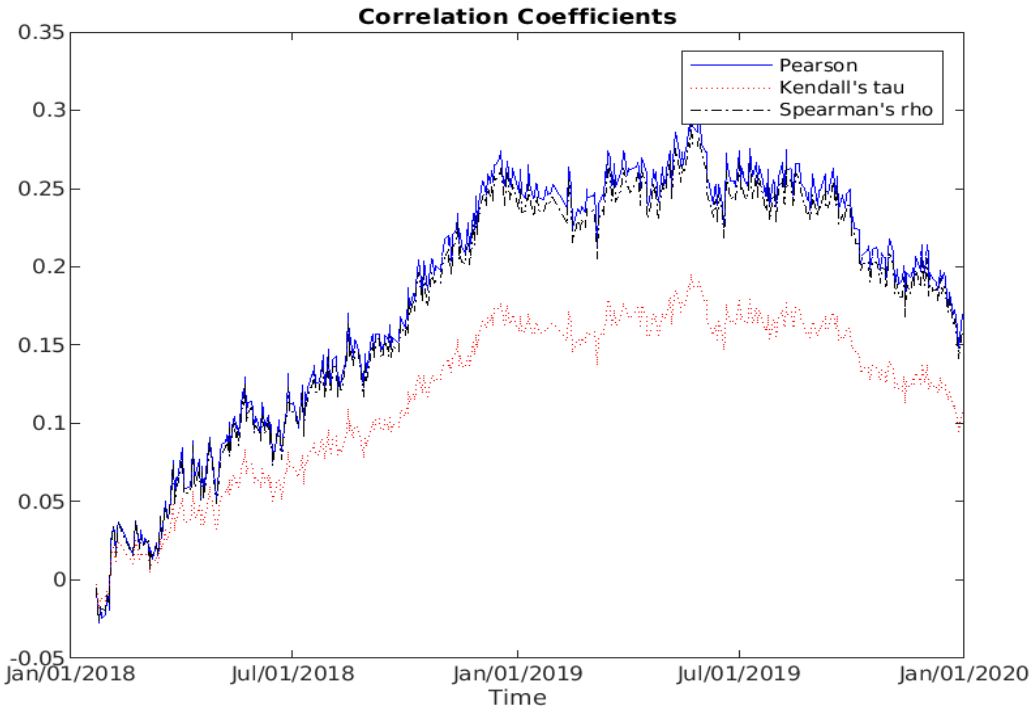

Figure 3: Estimated dependence coefficients

To gain further insight into the suggested asymmetry, we explore the quantile dependence between the two returns. The qth quantile dependence of a copula distribution $C(U, V)$ is defined as,

$$\tau_q = \begin{cases} \Pr\left[U \leq q | V \leq q\right] = \frac{C(q,q)}{q}, & \text{if } 0 < q \leq 1/2, \\ \Pr\left[U > q | V > q\right] = \frac{1-2q+C(q,q)}{1-q}, & \text{if } 1/2 < q < 1. \end{cases}$$

Given the posterior copula distribution, we use the sampling method described in Section 3.3 to draw a random sample $(U_i, V_i), i = 1, 2, \cdots, N$, and denote their associated importance weights by

W_i. The posterior quantile dependence is then calculated as

$$\hat{\tau}_q = \begin{cases} \frac{1}{q}\sum_{i=1}^{N} 1\{U_i \leq q, V_i \leq q\} W_i & \text{if } 0 < q \leq 1/2, \\ \frac{1}{(1-q)}\sum_{i=1}^{N} 1\{U_i > q, V_i > q\} W_i & \text{if } 1/2 < q < 1. \end{cases}$$

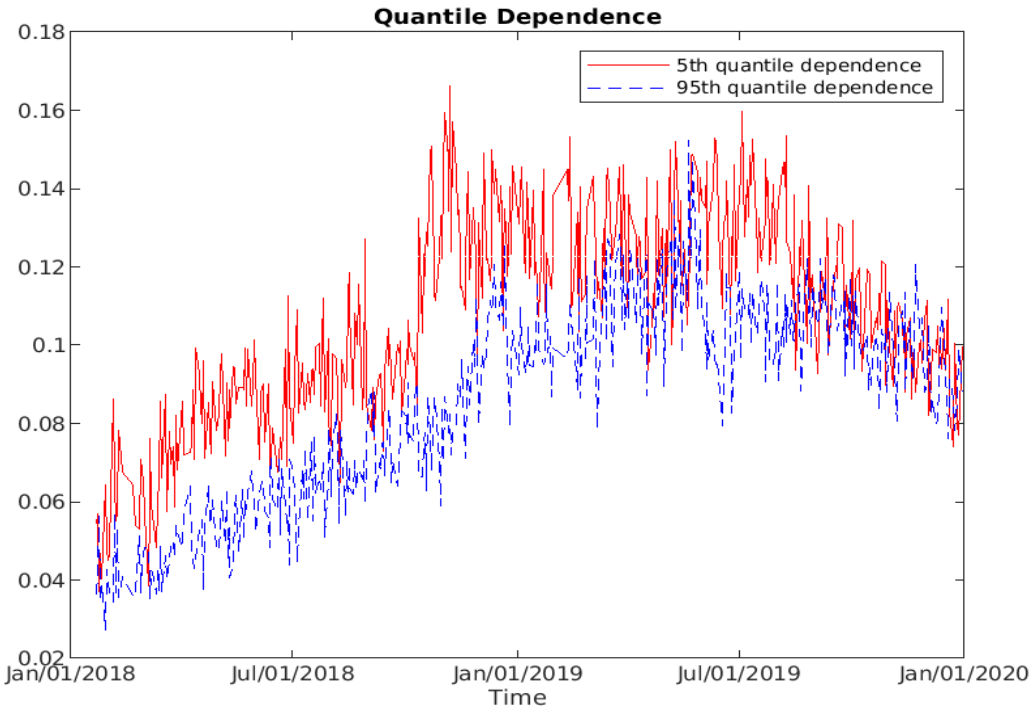

Figure 4: 5^{th} and 95^{th} quantile dependence.

Figure 4 depicts the estimated time-varying 5^{th} (red) and 95^{th} (blue) quantile dependence of NYSE and TAIEX based on a rolling windows of 250 days. The 5^{th} quantile dependence is generally stronger than the 95^{th} quantile dependence during the sample period, confirming a stronger dependence during market downturns. In particular, the lower tail dependence is noticeably higher during the global stock market crisis at the end of 2018. As the stock market recovered in the first half of 2019, the lower tail dependence gradually declined while the higher tail dependence plateaued. The two series converged and continued declining throughout the bull markets in the second half of 2019.

We further conduct a tail asymmetry analysis as described in Rosco and Joe (2013) and Joe (2014). Denote by $U \overset{d}{=} V$ the notion that U and V share a common distribution. For a reflection

symmetric copula, $(U, V) \stackrel{d}{=} (1-U, 1-V)$ and $U+V-1 \stackrel{d}{=} 1-(U+V)$. It follows that $U+V-1$ is symmetric about 0. If instead a copula C is skewed to the upper (lower) tail, $U+V-1$ is right-skewed (left-skewed). The estimated $U+V-1$ for our sample are reported in Figure 5. The majority of them are negative, again confirming the asymmetric dependence between the two returns.

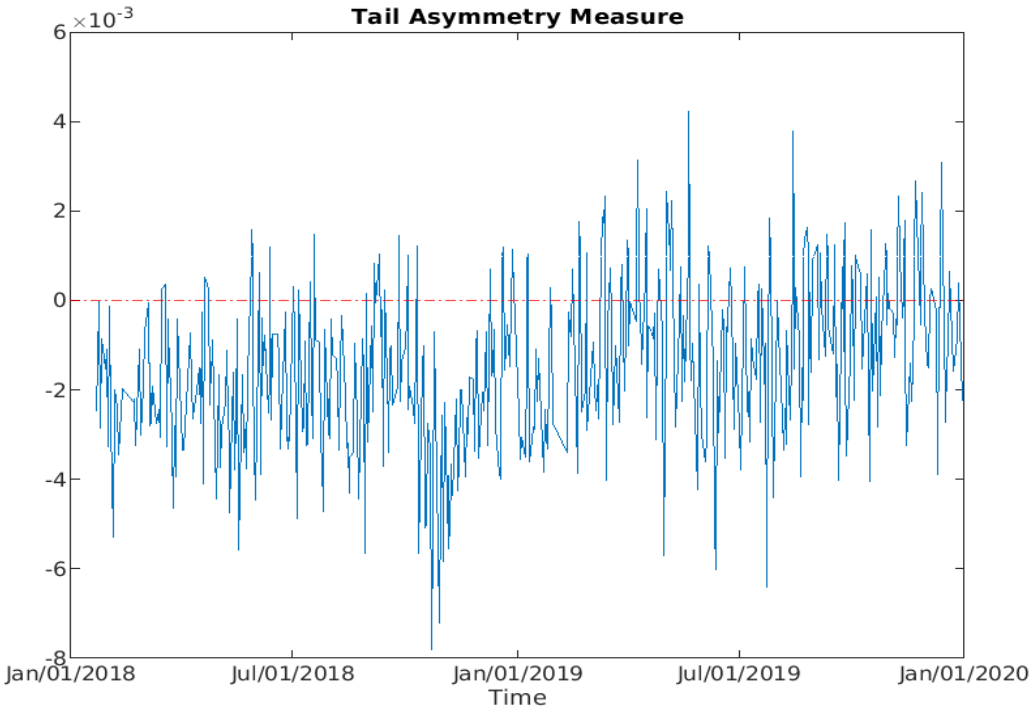

Figure 5: Tail asymmetry measure: $U+V-1$.

5.2 Option Pricing

Multivariate contingent claims (MVCCs), such as options, are derivatives whose payoffs depend on more than one underlying asset. The expected payoff of an MVCC is a function of the multivariate distribution of the underlying assets. Thus, the joint density function plays a crucial role in the valuation of an MVCC. Below we demonstrate how to utilize the proposed copula density estimator and sampler for option pricing.

The arbitrage-free option pricing is based on the joint risk-neutral density of underlying assets. It can be written as the product of risk-neutral marginal densities and a risk-neutral dependence

function as follows: $f^*(x,y) = f_X^*(x)f_Y^*(y)c^*(u^*,v^*)$, where $u^* = F_X^*(x)$, $v^* = F_Y^*(y)$, and the superscript $*$ represents risk-neutral density (RND) with distribution Q. This distribution differs from the real world return distribution P with density $f(x,y) = f_X(x)f_Y(y)c(u,v)$, where $u = F_X(x)$ and $v = F_Y(y)$. The latter can be estimated directly from historical finance data.

Following the common practice in the finance literature (see e.g. Cherubini and Luciano (2002), Rosenberg (2003), and Chiou and Tsay (2008)), we assume that the risk-neutral and real world distributions share the same dependence structure and seek the risk-neutral densities of the marginal processes. Let $S_{i,t}$ be the price for asset i at date t and $\sigma_{i,t}$ be the conditional standard deviation of the log returns. The dynamic of each log return process is assumed to follow the model

$$\begin{aligned} &\log\left(\frac{S_{i,t}}{S_{i,t-1}}\right) = r_f + \lambda_i\sigma_{i,t} - \frac{1}{2}\sigma_{i,t}^2 + a_{i,t}, \\ &a_{i,t} = \sigma_{i,t}\varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim N(0,1) \quad \text{under measure P (real world)}, \\ &\sigma_{i,t}^2 = \gamma_i + \alpha_i a_{i,t-1}^2 + \beta_i\sigma_{i,t-1}^2, \end{aligned} \tag{16}$$

where r_f is the risk-free interest rate and λ_i is the unit risk premium. Based on the locally risk-neutral valuation relationship (LRNVR) derived by Duan (1995), the dynamic process under risk-neutral measure Q is given by

$$\begin{aligned} &\log\left(\frac{S_{i,t}}{S_{i,t-1}}\right) = r_f - \frac{1}{2}\sigma_{i,t}^2 + \xi_{i,t}, \\ &\xi_{i,t} = \sigma_{i,t}\varepsilon_{i,t}^*, \quad \varepsilon_{i,t}^* \sim N(0,1) \quad \text{under measure Q (risk-neutral world)}, \\ &\sigma_{i,t}^2 = \gamma_i + \alpha_i(\varepsilon_{i,t-1}^* - \lambda_i)^2 a_{i,t-1}^2 + \beta_i\sigma_{i,t-1}^2. \end{aligned} \tag{17}$$

Our goal is to calculate the option price at maturity T, that is, the option expires in T days after December 31, 2019, the last day of our sample period. Below we set the time index $t = 0$ for the end of the sample period. The pricing procedure proceeds as follows: for $t = 0, 1, \ldots, T-1$,

1. Use the GARCH model (15) to estimate the marginal processes and the LGP model to estimate their copula based on the immediately preceding 250 observations. Combine the estimated marginals and copula densities to obtain the joint real world distribution P, which is to be updated sequentially.

2. (a) Draw a random sample $\boldsymbol{\varepsilon}^*_{t+1} = (\varepsilon_{1,t+1}, \varepsilon_{2,t+1})$ from the estimated copula.[1] (b) Apply model (15) to forecast the conditional volatility $\boldsymbol{h}_{t+1} = (h_{1,t+1}, h_{2,t+1})$ for both returns at time t under P, which are then used in conjunction with $\boldsymbol{\varepsilon}^*_{t+1}$ to predict the returns $\boldsymbol{X}_{t+1} = (X_{1,t+1}, X_{2,t+1})$. (c) Use models (16) and (17) to forecast volatility $\boldsymbol{\sigma}_1 = (\sigma_{1,1}, \sigma_{2,1})$ under the RND Q.

3. Treat $\boldsymbol{X}_{t+1}$ from step 2 as the realized return and combine it with the preceding 249 observations to form a new rolling window of 250 observations.

At the end of the iterations, we calculate one instance of the simulated asset price at maturity T by: $S_{i,T} = S_{i,0} \exp\left(r_f T - \frac{1}{2}\sum_{t=1}^{T} \sigma_{i,t}^2 + \sum_{t=1}^{T} \sigma_{i,t}\varepsilon^*_{i,t}\right)$, where $S_{i,0}$ is the observation at the end of sampling period for underlying asset $i = 1, 2$. This procedure is repeated a large number of times, say J. Finally we estimate the option price with

$$P_T = \exp(-r_f T)\frac{1}{J}\sum_{j=1}^{J} g(S^j_{1,T}, S^j_{2,T}),$$

where $g(\cdot)$ is the payoff function of the option under consideration.

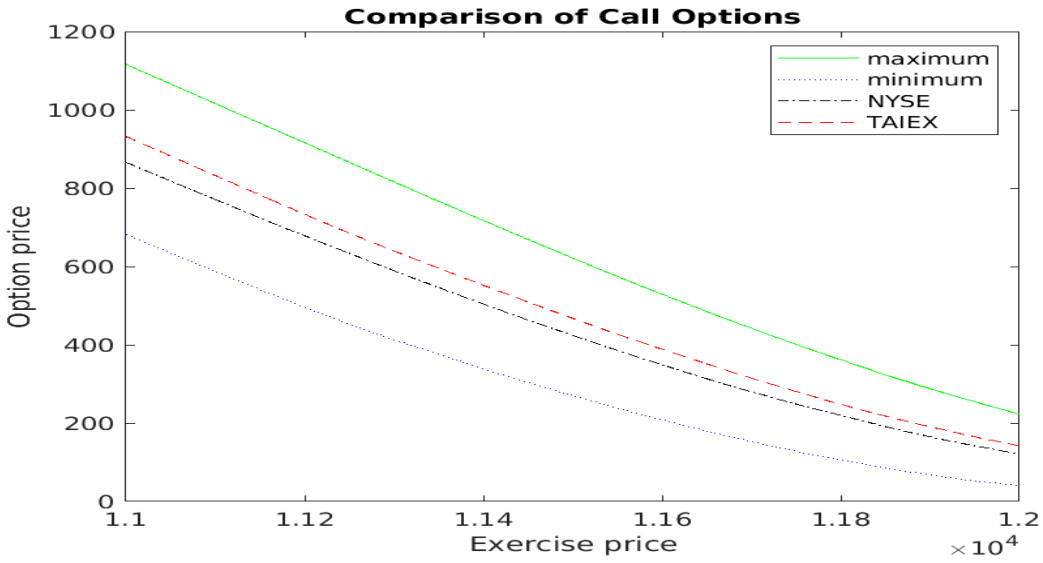

Figure 6: Call option prices on the NYSE, TAIEX indexes and maximum and minimum of these two.

[1] We use the proposed posterior sampler to draw a weighted samples and then apply the method of Sampling-Importance Resampling (SIR) to obtain an independent random samples; see e.g. Gelman et al. (2013).

In addition to call options on the individual index of NYSE and TAIEX, we also consider two multi-index call options: the "maximum" call option with payoff $\max[\max(\text{NYSE}, \text{TAIEX}) - P_{\text{ex}}, 0]$ and the"minimum" call option with a payoff $\max[\min(\text{NYSE}, \text{TAIEX}) - P_{\text{ex}}, 0]$, where P_{ex} is the exercise price. Since there is a roughly 2000 point average difference between the NYSE and TAIEX indexes during the sample period, we subtract 2000 from the NYSE index to make the two returns comparable in this analysis. Figure 6 plots the estimated prices of the four options with a wide range of exercise prices from 11,000 to 12,000 at a maturity date $T = 30$, where the risk-free rate is fixed at 0%. As expected, the maximum and minimum options command the highest and lowest prices among the four options.

We next examine option prices under some alternative dependence structures to explore how the dependence among assets influence the price of their induced options. In addition to the estimated dependence, we consider three hypothetical scenarios: the 'positive', 'negative' and 'independent' scenarios corresponding to a constant dependence at Kendall's $\tau = 0.99, -0.99$ and 0 during the entire course of simulations.[2]

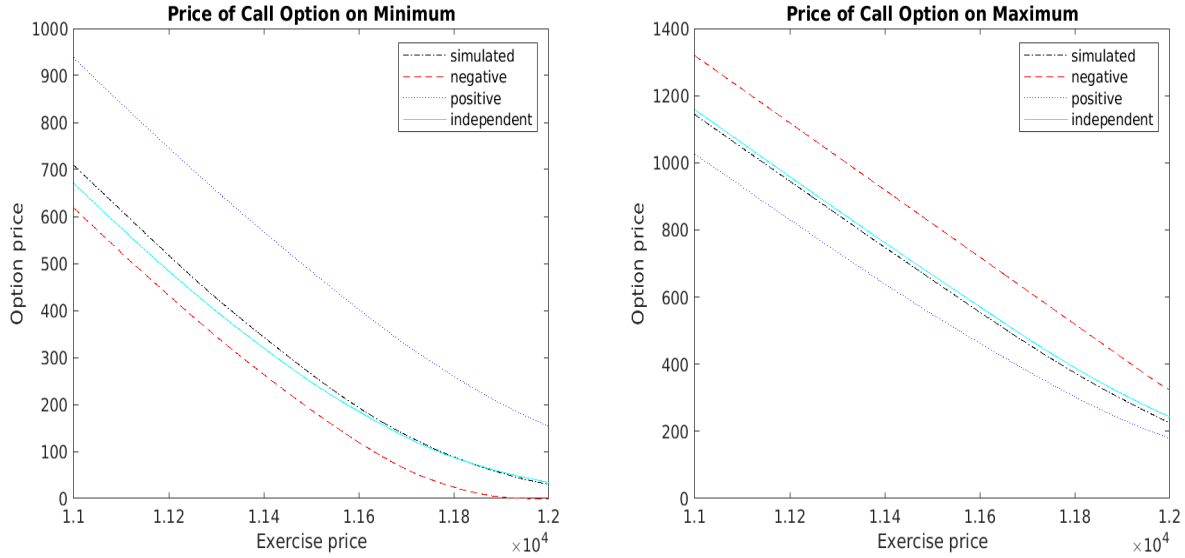

Figure 7: Simulated prices of call option on minimum(left) and maximum (right) payoff under different dependence scenarios.

Figure 7 displays the call option prices for the minimum and maximum of the two underlying

[2]Following Chiou and Tsay (2008), we use the Frank copula to construct these hypothetical dependence in our simulations.

indexes under these four scenarios. The results make intuitive sense. For the call option on minimum, the negative dependence is associated with the lowest option price because when the payoff is determined by the lower of two indexes, a negative dependence is particularly undesirable as one of the two tends to be low. In contrast, the highest option prices are observed under the positive dependence scenario because there is a chance that both indexes go up such that the lower of the two is still high. Analogously, for the maximum option, the highest/lowest option prices are observed under the negative/positively dependence.

6 Concluding remarks

This study develops a nonparametric Bayesian copula density estimator based on the method of Logistic Gaussian Process density estimation. A sampling method to facilitate sampling from the posterior copula distribution with ease and efficiency is also proposed. Monte Carlo simulations demonstrate the good overall and tail performance of this estimator, and the efficiency of the posterior sampler. Computer codes for our estimators will be made publicly available online. We expect the proposed methods to be a useful addition to the multivariate distribution estimation toolbox. Since the posterior distribution is not tractable, we use Laplace approximation for inference. We have experimented with Markov Chain Monte Carlo inference and found the scalability of the MCMC inference rather challenging. Further exploration in this direction may be worthwhile. We focus on bivariate copula in this study. In principle the estimators proposed here can be generalized to higher dimensions in a straightforward manner, but the computation cost may be prohibitive. A more practical alternative is to adopt the approach of vine-copulas to construct a high dimensional hierarchical copula with the bivariate LGP density as its building blocks. We leave these possible extensions to future studies.

References

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

Chen, X. and Fan, Y. (2006a). Estimation and model selection of semiparametric copula-based

multivariate dynamic models under copula misspecification. *Journal of econometrics*, 135(1-2):125–154.

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

Chen, X., Xiao, Z., and Wang, B. (2020). Copula-based time series with filtered nonstationarity. *Journal of Econometrics*.

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

Chiou, S. C. and Tsay, R. S. (2008). A copula-based approach to option pricing and risk assessment. *Journal of Data Science*, 6(3):273–301.

Creal, D. D. and Tsay, R. S. (2015). High dimensional dynamic stochastic copula models. *Journal of Econometrics*, 189(2):335–345.

Daniel, T. (1986). A bayesian approach to density estimation. *Biometrika*, 73(1):65–75.

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

Fan, Y. and Patton, A. J. (2014). Copulas in econometrics. *Annual Review of Economics*, 6(1):179–200.

Fermanian, J.-D. (2017). Recent developments in copula models. *Econometrics*, 5(3).

Fermanian, J.-D. and Scaillet, O. (2003). Nonparametric estimation of copulas for time series. *Journal of Risk*, 5(4):25–54.

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

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian data analysis*. CRC press.

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.

Geweke, J. (1989). Bayesian inference in econometric models using monte carlo integration. *Econometrica*, pages 1317–1339.

Geweke, J. (1996). Monte carlo simulation and numerical integration. volume 1 of *Handbook of Computational Economics*, pages 731–800. Elsevier.

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

Hjort, N. L. and Glad, I. K. (1995). Nonparametric density estimation with a parametric start. *The Annals of Statistics*, pages 882–904.

Joe, H. (1997). *Multivariate models and multivariate dependence concepts*. CRC Press.

Joe, H. (2014). *Dependence modeling with copulas*. CRC press.

Lee, T.-H. and Long, X. (2009). Copula-based multivariate garch model with uncorrelated dependent errors. *Journal of Econometrics*, 150(2):207–218.

Lenk, P. J. (1988). The logistic normal distribution for bayesian, nonparametric, predictive densities. *Journal of the American Statistical Association*, 83(402):509–516.

Lenk, P. J. (1991). Towards a practicable bayesian nonparametric density estimator. *Biometrika*, 78(3):531–543.

Leonard, T. (1978). Density estimation, stochastic processes and prior information. *Journal of the Royal Statistical Society: Series B (Methodological)*, 40(2):113–132.

Liu, J. S. (2008). *Monte Carlo strategies in scientific computing*. Springer Science & Business Media.

Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media.

Ning, S. and Shephard, N. (2018). A nonparametric bayesian approach to copula estimation. *Journal of Statistical Computation and Simulation*, 88(6):1081–1105.

Patton, A. J. (2009). *Copula–Based Models for Financial Time Series*, pages 767–785. Springer Berlin Heidelberg, Berlin, Heidelberg.

Rasmussen, C. E. and Williams, C. K. (2006). *Gaussian processes for machine learning*. MIT press Cambridge, MA.

Riihimäki, J., Vehtari, A., et al. (2014). Laplace approximation for logistic gaussian process density estimation and regression. *Bayesian analysis*, 9(2):425–448.

Rosco, J. and Joe, H. (2013). Measures of tail asymmetry for bivariate copulas. *Statistical Papers*, 54(3):709–726.

Rosenberg, J. V. (2003). Non-parametric pricing of multivariate contingent claims. *The Journal of Derivatives*, 10(3):9–26.

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8:229–231.

Smith, M. S. (2013). *Bayesian approaches to copula modelling*. Oxford University Press, Oxford.

Tokdar, S. T. (2007). Towards a faster implementation of density estimation with logistic gaussian process priors. *Journal of Computational and Graphical Statistics*, 16(3):633–655.

Tokdar, S. T. and Ghosh, J. K. (2007). Posterior consistency of logistic gaussian process priors in density estimation. *Journal of statistical planning and inference*, 137(1):34–42.

van der Vaart, A. W., van Zanten, J. H., et al. (2009). Adaptive bayesian estimation using a gaussian random field with inverse gamma bandwidth. *The Annals of Statistics*, 37(5B):2655–2675.

Wen, K. and Wu, X. (2020). Transformation-kernel estimation of copula densities. *Journal of Business & Economic Statistics*, 38(1):148–164.

Appendix

We gather in the appendix the derivations and technical details of some results in the text.

A Laplace Approximation

The Laplace method utilizes a Gaussian approximation $q(\boldsymbol{f}|\boldsymbol{Y},\boldsymbol{S},\boldsymbol{\theta})$ to the conditional posterior $p(\boldsymbol{f}|\boldsymbol{Y},\boldsymbol{S},\boldsymbol{\theta})$ in (7). Applying a second order Taylor expansion of $\log p(\boldsymbol{f}|\boldsymbol{Y},\boldsymbol{S},\boldsymbol{\theta})$ around the mode of the posterior, we obtain the Gaussian approximation (9).

Since the marginal likelihood is independent of $\boldsymbol{f}$, we only need to consider the un-normalized conditional posterior when maximizing w.r.t $\boldsymbol{f}$. So we define the objective function as

$$\begin{aligned}\Psi(\boldsymbol{f}) &= \log p(\boldsymbol{Y}|\boldsymbol{f}) + \log p(\boldsymbol{f}|\boldsymbol{S},\boldsymbol{\theta}) \\ &= \log p(\boldsymbol{Y}|\boldsymbol{f}) - \frac{1}{2}\boldsymbol{f}^T\boldsymbol{A}^{-1}\boldsymbol{f} - \frac{1}{2}\log|\boldsymbol{A}| - \frac{m^2}{2}\log 2\pi,\end{aligned}$$

where $\boldsymbol{A} = \boldsymbol{K} + \boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^T$ and m^2 is the total number of cells. Differentiating $\Psi(\boldsymbol{f})$ w.r.t $\boldsymbol{f}$ we obtain

$$\nabla\Psi(\boldsymbol{f}) = \nabla\log p(\boldsymbol{Y}|\boldsymbol{f}) - \boldsymbol{A}^{-1}\boldsymbol{f} \tag{A.1}$$

$$\nabla\nabla\Psi(\boldsymbol{f}) = \nabla\nabla\log p(\boldsymbol{Y}|\boldsymbol{f}) - \boldsymbol{A}^{-1} = -\boldsymbol{W} - \boldsymbol{A}^{-1}, \tag{A.2}$$

where $\boldsymbol{W} = -\nabla\nabla\log p(\boldsymbol{Y}|\boldsymbol{f})$.

Denote by $\hat{\boldsymbol{f}}$ the mode of $\Psi(\boldsymbol{f})$, we have

$$\nabla\Psi(\hat{\boldsymbol{f}}) = \boldsymbol{0} \Rightarrow \hat{\mathbf{f}} = \boldsymbol{A}\big(\nabla\log p(\boldsymbol{Y}|\hat{\boldsymbol{f}})\big).$$

We solve for $\hat{\boldsymbol{f}}$ via Newton's iteration

$$\begin{aligned}\boldsymbol{f}^{\text{new}} = \boldsymbol{f} - (\nabla\nabla\Psi)^{-1}\nabla\Psi &= \boldsymbol{f} + (\boldsymbol{W} + \boldsymbol{A}^{-1})^{-1}[\nabla\log p(\boldsymbol{Y}|\boldsymbol{f}) - \boldsymbol{A}^{-1}\boldsymbol{f}] \\ &= (\boldsymbol{W} + \boldsymbol{A}^{-1})^{-1}[\nabla\log p(\boldsymbol{y}|\boldsymbol{f}) + \boldsymbol{W}\boldsymbol{f}].\end{aligned}$$

Given the posterior mode $\hat{\boldsymbol{f}}$, the Laplace approximation to the conditional posterior can be written as a Gaussian distribution with mean $\hat{\boldsymbol{f}}$ and covariance matrix given by the negative

inverse Hessian of $\Psi(\boldsymbol{f})$, which is given in (9).

B Marginal Likelihood

The marginal likelihood, conditional on $\boldsymbol{\theta}$, is given by

$$p(\boldsymbol{Y}|\boldsymbol{S},\boldsymbol{\theta}) = \int p(\boldsymbol{Y}|\boldsymbol{f})p(\boldsymbol{f}|\boldsymbol{S},\boldsymbol{\theta})d\boldsymbol{f} = \int \exp\big(\Psi(\boldsymbol{f})\big)d\boldsymbol{f}.$$

We apply a Taylor expansion on $\Psi(\boldsymbol{f})$ locally around $\hat{\boldsymbol{f}}$, yielding $\Psi(\boldsymbol{f}) \simeq \Psi(\hat{\boldsymbol{f}}) - \frac{1}{2}(\boldsymbol{f}-\hat{\boldsymbol{f}})^T\boldsymbol{\Gamma}(\boldsymbol{f}-\hat{\boldsymbol{f}})$, where $\boldsymbol{\Gamma} = -\nabla\nabla\Psi(\boldsymbol{f})_{|\boldsymbol{f}=\hat{\boldsymbol{f}}}$. Thus we can approximate the marginal likelihood through

$$p(\boldsymbol{Y}|\boldsymbol{S},\boldsymbol{\theta}) \simeq q(\boldsymbol{Y}|\boldsymbol{S},\boldsymbol{\theta}) = \exp\big(\Psi(\hat{\boldsymbol{f}})\big)\int \exp\big(-\frac{1}{2}(\boldsymbol{f}-\hat{\boldsymbol{f}})^T\boldsymbol{\Gamma}(\boldsymbol{f}-\hat{\boldsymbol{f}})\big)d\boldsymbol{f}. \tag{B.1}$$

It turns out the integral could be evaluated analytically as

$$\int \exp\big(-\frac{1}{2}(\mathbf{f}-\hat{\mathbf{f}})^T\boldsymbol{\Gamma}(\mathbf{f}-\hat{\mathbf{f}})\big)d\mathbf{f} = \sqrt{(2\pi)^{m^2}|\boldsymbol{\Gamma}^{-1}|}. \tag{B.2}$$

Plugging (A.1), (A.2) and (B.2) into the approximated marginal likelihood (B.1), we obtain an approximation to the log marginal likelihood

$$\begin{aligned}\log q(\boldsymbol{Y}|\boldsymbol{S},\boldsymbol{\theta}) &= \log p(\boldsymbol{Y}|\hat{\boldsymbol{f}}) - \frac{1}{2}\hat{\boldsymbol{f}}^T\boldsymbol{A}^{-1}\hat{\boldsymbol{f}} - \frac{1}{2}\log|\boldsymbol{A}| - \frac{m^2}{2}\log 2\pi + \frac{m^2}{2}\log 2\pi - \frac{1}{2}\log|\boldsymbol{A}^{-1}+\boldsymbol{W}| \\ &= \log p(\boldsymbol{Y}|\hat{\boldsymbol{f}}) - \frac{1}{2}\hat{\boldsymbol{f}}^T\boldsymbol{A}^{-1}\hat{\boldsymbol{f}} - \frac{1}{2}\log|\boldsymbol{I}+\boldsymbol{W}^{1/2}\boldsymbol{A}\boldsymbol{W}^{1/2}|.\end{aligned}$$

C Optimization of approximating distribution

Below we describe the procedure to locate the coefficients for the bivariate approximating distribution $\mathcal{N}(\boldsymbol{s}|\tilde{\mu},\tilde{\Sigma})$ that minimize the sample ISE (14).

First write $\mu = \begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}$ and $\Sigma = \begin{bmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{bmatrix}$. To ensure the positive definiteness of Σ, an unconstrained optimization is conducted with respect to $(\mu_1, \mu_2, \log\sigma_1^2, \log\sigma_2^2, \Phi^{-1}(\rho/2+1/2))$ using Newton's method. The initial values of (μ_1,μ_2) is set at the mode of the posterior distribution of $g(\boldsymbol{s})$, and those for $(\rho,\sigma_1^2,\sigma_2^2)$ are set according to the covariance of the transformed observations.

Convergence is obtained typically after a few iterations. A small positive diagonal 'ridge' is added to $\tilde{\Sigma}$, if needed, to ensure it is positive definite and well-conditioned.