HEC MONTRÉAL

Estimation et test d'adéquation pour des modèles de copule à changement de régime, avec application

par

Mamadou Yamar Thioub

superviseur

Bruno N. Rémillard

co-superviseur

Bouchra R. Nasri

Sciences de la gestion
(Ingénierie financière)

Mémoire présenté en vue de l'obtention
du grade de maîtrise ès sciences en gestion
(M.Sc.)

Décembre 2018

Résumé

Considérons plusieurs séries temporelles univariées, et pour chacune supposons que l'on choisisse un modèle dynamique paramétrique approprié. On obtient alors des termes d'erreurs sériellement indépendants. L'objectif principal de ce mémoire est de modéliser la dépendance entre ces termes d'erreur par des copules avec changement de régime, ce qui est une nette amélioration des méthodes habituellement utilisées en pratique, car dans notre cadre, la dépendance peut varier dans le temps. De plus, nous présentons une méthode d'estimation des paramètres du modèle, et nous proposons un test d'adéquation qui permettra de sélectionner le nombre de régimes optimal. Comme application, nous évaluons le prix d'une option de vente européenne sur le rendement maximal de deux titres en tenant d'un modèle de copule à changement de régime. Finalement, afin de faciliter l'utilisation future de la méthodologie proposée, nous avons construit une librairie de fonctions basée sur le progiciel R, qui s'intitule HMMcopula, et qui est disponible gratuitement sur CRAN.

Mots-clés: Test d'adéquation; séries temporelles; copules; modèles dynamiques.

Abstract

Consider several time series and for each of them, one fits an appropriate dynamic parametric model. This produces serially independent error terms for each time series. The main purpose of this thesis is to model the dependence between these error terms using regime switching copula models which is an improvement over existing models since this setting allows for a time-varying dependence. Moreover, we present a method for estimating the parameters, and we propose a goodness-of-fit test that can be used to select the optimal number of regimes. An application is proposed to evaluate an European put-on-max option on the returns of two assests. Finally, in order to facilitate the use of our methodology, we have built a R package HMMcopula which is available on CRAN.

Keywords: Goodness-of-fit; time series; copulas; dynamic models; generalized error models.

Table des matières

Liste des figures

Liste des tableaux

Remerciements

Ce travail de recherche n'aurait pas pu être complété sans l'immense soutien du professeur Bruno Rémillard et de la Dre Bouchra Nasri, qui m'ont offert leur soutien. Je voudrais également remercier les professeurs Jean-François Plante et Chantal Labbé pour avoir eu la gentillesse d'arbitrer mon mémoire. Je tiens aussi à exprimer ma gratitude à ma famille, qui a été très compréhensive et m'a aidé tout au long de mes études.

Introduction

La modélisation de la dépendance entre les séries temporelles est un problème très important. Considérons l'exemple de la crise financière de 2008, où la contagion liée à la dépendance entre plusieurs actifs risqués a été clairement sous-estimée. De plus, il est naturel que la dépendance entre les séries temporelles varie dans le temps, et qu'elle devient potentiellement plus élevée en période de crise par exemple. Quelques méthodes de modélisation de la dépendance temporelle ont été proposées. Récemment, Adams et al. (2017) ont utilisé des modèles de corrélation conditionnelle dynamique DCC-GARCH, Engle (2002) afin de modéliser la dépendance entre des séries temporelles multivariées. L'un des points faibles de ces modèles est qu'ils utilisent la corrélation, ce qui n'est pas nécessairement une bonne façon d'estimer la dépendance. Afin de pallier ce problème nous proposons d'utiliser les copules.

À notre connaissance, le premier article traitant de copule dépendante du temps est celui de van den Goorbergh et al. (2005). Dans ce travail, les auteurs ont utilisé des copules dont le paramètre est une fonction des volatilités, afin de modéliser la dépendance entre les innovations de deux séries temporelles suivant un modèle dynamique de type GARCH. Leur modèle est un cas particulier du single-index copula récemment développé par Fermanian and Lopez (2015). Une autre approche est celle proposée par Nasri et al. (2017). Dans ce cas, les paramètres de la copule entre la variable dépendante et les variables indépendantes ne dépendent que d'une fonction du temps. Finalement, Fink et al. (2017) ont proposé des copules à changement de régime. Ils ont aussi appliqué leur modèle aux innovations de séries univariées de type GARCH. Cependant, ils n'ont pas proposé de test d'adéquation formel, et leur sélection du nombre de régimes est basée sur la comparaison des vraisemblances, ce qui n'est pas recommandé, Cappé et al. (2005), à cause de la présence possible de singularités et de paramètres non-identifiables.

En se basant sur les travaux de Fink et al. (2017) et Nasri et al. (2017), nous proposons un test d'adéquation pour les copules à changement de régime, ce qui n'a pas encore été fait à notre connaissance. Nous obtenons ainsi un moyen intéressant de sélectionner le nombre de régimes adéquat. Notons que l'on considère aussi des modèles dymaniques univariés plus généraux que les modèles à volatilité stochastique généralement utilisés dans la littérature.

Le reste de ce mémoire, basé sur un article soumit, est organisé comme suit. À la section 1.2, nous décrivons le modèle des séries temporelles et nous définissons les copules à changement de régime. À la section 1.3, nous donnons les détails de l'estimation des paramètres, nous proposons un test d'adéquation, ainsi qu'une procédure de sélection du nombre de régimes. À la section 1.4, nous proposons une application pour l'évaluation des options, similairement à celle proposée dans van den Goorbergh et al. (2005). Finalement, nous avons construit une librairie de fonctions basée sur le progiciel R, qui s'intitule HMMcopula. Cette librairie est disponible sur CRAN. Sa documentation constitue le deuxième chapitre de ce mémoire.

Chapter 1

Estimation and goodness-of-fit for regime switching copula models with application

Abstract

The goal of this paper is to propose inference methods for regime switching copula models. We present an algorithm for the estimation of parameters and we propose a goodness-of-fit test for the adequacy of these models. More precisely, it is assumed that we have a multivariate time series, and for each univariate time series, we fit a generalized error model, producing estimations of independent error terms. In fact, we assume that the copula associated with the error terms is a regime-switching copula. We show how to estimate the parameters of the model copula and we propose a goodness-if-fit test. The latter is then used to select the number of regimes. An application is proposed to evaluate an European put-on-max option on the returns of two assests. Finally, in order

to facilitate the use of our methodology, we have built a R package HMMcopula which is available on CRAN.

1.1 Introduction

Modeling dependence between time series in a very important problem. Just consider for example the 2008 financial crisis, where contagion, related to the dependence between several risky assets, was clearly underestimated. Also, we must take into account the fact that the dependence may vary with time, potentially increasing in crisis periods.

Some ways to take into account time-varying dependence have been proposed. Recently, Adams et al. (2017) fitted DCC-GARCH models (Engle, 2002) to multivariate time series, which is a bit restrictive in terms of dependence since it is based on the multivatiate Gaussian distribution. To overcome this limitation, and because copulas are specially designed to model dependence, it is no wonder that many time-varying dependence models are based on copulas.

To our knowledge, one of the first papers involving time-dependent copulas was van den Goorbergh et al. (2005), where the authors fitted a copula family to the residuals of two GARCH time series, with a parameter expressed as a function of the volatilities. Note that this is a special case of what is now known as single-index copula (Fermanian and Lopez, 2015). One can also use the methodology proposed in Nasri et al. (2017), where the copula between a response variable and covariates has parameters depending on a function of time only. Another possibility is to use regime-switching copulas. This has been proposed by Fink et al. (2017), who fitted their model to residuals of univariate GARCH series. They did not propose a formal test of goodness-of-fit, and the selection of the number of regimes is based on comparisons of likelihoods, which is not recommended (Cappé et al., 2005) as a selection method, due to possible singularities and non-identifiable parameters.

Building on Fink et al. (2017) and Rémillard et al. (2017), we propose a goodness-of-fit test for regime-switching copulas, which was not done before in this setting. As a by-product, we obtain an interesting way to select the number of regimes.

More precisely, in Section 1.2, we describe the model for the time series and we define regime-switching copulas. In Section 1.3, we detail the estimation procedure, the goodness-of-fit test and the selection of the number of regimes. In Section 1.4, we give an example of application for option pricing, along the same lines as van den Goorbergh et al. (2005). We have built a R package for regime switching copula models, HMMcopula available at CRAN, the documentation of the package is shown at the end of this document.

1.2 Multivariate time series and regime-switching copulas

When copula-based models are used for modeling dependence between many time series, individual series are often estimated first, and a copula is used to capture the dependence between serially independent innovations; see, e.g., van den Goorbergh et al. (2005), Chen and Fan (2006), Patton (2006), Rémillard (2017).

To fix ideas, let $\mathbf{X}_t = (X_{1t}, \ldots, X_{dt})$, be a multivariate time series. For each $j \in \{1, \ldots, d\}$, let $\mathcal{F}_{j,t-1}$ contains information from the past of $X_{j1}, \ldots X_{j,t-1}$, and possibly information from covariates as well. Further set $\mathcal{F}_t = \vee_{j=1}^{d} \mathcal{F}_{j,t}$.

In many papers, the serial dependence in each univariate time series is modeled by a GARCH-type model of the form

$$X_{jt} = \mu_{jt}(\boldsymbol{\alpha}) + \sigma_{jt}(\boldsymbol{\alpha})\varepsilon_{jt}, \quad \varepsilon_{jt} \sim F_j, \quad j \in \{1, \dots, d\}, \tag{1.1}$$

where μ_{jt} and σ_{jt} are $\mathcal{F}_{j,t-1}$-measurable, and the innovations ε_{jt} are independent of $\mathcal{F}_{j,t-1}$. Furthermore, for a fixed $j \in \{1, \dots, d\}$, the univariate innovations are i.i.d. This is the case in van den Goorbergh et al. (2005), where they assumed that the copula of $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \dots, \varepsilon_{dt})^\top$ given $\mathcal{F}_{t-1}$ was of the form $C_{\boldsymbol{\theta}_t}$, for a known copula family $C_{\boldsymbol{\theta}}$, and $\boldsymbol{\theta}_t$ is a function of $\boldsymbol{\sigma}_t = (\sigma_{1t}, \dots, \sigma_{dt})$.

However not all dynamic time series models have innovations. Take for example univariate time series modeled by hidden Markov models. In order to cover most dynamic models, we proceed as in Nasri and Remillard (2018). First, each univariate time series is a "generalized error model" (Du, 2016). This means that for each $j \in \{1, \dots, d\}$, there exists $G_{\boldsymbol{\alpha},jt}$, a continuous, increasing , and $\mathcal{F}_{j,t-1}$-measurable function so that $\varepsilon_{jt} = G_{\boldsymbol{\alpha},jt}(X_{jt})$ are i.i.d. with continuous distribution function F_j and density f_j, for some $\boldsymbol{\alpha} \in \mathcal{A}$. Note that these dynamic models contain the innovation models, the HMM models, and the models in Bai (2003) as particular cases. Second, to introduce the dependence between the time series, choose a sequence of $\mathcal{F}_{t-1}$-measurable copulas C_t, so that the joint conditional distribution function K_t of $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \dots, \varepsilon_{dt})$ given $\mathcal{F}_{t-1}$ is

$$K_t(\mathbf{x}) = C_t\{\mathbf{F}(\mathbf{x})\}, \quad \mathbf{F}(\mathbf{x}) = (F_1(x_1), \dots, F_d(x_d))^\top, \quad \mathbf{x} = (x_1, \dots, x_d)^\top \in \mathbb{R}^d. \tag{1.2}$$

In fact, $\mathbf{U}_t = \mathbf{F}(\boldsymbol{\varepsilon}_t) \sim C_t$, for every $t \in \{1, \dots, n\}$.

This way of modeling dependence between several time series is usually applied to innovations of stochastic volatility models as described by (1.1). In this case, one could take $G_{\boldsymbol{\alpha},jt}(x_j) = \frac{x_j - \mu_{jt}(\boldsymbol{\alpha})}{\sigma_{jt}(\boldsymbol{\alpha})}$, $j \in \{1, \ldots, d\}$, and then $\varepsilon_{jt} = G_{\boldsymbol{\alpha},jt}(X_{jt}) \sim F_j$. We could also take $G_{\boldsymbol{\alpha},jt}(x_j) = F_j\left(\frac{x_j - \mu_{jt}(\boldsymbol{\alpha})}{\sigma_{jt}(\boldsymbol{\alpha})}\right)$, $j \in \{1, \ldots, d\}$. In this case, the distribution of the error terms $\varepsilon_{jt} = G_{\boldsymbol{\alpha},jt}(X_{jt})$ is uniform, and we recover the setting of Bai (2003).

One could also consider Gaussian HMM models for univariate time series, say X_{1t}. This means that there exists a Markov chain $\tau_t^{(1)}$ on $\{1, \ldots, m\}$ with transition matrix Q so that given $\tau_1^{(1)} = i_1, \ldots, \tau_n^{(1)} = i_n$, $X_{11}, \ldots, X_{1n}$ are independent, and $X_{1t} \sim N(\mu_{i_t}, \sigma_{it}^2)$.

According to Rémillard et al. (2017), if $\eta_{t-1}(k)$ is the probability of being in regime $k \in \{1, \ldots, m\}$ at time $t-1$ given the past observations $X_{11}, \ldots, X_{1,t-1}$, then the conditional distribution F_t of X_{1t} given the past is

$$F_t(x) = \sum_{k=1}^{m} F^{(k)}(x) W_{t-1}(k), \tag{1.3}$$

with density $f_t(x) = \sum_{k=1}^{m} f^{(k)}(x) W_{t-1}(k)$, where $W_{t-1}(k) = \sum_{j=1}^{m} \eta_{t-1}(j) Q_{jk}$, $\eta_{t-1}(j)$ is the probability of being in regime j at time $t-1$ given the past observations, as given by formula (1.9), and $f^{(k)}$ is the density of a Gaussian distribution with mean μ_k and variance σ_k^2. It then follows that the sequence $U_{1t} = F_t(X_{1t})$ are i.i.d. uniform random variables.

After choosing the generalized error models for the univariate time series, it is time to choose the copula model for the series $\mathbf{U}_t$. From now on, one assumes that the time series $\mathbf{U}_t$ follows a HMM copula model. This means that there exists a finite Markov chain τ_t on $\{1, \ldots, \ell\}$ with transition matrix P, so that given $\tau_1 = i_1, \ldots, \tau_n = i_n$, $\mathbf{U}_1, \ldots, \mathbf{U}_n$

are independent, and $\mathbf{U}_t \sim C_{\boldsymbol{\beta}_{i_t}}$, $t \in \{1,\dots,n\}$, where $C_{\boldsymbol{\beta}}$, $\boldsymbol{\beta} \in \mathcal{B}$, is a given parametric copula family. Also one assumes the usual smoothness conditions on the associated densities $c_{\boldsymbol{\beta}}$ so that the maximum likelihood estimator exists.

Note that for a given $j \in \{1,\dots,d\}$, one needs that the values U_{jt}, $t \in \{1,\dots,n\}$, are i.i.d. uniform. This is indeed true as proven in the following theorem.

Theorem 1. *Suppose that the multivariate time series $\mathbf{U}_t$ is generated from a regime-switching copula model, with copulas $C_{\boldsymbol{\beta}_k}$, $k \in \{1,\dots,\ell\}$, and transition matrix P. For a given $j \in \{1,\dots,d\}$, the values U_{jt}, $t \in \{1,\dots,n\}$, are i.i.d. uniform.*

Proof. For simplicity, suppose that $j = 1$. According to formula (1.3),

$$\mathbb{P}(U_{1t} \le u_1,\dots,U_{dt} \le u_d | \mathbf{U}_1,\dots,\mathbf{U}_{t-1}) = \sum_{k=1}^{\ell} W_{t-1}(k) C_{\boldsymbol{\beta}_k}(u_1,\dots,u_d).$$

From the properties of copulas, one gets that

$$\mathbb{P}(U_{1t} \le u_1 | \mathbf{U}_1,\dots,\mathbf{U}_{t-1}) = \sum_{k=1}^{\ell} W_{t-1}(k) C_{\boldsymbol{\beta}_k}(u_1,1,\dots,1) = \sum_{k=1}^{\ell} W_{t-1}(k) u_1 = u_1.$$

As a result, one may conclude that $U_{11},\dots,U_{1n}$ are i.i.d. uniform. □

Since the generalized errors $\boldsymbol{\varepsilon}_t$ are not observable, $\boldsymbol{\alpha}$ being unknown, the latter must be estimated by a consistent estimator $\boldsymbol{\alpha}_n$. One can then compute the pseudo-observations $\mathbf{e}_{n,t} = (e_{n,1t},\dots,e_{n,dt})^\top = \mathbf{G}_{\boldsymbol{\alpha}_n,t}(\mathbf{X}_t)$, where $e_{n,jt} = G_{\boldsymbol{\alpha}_n,jt}(X_{jt})$, $j \in \{1,\dots,d\}$ and $t \in \{1,\dots,n\}$. Using these pseudo-observations might be a problem, but in Nasri and Remillard (2018), it was shown that using the normalized ranks of these pseudo-observations, one can estimate the parameters $\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_\ell$ and P, as if one was observing

$\mathbf{U}_1, \ldots, \mathbf{U}_n$. The same applies to the goodness-of-fit test that will be defined in Section 1.3.2.

Before ending this section, one needs to detail how one can simulate the multivariate time series.

1.2.1 Simulation of the multivariate time series

To simulate the multivariate time series, it suffices to generate $\mathbf{U}_t = (U_{1t}, \ldots, U_{dt})$ according to the regime-switching copula model, set $\varepsilon_{jt} = F_j^{-1}(U_{jt})$, and then compute $X_{jt} = G_{jt}^{-1}(\varepsilon_{jt})$, $j \in \{1, \ldots, d\}$, and $t \in \{1, \ldots, n\}$. It then follows that

$$P(X_{jt} \leq x_j | \mathcal{F}_{t-1}) = P(X_{jt} \leq x_j | \mathcal{F}_{j,t-1}) = F_j \left\{ G_{jt}(x_j) \right\}, \quad j \in \{1, \ldots, d\}, \tag{1.4}$$

which is a natural condition to impose; otherwise it would not be possible to estimate the parameters of a given model for X_{1t} without observing the other series. See, e.g., Fermanian and Wegkamp (2012).

1.3 Estimation and goodness-of-fit test

We first present general regime-switching which can be applied to univariate time series or copula. Then, we describe an estimation procedure and goodness-of fit for regime-switching copula models. Finally, we propose a procedure for selecting the optimal number of regimes.

1.3.1 General regime-switching models

Let τ_t be a homogeneous discrete-time Markov chain on $S = \{1, \ldots, \ell\}$, with transition probability matrix P on $S \times S$. Given $\tau_1 = k_1, \ldots, \tau_n = k_n$, the observations $\mathbf{Y} = (Y_1, \ldots, Y_n)$ are independent with densities $g_{\boldsymbol{\beta}_{k_t}}$, $t \in \{1, \ldots, n\}$.

Set $\boldsymbol{\theta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_\ell, P)$. Then the joint density of $\tau = (\tau_1, \ldots, \tau_n)$ and $\mathbf{Y}$ is

$$f_{\boldsymbol{\theta}}(\tau, \mathbf{Y}) = \left(\prod_{t=1}^{n} P_{\tau_{t-1}, \tau_t} \right) \times \prod_{t=1}^{n} g_{\boldsymbol{\beta}_{\tau_t}}(Y_t), \tag{1.5}$$

so one can write

$$\log f_{\theta}(\tau, \mathbf{Y}) = \sum_{t=1}^{n} \log P_{\tau_{t-1}, \tau_t} + \sum_{t=1}^{n} \log g_{\boldsymbol{\beta}_{\tau_t}}(Y_t). \tag{1.6}$$

Because the regimes τ_t are not observable, an easy way to estimate the parameters is to use the EM algorithm (Dempster et al., 1977), which proceeds in two steps: expectation (E step), where $Q_{\mathbf{y}}(\tilde{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}}\{\log f_{\tilde{\boldsymbol{\theta}}}(\tau, \mathbf{Y}) | \mathbf{Y} = \mathbf{y}\}$ is computed, and maximization (M step), where one computes

$$\boldsymbol{\theta}^{(k+1)} = \arg\max_{\boldsymbol{\theta}} Q_y\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}\right),$$

starting from an initial value $\boldsymbol{\theta}^{(0)}$. As $k \to \infty$, $\boldsymbol{\theta}^{(k)}$ converges to the maximum likelihood estimator of the density of $\mathbf{Y}$. The formulas for the E-M steps are given in Appendix 1.6.

As a particular case of regime-switching models, if $P_{ij} = \nu_j$, then one gets mixture models. In this case $\tau_1, \ldots, \tau_n$ are i.i.d. The simplified formulas for the E-M steps are also given in Appendix 1.6.

Application to copulas

For the regime-switching or mixture models, the density $g_{\boldsymbol{\beta}}$ is the density of a parametric family of copulas $C_{\boldsymbol{\beta}}$, with $\boldsymbol{\beta} \in \mathcal{B}$. However $Y_1, \ldots, Y_n$ are not observable so they must be replaced by the normalized ranks of the pseudo-observations $\mathbf{e}_{n,t}$, i.e., $Y_{jt} = \text{rank}(e_{n,jt})/(n+1)$. As proven in Nasri and Remillard (2018), taking the normalized ranks of pseudo-observations do not affect the asymptotic properties of the estimators.

1.3.2 Goodness-of-fit

In this section, we propose a methodology to perform a goodness-of-fit test on a multivariate time series, by using the Rosenblatt's transform, as defined next.

First, following Rémillard (2013), under the general regime-switching model described in Section 1.3.1, the conditional density f_t of Y_t given $Y_1, \ldots, Y_{t-1}$ can be expressed as a mixture via.

$$f_t(y_t|y_1, \ldots, y_{t-1}) = \sum_{i=1}^{\ell} f^{(i)}(y_t) \sum_{j=1}^{\ell} \eta_{t-1}(j) P_{ji} = \sum_{i=1}^{\ell} f^{(i)}(y_t) W_{t-1}(i), \tag{1.7}$$

where $f^{(i)} = g_{\boldsymbol{\beta}_i}$ and

$$W_{t-1}(i) = \sum_{j=1}^{\ell} \eta_{t-1}(j) P_{ji}, \quad i \in \{1 \ldots \ell\}, \tag{1.8}$$

$$\eta_t(j) = \frac{f^{(j)}(y_t)}{Z_{t|t-1}} \sum_{i=1}^{\ell} \eta_{t-1}(i) P_{ij}, \quad j \in \{1, \ldots, \ell\}, \tag{1.9}$$

$$Z_{t|t-1} = \sum_{j=1}^{\ell} f^{(j)}(y_t) \sum_{i=1}^{\ell} \eta_{t-1}(i) P_{ij}. \tag{1.10}$$

Note that formulas (1.7)–(1.10) also hold for univariate Gaussian HMM; in this case, $f^{(j)}$ is the Gaussian density with mean μ_j and variance σ_j^2.

Next, let $i \in \{1, \ldots, \ell\}$ be fixed and suppose that $Z = (Z_1, \ldots, Z_d)$ has density $f^{(i)}$. For any $q \in \{1, \ldots, d\}$, denote by $f_{1:q}^{(i)}$ the density of $(Z_1, \ldots, Z_q)$. Also, let $f_q^{(i)}$ be the conditional density of Z_q given $Z_1, \ldots, Z_{q-1}$. Further denote by $F_q^{(i)}$ the distribution function corresponding to density $f_q^{(i)}$.

The Rosenblatt's transform Ψ_t corresponding to the density (1.7) conditional on $y_1, \ldots, y_{t-1} \in \mathbb{R}^d$ is given by

$$\Psi_t^{(1)}(y_{1t}) = \sum_{i=1}^{\ell} W_{t-1}(i) F_1^{(i)}(y_{1t}), \tag{1.11}$$

and for $q \in \{2, \ldots, d\}$,

$$\Psi_t^{(q)}(y_{1t}, \ldots, y_{qt}) = \frac{\sum_{i=1}^{\ell} W_{t-1}(i) f_{1:q-1}^{(i)}(y_{1t}, \ldots, y_{q-1,t}) F_q^{(i)}(y_{tq})}{\sum_{i=1}^{\ell} W_{t-1}(i) f_{1:q-1}^{(i)}(y_{1t}, \ldots, y_{q-1,t})}. \tag{1.12}$$

Suppose now that $\mathbf{U}_1, \ldots, \mathbf{U}_n$ is a size n sample of a d-dimensional vector drawn from a joint continuous distribution $\mathbf{P}$ belonging to a parametric family of regime-switching copula models with ℓ regimes. Formally, the hypothesis to be tested is

$$\mathcal{H}_0 : \mathbf{P} \in \mathcal{P} = \{\mathbf{P}_\theta; \theta \in \mathcal{O}\} \quad vs \quad \mathcal{H}_0 : \mathbf{P} \notin \mathcal{P}.$$

Under the null hypothesis, it follows that,

$$\mathbf{V}_1 = \Psi_1(\mathbf{U}_1, \theta), \mathbf{V}_2 = \Psi_2(\mathbf{U}_1, \mathbf{U}_2, \theta), \ldots, \mathbf{V}_n = \Psi_n(\mathbf{U}_1, \ldots, \mathbf{U}_n, \theta)$$

are independent and uniformly distributed over $(0,1)^d$, where $\Psi_1(\cdot, \theta), \ldots, \Psi_n(\cdot, \theta)$ are the Rosenblatt's transforms for the true parameters $\theta \in \mathcal{O}$.

However, $\boldsymbol{\theta}$ must be estimated, say by $\boldsymbol{\theta}_n$. Also, the random vectors $\mathbf{U}_1, \ldots, \mathbf{U}_n$ are not observable, so they must be replaced by the normalized ranks $\mathbf{u}_{n,t}$ of the pseudo-observations $\mathbf{e}_{n,t}$, $t \in \{1, \ldots, n\}$. Then, the pseudo-observations $\mathbf{V}_{n,t} = \Psi_t(\mathbf{u}_{n,t}, \boldsymbol{\theta}_n)$ are no longer i.i.d., however they are close to be i.i.d.

Test statistics

Following Nasri and Remillard (2018), define the empirical process

$$D_n(\mathbf{u}) = \frac{1}{n} \sum_{t=1}^{n} \prod_{j=1}^{d} \mathbb{1}\left(V_{n,jt} \le u_j\right), \quad \mathbf{u} = (u_1, \ldots, u_d) \in [0,1]^d. \tag{1.13}$$

Following, Genest et al. (2009), to test $\mathcal{H}_0$ against $\mathcal{H}_1$, it is suggested to use the Cramér-von Mises type statistic because it seems to be much more powerful and easier to compute than the Kolmogorov-Smirnov type statistic. The Cramér-von Mises type statistic is given by:

$$S_n = B_n(\mathbf{V}_{n,1}, \ldots, \mathbf{V}_{n,n}) = n \int_{[0,1]^d} \left\{ D_n(\mathbf{u}) - \prod_{j=1}^{d} u_j \right\}^2 d\mathbf{u}.$$

As a result,

$$S_n = \frac{1}{n} \sum_{t=1}^{n} \sum_{i=1}^{n} \prod_{q=1}^{d} \{1 - \max(V_{n,qt}, V_{n,qi})\} - \frac{1}{2^{d-1}} \sum_{t=1}^{n} \prod_{q=1}^{d} \left(1 - V_{n,qt}^2\right) + \frac{n}{3^d}. \quad (1.14)$$

We can interpret S_n as the distance of our empirical distribution and the independence copula. Since $\mathbf{V}_{n,t}$, $t \in \{1, \ldots, n\}$ are almost uniformly distributed over $[0,1]^d$ under $\mathcal{H}_0$, large values of S_n lead to the rejection of the null hypothesis. Unfortunately, the limiting distribution of the test statistic will depend on the unknown parameter $\boldsymbol{\theta}$, but it does not depend on the estimated parameters of the univariate time series. This remarkable result was proven in Nasri and Remillard (2018). Since it is impossible to construct the tables, we use the parametric bootstrap, to compute P-values. Its validity has been proven for a wide range of assumptions in Nasri and Remillard (2018).

Parametric bootstrap

Simply stated, if a goodness-of-fit test is based on a statistic S_n of the observations $R_1, \ldots, R_n$ with distribution $\mathbf{P}_\theta$, for some unknown parameter θ estimated by θ_n, the parametric bootstrap approach consists in generating a large number B of sequences $Y_1^{(k)}, \ldots, Y_n^{(k)}$, with distribution $\mathbf{P}_{\theta_n}$, $k = 1, \ldots, B$, evaluating each time the goodness-of-fit statistic $S_n^{(k)}$, and then approximating the p-value by the percentage of value $S_n^{(k)}$ greater than S_n, assuming that the null hypothesis is rejected for large values of S_n.

Hence, in order to perform the goodness of fit test, we use the following algorithm:

Algorithm 1. *For a given number of regimes ℓ, get estimator $\boldsymbol{\theta}_n$ of $\boldsymbol{\theta}$ using the EM algorithm described in 1.3.1, applied to the pseudo-observations $u_{n,t}$, $t \in \{1, \ldots, n\}$. Then compute the statistic $S_n = B_n(\mathbf{V}_{n,1}, \ldots, \mathbf{V}_{n,n})$, using the pseudo-observations $\mathbf{V}_{n,t} = \Psi_t(\mathbf{u}_{n,t}, \boldsymbol{\theta}_n)$, $t \in \{1, \ldots, n\}$.*

Then for $k = 1, \ldots, B$, B large enough, repeat the following steps:

- *Generate a random sample $\mathbf{U}_1^*, \ldots, \mathbf{U}_n^*$ from distribution $\boldsymbol{P}_{\boldsymbol{\theta}_n}$, i.e., from a regime-switching copula model with parameter $\boldsymbol{\theta}_n$.*
- *Get the estimator $\boldsymbol{\theta}_n^*$ from $\mathbf{U}_1^*, \ldots, \mathbf{U}_n^*$.*
- *Compute the normalized ranks $\mathbf{u}_{n,1}^*, \ldots, \mathbf{u}_{n,n}^*$ from $\mathbf{U}_1^*, \ldots, \mathbf{U}_n^*$.*
- *Compute the pseudo-observations $\mathbf{V}_{n,t}^* = \boldsymbol{\Psi}_t\left(\mathbf{u}_{n,t}^*, \boldsymbol{\theta}_n^*\right)$, $t \in \{1, \ldots, n\}$ and calculate $S_n^{(k)} = B_n\big(\mathbf{V}_{n,1}^*, \ldots, \mathbf{V}_{n,n}^*\big)$.*

Then, an approximate p-value for the test based on Cramér-von Mises statistic S_n is given by

$$\frac{1}{B} \sum_{k=1}^{B} \mathbb{1}\big(S_n^{(k)} > S_n\big).$$

Choosing the number of regimes

The goodness-of-fit test methodology produces p-value from a Cramér-von Mises type statistic, for a given number of regimes ℓ. As suggested in Rémillard et al. (2017), it makes sense to choose the optimal number of regimes, ℓ^*, as the first ℓ for which the p-value is larger than 5%.

In some articles, it is suggested to used a criterion based on likelihoods to choose the number of regimes, but Cappé et al. (2005) strongly advised against it.

1.4 Application to option pricing

In this application, we want to model dependence between the daily returns on the Amazon (amzn) and the Apple (aapl) from January 1, 2015 to June 29, 2018. The sample size is 881 observations for each time series. The daily log returns of the stocks are shown in Figure 1.1.

The first step is to fit dynamic models to both time series. To this end, we choose to fit HMM Gaussian models. Using the selection procedure described in Section 1.3.2, we obtain a Gaussian HMM with 3 regimes for the the daily log-returns of Amazon, while for the daily log-returns of Apple, we obtain a Gaussian HMM with 3 regimes. Here, the p-values are 38.8% and 15.1% respectively, using 1000 bootstrap samples. The estimated parameters for both time series are given in Table 1.1, while a graph of the normalized ranks $\mathbf{u}_{n,t}$ is given in Figure 1.2.

Table 1.1: Estimated parameters for the log-returns of Amazon and Apple, using Gaussian HMM. Here, ν is the stationary distribution of the regimes.

	Amazon			Apple		
	Regime			Regime		
Parameter	1	2	3	1	2	3
$\mu \times 10^{-2}$	4.41	0.19	-0.12	0.043277	-0.28	0.22
$\sigma \times 10^{-3}$	3.3	0.10268	0.52519	0.42624	0.0061731	0.077751
ν	0.0201	0.7198	0.2601	0.3934	0.1007	0.5059
Q	$\begin{pmatrix} 0.1572 & 0.4450 & 0.3978 \\ 0.0038 & 0.9662 & 0.03 \\ 0.0545 & 0.0606 & 0.8849 \end{pmatrix}$			$\begin{pmatrix} 0.8788 & 0.0001 & 0.1211 \\ 0.4154 & 0.0674 & 0.5172 \\ 0.0098 & 0.1859 & 0.8043 \end{pmatrix}$		

From now on, let X_{1t} denotes the returns of Amazon and let X_{2t} denotes the returns of Apple. Further let F_{1t} and F_{2t} be the conditional distributions of X_{1t} and X_{2t} given the past observations, as defined by (1.3). Further set $U_{1t} = F_{1t}(X_{1t})$ and $U_{2t} = F_{2t}(X_{2t})$. As defined in Section 1.3, let $u_{n,jt} = F_{n,jt}(X_{jt})$, $j = 1, 2$, be pseudo-observations, where $F_{n,jt}$ is the conditional distribution function computed with the parameters of Table 1.1.

We perform goodness of fit test, based on $B = 1000$ replications to select the proper number of regimes and the copula family that better describe the dependence structure between the series. The results are shown in the Table 1.2. Hence the dependence structure of the data is best describe by regime-switching Gaussian copula with two regimes. The estimated parameters of the copula are shown in the Table 1.3.

Table 1.2: P-value (as percentage) of the different copula families, HMM.

Copula	One regime	Two regimes
Gaussian	0	9.9
Clayton	0	0.8
Gumbel	0	0
Frank	0.4	0.1
Student	4.3	11

1.4.1 Bivariate option pricing

In order to price an option with payoff Φ over n trading days, one performs a Monte Carlo simulation of size N. The following steps illustrate the procedure in the case of a general regime-switching copula with ℓ regimes, where each univariate time series is modeled by a Gaussian HMM with ℓ_j regimes and parameters $\mu_{j1}, \ldots, \mu_{j\ell_j}$, $\sigma_{j1}, \ldots, \sigma_{j\ell_j}$, $P^{(j)}$.

Table 1.3: Estimated parameters for the regime-switching Gaussian copula with two regimes. Here, ν is the stationary distribution of the regimes.

Parameter	Regime 1	Regime 2
τ	0.0859	0.5816
θ	0.1346	0.7917
ν	0.5211	0.4789
Q	$\begin{pmatrix} 0.7414 & 0.2586 \\ 0.2812 & 0.7188 \end{pmatrix}$	

1. Step 1 : Generate $\mathbf{U}_t$, $t \in \{1, \ldots, n\}$, from the regime-switching copula model.

2. Step 2 : For $j = 1, 2$, compute the conditional distribution function F_{jt} from formula (1.3), where under the risk neutral measure, $\mu_{jk} = r - \frac{\sigma_{jk}^2}{2}$, $k = 1, \ldots, \ell_j$. The other parameters are the same. For computing F_{j1}, one can use the probabilities $\eta_j(k)$ of being in regime k, as estimated from the historical data. For $t > 1$, the values $X_{j1}, \ldots, X_{j,t-1}$ are used to compute $\eta_{j,t-1}$, as given by formula (1.9).

3. Step 3 : For $t \in \{1, \ldots, n\}$, and $j = 1, 2$, set $X_{jt} = F_{jt}^{-1}(U_{jt})$. This way, under the risk neutral measure, for $j = 1, 2$, one still have Gaussian HMM for the returns, and in addition, $e^{-rt} S_{jt} = e^{\sum_{i=1}^{t}(X_{ji} - r)}$ is a martingale.

4. Step 5 : Once we have a path of returns, we redo steps 1 to 4 N times.

5. Step 6 : The value of the option is then approximated by the discounted value of the payoff evaluated at (S_{1n}, S_{2n}) for all N terminal values.

We applied the proposed model to the pricing of a bivariate claim, namely a put-on-max written on the stocks of Apple and Amazon. The payoff of this option is given by

$$max\{K - max(S_1, S_2)\},$$

where S_1 and S_2 are the gross return of the assets and K is the strike price. We use 10000 simulations, the maturity of the option is 20 days and the risk free rate is 4%, Figure 1.3 shows the price of the option as a function of the strike. The price of the option using an one regime Gaussian copula is also shown, the Kendall tau is taken at 0.32, which is our sample estimated Kendall tau. The prices given by the regime-switching Gaussian copula with two regimes are lower that those given by the one regime Gaussian copula, which suggest that the average dependence parameter is lower for the regime-switching Gaussian copula with two regimes.

1.5 Conclusion

In this paper, we proposed a methodology to select a number of regimes for the HMM copula models based on goodness-of fit test. Note that this methodology can also been used for the univariate HMM and the mixture copula models. We gave algorithms for the estimation and goodness-of-fit of regime switching copula models. The model was applied to the pricing of an put-on-max claim, and can also be applied to a wide range of multivariate options. The empirical results emphasize the importance of choosing the appropriate regime for the copula model, this choice can lead to difference in the price of the option.

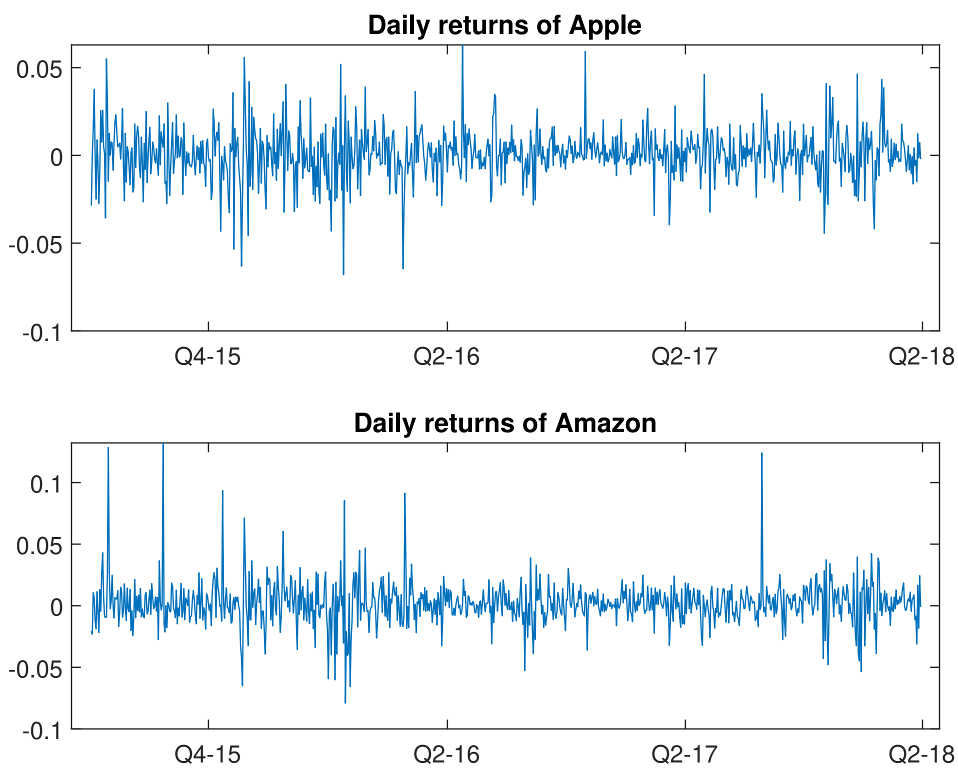

Figure 1.1: Daily log returns of Apple and Amazon.

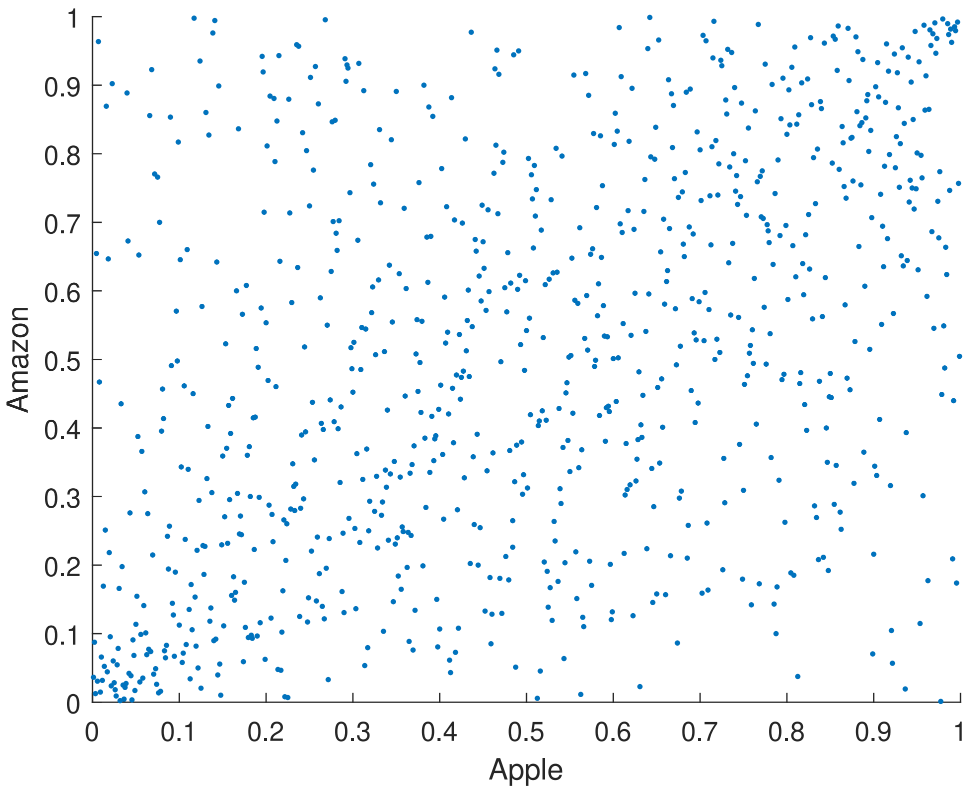

Figure 1.2: Scatter plot of the normalized ranks of Apple and Amazon from january 2015 to july 2018.

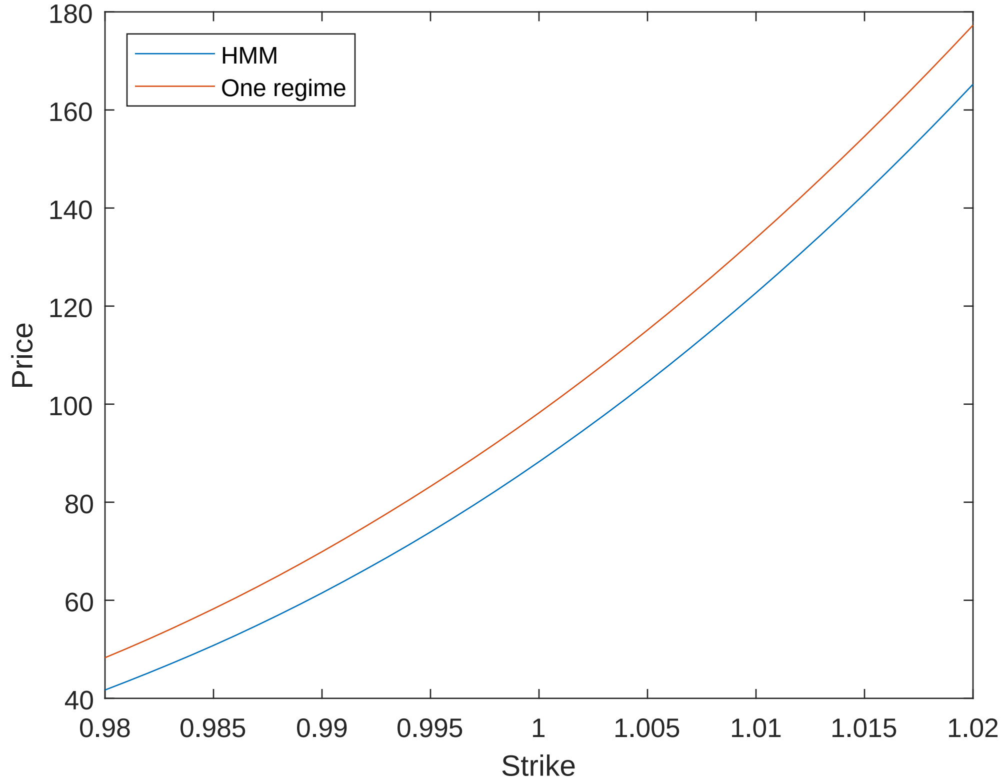

Figure 1.3: 1-month maturity put-on-max prices in basis points as a function of the strike under regime switching Gaussian copula with two regimes and one regime Gaussian copula.

Appendix

1.6 Estimation for general regime-switching models

1.6.1 E-Step

Set $\tilde{\boldsymbol{\theta}} = (\tilde{\boldsymbol{\beta}}_1, \ldots, \tilde{\boldsymbol{\beta}}_\ell, \tilde{P})$. Then, according to (Rémillard, 2013, Appendix 10.A),

$$
\begin{aligned}
Q_y(\tilde{\boldsymbol{\theta}}, \boldsymbol{\theta}) &= \mathbb{E}_{\boldsymbol{\theta}}\{\log f_{\tilde{\boldsymbol{\theta}}}(\tau, Y) | Y = y\} \\
&= \sum_{t=1}^{n} \sum_{j \in S} \sum_{k \in S} \mathbb{P}_{\boldsymbol{\theta}}(\tau_{t-1} = j, \tau_t = k | Y = y) \log \tilde{P}_{jk} \\
&\qquad + \sum_{t=1}^{n} \sum_{j \in S} \mathbb{P}_{\boldsymbol{\theta}}(\tau_t = j | Y = y) \log g_{\tilde{\boldsymbol{\beta}}_j}(y_t) \\
&= \sum_{t=1}^{n} \sum_{j \in S} \sum_{k \in S} \Lambda_{\boldsymbol{\theta},t}(j,k) \log \tilde{P}_{jk} + \sum_{t=1}^{n} \sum_{j \in S} \lambda_{\boldsymbol{\theta},t}(j) \log g_{\tilde{\boldsymbol{\beta}}_j}(y_t),
\end{aligned}
$$

where $\lambda_{\boldsymbol{\theta},t}(j) = P(\tau_t = j | Y = y)$ and $\Lambda_{\boldsymbol{\theta},t}(j,k) = P(\tau_{t-1} = j, \tau_t = k | Y = y)$, for all $t \in \{1, \ldots, n\}$ and $j, k \in S$. Next, define for all $j \in S$, $\bar{\eta}_{\boldsymbol{\theta},n}(j) = 1/\ell$, $\eta_{\boldsymbol{\theta},0}(j) = 1/\ell$,

$$
\begin{aligned}
\bar{\eta}_{\boldsymbol{\theta},t}(j) &= Prob(\tau_t = j | y_{t+1}, \ldots, y_n), \quad t = 1, \ldots, n-1, \\
\eta_{\boldsymbol{\theta},t}(j) &= Prob(\tau_t = j | y_1, \ldots, y_t), \quad t = 1, \ldots, n.
\end{aligned}
$$

It follows easily that

$$
\eta_t(j) = \frac{g_{\boldsymbol{\beta}_j}(y_t)}{Z_{t|t-1}} \sum_{i=1}^{\ell} \eta_{t-1}(i) P_{ij}, \quad t = 1, \ldots, n,
$$

where

$$Z_{t|t-1} = \sum_{j=1}^{\ell} g_{\boldsymbol{\beta}_j}(y_t) \sum_{i=1}^{\ell} \eta_{t-1}(i) P_{ij}.$$

Next, for all $i \in \{1, \ldots, l\}$, and for all $t = 0, \ldots, n-1$,

$$\begin{aligned}
\bar{\eta}_{\boldsymbol{\theta},t}(i) &= \frac{\sum_{\beta=1}^{\ell} \bar{\eta}_{\boldsymbol{\theta},t+1}(\beta) P_{i\beta} g_{\boldsymbol{\beta}_\beta}(y_{t+1})}{\sum_{k=1}^{\ell} \sum_{\beta=1}^{\ell} \bar{\eta}_{\boldsymbol{\theta},t+1}(\beta) P_{k\beta} g_{\boldsymbol{\beta}_\beta}(y_{t+1})}, \\
\lambda_{\boldsymbol{\theta},t}(i) &= \frac{\eta_{\boldsymbol{\theta},t}(i) \bar{\eta}_{\boldsymbol{\theta},t}(i)}{\sum_{k=1}^{l} \eta_{\boldsymbol{\theta},t}(k) \bar{\eta}_{\boldsymbol{\theta},t}(k)}.
\end{aligned}$$

Hence, for all $i, j \in \{1, \ldots, l\}$, and for all $t = 1, \ldots, n$,

$$\Lambda_{\boldsymbol{\theta},t}(i,j) = \frac{P_{ij} \eta_{\boldsymbol{\theta},t-1}(i) \bar{\eta}_{\boldsymbol{\theta},t}(j) g_{\boldsymbol{\beta}_j}(y_t)}{\sum_{k=1}^{l} \sum_{\beta=1}^{l} P_{k\beta} \eta_{\boldsymbol{\theta},t-1}(k) \bar{\eta}_{\boldsymbol{\theta},t}(\beta) g_{\boldsymbol{\beta}_\beta}(y_t)}.$$

As a result, for all $i \in \{1, \ldots, l\}$, and for every $t = 1, \ldots, n$,

$$\sum_{j=1}^{l} \Lambda_{\boldsymbol{\theta},t}(i,j) = \lambda_{\boldsymbol{\theta},t-1}(i).$$

1.6.2 M-Step

For this step, given $\boldsymbol{\theta}^{(k)}$, $\boldsymbol{\theta}^{(k+1)}$ is defined as

$$\boldsymbol{\theta}^{(k+1)} = \arg\max_{\boldsymbol{\theta}} Q_y \left(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)} \right).$$

Setting $\lambda_t^{(k)}(i) = \lambda_{\boldsymbol{\theta}^{(k)},t}(i)$ and $\Lambda_t^{(k)}(i,j) = \Lambda_{\boldsymbol{\theta}^{(k)},t}(i,j)$, it follows from Section 1.6.1 that

$$\begin{aligned}
\boldsymbol{\theta}^{(k+1)} &= \arg\max_{\boldsymbol{\theta}} \sum_{t=1}^{n} \sum_{i,j \in S} \Lambda_t^{(k)}(i,j) \log P_{ij} \\
&\quad + \sum_{t=1}^{n} \sum_{i \in S} \lambda_t^{(k)}(i) \log g_{\boldsymbol{\beta}_i}(y_t),
\end{aligned}$$

Using the Lagrange multipliers techniques, the function to maximize is $h(\boldsymbol{\theta}, \psi)$, where $\psi = (\psi_1, \ldots, \psi_\ell)$, and

$$\begin{aligned} h(\boldsymbol{\theta}, \psi) &= \sum_{t=1}^{n} \sum_{i,j \in S} \Lambda_t^{(k)}(i,j) \log P_{ij} \\ &\quad + \sum_{t=1}^{n} \sum_{i \in S} \lambda_t^{(k)}(i) \log g_{\boldsymbol{\beta}_i}(y_t) \\ &\quad\quad + \sum_{i=1}^{l} \psi_i \left(1 - \sum_{j=1}^{\ell} P_{ij} \right). \end{aligned}$$

For $i, j \in S$ we have,

$$\frac{\partial h}{\partial P_{i,j}} = \sum_{t=1}^{n} \Lambda_t^{(k)}(i,j) \frac{1}{P_{ij}} - \psi_i$$

As a result, for any $i, j \in S$, the partial derivative of h with respect to P_{ij} is zero if and only if

$$\psi_i P_{ij} = \sum_{t=1}^{n} \Lambda_t^{(k)}(i,j).$$

Summing over j yields that

$$\begin{aligned} \psi_i &= \sum_{j=1}^{\ell} \psi_i P_{ij} = \sum_{j=1}^{\ell} \sum_{t=1}^{n} \Lambda_t^{(k)}(i,j) \\ &= \sum_{t=1}^{n} \lambda_{t-1}^{(k)}(i) = \sum_{t=1}^{n} \lambda_{\boldsymbol{\theta}^{(k)}, t-1}(i). \end{aligned}$$

Hence

$$P_{ij}^{(k+1)} = \sum_{t=1}^{n} \Lambda_t^{(k)}(i,j) \Big/ \sum_{t=1}^{n} \lambda_{t-1}^{(k)}(i).$$

Also, maximize h with respect to $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_\ell$ amounts to maximize

$$\sum_{t=1}^{n} \sum_{i=1}^{\ell} \lambda_t^{(k)}(i) \log g_{\boldsymbol{\beta}_i}(y_t)$$

with respect to $\boldsymbol{\beta}_i$, for all $i \in S$.

1.7 Estimation for general mixture models

This model is a particular case of regime-switching where $P_{ij} = \nu_j$, $j \in \{1, \ldots, \ell\}$. So, under this model, τ_t is a sequence of i.i.d. observations with distribution $\nu = (\nu_1, \ldots, \nu_\ell)$. The algorithm described previously can then be simplified. To this end, set $\boldsymbol{\theta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_\ell, \nu)$.

The joint density of $\tau = (\tau_1, \ldots, \tau_n)$ and Y is

$$f_{\boldsymbol{\theta}}(\tau, Y) = \left(\prod_{t=1}^{n} \nu_{\tau_t} \right) \times \prod_{t=1}^{n} g_{\boldsymbol{\beta}_{\tau_t}}(Y_t),$$

yielding

$$\log f_{\boldsymbol{\theta}}(\tau, Y) = \sum_{t=1}^{n} \log P_{\tau_{t-1},\tau_t} + \sum_{t=1}^{n} \log g_{\boldsymbol{\beta}_{\tau_t}}(Y_t).$$

1.7.1 E-Step

Set $\tilde{\boldsymbol{\theta}} = (\tilde{\boldsymbol{\beta}}_1, \ldots, \tilde{\boldsymbol{\beta}}_\ell, \tilde{\nu})$. Then, according to the previous computations,

$$\begin{aligned} Q_y(\tilde{\boldsymbol{\theta}}, \boldsymbol{\theta}) &= \mathbb{E}_{\boldsymbol{\theta}}\{\log f_{\tilde{\boldsymbol{\theta}}}(\tau, Y) | Y = y\} \\ &= \sum_{t=1}^{n} \sum_{j \in S} \lambda_{\boldsymbol{\theta},t}(j) \left(\log \tilde{\nu}_j + \log g_{\tilde{\boldsymbol{\beta}}_j}(Y_t) \right), \end{aligned} \tag{1.15}$$

where

$$\lambda_{\boldsymbol{\theta},t}(j) = P_{\boldsymbol{\theta}}(\tau_t = j | Y = y) = \frac{\tilde{\nu}_j g_{\tilde{\boldsymbol{\beta}}_{\tau_t}}(y_t)}{\sum_{k=1}^{\ell} \tilde{\nu}_k g_{\tilde{\boldsymbol{\beta}}_k}(y_t)}$$

for all $t \in \{1, \ldots, n\}$ and $j \in S$.

1.7.2 M-Step

For this step, given $\boldsymbol{\theta}^{(k)}$, $\boldsymbol{\theta}^{(k+1)}$ is defined as

$$\boldsymbol{\theta}^{(k+1)} = \arg\max_{\boldsymbol{\theta}} Q_y \left(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)} \right).$$

Setting $\lambda_t^{(k)}(i) = \lambda_{\boldsymbol{\theta}^{(k)},t}(i)$, one obtains

$$
\begin{aligned}
Q_y(\tilde{\boldsymbol{\theta}}, \boldsymbol{\theta}) &= \sum_{t=1}^{n} \sum_{j=1}^{\ell} \lambda_t^{(k)}(j) \left(\log \tilde{\nu}_j + \log g_{\tilde{\boldsymbol{\beta}}_j}(Y_t) \right) \\
&= \sum_{t=1}^{n} \sum_{j=1}^{\ell} \lambda_t^{(k)}(j) \log \tilde{\nu}_j + \sum_{t=1}^{n} \sum_{j=1}^{\ell} \lambda_t^{(k)}(j) \log g_{\tilde{\boldsymbol{\beta}}_j}(Y_t)
\end{aligned}
$$

for $j \in S$ we have,

$$\frac{\partial Q_y}{\partial \tilde{\nu}_j} = \frac{n}{\tilde{\nu}_j} \sum_{t=1}^{n} \lambda_t^{(k)}(j).$$

Hence

$$\tilde{\nu}_j{}^{(k+1)} = \frac{\sum_{t=1}^{n} \lambda_t^{(k)}(j)}{n}, \quad j \in \{1, \dots, l\}.$$

and

$$\tilde{\boldsymbol{\beta}}_j{}^{(k+1)} = \arg\max_{\tilde{\boldsymbol{\beta}}_j} \sum_{t=1}^{n} \sum_{j=1}^{\ell} \lambda_t^{(k)}(j) \log g_{\tilde{\boldsymbol{\beta}}_j}(y_t).$$

Conclusion

Dans ce mémoire, nous avons présenté un algorithme pour l'estimation des paramètres d'un modèle de copule avec changement de régime, et nous avons proposé un nouveau test d'adéquation à partir duquel nous avons un critère de sélection pour le nombre de régimes. Dans le but de faciliter l'utilisation de la méthodologie proposée, nous avons construit une librairie de fonctions basée sur le progiciel R, HMMcopula, qui est disponible sur CRAN.

Comme application en ingénierie financière, nous avons considéré le problème de l'évaluation d'une option de vente sur le maximum des rendements des actions de Amazon et de Apple. Nous avons modélisé séparément chacune de ces séries de rendements par un modèle gaussien à changement de régime. Notre test d'adéquation nous a permis de détecter 3 régimes sur chacune des séries de rendements. Nous avons ensuite transformé ces données en observations presque indépendantes (pseudo-observations), et nous avons ensuite modélisé la dépendance entre ces pseudo-observations. Notre procédure nous a suggéré de prendre une copule gaussienne à deux régimes. Les résultats empiriques démontrent l'importance du choix approprié du nombre de régimes. Il est intéressant de noter que notre procédure peut être utilisée pour la construction de portefeuilles et la gestion des risques de ces derniers, par exemple le calcul de la valeur à risque (VAR), et autres mesures de risque.

Bibliography

Adams, Z., Füss, R., and Glück, T. (2017). Are correlations constant? empirical and theoretical results on popular correlation models in finance. *Journal of Banking & Finance*, 84:9–24.

Bai, J. (2003). Testing parametric conditional distributions of dynamic models. *The Review of Economics and Statistics*, 85(3):531–549.

Cappé, O., Moulines, E., and Rydén, T. (2005). *Inference in Hidden Markov Models*. Springer Series in Statistics. Springer, New York.

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

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B*, 39:1–38.

Du, Z. (2016). Nonparametric bootstrap tests for independence of generalized errors. *The Econometrics Journal*, 19(1):55–83.

Engle, R. (2002). Dynamic conditional correlation. *Journal of Business & Economic Statistics*, 20(3):339–350.

Fermanian, J.-D. and Lopez, O. (2015). Single-index copulae. *ArXiv e-prints*.

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

Fink, H., Klimova, Y., Czado, C., and Stöber, J. (2017). Regime switching vine copula models for global equity and volatility indices. *Econometrics*, 5(1):1–38.

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

Nasri, B., Rémillard, B., and Bouezmarni, T. (2017). Copula-based nonstationary conditional quantiles. Technical Report G-2017-102, GERAD.

Nasri, B. R. and Remillard, B. (2018). Copula-based dynamic models for multivariate time series. *SSRN eLibrary*.

Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review*, 47(2):527–556.

Rémillard, B. (2013). *Statistical Methods for Financial Engineering*. Chapman and Hall/CRC Financial Mathematics Series. Taylor & Francis.

Rémillard, B. (2017). Goodness-of-fit tests for copulas of multivariate time series. *Econometrics*, 5(1):13.

Rémillard, B., Hocquard, A., Lamarre, H., and Papageorgiou, N. A. (2017). Option pricing and hedging for discrete time regime-switching model. *Modern Economy*, 8:1005–1032.

van den Goorbergh, R. W. J., Genest, C., and Werker, B. J. M. (2005). Bivariate option pricing using dynamic copula models. *Insurance: Mathematics and Economics*, 37:101–114.

Package 'HMMcopula'

December 1, 2018

Type Package

Title Markov Regime Switching Copula Models Estimation and Goodness of Fit

Version 1.0.3

Author Mamadou Yamar Thioub <mamadou-yamar.thioub@hec.ca>,
Bouchra Nasri <bouchra.nasri@mail.mcgill.ca>,
Romanic Pieugueu <romanic.pieugueu@gerad.ca>, and
Bruno Remillard <bruno.remillard@hec.ca>

Maintainer Mamadou Yamar Thioub <mamadou-yamar.thioub@hec.ca>

Description R functions to estimate and perform goodness of fit test for several Markov regime switching and mixture bivariate copula models.
The goodness of fit test is based on a Cramer von Mises statistic and uses the Rosenblatt transform and parametric bootstrap to estimate the p-value.
The estimation of the copula parameters are based on the pseudo-maximum likelihood method using pseudo-observations defined as normalized ranks.

License GPL (>= 2)

Encoding UTF-8

LazyData true

Depends matrixcalc, mvtnorm, foreach, doParallel, copula

RoxygenNote 6.1.1

NeedsCompilation no

Repository CRAN

Date/Publication 2018-12-01 11:30:04 UTC

R topics documented:

`dilog` *Dilogarithm function*

Description

This function computes the dilogarithm of a number.

Usage

```
dilog(x)
```

Arguments

`x`	a real number

Value

`out`	dilogarithm

EstHMMCop	*Estimation of bivariate Markov regime switching bivariate copula model*

Description

This function estimates parameters from a bivariate Markov regime switching bivariate copula model

Usage

```
EstHMMCop(y, reg, family, max_iter, eps)
```

Arguments

y	(nx2) data matrix (observations or residuals) that will be transformed to pseudo-observations
reg	number of regimes
family	'gaussian' , 't' , 'clayton' , 'frank' , 'gumbel'
max_iter	maximum number of iterations of the EM algorithm
eps	precision (stopping criteria); suggestion 0.0001.

Value

theta	(1 x reg) estimated parameter of the copula according to CRAN copula package (except for Frank copula, where theta = log(theta_R_Package)) for each regime (except for degrees of freedom)
dof	estimated degree of freedom, only for the Student copula
Q	(reg x reg) estimated transition matrix
eta	(n x reg) conditional probabilities of being in regime k at time t given observations up to time t
tau	estimated Kendall tau for each regime
U	(n x 2) matrix of Rosenblatt transforms
cvm	Cramer-von-Mises statistic for goodness-of-fit
W	regime probabilities for the conditional distribution given the past Kendall's tau

Examples

```
Q <- matrix(c(0.8, 0.3, 0.2, 0.7),2,2) ; kendallTau <- c(0.3 ,0.7) ;
data <- SimHMMCop(Q, 'clayton', kendallTau, 10)$SimData;
estimations <- EstHMMCop(data,2,'clayton',10000,0.0001)
```

EstKendallTau *Sample Kendall's tau Estimation*

Description

This function estimates the sample Kendall's tau of a bivariate data matrix

Usage

```
EstKendallTau(X)
```

Arguments

X	(n x 2) matrix

Value

KendallTau	estimated sample Kendall's tau of the data

EstMixtureCop *Estimation of bivariate mixture bivariate copula model*

Description

This function estimates parameters from a mixture bivariate copula model

Usage

```
EstMixtureCop(y, reg, family, max_iter, eps)
```

Arguments

y	(nx2) data matrix (observations or residuals) that will be transformed to pseudo-observations
reg	number of regimes
family	'gaussian' , 't' , 'clayton' , 'frank' , 'gumbel'
max_iter	maximum number of iterations of the EM algorithm
eps	precision (stopping criteria); suggestion 0.0001.

Value

theta	(1 x reg) estimated parameter of the copula according to CRAN copula package (except for Frank copula, where theta = log(theta_R_Package)) for each component (except for degrees of freedom)
dof	estimated degree of freedom, only for the Student copula
Q	(1 x reg) estimated weights vector
eta	(n x reg) conditional probabilities of being in regime k at time t given observations up to time t
tau	estimated Kendall tau for each regime
U	(n x 2) matrix of Rosenblatt transforms
cvm	Cramer-von-Mises statistic for goodness-of-fit

GofHMMCop — *Goodness-of-fit of Markov regime switching bivariate copula model*

Description

This function performs goodness-of-fit test of a Markov regime switching bivariate copula model

Usage

```
GofHMMCop(R, reg, family, max_iter, eps, n_sample, n_cores)
```

Arguments

R	(n x 2) data matrix that will be transformed to pseudo-observations
reg	number of regimes
family	'gaussian' , 't' , 'clayton' , 'frank' , 'gumbel'
max_iter	maxmimum number of iterations of the EM algorithm
eps	precision (stopping criteria); suggestion 0.0001
n_sample	number of bootstrap; suggestion 1000
n_cores	number of cores to use in the parallel computing

Value

pvalue	pvalue (significant when the result is greater than 5)
theta	(1 x reg) estimated parameter of the copula according to CRAN copula package (except for Frank copula, where theta = log(theta_R_Package)) for each regime (except for degrees of freedom)
dof	estimated degree of freedom, only for the Student copula
Q	(reg x reg) estimated transition matrix

eta	(n x reg) conditional probabilities of being in regime k at time t given observations up to time t
tau	estimated Kendall tau for each regime
U	(n x 2) matrix of Rosenblatt transforms
cvm	Cramer-von-Mises statistic for goodness-of-fit
W	regime probabilities for the conditional distribution given the past Kendall's tau

---

GofMixtureCop *Goodness-of-fit of mixture bivariate copula model*

Description

This function performs goodness-of-fit test of a mixture bivariate copula model

Usage

```
GofMixtureCop(R, reg, family, max_iter, eps, n_sample, n_cores)
```

Arguments

R	(nx2) data matrix (observations or residuals) that will be transformed to pseudo-observations
reg	number of regimes
family	'gaussian' , 't' , 'clayton' , 'frank' , 'gumbel'
max_iter	maxmimum number of iterations of the EM algorithm
eps	precision (stopping criteria); suggestion 0.0001
n_sample	number of bootstrap; suggestion 1000
n_cores	number of cores to use in the parallel computing

Value

pvalue	pvalue (significant when the result is greater than 5)
theta	(1 x reg) estimated parameter of the copula according to CRAN copula package (except for Frank copula, where theta = log(theta_R_Package)) for each component (except for degrees of freedom)
dof	estimated degree of freedom, only for the Student copula
Q	(1 x reg) estimated weights vector
eta	(n x reg) conditional probabilities of being in regime k at time t given observations up to time t
tau	estimated Kendall tau for each regime
U	(n x 2) matrix of Rosenblatt transforms
cvm	Cramer-von-Mises statistic for goodness-of-fit

KendallTau *Kendall's tau of a copula*

Description

This function computes the Kendall's tau of a copula family with a unconstrainted parameter alpha.

Usage

```
KendallTau(family, alpha)
```

Arguments

`family`	"gaussian" , "t" , "clayton" , "frank" , "gumbel"
`alpha`	unconstrainted parameters of the copula family

Value

`tau`	estimated Kendall's tau

ParamCop *Theta estimation*

Description

This function computes the parameter of the copula according to CRAN copula package (except for Frank copula, where theta = log(theta_R_Package)), corresponding to the unconstrainted parameters alpha.

Usage

```
ParamCop(family, alpha)
```

Arguments

`family`	"gaussian" , "t" , "clayton" , "frank" , "gumbel"
`alpha`	unconstrainted parameters of the copula family

Value

`theta`	matlab parameters

ParamTau — *Alpha estimation*

Description

This function computes the unconstrainted parameter alpha for given Kendall's tau value

Usage

```
ParamTau(family, tau)
```

Arguments

`family`	'gaussian' , 't' , 'clayton' , 'frank' , 'gumbel'
`tau`	Kendall's tau of the copula family

Value

`alpha`	estimated unconstrainted parameter

RosenblattClayton — *Rosenblatt transform for Clayton copula*

Description

This function computes the Rosenblatt transform fot the Clayton copula

Usage

```
RosenblattClayton(u, theta)
```

Arguments

`u`	(n x d) matrix of pseudos-observations (normalized ranks)
`theta`	parameter of the Clayton copula

Value

`R`	Rosenblatt transform

`RosenblattFrank` *Rosenblatt transform for Frank copula*

Description

This function computes the Rosenblatt transform fot the Frank copula

Usage

```
RosenblattFrank(U, theta)
```

Arguments

`U`	(n x d) matrix of pseudos-observations (normalized ranks)
`theta`	parameter of the Frank copula

Value

`R`	Rosenblatt transform

`RosenblattGaussian` *Rosenblatt transform for Gaussian copula*

Description

This function computes the Rosenblatt transform fot the Gaussian copula

Usage

```
RosenblattGaussian(u, rho)
```

Arguments

`u`	(n x d) matrix of pseudos-observations (normalized ranks)
`rho`	(d x d) correlation matrix, or the correlation coefficient (if, d = 2)

Value

`R`	Rosenblatt transform

RosenblattGumbel *Rosenblatt transform for Gumbel copula*

Description

This function computes the Rosenblatt transform fot the Gumbel copula

Usage

```
RosenblattGumbel(U, theta)
```

Arguments

U	(n x d) matrix of pseudos-observations (normalized ranks)
theta	parameter of the Gumbel copula

Value

R	Rosenblatt transform

RosenblattStudent *Rosenblatt transform for Student copula*

Description

This function computes the Rosenblatt transform fot the Student copula

Usage

```
RosenblattStudent(u, rho, nu)
```

Arguments

u	(n x d) matrix of pseudos-observations (normalized ranks)
rho	(d x d) correlation matrix
nu	degrees of freedom

Value

R	Rosenblatt transform

`SimHMMCop` *Simulation of bivariate Markov regime switching copula model*

Description

This function simulates observation from a bivariate Markov regime switching copula model

Usage

```
SimHMMCop(Q, family, KendallTau, n, DoF)
```

Arguments

`Q`	Transition probality matrix (d x d);
`family`	'gaussian' , 't' , 'clayton' , 'frank' , 'gumbel'
`KendallTau`	Kendall's rank correlation
`n`	number of simulated vectors
`DoF`	degree of freedom only for the Student copula

Value

`SimData`	Simulated Data
`MC`	Markov chain regimes
`alpha`	parameters alpha

Examples

```
Q <- matrix(c(0.8, 0.3, 0.2, 0.7),2,2) ; kendallTau <- c(0.3 ,0.7) ;
simulations <- SimHMMCop(Q, 'gumbel', kendallTau, 300)
```

`SimMarkovChain` *Markov chain simulation*

Description

This function generates a Markov chain X(1), ..., X(n) with transition matrix Q, starting from a state eta0 or the uniform distribution on 1,..., r

Usage

```
SimMarkovChain(Q, n, eta0)
```

Arguments

Q	Transition probality matrix (d x d)
n	number of simulated vectors
eta0	variable eta

SimMixtureCop *Simulation of bivariate mixture copula model*

Description

This function simulates observation from a bivariate mixture copula model

Usage

```
SimMixtureCop(Q, family, KendallTau, n, DoF)
```

Arguments

Q	Weights vector (1 x component);
family	'gaussian' , 't' , 'clayton' , 'frank' , 'gumbel'
KendallTau	Kendall's rank correlation
n	number of simulated vectors
DoF	vector of degree of freedom only for the Student copula

Value

SimData	Simulated Data
MC	Markov chain regimes
alpha	parameters alpha

Examples

```
Q <- matrix(c(0.8, 0.2),1,2) ; kendallTau <- c(0.3 ,0.7) ;
simulations <- SimMixtureCop(Q, 'gaussian', kendallTau, 300)
```

SnB — *Cramer-von Mises statistic SnB for GOF based on the Rosenblatt transform*

Description

This function computes the Cramer-von Mises statistic SnB for GOF based on the Rosenblatt transform

Usage

```
SnB(E)
```

Arguments

E	(n x d) matrix of pseudos-observations (normalized ranks)

Value

Sn	Cramer-von Mises statistic

Tau2Rho — *Spearman's rho*

Description

This function estimates the Spearman's rho corresponding to a constrainted (matlab) parameter theta for a copula family.

Usage

```
Tau2Rho(family, theta)
```

Arguments

family	'gaussian' , 't' , 'clayton' , 'frank' , 'gumbel'
theta	parameter of the copula according to CRAN copula package (except for Frank copula, where theta = log(theta_R_Package))

Value

rho	estimated Spearman's rho

Index