

Chu, Ba; Satchell, Stephen

Article

Recovering the most entropic copulas from preliminary knowledge of dependence

Econometrics

Provided in Cooperation with:

MDPI – Multidisciplinary Digital Publishing Institute, Basel

Suggested Citation: Chu, Ba; Satchell, Stephen (2016) : Recovering the most entropic copulas from preliminary knowledge of dependence, *Econometrics*, ISSN 2225-1146, MDPI, Basel, Vol. 4, Iss. 2, pp. 1-21,
<https://doi.org/10.3390/econometrics4020020>

This Version is available at:

<https://hdl.handle.net/10419/171871>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by/4.0/>

Article

Recovering the Most Entropic Copulas from Preliminary Knowledge of Dependence

Ba Chu ^{1,*} and Stephen Satchell ^{2,3}

¹ Department of Economics, Carleton University, B-857 Loeb Building, 1125 Colonel By Drive, Ottawa, ON K1S 5B6, Canada

² Trinity College, University of Cambridge, Cambridge CB2 1TQ, UK; ses11@cam.ac.uk

³ The University of Sydney Business School, New South Wales NSW 2006, Australia

* Correspondence: ba.chu@carleton.ca; Tel.: +1-613-520-2600 (ext. 1546); Fax: +1-613-520-3906

Academic Editors: Fredj Jawadi, Tony S. Wirjanto, Marc S. Paoletta and Nuttanan Wichitaksorn

Received: 7 June 2015; Accepted: 1 March 2016; Published: 29 March 2016

Abstract: This paper provides a new approach to recover relative entropy measures of contemporaneous dependence from limited information by constructing the most entropic copula (MEC) and its canonical form, namely the most entropic canonical copula (MECC). The MECC can effectively be obtained by maximizing Shannon entropy to yield a proper copula such that known dependence structures of data (e.g., measures of association) are matched to their empirical counterparts. In fact the problem of maximizing the entropy of copulas is the dual to the problem of minimizing the Kullback-Leibler cross entropy (KLCE) of joint probability densities when the marginal probability densities are fixed. Our simulation study shows that the proposed MEC estimator can potentially outperform many other copula estimators in finite samples.

Keywords: entropy; relative entropy measure of joint dependence; copula; most entropic copula; canonical; kullback-Leibler cross entropy

JEL: C190; C590; C130

1. Introduction

There has been a substantial literature on estimation and inference of relative entropy measures of joint dependence as measures of serial correlation. These particular measures of dependence were first proposed by Joe [1] and extended by Granger and Lin [2]. Relative entropy based measures of dependence have so far received much interest in econometrics because they provide very general concepts for gauging joint dependence; and they can be used for a set of variables that can be a mixture of continuous, ordinal-categorical, and nominal-categorical variables. Interested readers are referred to [3–5] for a concise review of important contributions in this area.

Econometricians have recently become interested in the computation of maximum entropy densities (see, e.g., Golan [6], Usta and Kantar [7], and references therein for the background and discussions regarding maximum entropy (ME) densities.) The ME densities are derived by maximization of an information criterion (the level of uncertainty) subject to mass and mean preserving constraints. The justification for using the ME in this context can be found in [8]. Rockinger and Jondeau [9] apply the ME method to determine the ME return distribution which is then utilized to extend Bollerslev's GARCH into autoregressive conditional skewness and kurtosis. Maasoumi and Racine [10] employ a metric entropy measure of dependence to examine the predictability of asset returns. Hang [11] uses the ME to determine flexible functional forms of regression functions subject to side conditions. Miller and Liu [12] propose a method to recover a joint distribution function by applying the KLCE distance while imposing a required degree of dependence through

the joint moments. An example is the normal distribution which is completely characterized by first and second moments. In this case, the minimum KLCE distribution is the multivariate Normal distribution where the dependence is specified through conventional linear correlation.

There has been a great deal of interest in copulas, especially in financial economics, as they have the potential to model and explain asymmetric dependence between random variables separately from their marginal distributions. For example, Patton [13] employs various families of copulas to investigate the inter-relationship between univariate skewnesses, asymmetric dependence between asset returns, and the optimal portfolios of assets. Rodriguez [14] models financial contagion using copulas. Chollete, Heinen, and Valdesogo [15] propose a multivariate regime-switching copula to capture asymmetric dependence and regime-switching in portfolio selection. Ning, Xu, and Wirjanto [16] investigate asymmetric pattern in volatility clustering by employing a semi-parametric copula approach. Detailed indications of various econometric aspects or applications of copulas in economics and finance can be found, for instance, in the survey papers by Patton [17] and Fan and Patton [18]. A comprehensive treatment of copula theory is presented in the monograph by Nelsen [19].

Given the broad context described above, we propose a theoretical framework to recover relative entropy measures of joint dependence from limited information by constructing a set of the most entropic copulas (MEC's), which can essentially be done by maximizing Shannon entropy subject to constraints on the uniform marginal distributions and other constraints on the copula-based measures of dependence (or the distance between the MEC and an arbitrary nested copula). In the class of MECs, there exists a simplified form, namely the most entropic canonical copula (MECC). Moreover, it can be shown that the proposed MEC approach and the KLCE approach in Miller and Liu [12] are dual in the sense that they can recover the same joint distribution. Applications of MEC's to economics include Chu [20], Dempster, Medova, and Yang [21], Friedman and Huang [22], Veremyev, Tsyurmasto, Uryasev, and Rockafellar [23], Zhao and Lin [24].

We shall now discuss the contributions of the current paper in relation to [20]. The similarity between the two papers is that rank correlations are employed as prior information about dependence in order to construct the MECC. This paper differs from [20] in several respects. First, in [20], Carleman's condition permits constraints on moments to be employed so as to ensure that the MEC satisfies all the properties of a copula while, in the present paper, constraints are explicitly imposed on marginal copula densities. Therefore the entropy maximization problem defined in [20] is merely a good approximation of the entropy maximization problem in this study. Second the main problem in [20] is the standard entropy maximization problem while the main problem in the present paper involves a continuum of constraints on the marginal distributions, which can be written as integrals with varying end-points that need to be smoothed out by using kernels. This kernel-smoother can generate MECs with smooth densities whilst the discrete approximation technique proposed by [21] can only allow for MECs with discrete densities. The feasibility and benefits of the proposed approach to construct MECs will then be demonstrated through a Monte-Carlo simulation study presented in Section 3.

Although our analysis is restricted to the bivariate case, the multivariate case is a straightforward extension. The remainder of the paper is organized in three sections. In Section 2, we formulate and approximate most entropic copulas (MECs). Next, we discuss the link between the MEC and the minimum KLCE density and the extent to which the MEC is more flexible than the KLCE method. We then compute the MEC and the MECC subject to marginal constraints and other constraints on various copula-based dependence measures such as Spearman's rho and tau. We also outline the large sampling properties of the relevant parameter estimators. We present these results in Theorems 2.1–2.4. A simulation study is presented in Section 3, demonstrating that the MEC fits data well when compared with other competing procedures (e.g., parametric copulas and kernel estimators). Derivation of statistical properties for the proposed copula estimator is rather challenging and will be left for future research. Finally, to facilitate reading of this paper, we collect all materials of technical flavour into the three main appendices at the end of this paper.

2. Recovering the Most Entropic Copulas

2.1. Maximum Entropy and Copula

This section provides a brief explanation of entropy and copula. We refer to [25] for a comprehensive review of entropy econometrics and [19] for important results concerning copulas.

Shannon entropy has been used as an information criterion to construct the probability densities for economic or financial variables such as stock returns, income, GDP, etc. (see, *inter alia*, [26–28]). A univariate ME density is generally obtained by maximizing Shannon entropy, $-\int p(x) \log p(x) dx$, with respect to $p(x)$ under probability and moment constraints. A bivariate ME density that is closest to a given reference density, say the product of two univariate densities, can be obtained by minimizing the KLCE under joint moment constraints (see, e.g., [1] and [12]):

$$\min_f KLCE(f : g) = \min_f \int_{\text{supp}(g_1) \times \text{supp}(g_2)} f(x, y) \log \frac{f(x, y)}{g_1(x)g_2(y)} dx dy \quad (1)$$

subject to

$$\int_{\text{supp}(g_1) \times \text{supp}(g_2)} h(x, y) f(x, y) dx dy = \mu_0,$$

where f is a bivariate density, g_1 and g_2 are some univariate densities, $\text{supp}(g_1) = \{x \in \mathbb{R} : g_1(x) \neq 0\}$, $\text{supp}(g_2) = \{y \in \mathbb{R} : g_2(y) \neq 0\}$, and h is an arbitrary function such that $\mu_0 < \infty$.

The copula is proposed by Sklar [29] as a method to construct joint distributions with given marginals. The advantage of copulas is that dependence between random variables can be parametrically specified entirely independently from their marginals. A bivariate copula is defined as a function $C(\cdot, \cdot)$ from $[0, 1]^2$ to $[0, 1]$ with the following properties: 1) for every $u, v \in [0, 1]$, it holds that $C(u, 0) = C(0, v) = 0$, $C(u, 1) = u$, and $C(1, v) = v$; 2) $C(u, v)$ is 2-increasing, i.e., $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$ for every $u_1, u_2, v_1, v_2 \in [0, 1]$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$ (see, e.g., [19], p. 8). Note that Property (2) always holds if $C(u, v)$ has a positive density $c(u, v)$, and Property (1) implies that a copula is a function with Uniform[0,1] marginals. Sklar's theorem links a copula, $C(u, v)$, to a joint distribution, $F(X, Y)$, via $F(X, Y) = C(G_1(X), G_2(Y))$, where G_1 and G_2 are the marginals.

We shall use measures of association and rank correlations to construct the MEC, which we discuss next. Measures of association are, unlike joint moments, invariant under nonlinear transformations of the underlying random variables, and thus they are natural measures of dependence for non-elliptical random variables (see Appendix A for formal definitions of measures of association). A measure of association is, in general, defined as $\tau = \int_{[0,1]^2} h(u, v) dC(u, v)$, where h is a bivariate function such that $|\tau| < +\infty$. This measure, based on C , is also referred to as the *copula-based measure of dependence*. In practice, τ can be estimated by the rank statistic $\hat{\tau} = \frac{1}{N} \sum_{i=1}^N h(\frac{R_i}{N}, \frac{S_i}{N})$, where (R_i, S_i) represents the ranks of (X_i, Y_i) in a sample of size N . An advantage of using rank statistics as nonparametric measures of nonlinear dependence is that they are robust—in the sense that they will be insensitive to contamination and maintain a high efficiency for heavier tailed elliptical distributions as well as for multivariate normal distributions (see, e.g., [30] for a detailed treatment of rank statistics). Examples of $\hat{\tau}$ include Spearman's rho and Blest's rank correlations (see, e.g., [31]), which are summarized in Table 1.

Table 1. Blest’s measures of rank correlation.

Measures of Association	Rank Correlation
Spearman’s rho: $\rho_S = 12 \int_{[0,1]^2} uv c(u, v) dudv - 3$, and $\rho_S \in [-1, 1]$,	$\hat{\rho}_S = \frac{12}{N^3 - N} \sum_{i=1}^N R_i S_i - 3 \frac{N+1}{N-1}$.
Blest’s measure I: $v_1 = 2 - 12 \int_{[0,1]^2} (1-u)^2 v c(u, v) dudv$, and $v_1 \in [-1, 1]$,	$\hat{v}_1 = \frac{2N+1}{N-1} - \frac{12}{N^2 - N} \sum_{i=1}^N \left(1 - \frac{R_i}{N+1}\right)^2 S_i$.
Blest’s measure II: $v_2 = 2 - 12 \int_{[0,1]^2} u(1-v)^2 c(u, v) dudv$, and $v_2 \in [-1, 1]$,	$\hat{v}_2 = \frac{2N+1}{N-1} - \frac{12}{N^2 - N} \sum_{i=1}^N R_i \left(1 - \frac{S_i}{N+1}\right)^2$.
Blest’s measure III: $\eta = 6 \int_{[0,1]^2} u^2 v^2 c(u, v) dudv - \frac{1}{5}$, and $\eta \in [0, 1]$,	$\hat{\eta} = \frac{6}{N^2 - N} \sum_{i=1}^N \left(\frac{R_i}{N+1}\right)^2 \left(\frac{S_i}{N+1}\right)^2 - \frac{(1/5)N+1}{N-1}$.
Blest’s measure IV: $\phi = \int_{[0,1]^2} [10(1-u)^3 v - 3u^2 v^2] c(u, v) dudv - 9/10$, and $\phi \in [-1, 1]$,	$\hat{\phi} = \frac{1}{N^2 - N} \sum_{i=1}^N \left[\left(1 - \frac{R_i}{N+1}\right)^3 \frac{S_i}{N+1} - \left(\frac{R_i}{N+1}\right)^2 \left(\frac{S_i}{N+1}\right)^2 \right] - \frac{0.9N+1}{N-1}$.

Nonetheless, it is worth mentioning that the definition of τ is somewhat restrictive since it does not include Kendall’s tau, for example.¹ Moreover, not every rank correlation can be formulated in terms of the above general rank statistic $\hat{\tau}$. For instance, the statistic \hat{R}_g , which was proposed by Gideon and Hollister [32] as a coefficient of rank correlation resistant to outliers even in a small sample, has the form:

$$\hat{R}_g = \frac{1}{[N/2]} \left(\max_i \sum_{s=1}^i \mathbf{1}(p_s < N + 1 - i) - \max_i \sum_{s=1}^N \mathbf{1}(R_s \leq i < S_s) \right),$$

where p_s is the value of S_i with the subscript i satisfying $R_i = s$, and $[\bullet]$ is the greatest integer notation. In addition, \hat{R}_g estimates a copula-based measure of dependence, $R_g = 2 \int_{[0,1]^2} [\sup_{w \in [0,1]} \mathbf{1}(u \leq w, v < 1 - w) - \sup_{w \in [0,1]} (\mathbf{1}(u \leq w) - \mathbf{1}(u \leq w, v < w))] c(u, v) dudv$.

In the present paper, we use the bivariate Shannon entropy of a copula, given by

$$\mathbf{W}(c) = - \int_{[0,1]^2} c(u, v) \log c(u, v) dudv, \text{ where } c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v}. \tag{2}$$

By Sklar’s theorem the Shannon entropy of a copula is then equivalent to the KLCE:

$$\mathbf{W}(c) = -KLCE(f : g).$$

Hence, minimization of the KLCE and maximization of the bivariate Shannon entropy are dual problems. Let $\hat{c}(u, v)$ denote the MEC. Then, in view of [1], the relative entropy measure of dependence (recovered from limited information) is given by $-\mathbf{W}(\hat{c})$. Generally speaking, a multivariate Shannon entropy can be defined in an obvious way, and this dual relationship holds. However, as pointed out in Friedman and Huang [22] the problem of maximizing a multivariate Shannon entropy of copulas can suffer from the curse of dimensionality because the number of constraints (on the marginal densities) needed for the MEC to satisfy all the properties of a copula increases as the problem involve more dimensions.

2.2. The Most Entropic Copula

We assume for the rest of this paper that the MEC is a differentiable function so that its copula density exists. The bivariate MEC (or the MEC) is obtained by maximizing the bivariate Shannon entropy (2) under two following constraints: (1) the marginals of $c(u, v)$ are Uniform[0,1]; and (2) the measures of association, defined in Section 2.1, are set equal to the corresponding rank correlations. We call this Problem EM.

Problem EM: Maximize $\mathbf{W}(c) = - \int_{[0,1]^2} c(u, v) \log c(u, v) dudv$ (3)

¹ We are indebted to a referee for pointing this out.

subject to

$$\int_{[0,1]^2} c(u, v) dudv = 1, \tag{4}$$

$$\int_{(0,u]} \int_{[0,1]} c(x, v) dx dv = u, \forall u \in [0, 1], \tag{5}$$

$$\int_{[0,1]} \int_{(0,v]} c(u, y) du dy = v, \forall v \in [0, 1], \tag{6}$$

$$\int_{[0,1]^2} h(u, v; \hat{\theta}_N) c(u, v) dudv = 0, \tag{7}$$

where (4) implies that $c(u, v)$ is a joint density on the unit circle; Equations (5) and (6) imply that the marginals of $c(u, v)$ are Uniform[0,1] distributions; Equation (7) imposes a constraint on the joint behavior of U and V . To give an example, let $h(u, v; \hat{\theta}_N) = 12uv - 3 - \hat{\rho}_S$, then the left-hand side of (7) becomes Spearman’s rho and $\hat{\theta}_N = \hat{\rho}_S$ (note that, in what follows, we sometimes omit ‘ N ’ for brevity) is the rank correlation associated with Spearman’s rho. To give another example, suppose that the true data generating copula, say $C_0(u, v)$, belongs to a family, \mathcal{C}_0 . Given this prior information, to recover a MECC from the data, one may randomly choose a copula, $C_1(u, v; \hat{\beta})$, from \mathcal{C}_0 , then use it to construct (7) with $h(u, v; \hat{\theta}) = 4C_1(u, v; \hat{\beta}) - 1 - \hat{\tau}$, where $\hat{\theta} = \{\hat{\beta}, \hat{\tau}\}'$ and $\hat{\tau}$ is an estimate of the difference between the probabilities of concordance and discordance (cf. Appendix A). By doing this, it is expected that some feature of the family \mathcal{C}_0 could be effectively incorporated into the MECC. Other examples of Equation (7) also include Blest’s coefficients or Gideon and Hollister’s (1987) coefficient, etc. Also note that we may have more than one constraint like (7). It is to be stressed at this point that some versions of the MEC problem may exhibit boundary solutions due to theoretical restrictions on the measures of dependence employed (e.g., the Hoeffding-Frechet bounds on correlation statistics). Consequently, the large-sample theory stated in Section 2.3 below only holds for interior solutions to the stated problem.²

For future reference, we shall denote by $\hat{c}(u, v) = c(u, v, \hat{\Lambda})$, where $\hat{\Lambda}$ is a vector of coefficients, as the MEC [that solves Problem EM]. The MECs (accordingly the MECC) can then be approximated by replacing the continuums of varying end-points in (5) and (6) by sets of definite integrals. We now present an approximate solution to Problem EM in Theorem 2.1 below.

THEOREM 2.1. *The MEC, $\hat{c}(u, v)$, can be approximated by an approximator, $\hat{c}_{n, N_h}(u, v)$, as follows:*

$$\hat{c}(u, v) = \lim_{\substack{n \rightarrow \infty \\ N_h \rightarrow \infty}} \hat{c}_{n, N_h}(u, v)$$

with

$$\hat{c}_{n, N_h}(u, v) = \frac{\mathcal{E}_{n, N_h}(u, v)}{\int_{[0,1]^2} \mathcal{E}_{n, N_h}(u, v) dudv}, \tag{8}$$

where

$$\begin{aligned} \mathcal{E}_{n, N_h}(u, v) = & \exp \left\{ - \sum_{k=0}^{2^n-1} \left[\hat{\lambda}_k (\Phi(N_h(k2^{-n} - u)) + \Phi(-N_h((k+1)2^{-n} - u))) \right. \right. \\ & \left. \left. + \hat{\gamma}_k (\Phi(N_h(k2^{-n} - v)) + \Phi(-N_h((k+1)2^{-n} - v))) \right] - \hat{\lambda}_{2^n} h(u, v, \hat{\theta}) - b_0 \hat{c}(u, v) \right\} \end{aligned} \tag{9}$$

² We are indebted to a referee for suggesting to us this point.

and $\widehat{\Lambda}_n = \{\widehat{\lambda}_0, \dots, \widehat{\lambda}_{2^n-1}, \widehat{\gamma}_0, \dots, \widehat{\gamma}_{2^n-1}\}$ contains the minimal values of the following potential function:

$$\begin{aligned}
 Q_{n,N_h}(\Lambda_n, \widehat{\theta}) &= \int_{[0,1]^2} \exp \left\{ - \sum_{k=0}^{2^n-1} \left[\lambda_k (\Phi(N_h(k2^{-n} - u)) + \Phi(-N_h((k+1)2^{-n} - u)) - 1 + 2^{-n}) \right. \right. \\
 &\quad \left. \left. + \gamma_k (\Phi(N_h(k2^{-n} - v)) + \Phi(-N_h((k+1)2^{-n} - v)) - 1 + 2^{-n}) \right] \right. \\
 &\quad \left. - \lambda_{2^n} h(u, v, \widehat{\theta}) - b_0 \tilde{c}(u, v) \right\} dudv \text{ for a given } b_0 \text{ and } \tilde{c}(u, v).
 \end{aligned} \tag{10}$$

Note that $\Phi(x) = \frac{1}{2\sqrt{\pi}} \int_{-\infty}^x \exp\{-\frac{1}{2}y^2\} dy$ is the standard normal cdf (arising from smoothing indicator functions, $\mathbb{I}(u \in [k2^{-n}, (k+1)2^{-n}])$, with the Gaussian kernel) and $\tilde{c}(u, v)$ is an arbitrary copula (which may involve a nuisance parameter that needs to be estimated).

In particular, the MEC, $\widehat{c}(u, v)$, can be symmetrized by letting λ_k be equal to γ_k ($\forall k = 1, \dots, (2^n - 1)$) and $h(u, v, \widehat{\theta})$ be a symmetric function.

Proof: The proof utilizes the standard method of Variational Calculus for maximization of functions in normed linear spaces (see, e.g., [33], p. 129). See Appendix D. ■

As we can see, the MEC density nests an arbitrary copula, $\tilde{c}(u, v)$, (cf. Equation (9)). Indeed, the MEC depends on both b_0 and $\tilde{c}(u, v)$, thus no uniqueness is obtained. However, we can obtain a canonical form, which is called the MECC, by setting b_0 to zero. This idea of a *canonical* model can be traced back to Jeffreys³ who proposed to use the *principle of simplicity* for deductive inference—that is, for any given set of data, there is usually an infinite number of possible laws that will “explain” the data precisely; and the simplest model should be chosen.

It is also worth noting at this point that, like the empirical copula, the MECC is a valid distribution function; however, it satisfies the Uniform[0,1] marginal constraints only asymptotically. In addition the potential function $Q_{n,N_h}(\Lambda, \widehat{\theta})$ in the above theorem is a multivariate convex function of Λ , which in general has a unique minimum because it is the product of (positive) univariate convex functions.

We can claim that the MECC, $\widehat{c}(u, v)$, is equivalent to a maximum likelihood estimator (MLE). Now, we need to verify this claim—given a bivariate sample (X_i, Y_i) for $i = 1, \dots, N$, the average maximum log-likelihood function is given by

$$\begin{aligned}
 \ell(\widehat{\Lambda}_n) &= \frac{1}{N} \sum_{i=1}^N \log \widehat{c}_{n,N_h}(u_i, v_i, \widehat{\Lambda}_n) \\
 &= \frac{1}{N} \sum_{i=1}^N \log \mathcal{E}_{n,N_h}(u_i, v_i) - \log \int_{[0,1]^2} \mathcal{E}_{n,N_h}(u, v) dudv,
 \end{aligned}$$

where $\widehat{c}_{n,N_h}(u_i, v_i) \doteq \widehat{c}_{n,N_h}(u_i, v_i, \widehat{\Lambda}_n)$ is defined in (8),

$$u_i = \frac{1}{N} \sum_{s=1}^N \mathbf{1}(X_s \leq X_i) = \frac{R_i}{N+1},$$

and

$$v_i = \frac{1}{N} \sum_{s=1}^N \mathbf{1}(Y_s \leq Y_i) = \frac{S_i}{N+1},$$

³ Jeffreys, H. (1961). *Theory of Probability*. Oxford: Clarendon, pp. 2–3.

in which R_i and S_i are the ranks of X_i and Y_i in the sample, respectively. Assuming that N is greater than n and that n is large enough, in view of (9) with $b_0 = 0$, we obtain the following representation:

$$\begin{aligned} \ell(\widehat{\Lambda}_n) &= -\frac{1}{N} \sum_{i=1}^N \left(\widehat{\lambda}_{-1} + \sum_{k=0}^{2^n-1} \left(\left[\widehat{\lambda}_k (\Phi(N_h(k2^{-n} - u_i)) + \Phi(-N_h((k+1)2^{-n} - u_i))) \right. \right. \right. \\ &\quad \left. \left. \left. + \widehat{\gamma}_k (\Phi(N_h(k2^{-n} - v_i)) + \Phi(-N_h((k+1)2^{-n} - v_i))) \right] + \widehat{\lambda}_{2^n} h(u_i, v_i) \right) \right) \\ &\approx -\left(\widehat{\lambda}_{-1} + \frac{2}{2^n} \sum_{k=0}^{2^n-1} (\widehat{\lambda}_k + \widehat{\gamma}_k) + \widehat{\lambda}_{2^n} \frac{1}{N} \sum_{i=1}^N h(u_i, v_i) \right) \\ &= -\mathbf{W}(\widehat{c}(u, v)), \end{aligned}$$

where $\widehat{\lambda}_{-1} = \log \int_{[0,1]^2} \mathcal{E}_{n,N_h}(u, v) dudv$; the approximation (\approx) follows because $\frac{1}{N} \sum_{i=1}^N (\Phi(N_h(k2^{-n} - u_i)) + \Phi(-N_h((k+1)2^{-n} - u_i))) \approx \frac{1}{2^n}$ for every $k = 0, \dots, (2^n - 1)$; and the last equality holds because $\int_{[0,1]^2} h(u, v, \widehat{\theta}) c(u, v) dudv$ is set equal to its consistent rank estimator, $\frac{1}{N} \sum_{i=1}^N h(R_i / (N + 1), S_i / (N + 1))$. Hence, the claim has been verified.

REMARK 2.1. To compute the MECC, we could use either a Monte-Carlo integration procedure or Gaussian quadratures to approximate the potential function (10) (see Appendix C for further details), and then employ a global optimization technique (for example the stochastic search algorithm proposed by Csendes [34]) to minimize this function.

In general, we can also approximate $\widehat{c}(u, v)$ by using a collection of equally-spaced partitions of the unit interval $[0, 1]$, and then, a high-order kernel smoothing of the indicator function. This is stated in Theorem 2.2:

THEOREM 2.2. The MEC, $\widehat{c}(u, v)$, can be approximated by an approximator, $\widehat{c}_{L,h}(u, v)$, as follows:

$$\widehat{c}(u, v) = \lim_{\substack{L \rightarrow \infty \\ h \rightarrow 0}} \widehat{c}_{L,h}(u, v)$$

with

$$\widehat{c}_{L,h}(u, v) = \frac{\mathcal{E}_{L,h}(u, v)}{\int_{[0,1]^2} \mathcal{E}_{L,h}(u, v) dudv'}$$

where

$$\begin{aligned} \mathcal{E}_{L,h}(u, v) &= \exp \left\{ -\frac{1}{L} \sum_{k=1}^L \left[\widehat{\lambda}_k \left(\frac{1}{h} \int_{\frac{k-1}{L}}^{\frac{k}{L}} K \left(\frac{u-w}{h} \right) dw \right) + \widehat{\gamma}_k \left(\frac{1}{h} \int_{\frac{k-1}{L}}^{\frac{k}{L}} K \left(\frac{v-w}{h} \right) dw \right) \right] \right. \\ &\quad \left. - \widehat{\lambda}_{2^n} h(u, v, \widehat{\theta}) - b_0 \widehat{c}(u, v) \right\} \end{aligned}$$

for some kernel function, $K(\bullet)$, in $\mathcal{K}^r(\mathbb{R})$, where $\mathcal{K}^r(\mathbb{R})$ is the space of symmetric, Lebesgue integrable, kernel functions of order, r , (cf. Definition B.1) and $\widehat{\Lambda}_L = \{\widehat{\lambda}_1, \dots, \widehat{\lambda}_L, \widehat{\gamma}_1, \dots, \widehat{\gamma}_L\}$ contains the minimal values of the following potential function:

$$\begin{aligned} \mathcal{Q}_{L,h}(\Lambda_L, \widehat{\theta}) &= \int_{[0,1]^2} \exp \left\{ -\frac{1}{L} \sum_{k=1}^L \left[\lambda_k \left(\frac{1}{h} \int_{\frac{k-1}{L}}^{\frac{k}{L}} K \left(\frac{u-w}{h} \right) dw - \frac{1}{L} \right) \right. \right. \\ &\quad \left. \left. + \gamma_k \left(\frac{1}{h} \int_{\frac{k-1}{L}}^{\frac{k}{L}} K \left(\frac{v-w}{h} \right) dw - \frac{1}{L} \right) \right] - \lambda_{2^n} h(u, v, \widehat{\theta}) - b_0 \widehat{c}(u, v) \right\} dudv \\ &\text{for a given } b_0 \text{ and } \widehat{c}(u, v). \end{aligned}$$

Proof: The proof is very similar to Theorem 2.1 combined with Lemma B.1. So we shall omit its details here. ■

2.3. Large Sample Properties with Unknown Parameters of Dependence

The approximate MECC densities are members of a statistical exponential family parametrized by the Lagrange multipliers. Since the true parameters of dependence Θ^0 in (7) are unknown, a random sample of size N is then used to form their consistent estimates $\hat{\Theta}_N$. Therefore, the sampling properties of $\hat{\Lambda}_N$ may be derived from the associated sampling properties of $\hat{\Theta}_N$. Let $Q_n(\Lambda, \Theta)$ represent the approximate potential function with the dependence parameters Θ as formulated in Section 2, where $\hat{\Lambda}_N$ and Λ^0 denote the minimal values of $Q_n(\Lambda, \Theta)$ for $\Theta = \hat{\Theta}_N$ and $\Theta = \Theta^0$ respectively. The Hessian matrices of $Q_n(\Lambda, \Theta)$ are $\mathcal{H}_{1,n}(\Lambda, \Theta) = \nabla_{\Lambda\Lambda'} Q_n(\Lambda, \Theta)$ and $\mathcal{H}_{2,n}(\Lambda, \Theta) = \nabla_{\Lambda\Theta'} Q_n(\Lambda, \Theta)$. The following assumptions are maintained

AS1. $\hat{\Theta}_N \xrightarrow{p} \Theta^0 \in \text{int}(\mathcal{M})$, where \mathcal{M} is some non-empty compact set; $\dim(\mathcal{M})$ is the number of dependence constraints. Further,

$$\mathcal{N} = \left\{ \Lambda \in \mathbb{R}^{\dim(\Lambda)} : \nabla_{\Lambda} Q_n(\Lambda, \Theta) = 0, \forall \Theta \in \mathcal{M} \right\}$$

is also a non-empty and compact set, where $\dim(\Lambda)$ is the number of the Lagrange multipliers in $Q_n(\Lambda, \Theta)$. Therefore, the number of marginal constraints is $\dim(\Lambda) - \dim(\mathcal{M})$.

AS2. The map from \mathcal{M} to \mathcal{N} is a diffeomorphism (i.e., one-to-one, continuous and onto in both directions).

AS3. $Q_n(\Lambda, \Theta)$ is a strictly convex function of Λ for all Θ and uniformly continuous (in probability) in Θ , i.e.

$$\sup_{\Lambda \in \mathcal{N}} \left| Q_n(\Lambda, \hat{\Theta}_N) - Q_n(\Lambda, \Theta^0) \right| \xrightarrow{p} 0, \text{ as } |\hat{\Theta}_N - \Theta^0| \xrightarrow{p} 0.$$

AS4. The vector of dependence parameter estimates is asymptotically normal such that

$$N^{1/2}(\hat{\Theta}_N - \Theta^0) \xrightarrow{d} N(0, \Psi), \tag{11}$$

where Ψ is an asymptotic variance-covariance matrix of $\hat{\Theta}_N$.

AS2 states that the relationship between \mathcal{M} and \mathcal{N} is one-to-one in both directions (i.e., for a given set of dependence parameter estimates $\hat{\Theta}_N$ in \mathcal{M} there exists uniquely a set of the Lagrange multipliers $\hat{\Lambda}_N$ in \mathcal{N} which contains a unique subset of the Lagrange multipliers determining the dependence constraints). This assumption ensures that the potential function has uniquely minimal values for a given set of parameters. Conversely, these minimal values are uniquely determined by a set of parameters. Regarding **AS4**, $\hat{\Theta}_N$ may be a set of sample moments after N draws from the kernel densities constructed from actual data. If all the moments exist and Carleman’s condition holds, then $\hat{\Theta}_N$ are consistent asymptotically normal estimates of Θ_0 (see, e.g., Hardel, Muller, Sperlich, and Werwatz [35]).

THEOREM 2.3. In view of **AS1–AS4**, we obtain

$$\hat{\Lambda}_N \xrightarrow{p} \Lambda^0.$$

$$N^{1/2}(\hat{\Lambda}_N - \Lambda^0) \xrightarrow{d} N(0, \mathcal{H}_{1,n}^{-1}(\Lambda^0, \Theta^0) \mathcal{H}_{2,n}(\Lambda^0, \Theta^0) \Psi \mathcal{H}_{2,n}'(\Lambda^0, \Theta^0) \mathcal{H}_{1,n}^{-1}(\Lambda^0, \Theta^0)).$$

Proof: See Appendix D. ■

If the dependence constraints are linear in their parameters, *i.e.*, $\mathbf{h}(u, v, \Theta) = \mathbf{h}(u, v) - \Theta$, we can redefine the potential function associated with the constraints of Problem EM as follows:

$$\begin{aligned}
 Q_n(\Lambda, \widehat{\Theta}) &= \int_{[0,1]^2} \exp \left\{ \lambda_{-1} + \sum_{k=0}^{2^n-1} \left[\lambda_k (\Phi(k - 2^n u) + \Phi(2^n u - k - 1) - 1 + 2^{-n}) \right. \right. \\
 &\quad \left. \left. + \gamma_k (\Phi(k - 2^n v) + \Phi(2^n v - k - 1) - 1 + 2^{-n}) \right] \right. \\
 &\quad \left. - \boldsymbol{\kappa}'(\mathbf{h}(u, v) - \widehat{\Theta}) \right\} dudv - \lambda_{-1},
 \end{aligned} \tag{12}$$

where $\Lambda = \{\lambda_{-1}, \lambda_0, \gamma_0, \dots, \lambda_k, \gamma_k, \dots, \lambda_{2^n-1}, \gamma_{2^n-1}, \boldsymbol{\kappa}'\}$, and λ_{-1} is the Lagrange multiplier for the constraint $\int_{[0,1]^2} c(u, v) dudv = 1$.

THEOREM 2.4. *If (12) satisfies AS1–AS4, then we have*

$$\begin{aligned}
 \widehat{\Lambda}_N &\xrightarrow{p} \Lambda^0. \\
 N^{1/2}(\widehat{\Lambda}_N - \Lambda^0) &\xrightarrow{d} N(0, \mathcal{H}_{1,n}^{-1}(\Lambda^0, \Theta^0) \mathbf{I} \boldsymbol{\Psi}' \mathcal{H}_{1,n}^{-1}(\Lambda^0, \Theta^0)'),
 \end{aligned}$$

where \mathbf{I} is a $\dim(\Lambda^0) \times \dim(\Theta^0)$ diagonal matrix.

Proof: Noting that $\mathcal{H}_{2,n}(\Lambda^0, \Theta) = \mathbf{I}$, the proof follows directly from Theorem 2.3. ■

Theorem 2.4 suggests that in general the efficiency of the estimators $\widehat{\Lambda}_N$ can be improved by using more marginal constraints. However, adding too many marginal constraints can decrease efficiency since this may increase the probability that the covariances of $\{u, v, h(u, v, \widehat{\Theta}_N)\}$ in $Q_n(\Lambda, \widehat{\Theta}_N)$ are negative. Thus, the Hessian matrix $\mathcal{H}_{1,n}(\Lambda^0, \Theta^0)$ contains some negative elements which may cause the asymptotic variance of $N^{1/2}(\widehat{\Lambda}_N - \Lambda^0)$ to increase overall. Theorems 2.3 and 2.4 can be used to develop tests of hypotheses about the “distance” between the MECC and another copula of the exponential function family.

3. Simulation

In this section, we perform some simulations to investigate the finite-sample properties of the MECC approximators (proposed above). We shall address three main issues in these simulations. First, the MECC can outperform the parametric copulas used in this study (the Gaussian copula, Student’s *t* copula, the Clayton copula, and the Gumbel copula) while its performance remains comparable to other nonparametric estimators (*i.e.*, the “shrunked” local linear (LLS) type kernel copula estimator and the “shrunked” mirror-reflection (MRS) kernel copula estimator proposed by Omelka, Gijbels, and Veraverbeke [36]). Second, an increase in the number of marginal constraints leads to an improvement in the performance of the MECC. Third the MECC, for the most part, becomes as stable as other parametric copulas as more marginal constraints are utilized.

To accomplish the above objectives, we choose Frank’s copula,

$$C(u, v; \theta) = -\frac{1}{\theta} \log \left(\frac{(1 - e^{-\theta}) - (1 - e^{-\theta u})(1 - e^{-\theta v})}{(1 - e^{-\theta})} \right),$$

where $\theta \in (-\infty, \infty) \setminus \{0\}$, as the true model whereby samples are generated. (See [37,38] for the statistical properties of Frank’s copula.) This copula is radially symmetric and close to the independence as θ approaches the origin, *i.e.*, $\lim_{\theta \rightarrow 0} C(u, v; \theta) = uv$. Later, we shall use two values, 0.1 and 0.8, for the true parameter θ ; these values, roughly speaking, correspond to the close-to-independence case and the weak dependence case respectively.

The simulation procedure is outlined as follows. First, we generate 100 samples of 5000 observations from Frank’s copula for each value of θ . With these samples in hand, we estimate four commonly-used parametric copulas, mentioned above, by using MLE method. We also estimate 12 MECCs (that is, $MECC(L, M)$ with combinations of $L = 4, 16, 64$ marginal constraints and $M = 1, 2, 3, 4$ joint moment constraints) by using our proposed method. To gauge the errors of these estimators, we shall use the integrated mean squared error (IMSE);

$$\begin{aligned} \int_0^1 \int_0^1 \left\{ E \left[(c(u, v; \theta) - \hat{c}(u, v))^2 \right] \right\} dudv &= \int_0^1 \int_0^1 |E[\hat{c}(u, v) - c(u, v; \theta)]|^2 dudv \\ &+ \int_0^1 \int_0^1 E \left[(\hat{c}(u, v) - E[\hat{c}(u, v)])^2 \right] dudv \\ &= \text{Int. Bias}^2 + \text{Int. Var.}, \end{aligned}$$

where $c(u, v; \theta)$ is the density of Frank’s copula; and $\hat{c}(u, v)$ represents an estimate using one of the above-mentioned parametric copulas or a MECC. Next, for each copula, we use the 100 samples of 5000 observations drawn from Frank’s copula to estimate the squared bias and the variance (as the functions of u and v). Both the integrated squared bias (Int. Bias²) and the integrated variance (Int. Var.) are then obtained by evaluating the estimated squared bias ($\widehat{\text{Bias}}^2(u, v) = \left| \hat{E}[\hat{c}(u, v) - c(u, v; \theta)] \right|^2$, where \hat{E} denotes the empirical mean calculated using 100 samples) and the estimated variance ($\widehat{\text{Var.}}(u, v) = \hat{E} \left[(\hat{c}(u, v) - E[\hat{c}(u, v)])^2 \right]$) at 10000 pseudo-random Uniform [0,1] points, then taking their individual averages, *i.e.*,

$$\begin{aligned} \text{Int. Bias}^2 &\approx \frac{1}{10000} \sum_{i=1}^{10000} \widehat{\text{Bias}}^2(u_i, v_i), \\ \text{Int. Var.} &\approx \frac{1}{10000} \sum_{i=1}^{10000} \widehat{\text{Var.}}(u_i, v_i), \end{aligned}$$

where $\{(u_i, v_i)\}_{i=1}^{10000}$ denotes a sample of 10000 points (drawn from the Uniform [0,1] distribution) whereby both $c(u, v, \theta)$ and $\hat{c}(u, v)$ are evaluated. To gauge the errors of the nonparametric copula estimators, we shall use the expressions for the asymptotic bias and variance given in [36,39]; the optimal bandwidth is obtained by minimizing the integrated asymptotic MSE [39]. We report our simulation results in Table 2.

First, it can be noticed from Table 2 that the MECCs significantly outperform elliptical copulas (*i.e.*, the Normal copula and Student’s t copula) in terms of Int. Bias² and IMSE. However, with a small number of marginal constraints the MECCs are mostly less stable than other parametric copulas; the only way to improve the stability (Int. Var.) of the MECCs is to increase the number of marginal constraints. For the close-to-independence case ($\theta = 0.1$), the asymmetric copulas (*i.e.*, the Clayton copula and the Gumbel copula) outperform the MECCs. The intuition for these asymmetric copulas to have small Int. Bias² and Int. Var. is that Frank’s copula, the Clayton copula, and the Gumbel copula all behave like the independence copula for $\theta = 0.1$. It is also interesting to note that the MECCs often outperform the LLS and MRS estimators in terms of Int. Bias whilst these nonparametric estimators outperform the MECCs in terms of Int. Var. The reason for the existence of non-zero Int. Bias in the LLS and MRS estimators is that the optimal bandwidth (being shrinked close to zero at the corners of the unit square) can keep the bias bounded, but does not completely remove the bias.

Table 2. IMSE for the MECC and parametric copulas: Frank copula as the true copula.

Copula	$\theta = 0.1$ (Kendall's $\tau = 0.011$)			$\theta = 0.8$ (Kendall's $\tau = 0.1$)		
	Int. Bias ²	Int. Var.	IMSE	Int. Bias ²	Int. Var.	IMSE
Normal	27.9150	$1.0416 \times e^{-13}$	27.9150	28.3512	$1.0641 \times e^{-13}$	28.3512
Clayton	0.0002	$2.8924 \times e^{-15}$	0.0002	0.0176	$2.9086 \times e^{-15}$	0.0176
Gumbel	0.0002	$2.8745 \times e^{-15}$	0.0002	0.0203	$2.6789 \times e^{-15}$	0.0203
Student's <i>t</i>	1203.8600	$2.9161 \times e^{-12}$	1203.8600	955.8010	$3.3985 \times e^{-12}$	955.8010
LLS kernel	0.6180	0.0004	0.6185	0.6290	0.0004	0.6294
MRS kernel	0.6180	0.0004	0.6183	0.6290	0.0004	0.6292
MECC(4,1)	0.0028	0.0342	0.0371	0.0083	0.0048	0.0132
MECC(16,1)	0.0149	0.0014	0.0164	0.0157	0.0016	0.0173
MECC(64,1)	0.0126	0.0007	0.0133	0.0158	0.0005	0.0163
MECC(4,2)	0.0145	0.0878	0.1023	0.0132	0.0674	0.0806
MECC(16,2)	0.0170	0.0070	0.0241	0.0182	0.0074	0.0256
MECC(64,2)	0.0112	0.0017	0.0130	0.0134	0.0040	0.0175
MECC(4,3)	0.2667	11.4610	11.7277	0.1528	6.8401	6.9929
MECC(16,3)	0.0192	0.0804	0.0997	0.0196	0.1781	0.1977
MECC(64,3)	0.0184	0.0159	0.0343	0.0104	0.1360	0.1464
MECC(4,4)	0.1858	10.3994	10.5852	0.6811	25.2689	25.9500
MECC(16,4)	0.0487	1.7230	1.7718	0.0755	2.1647	2.2402
MECC(64,4)	0.0302	0.2109	0.2411	0.0230	0.1424	0.1655

Note: MECC(*L*, *M*) denotes the MECC estimated by using *L* marginal constraints and *M* moment constraints. All the figures are rounded to four decimal places. LLS is the “shrunked” version of the local linear-type kernel estimator of a copula [36,39]. MRS is the “shrunked” version of the mirror-reflection kernel estimator of a copula [36,40].

Second, when $\theta = 0.8$ the data will become less independent, leading to a significant increase in Int. Bias² pertaining to the estimation of the Clayton copula and Gumbel copula by using samples drawn from Frank's copula. In this case, MECC(4,1), MECC(16,1), MECC(64,1), MECC(4,2), MECC(64,2), and MECC(64,3) all show significant improvements in Int. Bias² over all the other estimators. It is also important to note at this point that, for a fixed number of marginal constraints, Int. Bias² and Int. Var. tend to deteriorate as one increases the number of joint moment constraints. To ameliorate this, it suffices to increase the number of marginal constraints as one adds one more joint moment constraint into the MEC problem. Indeed, as shown in Table 2, for one joint moment constraint, one merely needs four marginal constraints to yield MECC(4,1) with minimum Int. Bias² and IMSE; meanwhile, for two joint moment constraints, one needs to use up to 64 marginal constraints to yield MECC(64,2) with minimum Int. Bias², Int. Var., and IMSE. Our final observation is that, for a fixed number of moment constraints, an increase in the number of marginal constraints will always lead to a significant reduction in Int. Var.

Finally, to check the general validity of the obtained simulation results, we also replicate the above simulation study using data generated from Clayton copulas. Table 3 shows that the good performance of the MECCs relative to other copula estimators is still carried over to this case when a sufficient number of marginal constraints is being used.

Table 3. IMSE for the MECC and parametric copulas: Clayton copula as the true copula.

Copula	$\theta = 0.1$ (Kendall's $\tau = 0.05$)			$\theta = 0.8$ (Kendall's $\tau = 0.3$)		
	Int. Bias ²	Int. Var.	IMSE	Int. Bias ²	Int. Var.	IMSE
Normal	28.0325	$1.0580 \times e^{-13}$	28.0325	33.8124	$1.2673 \times e^{-13}$	33.8124
Frank	0.0137	0.0035	0.0172	0.1751	$6.4080 \times e^{-11}$	0.1751
Gumbel	0.0990	$4.6433 \times e^{-7}$	0.0990	4.1741	$3.9633 \times e^{-7}$	4.1741
Student's <i>t</i>	73.7077	4.6044	78.3122	23.1521	0.8711	24.0233
LLS kernel	0.6251	0.0005	0.6256	0.7183	0.0003	0.7186
MRS kernel	0.6253	0.0005	0.6256	0.7199	0.0003	0.7201
MECC(4,1)	0.0363	0.0907	0.1271	0.1006	0.1244	0.2251
MECC(16,1)	0.0245	0.0014	0.0259	0.0329	0.0026	0.0356
MECC(64,1)	0.0070	0.0009	0.0079	0.0188	0.0011	0.0200
MECC(4,2)	0.0149	0.6796	0.6946	0.2090	3.2983	3.5073
MECC(16,2)	0.0226	0.0049	0.0275	0.0821	0.0142	0.0964
MECC(64,2)	0.1801	0.1165	0.2966	0.2485	0.1658	0.4143
MECC(4,3)	0.2833	9.1560	9.4393	0.4628	17.1669	17.6297
MECC(16,3)	0.0304	0.4528	0.4833	0.4002	0.2889	0.6891
MECC(64,3)	0.0110	0.1303	0.1414	0.2430	0.1982	0.4412
MECC(4,4)	0.2993	14.8533	15.1526	1.8968	41.9359	43.8328
MECC(16,4)	0.0596	1.5987	1.6583	0.2976	2.1513	2.4489
MECC(64,4)	0.0608	0.2037	0.2645	0.2659	0.3957	0.6616

Note: MECC(*L*, *M*) denotes the MECC estimated by using *L* marginal constraints and *M* moment constraints. All the figures are rounded to four decimal places. LLS is the “shrunked” version of the local linear-type kernel estimator of a copula [36,39]. MRS is the “shrunked” version of the mirror-reflection kernel estimator of a copula [36,40].

4. Conclusions

We propose to employ the entropy-maximization principle to recover copulas from limited information regarding contemporaneous dependence between random variables. The main results of this article are twofold. First, we provide an entropy approach to recover relative entropy measures of joint dependence that are independent of marginal distributions by constructing most entropic copulas (MECs), in particular, their canonical forms, namely most entropic canonical copulas (MECC). Second, as a consequence of the MEC, we can construct ME joint distributions with a fixed dependence structure given by a MEC. Our method is shown to incorporate Miller and Liu [12]’s approach and can handle both moment-based and copula-based measures of dependence. Simulation results confirm that the accuracy of the approximate MECC can effectively be improved by increasing the number of side constraints.

Acknowledgments: We would like to express our sincere thanks to the guest editors (Professors Fredj Jawadi, Tony S. Wirjanto, Marc S. Paoella, and Nuttanan Wichitaksorn) and three anonymous referees for many valuable comments and constructive suggestions that help us to substantially improve this paper.

Author Contributions: Both authors contributed equally to the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix

A. Known Results

DEFINITION A.1 (Adapted and modified from Nelsen [19], Chapter 5). Let τ denote the difference between the probabilities of concordance and discordance of (X_1, Y_1) and (X_2, Y_2) , i.e., let

$$\tau = P\{(X_1 - X_2)(Y_1 - Y_2) > 0\} - P\{(X_1 - X_2)(Y_1 - Y_2) < 0\}, \quad (\text{A1})$$

where (X_1, Y_1) and (X_2, Y_2) are independent vectors of continuous random variables with joint distributions $F_1(X, Y)$ and $F_2(X, Y)$ respectively which have common marginals $G_1(X)$ (of X_1, X_2) and $G_2(Y)$ (of Y_1, Y_2). When (X_1, Y_1) and (X_2, Y_2) have the same joint distribution function $F(X, Y)$, τ is Kendall's tau (τ_K). The other measures of dependence such as Spearman's rho and Gini's gamma can be defined similarly.

THEOREM A.1 (Nelsen [19], Chapter 5). Let C_1 and C_2 denote the copulas of (X_1, Y_1) and (X_2, Y_2) respectively so that $F_1(X, Y) = C_1(G_1(X), G_2(Y))$ and $F_2(X, Y) = C_2(G_1(X), G_2(Y))$, then

$$\tau = Q(C_1, C_2) = 4 \int_{[0,1]^2} C_2(u, v) dC_1(u, v) - 1. \tag{A2}$$

If $C_2(u, v)$ is the Faliier-Gumbel-Mogernstern (FGM) copula, i.e., $C_2(u, v) = uv + \theta uv(1 - u - v - uv)$ then $\tau = \tau_{FGM}$. Note that we choose the FGM copula as a reference copula because it consists of quadrants of u, v which then enter τ_{FGM} as uniform joint moments.

If $C_2(u, v) = uv$ then $\tau = \rho_S$ (Spearman's rho).

If $C_2(u, v) = 2(|u + v - 1| - |u - v|)$ then $\tau = \gamma_G$ (Gini's gamma).

B. Auxiliary Results

DEFINITION B.1. A kernel function $K : \mathbb{R} \rightarrow \mathbb{R}$ of real order $r > 0$ is a symmetric, Lebesgue integrable, function such that

- (i) $\int_{\mathbb{R}} K(y) dy = 1,$
- (ii) $\int_{\mathbb{R}} y^j K(y) dy = 0$ for $j = 1, \dots, [r],$ and
- (iii) $\int_{\mathbb{R}} |y|^r |K(y)| dy < \infty,$ where $[r]$ is the integer part of $r.$

LEMMA B.1. Let $g(x)$ represent a measurable function of \mathbb{R}^n such that

- i $\int |g(x)| dx < \infty,$
- ii $\lim_{\|x\| \rightarrow 0} \|x\|^n |g(x)| = 0,$
- iii $\sup |g(x)| < \infty,$

where $\|x\|$ is the Euclidean norm of x . Let $f(x)$ be another function on \mathbb{R}^n such that $\int |f(x)| < \infty$. Then, at every point, x_0 , of continuity of f ,

$$\frac{1}{h_T^n} \int_{\mathbb{R}^n} g\left(\frac{x_0 - w}{h_T}\right) f(w) dw \rightarrow f(x_0) \int g(w) dw$$

as $T \rightarrow \infty$, where h_T is a sequence of positive constants such that $h \rightarrow 0$ as $T \rightarrow \infty$.

Proof: See ([41], p. 362). ■

LEMMA B.2. Let $\Omega = [0, 1]$, let P denote Lebesgue measure, and let $f = f(x) \in L^1(\Omega)$. Put

$$f_n(x) = 2^n \int_{[k2^{-n}, (k+1)2^{-n}]} f(y) dy, \quad x \in [k2^{-n}, (k+1)2^{-n}). \tag{B1}$$

Then $f_n(x) \xrightarrow{P-as} f(x)$, where $\{k2^{-n}\}$ is a compact dyadic sequence dense in Ω . (Note that a sequence is defined to be dense in an interval if for every point of the interval, there exist a point of the sequence which is arbitrarily close to it. (See [42], p. 515.))

LEMMA B.3. [DuBois-Reymond's lemma] Let a function $b(t)$ be continuous on the interval $[t_0, t_1]$. Assuming that the following equality holds for any continuous function $v(t)$ with mean value zero (i.e. $\int_{t_0}^{t_1} v(t) dt = 0$)

$$\int_{t_0}^{t_1} b(t)v(t) dt = 0. \tag{B2}$$

Then, $b(t) = b_0 = \text{const}$. Vice versa, if $b(t) = \text{const}$ then $\int_{t_0}^{t_1} b(t)v(t) dt = 0$. (See [43], p. 400)

LEMMA B.4. The indicator function $\mathbf{1}_{y>x}(y)$ can be approximated by a continuous function $\Phi_N(y, x)$, where $\Phi_N(y, x)$ is given by

$$\Phi_N(y, x) = \frac{N}{2\pi} \int_{-\infty}^y \exp\{-(v-x)^2 N^2/2\} dv; \tag{B3}$$

$\Phi_N(y, x)$ has the following properties:

$$\begin{aligned} \lim_{N \rightarrow \infty} \Phi_N(y, x) &\implies \mathbf{1}_{y>x}(y), \\ \lim_{N \rightarrow \infty} \frac{\partial \Phi_N(y, x)}{\partial y} &\implies \delta(y-x), \end{aligned}$$

where $\delta(\bullet)$ is Dirac's delta function. (See [44] (p. 30))

C. Approximation of Potential Functions

We now present a Gaussian-Legendre quadrature method to approximate the potential function (10) for the MECC. Using affine transformations, $x_1(u) : [0, 1] \rightarrow [-1, 1]$ with $x_1 = 2u - 1$ and $x_2(v) : [0, 1] \rightarrow [-1, 1]$ with $x_2 = 2v - 1$, (10) can be rewritten as follows:

$$\begin{aligned} Q_{n,N_h}(\mathbf{\Lambda}, \hat{\theta}) &= \frac{1}{4} \int_{[-1,1]^2} \exp \left\{ \sum_{k=0}^{2^n-1} \left[\lambda_k \left(\Phi(k - 2^{n-1}(x_1 + 1)) + \Phi(2^{n-1}(x_1 + 1) - k - 1) - 1 + 2^{-n} \right) \right. \right. \\ &\quad \left. \left. + \gamma_k \left(\Phi(k - 2^{n-1}(x_2 + 1)) + \Phi(2^{n-1}(x_2 + 1) - k - 1) - 1 + 2^{-n} \right) \right] \right. \\ &\quad \left. - \lambda_{2^n} \left(h \left(\frac{x_1 + 1}{2}, \frac{x_2 + 1}{2} \right) - \hat{\theta} \right) \right\} dx_1 dx_2 \\ &= \frac{1}{4} \int_{[-1,1]^2} \exp\{-\mathbf{\Lambda}' \mathbf{\Psi}(\mathbf{X})\} d\mathbf{X}, \tag{C1} \end{aligned}$$

where $\mathbf{X} = \{x_1, x_2\}$, $\mathbf{\Lambda}' = \{\lambda_0, \gamma_0, \dots, \lambda_k, \gamma_k, \dots, \lambda_{2^n-1}, \gamma_{2^n-1}, \lambda_{2^n}\}$, and $\mathbf{\Psi}(\mathbf{X})$ has an obvious meaning.

The function $\exp\{-\mathbf{\Lambda}' \mathbf{\Psi}(\mathbf{X})\}$ can be expanded into a series of the orthogonal Legendre polynomials, that is,

$$\exp\{-\mathbf{\Lambda}' \mathbf{\Psi}(\mathbf{X})\} = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} a_{nm} P_{nm}(\mathbf{X}), \tag{C2}$$

where $P_{nm}(\mathbf{X}) = P_n(x_1)P_m(x_2)$ are products of two Legendre orthogonal polynomials (see, e.g., [45] for further details of the Legendre polynomials),

$$a_{nm} = \frac{(2n+1)(2m+1)}{4} \int_{[-1,1]^2} \exp\{-\mathbf{\Lambda}' \mathbf{\Psi}(\mathbf{X})\} P_{nm}(\mathbf{X}) d\mathbf{X},$$

and

$$a_{00} = \frac{1}{4} \int_{[-1,1]^2} \exp\{-\mathbf{\Lambda}' \mathbf{\Psi}(\mathbf{X})\} d\mathbf{X}.$$

Now, let $\mathbf{X}_{i,j} = (x_{1i}, x_{2j})$, $\forall i = 1, \dots, N$ and $j = 1, \dots, M$, be the roots of the polynomials $P_N(x_1) = 0$ and $P_M(x_2) = 0$ respectively – $\mathbf{X}_{i,j}$ are also called the abscissae of the Legendre polynomials – then, choose weights, ω_{ij} , satisfying the following $M \times N$ relations:

$$\begin{cases} \sum_{i=1}^N \sum_{j=1}^M \omega_{ij} P_{00}(\mathbf{X}_{i,j}) = \sum_{i=1}^N \sum_{j=1}^M \omega_{ij} = 1, \\ \sum_{i=1}^N \sum_{j=1}^M \omega_{ij} P_{kh}(\mathbf{X}_{i,j}) = 0, \omega_{ij} \geq 0, \end{cases} \tag{C3}$$

where $(k, h) \in (1, \dots, N) \otimes (1, \dots, M)$. We obtain:

$$\begin{aligned} \sum_{i=1}^N \sum_{j=1}^M \omega_{ij} \exp\{-\Lambda' \Psi(\mathbf{X}_{ij})\} &= \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} a_{nm} \sum_{i=1}^N \sum_{j=1}^M \omega_{ij} P_{nm}(\mathbf{X}_{ij}) \\ &= a_{00} + \sum_{n=N+1}^{\infty} \sum_{m=M+1}^{\infty} a_{nm} \sum_{i=1}^N \sum_{j=1}^M \omega_{ij} P_{nm}(\mathbf{X}_{ij}). \end{aligned} \tag{C4}$$

Hence,

$$\begin{aligned} a_{00} = Q_{n, N_h}(\Lambda, \hat{\theta}) &= \sum_{i=1}^N \sum_{j=1}^M \omega_{ij} \exp\{-\Lambda' \Psi(\mathbf{X}_{ij})\} - \sum_{n=N+1}^{\infty} \sum_{m=M+1}^{\infty} a_{nm} \sum_{i=1}^N \sum_{j=1}^M \omega_{ij} P_{nm}(\mathbf{X}_{ij}) \\ &= \underbrace{\sum_{i=1}^N \sum_{j=1}^M \omega_{ij} \exp\{-\Lambda' \Psi(\mathbf{X}_{ij})\}}_{\text{The Approximation}} + R_{NM}, \end{aligned} \tag{C5}$$

where $R_{NM} = -\sum_{n=N+1}^{\infty} \sum_{m=M+1}^{\infty} a_{nm} \sum_{i=1}^N \sum_{j=1}^M \omega_{ij} P_{nm}(\mathbf{X}_{ij})$ is an error term, and (M, N) are large enough.

We now present a Quasi-Newton algorithm to minimize (C5).

(1) Given a starting point Λ_0 , a convergence tolerance $\epsilon > 0$ and an initial step length $\alpha_0 > 0$, $(\rho, c) \in [0, 1]^2$, an initial inverse Hessian matrix \mathbf{H}_0 and the numbers of Gaussian quadratures N, M .

(2) While $\|\nabla Q_n(\Lambda_k, \hat{\Theta})\|_{(2n+1 \times 1)} > \epsilon$:

$$\mathbf{p}_k = -\mathbf{H}_k \nabla Q_n(\Lambda_k, \hat{\Theta}) \text{ (search direction).}$$

$2n+1 \times 1$

$$\text{repeat until } Q_n(\Lambda_k + \alpha \mathbf{p}_k, \hat{\Theta}) \leq Q_n(\Lambda_k, \hat{\Theta}) + c\alpha \nabla' Q_n(\Lambda_k, \hat{\Theta}) \mathbf{p}_k$$

$$\alpha \leftarrow \rho \alpha$$

Stop with $\alpha_k = \alpha$ (step length satisfies the Goldstein condition (see, e.g., [46])

$$\text{Set } \Lambda_{k+1} = \Lambda_k - \alpha_k \mathbf{p}_k$$

$2n+1 \times 1$

(3) Define $\mathbf{s}_k = \Lambda_{k+1} - \Lambda_k$ and $\mathbf{y}_k = \nabla Q_n(\Lambda_{k+1}, \hat{\Theta}) - \nabla Q_n(\Lambda_k, \hat{\Theta})$

$2n+1 \times 1$

$2n+1 \times 1$

Compute the updated Hessian matrix:

$$H_{k+1} = (\mathbf{I} - \rho_k \mathbf{s}_k \mathbf{y}_k') H_k (\mathbf{I} - \rho_k \mathbf{y}_k \mathbf{s}_k') + \rho_k \mathbf{s}_k \mathbf{s}_k', \text{ where } \rho_k = \frac{1}{\mathbf{y}_k' \mathbf{s}_k}$$

$$k \leftarrow k + 1$$

end;

where $\nabla Q_n(\Lambda, \hat{\Theta})$ is the first-order gradient vector of $Q_n(\Lambda, \hat{\Theta})$ and $\|\bullet\|$ is a standard matrix norm.

To compute the MECC, we used a stochastic search algorithm to minimize (C5) whilst setting $M = N = 30$.

D. Proofs

Proof of Theorem 2.1: Since (5)–(6) are continuums of constraints with varying end-points, we need to replace these continuums with sets of definite integrals:

$$\int_{[a,b]} \int_{[0,1]} c(u, v) dudv = \int_{[0,1]} \int_{[a,b]} c(u, v) dudv = b - a, \tag{D1}$$

where a and b are arbitrary numbers in $[0, 1]$. Using a dense dyadic sequence in $[0, 1]$, (D1) can be approximated by

$$\sum_{k=k_1}^{k=k_2} \int_{[k2^{-n}, (k+1)2^{-n}]} \int_{[0,1]} c(u, v) dudv = \sum_{k=k_1}^{k=k_2} \int_{[0,1]} \int_{[k2^{-n}, (k+1)2^{-n}]} c(u, v) dudv = \frac{k_2 - k_1}{2^n},$$

where k_1 and k_2 are chosen such that $|a - k_1 2^{-n}| \leq \epsilon$ and $|b - k_2 2^{-n}| \leq \epsilon$, where ϵ is small enough. Hence, (D1) is equivalent to

$$\int_{[k2^{-n}, (k+1)2^{-n}]} \int_{[0,1]} c(u, v) dudv = \int_{[0,1]} \int_{[k2^{-n}, (k+1)2^{-n}]} c(u, v) dudv = \frac{1}{2^n} \quad \forall k = 0, 1, 2, \dots, (2^n - 1), \text{ and } n \text{ is large enough.} \tag{D2}$$

The Lagrangian function of Problem EM can be formulated as follows:

$$\begin{aligned} \mathcal{L}(c, \Lambda_n; \hat{\theta}) &= - \int_{[0,1]^2} c(u, v) \log c(u, v) dudv - \lambda_{-1} \left[\int_{[0,1]^2} c(u, v) dudv - 1 \right] \\ &\quad - \sum_{k=0}^{2^n-1} \left\{ \lambda_k \int_{[k2^{-n}, (k+1)2^{-n}]} \int_{[0,1]} [c(u, v) - 2^{-n}] + \gamma_k \int_{[0,1]} \int_{[k2^{-n}, (k+1)2^{-n}]} [c(u, v) - 2^{-n}] \right\} \\ &\quad - \lambda_{2^n} \int_{[0,1]^2} h(u, v, \hat{\theta}) c(u, v) dudv \\ &= - \int_{[0,1]^2} \left\{ c(u, v) \log c(u, v) + \lambda_{-1} [c(u, v) - 1] \right. \\ &\quad + \sum_{k=0}^{2^n-1} \left(\lambda_k \mathbb{I}(u \in [k2^{-n}, (k+1)2^{-n}]) + \gamma_k \mathbb{I}(v \in [k2^{-n}, (k+1)2^{-n}]) \right) [c(u, v) - 2^{-n}] \\ &\quad \left. + \lambda_{2^n} h(u, v, \hat{\theta}) c(u, v) \right\} dudv. \end{aligned} \tag{D3}$$

Taking the first derivative of $\mathcal{L}(c, \Lambda_n; \hat{\theta})$ with respect to c leads to

$$\int_{[0,1]^2} \left\{ \log c(u, v) + (1 + \lambda_{-1}) + \sum_{k=0}^{2^n-1} \left[\lambda_k \mathbb{I}(u \in [k2^{-n}, (k+1)2^{-n}]) + \gamma_k \mathbb{I}(v \in [k2^{-n}, (k+1)2^{-n}]) \right] + \lambda_{2^n} h(u, v, \hat{\theta}) \right\} dudv = 0. \tag{D4}$$

Define $b_n(u, v) \doteq \log c(u, v) + (1 + \lambda_{-1}) + \sum_{k=0}^{2^n-1} \left[\lambda_k \mathbb{I}(u \in [k2^{-n}, (k+1)2^{-n}]) + \gamma_k \mathbb{I}(v \in [k2^{-n}, (k+1)2^{-n}]) \right] + \lambda_{2^n} h(u, v, \hat{\theta})$, then applying Lemma B.3 to the function

$$b(u, v) = \frac{b_n(u, v)}{\bar{c}(u, v) - 1}, \tag{D5}$$

where $\tilde{c}(u, v)$ is an arbitrary copula density such that $\int_{[0,1]^2} (\tilde{c}(u, v) - 1) dudv = 0$, we obtain the following representation:

$$\hat{c}_{n, N_h}(u, v) = \exp \left\{ - (1 + \lambda_{-1} - b_0) - \sum_{k=0}^{2^n-1} \left[\lambda_k \mathbb{I}(u \in [k2^{-n}, (k+1)2^{-n}]) + \gamma_k \mathbb{I}(v \in [k2^{-n}, (k+1)2^{-n}]) \right] - \lambda_{2^n} h(u, v, \hat{\theta}) - b_0 \tilde{c}(u, v) \right\}, \tag{D6}$$

and b_0 is a generic constant. Now, by substituting (D6) into (4) the leading term, $1 + \lambda_{-1} - b_0$, is canceled out, then we obtain:

$$\hat{c}_{n, N_h}(u, v) = \frac{\mathcal{E}_{n, N_h}(u, v)}{\int_{[0,1]^2} \mathcal{E}_{n, N_h}(u, v) dudv}, \tag{D7}$$

where

$$\mathcal{E}_{n, N_h}(u, v) = \exp \left\{ - \sum_{k=0}^{2^n-1} \left[\hat{\lambda}_k \mathbb{I}(u \in [k2^{-n}, (k+1)2^{-n}]) + \hat{\gamma}_k \mathbb{I}(v \in [k2^{-n}, (k+1)2^{-n}]) \right] - \hat{\lambda}_{2^n} h(u, v, \hat{\theta}) - b_0 \tilde{c}(u, v) \right\}.$$

The Lagrangian multipliers $\hat{\Lambda}_n = \{ \hat{\lambda}_0, \dots, \hat{\lambda}_{2^n}, \hat{\gamma}_0, \dots, \hat{\gamma}_{2^n-1} \}$ can be solved out by substituting (D7) into (5), (6), and (7), leading to the following system of equations:

$$\begin{aligned} \frac{1}{\int_{[0,1]^2} \mathcal{E}_{n, N_h}(u, v) dudv} \int_{[0,1]^2} \mathbb{I}(u \in [k2^{-n}, (k+1)2^{-n}]) \mathcal{E}_{n, N_h}(u, v) dudv &= 2^{-n}, \\ \frac{1}{\int_{[0,1]^2} \mathcal{E}_{n, N_h}(u, v) dudv} \int_{[0,1]^2} \mathbb{I}(v \in [k2^{-n}, (k+1)2^{-n}]) \mathcal{E}_{n, N_h}(u, v) dudv &= 2^{-n}, \\ \frac{1}{\int_{[0,1]^2} \mathcal{E}_{n, N_h}(u, v) dudv} \int_{[0,1]^2} h(u, v, \hat{\theta}) \mathcal{E}_{n, N_h}(u, v) dudv &= 0 \end{aligned} \tag{D8}$$

for all $k = 0, \dots, (2^n - 1)$. Since (D7) can be rewritten as

$$\begin{aligned} \hat{c}_{n, N_h}(u, v) &= \frac{- \sum_{k=0}^{2^n-1} (\hat{\lambda}_k 2^{-n} + \hat{\gamma}_k 2^{-n})}{\int_{[0,1]^2} \mathcal{E}_{n, N_h}(u, v) dudv} \exp \left\{ - \sum_{k=0}^{2^n-1} \left[\hat{\lambda}_k (\mathbb{I}(u \in [k2^{-n}, (k+1)2^{-n}]) - 2^{-n}) \right. \right. \\ &\quad \left. \left. + \hat{\gamma}_k (\mathbb{I}(v \in [k2^{-n}, (k+1)2^{-n}]) - 2^{-n}) \right] - \hat{\lambda}_{2^n} h(u, v, \hat{\theta}) - b_0 \tilde{c}(u, v) \right\}, \end{aligned}$$

we can define the potential function as follows:

$$\begin{aligned} Q_{n, N_h}(\Lambda_n, \hat{\theta}) &= \int_{[0,1]^2} \exp \left\{ - \sum_{k=0}^{2^n-1} \left[\lambda_k (\mathbb{I}(u \in [k2^{-n}, (k+1)2^{-n}]) - 2^{-n}) \right. \right. \\ &\quad \left. \left. + \gamma_k (\mathbb{I}(v \in [k2^{-n}, (k+1)2^{-n}]) - 2^{-n}) \right] - \lambda_{2^n} h(u, v, \hat{\theta}) - b_0 \tilde{c}(u, v) \right\} dudv. \end{aligned}$$

Then, (D8) is equivalent to the following system of equations:

$$\begin{aligned} \frac{\partial}{\partial \lambda_k} Q_{n, N_h}(\Lambda_n, \hat{\theta}) &= 0, \\ \frac{\partial}{\partial \gamma_k} Q_{n, N_h}(\Lambda_n, \hat{\theta}) &= 0, \\ \frac{\partial}{\partial \lambda_{2^n}} Q_{n, N_h}(\Lambda_n, \hat{\theta}) &= 0 \end{aligned} \tag{D9}$$

for all $k = 0, \dots, (2^n - 1)$. Also note that, since the second order derivatives of $Q_{n,N_h}(\Lambda_n, \hat{\theta})$ is the covariance matrix of $\left\{ \mathbb{I}(u \in [k2^{-n}, (k+1)2^{-n}]), \mathbb{I}(v \in [k2^{-n}, (k+1)2^{-n}]), h(u, v, \hat{\theta}) \right\}_{k=0}^{2^n-1}$, thus $Q_{n,N_h}(\Lambda_n, \hat{\theta})$ is positive definite. It follows that the solutions to (D9) are the minimum values of $Q_{n,N_h}(\Lambda_n, \hat{\theta})$, which depend on $\hat{\theta}$, b_0 , and $\tilde{c}(u, v)$.

Since the potential function $Q_{n,N_h}(\Lambda_n, \hat{\theta})$ and the MEC (D7) are non-smooth, following common practice, they need to be smoothed out. We can obtain their smoothings by using a continuous approximation to the indicator function, $\sum_{k=0}^{2^n-1} \lambda_k \mathbb{I}(u \in [k2^{-n}, (k+1)2^{-n}])$, for a sufficiently large n . An application of Lemma B.4 yields

$$\begin{aligned} \mathbb{I}(u \in [k2^{-n}, (k+1)2^{-n}]) &= 1 - \mathbb{I}(u < k2^{-n}) - \mathbb{I}(u > (k+1)2^{-n}) \\ &= 1 - \lim_{N_h \rightarrow \infty} \frac{N_h}{\sqrt{2\pi}} \int_{-\infty}^{k2^{-n}} \exp\{-(x-u)^2 N_h^2 / 2\} dx \\ &\quad - \lim_{N_h \rightarrow \infty} \frac{N_h}{\sqrt{2\pi}} \int_{-\infty}^{-(k+1)2^{-n}} \exp\{-(x+u)^2 N_h^2 / 2\} dx \\ &= 1 - \lim_{N_h \rightarrow \infty} \Phi(N_h(k2^{-n} - u)) - \lim_{N_h \rightarrow \infty} \Phi(-N_h((k+1)2^{-n} - u)). \end{aligned}$$

We then immediately obtain:

$$\begin{aligned} Q_{n,N_h}(\Lambda_n, \hat{\theta}) &= \int_{[0,1]^2} \exp \left\{ \sum_{k=0}^{2^n-1} \left[\lambda_k \left(\Phi(N_h(k2^{-n} - u)) + \Phi(-N_h((k+1)2^{-n} - u)) - 1 + 2^{-n} \right) \right. \right. \\ &\quad \left. \left. + \gamma_k \left(\Phi(N_h(k2^{-n} - v)) + \Phi(-N_h((k+1)2^{-n} - v)) - 1 + 2^{-n} \right) \right] \right. \\ &\quad \left. - \lambda_{2^n} h(u, v, \hat{\theta}) - b_0 \tilde{c}(u, v) \right\} dudv, \end{aligned} \tag{D10}$$

and

$$\begin{aligned} \mathcal{E}_{n,N_h}(u, v) &\approx \exp \left\{ \sum_{k=0}^{2^n-1} \left[\hat{\lambda}_k \left(\Phi(N_h(k2^{-n} - u)) + \Phi(-N_h((k+1)2^{-n} - u)) \right) \right. \right. \\ &\quad \left. \left. + \hat{\gamma}_k \left(\Phi(N_h(k2^{-n} - v)) + \Phi(-N_h((k+1)2^{-n} - v)) \right) \right] \right. \\ &\quad \left. - \hat{\lambda}_{2^n} h(u, v, \hat{\theta}) - b_0 \tilde{c}(u, v) \right\}, \end{aligned} \tag{D11}$$

where $\hat{\Lambda}_n$ are the minimum values of (D10). In particular, $\hat{c}_{n,N_h}(u, v) = \frac{\mathcal{E}_{n,N_h}(u, v)}{\int_0^1 \int_0^1 \mathcal{E}_{n,N_h}(u, v) dudv}$ can be symmetrized by letting $\lambda_k = \gamma_k$ for every $k = 0, \dots, (2^n - 1)$ and letting $h(u, v, \hat{\theta})$ be a symmetric function.

Finally, to complete this proof, we still need to prove that the MEC approximator, $\hat{C}_{n,N_h}(u, v) = \int_0^u \int_0^v \hat{c}_{n,N_h}(u, v) dudv$, is 2-increasing. Let's denote by $[u_1, u_1 + \Delta] \times [v_1, v_1 + \Delta]$ a rectangle in $[0, 1]^2$, we immediately establish that, since $\hat{c}_{n,N_h}(u, v)$ is a (positive) exponential function, the mass of the rectangle, $\hat{C}_{n,N_h}(u_1 + \Delta, v_1 + \Delta) - \hat{C}_{n,N_h}(u_1 + \Delta, v_1) - \hat{C}_{n,N_h}(u_1, v_1 + \Delta) + \hat{C}_{n,N_h}(u_1, v_1) = \int_{u_1}^{u_1+\Delta} \int_{v_1}^{v_1+\Delta} \hat{c}_{n,N_h}(u, v) dudv$, is thus nonnegative. Now, we can obtain the MECs by letting n and N_h become sufficiently large. ■

Proof of Theorem 2.3: For all $\Theta_T \in \mathcal{M}$, $Q_n(\Lambda, \Theta_T)$ has a unique finite supremum for all T in view of AS1 and AS2. AS2 and $\hat{\Theta}_T \xrightarrow{p} \Theta^0$ implies that $Prob\{\hat{\Lambda}_T \in \partial \mathcal{N}\} \rightarrow 0$. Thus, $Q_n(\Lambda, \hat{\Theta}_T)$ has

a unique interior supremum $\widehat{\Lambda}_T$ for a sufficiently large T . Let Λ^0 denote the unique supremum of $Q_n(\Lambda, \Theta^0)$. In view of **AS3**, $\widehat{\Theta}_T \xrightarrow{p} \Theta^0$ implies $\widehat{\Lambda}_T \xrightarrow{p} \Lambda^0$.

An application of the mean-value theorem yields:

$$\nabla Q_n(\widehat{\Lambda}_T, \Theta^0) = \nabla Q_n(\Lambda^0, \Theta^0) + \mathcal{H}_{1,n}(\widehat{\Lambda}_T^*, \Theta^0)(\widehat{\Lambda}_T - \Lambda^0),$$

where $\min(\widehat{\Lambda}_T, \Lambda^0) < \widehat{\Lambda}_T^* < \max(\widehat{\Lambda}_T, \Lambda^0)$. Thus, we have

$$\widehat{\Lambda}_T - \Lambda^0 = \mathcal{H}_{1,n}^{-1}(\widehat{\Lambda}_T^*, \Theta^0) \nabla Q_n(\widehat{\Lambda}_T, \Theta^0).$$

Another application of the mean-value theorem yields:

$$\nabla Q_n(\widehat{\Lambda}_T, \widehat{\Theta}_T) = \nabla Q_n(\widehat{\Lambda}_T, \Theta^0) + \mathcal{H}_{2,n}(\widehat{\Lambda}_T, \widehat{\Theta}_T^*)(\widehat{\Theta}_T - \Theta^0),$$

where $\min(\widehat{\Theta}_T, \Theta^0) < \widehat{\Theta}_T^* < \max(\widehat{\Theta}_T, \Theta^0)$. Thus, we obtain

$$T^{1/2}(\widehat{\Lambda}_T - \Lambda^0) = -T^{1/2} \mathcal{H}_{1,n}^{-1}(\widehat{\Lambda}_T^*, \Theta^0) \mathcal{H}_{2,n}(\widehat{\Lambda}_T, \widehat{\Theta}_T^*)(\widehat{\Theta}_T - \Theta^0).$$

Since $\widehat{\Theta}_T \xrightarrow{p} \Theta^0$ implies $\widehat{\Lambda}_T \xrightarrow{p} \Lambda^0$, the continuous mapping theorem yields:

$$\mathcal{H}_{1,n}(\widehat{\Lambda}_T^*, \Theta^0) \xrightarrow{p} \mathcal{H}_{1,n}(\Lambda^0, \Theta^0),$$

$$\mathcal{H}_{2,n}(\widehat{\Lambda}_T, \widehat{\Theta}_T^*) \xrightarrow{p} \mathcal{H}_{2,n}(\Lambda^0, \Theta^0).$$

Hence, Slutsky's theorem and **AS4** yield

$$T^{1/2}(\widehat{\Lambda}_T - \Lambda^0) \xrightarrow{d} N(0, \mathcal{H}_{1,n}^{-1}(\Lambda^0, \Theta^0) \mathcal{H}_{2,n}(\Lambda^0, \Theta^0) \Psi \mathcal{H}_{2,n}'(\Lambda^0, \Theta^0) \mathcal{H}_{1,n}^{-1}(\Lambda^0, \Theta^0)).$$

■

References

1. Joe, H. Relative entropy measures of multivariate dependence. *J. Am. Stat. Assoc.* **1989**, *84*, 157–164.
2. Granger, C.; Lin, J.L. Using the mutual information coefficient to identify lags in nonlinear models. *J. Time Ser. Anal.* **1994**, *15*, 371–384.
3. De la Peña, V.H.; Ibragimov, R.; Sharakhmetov, S. Characterizations of joint distributions, copulas, information, dependence and decoupling, with applications to time series. In *Optimality, Institute of Mathematical Statistics Lecture Notes—Monograph Series 49*; The Institute of Mathematical Statistics: Beachwood, OH, USA, 2006; pp. 183–209.
4. Golan, A. Information and entropy econometrics: Editor's view. *J. Econom.* **2002**, *107*, 1–15.
5. Sarabia, J.M.; Gomez-Deniz, E. Construction of multivariate distributions: A review of some recent results. *Stat. Oper. Res. Trans.* **2008**, *32*, 3–36.
6. Golan, A. Information and entropy econometrics—Volume overview and synthesis. *J. Econom.* **2007**, *138*, 379–387.
7. Usta, I.; Kantar, Y.M. On the performance of the flexible maximum entropy distributions within partially adaptive estimation. *Comput. Stat. Data Anal.* **2011**, *55*, 2172–2182.
8. Jaynes, E.T. Information theory and statistical mechanics. *Phys. Rev.* **1957**, *106*, 620–630.
9. Rockinger, M.; Jondeau, E. Entropy densities with an application to autoregressive conditional skewness and kurtosis. *J. Econom.* **2002**, *106*, 119–142.
10. Maasoumi, E.; Racine, J. Entropy and predictability of stock market returns. *J. Econom.* **2002**, *107*, 291–312.
11. Hang, R.K. Maximum entropy estimation of density and regression functions. *J. Econom.* **1993**, *56*, 397–400.
12. Miller, D.J.; Liu, W. On the recovery of joint distributions from limited information. *J. Econom.* **2002**, *107*, 259–274.

13. Patton, A.J. On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. *J. Financ. Econom.* **2004**, *2*, 130–168.
14. Rodriguez, J.C. Measuring financial contagion: A copula approach. *J. Empir. Financ.* **2007**, *14*, 401–423.
15. Chollete, L.; Heinen, A.; Valdesogo, A. Modeling international financial returns with a multivariate regime-switching copula. *J. Financ. Econom.* **2009**, *7*, 437–480.
16. Ning, C.; Xu, D.; Wirjanto, T.S. Is volatility clustering of asset returns asymmetric? *J. Bank. Financ.* **2015**, *52*, 62–76.
17. Patton, A.J. Copula-based models for financial time series. In *Handbook of Financial Time Series*; Mikosch, T., Kreiß, J.-P., Davis, R.A., Andersen, T.G., Eds.; Springer-Verlag: Berlin, Heidelberg, Germany, 2009; pp. 767–785.
18. Fan, Y.; Patton, A.J. Copulas in econometrics. *Annu. Rev. Econ.* **2014**, *6*, 179–200.
19. Nelsen, R.B. *An Introduction to Copulas*; Springer-Verlag: New York, NY, USA, 1998.
20. Chu, B. Recovering copulas from limited information and an application to asset allocation. *J. Bank. Financ.* **2011**, *35*, 1824–1842.
21. Dempster, M.A.H.; Medova, E.A.; Yang, S.W. Empirical copulas for CDO tranche pricing using relative entropy. *Int. J. Theor. Appl. Financ.* **2007**, *10*, 679–701.
22. Friedman, C.A.; Huang, J. Most Entropic Copulas: General Form, and Calibration to High-Dimensional Data in an Important Special Case. *SSRN Electron. J.* **2010**, doi:10.2139/ssrn.1725517.
23. Veremyev, A.; Tsyurmasto, P.; Uryasev, S.; Rockafellar, R.T. Calibrating probability distributions with convex-concave-convex functions: Application to CDO pricing. *Comput. Manag. Sci.* **2014**, *11*, 341–364.
24. Zhao, N.; Lin, W.T. A copula entropy approach to correlation measurement at the country level. *Appl. Math. Comput.* **2011**, *218*, 628–642.
25. Golan, A.; Judge, G.; Miller, D. *Maximum Entropy Econometrics: Robust Estimation with Limited Data*; John Wiley & Sons: New York, NY, USA, 1996.
26. Wu, X. Calculation of maximum entropy densities with application to income distribution. *J. Econom.* **2003**, *115*, 347–354.
27. Wu, X.; Perloff, J.M. GMM estimation of a maximum entropy distribution with interval data. *J. Econom.* **2007**, *138*, 532–546.
28. Zellner, A.; Highfield, R.A. Calculation of maximum entropy distributions and approximation of marginal posterior distributions. *J. Econom.* **1988**, *37*, 195–209.
29. Sklar, A. *Fonctions de Repartition n Dimensions et Leurs Marges*; Publications de l'Institut Statistique de l'Universit de Paris: Paris, France, 1959; pp. 229–231.
30. Hajek, J.; Sidak, Z. *Theory of Rank Tests*; Academic Press: New York, NY, USA, 1967.
31. Genest, C.; Plante, J.-F. On Blest's measure of rank correlation. *Can. J. Stat.* **2003**, *31*, 1–18.
32. Gideon, R.A.; Hollister, R.A. A rank correlation coefficient resistant to outliers. *J. Am. Stat. Assoc.* **1987**, *82*, 656–666.
33. Troutman, J.L. *Variational Calculus and Optimal Control*, 2nd ed.; Springer: New York, NY, USA; Berlin, Heidelberg, Germany, 1996.
34. Csendes, T. Nonlinear parameter estimation by global optimization—Efficiency and reliability. *Acta Cybern.* **1988**, *8*, 361–370.
35. Hardel, W.; Muller, M.; Sperlich, S.; Werwatz, A. *Nonparametric and Semiparametric Models*; Springer Series in Statistics; Springer-Verlag: Berlin, Heidelberg, Germany; New York, NY, USA, 2004.
36. Omelka, M.; Gijbels, I.; Veraverbeke, N. Improved kernel estimation of copulas: Weak convergence and goodness-of-fit testing. *Ann. Stat.* **2009**, *37*, 3023–3058.
37. Genest, C. Frank's family of bivariate distributions. *Biometrika* **1987**, *74*, 549–555.
38. Nelsen, R.B. Properties of a one-parameter family of bivariate distributions with specified marginals. *Commun. Stat. Theory Methods* **1986**, *15*, 3277–3285.
39. Chen, S.X.; Huang, T.-M. Nonparametric estimation of copula functions for dependence Modelling. *Can. J. Stat.* **2007**, *35*, 265–282.
40. Gijbels, I.; Mielniczuk, J. Estimating the density of a copula function. *Commun. Stat. - Theory Methods* **1990**, *19*, 445–464.
41. Pagan, A.; Ullah, A. *Nonparametric Econometric*, 1st ed.; Themes in Modern Econometrics; Cambridge University Press: Cambridge, UK, 1999.

42. Shiryaev, A.N. *Probability*, 2nd ed.; Graduate text in mathematics; Springer-Verlag: New York, NY, USA; Berlin, Heidelberg, Germany, 1995; Volume 95.
43. Ioffe, A.D.; Tihomirov, V.M. *Theory of Extremal Problems*; Lions, J.L., Papanocolaou, G., Rockafellar, R.T., Eds.; Studies in mathematics and its applications; North Holland Publishing Company: Amsterdam, The Netherlands; New York, NY, USA; Oxford, UK, 1979; Volume 6.
44. Kutoyants, Y.A. *Statistical Inference for Ergodic Diffusion Processes*; Springer series in statistics; Springer-Verlag: London, UK; Berlin, Heidelberg, Germany, 2004.
45. Abramowitz, M.; Stegun, I.A. *Handbook of Mathematical Functions*; Dover Publications: New York, NY, USA, 1972.
46. Bertsekas, D.P. *Convex Analysis and Optimization*; Athena Scientific: Belmont, MA, USA, 2003.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).