Model and Moment Selection in Factor Copula Models

Fang Duan*†, Hans Manner‡, and Dominik Wied§

†TU Dortmund University, Faculty of Statistics & Ruhr Graduate School in Economics

‡University of Graz, Department of Economics

§University of Cologne, Institute of Econometrics and Statistics

November 11, 2019

Abstract

This paper develops a simultaneous model and moment selection procedure for factor copula models. Since the density of the factor copula is generally not known in closed form, widely used likelihood or moment based model selection criteria cannot be directly applied on factor copulas. The new approach is inspired by the methods for GMM proposed by Andrews (1999) and Andrews & Lu (2001). The consistency of the procedure is proved and Monte Carlo simulations show its good performance in finite samples in different scenarios of sample sizes and dimensions. The impact of the choice of moments in selected regions of the support on model selection and Value-at-Risk prediction are further examined by simulation and an application to a portfolio consisting of ten stocks in the DAX30 index.

Keywords Model Selection; Moment Selection; Factor Copula Model; Value-at-Risk

JEL Classification C38; C52; C58

*Address: Fakultät Statistik, Technische Universität Dortmund, Vogelpothsweg 87, 44227 Dortmund, Germany, Email: duan@statistik.tu-dortmund.de. Fang Duan gratefully acknowledges funding by Ruhr Graduate School in Economics (RGS Econ) and RAG Foundation.

1 Introduction

Copula models are superior in multivariate modeling compared to standard multivariate distributions in terms of their flexibility, e.g., one can construct the joint distribution by separately specifying the marginal distributions and the dependence structure that links the marginal distributions. In financial market modeling, it is of interest to analyze the dependence structure of various financial assets. For example, risk managers would like to know how asset returns co-move during a financial crisis. In such a period one might observe non-zero tail dependence between assets and certain Archimedean or elliptical copulas such as the Clayton or the t copula could be sufficient to capture the hidden trait of fat tails, i.e., the possibility of correlated crashes or booms. In addition, since such dependence might not be symmetric, asymmetric copulas such as the Clayton or the rotated Gumbel copula that allow for lower tail dependence may be favored for the scenario that the dependence between assets is stronger in crashes than in booms. The copula approach is also applied to forecast the Value-at-Risk (VaR) of a portfolio using non-normal joint distributions.

Since one may be interested in the dependence among variables in a large set instead of merely a small number of variables, one might run into problems with the specification of a sufficiently flexible model and the estimation of a large number of parameters in the dependence model. The factor copula model stands out as an elegant tool, as it is capable of representing high dimensional variables with a smaller number of latent factors. In addition to the ability of dimensionality-reduction, factor copula models can impose an informative structure to provide economic interpretations in high dimensional modeling. For example, the movements of the return series in a stock market may possibly be associated with common factors such as the macroeconomic state.

Research on factor copulas has become active in the past decade. The model was originally introduced in credit risk modeling of a portfolio. Andersen & Sidenius (2004) extended the Gaussian copula in a non-linear way to capture the heavy upper tails in portfolio loss distributions as well as correlation skews in CDO tranches. The Gaussian factor copula with a single factor was extended by van der Voort (2005) to a model that avoids the correlation effect of instantaneous defaults by additionally modeling external default risks. One of the most general factor copula models was presented by Joe (2014). This factor copula can be seen as a conditional independence model in which observed variables are conditionally independent given latent factors. This family of factor copulas was further developed by Krupskii & Joe (2013) and Nikoloulopoulos & Joe (2015). More recently, Ivanov et al. (2017) fit a copula autoregressive model to estimate unobservable factors in a dynamic factor model. Oh & Patton (2017) proposed the factor copula model with a linearly additive structure, which serves as a descendant of the model proposed by Hull & White (2004) for the evaluation of default probabilities and default correlations. Due to the fact that a closed form likelihood is not available, the model is estimated using the simulated method of moments, matching different dependence measures in the sample to the ones implied by the model. We focus on the setup in Oh & Patton (2017) throughout this paper.

The main contribution of this paper is to propose a simultaneous model and moment selection procedure for various specifications of factor copula models along with different combinations of moment conditions. Regarding the model selection task, factor copula models can be specified flexibly, i.e., latent factors and the idiosyncratic terms possibly follow distinct distributions as, e.g., in the *Skew t-t* factor model. Given an asymmetric and fat-tailed data generating process (DGP), the *Skew t-t* factor copula is expected to provide a better fit to the data than a model with only Gaussian factors. The issue is whether one can consistently select the true model or an approximately true model from a pool of candidate models. As far as we know, only the specification test of factor copula models proposed by Oh & Patton (2013) partially answered this question in the existing literature, but their approach is less helpful in the comparison of a set of nested models. That is, their approach could fail to detect the most parsimonious model which potentially performs better in terms of prediction compared to highly parametrized models.

Since the Simulated Method of Moments (SMM) serves as a popular method in the estimation of the parameters in factor copula models, the selection of correct moments, i.e., whether the moment conditions hold asymptotically, becomes crucial. In addition, the choice of moment conditions has an impact on model selection. In fact, it is unrealistic to take infinitely many moment conditions into account. One might be interested in characterizing a certain group of moments which are of particular interest in specific applications, e.g., the lower tail of the distribution when one deals with problems such as VaR estimation or prediction. To this end, we propose a feasible procedure of simultaneous selection of moments and models for factor copulas. From a practical point of view in risk management, we investigate the interplay between the selection of moments and models in backtesting VaR forecasts. Considering that the true model is generally unknown in empirical work, we point out the possibility of obtaining reasonable risk forecasts based on misspecified models using a specific set of moments in the lower tail of the distribution.

The rest of this paper is organized as follows. We briefly introduce factor copula models in Section 2. In Section 3, we review the estimation method for factor copulas and necessary assumptions for consistent parameter estimation, after which we discuss the model and moment selection criterion and its consistency in Section 4. In Section 5 we present simulation studies, followed by an empirical application of the procedure in Section 6. Lastly, we offer our conclusions in Section 7.

2 Factor Copula Models

Factor copulas are constructed on a set of latent variables with a factor structure. Each of the latent variables is modeled as a linear combination of a smaller number of latent factors and an idiosyncratic term for individual characteristics. Consider the following model with one common factor:

$$\begin{aligned} X_i &= \beta_i Z + \epsilon_i, i = 1, 2, \dots, N, \\ Z &\sim F_Z(\boldsymbol{\theta}_Z), \\ \epsilon_i &\sim \text{iid } F_\epsilon(\boldsymbol{\theta}_\epsilon), \text{ and } \epsilon_i \perp Z, \forall i \\ \boldsymbol{X} = (X_1, \cdots, X_N)' &\sim F_{\boldsymbol{X}} = \boldsymbol{C}(G_1(\boldsymbol{\theta}), \dots, G_N(\boldsymbol{\theta}); \boldsymbol{\theta}), \end{aligned} \tag{1}$$

where Z is the common factor. β_i is the factor loading of the common factor associated with i-th latent variable $X_i, i = 1, \dots, N$, whereas $\boldsymbol{\theta} = (\boldsymbol{\theta}'_Z, \boldsymbol{\theta}'_\epsilon, \beta_1, \dots, \beta_N)'$ denotes the vector of the copula parameters. Let p be the number of parameters to be estimated. The corresponding factor copula $\boldsymbol{C}(\boldsymbol{\theta})$ is used as the model of the copula of the observable variables $\boldsymbol{Y}$. Note that the marginal distributions of the latent variables X_i are not necessarily identical to those of the observable variables Y_i. The latent factor structure is employed to specify the copula and the estimation of the marginal distributions is left to, e.g., semiparametric models.

Although we impose a factor structure to obtain a more parsimonious copula model in high dimensions, a potential challenge in the estimation of such a model remains: N factor loadings have to be estimated, which is still a difficult task in large dimensions. In response, Oh & Patton (2017) suggested that one can sort the latent variables into groups and formulate a so-called block equi-dependence model. The intuition of such a clustering strategy can be implied by the scenario that, e.g., the stock prices in the same industry would have an analogous reaction to the latent factor representing the macroeconomic status. The aforementioned specifications could be extended to include multiple latent factors. A similar setting of this multiple factor structure in the general heterogeneous factor model can be found in Ansari et al. (2002). The nonlinear factor copula model relaxes the assumption of linearity and additivity; see Oh & Patton (2017) for discussions in greater detail. Throughout the paper, we restrict attention to the equi-dependence model with a single common factor as in specification (1) with identical factor loadings for all variables.

In our context, the factor copula of $\boldsymbol{X}$ is generally not known in closed form. A Gaussian copula is implied by specification (1) when the distributions of the common factor Z and the idiosyncratic term $\epsilon_i, i = 1, \ldots, N$, are Gaussian, or equivalently the joint distribution of $\boldsymbol{X}$ is multivariate Gaussian. However, the joint distribution of $\boldsymbol{X}$ and thus the copula of $\boldsymbol{X}$ are typically not known in closed form if Z and ϵ_i come from different families of distribution. To specify a factor copula, it is crucial to choose the distributions of the common factors and the idiosyncratic term in the first place. Among assorted options of the distributions of common and idiosyncratic variables, a favorable choice is the skewed Student t distribution, see Hansen (1994). This distribution is useful for modeling tail dependence and asymmetric dependence because it allows for tail thickness and skewness:

$$t(z|\nu,\lambda) = \begin{cases} bc(1 + \frac{1}{\nu-2}(\frac{bz+a}{1-\lambda})^2)^{-(\nu+1)/2} & z < -a/b, \\ bc(1 + \frac{1}{\nu-2}(\frac{bz+a}{1+\lambda})^2)^{-(\nu+1)/2} & z \geq -a/b, \end{cases}$$

where $2 < \nu < \infty$, $-1 < \lambda < 1$. a, b and c are constants: $a = 4\lambda c(\frac{\nu+2}{\nu-1})$, $b^2 = 1+3\lambda^2-a^2$

and $c = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi(\nu-2)}\Gamma(\nu/2)}$. It is parameterized by two parameters: ν for tail thickness and λ for skewness. A lower ν indicates fatter tails. When $\lambda = 0$, it collapses to the standard Student t distribution. $\lambda > 0$ indicates that the variable is skewed to the right, whereas $\lambda < 0$ implies the left skewed variable.

To show how the tail and the skewness parameters affect the dependence structure implied by the model, Figure 1 shows $T = 5000$ bivariate random samples from a *t-t* and a *skewt-t* model with factor loading $\beta = 1$, shape parameter $\nu = 2$ and skewness parameter $\lambda = -0.5$. For the *t-t* model in Figure 1(a) two clusters emerge at both tails in a symmetric way. The asymmetric dependence structure is apparent in Figure 1(b) as the skewness parameter comes into play.

Figure 1 about here

3 Estimation

3.1 Dependence Measures

Since it is necessary to choose dependence measures which can be used in the SMM estimation step, we firstly introduce two popular choices of dependence measures: Spearman's rho and quantile dependence; see Nelsen (2006), Genest & Favre (2007) and Joe (2014). Consider a pair of variables η_i and η_j. Spearman's rank correlation is defined as

$$\rho_S^{ij} = 12 \int_0^1 \int_0^1 uv dC_{ij}(u, v) - 3, \tag{2}$$

where $u = F_i(\eta_i)$, $v = F_j(\eta_j)$ and $C_{ij}(u, v)$ is the copula of the pair (η_i, η_j). In contrast to Spearman's rho which can be considered as a global measure of the dependence structure, quantile dependence focuses on the dependence structure in specific regions of the support of a distribution. The quantile dependence between the pair (η_i, η_j) at quantile q is defined as the conditional probability that $F_i(\eta_i)$ is smaller (greater) than q given that $F_j(\eta_j)$ is smaller (greater) than q for the lower (upper) tail. If q goes to zero, then the quantile dependence is the probability that $F_i(\eta_i)$ is extremely small given that $F_j(\eta_j)$ is also extremely small. It is practically useful for measuring concurrent extreme events in financial markets. The quantile dependence at quantile

q between η_i and η_j is defined as

$$\lambda_q^{ij} = \begin{cases} P[F_i(\eta_i) \leq q | F_j(\eta_j) \leq q], & \text{for } q \in (0, 0.5], \\ P[F_i(\eta_i) > q | F_j(\eta_j) > q], & \text{for } q \in (0.5, 1). \end{cases} \tag{3}$$

The estimators of $\hat{\rho}_S^{ij}$ and $\hat{\lambda}_q^{ij}$ are

$$\begin{aligned} \hat{\rho}_S^{ij} &= \tfrac{12}{T} \textstyle\sum_{t=1}^{T} \hat{F}_i(\hat{\eta}_{it}) \hat{F}_j(\hat{\eta}_{jt}) - 3, \\ \hat{\lambda}_q^{ij} &= \begin{cases} \frac{1}{Tq} \sum_{t=1}^{T} \mathbf{1}_{[\hat{F}_i(\hat{\eta}_{it}) \leq q, \hat{F}_j(\hat{\eta}_{jt}) \leq q]}, & \text{for } q \in (0, 0.5], \\ \frac{1}{T(1-q)} \sum_{t=1}^{T} \mathbf{1}_{[\hat{F}_i(\hat{\eta}_{it}) > q, \hat{F}_j(\hat{\eta}_{jt}) > q]}, & \text{for } q \in (0.5, 1), \end{cases} \end{aligned} \tag{4}$$

where $\hat{F}_i(\hat{\eta}_{it})$ and $\hat{F}_j(\hat{\eta}_{jt})$ are the empirical distribution functions of the variables $\hat{\eta}_{it}$ and $\hat{\eta}_{jt}$, respectively. Following the definition in Oh & Patton (2017), linear combinations of these dependence measures can be used as moment conditions in the SMM estimation. In the equi-dependence model, the vector of dependence measures is the average over the pairwise measures

$$\boldsymbol{m}_T = \begin{pmatrix} \rho_S & \lambda_{q_1} & \dots & \lambda_{q_{r-1}} \end{pmatrix} = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \begin{pmatrix} \rho_S^{ij} & \lambda_{q_1}^{ij} & \dots & \lambda_{q_{r-1}}^{ij} \end{pmatrix}', \tag{5}$$

where $q_1, \dots, q_{r-1} \in (0, 1)$ denote the quantiles of interest.

3.2 Simulated Method of Moments

The factor copula does not generally have a closed-form likelihood and Maximum Likelihood Estimation (MLE) is therefore only applicable in a few cases; see, e.g., Krupskii et al. (2018). In cases where MLE is not possible, however, the Simulated Method of Moments (SMM) can be applied; see McFadden (1989). Oh & Patton (2013) integrated this method into the estimation of factor copula models. Consider the following data generating process (DGP) which allows for a time-varying conditional mean and conditional variance for each variable. This setting can also be found in Chen & Fan (2006), Oh & Patton (2013) and Rémillard (2017):

$$\boldsymbol{Y}_t = \boldsymbol{\mu}_t(\boldsymbol{\phi}) + \boldsymbol{\sigma}_t(\boldsymbol{\phi}) \boldsymbol{\eta}_t, \tag{6}$$

with

$$\begin{aligned} \boldsymbol{\mu}_t(\boldsymbol{\phi}) &= [\mu_{1,t}(\boldsymbol{\phi}), \dots, \mu_{N,t}(\boldsymbol{\phi})]', \\ \boldsymbol{\sigma}_t(\boldsymbol{\phi}) &= \operatorname{diag}[\sigma_{1,t}(\boldsymbol{\phi}), \dots, \sigma_{N,t}(\boldsymbol{\phi})]', \\ \boldsymbol{\eta}_t &\sim_{iid} \boldsymbol{F}_\eta = \boldsymbol{C}(F_1, \dots, F_N; \boldsymbol{\theta}), \end{aligned}$$

where $\boldsymbol{\mu}_t(\boldsymbol{\phi})$ and $\boldsymbol{\sigma}_t(\boldsymbol{\phi})$ are $\mathcal{F}_{t-1}$ measurable and independent of $\boldsymbol{\eta}_t$, and $\mathcal{F}_{t-1}$ is the σ-Algebra containing the past information of $\{\boldsymbol{Y}_{t-1}, \boldsymbol{Y}_{t-2}, \ldots\}$. The parameter vector $\boldsymbol{\phi}$ controls the dynamics in the marginal distributions. When this parameter vector is known or both $\boldsymbol{\mu}_t(\boldsymbol{\phi})$ and $\boldsymbol{\sigma}_t(\boldsymbol{\phi})$ are constant, this model describes i.i.d time series. The estimation of the copula parameters $\boldsymbol{\theta}$ is based on the estimated standardized residuals $\hat{\boldsymbol{\eta}}_t = \boldsymbol{\sigma}_t^{-1}(\hat{\boldsymbol{\phi}})(\boldsymbol{Y}_t - \boldsymbol{\mu}_t(\hat{\boldsymbol{\phi}})), t = 1, \ldots, T$, and simulations from the multivariate distribution $\boldsymbol{F}_x$. Define $\hat{\boldsymbol{m}}_T$ as the $r \times 1$ vector of dependence measures computed from the estimated standardized residuals $\hat{\boldsymbol{\eta}}_t, t = 1, \ldots, T$, and $\tilde{\boldsymbol{m}}_S(\boldsymbol{\theta})$ as a vector of dependence measures obtained from $\boldsymbol{X}_s, s = 1, \ldots, S$, i.e., S simulations from the parametric joint distribution $\boldsymbol{F}_x(\boldsymbol{\theta})$. The SMM estimator is defined as

$$\hat{\boldsymbol{\theta}}_{T,S} = \arg\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \boldsymbol{g}_{T,S}(\boldsymbol{\theta})' \hat{\boldsymbol{W}}_T \boldsymbol{g}_{T,S}(\boldsymbol{\theta}), \tag{7}$$

where $\boldsymbol{g}_{T,S}(\boldsymbol{\theta})$ denotes the difference between $\hat{\boldsymbol{m}}_T$ and $\tilde{\boldsymbol{m}}_S(\boldsymbol{\theta})$,

$$\boldsymbol{g}_{T,S}(\boldsymbol{\theta}) = \hat{\boldsymbol{m}}_T - \tilde{\boldsymbol{m}}_S(\boldsymbol{\theta}). \tag{8}$$

$\hat{\boldsymbol{W}}_T$ is a positive definite weighting matrix. Intuitively, one needs to find the vector of simulated dependence measures with an appropriate parameter vector $\hat{\boldsymbol{\theta}}_{T,S}$ which is able to represent the dependence structure based on the empirical data. Next we restate some necessary assumptions for the consistency of the SMM estimator.

Assumption 1 *The sample rank correlation and quantile dependence converge in probability to their theoretical counterparts, respectively, i.e.,* $\hat{\rho}_S^{ij} \to_p \rho_S^{ij}$ *and* $\hat{\lambda}_q^{ij} \to_p \lambda_q^{ij}$*, for pair* i, j*,* $\forall q \in (0, 1)$*.*

This assumption implies weak convergence of $\boldsymbol{g}_{T,S}(\boldsymbol{\theta})$:

$$\boldsymbol{g}_{T,S}(\boldsymbol{\theta}) \equiv \hat{\boldsymbol{m}}_T - \tilde{\boldsymbol{m}}_S(\boldsymbol{\theta}) \to_p \boldsymbol{g}_0(\boldsymbol{\theta}) \equiv \boldsymbol{m}_0(\boldsymbol{\theta}_0) - \boldsymbol{m}_0(\boldsymbol{\theta}), \forall \boldsymbol{\theta} \in \boldsymbol{\Theta}, T, S \to \infty, \tag{9}$$

where $\boldsymbol{\theta}_0$ denotes the true value of the parameter vector. This assumption requires that $\boldsymbol{F}_\eta$ and $\boldsymbol{F}_x$ are continuous, and the partial derivative of each bivariate marginal copula C_{ij} of $\boldsymbol{C}$ is continuous with respect to u_i and u_j, as in Assumption 1 in Oh & Patton (2013). When one employs standardized residuals in the estimation of the copula, it is necessary to control the estimation error arising from the estimation of the conditional means and conditional variances. Assumption 2 in Oh & Patton (2013) or equivalently assumptions A1-A6 in Rémillard (2017) enable us to prove the weak convergence of the empirical copula process based on standardized residuals to its theoretical limit. We make one additional assumption for the consistency of the SMM estimator, which is identical to Assumption 3 in Oh & Patton (2013):

Assumption 2 *(i)* $\boldsymbol{g}_0(\boldsymbol{\theta}) \neq 0$ *for* $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$.

(ii) $\boldsymbol{\Theta}$ *is compact.*

(iii) Every bivariate marginal copula $C_{ij}(u_i, u_j; \boldsymbol{\theta})$ *of* $\boldsymbol{C}(\boldsymbol{\theta})$ *on* $(u_i, u_j) \in (0,1) \times (0,1)$ *is Lipschitz continuous on* $\boldsymbol{\Theta}$.

(iv) $\hat{\boldsymbol{W}}_T \to_p \boldsymbol{W}_0$, *where* $\hat{\boldsymbol{W}}_T$ *is* $O_p(1)$ *and* $\boldsymbol{W}_0$ *is some positive definite weighting matrix.*

Under Assumptions 1 and 2, the SMM estimator is a consistent estimator:

$$\hat{\boldsymbol{\theta}}_{T,S} \to_p \boldsymbol{\theta}_0, \text{ as } T, S \to \infty.$$

Oh & Patton (2013) also established the asymptotic normality of the SMM estimator under correct model specification.

In the next section we propose a model and moment selection procedure for factor copula models estimated by SMM. Since it is possible that some or all of the model candidates are misspecified, it is necessary to consider the SMM estimator under model misspecification. Oh & Patton (2013) also considered the misspecified case, but not the asymptotic distribution in this situation, since the pseudo-true SMM estimator depends on the limit distribution of the weighting matrix, and an alternative procedure for establishing stochastic equicontinuity of the objective function would be required. They adapted the definition of nonlocal misspecification in the GMM context in Hall & Inoue (2003) to the SMM world. This type of misspecification is defined as the case that there exists no such $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ that $\boldsymbol{g}_0(\boldsymbol{\theta}) = 0$.

Definition 1 *The pseudo-true value of the SMM estimator under misspecification is defined as* $\boldsymbol{\theta}_*(\boldsymbol{W}_0) = \arg\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \boldsymbol{g}_0(\boldsymbol{\theta})' \boldsymbol{W}_0 \boldsymbol{g}_0(\boldsymbol{\theta})$.

We recapitulate the additional assumptions for the consistency of the SMM estimator under misspecification:

Assumption 3 *(i) Under nonlocal misspecification of model,* $||\boldsymbol{g}_0(\boldsymbol{\theta})|| > 0, \forall \boldsymbol{\theta} \in \boldsymbol{\Theta}$.

(ii) $\exists \boldsymbol{\theta}_*(\boldsymbol{W}_0) \in \boldsymbol{\Theta}$ *such that* $\boldsymbol{g}_0(\boldsymbol{\theta}_*(\boldsymbol{W}_0))'\boldsymbol{W}_0\boldsymbol{g}_0(\boldsymbol{\theta}_*(\boldsymbol{W}_0)) < \boldsymbol{g}_0(\boldsymbol{\theta})'\boldsymbol{W}_0\boldsymbol{g}_0(\boldsymbol{\theta}), \forall \boldsymbol{\theta} \in \boldsymbol{\Theta} \backslash \{\boldsymbol{\theta}_*(\boldsymbol{W}_0)\}$.

Under Assumptions 1,2 and 3, we have $\hat{\boldsymbol{\theta}}_{T,S} \to_p \boldsymbol{\theta}_*(\boldsymbol{W}_0)$ as $T, S \to \infty$.

4 A Model and Moment Selection Procedure

4.1 Existing Approaches

In general, in order to handle the model selection problem in a parametric econometric model, one tends to assess the goodness-of-fit and penalize the excessive usage of parameters among competing models using likelihood-based information criteria; see Hastie et al. (2009) and Hannan & Quinn (1979). It is a common approach to apply the Akaike information criterion (AIC) in the selection of parametric copula models, see, e.g., Dias & Embrechts (2004). In a simulation study, Manner (2007) showed good performance of the AIC model selection procedure for copula models which possess symmetric and asymmetric tail behavior. Chen & Fan (2006) extended a pseudo likelihood ratio test for model selection between two semiparametric copula-based multivariate dynamic models under misspecification. Another appealing approach to model selection relies on Goodness-of-Fit (GoF) tests, see Patton (2012) for a general review. Two widely used GoF tests are the Kolmogorov-Smirnov (KS) and the Crámer-von-Mises (CvM) tests. Rémillard (2017) pointed out that one can use a CvM-type statistic, which is more powerful and easier to compute compared to the KS statistic, to test if the parametric family of copulas under consideration is correctly specified. Note that a distinction needs to be made between model comparison and selection, and GoF testing, which are two different problems. In this paper we are concerned with the problem of model comparison and selection, but we base our approach on a statistic that is originally used for model evaluation.

Unfortunately, because the density of the factor copula and its corresponding likelihood is not known in closed form, the aforementioned methods can not be directly borrowed to deal with the model selection problem. Oh & Patton (2013) suggest that the J test of over-identifying restrictions can be used as a specification test of the factor copula model when the number of moments is greater than the number of copula parameters. The statistic and its limit distribution are given as follows:

$$J_{T,S} := \min(T,S)\boldsymbol{g}_{T,S}(\hat{\boldsymbol{\theta}}_{T,S})'\hat{\boldsymbol{W}}_T\boldsymbol{g}_{T,S}(\hat{\boldsymbol{\theta}}_{T,S}) \to_d \boldsymbol{u}'\boldsymbol{A}_0'\boldsymbol{A}_0\boldsymbol{u}, \quad T,S\to\infty, \tag{10}$$

where $\boldsymbol{u}\sim N(0,\boldsymbol{I})$ and $\boldsymbol{A}_0 := \boldsymbol{W}_0^{1/2}\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{R}_0$. $\boldsymbol{\Sigma}_0$ denotes the asymptotic variance of $\hat{\boldsymbol{m}}_T$ and

$$\boldsymbol{R}_0 = \boldsymbol{I} - \boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{G}_0(\boldsymbol{G}_0'\boldsymbol{W}_0\boldsymbol{G}_0)^{-1}\boldsymbol{G}_0'\boldsymbol{W}_0\boldsymbol{\Sigma}_0^{1/2},$$

where $\boldsymbol{G}_0$ is the derivative of $\boldsymbol{g}_0(\boldsymbol{\theta})$. If one uses the efficient weighting matrix, i.e., $\hat{\boldsymbol{W}}_T = \hat{\boldsymbol{\Sigma}}_{T,B}^{-1}$, where $\hat{\boldsymbol{\Sigma}}_{T,B}$ is the iid bootstrap estimator of the covariance matrix $\boldsymbol{\Sigma}_0$, the limit distribution turns out to be the chi-squared distribution χ^2_{r-p}. Recall that r is the number of moments used in the estimation and p is the number of estimated parameters. If a weighting matrix different from the efficient weighting matrix, such as the identity matrix, is employed, the asymptotic distribution of the J test statistics is nonstandard and the corresponding critical values can be obtained via simulation. Unfortunately, the J test appears to be less helpful to choose less parametrized model, which could perform better for the task of prediction.

When it comes to moment selection for factor copula models, the literature is rather scarce. Clearly in our context, moment conditions are constructed on dependence measures, i.e., moments under consideration are based on the linear combination of Spearman's rho and quantile dependence. Some moment conditions may be correct whereas others may be incorrect. The selection of correct moments has an interplay with model selection: under the true model, with increasing inclusion of moments, the procedure we propose below tends to select all moment conditions asymptotically. If the model is not correctly specified, moment conditions which are continuously included would hardly jointly hold and only some of them will be selected. On the one hand, it would be valuable to determine a certain set of 'useful' moment conditions, e.g., quantile dependence at certain quantiles, to consistently and efficiently select the true model. On the other hand, if the set of moments is predetermined in a specific region of the support, e.g., quantile dependence measures restricted to certain quantiles, whether the true model is selected or not remains an issue. We would like to seek a method for the consistent selection of correct moments in addition to the selection of the true model. In the case that all candidate models are misspecified, on the other hand, we would like our method to select the combination of model and moment conditions that is most appropriate for the data of interest.

4.2 A New Selection Procedure

Our proposed procedure is inspired by the moment selection procedure for Generalized Methods of Moments (GMM) originally proposed by Andrews (1999). The simultaneous selection procedure for moments and models in GMM by Andrews & Lu (2001) also sheds light on the possibility of the model and moment selection of factor copula models constructed on J test statistics. Our selection criterion measures the degree of model fit, penalizing a larger number of model parameters, and rewards the selection vectors that use more over-identifying restrictions in the model. This is an analogous procedure to classical likelihood based information criteria. It measures the model fit based on the J statistic in (10) with a bonus term rewarding the selection of more moment conditions and fewer parameters.

Following Andrews & Lu (2001), the starting point is a finite set of r moments and a p-dimensional vector of parameters $\boldsymbol{\theta}$, which corresponds to different models. The crucial component of the new procedure are pairs of model and moment selection vectors $(\boldsymbol{b}, \boldsymbol{c}) \in \{0,1\}^p \times \{0,1\}^r$, which only contain zeros and ones. For instance, if the i-th element of $\boldsymbol{c}$ equals one, then the i-th moment condition is included in the SMM estimation procedure, while it is excluded if it is equal to zero. Similarly, if the i-th element of $\boldsymbol{b}$ equals one, then the i-th element of the parameter vector $\boldsymbol{\theta}$ is estimated, while it is set to zero and not estimated if the i-th element of $\boldsymbol{b}$ equals zero.

As an example, the selection vector $\boldsymbol{c} = (1,1,0,0,0,0,1,1)'$ of the moments $\boldsymbol{m} = (\lambda_{0.05}, \lambda_{0.1}, \lambda_{0.25}, \lambda_{0.45}, \lambda_{0.55}, \lambda_{0.75}, \lambda_{0.9}, \lambda_{0.95})'$ only chooses quantile dependence measures at both tails to construct moment conditions. Clearly, the construction allows for nested models. For example, $\boldsymbol{b} = (1,1)'$ of $\boldsymbol{\theta} = (\nu^{-1}, \lambda)'$ gives a *Skew t-t* factor copula model whereas in this case $\boldsymbol{b} = (1,0)'$ implies a *t-t* factor copula. On the other hand, this setup also incorporates non-nested models. It is convenient to stack parameter vectors from two non-nested models into a single vector, and let the model selection vector assign the corresponding positions of parameters in the first model with ones and the rest with zeros to select out the first model and vice versa.

Define the set

$$\mathcal{M} = \{(\boldsymbol{b}, \boldsymbol{c}) \in \{0,1\}^p \times \{0,1\}^r\}.$$

Furthermore, let $\boldsymbol{\theta}^b$ be the element-wise product of $\boldsymbol{\theta}$ and $\boldsymbol{b}$, and $\boldsymbol{g}^c_{T,S}(\boldsymbol{\theta})$ are the moment conditions specified by $\boldsymbol{c}$. Let $|\boldsymbol{b}| = \sum_{i=1}^p b_i$ denote the number of parameters to be estimated in the vector $\boldsymbol{b}$, and $|\boldsymbol{c}| = \sum_{j=1}^r c_j$ denote the number of moments selected in the vector $\boldsymbol{c}$. We make the following assumption for ensuring that the model parameters are overidentified. This appears to be natural as otherwise one could match every moment condition and it would not be possible to identify wrong models. Moreover, this assumption guarantees that we consider at least one moment condition.

Assumption 4 *The number of moments is greater than the number of parameters, i.e.,* $|\boldsymbol{c}| - |\boldsymbol{b}| > 0$.

The definition of selection vectors and a series of sets consisting of these vectors are borrowed from Andrews & Lu (2001). Denote the selection vector of correct moments as $\boldsymbol{c}^0$ and the selection vector of the correct model as $\boldsymbol{b}^0$, respectively. $\boldsymbol{c}^0$ selects the moments for which $\boldsymbol{g}^c_0(\boldsymbol{\theta}) = 0$, as $T, S \to \infty$ and $\boldsymbol{b}^0$ selects the model associated with the true DGP. We initially define a subset containing all selection vectors of models and moments such that $\boldsymbol{g}_{T,S}(\boldsymbol{\theta})$ equals zero asymptotically and that the parameters are overidentified:

$$\mathcal{M}_1 = \{(\boldsymbol{b}, \boldsymbol{c}) \in \mathcal{M} : |\boldsymbol{c}| - |\boldsymbol{b}| > 0, \boldsymbol{g}^c_{T,S}(\boldsymbol{\theta}^b) \to_p \mathbf{0}, \text{ as } T, S \to \infty, \text{ for some } \boldsymbol{\theta}^{\boldsymbol{b}} \in \boldsymbol{\Theta}\}.$$

We further define a subset of $\mathcal{M}_1$ containing the model and moment selection vectors with maximum number of over-identifying restrictions:

$$\mathcal{M}_2 = \{(\boldsymbol{b}, \boldsymbol{c}) \in \mathcal{M}_1 : |\boldsymbol{c}| - |\boldsymbol{b}| \geq |\tilde{\boldsymbol{c}}| - |\tilde{\boldsymbol{b}}|, \forall (\tilde{\boldsymbol{b}}, \tilde{\boldsymbol{c}}) \in \mathcal{M}_1\}.$$

Next, consider the parameter space of $(\hat{\boldsymbol{b}}, \hat{\boldsymbol{c}})$, denoted as $\mathcal{P} \subset \mathcal{M}$, where $(\hat{\boldsymbol{b}}, \hat{\boldsymbol{c}})$ is a generic estimator of model and moment selection vectors. This is the set of model and moment selection vectors that is chosen by the researcher in practice, having a limited set of model specifications and moment conditions in mind. The parameter space $\mathcal{P}$ is a smaller space than $\mathcal{M}$ ensuring that the estimator $(\hat{\boldsymbol{b}}, \hat{\boldsymbol{c}})$ has good finite sample behavior and is computationally feasible. For instance, when moment conditions are comprised of quantile dependencies $\lambda_q, q \in (0,1)$, it is infeasible to consider infinitely many quantile dependencies on the grid over q. Using a large number of quantile dependence measures would lead to a cumbersome computational burden in the estimation of $(\hat{\boldsymbol{b}}, \hat{\boldsymbol{c}})$. The parameter space $\mathcal{P}$ can also incorporate the information that certain moments are known to be correct and certain parameters in the models are known to be non-zero.

Define $\mathcal{MP}_1 = \mathcal{P} \cap \mathcal{M}_1$ as the set of pairs of selection vectors in the parameter space $\mathcal{P}$ that select the models and moments that are asymptotically equal to zero evaluated at some parameter vectors. We further define a set of selection vectors in $\mathcal{MP}_1$ as

$$\mathcal{MP}_2 = \{(\hat{\boldsymbol{b}}, \hat{\boldsymbol{c}}) \in \mathcal{MP}_1 : |\hat{\boldsymbol{c}}| - |\hat{\boldsymbol{b}}| \geq |\tilde{\boldsymbol{c}}| - |\tilde{\boldsymbol{b}}|, \forall (\tilde{\boldsymbol{b}}, \tilde{\boldsymbol{c}}) \in \mathcal{MP}_1\},$$

which is the set of selection vectors in $\mathcal{MP}_1$ with the maximum number of over-identifying restrictions.

The sets $\mathcal{MP}_1$ and $\mathcal{MP}_2$ can be both empty or non-empty. In the former case, there is no pair $(\hat{\boldsymbol{b}}, \hat{\boldsymbol{c}})$ in the parameter space with over-identified parameters which fits to the data generating process, i.e., which ensures that $\boldsymbol{g}_{T,S}(\boldsymbol{\theta})$ equals zero asymptotically. We will consider this case separately, but first focus on the standard non-empty case.[1]

The selection procedure in Andrews & Lu (2001) is constructed on the J test statistic. Similarly, our simultaneous model and moment selection criterion is defined as:

$$MSC_{T,S}(\boldsymbol{\theta}^{\boldsymbol{b}}, \boldsymbol{c}) = J^{\boldsymbol{c}}_{T,S}(\boldsymbol{\theta}^{\boldsymbol{b}}) - (|\boldsymbol{c}| - |\boldsymbol{b}|)\kappa_{T,S}, \tag{11}$$

where $J^{\boldsymbol{c}}_{T,S}(\boldsymbol{\theta}^{\boldsymbol{b}}) = \min(T,S)\boldsymbol{g}^{\boldsymbol{c}}_{T,S}(\boldsymbol{\theta}^{\boldsymbol{b}})'\hat{\boldsymbol{W}}^{c}_{T}\boldsymbol{g}^{\boldsymbol{c}}_{T,S}(\boldsymbol{\theta}^{\boldsymbol{b}})$. $|\boldsymbol{c}|$ denotes the number of moment conditions used in the SMM, which are based on the selected dependence measures under consideration such as Spearman's rho and quantile dependence $\lambda_q, q \in (0,1)$,

[1]Note that even if $\boldsymbol{g}_{T,S}(\boldsymbol{\theta})$ equals zero asymptotically, the moment conditions could have been derived from a model which is different from the actual data generating process. This can arise if two different models lead to the same moment conditions, as it is the case in certain copula models, see Fredricks & Nelsen (2007). It is the task of the researcher to choose the moment conditions such that sufficiently accurate statements about the data generating process are possible.

the dependence measured used throughout this paper. The penalizing factor $\kappa_{T,S}$ can be specified in the following three ways, which are analogous to classical information-based criteria adapted to our setting:

$$\begin{aligned} \text{SMM-AIC} : \kappa_{T,S} &= 2, \\ \text{SMM-BIC} : \kappa_{T,S} &= \ln(\min(T,S)), \\ \text{SMM-HQIC} : \kappa_{T,S} &= Q \ln\ln(\min(T,S)) \text{ for some } Q > 2. \end{aligned}$$

The estimator of the vector of the model and moment selection based on MSC is determined by minimizing the MSC criterion, i.e.,

$$(\hat{\boldsymbol{b}}_{MSC}, \hat{\boldsymbol{c}}_{MSC}) := \arg \min_{(\boldsymbol{b},\boldsymbol{c})\in\mathcal{P}} \text{MSC}_{T,S}(\boldsymbol{\theta}^{\boldsymbol{b}}, \boldsymbol{c}). \tag{12}$$

The intuition behind this criterion is as follows. In the case of a correctly specified model, the J statistic converges to a random variable by construction and the more parsimonious model with more correct moments will be preferred by the procedure. In contrast, when the model is misspecified, the J statistic diverges asymptotically with the divergence being faster in the case of more incorrect moments, which enlarge the J statistic.

The consistency of the MSC estimator for the selection vector requires the following two assumptions.

Assumption 5 $\kappa_{T,S} = o(\min\{T,S\})$ *and* $\kappa_{T,S} \to \infty$*, as* $T, S \to \infty$.

Assumption 6 $\mathcal{MP}_2 = \{(\boldsymbol{b}^0, \boldsymbol{c}^0)\}$*, where* $\boldsymbol{b}^0$ *is the model selection vector for the true model.*

Assumption 5 ensures that the penalty term diverges to ∞. Assumption 6 guarantees that the set contains one single element, i.e., the selection vector of the true model and correct moments. It rules out the case that $\mathcal{MP}_2$ is empty but also that it contains multiple elements. For example, given a certain set of moment conditions, suppose there exists a selection vector $(\boldsymbol{b}^*, \boldsymbol{c}^*) \in \mathcal{MP}_1$ with $(\boldsymbol{b}^*, \boldsymbol{c}^*) \neq (\boldsymbol{b}^0, \boldsymbol{c}^0)$, but this selection vector has the same number of over-identifying conditions as the one which represents the true model, i.e., $|\boldsymbol{c}^*| - |\boldsymbol{b}^*| = |\boldsymbol{c}^0| - |\boldsymbol{b}^0|$. Both $(\boldsymbol{b}^*, \boldsymbol{c}^*)$ and $(\boldsymbol{b}^0, \boldsymbol{c}^0)$ are included in $\mathcal{MP}_2$ and the selection procedure could fail to discriminate between those two vectors. Assumption 6 also implies that the selection vector $(\boldsymbol{b}^0, \boldsymbol{c}^0)$, which points to the true model along with a set of correct moments, lies in $\mathcal{P}$. Hence, $\mathcal{M}$ must also contain $(\boldsymbol{b}^0, \boldsymbol{c}^0)$.

Proposition 1 *Suppose Assumptions 1 to 6 hold.* $(\hat{\boldsymbol{b}}_{MSC}, \hat{\boldsymbol{c}}_{MSC}) = (\boldsymbol{b}^0, \boldsymbol{c}^0) \in \mathcal{P}, wp \to 1$*, as* $T, S \to \infty$.

The proof of Proposition 1 can be found in the appendix. $wp \to 1$ denotes that the probability goes to 1 as $T, S \to \infty$. Clearly, the SMM-AIC procedure is not consistent since $\kappa_{T,S} = 2$ does not satisfy Assumption 5. The simultaneous selection procedure asymptotically selects the true model and the largest possible number of correct moment conditions. If Assumption 6 does not hold and Assumption 4 holds, two possible cases emerge: $\mathcal{MP}_2$ could be either non-empty or empty. In the first case, the set $\mathcal{MP}_2$ contains multiple elements. For example, if $\mathcal{MP}_1$ only contains two vectors, i.e., $\mathcal{MP}_1 = \{(\boldsymbol{b}, \boldsymbol{c}_1), (\boldsymbol{b}, \boldsymbol{c}_2)\}$ with $|\boldsymbol{c}_1| - |\boldsymbol{b}| = |\boldsymbol{c}_2| - |\boldsymbol{b}|$, then both two vectors are included in $\mathcal{MP}_2$. In this case, we have the following corollary:

Corollary 1 *Suppose Assumptions 1 to 5 hold and* $\mathcal{MP}_2 \neq \emptyset$*,* $(\hat{\boldsymbol{b}}_{MSC}, \hat{\boldsymbol{c}}_{MSC}) \in \mathcal{MP}_2$*,* $wp \to 1$*, as* $T, S \to \infty$.

Corollary 1 is self-evident from the proof of Proposition 1. In the second case, $\mathcal{MP}_2 = \emptyset$, then $\mathcal{MP}_1$ is empty and $\{(\boldsymbol{b}^0, \boldsymbol{c}^0)\} \notin \mathcal{P}$. This means that there exists no selection vector $(\boldsymbol{b}, \boldsymbol{c}) \in \mathcal{P}$ such that $\boldsymbol{g}_{T,S}^c(\boldsymbol{\theta}^b) \to_p \mathbf{0}$, as $T, S \to \infty$. We adapt the set $\mathcal{MP}_1$ to $\mathcal{MP}_1' = \{(\boldsymbol{b}, \boldsymbol{c}) \in \mathcal{P} : \inf_{\boldsymbol{\theta}^b \in \boldsymbol{\Theta}} \boldsymbol{g}_0^c(\boldsymbol{\theta}^b)' \boldsymbol{W}_0^c \boldsymbol{g}_0^c(\boldsymbol{\theta}^b) \leq \inf_{\boldsymbol{\theta}^{b_*} \in \boldsymbol{\Theta}} \boldsymbol{g}_0^{c_*}(\boldsymbol{\theta}^{b_*})' \boldsymbol{W}_0^{c_*} \boldsymbol{g}_0^{c_*}(\boldsymbol{\theta}^{b_*}), \forall (\boldsymbol{b}_*, \boldsymbol{c}_*) \in \mathcal{P}\}$, where $\min(T,S)^{-1} MSC_{T,S}(\boldsymbol{\theta}^b, \boldsymbol{c}) \to_p \inf_{\boldsymbol{\theta}^b \in \boldsymbol{\Theta}} \boldsymbol{g}_0^c(\boldsymbol{\theta}^b)' \boldsymbol{W}_0^c \boldsymbol{g}_0^c(\boldsymbol{\theta}^b)$, for $(\boldsymbol{b}, \boldsymbol{c}) \in \mathcal{P}$, as $T, S \to \infty$. If $(\boldsymbol{b}, \boldsymbol{c}) \in \mathcal{P}$ but $(\boldsymbol{b}, \boldsymbol{c}) \notin \mathcal{MP}_1$, $\inf_{\boldsymbol{\theta}^b \in \boldsymbol{\Theta}} \boldsymbol{g}_0^c(\boldsymbol{\theta}^b)' \boldsymbol{W}_0^c \boldsymbol{g}_0^c(\boldsymbol{\theta}^b) > 0$. In this case, we have the following corollary:

Corollary 2 *Suppose Assumptions 1 to 5 hold and* $\mathcal{MP}_2 = \emptyset$*,* $(\hat{\boldsymbol{b}}_{MSC}, \hat{\boldsymbol{c}}_{MSC}) \in \mathcal{MP}_1'$*,* $wp \to 1$*, as* $T, S \to \infty$.

Therefore, in the case that for none of the model candidates the moment conditions hold asymptotically, our procedure selects the model that asymptotically minimizes the (weighted) distance between true and the model implied moment conditions, i.e., $\boldsymbol{g}_0^c(\boldsymbol{\theta}^b)' \boldsymbol{W}_0^c \boldsymbol{g}_0^c(\boldsymbol{\theta}^b)$. In applied work, when it is likely that all candidate models are misspecified in some sense, one may therefore expect our criterion to select a model/moment combination that fits the data as well as possible with respect to the considered moments conditions, while maintaining parsimony in the parametrization of the model.

5 Simulation Study

We investigate the finite sample performance of the selection procedure in this section. Section 5.1 focuses on the in-sample performance. A comparison of different model and moment combinations is conducted and we try to identify which specific moment conditions are useful for the task of model selection. Section 5.2 is concerned with the out-of-sample forecasting performance of the Value-at-Risk of different combinations of models and moments.

5.1 Estimation and In-sample Fit

Consider the following factor copula model as the data generating process:

$$\begin{aligned} X_i &= \beta Z + \epsilon_i, i = 1, 2, \cdots, N, \\ Z &\sim \text{Skew } t(\nu^{-1}, \lambda), \\ \epsilon_i &\sim \text{iid } t(\nu^{-1}), \text{ and } \epsilon_i \perp Z, \forall i \\ (X_1, \cdots, X_N)' &\sim F_X = C(G_1, \cdots, G_N). \end{aligned} \tag{13}$$

All factor loadings are restricted to be identical, and the true coefficient is specified as $\beta_0 = 1$. This corresponds to an equi-dependence model and the variance of the common factor is half of the variance of each variable $X_i, i = 1, \ldots, N$. We consider time-varying conditional means and variances in the marginal distributions of the variables, i.e., we assume an AR(1)-GARCH(1,1) process,

$$\begin{aligned} Y_{i,t} &= \phi_0 + \phi_1 Y_{i,t-1} + \epsilon_{i,t}, \epsilon_{i,t} = \sigma_{i,t}\eta_{i,t}, \quad t = 1, \ldots, T, \quad i = 1, \ldots, N, \\ \sigma^2_{i,t} &= \omega + \gamma\sigma^2_{i,t-1} + \alpha\epsilon^2_{i,t-1}, \\ \boldsymbol{\eta}_t &= [\eta_{1,t}, \ldots, \eta_{N,t}] \sim_{iid} F_{\boldsymbol{\eta}} = \boldsymbol{C}(\Phi, \Phi, \ldots, \Phi), \end{aligned} \tag{14}$$

where $\boldsymbol{C}$ is implied by the equi-dependent factor copula model above and Φ is the cumulative distribution function of the standard normal distribution. In the estimation step, the AR-GARCH parameters are estimated by MLE in the first stage and the standardized residuals $\hat{\eta}_{i,t}$ are obtained after the conditional mean and variance are filtered out. These quantities are used to estimate copula parameters in the second stage. We use $S = 25 \times T$ simulations in the calculation of simulated dependence measures in the SMM estimation. The procedure is briefly summarized as:

1. Generate a $T \times N$ matrix of pseudo random numbers uniformly distributed in $[0, 1]$ from a given factor copula.

2. Transform each margin by the inverse AR(1)-GARCH(1,1) filter with Gaussian innovations and parameter vector

$$\boldsymbol{\phi} = (\phi_0, \phi_1, \omega, \alpha, \gamma)' = (0.01, 0.05, 0.05, 0.1, 0.85)'$$

   to obtain the simulated data.

3. Estimate an AR(1)-GARCH(1,1) model for each marginal series and calculate the corresponding standardized residuals $\hat{\boldsymbol{\eta}}_t$. Obtain the probability integral transformation of the residuals $\hat{\boldsymbol{u}}_t$ using the empirical distribution function (EDF).

4. Based on the transformed residuals $\hat{\boldsymbol{u}}_t$, each candidate pair of model and moments is estimated by SMM with the identity weighting matrix $\boldsymbol{W} = \boldsymbol{I}_r$. Then calculate the quantities of the simultaneous selection procedure (MSC) evaluated at the estimated parameter vector $\hat{\boldsymbol{\theta}}$.

5. Repeat the steps 1-4 $R = 200$ times and record the empirical selection frequencies of all combinations of models and moments based on the selection procedure.

All parameters of the factor copula model are treated as unknowns in the model estimation and we are considering the case of nested models. Note that the shape parameters ν in the latent factor and the idiosyncratic term are specified to be identical, and thus we have $\boldsymbol{\theta} = (\beta, \nu^{-1}, \lambda)'$. Four factor copula models are considered as true underlying DGPs in this section. When both the tail parameter[2] ν^{-1} and the skewness parameter λ are zero, corresponding to the selection vector $\boldsymbol{b}_1 = (1, 0, 0)'$, this corresponds to the *normal-normal* factor copula (abbreviated as *n-n*

[2]The shape parameter ν is reparametrized as ν^{-1} in order to be able to get more accurate estimation results.

thereafter). The *t-t* factor copula with $\boldsymbol{\theta} = (1, 0.25, 0)'$ is obtained by the selection vector $\boldsymbol{b}_2 = (1, 1, 0)'$. The case $\boldsymbol{b}_3 = (1, 1, 1)'$ implies the *skew t-t* factor copula with $\boldsymbol{\theta} = (1, 0.25, -0.5)'$. The last DGP we consider is the *skew normal-normal* factor copula (abbreviated as *skewn-n* thereafter) with $\boldsymbol{\theta} = (1, 0, -0.5)'$ specified by $\boldsymbol{b}_4 = (1, 0, 1)'$. Spearman's rho and quantile dependence λ_q at quantiles $q \in (0, 1)$ are adopted as the dependence measures. The sets of moment conditions under consideration are $m_1 = \{\lambda_{0.01}, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.15}\}$, $m_2 = \{\rho_S, \lambda_{0.01}, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.15}\}$, $m_3 = \{\rho_S, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.9}, \lambda_{0.95}\}$, $m_4 = \{\rho_S, \lambda_{0.25}, \lambda_{0.45}, \lambda_{0.55}, \lambda_{0.75}\}$, $m_5 = \{\rho_S, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.25}, \lambda_{0.45}, \lambda_{0.55}, \lambda_{0.75}, \lambda_{0.9}, \lambda_{0.95}\}$. The choices of the sample length are $T = 500, 1000, 2000$[3] and we consider the dimensions $N = 5, 10$. The results in all scenarios are based on $R = 200$ Monte Carlo replications.

Table 1 reveals the selection frequency for various model specifications, i.e., the fraction of times that a specific combination of the four candidate models and five sets of moments is chosen according to the SMM-BIC criterion[4]. Not surprisingly, the BIC selection procedure rewards the additional usage of moment conditions and it selects the models combined with largest moment set m_5. Panel A together with Panel B in Table 1 show the dimension effect on the selection frequencies whereas Table 1 together with Tables S.A.1 to S.A.4 in web appendix show the effect of the sample size. The selection procedure generally produces higher correct selection rates when the sample size or the dimension is increased. The probabilities of choosing the true model are closer to 1 when the sample size increases from $T = 500$ to $T = 2000$ or the dimension increases from $N = 5$ to $N = 10$. Given the property of the equi-dependence model, there is only one unknown factor loading to be estimated. Increasing the dimensionality does not increase the number of unknown parameters, but increases the effective sample size for an equally complex model. Therefore, the selection procedure becomes more powerful not only with an increasing sample size but also with an increasing dimensionality.

Table 1 about here

[3]Here we only present the results for the case $T = 1000$. The results for $T = 500, 2000$ can be found in Tables S.A.1 to S.A.4 in the supplemental appendix to this article.

[4]The results based on SMM-AIC and SMM-HQIC are available in Tables S.A.5 to S.A.12 in the supplemental appendix to this article.

Furthermore, Table 2 presents the empirical selection frequencies of the candidate models for a predetermined choice of moments for $N = 5$ and $N = 10$ dimensions. The choice of moment conditions affects the performance of the selection procedure in the model selection. We illustrate the importance of the choice of quantile dependence measures from the following three comparisons:

First, comparing the candidate sets of moments m_3 and m_4, the former solely utilizes Spearman's rho and tail quantile dependence measures, the latter merely consists of Spearman's rho and quantile dependence measures in the center of the distribution. One can observe in Table 2 that the procedure works worse with the quantile dependence measures in the middle area m_4. No matter which model is the true model, the selection procedure highly prefers *n-n* model. For example, when the *t-t* factor copula is the true model, the selection frequencies of the *n-n* factor copula are close to 1 regardless of the dimension and the sample size, whereas the selection rates of the *t-t* factor copula remain zero. This is not surprising due to the fact that only the quantiles in the central region are considered, which do not capture the dependence structure in the tails.

Second, considering the candidate sets of moments m_1 and m_3, the first set only includes quantile dependence measures at the left tail, whereas the second set considers quantile dependence measures at both tails. When m_1 is used in the estimation and the true DGP is an asymmetric model, i.e., the *skewn-n* or *skewt-t* factor copula, the selection procedure points to the *n-n* and *t-t* factor copulas, their symmetric counterparts, respectively. This is due to the fact that the set m_1 excludes the dependence measures at right tail and is not able to capture the asymmetric dependence structure in the data.

Third, compared with m_1, the set m_2 additionally includes Spearman's rho compared to the set m_1. Although both sets exclude quantile dependence measures at the right tail, the problem of merely capturing symmetric dependence structure is resolved with increasing sample size due to the inclusion of ρ_S .

Table 2 about here

To sum up, the simultaneous selection of models and moments consistently selects the true model combined with the moment set with the largest number of moment conditions. The selection frequencies of the correct model improve with higher dimensions or larger sample sizes except in the case when one only utilizes quantile dependence measures at the central region of the support in the selection of the models.

5.2 Value-at-Risk Prediction

Precise estimates and forecasts of the Value-at-Risk (VaR) are often of interest in risk management. We use the following simulations to compare the performance of the correctly specified model and misspecified models in terms of their VaR forecasting accuracy. The usage of moment based estimation with factor copulas makes it possible to solely concentrate on the representation of the dependence structure in a specific region of the support instead of the global area. Hence, it is possible to restrict the choice of dependence measures to certain quantiles when one deals with a problem such as estimating the VaR. Given that the underlying DGP is generally unknown, the simulation design is to additionally answer the question whether a wrong model could achieve an acceptable or even better performance in the forecasting task, especially when the misspecified model has fewer parameters.

In this section, we consider the same models as in previous section, but restrict the sets of moment conditions to $m_1 = \{\lambda_{0.01}, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.15}\}$, $m_2 = \{\rho_S, \lambda_{0.01}, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.15}\}$, and $m_5 = \{\rho_S, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.25}, \lambda_{0.45}, \lambda_{0.55}, \lambda_{0.75}, \lambda_{0.9}, \lambda_{0.95}\}$. In contrast to the previous section, the data generating process is the *skewt-t* factor copula with parameters $\beta = 1$, $\nu^{-1} = 0.25$ and $\lambda = -0.5$. The marginal dependence structure is the same as in Section 5.1. The equal weighted portfolio consists of N series of returns and we consider $N = 5$ and $N = 10$. The portfolio return at time t is determined by

$$Y_{p,t} = \log(1 + \frac{1}{N}\sum_{i=1}^{N}(\exp(Y_{i,t}) - 1)), \tag{15}$$

where $Y_{i,t}$ is the simulated return of the i-th stock. The simulation procedure is as follows.

1. Simulate a sample of $T = 1500$ returns $\boldsymbol{Y}$ from the true factor copula model with a Gaussian AR-GARCH model for the margins. We select the first $T_e = 1000$ observations as the estimation window whereas the remaining $T_s = 500$ observations are left for testing the VaR forecasts.

2. For each t in the test window we compute the true VaR obtain using $B_1 = 1,000,000$ random draws generated from the true DGP, i.e., the *skewt-t* factor copula model using the true parameters with Gaussian margins using the true conditional mean $\boldsymbol{\mu}_t$ and variance $\boldsymbol{\sigma}_t^2$.

3. At each time point t in the test window, we estimate the dynamic parameter vector $\boldsymbol{\phi}_{t-1}$ with a rolling window of length $l = 1000$ based on the information up to $t-1$. This is used to predict the conditional mean vector $\hat{\boldsymbol{\mu}}_t(\hat{\boldsymbol{\phi}}_{t-1})$ and conditional volatility vector $\hat{\boldsymbol{\sigma}}_t(\hat{\boldsymbol{\phi}}_{t-1})$ at time t.

4. Only using the data from the estimation window, estimate the parameters $\boldsymbol{\theta} = (\beta, \nu^{-1}, \lambda)'$ of the *n-n*, *skewn-n*, *t-t* and *skewt-t* factor copula models using the three candidate sets of moment conditions, respectively.[5]

5. The predicted VaRs are computed using $B_2 = 100,000$ simulations from the estimated factor copula models and the predicted conditional means and variances from Step 3.

6. Repeat steps 1-5 for $R = 100$ times, record the average coverage rate of the VaR violations, the average percentage bias, the average percentage RMSE and the average loss of VaR forecasts in test window. The Diebold & Mariano (1995) test is used to obtain a pairwise comparison of the VaR forecasting accuracy of the four candidate models along with the three sets of moment conditions.

The average coverage rate $\overline{\text{Cov}}$ is the ratio of number of VaR exceedances and the length of test window averaged over the R Monte Carlo replications. Since the 'true' VaR changes over time, we compute the average of %bias and %RMSE of the VaR forecasts. The average %bias of the VaR forecasts is calculated via

$$\overline{\%\text{bias}} = \frac{1}{T_s}\sum_{t=1}^{T_s} 100 \frac{1}{R}\sum_{r=1}^{R} \frac{\text{VaR}_{r,t} - \text{VaR}^0_{r,t}}{|\text{VaR}^0_{r,t}|},$$

whereas the average %RMSE of the VaR forecasts follows

$$\overline{\%\text{RMSE}} = \frac{1}{T_s}\sum_{t=1}^{T_s} 100 \sqrt{\frac{1}{R}\sum_{r=1}^{R}\Big(\frac{\text{VaR}_{r,t} - \text{VaR}^0_{r,t}}{\text{VaR}^0_{r,t}}\Big)^2}.$$

[5]Note that re-estimating the copula parameters over the rolling window was computationally infeasible.

The loss function used to make the pairwise comparisons of the VaR prediction using the Diebold & Mariano (1995) test is chosen as (see Gneiting 2011)

$$L(Y_{p,t}, \mathrm{VaR}_{\alpha,t}) = \alpha(Y_{p,t} - \mathrm{VaR}_{\alpha,t})(1 - I_{[Y_{p,t}<\mathrm{VaR}_{\alpha,t}]}) + (1-\alpha)(\mathrm{VaR}_{\alpha,t} - Y_{p,t})I_{[Y_{p,t}<\mathrm{VaR}_{\alpha,t}]}. \tag{16}$$

The average loss of the VaR forecasts over the test horizon and the R Monte Carlo replications is computed by

$$\overline{\mathrm{Loss}} = \frac{1}{T_s}\sum_{t=1}^{T_s}\frac{1}{R}\sum_{r=1}^{R} L(Y_{p,t}^r, \mathrm{VaR}_{\alpha,t}^r).$$

The upper panels in Tables 3 and 4 list the average coverage rate, average %bias and average %RMSE of the 5%-VaR and 1%-VaR forecasts for $N = 10$.[6] The misspecified models perform reasonably well in terms of the coverage rates as well as in terms of the %bias and %RMSE when a small set of moments is used. For example, looking at the 5%-VaR forecasts, the average coverage rate obtained from the *t-t* factor copula is close to the nominal level 5% when m_1 is used. The average %RMSE is only slightly larger than that from the true model. If m_1 is selected, the *t-t* model achieves a smaller average %bias than the *skewt-t* model. However, when the quantile dependence at the upper tail is included into the set of moments, the misspecified models perform much worse than the true model in terms of these three metrics. Overall, however, the most accurate average coverage rate, lowest average %bias and lowest average %RMSE are achieved by the *skewt-t* model. The results for the 1%-VaR forecasts in Table 4 show similar results.

Table 3 and 4 about here

In the lower panels of Tables 3 and 4, the Diebold-Mariano (DM) test provides additional information on the pairwise comparisons between the models in terms of the VaR forecasting error. Regarding the 5%-VaR forecasts, when quantiles only at the lower tail are taken into account, the *n-n* factor copula model surpasses any other possible combinations of models and sets of moments. When Spearman's rho is additionally included in the set of moments, the *n-n* factor copula continues to be the

[6]The results of $N = 5$ case are available in Tables S.A.13 and S.A.14 in the supplemental appendix to this article.

preferred model except when the candidate model is the *skewn-n* factor copula along with m_1 or m_2. This could be explained by the fact that *n-n* factor copula is the most parsimonious model among all candidate models. The *t-t* factor copula along with the set m_1 is preferred in comparison with the *skewt-t* factor copula estimated using m_1. When any other set of moments is utilized, the *t-t* factor copula is strongly dominated by the *skewt-t* factor model along with all possible choices of moments. When the *skewt-t*, *t-t* and *n-n* models are paired with the moment set m_5, which also includes Spearman's rho and the upper tail quantiles $\{\lambda_{0.9}, \lambda_{0.95}\}$, the *skewt-t* model eventually outperforms both the *t-t* and *n-n* models. The *skewn-n* model generally dominates the *t-t* and *skewt-t* factor copula models regardless of the choice of moment conditions.

When it comes to the evaluation of the 1%-VaR forecasts in Table 4, apart from the result that the fraction of indifferent decisions between two models based on DM test increases compared to the case of 5%-VaR forecasts, the combinations involving the *n-n* factor copula tend outscore other combinations of models and moments less frequently. For instance, the *t-t* factor copula model is able to produce better forecasts than the *n-n* model when both of them use lower tail quantiles, i.e., the moment sets m_1 and m_2. In addition, the *skewt-t* model along with any choice of moment sets outperforms the combination of the *n-n* model and the set m_1 in approximately 20% of the replications. For the full set of moment conditions m_5 the *skewt-t* model dominates the *n-n* model. The worse performance of *n-n* model compared to the 5%-VaR can be explained by its lack of tail dependence. When both the *skewt-t* and *t-t* models use the moment set m_1, the *t-t* model achieves a lower %RMSE than the true model and is selected more often by the Diebold-Mariano statistic. Although the *t-t* factor copula is not able to capture the asymmetric characteristics of the true DGP, it is less parametrized and its performance in VaR prediction is quite acceptable even though it is misspecified. Finally, the *skewt-t* factor copula outperforms the *skewn-n* factor copula regardless of the choice of moment conditions.

6 Empirical Illustration

In this section, the model and moment selection procedure is applied to determine an appropriate dependence model for daily log returns of ten stocks in the DAX 30 index: *SAP*, *Siemens*, *Deutsche Telekom*, *Bayer*, *Allianz*, *Daimler*, *BASF*, *BMW*, *Continental* and *Deutsche Post*. These companies possessed the ten highest market values in the index on May 31, 2016. The sample ranges from January 1, 2010 to May 31, 2016, and $T = 1629$ observations are obtained after the exclusion of non-trading and settlement days. The descriptive statistics can be found in Panel A of Table 5. The means of the log returns (in percentages) are close to zero, and all stock returns are left-skewed and leptokurtic.

Table 5 about here

To estimate the factor copula model, it is necessary to filter out the conditional mean and variance in each of the ten series in the first step. We consider the AR(1)-GJR-GARCH(1,1,1) model with skewed Student t distributed innovations, see Glosten et al. (1993).

$$\begin{aligned} r_{i,t} &= \phi_{0,i} + \phi_{1,i} r_{i,t-1} + \epsilon_{i,t}, \quad \epsilon_{i,t} = \sigma_{i,t}\eta_{i,t}, \\ \sigma^2_{i,t} &= \omega_i + \alpha_i \epsilon^2_{i,t-1} + \beta_i \epsilon^2_{i,t-1} \mathbf{1}_{\epsilon_{i,t-1} \leq 0} + \gamma_i \sigma^2_{i,t-1}, \end{aligned} \tag{17}$$

where $r_{i,t}$ denotes the log return of stock i for $i = 1, \ldots, 10$. Unlike in Oh & Patton (2013) and Oh & Patton (2017), the market returns are not considered in modeling time-varying conditional means and variances. The parameters determining the dynamics of the margin are estimated via MLE and the results are presented in Panel B of Table 5. The estimated residuals $\hat{\eta}_{it}$ are obtained using the estimated parameters and the transformation of the residuals $\hat{\eta}_{it}$ to uniform series is accomplished by the empirical distribution function (EDF).

Table 6 presents the sample dependence measures between the ten stocks. The upper panel shows the sample Spearman's rho of the uniform residuals, which ranges from 0.402 to 0.796. The lower panel presents quantile dependence measures at the tails. The elements in the upper triangular matrix are computed as the average of the 1% and 99% quantile dependence measures $\lambda_{0.01}$ and $\lambda_{0.99}$. For some cases, we observe rather weak tail dependence, for example, zero dependence between *Deutsche Telekom* and *Continental*. However, substantial tail dependence is observed between *Bayer* and *BASF*, or between *BMW* and *Daimler*. This could be explained by the fact that stocks belonging to the same industry sector behave similarly in crashes or booms. The difference between the 90% and 10% quantile dependence presented in the lower triangular matrix reveals some degree of asymmetry in the dependence. Since all elements are negative, this left skewness indicates higher dependence in crashes than during booms.

Table 6 about here

Next, the various factor copula models are estimated. The *n-n* factor copula corresponds to the Gaussian Copula and the *skewn-n* factor copula allows for the asymmetric dependence but imposes zero tail dependence. The *t-n* and *t-t* factor copulas imply symmetric tail dependence whereas the *skewt-n* and *skewt-t* factor copulas allow tail dependence in both tails and an asymmetric dependence structure. Figure 2 shows the sample and fitted quantile dependence measures by three of those factor copula models estimated using the sets of moment conditions m_1, m_2, m_3, m_4 and m_5 (as defined in Section 5.1), respectively. In general, all presented model/moment combinations fit the lower tail of the distribution fairly well, with the exception of the *skewt-t* in combination with m_4. However, the fit in the upper tail is considerably worse when the models are estimated with m_1, m_2 and m_4, which do not contain quantile dependence in this region of the support. Overall, it appears that the less parametrized models *n-n* and *skewn-n* fit the data fairly well when used together with the appropriate moments.

Figure 2 about here

The first three columns in Table 7 contain the estimated parameters for the 30 model/ moment combinations. Note that the identity matrix $\boldsymbol{I}$ is fixed as the weighting matrix in the SMM estimation. The p-values of the over-identifying test proposed by Oh & Patton (2013) are additionally provided. The limit distribution of this test is approximated via $B = 1000$ bootstrap replications. The step size $\epsilon_{T,S}$ utilized in the computation of the numerical derivative of $\boldsymbol{g}^c_{T,S}(\boldsymbol{\theta}^b)$ at $\boldsymbol{\theta}^b = \hat{\boldsymbol{\theta}}_{T,S}$ is 0.005. The results of the J tests show that the *skewt-t* factor copula is clearly rejected for the sets of lower-tail-moments m_1 and m_2, but not rejected for the sets m_3, m_4 and m_5 at the 5% significance level, whereas the *n-n* factor copula is not rejected when m_1 and m_2 are used, but it is rejected when m_3 and m_5 are used. Only the *skewt-n* and *skewn-n* models are not rejected at the 5% significance level for any set of moments. This information is not very useful in comparing and ranking various model candidates and the choice of moment conditions. To this end, our selection procedure is applied.

Table 7 about here

The selection decisions are presented in the last three columns in Table 7. The SMM-AIC, SMM-BIC and SMM-HQIC criteria agree on the *skewn-n* factor copula model along with the moment set with the largest number of overidentifying conditions, i.e., m_5, as the best choice out of all combinations of models and moments. It is followed by the *skewt-n* factor copula and the *skewt-t* factor copula with the same moment set. It highlights the importance of the skewness parameter in these factor copula models. The best model implies an asymmetric dependence but zero tail dependence in the empirical data. The estimators of the tail parameter $\hat{\nu}^{-1}$ from the second and the third best model are close to zero, which further validates that the empirical data possesses little tail dependence. Models allowing for tail dependence therefore perform worse. When the set of moments is predetermined as m_2, m_3 or m_5, the *skewn-n* factor copula dominates the other models. Restricted to m_1 and m_4, the *n-n* model is selected as the best model.

Lastly, one-day ahead VaR forecasting and backtesting is performed on the portfolio of these ten stocks. Each stock is assigned an equal weight in this portfolio. For the sake of simplicity, we fix the set m_5 as the selected set of moments. In order to evaluate the accuracy of the VaR forecasts at the 5% and 1% level, we backtest the six factor copula models as follows.

1. The whole sample is initially split into the estimation period with sample length $T_e = 1000$ and the testing period with sample length $T_s = 629$.

2. For each time t in the testing period, the past T_e data points, i.e., the sample from $t-T_e+1$ to $t-1$, are utilized in the estimation of the copula parameters $\boldsymbol{\theta}_{t-1}$

for all six models and the parameters $\boldsymbol{\phi}_{t-1}$ of the margins. The one-step-ahead forecasts of the conditional means and variances are obtained by the plug-in method based on the information up to $t-1$.

3. To approximate the distribution of portfolio returns at time t, we simulate $B = 100,000$ standardized residuals $\hat{\boldsymbol{\eta}}_t$ which are determined by $\hat{\boldsymbol{\theta}}_{t-1}$ in each factor copula model along with $\hat{\boldsymbol{\phi}}_{t-1}$ in the AR(1)-GJR-GARCH(1,1,1) models for the margins. For each copula model, we use $\boldsymbol{r}_t^s = \hat{\boldsymbol{\mu}}_t(\hat{\boldsymbol{\phi}}_{t-1}) + \hat{\boldsymbol{\sigma}}_t(\hat{\boldsymbol{\phi}}_{t-1})\hat{\boldsymbol{\eta}}_t$ to generate the simulated distribution of the return at time t. An equally weighted portfolio is constructed from these simulated returns.

4. The 5%- and 1%-VaR forecasts at time t are computed based on the simulated distribution of portfolio returns. The number of VaR violations and the corresponding coverage rate are obtained by comparing the VaR forecasts and the actual returns.

To evaluate the forecasting performance of the competing models, three tests are performed to test the coverage level and the distribution of the violations: the unconditional coverage test by Kupiec (1995), the conditional coverage test by Christoffersen (1998) and the dynamic quantile test by Engle & Manganelli (2004). The results are available in Table 8. For the 5%-VaR forecasts all average coverage rates are generally slightly larger than the nominal VaR level. None of the factor copula models is rejected by the unconditional coverage test of Kupiec at the 5% significance level. The conclusions from the conditional coverage test stay in line with the unconditional coverage test. The dynamic quantile test provides further information on the distribution of the VaR violations. It shows that the violations are independent to each other for the VaR forecasts from all factor copula models. When it comes to 1%-VaR forecasts, none of models is rejected by any of the three tests. Given the fact that the testing window in our analysis is rather short, the occurrence of VaR violations at 1% level can be considered as relatively rare events. One could prolong the testing window to check if the difference of the performance of the VaR forecasts turns out to be significant.

Table 8 about here

The Diebold-Mariano test is used to compare the performance of the VaR forecasts between pairs of models. Considering the 5%-VaR forecasts, Table 9 shows that the skewed factor copula models are preferred and smaller models mostly outperform candidate models with more parameters. For instance, the *skewn-n* factor copula stands out in the skewed family, since it solely focuses on modeling the asymmetric dependence structure and has few parameters. This also coincides with the results from our selection procedure. When it comes to the 1%-VaR forecasts, the family of skewed models remains the most favorable choice. In addition, the Diebold-Mariano test tends to be indifferent in more occasions compared to the 5%-VaR analysis. For example, the forecasting performance of *skewt-t* and *skewt-n* are as good as for the *skewn-n* model for the 1%-VaR.

Table 9 about here

7 Conclusions

Since most factor copula models do not have a closed-form likelihood, this causes additional difficulties in the selection task of models and moments. In this paper, by integrating the specification test for factor copula models proposed by Oh & Patton (2013) into the model and moment selection framework in Andrews & Lu (2001), we obtain a simultaneous model and moment selection procedure for factor copula models. The consistency of the selection procedure is mathematically shown and its finite sample performance is validated by Monte Carlo simulations under various scenarios. We find that the choice of quantile dependence measures plays a crucial role in the selection of models. If one is merely concerned with VaR forecasts, a smaller model with quantile dependence measures at the lower tail may achieve a good forecasting performance. We also perform the selection procedure for the dependence modeling based on a real data set. The empirical findings corroborate the performance of the selection procedure in the sense that it chooses the candidate models with the largest number of over-identifying restrictions. Here, we only consider a rather simple model, i.e., the equi-dependence factor copula model. In the future research, it would be valuable to examine the performance of the proposed selection methods on other models such as blockwise dependence models. A further application of the method would be the choice of the number of factors.

References

Andersen, L. B. G. & Sidenius, J. (2004), 'Extensions to the Gaussian copula: Random recovery and random factor loadings', *Journal of Credit Risk* **1**(1), 29–70.

Andrews, D. W. K. (1999), 'Consistent moment selection procedures for generalized method of moments estimation', *Econometrica* **67**(3), 543–564.

Andrews, D. W. K. & Lu, B. (2001), 'Consistent model and moment selection procedures for GMM estimation with application to dynamic panel data models', *Journal of Econometrics* **101**(1), 123–164.

Ansari, A., Jedidi, K. & Dube, L. (2002), 'Heterogeneous factor analysis models: A Bayesian approach', *Psychometrika* **67**(1), 49–77.

Chen, X. & Fan, Y. (2006), 'Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification', *Journal of Econometrics* **135**(1-2), 125–154.

Christoffersen, P. F. (1998), 'Evaluating interval forecasts', *International Economic Review* **39**(4), 841–62.

Dias, A. & Embrechts, P. (2004), Dynamic copula models for multivariate high-frequency data in finance, Working paper, Warwick Business School, Finance Group.

Diebold, F. & Mariano, R. (1995), 'Comparing predictive accuracy', *Journal of Business & Economic Statistics* **13**(3), 253–63.

Engle, R. F. & Manganelli, S. (2004), 'CAViaR: Conditional autoregressive value at risk by regression quantiles', *Journal of Business & Economic Statistics* **22**(4), 367–381.

Fredricks, G. & Nelsen, R. (2007), 'On the relationship between Spearman's rho and Kendall's tau for pairs of continuous random variables', *Journal of Statistical Planning and Inference* **137**, 2143–50.

Genest, C. & Favre, A. C. (2007), 'Everything you always wanted to know about copula modeling but were afraid to ask', *Journal of Hydrologic Engineering* **12**(4), 347–368.

Glosten, L. R., Jagannathan, R. & Runkle, D. E. (1993), 'On the relation between the expected value and the volatility of the nominal excess return on stocks', *Journal of Finance* **48**(5), 1779–1801.

Gneiting, T. (2011), 'Quantiles as optimal point forecasts', *International Journal of Forecasting* **27**, 197–207.

Hall, A. R. & Inoue, A. (2003), 'The large sample behaviour of the generalized method of moments estimator in misspecified models', *Journal of Econometrics* **114**(2), 361–394.

Hannan, E. J. & Quinn, B. G. (1979), 'The determination of the order of an autoregression', *Journal of the Royal Statistical Society Series B* **41**(2), 190–195.

Hansen, B. (1994), 'Autoregressive conditional density estimation', *International Economic Review* **35**(3), 705–30.

Hastie, T., Tibshirani, R. & Friedman, J. (2009), *The Elements of Statistical Learning: Data mining, Inference and Prediction*, 2 edn, Springer.

Hull, J. & White, A. (2004), 'Valuation of a CDO and an n-th to default CDS without Monte Carlo simulation', *Journal of Derivatives* **12**(2), 8–23.

Ivanov, E., Min, A. & Ramsauer, F. (2017), 'Copula-based factor models for multivariate asset returns', *Econometrics* **5**(20), 1–24.

Joe, H. (2014), *Dependence Modeling with Copulas*, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Taylor & Francis.

Krupskii, P., Huser, R. & Genton, M. G. (2018), 'Factor copula models for replicated spatial data', *Journal of the American Statistical Association* **113**(521), 467–479.

Krupskii, P. & Joe, H. (2013), 'Factor copula models for multivariate data', *Journal of Multivariate Analysis* **120**, 85–101.

Kupiec, P. H. (1995), 'Techniques for verifying the accuracy of risk measurement models', *The Journal of Derivatives* **3**(2), 73–84.

Manner, H. (2007), Estimation and model selection of copulas with an application to exchange rates, Research Memorandum 056, Maastricht University, Maastricht Research School of Economics of Technology and Organization (METEOR).

McFadden, D. (1989), 'A method of simulated moments for estimation of discrete response models without numerical integration', *Econometrica* **57**(5), 995–1026.

Nelsen, R. B. (2006), *An Introduction to Copulas*, 2 edn, Springer, New York.

Newey, W. K. & McFadden, D. (1994), Large sample estimation and hypothesis testing, Vol. 4 of *Handbook of Econometrics*, Amsterdam: North-Holland, pp. 2111–2245.

Nikoloulopoulos, A. & Joe, H. (2015), 'Factor copula models for item response data', *Psychometrika* **80**(1), 126–150.

Oh, D. H. & Patton, A. J. (2013), 'Simulated method of moments estimation for copula-based multivariate models', *Journal of the American Statistical Association* **108**(502), 689–700.

Oh, D. H. & Patton, A. J. (2017), 'Modeling dependence in high dimensions with factor copulas', *Journal of Business & Economic Statistics* **35**(1), 139–154.

Patton, A. J. (2012), 'A review of copula models for economic time series', *Journal of Multivariate Analysis* **110**, 4 – 18.

Rémillard, B. (2017), 'Goodness-of-fit tests for copulas of multivariate time series', *Econometrics* **5**(13), 1–23.

van der Voort, M. (2005), Factor copulas: Totally external defaults, Working paper, Erasmus University Rotterdam.

A Proofs

A.1 Proof of Proposition 1

Proof According to Oh & Patton (2013), Assumption 1 and Assumption 2(iii) imply the stochastic Lipschitz continuity of $\boldsymbol{g}_{T,S}(\boldsymbol{\theta})$. Lemma 2.9 in Newey & McFadden (1994) shows that the stochastic Lipschitz continuity of $\boldsymbol{g}_{T,S}(\boldsymbol{\theta})$ guarantees the stochastic equicontinuity of $\boldsymbol{g}_{T,S}(\boldsymbol{\theta})$. By Lemma 2 in Oh & Patton (2013), when Assumption 2 is fulfilled, then $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} ||\boldsymbol{g}_{T,S}(\boldsymbol{\theta}) - \boldsymbol{g}_0(\boldsymbol{\theta})|| \to_p 0$ as $T, S \to \infty$ if and only if $\boldsymbol{g}_{T,S}(\boldsymbol{\theta}) \to_p \boldsymbol{g}_0(\boldsymbol{\theta}), \forall \boldsymbol{\theta} \in \boldsymbol{\Theta}$, and $\boldsymbol{g}_{T,S}(\boldsymbol{\theta})$ is stochastically equicontinuous. Given Assumption 2(iv), Lemma 3 in Oh & Patton (2013) shows that uniform convergence of $\boldsymbol{g}_{T,S}(\boldsymbol{\theta})$ implies uniform covergence of $Q_{T,S}(\boldsymbol{\theta})$, i.e., if $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} ||\boldsymbol{g}_{T,S}(\boldsymbol{\theta}) - \boldsymbol{g}_0(\boldsymbol{\theta})|| \to_p 0$, as $T, S \to \infty$, then $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |Q_{T,S}(\boldsymbol{\theta}) - Q_0(\boldsymbol{\theta})| \to_p 0$, as $T, S \to \infty$. Let $\boldsymbol{g}^c_{T,S}(\boldsymbol{\theta}^b)$ be the moment conditions selected by the moment selection vector $\boldsymbol{c}$ evaluated at $\boldsymbol{\theta}^b$, which is the SMM estimator obtained from the model selected by model selection vector $\boldsymbol{b}$. Then, for $(\boldsymbol{b}, \boldsymbol{c}) \in \mathcal{P}$, we have

$$\min(T,S)^{-1} J^c_{T,S}(\boldsymbol{\theta}^b) := \boldsymbol{g}^c_{T,S}(\boldsymbol{\theta}^b)' \hat{\boldsymbol{W}}^c_T \boldsymbol{g}^c_{T,S}(\boldsymbol{\theta}^b) \to_p \boldsymbol{g}^c_0(\boldsymbol{\theta}^b)' \boldsymbol{W}^c_0 \boldsymbol{g}^c_0(\boldsymbol{\theta}^b), \text{ as } T, S \to \infty. \tag{A.1}$$

For any $(\boldsymbol{b}, \boldsymbol{c}) \in \mathcal{MP}_1$, there exists $\boldsymbol{\theta}^b \in \boldsymbol{\Theta}$ such that $\boldsymbol{g}^c_0(\boldsymbol{\theta}^b) = \boldsymbol{0}$. Therefore, we have

$$\min(T,S)^{-1} J^c_{T,S}(\boldsymbol{\theta}^b) \to 0, \text{ as } T, S \to \infty.$$

Together with Assumption 5, $\kappa_{T,S} = o(\min(T,S))$, the above equation yields, for $(\boldsymbol{b}, \boldsymbol{c}) \in \mathcal{MP}_1$,

$$\min(T,S)^{-1} MSC_{T,S}(\boldsymbol{\theta}^b, \boldsymbol{c}) = \min(T,S)^{-1} J^c_{T,S}(\boldsymbol{\theta}^b) - \min(T,S)^{-1}(|\boldsymbol{c}| - |\boldsymbol{b}|)\kappa_{T,S} \to_p 0.$$

For $(\boldsymbol{b}, \boldsymbol{c}) \in \mathcal{P}$, but $(\boldsymbol{b}, \boldsymbol{c}) \notin \mathcal{MP}_1$, it holds that $\forall \boldsymbol{\theta}^b \in \boldsymbol{\Theta}$, $\boldsymbol{g}^c_0(\boldsymbol{\theta}^b) \neq \boldsymbol{0}$. In addition, $\boldsymbol{W}^c_0$ is positive definite by Assumption 2(iv). Then Equation (A.1) and Assumption 5(i) give:

$$\min(T,S)^{-1} J^c_{T,S}(\boldsymbol{\theta}^b, \boldsymbol{c}) \to_p \boldsymbol{g}^c_0(\boldsymbol{\theta}^b)' \boldsymbol{W}^c_0 \boldsymbol{g}^c_0(\boldsymbol{\theta}^b) > 0$$

and therefore, for $(\boldsymbol{b}, \boldsymbol{c}) \in \mathcal{P}$, but $(\boldsymbol{b}, \boldsymbol{c}) \notin \mathcal{MP}_1$,

$$\min(T,S)^{-1} J^c_{T,S}(\boldsymbol{\theta}^b, \boldsymbol{c}) - \min(T,S)^{-1}(|\boldsymbol{c}| - |\boldsymbol{b}|)\kappa_{T,S} \to_p \boldsymbol{g}^c_0(\boldsymbol{\theta}^b)' \boldsymbol{W}^c_0 \boldsymbol{g}^c_0(\boldsymbol{\theta}^b) > 0 \tag{A.2}$$

as $T, S \to \infty$. The combination of Equations (A.1) and (A.2) implies $(\hat{\boldsymbol{b}}_{MSC}, \hat{\boldsymbol{c}}_{MSC}) \in \mathcal{MP}_1, wp \to 1$ given the definition of $(\hat{\boldsymbol{b}}_{MSC}, \hat{\boldsymbol{c}}_{MSC}) := \arg\min_{(\boldsymbol{b},\boldsymbol{c})\in\mathcal{P}} \mathrm{MSC}_{T,S}(\boldsymbol{\theta}^{\boldsymbol{b}}, \boldsymbol{c})$. Suppose that $(\boldsymbol{b}_i, \boldsymbol{c}_i), (\boldsymbol{b}_j, \boldsymbol{c}_j) \in \mathcal{MP}_1$, $i \neq j$, but $(\boldsymbol{b}_i, \boldsymbol{c}_i) \in \mathcal{MP}_2$ and $(\boldsymbol{b}_j, \boldsymbol{c}_j) \notin \mathcal{MP}_2$, then we have $(|\boldsymbol{c}_i| - |\boldsymbol{b}_i|) - (|\boldsymbol{c}_j| - |\boldsymbol{b}_j|) =: c > 0$ by the definition of the set $\mathcal{MP}_2$.

Now, define $MSC_{T,S}(\boldsymbol{\theta}^{b_i}, \boldsymbol{c}_i) = J^{c_i}_{T,S}(\boldsymbol{\theta}^{b_i}) - (|\boldsymbol{c}_i| - |\boldsymbol{b}_i|)\kappa_{T,S}$ and $MSC_{T,S}(\boldsymbol{\theta}^{b_j}, \boldsymbol{c}_j) = J^{c_j}_{T,S}(\boldsymbol{\theta}^{b_j}) - (|\boldsymbol{c}_j| - |\boldsymbol{b}_j|)\kappa_{T,S}$. Then, with Assumption 5, we have

$$\begin{aligned} & \frac{1}{\kappa_{T,S}} \left(MSC_{T,S}(\boldsymbol{\theta}^{b_i}, \boldsymbol{c}_i) - MSC_{T,S}(\boldsymbol{\theta}^{b_j}, \boldsymbol{c}_j) \right) \\ = & \frac{1}{\kappa_{T,S}} \left(J^{c_i}_{T,S}(\boldsymbol{\theta}^{b_i}) - J^{c_j}_{T,S}(\boldsymbol{\theta}^{b_j}) \right) - c \to_p -c \leq 0 \end{aligned}$$

as $T, S \to \infty$. Therefore $(\hat{\boldsymbol{b}}_{MSC}, \hat{\boldsymbol{c}}_{MSC}) \in \mathcal{MP}_2, wp \to 1$. Assumption 6 ensures that $\mathcal{MP}_2 = \{(\boldsymbol{b}^0, \boldsymbol{c}^0)\}$, then this directly leads to $(\hat{\boldsymbol{b}}_{MSC}, \hat{\boldsymbol{c}}_{MSC}) = (\boldsymbol{b}^0, \boldsymbol{c}^0), wp \to 1$. ■

A.2 Proof of Corollary 2

Proof If $\mathcal{MP}_1 \neq \emptyset$, then by definition of $\mathcal{MP}_2$, $\mathcal{MP}_2$ must contain at least one element, i.e., $\mathcal{MP}_2 \neq \emptyset$. Therefore, if $\mathcal{MP}_2 = \emptyset$, then $\mathcal{MP}_1 = \emptyset$. For $(\hat{\boldsymbol{b}}_{MSC}, \hat{\boldsymbol{c}}_{MSC}) \in \mathcal{P}$,

$$\begin{aligned} & \min(T,S)^{-1} MSC_{T,S}(\hat{\boldsymbol{b}}_{MSC}, \hat{\boldsymbol{c}}_{MSC}) \\ = & \min(T,S)^{-1} J^{\hat{\boldsymbol{c}}_{MSC}}_{T,S}(\hat{\boldsymbol{b}}_{MSC}, \hat{\boldsymbol{c}}_{MSC}) - \min(T,S)^{-1}(|\hat{\boldsymbol{c}}_{MSC}| - |\hat{\boldsymbol{b}}_{MSC}|)\kappa_{T,S}. \end{aligned}$$

According to equation (A.2) in the proof of Proposition 1 in Section A.1, for $(\hat{\boldsymbol{b}}_{MSC}, \hat{\boldsymbol{c}}_{MSC}) \in \mathcal{P}$, $(\hat{\boldsymbol{b}}_{MSC}, \hat{\boldsymbol{c}}_{MSC}) \notin \mathcal{MP}_1$, we have

$$\min(T,S)^{-1} MSC_{T,S}(\hat{\boldsymbol{b}}_{MSC}, \hat{\boldsymbol{c}}_{MSC}) \to_p \boldsymbol{g}_0^{\hat{\boldsymbol{c}}_{MSC}}(\boldsymbol{\theta}^{\hat{\boldsymbol{b}}_{MSC}})' \boldsymbol{W}_0^{\hat{\boldsymbol{c}}_{MSC}} \boldsymbol{g}_0^{\hat{\boldsymbol{c}}_{MSC}}(\boldsymbol{\theta}^{\hat{\boldsymbol{b}}_{MSC}}).$$

By definition of $(\hat{\boldsymbol{b}}_{MSC}, \hat{\boldsymbol{c}}_{MSC}) := \arg\min_{(\boldsymbol{b},\boldsymbol{c})\in\mathcal{P}} \mathrm{MSC}_{T,S}(\boldsymbol{\theta}^{\boldsymbol{b}}, \boldsymbol{c})$, for $\boldsymbol{\theta}^{\boldsymbol{b}} \in \boldsymbol{\Theta}$. Hence, $(\hat{\boldsymbol{b}}_{MSC}, \hat{\boldsymbol{c}}_{MSC}) \in \mathcal{MP}'_1$. ■

B Tables

Table 1: Simulation results: empirical selection frequencies among all model/moment combinations based on the BIC for $T = 1000$ and $N = 5, 10$ dimensions

	$\hat{C}$ \ C_0	Panel A: $N = 5$ n-n	t-t	skewn-n	skewt-t	Panel B: $N = 10$ n-n	t-t	skewn-n	skewt-t
m_1	n-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
m_2	n-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
m_3	n-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
m_4	n-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
m_5	n-n	**0.980**	0.435	0.060	0.000	**0.995**	**0.480**	0.020	0.000
	t-t	0.000	**0.480**	0.000	0.000	0.000	0.460	0.000	0.000
	skewn-n	0.020	0.035	**0.940**	0.270	0.005	0.015	**0.980**	0.205
	skewt-t	0.000	0.050	0.000	**0.730**	0.000	0.045	0.000	**0.795**

Note: The vector of true model parameters is $(\beta, \nu^{-1}, \lambda)' = (1, 0.25, -0.5)'$. The element-wise product of $\boldsymbol{\theta}$ and the selection vector $\boldsymbol{b}$ gives one specific candidate model. The *n-n*, *skewn-n*, *t-t* and *skewt-t* factor copula models are the candidate models used in SMM estimation. The candidate sets of moments are $m_i, i = 1, \ldots, 5$. The marginal distributions of the simulated data follow an AR(1)-GARCH(1,1) process, see equation (14). The sample length is $T = 1000$. The number of Monte Carlo replications is $R = 200$. $\hat{C}$ (rowwise) denotes the model/moment combination selected by the SMM-BIC selection procedure, C_0 (columnwise) denotes the DGP. The highest selection frequencies are marked in bold.

Table 2: Simulation results: empirical selection frequencies of models given predetermined sets of moments based on the BIC for $T = 1000$ and $N = 5, 10$ dimensions

		Panel A: $N = 5$				Panel B: $N = 10$			
	$\hat{C}$ \ C_0	n-n	t-t	skewn-n	skewt-t	n-n	t-t	skewn-n	skewt-t
m_1	n-n	**0.985**	0.470	**0.945**	0.195	**0.985**	0.450	**0.985**	0.205
	t-t	0.015	**0.525**	0.045	**0.795**	0.015	**0.550**	0.015	**0.790**
	skewn-n	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.000
	skewt-t	0.000	0.005	0.000	0.010	0.000	0.000	0.000	0.005
m_2	n-n	**0.965**	0.250	**0.705**	0.000	**0.985**	0.275	**0.795**	0.000
	t-t	0.035	**0.670**	0.160	**0.655**	0.015	**0.670**	0.090	**0.635**
	skewn-n	0.000	0.080	0.135	0.075	0.000	0.055	0.115	0.065
	skewt-t	0.000	0.000	0.000	0.270	0.000	0.000	0.000	0.300
m_3	n-n	**0.980**	**0.535**	0.090	0.000	**0.995**	**0.610**	0.040	0.000
	t-t	0.000	0.400	0.000	0.000	0.000	0.325	0.000	0.000
	skewn-n	0.020	0.035	**0.910**	0.305	0.005	0.030	**0.960**	0.295
	skewt-t	0.000	0.030	0.000	**0.695**	0.000	0.035	0.000	**0.705**
m_4	n-n	**1.000**	**1.000**	**1.000**	**0.925**	**1.000**	**1.000**	**1.000**	**0.870**
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	skewn-n	0.000	0.000	0.000	0.070	0.000	0.000	0.000	0.130
	skewt-t	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000
m_5	n-n	**0.980**	0.435	0.060	0.000	**0.995**	**0.480**	0.020	0.000
	t-t	0.000	**0.480**	0.000	0.000	0.000	0.460	0.000	0.000
	skewn-n	0.020	0.035	**0.940**	0.270	0.005	0.015	**0.980**	0.205
	skewt-t	0.000	0.050	0.000	**0.730**	0.000	0.045	0.000	**0.795**

Note: The vector of true model parameters is $(\beta, \nu^{-1}, \lambda)' = (1, 0.25, -0.5)'$. The element-wise product of $\boldsymbol{\theta}$ and the selection vector $\boldsymbol{b}$ gives one specific candidate model. The *n-n*, *skewn-n*, *t-t* and *skewt-t* factor copula models are the candidate models used in SMM estimation. The candidate sets of moments are $m_i, i = 1, \ldots, 5$. The marginal distributions of the simulated data follow an AR(1)-GARCH(1,1) process, see equation (14). The sample length is $T = 1000$. The number of Monte Carlo replications is $R = 200$. $\hat{C}$ (rowwise) denotes the model selected by the SMM-BIC procedure given $m_i, i = 1, \ldots, 5$, respectively. C_0 (columnwise) denotes the DGP. The highest selection frequencies are marked in bold.

Table 3: Simulation results: evaluation of 5%-VaR forecasts and backtesting for $N = 10$ dimensions

		n-n			skewn-n			t-t			skewt-t		
		m_1	m_2	m_5	m_1	m_2	m_5	m_1	m_2	m_5	m_1	m_2	m_5
		Panel A: Metrics											
	$\overline{\text{Cov}}$	0.039	0.046	0.068	0.041	0.045	0.046	0.050	0.058	0.072	0.050	0.054	0.051
	$\overline{\text{\%Bias}}$	-11.629	-3.762	14.592	-9.588	-5.203	-3.155	0.018	6.811	17.610	0.470	3.534	0.766
	$\overline{\text{\%RMSE}}$	12.868	6.675	14.934	11.651	6.963	5.100	5.074	8.350	17.900	5.040	6.668	4.096
	$\overline{\text{Loss}}$	0.092	0.091	0.092	0.092	0.091	0.091	0.091	0.091	0.093	0.091	0.091	0.091
		Panel B: Diebold-Mariano test											
n-n	m_1	-	88|0	100|0	56|23	83|2	88|2	91|3	99|0	100|0	92|3	97|0	96|0
	m_2	-	-	100|0	8|79	25|71	52|42	72|24	99|0	100|0	75|21	88|10	85|12
	m_5	-	-	-	0|100	0|100	0|100	0|100	0|100	96|2	0|100	0|100	0|100
skewn-n	m_1	-	-	-	-	66|19	77|14	85|10	99|0	100|0	89|5	95|2	96|0
	m_2	-	-	-	-	-	72|24	80|14	99|0	100|0	85|11	97|1	97|0
	m_5	-	-	-	-	-	-	81|12	99|0	100|0	87|8	96|3	97|0
t-t	m_1	-	-	-	-	-	-	-	100|0	100|0	66|30	70|24	53|43
	m_2	-	-	-	-	-	-	-	-	100|0	1|98	10|89	2|98
	m_5	-	-	-	-	-	-	-	-	-	0|100	0|100	0|100
skewt-t	m_1	-	-	-	-	-	-	-	-	-	-	67|31	46|50
	m_2	-	-	-	-	-	-	-	-	-	-	-	27|70
	m_5	-	-	-	-	-	-	-	-	-	-	-	-

Note: The true DGP is the skewt-t model with parameter vector $(\beta, \nu^{-1}, \lambda)' = (1, 0.25, -0.5)'$. The *n-n*, *skewn-n*, *t-t* and *skewt-t* factor copula models are the candidate models. The candidate sets of moments are m_1, m_2 and m_5. The marginal distributions of the simulated data follow an AR(1)-GARCH(1,1) process, see equation (14). $\overline{\text{Cov}}$ denotes the average coverage rate of the VaR violations over $R = 100$ Monte Carlo replications. $\overline{\text{\%Bias}}$ and $\overline{\text{\%RMSE}}$ are the average percentage bias and the average percentage RMSE of the VaR forecasts over the $T_s = 500$ forecasting horizon. $\overline{\text{Loss}}$ is the average loss of the VaR forecasts using (16) over T_s and R Monte Carlo replications. The left (right) number in each cell in the lower panel reports the number of times that the row model significantly outperforms (underperforms) the column model, using a Diebold-Mariano test at the 5% significance level.

Table 4: Simulation results: evaluation of 1%-VaR forecasts and backtesting for $N = 10$ dimensions

		n-n			skewn-n			t-t			skewt-t		
		m_1	m_2	m_5	m_1	m_2	m_5	m_1	m_2	m_5	m_1	m_2	m_5
		Panel A: Metrics											
	$\overline{\text{Cov}}$	0.012	0.016	0.027	0.011	0.014	0.018	0.011	0.011	0.020	0.011	0.011	0.010
	$\overline{\text{\%Bias}}$	4.204	10.079	23.920	3.012	7.989	13.761	2.263	2.790	16.157	2.209	1.714	1.196
	$\overline{\text{\%RMSE}}$	5.913	10.911	24.044	5.704	9.001	13.983	5.035	5.136	16.684	5.259	4.851	4.380
	$\overline{\text{Loss}}$	0.025	0.025	0.028	0.025	0.025	0.025	0.025	0.025	0.026	0.025	0.025	0.025
		Panel B: Diebold-Mariano test											
n-n	m_1	-	51|0	86|0	9|23	45|0	57|0	4|25	6|23	60|0	5|23	2|29	5|22
	m_2	-	-	90|0	0|50	6|46	55|11	0|47	0|48	64|1	0|47	0|46	0|44
	m_5	-	-	-	0|82	0|89	0|93	0|83	0|82	1|94	0|81	0|80	0|82
skewn-n	m_1	-	-	-	-	45|1	55|0	6|21	11|14	58|0	13|16	4|22	7|16
	m_2	-	-	-	-	-	61|1	1|43	1|43	69|0	1|42	1|41	3|35
	m_5	-	-	-	-	-	-	0|53	0|54	57|15	0|53	0|53	0|52
t-t	m_1	-	-	-	-	-	-	-	13|6	56|0	14|10	3|15	9|19
	m_2	-	-	-	-	-	-	-	-	59|0	9|13	2|21	10|15
	m_5	-	-	-	-	-	-	-	-	-	0|57	0|56	0|54
skewt-t	m_1	-	-	-	-	-	-	-	-	-	-	4|17	10|16
	m_2	-	-	-	-	-	-	-	-	-	-	-	12|15
	m_5	-	-	-	-	-	-	-	-	-	-	-	-

Note: The true DGP is the skewt-t model with parameter vector $(\beta, \nu^{-1}, \lambda)' = (1, 0.25, -0.5)'$. The *n-n*, *skewn-n*, *t-t* and *skewt-t* factor copula models are the candidate models. The candidate sets of moments are m_1, m_2 and m_5. The marginal distributions of the simulated data follow an AR(1)-GARCH(1,1) process, see equation (14). $\overline{\text{Cov}}$ denotes the average coverage rate of the VaR violations over $R = 100$ Monte Carlo replications. $\overline{\text{\%Bias}}$ and $\overline{\text{\%RMSE}}$ are the average percentage bias and the average percentage RMSE of the VaR forecasts over the $T_s = 500$ forecasting horizon. $\overline{\text{Loss}}$ is the average loss of the VaR forecasts using (16) over T_s and R Monte Carlo replications. The left (right) number in each cell in the lower panel reports the number of times that the row model significantly outperforms (underperforms) the column model, using a Diebold-Mariano test at the 5% significance level.

Table 5: Summary Statistics and parameter estimates for the margins

	SAP	SIE	BAYN	ALV	BAS	DAI	DTE	BMW	DPW	CON
					Panel A: Descriptive Statistics					
Mean	0.049	0.027	0.026	0.031	0.028	0.031	0.026	0.053	0.040	0.104
Std dev	1.374	1.445	1.663	1.629	1.623	1.898	1.484	1.899	1.519	2.212
Skewness	-0.487	-0.243	-0.161	-0.117	-0.235	-0.221	-0.285	-0.272	-0.086	-0.050
Kurtosis	5.464	4.821	4.447	6.874	4.286	4.464	7.499	5.449	4.294	5.389
					Panel B: Parameter estimates for the margins					
ϕ_0	0.051	0.026	0.026	0.029	0.026	0.028	0.025	0.050	0.037	0.098
ϕ_1	-0.028	0.010	-0.026	0.060	0.033	0.070	-0.006	0.049	0.028	0.040
ω	0.066	0.032	0.049	0.083	0.061	0.079	0.026	0.035	0.080	0.064
α	0.010	0.005	0.000	0.004	0.011	0.029	0.043	0.011	0.000	0.001
γ	0.074	0.068	0.080	0.176	0.082	0.074	0.020	0.058	0.096	0.092
β	0.916	0.945	0.944	0.876	0.924	0.912	0.938	0.951	0.917	0.937
ν	6.107	7.416	7.655	6.098	10.280	10.189	4.643	8.169	7.927	13.688
λ	-0.149	-0.066	-0.032	-0.080	-0.111	-0.089	-0.050	-0.042	-0.037	-0.051

Note: Panel A presents the descriptive statistics for each stock return in the portfolio. Panel B contains the estimates of the parameters of the AR(1)-GJR-GARCH(1,1) models, see equation (17).

Table 6: Sample dependence measures

	SAP	SIE	BAYN	ALV	BAS	DAI	DTE	BMW	DPW	CON
					Panel A: Spearman's rho					
SAP	-	0.562	0.547	0.540	0.562	0.532	0.466	0.513	0.496	0.482
SIE	-	-	0.622	0.634	0.707	0.641	0.539	0.629	0.597	0.565
BAYN	-	-	-	0.604	0.669	0.592	0.556	0.563	0.564	0.507
ALV	-	-	-	-	0.660	0.600	0.580	0.567	0.611	0.540
BAS	-	-	-	-	-	0.653	0.534	0.628	0.590	0.589
DAI	-	-	-	-	-	-	0.515	0.796	0.581	0.693
DTE	-	-	-	-	-	-	-	0.462	0.502	0.402
BMW	-	-	-	-	-	-	-	-	0.556	0.696
DPW	-	-	-	-	-	-	-	-	-	0.562
					Panel B: Quantile dependence λ_q					
SAP	-	0.153	0.153	0.215	0.215	0.123	0.092	0.153	0.123	0.123
SIE	-0.123	-	0.184	0.215	0.184	0.184	0.031	0.215	0.215	0.246
BAYN	-0.135	-0.086	-	0.215	0.460	0.246	0.123	0.215	0.215	0.153
ALV	-0.104	-0.068	-0.092	-	0.307	0.215	0.092	0.246	0.184	0.153
BAS	-0.092	-0.080	-0.117	-0.074	-	0.276	0.153	0.276	0.184	0.153
DAI	-0.061	-0.123	-0.086	-0.074	-0.055	-	0.061	0.430	0.276	0.276
DTE	-0.061	-0.129	-0.104	-0.092	-0.068	-0.068	-	0.092	0.092	0.000
BMW	-0.092	-0.055	-0.141	-0.153	-0.068	-0.080	-0.098	-	0.123	0.246
DPW	-0.086	-0.178	-0.184	-0.092	-0.098	-0.117	-0.092	-0.080	-	0.215
CON	-0.110	-0.160	-0.110	-0.074	-0.049	-0.080	-0.074	-0.080	-0.068	-

Note: This table contains the sample dependence measures between the ten stocks. The numbers in Panel A are the estimates of Spearman's rho. In Panel B, the upper triangular part shows the average of the 1% and 99% quantile dependence measures, i.e., $(\hat{\lambda}_{0.01} + \hat{\lambda}_{0.99})/2$; the lower triangular part presents the difference between 90% and 10% quantile dependence measures, i.e., $\hat{\lambda}_{0.9} - \hat{\lambda}_{0.1}$.

Table 7: Parameter estimates, J test and selection criteria for combinations of factor copula models and candidate sets of moments

Moments	Models	$\hat{\beta}$	$\hat{\nu}^{-1}$	$\hat{\lambda}$	p-value	AIC	BIC	HQIC
m_1	n-n	1.478	-	-	0.950	-5.781	-21.969	-11.847
	t-t	1.498	0.010	-	0.889	-3.939	-14.730	-7.982
	t-n	1.496	0.011	-	0.942	-3.969	-14.761	-8.013
	skewn-n	1.709	-	0.234	0.519	-3.004	-13.795	-7.048
	skewt-t	2.210	0.119	0.434	0.002	-1.349	-6.744	-3.370
	skewt-n	1.477	0.012	-0.017	0.568	-1.968	-7.364	-3.990
m_2	n-n	1.397	-	-	0.201	2.352	-19.231	-5.735
	t-t	1.323	0.102	-	0.112	0.193	-15.994	-5.872
	t-n	1.358	0.064	-	0.065	0.692	-15.495	-5.373
	skewn-n	1.227	-	-0.601	1.000	-5.993	-22.180	-12.059
	skewt-t	1.427	0.286	0.203	0.000	12.968	2.176	8.924
	skewt-n	1.232	0.011	-0.344	0.369	-3.361	-14.153	-7.405
m_3	n-n	1.267	-	-	0.013	12.876	-8.707	4.789
	t-t	1.232	0.094	-	0.018	13.094	-3.094	7.028
	t-n	1.252	0.071	-	0.018	13.457	-2.730	7.391
	skewn-n	1.293	-	-0.267	0.466	-4.005	-20.192	-10.071
	skewt-t	1.247	0.080	-0.196	0.366	-3.486	-14.277	-7.529
	skewt-n	1.264	0.056	-0.205	0.316	-3.417	-14.209	-7.461
m_4	n-n	1.254	-	-	0.157	-5.192	-26.775	-13.279
	t-t	1.334	0.438	-	0.304	-4.514	-20.702	-10.580
	t-n	1.535	0.276	-	0.153	-3.316	-19.503	-9.382
	skewn-n	1.256	-	-0.316	0.181	-4.606	-20.793	-10.671
	skewt-t	1.357	0.442	-0.131	0.958	-3.987	-14.778	-8.030
	skewt-n	1.691	0.326	-0.139	0.094	-2.705	-13.496	-6.748
m_5	n-n	1.270	-	-	0.029	6.934	-36.232	-9.240
	t-t	1.262	0.071	-	0.025	7.713	-30.057	-6.440
	t-n	1.273	0.012	-	0.026	8.466	-29.304	-5.687
	skewn-n	1.297	-	-0.266	0.521	$\mathbf{-11.393}$	$\mathbf{-49.163}$	$\mathbf{-25.546}$
	skewt-t	1.275	0.057	-0.202	0.415	-10.533	-42.907	-22.664
	skewt-n	1.285	0.037	-0.208	0.345	-10.335	-42.709	-22.466

Note: This table presents the parameter estimates, p-values of the J test for overidentification, and the SMM-AIC, SMM-BIC and SMM-HQIC selection criteria for all combinations of 6 factor copula models and 5 sets of moments. The constant $Q = 2.01$ is selected for the SMM-HQIC criterion. The minimum values for each of the three criteria among the 30 combinations are marked in bold.

Table 8: Empirical evaluation of the VaR forecasting performance

	Coverage	Kupiec p value	Christoffersen p value	Dynamic Quantile p value
		Panel A: 5%-VaR		
n-n	0.064	0.133	0.305	0.273
t-t	0.067	0.066	0.179	0.151
t-n	0.064	0.133	0.305	0.273
skewn-n	0.064	0.133	0.305	0.273
skewt-t	0.064	0.133	0.305	0.273
skewt-n	0.062	0.182	0.377	0.345
		Panel B: 1%-VaR		
n-n	0.013	0.511	0.725	0.738
t-t	0.011	0.780	0.888	0.918
t-n	0.011	0.780	0.888	0.918
skewn-n	0.010	0.907	0.938	0.966
skewt-t	0.010	0.907	0.938	0.966
skewt-n	0.010	0.907	0.938	0.966

Note: This table presents the coverage rates of the VaR violations and p-values from the Kupiec test, the Christoffersen test and the dynamic quantile test. The upper and lower panels contain the results for 5%-VaR and 1%-VaR forecasts from six factor copula models, respectively. The results are based on the set of moment conditions m_5.

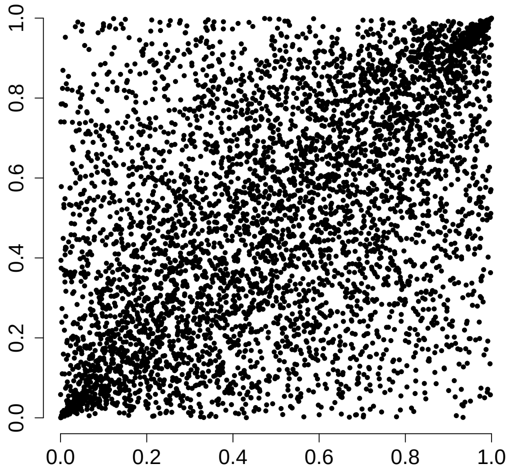
Table 9: Comparison of the relative forecasting accuracy of the 5%-VaR and 1%-VaR by the Diebold-Mariano test

	n-n	t-t	skewt-t	skewn-n	skewt-n	t-n
			Panel A: 5%-VaR			
$\overline{\text{Loss}} * 10^2$	0.172	0.174	0.181	0.177	0.181	0.175
			Diebold-Mariano test			
n-n	-	-3.808	1.370	3.881	3.488	-3.812
t-t	-	-	3.882	4.105	4.072	3.491
skewt-t	-	-	-	4.097	3.974	-3.628
skewn-n	-	-	-	-	-3.709	-4.116
skewt-n	-	-	-	-	-	-4.100
			Panel B: 1%-VaR			
$\overline{\text{Loss}} * 10^2$	0.042	0.043	0.043	0.042	0.043	0.042
			Diebold-Mariano test			
n-n	-	1.543	2.055	2.060	2.063	1.903
t-t	-	-	2.055	2.049	2.067	1.800
skewt-t	-	-	-	-1.396	2.014	-1.937
skewn-n	-	-	-	-	1.713	-1.754
skewt-n	-	-	-	-	-	-2.003

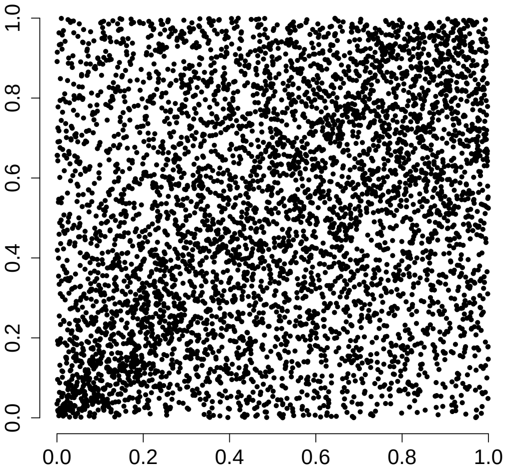
Note: $\overline{\text{Loss}}$ is the average loss of the VaR forecasts based on (16) over the testing preiod T_s. The other entries are the Diebold-Mariano test statistics. A t-statistic less than -1.96 (greater than $+1.96$) indicates that the row model has significantly lower (higher) average loss than the column model at the 5% significance level. The results are based on the set of moment conditions m_5.

C Figures

Figure 1: Scatter plot of observations generated by the (a) $t(2,0) - t(2,0)$ and the (b) $Skewt(2, -0.5) - t(2,0)$ factor copula model

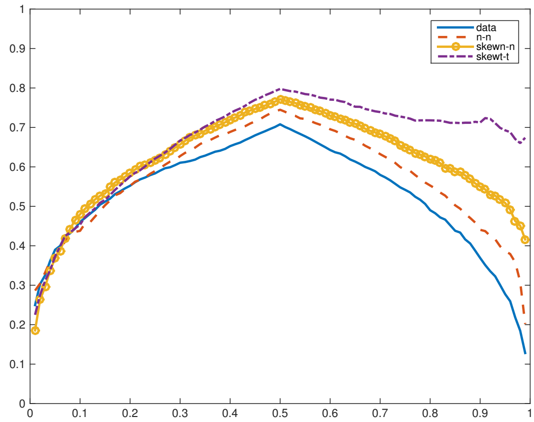


(a) t(2,0)-t(2,0)

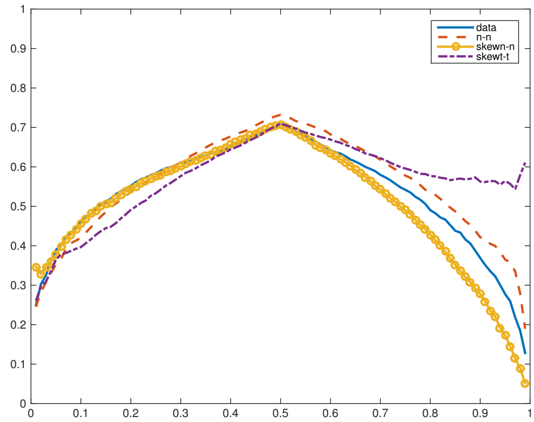


(b) Skew t(2,-0.5)-t(2,0)

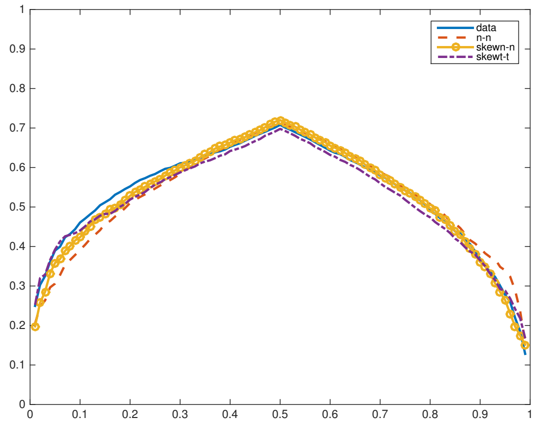
Figure 2: Sample and fitted quantile dependence measures using m_1, m_2, m_3, m_4 and m_5



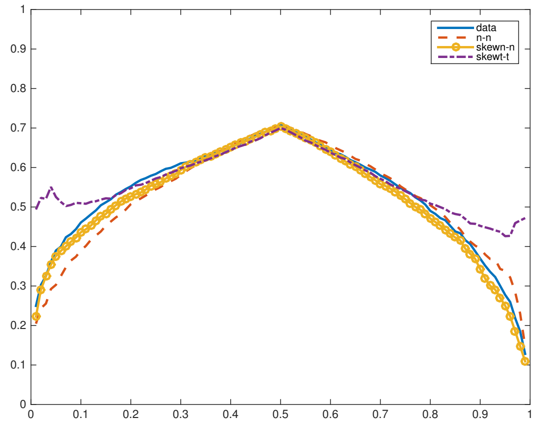
(a) $m_1 = \{\lambda_{0.01}, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.15}\}$



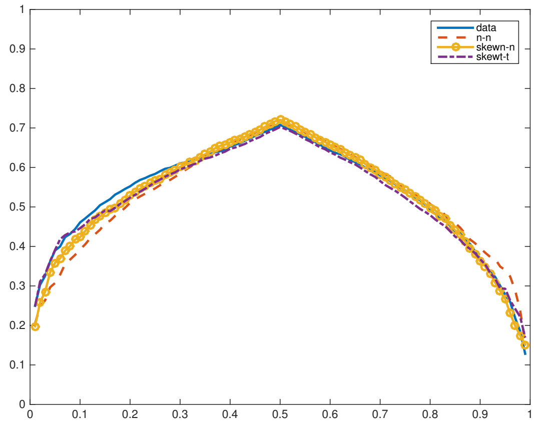
(b) $m_2 = \{\rho_S, \lambda_{0.01}, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.15}\}$



(c) $m_3 = \{\rho_S, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.9}, \lambda_{0.95}\}$



(d) $m_4 = \{\rho_S, \lambda_{0.25}, \lambda_{0.45}, \lambda_{0.55}, \lambda_{0.75}\}$



(e) $m_5 = m_3 \cup m_4$

Note: This figure plots the sample quantile dependence measures based on the empirical data (dark blue solid line) and fitted quantile dependence measures based on three of the six estimated factor copulas, i.e., n-n copula (red dashed line), skewn-n copula (orange circle solid line) and skewt-t copula (purple dash-dot line).