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 (2003), 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, \( \hat{c}(u, v) \), 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 MacKinnon (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). 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. 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 includes brief descriptions of entropy and copula. Section defines the MECs and the MECC; and also proposes its approximator with asymptotic theory. Some simulations and an empirical application are presented in Sections and , confirming that our approach is feasible. Section 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)):

\[
\min_{f} KLCE(f : g) = \min_{f} \int_{\mathbb{R}^2} f(X, Y) \log \frac{f(X, Y)}{g_1(X)g_2(Y)} dX dY
\]

subject to

\[
\int_{\mathbb{R}^2} h(X, Y) f(X, Y) dX dY = \mu_0,
\]

where \(f\) is a bivariate density, \(g_1\) and \(g_2\) are some univariate densities, 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 para-
metrically 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, \( \tau \) can be estimated by the rank statistic

\[
\hat{\tau} = \frac{1}{T} \sum_{t=1}^{T} h\left(\frac{R_t}{T}, \frac{S_t}{T}\right),
\]

where \( T \) denotes the rank of \( (X_t, Y_t) \) in the sample of size \( T \). A proof 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:

\[
\hat{\tau} = \frac{1}{T} \sum_{t=1}^{T} J\left(\frac{R_t}{T}\right)J\left(\frac{S_t}{T}\right),
\]

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 \( \hat{\tau} = \frac{12}{T(T^2 - 1)} \sum_{t=1}^{T} \left( R_t - \frac{T+1}{2} \right) \left( S_t - \frac{T+1}{2} \right) \) 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 \( \hat{\tau} \). For instance, the statistic \( \hat{R}_g \), 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

<table>
<thead>
<tr>
<th>Spearman’s rho: $\rho_S = 12 \int_{[0,1]^2} uv c(u, v) dudv - 3$, and $\rho_S \in [-1, 1]$</th>
<th>Rank correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blest’s measure $I$: $\nu_1 = 2 - 12 \int_{[0,1]^2} (1-u)^2 v c(u, v) dudv$, and $\nu_1 \in [-1, 1]$</td>
<td>$\hat{\rho}<em>S = \frac{12}{T^3 - T} \sum</em>{t=1}^{T} R_t S_t - \frac{3T+1}{T-1}$</td>
</tr>
<tr>
<td>Blest’s measure $II$: $\nu_2 = 2 - 12 \int_{[0,1]^2} u(1-v)^2 c(u, v) dudv$, and $\nu_2 \in [-1, 1]$</td>
<td>$\hat{\nu}<em>1 = \frac{2T+1}{T-1} - \frac{12}{T^2 - T} \sum</em>{t=1}^{T} (1 - \frac{R_t}{T+1})^2 S_t$</td>
</tr>
<tr>
<td>Blest’s measure $III$: $\eta = 6 \int_{[0,1]^2} u^2 v^2 c(u, v) dudv - \frac{1}{5}$, and $\eta \in [0, 1]$</td>
<td>$\hat{\nu}<em>2 = \frac{2T+1}{T-1} - \frac{12}{T^2 - T} \sum</em>{t=1}^{T} R_t \left(1 - \frac{S_t}{T+1}\right)^2$</td>
</tr>
<tr>
<td>Blest’s measure $IV$: $\phi = \int_{[0,1]^2} [10(1-u)^3 v - 3u^2 v^2] c(u, v) dudv - 9/10$, and $\phi \in [-1, 1]$</td>
<td>$\hat{\phi} = \frac{1}{T^2 - T} \sum_{t=1}^{T} \left[\left(1 - \frac{R_t}{T+1}\right)^3 \frac{S_t}{T+1} - \left(\frac{R_t}{T+1}\right)^2 \frac{S_t}{T+1}\right] - \frac{0.9T+1}{T-1}$</td>
</tr>
</tbody>
</table>

has the form:

\[
\hat{R}_g = \frac{1}{[T/2]} \left( \max_{t} \sum_{s=1}^{t} 1(p_s < T + 1 - t) - \max_{t} \sum_{s=1}^{T} 1(R_s \leq t < S_s) \right),
\]

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, $\hat{R}_g$ estimates a copula-based measure of dependence, $R_g = 2 \int_{[0,1]^2} [\sup_{w \in [0,1]} 1(u \leq w, v < 1 - w) - \sup_{w \in [0,1]} (1(u \leq w) - 1(u \leq w, v < w))] c(u, v) dudv$.

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

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

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

\[
W(c) = -KLCE(f : g).
\]

Hence, minimization of the KLCE and maximization of the bivariate Shannon entropy are dual problems. Let $\hat{c}(u, v)$ denote the MEC. Then, in view of [1989], the relative entropy measure of dependence [recovered from limited information] is given by $-W(\hat{c})$. 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 \( H \) 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.

**Problem A:**

Maximizing \( W(c) = -\int_{[0,1]^2} c(u, v) \log c(u, v) du dv \) \( (3.1) \)

subject to

\[
\begin{align*}
\int_{[0,1]^2} c(u, v) du dv &= 1, \\
\int_{[0,1]} \int_{[0,1]} c(x, v) dx dv &= u, \ \forall \ u \in [0,1], \\
\int_{[0,1]} \int_{(0,v]} c(u, y) du dy &= v, \ \forall \ v \in [0,1], \\
\int_{[0,1]^2} h(u, v; \hat{\theta}_T)c(u, v) du dv &= 0,
\end{align*}
\]

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 \( h(u, v; \hat{\theta}_T) = 12uv - 3 - \hat{\rho}_S \), then the left-hand side of \( (3.5) \) becomes Spearman’s rho, as defined in Section 2, and \( \hat{\theta}_T = \hat{\rho}_S \) (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, \( C_1(u, v; \beta) \), from \( \mathcal{C}_0 \), then use it to construct \( (3.5) \) with \( h(u, v; \hat{\theta}) = 4C_1(u, v; \beta) - 1 - \hat{\tau} \), where \( \hat{\theta} = \{\hat{\beta}, \hat{\tau}\}' \) and \( \hat{\tau} \) is an estimate of the difference between the probabilities of concordance.
and discordance (cf. Appendix A). By doing this way, it is expected that some feature of the family $C_0$ could be effectively incorporated into the MECC. Other examples of Eq. (3.3) also include Blest’s coefficients or Gideon and Hollister’s (1987) coefficient, etc. Also note that we may have more than one constraint like (3.3).

For future references, we shall denote by $\tilde{c}(u, v) = c(u, v, \Lambda)$, where $\Lambda$ 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, $\tilde{c}(u, v)$, can be approximated by an approximator, $\tilde{c}_{n,N_h}(u, v)$, as follows:

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

with

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

where

$$\mathcal{E}_{n,N_h}(u, v) = \exp \left\{ - \sum_{k=0}^{2^n-1} \left[ \tilde{\lambda}_k \left( \Phi(N_h(k2^{-n} - u)) + \Phi(-N_h((k+1)2^{-n} - u)) \right) + \tilde{\gamma}_k \left( \Phi(N_h(k2^{-n} - v)) + \Phi(-N_h((k+1)2^{-n} - v)) \right) \right] - \tilde{\Lambda}_n h(u, v, \tilde{\theta}) - b_0 \tilde{c}(u, v) \right\}, \quad (3.7)$$

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

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

for a given $b_0$ and $\tilde{c}(u, v)$. (3.8)

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

In particular, the MEC, \( \tilde{c}(u, v) \), can be symmetrized by letting \( \lambda_k \) be equal to \( \gamma_k \) (\( \forall k = 1, \ldots, (2^n - 1) \)) and \( h(u, v, \tilde{\theta}) \) be a symmetric function.

Proof. The proof is presented in the appendix.

As we can see, the MEC density nests an arbitrary copula, \( \tilde{c}(u, v) \), (cf. Eq. (3.7)). Indeed, the MEC depends on both \( b_0 \) and \( \tilde{c}(u, v) \), thus no uniqueness is obtained. However, we can obtain a canonical form, which is called the MECC, by setting \( b_0 \) to zero. This idea of 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 \( Q_{n,N_h}(\Lambda, \tilde{\theta}) \) in the above theorem is a multivariate convex function of \( \Lambda \), thus in general has a unique minimum because it is the product of [positive] univariate convex functions.

We can claim that the MECC, \( \tilde{c}(u, v) \), 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

\[
\ell(\tilde{\Lambda}_n) = \frac{1}{T} \sum_{t=1}^{T} \log \tilde{c}_{n,N_h}(u_t, v_t, \tilde{\Lambda}_n)
\]

\[
= \frac{1}{T} \sum_{t=1}^{T} \log \mathcal{E}_{n,N_h}(u_t, v_t) - \log \int_{[0,1]^2} \mathcal{E}_{n,N_h}(u, v) dudv,
\]

where \( \tilde{c}_{n,N_h}(u_t, v_t) = \tilde{c}_{n,N_h}(u_t, v_t, \tilde{\Lambda}_n) \) is defined in (3.8),

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

and

\[ v_t = \frac{1}{T} \sum_{s=1}^{T} 1(Y_s \leq Y_t) = \frac{S_t}{T + 1}, \]

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 (3.4) with \( b_0 = 0 \), we obtain the following representation:

\[
\ell(\hat{\lambda}_n) = -\frac{1}{T} \sum_{t=1}^{T} \left( \hat{\lambda}_{-1} + \sum_{k=0}^{2^n-1} \left( \hat{\lambda}_k (\Phi(N_h(k2^{-n} - u_t)) + \Phi(-N_h((k + 1)2^{-n} - u_t))) + \hat{\gamma}_k \Phi(N_h(2^{-n} - v_t)) + \Phi(-N_h((k + 1)2^{-n} - v_t))) \right) \right)
\]

\[
\approx -\left( \hat{\lambda}_{-1} + \frac{2}{n} \sum_{k=0}^{2^n-1} (\hat{\lambda}_k + \hat{\gamma}_k) + \hat{\lambda}_{2^n} \frac{1}{T} \sum_{t=1}^{T} h(u_t, v_t) \right)
\]

\[
= -W(\hat{c}(u, v)),
\]

where \( \hat{\lambda}_{-1} = \log \int_{[0,1]^2} e_{n,h}(u,v) du dv; \) the approximation \((\approx)\) follows because \( \frac{1}{T} \sum_{t=1}^{T} (\Phi(N_h(k2^{-n} - u_t)) + \Phi(-N_h((k + 1)2^{-n} - u_t)) \approx \frac{1}{2^n} \) for every \( k = 0, \ldots, (2^n - 1); \) and the last equality holds because \( \int_{[0,1]^2} h(u,v, \hat{\theta}) c(u,v) du dv \) 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 (1983)) to minimize this function.

In general, we can also approximate \( \hat{c}(u,v) \) 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, \( \hat{c}(u,v) \), can be approximated by an approximator, \( \hat{c}_{N,h}(u,v) \), as follows:

\[
\hat{c}(u,v) = \lim_{N \to \infty} \hat{c}_{N,h}(u,v)
\]
with
\[
\tilde{c}_{N,h}(u, v) = \frac{E_{N,h}(u, v)}{\int_{[0,1]^2} E_{N,h}(u,v) du dv},
\]
where
\[
E_{N,h}(u,v) = \exp \left\{ -\frac{1}{N} \sum_{k=1}^{N} \left[ \hat{\lambda}_k \left( \frac{1}{h} \int_{k-1/N}^{k/N} K \left( \frac{u-w}{h} \right) dw \right) + \hat{\gamma}_k \left( \frac{1}{h} \int_{k-1/N}^{k/N} K \left( \frac{v-w}{h} \right) dw \right) \right] 
- \hat{\lambda}_{2^n} h(u,v,\hat{\theta}) - b_0 \tilde{c}(u,v) \right\}
\]
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 \( \hat{\Lambda}_N = \{ \hat{\lambda}_1, \ldots, \hat{\lambda}_N, \hat{\gamma}_1, \ldots, \hat{\gamma}_N \} \) contains the minimal values of the following potential function:
\[
Q_{N,h}(\Lambda_N, \hat{\theta}) = \int_{[0,1]^2} \exp \left\{ -\frac{1}{N} \sum_{k=1}^{N} \left[ \lambda_k \left( \frac{1}{h} \int_{k-1/N}^{k/N} K \left( \frac{u-w}{h} \right) dw - \frac{1}{N} \right) \right] 
+ \gamma_k \left( \frac{1}{h} \int_{k-1/N}^{k/N} K \left( \frac{v-w}{h} \right) dw - \frac{1}{N} \right) \right] 
- \lambda_{2^n} h(u,v,\hat{\theta}) - b_0 \tilde{c}(u,v) \right\} du dv
\]
for a given \( b_0 \) and \( \tilde{c}(u,v) \).

Proof. The proof is reminiscent of that of Theorem 3.1 combining with Lemma 4. So we shall omit its details here. \( \Box \)

### 3.2 Convergence Rates of the MECC

For the estimator \( \tilde{c}_{N,h}(u,v) \) 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 \( \tilde{c}_T(u,v) \) signify the dependence of \( N, h, \) and \( \tilde{c}_{N,h}(u,v) \), 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:
The space of the MECs:

\[ \mathcal{C} \doteq \left\{ c \in W_1^r([0,1]^2) : c \in (0,\infty), \int_0^1 \int_0^1 c(u,v)dvdu = 1; \int_0^1 c(u,v)du = 1; \int_0^1 c(u,v)dv = 1 \text{ for every } u, v \in [0,1] \right\}, \]

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:

\[ \mathcal{C} = \left\{ c \in W_1^r([0,1]^2) : c \in (0,\infty), \ E[1] = 1, \ \int_{\epsilon}^1 \lambda(u) \left\{ \int_{u-\epsilon}^u \int_{0}^1 c(w,v)dwdv - \epsilon \right\} = 0, \ \int_{\epsilon}^1 \gamma(v) \left\{ \int_{0}^v \int_{v-\epsilon}^c c(u,w)dudw - \epsilon \right\} = 0, \text{ and } E[h(U,V,\theta_0)] = 0, \text{ where } \lambda(u), \gamma(v) \in C^*(\{0,1\}) \right\}, \]

where \( \epsilon \) 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 \( c_0 = \text{arg sup}_{c \in \mathcal{C}} \{-E[\log c(U,V)]\} \).

- The approximation space of copulas:

\[ \mathcal{C}_T = \left\{ c \in W_1^r([0,1]^2) : c \in (0,\infty), \ E[1] = 1, \ \frac{1}{N_T} \sum_{k=1}^{N_T} \lambda_k \left\{ \int_{k-1/N_T}^{k/N_T} \int_0^1 c(w,v)dwdv - \frac{1}{N_T} \right\} = 0, \right\}, \]

and \( \frac{1}{N_T} \sum_{k=1}^{N_T} \gamma_k \left\{ \int_0^1 \int_{k-1/N_T}^{k/N_T} c(u,w)dudw - \frac{1}{N_T} \right\} = 0 \) with \( \lambda_k, \gamma_k \in B_\lambda \times B_\gamma \),

where the \( \lambda_k \)’s and the \( \gamma_k \)’s represent some values of \( \lambda(u) \) and \( \gamma(v) \) respectively on \( \left[\frac{k-1}{N_T}, \frac{k}{N_T}\right] \); and \( B_\lambda \) and \( B_\gamma \) are some bounded sets of \( \mathbb{R} \).

- The kernel-smoothed approximation space of copulas:

\[ \mathcal{C}_T^* = \left\{ c \in W_1^r([0,1]^2) : c \in (0,\infty), \ E[1] = 1, \ \frac{1}{N_T} \sum_{k=1}^{N_T} \lambda_k \left\{ \frac{1}{h_T} \int_{k-1/N_T}^{1/k/N_T} K\left( \frac{u-w}{h_T} \right) dw \right\} = 0, \right\}, \]

and \( \frac{1}{N_T} \sum_{k=1}^{N_T} \gamma_k \left\{ \frac{1}{h_T} \int_{k-1/N_T}^{1/k/N_T} K\left( \frac{v-w}{h_T} \right) dw \right\} = 0 \) with \( \lambda_k, \gamma_k \in B_\lambda \times B_\gamma \).
• The kernel-smoothed approximation space of the MECs:

\[ C^*_T(\hat{\theta}) = \left\{ c \in \mathcal{C}_T^* : E[h(U, V, \hat{\theta})] = 0 \right\}. \]

Theorem 3.3 (below) provides an optimal rate of convergence for \( \hat{\theta}_T(u, v) \) in terms of \( \| \hat{\theta}_T - c_0 \|_\mathcal{C} \), where \( \| \cdot \|_\mathcal{C} \) is a pseudo-metric on \( \mathcal{C} \). 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 5.3).

(A2) **Marginal-Weighted Functions:** \( \{ \lambda(u), \gamma(v) \} \in C^r([0, 1]) \times C^r([0, 1]) \) 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:** \( \sup_{\theta \in \Theta} \| E[|D_\theta^2 h(U, V, \theta)|] \|_E < \infty \), where the vector \( D_\theta^2 h(u, v, \theta) \) contains the first-order partial derivatives of \( h(u, v, \theta) \) with respect to \( \theta \); \( \Theta \) is a compact finite-dimensional set; and \( \| \bullet \|_E \) is the usual Euclidean norm.

(B2) \( \sqrt{T} \)-**Consistency of the Estimates of Parameters:** \( \hat{\theta}_T - \theta_0 = O_p \left( T^{-1/2} \right) \).

(B3) **Rate of Deviations:** \( \inf_{\{c, \Lambda\} \in \{W^1_{[0,1]^2}, B_{x*} \times B_{y*} \times C^*[0,1], C^*[0,1]\}} \{ \mathcal{L}(c_0, \Lambda_0, \theta_0) - \mathcal{L}(c, \Lambda, \theta_0) \} \geq 2D_2 \varepsilon^{1/(n - \omega)} \) – where \( D_1 \) and \( D_2 \) denote some generic positive constants, \( B_{x*} \) and \( B_{y*} \) are bounded sets in \( \mathbb{R} \), \( \| \Lambda - \Lambda_0 \|_{EC} \leq \sqrt{\| \Lambda^* - \Lambda_0 \|^2 + \| \gamma^* - \gamma_0 \|^2} + \| \lambda - \lambda_0 \|_{C^*} + \| \gamma - \gamma_0 \|_{C^*} \) 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:

\[
\mathcal{L}(c, \Lambda; \theta_0) = -E[\log c(U, V)] - \lambda^*(E[1] - 1) - \gamma^*E[h(U, V, \theta_0)] - \int_{1}^{1} \lambda(u) \{E[I(w \in [u - \epsilon, u])] - \epsilon \} du - \int_{1}^{1} \gamma(v) \{E[I(w \in [v - \epsilon, v])] - \epsilon \} dv,
\]
where $\Lambda = \{\lambda^*, \gamma^*, \lambda(u), \gamma(v)\}'$ and all these expectations are taken with respect to $c \in W_T^r([0, 1]^2)$.

**Assumption C:**

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

**Assumption D:**

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

(D2) Bandwidth: $h_T = o \left( T^{-\frac{\ell}{2-\alpha}} \right)$.

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

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

and

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

**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 \( \theta \) 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 \( \theta \); 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 \( \theta \). 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):

\[
\int_0^1 \int_0^1 \left\{ E \left[ \left( c(u, v; \theta) - \hat{c}_T(u, v) \right)^2 \right] \right\} \, du \, dv = \int_0^1 \int_0^1 \left| E \left[ \hat{c}_T(u, v) - c(u, v; \theta) \right] \right|^2 \, du \, dv
\]

\[ + \int_0^1 \int_0^1 E \left[ (\hat{c}_T(u, v) - E[\hat{c}_T(u, v)])^2 \right] \, du \, dv
\]

\[ = \text{Int. Bias}^2 + \text{Int. Var.}, \]

where \( c(u, v; \theta) \) is the density of Frank’s copula; and \( \hat{c}_T(u, v) \) 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 \( \text{Int. Bias}^2 \) and \( \text{Int. Var.} \) are then obtained by evaluating the estimated squared bias, \( \widehat{\text{Bias}}^2(u, v) = \left| \text{E} [\hat{c}_T(u, v) - c(u, v; \theta)] \right|^2 \) and the estimated variance, \( \widehat{\text{Var.}}(u, v) = \text{E} \left[ (\hat{c}_T(u, v) - E[\hat{c}_T(u, v)])^2 \right] \), at 10000 pseudo-random Uniform[0,1] points, then taking
their individual averages, i.e.,

\[
\text{Int. Bias}^2 \approx \frac{1}{10000} \sum_{i=1}^{10000} \text{Bias}^2(u_i, v_i),
\]

\[
\text{Int. Var.} \approx \frac{1}{10000} \sum_{i=1}^{10000} \text{Var.}(u_i, v_i),
\]

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 \(\tilde{c}_T(u, v)\) are evaluated. We report our simulation results in Table I.

First, it can be noticed from Table I 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 I, 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, 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 [2003], 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:

\[
\omega_{t+1} = \arg \max_{\omega \in \mathcal{W}} E_{\tilde{F}_{t+1}}[U(1 + \omega_1 r_{1,t+1} + \omega_2 r_{2,t+1})]
\]

\[
= \arg \max_{\omega \in \mathcal{W}} \int \int U(1 + \omega_1 x + \omega_2 y) \tilde{g}_{1,t+1}(x) \tilde{g}_{2,t+1}(y) \tilde{c}_{t+1}(\tilde{G}_{1,t+1}(x), \tilde{G}_{2,t+1}(y)) dxdy,
\]

where \( \tilde{G}_{1,t+1} \) and \( \tilde{G}_{2,t+1} \) are conditional marginal distribution forecasts, \( \tilde{g}_{1,t+1}(x) \) and \( \tilde{g}_{2,t+1}(y) \) are conditional marginal density forecasts, \( \tilde{c}_{t+1} \) 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 \( U(x) \) is the CRRA utility function given by

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

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

\[4\text{The conditional Sklar theorem asserts that } F_{t+1}(X,Y|F_t) = C_{t+1}(G_{1,t+1}(X|F_t), G_{2,t+1}(Y|F_t)|F_t), \text{ where } F_t \text{ is a set of information available up to } t \text{ (see, e.g., Patton [2004, 2006]).}\]
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 $F_{t+1}(X, Y | F_t)$ 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. 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 3 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:

\[
\rho^+(c) = \text{corr}(r_{1,t}, r_{2,t} | r_{1,t} > c, r_{2,t} > c), \\
\rho^-(c) = \text{corr}(r_{1,t}, r_{2,t} | r_{1,t} < -c, r_{2,t} < -c).
\]

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 3).

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:

\[
\begin{align*}
    r_{i,t} &= \alpha_{i,0} + \alpha_{i,1}r_{i,t-1} + \alpha_{i,2}r_{i,t-1}^f + \alpha_{i,3}DY_{i,t-1} + \alpha_{i,4}PER_{i,t-1} + \alpha_{i,5}TS_{t-1} + \alpha_{i,6}DS_{t-1} + \alpha_{i,7}Cay_{t-1} \\
    &+ \sigma_{i,t}\epsilon_{i,t} 
\end{align*}
\]  

(5.3)

with

\[
\begin{align*}
    \sigma_{i,t}^2 &= \alpha_{i,8} + \alpha_{i,9}\sigma_{i,t-1}^2 + \alpha_{i,10}\sigma_{i,t-1}^2 + \alpha_{i,11}\epsilon_{i,t-1}^2 + \alpha_{i,12}\epsilon_{i,t-1}^2(\epsilon_{i,t-1} > 0) + \alpha_{i,13}\epsilon_{i,t-1}^2(\epsilon_{i,t-1} < 0) + \alpha_{i,12}\epsilon_{i,t-1}^2 \\
    &+ \alpha_{i,13}DY_{i,t-1} + \alpha_{i,14}PER_{i,t-1} + \alpha_{i,15}TS_{t-1} + \alpha_{i,16}DS_{t-1} + \alpha_{i,17}Cay_{t-1}, 
\end{align*}
\]  

(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 \(\kappa_{i,t} = 2.1 + (Z_{i,t-1}^\prime \gamma_{i})^2\) and \(\lambda_{i,t} = \Lambda(Z_{i,t-1}^\prime \gamma_{i})\) with \(Z_{i,t}^\prime = \{r_{i,t}^f, DY_{i,t}, PER_{i,t}, TS_{t}, DS_{t}, Cay_{t}\}\) and \(\Lambda(x) = (1 - e^{-x})/(1 + e^{-x})\]. Moreover, \(r_{i,t}^f\), \(DY\), \(PER\), \(TS\), \(DS\), 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

\[
\left( \frac{r_{1,t} - \mu_{1,t}}{\sigma_{1,t}}, \frac{r_{2,t} - \mu_{2,t}}{\sigma_{2,t}} \right) \sim C(G(\epsilon_{1,t}|\kappa_{1,t}, \lambda_{1,t}), G(\epsilon_{2,t}|\kappa_{2,t}, \lambda_{2,t}); \delta_t)
\]

with \( \delta_t = \Gamma(\beta_0 + \beta_1 r_{t-1}^f + \beta_2 D Y_{t-1} + \beta_3 P E R_{t-1} + \beta_4 T S_{t-1} + \beta_5 D S_{t-1} + \beta_6 C a y_{t-1}) \),

where \( C \) denotes the Normal, Clayton, or Gumbel copula; \( \mu_{i,t} = E[r_{i,t}|F_{t-1}] \), 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 \( \{\hat{\rho}_S, \hat{\nu}_1, \hat{\nu}_2, \hat{\eta}\} \) 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

\[
\begin{align*}
\frac{\rho_{S,t}}{\rho_S} &= \Lambda(\theta_0 + \theta_1 r_{t-1}^f + \theta_2 D Y_{t-1} + \theta_3 P E R_{t-1} + \theta_4 T S_{t-1} + \theta_5 D S_{t-1} + \theta_6 C a y_{t-1}), \\
\frac{\nu_{1,t}}{\nu_1} &= \Lambda(\theta_7 + \theta_8 r_{t-1}^f + \theta_9 D Y_{t-1} + \theta_{10} P E R_{t-1} + \theta_{11} T S_{t-1} + \theta_{12} D S_{t-1} + \theta_{13} C a y_{t-1}), \\
\frac{\nu_{2,t}}{\nu_2} &= \Lambda(\theta_{14} + \theta_{15} r_{t-1}^f + \theta_{16} D Y_{t-1} + \theta_{17} P E R_{t-1} + \theta_{18} T S_{t-1} + \theta_{19} D S_{t-1} + \theta_{20} C a y_{t-1}), \\
\frac{\eta_t}{\eta} &= \Lambda(\theta_{21} + \theta_{22} r_{t-1}^f + \theta_{23} D Y_{t-1} + \theta_{24} P E R_{t-1} + \theta_{25} T S_{t-1} + \theta_{26} D S_{t-1} + \theta_{27} C a y_{t-1}).
\end{align*}
\]

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 (5.3)-(5.5) 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 1), the MECC portfolio weights are often more extreme and more volatile than the Normal portfolio weights (cf. Figure 1), 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 5). 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 5). 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


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 $[0,1]^2$:

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

such that

1. $(D1)$ $C(u,0) = C(0,v) = 0$,
2. $(D2)$ $C(u,1) = u$ and $C(1,v) = v$ for every $(u,v)$ of $[0,1]^2$, and
3. $(D3)$ $C$ 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 $\tau$ denote the difference between the probabilities of concordance and discordance of $(X_1,Y_1)$ and $(X_2,Y_2)$ as follows:

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

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 $F_1 = F_2$, $\tau$ is Kendall’s tau, $\tau_K$. 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

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

For instance, if $C_2(u, v) = uv$, then $\tau$ is Spearman’s rho, $\rho_S$; if $C_2(u, v) = 2(|u + v - 1| - |u - v|)$, then $\tau$ is Gini’s gamma; and if $C_1(u, v) = C_2(u, v)$, then $\tau$ 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 $r$. 

32
APPENDIX B: APPROXIMATION OF POTENTIAL FUNCTIONS (Supplementary Material)

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

\[
Q_{n,N,h}(\Lambda, \widehat{\theta}) = \frac{1}{4} \int_{[-1,1]^2} \exp \left\{ \sum_{k=0}^{2n-1} \left[ \lambda_k (\Phi(k - 2^{n-1}(x_1 + 1)) + \Phi(2^{n-1}(x_1 + 1) - k) - 1 + 2^{-n}) \right. \\
+ \gamma_k (\Phi(k - 2^{n-1}(x_2 + 1)) + \Phi(2^{n-1}(x_2 + 1) - k) - 1 + 2^{-n}) \right. \\
- \lambda_{2n} \left( h\left( \frac{x_1 + 1}{2}, \frac{x_2 + 1}{2} \right) - \widehat{\theta} \right) \left. \right]\} dx_1 dx_2
\]

\[
= \frac{1}{4} \int_{[-1,1]^2} \exp\{-\Lambda'\Psi(X)\}dX,
\]

(A.3)

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

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

\[
\exp\{-\Lambda'\Psi(X)\} = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} a_{nm} P_{nm}(X),
\]

(A.4)

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

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

and

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

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

$$\begin{align*}
\sum_{i=1}^{N} \sum_{j=1}^{M} \omega_{ij} P_{00}(X_{ij}) &= \sum_{i=1}^{N} \sum_{j=1}^{M} \omega_{ij} = 1, \\
\sum_{1}^{N} \sum_{1}^{M} \omega_{ij} P_{kh}(X_{ij}) &= 0, \quad \omega_{ij} \geq 0,
\end{align*}$$

(A.5)

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

$$\begin{align*}
\sum_{i=1}^{N} \sum_{j=1}^{M} \omega_{ij} \exp\{-\Lambda^\prime \Psi(X_{ij})\} &= \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} a_{nm} \sum_{i=1}^{N} \sum_{j=1}^{M} \omega_{ij} P_{nm}(X_{ij}) \\
&= a_{00} + \sum_{n=N+1}^{\infty} \sum_{m=M+1}^{\infty} a_{nm} \sum_{i=1}^{N} \sum_{j=1}^{M} \omega_{ij} P_{nm}(X_{ij}).
\end{align*}$$

(A.6)

Hence,

$$a_{00} = Q_{n,N_h}(\Lambda, \hat{\theta}) = \sum_{i=1}^{N} \sum_{j=1}^{M} \omega_{ij} \exp\{-\Lambda^\prime \Psi(X_{ij})\} - \sum_{n=N+1}^{\infty} \sum_{m=M+1}^{\infty} a_{nm} \sum_{i=1}^{N} \sum_{j=1}^{M} \omega_{ij} P_{nm}(X_{ij})$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{M} \omega_{ij} \exp\{-\Lambda^\prime \Psi(X_{ij})\} + \text{The Approximation} \quad \text{(A.7)}$$

where $R_{NM} = -\sum_{n=N+1}^{\infty} \sum_{m=M+1}^{\infty} a_{nm} \sum_{i=1}^{N} \sum_{j=1}^{M} \omega_{ij} P_{nm}(X_{ij})$ 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 $M = N = 30$. 

34
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

$$f_n(x) = 2^n \int_{\lfloor k2^{-n}, (k+1)2^{-n} \rfloor} f(y)dy \text{ for some } x \in \lfloor k2^{-n}, (k+1)2^{-n} \rfloor,$$

(A.8)

where $\{k2^{-n}\}$ is a compact, dense dyadic sequence in $\Omega$. [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, $f_n(x) \xrightarrow{P, \text{ as}} f(x)$.


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

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

(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:

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

Then $b(t) = 0$.

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

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

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

(A.10)
which has the following properties:

$$\lim_{N \to \infty} \Phi_N(y, x) \implies 1_{y>x}(y)$$

and

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

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


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

i. $$\int |g(x)| dx < \infty,$$

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

iii. $$\sup |g(x)| < \infty,$$

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

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

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


**LEMMA 6.** Let $$f : [a, b] \to \mathbb{R}$$ 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 \textcite{Barnett et al. (2010), p. 48}.

\textbf{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

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

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

Proof. See \textcite{Gine and Nickl (2008), p. 359}.

\textbf{Lemma 8} (‘Chaining’ inequality). Let $\phi : \mathcal{X} \rightarrow \mathbb{R}^{+}$, 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 $\varepsilon$ and $\varepsilon$ and any sequence of positive $\epsilon_{j}$ with $\varepsilon + \sum_{j=0}^{\infty} \epsilon_{j} \leq 1$,

$$ \mathbb{P} \left( \sup_{x \in \mathcal{X}} \phi(x) > z \right) \leq \mathbb{P} \left( \phi(x_{0}) > \varepsilon z \right) + \sum_{j=0}^{\infty} |\mathcal{X}_{j+1}| \sup_{d(x, u) \leq \delta_{j}} \mathbb{P} \left( |\phi(x) - \phi(u)| > \varepsilon_{j} z \right). $$

Proof. See \textcite{Cranston et al. (2000), p. 1858}.

\textbf{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:

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

where $a$ and $b$ are arbitrary numbers in $[0, 1]$. Using a dense dyadic sequence in $[0, 1]$, (A.11) can be approximated by

$$ \sum_{k=k_{1}}^{k_{2}} \int_{[k2^{-n}, (k+1)2^{-n}]} \int_{[0, 1]} c(u, v)dv \, du = \sum_{k=k_{1}}^{k_{2}} \int_{[0, 1]} \int_{[k2^{-n}, (k+1)2^{-n}]} c(u, v)dv \, du = \frac{k_{2} - k_{1}}{2^{n}}. $$
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 \( \epsilon \) is small enough.

Hence, (A.11) is equivalent to

\[
\int_{[k2^{-n},(k+1)2^{-n}]} \int_{[0,1]} c(u,v) \, du \, dv = \int_{[0,1]} \int_{[k2^{-n},(k+1)2^{-n}]} c(u,v) \, du \, dv = \frac{1}{2^n}
\]

\( \forall k = 0, 1, 2, \ldots, (2^n - 1) \), and \( n \) is large enough. (A.12)

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

\[
\mathcal{L}(c, \Lambda_n; \hat{\theta}) = -\int_{[0,1]^2} c(u,v) \log c(u,v) \, du \, dv - \lambda_1 \left[ \int_{[0,1]^2} c(u,v) \, du \, dv - 1 \right] \\
- \sum_{k=0}^{2^n-1} \left\{ \lambda_k \int_{[k2^{-n},(k+1)2^{-n}]} \int_{[0,1]} [c(u,v) - 2^{-n}] + \gamma_k \int_{[0,1]} \int_{[k2^{-n},(k+1)2^{-n}]} [c(u,v) - 2^{-n}] \right\} \\
- \lambda_{2^n} \int_{[0,1]^2} h(u,v,\hat{\theta})c(u,v) \, du \, dv \\
= -\int_{[0,1]^2} \left\{ c(u,v) \log c(u,v) + \lambda_1[c(u,v) - 1] \\
+ \sum_{k=0}^{2^n-1} \left( \lambda_k \mathbb{I}(u \in [k2^{-n}, (k+1)2^{-n}]) + \gamma_k \mathbb{I}(v \in [k2^{-n}, (k+1)2^{-n}]) \right) [c(u,v) - 2^{-n}] \\
+ \lambda_{2^n} h(u,v,\hat{\theta})c(u,v) \right\} \, du \, dv. \tag{A.13}
\]

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

\[
\int_{[0,1]^2} \left\{ \log c(u,v) + (1 + \lambda_1) + \sum_{k=0}^{2^n-1} \left[ \lambda_k \mathbb{I}(u \in [k2^{-n}, (k+1)2^{-n}]) \right. \\
+ \gamma_k \mathbb{I}(v \in [k2^{-n}, (k+1)2^{-n}]) \right] + \lambda_{2^n} h(u,v,\hat{\theta}) \right\} \, du \, dv = 0. \tag{A.14}
\]

Define \( b_n(u,v) = \log c(u,v) + (1 + \lambda_1) + \sum_{k=0}^{2^n-1} \left[ \lambda_k \mathbb{I}(u \in [k2^{-n}, (k+1)2^{-n}]) + \gamma_k \mathbb{I}(v \in [k2^{-n}, (k+1)2^{-n}]) \right] + \lambda_{2^n} h(u,v,\hat{\theta}) \), then applying Lemma to the function

\[
b(u,v) = \frac{b_n(u,v)}{c(u,v) - 1}, \tag{A.15}
\]
where $\tilde{c}(u, v)$ is an arbitrary copula density such that $\int_{[0,1]^2} (\tilde{c}(u, v) - 1) du dv = 0$, we obtain the following representation:

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

$$\left. \quad - \lambda_{2^n} h(u, v; \hat{\theta}) - b_0 \tilde{c}(u, v) \right\}, \quad (A.16)$$

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:

$$\tilde{c}_{n,N_h}(u, v) = \frac{\mathcal{E}_{n,N_h}(u, v)}{\int_{[0,1]^2} \mathcal{E}_{n,N_h}(u, v) du dv}, \quad (A.17)$$

where

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

$$\left. \quad - b_0 \tilde{c}(u, v) \right\}. $$

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

$$\int_{[0,1]^2} \mathcal{E}_{n,N_h}(u, v) du dv \int_{[0,1]^2} \mathbb{I}(u \in [k2^{-n}, (k+1)2^{-n}]) \mathcal{E}_{n,N_h}(u, v) du dv = 2^{-n},$$

$$\int_{[0,1]^2} \mathcal{E}_{n,N_h}(u, v) du dv \int_{[0,1]^2} \mathbb{I}(v \in [k2^{-n}, (k+1)2^{-n}]) \mathcal{E}_{n,N_h}(u, v) du dv = 2^{-n},$$

$$\int_{[0,1]^2} \mathcal{E}_{n,N_h}(u, v) du dv \int_{[0,1]^2} h(u, v; \hat{\theta}) \mathcal{E}_{n,N_h}(u, v) du dv = 0,$$

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

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

$$\left. \quad + \hat{\gamma}_k \mathbb{I}(v \in [k2^{-n}, (k+1)2^{-n}]) - 2^{-n} \right] - \lambda_{2^n} h(u, v; \hat{\theta}) - b_0 \tilde{c}(u, v) \right\}, \quad (A.18)$$

for all $k = 0, \ldots, (2^n - 1)$. Since (A.17) can be rewritten
we can define the potential function as follows:

\[
Q_{n,Nh}(\boldsymbol{\Lambda}_n, \hat{\theta}) = \int_{[0,1]^2} \exp \left\{ - \sum_{k=0}^{2^n-1} \left[ \lambda_k (\mathbb{I}(u \in [k2^{-n}, (k+1)2^{-n}]) - 2^{-n}) + \gamma_k (\mathbb{I}(v \in [k2^{-n}, (k+1)2^{-n}]) - 2^{-n}) \right] - \lambda_{2^n} h(u, v, \hat{\theta}) - b_0 \tilde{c}(u, v) \right\} dudv.
\]

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

\[
\begin{align*}
\frac{\partial}{\partial \lambda_k} Q_{n,Nh}(\boldsymbol{\Lambda}_n, \hat{\theta}) &= 0, \quad \text{for all } k = 0, \ldots, (2^n - 1). \\
\frac{\partial}{\partial \gamma_k} Q_{n,Nh}(\boldsymbol{\Lambda}_n, \hat{\theta}) &= 0, \\
\frac{\partial}{\partial \lambda_{2^n}} Q_{n,Nh}(\boldsymbol{\Lambda}_n, \hat{\theta}) &= 0
\end{align*}
\]

for all \(k = 0, \ldots, (2^n - 1)\). Also note that, since the second order derivatives of \(Q_{n,Nh}(\boldsymbol{\Lambda}_n, \hat{\theta})\) is the covariance matrix of \(\left\{ \mathbb{I}(u \in [k2^{-n}, (k+1)2^{-n}]), \mathbb{I}(v \in [k2^{-n}, (k+1)2^{-n}]), h(u, v, \hat{\theta}) \right\}_{k=0}^{2^n-1}\), thus \(Q_{n,Nh}(\boldsymbol{\Lambda}_n, \hat{\theta})\) is positive definite. It follows that the solutions to \((A.18)\) are the minimum values of \(Q_{n,Nh}(\boldsymbol{\Lambda}_n, \hat{\theta})\), which depend on \(\hat{\theta}, b_0,\) and \(\tilde{c}(u, v)\).

Since the potential function \(Q_{n,Nh}(\boldsymbol{\Lambda}_n, \hat{\theta})\) 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 \((A.9)\) yields

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

We then immediately obtain:

\[
Q_{n,Nh}(\boldsymbol{\Lambda}_n, \hat{\theta}) = \int_{[0,1]^2} \exp \left\{ \sum_{k=0}^{2^n-1} \lambda_k \left( \Phi(N_h(k2^{-n} - u)) + \Phi(-N_h((k+1)2^{-n} - u)) - 1 + 2^{-n} \right) \right\}
\]
\[ + \gamma_k \left( \Phi(N_h(k2^{-n} - v)) + \Phi(-N_h((k + 1)2^{-n} - v)) - 1 + 2^{-n} \right), \]

\[- \lambda_{2^n} h(u, v, \hat{\theta}) - b_0 \bar{c}(u, v) \bigg\} \, du \, dv, \tag{A.20} \]

and

\[ E_{n,N_h}(u, v) \approx \exp \left\{ \sum_{k=0}^{2^n-1} \hat{\lambda}_k \left( \Phi(N_h(k2^{-n} - u)) + \Phi(-N_h((k + 1)2^{-n} - u)) \right) \right. \]

\[ + \gamma_k \left( \Phi(N_h(k2^{-n} - v)) + \Phi(-N_h((k + 1)2^{-n} - v)) \right) \]

\[- \hat{\lambda}_{2^n} h(u, v, \hat{\theta}) - b_0 \bar{c}(u, v) \bigg\}, \tag{A.21} \]

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

Finally, to complete this proof, we still need to prove that the MEC approximator, \( \hat{C}_{n,N_h}(u, v) = \int_0^u \int_0^v \hat{c}_{n,N_h}(u, v) \, du \, dv \), is 2-increasing (cf. Definition 3.3). Let’s denote by \([u_1, u_1 + \Delta] \times [v_1, v_1 + \Delta]\) a rectangle in \([0, 1]^2\), we immediately establish that, since \( \hat{c}_{n,N_h}(u, v) \) is a [positive] exponential function, the mass of the rectangle, \( \hat{C}_{n,N_h}(u_1 + \Delta, v_1 + \Delta) - \hat{C}_{n,N_h}(u_1 + \Delta, v_1) - \hat{C}_{n,N_h}(u_1, v_1 + \Delta) + \hat{C}_{n,N_h}(u_1, v_1) = \int_{u_1}^{u_1 + \Delta} \int_{v_1}^{v_1 + \Delta} \hat{c}_{n,N_h}(u, v) \, du \, dv \), is thus nonnegative. Now, we can obtain the MECs by letting \( n \) and \( N_h \) become sufficiently large.

\[ \square \]

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

\[ \mathfrak{C} = \left\{ c \in W^r_1([0, 1]^2) : c \in (0, \infty); \int_0^1 \int_0^1 c(u, v) \, du \, dv = 1; \int_0^1 c(u, v) \, du = 1; \text{ and } \int_0^1 c(u, v) \, dv = 1 \right\}, \]

where \( W^r_1([0, 1]^2) \) 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:

\[
\int_{\epsilon}^{1} \lambda(u) \left\{ \int_{u-\epsilon}^{u} \int_{0}^{1} c(w, v) dwdv - \epsilon \right\} du = 0
\]

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

\[
\int_{\epsilon}^{1} \gamma(v) \left\{ \int_{0}^{1} \int_{v-\epsilon}^{v} c(u, w) dudw - \epsilon \right\} dv = 0,
\]

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

\[
\mathcal{L}(c, \Lambda; \theta_0) = -E[\log c(U, V)] - \lambda^* (E[1] - 1) - \gamma^* E[h(U, V, \theta_0)] - \int_{\epsilon}^{1} \lambda(u) (E[\Pi(w \in [u-\epsilon, u])] - \epsilon) du
\]

\[
- \int_{\epsilon}^{1} \gamma(v) (E[\Pi(w \in [v-\epsilon, v])] - v) dv,
\]

where \( \Lambda = \{\lambda^*, \gamma^*, \lambda(u), \gamma(v)\} \); and all the expectations are hereafter taken with respect to \( c \) belonging to the space \( W^r_1([0, 1]^2) \), unless otherwise indicated. The MEC is now defined as

\[
c_0(u, v) = \arg\sup_{c \in \mathcal{C}} \{-E[\log c(U, V)]\}
\]

\[
= \arg\sup_{c \in \mathcal{C}, \gamma^* \in B_{\gamma^*}} \{-E[\log c(U, V)] - \gamma^* E[h(U, V, \theta_0)]\}
\]

\[
= \arg\sup_{c \in W^r_1([0, 1]^2), \Lambda \in B_{\Lambda^*} \times B_{\gamma^*} \times C^*([0, 1]) \times C^*([0, 1])} \mathcal{L}(c, \Lambda; \theta_0), \quad (A.23)
\]

where \( B_{\lambda^*} \) and \( B_{\gamma^*} \) are bounded sets in \( \mathbb{R} \);

\[
\mathcal{C} = \left\{ c \in W^r_1([0, 1]^2) : c \in (0, \infty), \ E[1] = 1, \ \int_{\epsilon}^{1} \lambda(u) \left\{ \int_{u-\epsilon}^{u} \int_{0}^{1} c(w, v) dwdv - \epsilon \right\} = 0, \right. \\
\left. \int_{\epsilon}^{1} \gamma(v) \left\{ \int_{0}^{1} \int_{v-\epsilon}^{v} c(u, w) dudw - \epsilon \right\} = 0, \ and \ E[h(U, V, \theta_0)] = 0, \ where \ \lambda(u), \gamma(v) \in C^*([0, 1]) \right\};
\]

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:

$$\mathcal{C}_T \doteq \left\{ c \in W^r_1([0, 1]^2) : c \in (0, \infty), \ E[1] = 1, \ \frac{1}{N_T} \sum_{k=1}^{N_T} \lambda_k \left\{ \int_{k-1/N_T}^{k/N_T} \int_0^1 c(w, v)dw \, dv - \frac{1}{N_T} \right\} = 0, \right.$$ 

and

$$\frac{1}{N_T} \sum_{k=1}^{N_T} \gamma_k \left\{ \int_{k-1/N_T}^{k/N_T} \int_0^1 c(u, w)du \, dw - \frac{1}{N_T} \right\} = 0 \text{ with } \{\lambda_k, \gamma_k\} \in B_\lambda \times B_\gamma,$$

where the $\lambda_k$'s and the $\gamma_k$'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

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

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

$$\mathcal{L}(c, \Lambda_T; \theta_0) = -E[\log c(U, V)] - \lambda^* (E[1] - 1) - \gamma^* E[h(U, V, \theta_0)]$$

$$- \frac{1}{N_T} \sum_{k=1}^{N_T} \lambda_k \left\{ \int_{k-1/N_T}^{k/N_T} \int_0^1 c(w, v)dw \, dv - \frac{1}{N_T} \right\} - \frac{1}{N_T} \sum_{k=1}^{N_T} \gamma_k \left\{ \int_{k-1/N_T}^{k/N_T} c(u, w)du \, dw - \frac{1}{N_T} \right\}$$

$$= -E[\log c(U, V)] - \lambda^* (E[1] - 1) - \gamma^* E[h(U, V, \theta_0)]$$

$$- \frac{1}{N_T} \sum_{k=1}^{N_T} \lambda_k \left\{ E \left[ \mathbb{I} \left( u \in \left[ \frac{k-1}{N_T}, \frac{k}{N_T} \right) \right) \right] - \frac{1}{N_T} \right\}$$

$$- \frac{1}{N_T} \sum_{k=1}^{N_T} \gamma_k \left\{ E \left[ \mathbb{I} \left( v \in \left[ \frac{k-1}{N_T}, \frac{k}{N_T} \right) \right) \right] - \frac{1}{N_T} \right\},$$

where $\Lambda_T = \{\lambda^*, \gamma^*, \lambda_1, \ldots, \lambda_{N_T}, \gamma_1, \ldots, \gamma_{N_T}\}$. By virtue of Lemma 3, a kernel-based smoothing of indicator functions yields

$$\mathcal{L}_T(c, \Lambda_T; \theta_0) = E[\log c(U, V)] - \lambda^* (E[1] - 1) - \gamma^* E[h(U, V, \theta_0)]$$

$$- \frac{1}{N_T} \sum_{k=1}^{N_T} \lambda_k \left\{ E \left[ \frac{1}{h_T} \int_{k-1/N_T}^{k/N_T} K \left( \frac{u - w}{h_T} \right) \, dw \right] - \frac{1}{N_T} \right\}$$

$$- \frac{1}{N_T} \sum_{k=1}^{N_T} \gamma_k \left\{ E \left[ \frac{1}{h_T} \int_{k-1/N_T}^{k/N_T} K \left( \frac{v - w}{h_T} \right) \, dw \right] - \frac{1}{N_T} \right\},$$
where $K(\cdot)$ is a continuously bounded kernel function satisfying Assumption (4). The smoothed approximation spaces are then defined as

$$
\mathcal{C}_T^* = \left\{ c \in W_1([0,1]^2) : c \in (0, \infty), \ E[1] = 1, \ \frac{1}{N_T} \sum_{k=1}^{N_T} \lambda_k \left\{ E \left[ \frac{1}{h_T} \int_{h_T}^{k/N_T} K \left( \frac{u-w}{h_T} \right) dw \right] \right\} - \frac{1}{N_T} = 0, \ and \ \frac{1}{N_T} \sum_{k=1}^{N_T} \gamma_k \left\{ E \left[ \frac{1}{h_T} \int_{h_T}^{k/N_T} K \left( \frac{v-w}{h_T} \right) dw \right] - \frac{1}{N_T} \right\} = 0 \right\}
$$

with $\{\lambda_k, \gamma_k\} \in B_\lambda \times B_\gamma$.

$$
C_T^*(\tilde{\theta}) = \left\{ c \in \mathcal{C}_T^* : E[h(U, V, \tilde{\theta})] = 0 \right\}.
$$

First, we shall show that

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

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

$$
\max_{\{c, \Lambda_T\} \in \{W_1([0,1]^2), B_\lambda \times B_\gamma \times B_\gamma^{N_T} \times B_\gamma^{N_T} \}} \mathcal{L}_T \left( c, \Lambda_T; \tilde{\theta} \right) \geq \max_{\{c, \Lambda\} \in \{W_1([0,1]^2), B_\lambda \times B_\gamma \times C^*([0,1]) \times C^*([0,1]) \}} \mathcal{L} \left( c, \Lambda; \theta_0 \right) - o_P(1), \quad (A.24)
$$

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

$$
\mathcal{L}_T(c, \Lambda_T; \tilde{\theta}) = \mathcal{L}(c, \Lambda, \theta_0) + \mathcal{L}(c, \Lambda_T; \theta_0) - \mathcal{L}(c, \Lambda; \theta_0) + \mathcal{L}_T \left( c, \Lambda_T; \tilde{\theta} \right)
$$

$$
\geq \mathcal{L}(c, \Lambda, \theta_0) - \mathcal{L}(c, \Lambda_T; \theta_0) - \mathcal{L}(c, \Lambda; \theta_0) - \mathcal{L}_T \left( c, \Lambda_T; \tilde{\theta} \right) - \mathcal{L}(c, \Lambda_T; \theta_0)
$$

$$
= \mathcal{L}(c, \Lambda, \theta_0) - \{T_1 + T_2 + T_3 \}.
$$
We now proceed to bound the above three terms: For $\mathcal{T}_1$, one has

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

where

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

and

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

Setting $\epsilon = 1/N_T$, it then follows from Lemma~\ref{lem:smooth} 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

$$
h(u, v, \hat{\theta}) = h(u, v, \theta_0) + D^\theta h(u, v, \theta_0) (\hat{\theta} - \theta_0),
$$

where $\theta$ lies on the line joining $\theta_0$ and $\hat{\theta}$ 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~\ref{lem:smooth} yields; for some $c \in W^1_1([0, 1]^2)$ and $K \in \mathcal{K}^r(\mathbb{R})$ with some integer, $r > 1$;

$$
E \left[ \frac{1}{h_T} \int_{\frac{k-1}{N_T}}^{\frac{k}{N_T}} K \left( \frac{u - w}{h_T} \right) dw - \mathbb{I} \left( u \in \left[ \frac{k-1}{N_T}, \frac{k}{N_T} \right) \right) \right] \right|.
$$
Thus, Eq. (A.24) follows.

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

**Step 1:** First we choose sequences of constants, \( \epsilon_T^{(1)} = T^{-\ell} \) 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 \( \mathcal{C}_T^* \), where \( \Pi_T \) signifies a projection operator; and \( \Pi_T \mathcal{A}_0 = \{\lambda_0^*, \gamma_0^*, \Pi_T \lambda_0(u), \Pi_T \gamma_0(v)\} \). We obtain

\[
\mathbb{P} \left( \| \hat{c}_T - c_0 \|_\epsilon \geq D \epsilon_T^{(1)} \right) \leq \mathbb{P} \left( \sup_{(c, \Lambda_T) \in \{W^1([0,1]^2), B_\Lambda \times B_\gamma N_T \times B_\gamma N_T\}} \mathcal{L}_T \left( c, \Lambda_T; \hat{\theta} \right) \geq \mathcal{L}_T \left( \Pi_T c_0, \Pi_T \mathcal{A}_0, \hat{\theta} \right) \right) \left\{ \begin{array}{l}
\| c - c_0 \|_\epsilon \geq D \epsilon_T^{(1)} \\
c \in \mathcal{C}_T^*
\end{array} \right\}
\]
\[
\mathcal{P} \left\{ \sup \left\{ \text{L}_T(c, \Lambda_T; \hat{\theta}) \geq \text{L}_T \left( \Pi_T c_0, \Pi_T \Lambda_0, \hat{\theta} \right) \right\} \right\},
\]

\[
\bigcap \left\{ \sup \left\{ \text{L}_T(c, \Lambda_T, \hat{\theta}) - \text{L}_T(c, \Lambda_T, \theta_0) > \eta_T^{(1)} \right\} \right\}
\]

\[
\bigcup \left\{ \sup \left\{ \text{L}_T(c, \Lambda_T, \hat{\theta}) - \text{L}_T(c, \Lambda_T, \theta_0) \leq \eta_T^{(1)} \right\} \right\}
\]

\[
\bigcup \left\{ \sup \left\{ \text{L}_T(c, \Lambda_T, \theta_0) \geq \text{L}_T(\Pi_T c_0, \Pi_T \Lambda_0, \theta_0) - 2\eta_T^{(1)} \right\} \right\}
\]

\[
= \mathcal{T}_1^* + \mathcal{T}_2^*.
\]

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

Note that \( \mathcal{T}_1^* = \mathcal{P} \left( \overline{\gamma} \sup_{c \in \mathcal{C}_T^*} |E[h(U, V, \hat{\theta}) - h(U, V, \theta_0)]| > \eta_T^{(1)} \right) \), where \( \overline{\gamma} \) is the largest value
of the bounded subset $B_{r^*}$. To bound this probability, we shall rely on the ‘chaining’ inequality of Cranston et al. [2001] stated as Lemma 8. We start by choosing a finite ‘net’, $\mathcal{C}_{T,j}^*$, in $\mathcal{C}_T^*$ such that, for each $c \in \mathcal{C}_T^*$, there is a $\pi_j c \in \mathcal{C}_{T,j}^*$ such that $\|\pi_j c - c\|_\varepsilon \leq r^{-j}$ for some integer, $r \geq 2$. Indeed, $|\mathcal{C}_{T,j}^*|$ is the number of copulas of $\mathcal{C}_T^*$ contained in $\mathcal{C}_{T,j}^*$. We denote by $N(r^{-j}, \mathcal{C}_T^*)$ the number of balls, $B_{r^{-j}}(c_0) = \{c \in \mathcal{C}_T^* : \|c - c_0\|_\varepsilon < r^{-j}\}$, that covers $\mathcal{C}_T^*$. Then, since the space $\mathcal{C}_T^*$ is bounded, one can always choose $\mathcal{C}_{T,j}^*$ such that $|\mathcal{C}_{T,j}^*| \leq N(r^{-j}, \mathcal{C}_T^*)$. By virtue of Lemma 8, it follows that

$$T_1^* \leq \mathbb{P}\left(\phi(c_0; \hat{\theta}, \theta_0) > \epsilon_\star \eta_T^{(1)}\right) + \sum_{j=0}^{\infty} N(r^{-j+1}, \mathcal{C}_T^*) \sup_{\|c_1 - c_2\|_\varepsilon \leq r^{-j}} \mathbb{P}\left(\phi(c_1; \hat{\theta}, \theta_0) - \phi(c_2; \hat{\theta}, \theta_0) > \epsilon_j \eta_T^{(1)}\right)$$

$$= T_{1,a}^* + T_{1,b}^*,$$

for generic positive constants, $\epsilon_\star$ and $\epsilon_0, \ldots, \epsilon_\infty$, such that $\epsilon_\star + \sum_{j=0}^{\infty} \epsilon_j < \infty$.

where $\phi(c; \hat{\theta}, \theta_0) = \left|E_c[h(U, V, \hat{\theta}) - h(U, V, \theta_0)]\right|$. The mean-value theorem yields

$$T_{1,a}^* = \mathbb{P}\left(|D_\theta E[h(U, V, \theta)]|_{\theta = \bar{\theta}} \mid \hat{\theta} - \theta_0\mid > \epsilon_\star \eta_T^{(1)}\right),$$

where $\bar{\theta}$ is some point on the line joining $\theta_0$ and $\hat{\theta}$ in the Euclidean space; and $D_\theta$ contains first-order partial derivatives. By virtue of Assumptions (3.1) and (3.2), we obtain $T_{1,a}^* \to 0$. Next, to bound $T_{1,b}^*$, note that, by the mean-value theorem,

$$\phi(c_1; \hat{\theta}, \theta_0) - \phi(c_2; \hat{\theta}, \theta_0) = \{|D_\theta E_{c_1}[h(U, V, \theta)]|_{\theta = \theta^*} - |D_\theta E_{c_2}[h(U, V, \theta)]|_{\theta = \theta^*}\} \mid \hat{\theta} - \theta_0\mid,$$

where $\theta^*$ and $\theta^{**}$ are some points on the line joining $\hat{\theta}$ and $\theta_0$. The Tchebyshev inequality gives that

$$\sup_{\|c_1 - c_2\|_\varepsilon \leq r^{-j}} \mathbb{P}\left(\mid \phi(c_1; \hat{\theta}, \theta_0) - \phi(c_2; \hat{\theta}, \theta_0)\mid > \epsilon_j \eta_T^{(1)}\right) \leq r^{-j} \left\{\sup_{\theta \in \Theta} \int_0^1 \int_0^1 |D_\theta h(u, v, \theta)| du dv \right\} \frac{E\left[|\hat{\theta} - \theta_0|\right]}{\epsilon_j \eta_T^{(1)}}.$$

It then follows that

$$T_{1,b}^* \leq \left\{\sup_{\theta \in \Theta} \int_0^1 \int_0^1 |D_\theta h(u, v, \theta)| du dv \right\} \frac{E\left[|\hat{\theta} - \theta_0|\right]}{\eta_T^{(1)}} \sum_{j=0}^{\infty} r^{-j} N(r^{-(j+1)}, \mathcal{C}_T^*)/\epsilon_j.$$
Setting $\epsilon_j = \frac{1}{r^{j-a}(\log(\frac{1}{r^{j-a}}))^{p}}$ 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

$$\mathcal{T}_{1,b}^{*} \leq \text{Const.} \times \frac{E[\hat{\theta} - \theta_0]}{\eta_T^{(1)}} \sum_{j=0}^{\infty} r^{-j-a} (\log (1/r^{-j-1}))^{p} N(r^{-j+1}, \mathcal{C}_{T}^{*})(r^{-j} - r^{-j-1}).$$

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

$$\mathcal{T}_{1,b}^{*} \leq \sup_{\theta \in \Theta} \int_{0}^{1} \int_{0}^{1} |D_{\theta}h(u, v, \theta)| \, du \, dv \cdot \frac{E[|\hat{\theta} - \theta_0|]}{\eta_T^{(1)}} \int_{0}^{1} \epsilon (\log 1/\epsilon)^{p} N(\epsilon, \mathcal{C}_{T}^{*}) \, d\epsilon.$$

Under Assumptions (E2) and (E3), we have that $\mathcal{T}_{1,b}^{*} \rightarrow 0$.

Next, for the term $\mathcal{T}_{2}^{*}$, an algebraic manipulation yields

$$\mathcal{T}_{2}^{*} \leq \Pi \left( \sup_{\{c, \Pi_{T} \Lambda \} \in \{W_{T}^{1}(0,1)^{2}, B_{\gamma} \times B_{\gamma} \times B_{\gamma}^{N_{\gamma} \times B_{\gamma}^{N_{\gamma}}}, \|c-c_{0}\|_{E} + \|\Lambda - \Lambda_{0}\|_{EC} \geq \Delta_{C}^{(1)} \}} \mathcal{L}_{T}(c, \Pi_{T} \Lambda, \theta_0) \geq \mathcal{L}_{T}(\Pi_{T} c_{0}, \Pi_{T} \Lambda_{0}, \theta_0) - 2\eta_T^{(1)} \right),$$

where $\|\Lambda - \Lambda_{0}\|_{EC} \geq \sqrt{\|\lambda^{*} - \lambda_{0}\|^{2} + \|\gamma^{*} - \gamma_{0}\|^{2} + \|\lambda - \lambda_{0}\|_{C^{*}}^{2} + \|\gamma - \gamma_{0}\|_{C^{*}}^{2}}$ with a pseudo-metric, $\| \bullet \|_{C^{*}}$ on $C^{*}([0,1])$. Then,

$$\mathcal{T}_{2}^{*} \leq \Pi \left( \sup_{\{c, \Pi_{T} \Lambda \} \in \{W_{T}^{1}(0,1)^{2}, B_{\gamma} \times B_{\gamma} \times B_{\gamma}^{N_{\gamma} \times B_{\gamma}^{N_{\gamma}}}, \|c-c_{0}\|_{E} + \|\Lambda - \Lambda_{0}\|_{EC} \geq \Delta_{C}^{(1)} \}} \left\{ \mathcal{L}_{T}(c, \Pi_{T} \Lambda, \theta_0) - \mathcal{L}_{T}(\Pi_{T} c_{0}, \Pi_{T} \Lambda_{0}, \theta_0) - \mathcal{L}(c, \Lambda, \theta_0) \right\} + \mathcal{L}(c_{0}, \Lambda_{0}, \theta_0) \right) \geq \inf_{\{c, \Lambda \} \in \{W_{T}^{1}(0,1)^{2}, B_{\gamma} \times B_{\gamma} \times C^{*}((0,1)] \times C^{*}((0,1]) \}} \left\{ \mathcal{L}(c_{0}, \Lambda_{0}, \theta_0) - \mathcal{L}(c, \Lambda, \theta_0) \right\} - 2\eta_T^{(1)} \right).$$
In light of Assumption \([\text{B3}]\), one can show that

\[
\mathcal{T}_2^* \leq \mathbb{I} \left( \sup_{\{c, \Pi_T \Lambda \} \in \{W^*_T(\{0,1\}^2), B_{\pi_t} \times B_{\pi_\theta} \times B_{\pi_T} \times B_{\pi_N} \}} \left\{ \mathcal{L}_T(c, \Pi_T \Lambda, \theta_0) - \mathcal{L}_T(P \Pi_T c_0, \Pi_T \Lambda_0, \theta_0) - \mathcal{L}(c, \Lambda, \theta_0) \right\} \right)
\]

\[
+ \mathcal{L}(c_0, \Lambda_0, \theta_0) \geq 2D_4 \varepsilon_T^{(1/\omega)},
\]

where \(D_4 > 0\).

It then follows that

\[
\mathcal{T}_2^* \leq \mathbb{I} \left( \sup_{\{c, \Pi_T \Lambda \} \in \{W^*_T(\{0,1\}^2), B_{\pi_t} \times B_{\pi_\theta} \times B_{\pi_T} \times B_{\pi_N} \}} \left\{ \int_{\frac{1}{N_T}} \lambda(u) \left( E \left[ \mathbb{I}(w \in [u - \frac{1}{N_T}, u]) \right] - \frac{1}{N_T} \right) \right\} \right)
\]

\[
- \frac{1}{N_T} \sum_{k=1}^{N_T} \lambda_k \left\{ E \left[ \frac{1}{h_T} \int_{\frac{k}{N_T}}^{\frac{k+1}{N_T}} K \left( \frac{u-w}{h_T} \right) dw \right] - \frac{1}{N_T} \right\}
\]

\[
+ \left\{ \int_{\frac{1}{N_T}}^{1} \rho(u) \left( E \left[ \mathbb{I}(w \in [u - \frac{1}{N_T}, u]) \right] - \frac{1}{N_T} \right) \right\}
\]

\[
- \left\{ \int_{\frac{1}{N_T}}^{1} \lambda_0(u) \left( E \left[ \mathbb{I}(w \in [u - \frac{1}{N_T}, u]) \right] - \frac{1}{N_T} \right) \right\}
\]

\[
- \frac{1}{N_T} \sum_{k=1}^{N_T} \lambda_{0,k} \left\{ E \left[ \frac{1}{h_T} \int_{\frac{k}{N_T}}^{\frac{k+1}{N_T}} K \left( \frac{u-w}{h_T} \right) dw \right] - \frac{1}{N_T} \right\}
\]

\[
- \left\{ \int_{\frac{1}{N_T}}^{1} \gamma_0(u) \left( E \left[ \mathbb{I}(w \in [u - \frac{1}{N_T}, u]) \right] - \frac{1}{N_T} \right) \right\}
\]

\[
- \frac{1}{N_T} \sum_{k=1}^{N_T} \gamma_{0,k} \left\{ E \left[ \frac{1}{h_T} \int_{\frac{k}{N_T}}^{\frac{k+1}{N_T}} K \left( \frac{v-w}{h_T} \right) dw \right] - \frac{1}{N_T} \right\}
\]

\[
= \mathbb{I} \left( \sup_{\{c, \Pi_T \Lambda \} \in \{W^*_T(\{0,1\}^2), B_{\pi_t} \times B_{\pi_\theta} \times B_{\pi_T} \times B_{\pi_N} \}} \left\{ |\mathcal{F}_1 + \mathcal{F}_2 - \mathcal{F}_3 - \mathcal{F}_4| \geq 2D_4 \varepsilon_T^{(1/\omega)} \right\} \right).
\]
An application of Lemmas 3 and 4 yields

\[
\sum_1 \leq A_1 N_T^{-r_1} + A_2 \sup_{v \in [0,1]} \left( \int_0^1 |D_v^c(c(u,v))| \, du \right) \left( \int \left| u \right|^r K(u) \, du \right) h_T^r,
\]

\[
\sum_2 \leq A_3 N_T^{-r_2} + A_4 \sup_{u \in [0,1]} \left( \int_0^1 |D_u^c(c(u,v))| \, dv \right) \left( \int \left| v \right|^r K(v) \, dv \right) h_T^r,
\]

\[
\sum_3 \leq A_5 N_T^{-r_1} + A_6 \sup_{v \in [0,1]} \left( \int_0^1 |D_v^c(c_0(u,v))| \, du \right) \left( \int \left| u \right|^r K(u) \, du \right) h_T^r,
\]

\[
\sum_4 \leq A_7 N_T^{-r_2} + A_8 \sup_{u \in [0,1]} \left( \int_0^1 |D_u^c(c_0(u,v))| \, dv \right) \left( \int \left| v \right|^r K(v) \, dv \right) h_T^r,
\]

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 \(N_T \min\left( T^{-\frac{r}{1+\omega}}, T^{-\frac{r}{2}} \right) \rightarrow \infty\) and \(h_T = o\left( T^{-\frac{r}{2}} \right)\) respectively, then

\[
\lim_{T \to \infty} \mathbb{I}_{\sum_1 + \sum_2 - \sum_3 - \sum_4 \geq 2D_1 e_T^{(1)/\omega}} = 0.
\]

This immediately implies that \(T_2^* \to 0\).

**Step 2:** We conjecture that the optimal rate of convergence for \(\tilde{c}_T\) also depends on the rate of deviation, \(\omega\), [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 \(e_T^{(2)} = T^{-\ell(1+\Omega_1)} \) and \(\eta_T^{(2)} = T^{-\ell^*}\) for some \(\ell^* \in \left( \frac{(1+\Omega_1)}{\omega}, \frac{1}{2} \right)\). Note that

\[
\mathbb{P}\left( \left\| \tilde{c}_T - c_0 \right\|_E \geq D e_T^{(2)} \right) = \mathbb{P}\left( \left\| \tilde{c}_T - c_0 \right\|_E \geq D e_T^{(1)} \right) + \mathbb{P}\left( D e_T^{(2)} \leq \left\| \tilde{c}_T - c_0 \right\|_E < D e_T^{(1)} \right).
\]

We have shown in Step 1 that \(\mathbb{P}\left( \left\| \tilde{c}_T - c_0 \right\|_E \geq D e_T^{(1)} \right) \rightarrow 0\); thus, in this step, we shall show that
\[
\mathbb{P}\left( D\varepsilon_T^{(2)} \leq \|\hat{c}_T - c_0\|_c < D\varepsilon_T^{(1)} \right) \rightarrow 0.
\]

\[
\mathbb{P}\left( D\varepsilon_T^{(2)} \leq \|\hat{c}_T - c_0\|_c < D\varepsilon_T^{(1)} \right) = \mathbb{P}\left( \sup_{\{c, \Lambda_T\} \in \{W_{\bar{T}}([0,1]^2), B_{\lambda_T} \times B_{\bar{T}} \times B_{\Lambda_T} \times B_{\bar{\Lambda}}^{NT} \times B_{\bar{\Lambda}}^{NT}\}} \mathcal{L}_T(c, \Lambda_T, \hat{\theta}) \geq \mathcal{L}_T(\Pi_T c_0, \Pi_T \Lambda_0, \hat{\theta}) \right)
\]

\[
\sup_{\{c, \Lambda_T\} \in \{W_{\bar{T}}([0,1]^2), B_{\lambda_T} \times B_{\bar{T}} \times B_{\Lambda_T} \times B_{\bar{\Lambda}}^{NT} \times B_{\bar{\Lambda}}^{NT}\}} \mathcal{L}_T(c, \Lambda_T, \theta_0) \geq \mathcal{L}_T(\Pi_T c_0, \Pi_T \Lambda_0, \theta_0) - 2\eta_T^{(2)}
\]

\[
\mathbb{P}(\mathcal{A}) = \mathbb{P}(\mathcal{A} \cap \mathcal{B} \cup \mathcal{A} \cap \mathcal{B}^c)
\]

\[
\leq \mathbb{P}(\mathcal{B}) + I
\]

\[
\sup_{\{c, \Lambda_T\} \in \{W_{\bar{T}}([0,1]^2), B_{\lambda_T} \times B_{\bar{T}} \times B_{\Lambda_T} \times B_{\bar{\Lambda}}^{NT} \times B_{\bar{\Lambda}}^{NT}\}} \mathcal{L}_T(c, \Lambda_T, \theta_0) \geq \mathcal{L}_T(\Pi_T c_0, \Pi_T \Lambda_0, \theta_0) - 2\eta_T^{(2)}
\]

\[
\mathcal{T}_1^{**} + \mathcal{T}_2^{**},
\]

where \( B = \left\{ \sup_{\{c, \Lambda_T\} \in \{W_{\bar{T}}([0,1]^2), B_{\lambda_T} \times B_{\bar{T}} \times B_{\Lambda_T} \times B_{\bar{\Lambda}}^{NT} \times B_{\bar{\Lambda}}^{NT}\}} | \mathcal{L}_T(c, \Lambda_T, \hat{\theta}) - \mathcal{L}_T(\Pi_T c_0, \Pi_T \Lambda_0, \hat{\theta}) \right\}
\]

\[
\mathcal{L}_T(\Pi_T c_0, \Pi_T \Lambda_0, \theta_0) - \mathcal{L}_T(c, \Lambda_T, \theta_0) > \eta_T^{(2)} \right\}. \text{ We now proceed to bound } \mathcal{T}_1^{**} \text{ and } \mathcal{T}_2^{**}.
\]

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

\[
\mathcal{L}_T(c, \Lambda_T, \hat{\theta}) - \mathcal{L}_T(c, \Lambda_T, \theta_0) = \gamma^* E_{\theta} \left[ D'_{\theta} h(U, V, \theta) \right]_{\theta=\hat{\theta}} \left( \hat{\theta} - \theta \right)
\]

and

\[
\mathcal{L}_T(\Pi_T c_0, \Pi_T \Lambda_0, \hat{\theta}) - \mathcal{L}_T(\Pi_T c_0, \Pi_T \Lambda_0, \theta_0) = \gamma^* E_{\Pi_T c_0} \left[ D'_{\theta} h(U, V, \theta) \right]_{\theta=\tilde{\theta}} \left( \hat{\theta} - \theta \right).
\]
In view of Assumptions (B1) and (B2), it immediately follows that, as the ball \( \{ c \in \mathcal{C}^* : D(c, 0)^{ε}_T \leq \| c - c_0 \|_ε \leq Dε_T^{(1)} \} \) shrinks to \( c_0 \), we have that \( \mathcal{T}_1^{**} \rightarrow 0 \).

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

\[
\mathcal{T}_2^{**} \leq \mathbb{I} \left( \sup_{c, Π_T Λ \in \{ W_T((0,1]^2), B_λ^* × B_{N_T}^* × B_{N_T}^* \}} \mathcal{L}_T(c, Π_T Λ, θ_0) - \mathcal{L}_T(Π_T c_0, Π_T Λ_0, θ_0) - \mathcal{L}(c, Λ, θ_0) - \mathcal{L}(c_0, Λ_0, θ_0) \right) \geq \mathcal{L}(c_0, Λ_0, θ_0)
\]

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

Repeat Step 2 an infinite number of times, by setting \( ε_T^{(i)} = T^{-ε(1+Ω+\cdots+Ω)} \) and \( η_T^{(i)} = T^{-ε(1)} \) for some \( ε^{(i)} \in \left( \frac{ε(1+Ω+\cdots+Ω)}{ω}, \frac{1}{2} \right) \) in Step \( i \), one can derive that \( \| \hat{c}_T - c_0 \|_ε = o \left( T^{-ωε} \right) \).
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 (5.3)-(5.9). Basically, we need to estimate the set of parameters \( \Xi = \{\theta_0, \ldots, \theta_{27}\} \). Let \( \hat{\Xi} \) denote an estimate of \( \Xi \), our estimation procedure comprises of two steps described below.

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

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

(b) Given these measures of association, compute the time-varying parameters of the MECC \( (\Lambda_t \forall t = 1, \ldots, T) \) 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:

\[
 MECC - \text{likelihood}(\Xi) = \frac{1}{T} \sum_{t=1}^T \log c(\hat{G}_{1,t}, \hat{G}_{2,t}, \Lambda_t),
\]

where \( c(u, v, \Lambda) \) is the MECC density.

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

In order to reduce the computational time of finding the global maximum value of \( MECC - \text{likelihood}(\Xi) \), we computed \( \Lambda_t \), 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, \( \Lambda_t = f(\rho_{S_t}, \nu_{1,t}, \nu_{2,t}, \eta_t) \), from the domains of \( \rho_S, \nu_1, \nu_2, \) and \( \eta \) to the domains of \( \Lambda \); 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

$$\max_x \max_{x \in X} |f(x) - g(x)|,$$

where $X$ 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 $\Lambda$ from $\{\rho_S, \nu_1, \nu_2, \eta\}$. 
<table>
<thead>
<tr>
<th>Copula</th>
<th>$\theta = 0.1$</th>
<th>$\theta = 0.8$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Int. Bias$^2$</td>
<td>Int. Var.</td>
</tr>
<tr>
<td>Normal</td>
<td>27.9150</td>
<td>1.0416*e$^{-15}$</td>
</tr>
<tr>
<td>Clayton</td>
<td>0.0002</td>
<td>2.8924*e$^{-15}$</td>
</tr>
<tr>
<td>Gumbel</td>
<td>0.0002</td>
<td>2.8745*e$^{-15}$</td>
</tr>
<tr>
<td>Student’s $t$</td>
<td>1203.8600</td>
<td>2.9161*e$^{-12}$</td>
</tr>
<tr>
<td>MECC(4,1)</td>
<td>0.0028</td>
<td>0.0342</td>
</tr>
<tr>
<td>MECC(16,1)</td>
<td>0.0149</td>
<td>0.0014</td>
</tr>
<tr>
<td>MECC(64,1)</td>
<td>0.0126</td>
<td>0.0007</td>
</tr>
<tr>
<td>MECC(4,2)</td>
<td>0.0145</td>
<td>0.0878</td>
</tr>
<tr>
<td>MECC(16,2)</td>
<td>0.0170</td>
<td>0.0070</td>
</tr>
<tr>
<td>MECC(64,2)</td>
<td>0.0112</td>
<td>0.0017</td>
</tr>
<tr>
<td>MECC(4,3)</td>
<td>0.2667</td>
<td>11.4610</td>
</tr>
<tr>
<td>MECC(16,3)</td>
<td>0.0192</td>
<td>0.0804</td>
</tr>
<tr>
<td>MECC(64,3)</td>
<td>0.0184</td>
<td>0.0159</td>
</tr>
<tr>
<td>MECC(4,4)</td>
<td>0.1858</td>
<td>10.3994</td>
</tr>
<tr>
<td>MECC(16,4)</td>
<td>0.0487</td>
<td>1.7230</td>
</tr>
<tr>
<td>MECC(64,4)</td>
<td>0.0302</td>
<td>0.2109</td>
</tr>
</tbody>
</table>

*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

<table>
<thead>
<tr>
<th></th>
<th>Full Sample</th>
<th>In-Sample</th>
<th>Out-of-Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Russell 1000 Growth</td>
<td>Russell 1000 Value</td>
<td>Russell 1000 Growth</td>
</tr>
<tr>
<td>Mean</td>
<td>0.0051</td>
<td>0.0072</td>
<td>0.0204</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.0497</td>
<td>0.0382</td>
<td>0.1295</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.4504</td>
<td>-0.8437</td>
<td>-0.0043</td>
</tr>
<tr>
<td>25% quantile</td>
<td>-0.0187</td>
<td>-0.0141</td>
<td>0.0420</td>
</tr>
<tr>
<td>75% quantile</td>
<td>0.0371</td>
<td>0.0302</td>
<td>1.4216</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>1.4665</td>
<td>2.1669</td>
<td>1.4216</td>
</tr>
<tr>
<td>Min</td>
<td>-0.1514</td>
<td>-0.1529</td>
<td>-0.0828</td>
</tr>
<tr>
<td>Max</td>
<td>0.1580</td>
<td>0.1103</td>
<td>0.1580</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov p-value</td>
<td>0.08091</td>
<td>0.0972</td>
<td>0.1027</td>
</tr>
<tr>
<td>Correlation</td>
<td></td>
<td>0.8091</td>
<td>0.8866</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Model</th>
<th>Average $p$–value</th>
<th>Minimum $p$–value</th>
<th>Maximum $p$–value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clayton-Gumbel</td>
<td>0.5303</td>
<td>0.0901</td>
<td>0.8889</td>
</tr>
<tr>
<td>Clayton-MECC</td>
<td>0.5729</td>
<td>0.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>Gumbel-MECC</td>
<td>0.5684</td>
<td>0.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Model</th>
<th>Average log-likelihood</th>
<th>Number of parameters</th>
<th>Average AIC</th>
<th>Average cAIC</th>
<th>Average BIC</th>
<th>Average HQIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>1.8379</td>
<td>7</td>
<td>-6.5136</td>
<td>12.3242</td>
<td>-5.9660</td>
<td>-6.3227</td>
</tr>
<tr>
<td>Clayton</td>
<td>1.3092</td>
<td>7</td>
<td>-5.4564</td>
<td>13.3815</td>
<td>-4.9087</td>
<td>-5.2654</td>
</tr>
<tr>
<td>Gumbel</td>
<td>1.3586</td>
<td>7</td>
<td>-5.5550</td>
<td>13.2829</td>
<td>-5.0763</td>
<td>-5.3640</td>
</tr>
</tbody>
</table>

Table 5: Descriptive statistics of realized portfolio returns

<table>
<thead>
<tr>
<th>Unconstrained</th>
<th>Normal</th>
<th>Clayton</th>
<th>Gumbel</th>
<th>MECC</th>
<th>Short-sales constrained</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.9806</td>
<td>0.5646</td>
<td>1.1362</td>
<td>0.8655</td>
<td>1.0078</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.3538</td>
<td>2.8754</td>
<td>0.8711</td>
<td>1.4105</td>
<td>0.0393</td>
</tr>
<tr>
<td>Sharpe Ratio</td>
<td>2.7717</td>
<td>0.1963</td>
<td>1.3044</td>
<td>0.6136</td>
<td>25.6147</td>
</tr>
<tr>
<td>Skewness</td>
<td>1.1703</td>
<td>-5.5290</td>
<td>1.6482</td>
<td>-8.3289</td>
<td>-0.1321</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>8.2222</td>
<td>75.4105</td>
<td>17.6369</td>
<td>73.6448</td>
<td>0.2316</td>
</tr>
<tr>
<td>10% quantile</td>
<td>0.5652</td>
<td>0.4243</td>
<td>0.8247</td>
<td>0.8107</td>
<td>0.9609</td>
</tr>
<tr>
<td>90% quantile</td>
<td>1.2675</td>
<td>1.8191</td>
<td>1.6606</td>
<td>1.2372</td>
<td>1.0567</td>
</tr>
</tbody>
</table>

|               |        |         |        |      |                         |
| Mean                   | 0.9772 | 0.9440  | 1.1070 | 0.9154 | 1.0102                  |
| Std. Dev.              | 0.2862 | 0.3453  | 0.4926 | 1.1942 | 0.0496                  |
| Sharpe Ratio           | 3.4139 | 2.7337  | 2.2474 | 0.7666 | 20.3604                 |
| Skewness               | -0.4918 | -0.5886 | 3.3726 | -7.3603 | -0.3709                |
| Kurtosis               | 3.0483 | 2.8632  | 18.5277 | 64.0690 | 0.4386                 |
| 10% quantile           | 0.6803 | 0.5352  | 0.7202 | 0.8321 | 0.9561                  |
| 90% quantile           | 1.2205 | 1.2638  | 1.3313 | 1.1283 | 1.0781                  |

|               |        |         |        |      |                         |
| Mean                   | 0.9867 | 0.9987  | 1.0182 | 1.0142 | 1.0091                  |
| Std. Dev.              | 0.0914 | 0.1510  | 0.2261 | 0.7752 | 0.0483                  |
| Sharpe Ratio           | 10.7918 | 6.6140  | 4.5028 | 1.3083 | 20.9019                 |
| Skewness               | -0.3631 | -1.6466 | 0.5243 | 1.5470 | -0.5260                |
| Kurtosis               | 1.6121 | 15.4838 | 5.0600 | 26.9495 | 0.8780                 |
| 10% quantile           | 0.8681 | 0.8918  | 0.8404 | 0.8794 | 0.9540                  |
| 90% quantile           | 1.0886 | 1.1398  | 1.1534 | 1.1394 | 1.0665                  |

| Mean                   | 0.9963 | 0.9886  | 1.0115 | 0.9696 | 1.0109                  |
| Std. Dev.              | 0.0794 | 0.0752  | 0.1230 | 0.5810 | 0.0431                  |
| Sharpe Ratio           | 12.5485 | 13.1501 | 8.2236 | 1.6688 | 23.4544                |
| Skewness               | -2.4794 | -0.3464 | 1.4049 | 0.0377 | -0.4331                |
| Kurtosis               | 12.7335 | 1.7569  | 10.3033 | 12.0053 | 0.8359                 |
| 10% quantile           | 0.9333 | 0.8951  | 0.9029 | 0.7356 | 0.9671                  |
| 90% quantile           | 1.0789 | 1.0742  | 1.0799 | 1.1578 | 1.0655                  |
Table 6: Descriptive statistics of optimal portfolios: *Normal copula*

<table>
<thead>
<tr>
<th></th>
<th>RRA = 3</th>
<th>RRA = 7</th>
<th>RRA = 15</th>
<th>RRA = 25</th>
<th>RRA = 3</th>
<th>RRA = 7</th>
<th>RRA = 15</th>
<th>RRA = 25</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Growth</td>
<td>Value</td>
<td>Growth</td>
<td>Value</td>
<td>Growth</td>
<td>Value</td>
<td>Growth</td>
<td>Value</td>
</tr>
<tr>
<td>Unconstrained</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.2969</td>
<td>-1.0153</td>
<td>-0.6238</td>
<td>1.0488</td>
<td>0.3463</td>
<td>1.0581</td>
<td>-0.2063</td>
<td>1.0704</td>
</tr>
<tr>
<td>25% quantile</td>
<td>-2.5725</td>
<td>-3.7589</td>
<td>-3.5001</td>
<td>-2.3879</td>
<td>-0.9182</td>
<td>-0.9342</td>
<td>-0.8802</td>
<td>-0.7516</td>
</tr>
<tr>
<td>Median</td>
<td>0.7327</td>
<td>-0.2068</td>
<td>-0.9591</td>
<td>1.0881</td>
<td>0.8061</td>
<td>1.6245</td>
<td>-0.6423</td>
<td>1.9719</td>
</tr>
<tr>
<td>75% quantile</td>
<td>3.1520</td>
<td>3.0401</td>
<td>3.0608</td>
<td>3.3116</td>
<td>1.0077</td>
<td>2.9448</td>
<td>0.7017</td>
<td>2.8729</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short-sales constrained</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.4714</td>
<td>0.4047</td>
<td>0.4256</td>
<td>0.5035</td>
<td>0.5250</td>
<td>0.4107</td>
<td>0.4708</td>
<td>0.4761</td>
</tr>
<tr>
<td>Min</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0017</td>
</tr>
<tr>
<td>25% quantile</td>
<td>0.0884</td>
<td>0.0660</td>
<td>0.0341</td>
<td>0.0633</td>
<td>0.0676</td>
<td>0.0495</td>
<td>0.1523</td>
<td>0.1466</td>
</tr>
<tr>
<td>Median</td>
<td>0.6595</td>
<td>0.2626</td>
<td>0.4177</td>
<td>0.3040</td>
<td>0.6870</td>
<td>0.4096</td>
<td>0.4390</td>
<td>0.4916</td>
</tr>
<tr>
<td>75% quantile</td>
<td>0.7537</td>
<td>0.7364</td>
<td>0.8461</td>
<td>0.8943</td>
<td>0.8732</td>
<td>0.8429</td>
<td>0.7961</td>
<td>0.7501</td>
</tr>
<tr>
<td>Max</td>
<td>0.9837</td>
<td>0.9919</td>
<td>0.9742</td>
<td>0.9963</td>
<td>0.9757</td>
<td>0.9562</td>
<td>0.9684</td>
<td>0.9443</td>
</tr>
</tbody>
</table>

*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

<table>
<thead>
<tr>
<th></th>
<th>RRA = 3</th>
<th>RRA = 7</th>
<th>RRA = 15</th>
<th>RRA = 25</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Growth</td>
<td>Value</td>
<td>Growth</td>
<td>Value</td>
</tr>
<tr>
<td>Mean</td>
<td>-0.0431</td>
<td>-3.2130</td>
<td>-0.5601</td>
<td>-4.5784</td>
</tr>
<tr>
<td>Min</td>
<td>-56.6288</td>
<td>-234.1930</td>
<td>-24.3194</td>
<td>-426.3080</td>
</tr>
<tr>
<td>25% quantile</td>
<td>-0.9662</td>
<td>-0.8578</td>
<td>-0.9711</td>
<td>-0.8836</td>
</tr>
<tr>
<td>Median</td>
<td>0.0000</td>
<td>0.1789</td>
<td>-0.5256</td>
<td>0.2652</td>
</tr>
<tr>
<td>75% quantile</td>
<td>0.9644</td>
<td>0.9176</td>
<td>0.8800</td>
<td>0.9579</td>
</tr>
<tr>
<td>Max</td>
<td>70.1992</td>
<td>8.5971</td>
<td>27.4618</td>
<td>18.7443</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>RRA = 3</th>
<th>RRA = 7</th>
<th>RRA = 15</th>
<th>RRA = 25</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Growth</td>
<td>Value</td>
<td>Growth</td>
<td>Value</td>
</tr>
<tr>
<td>Mean</td>
<td>0.4176</td>
<td>0.4258</td>
<td>0.4099</td>
<td>0.4821</td>
</tr>
<tr>
<td>Min</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>25% quantile</td>
<td>0.0001</td>
<td>0.0227</td>
<td>0.0243</td>
<td>0.0543</td>
</tr>
<tr>
<td>Median</td>
<td>0.2854</td>
<td>0.2958</td>
<td>0.1989</td>
<td>0.4689</td>
</tr>
<tr>
<td>75% quantile</td>
<td>0.8659</td>
<td>0.8283</td>
<td>0.8885</td>
<td>0.9303</td>
</tr>
<tr>
<td>Max</td>
<td>0.9987</td>
<td>0.9928</td>
<td>0.9949</td>
<td>0.9975</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>RRA</th>
<th>3</th>
<th>7</th>
<th>15</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>p-value*</td>
<td>p-value**</td>
<td>p-value*</td>
<td>p-value**</td>
</tr>
<tr>
<td>Clayton vs. Normal</td>
<td>0.0000</td>
<td>1(^2)</td>
<td>0.1438</td>
<td>0.8561(^0)</td>
<td>0.0173</td>
</tr>
<tr>
<td>Gumbel vs. Normal</td>
<td>0.0000</td>
<td>1(^2)</td>
<td>0.1152</td>
<td>0.8847(^0)</td>
<td>0.0222</td>
</tr>
<tr>
<td>MECC vs. Normal</td>
<td>1(^1)</td>
<td>0.0000</td>
<td>1(^1)</td>
<td>0.0000</td>
<td>1(^1)</td>
</tr>
<tr>
<td>Gumbel vs. Clayton</td>
<td>0.1542</td>
<td>0.8457(^0)</td>
<td>0.6789(^0)</td>
<td>0.7445(^0)</td>
<td>0.2554</td>
</tr>
<tr>
<td>MECC vs. Clayton</td>
<td>1(^1)</td>
<td>0.0000</td>
<td>1(^1)</td>
<td>0.0000</td>
<td>1(^1)</td>
</tr>
<tr>
<td>MECC vs. Gumbel</td>
<td>1(^1)</td>
<td>0.0000</td>
<td>1(^1)</td>
<td>0.0000</td>
<td>1(^1)</td>
</tr>
</tbody>
</table>

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 \(\mu_i\). 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

<table>
<thead>
<tr>
<th></th>
<th>RRA</th>
<th>3</th>
<th>7</th>
<th>15</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>p-value</td>
<td>p-value</td>
<td>p-value</td>
<td>p-value</td>
</tr>
<tr>
<td>Bootstrapping reality check p-value</td>
<td>0.09161</td>
<td>0.09671</td>
<td>0.09651</td>
<td>0.95720</td>
<td></td>
</tr>
<tr>
<td>Nave p-value</td>
<td>0.09161</td>
<td>0.09671</td>
<td>0.09651</td>
<td>0.25883</td>
<td></td>
</tr>
<tr>
<td>Bootstrapping reality check p-value</td>
<td>0.09821</td>
<td>0.96740</td>
<td>0.97870</td>
<td>0.96260</td>
<td></td>
</tr>
<tr>
<td>Nave p-value</td>
<td>0.09821</td>
<td>0.26353</td>
<td>0.11841</td>
<td>0.25933</td>
<td></td>
</tr>
</tbody>
</table>

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 $\rho^+$ and $\rho^-$ represent the positive and negative exceedance correlations, respectively.
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 $\rho^+$ and $\rho^-$ 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). \( \rho \) is the Spearman’s rho, \( \nu_1, \nu_2, \kappa, \) and \( \phi \) 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.