Econometric Theory, **35**, 2019, 777–815.
doi:10.1017/S0266466618000270

DETECTING FINANCIAL DATA DEPENDENCE STRUCTURE BY AVERAGING MIXTURE COPULAS

GUANNAN LIU
School of Economics, Xiamen University

WEI LONG
Tulane University

XINYU ZHANG
Academy of Mathematics and Systems Science, Chinese Academy of Sciences

Qingdao University

QI LI
Texas A&M University
Capital University of Economics and Business

A mixture copula is a linear combination of several individual copulas that can be used to generate dependence structures not belonging to existing copula families. Because different pairs of markets may exhibit quite different dependence structures in empirical studies, mixture copulas are useful in modeling the dependence in financial data. Therefore, rather than selecting a single copula based on certain criteria, we propose using a model averaging approach to estimate financial data dependence structures in a mixture copula framework. We select weights (for averaging) by a J-fold Cross-Validation procedure. We prove that the model averaging estimator is asymptotically optimal in the sense that it minimizes the squared estimation loss. Our simulation results show that the model averaging approach outperforms some competing methods when the working mixture model is misspecified. Using 12 years of data on daily returns from four developed economies' stock indexes, we

We would like to thank the Editor, Peter C.B. Phillips, Co-Editor, Dennis Kristensen, and three anonymous referees for their insightful comments that greatly improved our article. We would also like to thank Yanqin Fan and Xiaohong Chen for their help during the revision process of this article. Liu's research is supported by the National Natural Science Foundation of China (Grant No. 71803160) and the Fundamental Research Funds for the Central Universities (project number 20720171061). Long's research is partially supported by the Carol Lavin Bernick Faculty Grants at Tulane University. Zhang and Li's research is partially supported by National Natural Science Foundation of China (projects 71522004, 11471324, and 71631008 for Zhang; 71722011 and 71601130 for Li). Address correspondence to Xinyu Zhang, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China; and Qingdao University, Qingdao, China; email: xinyu@amss.ac.cn.

© Cambridge University Press 2018

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

show that the model averaging approach more accurately estimates their dependence structures than some competing methods.

1. INTRODUCTION

Nelsen (2006) provided an introduction to the methodology of copulas, defined as "functions that join or couple multivariate distribution functions to their one-dimensional marginal distribution functions". Specifically, let $\boldsymbol{X} = (X_1, \ldots, X_p)^\top$ be a random vector and the respective marginal cumulative distribution functions be defined as F_i for $i \in \{1, \ldots, p\}$. Then there exists a copula $C : [0,1]^p \to [0,1]$ such that $\forall\ \boldsymbol{x} = (x_1, \ldots, x_p) \in \mathbb{R}^p$, $F(\boldsymbol{x}) = C\{F_1(x_1), \ldots, F_p(x_p)\}$ (see Sklar, 1959). Therefore, the copula is flexible: it does not constrain the selection of marginal distributions: one can couple various margins together via a copula to obtain a flexible distribution function. In this article we propose using a model averaging approach to estimate a mixture copula model, which is a linear combination of multiple individual copulas.

Copula models are used primarily to study dependence patterns among variables, e.g., the comovements among international equity markets. They have been applied extensively in a number of empirical studies. For example, estimating default correlations (Li, 2000); examining the difference in dependence structures between developed and developing economies (Chollete, Peña, and Lu, 2005; Chollete, Heinen, and Valdesogo, 2009; and Aloui, Aïssa, and Nguyen, 2011); studying structural breaks in exchange rates (Patton, 2006); examining financial contagions (Rodriguez, 2007); and investigating the cross-state housing prices during the subprime mortgage crisis (Zimmer, 2012). One fundamental issue for all of these studies is how to select an appropriate copula to satisfactorily describe the dependence structure among the variables under study. Almost all of the work mentioned above presumptively builds a candidate set and relies on certain statistical criteria to select one copula. Then, they use that selected copula to measure the degree of dependence among variables. For example, in studying the home price dependence among different states in the United States during the housing crisis, Zimmer (2012) used both the Bayesian Information Criterion (BIC) and the Vuong test. He showed that in that context, a Clayton–Gumbel mixture copula provides a better estimate of dependence structures than a Gaussian, Clayton, or Gumbel copula.[1]

In practice, there are many types of copulas. One can fit the data being studied to each existing copula family and then select the most appropriate copula. However, this strategy is not feasible because one can always create a new copula by making certain transformations to an existing copula.[2] Thus, most empirical

[1] Other methods include comparing which copula gives the largest log-likelihood function value. Interested readers are referred to Manner and Reznikova (2012), Patton (2012) and Fan and Patton (2014) for details.

[2] For example, Patton (2006) introduces a symmetrized Joe–Clayton copula by taking a particular Laplace transformation on the BB7 copula of Joe (1997).

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

researchers rely on several common copulas—for example, Gaussian, Clayton, and Gumbel—to construct their candidate sets. The argument for this practice is that the candidate set should be general enough to capture most of the possible dependence patterns in the real world. Even though this strategy is relatively easy to implement, it still has a cost: one needs to assume that the observations are generated from one of the copulas included in the candidate set. In other words, one selects the most "appropriate" copula from the candidate set based on some criteria, and then uses it to describe the dependence pattern and to evaluate the degree of dependence of the data under analysis. If this candidate set includes the true copula that generates the observations, then the estimating procedure just described should be effective and efficient. However, the true data generating copula model is usually unknown to researchers. In practice, it is highly likely that the constructed candidate set fails to include the true copula. Under such a circumstance, the selected candidate copula based on certain criteria (for example, by minimizing BIC) may fail to provide an adequate description of the true dependence structure.

To take advantage of different copula shapes, Chollete et al. (2005) and Hu (2006) introduced mixture copula models. In their analyses, a mixture copula is formulated as a weighted average of several individual copulas, with the weights constrained between 0 and 1 and the weights summing to 1. A mixture copula is more flexible than an individual copula because it nests several individual copulas with quite different dependence structures. As we shall see, a mixture copula can generate dependence structures that do not belong to any existing individual copula. By combining several widely used individual copulas, one can build a parsimonious yet flexible mixture copula to capture various dependence patterns in financial data, such as zero and nonzero tail dependence or symmetric and asymmetric tail dependence. In their analysis, Chollete et al. (2005) and Hu (2006) considered a mixture model that included the Gaussian, Gumbel, and rotated Gumbel copulas to evaluate the dependence structures among stock indexes in developed economies. They found strong left-tail dependence, because the weight associated with the rotated Gumbel copula tends to be nonzero. The Gumbel copula tends to be filtered out due to its small weight. They concluded that stock markets in developed economies tend to decrease simultaneously. This is consistent with Longin and Solnik (2001), who found that equity returns tend to take on joint negative extremes. In a more recent work, Cai and Wang (2014) introduced a penalized likelihood method to estimate weights and copula parameters simultaneously. They used the Smoothly Clipped Absolute Deviation (SCAD) method proposed by Fan and Li (2001) as the penalty. They further established the asymptotic result for their proposed estimator and used simulations to demonstrate that their proposed method yields satisfactory estimation results on weights and copula parameters. In other words, different dependence structures are captured well by their estimated mixture copulas via the penalized likelihood method.

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

This article contributes to the existing literature by providing a model averaging method to estimate mixture copula models: instead of choosing one appropriate copula model through certain criteria such as AIC or BIC, we first fit observations to each individual candidate copula separately. Sometimes all individual copulas lead to a poor fit. Therefore, we also fit the data to the composite of all the candidate copulas. We average over the estimates of each individual copula and their composite, then select the associated weights by minimizing a leave-one-group-out cross-validation criterion. This approach is similar to the Jackknife model averaging proposed by Hansen and Racine (2012). We obtain solutions through a standard application of quadratic programming technique, because the leave-one-group-out cross-validation criterion is a quadratic function of weights. Under certain regularity conditions, we prove that our model averaging estimator is asymptotically optimal in terms of achieving the infeasible lowest possible squared estimation loss. In other words, the distance (measured by the estimated squared loss) between the estimated mixture copula and the unknown true model is asymptotically minimized. This is extremely important when the working model is misspecified, i.e., when observations are generated from copulas that are not included in the model. Cai and Wang (2014) argued that when the working model is misspecified, their penalized likelihood method will select a (mixture) copula exhibiting a dependence pattern similar to the unknown true copula. For example, suppose we use a mixture copula model that contains the Gaussian and the rotated Gumbel copulas to fit a dataset generated from a mixture of the Gaussian and Clayton copulas. Because the rotated Gumbel copula also exhibits the left-tail dependence like the Clayton copula, the Cai and Wang (2014) method assigns a certain weight to the rotated Gumbel to guarantee that the left-tail dependence pattern is captured by the mixture copula model. Cai and Wang (2014) termed the chosen mixture copula selected by minimizing a penalized criterion function as the "best approximated" copula. The model averaging method provides a more general criterion for the *best approximated* copula when the working model is misspecified: the optimal mixture model is constructed to minimize its distance (the estimated squared loss) to the true model, so that it can best describe the dependence pattern among variables. Because applied researchers prefer parsimonious working models, the misspecification problem should be common. Therefore, the model averaging method is an alternative way to estimate mixture copula models and leads to the estimated squared loss being asymptotically minimized.

In the empirical section of this article, we use the model averaging method to study the dependence structures of the daily returns of equity indexes in four developed economies (the United States, the United Kingdom, Hong Kong and Japan). Our estimation results support the superiority of the model averaging method over some existing methods. Compared with the Cai and Wang (2014) penalized likelihood method, the standard copula selection method (such as BIC), and the quasi maximum likelihood estimation method (QMLE) of Hu (2006), our model averaging method exhibits the smallest estimated squared loss in out-of-

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

sample predictions. This empirical application suggests that the model averaging method is a useful tool in analyzing the dependence structures of stock markets and risk management.

The rest of this article is organized as follows. In Section 2, we introduce the definition of the mixture copula model. Section 3 describes how to implement the model averaging method in estimating a mixture copula model. In Section 4, we use simulations to compare estimated squared losses under the model averaging method, the Cai and Wang (2014) penalized likelihood method, the BIC method and the QMLE method. Section 5 presents a real data example. Section 6 concludes. Regularity conditions and the proof of the optimality of the model averaging method are presented in Appendix A.

2. MIXTURE COPULA MODELS: A BRIEF INTRODUCTION

Suppose we have a series of p-dimensional vectors of random variables $\{\mathbf{X}_t\}_{t=1}^T$, where p is a finite positive integer and $\mathbf{X}_t = \left(X_{t1}, \ldots, X_{tp}\right)^\top$. Let $F^0(\mathbf{x})$ and $f^0(\mathbf{x})$ be the joint distribution and the density function of $\mathbf{X}$ evaluated at $\mathbf{x} \in \mathcal{R}^p$, and $F_i^0(x_i)$ and $f_i^0(x_i)$ be the marginal distribution and the density function of X_i evaluated at $x_i \in \mathcal{R}$, respectively, where $1 \le i \le p$.

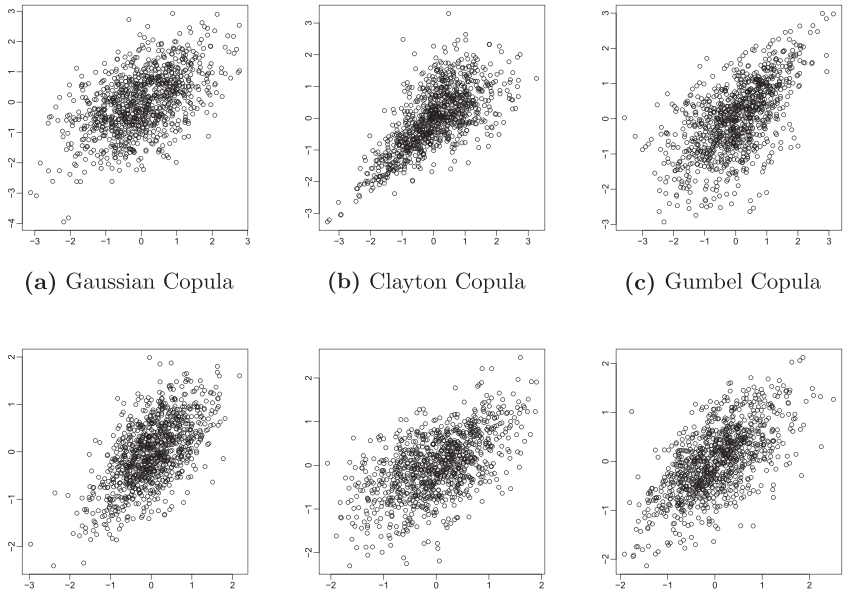
As in Hu (2006) and Cai and Wang (2014), a mixture copula model is a linear combination of several individual copulas. Specifically, a mixture copula model can be written as

$$C(\mathbf{u}; \boldsymbol{\theta}, \omega) = \sum_{k=1}^{L} \omega_k C_k(\mathbf{u}; \boldsymbol{\theta}_k) = \sum_{k=1}^{L} \omega_k C_k \left\{ F_1^0(x_1), \ldots, F_p^0(x_p); \boldsymbol{\theta}_k \right\}, \tag{1}$$

where $\{C_1(\cdot), \ldots, C_L(\cdot)\}$ is a set of candidate copulas with a vector of unknown parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_L^\top)^\top$ and a p-dimensional marginal distribution $\mathbf{u} = (F_1^0(\cdot), \ldots, F_p^0(\cdot))$. Let $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_L)^\top$ denote the weight parameters with $0 \le \omega_l \le 1$ and $\sum_{l=1}^L \omega_l = 1$. In equation (1), both the copula parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_L^\top)^\top$ and the weights $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_L)^\top$ control the shape of the mixture copula's dependence structure.

One may want to include many existing individual copulas into a mixture model to capture every possible dependence pattern. But in application this may lead to a complicated mixture model, and the computational burden of such a large mixture model can be heavy. In practice, a combination of several widely used copulas can well capture the dependence structure. For example, each panel in Figure 1 is a scatter-plot of an independently and identically distributed (i.i.d.) sample of size 1,000 generated from three types of widely used copulas—Gaussian, Clayton, and Gumbel. For each individual copula, its two marginal distributions are the standard normal distributions and its corresponding copula parameter is calibrated to imply a strong dependence with Kendall's $\tau = 0.5$. From Figure 1(a)–(c), it can be seen that the Clayton copula displays strong dependence at the left tail, whereas the Gumbel copula exhibits strong right tail dependence.

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press



(a) Gaussian Copula (b) Clayton Copula (c) Gumbel Copula

(d) Gaussian-Clayton Mixture (e) Gaussian-Gumbel Mixture (f) Clayton-Gumbel Mixture

FIGURE 1. Scatter plots of the Gaussian, Clayton, and Gumbel copulas and their mixture. All have standard normal margins and the Kendall's $\tau = 0.5$. For the mixture copulas, $\omega_k = 0.5$.

Unlike the Clayton and Gumbel copulas which exhibit asymmetric dependence structures, the Gaussian copula is symmetric and the stronger dependence appears at the center. Figure 1(d)–(f) present scatter plots from three mixture copulas with equal weights for each component. Figure 1 clearly demonstrates that a mixture of the Gaussian, Clayton, and Gumbel copulas can exhibit asymmetric tail dependence. Therefore, the flexibility of a mixture copula stems from its ability to nest various copula shapes.

For our model averaging method, besides estimating the parameters of each copula and the marginal densities, we also need to estimate a vector of weight parameters ($\boldsymbol{\omega}$) introduced by mixture copula models. Chollete et al. (2005) and Hu (2006) independently proposed a two-stage semiparametric method to estimate a mixture copula model. In the first stage, the marginal distributions are estimated nonparametrically to avoid the misspecification of marginals. In the second stage, the estimated marginals or the empirical CDFs are plugged into the copula and then the copula parameter is estimated by the maximum likelihood method. For a mixture copula model, to facilitate the estimation of weight parameters for each nested copula, an (iterative) EM algorithm is implemented. Cai and Wang (2014) provided the asymptotic results for mixture copula estimators. They pro-

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

posed a data-driven copula selection method via the penalized likelihood with a shrinkage operator so that parameter estimation and model selection are achieved simultaneously. In our framework, the model averaging method estimates a mixture copula based on a criterion function that minimizes an estimated squared loss function. In the next section, we discuss the estimation procedure and prove that the model averaging estimator is asymptotically optimal in the sense of minimizing the estimated squared loss.

3. THEORETICAL MODEL

We consider a class of Semiparametric COpula-based Multivariate DYnamic (SCOMDY) models proposed in Chen and Fan (2006b). Define $\{\mathbf{Y}_t^\top, \mathbf{Z}_t^\top\}_{t=1}^T$ as a vector stochastic process where $\mathbf{Y}_t$ is of dimension p, and $\mathbf{Z}_t$ is a vector of predetermined or exogenous variables. Denote $\mathcal{F}_{t-1}$ as the information set at time t, i.e., $\mathcal{F}_{t-1}$ is the sigma-field generated by $\{\mathbf{Z}_t, \mathbf{Z}_{t-1}, \ldots; \mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \ldots\}$. The class of SCOMDY models is specified as follows:

$$\mathbf{Y}_t = \mu_t(\beta_{01}) + \sqrt{H_t(\beta_0)}\varepsilon_t,$$

where $\mu_t(\beta_{01}) = E(\mathbf{Y}_t \mid \mathcal{F}_{t-1})$, and $H_t(\beta_0) = diag(h_{1,t}(\beta_0), \ldots, h_{p,t}(\beta_0))$, in which

$$h_{i,t}(\beta_0) = E[(Y_{it} - \mu_{it}(\beta_{01}))^2 \mid \mathcal{F}_{t-1}], \qquad i = 1, \ldots, p.$$

The unknown parameters β_{01} and β_0 are of fixed dimensions and $\beta_0 = (\beta_{01}^\top, \beta_{02}^\top)^\top$, where β_{01} and β_{02} do not have common elements. We further assume that the standardized innovations $\{\varepsilon_t \equiv (\varepsilon_{1t}, \ldots, \varepsilon_{pt})^\top\}$ are independent of $\mathcal{F}_{t-1}$. For each $i \in \{1, \ldots, p\}$, $\{\varepsilon_{it}\}_{t=1}^T$ are i.i.d. over t with $E(\varepsilon_{it}) = 0$ and $E(\varepsilon_{it}^2) = 1$. Moreover, different components of $\{\varepsilon_t \equiv (\varepsilon_{1t}, \ldots, \varepsilon_{pt})^\top\}$ are allowed to be contemporaneously correlated. That is, $\{\varepsilon_{it}, i = 1, \ldots, p; t = 1, \ldots, T\}$ are independent across the t-index, but we allow for contemporaneous dependence across the i-index. The SCOMDY models specified here can cover many commonly used specifications such as ARCH, GARCH, and vector autoregressions (VAR); see Chen and Fan (2006b) for a detailed discussion on this.

Our target is to estimate the joint distribution of ε_t based on a mixture copula model. Since $\{\varepsilon_t\}$ are unobservable, following Chen and Fan (2006b), we first estimate β_0 by a moment-based method, and thereby obtain estimator $\widehat{\beta} = (\widehat{\beta}_1^\top, \widehat{\beta}_2^\top)^\top$ and residuals $\{\widehat{\varepsilon}_t\} = [\mathbf{Y}_t - \mu_t(\widehat{\beta}_1)]/\sqrt{H_t(\widehat{\beta})}$. Given the residuals $\{\widehat{\varepsilon}_t\}$, we estimate the marginal distributions of ε_{it}, $i = 1, \ldots, p$, through the rescaled empirical distribution of the residuals:

$$\widetilde{F}_i(x_i) = \frac{1}{T+1}\sum_{t=1}^{T} I\{\widehat{\varepsilon}_{it} \le x_i\}, \quad i = 1, \ldots, p.$$

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

Suppose that we have $K-1$ individual copulas

$$C_k(\mathbf{u};\boldsymbol{\theta}_k) = C_k\left\{F_1^0(x_1),\ldots,F_p^0(x_p);\boldsymbol{\theta}_k\right\}, \qquad k=1,\ldots,K-1, \tag{2}$$

where $F_i^0(\cdot)$ is the true (but unknown) marginal distribution of ε_{it}, $i=1,\ldots,p$, $\mathbf{u}=(F_1^0(x_1),\ldots,F_p^0(x_p))$ is an arbitrary point in $[0,1]^p$, and $\boldsymbol{\theta}_k$ is a finite dimensional parameter associated with the k^{th} copula. In this article we assume that K is fixed, because K is often small in applications. We maintain the fixed K assumption and leave the theoretical investigation of allowing for $K\to\infty$ as $T\to\infty$ to future research.

When $\{\varepsilon_t\}_{t=1}^{\mathrm{T}}$ is thought to be generated from the k^{th} copula, we can estimate $C_k(\mathbf{u};\boldsymbol{\theta}_k)$ by the quasi maximum likelihood estimation (QMLE) method after replacing the unknown margins with the estimators, $\widetilde{F}_i(\cdot)$, $i=1,\ldots,p$, as stated in Chen and Fan (2006b). Let $\widetilde{\mathbf{u}}=(\widetilde{F}_1(x_1),\ldots,\widetilde{F}_p(x_p))$ and denote the resulting estimator as

$$C_k(\widetilde{\mathbf{u}};\widehat{\boldsymbol{\theta}}_k) = C_k\left\{\widetilde{F}_1(x_1),\ldots,\widetilde{F}_p(x_p);\widehat{\boldsymbol{\theta}}_k\right\}, \qquad k=1,\ldots,K-1. \tag{3}$$

We note that if the candidate set only consists of individual copulas while the true copula is a mixture of them and is not close to any of them, an averaging estimator based on these individual copulas can lead to a poorer performance than the QMLE based on a mixture copula. Therefore, we add the mixture copula to the model averaging copula candidate set, i.e., we add a mixture of the $K-1$ copulas to the candidate set. Consequently, our candidate set contains K copulas in total: the first $K-1$ are the individual copulas in which the unknown parameters are estimated by the QMLE method and the last one is a mixture of the first $K-1$ copulas with the parameters and the weights simultaneously estimated by the QMLE method. We denote C_K as the K^{th} copula. Let

$$C_K(\widetilde{\mathbf{u}};\widehat{\boldsymbol{\theta}}_K) = C_K\left\{\widetilde{F}_1(x_1),\ldots,\widetilde{F}_p(x_p);\widehat{\boldsymbol{\theta}}_K\right\} \equiv \sum_{k=1}^{K-1}\breve{\omega}_k C_k\left\{\widetilde{F}_1(x_1),\ldots,\widetilde{F}_p(x_p);\breve{\boldsymbol{\theta}}_k\right\},$$

where $\breve{\omega}_1,\ldots,\breve{\omega}_{K-1}$ and $\breve{\boldsymbol{\theta}}_1,\ldots,\breve{\boldsymbol{\theta}}_{K-1}$ are the QMLE estimates, and $\widehat{\boldsymbol{\theta}}_K \equiv (\breve{\omega}_1,\ldots,\breve{\omega}_{K-1},\breve{\boldsymbol{\theta}}_1^{\top},\ldots,\breve{\boldsymbol{\theta}}_{K-1}^{\top})^{\top}$. Here $\breve{\omega}_k$ is constrained to be between 0 and 1 and the summation equals 1. For each individual copula k, $\breve{\boldsymbol{\theta}}_k$ also has its own constraint. For example, the parameter (i.e., the correlation coefficient) for a Gaussian copula should be between -1 and 1. We use the rescaled empirical distributions of the residuals to replace the unknown marginals, and use the QMLE method to estimate the finite dimensional parameters.

Let $C_0(\mathbf{u};\boldsymbol{\theta}_0) = C_0\{F_1^0(x_1),\ldots,F_p^0(x_p);\boldsymbol{\theta}_0\}$ be the true copula. Note that $C_0(\mathbf{u};\boldsymbol{\theta}_0)$ can be outside the set $\{C_1(\mathbf{u};\boldsymbol{\theta}_1),\ldots,C_{K-1}(\mathbf{u};\boldsymbol{\theta}_{K-1})\}$ and it may

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

not be a mixture copula based on $\{C_1(\mathbf{u};\boldsymbol{\theta}_1),\ldots,C_{K-1}(\mathbf{u};\boldsymbol{\theta}_{K-1})\}$. The goal of this article is to approximate $C_0(\mathbf{u};\boldsymbol{\theta}_0)$ by the model averaging approach. Write $\mathbf{w}=(w_1,\ldots,w_K)^\top$ as a weight vector belonging to the set[3]

$$\mathcal{W}=\left\{\mathbf{w}\in[0,1]^K:\sum\nolimits_{k=1}^{K}w_k=1\right\}.$$

Then, the model averaging method is to use the following weighted average of all candidate copulas to approximate the true unknown copula

$$C(\widetilde{\mathbf{u}};\widehat{\boldsymbol{\theta}},\mathbf{w})=\sum_{k=1}^{K}w_kC_k(\widetilde{\mathbf{u}};\widehat{\boldsymbol{\theta}}_k), \tag{4}$$

where $\widehat{\boldsymbol{\theta}}=\left(\widehat{\boldsymbol{\theta}}_1^\top,\ldots,\widehat{\boldsymbol{\theta}}_K^\top\right)^\top$.

We would like to make a comment on the model averaging estimator above. In this article we will omit the $(\widetilde{F}_1(x_1),\ldots,\widetilde{F}_p(x_p))$ argument in copula functions for brevity. The K^{th} copula is a "mixture" of the first $(K-1)$ individual copulas which are, respectively, estimated as $\sum_{k=1}^{K-1}\breve{\omega}_kC_k\left(\breve{\boldsymbol{\theta}}_k\right)$, where $\breve{\omega}_k$ and $\breve{\boldsymbol{\theta}}_k$, $k=1,\ldots,K-1$, are the QMLE estimates of ω_k and $\boldsymbol{\theta}_k$ based on the mixture copula model. At first inspection, one may get the impression that the K^{th} copula is redundant because it looks like a linear combination of the first $K-1$ copulas. However, this is not the case because $\breve{\boldsymbol{\theta}}_k$ differs from $\widehat{\boldsymbol{\theta}}_k$ in general. Hence, the K^{th} mixture copula is not a linear combination of the first $K-1$ copulas. In fact, simulations (not reported here to save space) show that by adding the K^{th} mixture copula to the candidate set, the model averaging estimator performs much better than an estimator that only averages over the first $K-1$ copulas.

A crucial question in the model averaging estimation is how to select the weight $\mathbf{w}$. In this article we use a J-fold ($J>1$ is a finite positive integer) Cross-Validation (CV) method to choose weights, which is similar to the Jackknife model averaging method (Hansen and Racine, 2012). Specifically, we divide the data set into J groups such that for each group, we have $M=T/J$ observations except that the last group may have more data (if T/J is not an integer). In the j^{th} group, we have observations $\widehat{\varepsilon}_{(j-1)M+1},\ldots,\widehat{\varepsilon}_{jM}$ for $j=1,\ldots,J$. Write $C_k(\widetilde{\mathbf{u}};\widehat{\boldsymbol{\theta}}_k^{(-j)})$ as the estimator of $C_k(\mathbf{u};\boldsymbol{\theta}_k)$ with the j^{th} group removed from the sample, i.e., it is the same as $C_k(\widetilde{\mathbf{u}};\widehat{\boldsymbol{\theta}}_k)$ except that one drops the j^{th} group in the

[3] Note that in Ando and Li (2014), $\sum_{k=1}^{K}w_k$ is not restricted to unity when selecting weights, but in other recent model averaging literature such as Cheng and Hansen (2015) and Zhang, Yu, Zou, and Liang (2016), the $\sum_{k=1}^{K}w_k=1$ restriction is imposed. In this article, we keep this restriction because less restriction on weights can make it harder to find the optimal weights. In Ando and Li (2014), the candidate models have no overlapping regressors, but in the current article, the first $K-1$ copulas are nested in the K^{th} copula, so using this restriction is appropriate for our method. In Appendix A, we compare model averaging methods with and without this restriction. We find that the out-of-sample prediction error with the restriction is smaller than that without imposing the restriction, but the difference is not statistically significant.

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

QMLE. Let $\widehat{\boldsymbol{\theta}}^{(-j)} = \left(\widehat{\boldsymbol{\theta}}_1^{(-j)\top}, \ldots, \widehat{\boldsymbol{\theta}}_K^{(-j)\top}\right)^\top$. Then similar to $C(\widetilde{\mathbf{u}}; \widehat{\boldsymbol{\theta}}, \mathbf{w})$ defined in (4), we define the weighted averaging estimator $C(\widetilde{\mathbf{u}}; \widehat{\boldsymbol{\theta}}^{(-j)}, \mathbf{w})$ with the j^{th} group excluded as

$$C(\widetilde{\mathbf{u}}; \widehat{\boldsymbol{\theta}}^{(-j)}, \mathbf{w}) = \sum_{k=1}^{K} w_k C_k(\widetilde{\mathbf{u}}; \widehat{\boldsymbol{\theta}}_k^{(-j)}), \qquad j = 1, \ldots, J.$$

An empirical estimator of $C_0(\mathbf{u}; \boldsymbol{\theta}_0)$ that only uses the j^{th} group data is denoted by

$$\widetilde{C}_{(j)}(\mathbf{x}) = \frac{1}{M} \sum_{m=1}^{M} I(\widehat{\varepsilon}_{(j-1)M+m} \leq \mathbf{x}), \tag{5}$$

where $\mathbf{x} = (x_1, \ldots, x_p)$ is an arbitrary point in $\mathcal{R}^p$, $I(\cdot)$ is an indicator function, and the comparison between $\widehat{\varepsilon}_{(j-1)M+m}$ and $\mathbf{x}$ means comparison componentwise in the p-dimensional vector. We emphasize that the superscript $(-j)$ denotes "leave the j^{th} group data out" and the subscript (j) means "only use the j^{th} group data".

Define $U_{0t} \equiv (F_1^0(\varepsilon_{1t}), \ldots, F_p^0(\varepsilon_{pt}))$ and $\widetilde{U}_t \equiv (\widetilde{F}_1(\widehat{\varepsilon}_{1t}), \ldots, \widetilde{F}_p(\widehat{\varepsilon}_{pt}))$, then the J-fold CV criterion is given by

$$CV_J(\mathbf{w}) = \sum_{j=1}^{J} \sum_{m=1}^{M} \left\{ C(\widetilde{U}_{(j-1)M+m}; \widehat{\boldsymbol{\theta}}^{(-j)}, \mathbf{w}) - \widetilde{C}_{(j)}(\widehat{\varepsilon}_{(j-1)M+m}) \right\}^2. \tag{6}$$

The weight $\mathbf{w}$ is selected via

$$\widehat{\mathbf{w}} = \text{argmin}_{\mathbf{w} \in \mathcal{W}} CV_J(\mathbf{w}),$$

and we estimate $C_0(\mathbf{u}; \boldsymbol{\theta}_0)$ by the model averaging estimator $C(\widetilde{\mathbf{u}}; \widehat{\boldsymbol{\theta}}, \widehat{\mathbf{w}})$ as defined in (4) with $\widehat{\mathbf{w}}$ replacing $\mathbf{w}$.

For ease of exposition, we summarize the notations used in the article. The $T \times 1$ vector of the true copula evaluated at $(\varepsilon_1, \ldots, \varepsilon_T)$ is denoted by

$$\mathbf{C}_0 = \{C_0(U_{01}; \boldsymbol{\theta}_0), \ldots, C_0(U_{0T}; \boldsymbol{\theta}_0)\}^\top. \tag{7}$$

For $k = 1, \ldots, K$, the $T \times 1$ vector of the copula estimated by the k^{th} candidate copula, using all observations (when $k = K$, the K^{th} copula is a composite of the candidate copulas) evaluated at $(\widehat{\varepsilon}_1, \ldots, \widehat{\varepsilon}_T)$ is

$$\widehat{\mathbf{C}}_k = \left\{C_k(\widetilde{U}_1; \widehat{\boldsymbol{\theta}}_k), \ldots, C_k(\widetilde{U}_T; \widehat{\boldsymbol{\theta}}_k)\right\}^\top. \tag{8}$$

The $T \times 1$ vector of the weighted average of the estimated candidate copulas $\left\{\widehat{\mathbf{C}}_1, \ldots, \widehat{\mathbf{C}}_K\right\}$ evaluated at $(\widehat{\varepsilon}_1, \ldots, \widehat{\varepsilon}_T)$ is denoted by

$$\widehat{\mathbf{C}}(\mathbf{w}) = \sum_{k=1}^{K} w_k \widehat{\mathbf{C}}_k = \left\{C(\widetilde{U}_1; \widehat{\boldsymbol{\theta}}, \mathbf{w}), \ldots, C(\widetilde{U}_T; \widehat{\boldsymbol{\theta}}, \mathbf{w})\right\}^\top. \tag{9}$$

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

The leave-M-out, $T \times 1$ vector of the copula estimated by using the k^{th} candidate copula evaluated at $(\widehat{\varepsilon}_1, \ldots \widehat{\varepsilon}_T)$ is

$$\overline{\mathbf{C}}_k = \left\{ C_k(\widetilde{U}_1; \widehat{\boldsymbol{\theta}}_k^{(-1)}), \ldots, C_k(\widetilde{U}_M; \widehat{\boldsymbol{\theta}}_k^{(-1)}), \ldots, C_k(\widetilde{U}_{T-M+1}; \widehat{\boldsymbol{\theta}}_k^{(-J)}), \ldots, C_k(\widetilde{U}_T; \widehat{\boldsymbol{\theta}}_k^{(-J)}) \right\}^\top. \tag{10}$$

The $T \times 1$ vector of the weighted average of $\overline{\mathbf{C}}_1, \ldots, \overline{\mathbf{C}}_K$ evaluated at $(\widehat{\varepsilon}_1, \ldots, \widehat{\varepsilon}_T)$ is

$$\begin{aligned} \overline{\mathbf{C}}(\mathbf{w}) &= \sum_{k=1}^{K} w_k \overline{\mathbf{C}}_k \\ &= \left\{ C(\widetilde{U}_1; \widehat{\boldsymbol{\theta}}^{(-1)}, \mathbf{w}), \ldots, C(\widetilde{U}_M; \widehat{\boldsymbol{\theta}}^{(-1)}, \mathbf{w}), \ldots, C(\widetilde{U}_T; \widehat{\boldsymbol{\theta}}^{(-J)}, \mathbf{w}) \right\}^\top. \end{aligned} \tag{11}$$

The $T \times 1$ vector of the rescaled empirical distribution function using M observations evaluated at $(\widehat{\varepsilon}_1, \ldots, \widehat{\varepsilon}_T)$ is

$$\widetilde{\mathbf{C}} = \left\{ \widetilde{C}_{(1)}(\widehat{\varepsilon}_1), \ldots, \widetilde{C}_{(1)}(\widehat{\varepsilon}_M), \ldots, \widetilde{C}_{(J)}(\widehat{\varepsilon}_{T-M+1}), \ldots, \widetilde{C}_{(J)}(\widehat{\varepsilon}_T) \right\}^\top. \tag{12}$$

Finally, the vector of the model averaging estimator of $\mathbf{C}_0$ evaluated at $(\widehat{\varepsilon}_1, \ldots, \widehat{\varepsilon}_T)$ is

$$\widehat{\mathbf{C}}(\widehat{\mathbf{w}}) = \left\{ C(\widetilde{U}_1; \widehat{\boldsymbol{\theta}}, \widehat{\mathbf{w}}), \ldots, C(\widetilde{U}_T; \widehat{\boldsymbol{\theta}}, \widehat{\mathbf{w}}) \right\}^\top. \tag{13}$$

Note that in general we use the "hat" notation to denote estimators that are based on semiparametric estimation methods (because the margins are estimated nonparametrically). We use the "tilde" notation to denote nonparametric (empirical function based) estimators.

PROPOSITION 1

(i). Under Conditions C.1–C.4 presented in Appendix A.1, the bias of the J-fold CV criterion is small relative to the expected CV squared loss. That is,

$$\sup_{\mathbf{w} \in \mathcal{W}} \frac{E \left\| \overline{\mathbf{C}}(\mathbf{w}) - \mathbf{C}_0 \right\|^2 - E \left\| \overline{\mathbf{C}}(\mathbf{w}) - \widetilde{\mathbf{C}} \right\|^2}{E \left\| \overline{\mathbf{C}}(\mathbf{w}) - \mathbf{C}_0 \right\|^2} = o(1).$$

(ii). Under an i.i.d. situation, $\{\varepsilon_{it}, i = 1, \ldots, p; t = 1, \ldots, T\}$ are independent across the t-index; the J-fold CV criterion is an unbiased estimator of the expected CV squared loss plus a term unrelated to $\mathbf{w}$. That is,

$$E \left\| \overline{\mathbf{C}}(\mathbf{w}) - \widetilde{\mathbf{C}} \right\|^2 = E \left\| \overline{\mathbf{C}}(\mathbf{w}) - \mathbf{C}_0 \right\|^2 + \{\text{a term unrelated to } \mathbf{w}\}.$$

See Appendix A.3 for the proof of Proposition 1. Let $\breve{\mathbf{h}}_k = \overline{\mathbf{C}}_k - \widetilde{\mathbf{C}}$ and $\breve{\mathbf{H}} = (\breve{\mathbf{h}}_1, \ldots, \breve{\mathbf{h}}_K)$. We can rewrite the J-fold CV criterion as

$$CV_J(\mathbf{w}) = \left\| \overline{\mathbf{C}}(\mathbf{w}) - \widetilde{\mathbf{C}} \right\|^2 = \mathbf{w}^\top \breve{\mathbf{H}}^\top \breve{\mathbf{H}} \mathbf{w},$$

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

which is quadratic in $\mathbf{w}$. Hence, the minimization of $CV_J(\mathbf{w})$ with respect to $\mathbf{w}$ can be implemented easily.

Define a quadratic loss function of the model averaging estimator as

$$L_T(\mathbf{w}) = \|\widehat{\mathbf{C}}(\mathbf{w}) - \mathbf{C}_0\|^2. \tag{14}$$

As in the literature on model selection and model averaging, for example Shao (1997) and Hansen (2007), our goal is to reduce the quadratic loss by using the model averaging method. The following theorem shows that our method minimizes the quadratic loss asymptotically.

THEOREM 1. *Under Conditions C.1–C.4 presented in Appendix A.1,*

$$\frac{L_T(\widehat{\mathbf{w}})}{\inf_{\mathbf{w}\in\mathcal{W}} L_T(\mathbf{w})} \to 1 \qquad \text{in probability} \quad (\text{as } T \to \infty). \tag{15}$$

The detailed proof of Theorem 1 is presented in Appendix A. Theorem 1 states that the model averaging estimator $\widehat{\mathbf{C}}(\widehat{\mathbf{w}})$ is asymptotically optimal: it achieves the infeasible lowest possible squared estimation loss. In addition, it is obvious that $\inf_{\mathbf{w}\in\mathcal{W}} L_T(\mathbf{w}) \le \inf_{k\in\{1,\ldots,K\}} \|\widehat{\mathbf{C}}_k - \mathbf{C}_0\|^2$, so we expect that the model averaging method can reduce estimation errors relative to using an individual candidate copula.

Next, we discuss how to choose appropriate values of K and J. For K, we use the Cai and Wang (2014) method to filter out candidates with zero weights. Then, we implement the proposed model averaging method using candidates preserved after the screening. For J, we follow the method proposed by Ma and Zhu (2012). Specifically, we choose J from a discrete set consisting of l candidates $J \in \{J_1, J_2, \ldots, J_l\}$. For each $s \in \{1, \ldots, l\}$, we use the block bootstrap method with a block size m to generate a bootstrap sample. We then repeat the process B times to obtain B bootstrap samples: $[\{\mathbf{Y}_t^\top, \mathbf{Z}_t^\top\}_{t=1}^T]^{[1]}, \ldots, [\{\mathbf{Y}_t^\top, \mathbf{Z}_t^\top\}_{t=1}^T]^{[B]}$. Using the model averaging procedure discussed above, we can calculate $\widehat{\mathbf{w}}_1^1, \ldots, \widehat{\mathbf{w}}_B^1, \ldots, \widehat{\mathbf{w}}_1^l, \ldots, \widehat{\mathbf{w}}_B^l$. For candidate J_s, we can calculate the sample variance matrix using $\widehat{\mathbf{w}}_1^s, \ldots, \widehat{\mathbf{w}}_B^s$. We denote it as V_{J_s} for $s = 1, \ldots, l$. Finally, we select the J^* that yields the minimum value of $trace(V_J)$, i.e., $J^* = \arg\min_{J\in\{J_1,\ldots,J_l\}} trace(V_J)$. We term this approach the data-driven method for selecting J. In our simulations, we choose $m = 10$ and $B = 100$.

4. NUMERICAL STUDIES

We compare the squared estimation losses obtained from the proposed model averaging method with three other existing methods: the Cai and Wang (2014) penalized likelihood method, the BIC method, which selects one copula by comparing each candidate's BIC value, and the QMLE method of Hu (2006). Two

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

types of simulations are used. In Type I simulations, the data are generated from the correctly specified mixture copula model. In Type II simulations, the working mixture copula model is misspecified. In other words, the data are generated from copulas excluded from the working mixture copula model.

4.1. Type I Simulations

In Type I simulations, we assume that the data are generated from copulas that are constituents of our working model. First, we consider a bivariate case where the data are generated by an AR(1)-GARCH(1,1) process:

$$y_{it} = \gamma_i y_{i,t-1} + e_{it}, \qquad i = 1, 2;\ t = 1, \ldots, T,$$

and $\gamma_1 = 0.05$, $\gamma_2 = 0.1$, $e_{it} = \sigma_{it}\epsilon_{it}$, ϵ_{it} has a standard normal marginal distribution with a dependence structure between ϵ_{1t} and ϵ_{2t}, governed by a given form of copula function. Furthermore,

$$\sigma_{it} = \sqrt{\alpha_{i0} + \alpha_{i1} e_{i,t-1}^2 + \beta_{i1}\sigma_{i,t-1}^2},$$

where $\alpha_{10} = 0.0001$, $\alpha_{11} = 0.04$, $\beta_{11} = 0.95$ for the first margin; and $\alpha_{20} = 0.0001$, $\alpha_{21} = 0.09$, $\beta_{21} = 0.90$ for the second margin. Our working model includes three commonly used copulas: the Gaussian, Clayton, and Gumbel copulas. The characteristics of these three copulas were discussed in Section 2, and the simulated scatter plots are displayed in Figure 1. We first use the Cai and Wang (2014) method to screen the three candidate copulas. Then, we decide which copula should be kept in the candidate set. Suppose that all three candidates have nonzero weights under the Cai and Wang (2014) method. The presumed mixture copula is

$$C(\boldsymbol{u}; \boldsymbol{\theta}, \boldsymbol{\omega}) = \omega_{Ga} C_{Ga}(\boldsymbol{u}; \boldsymbol{\theta}_1) + \omega_{Cl} C_{Cl}(\boldsymbol{u}; \boldsymbol{\theta}_2) + \omega_{Gu} C_{Gu}(\boldsymbol{u}; \boldsymbol{\theta}_3),$$

where C_{Ga}, C_{Cl} and C_{Gu} stand for the Gaussian, the Clayton, and the Gumbel copulas, respectively, and $\boldsymbol{u} = (F_1(\cdot), F_2(\cdot))$ denotes the two marginals. By fitting the data into the Gaussian, Clayton, and Gumbel copulas separately, we obtain their maximum likelihood estimates $\widehat{\boldsymbol{\theta}}_1$, $\widehat{\boldsymbol{\theta}}_2$, and $\widehat{\boldsymbol{\theta}}_3$.

In Section 3 we argued that when the true model is a mixture copula, an averaging estimator based on an individual copula leads to a poor fit. For that reason, we include a linear combination of the three individual copulas in our working copula candidate set. Specifically, let $C_{QMLE}(\tilde{\boldsymbol{u}}; \widehat{\boldsymbol{\theta}}_4) = \breve{\omega}_1 C_{Ga}(\tilde{\boldsymbol{u}}; \breve{\boldsymbol{\theta}}_1) + \breve{\omega}_2 C_{Cl}(\tilde{\boldsymbol{u}}; \breve{\boldsymbol{\theta}}_2) + \breve{\omega}_3 C_{Gu}(\tilde{\boldsymbol{u}}; \breve{\boldsymbol{\theta}}_3)$, where $\tilde{\boldsymbol{u}} = (\tilde{F}_1(\cdot), \tilde{F}_2(\cdot))$ is a rescaled empirical CDF estimate of $\boldsymbol{u}$, and $\breve{\omega}_1, \breve{\omega}_2, \breve{\omega}_3$ and $\breve{\boldsymbol{\theta}}_1, \breve{\boldsymbol{\theta}}_2, \breve{\boldsymbol{\theta}}_3$ are the QMLE of $\omega_1, \omega_2, \omega_3$ and $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3$, respectively. Let $\widehat{\boldsymbol{\theta}}_4 = (\breve{\omega}_1, \breve{\omega}_2, \breve{\omega}_3, \breve{\boldsymbol{\theta}}_1^\top, \breve{\boldsymbol{\theta}}_2^\top, \breve{\boldsymbol{\theta}}_3^\top)^\top$. Our method then averages the four components: three individual candidate copulas plus a QMLE of their linear combination. We choose $w_{Ga}, w_{Cl}, w_{Gu}, w_{QMLE}$ in the working

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

mixture copula model

$$C(\tilde{u},\tilde{v};\widehat{\boldsymbol{\theta}},\mathbf{w}) = w_{Ga}C_{Ga}(\tilde{u},\tilde{v};\widehat{\boldsymbol{\theta}}_1) + w_{Cl}C_{Cl}(\tilde{u},\tilde{v};\widehat{\boldsymbol{\theta}}_2) + w_{Gu}C_{Gu}(\tilde{u},\tilde{v};\widehat{\boldsymbol{\theta}}_3) + w_{QMLE}C_{QMLE}(\tilde{u},\tilde{v};\widehat{\boldsymbol{\theta}}_4)$$

to minimize the squared estimation loss $L_T(\mathbf{w})$ as defined in Section 3. For expositional simplicity, the model averaging method with the data-driven method to select K and J is denoted as MA^*.

We consider two sample sizes, $T = 200$ and 500, and focus on out-of-sample forecasting evaluations among the different estimation methods with the in-sample data size equal to the out-of-sample data size. In other words, for $T = 200$ (500), we use a sample of 200 (500) to estimate a model, and then evaluate the estimated squared loss using another 200 (500) out-of-sample observations. We consider four different values for $J \in \{2,4,5,10\}$. For each J, we use the block bootstrap size $m = 10$ to generate $B = 100$ bootstrap samples. We then use the method discussed at the end of Section 3 to select J^*. All the simulations are repeated 2000 times. We simulate three mixture copulas with two components, and an additional mixture copula with three components. Specifically, we consider the following four cases for weights:

$$\text{Case 1:}\quad \omega_{Ga} = 1/2, \omega_{Cl} = 1/2, \omega_{Gu} = 0;$$

$$\text{Case 2:}\quad \omega_{Ga} = 1/2, \omega_{Cl} = 0, \omega_{Gu} = 1/2;$$

$$\text{Case 3:}\quad \omega_{Ga} = 0, \omega_{Cl} = 1/2, \omega_{Gu} = 1/2;$$

$$\text{Case 4:}\quad \omega_{Ga} = 1/3, \omega_{Cl} = 1/3, \omega_{Gu} = 1/3\ .$$

For each case of these weights, we consider two sets of copula parameters:

$$\text{Parameter setting 1:}\quad \theta_{Ga} = 0.5, \theta_{Cl} = 5.8, \theta_{Gu} = 5.1;$$

$$\text{Parameter setting 2:}\quad \theta_{Ga} = 0.7, \theta_{Cl} = 7.8, \theta_{Gu} = 7.1.$$

Therefore, we have $4 \times 2 = 8$ DGPs in total.

Table 1 displays the out-of-sample mean squared estimation losses across the four methods: the proposed model average method (MA^*); the Cai and Wang (2014) penalized likelihood (CW); the BIC method, which selects one from a set of candidates based on BIC values (BIC); and the QMLE method of Hu (2006). To save space, we only present the ratios of the out-of-sample prediction errors of CW to MA^* (CW/MA^*), BIC to MA^* (BIC/MA^*), and QMLE to MA^* (QMLE/MA^*). Hence, MA^* is superior to CW, BIC, and QMLE if these ratios are all greater than one. To further check whether such superiority is statistically significant, we implement the Diebold–Mariano (DM) test and report the p-value of the DM test for each case.

The results in Table 1 show that the mixture copula models outperforms the models using an individual copula (BIC). Among the four methods, the BIC

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

TABLE 1. Ratios of squared prediction losses for Type I simulations. The data are generated from linear combinations of the Gaussian, Clayton, and Gumbel copulas with the associated weights denoted by $\omega_{Ga}, \omega_{Cl}, \omega_{Gu}$. $\theta_{Ga}, \theta_{Cl}, \theta_{Gu}$ represent the copula parameters for the Gaussian, Clayton, and Gumbel copulas. MA^* represents the model averaging method with data-driven method selected J and K. CW represents the Cai and Wang (2014) penalized maximum likelihood. BIC represents the baseline approach based on the BIC comparisons. $QMLE$ represents the quasi maximum likelihood estimation method. Each simulation is repeated 2,000 times. Values in parentheses indicate the p-values obtained through the Diebold–Mariano Test

Type I simulations: Out-of-sample forecast (sample size = 200)						
	$\theta_{Ga}=0.5, \theta_{Cl}=5.8, \theta_{Gu}=5.1$			$\theta_{Ga}=0.7, \theta_{Cl}=7.8, \theta_{Gu}=7.1$		
	CW/MA^*	$QMLE/MA^*$	BIC/MA^*	CW/MA^*	$QMLE/MA^*$	BIC/MA^*
$\omega_{Ga}=0.5, \omega_{Cl}=0.5, \omega_{Gu}=0$	0.9931	1.0821	1.2043	0.9943	1.0526	1.1832
	(0.61)	(0.08)	(0.00)	(0.23)	(0.11)	(0.01)
$\omega_{Ga}=0.5, \omega_{Cl}=0, \omega_{Gu}=0.5$	0.9831	1.0574	1.1937	0.9922	1.0169	1.1491
	(0.59)	(0.11)	(0.01)	(0.63)	(0.87)	(0.04)
$\omega_{Ga}=0, \omega_{Cl}=0.5, \omega_{Gu}=0.5$	0.9951	1.0643	1.1739	0.9816	1.0713	1.1756
	(0.49)	(0.13)	(0.02)	(0.41)	(0.07)	(0.01)
$\omega_{Ga}=1/3, \omega_{Cl}=1/3, \omega_{Gu}=1/3$	1.0022	1.0754	1.1928	0.9983	1.0983	1.1749
	(0.64)	(0.07)	(0.01)	(0.89)	(0.05)	(0.00)
Type I simulations: Out-of-sample forecast (sample size = 500)						
	$\theta_{Ga}=0.5, \theta_{Cl}=5.8, \theta_{Gu}=5.1$			$\theta_{Ga}=0.7, \theta_{Cl}=7.8, \theta_{Gu}=7.1$		
	CW/MA^*	$QMLE/MA^*$	BIC/MA^*	CW/MA^*	$QMLE/MA^*$	BIC/MA^*
$\omega_{Ga}=0.5, \omega_{Cl}=0.5, \omega_{Gu}=0$	1.0006	1.0679	1.1923	0.9926	1.0538	1.1749
	(0.71)	(0.08)	(0.00)	(0.83)	(0.12)	(0.01)
$\omega_{Ga}=0.5, \omega_{Cl}=0, \omega_{Gu}=0.5$	1.0135	1.0365	1.2173	1.0219	1.0766	1.2207
	(0.19)	(0.14)	(0.00)	(0.14)	(0.09)	(0.00)
$\omega_{Ga}=0, \omega_{Cl}=0.5, \omega_{Gu}=0.5$	0.9908	1.0749	1.1697	0.9877	1.0621	1.2134
	(0.91)	(0.08)	(0.02)	(0.16)	(0.13)	(0.00)
$\omega_{Ga}=1/3, \omega_{Cl}=1/3, \omega_{Gu}=1/3$	0.9945	1.0693	1.2158	0.9833	1.0421	1.2036
	(0.54)	(0.09)	(0.00)	(0.67)	(0.12)	(0.00)

method yields the largest prediction errors for all cases (BIC/MA^* is greater than one and is larger than CW/MA^* and QMLE/MA^*). The performances of the CW, QMLE and MA^* are similar because the ratios of CW/MA^* and QMLE/MA^* are all close to one, and the p-values from the DM test are large in most cases.

In terms of the out-of-sample prediction error, the results obtained from the Type I simulations exhibit the superiority of the mixture copula model. Furthermore, for estimation of the mixture copula model, the model averaging method performs similarly to the Cai and Wang (2014) method and the QMLE method.

4.2. Type II Simulations

In the Type II simulations, the working model is misspecified. The Type II simulations examine how the proposed model averaging method performs when the data are generated from copulas not included in the candidate set, which is common in empirical studies when the true model is unknown to researchers.

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

Similar to the Type I simulations, the constructed working mixture model in Type II simulations also contains the Gaussian, Clayton and Gumbel copulas. However, the observations are generated from a linear combination of the Frank, Survival Joe (SJ), and Joe copulas. These three copulas are also widely used in empirical studies. The Frank copula is similar to the Gaussian copula in that it does not exhibit tail dependence, but it instead has relatively stronger dependence in the center of the distribution. The Joe copula, like the Gumbel copula, exhibits right tail dependence. The Survival Joe copula is a 180° rotation of the Joe copula and like the Clayton copula exhibits left tail dependence. In the first three cases, the data are generated from a linear combination of two candidate copulas. In the fourth case, the data are generated using all three candidate copulas with equal weights. That is, we consider the following four cases for weights:

$$\text{Case 1:}\quad \omega_{Frank} = 1/2, \omega_{SJ} = 1/2, \omega_{Joe} = 0;$$

$$\text{Case 2:}\quad \omega_{Frank} = 1/2, \omega_{SJ} = 0, \omega_{Joe} = 1/2;$$

$$\text{Case 3:}\quad \omega_{Frank} = 0, \omega_{SJ} = 1/2, \omega_{Joe} = 1/2;$$

$$\text{Case 4:}\quad \omega_{Frank} = 1/3, \omega_{SJ} = 1/3, \omega_{Joe} = 1/3,$$

and we use the following two sets of copula parameters:

$$\text{Parameter setting 1:}\quad \theta_{Frank} = 5.5, \theta_{SJ} = 4.8, \theta_{Joe} = 4.5;$$

$$\text{Parameter setting 2:}\quad \theta_{Frank} = 7.5, \theta_{SJ} = 6.8, \theta_{Joe} = 6.5.$$

Table 2 presents the out-of-sample prediction results of the four competing methods under Type II simulations. The number of out-of-sample observations is the same as the number of in-sample observations. Table 2 shows that MA^* yields more accurate out-of-sample predictions than CW, BIC, and QMLE , because the ratios of CW to MA^*, BIC to MA^*, and QMLE to MA^* are all greater than one. Therefore, the model averaging method outperforms CW, BIC, and QMLE when the working mixture copula is misspecified. This finding has important implications for empirical studies because the true copula is always unknown to researchers and in practice the misspecification of a working mixture model is common.

Next, we compare the performances of the four competing methods by extending the simple 2-dimensional copula to 3- and 4-dimensional copulas. We generate the 3rd and the 4th marginal distributions via

$$y_{it} = \gamma_i y_{i,t-1} + e_{it}, \qquad i = 3,4;\ t = 1,\ldots,T,$$

where $\gamma_3 = 0.09$, $\gamma_4 = 0.15$, $e_{it} = \sigma_{it}\epsilon_{it}$, ϵ_{it} has a standard normal marginal distribution and the dependence structure among ϵ_{1t}, ϵ_{2t}, ϵ_{3t} and ϵ_{4t} is governed by a given form of copula function. Furthermore,

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

TABLE 2. Ratios of squared prediction losses for Type II simulations. The data are generated from linear combinations of the Frank, Survival Joe, and Joe copulas with the associated weights denoted by $\omega_F, \omega_{SJ}, \omega_J$. $\theta_F, \theta_{SJ}, \theta_J$ represent the copula parameters for the Frank, Survival Joe, and Joe copulas. MA^* represents the model averaging method with data-driven method selected J and K. represents the Cai and Wang (2014) penalized maximum likelihood. BIC represents the baseline approach based on the BIC comparisons. $QMLE$ represents the quasi maximum likelihood estimation method. Each simulation is repeated 2000 times. Values in parentheses indicate the p-values obtained through the Diebold–Mariano Test

Type II simulations: Out-of-sample forecast (sample size = 200)						
	$\theta_F = 5.5, \theta_{SJ} = 4.8, \theta_J = 4.5$			$\theta_F = 7.5, \theta_{SJ} = 6.8, \theta_J = 6.5$		
	CW/MA^*	$QMLE/MA^*$	BIC/MA^*	CW/MA^*	$QMLE/MA^*$	BIC/MA^*
$\omega_F = 0.5, \omega_{SJ} = 0.5, \omega_J = 0$	1.0801	1.1728	1.2159	1.0739	1.1642	1.2207
	(0.06)	(0.01)	(0.00)	(0.07)	(0.02)	(0.00)
$\omega_F = 0.5, \omega_{SJ} = 0, \omega_J = 0.5$	1.0933	1.1324	1.2283	1.0845	1.1175	1.1934
	(0.05)	(0.05)	(0.00)	(0.06)	(0.03)	(0.01)
$\omega_F = 0, \omega_{SJ} = 0.5, \omega_J = 0.5$	1.1291	1.1467	1.1908	1.1022	1.1236	1.1968
	(0.04)	(0.04)	(0.00)	(0.05)	(0.04)	(0.01)
$\omega_F = 1/3, \omega_{SJ} = 1/3, \omega_J = 1/3$	1.1327	1.1055	1.1997	1.1237	1.0978	1.2114
	(0.04)	(0.04)	(0.01)	(0.04)	(0.05)	(0.00)
Type II simulations: Out-of-sample forecast (sample size = 500)						
	$\theta_F = 5.5, \theta_{SJ} = 4.8, \theta_J = 4.5$			$\theta_F = 7.5, \theta_{SJ} = 6.8, \theta_J = 6.5$		
	CW/MA^*	$QMLE/MA^*$	BIC/MA^*	CW/MA^*	$QMLE/MA^*$	BIC/MA^*
$\omega_F = 0.5, \omega_{SJ} = 0.5, \omega_J = 0$	1.1187	1.2001	1.2567	1.1304	1.2213	1.2265
	(0.04)	(0.01)	(0.01)	(0.05)	(0.01)	(0.00)
$\omega_F = 0.5, \omega_{SJ} = 0, \omega_J = 0.5$	1.1531	1.2093	1.2317	1.1613	1.2764	1.2618
	(0.03)	(0.00)	(0.00)	(0.02)	(0.00)	(0.00)
$\omega_F = 0, \omega_{SJ} = 0.5, \omega_J = 0.5$	1.1664	1.2231	1.2504	1.1562	1.2376	1.2703
	(0.02)	(0.00)	(0.00)	(0.02)	(0.00)	(0.00)
$\omega_F = 1/3, \omega_{SJ} = 1/3, \omega_J = 1/3$	1.1981	1.1944	1.2576	1.1683	1.2247	1.2629
	(0.00)	(0.00)	(0.00)	(0.02)	(0.00)	(0.00)

,

$$\sigma_{it} = \sqrt{\alpha_{i0} + \alpha_{i1} e_{i,t-1}^2 + \beta_{i1} \sigma_{i,t-1}^2},$$

where $\alpha_{30} = 0.0001$, $\alpha_{31} = 0.05$ and $\beta_{31} = 0.94$ for the third marginal; and $\alpha_{40} = 0.0001$, $\alpha_{41} = 0.08$ and $\beta_{41} = 0.91$ for the fourth marginal. The remaining pieces of this setup are similar to the 2-dimensional copula case discussed earlier. Both the in-sample (estimation) size and the out-of-sample (prediction) size are $T = 500$. For brevity, we only conduct Type II simulations for the 3-dimensional copula and the 4-dimensional copula. Our working mixture model still contains the Gaussian, Clayton, and Gumbel copulas, but the observations are generated from a linear combination of the Frank, Survival Joe (SJ), and Joe copulas.

Table 3 presents the ratios of the mean squared out-of-sample prediction losses of the model averaging method to those of the three competing methods for the 3-dimensional copula and 4-dimensional copula cases. If the ratio is greater than 1, it implies that MA^* outperforms the competing methods. As docu-

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

TABLE 3. Ratios of squared out-of-sample prediction losses for Type II simulations with the 3-dimensional copula and the 4-dimensional copula. The data are generated from linear combinations of the Frank, Survival Joe, and Joe copulas with the associated weights denoted by $\omega_F, \omega_{SJ}, \omega_J$. $\theta_F, \theta_{SJ}, \theta_J$ represent the copula parameters for the Frank, Survival Joe, and Joe copulas. MA^* represents the model averaging method with data-driven method selected J and K. CW represents the Cai and Wang (2014) penalized maximum likelihood. BIC represents the baseline approach based on the BIC comparisons. $QMLE$ represents the quasi maximum likelihood estimation method. Each simulation is repeated 2000 times. Values in parentheses indicate the p-values obtained through the Diebold–Mariano Test

3-component copula: Out-of-sample forecast (sample size = 500)						
	$\theta_F = 5.5, \theta_{SJ} = 4.8, \theta_J = 4.5$			$\theta_F = 7.5, \theta_{SJ} = 6.8, \theta_J = 6.5$		
	CW/MA^*	$QMLE/MA^*$	BIC/MA^*	CW/MA^*	$QMLE/MA^*$	BIC/MA^*
$\omega_F = 0.5, \omega_{SJ} = 0.5, \omega_J = 0$	1.1631	1.3859	2.0968	1.1954	1.2207	2.5491
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$\omega_F = 0.5, \omega_{SJ} = 0, \omega_J = 0.5$	1.1927	1.2157	1.9804	1.1368	1.2544	1.8093
	(0.01)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)
$\omega_F = 0, \omega_{SJ} = 0.5, \omega_J = 0.5$	1.1179	1.1038	1.4353	1.1277	1.1184	1.3965
	(0.05)	(0.04)	(0.00)	(0.03)	(0.04)	(0.00)
$\omega_F = 1/3, \omega_{SJ} = 1/3, \omega_J = 1/3$	1.0915	1.1179	1.4843	1.1302	1.1408	1.3951
	(0.04)	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)
4-component copula: Out-of-sample forecast (sample size = 500)						
	$\theta_F = 5.5, \theta_{SJ} = 4.8, \theta_J = 4.5$			$\theta_F = 7.5, \theta_{SJ} = 6.8, \theta_J = 6.5$		
	CW/MA^*	$QMLE/MA^*$	BIC/MA^*	CW/MA^*	$QMLE/MA^*$	BIC/MA^*
$\omega_F = 0.5, \omega_{SJ} = 0.5, \omega_J = 0$	1.1326	1.2605	2.9317	1.1429	1.2063	2.8970
	(0.07)	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)
$\omega_F = 0.5, \omega_{SJ} = 0, \omega_J = 0.5$	1.1018	1.1916	1.4381	1.0988	1.1543	1.3806
	(0.05)	(0.00)	(0.00)	(0.04)	(0.00)	(0.00)
$\omega_F = 0, \omega_{SJ} = 0.5, \omega_J = 0.5$	1.1942	1.1139	1.4026	1.2237	1.1435	1.4109
	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)	(0.00)
$\omega_F = 1/3, \omega_{SJ} = 1/3, \omega_J = 1/3$	1.1079	1.1834	1.3072	1.1406	1.2267	1.3869
	(0.03)	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)

mented in Table 3, MA^* outperforms the other three methods in all considered cases.

In Appendix A, we run additional simulations to compare the performances of: (i) the original model averaging method with fixed K and J (no selections of K and J; denoted as MA hereafter) and MA without the $\sum_{k=1}^{K} w_k = 1$ constraint; (ii) MA and MA with the screening step to select K (MAS); (iii) MA with the proposed data-driven method selecting J and MA with $J \in \{2, 4, 5, 10\}$. The simulation results are documented in Appendix A. Table A1 reveals that the out-of-sample predictions between MA and MA without the $\sum_{k=1}^{K} w_k = 1$ constraint are not significantly different from each other because all the ratios are close to 1. Table A2 demonstrates that even though the ratio of MAS to MA is less than 1, the difference between the two methods is not statistically significant according to the DM test. This implies that the additional screening step does not substantially improve the performance of MA. Table A3 shows that MA with the data-driven method selecting J significantly outperforms MA with a fixed J in most cases.

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

This demonstrates the utility of our proposed J selection strategy in improving the performance of MA.

5. AN EMPIRICAL STUDY

We use actual data to examine how the model averaging method performs in estimating a mixture copula model. Specifically, we look at the daily returns of Morgan Stanley Capital International (MSCI) equity indexes for four developed economies: the United States (US), the United Kingdom (UK), Hong Kong (HK), and Japan (JP). These daily data span 12 years, from August 2002 to December 2014, with a total of 3,220 observations. We download these equity indexes from Datastream and calculate their log returns. For comparison purposes, we convert the currency for the non-U.S. daily indexes into U.S. dollars, based on their respective contemporaneous exchange rates.

We then split the data into two equal parts: we use the first 1,610 observations (training set), ranging from August 2002 to October 2008, to fit a model, and the remaining 1,610 observations (testing set) to examine the out-of-sample prediction accuracy across the competing methods. Table 4 displays the summary statistics for the daily log-returns of the MSCI indexes of the four markets. Over the six years between 2002 and 2008, the HK market had the highest average daily return while the UK market had the highest median daily return. The skewness was negative for all markets, indicating higher probabilities of having extreme daily losses than extreme daily gains. For the US, UK, and HK markets, the kurtosis was greater than 3; it was less than 3 for the JP market. These statistics indicate that in practice it can be difficult to correctly specify each marginal distribution and that nonparametric methods should be used to estimate the marginals.

Table 5 presents the linear correlation coefficients and the Kendall's τ (in parentheses) for each pair. We observe that the HK-JP pair has the strongest correlation, or degree of dependence, based on both the correlation coefficients and the

TABLE 4. The summary statistics for daily log-returns of the MSCI Indexes in the United States, United Kingdom, Hong Kong, and Japan markets

	US	UK	HK	JP
Mean	0.0186	0.0084	0.0295	0.0070
Median	0.0307	0.0492	0.0117	0.0174
Min	−9.2002	−10.494	−9.0137	−7.3300
Max	5.2785	9.2065	10.192	4.2476
S.D.	1.0752	1.2243	1.3534	1.3763
Skewness	−0.7713	−0.4470	−0.1229	−0.4189
Kurtosis	7.7132	8.1975	7.3170	1.7011

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

TABLE 5. The linear correlation coefficients and the Kendall's τs (Kendall's τs are in parentheses) across the United States, United Kingdom, Hong Kong, and Japan markets

	UK	JP	HK
US	0.4304 (0.271)	0.0274 (0.037)	0.1147 (0.072)
UK		0.1995 (0.140)	0.3753 (0.209)
JP			0.4880 (0.306)

Kendall's τ. Figure 2 shows the pairwise scatter plots. For each pair, daily returns appear to be positively correlated, especially for the US-UK, JP-HK, and UK-HK pairs. Figure 2 also displays a violation of the elliptical multivariate distributions, because asymmetry can be observed in all of the pairs. Figure 2 further confirms the existence of a large number of extreme observations at the lower left corner for the UK-HK, JP-HK, and JP-UK pairs. Simply choosing the most "appropriate" copula from a candidate set may not reveal the characteristics of the joint distribution. To take advantage of the flexibility of each individual copula, we consider a mixture copula model.

Hu (2006) demonstrated that data with conditional heteroscedasticity can lead to underestimation of the degree of dependence because of the clustering of large volatilities. Our preliminary examination indicates the existence of both autocorrelation and conditional heteroscedasticity in the daily returns for the four economies. To filter both effects out, we specify an AR(1)-GARCH(1,1) model, a special case of the class of SCOMDY models presented in Section 3. The filtered daily percentage changes (the residuals) are substituted into the working mixture copula model. According to Theorem 1, the model averaging method generates an asymptotically optimal estimate that minimizes the squared estimation loss.

Next, we fit the filtered daily returns in the four markets into a mixture copula model that includes the Student's t, Clayton, and Gumbel copulas.[4] We use the model averaging method, the Cai and Wang (2014) penalized likelihood method, the BIC method, and the QMLE method. To compare the performances across the four methods, we follow a procedure suggested by Genest and Rivest (1993). Specifically, we construct five 7×7 cross-classification tables for each of the six pairs. For each pair, the values in the first cross-classification table are computed using the observed data, while the other four tables are computed using the four competing methods. Let G represent a table and $G_{i,j}$ be the value in the cell (i, j) (the ith row and jth column of table G) for a pair of markets (call them markets

[4] We thank a referee for suggesting the use of the Student's t copula. Compared with the Gaussian copula, the Student's t copula is more suitable for financial data because it has the tail dependence property and is able to capture correlations in the extreme market movements.

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

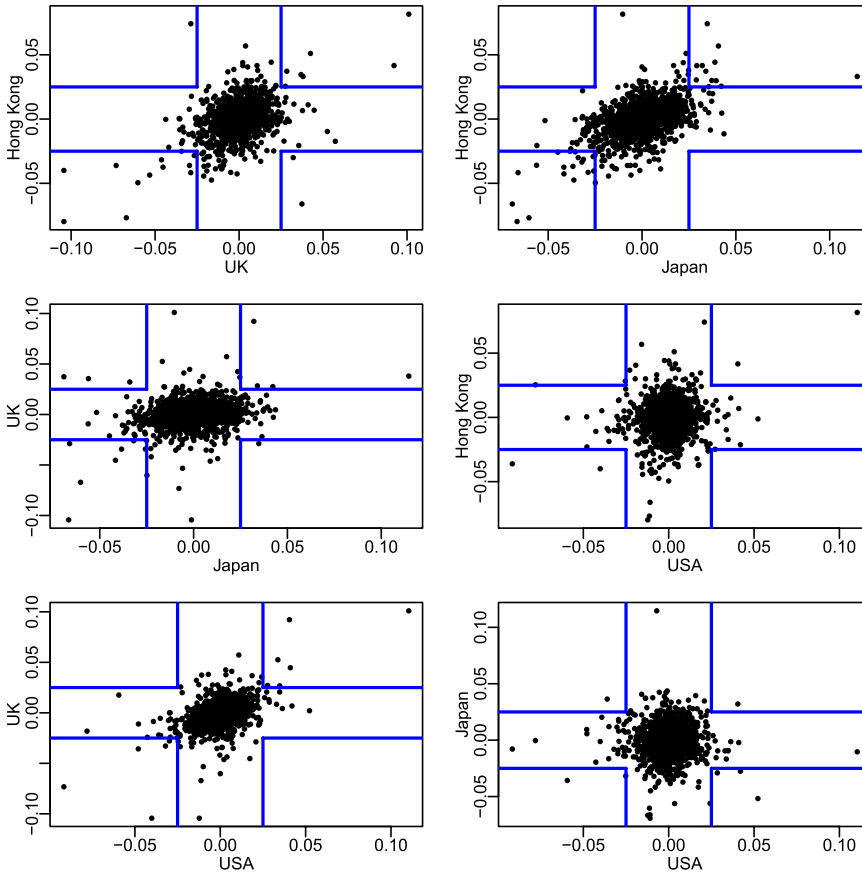


FIGURE 2. Scatter plots for daily returns of the MSCI Indexes.

1 and 2), where $i, j = 1, \ldots, 7$. Let u_i and v_j be the upper bounds for determining counts $G_{i,j}$ in the cell (i, j), where u_i and v_j are defined as the $i/7$ and $j/7$ percentiles of the two return series, respectively $(i, j = 1, \ldots, 7)$. Then a pair of observations (u, v) belongs to the cell (i, j) if $u_{i-1} < u \leq u_i$ and $v_{j-1} < v \leq v_j$. Thus, $G_{i,j}$ represents the number of times the daily return of market 1 falls between the $(i-1)/7$ and the $i/7$ percentile of its data range, and at the same time, the daily return of market 2 is within the $(j-1)/7$ and the $j/7$ percentile of its data range. For example, the number recorded in the cell $(3, 2)$ indicates the number of times that the daily return percentage changes in the first market are between the 29th (2/7) and the 43rd (3/7) percentile of its data range, while those of the second market are within the 14th (1/7) and 29th (2/7) percentile of its data range. If the two markets are positively correlated, we should find that most observations lie on the principal diagonal. If they are negatively correlated, most observations should lie on the diagonal that is perpendicular to the principal

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

one. If they are independent of each other, the number of observations in different cells should be close to each other.

We take the UK-HK pair as an example. For the observed frequencies, the cell at the top left represents the number of times that indexes in the UK and Hong Kong markets were both below the 14th (1/7) percentile of their respective ranges. In other words, the number of times when both markets simultaneously faced downturn risks. Similarly, the cell at the bottom right shows the frequency of both daily returns being above the 86th (6/7) percentile. During the period between August 2002 and October 2008 (a total of 1,610 observations), there were 83 times when the daily returns of the MSCI indexes in the UK and Hong Kong markets were both below their 14th percentile. By contrast, there were 64 times when the daily returns in the UK and Hong Kong markets were both above their 86th percentile. Thus, the markets in the UK and Hong Kong exhibit a higher probability of downward comovements than upward comovements. The US-HK, UK-JP, and JP-HK markets display similar left-tail dependence structures. This pattern was not significant in the US-JP and US-UK markets.

Next we evaluate the fit of an estimation method by comparing the sum of the squared differences between the estimated count and the observed count over all of the $7 \times 7 = 49$ cells. Here we describe how to compute the estimated count in each cell. Let $\widehat{C}$ denote an estimated copula. The estimated count in the cell (i, j), for example, can be obtained by multiplying the probability $\widehat{C}(u_i, v_j) - \widehat{C}(u_i, v_{j-1}) - \widehat{C}(u_{i-1}, v_j) + \widehat{C}(u_{i-1}, v_{j-1})$ to the sample size, 1,610. Let $G_{i,j}$ and $G_{i,j}^{MA^*}$, respectively, denote the observed frequency and the estimated frequency, according to the model averaging method in the cell (i, j). Then we define the estimated squared error as

$$Q_{MA^*} = \frac{1}{k^2} \sum_{i=1}^{k} \sum_{j=1}^{k} (G_{i,j} - G_{i,j}^{MA^*})^2. \tag{16}$$

In our case, $k = 7$. Similarly, for the CW, BIC, and QMLE methods, we can calculate the estimated squared errors and denote them as Q_{CW}, Q_{BIC} and Q_{QMLE}, respectively. The estimated squared errors and the ratios are documented in Table 6. In all cases, the model averaging method exhibits the smallest estimated squared errors, while the BIC method yields the largest estimated squared errors. Again, this empirical example shows that the model average method yields more accurate estimates for dependence structures than other competing methods.

We now consider the out-of-sample prediction performances of the four competing methods based on daily observations between October 2008 and December 2014. The estimated frequencies are based on our estimated mixture copula model under MA^*, CW, BIC, and QMLE, respectively. The out-of-sample prediction errors are calculated similarly to the in-sample estimation losses. For each pair, we present the prediction squared errors of each competing method and the ratios of CW to MA^*, BIC to MA^*, and QMLE to MA^*. Our results are shown in

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

TABLE 6. Means of the in-sample estimation errors based on MA^*, CW, BIC, and $QMLE$. MA^* represents the model averaging method with data-driven method selected J and K. CW represents the Cai and Wang (2014) penalized maximum likelihood. BIC represents the baseline approach based on the BIC comparisons. $QMLE$ represents the quasi maximum likelihood estimation method

	Q_{MA^*}	Q_{CW}	Q_{BIC}	Q_{QMLE}	Q_{CW}/Q_{MA^*}	Q_{BIC}/Q_{MA^*}	Q_{QMLE}/Q_{MA^*}
US-UK	17.59	22.09	28.32	21.13	1.256	1.610	1.201
US-HK	31.17	35.91	36.17	34.49	1.152	1.160	1.107
US-JP	27.12	29.24	29.61	28.74	1.078	1.092	1.060
UK-HK	30.19	33.77	35.43	33.16	1.119	1.174	1.098
UK-JP	20.04	25.21	25.58	24.41	1.258	1.276	1.218
HK-JP	29.17	31.41	35.04	33.11	1.077	1.201	1.135

Table 7. It is obvious that the model averaging method exhibits the smallest squared prediction errors, indicating its superior performance in out-of-sample prediction.

Note that the dependent structures among different markets may change drastically over time, especially during periods when global financial markets experience substantial fluctuation. Ideally, one should allow the parameters in a mixture model to change over time to capture the dynamics among the international financial markets. Although some approaches to time-varying copulas have been developed (see, for example, Patton, 2006 and Manner and Reznikova, 2012), extending the semiparametric (with nonparametric marginals) model averaging method to a time-varying copula framework is challenging (e.g., Fan and Patton, 2014). Therefore, we leave that for future research. To deal in part with the time-varying nature of the dependence structures among the returns from the four developed markets, we use a rolling window method to examine the performances of MA^*, CW, BIC, and QMLE. Specifically, we first estimate MA^*,

TABLE 7. Means of the out-of-sample prediction errors based on MA^*, CW, BIC, and $QMLE$. MA^* represents the model averaging method with data-driven method selected J and K. CW represents the Cai and Wang (2014) penalized maximum likelihood. BIC represents the baseline approach based on the BIC comparisons. $QMLE$ represents the quasi maximum likelihood estimation method

	Q_{MA^*}	Q_{CW}	Q_{BIC}	Q_{QMLE}	Q_{CW}/Q_{MA^*}	Q_{BIC}/Q_{MA^*}	Q_{QMLE}/Q_{MA^*}
US-UK	112.376	115.403	128.514	119.027	1.027	1.144	1.059
US-HK	42.3517	44.8726	44.1107	44.6314	1.060	1.042	1.054
US-JP	19.0015	20.6804	22.0029	22.6441	1.088	1.158	1.192
UK-HK	30.4812	32.9507	35.1903	34.6621	1.081	1.154	1.137
UK-JP	19.0735	22.2811	21.9532	23.2734	1.168	1.151	1.220
HK-JP	33.7658	35.1526	36.4483	36.0025	1.041	1.079	1.066

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

TABLE 8. Means of the out-of-sample prediction errors of MA^*, CW, BIC, and $QMLE$ based on the rolling window method (fixed window width). MA^* represents the model averaging method with data-driven method selected J and K. CW represents the Cai and Wang (2014) penalized maximum likelihood. BIC represents the baseline approach based on the BIC comparisons. $QMLE$ represents the quasi maximum likelihood estimation method

	$Q^{rw}_{MA^*}$	Q^{rw}_{CW}	Q^{rw}_{BIC}	Q^{rw}_{QMLE}	$Q^{rw}_{CW}/Q^{rw}_{MA^*}$	$Q^{rw}_{BIC}/Q^{rw}_{MA^*}$	$Q^{rw}_{QMLE}/Q^{rw}_{MA^*}$
US-UK	32.2715	35.9321	39.2237	38.0735	1.113	1.215	1.180
US-HK	27.0634	28.9637	30.2680	30.1159	1.070	1.118	1.113
US-JP	14.5301	16.5926	17.1104	16.6603	1.142	1.178	1.147
UK-HK	17.1756	20.2027	21.3544	20.1138	1.176	1.243	1.171
UK-JP	17.0035	21.0038	21.1917	20.5104	1.235	1.246	1.206
HK-JP	34.1824	35.1027	39.0723	38.8361	1.027	1.143	1.136

CW, BIC, and QMLE based on observations from 2002 to 2008, and then predict the dependence structures among the US, UK, JP, and HK markets in 2009. Next, we re-estimate MA^*, CW, BIC, and QMLE by using the observations from 2003 to 2009 and then predict the dependence in 2010. We repeat this procedure and make the final prediction about the dependence structures in 2014 using observations from 2007 to 2013. In other words, we update the mixture model by using the most recent six years' observations. We define the squared estimation error of MA^* based on the rolling window method (RW) as

$$Q^{rw}_{MA^*} = \frac{1}{k^2}\sum_{i=1}^{k}\sum_{j=1}^{k}\sum_{l=2009}^{2014}(G_{l,i,j} - G^{MA^*}_{l,i,j})^2,$$

and in our case, $k = 7$, where $G_{l,i,j}$ and $G^{MA^*}_{l,i,j}$ are the frequency observed and frequency estimated by the model averaging method in the cell (i, j) for year l. In the same manner, we calculate the estimated squared errors for CW, BIC, and QMLE methods, which are, respectively, denoted as Q^{rw}_{CW}, Q^{rw}_{BIC} and Q^{rw}_{QMLE}. The estimated squared errors and the ratios are presented in Table 8.

We have two observations from Table 8. First, the model averaging method still outperforms CW, BIC, and QMLE in terms of squared estimation loss based on the rolling window method. Second, the estimated losses in Table 8 are smaller than the results in Table 7 (except for the HK-JP pair). This is expected, because we can improve prediction accuracy by updating the copula model using the rolling window method.

Finally, we compare the performances of the four methods, MA^*, CW, BIC, and QMLE by including all four markets in the 4-dimensional copula model. The calculation procedure is the same as for the 2-dimensional copula. However, when calculating the squared prediction errors, we do not split the data into the 7^4 cross-classification table because this will lead to insufficient observations in each cell. Instead, we consider the results from the 2^4, 3^4, and 4^4 cross-classification

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

TABLE 9. Means of the out-of-sample prediction errors based on MA^*, CW, BIC, and $QMLE$ with a 4-dimensional mixture copula. The mixture copula includes the Student's t, Clayton, and Gumbel copulas. The first row considers $2^4 = 16$ cells, the second row considers $3^4 = 81$ cells and the third row considers $4^4 = 256$ cells. MA^* represents the model averaging method with data-driven method selected J and K. CW represents the Cai and Wang (2014) penalized maximum likelihood. BIC represents the baseline approach based on the BIC comparisons. $QMLE$ represents the quasi maximum likelihood estimation method

	Q_{MA^*}	Q_{CW}	Q_{BIC}	Q_{QMLE}	Q_{CW}/Q_{MA^*}	Q_{BIC}/Q_{MA^*}	Q_{QMLE}/Q_{MA^*}
US-UK-HK-JP (2^4)	39.0027	47.2095	154.5832	45.3945	1.210	3.963	1.164
US-UK-HK-JP (3^4)	48.4601	55.1538	163.2781	61.1347	1.138	3.369	1.262
US-UK-HK-JP (4^4)	41.1866	44.7911	149.4092	47.0136	1.088	3.628	1.141

tables, respectively. For example, for the 2^4 cross-classification table, we order the returns of each index and divide them into two parts: one part is below the median return (we use 1 to denote this) and the other is above the median return (we use 2 to denote this). Then we can label the $2^4 = 16$ cells as (i_1, i_2, i_3, i_4) for $i_j = 1, 2$ and $j = 1, 2, 3, 4$. The cell (1,1,1,1) denotes the case where all four indexes are below their respective median returns. Similarly, (1,1,2,2) denotes the case where the first two indexes (say, the United States and the United Kingdom) are below their medians while the last two indexes (say, Hong Kong and Japan) are above their medians. Let G_{i_1,i_2,i_3,i_4} and $G^{MA^*}_{i_1,i_2,i_3,i_4}$ denote the frequency observed and the frequency estimated by the model averaging method in the cell (i_1, i_2, i_3, i_4), respectively. Then, we define the estimated squared error as

$$Q_{2^4,MA^*} = \frac{1}{2^4} \sum_{i_1=1}^{2} \sum_{i_2=1}^{2} \sum_{i_3=1}^{2} \sum_{i_4=1}^{2} (G_{i_1,i_2,i_3,i_4} - G^{MA^*}_{i_1,i_2,i_3,i_4})^2. \tag{17}$$

The 3^4 and 4^4 cross-classification tables are defined similarly. The corresponding squared prediction errors of the four methods for each cross-classification table are reported in different rows of Table 9. There we observe that the model averaging method outperforms the other three methods in terms of squared prediction error.

6. CONCLUSION

In this article we propose using a model averaging method to estimate a mixture copula model. Unlike the BIC method, which selects only one individual copula based on a comparison of the BIC values, the model averaging method optimally chooses weights (associated with the components of individual copulas) so as to minimize the estimated squared loss. Our simulation studies show that the model averaging method performs similarly to the penalized likelihood method proposed

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

by Cai and Wang (2014) and the QMLE method in Hu (2006) when the observations are generated from copulas included in the working mixture copula model. However, when the working mixture copula model is misspecified (i.e., when observations are generated from copulas not included in the working mixture model) the model averaging method outperforms the Cai and Wang (2014) penalized likelihood method, the BIC method, and the QMLE method. Our empirical example shows that the model averaging method provides satisfactory estimates of the dependence structures among the stock markets in four developed economies. Thus, the model averaging method can be a useful tool for estimating mixture copula models. It can be used by practitioners in the financial industry for portfolio diversification and risk management.

REFERENCES

Aloui, R., M. Aïssa, & D. Nguyen (2011) Global financial crisis, extreme interdependences, and contagion effects: The role of economic structure? *Journal of Banking & Finance* 35, 130–141.

Ando, T. & K. Li (2014) A model-averaging approach for high-dimensional regression. *Journal of the American Statistical Association* 109, 254–265.

Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.

Cai, Z. & X. Wang. (2014) Selection of mixed copula model via penalized likelihood. *Journal of the American Statistical Association* 109, 788–801.

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

Chen, X. & Y. Fan (2006b) Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *Journal of Econometrics* 135, 125–154.

Cheng, X. & B. Hansen (2015) Forecasting with factor-augmented regression: A frequentist model averaging approach. *Journal of Econometrics* 186, 280–293.

Chollete, L., V. Peña, & C. Lu. (2005) Comovement of international financial markets. Unpublished manuscript.

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

Fan, J. & R. Li (2001) Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* 96, 1348–1360.

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

Gao, Y., X. Zhang, S. Wang, T. Chong, & G. Zou (2018). Frequentist model averaging for threshold models. *Annals of the Institute of Statistical Mathematics*, first published online 12 January 2018. doi:10.1007/s10463-017-0642-9.

Genest, C. & L. Rivest (1993) Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association* 88, 1034–1043.

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

Hansen, B. & J. Racine (2012) Jackknife model averaging. *Journal of Econometrics* 167, 38–46.
Hansen, B. (2007) Least squares model averaging. *Econometrica* 75, 1175–1189.
Hu, L. (2006) Dependence patterns across financial markets: A mixed copula approach. *Applied Financial Economics* 16, 717–729.
Joe, H. (1997) *Multivariate Models and Dependence Concepts*. Chapman & Hall.
Li, D. (2000) On default correlation: A copula function approach. *Journal of Fixed Income* 9, 43–54.
Longin, F. & B. Solnik (2001) Extreme correlation of international equity markets. *The Journal of Finance* 56, 649–676.
Ma, Y. & L. Zhu (2012) A semiparametric approach to dimension reduction. *Journal of the American Statistical Association* 497, 168–179.
Manner, H. & O. Reznikova (2012) A survey on time-varying copulas: Specification, simulations and application. *Econometric Reviews* 31, 654–687.
Nelsen, R. (2006) *An Introduction to Copulas*, 2nd ed. Springer.
Patton, A. (2006) Modelling asymmetric exchange rate dependence. *International Economic Review* 47, 527–556.
Patton, A. (2012) A review of copula models for economic time series. *Journal of Multivariate Analysis* 110, 4–18.
Rodriguez, J. (2007) Measuring financial contagion: A copula approach. *Journal of Empirical Finance* 14, 401–423.
Shao, J. (1997) An asymptotic theory for linear model selection. *Statistica Sinica* 7, 221–264.
Sklar, A. (1959) Fonctions de répartition à n dimensions et leurs marges. *Publication de l'Institut de Statistique de l'Universite de Paris* 8, 229–231.
Zhang, X. (2010) Model Averaging and its Applications. Ph.D. thesis, Academy of Mathematics and Systems Science, Chinese Academy of Sciences.
Zhang, X., A.T.K. Wan, & G. Zou (2013) Model averaging by Jackknife criterion in models with dependent data. *Journal of Econometrics* 174, 82–94.
Zhang, X., D. Yu, G. Zou, & H. Liang (2016) Optimal model averaging estimation for generalized linear models and generalized linear mixed-effects models. *Journal of the American Statistical Association* 111, 1775–1790.
Zimmer, D.M. (2012) The role of copulas in the housing crisis. *Review of Economics and Statistics* 94, 607–620.

APPENDIX

A.1. Notations and Conditions

Let $\widehat{\boldsymbol{\theta}} = \left(\widehat{\boldsymbol{\theta}}_1^\top, \ldots, \widehat{\boldsymbol{\theta}}_K^\top\right)^\top$, where K is a fixed positive integer, and denote the fixed dimension of $\widehat{\boldsymbol{\theta}}$ by $\varkappa$. Also, let $\boldsymbol{\theta}_k^*$ be the pseudo true value defined by

$$\boldsymbol{\theta}_k^* = \arg\max_{\boldsymbol{\theta}_k} E\left[\log\, c_k(F_1^0(\varepsilon_{1t}), \ldots, F_p^0(\varepsilon_{pt}); \boldsymbol{\theta}_k)\right], \qquad k = 1, \ldots, K,$$

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

where $c_k(F_1^0,\ldots,F_p^0;\boldsymbol{\theta}_k) \equiv \partial^p C_k(F_1^0,\ldots,F_p^0;\boldsymbol{\theta}_k)/\partial F_1^0,\ldots,\partial F_p^0$ is the copula density. Denote $\boldsymbol{\theta}^* = \left(\boldsymbol{\theta}_1^{*\top},\ldots,\boldsymbol{\theta}_K^{*\top}\right)^\top$. Let $\mathbf{C}^*(\mathbf{w}) = \widehat{\mathbf{C}}(\mathbf{w})\,|_{\widehat{\boldsymbol{\theta}}=\boldsymbol{\theta}^*}$, $\nu_t(\mathbf{w}) = \partial C(\widetilde{U}_t;\widehat{\boldsymbol{\theta}}_k,\mathbf{w})/\partial\widehat{\boldsymbol{\theta}}|_{\widehat{\boldsymbol{\theta}}=\bar{\boldsymbol{\theta}}_t}$, for $t=1,\ldots,T$, where $\bar{\boldsymbol{\theta}}_t$ is between $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^*$, $\mathbf{Q}(\mathbf{w}) = \{\nu_1(\mathbf{w}),\ldots,\nu_T(\mathbf{w})\}^\top$, $L_T^*(\mathbf{w}) = \|\mathbf{C}^*(\mathbf{w}) - \mathbf{C}_0\|^2$, and $\xi_T = \inf_{\mathbf{w}\in\mathcal{W}} L_T^*(\mathbf{w})$. We assume that J is fixed and $M\to\infty$ as $T\to\infty$.

In addition, for the k^{th} copula, we denote $C_k\left\{v_1,\ldots,v_p;\boldsymbol{\theta}_k\right\} \equiv C_k(\mathbf{v};\boldsymbol{\theta}_k)$ and $c_k\left\{v_1,\ldots,v_p;\boldsymbol{\theta}_k\right\} \equiv c_k(\mathbf{v};\boldsymbol{\theta}_k)$. Also, we denote $l_k(\mathbf{v};\boldsymbol{\theta}_k) \equiv \log c_k(\mathbf{v};\boldsymbol{\theta}_k)$, $l_{\boldsymbol{\theta},k}(\mathbf{v};\boldsymbol{\theta}_k) \equiv \partial l_k(\mathbf{v};\boldsymbol{\theta}_k)/\partial\,\boldsymbol{\theta}_k$, $l_{\boldsymbol{\theta\theta},k}(\mathbf{v};\boldsymbol{\theta}_k) \equiv \partial^2 l_k(\mathbf{v};\boldsymbol{\theta}_k)/\partial\boldsymbol{\theta}_k\partial\boldsymbol{\theta}_k^\top$, $l_{\boldsymbol{\theta j},k}(\mathbf{v};\boldsymbol{\theta}_k) \equiv \partial^2 l_k(\mathbf{v};\boldsymbol{\theta}_k)/\partial u_j\partial\boldsymbol{\theta}_k^\top$ and $U_t \equiv (F_1(\varepsilon_{1t}),\ldots,F_p(\varepsilon_{pt}))$.

To prove the asymptotic optimality as stated in Theorem 1, we assume that the following regularity conditions hold.

Condition C.1. $\boldsymbol{\theta}^*$ is a finite dimensional vector with constant components and it takes values in a compact subset of $\mathcal{R}^\varkappa$, $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* = O_p(T^{-1/2})$.

Condition C.2. Uniformly in $\mathbf{w}\in\mathcal{W}$, the elements of $T\times\varkappa$ matrix $\mathbf{Q}(\mathbf{w})$ are uniformly bounded.

Condition C.3. Uniformly in $\mathbf{w}\in\mathcal{W}$, $T^{-1/2}\|\widehat{\mathbf{C}}(\mathbf{w}) - \overline{\mathbf{C}}(\mathbf{w})\|^2 = O_p(1)$, $T^{-1/2}\{\widehat{\mathbf{C}}(\mathbf{w}) - \overline{\mathbf{C}}(\mathbf{w})\}^\top\{\widehat{\mathbf{C}}(\mathbf{w}) - \widetilde{\mathbf{C}}\} = O_p(1)$, and $T^{-1/2}\{\widehat{\mathbf{C}}(\mathbf{w}) - \overline{\mathbf{C}}(\mathbf{w})\}^\top(\mathbf{C}_0 - \widetilde{\mathbf{C}}) = O_p(1)$.

Condition C.4. There exists a sequence $c_T\to 0$ such that $T\xi_T^{-2} \le c_T$ almost surely, where $\xi_T = \inf_{\mathbf{w}\in\mathcal{W}} L_T^*(\mathbf{w})$.

Remark 2. Condition C.1 requires that $\boldsymbol{\theta}^*$ takes values in a compact set. It rules out some cases such as when the true distribution is normal, while one fits a t_ν-distribution model and estimates the degree of freedom ν (as $\nu^* = \infty$). Condition C.1 also requires that the convergence rate of $\widehat{\boldsymbol{\theta}}$ to the pseudo true value $\boldsymbol{\theta}^*$ is $O_p(T^{-1/2})$. Chen and Fan (2006b) showed that Condition C.1 holds true under quite general regularity conditions including: (i) $\boldsymbol{\theta}_k^*$ are in the interior of the parameter space for $k=1,\ldots,K$, (ii) $\left\{\mathbf{Y}_t',\mathbf{Z}_t'\right\}_{t=1}^{\mathrm{T}}$ is stationary β mixing with the appropriate decay rate, (iii) $l_{\boldsymbol{\theta},k}(\mathbf{u};\boldsymbol{\theta}_k)$, $l_{\boldsymbol{\theta\theta},k}(\mathbf{u};\boldsymbol{\theta}_k)$ and $l_{\boldsymbol{\theta j},k}(\mathbf{u};\boldsymbol{\theta}_k)$ satisfy some standard smoothness conditions for $k=1,\ldots,K$, and (iv) $l_{\boldsymbol{\theta},k}(U_t;\,\boldsymbol{\theta}_k)$, $l_{\boldsymbol{\theta\theta},k}(U_t;\boldsymbol{\theta}_k)$ and $l_{\boldsymbol{\theta j},k}(U_t;\boldsymbol{\theta}_k)$ satisfy some appropriate moment conditions for $k=1,\ldots,K$. Condition C.2 puts a restriction of the derivative of copulas and requires that the derivatives are bounded uniformly.

With Conditions C.1 and C.2, we obtain that uniformly in $w\in\mathcal{W}$,

$$\begin{aligned}
&T^{-1/2}\left\|\mathbf{Q}(\mathbf{w})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)\right\|^2\\
&\le T^{-1/2}\left\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right\|^2\lambda_{\max}\left\{\mathbf{Q}^{\mathrm{T}}(\mathbf{w})\mathbf{Q}(\mathbf{w})\right\}\\
&= T^{-1/2}O_p(T^{-1})O_p(T)\\
&= O_p(T^{-1/2}),
\end{aligned}\tag{A.1}$$

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

where $\lambda_{\max}(\cdot)$ denotes the maximum eigenvalue of a matrix (since C.2 implies that $\lambda_{\max}\left\{\mathbf{Q}^{\mathrm{T}}(\mathbf{w})\mathbf{Q}(\mathbf{w})\right\} = O_p(T)$), and

$$T^{-1/2}\left\{\mathbf{C}^*(\mathbf{w}) - \mathbf{C}_0\right\}^\top \mathbf{Q}(\mathbf{w})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) = T^{-1/2} O_p(T^{1/2}) = O_p(1), \tag{A.2}$$

where we also used the fact that the elements of vector $|\mathbf{C}^*(\mathbf{w}) - \mathbf{C}_0|$ are uniformly bounded by 2.

Condition C.3 requires that as $T \to \infty$, the difference between the loss of the regular and leave-M-out estimators decreases at some rate. This is similar to condition (A.10) of Andrews (1991) and condition (A.5) of Hansen and Racine (2012). Let $\widehat{\boldsymbol{\theta}}^{(-j)}$ be the estimator of $\boldsymbol{\theta}$ with the j^{th} group removed from the sample, and $\boldsymbol{\nu}_t^{(-j)}(\mathbf{w}) = \partial \widehat{C}^{(-j)}(\widehat{\varepsilon}_t, \mathbf{w})/\partial \widehat{\boldsymbol{\theta}}^{(-j)}|_{\widehat{\boldsymbol{\theta}}^{(-j)} = \bar{\boldsymbol{\theta}}_t^{(-j)}}$ for $t = 1, \ldots, T$, where $\bar{\boldsymbol{\theta}}_t^{(-j)}$ is between $\widehat{\boldsymbol{\theta}}^{(-j)}$ and $\boldsymbol{\theta}^*$, and $\overline{\mathbf{Q}}(\mathbf{w}) = \{\widetilde{\boldsymbol{\nu}}_1^{(-1)}(\mathbf{w}), \ldots, \boldsymbol{\nu}_T^{(-J)}(\mathbf{w})\}^{\mathrm{T}}$. From Proposition 3.2 of Chen and Fan (2006b) and Assumptions D and N in that article, the estimator $\widehat{\boldsymbol{\theta}}$ converges to $\boldsymbol{\theta}^*$ with a root-T rate and the estimator $\widehat{\boldsymbol{\theta}}^{(-j)}$ converges to $\boldsymbol{\theta}^*$ with rate root-$(T - M)$, which is uniformly over $j = 1, \ldots, J$ because J is a finite constant. When the elements of the $T \times \kappa$ matrix $\overline{\mathbf{Q}}(\mathbf{w})$ are uniformly bounded, similar to the derivations in (A.1) and (A.2), we obtain that uniformly over $\mathbf{w} \in \mathcal{W}$, $T^{-1/2}\|\widehat{\mathbf{C}}(\mathbf{w}) - \overline{\mathbf{C}}(\mathbf{w})\|^2 = O_p(T^{-1/2})$, $T^{-1/2}\{\widehat{\mathbf{C}}(\mathbf{w}) - \overline{\mathbf{C}}(\mathbf{w})\}^\top\{\widehat{\mathbf{C}}(\mathbf{w}) - \widetilde{\mathbf{C}}\} = O_p(1)$, and $T^{-1/2}\{\widehat{\mathbf{C}}(\mathbf{w}) - \overline{\mathbf{C}}(\mathbf{w})\}^\top(\mathbf{C}_0 - \widetilde{\mathbf{C}}) = O_p(1)$.

Condition C.4 requires that ξ_T grows at a rate faster than $T^{1/2}$, which implies that all candidate copulas are misspecified. This is similar to the third part of condition (A7) in Zhang, Wan, and Zou (2013) and condition 7 in Ando and Li (2014). The assumption that all candidate models are misspecified is a common condition used in proving optimality properties of model averaging estimators.

Here we would like to comment on the case when the candidate mixture copula is not misspecified, i.e., the true copula is the mixture copula $C_K(\mathbf{u}; \boldsymbol{\theta}_K) = \sum_{k=1}^{K-1} \omega_k C_k\left\{F_1(x_1), \ldots, F_p(x_p); \boldsymbol{\theta}_k\right\}$ with $\sum_{k=1}^{K-1} \omega_k = 1$ and $\omega_k > 0$ for $k = 1, \ldots, K-1$, we can show that under some regularity conditions, the estimated weight of $C_K(\widetilde{\mathbf{u}}; \widehat{\boldsymbol{\theta}}_K)$,

$$\widehat{w}_K \to 1, \tag{A.3}$$

in probability, as $T \to \infty$ (see Appendix A.4 for the required regularity conditions and a proof of this result), thus our method will have the same large sample property as that of QMLE. This result is expected and reasonable, and means that our method is adaptable to the "lucky" case that the true copula is the mixture copula. The simulation results in Table 1 also indicate that under this case, our MA method performs similarly to the QMLE method. Let $\mathbf{w}^* = (0, \ldots, 0, 1)^\top$. When the candidate mixture copula is not misspecified, we have $L_T^*(\mathbf{w}^\star) = \|\mathbf{C}^*(\mathbf{w}^\star) - \mathbf{C}_0\|^2 = O_p(1)$ because of the convergence of QMLE to the true value. Therefore, Condition 4 is violated when the mixture copula contains the true copula.

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

A.2. Proof of Theorem 1

The result below follows from various definitions given in the article (and some standard manipulations)

$$
\begin{aligned}
CV_J(\mathbf{w}) &= \left\|\overline{\mathbf{C}}(\mathbf{w})-\widetilde{\mathbf{C}}\right\|^2 \\
&= \left\|\left\{\widehat{\mathbf{C}}(\mathbf{w})-\mathbf{C}_0\right\}-\left\{\widehat{\mathbf{C}}(\mathbf{w})-\overline{\mathbf{C}}(\mathbf{w})\right\}+\left(\mathbf{C}_0-\widetilde{\mathbf{C}}\right)\right\|^2 \\
&= L_T(\mathbf{w})+\left\|\widehat{\mathbf{C}}(\mathbf{w})-\overline{\mathbf{C}}(\mathbf{w})\right\|^2+\left\|\mathbf{C}_0-\widetilde{\mathbf{C}}\right\|^2-2\left\{\widehat{\mathbf{C}}(\mathbf{w})-\overline{\mathbf{C}}(\mathbf{w})\right\}^\top\left(\mathbf{C}_0-\widetilde{\mathbf{C}}\right) \\
&\quad -2\left\{\widehat{\mathbf{C}}(\mathbf{w})-\overline{\mathbf{C}}(\mathbf{w})\right\}^\top\left\{\widehat{\mathbf{C}}(\mathbf{w})-\mathbf{C}_0\right\}+2\left\{\widehat{\mathbf{C}}(\mathbf{w})-\mathbf{C}_0\right\}^\top\left(\mathbf{C}_0-\widetilde{\mathbf{C}}\right) \\
&= L_T(\mathbf{w})+\left\|\widehat{\mathbf{C}}(\mathbf{w})-\overline{\mathbf{C}}(\mathbf{w})\right\|^2-2\left\{\widehat{\mathbf{C}}(\mathbf{w})-\overline{\mathbf{C}}(\mathbf{w})\right\}^\top\left\{\widehat{\mathbf{C}}(\mathbf{w})-\widetilde{\mathbf{C}}\right\} \\
&\quad +2\left\{\widehat{\mathbf{C}}(\mathbf{w})-\overline{\mathbf{C}}(\mathbf{w})\right\}^\top\left(\mathbf{C}_0-\widetilde{\mathbf{C}}\right)+2\overline{\mathbf{C}}(\mathbf{w})^\top\left(\mathbf{C}_0-\widetilde{\mathbf{C}}\right)-\left(\mathbf{C}_0+\widetilde{\mathbf{C}}\right)^\top\left(\mathbf{C}_0-\widetilde{\mathbf{C}}\right) \\
&\equiv L_T(\mathbf{w})+\Xi_T(\mathbf{w})-\left(\mathbf{C}_0+\widetilde{\mathbf{C}}\right)^\top\left(\mathbf{C}_0-\widetilde{\mathbf{C}}\right), \qquad \textbf{(A.4)}
\end{aligned}
$$

where the last term is unrelated to the weight vector $\mathbf{w}$, and

$$
\begin{aligned}
L_T(\mathbf{w}) &= \left\|\widehat{\mathbf{C}}(\mathbf{w})-\mathbf{C}_0\right\|^2 \\
&= \left\|\left\{\widehat{\mathbf{C}}(\mathbf{w})-\mathbf{C}^*(\mathbf{w})\right\}+\left\{\mathbf{C}^*(\mathbf{w})-\mathbf{C}_0\right\}\right\|^2 \\
&= \left\|\widehat{\mathbf{C}}(\mathbf{w})-\mathbf{C}^*(\mathbf{w})\right\|^2+\left\|\mathbf{C}^*(\mathbf{w})-\mathbf{C}_0\right\|^2+2\left\{\mathbf{C}^*(\mathbf{w})-\mathbf{C}_0\right\}^\top\left\{\widehat{\mathbf{C}}(\mathbf{w})-\mathbf{C}^*(\mathbf{w})\right\} \\
&= L_T^*(\mathbf{w})+\left\|\widehat{\mathbf{C}}(\mathbf{w})-\mathbf{C}^*(\mathbf{w})\right\|^2+2\left\{\mathbf{C}^*(\mathbf{w})-\mathbf{C}_0\right\}^\top\left\{\widehat{\mathbf{C}}(\mathbf{w})-\mathbf{C}^*(\mathbf{w})\right\} \\
&\equiv L_T^*(\mathbf{w})+\Pi_T(\mathbf{w}). \qquad \textbf{(A.5)}
\end{aligned}
$$

We first prove that if

$$
\sup_{\mathbf{w}\in\mathcal{W}}\frac{\left|L_T(\mathbf{w})-L_T^*(\mathbf{w})\right|}{L_T^*(\mathbf{w})}=o_p(1), \qquad \textbf{(A.6)}
$$

and $CV_J(\mathbf{w})$ can be written as

$$
CV_J(\mathbf{w})=L_T(\mathbf{w})+a_T(\mathbf{w})+b_T \qquad \textbf{(A.7)}
$$

with the term $a_T(\mathbf{w})$ satisfies

$$
\sup_{\mathbf{w}\in\mathcal{W}}\frac{|a_T(\mathbf{w})|}{L_T^*(\mathbf{w})}=o_p(1) \qquad \textbf{(A.8)}
$$

and the term b_T is unrelated to $\mathbf{w}$, then Theorem 1 holds true. Our proof follows the similar arguments as in Zhang (2010) and Gao, Zhang, Wang, Chong, and Zou (2018). From (A.6), we know that

$$
\inf_{\mathbf{w}\in\mathcal{W}}\frac{L_T(\mathbf{w})}{L_T^*(\mathbf{w})}\geq-\sup_{\mathbf{w}\in\mathcal{W}}\frac{\left|L_T(\mathbf{w})-L_T^*(\mathbf{w})\right|}{L_T^*(\mathbf{w})}\to 1, \qquad \textbf{(A.9)}
$$

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

in probability as $T \to \infty$. By the definition of $\widehat{\mathbf{w}}$ and (A.7), we have

$$\inf_{\mathbf{w}\in\mathcal{W}} \{L_T(\mathbf{w}) + a_T(\mathbf{w})\} = L_T(\widehat{\mathbf{w}}) + a_T(\widehat{\mathbf{w}}). \tag{A.10}$$

By Condition C.4, we know that there exists a non-negative sequence ν_T and a sequence of vectors $\mathbf{w}(T) \in \mathcal{W}$ such that as $T \to \infty$,

$$\nu_T \xi_T^{-1} \to 0 \quad \text{and} \quad \inf_{\mathbf{w}\in\mathcal{W}} L_T(\mathbf{w}) = L_T\{\mathbf{w}(T)\} - \nu_T. \tag{A.11}$$

From (A.9) and (A.11), we have

$$\inf_{\mathbf{w}\in\mathcal{W}} \frac{|L_T(\mathbf{w}) - \nu_T|}{L_T^*(\mathbf{w})} \geq \inf_{\mathbf{w}\in\mathcal{W}} \frac{L_T(\mathbf{w})}{L_T^*(\mathbf{w})} - \frac{\nu_T}{\xi_T} \to 1, \tag{A.12}$$

in probability as $T \to \infty$. From (A.8)–(A.12), we obtain that, for any $\delta > 0$,

$$\begin{aligned}
&\Pr\left\{\left|\frac{\inf_{\mathbf{w}\in\mathcal{W}} L_T(\mathbf{w})}{L_T(\widehat{\mathbf{w}})} - 1\right| > \delta\right\} \\
&= \Pr\left\{\frac{L_T(\widehat{\mathbf{w}}) - \inf_{\mathbf{w}\in\mathcal{W}} L_T(\mathbf{w})}{L_T(\widehat{\mathbf{w}})} > \delta\right\} \\
&= \Pr\left\{\frac{\inf_{\mathbf{w}\in\mathcal{W}} (L_T(\mathbf{w}) + a_T(\mathbf{w})) - a_T(\widehat{\mathbf{w}}) - \inf_{\mathbf{w}\in\mathcal{W}} L_T(\mathbf{w})}{L_T(\widehat{\mathbf{w}})} > \delta\right\} \\
&\leq \Pr\left\{\frac{L_T\{\mathbf{w}(T)\} + a_T\{\mathbf{w}(T)\} - a_T(\widehat{\mathbf{w}}) - L_T\{\mathbf{w}(T)\} + \nu_T}{L_T(\widehat{\mathbf{w}})} > \delta\right\} \\
&\leq \Pr\left\{\frac{|a_T\{\mathbf{w}(T)\}|}{\inf_{\mathbf{w}\in\mathcal{W}} L_T(\mathbf{w})} + \frac{|a_T(\widehat{\mathbf{w}})|}{L_T(\widehat{\mathbf{w}})} + \frac{\nu_T}{L_T(\widehat{\mathbf{w}})} > \delta\right\} \\
&\leq \Pr\left\{\sup_{\mathbf{w}\in\mathcal{W}} \frac{|a_T(\mathbf{w})|}{L_T(\mathbf{w}) - \nu_T} + \sup_{\mathbf{w}\in\mathcal{W}} \frac{|a_T(\mathbf{w})|}{L_T(\mathbf{w})} + \sup_{\mathbf{w}\in\mathcal{W}} \frac{\nu_T}{L_T(\mathbf{w})} > \delta\right\} \\
&\leq \Pr\left\{\sup_{\mathbf{w}\in\mathcal{W}} \frac{|a_T(\mathbf{w})|}{L_T^*(\mathbf{w})} \sup_{\mathbf{w}\in\mathcal{W}} \frac{L_T^*(\mathbf{w})}{|L_T(\mathbf{w}) - \nu_T|} + \sup_{\mathbf{w}\in\mathcal{W}} \frac{|a_T(\mathbf{w})|}{L_T^*(\mathbf{w})} \sup_{\mathbf{w}\in\mathcal{W}} \frac{L_T^*(\mathbf{w})}{L_T(\mathbf{w})}\right. \\
&\qquad \left. + \sup_{\mathbf{w}\in\mathcal{W}} \frac{\nu_T}{L_T^*(\mathbf{w})} \sup_{\mathbf{w}\in\mathcal{W}} \frac{L_T^*(\mathbf{w})}{L_T(\mathbf{w})} > \delta\right\} \\
&= \Pr\left\{\sup_{\mathbf{w}\in\mathcal{W}} \frac{|a_T(\mathbf{w})|}{L_T^*(\mathbf{w})} \left[\inf_{\mathbf{w}\in\mathcal{W}} \frac{|L_T(\mathbf{w}) - \nu_T|}{L_T^*(\mathbf{w})}\right]^{-1} + \sup_{\mathbf{w}\in\mathcal{W}} \frac{|a_T(\mathbf{w})|}{L_T^*(\mathbf{w})} \left[\inf_{\mathbf{w}\in\mathcal{W}} \frac{L_T(\mathbf{w})}{L_T^*(\mathbf{w})}\right]^{-1}\right. \\
&\qquad \left. + \frac{\nu_T}{\inf_{\mathbf{w}\in\mathcal{W}} L_T^*(\mathbf{w})} \left[\inf_{\mathbf{w}\in\mathcal{W}} \frac{L_T(\mathbf{w})}{L_T^*(\mathbf{w})}\right]^{-1} > \delta\right\} \\
&\to 0.
\end{aligned}$$

Hence from (A.4)–(A.8), Theorem 1 is valid if the following hold:

$$\sup_{\mathbf{w}\in\mathcal{W}} \frac{|\Xi_T(\mathbf{w})|}{L_T^*(\mathbf{w})} = o_p(1) \tag{A.13}$$

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

and

$$\sup_{\mathbf{w}\in\mathcal{W}} \frac{|\Pi_T(\mathbf{w})|}{L_T^*(\mathbf{w})} = o_p(1). \tag{A.14}$$

Using Taylor expansion,

$$\widehat{\mathbf{C}}(\mathbf{w}) - \mathbf{C}^*(\mathbf{w}) = \mathbf{Q}(\mathbf{w})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*), \tag{A.15}$$

where $\mathbf{Q}(\mathbf{w}) = \mathbf{Q}(\mathbf{w}; \overline{\boldsymbol{\theta}}_1, \ldots, \overline{\boldsymbol{\theta}}_T)$, $\overline{\boldsymbol{\theta}}_t$ is between the line segment of $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^*$ (see the first paragraph of Appendix A.1 for the definition of $\mathbf{Q}(\mathbf{w})$). From (A.15), (A.1), (A.2), and Condition C.4, we have

$$\begin{aligned}
&\sup_{\mathbf{w}\in\mathcal{W}} \frac{|\Pi_T(\mathbf{w})|}{L_T^*(\mathbf{w})} \\
&\le \xi_T^{-1} \sup_{\mathbf{w}\in\mathcal{W}} \left\|\widehat{\mathbf{C}}(\mathbf{w}) - \mathbf{C}^*(\mathbf{w})\right\|^2 + 2\xi_T^{-1} \sup_{\mathbf{w}\in\mathcal{W}} \left|\left\{\mathbf{C}^*(\mathbf{w}) - \mathbf{C}_0\right\}^\top \left\{\widehat{\mathbf{C}}(\mathbf{w}) - \mathbf{C}^*(\mathbf{w})\right\}\right| \\
&= \frac{T^{1/2}}{\xi_T} T^{-1/2} \sup_{\mathbf{w}\in\mathcal{W}} \left\|\mathbf{Q}(\mathbf{w})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)\right\|^2 + 2\frac{T^{1/2}}{\xi_T} T^{-1/2} \sup_{\mathbf{w}\in\mathcal{W}} \left|\left\{\mathbf{C}^*(\mathbf{w}) - \mathbf{C}_0\right\}^\top \mathbf{Q}(\mathbf{w})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)\right| \\
&= o_p(1),
\end{aligned}$$

which is (A.14).

Similarly, from Conditions C.3 and C.4, we have

$$\begin{aligned}
&\sup_{\mathbf{w}\in\mathcal{W}} \frac{\left|\left\|\widehat{\mathbf{C}}(\mathbf{w}) - \overline{\mathbf{C}}(\mathbf{w})\right\|^2 - 2\left\{\widehat{\mathbf{C}}(\mathbf{w}) - \overline{\mathbf{C}}(\mathbf{w})\right\}^\top \left\{\widehat{\mathbf{C}}(\mathbf{w}) - \widetilde{\mathbf{C}}\right\} + 2\left\{\widehat{\mathbf{C}}(\mathbf{w}) - \overline{\mathbf{C}}(\mathbf{w})\right\}^\top (\mathbf{C}_0 - \widetilde{\mathbf{C}})\right|}{L_T^*(\mathbf{w})} \\
&= o_p(1).
\end{aligned} \tag{A.16}$$

Denote $F_0(\cdot)$ as the true distribution function of $\varepsilon_t \equiv (\varepsilon_{1t}, \ldots, \varepsilon_{pt})'$, $\widetilde{C}_T(\mathbf{x}) = T^{-1}\sum_{s=1}^T I\{\widehat{\varepsilon}_s \le \mathbf{x}\}$ and $C_T(\mathbf{x}) = \frac{1}{T}\sum_{s=1}^T I\{\varepsilon_s \le \mathbf{x}\}$.

Next, we want to show that

$$T^{-1}\sum_{t=1}^T \left|F_0(\varepsilon_t) - \widetilde{C}_T(\widehat{\varepsilon}_t)\right| = O_p(T^{-1/2}). \tag{A.17}$$

By adding/subtracting terms we have

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^T \left|F_0(\varepsilon_t) - \widetilde{C}_T(\widehat{\varepsilon}_t)\right| \\
&= \frac{1}{T}\sum_{t=1}^T \left|F_0(\varepsilon_t) - C_T(\varepsilon_t) + C_T(\varepsilon_t) - \widetilde{C}_T(\varepsilon_t) + \widetilde{C}_T(\varepsilon_t) - \widetilde{C}_T(\widehat{\varepsilon}_t)\right| \\
&\le \frac{1}{T}\sum_{t=1}^T |F_0(\varepsilon_t) - C_T(\varepsilon_t)| + \frac{1}{T}\sum_{t=1}^T \left|C_T(\varepsilon_t) - \widetilde{C}_T(\varepsilon_t)\right| \\
&\quad + \frac{1}{T}\sum_{t=1}^T \left|\widetilde{C}_T(\varepsilon_t) - \widetilde{C}_T(\widehat{\varepsilon}_t)\right|.
\end{aligned} \tag{A.18}$$

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

Since $\{\varepsilon_t\}$ are i.i.d., it is well established that

$$\sqrt{T}\sup_{\mathbf{x}\in\mathcal{R}^p}|F_0(\mathbf{x})-C_T(\mathbf{x})|=O_p(1). \tag{A.19}$$

Therefore, the first term of (A.18) is $O_p(T^{-1/2})$. By the same proof method as used in Chen and Fan (2006b, Lemma A.1 (3)), one can show that

$$\sqrt{T}\sup_{\mathbf{x}\in\mathcal{R}^p}|C_T(\mathbf{x})-\widetilde{C}_T(\mathbf{x})|=O_p(1), \tag{A.20}$$

so the second term of (A.18) is also $O_p(T^{-1/2})$. Finally the third term of (A.18) can be written as

$$\begin{aligned}
&T^{-1}\sum_{t=1}^{T}\left|\widetilde{C}_T(\varepsilon_t)-\widetilde{C}_T(\widehat{\varepsilon}_t)\right|\\
&=T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}|I\{\widehat{\varepsilon}_s\le\varepsilon_t\}-I\{\widehat{\varepsilon}_s\le\widehat{\varepsilon}_t\}|\\
&=T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}|1-I\{\widehat{\varepsilon}_s>\varepsilon_t\}-1+I\{\widehat{\varepsilon}_s>\widehat{\varepsilon}_t\}|\\
&=T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}|I\{\widehat{\varepsilon}_t<\widehat{\varepsilon}_s\}-I\{\varepsilon_t<\widehat{\varepsilon}_s\}|\\
&=T^{-1}\sum_{s=1}^{T}\left|\widetilde{C}_T(\widehat{\varepsilon}_s)-C_T(\widehat{\varepsilon}_s)\right|=O_p(T^{-1/2})
\end{aligned} \tag{A.21}$$

by (A.20). Hence, we have shown that the third term of (A.18) is $O_p(T^{-1/2})$.

From $c_T\to 0$, (A.17), and the fact that all elements of the vector $|\widehat{\mathbf{C}}_k|$ are bounded by 1, we obtain that for any $j\in\{1,\ldots,J\}$,

$$\begin{aligned}
&c_T^{1/2}M^{-1/2}\left|\sum_{m=1}^{M}C_k(\widetilde{U}_{(j-1)M+m};\widehat{\boldsymbol{\theta}}_k^{(-j)})\left\{C_0(U_{0,(j-1)M+m};\boldsymbol{\theta}_0)-\widetilde{C}_{(j)}(\widehat{\varepsilon}_{(j-1)M+m})\right\}\right|\\
&\le c_T^{1/2}M^{1/2}M^{-1}\sum_{m=1}^{M}\left|C_0(U_{0,(j-1)M+m};\boldsymbol{\theta}_0)-\widetilde{C}_{(j)}(\widehat{\varepsilon}_{(j-1)M+m})\right|\\
&=c_T^{1/2}M^{1/2}M^{-1}\sum_{m=1}^{M}\left|F_0(\varepsilon_{(j-1)M+m})-\widetilde{C}_{(j)}(\widehat{\varepsilon}_{(j-1)M+m})\right|\\
&=o_p(1),
\end{aligned} \tag{A.22}$$

which, along with Condition C.4 and the assumption that K and J are fixed, implies that

$$\begin{aligned}
&\xi_T^{-1}\sup_{\mathbf{w}\in\mathcal{W}}\left|\overline{\mathbf{C}}(\mathbf{w})^{\top}(\mathbf{C}_0-\widetilde{\mathbf{C}})\right|\\
&=\xi_T^{-1}\sup_{\mathbf{w}\in\mathcal{W}}\left|\sum_{k=1}^{K}w_k\overline{\mathbf{C}}_k^{\top}(\mathbf{C}_0-\widetilde{\mathbf{C}})\right|
\end{aligned}$$

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

$$
\begin{aligned}
&\le \sum_{k=1}^{K} \frac{T^{1/2}}{\xi_T} T^{-1/2} \left| \overline{\mathbf{C}}_k^{\top} (\mathbf{C}_0 - \widetilde{\mathbf{C}}) \right| \\
&= \sum_{k=1}^{K} J^{-1/2} \frac{T^{1/2}}{\xi_T} \left| \sum_{j=1}^{J} M^{-1/2} \sum_{m=1}^{M} C_k(\widetilde{U}_{(j-1)M+m}; \widehat{\boldsymbol{\theta}}_k^{(-j)}) \left\{ C_0(U_{0,(j-1)M+m}; \boldsymbol{\theta}_0) - \widetilde{C}_{(j)}(\widehat{\varepsilon}_{(j-1)M+m}) \right\} \right| \\
&\le \sum_{k=1}^{K} J^{-1/2} c_T^{1/2} \left| \sum_{j=1}^{J} M^{-1/2} \sum_{m=1}^{M} C_k(\widetilde{U}_{(j-1)M+m}; \widehat{\boldsymbol{\theta}}_k^{(-j)}) \left\{ C_0(U_{0,(j-1)M+m}; \boldsymbol{\theta}_0) - \widetilde{C}_{(j)}(\widehat{\varepsilon}_{(j-1)M+m}) \right\} \right| \\
&= \sum_{k=1}^{K} J^{-1/2} \left| \sum_{j=1}^{J} c_T^{1/2} M^{-1/2} \sum_{m=1}^{M} C_k(\widetilde{U}_{(j-1)M+m}; \widehat{\boldsymbol{\theta}}_k^{(-j)}) \left\{ C_0(U_{0,(j-1)M+m}; \boldsymbol{\theta}_0) - \widetilde{C}_{(j)}(\widehat{\varepsilon}_{(j-1)M+m}) \right\} \right| \\
&= o_p(1), \qquad \textbf{(A.23)}
\end{aligned}
$$

where the second '$\le$' holds almost surely. From (A.16) and (A.23), we obtain (A.13). This completes the proof for Theorem 1.

A.3. Proof of Proposition 1

First, we show that Proposition 1(i) holds without imposing the i.i.d assumption. From the steps in deriving (A.23), we know that uniformly in $\mathbf{w} \in \mathcal{W}$,

$$
\xi_T^{-1} \left\{ \left\| \overline{\mathbf{C}}(\mathbf{w}) - \mathbf{C}_0 \right\|^2 - \left\| \overline{\mathbf{C}}(\mathbf{w}) - \widetilde{\mathbf{C}} \right\|^2 \right\} = \xi_T^{-1} \left\{ 2\overline{\mathbf{C}}(\mathbf{w}) - \widetilde{\mathbf{C}} - \mathbf{C}_0 \right\}^{\top} (\widetilde{\mathbf{C}} - \mathbf{C}_0) = o_p(1).
$$

From Conditions C.1–C.3, (A.6), and the steps used in the proof of (A.14), we have that uniformly in $\mathbf{w} \in \mathcal{W}$,

$$
\begin{aligned}
\left\| \overline{\mathbf{C}}(\mathbf{w}) - \mathbf{C}_0 \right\|^2 &= L_T(\mathbf{w}) + \left\| \widehat{\mathbf{C}}(\mathbf{w}) - \overline{\mathbf{C}}(\mathbf{w}) \right\|^2 - 2 \left\{ \widehat{\mathbf{C}}(\mathbf{w}) - \overline{\mathbf{C}}(\mathbf{w}) \right\}^{\top} \left\{ \widehat{\mathbf{C}}(\mathbf{w}) - \mathbf{C}_0 \right\} \\
&= L_T^*(\mathbf{w}) + o_p(\xi_T).
\end{aligned}
$$

Thus, uniformly in $\mathbf{w} \in \mathcal{W}$,

$$
\begin{aligned}
\frac{E \left\| \overline{\mathbf{C}}(\mathbf{w}) - \mathbf{C}_0 \right\|^2 - E \left\| \overline{\mathbf{C}}(\mathbf{w}) - \widetilde{\mathbf{C}} \right\|^2}{E \left\| \overline{\mathbf{C}}(\mathbf{w}) - \mathbf{C}_0 \right\|^2} &= \frac{E \left[\xi_T^{-1} \left\{ \left\| \overline{\mathbf{C}}(\mathbf{w}) - \mathbf{C}_0 \right\|^2 - \left\| \overline{\mathbf{C}}(\mathbf{w}) - \widetilde{\mathbf{C}} \right\|^2 \right\} \right]}{E \left\{ \xi_T^{-1} L_T^*(\mathbf{w}) + o_p(1) \right\}} \\
&= \frac{E \left[o_p(1) \right]}{E \xi_T^{-1} L_T^*(\mathbf{w}) + E[o_p(1)]}. \qquad \textbf{(A.24)}
\end{aligned}
$$

Under the assumption that the above $o_p(1)$ term is uniformly integrable, then (A.24) is $o(1)$ uniformly in $\mathbf{w} \in \mathcal{W}$, which is the first part of Proposition 1.

Next, we consider the i.i.d. data case, i.e., the second part of Proposition 1. We denote $V_t = (F_1^0(X_{1t}), \ldots, F_p^0(X_{pt}))$ and $\widetilde{V}_t = (\widetilde{F}_1(X_{1t}), \ldots, \widetilde{F}_p(X_{pt}))$, then it is easy to see that

$$
\begin{aligned}
&E \left\| \overline{\mathbf{C}}(\mathbf{w}) - \mathbf{C}_0 \right\|^2 - E \left\| \overline{\mathbf{C}}(\mathbf{w}) - \widetilde{\mathbf{C}} \right\|^2 \\
&= 2E \left\{ \overline{\mathbf{C}}^{\top}(\mathbf{w}) (\widetilde{\mathbf{C}} - \mathbf{C}_0) \right\} - E \left\{ (\widetilde{\mathbf{C}} + \mathbf{C}_0)^{\top} (\widetilde{\mathbf{C}} - \mathbf{C}_0) \right\}
\end{aligned}
$$

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

$$= 2\sum_{k=1}^{K} w_k \sum_{j=1}^{J}\sum_{m=1}^{M} E\left[C_k(\widetilde{V}_{(j-1)M+m};\widehat{\boldsymbol{\theta}}_k^{(-j)})\{\widetilde{C}_{(j)}(X_{(j-1)M+m}) - C_0(V_{0,(j-1)M+m};\boldsymbol{\theta}_0)\}\right]$$
$$-E\left\{(\widetilde{\mathbf{C}}+\mathbf{C}_0)^{\mathrm{T}}(\widetilde{\mathbf{C}}-\mathbf{C}_0)\right\}. \tag{A.25}$$

Let $X_t = (X_{1t},\ldots,X_{pt})^\top$ and $\mathcal{X}^{(-j)} = \{X_1,\ldots,X_{(j-1)M}, X_{jM+1},\ldots,X_T\}$. Then from

$$E\left\{\widetilde{C}_{(j)}(\mathbf{x})\right\} = C_0(\mathbf{u};\boldsymbol{\theta}_0),$$

we know that for any $m \in \{1,\ldots,M\}$ and $j \in \{1,\ldots,J\}$,

$$\begin{aligned}
&E\left[C_k(\widetilde{V}_{(j-1)M+m};\widehat{\boldsymbol{\theta}}_k^{(-j)})\{\widetilde{C}_{(j)}(X_{(j-1)M+m}) - C_0(V_{0,(j-1)M+m};\boldsymbol{\theta}_0)\} \mid \mathcal{X}^{(-j)}\right]\\
&= E_{X_{(j-1)M+m}}\Big(E\Big[C_k(\widetilde{V}_{(j-1)M+m};\widehat{\boldsymbol{\theta}}_k^{(-j)})\\
&\qquad\times\{\widetilde{C}_{(j)}(X_{(j-1)M+m}) - C_0(V_{0,(j-1)M+m};\boldsymbol{\theta}_0)\}|X_{(j-1)M+m},\ \mathcal{X}^{(-j)}\Big]\Big)\\
&= E_{X_{(j-1)M+m}}\Big[C_k(\widetilde{V}_{(j-1)M+m};\widehat{\boldsymbol{\theta}}_k^{(-j)})\\
&\qquad\times E\left\{\widetilde{C}_{(j)}(X_{(j-1)M+m}) - C_0(V_{0,(j-1)M+m};\boldsymbol{\theta}_0)|X_{(j-1)M+m},\ \mathcal{X}^{(-j)}\right\}\Big]\\
&= 0,
\end{aligned} \tag{A.26}$$

where the expectation $E_{X_{(j-1)M+m}}$ is taken with respect to the randomness of $X_{(j-1)M+m}$. The above two formulas (A.25) and (A.26) imply that

$$E\left\|\overline{\mathbf{C}}(\mathbf{w}) - \mathbf{C}_0\right\|^2 - E\left\|\overline{\mathbf{C}}(\mathbf{w}) - \widetilde{\mathbf{C}}\right\|^2 = -E\left\{(\widetilde{\mathbf{C}}+\mathbf{C}_0)^{\mathrm{T}}(\widetilde{\mathbf{C}}-\mathbf{C}_0)\right\},$$

where the right-hand-side quantity does not depend on $\mathbf{w}$, completing the proof for the claim that the J-fold CV criterion is an unbiased estimator of the expected CV squared loss plus a term unrelated to $\mathbf{w}$.

A.4. Proof of (A.3)

Let $\boldsymbol{\psi} = (\psi_1,\ldots,\psi_{K-1})^\top$ belonging to $\mathcal{Q} = \left\{\boldsymbol{\psi} \in [0,1]^{K-1} : \sum_{k=1}^{K-1}\psi_k = 1\right\}$, $\mathbf{C}^*(\boldsymbol{\psi}) = \sum_{k=1}^{K-1}\psi_k \widehat{\mathbf{C}}_k\,|_{\widehat{\boldsymbol{\theta}}_k=\boldsymbol{\theta}_k^*}$, $L_T^*(\boldsymbol{\psi}) = \|\mathbf{C}^*(\boldsymbol{\psi}) - \mathbf{C}_0\|^2$, and $\widetilde{\xi}_T = \inf_{\boldsymbol{\psi}\in\mathcal{Q}} L_T^*(\boldsymbol{\psi})$.

Condition C.5. There exists a sequence $\widetilde{c}_T \to 0$ such that $T\widetilde{\xi}_T^{-2} \le \widetilde{c}_T$ almost surely.

Condition C.5 is a counterpart of Condition C.4 and requires that $\widetilde{\xi}_T$ grows at a rate faster than $T^{1/2}$, which implies the true copula cannot be approximated accurately by combining estimated single copulas. This is reasonable because the true copula is a mixture copula with $\omega_k > 0$.

Next, we prove (A.3) by using Conditions C.1, C.2, C.3 and C.5. Let $\mathbf{w}^\star = (0,\ldots,0,1)^\top$ and $\widehat{\boldsymbol{\psi}} = (\widehat{w}_1/(1-\widehat{w}_K),\ldots,\widehat{w}_{K-1}/(1-\widehat{w}_K))^\top$. From Conditions C.1-C.3, and the steps used in the proof of (A.17), we have

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

$$
\begin{aligned}
CV_J(\mathbf{w}^\star) &= \left\|\overline{\mathbf{C}}(\mathbf{w}^\star) - \widetilde{\mathbf{C}}\right\|^2 \\
&= \left\|\overline{\mathbf{C}}_K - \widetilde{\mathbf{C}}\right\|^2 \\
&= \left\|\overline{\mathbf{C}}_K - \widehat{\mathbf{C}}_K + \widehat{\mathbf{C}}_K - \mathbf{C}_0 + \mathbf{C}_0 - \widetilde{\mathbf{C}}\right\|^2 \\
&= O_p(1). \qquad \textbf{(A.27)}
\end{aligned}
$$

From Condition C.3, and the steps used in the proof of (A.14), we have

$$
\begin{aligned}
CV_J(\widehat{\mathbf{w}}) &= \left\|\overline{\mathbf{C}}(\widehat{\mathbf{w}}) - \widetilde{\mathbf{C}}\right\|^2 \\
&= \left\|\left\{\widehat{\mathbf{C}}(\widehat{\mathbf{w}}) - \mathbf{C}_0\right\} - \left\{\widehat{\mathbf{C}}(\widehat{\mathbf{w}}) - \overline{\mathbf{C}}(\widehat{\mathbf{w}})\right\} + (\mathbf{C}_0 - \widetilde{\mathbf{C}})\right\|^2 \\
&= L_T(\widehat{\mathbf{w}}) + \left\|\widehat{\mathbf{C}}(\widehat{\mathbf{w}}) - \overline{\mathbf{C}}(\widehat{\mathbf{w}})\right\|^2 + \|\mathbf{C}_0 - \widetilde{\mathbf{C}}\|^2 - 2\left\{\widehat{\mathbf{C}}(\widehat{\mathbf{w}}) - \overline{\mathbf{C}}(\widehat{\mathbf{w}})\right\}^\top (\mathbf{C}_0 - \widetilde{\mathbf{C}}) \\
&\quad -2\left\{\widehat{\mathbf{C}}(\widehat{\mathbf{w}}) - \overline{\mathbf{C}}(\widehat{\mathbf{w}})\right\}^\top \left\{\widehat{\mathbf{C}}(\widehat{\mathbf{w}}) - \mathbf{C}_0\right\} + 2\left\{\widehat{\mathbf{C}}(\widehat{\mathbf{w}}) - \mathbf{C}_0\right\}^\top (\mathbf{C}_0 - \widetilde{\mathbf{C}}) \\
&= L_T^*(\widehat{\mathbf{w}}) + \Pi_T(\widehat{\mathbf{w}}) + \left\|\widehat{\mathbf{C}}(\widehat{\mathbf{w}}) - \overline{\mathbf{C}}(\widehat{\mathbf{w}})\right\|^2 - 2\left\{\widehat{\mathbf{C}}(\widehat{\mathbf{w}}) - \overline{\mathbf{C}}(\widehat{\mathbf{w}})\right\}^\top (\mathbf{C}_0 - \widetilde{\mathbf{C}}) \\
&\quad -2\left\{\widehat{\mathbf{C}}(\widehat{\mathbf{w}}) - \overline{\mathbf{C}}(\widehat{\mathbf{w}})\right\}^\top \left\{\widehat{\mathbf{C}}(\widehat{\mathbf{w}}) - \mathbf{C}_0\right\} + 2\left\{\widehat{\mathbf{C}}(\widehat{\mathbf{w}}) - \mathbf{C}_0\right\}^\top (\mathbf{C}_0 - \widetilde{\mathbf{C}}) + \|\mathbf{C}_0 - \widetilde{\mathbf{C}}\|^2 \\
&= L_T^*(\widehat{\mathbf{w}}) + O_p(T^{1/2}) = \left\|\mathbf{C}^*(\widehat{\mathbf{w}}) - \mathbf{C}_0\right\|^2 + O_p(T^{1/2}) \\
&= \left\|(1-\widehat{\omega}_K)[\mathbf{C}^*(\widehat{\psi}) - \mathbf{C}_0] + \widehat{\omega}_K(\widehat{\mathbf{C}}_K\,|_{\widehat{\boldsymbol{\theta}}=\boldsymbol{\theta}^*} - \mathbf{C}_0)\right\|^2 + O_p(T^{1/2}) \\
&= (1-\widehat{\omega}_K)^2\left\|\mathbf{C}^*(\widehat{\psi}) - \mathbf{C}_0\right\|^2 + 2(1-\widehat{\omega}_K)\widehat{\omega}_K[\mathbf{C}^*(\widehat{\psi}) - \mathbf{C}_0](\widehat{\mathbf{C}}_K\,|_{\widehat{\boldsymbol{\theta}}=\boldsymbol{\theta}^*} - \mathbf{C}_0) \\
&\quad +\widehat{\omega}_K^2\left\|\widehat{\mathbf{C}}_K\,|_{\widehat{\boldsymbol{\theta}}=\boldsymbol{\theta}^*} - \mathbf{C}_0\right\|^2 + O_p(T^{1/2}) \\
&= (1-\widehat{\omega}_K)^2\left\|\mathbf{C}^*(\widehat{\psi}) - \mathbf{C}_0\right\|^2 + O_p(T^{1/2}) + O_p(1) + O_p(T^{1/2}) \\
&= (1-\widehat{\omega}_K)^2\left\|\mathbf{C}^*(\widehat{\psi}) - \mathbf{C}_0\right\|^2 + O_p(T^{1/2}), \qquad \textbf{(A.28)}
\end{aligned}
$$

as $T \to \infty$. Since $\widehat{\mathbf{w}} = \text{argmin}_{\mathbf{w}\in\mathcal{W}} CV_J(\mathbf{w})$, we have $CV_J(\mathbf{w}^\star) \geq CV_J(\widehat{\mathbf{w}})$, which along with the above two results, implies that

$$
(1-\widehat{\omega}_K)^2 = \left\|\mathbf{C}^*(\widehat{\psi}) - \mathbf{C}_0\right\|^{-2} O_p(T^{1/2}),
$$

which, along with Condition C.4, implies (A.3).

A.5. More Simulation Results

In this section we present some additional simulation results. First, we compare the out-of-sample forecast performances between MA and MA without $\sum_{k=1}^K w_k = 1$ constraint (denote this as MA_{un}). For simplicity, we do not add the screening step and fix $J = 4$ in all the simulations. In Table A1, the out-of-sample prediction performances of MA and MA_{un} do not exhibit significant difference because the ratios are all close to 1.

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

Table A2 reports the results of comparisons between MA and MA with the screening step to select K (denoted as MAS). For simplicity, we still fix $J = 4$ in this simulation. Table A2 demonstrates that even though the ratio between MAS and MA is less than 1, such a difference is not statistically significant according to the DM test, indicating that the additional screening step does not effectively improve the performance of MA. Intuitively, when the mixture model is correctly specified, both MA and CW will impose zero weight to the unrelated candidate so the performances between MA and MAS should be quite similar. When the model is misspecified, many candidate copulas often pass the screening step and enter the following MA step. Therefore, one would observe similar performances between MA and MAS no matter whether the true copula is contained in the mixture copula or not.

Table A3 compares the performances between MA with data-driven method selected J and MA with J, respectively, equal to 2, 4, 5, and 10. As one can see, MA with data-driven method selected J outperforms MA with fixed J in most cases and the difference is often statistically significant. This shows that the proposed J selection strategy is useful in improving the performance of MA.

TABLE A1. Ratios of squared prediction losses of MA_{un} (without the $\sum_{k=1}^{K} w_k = 1$ constraint) to MA (with the $\sum_{k=1}^{K} w_k = 1$ constraint) for Type I and Type II simulations. Each simulation is repeated 300 times. J is fixed to 4 for all the simulations. Values in parentheses indicate the p-values obtained through the Diebold–Mariano Test

Type I simulation: Out-of-sample forecast (sample size = 500)		
	$\theta_{Ga} = 0.5, \theta_{Cl} = 5.8, \theta_{Gu} = 5.1$	$\theta_{Ga} = 0.7, \theta_{Cl} = 7.8, \theta_{Gu} = 7.1$
	MA_{un}/MA	MA_{un}/MA
$\omega_{Ga} = 0.5, \omega_{Cl} = 0.5, \omega_{Gu} = 0$	1.0135	1.0397
	(0.79)	(0.77)
$\omega_{Ga} = 0.5, \omega_{Cl} = 0, \omega_{Gu} = 0.5$	1.0272	1.0389
	(0.27)	(0.24)
$\omega_{Ga} = 0, \omega_{Cl} = 0.5, \omega_{Gu} = 0.5$	1.0347	1.0273
	(0.33)	(0.52)
$\omega_{Ga} = 1/3, \omega_{Cl} = 1/3, \omega_{Gu} = 1/3$	0.9536	1.0141
	(0.83)	(0.70)
Type II simulation: Out-of-sample forecast (sample size = 500)		
	$\theta_F = 5.5, \theta_{SJ} = 4.8, \theta_J = 4.5$	$\theta_F = 7.5, \theta_{SJ} = 6.8, \theta_J = 6.5$
	MA_{un}/MA	MA_{un}/MA
$\omega_F = 0.5, \omega_{SJ} = 0.5, \omega_J = 0$	1.0242	1.0367
	(0.23)	(0.68)
$\omega_F = 0.5, \omega_{SJ} = 0, \omega_J = 0.5$	1.0241	1.0135
	(0.47)	(0.52)
$\omega_F = 0, \omega_{SJ} = 0.5, \omega_J = 0.5$	1.0206	1.0109
	(0.84)	(0.95)
$\omega_F = 1/3, \omega_{SJ} = 1/3, \omega_J = 1/3$	1.0137	1.0175
	(0.73)	(0.87)

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

TABLE A2. Ratios of squared prediction losses of MAS to MA for Type I and Type II simulations. MAS denotes the model averaging method with the screening step to select K. Each simulation is repeated 300 times. J is fixed at 4 for all simulations. Values in parentheses indicate the p-values obtained through the Diebold–Mariano Test

Type I simulation: Out-of-sample forecast (sample size = 500)		
	$\theta_{Ga}=0.5,\theta_{Cl}=5.8,\theta_{Gu}=5.1$	$\theta_{Ga}=0.7,\theta_{Cl}=7.8,\theta_{Gu}=7.1$
	MAS/MA	MAS/MA
$\omega_{Ga}=0.5,\omega_{Cl}=0.5,\omega_{Gu}=0$	0.9754	0.9932
	(0.88)	(0.90)
$\omega_{Ga}=0.5,\omega_{Cl}=0,\omega_{Gu}=0.5$	0.9745	0.9941
	(0.90)	(0.92)
$\omega_{Ga}=0,\omega_{Cl}=0.5,\omega_{Gu}=0.5$	0.9875	0.9896
	(0.91)	(0.88)
$\omega_{Ga}=1/3,\omega_{Cl}=1/3,\omega_{Gu}=1/3$	0.9846	0.9837
	(0.89)	(0.93)
Type II simulation: Out-of-sample forecast (sample size = 500)		
	$\theta_F=5.5,\theta_{SJ}=4.8,\theta_J=4.5$	$\theta_F=7.5,\theta_{SJ}=6.8,\theta_J=6.5$
	MAS/MA	MAS/MA
$\omega_F=0.5,\omega_{SJ}=0.5,\omega_J=0$	0.9871	0.9803
	(0.94)	(0.90)
$\omega_F=0.5,\omega_{SJ}=0,\omega_J=0.5$	0.9974	0.9923
	(0.89)	(0.91)
$\omega_F=0,\omega_{SJ}=0.5,\omega_J=0.5$	0.9857	0.9908
	(0.95)	(0.89)
$\omega_F=1/3,\omega_{SJ}=1/3,\omega_J=1/3$	0.9942	0.9936
	(0.94)	(0.92)

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466618000270 Published online by Cambridge University Press

TABLE A3. Ratios of squared prediction losses of MA_{J*} (with data-driven method selected J^*) to MA_J (with a fixed J, $J = 2, 4, 5, 10$) for Type I and Type II simulations. Each simulation is repeated 300 times. Values in parentheses indicate the p-values obtained through the Diebold–Mariano Test

Type I simulation: Out-of-sample forecast (sample size = 500)								
	$\theta_{Ga}=0.5, \theta_{Cl}=5.8, \theta_{Gu}=5.1$				$\theta_{Ga}=0.7, \theta_{Cl}=7.8, \theta_{Gu}=7.1$			
	$MA_{J^*}/MA_{J=2}$	$MA_{J^*}/MA_{J=4}$	$MA_{J^*}/MA_{J=5}$	$MA_{J^*}/MA_{J=10}$	$MA_{J^*}/MA_{J=2}$	$MA_{J^*}/MA_{J=4}$	$MA_{J^*}/MA_{J=5}$	$MA_{J^*}/MA_{J=10}$
$\omega_{Ga}=0.5, \omega_{Cl}=0.5, \omega_{Gu}=0$	0.9769	0.8841	1.0073	0.9765	0.9348	0.9339	0.8931	0.9107
	(0.12)	(0.04)	(0.84)	(0.12)	(0.05)	(0.05)	(0.04)	(0.08)
$\omega_{Ga}=0.5, \omega_{Cl}=0, \omega_{Gu}=0.5$	0.9337	0.8124	0.8675	0.9526	1.0011	0.8357	0.9431	0.8972
	(0.07)	(0.03)	(0.04)	(0.06)	(0.31)	(0.02)	(0.05)	(0.03)
$\omega_{Ga}=0, \omega_{Cl}=0.5, \omega_{Gu}=0.5$	0.9179	0.8325	1.0041	0.9704	0.9321	0.7159	0.9327	0.9172
	(0.06)	(0.03)	(0.61)	(0.11)	(0.08)	(0.01)	(0.07)	(0.06)
$\omega_{Ga}=1/3, \omega_{Cl}=1/3, \omega_{Gu}=1/3$	0.9769	0.9528	0.8709	0.9132	0.9537	0.9921	0.9069	0.8257
	(0.13)	(0.10)	(0.05)	(0.07)	(0.10)	(0.31)	(0.08)	(0.02)
Type II simulation: Out-of-sample forecast (sample size = 500)								
	$\theta_F=5.5, \theta_{SJ}=4.8, \theta_J=4.5$				$\theta_F=7.5, \theta_{SJ}=6.8, \theta_J=6.5$			
	$MA_{J^*}/MA_{J=2}$	$MA_{J^*}/MA_{J=4}$	$MA_{J^*}/MA_{J=5}$	$MA_{J^*}/MA_{J=10}$	$MA_{J^*}/MA_{J=2}$	$MA_{J^*}/MA_{J=4}$	$MA_{J^*}/MA_{J=5}$	$MA_{J^*}/MA_{J=10}$
$\omega_F=0.5, \omega_{SJ}=0.5, \omega_J=0$	0.8851	0.9501	0.8912	0.8537	0.8731	0.9269	0.9104	0.8922
	(0.03)	(0.08)	(0.04)	(0.03)	(0.02)	(0.06)	(0.05)	(0.04)
$\omega_F=0.5, \omega_{SJ}=0, \omega_J=0.5$	0.9177	0.9025	0.8241	0.9030	0.9511	0.8546	0.9279	0.8732
	(0.08)	(0.07)	(0.02)	(0.05)	(0.08)	(0.04)	(0.06)	(0.03)
$\omega_F=0, \omega_{SJ}=0.5, \omega_J=0.5$	1.0091	0.9279	0.8397	0.9012	1.0017	0.9346	0.8162	0.9341
	(0.35)	(0.06)	(0.03)	(0.05)	(0.81)	(0.06)	(0.01)	(0.07)
$\omega_F=1/3, \omega_{SJ}=1/3, \omega_J=1/3$	0.8455	0.8703	0.8768	0.9347	0.9607	0.9225	0.8930	0.9134
	(0.03)	(0.04)	(0.05)	(0.07)	(0.06)	(0.06)	(0.04)	(0.05)