Model and Moment Selection in Factor Copula Models

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

†Technical University of Dortmund & Ruhr Graduate School in Economics

‡University of Graz, Department of Economics

§University of Cologne, Institute of Econometrics and Statistics

June 4, 2018

Abstract

This paper develops a simultaneous model and moment selection procedure for factor copula models. Since the density of the factor copula is 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. Our selection procedure shows considerable selection frequencies of the true models 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; Dependence Modeling

JEL Classification C38; C52; C58

*Fang Duan gratefully acknowledges funding by Ruhr Graduate School in Economics (RGS Econ), Address: Technical University Dortmund, Otto-Hahn-Str.6, 44227, Dortmund, Germany, Email: duan@statistik.tu-dortmund.de

1 Introduction

Copula models are superior in multivariate modeling compared to standard multivariate distributions in terms of their flexibility, i.e., one can construct the joint distribution by separately specifying the marginal distributions and the dependence structure that link 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 Archimedean or elliptical copulas 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 copula that allows 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 of Value-at-Risk (VaR) of a portfolio using a non-normal joint distribution.

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 a problem with 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 to represent 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 movement of the return series in a stock market may possibly be associated with common factors such as the macroeconomic status.

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 tail in portfolio loss distributions as well as correlation skews in CDO tranches. van der Voort (2005) extended the factor Gaussian copula with a single factor to the one which 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 that 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 estimated 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 one proposed by Hull & White (2004) in the evaluation of default probabilities and default correlations. The model is estimated using the simulated method of moments due to the fact that a closed form likelihood is not available. 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, for example, in the *Skew t-t* factor model. Given an asymmetric and fat-tailed data generating process (DGP), the factor copula with the *Skew t-t* factor probably outperforms a model merely with a Gaussian factor. The remaining problem is whether one could consistently select the true model or an approximately true model from a pool of model candidates. 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 probably less helpful in the comparison between a pair 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 parameterized models. When it comes to the moment selection, since the Simulated Method of Moments (SMM) serves as a popular method in the estimation of 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. Indeed, it is unrealistic to take infinitely many moment conditions into account. One might be interested in characterizing a certain group of moments which plays a significant role in model selection, especially when one deals with problems such as VaR estimation or prediction. To this end, we propose a plausible procedure of simultaneous selection of moments and models for factor copulas. From the practical point of view in risk management, we further investigate the interplay between selection of moments and models in backtesting VaR forecasts. Considering that the true model in empirical data is generally unknown, we point out the possibility of obtaining acceptable risk forecasts based on misspecified models with the usage of a specific set of moments.

The rest of this paper is organized as follows. We briefly introduce factor copula models in Section 2. In Section 3, we shortly review the estimation method for factor copulas and necessary assumptions for the consistent copula parameter estimator, after which we discuss the model and moment selection criterion and its consistency in Section 4. In Section 5 we present finite sample simulation studies, followed by an 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:

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

where Z is the common factor. β_i is the factor loading of the common factor associated with i-th latent variable $X_i, i = 1, \ldots, N$, whereas $\boldsymbol{\theta} = (\boldsymbol{\theta}'_Z, \boldsymbol{\theta}'_\epsilon, \beta_1, \ldots, \beta_N)'$ denotes the vector of the copula parameters. 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 $X_i, i = 1, \ldots, N$ are not necessary to be identical to those of the observable variables $Y_i, i = 1, \ldots, N$, respectively. The latent factor structure is employed to model the copula and the estimation of the marginal distributions is left to, for example, semiparametric models. Although we impose the factor structure to overcome the difficulty of the estimation of copula parameters in high dimensions, one of the potential challenges of the estimation in such a specification remains: N factor loadings have to be estimated, which would still cause "curse of dimensionality". In response, Oh & Patton (2017) suggested that one can sort the latent variables into groups and formulate a so-called block equi-dependence model:

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

where $\boldsymbol{\theta} = (\boldsymbol{\theta}'_Z, \boldsymbol{\theta}'_\epsilon, \beta_{B_1}, \ldots, \beta_{B_N})'$. B_i denotes the block to which X_i belongs and β_{B_i} is the corresponding loading coefficient associated with the common factor for variable X_i in block B_i. The intuition of such a clustering strategy can be implied by the scenario that, for example, 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 excessive number of the latent factors. The multi-factor copula model allows for more than one latent factor in the linear structure:

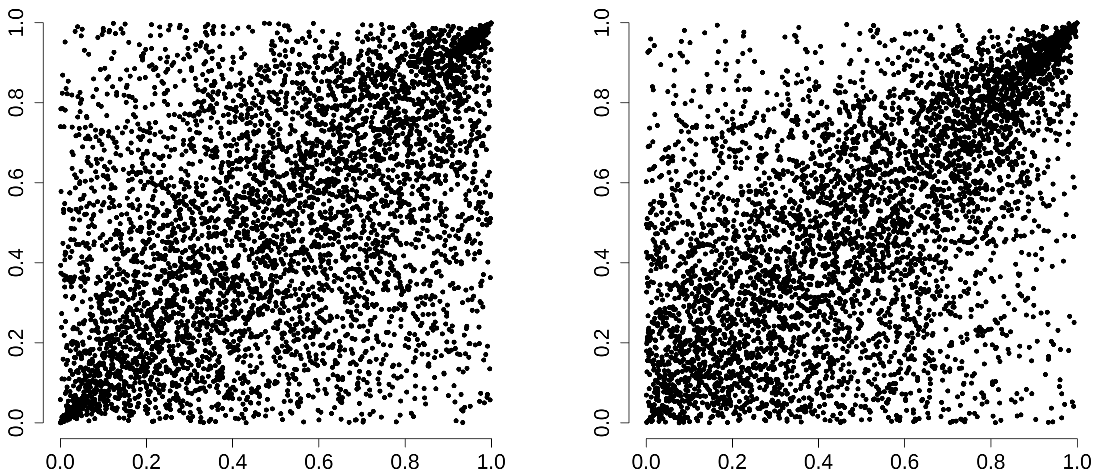
$$\begin{aligned} X_i &= \textstyle\sum_{k=1}^K \beta_{ik} Z_k + \epsilon_i, i = 1, 2, \cdots, N \\ Z_k &\sim F_{Z_k}(\boldsymbol{\theta}_Z), Z_k \perp Z_l, \forall k \neq l \\ \epsilon_i &\sim \text{iid } F_\epsilon(\boldsymbol{\theta}_\epsilon), \text{ and } \epsilon_i \perp Z_k, \forall i, k \\ \boldsymbol{X} = (X_1, \cdots, X_N)' &\sim F_{\boldsymbol{X}} = \boldsymbol{C}(G_1(\boldsymbol{\theta}), \ldots, G_N(\boldsymbol{\theta}); \boldsymbol{\theta}), \end{aligned} \tag{3}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}'_Z, \boldsymbol{\theta}'_\epsilon, \boldsymbol{\beta}'_1, \ldots, \boldsymbol{\beta}'_N)'$ and $\boldsymbol{\beta}_i = (\beta_{i1}, \ldots, \beta_{iK})', i = 1, \ldots, N$. $Z_k, k = 1, \ldots, K$ are allowed to follow different distributions. A similar setting of the 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.

In our context, the factor copula model is generally not known in closed form. For example, 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 not known in closed form if Z and $\epsilon_i, i = 1, \ldots, N$ come from different families of distribution. To specify a factor copula, it is crucial to choose the distributions of the common factors and the error term in the first place. Among assorted options of the distributions of common and idiosyncratic variables, a favorable choice is skewed student's 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, λ for skewness. The lower ν indicates fatter tails. When $\lambda = 0$, it collapses to the standard student's 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-thickness parameter and



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

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

the skewness parameter affect the dependence structure of the generated data after probability integral transformation, Figure 1 is generated based on $T = 5000$ bivariate random samples with the factor loading $\beta = 1$, shape parameter $\nu = 2$ and skewness parameter $\lambda = -0.5$ in the factor structure. Since the data points in Figure 1(a) are generated by factor t-t model, two clusters emerge at both tails. The asymmetric clusters at the tails are apparent in Figure 1(b) as the skewness parameter comes into play.

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, its Spearman's rank correlation is defined as

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

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, the 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) of order 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 modeling concurrent extreme events in financial markets. The quantile dependence of order q between η_i and η_j is defined as

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

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

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

where $\hat{F}_i(\hat{\eta}_{it})$ and $\hat{F}_j(\hat{\eta}_{jt})$ are 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. Let δ_{ij} denotes a generic dependence measure between variable i and variable j such as ρ_S^{ij} or λ_q^{ij}. Then the corresponding matrix of pairwise dependence measures is

$$D = \begin{pmatrix} 1 & \delta & \cdots & \delta_{1N}\\ \delta_{12} & 1 & \cdots & \delta_{2N}\\ \vdots & \vdots & \ddots & \vdots\\ \delta_{N1} & \delta_{N2} & \cdots & 1 \end{pmatrix}.$$

In the equi-dependence model, after we take the average of the upper off-diagonal pairwise elements in D, the vector of dependence measures is

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

where $q_1, \ldots, q_{r-1} \in (0, 1)$ denote the quantiles. In the block equi-dependence model, one assumes that any pair of variables in a block has the same dependence as any other pair of variables in the same block. This allows for the constructing moment conditions based on the average of within-block and between-block dependence measures; see Oh & Patton (2017) for details.

3.2 Simulated Method of Moments

If the density of the factor copula is known in closed form, or copula parameters have a direct relation with population moments or dependence measures, then one could use the sample counterpart of dependence measures to estimate the model parameters. However, the factor copula does not have a closed-form likelihood and the Maximum Likelihood Estimation (MLE) is therefore generally not applicable. If such circumstances occur, the estimation method can be adapted to the Simulated Method of

Moments (SMM), 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 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):

$$\boldsymbol{Y}_t = \boldsymbol{\mu}_t(\boldsymbol{\phi}) + \boldsymbol{\sigma}_t(\boldsymbol{\phi})\boldsymbol{\eta}_t, \tag{4}$$

where

$$\begin{aligned}
\boldsymbol{\mu}_t(\boldsymbol{\phi}) &= [\mu_{1,t}(\boldsymbol{\phi}), \ldots, \mu_{N,t}(\boldsymbol{\phi})]', \\
\boldsymbol{\sigma}_t(\boldsymbol{\phi}) &= \text{diag}[\sigma_{1,t}(\boldsymbol{\phi}), \ldots, \sigma_{N,t}(\boldsymbol{\phi})]', \\
\boldsymbol{\eta}_t &\sim_{iid} \boldsymbol{F}_\eta = \boldsymbol{C}(F_1, \ldots, F_N; \boldsymbol{\theta}_0),
\end{aligned}$$

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 marginal distributions. When the dynamic parameter vector is known or both $\boldsymbol{\mu}_t(\boldsymbol{\phi})$ and $\boldsymbol{\sigma}_t(\boldsymbol{\phi})$ are constant, this model fits the i.i.d data series. The estimation of the parameter vector $\boldsymbol{\theta}_0$ is constructed on estimated standardized residuals $\hat{\boldsymbol{\eta}}_t = \boldsymbol{\sigma}_t^{-1}(\hat{\boldsymbol{\phi}})(\boldsymbol{Y}_t - \boldsymbol{\mu}_t(\hat{\boldsymbol{\phi}}))), t = 1, \ldots, T$ plus simulations from the multivariate distribution $\boldsymbol{F}_x$. Define $\hat{\boldsymbol{m}}_T$ as a $r \times 1$ 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:

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

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

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

$\hat{\boldsymbol{W}}_T$ 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 construction of a consistent 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 $\boldsymbol{g}_{T,S}(\boldsymbol{\theta})$:

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

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 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 arisen 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 specification, whereas the misspecified case is untouched, since the pseudo-true SMM estimator depends on the limit distribution of the weighting matrix, and an alternative procedure for construction of stochastic equicontinuity of the objective function is required.

4 Model and Moment Selection Procedure

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 model selection of parametric copula models, see, e.g., Dias & Embrechts (2004). Manner (2007) provided a simulation study of the AIC model selection procedure for copula models which possess symmetric and asymmetric tail behavior, and applied such models in the modeling of exchange rates. Chen & Fan (2006) extended a pseudo likelihood ratio test for model selection between two semiparametric copula-based multivariate dynamic models under misspecification. Another appealing class of model selection procedures relies on Goodness-of-Fit (GoF) tests, see Patton (2012) for a general review. Two widely favored GoF tests are the Kolmogorov-Smirnov(KS) and the Crámer-von-Mises (CvM) tests. Rémillard (2017) pointed out that one can use CvM-type statistic, which is powerful and easier to compute compared to KS statistic, to test if the copula candidate belongs to a specific family of parametric copulas.

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. Most recently, Oh & Patton (2013) suggested 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:

$$J_{T,S} \coloneqq \min(T,S)\boldsymbol{g}_{T,S}(\hat{\boldsymbol{\theta}}_{T,S})'\hat{\boldsymbol{W}}_T\boldsymbol{g}_{T,S}(\hat{\boldsymbol{\theta}}_{T,S}) \to_d \boldsymbol{u}'\boldsymbol{A}_0'\boldsymbol{A}_0\boldsymbol{u}, \quad T,S\to\infty. \tag{5}$$

where $\boldsymbol{A}_0 \coloneqq \boldsymbol{W}_0^{1/2}\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{R}_0$, $\boldsymbol{\Sigma}_0$ denotes the asymptotic variance of $\hat{\boldsymbol{m}}_T$ and

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

and $\boldsymbol{u} \sim N(0,\boldsymbol{I})$. If one uses efficient weighting matrix $\hat{\boldsymbol{\Sigma}}_{T,B}^{-1}$ as $\hat{\boldsymbol{W}}_T$, the limit distributions turns out to be the chi-squared distribution $\chi^2_{r,p}$. If a weighting matrix different from the efficient weighting matrix, such as the identity matrix, is employed, the asymptotic distribution of J test statistics is nonstandard and the corresponding critical values can be obtained via simulation. However, this approach 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 rank correlation and quantile dependence, some of moments are correct whereas the rest of moments are incorrect. The selection of correct moments has an interplay with the model selection: under the true model, with increasing inclusion of moments, the procedure tends to select all moment conditions asymptotically. If the model is not true, 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 select the true model. One the other hand, if the set of moments are predetermined in a specific region of the support, e.g., quantile dependence measures are restricted at certain quantiles, whether the true model is selected or not remains a problem. We would like to seek a method for the consistent selection of correct moments in addition to the selection of the true model.

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. We propose a simultaneous model and moment selection procedure for the factor copula model in the SMM context. Our selection criterion measures the degree of the model specification and rewards the selection vectors that use a higher number of over-identifying restrictions in the model. This is an analogous procedure of classic model selection criteria, which utilizes the log-likelihood as well as a bonus term rewarding the selection vectors that utilize more moment conditions and less parameters.

Since it is possible that the model candidate is misspecified, it is necessary to develop a SMM estimator under model misspecification. Oh & Patton (2013) transplanted 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 SMM estimator* $\boldsymbol{\theta}_*(\boldsymbol{W}_0) = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \boldsymbol{g}_0(\boldsymbol{\theta})'\boldsymbol{W}_0\boldsymbol{g}_0(\boldsymbol{\theta})$.

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 $T, S \to \infty$.

Following Andrews & Lu (2001), let $(\boldsymbol{b}, \boldsymbol{c}) \in \mathbb{R}^p \times \mathbb{R}^r$ be a pair of model and moment selection vectors, which only contain zeros and ones. For instance, if the i-th element of $\boldsymbol{c}$ equals to one, then the i-th moment condition is inserted into the SMM estimation procedure whereas the i-th moment condition is excluded in the estimation when the corresponding i-th position in $\boldsymbol{c}$ takes the value zero. Similarly, if the i-th element of $\boldsymbol{b}$ equals to one, then the i-th element of the parameter vector $\boldsymbol{\theta}$ is to be included in the model and be estimated, whereas if the i-th element in the vector $\boldsymbol{b}$ is zero, then the i-th element of the parameter vector $\boldsymbol{\theta}$ is set to be zero and is not estimated. Define

$$\begin{aligned} \mathcal{M} &= \{(\boldsymbol{b}, \boldsymbol{c}) \in \mathbb{R}^p \times \mathbb{R}^r : b_i = 0 \text{ or } 1, \forall 1 \le i \le p, c_j = 0 \text{ or } 1, \forall 1 \le j \le r, \\ &\quad \text{where } \boldsymbol{b} = (b_1, \ldots, b_p)', \boldsymbol{c} = (c_1, \ldots, c_r)'\}. \end{aligned}$$

As a result, $\boldsymbol{\theta}^b$ is the element-wise product of $\boldsymbol{\theta}$ and $\boldsymbol{b}$, and $\boldsymbol{g}^c_{T,S}(\boldsymbol{\theta})$ is moment conditions specified by $\boldsymbol{c}$. There exists a correspondence between the model selection vector $\boldsymbol{b}$ and the corresponding model candidates. Let $\mathcal{D}$ be the set of l models such that $\mathcal{D} = \{d_0, d_1, \ldots, d_{l-1}\}$, where d_0 denotes the true model, the $d_1, \ldots, d_{l-1}$ are other possible model candidates. One pair of selection vectors $(\boldsymbol{b}, \boldsymbol{c}) \in \mathcal{M}$ uniquely corresponds to $d_i \in \mathcal{D}, i = 0, \ldots, l-1$ with the vector of moment selection $\boldsymbol{c}$. For example, $\boldsymbol{b} = (1, 1)'$ of $\boldsymbol{\theta} = (\nu^{-1}, \lambda)'$ gives a *Skew t-t* factor copula model whereas $\boldsymbol{b} = (1, 0)'$ of $\boldsymbol{\theta} = (\nu^{-1}, \lambda)'$ implies a *t-t* factor copula. Similarly, 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. This setup certainly 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 assigns 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.

Denote the selection vector of correct moments as $\boldsymbol{c}^0$ and the selection vector of correct model as $\boldsymbol{b}^0$, respectively. $\boldsymbol{c}^0$ selects the moments for which $\boldsymbol{g}_0^c(\boldsymbol{\theta}) = 0$, 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:

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

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

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

where $|\boldsymbol{b}|$ denotes the number of parameters to be estimated in the vector $\boldsymbol{b}$ such that $|\boldsymbol{b}| = \sum_{i=1}^p b_i$, and $|\boldsymbol{c}| = \sum_{j=1}^r c_j$ denotes the number of moments selected in the vector $\boldsymbol{c}$. The parameter space of $(\hat{\boldsymbol{b}}, \hat{\boldsymbol{c}})$ is denoted as $\mathcal{P} \subset \mathcal{M}$, where $(\hat{\boldsymbol{b}}, \hat{\boldsymbol{c}})$ is a generic estimator of model and moment selection vectors. The parameter space $\mathcal{P}$ is smaller space than $\mathcal{M}$ such that the estimator vector $(\hat{\boldsymbol{b}}, \hat{\boldsymbol{c}})$ has a good finite sample behavior and it is computationally easy. For instance, when moment conditions are comprised of quantile dependences $\lambda_q, q \in (0, 1)$, it is infeasible to consider infinitely many quantile dependences on the grid of q. A huge amount of the utilization of quantile dependences would lead a cumbersome computational burden in the estimation of $(\hat{\boldsymbol{b}}, \hat{\boldsymbol{c}})$. The parameter space $\mathcal{P}$ should 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 with maximum number of over-identifying restrictions. The simultaneous model and moment selection criterion is defined as:

$$MSC_{T,S}(\boldsymbol{\theta^b}, \boldsymbol{c}) = J^{\boldsymbol{c}}_{T,S}(\boldsymbol{\theta^b}) - (|\boldsymbol{c}| - |\boldsymbol{b}|)\kappa_{T,S}, \tag{6}$$

where $J^{\boldsymbol{c}}_{T,S}(\boldsymbol{\theta^b}) = \min(T,S)\boldsymbol{g}^{\boldsymbol{c}}_{T,S}(\boldsymbol{\theta^b})'\hat{\boldsymbol{W}}^c_T\boldsymbol{g}^{\boldsymbol{c}}_{T,S}(\boldsymbol{\theta^b})$. $|\boldsymbol{c}|$ 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 ρ_S and quantile dependence $\lambda_q, q \in (0,1)$ throughout this paper. As exemplified, the penalizing factor $\kappa_{T,S}$ can be specified in the following three versions which are analogous to classical information-based criteria for the model and moment selection:

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

The estimator of the vector of the model and moment selection based on MSC is decided by minimizing the MSC criterion:

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

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 a higher number of 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 that more incorrect moments, which enlage the J statistic, are included. Therefore, the model and moment selection vector is determined through minimizing the MSC criterion.

The consistency of the MSC estimator of selection vector requires the following two assumptions being fulfilled:

Assumption 4 *$\mathcal{M}_2 = \{(\boldsymbol{b}^0, \boldsymbol{c}^0)\}$, where $\boldsymbol{b}^0$ is the model selection vector for the true model.*

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

(ii) The number of moments is greater than that of the parameter vector, i.e., $|\boldsymbol{c}| - |\boldsymbol{b}| > 0$.

Assumption 4 guarantees that the set contains one single element, i.e., the selection vector of the true model and the correct moments. However, this assumption may not hold under certain circumstances, for example, if the selected moment conditions are 'insufficient' for the selection of the true model. More specifically, given a certain set of moment conditions, there exists a selection vector $(\boldsymbol{b}^*, \boldsymbol{c}^*) \in \mathcal{M}_1$ associated with a model which is not true, but has the same number of over-identifying conditions as the selection vector which represents the true model. Both $(\boldsymbol{b}^*, \boldsymbol{c}^*)$ and $(\boldsymbol{b}^0, \boldsymbol{c}^0)$ naturally enter into $\mathcal{M}_2$ and the selection procedure could fail to discriminate between those two vectors. Alternatively, there exists a selection vector $(\boldsymbol{b}^{**}, \boldsymbol{c}^{**}) \in \mathcal{M}_1$ representing a model which is not the true model but less parametrized compared to the selection vector of the true model. Then such a vector $(\boldsymbol{b}^{**}, \boldsymbol{c}^{**}) \in \mathcal{M}_1$ is the only possible element contained in $\mathcal{M}_2$ and the selection procedure opts for this selection vector instead of the true one. If the latter circumstance occurs, one might redefine the 'correct' selection vector $(\boldsymbol{b}^0, \boldsymbol{c}^0)$ such that the 'correct' model selection vector $\boldsymbol{b}^0$ out of the pair $(\boldsymbol{b}^0, \boldsymbol{c}^0) \in \mathcal{M}_2$ is the model selection vector which points to the least parametrized model when moment conditions hold asymptotically. Assumption 4 could be rewritten in an adapted way: $\mathcal{M}_2 = \{(\boldsymbol{b}^0, \boldsymbol{c}^0)\}$ with $|\boldsymbol{b}^0| = \min\{|\boldsymbol{b}| : (\boldsymbol{b}, \boldsymbol{c}) \in \mathcal{M}_2\}$. The 'correct' selection vector $(\boldsymbol{b}^0, \boldsymbol{c}^0)$ gives us the least parametrized model out of all selection vectors with the maximized number of over-identifying conditions. The above discussion only points out the consequence of the failure of Assumption 4 and the possibility that the consistency of the selection procedure can be constructed in a more general way. The second part of the proposition below is formulated for the case that Assumption 4 holds.

Proposition 1 *(a) Suppose Assumption 1, 2, 3 and 5 hold. Then, $(\hat{\boldsymbol{b}}_{MSC}, \hat{\boldsymbol{c}}_{MSC}) \in \mathcal{MP}_2, wp \to 1$.*

(b) If additionally Assumption 4 holds, then $(\hat{\boldsymbol{b}}_{MSC}, \hat{\boldsymbol{c}}_{MSC}) = (\boldsymbol{b}^0, \boldsymbol{c}^0) \in \mathcal{P}, wp \to 1$.

The proof is found in the appendix. $wp \to 1$ denotes with the probability goes to 1 as $T, S \to \infty$. The simultaneous selection procedure asymptotically selects the true model and possibly the largest number of correct moment conditions, i.e., $\boldsymbol{c}_0 = \{\boldsymbol{c} : \max |\boldsymbol{c}|, (\boldsymbol{b}, \boldsymbol{c}) \in \mathcal{MP}_1, \text{ for some } \boldsymbol{b}\}$. Note that even when Assumption 4 fails, the estimator of the selection vector belongs to $\mathcal{MP}_2$, i.e., the set of selection vectors with maximized number of over-identifying restrictions from all selection vectors that moment conditions asymptotically hold under some model specifications, with probability going to 1, as $T, S \to \infty$.

5 Simulation Study

We investigate the finite sample performance of the selection procedure in this section. Section 5.1 focuses on the performance of the in-sample fit of models along with moments based on the procedure. A horse race of these models is conducted under various scenarios of dimensionalities and sample sizes. In addition, we try to identify the useful moments in the model selection. Section 5.2 is concerned with the out-of-sample forecasting performance among combinations of models and moments based on the simultaneous selection procedure.

5.1 In-sample Fit

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

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

All factor loadings are restricted to be identical, i.e., the true coefficient $\beta_0 = 1$. This corresponds to an equi-dependence model and the variance of the common factor is half of each variable $X_i, i = 1, \ldots, N$. We consider a dependence structure in the marginal distributions of the variables, i.e., we assume that it follows a AR(1)-GARCH(1,1) process:

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

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 dynamic 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 and factor coefficients in the second stage. The procedure is briefly summarized as:

1. Generate a $(T \times N)$ matrix of standard uniformly distributed pseudo observations from a given factor copula.

2. Transform each margin (a column in the data matrix) by the inverse AR(1)-GARCH(1,1) filter with Gaussian innovations and the dynamic parameters in the marginal distributions $[\phi_0, \phi_1, \omega, \alpha, \gamma] = [0.01, 0.05, 0.05, 0.1, 0.85]$ to obtain the true data sample.

3. Estimate each marginal series of the sample with AR(1)-GARCH(1,1) model and calculate the corresponding residuals $\hat{\boldsymbol{\eta}}_t$, then obtain the probability transformation of residuals $\hat{\boldsymbol{u}}_t$ with the empirical distribution function.

4. Use factor copula model candidates fit $\hat{\boldsymbol{u}}_t$, the unknown parameters are estimated by SMM with the identity weighting matrix. Then calculate the quantity of the model selection procedure MSC.

5. Repeat the steps 1-4 for $R = 200$ times and record the empirical selection frequencies of model and moment candidates based on the selection procedures.

Two further model specifications are considered. In the first scenario we only consider the factor loading β as the unknown parameter but fix all copula parameters as constants, this yields a non-nested model; in the second scenario both the factor loading and the copula parameters are treated as unknowns, the selection problem ends up with the one for the nested model. The true underlying DGP of four model candidates are summarized in Table 1. First, note that shape parameters in the latent factor and the

Table 1: The true DGP of the Equi-dependence model

	Model	β	$F(Z)$	$F(\epsilon)$	$\boldsymbol{b}$	$\boldsymbol{\theta}$
	normal-normal	1	$t(0,0)$	$t(0)$	$(1,0,0)'$	$\begin{pmatrix} \beta \\ \nu^{-1} \\ \lambda \end{pmatrix}$
Equidependence	Skew normal - normal	1	$t(0,-0.5)$	$t(0)$	$(1,0,1)'$	
(nested)	t - t	1	$t(0.25,0)$	$t(0.25)$	$(1,1,0)'$	
	Skew t - t	1	$t(0.25,-0.5)$	$t(0.25)$	$(1,1,1)'$	
	normal-normal	1	$t(0,0)$	$t(0)$	$(1,0,0,0)'$	$\begin{pmatrix} \beta_{\text{normal}} \\ \beta_{\text{skew n-n}} \\ \beta_{\text{t-t}} \\ \beta_{\text{skew t-t}} \end{pmatrix}$
Equidependence	Skew normal - normal	1	$t(0,-0.5)$	$t(0)$	$(0,1,0,0)'$	
(non-nested)	t - t	1	$t(0.25,0)$	$t(0.25)$	$(0,0,1,0)'$	
	Skew t - t	1	$t(0.25,-0.5)$	$t(0.25)$	$(0,0,0,1)'$	

idiosyncratic term are set to be identical. When we set the shape parameter $\nu^{-1} \to 0$ and the skewness parameter $\lambda = 0$, it corresponds to the *normal-normal* factor copula (abbreviated as *n-n* thereafter). Next, when we set $\nu^{-1} = 0.25$ and $\lambda = 0$ we obtain the *t-t* factor copula. If one further includes the skewness by setting $\nu^{-1} = 0.25$ and $\lambda = -0.5$, this implies the *skew t-t* factor copula model. If one only allows for the skewness, one obtains the *skew normal-normal* factor copula model (abbreviated as *skewn-n* thereafter). Spearman's rho ρ_S combined with quantile dependence λ_q at quantiles $q \in (0,1)$ are adopted as the dependence measures. The sets of moment conditions under consideration are $m_1 = \{\rho_S, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.9}, \lambda_{0.95}\}$, $m_2 = \{\rho_S, \lambda_{0.25}, \lambda_{0.45}, \lambda_{0.55}, \lambda_{0.75}\}$, $m_3 = \{\rho_S, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.25}, \lambda_{0.45}, \lambda_{0.55}, \lambda_{0.75}, \lambda_{0.9}, \lambda_{0.95}\}$. The choices of the sample length are $T = 500, 1000, 2000$ and three possibilities of dimensionality are $N = 3, 5, 10$. The results in all scenarios are based on $R = 200$ Monte Carlo replications.

Tables 7, 8 and 9 reveal the selection percentage for nested models, i.e., the fraction of times that a specific model is chosen among four model candidates along with three sets of moments, according to the criteria SMM-AIC, SMM-BIC and SMM-HQIC, respectively. All three versions of the selection procedure choose the models combined with the moment set with the largest number of moment conditions, i.e., m_3. Not surprisingly, the selection procedure rewards the additional usage of moment conditions. Table 13 shows the sample size effect and the dimension effect on the selection frequencies given moment conditions which only contain quantile dependence measures at the tails: 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 $N = 3$ to $N = 10$. Given the property of the equi-dependence model, there is only one unknown factor loading to be estimated. The increase of the dimensionality does not increase the number of unknown parameters, but simply offers the additional information available for unknown parameters. Therefore the selection procedure becomes more powerful not only with the increase of the sample size but also with the dimensionality. Furthermore, the choice of moment conditions affects the performance of selection procedure in the model selection. That is, the procedure behaves differently when it utilizes quantile dependence measures only at the middle area (m_2) and when it only uses tail quantile dependence measures (m_1). The cross comparison between Table 13 and Table 14 highlights the importance of the choice of quantile dependence measures: the procedure works worse with quantile dependence measures in the middle area (m_2). 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 dimensions and sample lengths, whereas the correct selection rates of the *t*-*t* factor copula stay close to zero. This is not surprising due to the fact that only the quantiles in the central region are considered and then it does not capture the dependence structure in the tail region. This could corresponds to the scenario when Assumption 4 is violated: The moments in use are such 'insufficient' that the model with the set with largest number of over-identifying restrictions stands out, but there is no guarantee that the selection procedure will select the true model.

When it comes to the selection performance for non-nested models, the selection procedure consistently select the true model along with the largest set of moment conditions in the simultaneous selection of moments and models, see Tables 10, 11 and 12. In contrary to nested models, when the quantile dependence measures are fixed within a certain region, the performance of the model selection is quite appealing regardless of the choice of quantiles, shown in Tables 15 and 16. Note that in this scenario, one only has to estimate the factor loading and leaves the factor parameters as fixed, i.e.,

$\nu^{-1} = 0.25$ and $\lambda = -0.5$. The correct model is consistently selected in each case as the probability of the correct selection goes to 1 as the dimension or the sample size increases. The first noticeable phenomenon is that the performance of the model selection is less dependent on the choice of the set of quantile dependence measures compared with the previous case, since the criteria with dependence measures at both tail and central quantiles yield almost equally good finite sample behaviors, although the one with tail quantiles provides slightly higher correct selection probabilities in general. The possible argument is that the non-nested model is less flexible, since the only unknown parameter is the factor loading coefficient and this leaves misspecified models no space to capture the characteristics of the true DGP through tunning the unknown parameter. In other words, misspecified models will drive the J test statistics to be considerably large and the choice of quantile moments becomes less crucial. In addition, the SMM-AIC, SMM-BIC and SMM-HQIC procedures give identical decisions under such a model specification. The aforementioned inner trait of non-nested models matters again for this issue, i.e., the J test statistics is sufficient to detect the misspecification and the penalizing term play a less important role.

To sum up, the simultaneous selection of models and moments consistently select the true model combined with the moment set with the largest number of moment conditions, and the selection frequencies of the correct model improve with higher dimensions or larger sample sizes except the circumstance when one only utilizes quantile dependence measures at the central region of the support in the model selection of nested equi-dependece models. If only the factor loading coefficient is allowed to be estimated in the model setup, then the selection procedure is less sensitive to both the choice of quantile dependence measures and the penalty term in selection criteria.

5.2 Prediction Performance

The precise estimation of the Value-at-Risk (VaR) and accurate VaR forecasts are often of interest in risk management. We use the following simulations to study the performance of the correctly specified model and misspecified models in terms of the VaR estimation and forecasting accuracy. The corresponding decisions made by selection procedure are also presented. Factor copulas make it possible to solely concentrate on the representation of the dependence structure in a specific region instead of the global area. Hence it is possible to restrict the choice of dependence measures at certain quantiles when one deals with a problem such as the VaR estimation or forecasts. Given that the underlying true model in reality is generally unknown, the simulation design is to additionally answer the question whether the wrong model could achieve an acceptable or even better performance in the forecasting task, especially when the misspecified model is less parameterized. The procedure is described as follows:

1. Initially, we fix the true DGP as *skewt-t* factor copula model parameterized as $\beta = 1$, $\nu^{-1} = 0.25$ and $\lambda = -0.5$. The marginal distribution is assumed as the i.i.d standard normal distribution. The time dependence structure in margins is excluded at the moment for the sake of simplicity. The portfolio consists of N series of returns and the possible dimensionalities are set to $N = 3, 5, 10$. The portfolio return at time t is determined by

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

   where $Y_{i,t}$ is the return of the i-th stock. We estimate the 'true' VaR at $\alpha \in \{0.05, 0.01\}$ based on $B_1 = 1000000$ random draws from the true DGP. The transformation from uniform random numbers to return series is accomplished by the inverse CDF Φ^{-1}.

2. Simulate a $T \times N$ series of returns from the true model with size $T = 1500$. The model candidates considered in the estimation are *n-n*, *t-t*, and *skewt-t* factor copulas. The choice of the set of moment conditions are $m_1 = \{\lambda_{0.01}, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.15}\}$ which only focuses on lower tail quantile dependence measures,

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

   which further includes Spearman's rho and

$$m_3 = \{\rho_S, \lambda_{0.01}, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.15}, \lambda_{0.85}, \lambda_{0.9}, \lambda_{0.95}, \lambda_{0.99}\}$$

   which considers quantile dependence measures at the upper tail in addition to m_2. We select the first $T_e = 1000$ sample as the estimation window for the vector of copula parameters $\boldsymbol{\theta}$, the rest of $T_s = 500$ sample is left for testing VaR forecasts. Please note that no rolling window is applied in the estimation. We estimate the factor copula model only once. The same vector of estimated copula parameters $\hat{\boldsymbol{\theta}}$ is applied in the generation of the distribution of returns at each time t in the testing period. To generate the VaR forecast, $B_2 = 100000$ residuals $\boldsymbol{\eta}$ from the estimated factor copula model based on $\hat{\boldsymbol{\theta}}$. The portfolio returns are calculated as defined above. The VaR forecasts at time t are the α quantile of the distribution of simulated returns with $\alpha \in \{0.05, 0.01\}$. The VaR_α is compared with the corresponding actual returns at each time period in the testing window and record the number of VaR violations.

3. Repeat step 2 for $R = 100$ times, record the average coverage rate of VaR violations, the percentage biasness and the RMSE of VaR forecasts. The Diebold & Mariano (1995) test is used to accomplish the pairwise comparison of the VaR forecasting accuracy among the three model candidates along with three sets of moment conditions. The loss function is defined as

$$L(Y_{p,t}, VaR_\alpha) = \alpha(Y_{p,t} - VaR_\alpha)(1 - I_{[Y_{p,t} < VaR_\alpha]}) + (1-\alpha)(VaR_\alpha - Y_{p,t}) I_{[Y_{p,t} < VaR_\alpha]}.$$

Starting from the selection frequencies in Table 2, we conduct a trial of the simultaneous selection of moments and models in Panel A. It shows a horse race among all possibilities of combinations of moments and models. Three selection criteria highly prefer the true model (*skewt-t* factor copula) with the largest set of moments (m_3), this becomes clearer when the dimensionality increases. It could be valuable to check the preference among models within each group of moments. As shown in Panel B, if the moment conditions only cover quantiles at the lower tail, the *t-t* factor copula model is chose around 70% of the time by the selection procedure regardless of the dimensionality as expected. Such a model beats the true model because of the lower parameterization, which is rewarded by the selection procedure. Furthermore, it behaves better in capturing the tail dependence compared to the *n-n* factor copula model. When ρ_S is added to the set of moments, the frequencies of the selection of the true model (*skewt-t* factor copula) becomes higher, although *t-t* factor copula, to some extend, still remains preferable. The *n-n* factor model is the most unfavorable model at this moment, since it serves as the most distinct model from the true one. When the complete moment set m_3 is used, the probabilities of the selection of the true model turn to be 99% or 100% and the *t*-*t* factor copula does not survive due to the failure of capturing asymmetric characteristics.

Tables 3 and 4 list the results of $VaR_{0.05}$ and $VaR_{0.01}$ forecasts, respectively. The wrong models are able to produce acceptable average coverage rates for VaR violations as well as the percentage biasness or the percentage RMSE in the estimation of the true VaR when a smaller set of moments is in use. For example, in $VaR_{0.05}$ forecasts for the three-dimensional series, the average coverage rate obtained from *t*-*t* factor copula are closer to the nominal level 5% when m_1 or m_2 is used. The relative bias and the RMSE are only slightly larger than that from the true model, respectively. However, when the quantiles at the upper tail enter into the set of moments, the misspecified models perform much worse than the true model in terms of the average coverage rate and the precision of the VaR estimation. The most accurate average coverage rate, the lowest percentage bias and the lowest RMSE are achieved by the true model with the most complete moments regardless of the dimensionality. The result of $VaR_{0.01}$ forecasts also verify this finding.

Lastly, the Diebold-Mariano (DM) test provides additional information on the comparison between a pair of models in terms of the VaR forecasting error. In Table 3, the dimensionality seems to play no significant role in the forecasting performance, so we take $N = 3$ as the representative example. When only quantiles at the lower tail are taken into account, the *n*-*n* factor copula model beats other model candidates as well as itself with other sets of moments. This could explained by the fact that *n*-*n* factor copula is the most parsimonious model among them. The DM test makes similar decisions for *t*-*t* factor copula under the same choice of moments when it is compared with *skewt*-*t* factor copula. When Spearman's rho is further considered, the DM test still votes frequently for the *n*-*n* factor copula, although the ratio of selecting the *n*-*n* copula with m_2 to selecting *t*-*t* factor copula (or *skewt*-*t*) with m_1 is lower than the ratio of selection frequencies in the competition between *n*-*n* factor model with m_1 and *t*-*t* factor copula (or *skewt*-*t*) with m_1. The *t*-*t* factor copula along with m_2 has been already dominated by *skewt*-*t* factor model no matter which set of moments is used in the estimation of *skewt*-*t* model. The test strongly recommends the *skewt*-*t* factor

Table 2: Empirical selection frequencies based on SMM-AIC, SMM-BIC and SMM-HQIC procedures

Panel A: overall comparison									
	n-n + m_1	n-n + m_2	n-n + m_3	t-t + m_1	t-t + m_2	t-t + m_3	skewt-t + m_1	skewt-t + m_2	skewt-t + m_3
					$N = 3$				
SMM-AIC	0.07	0	0	0.15	0.03	0	0.02	0.07	0.66
SMM-BIC	0.06	0	0	0	0	0	0	0	0.94
SMM-HQIC	0.06	0	0	0.05	0	0	0.01	0.02	0.86
					$N = 5$				
SMM-AIC	0.02	0	0	0.06	0.01	0	0.01	0.02	0.88
SMM-BIC	0.04	0	0	0	0	0	0	0	0.96
SMM-HQIC	0.04	0	0	0.03	0	0	0	0.01	0.92
					$N = 10$				
SMM-AIC	0.01	0	0	0.05	0	0	0	0.03	0.91
SMM-BIC	0.01	0	0	0	0	0	0	0	0.99
SMM-HQIC	0.01	0	0	0.02	0	0	0	0	0.97
Panel B: comparison with respect to sets of moments									
		m_1			m_2			m_3	
	n-n	t-t	skewt-t	n-n	t-t	skewt-t	n-n	t-t	skewt-t
					$N = 3$				
SMM-AIC	0.11	0.68	0.21	0.01	0.29	0.70	0.00	0.00	1.00
SMM-BIC	0.25	0.70	0.05	0.03	0.51	0.46	0.01	0.00	0.99
SMM-HQIC	0.15	0.72	0.13	0.01	0.42	0.57	0.00	0.00	1.00
					$N = 5$				
SMM-AIC	0.11	0.70	0.19	0.00	0.25	0.75	0.00	0.00	1.00
SMM-BIC	0.19	0.78	0.03	0.01	0.53	0.46	0.00	0.00	1.00
SMM-HQIC	0.14	0.75	0.11	0.00	0.38	0.62	0.00	0.00	1.00
					$N = 10$				
SMM-AIC	0.10	0.69	0.21	0.00	0.27	0.73	0.00	0.00	1.00
SMM-BIC	0.21	0.73	0.06	0.02	0.47	0.51	0.00	0.00	1.00
SMM-HQIC	0.15	0.72	0.13	0.00	0.35	0.65	0.00	0.00	1.00

Note: The true model is assumed as the *skewt-t* factor copula with $\beta = 1$, $\nu^{-1} = 0.25$ and $\lambda = -0.5$, the involved sets of moments are $m_1 = \{\lambda_{0.01}, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.15}\}$, $m_2 = \{\rho_S, \lambda_{0.01}, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.15}\}$, $m_3 = \{\rho_S, \lambda_{0.01}, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.15}, \lambda_{0.85}, \lambda_{0.9}, \lambda_{0.95}, \lambda_{0.99}\}$, the replication number is $R = 100$.

copula model if the complete moment set m_3 is used. When it comes to the evaluation of $VaR_{0.01}$ forecasts in Table 4, apart from the result that DM test makes indifferent decisions between two models much more frequently than that in the $VaR_{0.05}$ forecasting, one significant message transmitted is that *n-n* factor model becomes harder to beat by other combinations of models and moments. For instance, the test prefers the *t-t* or *skewt-t* factor copula over *n-n* factor model for $N = 5$ or $N = 10$, even if only m_1 or m_2 is employed. The *n-n* factor model cannot capture tail dependence, so it does not achieve the equivalently good performance in the VaR forecasting at such a low quantile compared with that in the $VaR_{0.05}$ forecasting. In addition, the performance of the *t-t* factor copula model is improved when one only uses tail quantiles, i.e., the moment set m_1. The model not only provides a lower percentage RMSE than that from the true model in the approximation of the true VaR, but also outperforms *skewt-t* factor copula in the VaR forecasting, which is confirmed by the DM test. Although the *t-t* factor copula is not able to capture the asymmetric characteristics from the true DGP, it is less parameterized and its performance in the VaR forecasting is quite acceptable even it is misspecified.

6 Empirical analysis

In this section, the procedure for the model and moment selection is applied to the dependence estimation of daily log returns of ten stocks in DAX 30 index: *SAP, Siemens, Deutsche Telekom, Bayer, Allianz, Daimler, BASF, BMW, Continental* and *Deutsche Post*, which possess the top ten highest market values in the index by May 31, 2016. The sample ranges from January 1, 2010 to May 31, 2016, $T = 1629$ observations are obtained after the exclusion of non-trading and settlement days. The descriptive statistics can be found in Panel A of Table 5, and it shows that the means of log returns are close to zero, all stocks are left-skewed and leptokurtic. To fit the factor copula model to uniform residuals, it is necessary to filter out the conditional mean and variance in each of ten series in the first step. We consider the AR(1)-GJR-GARCH(1,1,1) model with Generalized Error Distribution (GED) innovations,

Table 3: $VaR_{0.05}$ forecasts

	n-n + m_1	**n-n** + m_2	**n-n** + m_3	**t-t** + m_1	**t-t** + m_2	**t-t** + m_3	**skewt-t** + m_1	**skewt-t** + m_2	**skewt-t** + m_3
					$N = 3$				
Avg.cov	0.0413	0.0470	0.0594	0.0492	0.0543	0.0655	0.0498	0.0523	0.0513
%Bias	7.5427	2.7452	-6.1487	1.0574	-2.6506	-10.2552	0.7040	-1.2980	-0.5228
%RMSE	8.6818	5.2631	6.8252	4.0398	4.5269	10.5360	3.9665	4.6553	3.8223
					Diebold-Mariano test				
n-n + m_1	-	88|0	96|0	89|3	95|0	99|0	85|6	94|0	94|0
n-n + m_2	-	-	99|0	71|24	95|2	99|0	69|27	91|6	89|7
n-n + m_3	-	-	-	0|99	5|94	94|6	1|98	3|96	1|98
t-t + m_1	-	-	-	-	97|0	99|0	64|32	85|12	63|34
t-t + m_2	-	-	-	-	-	100|0	5|92	16|81	7|90
t-t + m_3	-	-	-	-	-	-	0|99	1|99	1|99
skewt-t + m_1	-	-	-	-	-	-	-	66|31	53|44
skewt-t + m_2	-	-	-	-	-	-	-	-	23|74
					$N = 5$				
Avg.cov	0.0402	0.0465	0.0609	0.0491	0.0549	0.0695	0.0499	0.0521	0.0504
%Bias	9.9040	3.5375	-8.4147	1.2028	-3.5976	-14.0159	0.5533	-1.2817	-0.0653
%RMSE	11.3094	6.6901	9.1756	5.0972	5.9483	14.2941	4.9990	5.5345	4.6703
					Diebold-Mariano test				
n-n + m_1	-	84|0	99|0	90|3	94|0	100|0	88|5	94|0	93|0
n-n + m_2	-	-	100|0	68|26	98|0	100|0	74|21	85|10	79|16
n-n + m_3	-	-	-	0|100	0|100	97|3	0|100	0|100	1|99
t-t + m_1	-	-	-	-	98|0	100|0	70|26	80|17	62|34
t-t + m_2	-	-	-	-	-	100|0	7|92	10|89	6|93
t-t + m_3	-	-	-	-	-	-	0|100	0|100	0|100
skewt-t + m_1	-	-	-	-	-	-	-	58|39	42|54
skewt-t + m_2	-	-	-	-	-	-	-	-	27|71
					$N = 10$				
Avg.cov	0.0382	0.0446	0.0608	0.0483	0.0558	0.0712	0.0493	0.0519	0.0495
%Bias	13.3401	5.5116	-9.5888	1.6389	-5.1437	-17.2859	0.7437	-1.4492	0.5733
%RMSE	14.9557	8.6956	10.5586	6.2388	7.8202	17.6259	6.0060	6.9778	5.4697
					Diebold-Mariano test				
n-n + m_1	-	79|0	99|0	85|4	94|0	100|0	87|5	92|0	92|0
n-n + m_2	-	-	100|0	70|24	99|0	100|0	77|18	84|11	84|11
n-n + m_3	-	-	-	0|100	3|97	100|0	2|98	3|97	0|100
t-t + m_1	-	-	-	-	99|0	100|0	78|20	75|23	56|42
t-t + m_2	-	-	-	-	-	100|0	6|94	9|91	1|99
t-t + m_3	-	-	-	-	-	-	0|100	0|100	0|100
skewt-t + m_1	-	-	-	-	-	-	-	55|43	44|54
skewt-t + m_2	-	-	-	-	-	-	-	-	27|71

Note: The involved sets of moments are $m_1 = \{\lambda_{0.01}, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.15}\}$, $m_2 = \{\rho_S, \lambda_{0.01}, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.15}\}$, $m_3 = \{\rho_S, \lambda_{0.01}, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.15}, \lambda_{0.85}, \lambda_{0.9}, \lambda_{0.95}, \lambda_{0.99}\}$, the replication number is $R = 100$, the first cell of each entry in the matrix of the result of Diebold-Mariano test stands for the number of times that the model at rows outperforms the model at column out of all replications, the second cell of each entry indicates the number of cases that the column model is preferred and the test is indifferent for remaining cases.

Table 4: $VaR_{0.01}$ forecasts

	n-n + m_1	**n-n** + m_2	**n-n** + m_3	**t-t** + m_1	**t-t** + m_2	**t-t** + m_3	**skewt-t** + m_1	**skewt-t** + m_2	**skewt-t** + m_3
					$N = 3$				
Avg. cov	0.0105	0.0130	0.0179	0.0102	0.0110	0.0165	0.0104	0.0103	0.0097
%Bias	-1.2465	-5.1481	-12.4378	-0.8554	-2.0435	-10.5572	-1.0263	-0.7265	-0.0532
%RMSE	3.7555	6.3516	12.6808	3.4756	3.9229	10.9004	3.9824	3.9622	3.3862
					Diebold-Mariano test				
n-n + m_1	-	35|0	55|0	13|17	25|7	52|0	17|13	11|19	4|23
n-n + m_2	-	-	61|0	0|36	1|36	55|0	0|36	0|36	0|36
n-n + m_3	-	-	-	0|55	0|55	9|60	0|54	0|54	0|53
t-t + m_1	-	-	-	-	30|1	51|0	22|8	14|15	6|20
t-t + m_2	-	-	-	-	-	53|0	4|28	4|26	1|27
t-t + m_3	-	-	-	-	-	-	0|50	0|50	0|50
skewt-t + m_1	-	-	-	-	-	-	-	9|20	1|26
skewt-t + m_2	-	-	-	-	-	-	-	-	4|23
					$N = 5$				
Avg.cov	0.0110	0.0139	0.0210	0.0102	0.0105	0.0178	0.0102	0.0100	0.0100
%Bias	-2.3587	-7.3387	-16.7444	-0.8830	-1.6957	-12.8966	-0.8678	-0.4398	-0.1812
%RMSE	4.8572	8.5836	16.9907	4.4664	4.4169	13.4249	4.7059	4.5079	4.4170
					Diebold-Mariano test				
n-n + m_1	-	46|0	63|0	5|28	12|21	52|0	8|25	4|29	2|28
n-n + m_2	-	-	69|0	0|43	0|42	60|1	0|43	0|43	0|42
n-n + m_3	-	-	-	0|60	0|62	4|76	0|61	0|60	0|60
t-t + m_1	-	-	-	-	23|8	52|0	23|8	6|24	4|27
t-t + m_2	-	-	-	-	-	52|0	9|22	4|27	3|27
t-t + m_3	-	-	-	-	-	-	0|52	0|51	0|51
skewt-t + m_1	-	-	-	-	-	-	-	5|26	5|26
skewt-t + m_2	-	-	-	-	-	-	-	-	14|16
					$N = 10$				
Avg.cov	0.0115	0.0149	0.0235	0.0105	0.0108	0.0192	0.0105	0.0102	0.0103
%Bias	-3.0669	-8.9361	-20.3080	-0.9601	-1.5368	-15.1078	-0.8501	-0.3468	-0.4547
%RMSE	5.9393	10.2721	20.5918	5.3620	5.1335	15.7843	5.6921	5.3703	5.4827
					Diebold-Mariano test				
n-n + m_1	-	46|0	70|0	3|35	6|32	57|0	3|34	1|36	1|36
n-n + m_2	-	-	75|0	0|45	0|45	63|1	0|45	0|45	0|45
n-n + m_3	-	-	-	0|67	0|67	2|81	0|67	0|67	0|66
t-t + m_1	-	-	-	-	19|16	54|0	24|10	6|26	8|25
t-t + m_2	-	-	-	-	-	56|0	23|12	8|26	13|22
t-t + m_3	-	-	-	-	-	-	0|54	0|53	0|54
skewt-t + m_1	-	-	-	-	-	-	-	7|27	10|25
skewt-t + m_2	-	-	-	-	-	-	-	-	23|10

Note: The involved sets of moments are $m_1 = \{\lambda_{0.01}, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.15}\}$, $m_2 = \{\rho_S, \lambda_{0.01}, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.15}\}$, $m_3 = \{\rho_S, \lambda_{0.01}, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.15}, \lambda_{0.85}, \lambda_{0.9}, \lambda_{0.95}, \lambda_{0.99}\}$, the replication number is $R = 100$, the first cell of each entry in the matrix of the result of Diebold-Mariano test stands for the number of times that the model at rows outperforms the model at column out of all replications, the second cell of each entry indicates the number of cases that the column model is preferred and the test is indifferent for remaining cases.

Table 5: Summary Statistics

	SAP	SIE	BAYN	ALV	BAS	DAI	DTE	BMW	DPW	CON
Panel A: Descriptive Statistics										
Mean	4.7309e-04	2.6166e-04	2.5217e-04	3.0627e-04	2.7687e-04	3.0000e-04	2.5559e-04	5.1796e-04	3.9404e-04	10.1341e-04
Std dev	0.0136	0.0143	0.0164	0.0161	0.0160	0.0187	0.0146	0.0187	0.0150	0.0218
Skewness	-0.4901	-0.2449	-0.1622	-0.1167	-0.2371	-0.2225	-0.2873	-0.2739	-0.0851	-0.0470
Kurtosis	5.6060	4.9473	4.5642	7.0550	4.3987	4.5817	7.6958	5.5914	4.4068	5.5301
Panel B: Parameter estimates in margins										
ϕ_0	5.0653e-04	2.5681e-04	2.6451e-04	2.8999e-04	2.6364e-04	2.8059e-04	2.4888e-04	5.0497e-04	3.7458e-04	9.8442e-04
ϕ_1	-0.0276	0.0098	-0.0264	0.0601	0.0332	0.0704	-0.0062	0.0493	0.0284	0.0396
ω	7.5460e-06	5.6271e-06	4.7663e-06	8.6512e-06	6.4722e-06	8.4571e-06	7.2499e-06	5.1333e-06	8.6553e-06	7.3451e-06
α	8.7998e-03	9.1089e-03	1.6508e-07	6.0675e-03	1.1343e-02	2.5459e-02	5.3061e-02	1.3367e-02	2.6959e-07	1.3203e-03
γ	0.0697	0.0781	0.0704	0.1690	0.0763	0.0692	0.0161	0.0623	0.0937	0.0889
β	0.9124	0.9236	0.9478	0.8734	0.9238	0.9157	0.9067	0.9407	0.9140	0.9357
ν	1.3524	1.4383	1.4576	1.3246	1.5649	1.5605	1.1715	1.5133	1.4938	1.647

see Glosten et al. (1993):

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

where $r_{i,t}$ denotes the log return of stock i from the ten stocks. Unlikely in Oh & Patton (2013) and Oh & Patton (2017), the market returns are not considered in modeling time-varying conditional means and variances at the moment. The parameters governing dynamics 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 residuals $\hat{\eta}_{it}$ to uniform series are accomplished by the empirical distribution function (EDF).

Table 17 presents sample dependence measures between the ten stocks. The upper panel gives the sample Spearman's rho of the uniform residuals, which ranges from 0.405 to 0.796. The lower panel reveals quantile dependence measures at the tails. The elements in the upper triangular matrix are computed by the average of 1% and 99% quantile dependence measures $\lambda_{0.01}$ and $\lambda_{0.99}$. On one hand, we have a relatively low tail dependence, for example, zero dependence between *Deutsche Telekom* and *Continental*. On the other hand, substantial tail dependences are observed between *Bayer* and *BASF*, or between *BMW* and *Daimler*. This could be explained by that the stocks which belong to the same industry sector would behave similarly in crashes or booms. The difference between 90% and 10% quantile dependence presented in the

Table 6: Estimation and model selection of factor copula models

Moments	Models	$\hat{\beta}$	$\hat{\nu}^{-1}$	$\hat{\lambda}$	p-val of J test	SMM-AIC	SMM-BIC	SMM-HQIC
$\{\lambda_{0.01}, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.15}\}$	**Factor n-n**	1.4896	-	-	0.9090	-5.6327	-21.8199	-11.6981
	Factor t-t	1.4986	0.0140	-	0.7180	-3.8266	-14.6180	-7.8702
	Factor skewt-t	2.1221	0.0988	0.4043	0.0050	-1.1892	-6.5849	-3.2110
	Factor skewn-n	1.6774	-	0.2234	0.3720	-3.2375	-14.0290	-7.2812
	Factor skewt-n	1.4512	0.0105	-0.0504	0.2120	-1.8911	-7.2868	-3.9129
	Factor t-n	1.5034	0.0101	-	0.7850	-3.8841	-14.6755	-7.9277
$\{\lambda_{0.01}, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.15}, \rho_S\}$	**Factor n-n**	1.3978	-	-	0.1920	2.5189	-19.0639	-5.5683
	Factor t-t	1.3375	0.0989	-	0.0610	0.4177	-15.7695	-5.6478
	Factor skewt-t	1.4223	0.2860	0.2009	0.0000	13.6265	2.8350	9.5828
	Factor skewn-n	1.2327	-	-0.6118	0.9940	-5.9566	-22.1437	-12.0220
	Factor skewt-n	1.2386	0.0101	-0.3262	0.3550	-3.2907	-14.0821	-7.3343
	Factor t-n	1.3669	0.0578	-	0.0540	0.9762	-15.2109	-5.0892
$\{\lambda_{0.01}, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.15}$ $\lambda_{0.85}, \lambda_{0.9}, \lambda_{0.95}, \lambda_{0.99}, \rho_S\}$	**Factor n-n**	1.2741	-	-	0.0580	21.6123	-21.5534	5.4378
	Factor t-t	1.2774	0.0101	-	0.0500	23.7195	-14.0505	9.5669
	Factor skewt-t	1.2997	0.0120	-0.2119	0.6010	-9.6025	-41.9769	-21.7334
	Factor skewn-n	1.3039	-	-0.2668	0.8940	-11.1498	-48.9198	-25.3025
	Factor skewt-n	1.2984	0.0120	-0.2097	0.5370	-9.6668	-42.0412	-21.7977
	Factor t-n	1.2771	0.0100	-	0.0430	23.9671	-13.8030	9.8144
$\{\lambda_{0.05}, \lambda_{0.1}, \lambda_{0.9}, \lambda_{0.95}, \rho_S\}$	**Factor n-n**	1.2673	-	-	0.0160	12.8856	-8.6973	4.7983
	Factor t-t	1.2497	0.0853	-	0.0200	12.7921	-3.3951	6.7267
	Factor skewt-t	1.2527	0.0864	-0.1930	0.3410	-3.4969	-14.2883	-7.5405
	Factor skewn-n	1.3075	-	-0.3014	0.1280	-3.5204	-19.7075	-9.5858
	Factor skewt-n	1.2681	0.0620	-0.2002	0.3430	-3.4318	-14.2233	-7.4755
	Factor t-n	1.2537	0.0704	-	0.0190	13.0199	-3.1672	6.9545

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 contains the parameter estimator from six factor copula models given each choice out of four sets of moment conditions. The selection decisions are presented at the last three columns. 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 at both tails. The *t-n* and *t-t* factor copulas imply the symmetric tail dependence whereas the *skewt-n* and *skewt-t* factor copulas allow tail dependence in both directions and an asymmetric dependence structure. In the SMM estimation, four possibilities of dependence measures are taken into consideration: $m_1 = \{\lambda_{0.01}, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.15}\}$, $m_2 = \{\lambda_{0.01}, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.15}, \rho_S\}$, $m_3 =$

$\{\lambda_{0.01}, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.15}, \lambda_{0.85}, \lambda_{0.9}, \lambda_{0.95}, \lambda_{0.99}, \rho_S\}$ and $m_4 = \{\lambda_{0.05}, \lambda_{0.1}, \lambda_{0.9}, \lambda_{0.95}, \rho_S\}$. The first two merely focus on the lower tail region, but the former excludes Spearman's rho ρ_S and the later includes it. The last two choices consider both tails but differ in the number of moment conditions. The identity matrix is employed as weighting matrix. The p-values in the J test are additionally provided. They are obtained via $B = 1000$ bootstrap replications in the approximation of the limit distribution. The step size $\epsilon_{T,S}$ utilized in the computation of the numerical derivative of $\boldsymbol{g}^c_{T,S}(\boldsymbol{\theta}^b)$ at $\hat{\boldsymbol{\theta}}_{T,S}$ is 0.005.

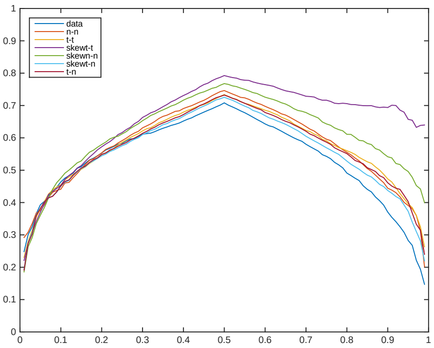
Please note that we only consider (nested) equi-dependence models, which allow for one factor loading, one skewness parameter and one tail parameter as possible unknowns, since this is a less restrictive model. The over-identifying test proposed by Oh & Patton (2013) shows that the *skewt-t* factor copula is strongly rejected given sets of lower-tail-moments m_1 and m_2, but not rejected given sets of both-tail-moments m_3 and m_4 at significance level 5%, whereas the *n-n* factor copula is not rejected when m_1 and m_2 are used, but strongly rejected in case of the employment of m_4. 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 comes into play. SMM-AIC, SMM-BIC and SMM-HQIC agree on *skewn-n* factor copula model combined with the set m_3 as the best choice out of all possible combinations, which is followed by *skewt-n* factor copula and *skewt-t* factor copula with the set m_3. Not surprisingly, the procedure prefers the model candidate with possibly larger number of over-identifying conditions. The best model implies an asymmetric dependence but zero tail dependence in the empirical data. The estimators of the tailedness parameter $\hat{\nu}^{-1}$ from the second best model and the third best model are close to zero, which further validates that the tail dependence is ignorable. The models with the excessive usage of the tailedness parameter are less favorable.

If we take a closer look at the selection of models in each specific category of moments, the selection procedures give us different decisions from the previous overall comparison, although three criteria still have the agreement in each category. Given the first choice of the set of moments m_1, the *n-n* factor copula is ranked in the first place and even the *t-n* and *t-t* factor copula models beat *skewn-n* factor copula. Since the moment set m_1 is only concerned about the lower part of tail quantiles and excludes

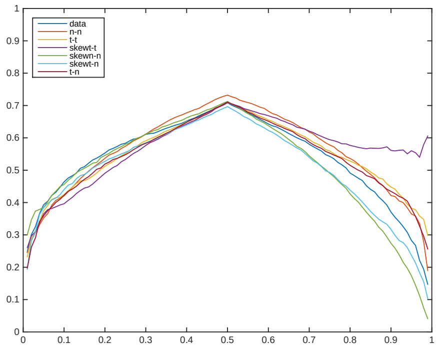
Spearman's rho, the ability of capturing asymmetry becomes in vain in this scenario. The *n-n* model has been able to provide considerable goodness-of-fit in this specific lower tail region and the selection procedure gives a higher reward to *n-n* model for its less usage of parameters. When Spearman's rho is put back to the basket of moments, i.e., the set of moments m_2, the *skewn-n* factor copula gets the position of the best model, whereas the *n-n* factor copula is suggested as the second best in this category. The ranking of the *n-n* factor copula changes when upper tail quantiles are added into the set of moments, i.e., m_3 and m_4. The models from the family of skewed factor copulas dominate, because quantiles at both tails are considered at the moment and the skewness parameter plays a crucial role. Figure 2 plots the sample and fitted quantile dependence measures to visually compare the performance of all model candidates and shows the effect of the choice of moment conditions on the model performance. In Figure 2(a), all models merely fit well in the lower tail region. The fit at the middle quantiles are improved when ρ_S is added as shown in Figure 2(b), but the fit at the upper tail region is still problematic. When quantiles at both tails enter in the set of moment conditions, all models fit better at upper and lower tail areas. However, *n-n*, *t-n* and *t-t* factor copulas which only offer symmetric dependence structure fit relatively worse than factor copulas from the skewed family, since the empirical data series possess the asymmetric characteristics. The graphical information stays consistent with our finding that the selection procedure picked *skewn-n* factor copula as the best model.

To wrap up the empirical section, (one-day ahead) Value-at-Risk forecasting and backtesting is performed on the portfolio of ten stocks. Each stock is assigned with an equal weight in such a portfolio, which can, of course, be relaxed. From a practical view of point in risk management, it is crucial to worry about the VaR forecasting ability of model candidates and the forecasting behavior when one employs insufficient moment conditions or only moments at the lower tail. In order to evaluate the accuracy of VaR forecasts at 0.05 and 0.01 level, we backtest six factor copula models combined with four sets of moment conditions as follows. The whole sample period is initially split into the estimation period $T_s = 1000$ and the testing period $T_e = 629$. At each time t in the testing period, the past T_s data points, i.e., sample from $t - T_s + 1$ to $t - 1$, are utilized in the estimation of copula models as well as dynamic models in margins. To

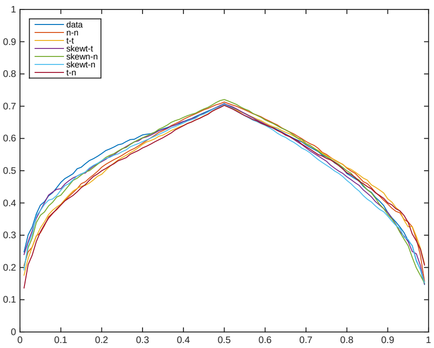
Figure 2: Sample and fitted quantile dependences



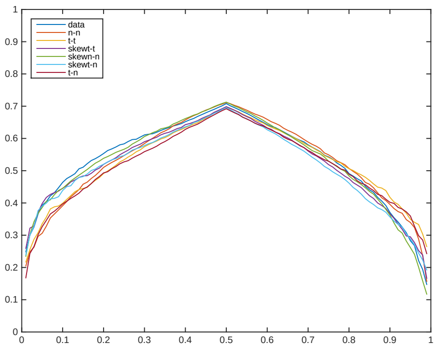
(a) $m_1 = \{\lambda_{0.01}, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.15}\}$



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



(c) $m_3 = m_2 \cup \{\lambda_{0.85}, \lambda_{0.9}, \lambda_{0.95}, \lambda_{0.99}\}$



(d) $m_4 = \{\lambda_{0.05}, \lambda_{0.1}, \lambda_{0.9}, \lambda_{0.95}, \rho_S\}$

approximate the distribution of portfolio return at each time t, we simulate $B = 100000$ standardized residuals $\hat{\boldsymbol{\eta}}_t$ which are determined by the factor copula model plus the AR(1)-GJR-GARCH(1,1,1) model. Then we use $\boldsymbol{r}_t^s = \boldsymbol{\mu}(\hat{\boldsymbol{\phi}}) + \boldsymbol{\sigma}(\hat{\boldsymbol{\phi}})\hat{\boldsymbol{\eta}}_t$ to generate the simulated distribution of the return at time t. Note that we estimate factor copula models only once at every 20 observations to avoid the cumbersome computation and we do not expect a frequently large variation in the dependence structure in this small period. The one-step-ahead forecast of $\hat{\sigma}_{i,t}$ is obtained via the plug-in method based on the information at time $t-1$ and the vector $\hat{\boldsymbol{\phi}}$. The 5% and 1% VaR forecasts at time t are computed based on the simulated distribution of returns. The number of violations and the corresponding coverage rate are obtained by comparing VaR forecasts and actual returns.

To evaluate the forecasting performance of various models, three tests are performed to test on the coverage level and the distribution of violations: the unconditional coverage test by Kupiec (1995), the conditional coverage test by Christoffersen (1998) and the CAViaR test by Engle & Manganelli (2004). The results are available in Table 18. The first column shows that coverage rates of $VaR_{0.05}$ forecasts are generally slightly larger than 0.05. The *n-n*, *t-t* and *t-n* factor copula models tend to have difficulty passing the unconditional coverage test, especially when the more complete set of moment conditions is used. For instance, the *n-n* factor copula with the sets m_3 or m_4. Skewed factor copula models outperform the former ones. The *skewn-n* copula achieves the best performance: it passes the unconditional coverage test regardless of the choice of moment conditions. The findings of the conditional coverage test mostly stay in line with the unconditional coverage test. The CAViaR test provides further information on the distribution of VaR violations. It shows that the violations are independent to each other in VaR forecasts from all skewed factor copula models except for the *skewt-t* factor copula with m_2. The series of violations does not follow the martingale difference process when VaR forecasts are made by heavy-tailed models with a larger set of moments, such as *t-t* or *t-n* factor model with m_3 and m_4. The models with sets of moments in the consideration work fairly well for the $VaR_{0.01}$ forecasting, which is confirmed by all three tests and coverage rates. This could be explained by the fact that the testing window selected in our analysis is rather short and 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 VaR models turns significant. Lastly, the Diebold-Mariano test is used to relatively compare VaR forecasts between a pair of models. As indicated in Table 19, for the $VaR_{0.05}$ forecasting, the less parameterized model generally outperforms other candidates when a smaller set of moments is used in the estimation. For instance, the *n-n* factor copula model beats all models except the *skewn-n* factor copula when quantiles only at the lower tail, for example in m_1, are employed. However, when the set of moments contains a higher number of moments, the skewed factor copula becomes preferable. The leading example is that the *skewn-n* factor copula model is considered as a better model when quantile dependence measures at both tails lie in the set of moments, i.e., m_3 and m_4. This model also stands out in the skewed family, since it has the least number of parameters. The Diebold-Mariano test is generally indifferent in comparison of the relative accuracy of $VaR_{0.01}$ forecasts.

7 Conclusion

Since the factor copula model does 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) to 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 decisions of models. If one is merely concerned with VaR forecasts, a less parameterized model with quantile dependence measures at the lower tail may achieve a good forecasting performance. In the end, we perform the selection procedure for the dependence modeling based on a real data set. The empirical findings corroborate the performance of the selection procedure such that it chooses the model candidates with the largest number of over-identifying restrictions. It also confirms the forecasting behavior of misspecified models with a specific choice of moment conditions as in the simulations. For now, we only consider a rather simpler model, i.e., the equi-dependence factor copula model. In the future research, it would be valuable to examine the performance of proposed selection methods on other models such as blockwise dependence models.

A Proofs

A.1 Proof of Proposition 1

Proof Assumption 1 & 2(iv) implies

$$\min(T,S)^{-1}J^c_{T,S}(\hat{\boldsymbol{\theta}}^b_{T,S}) := \boldsymbol{g}^c_{T,S}(\hat{\boldsymbol{\theta}}^b_{T,S})'\hat{\boldsymbol{W}}^c_T\boldsymbol{g}^c_{T,S}(\hat{\boldsymbol{\theta}}^b_{T,S}) \rightarrow_p \boldsymbol{g}^c_0(\hat{\boldsymbol{\theta}}^b_{T,S})'\boldsymbol{W}^c_0\boldsymbol{g}^c_0(\hat{\boldsymbol{\theta}}^b_{T,S})$$

where $\boldsymbol{g}^c_{T,S}(\hat{\boldsymbol{\theta}}^b_{T,S})$ is the moment conditions selected by moment selection vector $\boldsymbol{c}$ evaluated at $\hat{\boldsymbol{\theta}}^b_{T,S}$, which is the SMM estimator obtained from the model selected by model selection vector $\boldsymbol{b}$. For any $(\boldsymbol{b},\boldsymbol{c}) \in \mathcal{MP}_1$, when Assumptions 1 and 2 hold, the consistency of $\hat{\boldsymbol{\theta}}^b_{T,S}$ implies $\boldsymbol{g}^c_0(\hat{\boldsymbol{\theta}}^b_{T,S}) = \boldsymbol{m}^c_0(\boldsymbol{\theta}_0) - \boldsymbol{m}^c_0(\hat{\boldsymbol{\theta}}^b_{T,S}) = 0$, as $T,S \rightarrow \infty$. Therefore, the first part in MSC criterion vanishes asymptotically. Together with Assumption 5(i): $\kappa_{T,S} = o(\min(T,S))$, it gives us the asymptotics of MSC under correct model specification as

$$\min(T,S)^{-1}MSC_{T,S}(\hat{\boldsymbol{\theta}}^b_{T,S},\boldsymbol{c}) = \boldsymbol{g}^c_{T,S}(\hat{\boldsymbol{\theta}}^b_{T,S})'\hat{\boldsymbol{W}}^c_T\boldsymbol{g}^c_{T,S}(\hat{\boldsymbol{\theta}}^b_{T,S}) - \min(T,S)^{-1}(|\boldsymbol{c}|-|\boldsymbol{b}|)\kappa_{T,S} \rightarrow_p 0$$

For any $(\boldsymbol{b},\boldsymbol{c}) \in \mathcal{MP}\backslash\mathcal{MP}_1$, following Assumption 3(i), there exists no such $\hat{\boldsymbol{\theta}}^b_{T,S} \in \boldsymbol{\Theta}$ fulfilling $\boldsymbol{g}^c_0(\hat{\boldsymbol{\theta}}^b_{T,S}) = 0$, as $T,S \rightarrow \infty$. When Assumption 1 2 and 3 hold, one has consistency of $\hat{\boldsymbol{\theta}}^b_{T,S}$ under misspecification, there exists pseudo-true SMM estimator $\boldsymbol{\theta}^b_*(\boldsymbol{W}^c_0)$ for model selected by vector $\boldsymbol{b}$ along with the moments selected by vector $\boldsymbol{c}$ such that $\boldsymbol{g}^c_0(\boldsymbol{\theta}^b_*(\boldsymbol{W}^c_0)) > 0$ in asymptotic. Then the quadratic form does not vanish as $T,S \rightarrow \infty$, i.e., $\boldsymbol{g}^c_0(\boldsymbol{\theta}^b_*(\boldsymbol{W}^c_0))\boldsymbol{W}^c_0\boldsymbol{g}^c_0(\boldsymbol{\theta}^b_*(\boldsymbol{W}^c_0)) > 0$ for some $\boldsymbol{\theta}^b_*(\boldsymbol{W}^c_0) \in \boldsymbol{\Theta}$. Hence the asymptotics of MSC criterion for the misspecified models:

$$\min(T,S)^{-1}\mathrm{MSC}_{T,S}(\hat{\boldsymbol{\theta}}^b_{T,S},\boldsymbol{c}) \rightarrow_p \boldsymbol{g}^c_0(\boldsymbol{\theta}^b_*(\boldsymbol{W}^c_0))\boldsymbol{W}^c_0\boldsymbol{g}^c_0(\boldsymbol{\theta}^b_*(\boldsymbol{W}^c_0)) > 0, \text{ as } T,S \rightarrow \infty$$

This implies $(\hat{\boldsymbol{b}}_{MSC},\hat{\boldsymbol{c}}_{MSC}) \in \mathcal{MP}_1, wp \rightarrow 1$. 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 $|\boldsymbol{c}_i| - |\boldsymbol{b}_i| > |\boldsymbol{c}_j| - |\boldsymbol{b}_j|$ by definition of set $\mathcal{MP}_2$, with Assumption 5, it implies

$$-(|\boldsymbol{c}_i| - |\boldsymbol{b}_i|)\kappa_{T,S} < -(|\boldsymbol{c}_j| - |\boldsymbol{b}_j|)\kappa_{T,S}$$

and then,

$$J^{c_i}_{T,S}(\hat{\boldsymbol{\theta}}^{b_i}_{T,S}) - (|\boldsymbol{c}_i|-|\boldsymbol{b}_i|)\kappa_{T,S} := \mathrm{MSC}_{T,S}(\hat{\boldsymbol{\theta}}^{b_i}_{T,S},\boldsymbol{c}_i) < \mathrm{MSC}_{T,S}(\hat{\boldsymbol{\theta}}^{b_j}_{T,S}) := J^{c_j}_{T,S}(\hat{\boldsymbol{\theta}}^{b_j}_{T,S}) - (|\boldsymbol{c}_j|-|\boldsymbol{b}_j|)\kappa_{T,S}$$

Therefore $(\hat{\boldsymbol{b}}_{MSC}, \hat{\boldsymbol{c}}_{MSC}) \in \mathcal{MP}_2, wp \to 1$. Assumption 4 ensures that $\mathcal{MP}_2 = \{(\boldsymbol{b}^0, \boldsymbol{c}^0)\}$, then this directly leads to $(\hat{\boldsymbol{b}}_{MSC}, \hat{\boldsymbol{c}}_{MSC}) = (\boldsymbol{b}^0, \boldsymbol{c}^0), wp \to 1$. ■

B Tables

Table 7: Empirical selection frequencies based on the selection procedure SMM-AIC, the Equi-dependence model with the marginal dependence, $\beta_0 = 1, \nu_0^{-1} = 0.25, \lambda_0 = -0.5$

	$\hat{d}$ / d_0	m_1				m_2				m_3			
		n-n	t-t	skewn-n	skewt-t	n-n	t-t	skewn-n	skewt-t	n-n	t-t	skewn-n	skewt-t
						$N = 3$							
$T = 500$	**n-n**	0.000	0.000	0.000	0.000	0.015	0.000	0.000	0.000	0.615	0.070	0.265	0.035
	t-t	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.225	0.435	0.130	0.200
	skewn-n	0.000	0.000	0.000	0.000	0.020	0.000	0.000	0.000	0.075	0.005	0.790	0.110
	skewt-t	0.000	0.000	0.000	0.000	0.015	0.000	0.000	0.005	0.005	0.000	0.275	0.700
$T = 1000$	**n-n**	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.625	0.080	0.270	0.015
	t-t	0.000	0.000	0.000	0.000	0.015	0.000	0.000	0.000	0.090	0.570	0.040	0.285
	skewn-n	0.000	0.000	0.000	0.000	0.020	0.000	0.000	0.000	0.025	0.000	0.895	0.060
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.105	0.890
$T = 2000$	**n-n**	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.530	0.055	0.395	0.010
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.615	0.020	0.360
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.935	0.055
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.015	0.000	0.000	0.005	0.970
						$N = 5$							
$T = 500$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.745	0.020	0.230	0.005
	t-t	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.265	0.430	0.100	0.200
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.035	0.000	0.920	0.045
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.245	0.750
$T = 1000$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.760	0.025	0.205	0.010
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.055	0.600	0.040	0.305
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.970	0.025
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.060	0.940
$T = 2000$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.790	0.015	0.195	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.675	0.000	0.325
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.960	0.040
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.000	0.000	0.000	0.985
						$N = 10$							
$T = 500$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.875	0.015	0.105	0.005
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.235	0.515	0.085	0.165
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.020	0.000	0.980	0.000
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.195	0.805
$T = 1000$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.895	0.010	0.095	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.035	0.690	0.015	0.260
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.990	0.010
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.025	0.975
$T = 2000$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.910	0.005	0.085	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.680	0.000	0.315
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.995	0.005
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.995

Note: The sets of moments are $m_1 = \{\rho_S, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.9}, \lambda_{0.95}\}$, $m_2 = \{\rho_S, \lambda_{0.25}, \lambda_{0.45}, \lambda_{0.55}, \lambda_{0.75}\}$, $m_3 = \{\rho_S, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.25}, \lambda_{0.45}, \lambda_{0.55}, \lambda_{0.75}, \lambda_{0.9}, \lambda_{0.95}\}$, the number of replications is $R = 200$, $\hat{d}$ denotes the models in the estimation and d_0 denotes the true model.

Table 8: Empirical selection frequencies based on the selection procedure SMM-BIC, the Equi-dependence model with the marginal dependence, $\beta_0 = 1, \nu_0^{-1} = 0.25, \lambda_0 = -0.5$

	$\hat{d}$ \ d_0	m_1				m_2				m_3			
		n-n	t-t	skewn-n	skewt-t	n-n	t-t	skewn-n	skewt-t	n-n	t-t	skewn-n	skewt-t
						$N = 3$							
$T = 500$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.940	0.000	0.060	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.630	0.265	0.070	0.035
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.380	0.000	0.600	0.020
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.555	0.435
$T = 1000$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.935	0.010	0.055	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.375	0.505	0.065	0.055
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.115	0.000	0.875	0.010
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.325	0.675
$T = 2000$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.945	0.005	0.050	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.140	0.770	0.020	0.070
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.995	0.005
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.040	0.960
						$N = 5$							
$T = 500$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.970	0.000	0.030	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.690	0.215	0.060	0.035
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.280	0.000	0.715	0.005
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.525	0.470
$T = 1000$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.980	0.000	0.020	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.435	0.480	0.035	0.050
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.060	0.000	0.940	0.000
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.270	0.730
$T = 2000$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.995	0.000	0.005	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.145	0.805	0.010	0.040
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.970
						$N = 10$							
$T = 500$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.995	0.000	0.005	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.755	0.160	0.055	0.030
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.220	0.000	0.780	0.000
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.525	0.470
$T = 1000$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.995	0.000	0.005	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.480	0.460	0.015	0.045
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.020	0.000	0.980	0.000
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.205	0.795
$T = 2000$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.115	0.850	0.000	0.035
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000

NOTE: The sets of moments are $m_1 = \{\rho_S, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.9}, \lambda_{0.95}\}$, $m_2 = \{\rho_S, \lambda_{0.25}, \lambda_{0.45}, \lambda_{0.55}, \lambda_{0.75}\}$, $m_3 = \{\rho_S, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.25}, \lambda_{0.45}, \lambda_{0.55}, \lambda_{0.75}, \lambda_{0.9}, \lambda_{0.95}\}$, the number of replications $R = 200$, $\hat{d}$ denotes the models in the estimation, d_0 denotes the true model.

Table 9: Empirical selection frequencies based on the selection procedure SMM-HQIC, the Equi-dependence model with the marginal dependence, $\beta_0 = 1, \nu_0^{-1} = 0.25, \lambda_0 = -0.5$

	d_0 \ $\hat{d}$	m_1 n-n	m_1 t-t	m_1 skewn-n	m_1 skewt-t	m_2 n-n	m_2 t-t	m_2 skewn-n	m_2 skewt-t	m_3 n-n	m_3 t-t	m_3 skewn-n	m_3 skewt-t
						$N = 3$							
$T = 500$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.790	0.015	0.190	0.005
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.410	0.415	0.105	0.070
	skewn-n	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.210	0.005	0.730	0.050
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.365	0.630
$T = 1000$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.805	0.015	0.180	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.210	0.565	0.070	0.155
	skewn-n	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.045	0.000	0.915	0.030
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.180	0.820
$T = 2000$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.780	0.025	0.195	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.755	0.025	0.190
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.970	0.025
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.990
						$N = 5$							
$T = 500$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.890	0.005	0.105	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.480	0.355	0.060	0.105
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	0.885	0.015
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.350	0.645
$T = 1000$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.910	0.005	0.080	0.005
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.190	0.625	0.035	0.150
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.000	0.975	0.010
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.115	0.885
$T = 2000$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.935	0.005	0.060	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.035	0.815	0.005	0.145
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
						$N = 10$							
$T = 500$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.960	0.000	0.040	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.445	0.385	0.095	0.075
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.045	0.000	0.955	0.000
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.295	0.700
$T = 1000$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.980	0.000	0.020	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.165	0.670	0.020	0.145
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.075	0.925
$T = 2000$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.980	0.000	0.020	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.025	0.850	0.000	0.125
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000

Note: The sets of moments are $m_1 = \{\rho_S, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.9}, \lambda_{0.95}\}$, $m_2 = \{\rho_S, \lambda_{0.25}, \lambda_{0.45}, \lambda_{0.55}, \lambda_{0.75}\}$, $m_3 = \{\rho_S, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.25}, \lambda_{0.45}, \lambda_{0.55}, \lambda_{0.75}, \lambda_{0.9}, \lambda_{0.95}\}$, the number of replications is $R = 200$, $\hat{d}$ denotes the models in the estimation and d_0 denotes the true model.

Table 10: Empirical selection frequencies based on the selection procedure SMM-AIC, the Equi-dependence model with the marginal dependence, $\beta_0 = 1$

		m_1				m_2				m_3			
	d_0 \ $\hat{d}$	n-n	t-t	skewn-n	skewt-t	n-n	t-t	skewn-n	skewt-t	n-n	t-t	skewn-n	skewt-t
							$N = 3$						
$T = 500$	**n-n**	0.000	0.000	0.000	0.000	0.035	0.010	0.010	0.000	0.620	0.195	0.125	0.005
	t-t	0.000	0.000	0.000	0.000	0.030	0.065	0.000	0.000	0.200	0.625	0.075	0.005
	skewn-n	0.000	0.000	0.000	0.000	0.005	0.000	0.020	0.005	0.065	0.030	0.830	0.045
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.020	0.055	0.000	0.005	0.120	0.800
$T = 1000$	**n-n**	0.000	0.000	0.000	0.000	0.065	0.015	0.000	0.000	0.775	0.100	0.045	0.000
	t-t	0.000	0.000	0.000	0.000	0.035	0.110	0.010	0.000	0.100	0.735	0.010	0.000
	skewn-n	0.000	0.000	0.000	0.000	0.005	0.000	0.025	0.000	0.040	0.000	0.910	0.020
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.090	0.000	0.000	0.030	0.875
$T = 2000$	**n-n**	0.000	0.000	0.000	0.000	0.055	0.010	0.005	0.000	0.890	0.025	0.015	0.000
	t-t	0.000	0.000	0.000	0.000	0.025	0.135	0.010	0.000	0.025	0.805	0.000	0.000
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.090	0.005	0.000	0.000	0.905	0.000
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.055	0.000	0.000	0.000	0.930
							$N = 5$						
$T = 500$	**n-n**	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.810	0.095	0.090	0.000
	t-t	0.000	0.000	0.000	0.000	0.025	0.050	0.005	0.000	0.160	0.700	0.055	0.005
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.000	0.940	0.030
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.005	0.000	0.085	0.905
$T = 1000$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.925	0.040	0.030	0.000
	t-t	0.000	0.000	0.000	0.000	0.015	0.075	0.005	0.000	0.080	0.820	0.005	0.000
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.005	0.015	0.000	0.970	0.000
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.000	0.000	0.025	0.945
$T = 2000$	**n-n**	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.990	0.005	0.000	0.000
	t-t	0.000	0.000	0.000	0.000	0.005	0.060	0.005	0.000	0.010	0.920	0.000	0.000
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.990	0.000
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.000	0.000	0.000	0.970
							$N = 10$						
$T = 500$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.900	0.065	0.035	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.065	0.000	0.000	0.140	0.745	0.045	0.005
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.000	0.975	0.010
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.070	0.920
$T = 1000$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.970	0.025	0.005	0.000
	t-t	0.000	0.000	0.000	0.000	0.005	0.055	0.005	0.000	0.040	0.890	0.005	0.000
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.000	0.000	0.020	0.965
$T = 2000$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.045	0.000	0.000	0.015	0.940	0.000	0.000
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.995

Note: The sets of moments are $m_1 = \{\rho_S, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.9}, \lambda_{0.95}\}$, $m_2 = \{\rho_S, \lambda_{0.25}, \lambda_{0.45}, \lambda_{0.55}, \lambda_{0.75}\}$, $m_3 = \{\rho_S, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.25}, \lambda_{0.45}, \lambda_{0.55}, \lambda_{0.75}, \lambda_{0.9}, \lambda_{0.95}\}$, the number of replications is $R = 200$, $\hat{d}$ denotes the models in the estimation and d_0 denotes the true model.

Table 11: Empirical selection frequency based on the selection procedure SMM-BIC, the Equi-dependence model with the marginal dependence, $\beta_0 = 1$

	$\hat{d}$ / d_0	m_1				m_2				m_3			
		n-n	t-t	skewn-n	skewt-t	n-n	t-t	skewn-n	skewt-t	n-n	t-t	skewn-n	skewt-t
						$N = 3$							
$T = 500$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.660	0.205	0.130	0.005
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.225	0.680	0.080	0.015
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.070	0.030	0.850	0.050
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.120	0.875
$T = 1000$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.845	0.105	0.050	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.115	0.865	0.020	0.000
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.040	0.005	0.935	0.020
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.045	0.955
$T = 2000$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.955	0.025	0.020	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.045	0.955	0.000	0.000
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.995	0.005
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
						$N = 5$							
$T = 500$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.815	0.095	0.090	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.170	0.750	0.065	0.010
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.000	0.940	0.030
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.085	0.910
$T = 1000$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.930	0.040	0.030	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.090	0.895	0.010	0.005
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.000	0.980	0.005
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.025	0.975
$T = 2000$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.995	0.005	0.000	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.990	0.000	0.000
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
						$N = 10$							
$T = 500$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.900	0.065	0.035	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.155	0.795	0.045	0.005
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.000	0.975	0.010
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.070	0.930
$T = 1000$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.970	0.025	0.005	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.040	0.950	0.010	0.000
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.020	0.980
$T = 2000$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.985	0.000	0.000
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000

Note: The sets of moments are $m_1 = \{\rho_S, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.9}, \lambda_{0.95}\}$, $m_2 = \{\rho_S, \lambda_{0.25}, \lambda_{0.45}, \lambda_{0.55}, \lambda_{0.75}\}$, $m_3 = \{\rho_S, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.25}, \lambda_{0.45}, \lambda_{0.55}, \lambda_{0.75}, \lambda_{0.9}, \lambda_{0.95}\}$, the number of replications is $R = 200$, $\hat{d}$ denotes the models in the estimation and d_0 denotes the true model.

Table 12: Empirical selection frequencies based on the selection procedure SMM-HQIC, the Equi-dependence model with the marginal dependence, $\beta_0 = 1$

		m_1				m_2				m_3			
	$\hat{d}$ \ d_0	n-n	t-t	skewn-n	skewt-t	n-n	t-t	skewn-n	skewt-t	n-n	t-t	skewn-n	skewt-t
							$N = 3$						
$T = 500$	**n-n**	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.650	0.205	0.130	0.005
	t-t	0.000	0.000	0.000	0.000	0.005	0.015	0.000	0.000	0.225	0.665	0.080	0.010
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.065	0.030	0.850	0.050
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.005	0.120	0.870
$T = 1000$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.845	0.105	0.050	0.000
	t-t	0.000	0.000	0.000	0.000	0.010	0.015	0.000	0.000	0.105	0.850	0.020	0.000
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.040	0.005	0.935	0.020
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.045	0.950
$T = 2000$	**n-n**	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.950	0.025	0.020	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.045	0.950	0.000	0.000
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.990	0.005
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
							$N = 5$						
$T = 500$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.815	0.095	0.090	0.000
	t-t	0.000	0.000	0.000	0.000	0.010	0.010	0.000	0.000	0.165	0.740	0.065	0.010
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.000	0.940	0.030
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.085	0.910
$T = 1000$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.930	0.040	0.030	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.090	0.890	0.010	0.000
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.000	0.980	0.005
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.025	0.975
$T = 2000$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.995	0.005	0.000	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.010	0.980	0.000	0.000
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
							$N = 10$						
$T = 500$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.900	0.065	0.035	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.015	0.000	0.000	0.155	0.780	0.045	0.005
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.000	0.975	0.010
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.070	0.930
$T = 1000$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.970	0.025	0.005	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.040	0.945	0.010	0.000
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.020	0.980
$T = 2000$	**n-n**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	t-t	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.015	0.980	0.000	0.000
	skewn-n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
	skewt-t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000

Note: The sets of moments are $m_1 = \{\rho_S, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.9}, \lambda_{0.95}\}$, $m_2 = \{\rho_S, \lambda_{0.25}, \lambda_{0.45}, \lambda_{0.55}, \lambda_{0.75}\}$, $m_3 = \{\rho_S, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.25}, \lambda_{0.45}, \lambda_{0.55}, \lambda_{0.75}, \lambda_{0.9}, \lambda_{0.95}\}$, the number of replications is $R = 200$, $\hat{d}$ denotes the models in the estimation and d_0 denotes the true model.

Table 13: Empirical selection frequencies based on selection procedures SMM-AIC, SMM-BIC and SMM-HQIC, the Equi-dependence model with the marginal dependence, $\beta_0 = 1, \nu_0^{-1} = 0.25, \lambda_0 = -0.5$

		SMM-AIC				SMM-BIC				SMM-HQIC			
	d_0 \ $\hat{d}$	n-n	t-t	skewn-n	skewt-t	n-n	t-t	skewn-n	skewt-t	n-n	t-t	skewn-n	skewt-t
						$N = 3$							
$T = 500$	**n-n**	0.655	0.045	0.295	0.005	0.930	0.000	0.070	0.000	0.820	0.005	0.175	0.000
	t-t	0.255	0.400	0.150	0.195	0.705	0.190	0.065	0.040	0.470	0.355	0.110	0.065
	skewn-n	0.140	0.005	0.770	0.085	0.415	0.000	0.580	0.005	0.295	0.000	0.695	0.010
	skewt-t	0.005	0.000	0.325	0.670	0.005	0.000	0.615	0.380	0.005	0.000	0.450	0.545
$T = 1000$	**n-n**	0.660	0.060	0.275	0.005	0.940	0.000	0.060	0.000	0.815	0.010	0.175	0.000
	t-t	0.080	0.585	0.055	0.280	0.460	0.420	0.060	0.060	0.255	0.530	0.060	0.155
	skewn-n	0.030	0.000	0.920	0.050	0.140	0.000	0.850	0.010	0.055	0.000	0.920	0.025
	skewt-t	0.000	0.000	0.125	0.875	0.000	0.000	0.410	0.590	0.000	0.000	0.225	0.775
$T = 2000$	**n-n**	0.595	0.035	0.370	0.000	0.955	0.000	0.045	0.000	0.835	0.005	0.160	0.000
	t-t	0.010	0.595	0.015	0.380	0.200	0.720	0.030	0.050	0.035	0.730	0.030	0.205
	skewn-n	0.000	0.000	0.965	0.035	0.010	0.000	0.980	0.010	0.000	0.000	0.990	0.010
	skewt-t	0.000	0.000	0.015	0.985	0.000	0.000	0.075	0.925	0.000	0.000	0.030	0.970
						$N = 5$							
$T = 500$	**n-n**	0.770	0.005	0.225	0.000	0.975	0.000	0.025	0.000	0.895	0.000	0.105	0.000
	t-t	0.285	0.435	0.095	0.185	0.730	0.175	0.070	0.025	0.530	0.290	0.085	0.095
	skewn-n	0.055	0.000	0.920	0.025	0.360	0.000	0.640	0.000	0.145	0.000	0.845	0.010
	skewt-t	0.005	0.000	0.235	0.760	0.005	0.000	0.630	0.365	0.005	0.000	0.375	0.620
$T = 1000$	**n-n**	0.780	0.025	0.190	0.005	0.980	0.000	0.020	0.000	0.920	0.005	0.075	0.000
	t-t	0.070	0.570	0.035	0.325	0.535	0.400	0.035	0.030	0.235	0.580	0.030	0.155
	skewn-n	0.010	0.000	0.980	0.010	0.090	0.000	0.910	0.000	0.015	0.000	0.985	0.000
	skewt-t	0.000	0.000	0.065	0.935	0.000	0.000	0.305	0.695	0.000	0.000	0.110	0.890
$T = 2000$	**n-n**	0.820	0.005	0.175	0.000	0.995	0.000	0.005	0.000	0.945	0.005	0.050	0.000
	t-t	0.000	0.660	0.000	0.340	0.205	0.750	0.010	0.035	0.035	0.805	0.005	0.155
	skewn-n	0.000	0.000	0.980	0.020	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000
	skewt-t	0.000	0.000	0.000	1.000	0.000	0.000	0.025	0.975	0.000	0.000	0.000	1.000
						$N = 10$							
$T = 500$	**n-n**	0.905	0.000	0.090	0.005	1.000	0.000	0.000	0.000	0.970	0.000	0.030	0.000
	t-t	0.255	0.495	0.100	0.150	0.820	0.095	0.060	0.025	0.535	0.300	0.100	0.065
	skewn-n	0.025	0.000	0.970	0.005	0.320	0.000	0.680	0.000	0.085	0.000	0.910	0.005
	skewt-t	0.000	0.000	0.185	0.815	0.005	0.000	0.610	0.385	0.005	0.000	0.380	0.615
$T = 1000$	**n-n**	0.910	0.005	0.085	0.000	0.995	0.000	0.005	0.000	0.980	0.000	0.020	0.000
	t-t	0.045	0.695	0.015	0.245	0.610	0.325	0.030	0.035	0.230	0.605	0.030	0.135
	skewn-n	0.000	0.000	1.000	0.000	0.040	0.000	0.960	0.000	0.000	0.000	1.000	0.000
	skewt-t	0.000	0.000	0.015	0.985	0.000	0.000	0.295	0.705	0.000	0.000	0.085	0.915
$T = 2000$	**n-n**	0.925	0.000	0.075	0.000	1.000	0.000	0.000	0.000	0.980	0.000	0.020	0.000
	t-t	0.000	0.695	0.000	0.305	0.165	0.795	0.000	0.040	0.025	0.855	0.000	0.120
	skewn-n	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000
	skewt-t	0.000	0.000	0.000	1.000	0.000	0.000	0.005	0.995	0.000	0.000	0.000	1.000

Note: The set of moments $m = \{\rho_S, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.9}, \lambda_{0.95}\}$ is relevant, the number of replications is $R = 200$, $\hat{d}$ denotes the models in the estimation and d_0 denotes the true model.

Table 14: Empirical selection frequencies based on selection procedures SMM-AIC, SMM-BIC and SMM-HQIC, the Equi-dependence model with the marginal dependence, $\beta_0 = 1, \nu_0^{-1} = 0.25, \lambda_0 = -0.5$

	d_0 \ $\hat{d}$	SMM-AIC				SMM-BIC				SMM-HQIC			
		n-n	t-t	skewn-n	skewt-t	n-n	t-t	skewn-n	skewt-t	n-n	t-t	skewn-n	skewt-t
								$N = 3$					
$T = 500$	**n-n**	0.995	0.000	0.005	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	t-t	0.990	0.000	0.010	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	skewn-n	0.895	0.000	0.075	0.030	1.000	0.000	0.000	0.000	0.975	0.000	0.025	0.000
	skewt-t	0.640	0.000	0.105	0.255	0.980	0.000	0.020	0.000	0.910	0.000	0.070	0.020
$T = 1000$	**n-n**	0.985	0.000	0.015	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	t-t	0.990	0.000	0.010	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	skewn-n	0.655	0.000	0.260	0.085	0.990	0.000	0.010	0.000	0.915	0.000	0.085	0.000
	skewt-t	0.200	0.000	0.165	0.635	0.910	0.000	0.080	0.010	0.610	0.000	0.150	0.240
$T = 2000$	**n-n**	0.980	0.000	0.020	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	t-t	0.995	0.005	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	skewn-n	0.365	0.000	0.540	0.095	0.965	0.000	0.035	0.000	0.710	0.000	0.285	0.005
	skewt-t	0.000	0.000	0.225	0.775	0.615	0.000	0.210	0.175	0.105	0.000	0.225	0.670
								$N = 5$					
$T = 500$	**n-n**	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	t-t	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	skewn-n	0.900	0.000	0.100	0.000	1.000	0.000	0.000	0.000	0.990	0.000	0.010	0.000
	skewt-t	0.640	0.000	0.155	0.205	0.985	0.000	0.015	0.000	0.910	0.000	0.085	0.005
$T = 1000$	**n-n**	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	t-t	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	skewn-n	0.645	0.000	0.340	0.015	1.000	0.000	0.000	0.000	0.925	0.000	0.075	0.000
	skewt-t	0.085	0.000	0.180	0.735	0.925	0.000	0.070	0.005	0.610	0.000	0.165	0.225
$T = 2000$	**n-n**	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	t-t	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	skewn-n	0.185	0.000	0.735	0.080	0.995	0.000	0.005	0.000	0.690	0.000	0.310	0.000
	skewt-t	0.000	0.000	0.180	0.820	0.695	0.000	0.180	0.125	0.035	0.000	0.180	0.785
								$N = 10$					
$T = 500$	**n-n**	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	t-t	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	skewn-n	0.945	0.000	0.055	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	skewt-t	0.615	0.000	0.240	0.145	1.000	0.000	0.000	0.000	0.870	0.000	0.130	0.000
$T = 1000$	**n-n**	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	t-t	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	skewn-n	0.635	0.000	0.365	0.000	1.000	0.000	0.000	0.000	0.975	0.000	0.025	0.000
	skewt-t	0.055	0.000	0.235	0.710	0.870	0.000	0.130	0.000	0.640	0.000	0.225	0.135
$T = 2000$	**n-n**	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	t-t	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	skewn-n	0.150	0.000	0.830	0.020	0.995	0.000	0.005	0.000	0.710	0.000	0.290	0.000
	skewt-t	0.000	0.000	0.175	0.825	0.745	0.000	0.175	0.080	0.005	0.000	0.175	0.820

Note: The set of moments $m = \{\rho_S, \lambda_{0.25}, \lambda_{0.45}, \lambda_{0.55}, \lambda_{0.75}\}$ is relevant, the number of replications is $R = 200$, $\hat{d}$ denotes the models in the estimation and d_0 denotes the true model.

Table 15: Empirical selection frequencies based on selection procedures SMM-AIC, SMM-BIC and SMM-HQIC, the Equi-dependence model with the marginal dependence, $\beta_0 = 1$

	d_0 \ $\hat{d}$	SMM-AIC				SMM-BIC				SMM-HQIC			
		n-n	t-t	skewn-n	skewt-t	n-n	t-t	skewn-n	skewt-t	n-n	t-t	skewn-n	skewt-t
						$N = 3$							
$T = 500$	**n-n**	0.655	0.180	0.160	0.005	0.655	0.180	0.160	0.005	0.655	0.180	0.160	0.005
	t-t	0.210	0.680	0.095	0.015	0.210	0.680	0.095	0.015	0.210	0.680	0.095	0.015
	skewn-n	0.085	0.030	0.825	0.060	0.085	0.030	0.825	0.060	0.085	0.030	0.825	0.060
	skewt-t	0.000	0.005	0.105	0.890	0.000	0.005	0.105	0.890	0.000	0.005	0.105	0.890
$T = 1000$	**n-n**	0.835	0.100	0.065	0.000	0.835	0.100	0.065	0.000	0.835	0.100	0.065	0.000
	t-t	0.095	0.880	0.025	0.000	0.095	0.880	0.025	0.000	0.095	0.880	0.025	0.000
	skewn-n	0.045	0.005	0.925	0.025	0.045	0.005	0.925	0.025	0.045	0.005	0.925	0.025
	skewt-t	0.000	0.000	0.040	0.960	0.000	0.000	0.040	0.960	0.000	0.000	0.040	0.960
$T = 2000$	**n-n**	0.955	0.020	0.025	0.000	0.955	0.020	0.025	0.000	0.955	0.020	0.025	0.000
	t-t	0.025	0.975	0.000	0.000	0.025	0.975	0.000	0.000	0.025	0.975	0.000	0.000
	skewn-n	0.015	0.000	0.980	0.005	0.015	0.000	0.980	0.005	0.015	0.000	0.980	0.005
	skewt-t	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000
						$N = 5$							
$T = 500$	**n-n**	0.805	0.080	0.115	0.000	0.805	0.080	0.115	0.000	0.805	0.080	0.115	0.000
	t-t	0.150	0.765	0.075	0.010	0.150	0.765	0.075	0.010	0.150	0.765	0.075	0.010
	skewn-n	0.035	0.000	0.935	0.030	0.035	0.000	0.935	0.030	0.035	0.000	0.935	0.030
	skewt-t	0.005	0.000	0.085	0.910	0.005	0.000	0.085	0.910	0.005	0.000	0.085	0.910
$T = 1000$	**n-n**	0.930	0.040	0.030	0.000	0.930	0.040	0.030	0.000	0.930	0.040	0.030	0.000
	t-t	0.075	0.905	0.015	0.005	0.075	0.905	0.015	0.005	0.075	0.905	0.015	0.005
	skewn-n	0.015	0.000	0.965	0.020	0.015	0.000	0.965	0.020	0.015	0.000	0.965	0.020
	skewt-t	0.000	0.000	0.025	0.975	0.000	0.000	0.025	0.975	0.000	0.000	0.025	0.975
$T = 2000$	**n-n**	0.995	0.005	0.000	0.000	0.995	0.005	0.000	0.000	0.995	0.005	0.000	0.000
	t-t	0.005	0.995	0.000	0.000	0.005	0.995	0.000	0.000	0.005	0.995	0.000	0.000
	skewn-n	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000
	skewt-t	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000
						$N = 10$							
$T = 500$	**n-n**	0.910	0.055	0.035	0.000	0.910	0.055	0.035	0.000	0.910	0.055	0.035	0.000
	t-t	0.155	0.790	0.050	0.005	0.155	0.790	0.050	0.005	0.155	0.790	0.050	0.005
	skewn-n	0.015	0.000	0.975	0.010	0.015	0.000	0.975	0.010	0.015	0.000	0.975	0.010
	skewt-t	0.000	0.000	0.070	0.930	0.000	0.000	0.070	0.930	0.000	0.000	0.070	0.930
$T = 1000$	**n-n**	0.970	0.025	0.005	0.000	0.970	0.025	0.005	0.000	0.970	0.025	0.005	0.000
	t-t	0.020	0.965	0.015	0.000	0.020	0.965	0.015	0.000	0.020	0.965	0.015	0.000
	skewn-n	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000
	skewt-t	0.000	0.000	0.020	0.980	0.000	0.000	0.020	0.980	0.000	0.000	0.020	0.980
$T = 2000$	**n-n**	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	t-t	0.010	0.990	0.000	0.000	0.010	0.990	0.000	0.000	0.010	0.990	0.000	0.000
	skewn-n	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000
	skewt-t	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000

Note: The set of moments $m = \{\rho_S, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.9}, \lambda_{0.95}\}$ is relevant, the number of replications is $R = 200$, $\hat{d}$ denotes the models in the estimation and d_0 denotes the true model.

Table 16: Empirical selection frequencies based on selection procedures SMM-AIC, SMM-BIC and SMM-HQIC, the Equi-dependence model with the marginal dependence, $\beta_0 = 1$

		SMM-AIC				SMM-BIC				SMM-HQIC			
	d_0 \ $\hat{d}$	n-n	t-t	skewn-n	skewt-t	n-n	t-t	skewn-n	skewt-t	n-n	t-t	skewn-n	skewt-t
								$N = 3$					
$T = 500$	**n-n**	0.605	0.215	0.175	0.005	0.605	0.215	0.175	0.005	0.605	0.215	0.175	0.005
	t-t	0.250	0.615	0.130	0.005	0.250	0.615	0.130	0.005	0.250	0.615	0.130	0.005
	skewn-n	0.130	0.075	0.575	0.220	0.130	0.075	0.575	0.220	0.130	0.075	0.575	0.220
	skewt-t	0.005	0.020	0.190	0.785	0.005	0.020	0.190	0.785	0.005	0.020	0.190	0.785
$T = 1000$	**n-n**	0.640	0.195	0.165	0.000	0.640	0.195	0.165	0.000	0.640	0.195	0.165	0.000
	t-t	0.205	0.715	0.080	0.000	0.205	0.715	0.080	0.000	0.205	0.715	0.080	0.000
	skewn-n	0.100	0.040	0.710	0.150	0.100	0.040	0.710	0.150	0.100	0.040	0.710	0.150
	skewt-t	0.000	0.000	0.100	0.900	0.000	0.000	0.100	0.900	0.000	0.000	0.100	0.900
$T = 2000$	**n-n**	0.810	0.120	0.070	0.000	0.810	0.120	0.070	0.000	0.810	0.120	0.070	0.000
	t-t	0.135	0.850	0.015	0.000	0.135	0.850	0.015	0.000	0.135	0.850	0.015	0.000
	skewn-n	0.015	0.005	0.955	0.025	0.015	0.005	0.955	0.025	0.015	0.005	0.955	0.025
	skewt-t	0.000	0.000	0.030	0.970	0.000	0.000	0.030	0.970	0.000	0.000	0.030	0.970
$T = 500$	**n-n**	0.710	0.175	0.115	0.000	0.710	0.175	0.115	0.000	0.710	0.175	0.115	0.000
	t-t	0.260	0.645	0.095	0.000	0.260	0.645	0.095	0.000	0.260	0.645	0.095	0.000
	skewn-n	0.100	0.035	0.715	0.150	0.100	0.035	0.715	0.150	0.100	0.035	0.715	0.150
	skewt-t	0.000	0.000	0.135	0.865	0.000	0.000	0.135	0.865	0.000	0.000	0.135	0.865
$T = 1000$	**n-n**	0.755	0.160	0.085	0.000	0.755	0.160	0.085	0.000	0.755	0.160	0.085	0.000
	t-t	0.160	0.780	0.060	0.000	0.160	0.780	0.060	0.000	0.160	0.780	0.060	0.000
	skewn-n	0.055	0.005	0.890	0.050	0.055	0.005	0.890	0.050	0.055	0.005	0.890	0.050
	skewt-t	0.000	0.000	0.040	0.960	0.000	0.000	0.040	0.960	0.000	0.000	0.040	0.960
$T = 2000$	**n-n**	0.920	0.060	0.020	0.000	0.920	0.060	0.020	0.000	0.920	0.060	0.020	0.000
	t-t	0.045	0.945	0.010	0.000	0.045	0.945	0.010	0.000	0.045	0.945	0.010	0.000
	skewn-n	0.015	0.000	0.970	0.015	0.015	0.000	0.970	0.015	0.015	0.000	0.970	0.015
	skewt-t	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000
$T = 500$	**n-n**	0.840	0.115	0.045	0.000	0.840	0.115	0.045	0.000	0.840	0.115	0.045	0.000
	t-t	0.120	0.830	0.050	0.000	0.120	0.830	0.050	0.000	0.120	0.830	0.050	0.000
	skewn-n	0.035	0.015	0.910	0.040	0.035	0.015	0.910	0.040	0.035	0.015	0.910	0.040
	skewt-t	0.000	0.000	0.060	0.940	0.000	0.000	0.060	0.940	0.000	0.000	0.060	0.940
$T = 1000$	**n-n**	0.915	0.055	0.030	0.000	0.915	0.055	0.030	0.000	0.915	0.055	0.030	0.000
	t-t	0.045	0.940	0.015	0.000	0.045	0.940	0.015	0.000	0.045	0.940	0.015	0.000
	skewn-n	0.005	0.000	0.985	0.010	0.005	0.000	0.985	0.010	0.005	0.000	0.985	0.010
	skewt-t	0.000	0.000	0.010	0.990	0.000	0.000	0.010	0.990	0.000	0.000	0.010	0.990
$T = 2000$	**n-n**	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	t-t	0.005	0.995	0.000	0.000	0.005	0.995	0.000	0.000	0.005	0.995	0.000	0.000
	skewn-n	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000
	skewt-t	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000

Note: The set of moments $m = \{\rho_S, \lambda_{0.25}, \lambda_{0.45}, \lambda_{0.55}, \lambda_{0.75}\}$ is relevant, the number of replications is $R = 200$, $\hat{d}$ denotes the models in the estimation and d_0 denotes the true model.

Table 17: Sample dependence measures

	SAP	**SIE**	**BAYN**	**ALV**	**BAS**	**DAI**	**DTE**	**BMW**	**DPW**	**CON**
					Panel A: estimates of Spearman's rho ρ_S					
SAP		0.563	0.548	0.541	0.563	0.533	0.470	0.513	0.497	0.483
SIE			0.623	0.634	0.708	0.642	0.540	0.629	0.599	0.567
BAYN				0.605	0.670	0.593	0.560	0.564	0.565	0.508
ALV					0.661	0.601	0.581	0.567	0.612	0.541
BAS						0.654	0.536	0.628	0.591	0.590
DAI							0.518	0.796	0.581	0.694
DTE								0.463	0.505	0.405
BMW									0.556	0.697
DPW										0.563
CON										
					Panel B: estimates of Quantile dependence λ_q					
SAP		0.215	0.153	0.215	0.215	0.153	0.092	0.153	0.123	0.123
SIE	-0.129		0.184	0.276	0.246	0.184	0.061	0.246	0.215	0.246
BAYN	-0.135	-0.104		0.246	0.430	0.246	0.092	0.215	0.215	0.153
ALV	-0.104	-0.086	-0.086		0.338	0.246	0.123	0.246	0.184	0.153
BAS	-0.098	-0.092	-0.123	-0.055		0.307	0.153	0.276	0.184	0.153
DAI	-0.055	-0.110	-0.086	-0.061	-0.061		0.123	0.460	0.246	0.276
DTE	-0.043	-0.135	-0.123	-0.098	-0.092	-0.074		0.092	0.061	0.000
BMW	-0.068	-0.031	-0.135	-0.147	-0.092	-0.074	-0.110		0.123	0.246
DPW	-0.092	-0.166	-0.190	-0.086	-0.104	-0.123	-0.098	-0.074		0.215
CON	-0.110	-0.153	-0.117	-0.074	-0.055	-0.061	-0.086	-0.092	-0.074	

Table 18: The evaluation of VaR forecasting performance, $m_1 = \{\lambda_{0.01}, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.15}\}$, $m_2 = \{\lambda_{0.01}, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.15}, \rho_S\}$, $m_3 = \{\lambda_{0.01}, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.15}, \lambda_{0.85}, \lambda_{0.9}, \lambda_{0.95}, \lambda_{0.99}, \rho_S\}$, $m_4 = \{\lambda_{0.05}, \lambda_{0.1}, \lambda_{0.9}, \lambda_{0.95}, \rho_S\}$

	$VaR_{0.05}$				$VaR_{0.01}$			
	Coverage	**Kupiec**	**Christoffersen**	**CAViaR**	**Coverage**	**Kupiec**	**Christoffersen**	**CAViaR**
Factor n-n + m_1	0.0588	0.3227	0.5187	0.4824	0.0095	0.9068	0.9377	0.9663
Factor n-n + m_2	0.0636	0.1326	0.2109	0.1600	0.0095	0.9068	0.9377	0.9663
Factor n-n + m_3	0.0700	0.0299	0.0801	0.0563	0.0127	0.5108	0.7246	0.7376
Factor n-n + m_4	0.0700	0.0299	0.0801	0.0563	0.0127	0.5108	0.7246	0.7376
Factor t-t + m_1	0.0652	0.0944	0.1744	0.1319	0.0095	0.9068	0.9377	0.9663
Factor t-t + m_2	0.0684	0.0448	0.1074	0.0782	0.0095	0.9068	0.9377	0.9663
Factor t-t + m_3	0.0715	0.0195	0.0578	0.0387	0.0111	0.7799	0.8877	0.9175
Factor t-t + m_4	0.0763	0.0048	0.0141	0.0065	0.0095	0.9068	0.9377	0.9663
Factor skewt-t + m_1	0.0636	0.1326	0.3046	0.2733	0.0095	0.9068	0.9377	0.9663
Factor skewt-t + m_2	0.0715	0.0195	0.0389	0.0204	0.0111	0.7799	0.8877	0.9175
Factor skewt-t + m_3	0.0668	0.0657	0.1392	0.1039	0.0095	0.9068	0.9377	0.9663
Factor skewt-t + m_4	0.0700	0.0299	0.0801	0.0563	0.0079	0.5919	0.8338	0.8625
Factor skewn-n + m_1	0.0604	0.2452	0.4503	0.4169	0.0095	0.9068	0.9377	0.9663
Factor skewn-n + m_2	0.0636	0.1326	0.2109	0.1600	0.0079	0.5919	0.8338	0.8625
Factor skewn-n + m_3	0.0636	0.1326	0.2109	0.1600	0.0095	0.9068	0.9377	0.9663
Factor skewn-n + m_4	0.0652	0.0944	0.1744	0.1319	0.0095	0.9068	0.9377	0.9663
Factor skewt-n + m_1	0.0620	0.1823	0.3770	0.3450	0.0095	0.9068	0.9377	0.9663
Factor skewt-n + m_2	0.0652	0.0944	0.1744	0.1319	0.0095	0.9068	0.9377	0.9663
Factor skewt-n + m_3	0.0652	0.0944	0.1744	0.1319	0.0095	0.9068	0.9377	0.9663
Factor skewt-n + m_4	0.0700	0.0299	0.0801	0.0563	0.0048	0.1421	0.3373	0.4195
Factor t-n + m_1	0.0636	0.1326	0.2109	0.1600	0.0095	0.9068	0.9377	0.9663
Factor t-n + m_2	0.0684	0.0448	0.1074	0.0782	0.0095	0.9068	0.9377	0.9663
Factor t-n + m_3	0.0715	0.0195	0.0578	0.0387	0.0095	0.9068	0.9377	0.9663
Factor t-n + m_4	0.0747	0.0078	0.0204	0.0099	0.0095	0.9068	0.9377	0.9663

Table 19: The relative comparison of forecasting accuracies of $VaR_{0.05}$ ($VaR_{0.01}$) by Diebold-Mariano Test

		n-n				t-t				skewt-t				skewn-n				skewt-n				t-n			
		m_1	m_2	m_3	m_4	m_1	m_2	m_3	m_4	m_1	m_2	m_3	m_4	m_1	m_2	m_3	m_4	m_1	m_2	m_3	m_4	m_1	m_2	m_3	m_4
n-n	m_1		1(0)	1(0)	1(0)	1(0)	1(0)	1(0)	1(0)	0(0)	1(0)	1(0)	1(0)	2(0)	1(0)	1(0)	1(0)	0(0)	1(0)	1(0)	1(0)	1(0)	1(0)	1(0)	1(0)
	m_2			1(0)	1(0)	0(0)	1(0)	1(0)	1(0)	0(0)	0(0)	0(0)	1(0)	2(0)	0(0)	0(0)	0(0)	2(0)	0(0)	0(0)	1(0)	2(0)	1(0)	1(0)	1(0)
	m_3				0(0)	2(0)	0(0)	1(0)	1(0)	2(0)	0(0)	2(0)	0(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	0(0)	1(0)
	m_4					2	0(0)	1(0)	1(0)	2(0)	0(0)	2(0)	0(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	0(0)	1(0)
t-t	m_1						1(0)	1(0)	1(0)	0(0)	1(0)	1(0)	1(0)	2(0)	0(0)	0(0)	1(0)	2(0)	0(0)	0(0)	1(0)	2(0)	1(0)	1(0)	1(0)
	m_2							1(0)	1(0)	2(0)	0(0)	2(0)	0(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	0(0)	1(0)
	m_3								1(2)	2(0)	0(0)	2(0)	0(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	1(0)
	m_4									2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)
skewt-t	m_1										1(0)	1(0)	1(0)	2(0)	0(0)	0(0)	1(0)	2(0)	0(0)	0(0)	1(0)	0(0)	1(0)	1(0)	1(0)
	m_2											0(0)	0(0)	2(0)	0(0)	0(0)	0(0)	2(0)	2(0)	0(0)	0(0)	2(0)	0(0)	0(0)	0(0)
	m_3												1(0)	2(0)	2(0)	0(0)	0(0)	2(0)	2(0)	0(0)	1(0)	2(0)	1(0)	1(0)	1(0)
	m_4													2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	2(0)	0(0)	1(0)
skewn-n	m_1														1(0)	1(0)	1(0)	0(0)	1(0)	1(0)	1(0)	1(0)	1(0)	1(0)	1(0)
	m_2															0(0)	0(0)	2(0)	0(0)	0(0)	1(0)	0(0)	1(0)	1(0)	1(0)
	m_3																0(0)	2(0)	0(0)	0(0)	1(0)	0(0)	1(0)	1(0)	1(0)
	m_4																	2(0)	0(0)	0(0)	1(0)	2(0)	1(0)	1(0)	1(0)
skewt-n	m_1																		1(0)	1(0)	1(0)	1(0)	1(0)	1(0)	1(0)
	m_2																			0(0)	1(0)	2(0)	1(0)	1(0)	1(0)
	m_3																				1(0)	2(0)	1(0)	1(0)	1(0)
	m_4																					2(0)	0(0)	1(0)	1(0)
t-n	m_1																						1(0)	1(0)	1(0)
	m_2																							1(0)	1(0)
	m_3																								1(0)
	m_4																								

Note: $m_1 = \{\lambda_{0.01}, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.15}\}$, $m_2 = \{\lambda_{0.01}, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.15}, \rho_S\}$, $m_3 = \{\lambda_{0.01}, \lambda_{0.05}, \lambda_{0.1}, \lambda_{0.15}, \lambda_{0.85}, \lambda_{0.9}, \lambda_{0.95}, \lambda_{0.99}, \rho_S\}$, $m_4 = \{\lambda_{0.05}, \lambda_{0.1}, \lambda_{0.9}, \lambda_{0.95}, \rho_S\}$, 1 indicates the scenario that row model outperforms column model, 2 indicates the case that model at column outperforms model at row, the indifference between two models is marked as 0.

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. & 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.

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

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), 841862.

Dias, A. & Embrechts, P. (2004), Dynamic copula models for multivariate high-frequency data in finance, Working papers, 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 valueat risk by regression quantiles', *Journal of Business & Economic Statistics* **22**(4), 367382.

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

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 (Methodological)* **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 cdo and an n-th default cds without monte carlo simulation’, *The journal of derivatives : the official publication of the International Association of Financial Engineers* **12**(2), 8–23.

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

Joe, H. (2014), *Dependence Modeling with Copulas*, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, 1 edn, Chapman and Hall/CRC.

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

Kupiec, P. H. (1995), ‘Techniques for verifying the accuracy of risk measurement models’, *The Journal of Derivatives* **3**(2), 7384.

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*, Springer-Verlag New York, Inc., Secaucus, NJ, USA.

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

Oh, D. H. & Patton, A. (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).

van der Voort, M. (2005), 'Factor copulas: Totally external defaults', *Working paper* .