The Most Entropic Canonical Copula with An Application to 'Style' Investment *

Ba Chu†

Carleton University

September 9, 2011

Abstract

We propose a new approach to recover relative entropy measures of dependence from limited information by constructing the most entropic copulas (MECs) and their canonical form, namely the most entropic canonical copula (MECC). In the empirical study, we focus on an application of the MECC theory to a 'style investing' problem for an investor with a constant relative risk aversion (CRRA) utility function allocating wealth between the Russell 1000 'growth' and 'value' indices. We found that, using the data in hand, the gains from using the MECC (vis-à-vis commonly used parametric copulas) to model the dependence between the indices' returns for our investment strategies are economically and statistically significant for the case with/without short-sales constraints.

JEL classification: C190; C590; C130.

Keywords: Entropy, Relative entropy measure of joint dependence, Copula, Most entropic copula, Canonical, Kullback-Leibler cross entropy, 'growth' index, 'value' index.

*The author would like to sincerely thank the editor (René Garcia), an associate editor, and two anonymous referees for their valuable comments that led to corrections, clarifications, and various improvements in this version of the paper. Special thanks also go to Lynda Khalaf for her interest in discussing every aspect of the work, always resulting in new thoughts and ideas. The author is particularly indebted to John Knight, Kim P. Huỳnh, and David T. Jacho-Chávez for having carefully read the previous manuscripts and provided many constructive suggestions. The remaining errors are the responsibilities of the author. The usual disclaimer applies.

†Department of Economics, Carleton University, B-857 Loeb Building, 1125 Colonel By Drive, Ottawa, ON K1S 5B6, Canada. Phone: +1 (613) 520 2600 (ext. 1546). Fax: +1 (613) 520 3906. E-mail: ba_chu@carleton.ca. Web Page: http://http-server.carleton.ca/~bchu/.

1 INTRODUCTION

Modeling dependence between asset returns is crucial in a number of financial applications. For a number of standard distributions in the elliptical family, dependence is simply measured by Pearson's correlation coefficient. In practice, however, asset returns do not always belong to the elliptical family, and thus the dependence structure does not always show out of the joint distribution function under consideration. It would be useful to separate the statistical properties of each return from their dependence structure. Copulas provide us with a viable way to achieve this goal. Literature on copula and its applications in economics is particularly extensive, and is relatively well-known to readers, so we shall not attempt to survey it here. Contributions in this field that are closely related to the topic of the present paper include, among many others, Patton (2004), Sancetta and Satchell (2004), Jondeau and Rockinger (2006), Patton (2006), Scaillet (2007), Dempster et al. (2007), and Chollete et al. (2009).

Using Shannon's (1948) entropy to construct probability distributions has been widely accepted in economics. (See, e.g., Hang (1993), Rockinger and Jondeau (2002), Golan (2002), Wu and Stengos (2005), and Golan (2007), to mention a few.) Closer in spirit to the present paper is the paper by Miller and Liu (2002), who use the Kullback-Leibler cross entropy (KLCE) distance to recover joint probability distributions while disposing only of information on certain joint moments. This paper constitutes a continuation of work on so-called maximum entropy distributions; in particular, we employ Shannon's (1948) entropy to recover copulas from limited information. In the present context, we may view our problem of recovering the most entropic copulas (MECs) as that of minimizing the KLCE distance for Uniform[0,1] random variables subject to the Uniform[0,1] marginal constraints.

To be precise the MEC is defined as a copula, $\widehat{c}(u, v)$, which, given an amount of prior information on dependence, maximizes the amount of information that we will be receiving (putting it simply, suppose that we do not assume any prior information at all, a natural candidate of the MEC is obviously uv – that is, the random variables are independent). Since copula is a joint uniform distribution, we can derive the MEC by maximizing the Shannon entropy of a joint distribution with its domain in a unit cube subject to the constraints that this joint distribution has Uniform[0,1]

marginals, and that the measures of association are equal to their nonparametric estimates, herein, conceived as rank correlations [e.g., Blest's rank correlations (see Blest (2000) and Genest and MacKay (1986))]. In addition, there exists, in the class of the MECs, a simple form called the most entropic canonical copula (MECC).

In empirical analysis, we applied the MECC theory to examine the problem of an investor with constant relative risk aversion (CRRA) investing in two 'style' equity portfolios (i.e., Russell 1000 'growth' and 'value' indices, which are typical of high risk-low return and low risk-high return indices respectively[1]). This problem is of practical interest because 'style investing', where portfolio managers rotate between 'value' and 'growth' stocks, has recently become popular in Wall Street. In the broad context of optimal asset allocation using copulas, our study is an extension of a recent empirical study of Patton (2004), who focuses on the 'large caps' and 'small caps' investment problem.

The impact of dynamic dependence among asset returns in a portfolio on the optimal weights is evident.[2] Thus, to effectively capture the dynamic dependence structure, which may be rather complex, is crucial for asset allocation. In this spirit, we demonstrate that the gains from using the MECC – measured by comparing the expected utility of a portfolio based on the MECC with that based on the Normal, Clayton or Gumbel copulas – are economically and statistically significant.

Although, in the present paper, we mostly concentrate on the bivariate MECC, the multivariate MECC can be feasibly computed by maximizing the multivariate Shannon entropy subject to prior information expressed via multivariate rank correlations. Hence, the results presented here are in line with those for the multivariate case. The rest of this paper is organized as follows. Section 2 includes brief descriptions of entropy and copula. Section 3 defines the MECs and the MECC; and also proposes its approximator with asymptotic theory. Some simulations and an empirical application are presented in Sections 4 and 5, confirming that our approach is feasible. Section 6 concludes this paper. Last but not least, to make the current paper as short and informative as possible, we shall delegate the results essential for the paper and results of technical flavor to

[1]The full-sample means and the full-sample standard deviations are 0.0051 and 0.0497 for the 'growth' index, and 0.0072 and 0.0382 for the 'value' index, respectively.

[2]For instance, in the mean-variance framework, the optimal proportions of a portfolio allocated to individual assets are functions of the expected returns, the variances, and the linear correlations.

the appendices of the paper. Also to keep the empirical section parsimonious and solid, we shall unfortunately allow ourself not to report some trivial statistical results. Nevertheless, they are all available from us upon request.

2 MAXIMUM ENTROPY AND COPULA

This section provides a brief explanation of entropy and copula. We refer to the Appendix for further details.

The *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*, Zellner and Highfield (1988), Wu (2003), and Wu and Perloff (2007)]. A univariate ME density is generally obtained by maximizing the 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., Joe (1989) and Miller and Liu (2002)):

$$\min_f KLCE(f:g) = \min_f \int_{\mathbb{R}^2} f(X,Y)\log\frac{f(X,Y)}{g_1(X)g_2(Y)}dXdY \tag{2.1}$$

subject to

$$\int_{\mathbb{R}^2} h(X,Y)f(X,Y)dXdY = \mu_0,$$

where f is a bivariate density, g_1 and g_2 are some univariate densities, and h is an arbitrary function such that $\mu_0 < \infty$.

Copula is proposed by Sklar (1959) 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. Throughout this paper, we denote a bivariate copula of (u,v) by $C(u,v)$ and its density by $c(u,v)$. Since, by definition, $C(1,1)=1$, $C(u,1)=u$, and $C(1,v)=v$, a copula, $C(u,v)$, must be a bivariate Uniform[0,1] distribution with the Uniform[0,1] marginals. In the same way, a multivariate copula is a multivariate Uniform[0,1]

distribution. By Sklar's theorem, the exact link between a copula, $C(u,v)$, and a joint distribution, $F(X,Y)$, is $F(X,Y) = C(G_1(X), G_2(Y))$, where G_1 and G_2 are the marginals.

As we have mentioned in the Introduction that we use measures of association and rank correlations in constructing the MEC, thus it is useful, at this stage, to discuss about those quantities. 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 the Appendix for formal definitions of measures of association). A measure of association is, in general, defined as $\tau = \int_0^1 h(u,v)dC(u,v)$, 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 $\widehat{\tau} = \frac{1}{T}\sum_{t=1}^{T} h(\frac{R_t}{T}, \frac{S_t}{T})$, where (R_t, S_t) denotes the rank of (X_t, Y_t) in the sample of size T. A pro of using rank statistics as nonparametric measures of nonlinear dependence is that they are robust – where "robust" means to be insensitive to contamination and to maintain a high efficiency for heavier tailed elliptical distributions as well as for multivariate normal distributions. For helpful ideas on the topic of rank statistics, we refer to the monograph by Hajek and Sidak (1967).

In particular, if $h(u,v)$ is a self-decomposable function, i.e., $h(u,v) = J(u)J(v)$, we obtain:

$$\widehat{\tau} = \frac{1}{T}\sum_{t=1}^{T} J(\frac{R_t}{T})J(\frac{S_t}{T}),$$

where J is also called a standardized score function (see, e.g., Puri et al. (1970)). For instance, when the score function is $J(u) = \sqrt{12}(u - 1/2)$, then $\widehat{\tau} = \frac{12}{T(T^2-1)}\sum_{t=1}^{T}\left(R_t - \frac{T+1}{2}\right)\left(S_t - \frac{T+1}{2}\right)$ is Spearman's rank correlation with asymptotic mean given by Spearman's rho. Other special cases of the above-mentioned rank statistic include Blest's rank correlations (see, e.g., Genest and Plante (2003)), which are summarized in the following table:

Nonetheless, it is worth mentioning that not every rank correlation can be formulated in terms of the above general rank statistic $\widehat{\tau}$. For instance, the statistic $\widehat{R}_g$, which was proposed by Gideon and Hollister (1987) as a coefficient of rank correlation resistant to outliers even in a small sample,

Measures of association	Rank correlation
Spearman's rho: $\rho_S = 12\int_{[0,1]^2} uvc(u,v)dudv - 3$, and $\rho_S \in [-1,1]$,	$\widehat{\rho}_S = \frac{12}{T^3-T}\sum_{t=1}^{T} R_t S_t - 3\frac{T+1}{T-1}$.
Blest's measure I: $\nu_1 = 2 - 12\int_{[0,1]^2}(1-u)^2 vc(u,v)dudv$, and $\nu_1 \in [-1,1]$,	$\widehat{\nu}_1 = \frac{2T+1}{T-1} - \frac{12}{T^2-T}\sum_{t=1}^{T}\left(1-\frac{R_t}{T+1}\right)^2 S_t$.
Blest's measure II: $\nu_2 = 2-12\int_{[0,1]^2} u(1-v)^2 c(u,v)dudv$, and $\nu_2 \in [-1,1]$,	$\widehat{\nu}_2 = \frac{2T+1}{T-1} - \frac{12}{T^2-T}\sum_{t=1}^{T} R_t\left(1-\frac{S_t}{T+1}\right)^2$.
Blest's measure III: $\eta = 6\int_{[0,1]^2} u^2v^2c(u,v)dudv - \frac{1}{5}$, and $\eta \in [0,1]$,	$\widehat{\eta} = \frac{6}{T^2-T}\sum_{t=1}^{T}\left(\frac{R_t}{T+1}\right)^2\left(\frac{S_t}{T+1}\right)^2 - \frac{(1/5)T+1}{T-1}$.
Blest's measure IV: $\phi = \int_{[0,1]^2}[10(1-u)^3 v - 3u^2v^2]c(u,v)dudv - 9/10$, and $\phi \in [-1,1]$),	$\widehat{\phi} = \frac{1}{T^2-T}\sum_{t=1}^{T}\left[\left(1-\frac{R_t}{T+1}\right)^3\frac{S_t}{T+1} - \left(\frac{R_t}{T+1}\right)^2\left(\frac{S_t}{T+1}\right)^2\right] - \frac{0.9T+1}{T-1}$.

has the form:

$$\widehat{R}_g = \frac{1}{[T/2]}\left(\max_t \sum_{s=1}^{t} \mathbf{1}(p_s < T+1-t) - \max_t \sum_{s=1}^{T} \mathbf{1}(R_s \le t < S_s)\right),$$

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

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

$$\mathbf{W}(c) = -\int_{[0,1]^2} c(u,v)\log c(u,v)dudv, \text{ where } c(u,v) = \frac{\partial^2 C(u,v)}{\partial u \partial v}. \tag{2.2}$$

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 $\widehat{c}(u,v)$ denote the MEC. Then, in view of Joe (1989), the relative entropy measure of dependence [recovered from limited information] is given by $-\mathbf{W}(\widehat{c})$. Generally speaking, a multivariate Shannon entropy can be defined in an obvious way, and this dual relationship holds.

3 PROPOSED METHOD

3.1 The Most Entropic Copula

The bivariate MEC (*or* the MEC) is obtained by maximizing the bivariate Shannon entropy (2.2) 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, are set equal to the corresponding rank correlations. We call this the Problem A.

$$\textbf{Problem A:} \qquad \text{Maximizing} \quad \mathbf{W}(c) = -\int_{[0,1]^2} c(u,v)\log c(u,v)dudv \tag{3.1}$$

subject to

$$\int_{[0,1]^2} c(u,v)dudv = 1, \tag{3.2}$$

$$\int_{(0,u]}\int_{[0,1]} c(x,v)dxdv = u, \ \forall\ u \in [0,1], \tag{3.3}$$

$$\int_{[0,1]}\int_{(0,v]} c(u,y)dudy = v, \ \forall\ v \in [0,1], \tag{3.4}$$

$$\int_{[0,1]^2} h(u,v;\widehat{\boldsymbol{\theta}}_T)c(u,v)dudv = 0, \tag{3.5}$$

where (3.2) implies that $c(u,v)$ is a joint density on the unit circle; Eqs. (3.3) and (3.4) imply that the marginals of $c(u,v)$ are Uniform[0,1] distributions; Eq. (3.5) imposes a constraint on the joint behavior of U and V. To give an example, let $h(u,v;\widehat{\boldsymbol{\theta}}_T) = 12uv - 3 - \widehat{\rho}_S$, then the left-hand side of (3.5) becomes Spearman's rho, as defined in Section 2, and $\widehat{\boldsymbol{\theta}}_T = \widehat{\rho}_S$ (note that, in what follows, we sometimes omit 'T' 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 pick up a copula, $C_1(u,v;\boldsymbol{\beta})$, from $\mathcal{C}_0$, then use it to construct (3.5) with $h(u,v;\widehat{\boldsymbol{\theta}}) = 4C_1(u,v;\widehat{\boldsymbol{\beta}}) - 1 - \widehat{\tau}$, where $\widehat{\boldsymbol{\theta}} = \{\widehat{\boldsymbol{\beta}}, \widehat{\tau}\}'$ and $\widehat{\tau}$ is an estimate of the difference between the probabilities of concordance

and discordance (cf. Appendix A). By doing this way, it is expected that some feature of the family $\mathcal{C}_0$ could be effectively incorporated into the MECC. Other examples of Eq. (3.5) also include Blest's coefficients or Gideon and Hollister's (1987) coefficient, etc. Also note that we may have more than one constraint like (3.5).

For future references, we shall denote by $\widehat{c}(u,v) = c(u,v,\widehat{\Lambda})$, where $\widehat{\Lambda}$ is a vector of coefficients, as the MEC [that solves Problem A]. The MECs (accordingly the MECC) can then be approximated by replacing the continuums of varying end-points in (3.3) and (3.4) by sets of definite integrals. We now present an approximate solution to Problem A in Theorem 3.1 below.

THEOREM 3.1. *The MEC, $\widehat{c}(u,v)$, can be approximated by an approximator, $\widehat{c}_{n,N_h}(u,v)$, as follows:*

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

with

$$\widehat{c}_{n,N_h}(u,v) = \frac{\mathcal{E}_{n,N_h}(u,v)}{\int_{[0,1]^2} \mathcal{E}_{n,N_h}(u,v)dudv}, \tag{3.6}$$

where

$$\begin{aligned}\mathcal{E}_{n,N_h}(u,v) &= \exp\Bigg\{-\sum_{k=0}^{2^n-1}\Bigg[\widehat{\lambda}_k\big(\Phi(N_h(k2^{-n}-u)) + \Phi(-N_h((k+1)2^{-n}-u))\big) \\ &+ \widehat{\gamma}_k\big(\Phi(N_h(k2^{-n}-v)) + \Phi(-N_h((k+1)2^{-n}-v))\big)\Bigg] - \widehat{\lambda}_{2^n} h(u,v,\widehat{\boldsymbol{\theta}}) - b_0\widetilde{c}(u,v)\Bigg\}\end{aligned} \tag{3.7}$$

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

$$\begin{aligned}Q_{n,N_h}(\boldsymbol{\Lambda}_n, \widehat{\boldsymbol{\theta}}) &= \int_{[0,1]^2} \exp\Bigg\{-\sum_{k=0}^{2^n-1}\Bigg[\lambda_k\big(\Phi(N_h(k2^{-n}-u)) + \Phi(-N_h((k+1)2^{-n}-u)) - 1 + 2^{-n}\big) \\ &+ \gamma_k\big(\Phi(N_h(k2^{-n}-v)) + \Phi(-N_h((k+1)2^{-n}-v)) - 1 + 2^{-n}\big)\Bigg] \\ &- \lambda_{2^n} h(u,v,\widehat{\boldsymbol{\theta}}) - b_0\widetilde{c}(u,v)\Bigg\} dudv \text{ for a given } b_0 \text{ and } \widetilde{c}(u,v).\end{aligned} \tag{3.8}$$

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 in-

dicator functions, $\mathbb{I}(u \in [k2^{-n}, (k+1)2^{-n}])$*, with the Gaussian kernel) and* $\widetilde{c}(u,v)$ *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 *(*$\forall\ k = 1, \ldots, (2^n - 1)$*) and* $h(u, v, \widehat{\boldsymbol{\theta}})$ *be a symmetric function.*

Proof. The proof is presented in the appendix. □

As we can see, the MEC density nests an arbitrary copula, $\widetilde{c}(u,v)$, (cf. Eq. (3.7)). Indeed, the MEC depends on both b_0 and $\widetilde{c}(u,v)$, 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 *canonical* model can be traced back to Jeffreys[3] 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 $Q_{n,N_h}(\boldsymbol{\Lambda}, \widehat{\boldsymbol{\theta}})$ in the above theorem is a multivariate convex function of $\boldsymbol{\Lambda}$, thus 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_t, Y_t) for $t = 1, \ldots, T$, the average maximum log-likelihood function is given by

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

where $\widehat{c}_{n,N_h}(u_t, v_t) \doteq \widehat{c}_{n,N_h}\left(u_t, v_t, \widehat{\boldsymbol{\Lambda}}_n\right)$ is defined in (3.6),

$$u_t = \frac{1}{T}\sum_{s=1}^{T} \mathbf{1}(X_s \leq X_t) = \frac{R_t}{T+1},$$

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

and

$$v_t = \frac{1}{T}\sum_{s=1}^{T}\mathbf{1}(Y_s \le Y_t) = \frac{S_t}{T+1},$$

in which R_t and S_t are the ranks of X_t and Y_t in the sample, respectively. Assuming that T is greater than n and that n is large enough, in view of (3.7) with $b_0 = 0$, we obtain the following representation:

$$\begin{aligned}
\boldsymbol{\ell}(\widehat{\boldsymbol{\Lambda}}_n) &= -\frac{1}{T}\sum_{t=1}^{T}\Bigg(\widehat{\lambda}_{-1} + \sum_{k=0}^{2^n-1}\Bigg(\Big[\widehat{\lambda}_k\big(\Phi(N_h(k2^{-n}-u_t)) + \Phi(-N_h((k+1)2^{-n}-u_t))\big) \\
&+ \widehat{\gamma}_k\big(\Phi(N_h(k2^{-n}-v_t)) + \Phi(-N_h((k+1)2^{-n}-v_t))\big)\Big] + \widehat{\lambda}_{2^n}h(u_t,v_t)\Bigg)\Bigg) \\
&\approx -\left(\widehat{\lambda}_{-1} + \frac{2}{2^n}\sum_{k=0}^{2^n-1}\left(\widehat{\lambda}_k + \widehat{\gamma}_k\right) + \widehat{\lambda}_{2^n}\frac{1}{T}\sum_{t=1}^{T}h(u_t,v_t)\right) \\
&= -\mathbf{W}(\widehat{c}(u,v)),
\end{aligned}$$

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

REMARK 3.1. *To compute the MECC, we could use either a Monte-Carlo integration procedure or Gaussian quadratures to approximate the potential function* (3.8) *(see Appendix for further details), and then employ a global optimization technique (the stochastic search algorithm proposed by Csendes (1988)) 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 3.2:

THEOREM 3.2. *The MEC, $\widehat{c}(u,v)$, can be approximated by an approximator, $\widehat{c}_{N,h}(u,v)$, as follows:*

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

with

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

where

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

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 6.3) and $\widehat{\boldsymbol{\Lambda}}_N = \{\widehat{\lambda}_1, \ldots, \widehat{\lambda}_N, \widehat{\gamma}_1, \ldots, \widehat{\gamma}_N\}$ contains the minimal values of the following potential function:

$$\begin{aligned}Q_{N,h}(\boldsymbol{\Lambda}_N, \widehat{\boldsymbol{\theta}}) &= \int_{[0,1]^2} \exp\left\{-\frac{1}{N}\sum_{k=1}^{N}\left[\lambda_k\left(\frac{1}{h}\int_{\frac{k-1}{N}}^{\frac{k}{N}} K\left(\frac{u-w}{h}\right)dw - \frac{1}{N}\right)\right.\right. \\ &+ \left.\left.\gamma_k\left(\frac{1}{h}\int_{\frac{k-1}{N}}^{\frac{k}{N}} K\left(\frac{v-w}{h}\right)dw - \frac{1}{N}\right)\right] - \lambda_{2^n} h(u,v,\widehat{\boldsymbol{\theta}}) - b_0\widetilde{c}(u,v)\right\}dudv \\ &\quad \textit{for a given } b_0 \textit{ and } \widetilde{c}(u,v).\end{aligned}$$

Proof. The proof is reminiscent of that of Theorem 3.1 combining with Lemma 5. So we shall omit its details here. □

3.2 Convergence Rates of the MECC

For the estimator $\widehat{c}_{N,h}(u,v)$ to converge to the true MEC density $c_0(u,v)$ (as defined below), one will need to allow that $N = 2^n$ grows with T and $h = 1/N_h$ shrinks with T at certain rates. (It is important to note that, in what follows, we shall let N_T, h_T, and $\widehat{c}_T(u,v)$ signify the dependence of N, h, and $\widehat{c}_{N,h}(u,v)$, respectively, on T.) For readability, we herein list some spaces of copulas and their approximation spaces, which are also defined in the main proof.

- The space of copulas:

$\mathfrak{C} \doteq \Big\{ c \in W_1^r([0,1]^2) : \ c \in (0,\infty); \ \int_0^1 \int_0^1 c(u,v)dudv = 1; \int_0^1 c(u,v)du = 1;$ $\int_0^1 c(u,v)dv = 1$ for every $u, v \in [0,1]\}$, where $W_1^r([0,1]^2)$ represents the space of functions, c, whose partial derivatives up to an order, r, exist and lie in $\ell^1(\mathbb{R})$.

- The space of the MECs:

$$\mathcal{C} \doteq \Bigg\{ c \in W_1^r([0,1]^2) : \ c \in (0,\infty), \ E[\mathbf{1}] = 1, \ \int_\epsilon^1 \lambda(u) \left\{ \int_{u-\epsilon}^u \int_0^1 c(w,v)dwdv - \epsilon \right\} = 0,$$
$$\int_\epsilon^1 \gamma(v) \left\{ \int_0^1 \int_{v-\epsilon}^v c(u,w)dudw - \epsilon \right\} = 0, \text{ and } E[h(U,V,\boldsymbol{\theta}_0)] = 0, \text{ where } \lambda(u), \gamma(v) \in C^*([0,1]) \Bigg\},$$

where ϵ is an arbitrarily small constant and $C^*([0,1])$ is a space of continuously smooth functions vanishing at the end points. Hereafter the true MEC density is defined as $c_0 = \arg\sup_{c \in \mathcal{C}} \{-E[\log c(U,V)]\}$.

- The approximation space of copulas:

$$\begin{aligned} \mathfrak{C}_T \ \doteq \ & \Bigg\{ c \in W_1^r([0,1]^2) : \ c \in (0,\infty), \ E[\mathbf{1}] = 1, \ \frac{1}{N_T} \sum_{k=1}^{N_T} \lambda_k \left\{ \int_{\frac{k-1}{N_T}}^{\frac{k}{N_T}} \int_0^1 c(w,v)dwdv - \frac{1}{N_T} \right\} = 0, \\ & \text{and } \frac{1}{N_T} \sum_{k=1}^{N_T} \gamma_k \left\{ \int_0^1 \int_{\frac{k-1}{N_T}}^{\frac{k}{N_T}} c(u,w)dudw - \frac{1}{N_T} \right\} = 0 \text{ with } \{\lambda_k, \gamma_k\} \in B_\lambda \times B_\gamma \Bigg\}, \end{aligned}$$

where the λ_k's and the γ_k's represent some values of $\lambda(u)$ and $\gamma(v)$ respectively on $\left[\frac{k-1}{N_T}, \frac{k}{N_T}\right)$; and B_λ and B_γ are some bounded sets of $\mathbb{R}$.

- The kernel-smoothed approximation space of copulas:

$$\begin{aligned} \mathfrak{C}_T^* \ \doteq \ & \Bigg\{ c \in W_1^r([0,1]^2) : \ c \in (0,\infty), \ E[\mathbf{1}] = 1, \ \frac{1}{N_T} \sum_{k=1}^{N_T} \lambda_k \left\{ E\left[\frac{1}{h_T} \int_{\frac{k-1}{N_T}}^{\frac{k}{N_T}} K\left(\frac{u-w}{h_T} \right) dw \right] \right. \\ & - \left. \frac{1}{N_T} \right\} = 0, \text{ and } \frac{1}{N_T} \sum_{k=1}^{N_T} \gamma_k \left\{ E\left[\frac{1}{h_T} \int_{\frac{k-1}{N_T}}^{\frac{k}{N_T}} K\left(\frac{v-w}{h_T} \right) dw \right] - \frac{1}{N_T} \right\} = 0 \\ & \text{with } \{\lambda_k, \gamma_k\} \in B_\lambda \times B_\gamma \Bigg\}. \end{aligned}$$

- The kernel-smoothed approximation space of the MECs:

$$\mathcal{C}_T^*(\widehat{\boldsymbol{\theta}}) \doteq \left\{ c \in \mathfrak{C}_T^* : \; E[h(U,V,\widehat{\boldsymbol{\theta}})] = 0 \right\}.$$

Theorem 3.3 (below) provides an optimal rate of convergence for $\widehat{c}_T(u,v)$ in terms of $\|\widehat{c}_T - c_0\|_{\mathfrak{C}}$, where $\|\bullet\|_{\mathfrak{C}}$ is a pseudo-metric on $\mathfrak{C}$. First, we specify a set of minimal assumptions needed for our results.

Assumption A:

(A1) Kernel Function: $K(\bullet) \in \mathcal{K}^r(\mathbb{R})$ for some integer, $r > 1$, where $\mathcal{K}^r(\mathbb{R})$ is the space of symmetric, Lebesgue integrable, kernel functions of order r. (cf. Definition 6.3).

(A2) Marginal-Weighted Functions: $\{\lambda(u), \gamma(v)\} \in C^*([0,1]) \times C^*([0,1])$ satisfy Hölder conditions with some exponents, r_1 and r_2, respectively (i.e., $|\lambda(u_1) - \lambda(u_2)| \leq H_1 |u_1 - u_2|^{r_1}$ and $|\gamma(v_1) - \gamma(v_2)| \leq H_2 |v_1 - v_2|^{r_2}$ with some generic positive constants, $H_1 < \infty$ and $H_2 < \infty$).

Assumption B:

(B1) Moment Condition: $\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| E[|D'_{\boldsymbol{\theta}} h(U,V,\boldsymbol{\theta})|] \right\|_E < \infty$, where the vector $D'_{\boldsymbol{\theta}} h(u,v,\boldsymbol{\theta})$ contains the first-order partial derivatives of $h(u,v,\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$; $\boldsymbol{\Theta}$ is a compact finite-dimensional set; and $\| \bullet \|_E$ is the usual Euclidean norm.

(B2) $\sqrt{T}$-Consistency of the Estimates of Parameters: $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = O_p\left(T^{-1/2}\right)$.

(B3) Rate of Deviations: $\inf_{\substack{\{c,\boldsymbol{\Lambda}\} \in \{W_1^r([0,1]^2), B_{\lambda^*} \times B_{\gamma^*} \times C^*([0,1]) \times C^*([0,1])\} \\ \|c - c_0\|_{\mathfrak{C}} + \|\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0\|_{EC} \geq D_1 \epsilon}} \{\mathcal{L}(c_0, \boldsymbol{\Lambda}_0, \boldsymbol{\theta}_0) - \mathcal{L}(c, \boldsymbol{\Lambda}, \boldsymbol{\theta}_0)\} \geq 2D_2 \epsilon^{1/\omega}$ – where D_1 and D_2 denote some generic positive constants, B_{λ^*} and B_{γ^*} are bounded sets in $\mathbb{R}$, $\|\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0\|_{EC} \doteq \sqrt{|\lambda^* - \lambda_0|^2 + |\gamma^* - \gamma_0|^2} + \|\lambda - \lambda_0\|_{C^*} + \|\gamma - \gamma_0\|_{C^*}$ with a pseudo-metric, $\| \bullet \|_{C^*}$, on $C^*([0,1])$, and $\omega \in (1, \infty)$ is some coefficient of deviations – represents the minimum rate of deviations associated with the true Lagrangian function:

$$\begin{aligned} \mathcal{L}(c, \boldsymbol{\Lambda}; \boldsymbol{\theta}_0) &= -E[\log c(U,V)] - \lambda^* (E[\mathbf{1}] - 1) - \gamma^* E[h(U,V,\boldsymbol{\theta}_0)] \\ &- \int_\epsilon^1 \lambda(u) \left(E[\mathbb{I}(w \in [u-\epsilon, u])] - \epsilon\right) du - \int_\epsilon^1 \gamma(v) \left(E[\mathbb{I}(w \in [v-\epsilon, v])] - \epsilon\right) dv, \end{aligned}$$

where $\mathbf{\Lambda} = \{\lambda^*, \gamma^*, \lambda(u), \gamma(v)\}'$ and all these expectations are taken with respect to $c \in W_1^r([0,1]^2)$.

Assumption C:

(C) Covering Number: $\int_0^1 \epsilon \left(\log 1/\epsilon\right)^p N\left(\epsilon, \mathfrak{C}_T^*\right) d\epsilon < \infty$ as $T \to \infty$ for some $p > 1$, where $N\left(\epsilon, \mathfrak{C}_T^*\right)$ is the number of open balls of diameter, ϵ, covering $\mathfrak{C}_T^*$.

Assumption D:

(D1) Number of Side Constraints: $N_T \min\left(T^{-\frac{\ell}{r_1\omega}}, T^{-\frac{\ell}{r_2\omega}}\right) \to \infty$, where $\ell \in (0, 1/2)$.

(D2) Bandwidth: $h_T = o\left(T^{-\frac{\ell}{\omega r}}\right)$.

THEOREM 3.3. *Under Assumptions A, B, C, and D, the MECC,* $\widehat{c}_T \doteq \arg\sup_{c \in \mathcal{C}_T^*(\widehat{\boldsymbol{\theta}})} \{-E\left[\log c(U,V)\right]\}$, *satisfies*

$$\sup_{c \in \mathcal{C}_T^*(\widehat{\boldsymbol{\theta}})} \{-E[\log c(U,V)]\} \geq \sup_{c \in \mathcal{C}} \{-E[\log c(U,V)]\} - o_P(1)$$

and

$$\|\widehat{c}_T - c_0\|_{\mathfrak{C}} = o_P\left(T^{-\omega\ell}\right).$$

Proof. The proof is presented in Appendix. □

4 SIMULATION

In this section, we perform simulations to investigate the finite-sample properties of the MECC density [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). It is to be stressed at this point that, because the MECC is parametrically constructed, it is natural to make comparisons with these commonly-used parametric copulas. 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 $\theta \in (-\infty,\infty)/\{0\}$, as the true model whereby samples are generated. [See Nelsen (1986) and Genest (1987) for the statistical properties of Frank's copula.] This copula is symmetric under rotations of 180° and close to the independence as θ approaches the origin, i.e. $\lim_{\theta\to 0} C(u,v;\theta) = uv$. 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 the MLE method and 12 MECCs (that is, $MECC(N,M)$ with combinations of $N = 4, 16, 64$ 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*);

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

where $c(u,v;\theta)$ is the density of Frank's copula; and $\widehat{c}_T(u,v)$ 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 Int. Bias^2 and Int. Var. are then obtained by evaluating the estimated squared bias, $\widehat{\text{Bias}}^2(u,v) = \left|\widehat{E}[\widehat{c}_T(u,v) - c(u,v;\theta)]\right|^2$ and the estimated variance, $\widehat{\text{Var.}}(u,v) = \widehat{E}\left[(\widehat{c}_T(u,v) - E[\widehat{c}_T(u,v)])^2\right]$, 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 $\widehat{c}_T(u,v)$ are evaluated. We report our simulation results in Table 1.

First, it can be noticed from Table 1 that the MECCs significantly outperform elliptical copulas (i.e., the Normal copula and Student's t copula) in terms of Int. Bias^2 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. Second, 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. Bias^2 and Int. Var. is that Frank's copula, the Clayton copula, and the Gumbel copula all belong to the family of Archimedean copulas; thus, when observations in a sample are close-to-independently generated from Frank's copula, the estimator using either the Clayton copula or the Gumbel copula is close-to-consistent so that these three copulas (including the estimated ones) will behave like the independence copula.

Second, when $\theta = 0.8$ the data will become less independent, leading to a significant increase in Int. Bias^2 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. Bias^2 over the other parametric copulas. It is also important to note at this point that, for a fixed number of marginal constraints, Int. Bias^2 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 1, for one joint moment constraint, one merely needs four marginal constraints to yield MECC(4,1) with minimum Int. Bias^2 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. Bias^2, 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.

5 APPLICATION TO ASSET ALLOCATION

We analyze a CRRA investor's problem of allocating wealth between two 'style' portfolios - a portfolio of 'growth' stocks and a portfolio of 'value' stocks. (See, e.g., Fama and French (1998), Basu (1977), Petkova and Zhang (2005), Barberis and Shleifer (2003), among many others, for a complete discussion on the characteristics of these portfolios.) This is a typical problem of optimal investment beyond linear risk and linear dependence. A precise formulation now follows. Let $r_{1,t}$ and $r_{2,t}$ be the 'growth' and 'value' returns with some conditional joint distribution, F_t, associated with a conditional copula, C_t, and conditional marginals, $G_{1,t}$ & $G_{2,t}$. In view of the conditional Sklar theorem the one-period ahead optimal weights $\omega^*_{t+1} = \{\omega_{1,t+1}, \omega_{2,t+1}\}$ can be written as follows:[4]

$$\begin{aligned}
\omega^*_{t+1} &= \arg\max_{\omega\in\mathcal{W}} E_{\widehat{F}_{t+1}}[\mathcal{U}(1+\omega_1 r_{1,t+1}+\omega_2 r_{2,t+1})] \\
&= \arg\max_{\omega\in\mathcal{W}} \int\int \mathcal{U}(1+\omega_1 x+\omega_2 y).\widehat{g}_{1,t+1}(x).\widehat{g}_{2,t+1}(y).\widehat{c}_{t+1}(\widehat{G}_{1,t+1}(x),\widehat{G}_{2,t+1}(y))dxdy,
\end{aligned}$$

where $\widehat{G}_{1,t+1}$ and $\widehat{G}_{2,t+1}$ are conditional marginal distribution forecasts, $\widehat{g}_{1,t+1}(x)$ and $\widehat{g}_{2,t+1}(y)$ are conditional marginal density forecasts, $\widehat{c}_{t+1}$ is a conditional copula density forecast, $\mathcal{W}$ is defined as some compact subset of $\mathbb{R}^2$ for unconstrained investors and as the set $\{(\omega_1,\omega_2)\in[0,1]^2 :\ \omega_1+\omega_2 \leq 1\}$ for short-sales constrained investors, and $\mathcal{U}(x)$ is the CRRA utility function given by

$$\mathcal{U}(x) = \begin{cases} (1-RRA)^{-1}x^{1-RRA}, & \text{if} \quad RRA \neq 1, \text{ where } RRA \text{ is a degree of risk aversion,} \\ \log(x), & \text{if} \quad RRA = 1. \end{cases}$$

Note that, in our computations, we use the method of Levin (1983) to approximate the double integral in the above expected utility function, instead of the Monte Carlo method. The former method

[4]The conditional Sklar theorem asserts that $F_{t+1}(X,Y|\mathcal{F}_t) = C_{t+1}(G_{1,t+1}(X|\mathcal{F}_t), G_{2,t+1}(Y|\mathcal{F}_t)|\mathcal{F}_t)$, where $\mathcal{F}_t$ is a set of information available up to t (see, e.g., Patton (2004, 2006)).

is very efficient because it guarantees that any double integral of bounded and piecewise continuous functions can be effectively approximated by optimal (*or* asymptotically optimal) quadratures, and that the approximation error is minimal. Also, to estimate the joint model $F_{t+1}(X, Y|\mathcal{F}_t)$ we apply the inference function for margins (IFM) method, which has been introduced by Joe and Xu (1996). The IFM method has been used, for instance, in Patton (2004) and Jondeau and Rockinger (2006).

Before going into details, there are three other implementation issues that warrant discussion at this point. First of all, due to some considerable computational constraints, we were compelled to ignore the ubiquitous issue of parameter and model uncertainty on investment decisions (see, e.g., Brandt (2004)). We are aware that this issue can be approached with the Bayesian estimation and model averaging method at a considerable computational cost, even so it is not certain that the Bayesian portfolios will outperform the $1/N$ portfolio (see DeMiguel et al. (2009)). Second, we considered only an one-period-ahead investment problem. This is a common problem in optimal asset allocation (see, e.g., Patton (2004), among many others). Third, we used Hansen's (1994) skewed t distribution to capture 4 time-varying central moments (means, variances, skewness, and kurtosis) of the standardized residuals of the TGARCH-type models of the 'value' and 'growth' returns; and the copulas (Normal, Clayton, Gumbel, and the MECC), which have parameter(s) depending on the latent underlying fundamentals of the returns, to model asymmetric dependence between the standardized residuals. While the Normal, Clayton or Gumbel copula has only one parameter, the MECC may have many parameters, depending on how much prior information we want to impose. Also note that this framework is a special case of the copula-based multivariate models [see, e.g., Patton (2004, 2006) and Jondeau and Rockinger (2006)].

We now use monthly (from June 1995 to July 2006) data of the Russell 1000 index[5]. This index comprises 1,000 large U.S. companies, which are determined by their market capitalizations and ranked by their adjusted book-to-price ratios. A probability methodology is then used to split the index into 'growth' and 'value' subindices – that is, each company is assigned a probability of being a 'growth' or a 'value' one. The 'growth' index thus contains the companies with greater-than-average growth orientations, higher-than-average P/B and P/E ratios, lower-than-average dividend yields,

[5]All the data used in the current paper were downloaded from *Bloomberg* and *Datastream*.

and higher-than-average forecast earnings growth; the 'value' index contains the companies which are not 'growth' ones. We use the period of June 1995 to September 1999 (*namely*, the in-sample period) to develop our models, and the remaining period (*namely*, the out-of-sample period) for out-of-sample evaluation.

Table 2 shows that the continuously compounded 'growth' and 'value' returns in excess of the three-months Treasury Bill (T-Bill) rate, a proxy of risk-free interest rate, (hereafter, we shall write 'returns' for brevity, and specify otherwise) are not normally distributed in the entire sample period. (The Kolmogorov-Smirnov test clearly rejects the normality for both series.) Moreover, the 'value' return has higher mean and lower variance than the 'growth' return in the entire sample and out-of-sample periods, but this pattern is *not* clear in the in-sample period; and these returns also display non-negligible linear correlation. In fact, this result is consistent with the claim by Ibbotson and Riepe (1997, June) that " 'value' trumps growth, with the caveat that 'growth' stocks can outperform 'value' stocks over an extended of time."

First of all, to get a preliminary idea about the degree of asymmetric dependence between these returns, we shall use measures of asymmetric linear correlation (e.g. the 'exceedance' correlations; see Longin and Solnik (2001) and Ang and Chen (2002)) and a test for asymmetries proposed by Hong et al. (2007). Let $\rho^+(c)$ and $\rho^-(c)$ denote the 'exceedance' correlations between the 'growth' return, $r_{1,t}$, and the 'value' return, $r_{2,t}$, at a threshold level c, defined as follows:

$$\rho^+(c) = \mathrm{corr}(r_{1,t}, r_{2,t} | r_{1,t} > c, r_{2,t} > c), \tag{5.1}$$

$$\rho^-(c) = \mathrm{corr}(r_{1,t}, r_{2,t} | r_{1,t} < -c, r_{2,t} < -c). \tag{5.2}$$

Figure 1 shows that the differences between the sample 'exceedance' correlations (of these returns) are non-negligible under various cutoff quantiles. We also conducted the test for asymmetry with 24 'exceedance' levels [not reported], and found that the p-value is close to zero, confirming that there is a strong evidence of asymmetric correlation. Furthermore, this asymmetric correlation varies significantly over time (cf. Figure 2).

To model the time-varying feature of a conditional marginal distribution, we use a TGARCH

model, which has also been employed by Patton (2004) and Jondeau and Rockinger (2003). After having implemented some loglikelihood-based model selection tests for the optimal lags of the returns and best predictive instruments [not reported], we end up with the following predictive TGARCH(1,1) model:

$$\begin{aligned} r_{i,t} &= \alpha_{i,0} + \alpha_{i,1} r_{i,t-1} + \alpha_{i,2} r^f_{t-1} + \alpha_{i,3} DY_{i,t-1} + \alpha_{i,4} PER_{i,t-1} + \alpha_{i,5} TS_{t-1} + \alpha_{i,6} DS_{t-1} + \alpha_{i,7} Cay_{t-1} \\ &+ \sigma_{i,t}\epsilon_{i,t} \end{aligned} \tag{5.3}$$

with

$$\begin{aligned} \sigma^2_{i,t} &= \alpha_{i,8} + \alpha_{i,9}\sigma^2_{i,t-1} + \alpha_{i,10}\sigma^2_{i,t-1}\epsilon^2_{i,t-1}\mathbf{1}(\epsilon_{i,t-1} > 0) + \alpha_{i,11}\sigma^2_{i,t-1}\epsilon^2_{i,t-1}\mathbf{1}(\epsilon_{i,t-1} < 0) + \alpha_{i,12} r^f_{t-1} \\ &+ \alpha_{i,13} DY_{i,t-1} + \alpha_{i,14} PER_{i,t-1} + \alpha_{i,15} TS_{t-1} + \alpha_{i,16} DS_{t-1} + \alpha_{i,17} Cay_{t-1}, \end{aligned} \tag{5.4}$$

where $i = 1$ (*growth*) or 2 (*value*). The distribution of the error $\epsilon_{i,t}$ is modeled with Hansen (1994)'s skewed t distribution [i.e., $\epsilon_{i,t} \sim G(\epsilon|\kappa_{i,t}, \lambda_{i,t})$, where $\kappa_{i,t} = 2.1 + \left(\boldsymbol{Z}'_{i,t-1}\boldsymbol{\gamma}_i\right)^2$ and $\lambda_{i,t} = \Lambda(\boldsymbol{Z}'_{i,t-1}\boldsymbol{\gamma}_i)$ with $\boldsymbol{Z}'_{i,t} = \{r^f_t, DY_{i,t}, PER_{i,t}, TS_t, DS_t, Cay_t\}$ and $\Lambda(x) = (1 - e^{-x})/(1 + e^{-x})$]. Moreover, r^f_t, DY, PER, TS, DS, and Cay represent the three-months T-bill rate, the dividend-price ratio, the price-earning ratio, the term-spread, the default-spread, and the log consumption-aggregate wealth ratio, respectively. These variables are commonly-used predictors of stock returns in the literature (see, e.g., Fama (1981), Campbell and Shiller (1988), Lettau and Ludvigson (2001), and Ait-Sahalia and Brandt (2001), among others).

We estimate (5.3) and (5.4) by the method of MLE. Due to the high dimensionality of this problem, we used a fast stochastic search algorithm[6] to find global maximum log-likelihood values. The algorithm consists of 4 steps: (1) transforming the domains of (multi-)parameters into unit cubes, (2) drawing 5000 vectors of transformed parameter values from these cubes, (3) ordering the samples descendingly with respect to the values of the objective function, then picking up 5 samples on the top, and (4) using those 5 samples as the starting values to run 5 local maximization algorithms (e.g., Quasi-Newton algorithm), the global maximum value is, then, the best local maximum value.

[6] C++ code to implement this method was downloaded from http://www.inf.u-szeged.hu/ csendes/

We found significant evidence [not reported] that the parameters of the error distributions, $\kappa_{i,t}$ and $\lambda_{i,t}$, are time-varying in the whole sample period. We also conduct various goodness-of-fit tests for our conditional-marginal-distribution models; and their results [not reported] suggest that the skewed t distribution is a good model for our data.

5.1 Copula models

The benchmark models are the conditional Normal copula, the conditional Clayton copula, and the conditional Gumbel copula. These models can be specified by

$$\left(\frac{r_{1,t}-\mu_{1,t}}{\sigma_{1,t}}, \frac{r_{2,t}-\mu_{2,t}}{\sigma_{2,t}}\right) \sim C(G(\epsilon_{1,t}|\kappa_{1,t},\lambda_{1,t}), G(\epsilon_{2,t}|\kappa_{2,t},\lambda_{2,t}); \delta_t)$$
$$\text{with } \delta_t = \Gamma(\beta_0 + \beta_1 r^f_{t-1} + \beta_2 DY_{t-1} + \beta_3 PER_{t-1} + \beta_4 TS_{t-1} + \beta_5 DS_{t-1} + \beta_6 Cay_{t-1}),$$

where C denotes the Normal, Clayton, or Gumbel copula; $\mu_{i,t} = E[r_{i,t}|\mathcal{F}_{t-1}]$, for $i = 1, 2$, denotes the conditional mean of $r_{i,t}$, defined in (5.3); and $\Gamma(x)$ is a function transforming x into a feasible copula parameter.

The *conditional MECC* model is defined in the following way: given the rank correlations between the standardized excess returns $\{\widehat{\rho}_S, \widehat{\nu}_1, \widehat{\nu}_2, \widehat{\eta}\}$ and a logistic function, $\Lambda(x)$, we specify

$$\left(\frac{r_{1,t}-\mu_{1,t}}{\sigma_{1,t}}, \frac{r_{2,t}-\mu_{2,t}}{\sigma_{2,t}}\right) \sim MECC(G_{1,t}, G_{2,t}; \rho_{S_t}, \nu_{1,t}, \nu_{2,t}, \eta_t), \tag{5.5}$$

where

$$\frac{\rho_{S_t}}{\widehat{\rho}_S} = \Lambda(\theta_0 + \theta_1 r^f_{t-1} + \theta_2 DY_{t-1} + \theta_3 PER_{t-1} + \theta_4 TS_{t-1} + \theta_5 DS_{t-1} + \theta_6 Cay_{t-1}), \tag{5.6}$$
$$\frac{\nu_{1,t}}{\widehat{\nu}_1} = \Lambda(\theta_7 + \theta_8 r^f_{t-1} + \theta_9 DY_{t-1} + \theta_{10} PER_{t-1} + \theta_{11} TS_{t-1} + \theta_{12} DS_{t-1} + \theta_{13} Cay_{t-1}), \tag{5.7}$$
$$\frac{\nu_{2,t}}{\widehat{\nu}_2} = \Lambda(\theta_{14} + \theta_{15} r^f_{t-1} + \theta_{16} DY_{t-1} + \theta_{17} PER_{t-1} + \theta_{18} TS_{t-1} + \theta_{19} DS_{t-1} + \theta_{20} Cay_{t-1}), \tag{5.8}$$
$$\frac{\eta_t}{\widehat{\eta}} = \Lambda(\theta_{21} + \theta_{22} r^f_{t-1} + \theta_{23} DY_{t-1} + \theta_{24} PER_{t-1} + \theta_{25} TS_{t-1} + \theta_{26} DS_{t-1} + \theta_{27} Cay_{t-1}). \tag{5.9}$$

Here note that, with an aim in our mind to keep the models parsimonious, we assume that the

parameters of the copulas under our consideration have the same predictive instruments [i.e., DY_t, PER_t (the dividend yield and the price-earning ratio of Russell 1000), TS_t, DS_t, and Cay_t]. In addition, we did not include the dynamics of Blest's measure IV, which is defined in Section 2, in the model (5.5)-(5.9) because including Blest's measure IV would result in a model with higher AIC and BIC [not reported].

We estimate the MECC model (5.5)-(5.9) by using the empirical likelihood (EL) method. This is a commonly used, efficient method when the log-likelihood function has neither a trackable nor close form (see, e.g., Rockinger and Jondeau (2002), among many others). (Further details are presented in the appendix.) Other parametric copula models are estimated by the conventional ML technique.

We, then, assess relative performance of the estimated copula models. Due to time variations in the rank correlations between the returns (Figure 3), we allow for fluctuations in the relative performance of the models by using moving windows to recursively estimate the models. In the sequel, we apply the out-of-sample fluctuation test (based on one-period forecasts of the (empirical) log-likelihoods of the copulas for all the moving windows), which is proposed by Giacomini and Rossi (2007) and Giacomini and White (2006), to test the null hypothesis that two models perform equally well in the out-of-sample period. The p-values of the test, using the critical values computed by Giacomini and Rossi (2007), are reported in Table 3. The null hypothesis is clearly rejected for the following pairs: Clayton vs. Gumbel, Clayton vs. MECC, and Gumbel vs. MECC (cf. Figure 4). Table 4 suggests that, in terms of average maximum ELs, the MECC provides the best performance while the Normal copula performs slightly better than the Clayton and Gumbel copulas. These results are also confirmed by the information criteria values reported.

5.2 Relative Performance of Investment Models

In this section, we report relative performances of eight different asset allocation strategies for four levels of RRA (3, 7, 15, and 25). These strategies are defined as the optimal weights invested in the 'growth' and 'value' indices for each period ahead by (i) using the Normal copula (the Normal strategy); (ii) using the Clayton copula (the Clayton strategy); (iii) using the Gumbel copula (the

Gumbel strategy); and (iv) using the MECC (the MECC strategy). Strategies (5)-(8) are the strategies (1)-(4) with short-sales constraints imposed.

Table 5 provides descriptive statistics of the continuously compounded portfolio returns computed for eight above-mentioned strategies. The scenarios turn out differently, depending on whether the short-sales constraint is imposed. When the RRA level is equal to 3, 7, or 15, the Sharpe ratios based on the Normal copula, in the unconstrained case, are highest; and in the short-sales constrained case, the Sharpe ratios based on the Clayton copula are highest. Moreover, the portfolio returns based on the Normal copula, in the short-sales constrained case, often have negative skewnesses and smallest kurtosis. Meanwhile, the portfolio returns based on the MECC with a short-sales constraint always exhibit strong, positive skewness and high, positive kurtosis, suggesting that using the MECC to model asymmetric dependence may potentially help investors with short-sales constraints to avoid negatively-skewed portfolios. The intuition for this empirical finding is as follows. The simulation study presented in Section 4 suggests that, when the dependence between two random variables is non-negligible, the MECC can provide better goodness-of-fit than other parametric copulas. It then implies that, in our investment context, the MECC can, for the most part, generate an improved forecast on the future dependence between the risk-adjusted excess returns of Russell 1000 Growth and Value. Moreover, investors with short-sale constraints tend to be more sensitive to (*or* more likely to be affected by) a future change in the dependence among assets in their portfolio than the ones without short-sale constraints; this is because the former are often subjected to, among other things, restricted investment opportunity sets. Hence the MECC can potentially improve the performance of the portfolios held by these investors.

Tables 6 and 7 report descriptive statistics of the optimal portfolio weights for four among eight above-mentioned strategies – the Normal strategies and the MECC strategies. As expected, an increase in the RRA level – the investor becomes more risk-averse – leads to an increase in the median of the portfolio weights invested in the 'value' index and a decrease in the median of those invested in the 'growth' index– this is especially true for the MECC model. These findings are consistent with the remark of Ibbotson and Riepe (1997, June) that "the superior historical performance of value leads us to tilt the equity allocation in that direction for all but the most

aggressive investors." Moreover, while the Normal and Gumbel portfolio weights display similar patterns in most periods (cf. Figure 6), the MECC portfolio weights are often more extreme and more volatile than the Normal portfolio weights (cf. Figure 5), suggesting that the MECC may capture hidden asymmetric dependence that the Normal or Gumbel copula fails to capture.

We also used a bootstrap test (via pairwise comparisons of bootstrap samples) to examine whether the differences in the numerical expected utilities of eight above-mentioned strategies are statistically significant. Indeed, the Normal strategy, in the *unconstrained* case, significantly outperforms both the Clayton and Gumbel ones for three levels of RRA (3, 15, and 25) except for the level of RRA 7, where these three strategies perform equally well (cf. Table 8). This may well be the case that the standardized returns are asymmetrically dependent in a complex way which is not sufficiently captured by the Normal, Gumbel or Clayton copula; in other words, these copulas may be misspecified, and the Normal strategy is better than the other two copulas. In addition, in the *constrained* case, while the Gumbel strategy significantly outperforms the Clayton strategy for three levels of RRA (3, 7, and 15), the Clayton and Gumbel strategies perform equally well for all the RRA levels. Nevertheless, the MECC strategy always outperforms other strategies, and the degree of outperformance is rather strong in the unconstrained case.

To compare all the models jointly, we applied the reality check test of White (2000). The null hypothesis is that the benchmark model performs as well as the best competing model. We used the Normal copula as a benchmark model and rejected the null hypothesis when the p-value is less than 10%. We found that, without short-sales constraints, the null hypothesis is always rejected for all RRA levels except 25 (cf. Table 9). Nevertheless, with short-sales constraints, the Normal model may perform as well as the best competing model.

6 CONCLUSION

In this paper, we have made two contributions to the growing literature of entropy and copula. First, we provided an analytical approach to recover the relative entropy measure of dependence from limited information by building the most entropic copula (MEC). Among the MECs, there

exists a canonical form called the most entropic canonical copula (MECC). Second, we showed that our approach and the minimum KLCE approach are dual, that is, the minimum KLCE joint distributions, obtained by minimizing the KLCE subject to constraints on joint moments and marginal distributions, can be constructed from a MECC and arbitrary sets of marginals.

We have illustrated the practicability of our approach by considering a 'style investing' problem for an investor with the CRRA utility allocating his/her wealth between the 'growth' and 'value' indices. The main finding is that, using the skewed-t distributions (to capture time-varying skewnesses and kurtoses), the MECC model significantly outperforms the Normal copula model in terms of investment gains, and the Normal copula model may outperform the Clayton or Gumbel copula model in certain cases. The most economically and statistically significant gains from using the MECC over the other copulas are brought about by the unconstrained portfolios.

We conclude this paper with a thoughtful remark: Since outliers are an ill-defined, albeit important concept without clear boundaries, the improved immunity to outliers offered by robust techniques is usually obtained at the expense of a considerable increase in computation. Admittedly, given some considerable merits the method of MEC, which is based on the premise that rank correlations are resistant against noise and outliers, is unfortunately not an exception. Hence, future researches should focus on computationally efficient, nonparametric estimation of the MECC. We believe that this is an absolutely feasible task.

REFERENCES

Abramowitz, M. and Stegun, I. A. (1972). *Handbook of Mathematical Functions*, Dover Publications.

Ait-Sahalia, Y. and Brandt, M. W. (2001). Variable selection for portfolio choice, *Journal of Finance* **56**: 1297–1355.

Ang, A. and Chen, J. (2002). Asymmetric correlations of equity portfolios, *Journal of Financial Economics* **63**: 443–494.

Barberis, N. and Shleifer, A. (2003). Style investing, *Journal of Financial Economics* **68**: 161–199.

Barnett, N. S., Cerone, P. and Dragomir, S. S. (2010). Inequalities for random variables over a finite interval. Lecture Notes.

Basu, S. (1977). Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis, *Journal of Finance* **3**: 663–682.

Beliakov, G. (2006). Interpolation of lipschitz functions, *Journal of Computational and Applied Mathematics* **196**: 20–44.

Blest, D. C. (2000). Rank correlation – an alternative measure, *Australia & New Zealand Journal of Statistics* **42**(1): 101–111.

Brandt, M. W. (2004). Portfolio choice problem, *in* Y. Ait-Sahalia and L. P. Hansen (eds), *Handbook of Financial Econometrics*, Elsevier Science, Amsterdam.

Campbell, J. Y. and Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors, *Review of Financial Studies* **1**: 195–228.

Cherubini, U., Luciano, E. and Vecchiato, W. (2004). *Copula Methods in Finance*, John Wiley & Sons, London, New York.

Chollete, L., Heinen, A. and Valdesogo, A. (2009). Modeling international financial returns with a multivariate regime-switching copula, *Journal of Financial Econometrics* **7**(4): 437–480.

Cranston, M., Scheutzow, M. and Steinsaltz, D. (2000). Linear bounds for stochastic dispersion, *Annals of Probability* **28**(4): 1852–1869.

Csendes, T. (1988). Nonlinear parameter estimation by global optimization - efficiency and reliability, *Acta Cybernetica* **8**: 361–370.

DeMiguel, V., Garlappi, L. and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the $1/n$ portfolio strategy?, *Review of Financial Studies* **22**(5): 1915–1953.

Dempster, M. A. H., Medova, E. A. and Yang, S. W. (2007). Empirical copulas for CDO tranche pricing using relative entropy, *International Journal of Theoretical and Applied Finance* **10**(4): 679–701.

Fama, E. (1981). Stock returns, real activity, inflation, and money, *American Economic Review* **71**: 545–565.

Fama, E. F. and French, K. R. (1998). Value versus growth: The international evidence, *Journal of Finance* **LIII**(6): 1975–1999.

Genest, C. (1987). Frank's family of bivariate distributions, *Biometrika* **74**: 549–555.

Genest, C. and MacKay, J. (1986). The joy of copulas: Bivariate distributions with uniform marginals, *The American Statistician* **40**(4): 280–283.

Genest, C. and Plante, J.-F. (2003). On blest's measure of rank correlation, *Canadian Journal of Statistics* **31**: 1–18.

Giacomini, R. and Rossi, B. (2007). Model selection and forecast comparison in unstable environment, *Working paper* .

Giacomini, R. and White, H. (2006). Tests of conditional predictive ability, *Econometrica* **74**: 1545–1578.

Gideon, R. A. and Hollister, R. A. (1987). A rank correlation coefficient resistant to outliers, *Journal of American Statistical Association* **82**(398): 656–666.

Giné, E. and Nickl, R. (2008). Uniform central limit theorems for kernel density estimators, *Probability Theory and Related Fields* **141**: 333–387.

Golan, A. (2002). Information and entropy econometrics: Editor's view, *Journal of Econometrics* **107**: 1–15.

Golan, A. (2007). Information and entropy econometrics – volume overview and synthesis, *Journal of Econometrics* **138**: 379–387.

Hajek, J. and Sidak, Z. (1967). *Theory of Rank Tests*, Academic Press, New York.

Hang, R. K. (1993). Maximum entropy estimation of density and regression functions, *Journal of Econometrics* **56**: 397–400.

Hansen, B. E. (1994). Autoregressive conditional density estimation, *International Economic Review* **35**: 705–729.

Hong, Y., Tu, J. and Zhou, G. (2007). Asymmetries in stock returns: Statistical tests and economic evaluation, *Review of Financial Studies* **20**: 1547–1581.

Ibbotson, R. G. and Riepe, M. W. (1997, June). Growth vs. value investing and the winner is..., *Journal of Financial Planning* pp. 64–71.

Ioffe, A. D. and Tihomirov, V. M. (1979). *Theory of Extremal Problems*, Vol. 6 of *Studies in mathematics and its applications*, North Holland Publishing Company., Amsterdam, New York, Oxford.

Joe, H. (1989). Relative entropy measures of multivariate dependence, *Journal of the American Statistical Association* **84**(405): 157–164.

Joe, H. and Xu, J. J. (1996). The estimation method of inference functions for margins for multivariate models, *Working paper, UBC* .

Jondeau, E. and Rockinger, M. (2003). Conditional volatility, skewness, and kurtosis: Existence, persistence and comovement, *Journal of Economic Dynamics and Control* **27**: 1699–1737.

Jondeau, E. and Rockinger, M. (2006). The copula-garch models of conditional dependencies: An international stock market applications, *Journal of International Money and Finance* **25**: 827–853.

Kutoyants, Y. A. (2004). *Statistical Inference for Ergodic Diffusion Processes*, Springer series in statistics, Springer-Verlag, London, Berlin, Heidelberg.

Lettau, M. and Ludvigson, S. (2001). Consumption, aggregate wealth, and expected stock returns, *Journal of Finance* **LVI**(3): 815–849.

Levin, M. (1983). On the approximate calculation of double integrals, *Mathematics of Computation* **40**: 273–282.

Longin, F. and Solnik, B. (2001). Extreme correlation of international equity markets, *Journal of Finance* **56**: 649–676.

Miller, D. J. and Liu, W. (2002). On the recovery of joint distributions from limited information, *Journal of Econometrics* **107**: 259–274.

Nelsen, R. B. (1986). Properties of a one-parameter family of bivariate distributions with specified marginals, *Communications in Statistics –Theory and Methods* **15**(11): 3277–3285.

Nelsen, R. B. (1998). *An Introduction to Copulas*, Springer-Verlag.

Pagan, A. and Ullah, A. (1999). *Nonparametric Econometric*, Themes in Modern Econometrics, first edn, Cambridge University Press.

Patton, A. J. (2004). On the out-of-sample importance of skewness and asymmetric dependence for asset allocation, *Journal of Financial Econometrics* **2**(1): 130–168.

Patton, A. J. (2006). Modelling asymmetric exchange rate dependence, *International Economic Review* **47**(2): 527–555.

Petkova, R. and Zhang, L. (2005). Is value riskier than growth, *Journal of Financial Economics* **78**: 187–202.

Puri, M. L., Sen, P. K. and Gokhale, D. V. (1970). On a class of rank order tests for independence in multivariate distributions, *Sankhya Ser. A* **32**: 271–298.

Rockinger, M. and Jondeau, E. (2002). Entropy densities with an application to autoregressive conditional skewness and kurtosis, *Journal of Econometrics* **106**: 119–142.

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

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

Shannon, C. E. (1948). The mathematical theory of communication, *Bell System Technical Journal* **Jul-Oct**: 3–91.

Shiryaev, A. N. (1995). *Probability*, Vol. 95 of *Graduate text in mathematics*, 2 edn, Springer-Verlag, New York, Berlin, Heidelberg.

Sklar, A. (1959). Fonctions de rpartition n dimensions et leurs marges, *Publications de l'Institut Statistique de l'Universit de Paris 8* pp. 229–231.

White, H. (2000). A reality check for data snooping, *Econometrica* **68**: 1097–1126.

Widder, D. V. (1971). *An Introduction to Transform Theory*, Academic Press, New York & London.

Wu, X. (2003). Calculation of maximum entropy densities with application to income distribution, *Journal of Econometrics* **115**: 347–354.

Wu, X. and Perloff, J. M. (2007). Gmm estimation of a maximum entropy distribution with interval data, *Journal of Econometrics* **138**: 532–546.

Wu, X. and Stengos, T. (2005). Partially adaptive estimation via the maximum entropy densities, *Econometrics Journal* **8**: 352–366.

Zellner, A. and Highfield, R. A. (1988). Calculation of maximum entropy distributions and approximation of marginal posterior distributions, *Journal of Econometrics* **37**: 195–209.

APPENDIX A: BASIC RESULTS

DEFINITION 6.1 (Adapted and modified from Nelsen, 1998, p. 10)**.** *A two-dimensional copula, $C(u, v)$, is a real function defined on the unit square $[0, 1]^2$:*

$$C: \ [0,1]^2 \Longrightarrow [0,1]$$

such that

(D1) $C(u, 0) = C(0, v) = 0$,

(D2) $C(u, 1) = u$ *and* $C(1, v) = v$ *for every* (u, v) *of* $[0, 1]^2$, *and*

(D3) C *is* 2-increasing *[i.e.,* $C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) + C(u_1, v_1) \geq 0$ *for every rectangle* $[u_1, u_2] \times [v_1, v_2]$, *where* $u_1 \leq u_2$ *and* $v_1 \leq v_2$, *whose vertices lie in* $[0, 1]^2$*.]*

Note that the requirement (3) in the above definition purports to guarantee that copulas fall within the Fréchet bounds [cf. (Cherubini et al., 2004, Theorem 2.4)].

DEFINITION 6.2 (Adapted and modified from Nelsen, 1998, chap. 5)**.** *Let τ denote the difference between the probabilities of concordance and discordance of (X_1, Y_1) and (X_2, Y_2) as follows:*

$$\tau = P\{(X_1 - X_2)(Y_1 - Y_2) > 0\} - P\{(X_1 - X_2)(Y_1 - Y_2) < 0\}, \tag{A.1}$$

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

The copula representations of Kendall's tau, Spearman's rho, and Gini's gamma are given in Theorem 6.1.

THEOREM 6.1. *Let C_1 and C_2 denote copulas such 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*

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

For instance, if $C_2(u,v) = uv$, then τ is Spearman's rho, ρ_S; if $C_2(u,v) = 2(|u+v-1| - |u-v|)$, then τ is Gini's gamma; and if $C_1(u,v) = C_2(u,v)$, then τ is Kendall's tau.

Proof. See Nelsen (1998, chap. 5). □

DEFINITION 6.3. *A kernel function $K: \mathbb{R} \to \mathbb{R}$ 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* $j = 1, \ldots, [r]$, *and*

(iii) $\int_{\mathbb{R}} |y|^r |K(y)|dy < \infty$, *where* $[r]$ *is the integer part of* r.

APPENDIX B: APPROXIMATION OF POTENTIAL FUNCTIONS (Supplementary Material)

We now present a Gaussian-Legendre quadrature method to approximate the potential function (3.8) for the MECC. Using affine transformations, $x_1(u): \ [0,1] \longrightarrow [-1,1]$ with $x_1 = 2u - 1$ and $x_2(v): \ [0,1] \longrightarrow [-1,1]$ with $x_2 = 2v - 1$, (3.8) can be rewritten as follows:

$$\begin{aligned} Q_{n,N_h}(\boldsymbol{\Lambda}, \widehat{\boldsymbol{\theta}}) &= \frac{1}{4}\int_{[-1,1]^2} \exp\left\{ \sum_{k=0}^{2^n-1} \left[\lambda_k \Big(\Phi\big(k - 2^{n-1}(x_1+1)\big) + \Phi\big(2^{n-1}(x_1+1) - k - 1\big) - 1 + 2^{-n} \Big) \right.\right. \\ &+ \left. \gamma_k \Big(\Phi\big(k - 2^{n-1}(x_2+1)\big) + \Phi\big(2^{n-1}(x_2+1) - k - 1\big) - 1 + 2^{-n} \Big) \right] \\ &- \left. \lambda_{2^n} \left(h\big(\frac{x_1+1}{2}, \frac{x_2+1}{2}\big) - \widehat{\boldsymbol{\theta}} \right) \right\} dx_1 dx_2 \\ &= \frac{1}{4}\int_{[-1,1]^2} \exp\{-\boldsymbol{\Lambda}'\boldsymbol{\Psi}(\boldsymbol{X})\} d\boldsymbol{X}, \end{aligned} \tag{A.3}$$

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

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

$$\exp\{-\boldsymbol{\Lambda}'\boldsymbol{\Psi}(\boldsymbol{X})\} = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} a_{nm} P_{nm}(\boldsymbol{X}), \tag{A.4}$$

where $P_{nm}(\boldsymbol{X}) = P_n(x_1)P_m(x_2)$ are products of two Legendre orthogonal polynomials [see, e.g., Abramowitz and Stegun (1972) for further details of the Legendre polynomials],

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

and

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

Now, let $\boldsymbol{X}_{i,j} = (x_{1i}, x_{2j})$, $\forall\ i = 1, \ldots, N$ and $j = 1, \ldots, M$, be the roots of the polynomials $P_N(x_1) = 0$ and $P_M(x_2) = 0$ respectively – $\boldsymbol{X}_{i,j}$ are also called the abscissae of the Legendre

polynomials – then, choose weights, ω_{ij}, satisfying the following M×N relations:

$$\begin{cases} \sum_{i=1}^{N}\sum_{j=1}^{M}\omega_{ij}P_{00}(\boldsymbol{X}_{i,j}) = \sum_{1}^{N}\sum_{1}^{M}\omega_{ij} = 1, \\ \sum_{1}^{N}\sum_{1}^{M}\omega_{ij}P_{kh}(\boldsymbol{X}_{i,j}) = 0,\ \omega_{ij} \geq 0, \end{cases} \tag{A.5}$$

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

$$\begin{aligned} \sum_{i=1}^{N}\sum_{j=1}^{M}\omega_{ij}\exp\{-\boldsymbol{\Lambda}'\boldsymbol{\Psi}(\boldsymbol{X}_{ij})\} &= \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}a_{nm}\sum_{i=1}^{N}\sum_{j=1}^{M}\omega_{ij}P_{nm}(\boldsymbol{X}_{ij}) \\ &= a_{00} + \sum_{n=N+1}^{\infty}\sum_{m=M+1}^{\infty}a_{nm}\sum_{i=1}^{N}\sum_{j=1}^{M}\omega_{ij}P_{nm}(\boldsymbol{X}_{ij}). \end{aligned} \tag{A.6}$$

Hence,

$$\begin{aligned} a_{00} = Q_{n,N_h}(\boldsymbol{\Lambda},\widehat{\boldsymbol{\theta}}) &= \sum_{i=1}^{N}\sum_{j=1}^{M}\omega_{ij}\exp\{-\boldsymbol{\Lambda}'\boldsymbol{\Psi}(\boldsymbol{X}_{ij})\} - \sum_{n=N+1}^{\infty}\sum_{m=M+1}^{\infty}a_{nm}\sum_{i=1}^{N}\sum_{j=1}^{M}\omega_{ij}P_{nm}(\boldsymbol{X}_{ij}) \\ &= \underbrace{\sum_{i=1}^{N}\sum_{j=1}^{M}\omega_{ij}\exp\{-\boldsymbol{\Lambda}'\boldsymbol{\Psi}(\boldsymbol{X}_{ij})\}}_{\text{The Approximation}} + R_{NM}, \end{aligned} \tag{A.7}$$

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

To compute the MECC, we used a stochastic search algorithm to minimize (A.7) whilst setting $M = N = 30$.

APPENDIX C: PROOFS

In our proofs, we will use the following lemmas:

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

$$f_n(x) = 2^n \int_{[k2^{-n},(k+1)2^{-n}]} f(y)dy \text{ for some } x \in [k2^{-n},(k+1)2^{-n}), \tag{A.8}$$

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

Proof. See Shiryaev (1995, p. 515). □

LEMMA 2 (DuBois-Reymond's lemma)**.** *Let $b(t)$ denote a continuous on $[t_0, t_1]$. Assuming that the following equality:*

$$\int_{t_0}^{t_1} b(t)v(t)dt = 0 \tag{A.9}$$

holds for any continuous function, $v(t)$, with $\int_{t_0}^{t_1} v(t)dt = 0$, then $b(t)$ must be a constant. Conversely, if $b(t)$ is a constant, then $\int_{t_0}^{t_1} b(t)v(t)dt = 0$.

Proof. See Ioffe and Tihomirov (1979, p. 400). □

LEMMA 3 (Lagrange's lemma)**.** *Let a function, $b(t)$, be continuous on the interval $[t_0, t_1]$. Assume that, for any continuously differentiable function $v(t)$ which vanishes at the end points of the interval $[t_0, t_1]$, the following equality holds:*

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

Then $b(t) = 0$.

Proof. See Ioffe and Tihomirov (1979, p. 103). □

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

$$\Phi_N(y,x) = \frac{N}{2\pi} \int_{-\infty}^{y} \exp\{-(v-x)^2 N^2/2\} dv, \tag{A.10}$$

which has the following properties:

$$\lim_{N \longrightarrow \infty} \Phi_N(y, x) \Longrightarrow \mathbf{1}_{y>x}(y)$$

and

$$\lim_{N \longrightarrow \infty} \frac{\partial \Phi_N(y, x)}{\partial y} \Longrightarrow \delta(y - x),$$

where $\delta(\bullet)$ is Dirac's delta function.

Proof. See Kutoyants (2004, p. 30). □

LEMMA 5. *Let $g(\boldsymbol{x})$ represent a measurable function of $\mathbb{R}^n$ such that*

i $\int |g(\boldsymbol{x})| d\boldsymbol{x} < \infty$,

ii $\lim_{\|\boldsymbol{x}\| \longrightarrow 0} \|\boldsymbol{x}\|^n |g(\boldsymbol{x})| = 0$,

iii $\sup |g(\boldsymbol{x})| < \infty$,

where $\|\boldsymbol{x}\|$ is the Euclidean norm of $\boldsymbol{x}$. Let $f(\boldsymbol{x})$ be another function on $\mathbb{R}^n$ such that $\int |f(\boldsymbol{x})| < \infty$. Then, at every point, $\boldsymbol{x}_0$, of continuity of f,

$$\frac{1}{h_T^n} \int_{\mathbb{R}^n} g\left(\frac{\boldsymbol{x}_0 - \boldsymbol{w}}{h_T}\right) f(\boldsymbol{w}) d\boldsymbol{w} \longrightarrow f(\boldsymbol{x}_0) \int g(\boldsymbol{w}) d\boldsymbol{w}$$

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

Proof. See Pagan and Ullah (1999, p. 362). □

LEMMA 6. *Let $f : [a, b] \longrightarrow \mathbb{R}$ be Hölder continuous, i.e., $|f(x) - f(y)| \leq H|x - y|^r$ for all $x, y \in [a, b]$, where $r \in (0, 1]$ and H is some positive constant. Then, for all $x \in [a, b]$,*

$$\left| f(x) - \frac{1}{b-a} \int_a^b f(w) dw \right| \leq \frac{H}{r+1} \left[\left(\frac{b-w}{b-a} \right)^{r+1} + \left(\frac{w-a}{b-a} \right)^{r+1} \right] (b-a)^r.$$

Proof. See Barnett et al. (2010, p. 48). □

LEMMA 7. *Let $K(x)$ be a kernel of order $r > 0$ and let $dP(x) = p(x)dx$ represent a probability measure with a bounded, $[r]$-times differentiable, density. Let $f(x)$ denote a bounded function. Then*

$$\left| E\left[\frac{1}{h}\int_{\mathbb{R}} K\left(\frac{X-x}{h}\right) f(x)dx - f(X)\right]\right| \leq 2\sup_x |f(x)| \left(\int_{\mathbb{R}} \left|D^{[r]}p(x)\right| dx\right)\left(\int_{\mathbb{R}} |y|^r |K(y)| dy\right) h^r,$$

where $[r]$ is the integer part of r.

Proof. See Giné and Nickl (2008, p. 359). □

LEMMA 8 ('Chaining' inequality)**.** *Let $\phi : \mathcal{X} \longrightarrow \mathbb{R}^+$, where $\mathcal{X}$ is a metric space endowed with the metric d, be an almost-surely continuous random function; and let $(\mathcal{X}_j, \delta_j)$ represent a sequence of metric skeletons: $\mathcal{X}_j$ is a finite 'net' in $\mathcal{X}$ such that, for every point $x \in \mathcal{X}$, there exists a point $x_j \in \mathcal{X}_j$ with $d(x, x_j) \leq \delta_j$ and $\sum_{j=0}^{\infty} \delta_j < \infty$. For any positive z and ϵ and any sequence of positive ϵ_j with $\epsilon + \sum_{j=0}^{\infty} \epsilon_j \leq 1$,*

$$\mathbb{P}\left(\sup_{x\in\mathcal{X}} \phi(x) > z\right) \leq \mathbb{P}\left(\phi(x_0) > \epsilon z\right) + \sum_{j=0}^{\infty} |\mathcal{X}_{j+1}| \sup_{d(x,y)\leq\delta_j} \mathbb{P}\left(|\phi(x) - \phi(y)| > \epsilon_j z\right).$$

Proof. See Cranston et al. (2000, p. 1858). □

Proof of Theorem 3.1. Since (3.3)-(3.4) are continuums of constraints with varying end-points, we need to replace these continuums with sets of definite integrals:

$$\int_{[a,b]}\int_{[0,1]} c(u,v)dudv = \int_{[0,1]}\int_{[a,b]} c(u,v)dudv = b - a, \tag{A.11}$$

where a and b are arbitrary numbers in $[0, 1]$. Using a dense dyadic sequence in $[0, 1]$, (A.11) 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, (A.11) is equivalent to

$$\int_{[k2^{-n},(k+1)2^{-n}]} \int_{[0,1]} c(u,v)dudv = \int_{[0,1]} \int_{[k2^{-n},(k+1)2^{-n}]} c(u,v)dudv = \frac{1}{2^n}$$
$$\forall k = 0, 1, 2, \ldots, (2^n - 1), \text{ and } n \text{ is large enough.} \tag{A.12}$$

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

$$\begin{aligned}
\mathcal{L}(c, \mathbf{\Lambda}_n; \widehat{\boldsymbol{\theta}}) &= -\int_{[0,1]^2} c(u,v) \log c(u,v) dudv - \lambda_{-1}\Big[\int_{[0,1]^2} c(u,v)dudv - 1\Big] \\
&- \sum_{k=0}^{2^n-1} \left\{ \lambda_k \int_{[k2^{-n},(k+1)2^{-n}]} \int_{[0,1]} \big[c(u,v) - 2^{-n}\big] + \gamma_k \int_{[0,1]} \int_{[k2^{-n},(k+1)2^{-n}]} \big[c(u,v) - 2^{-n}\big] \right\} \\
&- \lambda_{2^n} \int_{[0,1]^2} h(u,v,\widehat{\boldsymbol{\theta}}) c(u,v) dudv \\
&= -\int_{[0,1]^2} \Bigg\{ c(u,v) \log c(u,v) + \lambda_{-1}[c(u,v) - 1] \\
&+ \sum_{k=0}^{2^n-1} \Big(\lambda_k \mathbb{I}(u \in [k2^{-n}, (k+1)2^{-n}]) + \gamma_k \mathbb{I}(v \in [k2^{-n}, (k+1)2^{-n}]) \Big) [c(u,v) - 2^{-n}] \\
&+ \lambda_{2^n} h(u,v,\widehat{\boldsymbol{\theta}}) c(u,v) \Bigg\} dudv.
\end{aligned} \tag{A.13}$$

Taking the first derivative of $\mathcal{L}(c, \mathbf{\Lambda}_n; \widehat{\boldsymbol{\theta}})$ with respect to c leads to

$$\begin{aligned}
\int_{[0,1]^2} \Bigg\{ \log c(u,v) &+ (1 + \lambda_{-1}) + \sum_{k=0}^{2^n-1} \Big[\lambda_k \mathbb{I}(u \in [k2^{-n}, (k+1)2^{-n}]) \\
&+ \gamma_k \mathbb{I}(v \in [k2^{-n}, (k+1)2^{-n}]) \Big] + \lambda_{2^n} h(u,v,\widehat{\boldsymbol{\theta}}) \Bigg\} dudv = 0.
\end{aligned} \tag{A.14}$$

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

$$b(u,v) = \frac{b_n(u,v)}{\widetilde{c}(u,v) - 1}, \tag{A.15}$$

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

$$\begin{aligned}\widehat{c}_{n,N_h}(u,v) &= \exp\Bigg\{-(1+\lambda_{-1}-b_0)-\sum_{k=0}^{2^n-1}\Big[\lambda_k\mathbb{I}(u\in[k2^{-n},(k+1)2^{-n}])+\gamma_k\mathbb{I}(v\in[k2^{-n},(k+1)2^{-n}])\Big]\\ &- \lambda_{2^n}h(u,v,\widehat{\boldsymbol{\theta}})-b_0\widetilde{c}(u,v)\Bigg\},\end{aligned} \tag{A.16}$$

and b_0 is a generic constant. Now, by substituting (A.16) into (3.2) the leading term, $1+\lambda_{-1}-b_0$, is canceled out, then we obtain:

$$\widehat{c}_{n,N_h}(u,v)=\frac{\mathcal{E}_{n,N_h}(u,v)}{\int_{[0,1]^2}\mathcal{E}_{n,N_h}(u,v)dudv}, \tag{A.17}$$

where

$$\begin{aligned}\mathcal{E}_{n,N_h}(u,v) &= \exp\Bigg\{-\sum_{k=0}^{2^n-1}\Big[\widehat{\lambda}_k\mathbb{I}(u\in[k2^{-n},(k+1)2^{-n}])+\widehat{\gamma}_k\mathbb{I}(v\in[k2^{-n},(k+1)2^{-n}])\Big]-\widehat{\lambda}_{2^n}h(u,v,\widehat{\boldsymbol{\theta}})\\ &- b_0\widetilde{c}(u,v)\Bigg\}.\end{aligned}$$

The Lagrangian multipliers $\widehat{\boldsymbol{\Lambda}}_n=\left\{\widehat{\lambda}_0,\ldots,\widehat{\lambda}_{2^n},\widehat{\gamma}_0,\ldots,\widehat{\gamma}_{2^n-1}\right\}$ can be solved out by substituting (A.17) into (3.3), (3.4), and (3.5), leading to the following system of equations:

$$\begin{aligned}\tfrac{1}{\int_{[0,1]^2}\mathcal{E}_{n,N_h}(u,v)dudv}\int_{[0,1]^2}\mathbb{I}(u\in[k2^{-n},(k+1)2^{-n}])\mathcal{E}_{n,N_h}(u,v)dudv &= 2^{-n},\\ \tfrac{1}{\int_{[0,1]^2}\mathcal{E}_{n,N_h}(u,v)dudv}\int_{[0,1]^2}\mathbb{I}(v\in[k2^{-n},(k+1)2^{-n}])\mathcal{E}_{n,N_h}(u,v)dudv &= 2^{-n},\\ \tfrac{1}{\int_{[0,1]^2}\mathcal{E}_{n,N_h}(u,v)dudv}\int_{[0,1]^2}h(u,v,\widehat{\boldsymbol{\theta}})\mathcal{E}_{n,N_h}(u,v)dudv &= 0\end{aligned} \tag{A.18}$$

for all $k=0,\ldots,(2^n-1)$. Since (A.17) can be rewritten as

$$\begin{aligned}\widehat{c}_{n,N_h}(u,v) &= \frac{-\sum_{k=0}^{2^n-1}(\widehat{\lambda}_k2^{-n}+\widehat{\gamma}_k2^{-n})}{\int_{[0,1]^2}\mathcal{E}_{n,N_h}(u,v)dudv}\exp\Bigg\{-\sum_{k=0}^{2^n-1}\Big[\widehat{\lambda}_k(\mathbb{I}(u\in[k2^{-n},(k+1)2^{-n}])-2^{-n})\\ &+ \widehat{\gamma}_k(\mathbb{I}(v\in[k2^{-n},(k+1)2^{-n}])-2^{-n})\Big]-\widehat{\lambda}_{2^n}h(u,v,\widehat{\boldsymbol{\theta}})-b_0\widetilde{c}(u,v)\Bigg\},\end{aligned}$$

we can define the potential function as follows:

$$
\begin{aligned}
Q_{n,N_h}(\boldsymbol{\Lambda}_n, \widehat{\boldsymbol{\theta}}) &= \int_{[0,1]^2} \exp\Bigg\{ -\sum_{k=0}^{2^n-1} \Big[\lambda_k (\mathbb{I}(u \in [k2^{-n}, (k+1)2^{-n}]) - 2^{-n}) \\
&+ \gamma_k (\mathbb{I}(v \in [k2^{-n}, (k+1)2^{-n}]) - 2^{-n}) \Big] - \lambda_{2^n} h(u, v, \widehat{\boldsymbol{\theta}}) - b_0 \widetilde{c}(u, v) \Bigg\} du dv.
\end{aligned}
$$

Then, (A.18) is equivalent to the following system of equations:

$$
\begin{aligned}
\tfrac{\partial}{\partial \lambda_k} Q_{n,N_h}(\boldsymbol{\Lambda}_n, \widehat{\boldsymbol{\theta}}) &= 0, \\
\tfrac{\partial}{\partial \gamma_k} Q_{n,N_h}(\boldsymbol{\Lambda}_n, \widehat{\boldsymbol{\theta}}) &= 0, \\
\tfrac{\partial}{\partial \lambda_{2^n}} Q_{n,N_h}(\boldsymbol{\Lambda}_n, \widehat{\boldsymbol{\theta}}) &= 0
\end{aligned}
\tag{A.19}
$$

for all $k = 0, \ldots, (2^n - 1)$. Also note that, since the second order derivatives of $Q_{n,N_h}(\boldsymbol{\Lambda}_n, \widehat{\boldsymbol{\theta}})$ is the covariance matrix of $\left\{ \mathbb{I}(u \in [k2^{-n}, (k+1)2^{-n}]), \mathbb{I}(v \in [k2^{-n}, (k+1)2^{-n}]), h(u, v, \widehat{\boldsymbol{\theta}}) \right\}_{k=0}^{2^n-1}$, thus $Q_{n,N_h}(\boldsymbol{\Lambda}_n, \widehat{\boldsymbol{\theta}})$ is positive definite. It follows that the solutions to (A.19) are the minimum values of $Q_{n,N_h}(\boldsymbol{\Lambda}_n, \widehat{\boldsymbol{\theta}})$, which depend on $\widehat{\boldsymbol{\theta}}$, b_0, and $\widetilde{c}(u, v)$.

Since the potential function $Q_{n,N_h}(\boldsymbol{\Lambda}_n, \widehat{\boldsymbol{\theta}})$ and the MEC (A.17) 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 4 yields

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

We then immediately obtain:

$$
Q_{n,N_h}(\boldsymbol{\Lambda}_n, \widehat{\boldsymbol{\theta}}) = \int_{[0,1]^2} \exp\Bigg\{ \sum_{k=0}^{2^n-1} \Bigg[\lambda_k \Big(\Phi(N_h(k2^{-n} - u)) + \Phi(-N_h((k+1)2^{-n} - u)) - 1 + 2^{-n} \Big)
$$

$$
\begin{aligned}
&+ \quad \gamma_k\Big(\Phi(N_h(k2^{-n}-v))+\Phi(-N_h((k+1)2^{-n}-v))-1+2^{-n}\Big)\Big] \\
&- \quad \lambda_{2^n}h(u,v,\widehat{\boldsymbol{\theta}})-b_0\widetilde{c}(u,v)\Big\}dudv, \qquad \text{(A.20)}
\end{aligned}
$$

and

$$
\begin{aligned}
\mathcal{E}_{n,N_h}(u,v) \;\approx\; & \exp\Bigg\{\sum_{k=0}^{2^n-1}\Bigg[\widehat{\lambda}_k\Big(\Phi(N_h(k2^{-n}-u))+\Phi(-N_h((k+1)2^{-n}-u))\Big) \\
&+ \widehat{\gamma}_k\Big(\Phi(N_h(k2^{-n}-v))+\Phi(-N_h((k+1)2^{-n}-v))\Big)\Bigg] \\
&- \widehat{\lambda}_{2^n}h(u,v,\widehat{\boldsymbol{\theta}})-b_0\widetilde{c}(u,v)\Bigg\}, \qquad \text{(A.21)}
\end{aligned}
$$

where $\widehat{\boldsymbol{\Lambda}}_n$ are the minimum values of (A.20). In particular, $\widehat{c}_{n,N_h}(u,v)=\frac{\mathcal{E}_{n,N_h}(u,v)}{\int_0^1\int_0^1\mathcal{E}_{n,N_h}(u,v)dudv}$ can be symmetrized by letting $\lambda_k=\gamma_k$ for every $k=0,\ldots,(2^n-1)$ and letting $h(u,v,\widehat{\boldsymbol{\theta}})$ be a symmetric function.

Finally, to complete this proof, we still need to prove that the MEC approximator, $\widehat{C}_{n,N_h}(u,v)=\int_0^u\int_0^v\widehat{c}_{n,N_h}(u,v)dudv$, is *2-increasing* (cf. Definition 6.1). 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 $\widehat{c}_{n,N_h}(u,v)$ is a [positive] exponential function, the mass of the rectangle, $\widehat{C}_{n,N_h}(u_1+\Delta,v_1+\Delta)-\widehat{C}_{n,N_h}(u_1+\Delta,v_1)-\widehat{C}_{n,N_h}(u_1,v_1+\Delta)+\widehat{C}_{n,N_h}(u_1,v_1)=\int_{u_1}^{u_1+\Delta}\int_{v_1}^{v_1+\Delta}\widehat{c}_{n,N_h}(u,v)dudv$, is thus nonnegative. Now, we can obtain the MECs by letting n and N_h become sufficiently large. □

Proof of Theorem 3.3. First of all, we shall define the following space of copulas:

$$
\mathfrak{C}\doteq\Big\{c\in W_1^r([0,1]^2):\; c\in(0,\infty);\; \int_0^1\int_0^1 c(u,v)dudv=1;\int_0^1 c(u,v)du=1;\text{ and }\int_0^1 c(u,v)dv=1 \\
\text{for every } u,v\in[0,1]\Big\},
$$

where $W_1^r([0,1]^2)$ represents the space of functions, c, whose partial derivatives up to an order, r, exist and lie in $\ell^1(\mathbb{R})$.

In view of the Lagrange lemma stated as Lemma 3, the continuums of marginal (side) constraints

with varying end-points on c, defined by (3.3) and (3.4), are equivalent to the following integral constraints:

$$\int_{\epsilon}^{1} \lambda(u) \left\{ \int_{u-\epsilon}^{u} \int_{0}^{1} c(w,v) dw dv - \epsilon \right\} du = 0$$

for an arbitrarily small positive constant, ϵ, where $\lambda(u)$ is a continuously smooth function satisfying $\lambda(0) = \lambda(1) = 0$, and

$$\int_{\epsilon}^{1} \gamma(v) \left\{ \int_{0}^{1} \int_{v-\epsilon}^{v} c(u,w) du dw - \epsilon \right\} dv = 0,$$

where $\gamma(v)$ is a continuously smooth function satisfying $\gamma(0) = \gamma(1) = 0$. Hence the Lagrangian function associated with **Problem A** is given by

$$\begin{aligned} \mathcal{L}\left(c, \mathbf{\Lambda}; \boldsymbol{\theta}_0\right) &= -E[\log c(U,V)] - \lambda^* \left(E[\mathbf{1}] - 1\right) - \gamma^* E[h(U,V,\boldsymbol{\theta}_0)] - \int_{\epsilon}^{1} \lambda(u) \left(E[\mathbb{I}(w \in [u-\epsilon, u])] - \epsilon\right) du \\ &- \int_{\epsilon}^{1} \gamma(v) \left(E[\mathbb{I}(w \in [v-\epsilon, v])] - \epsilon\right) dv, \end{aligned} \tag{A.22}$$

where $\mathbf{\Lambda} = \{\lambda^*, \gamma^*, \lambda(u), \gamma(v)\}'$; and all the expectations are hereafter taken with respect to c belonging to the space $W_1^r([0,1]^2)$, unless otherwise indicated. The MEC is now defined as

$$\begin{aligned} c_0(u,v) &\doteq \arg\sup_{c \in \mathcal{C}} \{-E[\log c(U,V)]\} \\ &= \arg \sup_{c \in \mathfrak{C}, \gamma^* \in B_{\gamma^*}} \{-E[\log c(U,V)] - \gamma^* E[h(U,V,\boldsymbol{\theta}_0)]\} \\ &= \arg \sup_{c \in W_1^r([0,1]^2), \mathbf{\Lambda} \in B_{\lambda^*} \times B_{\gamma^*} \times C^*([0,1]) \times C^*([0,1])} \mathcal{L}\left(c, \mathbf{\Lambda}; \boldsymbol{\theta}_0\right), \end{aligned} \tag{A.23}$$

where B_{λ^*} and B_{γ^*} are bounded sets in $\mathbb{R}$;

$$\mathcal{C} \doteq \left\{ c \in W_1^r([0,1]^2) : \ c \in (0,\infty), \ E[\mathbf{1}] = 1, \ \int_{\epsilon}^{1} \lambda(u) \left\{ \int_{u-\epsilon}^{u} \int_{0}^{1} c(w,v) dw dv - \epsilon \right\} = 0, \right.$$
$$\left. \int_{\epsilon}^{1} \gamma(v) \left\{ \int_{0}^{1} \int_{v-\epsilon}^{v} c(u,w) du dw - \epsilon \right\} = 0, \text{ and } E[h(U,V,\boldsymbol{\theta}_0)] = 0, \text{ where } \lambda(u), \gamma(v) \in C^*([0,1]) \right\};$$

and $C^*([0,1])$ denotes a space of continuously smooth functions vanishing at the end points. By

using a dense sequence in $[0,1]$, one can define the following approximation spaces:

$$\begin{aligned}\mathfrak{C}_T \;\doteq\; \Bigg\{ & c \in W_1^r([0,1]^2):\; c \in (0,\infty),\; E[\mathbf{1}] = 1,\; \frac{1}{N_T}\sum_{k=1}^{N_T} \lambda_k \left\{ \int_{\frac{k-1}{N_T}}^{\frac{k}{N_T}} \int_0^1 c(w,v)dwdv - \frac{1}{N_T} \right\} = 0, \\ & \text{and } \frac{1}{N_T}\sum_{k=1}^{N_T} \gamma_k \left\{ \int_0^1 \int_{\frac{k-1}{N_T}}^{\frac{k}{N_T}} c(u,w)dudw - \frac{1}{N_T} \right\} = 0 \text{ with } \{\lambda_k, \gamma_k\} \in B_\lambda \times B_\gamma \Bigg\},\end{aligned}$$

where the λ_k's and the γ_k's hereafter represent some values of $\lambda(u)$ and $\gamma(v)$ respectively on $\left[\frac{k-1}{N_T}, \frac{k}{N_T}\right)$, and

$$\mathcal{C}_T(\widehat{\boldsymbol{\theta}}) \;\doteq\; \left\{ c \in \mathfrak{C}_T:\; E[h(U,V,\widehat{\boldsymbol{\theta}})] = 0 \right\},$$

where B_λ and B_γ are bounded sets in $\mathbb{R}$. It then follows that the approximated Lagrangian function is given by

$$\begin{aligned}\mathcal{L}\left(c, \boldsymbol{\Lambda}_T; \boldsymbol{\theta}_0\right) &= -E[\log c(U,V)] - \lambda^* \left(E[\mathbf{1}] - 1\right) - \gamma^* E[h(U,V,\boldsymbol{\theta}_0)] \\ &- \frac{1}{N_T}\sum_{k=1}^{N_T} \lambda_k \left\{ \int_{\frac{k-1}{N_T}}^{\frac{k}{N_T}} \int_0^1 c(w,v)dwdv - \frac{1}{N_T} \right\} - \frac{1}{N_T}\sum_{k=1}^{N_T} \gamma_k \left\{ \int_0^1 \int_{\frac{k-1}{N_T}}^{\frac{k}{N_T}} c(u,w)dudw - \frac{1}{N_T} \right\} \\ &= -E[\log c(U,V)] - \lambda^* \left(E[\mathbf{1}] - 1\right) - \gamma^* E[h(U,V,\boldsymbol{\theta}_0)] \\ &- \frac{1}{N_T}\sum_{k=1}^{N_T} \lambda_k \left\{ E\left[\mathbb{I}\left(u \in \left[\frac{k-1}{N_T}, \frac{k}{N_T}\right)\right)\right] - \frac{1}{N_T} \right\} \\ &- \frac{1}{N_T}\sum_{k=1}^{N_T} \gamma_k \left\{ E\left[\mathbb{I}\left(v \in \left[\frac{k-1}{N_T}, \frac{k}{N_T}\right)\right)\right] - \frac{1}{N_T} \right\},\end{aligned}$$

where $\boldsymbol{\Lambda}_T = \{\lambda^*, \gamma^*, \lambda_1, \ldots, \lambda_{N_T}, \gamma_1, \ldots, \gamma_{N_T}\}'$. By virtue of Lemma 5, a kernel-based smoothing of indicator functions yields

$$\begin{aligned}\mathcal{L}_T\left(c, \boldsymbol{\Lambda}_T; \boldsymbol{\theta}_0\right) &= E[\log c(U,V)] - \lambda^* \left(E[\mathbf{1}] - 1\right) - \gamma^* E[h(U,V,\boldsymbol{\theta}_0)] \\ &- \frac{1}{N_T}\sum_{k=1}^{N_T} \lambda_k \left\{ E\left[\frac{1}{h_T} \int_{\frac{k-1}{N_T}}^{\frac{k}{N_T}} K\left(\frac{u-w}{h_T}\right) dw \right] - \frac{1}{N_T} \right\} \\ &- \frac{1}{N_T}\sum_{k=1}^{N_T} \gamma_k \left\{ E\left[\frac{1}{h_T} \int_{\frac{k-1}{N_T}}^{\frac{k}{N_T}} K\left(\frac{v-w}{h_T}\right) dw \right] - \frac{1}{N_T} \right\},\end{aligned}$$

where $K(\cdot)$ is a continuously bounded kernel function satisfying Assumption (A1). The smoothed approximation spaces are then defined as

$$\begin{aligned}
\mathfrak{C}_T^* &\doteq \left\{ c \in W_1^r([0,1]^2) : \ c \in (0,\infty), \ E[\mathbf{1}] = 1, \ \frac{1}{N_T}\sum_{k=1}^{N_T} \lambda_k \left\{ E\left[\frac{1}{h_T} \int_{\frac{k-1}{N_T}}^{\frac{k}{N_T}} K\left(\frac{u-w}{h_T}\right) dw \right] \right.\right. \\
&- \left.\left. \frac{1}{N_T} \right\} = 0, \text{ and } \frac{1}{N_T}\sum_{k=1}^{N_T} \gamma_k \left\{ E\left[\frac{1}{h_T} \int_{\frac{k-1}{N_T}}^{\frac{k}{N_T}} K\left(\frac{v-w}{h_T}\right) dw \right] - \frac{1}{N_T} \right\} = 0 \right. \\
&\quad \text{with } \left. \{\lambda_k, \gamma_k\} \in B_\lambda \times B_\gamma \right\}. \\
\mathcal{C}_T^*(\widehat{\boldsymbol{\theta}}) &\doteq \left\{ c \in \mathfrak{C}_T^* : \ E[h(U,V,\widehat{\boldsymbol{\theta}})] = 0 \right\}.
\end{aligned}$$

First, we shall show that

$$\sup_{c \in \mathcal{C}_T^*(\widehat{\boldsymbol{\theta}})} \{-E[\log c(U,V)]\} \geq \sup_{c \in \mathcal{C}} \{-E[\log c(U,V)]\} - o_P(1).$$

In view of Eq. (A.23), it is sufficient to show that

$$\begin{aligned}
&\max_{\{c, \boldsymbol{\Lambda}_T\} \in \left\{ W_1^r([0,1]^2), B_{\lambda^*} \times B_{\gamma^*} \times B_\lambda^{N_T} \times B_\gamma^{N_T} \right\}} \mathcal{L}_T\left(c, \boldsymbol{\Lambda}_T; \widehat{\boldsymbol{\theta}}\right) \\
&\qquad \geq \max_{\{c, \boldsymbol{\Lambda}\} \in \left\{ W_1^r([0,1]^2), B_{\lambda^*} \times B_{\gamma^*} \times C^*([0,1]) \times C^*([0,1]) \right\}} \mathcal{L}\left(c, \boldsymbol{\Lambda}; \boldsymbol{\theta}_0\right) - o_P(1), \quad \text{(A.24)}
\end{aligned}$$

where $B_\lambda^{N_T} \times B_\gamma^{N_T}$ is a bounded set in $\mathbb{R}^{N_T} \times \mathbb{R}^{N_T}$. Note that

$$\begin{aligned}
\mathcal{L}_T(c, \boldsymbol{\Lambda}_T; \widehat{\boldsymbol{\theta}}) &= \mathcal{L}\left(c, \boldsymbol{\Lambda}, \boldsymbol{\theta}_0\right) + \mathcal{L}\left(c, \boldsymbol{\Lambda}_T; \boldsymbol{\theta}_0\right) - \mathcal{L}\left(c, \boldsymbol{\Lambda}; \boldsymbol{\theta}_0\right) + \mathcal{L}_T\left(c, \boldsymbol{\Lambda}_T; \widehat{\boldsymbol{\theta}}\right) \\
&- \mathcal{L}_T\left(c, \boldsymbol{\Lambda}_T; \boldsymbol{\theta}_0\right) + \mathcal{L}_T\left(c, \boldsymbol{\Lambda}_T; \boldsymbol{\theta}_0\right) - \mathcal{L}\left(c, \boldsymbol{\Lambda}_T; \boldsymbol{\theta}_0\right) \\
&\geq \mathcal{L}\left(c, \boldsymbol{\Lambda}, \boldsymbol{\theta}_0\right) - \left|\mathcal{L}\left(c, \boldsymbol{\Lambda}_T; \boldsymbol{\theta}_0\right) - \mathcal{L}\left(c, \boldsymbol{\Lambda}; \boldsymbol{\theta}_0\right)\right| - \left|\mathcal{L}_T\left(c, \boldsymbol{\Lambda}_T; \widehat{\boldsymbol{\theta}}\right) - \mathcal{L}_T\left(c, \boldsymbol{\Lambda}_T; \boldsymbol{\theta}_0\right)\right| \\
&- \left|\mathcal{L}_T\left(c, \boldsymbol{\Lambda}_T; \boldsymbol{\theta}_0\right) - \mathcal{L}\left(c, \boldsymbol{\Lambda}_T; \boldsymbol{\theta}_0\right)\right| \\
&= \mathcal{L}\left(c, \boldsymbol{\Lambda}, \boldsymbol{\theta}_0\right) - \left\{\mathcal{T}_1 + \mathcal{T}_2 + \mathcal{T}_3\right\}.
\end{aligned}$$

We now proceed to bound the above three terms: For $\mathcal{T}_1$, one has

$$\begin{aligned}
\mathcal{T}_1 &\leq \left| \frac{1}{N_T} \sum_{k=1}^{N_T} \lambda_k \left\{ E\left[\mathbb{I}\left(u \in \left[\frac{k-1}{N_T}, \frac{k}{N_T} \right) \right) \right] - \frac{1}{N_T} \right\} - \int_{\frac{1}{N_T}}^{1} \lambda(u) \left\{ E\left[\mathbb{I}\left(w \in \left[u - \frac{1}{N_T}, u \right] \right) \right] - \frac{1}{N_T} \right\} \right| \\
&+ \left| \frac{1}{N_T} \sum_{k=1}^{N_T} \gamma_k \left\{ E\left[\mathbb{I}\left(v \in \left[\frac{k-1}{N_T}, \frac{k}{N_T} \right) \right) \right] - \frac{1}{N_T} \right\} - \int_{\frac{1}{N_T}}^{1} \gamma(v) \left\{ E\left[\mathbb{I}\left(w \in \left[v - \frac{1}{N_T}, v \right] \right) \right] - \frac{1}{N_T} \right\} \right| \\
&= \mathcal{T}_{1;a} + \mathcal{T}_{1;b},
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{T}_{1;a} \leq \frac{1}{N_T} \sum_{k=1}^{N_T} \Bigg| & \lambda_k \left\{ E\left[\mathbb{I}\left(u \in \left[\frac{k-1}{N_T}, \frac{k}{N_T} \right) \right) \right] - \frac{1}{N_T} \right\} \\
& - N_T \int_{\frac{k-1}{N_T}}^{\frac{k}{N_T}} \lambda\left(u + \frac{1}{N_T} \right) \left\{ E\left[\mathbb{I}\left(w \in \left[u, u + \frac{1}{N_T} \right] \right) \right] - \frac{1}{N_T} \right\} du \Bigg|
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{T}_{1;b} \leq \frac{1}{N_T} \sum_{k=1}^{N_T} \Bigg| & \gamma_k \left\{ E\left[\mathbb{I}\left(v \in \left[\frac{k-1}{N_T}, \frac{k}{N_T} \right) \right) \right] - \frac{1}{N_T} \right\} \\
& - N_T \int_{\frac{k-1}{N_T}}^{\frac{k}{N_T}} \gamma\left(v + \frac{1}{N_T} \right) \left\{ E\left[\mathbb{I}\left(w \in \left[v, v + \frac{1}{N_T} \right] \right) \right] - \frac{1}{N_T} \right\} dv \Bigg| .
\end{aligned}$$

Setting $\epsilon = 1/N_T$, it then follows from Lemma 6 that, under Assumption (A2), $\mathcal{T}_{1;a} = O\left(N_T^{-r_1}\right)$ and $\mathcal{T}_{1;b} = O\left(N_T^{-r_1}\right)$ for some $r \in (0, 1]$. Hence, it immediately follows that $\mathcal{T}_1 = O\left(\max(N_T^{-r_1}, N_T^{-r_2})\right)$.

For the second term $\mathcal{T}_2$, the mean-value theorem yields $h(u, v, \widehat{\boldsymbol{\theta}}) = h(u, v, \boldsymbol{\theta}_0) + D'_{\boldsymbol{\theta}} h(u, v, \overline{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$, where $\overline{\boldsymbol{\theta}}$ lies on the line joining $\boldsymbol{\theta}_0$ and $\widehat{\boldsymbol{\theta}}$ in the Euclidean space. Under Assumptions (B1) and (B2), we obtain $\mathcal{T}_2 = O_P\left(T^{-1/2}\right)$.

For the third term $\mathcal{T}_3$, an application of Lemma 7 yields; for some $c \in W_1^r([0, 1]^2)$ and $K \in \mathcal{K}^r(\mathbb{R})$ with some integer, $r > 1$;

$$\left| E\left[\frac{1}{h_T} \int_{\frac{k-1}{N_T}}^{\frac{k}{N_T}} K\left(\frac{u - w}{h_T} \right) dw - \mathbb{I}\left(u \in \left[\frac{k-1}{N_T}, \frac{k}{N_T} \right) \right) \right] \right|$$

$$\leq 2 \times \text{Const.} \times \sup_{v\in[0,1]} \left(\int_{\mathbb{R}} \left| D_u^{[r]} c(u,v) \right| du \right) \left(\int_{\mathbb{R}} |u|^r K(u) du \right) h_T^r$$

and

$$\left| E\left[\frac{1}{h_T} \int_{\frac{k-1}{N_T}}^{\frac{k}{N_T}} K\left(\frac{v-w}{h_T} \right) dw - \mathbb{I}\left(v \in \left[\frac{k-1}{N_T}, \frac{k}{N_T} \right) \right) \right] \right|$$
$$\leq 2 \times \text{Const.} \times \sup_{v\in[0,1]} \left(\int_{\mathbb{R}} \left| D_v^{[r]} c(u,v) \right| dv \right) \left(\int_{\mathbb{R}} |v|^r K(v) dv \right) h_T^r.$$

It then follows that $\mathcal{T}_3 = O(h_T^r)$. Gathering all the terms, one can derive that

$$\mathcal{L}_T(c, \mathbf{\Lambda}_T, \widehat{\boldsymbol{\theta}}) \geq \mathcal{L}(c, \mathbf{\Lambda}, \boldsymbol{\theta}_0) - O_P\left(\max\left(N_T^{-r_1}, N_T^{-r_2}, h_T^r, T^{-1/2} \right) \right).$$

Thus, Eq. (A.24) follows.

Next, we shall derive the convergence rate for $\widehat{c}_T(u,v) = \arg\sup_{c\in\mathcal{C}_T^*(\widehat{\boldsymbol{\theta}})} \{-E[\log c(U,V)]\}$. This process consists of 2 main steps:

Step 1: First we choose sequences of constants, $\epsilon_T^{(1)} \doteq T^{-\ell}$ and $\eta_T^{(1)} = D_3 T^{-\ell^*}$, where D_3 is some generic constant chosen later, $\ell \in (0, 1/2)$, and $\ell^* \in (\ell/\omega, 1/2)$. Let $\Pi_T c_0$ denote the image of the MEC $c_0(u,v)$, defined by Eq. (A.23), on the approximation space $\mathfrak{C}_T^*$, where Π_T signifies a projection operator; and $\Pi_T \mathbf{\Lambda}_0 = \{\lambda_0^*, \gamma_0^*, \Pi_T \lambda_0(u), \Pi_T \gamma_0(v)\}$. We obtain

$$\mathbb{P}\left(\|\widehat{c}_T - c_0\|_{\mathfrak{C}} \geq D\epsilon_T^{(1)} \right) \leq \mathbb{P}\left(\sup_{\substack{\{c,\mathbf{\Lambda}_T\}\in\{W_1^r([0,1]^2), B_{\lambda^*}\times B_{\gamma^*}\times B_\lambda^{N_T}\times B_\gamma^{N_T}\} \\ \begin{cases} \|c-c_0\|_{\mathfrak{C}} \geq D\epsilon_T^{(1)} \\ c \in \mathfrak{C}_T^* \end{cases}} \mathcal{L}_T\left(c, \mathbf{\Lambda}_T; \widehat{\boldsymbol{\theta}} \right) \geq \mathcal{L}_T\left(\Pi_T c_0, \Pi_T \mathbf{\Lambda}_0, \widehat{\boldsymbol{\theta}} \right) \right)$$

$$
\begin{aligned}
&= \mathbb{P}\left(\left\{\sup_{\substack{\{c,\boldsymbol{\Lambda}_T\}\in\{W_1^r([0,1]^2),B_{\lambda^*}\times B_{\gamma^*}\times B_\lambda^{N_T}\times B_\gamma^{N_T}\}\\ \begin{cases}\|c-c_0\|_{\mathfrak{C}}\geq D\epsilon_T^{(1)}\\ c\in\mathfrak{C}_T^*\end{cases}}} \mathcal{L}_T\left(c,\boldsymbol{\Lambda}_T;\widehat{\boldsymbol{\theta}}\right)\geq\mathcal{L}_T\left(\Pi_T c_0,\Pi_T\boldsymbol{\Lambda}_0,\widehat{\boldsymbol{\theta}}\right)\right\}\right.\\
&\quad\left.\bigcap\left\{\sup_{\{c,\boldsymbol{\Lambda}_T\}\in\{\mathfrak{C}_T^*,B_{\lambda^*}\times B_{\gamma^*}\times B_\lambda^{N_T}\times B_\gamma^{N_T}\}}\left|\mathcal{L}_T(c,\boldsymbol{\Lambda}_T,\widehat{\boldsymbol{\theta}})-\mathcal{L}_T(c,\boldsymbol{\Lambda}_T,\boldsymbol{\theta}_0)\right|>\eta_T^{(1)}\right\}\right.\\
&\quad\bigcup\left\{\sup_{\substack{\{c,\boldsymbol{\Lambda}_T\}\in\{W_1^r([0,1]^2),B_{\lambda^*}\times B_{\gamma^*}\times B_\lambda^{N_T}\times B_\gamma^{N_T}\}\\ \begin{cases}\|c-c_0\|_{\mathfrak{C}}\geq D\epsilon_T^{(1)}\\ c\in\mathfrak{C}_T^*\end{cases}}} \mathcal{L}_T\left(c,\boldsymbol{\Lambda}_T;\widehat{\boldsymbol{\theta}}\right)\geq\mathcal{L}_T\left(\Pi_T c_0,\Pi_T\boldsymbol{\Lambda}_0,\widehat{\boldsymbol{\theta}}\right)\right\}\right.\\
&\quad\left.\bigcap\left\{\sup_{\{c,\boldsymbol{\Lambda}_T\}\in\{\mathfrak{C}_T^*,B_{\lambda^*}\times B_{\gamma^*}\times B_\lambda^{N_T}\times B_\gamma^{N_T}\}}\left|\mathcal{L}_T(c,\boldsymbol{\Lambda}_T,\widehat{\boldsymbol{\theta}})-\mathcal{L}_T(c,\boldsymbol{\Lambda}_T,\boldsymbol{\theta}_0)\right|\leq\eta_T^{(1)}\right\}\right)\\
&\leq\mathbb{P}\left(\left\{\sup_{\{c,\boldsymbol{\Lambda}_T\}\in\{\mathfrak{C}_T^*,B_{\lambda^*}\times B_{\gamma^*}\times B_\lambda^{N_T}\times B_\gamma^{N_T}\}}\left|\mathcal{L}_T(c,\boldsymbol{\Lambda}_T,\widehat{\boldsymbol{\theta}})-\mathcal{L}_T(c,\boldsymbol{\Lambda}_T,\boldsymbol{\theta}_0)\right|>\eta_T^{(1)}\right\}\right)\\
&\quad+\mathbb{I}\left(\sup_{\substack{\{c,\boldsymbol{\Lambda}_T\}\in\{W_1^r([0,1]^2),B_{\lambda^*}\times B_{\gamma^*}\times B_\lambda^{N_T}\times B_\gamma^{N_T}\}\\ \begin{cases}\|c-c_0\|_{\mathfrak{C}}\geq D\epsilon_T^{(1)}\\ c\in\mathfrak{C}_T^*\end{cases}}} \mathcal{L}_T(c,\boldsymbol{\Lambda}_T,\boldsymbol{\theta}_0)\geq\mathcal{L}_T(\Pi_T c_0,\Pi_T\boldsymbol{\Lambda}_0,\boldsymbol{\theta}_0)-2\eta_T^{(1)}\right)\\
&\qquad=\mathcal{T}_1^*+\mathcal{T}_2^*.
\end{aligned}
$$

We now proceed to bound $\mathcal{T}_1^*$ and $\mathcal{T}_2^*$:

Note that $\mathcal{T}_1^*=\mathbb{P}\left(\overline{\gamma^*}\sup_{c\in\mathfrak{C}_T^*}\left|E[h(U,V,\widehat{\boldsymbol{\theta}})-h(U,V,\boldsymbol{\theta}_0)]\right|>\eta_T^{(1)}\right)$, where $\overline{\gamma^*}$ is the largest value

of the bounded subset B_{γ^*}. To bound this probability, we shall rely on the 'chaining' inequality of Cranston et al. (2000) stated as Lemma 8. We start by choosing a finite 'net', $\mathfrak{C}^*_{T,j}$, in $\mathfrak{C}^*_T$ such that, for each $c \in \mathfrak{C}^*_T$, there is a $\pi_j c \in \mathfrak{C}^*_{T,j}$ such that $\|\pi_j c - c\|_{\mathfrak{C}} \leq r^{-j}$ for some integer, $r \geq 2$. Indeed, $\left|\mathfrak{C}^*_{T,j}\right|$ is the number of copulas of $\mathfrak{C}^*_T$ contained in $\mathfrak{C}^*_{T,j}$. We denote by $N(r^{-j}, \mathfrak{C}^*_T)$ the number of balls, $B_{r^{-j}}(c_0) \doteq \{c \in \mathfrak{C}^*_T : \ \|c - c_0\|_{\mathfrak{C}} < r^{-j}\}$, that covers $\mathfrak{C}^*_T$. Then, since the space $\mathfrak{C}^*_T$ is bounded, one can always choose $\mathfrak{C}^*_{T,j}$ such that $\left|\mathfrak{C}^*_{T,j}\right| \leq N(r^{-j}, \mathfrak{C}^*_T)$. By virtue of Lemma 8, it follows that

$$\begin{aligned}
\mathcal{T}^*_1 &\leq \mathbb{P}\left(\phi(c_0; \widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) > \epsilon_* \eta^{(1)}_T\right) + \sum_{j=0}^{\infty} N(r^{-(j+1)}, \mathfrak{C}^*_T) \sup_{\|c_1 - c_2\|_{\mathfrak{C}} \leq r^{-j}} \mathbb{P}\left(\left|\phi(c_1; \widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) - \phi(c_2; \widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0)\right| > \epsilon_j \eta^{(1)}_T\right) \\
&= \mathcal{T}^*_{1;a} + \mathcal{T}^*_{1;b}, \text{ for generic positive constants, } \epsilon_* \text{ and } \epsilon_0, \ldots, \epsilon_\infty, \text{ such that } \epsilon_* + \textstyle\sum_{j=0}^{\infty} \epsilon_j < \infty,
\end{aligned}$$

where $\phi(c; \widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) = \left|E_c[h(U, V, \widehat{\boldsymbol{\theta}}) - h(U, V, \boldsymbol{\theta}_0)]\right|$. The mean-value theorem yields

$$\mathcal{T}^*_{1;a} = \mathbb{P}\left(\left|D_{\boldsymbol{\theta}} E[h(U, V, \boldsymbol{\theta})]\right|_{\boldsymbol{\theta} = \overline{\boldsymbol{\theta}}} \left|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right| > \epsilon_* \eta^{(1)}_T\right),$$

where $\overline{\boldsymbol{\theta}}$ is some point on the line joining $\boldsymbol{\theta}_0$ and $\widehat{\boldsymbol{\theta}}$ in the Euclidean space; and $D_{\boldsymbol{\theta}}$ contains first-order partial derivatives. By virtue of Assumptions (B1) and (B2), we obtain $\mathcal{T}^*_{1;a} \longrightarrow 0$. Next, to bound $\mathcal{T}^*_{1;b}$, note that, by the mean-value theorem,

$$\phi(c_1; \widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) - \phi(c_2; \widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) = \left\{\left|D_{\boldsymbol{\theta}} E_{c_1}[h(U, V, \boldsymbol{\theta})]\right|_{\boldsymbol{\theta} = \boldsymbol{\theta}^*} - \left|D_{\boldsymbol{\theta}} E_{c_2}[h(U, V, \boldsymbol{\theta})]\right|_{\boldsymbol{\theta} = \boldsymbol{\theta}^{**}}\right\} \left|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right|,$$

where $\boldsymbol{\theta}^*$ and $\boldsymbol{\theta}^{**}$ are some points on the line joining $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$. The Tchebyshev inequality gives that

$$\sup_{\|c_1 - c_2\|_{\mathfrak{C}} \leq r^{-j}} \mathbb{P}\left(\left|\phi(c_1; \widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) - \phi(c_2; \widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0)\right| > \epsilon_j \eta^{(1)}_T\right) \leq r^{-j} \left\{\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \int_0^1 \int_0^1 \left|D_{\boldsymbol{\theta}} h(u, v, \boldsymbol{\theta})\right| du dv\right\} \frac{E\left[\left|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right|\right]}{\epsilon_j \eta^{(1)}_T}.$$

It then follows that

$$\mathcal{T}^*_{1;b} \leq \left\{\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \int_0^1 \int_0^1 \left|D_{\boldsymbol{\theta}} h(u, v, \boldsymbol{\theta})\right| du dv\right\} \frac{E\left[\left|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right|\right]}{\eta^{(1)}_T} \sum_{j=0}^{\infty} r^{-j} N(r^{-(j+1)}, \mathfrak{C}^*_T)/\epsilon_j.$$

Setting $\epsilon_j = \frac{1}{r^{-j-a}\left(\log\left(\frac{1}{r^{-j-1}}\right)\right)^p}$ for some $p > 1$ and $a > 1$, by the condensation test (see, e.g., Widder (1971, p. 21)) the sum $\sum_{j=0}^{\infty} \epsilon_j < \infty$. Then, we have

$$\mathcal{T}_{1;b}^* \leq \text{Const.} \times \frac{E\left[|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0|\right]}{\eta_T^{(1)}} \sum_{j=0}^{\infty} r^{-j-a} \left(\log\left(1/r^{-j-1}\right)\right)^p N(r^{-(j+1)}, \mathfrak{C}_T^*)(r^{-j} - r^{-j-1}).$$

Let ϵ represent an arbitrarily small constant such that $r^{-j-a} < \epsilon < r^{-j-1}$. We have $N(r^{-(j+1)}, \mathfrak{C}_T^*) < N(\epsilon, \mathfrak{C}_T^*)$ and whereby,

$$\mathcal{T}_{1;b}^* \leq \left\{\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \int_0^1 \int_0^1 |D_{\boldsymbol{\theta}} h(u, v, \boldsymbol{\theta})|\, dudv\right\} \frac{E\left[|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0|\right]}{\eta_T^{(1)}} \int_0^1 \epsilon \left(\log 1/\epsilon\right)^p N(\epsilon, \mathfrak{C}_T^*) d\epsilon.$$

Under Assumptions (B2) and (C), we have that $\mathcal{T}_{1;b}^* \longrightarrow 0$.

Next, for the term $\mathcal{T}_2^*$, an algebraic manipulation yields

$$\mathcal{T}_2^* \leq \mathbb{I}\left(\sup_{\substack{\{c,\Pi_T\boldsymbol{\Lambda}\}\in\{W_1^r([0,1]^2), B_{\lambda^*}\times B_{\gamma^*}\times B_{\lambda}^{N_T}\times B_{\gamma}^{N_T}\} \\ \|c-c_0\|_{\mathfrak{C}}+\|\boldsymbol{\Lambda}-\boldsymbol{\Lambda}_0\|_{EC}\geq D\epsilon_T^{(1)}}} \mathcal{L}_T(c, \Pi_T\boldsymbol{\Lambda}, \boldsymbol{\theta}_0) \geq \mathcal{L}_T(\Pi_T c_0, \Pi_T \Lambda_0, \boldsymbol{\theta}_0) - 2\eta_T^{(1)}\right),$$

where $\|\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0\|_{EC} \doteq \sqrt{|\lambda^* - \lambda_0|^2 + |\gamma^* - \gamma_0|^2} + \|\lambda - \lambda_0\|_{C^*} + \|\gamma - \gamma_0\|_{C^*}$ with a pseudo-metric, $\| \bullet \|_{C^*}$, on $C^*([0, 1])$. Then,

$$\mathcal{T}_2^* \leq \mathbb{I}\left(\sup_{\substack{\{c,\Pi_T\boldsymbol{\Lambda}\}\in\{W_1^r([0,1]^2), B_{\lambda^*}\times B_{\gamma^*}\times B_{\lambda}^{N_T}\times B_{\gamma}^{N_T}\} \\ \|c-c_0\|_{\mathfrak{C}}+\|\boldsymbol{\Lambda}-\boldsymbol{\Lambda}_0\|_{EC}\geq D\epsilon_T^{(1)}}} \Big\{\mathcal{L}_T(c, \Pi_T\boldsymbol{\Lambda}, \boldsymbol{\theta}_0) - \mathcal{L}_T(\Pi_T c_0, \Pi_T \Lambda_0, \boldsymbol{\theta}_0) - \mathcal{L}(c, \boldsymbol{\Lambda}, \boldsymbol{\theta}_0)\right.$$

$$\left. + \mathcal{L}(c_0, \boldsymbol{\Lambda}_0, \boldsymbol{\theta}_0)\Big\} \geq \inf_{\substack{\{c,\boldsymbol{\Lambda}\}\in\{W_1^r([0,1]^2), B_{\lambda^*}\times B_{\gamma^*}\times C^*([0,1])\times C^*([0,1])\} \\ \|c-c_0\|_{\mathfrak{C}}+\|\boldsymbol{\Lambda}-\boldsymbol{\Lambda}_0\|_{EC}\geq D\epsilon_T^{(1)}}} \Big\{\mathcal{L}(c_0, \boldsymbol{\Lambda}_0, \boldsymbol{\theta}_0) - \mathcal{L}(c, \boldsymbol{\Lambda}, \boldsymbol{\theta}_0)\Big\} - 2\eta_T^{(1)}\right).$$

In light of Assumption (B3), one can show that

$$\mathcal{T}_2^* \leq \mathbb{I}\left(\sup_{\substack{\{c,\Pi_T\boldsymbol{\Lambda}\}\in\{W_1^r([0,1]^2),B_{\lambda^*}\times B_{\gamma^*}\times B_\lambda^{N_T}\times B_\gamma^{N_T}\} \\ \|c-c_0\|_{\mathfrak{C}}+\|\boldsymbol{\Lambda}-\boldsymbol{\Lambda}_0\|_{EC}\geq D\epsilon_T^{(1)}}} \Big\{ \mathcal{L}_T(c,\Pi_T\boldsymbol{\Lambda},\boldsymbol{\theta}_0) - \mathcal{L}_T(\Pi_T c_0,\Pi_T\Lambda_0,\boldsymbol{\theta}_0) - \mathcal{L}(c,\boldsymbol{\Lambda},\boldsymbol{\theta}_0) \right.$$
$$\left. + \mathcal{L}(c_0,\boldsymbol{\Lambda}_0,\boldsymbol{\theta}_0)\Big\} \geq 2D_4{\epsilon_T^{(1)}}^{1/\omega}\right), \text{ where } D_4>0.$$

It then follows that

$$\begin{aligned}
\mathcal{T}_2^* \;\leq\; & \mathbb{I}\left(\sup_{\substack{\{c,\Pi_T\boldsymbol{\Lambda}\}\in\{W_1^r([0,1]^2),B_{\lambda^*}\times B_{\gamma^*}\times B_\lambda^{N_T}\times B_\gamma^{N_T}\} \\ \|c-c_0\|_{\mathfrak{C}}+\|\boldsymbol{\Lambda}-\boldsymbol{\Lambda}_0\|_{EC}\geq D\epsilon_T^{(1)}}} \left| \left\{ \int_{\frac{1}{N_T}}^1 \lambda(u)\left(E\left[\mathbb{I}(w\in\left[u-\frac{1}{N_T},u\right)\right)\right]-\frac{1}{N_T}\right) \right.\right. \\
- & \frac{1}{N_T}\sum_{k=1}^{N_T}\lambda_k\left\{E\left[\frac{1}{h_T}\int_{\frac{k-1}{N_T}}^{\frac{k}{N_T}} K\left(\frac{u-w}{h_T}\right)dw\right]-\frac{1}{N_T}\right\}\Bigg\} \\
+ & \left\{\int_{\frac{1}{N_T}}^1 \gamma(v)\left(E\left[\mathbb{I}(w\in\left[v-\frac{1}{N_T},v\right)\right)\right]-\frac{1}{N_T}\right)\right. \\
- & \frac{1}{N_T}\sum_{k=1}^{N_T}\gamma_k\left\{E\left[\frac{1}{h_T}\int_{\frac{k-1}{N_T}}^{\frac{k}{N_T}} K\left(\frac{v-w}{h_T}\right)dw\right]-\frac{1}{N_T}\right\}\Bigg\} \\
- & \left\{\int_{\frac{1}{N_T}}^1 \lambda_0(u)\left(E_{c_0}\left[\mathbb{I}(w\in[u-\frac{1}{N_T},u))\right]-\frac{1}{N_T}\right)\right. \\
- & \frac{1}{N_T}\sum_{k=1}^{N_T}\lambda_{0,k}\left\{E_{c_0}\left[\frac{1}{h_T}\int_{\frac{k-1}{N_T}}^{\frac{k}{N_T}} K\left(\frac{u-w}{h_T}\right)dw\right]-\frac{1}{N_T}\right\}\Bigg\} \\
- & \left\{\int_{\frac{1}{N_T}}^1 \gamma_0(v)\left(E_{c_0}\left[\mathbb{I}(w\in[v-\frac{1}{N_T},v))\right]-\frac{1}{N_T}\right)\right. \\
- & \left.\left.\frac{1}{N_T}\sum_{k=1}^{N_T}\gamma_{0,k}\left\{E_{c_0}\left[\frac{1}{h_T}\int_{\frac{k-1}{N_T}}^{\frac{k}{N_T}} K\left(\frac{v-w}{h_T}\right)dw\right]-\frac{1}{N_T}\right\}\Bigg\}\right| \geq 2D_4{\epsilon_T^{(1)}}^{1/\omega}\right) \\
= \; & \mathbb{I}\left(\sup_{\substack{\{c,\Pi_T\boldsymbol{\Lambda}\}\in\{W_1^r([0,1]^2),B_{\lambda^*}\times B_{\gamma^*}\times B_\lambda^{N_T}\times B_\gamma^{N_T}\} \\ \|c-c_0\|_{\mathfrak{C}}+\|\boldsymbol{\Lambda}-\boldsymbol{\Lambda}_0\|_{EC}\geq D\epsilon_T^{(1)}}} |\mathfrak{T}_1+\mathfrak{T}_2-\mathfrak{T}_3-\mathfrak{T}_4| \geq 2D_4{\epsilon_T^{(1)}}^{1/\omega}\right).
\end{aligned}$$

An application of Lemmas 6 and 7 yields

$$\begin{aligned}
\mathfrak{T}_1 &\leq A_1 N_T^{-r_1} + A_2 \sup_{v\in[0,1]} \left(\int_0^1 |D_u^r c(u,v)|\, du\right) \left(\int_{\mathbb{R}} |u|^r K(u) du\right) h_T^r, \\
\mathfrak{T}_2 &\leq A_3 N_T^{-r_2} + A_4 \sup_{u\in[0,1]} \left(\int_0^1 |D_v^r c(u,v)|\, dv\right) \left(\int_{\mathbb{R}} |v|^r K(v) dv\right) h_T^r, \\
\mathfrak{T}_3 &\leq A_5 N_T^{-r_1} + A_6 \sup_{v\in[0,1]} \left(\int_0^1 |D_u^r c_0(u,v)|\, du\right) \left(\int_{\mathbb{R}} |u|^r K(u) du\right) h_T^r, \\
\mathfrak{T}_4 &\leq A_7 N_T^{-r_2} + A_8 \sup_{u\in[0,1]} \left(\int_0^1 |D_v^r c_0(u,v)|\, dv\right) \left(\int_{\mathbb{R}} |v|^r K(v) dv\right) h_T^r,
\end{aligned}$$

where $A_1, \ldots, A_8$ are arbitrarily generic positive constants. In view of Assumptions (D1) and (D2), one can choose N_T and h_T such that $N_T \min\left(T^{-\frac{\ell}{r_1\omega}}, T^{-\frac{\ell}{r_2\omega}}\right) \to \infty$ and $h_T = o\left(T^{-\frac{\ell}{\omega r}}\right)$ respectively, then

$$\lim_{T\longrightarrow\infty} \mathbb{I}\left(\sup_{\substack{\{c,\Pi_T\mathbf{\Lambda}\}\in\{W_1^r([0,1]^2), B_{\lambda^*}\times B_{\gamma^*}\times B_\lambda^{N_T}\times B_\gamma^{N_T}\} \\ \|c-c_0\|_{\mathfrak{C}}+\|\mathbf{\Lambda}-\mathbf{\Lambda}_0\|_{EC}\geq D\epsilon_T^{(1)}}} |\mathfrak{T}_1+\mathfrak{T}_2-\mathfrak{T}_3-\mathfrak{T}_4| \geq 2D_4{\epsilon_T^{(1)}}^{1/\omega}\right) = 0.$$

This immediately implies that $\mathcal{T}_2^* \to 0$.

Step 2: We conjecture that the optimal rate of convergence for $\widehat{c}_T$ also depends on the rate of deviation, ω, [from the true MEC] of the Lagrangian function $\mathcal{L}$. First, we choose a sequence of positive constants, $\{\Omega_1, \ldots, \Omega_\infty\}$, such that $\sum_{i=1}^\infty \Omega_i = \omega$; and let $\epsilon_T^{(2)} = T^{-\ell(1+\Omega_1)}$ and $\eta_T^{(2)} = T^{-\ell^*}$ for some $\ell^* \in \left(\frac{\ell(1+\Omega_1)}{\omega}, \frac{1}{2}\right)$. Note that

$$\mathbb{P}\left(\|\widehat{c}_T - c_0\|_{\mathfrak{C}} \geq D\epsilon_T^{(2)}\right) = \mathbb{P}\left(\|\widehat{c}_T - c_0\|_{\mathfrak{C}} \geq D\epsilon_T^{(1)}\right) + \mathbb{P}\left(D\epsilon_T^{(2)} \leq \|\widehat{c}_T - c_0\|_{\mathfrak{C}} < D\epsilon_T^{(1)}\right).$$

We have shown in Step 1 that $\mathbb{P}\left(\|\widehat{c}_T - c_0\|_{\mathfrak{C}} \geq D\epsilon_T^{(1)}\right) \to 0$; thus, in this step, we shall show that

$$\mathbb{P}\left(D\epsilon_T^{(2)} \leq \|\widehat{c}_T - c_0\|_{\mathfrak{C}} < D\epsilon_T^{(1)}\right) \to 0.$$

$$\mathbb{P}\left(D\epsilon_T^{(2)} \leq \|\widehat{c}_T - c_0\|_{\mathfrak{C}} < D\epsilon_T^{(1)}\right) = \mathbb{P}\left(\sup_{\substack{\{c,\mathbf{\Lambda}_T\}\in\{W_1^r([0,1]^2),B_{\lambda^*}\times B_{\gamma^*}\times B_\lambda^{N_T}\times B_\gamma^{N_T}\} \\ \begin{cases} D\epsilon_T^{(2)} \leq \|c-c_0\|_{\mathfrak{C}} < D\epsilon_T^{(1)} \\ c\in\mathfrak{C}_T^* \end{cases}}} \mathcal{L}_T(c,\mathbf{\Lambda}_T,\widehat{\boldsymbol{\theta}}) \geq \mathcal{L}_T(\Pi_T c_0, \Pi_T\mathbf{\Lambda}_0, \widehat{\boldsymbol{\theta}})\right)$$

$$= \mathbb{P}\left(\mathcal{A}\right) = \mathbb{P}\left(\mathcal{A}\bigcap\mathcal{B}\bigcup\mathcal{A}\bigcap\mathcal{B}^c\right)$$

$$\leq \mathbb{P}\left(\mathcal{B}\right) + \mathbb{I}\left(\sup_{\substack{\{c,\mathbf{\Lambda}_T\}\in\{W_1^r([0,1]^2),B_{\lambda^*}\times B_{\gamma^*}\times B_\lambda^{N_T}\times B_\gamma^{N_T}\} \\ \begin{cases} D\epsilon_T^{(2)} \leq \|c-c_0\|_{\mathfrak{C}} < D\epsilon_T^{(1)} \\ c\in\mathfrak{C}_T^* \end{cases}}} \mathcal{L}_T(c,\mathbf{\Lambda}_T,\boldsymbol{\theta}_0) \geq \mathcal{L}_T(\Pi_T c_0, \Pi_T\mathbf{\Lambda}_0, \boldsymbol{\theta}_0) - 2\eta_T^{(2)}\right)$$

$$= \mathcal{T}_1^{**} + \mathcal{T}_2^{**},$$

where $\mathcal{B} \doteq \left\{\sup_{\substack{\{c,\mathbf{\Lambda}_T\}\in\{W_1^r([0,1]^2),B_{\lambda^*}\times B_{\gamma^*}\times B_\lambda^{N_T}\times B_\gamma^{N_T}\} \\ \begin{cases} D\epsilon_T^{(2)} \leq \|c-c_0\|_{\mathfrak{C}} < D\epsilon_T^{(1)} \\ c\in\mathfrak{C}_T^* \end{cases}}} \left|\mathcal{L}_T(c,\mathbf{\Lambda}_T,\widehat{\boldsymbol{\theta}}) - \mathcal{L}_T(\Pi_T c_0, \Pi_T\mathbf{\Lambda}_0, \widehat{\boldsymbol{\theta}})\right.\right.$
$\left.\left. + \mathcal{L}_T(\Pi_T c_0, \Pi_T\mathbf{\Lambda}_0, \boldsymbol{\theta}_0) - \mathcal{L}_T(c,\mathbf{\Lambda}_T,\boldsymbol{\theta}_0)\right| > \eta_T^{(2)}\right\}$. We now proceed to bound $\mathcal{T}_1^{**}$ and $\mathcal{T}_2^{**}$.

As for $\mathcal{T}_1^{**}$, by the mean-value theorem, we have

$$\mathcal{L}_T(c,\mathbf{\Lambda}_T,\widehat{\boldsymbol{\theta}}) - \mathcal{L}_T(c,\mathbf{\Lambda}_T,\boldsymbol{\theta}_0) = \overline{\gamma}^* E_c\left[D_\theta' h(U,V,\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\overline{\boldsymbol{\theta}}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})\right]$$

and

$$\mathcal{L}_T(\Pi_T c_0, \Pi_T\mathbf{\Lambda}_0, \widehat{\boldsymbol{\theta}}) - \mathcal{L}_T(\Pi_T c_0, \Pi_T\mathbf{\Lambda}_0, \boldsymbol{\theta}_0) = \overline{\gamma}^* E_{\Pi_T c_0}\left[D_\theta' h(U,V,\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\overline{\overline{\boldsymbol{\theta}}}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})\right].$$

In view of Assumptions (B1) and (B2), it immediately follows that, as the ball $\left\{c \in \mathfrak{C}^* : \ D\epsilon_T^{(2)} \leq \|c - c_0\|_{\mathfrak{C}} < D\epsilon_T^{(1)}\right\}$ shrinks to c_0, we have that $\mathcal{T}_1^{**} \to 0$.

As for $\mathcal{T}_2^{**}$, first we note that

$$\begin{aligned}
\mathcal{T}_2^{**} \ &\leq \ \mathbb{I}\Bigg(\sup_{\substack{\{c,\Pi_T\mathbf{\Lambda}\}\in\{W_1^r([0,1]^2),B_{\lambda^*}\times B_{\gamma^*}\times B_\lambda^{N_T}\times B_\gamma^{N_T}\} \\ \|c-c_0\|_{\mathfrak{C}}+\|\mathbf{\Lambda}-\mathbf{\Lambda}_0\|_{EC}\geq\epsilon_T(2)}} \mathcal{L}_T(c,\Pi_T\mathbf{\Lambda},\boldsymbol{\theta}_0) - \mathcal{L}_T(\Pi_T c_0,\Pi_T\mathbf{\Lambda}_0,\boldsymbol{\theta}_0) \\
&- \ \{\mathcal{L}(c,\mathbf{\Lambda},\boldsymbol{\theta}_0) - \mathcal{L}(c_0,\mathbf{\Lambda}_0,\boldsymbol{\theta}_0)\} \geq \inf_{\substack{\{c,\Pi_T\mathbf{\Lambda}\}\in\{W_1^r([0,1]^2),B_{\lambda^*}\times B_{\gamma^*}\times B_\lambda^{N_T}\times B_\gamma^{N_T}\} \\ \|c-c_0\|_{\mathfrak{C}}+\|\mathbf{\Lambda}-\mathbf{\Lambda}_0\|_{EC}\geq\epsilon_T^{(2)}}} \{\mathcal{L}(c_0,\mathbf{\Lambda}_0,\boldsymbol{\theta}_0) \\
&- \ \mathcal{L}(c,\mathbf{\Lambda},\boldsymbol{\theta}_0)\} - 2\eta_T^{(2)}\Bigg).
\end{aligned}$$

Next, by the same argument used to bound $\mathcal{T}_2^*$, we can show that $\mathcal{T}_2^{**} \to 0$.

Repeat Step 2 an infinite number of times, by setting $\epsilon_T^{(i)} = T^{-\ell(1+\Omega_1+\cdots+\Omega_i)}$ and $\eta_T^{(i)} = T^{-\ell^{*(i)}}$ for some $\ell^{*(i)} \in \left(\frac{\ell(1+\Omega_1+\cdots+\Omega_i)}{\omega}, \frac{1}{2}\right)$ in Step i, one can derive that $\|\widehat{c}_T - c_0\|_{\mathfrak{C}} = o\left(T^{-\omega\ell}\right)$. $\square$

APPENDIX D: ESTIMATION OF THE DYNAMIC MECC MODEL

In this section, we briefly present a numerical method based on empirical likelihoods to estimate the dynamic MECC model specified in (5.5)-(5.9). Basically, we need to estimate the set of parameters $\Xi = \{\theta_0, \ldots, \theta_{27}\}$. Let $\widehat{\Xi}$ denote an estimate of Ξ, our estimation procedure comprises of two steps described below.

Step 1: Let $MECC - likelihood(\Xi)$ denote the empirical log-likelihood function of Ξ, which can be computed in two sub-steps:

(a) Given a value, Ξ, in view of (5.6)-(5.9), we can compute measures of association, $\{\rho_{S_t}, \nu_{1,t}, \nu_{2,t}, \eta_t\}_{t=1}^{T}$.

(b) Given these measures of association, compute the time-varying parameters of the MECC ($\mathbf{\Lambda}_t \ \forall \ t = 1, \ldots, T$) by using the stochastic optimization algorithm to optimize the potential functions defined in Corollary 3.3. Hence, we can immediately obtain the empirical log-likelihood function: $MECC - likelihood(\Xi) = \frac{1}{T}\sum_{t=1}^{T} \log c(\widehat{G}_{1,t}, \widehat{G}_{2,t}, \mathbf{\Lambda}_t)$, where $c(u, v, \mathbf{\Lambda})$ is the MECC density.

Step 2: In the sequel, use the stochastic optimization algorithm to find a global maximum value, $\widehat{\Xi}$, for $MECC - likelihood(\Xi)$. This step requires the algorithm to go back to **Step 1** to recursively compute the MECC for each Ξ until a global maximum point is reached.

In order to reduce the computational time of finding the global maximum value of $MECC - likelihood(\Xi)$, we computed $\mathbf{\Lambda}_t$, for $t = 1, \ldots, T$, in **Step 1**(b) by using an efficient interpolation method proposed by Beliakov (2006). This method requires constructing a common efficient interpolant, $g(\rho_{S_t}, \nu_{1,t}, \nu_{2,t}, \eta_t)$, which approximates the mapping, $\mathbf{\Lambda}_t = f(\rho_{S_t}, \nu_{1,t}, \nu_{2,t}, \eta_t)$, from the domains of ρ_S, ν_1, ν_2, and η to the domains of $\mathbf{\Lambda}$; and it is assumed that this mapping satisfies the following Lipschitz condition with a Lipschitz constant, M:

$$|f(x) - f(z)| \leq Md(x, z)$$

for all x and z, where $d(x, z)$ is a distance function. Assume that the mapping $f(x)$ exists, then $g(x)$ provides the best uniform approximation to $f(x)$ in the worst case scenario – that is, $g(x)$ minimizes the maximum possible error given by

$$\max_{f} \max_{x \in \boldsymbol{X}} |f(x) - g(x)|, \text{ where } \boldsymbol{X} \text{ is the domain of } x.$$

In our empirical application, to construct $g(x)$, we first draw 50,000 sample values of measures of association, $\{\rho_{S_i}, \nu_{1,i}, \nu_{2,i}, \eta_i\}_{i=1}^{50,000}$, from their domains, then compute the corresponding values of the MECC parameters, $\{\Lambda_i\}_{i=1}^{50,000}$, by minimizing the potential functions. Next, given this sample of scattered data, $\{\rho_{S_i}, \nu_{1,i}, \nu_{2,i}, \eta_i, \Lambda_i\}_{i=1}^{50,000}$, we used the C++ interface developed by Beliakov (2006) to find $g(x)$. Thereby, we could substantially reduce our computational time by avoiding a rather time-consuming task of finding global optimal values (**Step 1**(b)) because the efficient interpolant $g(x)$ facilitates prompt computations of $\boldsymbol{\Lambda}$ from $\{\rho_S, \nu_1, \nu_2, \eta\}$.

Table 1: *IMSE* for the MECC & parametric copulas

	$\theta = 0.1$			$\theta = 0.8$		
Copula	Int. Bias^2	Int. Var.	IMSE	Int. Bias^2	Int. Var.	IMSE
Normal	27.9150	$1.0416{*}e^{-13}$	27.9150	28.3512	$1.0641{*}e^{-13}$	28.3512
Clayton	0.0002	$2.8924{*}e^{-15}$	0.0002	0.0176	$2.9086{*}e^{-15}$	0.0176
Gumbel	0.0002	$2.8745{*}e^{-15}$	0.0002	0.0203	$2.6789{*}e^{-15}$	0.0203
Student's t	1203.8600	$2.9161{*}e^{-12}$	1203.8600	955.8010	$3.3985{*}e^{-12}$	955.8010
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(N, M) denotes the MECC estimated by using N marginal constraints and M moment constraints. All the figures are rounded to four decimal places.

Table 2: Descriptive statistics

	Full Sample		In-Sample		Out-of-Sample	
	Russell 1000 Growth	Russell 1000 Value	Russell 1000 Growth	Russell 1000 Value	Russell 1000 Growth	Russell 1000 Value
Mean	0.0051	0.0072	0.0204	0.0162	-0.0041	0.0017
Std. Dev.	0.0497	0.0382	0.0435	0.0371	0.0511	0.0380
Skewness	-0.4504	-0.8437	0.1295	-0.6131	-0.5721	-1.0374
25% quantile	-0.0187	-0.0141	-0.0043	0.0023	-0.0267	-0.0191
75% quantile	0.0371	0.0302	0.0420	0.0409	0.0282	0.0244
Kurtosis	1.4665	2.1669	1.4216	1.1533	1.2462	2.8422
Min	-0.1514	-0.1529	-0.0828	-0.0802	-0.1514	-0.1529
Max	0.1580	0.1103	0.1580	0.1103	0.1211	0.0761
Kolmogorov-Smirnov p-value	0.0802	0.0972	0.1027	0.1480	0.0999	0.0776
Correlation	0.8091		0.8866		0.7621	

Note: "Kolmogorov-Smirnov" refers to the test for normality of the unconditional return distributions. The full-sample period runs from June 1995 to July 2006, the in-sample period from June 1995 to September 1999, and the out-of-sample period from October 1999 to July 2006.

Table 3: Results from the out-of-sample copula fluctuation test

Model	Average $p-$value	Minimum $p-$value	Maximum $p-$value
Clayton-Gumbel	0.5303	0.0901	0.8889
Clayton-MECC	0.5729	0.0000	1.0000
Gumbel-MECC	0.5684	0.0000	1.0000

Table 4: Results from the copula specification search using the information criteria

Model	Average log-likelihood	Number of parameters	Average **AIC**	Average **cAIC**	Average **BIC**	Average **HQIC**
Normal	1.8379	7	-6.5136	12.3242	-5.9660	-6.3227
Clayton	1.3092	7	-5.4564	13.3815	-4.9087	-5.2654
Gumbel	1.3586	7	-5.5550	13.2829	-5.0073	-5.3640
MECC	5.7761	28	-9.7143	-261.5523	-11.4527	-9.6692

Table 5: Descriptive statistics of realized portfolio returns

	Unconstrained				**Short-sales constrained**			
	Normal	Clayton	Gumbel	MECC	Normal	Clayton	Gumbel	MECC
				RRA = 3				
Mean	0.9806	0.5646	1.1362	0.8655	1.0078	1.0078	1.0095	1.0077
Std. Dev.	0.3538	2.8754	0.8711	1.4105	0.0393	0.0280	0.0376	0.0477
Sharpe Ratio	2.7717	0.1963	1.3044	0.6136	25.6147	35.9540	26.8645	21.1357
Skewness	1.1703	-8.5290	1.6482	-8.3289	-0.1321	1.1445	0.1172	1.4038
Kurtosis	8.2222	75.4105	17.6369	73.6448	0.2316	4.0581	1.1847	5.6840
10% quantile	0.5652	0.4243	0.8247	0.8107	0.9609	0.9893	0.9732	0.9579
90% quantile	1.2675	1.1819	1.6606	1.2372	1.0567	1.0421	1.0626	1.0548
				RRA = 7				
Mean	0.9772	0.9440	1.1070	0.9154	1.0102	1.0078	1.0070	1.0097
Std. Dev.	0.2862	0.3453	0.4926	1.1942	0.0496	0.0355	0.0420	0.0509
Sharpe Ratio	3.4139	2.7337	2.2474	0.7666	20.3604	28.4240	23.9917	19.8200
Skewness	-0.4918	-0.5886	3.3726	-7.3603	-0.3709	0.5856	-0.0857	1.7248
Kurtosis	3.0483	2.8632	18.5277	64.0690	0.4386	3.7046	0.2711	7.1301
10% quantile	0.6803	0.5352	0.7202	0.8321	0.9561	0.9807	0.9491	0.9605
90% quantile	1.2205	1.2638	1.3313	1.1283	1.0781	1.0481	1.0662	1.0635
				RRA = 15				
Mean	0.9867	0.9987	1.0182	1.0142	1.0091	1.0094	1.0098	1.0081
Std. Dev.	0.0914	0.1510	0.2261	0.7752	0.0483	0.0416	0.0448	0.0520
Sharpe Ratio	10.7918	6.6140	4.5028	1.3083	20.9019	24.2914	22.5284	19.3823
Skewness	-0.3631	-1.6466	0.5243	1.5470	-0.5260	0.1593	-0.2293	1.8270
Kurtosis	1.6121	15.4838	5.0600	26.9495	0.8780	0.1635	1.0685	8.6756
10% quantile	0.8681	0.8918	0.8404	0.8794	0.9540	0.9573	0.9670	0.9533
90% quantile	1.0886	1.1398	1.1534	1.1394	1.0665	1.0667	1.0629	1.0596
				RRA = 25				
Mean	0.9963	0.9886	1.0115	0.9696	1.0109	1.0109	1.0083	1.0060
Std. Dev.	0.0794	0.0752	0.1230	0.5810	0.0431	0.0419	0.0412	0.0513
Sharpe Ratio	12.5485	13.1501	8.2236	1.6688	23.4544	24.1502	24.5012	19.6174
Skewness	-2.4794	-0.3464	1.4049	0.0377	-0.4331	-0.1296	-0.3046	1.7097
Kurtosis	12.7335	1.7569	10.3033	12.0053	0.8359	1.0711	1.3174	10.5269
10% quantile	0.9133	0.8951	0.9029	0.7356	0.9671	0.9687	0.9626	0.9466
90% quantile	1.0789	1.0742	1.0799	1.1578	1.0655	1.0613	1.0598	1.0561

Table 6: Descriptive statistics of optimal portfolios: *Normal copula*

	RRA = 3		RRA =7		RRA = 15		RRA = 25	
	Growth	Value	Growth	Value	Growth	Value	Growth	Value
			Unconstrained					
Mean	0.2969	-1.0153	-0.6238	1.0488	0.3463	1.0581	-0.2063	1.0704
Min	-15.6632	-22.8090	-15.3314	-10.1389	-4.3280	-3.8435	-2.4541	-2.0194
25% quantile	-2.5726	-3.7589	-3.5001	-2.3879	-0.9182	-0.9342	-0.8802	-0.7516
Median	0.7327	-0.2068	-0.9591	1.0881	0.8061	1.6245	-0.6423	1.6719
75% quantile	3.1520	3.0401	3.0608	3.3116	1.0077	2.9448	0.7017	2.8729
Max	16.0126	12.7075	7.2890	31.4231	6.3339	3.7937	2.3882	3.0953
			Short-sales constrained					
Mean	0.4714	0.4047	0.4256	0.5035	0.5250	0.4107	0.4708	0.4761
Min	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0017
25% quantile	0.0884	0.0660	0.0341	0.0633	0.0676	0.0495	0.1523	0.1466
Median	0.6595	0.2626	0.4177	0.3040	0.6870	0.4006	0.4300	0.4916
75% quantile	0.7537	0.7364	0.8461	0.8943	0.8732	0.8429	0.7961	0.7501
Max	0.9837	0.9919	0.9742	0.9963	0.9757	0.9562	0.9684	0.9443

Note: This table presents some summary statistics of the optimal portfolio weights over the out-of-sample period for an investor, using the Normal copula, with the relative risk aversion degrees of 3, 7, 15, and 25, with/without a short-sales constraint.

Table 7: Descriptive statistics of optimal portfolios: *MECC*

	RRA = 3		RRA =7		RRA = 15		RRA = 25	
	Growth	Value	Growth	Value	Growth	Value	Growth	Value
				Unconstrained				
Mean	-0.0431	-3.2130	-0.5601	-4.5784	-1.1203	2.0668	1.0984	0.8952
Min	-56.6288	-234.1930	-24.3194	-426.3080	-42.1233	-37.1485	-25.2172	-39.5055
25% quantile	-0.9662	-0.8578	-0.9711	-0.8836	-0.9515	-0.8835	-0.8620	-0.8993
Median	0.0000	0.1789	-0.5256	0.2652	-0.1845	2.3832	0.1948	3.2479
75% quantile	0.9644	0.9176	0.8800	0.9579	0.9070	4.8914	0.9333	5.7697
Max	70.1992	8.5971	27.4618	18.7443	15.8168	28.5662	25.4485	113.9150
				Short-sales constrained				
Mean	0.4176	0.4258	0.4099	0.4821	0.4932	0.4610	1.0984	0.8952
Min	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-25.2172	-39.5055
25% quantile	0.0001	0.0227	0.0243	0.0543	0.0367	0.0884	-0.8620	-0.8993
Median	0.2854	0.2958	0.1989	0.4689	0.6620	0.4347	0.1948	1.2479
75% quantile	0.8659	0.8283	0.8885	0.9303	0.8909	0.9023	0.9333	8.7697
Max	0.9987	0.9928	0.9949	0.9975	0.9986	0.9921	25.4485	113.9150

Note: This table presents some summary statistics of the optimal portfolio weights over the out-of-sample period for an investor, using the MECC, with the relative risk aversion degrees of 3, 7, 15, and 25, with/without a short-sales constraint.

Table 8: Pairwise comparisons of models' performance

	RRA							
	3		7		15		25	
	p-value*	p-value**	p-value*	p-value**	p-value*	p-value**	p-value*	p-value**
				Unconstrained				
Clayton vs. Normal	0.0000	1^2	0.1438	0.8561^0	0.0173	0.9826^2	0.0296	0.9703^2
Gumbel vs. Normal	0.0000	1^2	0.1152	0.8847^0	0.0222	0.9777^2	0.0650	0.9349^2
MECC vs. Normal	1^1	0.0000	1^1	0.0000	1^1	0.0000	1^1	0.0000
Gumbel vs. Clayton	0.1542	0.8457^0	0.6789^0	0.3210	0.7445^0	0.2554	0.7347^0	0.2652
MECC vs. Clayton	1^1	0.0000	1^1	0.0000	1^1	0.0000	1^1	0.0000
MECC vs. Gumbel	1^1	0.0000	1^1	0.0000	1^1	0.0000	1^1	0.0000
				Short-sales constrained				
Clayton vs. Normal	0.0000	1^2	0.0104	0.9895^2	0.0008	0.9992^2	0.0108	0.9891^2
Gumbel vs. Normal	0.0360	0.963996^2	0.0036	0.9964^2	0.0015	0.9985^2	0.0281	0.9718^2
MECC vs. Normal	0.9763^1	0.0236	1^1	0.0000	1^1	0.0000	1^1	0.0000
Gumbel vs. Clayton	1^1	0.0000	0.9118^1	0.0881	0.9353^1	0.0646	0.8834^0	0.1165
MECC vs. Clayton	0.9843^1	0.0156	1^1	0.0000	1^1	0.0000	1^1	0.0000
MECC vs. Gumbel	0.9822^1	0.0177	1^1	0.0000	1^1	0.0000	1^1	0.0000

Note: This table presents the results of pairwise comparisons of the optimal portfolios based on the Normal copula, the Clayton copula, the Gumbel copula, and the MECC. Let the performance measure of portfolio i be μ_i. The p-value* denotes the bootstrap probability that $\mu_i - \mu_j$ is greater than zero. The p-value** denotes the bootstrap probability that $\mu_i - \mu_j$ is less than zero. The tests were conducted at the 10% significance level. The subscript, "0", indicates that the test was inconclusive. The subscript, "1", indicates that the first model outperforms the second one. The subscript, "2", indicates that the second model outperforms the first one. The performance measure being used is the sample mean of realized utilities.

Table 9: Results obtained from White's reality check

	RRA			
	3	7	15	25
		Unconstrained		
Bootstrapping reality check p-value	0.09161	0.09671	0.09651	0.95720
Nave p -value	0.09161	0.09671	0.09651	0.25883
		Short-sales constrained		
Bootstrapping reality check p-value	0.09821	0.96740	0.97870	0.96260
Nave p -value	0.09821	0.26353	0.11841	0.25933

Note: The "bootstrap reality check p-value" is the one corresponding to the best model with respect to the Normal copula model. The "naive p-value" is the bootstrap reality check p-value computed by treating the best model as if it were the only model considered.

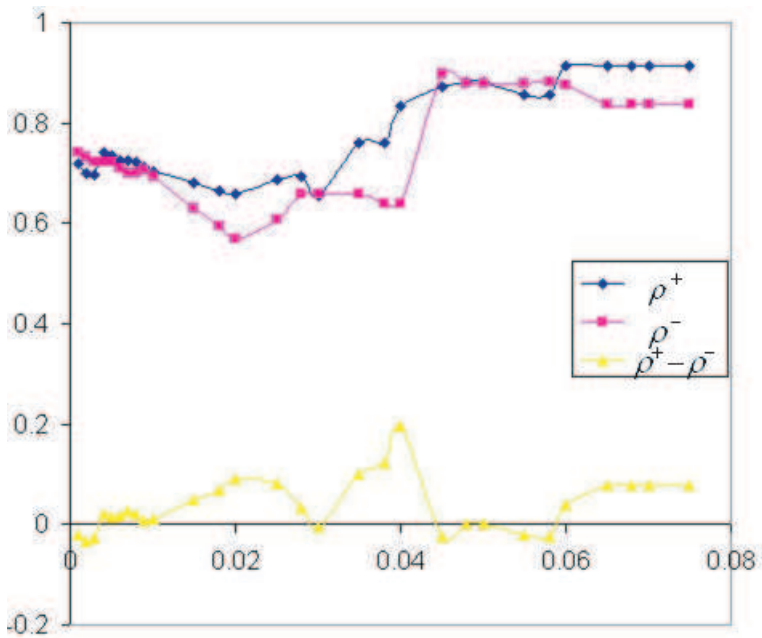
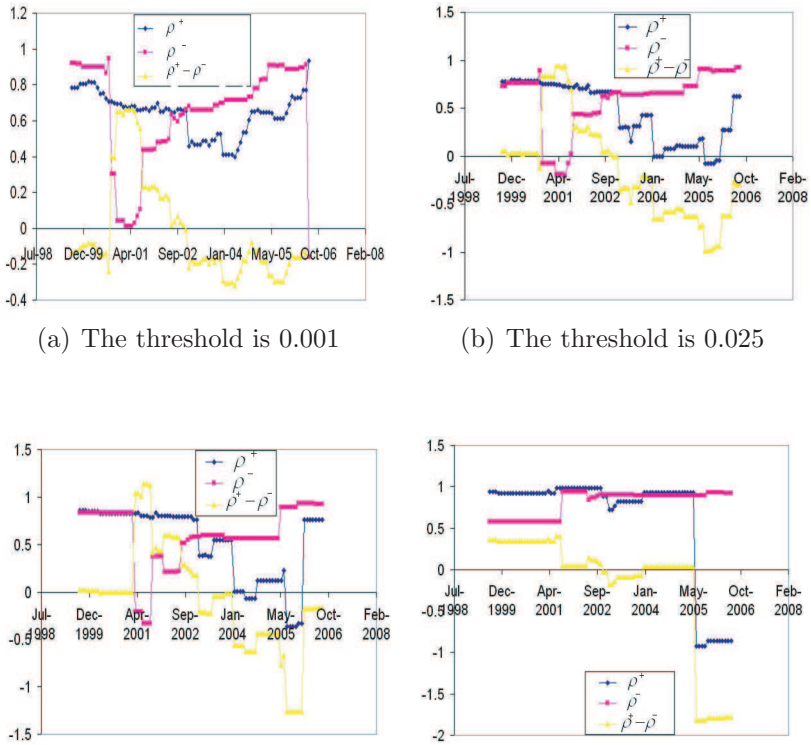


Figure 1: Asymmetric correlation between the excess returns on Russell 1000 Growth and Value. The horizontal axis shows cutoff quantiles, the vertical axis shows *exceedance* correlations between these returns, and ρ^+ and ρ^- represent the positive and negative *exceedance* correlations, respectively.



(a) The threshold is 0.001

(b) The threshold is 0.025

(c) The threshold is 0.038

(d) The threshold is 0.05

Figure 2: Moving-window asymmetric correlation between the excess returns on Russell 1000 Growth and Value. The horizontal axis shows time periods, the vertical axis shows *exceedance* correlations between these returns, and ρ^+ and ρ^- denote the positive and negative *exceedance* correlations, respectively.

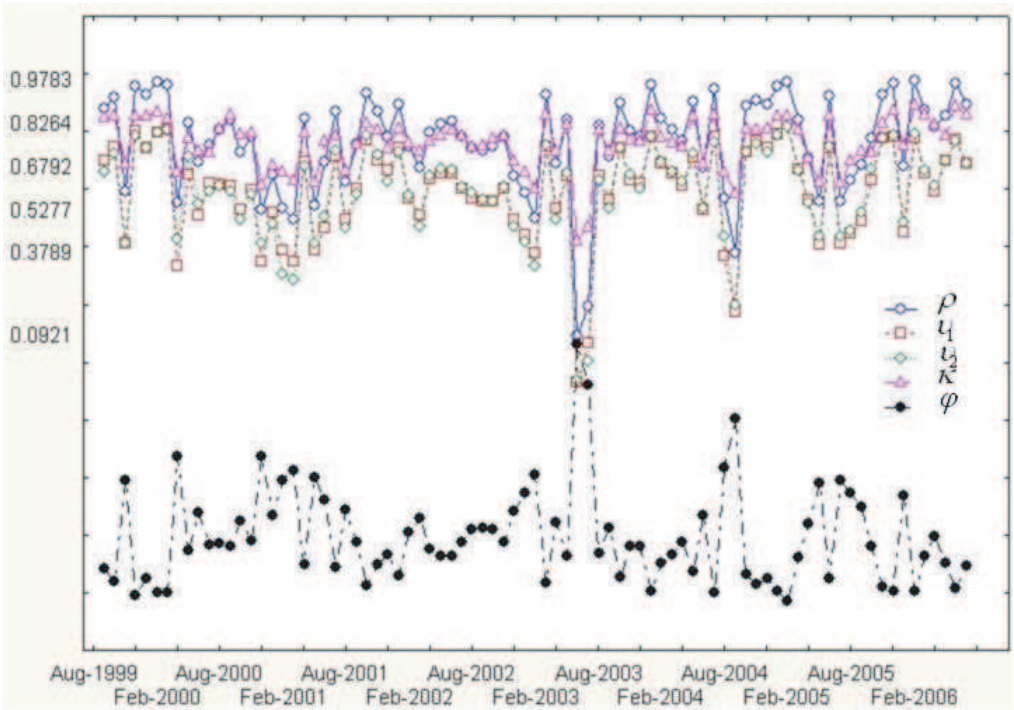


Figure 3: Moving-window rank correlations in the out-of-sample period (October 1999 to July 2006). ρ is the Spearman's rho, ν_1, ν_2, κ, and ϕ are, as defined in Section 2, the second, third, fourth, and fifth rank correlations respectively. The horizontal axis shows the out-of-sample period, and the vertical axis shows rank correlations.

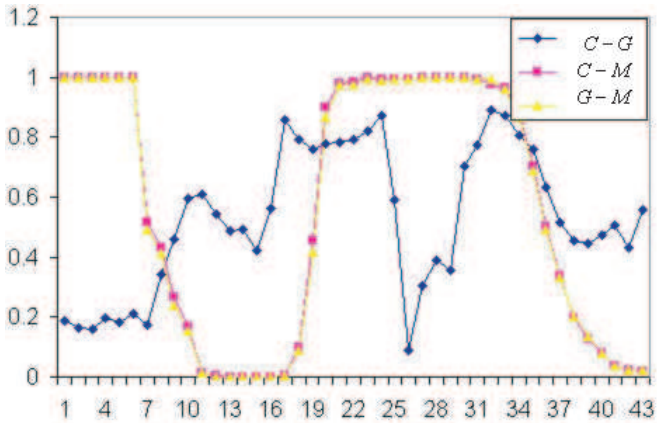


Figure 4: The p-values of the out-of-sample fluctuation test for equal performance of two copulas over time. The horizontal axis shows the orders of out-of-sample data and the vertical axis shows the p-values. Let "C", "G" and "M" denote the Clayton copula, the Gumbel copula and the MECC, respectively.

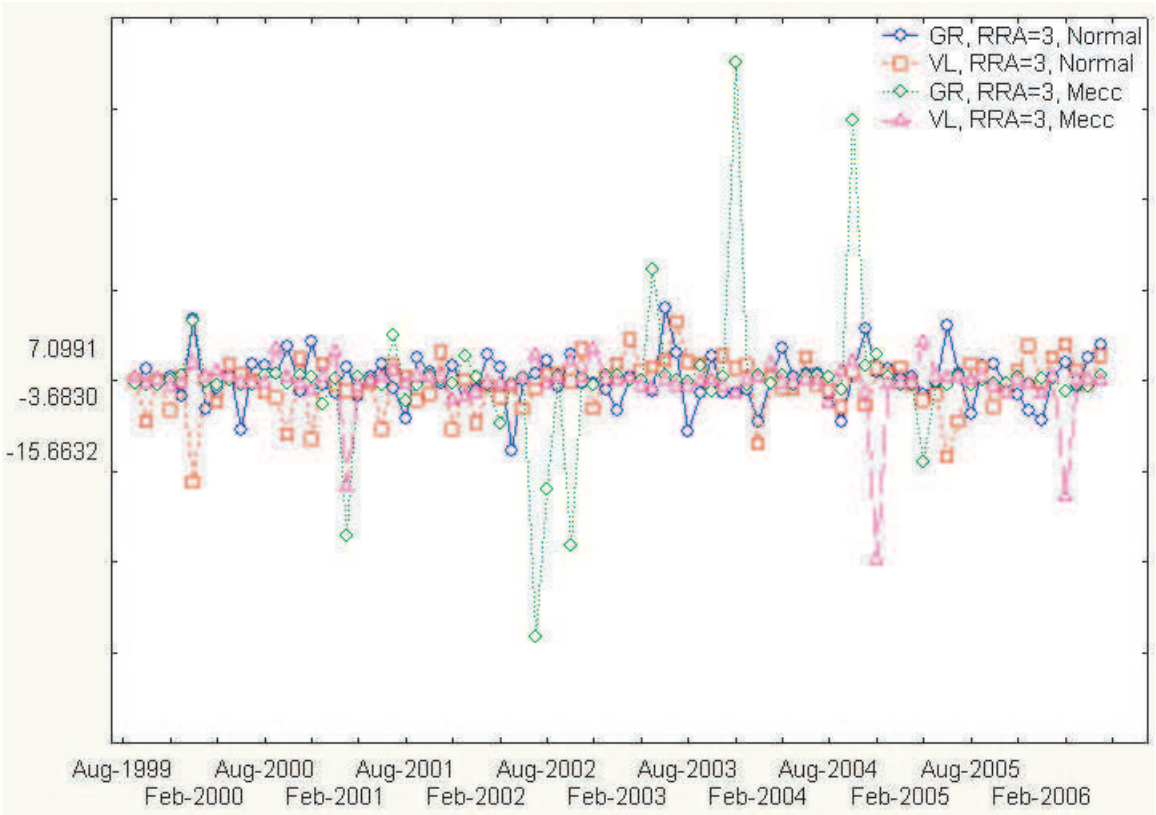


Figure 5: Optimal *unconstrained* Normal and MECC portfolios weights for an investor with the degree of relative risk aversion equal to 3 over the out-of-sample period (October 1999 to July 2006). Note that "GR" stands for the weight put in Russell 1000 Growth and "VL" stands for the weight put in Russell 1000 Value. The horizontal axis shows the out-of-sample period, and the vertical axis shows optimal portfolio weights.

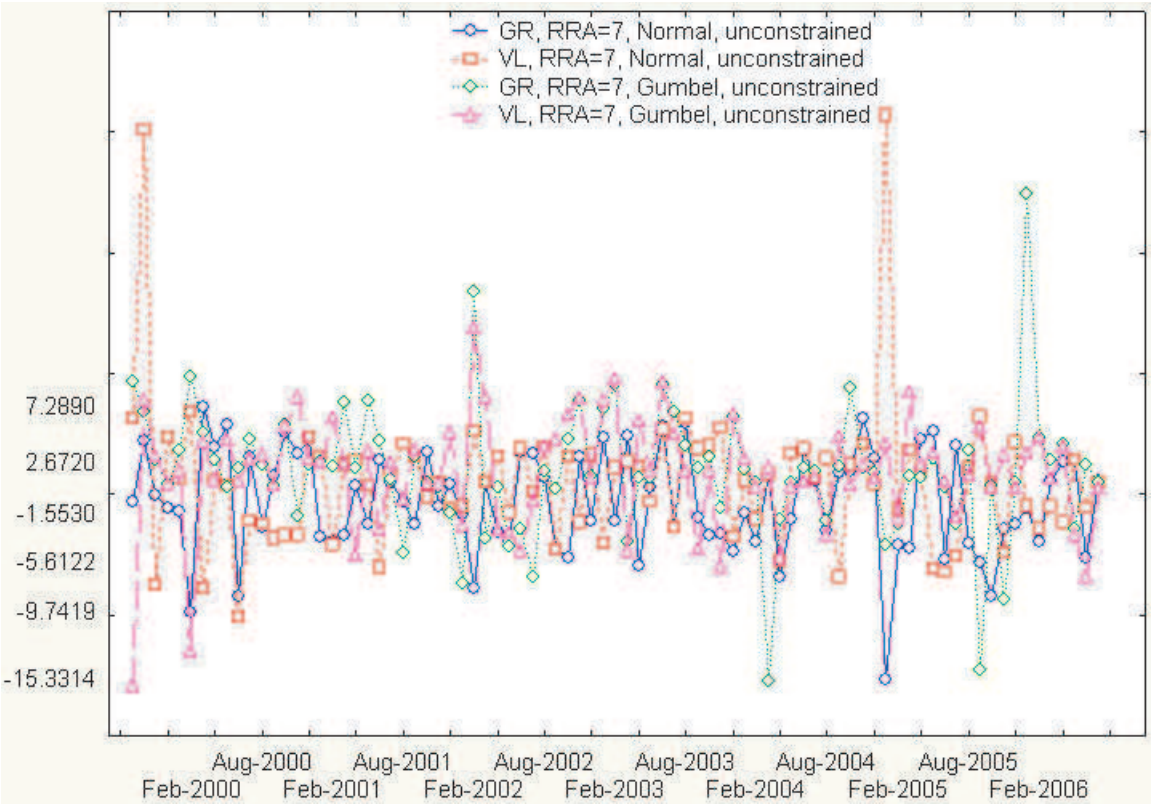


Figure 6: Optimal *unconstrained* Normal and Gumbel portfolios weights for an investor with the degree of relative risk aversion equal to 7 over the out-of-sample period (October 1999 to July 2006). Note that "GR" stands for the weight put in Russell 1000 Growth and "VL" stands for the weight put in Russell 1000 Value. The horizontal axis shows the out-of-sample period, and the vertical axis shows optimal portfolio weights.