Dynamic Factor Copula Models with Estimated Cluster Assignments*

Dong Hwan Oh
Federal Reserve Board

Andrew J. Patton
Duke University

3 November 2020

First Draft: Comments Welcome.

Abstract

This paper proposes a dynamic multi-factor copula for use in high dimensional time series applications. A novel feature of our model is that the assignment of individual variables to groups is estimated from the data, rather than being pre-assigned using SIC industry codes, market capitalization ranks, or other *ad hoc* methods. We adapt the k-means clustering algorithm for use in our application and show that it has excellent finite-sample properties. Applying the new model to returns on 110 US equities, we find around 20 clusters to be optimal. In out-of-sample forecasts, we find that a model with as few as five estimated clusters significantly outperforms an otherwise identical model with 21 clusters formed using two-digit SIC codes.

Keywords: correlation, tail risk, multivariate density forecast.

J.E.L. codes: C32, C38, C58.

*The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors. We are grateful to Forrest Denson for his excellent computing support. Email: donghwan.oh@frb.gov, andrew.patton@duke.edu. A Matlab toolbox to implement the methods proposed in this paper will be available at http://public.econ.duke.edu/∼ap172.

1 Introduction

Models for the dependence structure of a large collection of variables play an important role in risk management and regulation, yet there is a relative paucity of such models. A key impediment is that these models need to be parsimonious enough to avoid the inevitable curse of dimensionality that arises in high-dimensional applications, yet flexible enough to capture the time-varying and potentially asymmetric nature of the dependence between economic variables.

We propose a multi-factor, high-dimensional, copula model where the assignment of individual variables to groups or clusters is estimated from the data. Existing approaches for similar problems (see Creal and Tsay, 2015, Bester and Hansen, 2016, and Opschoor *et al.*, 2020, for example) use pre-specified cluster assignments, based on SIC industry codes, or market capitalization deciles, or similar. In the absence of a computationally feasible data-driven alternative such approaches are reasonable, however it is not obvious that such assignments are optimal empirically. We propose a method based on k-means clustering (see, e.g., Hastie *et al.*, 2009) to estimate the optimal assignments of variables to clusters, and we model dynamics in the conditional copula using a GAS model (Creal *et al.*, 2013; Harvey 2013).

The estimation of the optimal cluster assignments for a high-dimensional dynamic copula model requires us to overcome two computational hurdles. Firstly, rather than the simulation-based factor copula model of Oh and Patton (2017), we adopt and extend the model of Opschoor et al. (2020), which has a closed-form likelihood and is thus much faster to estimate. Our extension enables us to capture asymmetric dependencies which can be important for equity returns, see Ang and Chen (2002), Hong *et al.* (2007) and Patton (2013) amongst many others. Secondly, we exploit the fact that the presence of clusters in the dynamic model implies the presence of clusters in the (misspecified) static version of the model. The static version of the model is naturally much faster to estimate than the dynamic version. These two techniques, combined with extensive use of parallel processing, make the estimation of optimal cluster assignments feasible.

We prove the consistency of the estimated cluster assignments under very mild conditions, and we find in realistically-designed simulations that our estimation method is remarkably accurate.

We apply the new model to daily returns on 110 U.S. equities over the period 2010-2019, and consider a range of choices for the number of clusters in the model. We find that the BIC optimal number of clusters is around 20, and moreover find that a model with just *five* estimated clusters outperforms an otherwise identical model based on 21 clusters formed using two-digit SIC groupings. In out-of-sample forecast comparisons, we find that the model with estimated cluster assignments significantly outperforms one with clusters formed using two-digit SIC codes.

This paper bridges two lines in the extant literature. Most directly, this paper is related to the literature on high-dimensional methods for financial risk measurement. Early work focused on improved methods for estimating large covariance matrices. For example, Fan *et al.* (2008, 2013) propose using a factor model where the number of factors grows with the number of variables, with the latter of these papers also accommodating approximate factor models. Tao *et al.* (2011) consider high-dimensional covariance matrix estimation based on a combination of high- and low-frequency data, also using a factor model. Hautsch *et al.* (2012) propose a method to estimate covariance matrices using high frequency data from assets with varying degrees of liquidity. More recent work in this area has included a focus on copula-based models, such as Creal and Tsay (2015) who proposed a high-dimensional stochastic copula with a factor structure, and Oh and Patton (2018) and Opschoor *et al.* (2020) who consider factor copulas with dynamics driven by a GAS specification. Christoffersen *et al.* (2018) propose a high-dimensional dynamic copula model with DCC (Engle, 2002) type dynamics. As far as we know, our paper is the first to consider a high-dimensional copula model with estimated group assignments.

This paper is also related to the fast-growing area of clustering and classification methods in economics and finance. Lin and Ng (2012) and Bonhomme and Manresa (2015) consider linear panel models with unknown group assignments which are estimated using k-means clustering. Su *et al.* (2016, 2019) consider panel models with group assignments estimated using a new type of LASSO estimator. The latter of these papers allows the parameters of the panel model to vary nonparametrically with time. Vogt and Linton (2020) also consider nonparametric regression for a panel of data with unknown group assignments. Francis *et al.* (2017) cluster countries by their business cycle patterns, and Patton and Weller (2019) consider clustering stocks by the risk premia they generate. This research area is very active and this review is surely incomplete already.

The remainder of the paper is structured as follows. In Section 2 we present the dynamic copula models considered in this paper, and in Section 3 we discuss how we can optimally assign variables to clusters. Section 4 presents the results of a simulation study of the finite-sample performance of the proposed model and estimation method. Section 5 applies the new methods to a collection of 110 stock returns. Section 6 concludes, and the appendix contains proofs and technical details. A web appendix contains additional analyses and material.

2 A dynamic skewed t factor copula model

A copula is an N-dimensional distribution function with $Unif\,(0,1)$ margins, and even when N is only moderately-sized the curse of dimensionality arises. A common approach to overcome this in other contexts is to impose some sort of factor structure, and recent work on high-dimensional copula models has moved in this direction, see Oh and Patton (2017, 2018), Creal and Tsay (2015) and Opschoor *et al.* (2020). An attractive feature of the latter two papers is that the copula likelihood is available in closed form. Motivated by previous work showing that equity returns exhibit asymmetric dependence (see, e.g., Ang and Chen, 2002, Hong *et al.*, 2007, and Patton, 2013), we consider an extension of the model proposed by Opschoor *et al.* (2020) to allow for asymmetric dependence, namely a skewed t factor copula:

$$
\begin{aligned}
u_{i,t} &= T_{skew}\left(x_{i,t};\nu,\zeta\right), \quad i=1,\cdots,N, && (1)\\
x_{i,t} &= \sqrt{W_t}\left(\widetilde{\boldsymbol{\lambda}}_{i,t}'\mathbf{z}_t+\sigma_{i,t}\epsilon_{i,t}\right)+\zeta W_t, && (2)\\
\text{where}\quad \mathbf{z}_t &\sim iid\ N\left(\mathbf{0},\mathbf{I}_k\right),\ \ \epsilon_{i,t}\sim iid\ \mathcal{N}\left(0,1\right), && (3)\\
W_t &\sim iid\ IG\left(\frac{\nu}{2},\frac{\nu}{2}\right),\ \ W_t\perp\mathbf{z}_t\perp\epsilon_{i,t} && (4)
\end{aligned}
$$

where $T_{skew}\left(\cdot;\nu,\zeta\right)$ denotes the univariate skewed t CDF of $x_{i,t}$, with degrees of freedom parameter $\nu\in(2,\infty]$ and asymmetry parameter $\zeta\in[-1,1]$.[1] $\widetilde{\boldsymbol{\lambda}}_{i,t}$ is a $k\times 1$ vector of scaled factor loadings,

[1] Creal and Tsay (2010) describe this copula but do not implement it or present results on its likelihood and scores. As that paper notes, the presence of skewness implies that $T_{skew}\left(\cdot;\nu,\zeta\right)$ is not available in closed form, and Creal

$\mathbf{z}_t$ is a $k \times 1$ vector of common latent factors and $\epsilon_{i,t}$ is an idiosyncratic shock, both Normally distributed, and W_t is an inverse gamma variable. We define the $k \times 1$ vector $\widetilde{\boldsymbol{\lambda}}_{i,t}$ and scalar $\sigma_{i,t}$ as

$$\widetilde{\boldsymbol{\lambda}}_{i,t} = \frac{\boldsymbol{\lambda}_{i,t}}{\sqrt{1 + \boldsymbol{\lambda}'_{i,t}\boldsymbol{\lambda}_{i,t}}}, \quad \sigma^2_{i,t} = \frac{1}{1 + \boldsymbol{\lambda}'_{i,t}\boldsymbol{\lambda}_{i,t}} \tag{5}$$

for a factor loading $\boldsymbol{\lambda}_{i,t}$ to maintain the unit variance of $\widetilde{\boldsymbol{\lambda}}'_{i,t}\mathbf{z}_t + \sigma_{i,t}\epsilon_{i,t}$. The skewed t copula nests the Student's t copula when $\zeta = 0$, and the Gaussian copula when $\zeta = 0$ and $\nu \to \infty$. Given this structure, the correlation matrix, $\mathbf{R}_t$, of $\left[\widetilde{\boldsymbol{\lambda}}'_{1,t}\mathbf{z}_t + \sigma_{1,t}\epsilon_{1,t}, \ldots, \widetilde{\boldsymbol{\lambda}}'_{N,t}\mathbf{z}_t + \sigma_{N,t}\epsilon_{N,t}\right]'$ is

$$\mathbf{R}_t = \widetilde{\mathbf{L}}'_t\widetilde{\mathbf{L}}_t + \mathbf{D}_t \tag{6}$$

where $\widetilde{\mathbf{L}}_t = \left[\widetilde{\boldsymbol{\lambda}}_{1,t}, \ldots, \widetilde{\boldsymbol{\lambda}}_{N,t}\right]$ and $\mathbf{D}_t = diag\left(\sigma^2_{1,t}, \ldots, \sigma^2_{N,t}\right)$. The skewed t copula then contains time-varying factor loadings $\left[\boldsymbol{\lambda}'_{1,t}, \ldots\cdots, \boldsymbol{\lambda}'_{N,t}\right]$ and static shape parameters $[\nu, \zeta]$. Creal and Tsay (2015) show that a factor copula structure of the sort in equation (2) facilitates the evaluation of the copula density even for high dimensions since the inverse and determinant of $\mathbf{R}_t$ are available in closed form and require only lower dimension inversions and determinant calculations:

$$\begin{aligned}
\mathbf{R}_t^{-1} &= \mathbf{D}_t^{-1} - \mathbf{D}_t^{-1}\widetilde{\mathbf{L}}'_t\left(\mathbf{I}_k + \widetilde{\mathbf{L}}_t\mathbf{D}_t^{-1}\widetilde{\mathbf{L}}'_t\right)^{-1}\widetilde{\mathbf{L}}_t\mathbf{D}_t^{-1} \\
|\mathbf{R}_t| &= \left|\mathbf{I}_k + \widetilde{\mathbf{L}}_t\mathbf{D}_t^{-1}\widetilde{\mathbf{L}}'_t\right| \cdot |\mathbf{D}_t|\,.
\end{aligned}$$

We consider a factor structure determined by a $(G+1) \times 1$ vector $\mathbf{z}_t$ of common latent factors and a loading matrix $\widetilde{\mathbf{L}}_t$. Specifically, we employ one common factor and G group-specific factors where the common factor is shared by all variables, and each group-specific factor is shared only by members of that cluster. We allow the factor loadings to vary across groups (corresponding to the most flexible model considered by Opschoor *et al.*, 2020). For example, assuming there are G

and Tsay (2010) omit it from their analysis. In Appendix A.1 we describe a simple and computationally tractable method to overcome this difficulty.

groups and each group has only two members, $\mathbf{z}_t$ and $\widetilde{\mathbf{L}}_t$ are determined by

$$\mathbf{z}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{I}_{G+1}\right)$$
$$\widetilde{\mathbf{L}}_t' = \begin{pmatrix} \widetilde{\lambda}_{M,1,t} & \widetilde{\lambda}_{1,1,t} & 0 & 0 & \cdots & 0 \\ \widetilde{\lambda}_{M,2,t} & 0 & \widetilde{\lambda}_{2,2,t} & 0 & \cdots & 0 \\ \widetilde{\lambda}_{M,3,t} & 0 & 0 & \widetilde{\lambda}_{3,3,t} & 0 & 0 \\ \vdots & \vdots & \vdots & 0 & \ddots & 0 \\ \widetilde{\lambda}_{M,G,t} & 0 & 0 & 0 & 0 & \widetilde{\lambda}_{G,G,t} \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{7}$$

where $\otimes$ denotes the Kronecker product.

Next, we formulate the dynamics of $2G$ distinct factor loadings based on the generalized autoregressive score model proposed by Creal *et al.* (2013) and Harvey (2013). Specifically, we model those dynamics by:

$$\begin{aligned} \lambda_{M,g,t+1} &= \omega_g^M + \alpha^M \frac{\partial \log \mathbf{c}_{Skewt,t}\left(\mathbf{x}_t; \mathbf{R}_t, \nu, \zeta\right)}{\partial \lambda_{M,g,t}} + \beta^M \lambda_{M,g,t}, \text{ for } g = 1, ..., G \\ \lambda_{g,g,t+1} &= \omega_g^C + \alpha^C \frac{\partial \log \mathbf{c}_{Skewt,t}\left(\mathbf{x}_t; \mathbf{R}_t, \nu, \zeta\right)}{\partial \lambda_{g,g,t}} + \beta^C \lambda_{g,g,t}, \text{ for } g = 1, ..., G \end{aligned} \tag{8}$$

where $\mathbf{x}_t = T_{skew}^{-1}\left(\mathbf{u}_t; \nu, \zeta\right)$, $\mathbf{c}_{Skewt,t}\left(\cdot; \mathbf{R}_t, \nu, \zeta\right)$ is the conditional skewed t copula density and $\left[\omega_1^M, \ldots, \omega_G^M, \omega_1^C, \ldots, \omega_G^C, \alpha^M, \beta^M, \alpha^C, \beta^C\right]'$ is the vector of parameters determining the dynamics of time varying factor loadings. Obviously the key component is the score of the conditional copula $\partial \log \mathbf{c}_{Skewt,t}\left(\mathbf{x}_t; \mathbf{R}_t, \nu, \zeta\right) / \partial \boldsymbol{\eta}_t$ where $\boldsymbol{\eta}_t$ is a $(2G \times 1)$ vector of all dynamic factor loadings:

$$\boldsymbol{\eta}_t = \left[\lambda_{M,1,t}, \ldots, \lambda_{M,G,t}, \lambda_{1,1,t}, \ldots, \lambda_{G,G,t}\right]'. \tag{9}$$

The skewed t copula density and the analytical derivation of its score are given in Appendix A.1. and A.2. respectively.

3 Clustering and factor copulas

3.1 Clustering via a misspecified model

While the closed-form density and GAS equations presented in the previous section greatly reduce the computational burden of estimating a dynamic high-dimensional copula model, this model is still too costly to use when combined with an EM algorithm to estimate group assignments from the data. In this section we show that the structure of our model is such that we can estimate group assignments based on a simpler, misspecified, model, overcoming this hurdle.

Firstly, consider a static skew t factor copula. The factor loading vectors ($\boldsymbol{\lambda}_i$) obey a cluster structure, in that all variables in the same cluster have the same loading vector. From equation (5) above, given the factor loadings we can obtain the normalized loadings and idiosyncratic variances, $\widetilde{\boldsymbol{\lambda}}_{i,t}$ and $\sigma^2_{i,t}$, and from those we obtain the correlation matrix:

$$\mathbf{R} = \widetilde{\mathbf{L}}'\widetilde{\mathbf{L}} + \mathbf{D}$$

where $\widetilde{\mathbf{L}} = \left[\widetilde{\boldsymbol{\lambda}}_1, \ldots, \widetilde{\boldsymbol{\lambda}}_N\right]$ and $\mathbf{D} = diag\left(\sigma^2_1, \ldots, \sigma^2_N\right)$. The cluster structure embedded in $\boldsymbol{\lambda}_i$ implies that $\mathbf{R}$ exhibits a block structure, which, as discussed above, can be used to speed up matrix inverse and determinant calculations. Further, we note that the block structure in $\mathbf{R}$ holds *regardless* of the shape parameters (ν, ζ). Thus a Normal factor copula, where the shape parameters are incorrectly fixed at $(\nu, \zeta) = (\infty, 0)$ will exhibit the same cluster structure as the more complicated skew t factor copula. This means that the cluster assignments implied by the Normal factor copula are identical to the skew t factor copula, permitting us to use the simpler model to estimate cluster assignments, with the usual caveat that these estimates are likely less precise than those based on the true model.

Next consider a time-varying skew t factor copula. In this case the time-varying correlation matrix $\mathbf{R}_t = \widetilde{\mathbf{L}}_t'\widetilde{\mathbf{L}}_t + \mathbf{D}_t$ obeys a block structure, and while the values taken by the elements of $\mathbf{R}_t$ vary over time, the block structure is constant due to the maintained assumption that group assignments are stable. The conditional marginal copula of any pair $(u_{i,t}, u_{j,t})$ is determined completely by $(R_{i,j,t}, \nu, \zeta)$, and any pair of variables (i, j) belonging to groups (g_1, g_2) will have the same distribution as any other pair (i', j') belonging to the same two groups. The unconditional marginal copula is just an integral of the conditional marginal copula, and so the unconditional rank correlation matrix, $\bar{\boldsymbol{\varrho}} \equiv Corr[\mathbf{u}_t]$, exhibits the same cluster structure as the conditional correlation matrix $\mathbf{R}_t$, opening up the possibility of using a constant Normal factor copula to estimate group assignments for a dynamic skew t factor copula.

One complication arises when using a static copula to determine group assignments for a dynamic DGP: since we are taking time series averages, it is possible that the unconditional rank correlation matrix $\bar{\boldsymbol{\varrho}}$ is more homogeneous than the conditional correlation matrix $\mathbf{R}_t$, making it harder to identify group assignments. That is, clusters may not be as well separated in the approximating model as in the true model. The concept of "well separatedness" is a finite-sample issue, and we examine it in detail in our simulation study. To preview our findings, our simulations indicate that this is not a significant concern here.

3.2 Estimation of cluster assignments and copula parameters

The main advantage of using a factor copula comes from the dimension reduction enabled by classifying variables into a relatively small number of clusters and assuming identical factor loadings within each cluster. In the existing literature, variables are clustered according to observable characteristics, such as SIC industry classifications. Given those cluster assignments, the factor copula can be estimated via maximum likelihood under standard conditions, however, the *ex ante* assignments of variables to clusters may not provide the best fit to the data.

We propose an iterative method which estimates cluster assignments, and copula parameters, directly from the data, exploiting an expectation-maximization (EM) algorithm. This algorithm cycles between (1) estimating copula parameters given cluster assignments and (2) estimating cluster assignments given the estimated copula parameters. Let $\Gamma = [\gamma_1, ..., \gamma_N]$ where $\gamma_i \in \{1, ..., G\}$ for $i = 1, ..., N$, denote the vector of cluster assignments, and let $\boldsymbol{\theta} = \left[\lambda_1^M, ..., \lambda_G^M, \lambda_1^C, ..., \lambda_G^C\right]$ be the vector of market and cluster-specific factor loadings used to obtain the correlation matrix parameter for the static Gaussian factor copula, with log-likelihood denoted $\log \mathbf{c}(\cdot)$. Given an estimate of the cluster assignment vector, $\hat{\Gamma}^{(s)}$ the log-likelihood of the copula model is maximized over the copula parameters $\boldsymbol{\theta}$ to yield:

$$\hat{\boldsymbol{\theta}}^{(s+1)} = \arg\max_{\boldsymbol{\theta}} \hat{Q}_T\left(\boldsymbol{\theta},\hat{\Gamma}^{(s)}\right) \tag{10}$$

$$\text{where } \hat{Q}_T(\boldsymbol{\theta},\Gamma) \equiv \frac{1}{T}\sum_{t=1}^{T}\log \mathbf{c}(\mathbf{u}_t;\boldsymbol{\theta},\Gamma) \tag{11}$$

Then, given copula parameter $\hat{\boldsymbol{\theta}}^{(s+1)}$, the log-likelihood is maximized over cluster assignments γ_i for $i = 1, ..., N$:

$$\hat{\gamma}_i^{(s+1)} = \arg\max_{g\in\{1,\dots,G\}} \hat{Q}_T\left(\hat{\boldsymbol{\theta}}^{(s+1)},\tilde{\Gamma}_{i,g}^{(s)}\right) \tag{12}$$

where $\tilde{\Gamma}_{i,g}^{(s)}$ is equal to $\hat{\Gamma}^{(s)}$ except that the i^{th} element is set equal to g.

The copula parameter in equation (10) is estimated through a typical gradient-based optimization, while we update the each variable's cluster assignment (equation 12) by re-optimizing the cluster assignments one variable at a time, motivated by the method underlying k-means clustering. This step requires only $G \times N$ likelihood evaluations, which makes the cluster assignment estimation feasible and fast.[2] The iteration between equation (10) and equation (12) continues until convergence. Convergence to a local optimum is guaranteed, and we use 100 randomly-chosen starting values to improve the accuracy of the estimator. Our simulation study below confirms this to be a sufficient number of starting values.[3] Denote the resulting estimates as $\left(\hat{\boldsymbol{\theta}}_T, \hat{\Gamma}_T\right)$.

We next provide conditions under which the estimated cluster assignments, $\hat{\Gamma}_T$, are consistent for the true cluster assignments, Γ_0. This is a non-standard estimation problem as the parameter Γ_0 is discrete: each of its N elements can take one of only G values.[4] Let $\mathcal{G}$ denote the parameter space for Γ. Since the labels attached to clusters are arbitrary (i.e., the objective function is invariant

[2] Other estimation algorithms for k-means type problems have been proposed in the computer science/machine learning literature. Given the very good finite-sample performance we find for the algorithm described here, when a sufficient number of starting values is used, we did not consider any alternatives.

[3] In our simulations we find that between 25% and 45% of starting values lead to the global optimum, emphasizing that more than a single starting value is required, and also that the problem of convergence to local optima can be overcome with a reasonable number of starting values.

[4] In some k-means applications, a cluster may contain just a single member. Our factor copula model, which has one common factor and G cluster-specific factors, is unidentified if any cluster has fewer than two members and so we restrict $\mathcal{G}$ to impose this condition.

to relabeling the clusters), there is a *set* of correct cluster labels, rather than just a singleton; let $\mathcal{G}_0$ denote this set. To state the assumptions we define the following:

$$\tilde{\boldsymbol{\theta}}^*(\Gamma) = \arg\min_{\boldsymbol{\theta}\in\Theta} \mathbb{E}\left[\log \mathbf{c}\left(\mathbf{u}_t;\boldsymbol{\theta},\Gamma\right)\right] \tag{13}$$

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}\in\Theta} \mathbb{E}\left[\log \mathbf{c}\left(\mathbf{u}_t;\boldsymbol{\theta},\Gamma_0\right)\right] \tag{14}$$

Note that the parameter $\boldsymbol{\theta}^*$ is a pseudo-true parameter: it is the optimal parameter for the misspecified static Gaussian copula model. We obtain this parameter as a by-product of estimating the cluster assignments, but we have no subsequent use for it.

Assumption 1: $\{\mathbf{u}_t\}$ is a stationary ergodic sequence.

Assumption 2: For each $\Gamma \in \mathcal{G}$, (a) $\left|\log \mathbf{c}\left(\mathbf{u}_t;\boldsymbol{\theta},\Gamma\right)\right|_1 < \infty \,\forall\, \boldsymbol{\theta} \in \Theta$, (b) $\left\|\nabla_\theta \log \mathbf{c}\left(\mathbf{u}_t;\tilde{\boldsymbol{\theta}}^*(\Gamma),\Gamma\right)\right\|_1 < \infty$, and (c) $\left\|\nabla_{\theta\theta} \log \mathbf{c}\left(\mathbf{u}_t;\boldsymbol{\theta},\Gamma\right)\right\|_1 < \infty \;\forall\; \boldsymbol{\theta} \in \Theta$.

Assumption 3: (a) For each $\Gamma \in \mathcal{G}$, $\limsup_{T\to\infty}\left[\hat{Q}_T\left(\tilde{\boldsymbol{\theta}}^*(\Gamma),\Gamma\right) - \hat{Q}_T\left(\boldsymbol{\theta},\Gamma\right)\right] > 0 \;\forall\; \boldsymbol{\theta} \in \Theta\backslash\eta_T(\varepsilon)$, where $\eta_T(\varepsilon)$ is an ε-neighborhood of $\tilde{\boldsymbol{\theta}}^*(\Gamma)$, and (b) $\limsup_{T\to\infty}\left[\hat{Q}_T\left(\boldsymbol{\theta}^*,\Gamma_0\right) - \hat{Q}_T\left(\boldsymbol{\theta},\Gamma\right)\right] > 0 \;\forall\; (\boldsymbol{\theta},\boldsymbol{\Gamma}) \in \{\Theta \setminus \eta_T(\varepsilon)\} \times \{\mathcal{G} \setminus \mathcal{G}_0\}$.

Assumption 1 allows for general forms of serial dependence in the data (e.g., mixing). Importantly, given that we expect the static Gaussian copula model to be misspecified, it does *not* require correct specification of the conditional copula. Assumption 2, combined with Assumption 1, ensures that the log-likelihood and its first and second derivatives each obey a law of large numbers. Assumption 3 is a standard "identifiable uniqueness" assumption required for estimation, see Definition 3.3 of White (1994). In our application, it requires that the clusters are "well separated." If the clusters are too close together, then identification of the clusters breaks down. A similar assumption is made in, e.g., Hahn and Moon (2010) and Bonhomme and Manresa (2015). The proof of the following theorem is in Appendix A.4.

Theorem 1 *Under Assumptions 1-3 we have* $\hat{\Gamma}_T \xrightarrow{p} \Gamma_0$ *as* $T \to \infty$, *up to re-labeling of the clusters.*

Results from related contexts suggest that if the series $\{\mathbf{u}_t\}$ generated by equation (1) satisfies certain mixing properties, a large deviations principle may be applied (e.g., see Hahn and Moon, 2010, Choirat and Seri, 2012, and Bonhomme and Manresa, 2015). This enables obtaining a rate result, refining the consistency result in Theorem 1. Specifically, estimated cluster assignments have been shown in some applications to be superconsistent, with estimation errors taking the form:

$$\Pr\left[\hat{\Gamma}_T \neq \Gamma_0\right] \leq C_1 \exp\left\{-C_2 T^{\kappa}\right\} \tag{15}$$

for some constants $C_1, C_2, \kappa > 0$.[5] A result of the form in equation (15) implies that estimation error in estimated cluster assignments vanishes much faster than the usual $\sqrt{T}$ rate, and the simulation results presented in the next section reveal that cluster assignments are indeed estimated extremely well. Unfortunately, general results on the mixing properties of GAS processes are not yet available in the literature, and so we do not pursue a result of this nature here.[6]

With the estimated the cluster assignments $\hat{\Gamma}_T$ in hand, we estimate the parameters of the skewed t copula with GAS dynamics:

$$\hat{\boldsymbol{\psi}}_T = \arg\max_{\boldsymbol{\psi}} \sum_{t=1}^{T} \log \mathbf{c}_{Skewt,t}\left(u_t; \boldsymbol{\psi} \middle| \hat{\Gamma}_T\right)$$

where $\boldsymbol{\psi} = \left[\omega_1^M, \ldots, \omega_G^M, \omega_1^C, \ldots, \omega_G^C, \alpha^M, \beta^M, \alpha^C, \beta^C, \nu, \zeta\right]'$. As the parameter $\boldsymbol{\psi}$ is large, we adopt a "variance targeting" approach to separately estimate the intercept parameters $\left[\omega_1^M, \ldots, \omega_G^C\right]$, leaving us with only six parameters that require difficult numerical optimization. Details on this method are described in Appendix A.3.

4 Simulation study

We investigate the finite-sample performance of the estimation method proposed above in a simulation study designed to match the key features of our empirical application below. We consider

[5] For example, Hahn and Moon (2010) provide conditions under which alpha mixing implies $\kappa = 1/2$, and phi mixing implies $\kappa = 1$. The constants C_1, C_2 vary with the specifics of the application.

[6] Related to the GAS context considered here, Carrasco and Chen (2002) and Hafner and Preminger (2009) show that univariate and multivariate GARCH processes, respectively, are beta mixing. Some results on the stationarity and ergodicity of univariate GAS processes are presented in Blasques et al. (2014).

a sample size of $T = 1000$ and a collection of $N = 100$ variables, and three different factor copulas: a Gaussian factor copula, a t factor copula, and a skew t factor copula, corresponding to $[\nu, \zeta] = [\infty, 0]$, $[5, 0]$, $[5, -0.1]$ respectively. For illustration, a sample of bivariate data from these three copulas, as well as a skew Normal copula which we omit from the simulation study, is presented in Figure 1. In all cases the linear correlation is 0.5, and to aid the interpretation we transform draws from these copulas using the inverse Normal CDF, and so these four distributions all have $N(0,1)$ marginal distributions. In the upper-left panel of Figure 1, we see the familiar bivariate Normal distribution, with low dependence in the tails and displaying radial symmetry. The upper-right panel displays the Student's t copula, which is also radially symmetric but exhibits tail dependence, which manifests in this figure as realizations that lie close to the main diagonal in the upper and lower joint tails. The lower two panels present asymmetric copulas, with dependence being stronger in the lower tail than the upper tail, particularly for the skew t copula which exhibits non-zero tail dependence.

[INSERT FIGURE 1 ABOUT HERE]

We consider two cases for the dynamics of the copula: the benchmark static case, in which the conditional copula is constant, and the case of empirical interest, where the parameters of the copula evolve according to the GAS model introduced in Section 2. In both cases we set the number of clusters to be $G = 10$, with an equal number of variables allocated to each cluster. In the static case, we assume that the loadings across groups on the market factor ranges from 0.25 to 2.50 in increments of 0.25, while the loadings on the group specific factors range from 2.5 to 0.25 in increments of -0.25. This implies that some groups are more influenced by the common market factor than their group factor, while the reverse is true for other groups, roughly mimicking the differences between industries like manufacturing and mining/construction. Naturally, in this case the GAS dynamic parameters $\left(\alpha^M, \beta^M, \alpha^C, \beta^C\right)$ are all zero.

In the dynamic case, we set the intercept parameters $\left(\omega_g^M, \omega_g^C\right)$ equal to 0.04 for all groups, which, combined with the common values for the GAS dynamic parameters $\left(\alpha^M, \beta^M, \alpha^C, \beta^C\right) = (0.02, 0.9, 0.02, 0.9)$, means that all groups have the same average loading on the market factor and on their group-specific factor. This homogeneity of loadings makes the estimation problem more difficult than if the loadings had different long-run averages, and is designed to further interrogate the ability of our clustering method to correctly assign variables to groups.

Table 1 presents the results for the static copula case. In Panel A we see that the estimated parameters are centered on the true values, for all three copulas, and the standard errors on the factor loadings increase slightly (on average) as we move from Gaussian to t to skew t copulas.

Panel B of Table 1 reports the striking result that in 100% of the simulations there were *zero* variables assigned to an incorrect group. That is, in every simulation the clustering algorithm was able to correctly allocate variables to their groups.[7] In the Gaussian case, the clustering step is done using the correct model (a static Gaussian copula) while in the other two cases the model used in the clustering step is misspecified. Panel B reveals that this misspecification leads to no errors in the classification of these variables.[8] This is consistent with the exponential convergence rate (see equation 15) found in other contexts for cluster assignment estimators.

Panel C of Table 1 reports the average estimation time (using a machine with 28 cores) and number of EM iterations required for convergence, and reveals no large differences in the difficulty of estimation across these models.

[INSERT TABLE 1 ABOUT HERE]

Table 2 presents the results for the dynamic copula case. We again see that the estimated parameters are centered on the true values, and in Panel B we see the remarkable result that the clustering algorithm described above is able to correctly assign every variable to its group in 100% of simulations. Recall that the estimated cluster assignments are based on a static Gaussian copula model, which is misspecified in all three cases considered in Table 2. That model is shown in Table 2 to be rich enough to reveal the true clusters in the data even though it is misspecified, confirming the discussion in Section 3.1. Panel C of Table 2 shows that the dynamic model is about 35%

[7] Recall that groups are identified only up to a re-labelling; we account for this when computing the accuracy of the estimated group assignments.

[8] The clustering algorithm is not, of course, infallible: its accuracy depends on the structure of the DGP and the data available. In situations where the clusters are close together relative to sampling variation, estimated cluster assignments will inevitably contain errors. In our realistically-calibrated simulation design, the clusters appear to be sufficiently well separated that cluster assignments can be very accurately estimated.

slower to estimate than the static model.[9]

Overall, the results in Tables 1 and 2 provide strong reassurance that the models and estimation methods proposed in Section 3 work well in finite samples, enabling us to take these to real data in the next section.

[INSERT TABLE 2 ABOUT HERE]

5 Empirical application

5.1 Data and summary statistics

We study daily equity returns over the period from January 4, 2010 to December 31, 2019, a total of $T = 2159$ trade days. Every stock that was ever a constituent of the S&P 100 index during this sample, and which traded for the full sample period, is included in the data set, yielding a total of $N = 110$ firms. A list of those firms, including their names, ticker symbols, and two-digit Standard Industrial Classification (SIC) codes, are provided in Table S1 in the supplemental appendix.

Table 3 presents summary statistics of the data and parameter estimates for the mean, variance and marginal distribution models. Panel A presents unconditional sample moments of the daily returns for each stock, and these moments are comparable to those observed in other studies. Given the skewness and kurtosis estimates reported in Panel A, our marginal distribution model combines an AR(1) for the conditional mean, GJR-GARCH(1,1) for the conditional variance, and a skewed t for the marginal distribution of the standardized residuals:

$$\begin{aligned} r_{i,t} &= \phi_{0i} + \phi_{1i} r_{i,t-1} + \epsilon_{i,t} \\ h_{i,t} &= \varpi_i + \beta_i h_{i,t-1} + \alpha_i \epsilon_{i,t-1}^2 + \kappa_i \epsilon_{i,t-1}^2 1\left\{\epsilon_{i,t-1} \leq 0\right\} \\ \frac{\epsilon_{i,t}}{\sqrt{h_{i,t}}} &\sim iid\ Skew\ t\left(\xi_i, \psi_i\right) \end{aligned}$$

where $h_{i,t}$ is the conditional variance at time t for firm i and $Skew\ t$ is the univariate skewed t

[9] The estimation of the cluster assignments takes about 92% of the total computation time; the average computation time with the true cluster assignments known is only around 0.6 hours.

distribution of Hansen (1994) with the tail parameter ξ_i and the asymmetry parameter ψ_i. Using quasi-maximum likelihood, we estimate the conditional mean and variance models, then given those estimated standardized residuals, we estimate the skewed t parameters. Panel B of Table 3 provides the estimation results of the marginal distribution model, and the values there are consistent with those reported in the empirical finance literature (see, e.g., Bollerslev, Engle, and Nelson 1994). The standardized residuals still indicate substantial skewness ($\hat{\psi} = -0.027$ on average) and kurtosis ($\hat{\xi} = 5.089$ on average). Given the marginal model parameters we obtain the probability integral transforms, u_{it}, used in the estimation of the copula.

Panel C of Table 3 presents Pearson's linear correlations and Spearman's rank correlations between those standardized residuals whose quantiles between 5% and 95% range from 0.17 to 0.49 and from 0.20 to 0.53, respectively, indicating heterogeneous pairwise dependence, and motivating our flexible factor copula specification presented in Section 2.

[INSERT TABLE 3 ABOUT HERE]

5.2 Estimated cluster assignments

We firstly use the method described in Sections 3 to estimate the group assignments for each variable. To determine the optimal number of groups, we use the BIC for the fitted static Gaussian copula model.[10] The value of the BIC for each choice of G is plotted in Figure 2, along with the values of the BIC obtained when using one-digit or two-digit SIC codes to determine group assignments. In our sample there are seven one-digit SIC groups and 21 two-digit SIC groups.[11] Figure 2 reveals that the BIC from a model using only four estimated group assignments dominates the seven one-digit SIC groups, and a model with just five estimated group assignments beats the 21-group model based on two-digit SIC codes. These rankings reveal the gains available from a

[10] The BIC is computed as $BIC\,(G) = -2\sum_{t=1}^{T} \log c\left(\mathbf{u}_t; \hat{\boldsymbol{\theta}}_T^{(G)}, \hat{\Gamma}_T^{(G)}\right) + 2G\log\,(T)$, where G denotes the number of clusters. We use the notation $(\hat{\boldsymbol{\theta}}_T^{(G)}, \hat{\Gamma}_T^{(G)})$ to emphasize that the parameters of the copula vary with G.

[11] Our model cannot accommodate groups with only one member, and when estimating with SIC-based clusters we address this by moving stocks that are a singleton in their group to the SIC group with which they have the highest correlation. Specifically, in the one-digit clustering model, Weyerhaeuser (WY) is the only stock in the one-digit SIC group 0, and we move it to SIC group 3. In the two-digit clustering model, FCX (10), NKE (30), WY (08), FDX (45) and V(61) are all singletons, and those are moved into the two-digit SIC groups 13, 37, 37, 42, and 60, respectively.

data-driven assignment of stocks to groups, rather than assignments based on SIC codes.[12]

The optimal number of estimated groups, according to the BIC, is 21, which is coincidentally the same as the number of two-digit SIC groups.[13] We note that the BIC curve is relatively flat near the optimum, indicating that choosing G between 20 and 25 leads to approximately the same fit; i.e., there is some robustness to the specific choice of G.

[INSERT FIGURE 2 ABOUT HERE]

Table 4 presents the estimated group assignments for the 110 stocks in our sample, along with each stock's SIC code. Some of the estimated groups line up closely with a two-digit SIC group. For example, the largest group (Group 1) is comprised of 13 stocks, ten of which have SIC code 28 ("Chemical & Allied Products" manufacturing). The three other stocks (Baxter, Medtronic and United Health) have different SIC codes, but are clearly broadly in the same category as the rest of this group. Group 5, as another example, looks clearly like a "Tech" group, and all but two members have SIC code 73 ("Business Services"). The two listed with other codes are Apple (listed as 35, "Industrial Machinery & Equipment" manufacturing) and Netflix (listed as 78, "Motion Pictures"). Despite the different SIC codes, most investors would agree that Apple and Netflix fit neatly in a cluster containing Google, Amazon and Ebay. Among the smaller clusters, we see some obvious pairs of stocks grouped together: AT&T and Verizon, Lowes and Home Depot; Mastercard and Visa; McDonalds and Starbucks.

Overall, the group assignments in Table 4 look economically plausible, in addition to representing a much better statistical fit according to the BIC. In Section 5.4 we conduct formal out-of-sample forecast comparison tests to determine whether the improved in-sample fit leads to significantly better out-of-sample forecasts.

[INSERT TABLE 4 ABOUT HERE]

[12] Opschoor *et al.* (2020) compare cluster assignments based on SIC codes with those based on some other common characteristics: market capitalization (size), the book-to-market ratio (value), and past returns (momentum). They find that SIC-based assignments easily dominates these alternatives.

[13] We used a set of 100 random starting values for Γ, the cluster assignment vector, in estimation, and did not use information from SIC codes at all in the EM-based model.

5.3 Estimated dependence time series

We now compare the fitted dependence time series from the two-digit SIC factor copula model and the factor copula model with estimated group assignments. We use rank correlations as a summary measure for the strength and direction of the dependence between assets implied by these models. With a fully-specified copula model such as the ones employed here, it is also possible to extract other dependence measures, such as tail dependence or probabilities of joint tail events, see e.g. the measures in Giesecke and Kim (2011) and Oh and Patton (2018).

The complete rank correlation matrix is 110×110, and even just focusing on the blocks implied by the factor structure embedded in the model the matrix is 21×21. As an initial summary measure, we firstly consider the conditional rank correlation for pairs in the same group. Figure 3 plots these for three groups, along with the two-digit SIC group that best matches the estimated group.[14,15] The top panel compares estimated group 3 with SIC group 13. We observe that the two conditional rank correlation paths track each other quite closely, but the rank correlations based on estimated group assignments appear to adjust more quickly to news, and the SIC-based estimates look somewhat like a rolling average of the path from the model with estimated group assignments. A similar picture arises in the middle panel, comparing estimated group 7 with SIC group 36. It appears that by getting group assignments that better match the data, the model is more quickly able to react to information that suggests dependence has gone up or down.

The lower panel of Figure 3 compares estimated group 9 and SIC group 49, and represents a particularly interesting comparison. Group 9 contains six members, and all of them are from SIC group 49 ("Electric, Gas, & Sanitary Services," in the "Transportation & Public Utilities" group). There is just one other SIC group 49 stock in our sample (Williams, ticker WMB), and this stock was estimated to belong to group 3, which is dominated by SIC group 13 members (SIC 13 is "Oil & Gas Extraction" in the "Mining" group). From the firm's description on its website, it conducts a mix of activities captured by these SIC labels, and it turns out that our cluster assignment

[14] Figures S1-S2 in the supplemental appendix present other comparisons of fitted rank correlations from the two models.

[15] For example, estimated Group 3 has eleven members, including all eight of the SIC group 13 stocks. Estimated Group 7 has seven members including all five members of SIC group 36. Estimated Group 9 has six members and all of them belong to SIC group 49; the single other SIC group 49 member was estimated to belong to Group 3.

algorithm estimates it to be a better match with mining firms than with utilities firms. The lower panel of Figure 3 shows that by removing just this one stock the within-group rank correlation rises from around 0.55 to around 0.68. Moreover, we again see that the conditional rank correlations are more dynamic in the model with estimated group assignments.

[INSERT FIGURE 3 ABOUT HERE]

The plots of conditional rank correlations in Figure 3 allow us to see differences in pairwise dependence implied by the two models. For a more complete depiction of the differences implied by the model in the upper panel of Figure 4 we plot the QLIKE distance measure between the full 110×110 rank correlation matrices implied by the two models.[16] When this measure is lower, the rank correlation matrices are more similar. We see that the difference is largest in mid 2011, and also large in late 2015, while it was relatively low in 2012. The middle panel of Figure 4 presents the normalized sum of the first 22 eigenvalues of the model-implied rank correlation matrices. Both of the models are based on a 22-factor model (one common factor and 21 group-specific factors), and the sum of the first 22 eigenvalues provides a summary for how informative the factors are.[17] We see that the sum is uniformly greater for the model with estimated group assignments than for the model based on SIC group assignments. Note that the period when the two sums are furthest apart corresponds to the period when the QLIKE distance is also the greatest, indicating that this is one reason for the increased QLIKE distance. The lower panel of Figure 4 plots cross-sectional dispersion in pairwise rank correlations. We see that this dispersion has been broadly increasing over the sample period, and that periods when the two models differ most in the degree of dispersion also correspond to times when the QLIKE distance is larger.

[INSERT FIGURE 4 ABOUT HERE]

[16] The QLIKE distance between two $(N \times N)$ matrices is $QLIKE(A, B) = \text{tr}\left(A^{-1}B\right) - \log\left|A^{-1}B\right| - N$.

[17] Figure S3 in the supplemental appendix presents corresponding results using just the largest eigenvalue, or the sum of the first three eigenvalues. The largest eigenvalues from each of the models are roughly equal, although similar to the pairwise rank correlation plots, the time series from the model with estimated group assignments appears more dynamic. The plot of the sum of the largest three eigenvalues reveals not only more dynamics, but a slight gap in the level, though it is not as large and not uniform as it is for the sum of the first 22 eigenvalues.

5.4 Out-of-sample forecast performance comparisons

Finally, we compare the out-of-sample (OOS) forecasts of the factor copula models using SIC-based group assignments with those using estimated group assignments. To do so, we split our sample period in half, using data from 2010 to 2014 to estimate the models, and data from 2015 to 2019 to evaluate the models. Given the computational complexity of the models, we estimate the models only once, on the last day of the in-sample period, and retain those parameters throughout the OOS period. We compare the models using their out-of-sample likelihoods, which is a consistent scoring rule for ranking density forecasts, see Gneiting and Raftery (2007). We test for the significance of the differences in OOS likelihoods using a Diebold and Mariano (1995)/Giacomini and White (2006) test, with a Newey-West (1987) estimator of the standard error based on 10 lags.

In Table 5 we use OOS forecast performance to determine the optimal shape of the copula (Gaussian, t, or skew t), as well as the optimal choice of dynamics (static vs. GAS).[18] We do this for a range of choices for the number of groups, to determine the robustness of the conclusions, and also for the two SIC-based group assignments. The left panel of Table 5 clearly indicates that including GAS dynamics in the model improves the fit: in all cases the t-statistic is positive, and the smallest t-statistic across all configurations is 6.5, indicating strong evidence in favor of the GAS model over the static model.

In the right panel of Table 5 we adopt GAS dynamics in all cases, and we compare the choice of copula shape across various choices the number of groups. We find in all cases that the t and skew t models out-perform the Gaussian factor copula, with t-statistics all greater than 7.7. This is consistent with previous work in the literature (see, e.g., Patton, 2004, 2013, and Amengual and Sentana, 2020) that the Normal copula is not a good description of equity return dependence. In the last column of Table 5 we compare the t and skew t copulas, and we find that the t-statistics are all negative, and generally significant, indicating that the estimation of the additional skewness parameter in the skew t copula leads to worse OOS performance than the symmetric t factor copula. This is in contrast with the in-sample parameter estimates (presented in Tables S2 and S3 in the

[18] In addition to being economically interesting in their own right, using OOS forecast performance to make these comparisons allows us to conduct formal statistical tests without having to make assumptions about the error rate in the cluster assignment estimation step (see Section 3.2) that cannot be verified.

supplemental appendix) where the copula asymmetry parameter is significantly negative.[19] These conflicting results can be reconciled by the fact that OOS forecast comparisons tend to carry a strong implicit penalty for estimation error, and so unless the new parameter is far from zero and precisely estimated, better forecasts may be obtained by setting it to zero.

[INSERT TABLE 5 ABOUT HERE]

In Table 6 we compare the OOS performance of t factor copulas with GAS dynamics that use different numbers of groups. Consistent with the BIC rankings of models presented in Figure 2, the model based on one-digit SIC groupings is beaten by every other model except the estimated group assignment model with only 3 groups. Similarly, the model based on two-digit SIC groupings is beaten by every other model except the estimated group assignment models with only 3 or 4 groups, and the one-digit SIC model. Amongst the models with estimated group assignments, the model with $G = 20$ groups performs the best in terms of OOS likelihoods, significantly beating every other model, including the $G = 21$ model which was selected as being optimal over the full sample according to the BIC. That the optimal model for OOS forecasting is smaller than the optimal model for in-sample fitting is consistent with the abovementioned predilection of OOS forecasts for parsimonious models.

[INSERT TABLE 6 ABOUT HERE]

6 Conclusion

This paper proposes a new dynamic factor copula model for use in high dimensional time series applications. Our model does not require variables to be grouped according to *ex ante* information, like SIC industry codes or similar; instead we estimate the optimal assignment of variables to groups from the data using a k-means type approach. Our clustering method exploits the fact that clusters

[19] These two tables present parameter estimates and standard errors for models using one-digit SIC codes (seven clusters) or using estimated cluster assignments (using the BIC-optimal number of clusters, 21). In both cases we see that the copula asymmetry parameter is significantly negative. Standard errors in this table are computed assuming that estimation error from cluster assignments is negligible.

can be estimated from a *static* version of the copula model, rather than the more computationally-challenging dynamic version that is of primary interest, making the clustering problem feasible. We show via an extensive simulation study that group assignments can be accurately estimated in finite samples. In an application to 110 U.S. equity returns over the period 2010-2019 we find evidence that a model with estimated group assignments significantly outperforms an otherwise identical model with group assignments determined using SIC codes. The improvement in fit appears to come from a better assignment of stocks that are labeled with one SIC code but comove more like stocks from a different SIC code, which allows the dynamic model to react to new information more quickly.

Appendix

A.1 The probability density of the skewed t copula

We adopt the skewed t copula discussed in Demarta and McNeil (2005) and Christoffersen *et al.* (2012). Specifically, it is the copula embedded in the multivariate skewed t distribution of $\mathbf{X}_t$, where:

$$\mathbf{X}_t = \sqrt{W_t}\mathbf{Z}_t + \boldsymbol{\zeta} W_t \tag{16}$$

where $\boldsymbol{\zeta}$ is a $N \times 1$ asymmetry parameter vector filled with an identical scalar ζ, W_t is an inverse gamma variable $W_t \sim IG\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$, $\mathbf{Z}_t$ is a $N \times 1$ normal variable $\mathbf{Z}_t \sim \mathcal{N}\left(0, \mathbf{R}_t\right)$, and W_t and $\mathbf{Z}_t$ are independent. The probability density function of the skewed t copula is given by

$$\begin{aligned} c\left(\mathbf{x}_t;\ \mathbf{R}_t, \nu, \zeta\right) \ &= \ \frac{2^{\frac{(\nu-2)(N-1)}{2}} K_{\frac{\nu+N}{2}}\left(\sqrt{\left(\nu + \mathbf{x}_t'\mathbf{R}_t^{-1}\mathbf{x}_t\right)\boldsymbol{\zeta}'\mathbf{R}_t^{-1}\boldsymbol{\zeta}}\right)\exp\left(\mathbf{x}_t'\mathbf{R}_t^{-1}\boldsymbol{\zeta}\right)}{\ddot{\Gamma}\left(\frac{\nu}{2}\right)^{1-N}\left|\mathbf{R}_t\right|^{\frac{1}{2}}\left(\sqrt{\left(\nu + \mathbf{x}_t'\mathbf{R}_t^{-1}\mathbf{x}_t\right)\boldsymbol{\zeta}'\mathbf{R}_t^{-1}\boldsymbol{\zeta}}\right)^{-\frac{\nu+N}{2}}\left(1+\frac{1}{\nu}\mathbf{x}_t'\mathbf{R}_t^{-1}\mathbf{x}_t\right)^{\frac{\nu+N}{2}}} \\ &\quad \times \prod_{i=1}^{N}\frac{\left(\sqrt{\left(\nu + x_{it}^2\right)\zeta^2}\right)^{-\frac{\nu+1}{2}}\left(1+\frac{x_{it}^2}{\nu}\right)^{\frac{\nu+1}{2}}}{K_{\frac{\nu+1}{2}}\left(\sqrt{\left(\nu + x_{it}^2\right)\zeta^2}\right)\exp\left(x_{it}\zeta\right)} \end{aligned} \tag{17}$$

where $\ddot{\Gamma}$ is the Gamma function, $K(\cdot)$ is the modified Bessel function of the third kind (also called the modified Bessel function of the second kind, or the modified Hankel function), and $\mathbf{x}_t =$

$[x_{1,t}, \cdots, x_{N,t}]' = \left[T_{skew}^{-1}(u_{1,t}; \nu, \zeta), \cdots, T_{skew}^{-1}(u_{N,t}; \nu, \zeta)\right]'$ are obtained by applying the inverse of the univariate skewed t distribution from equation (16) defined by

$$T_{skew}(y; \nu, \zeta) = \int_{-\infty}^{y} \frac{2^{1-0.5(\nu+1)} K_{\frac{\nu+1}{2}}\left(\sqrt{(\nu + x^2)\zeta^2}\right) \exp(x\zeta)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}\left(\sqrt{(\nu + x^2)\zeta^2}\right)^{-\frac{\nu+1}{2}}\left(1 + \frac{x^2}{\nu}\right)^{\frac{\nu+1}{2}}} dx.$$

Since $T_{skew}^{-1}(\cdot; \nu, \zeta)$ is not available in closed form, we generate 1,000,000 random draws from equation (16) and use linear interpolation to approximate $T_{skew}^{-1}(\cdot; \nu, \zeta)$ on $(0, 1)$. Note that $T_{skew}(\cdot; \nu, \zeta)$ is identical across the cross sectional dimension and also over time because the shape parameters of this distribution are assumed constant, this means that we can approximate T_{skew}^{-1} just once per parameter set $[\nu, \zeta]$ and apply it to all copula inputs $\{\mathbf{u}_t\}_{t=1}^{T}$, making the likelihood evaluation of the skewed t copula very fast.

A.2 Derivation of the score

From equation (17), the log-likelihood of the skewed t copula is obtained by

$$\begin{aligned} \log \mathbf{c}_{Skewt,t}(\mathbf{x}_t; \mathbf{R}_t, \nu, \zeta) &= -\frac{1}{2}\log|\mathbf{R}_t| - \frac{\nu+N}{2}\log\left(1 + \frac{1}{\nu}\mathbf{x}_t'\mathbf{R}_t^{-1}\mathbf{x}_t\right) \\ &\quad + \log\left(K_{\frac{\nu+N}{2}}\left(\sqrt{(\nu + \mathbf{x}_t'\mathbf{R}_t^{-1}\mathbf{x}_t)\boldsymbol{\zeta}'\mathbf{R}_t^{-1}\boldsymbol{\zeta}}\right)\right) \\ &\quad + \mathbf{x}_t'\mathbf{R}_t^{-1}\boldsymbol{\zeta} + \frac{\nu+N}{2}\log\left(\sqrt{(\nu + \mathbf{x}_t'\mathbf{R}_t^{-1}\mathbf{x}_t)\boldsymbol{\zeta}'\mathbf{R}_t^{-1}\boldsymbol{\zeta}}\right) + const(\nu, \zeta) \end{aligned} \tag{18}$$

where $const(\nu, \zeta)$ contains any components that do not depend on $\mathbf{R}_t$, and recall that

$$\mathbf{R}_t = \widetilde{\mathbf{L}}_t'\widetilde{\mathbf{L}}_t + \mathbf{D}_t, \quad \widetilde{\mathbf{L}}_t = \left[\widetilde{\boldsymbol{\lambda}}_{1,t}, \ldots, \widetilde{\boldsymbol{\lambda}}_{N,t}\right], \quad \mathbf{D}_t = diag\left(\sigma_{1,t}^2, \ldots, \sigma_{N,t}^2\right)$$

where

$$\widetilde{\boldsymbol{\lambda}}_{i,t} = \frac{\boldsymbol{\lambda}_{i,t}}{\sqrt{1 + \boldsymbol{\lambda}_{i,t}'\boldsymbol{\lambda}_{i,t}}}, \quad \sigma_{i,t}^2 = \frac{1}{1 + \boldsymbol{\lambda}_{i,t}'\boldsymbol{\lambda}_{i,t}}$$

To derive the score, we first define $\mathbf{L}_t = (\boldsymbol{\lambda}_{1,t}, \ldots, \boldsymbol{\lambda}_{N,t}) \in \mathbb{R}^{k \times N}$ where $\boldsymbol{\lambda}_{i,t}$ is a $k \times 1$ vector of factor loadings for variable i. In the example of equation (7), the number of factors denoted by k

is $G+1$ and if the variable i belongs to Group 3, then $\boldsymbol{\lambda}_{i,t} = [\lambda_{M,3,t}, 0, 0, \lambda_{3,3,t}, 0, \ldots, 0]'$. By the chain rule, the derivative of equation (18) with respect to $\boldsymbol{\eta}_t$ can be written as product of three factors

$$\underbrace{\frac{\partial \log \mathbf{c}_{Skewt,t}\left(\mathbf{x}_t; \mathbf{R}_t, \nu, \zeta\right)}{\partial \boldsymbol{\eta}_t'}}_{(1\times 2G)} = \underbrace{\frac{\partial \log \mathbf{c}_{Skewt,t}\left(\mathbf{x}_t; \mathbf{R}_t, \nu, \zeta\right)}{\partial vec\left(\mathbf{R}_t\right)'}}_{(1\times N^2)} \cdot \underbrace{\frac{\partial vec\left(\mathbf{R}_t\right)}{\partial vec\left(\mathbf{L}_t\right)'}}_{(N^2\times(G+1)N)} \cdot \underbrace{\frac{\partial vec\left(\mathbf{L}_t\right)}{\partial \boldsymbol{\eta}_t'}}_{((G+1)N\times 2G)} \tag{19}$$

where $vec\left(\cdot\right)$ stacks the columns of the matrix on top of one another to form a vector. The first factor of equation (19) can be written as

$$\frac{\partial \log \mathbf{c}_{Skewt,t}\left(\mathbf{x}_t; \mathbf{R}_t, \nu, \zeta\right)}{\partial vec\left(\mathbf{R}_t\right)'} = \left(vec\left(\frac{\partial \log \mathbf{c}_{Skewt,t}\left(\mathbf{x}_t; \mathbf{R}_t, \nu, \zeta\right)}{\partial \mathbf{R}_t}\right)\right)'$$

so we focus on

$$\begin{aligned}\frac{\partial \log \mathbf{c}_{Skewt,t}\left(\mathbf{x}_t; \mathbf{R}_t, \nu, \zeta\right)}{\partial \mathbf{R}_t} &= -\frac{1}{2}\cdot\frac{\partial \log\left|\mathbf{R}_t\right|}{\partial \mathbf{R}_t} - \frac{\nu+N}{2}\frac{\partial \log\left(1+\frac{1}{\nu}\mathbf{x}_t'\mathbf{R}_t^{-1}\mathbf{x}_t\right)}{\partial \mathbf{R}_t} \\ &+\frac{\partial \log\left(K_{\frac{\nu+N}{2}}\left(\sqrt{\left(\nu+\mathbf{x}_t'\mathbf{R}_t^{-1}\mathbf{x}_t\right)\boldsymbol{\zeta}'\mathbf{R}_t^{-1}\boldsymbol{\zeta}}\right)\right)}{\partial \mathbf{R}_t} \\ &+\frac{\partial \mathbf{x}_t'\mathbf{R}_t^{-1}\boldsymbol{\zeta}}{\partial \mathbf{R}_t} + \frac{\nu+N}{2}\cdot\frac{\partial \log\left(\sqrt{\left(\nu+\mathbf{x}_t'\mathbf{R}_t^{-1}\mathbf{x}_t\right)\boldsymbol{\zeta}'\mathbf{R}_t^{-1}\boldsymbol{\zeta}}\right)}{\partial \mathbf{R}_t}\end{aligned}$$

Two useful formulas from the matrix differentials are

$$\frac{d\left(\mathbf{v}'\mathbf{M}^{-1}\mathbf{w}\right)}{d\mathbf{M}} = -\left(\mathbf{M}^{-1}\right)'\mathbf{v}\mathbf{w}'\left(\mathbf{M}^{-1}\right)' \text{ and } \frac{d\log\left|\mathbf{M}\right|}{d\mathbf{M}} = \mathbf{M}^{-1}$$

where $\mathbf{M}$ is a symmetric non-singular matrix, and $\mathbf{v}$ and $\mathbf{w}$ are vectors conformable with $\mathbf{M}$. With those formulas we calculate each component separately to obtain

$$\begin{aligned}\frac{\partial \log \mathbf{c}_{Skewt,t}\left(\mathbf{x}_t; \mathbf{R}_t, \nu, \zeta\right)}{\partial \mathbf{R}_t} &= -\frac{1}{2}\mathbf{R}_t^{-1} + \frac{\nu+N}{2}\frac{\mathbf{R}_t^{-1}\mathbf{x}_t\mathbf{x}_t'\mathbf{R}_t^{-1}}{\left(\nu+\mathbf{x}_t'\mathbf{R}_t^{-1}\mathbf{x}_t\right)} + \frac{K'_{\frac{\nu+N}{2}}\left(A_t\right)}{K_{\frac{\nu+N}{2}}\left(A_t\right)}\cdot\frac{B_t}{2A_t} \\ &-\mathbf{R}_t^{-1}\mathbf{x}_t\boldsymbol{\zeta}'\mathbf{R}_t^{-1} + \left(\frac{\nu+N}{4}\right)\frac{B_t}{A_t^2}\end{aligned} \tag{20}$$

where

$$
\begin{aligned}
A_t &= \sqrt{\left(\nu + \mathbf{x}_t' \mathbf{R}_t^{-1} \mathbf{x}_t\right) \boldsymbol{\zeta}' \mathbf{R}_t^{-1} \boldsymbol{\zeta}} \\
B_t &= -\left(\nu + \mathbf{x}_t' \mathbf{R}_t^{-1} \mathbf{x}_t\right) \left(\mathbf{R}_t^{-1} \boldsymbol{\zeta}\boldsymbol{\zeta}' \mathbf{R}_t^{-1}\right) - \left(\mathbf{R}_t^{-1} \mathbf{x}_t \mathbf{x}_t' \mathbf{R}_t^{-1}\right) \left(\boldsymbol{\zeta}' \mathbf{R}_t^{-1} \boldsymbol{\zeta}\right) \\
K'_{\frac{\nu+N}{2}}(\cdot) &= -\frac{1}{2}\left[K_{\frac{\nu+N}{2}-1}(\cdot) + K_{\frac{\nu+N}{2}+1}(\cdot)\right].
\end{aligned}
$$

As $\zeta \to 0$, equation (20) boils down to the derivative of the log density of the Student t copula,

$$
\frac{\partial \log \mathbf{c}_{Student-t,t}\left(\mathbf{x}_t; \mathbf{R}_t, \nu\right)}{\partial \mathbf{R}_t} = -\frac{1}{2}\mathbf{R}_t^{-1} + \frac{\nu + N}{2} \frac{\mathbf{R}_t^{-1} \mathbf{x}_t \mathbf{x}_t' \mathbf{R}_t^{-1}}{\left(\nu + \mathbf{x}_t' \mathbf{R}_t^{-1} \mathbf{x}_t\right)}
$$

and in addition, as $\nu \to \infty$, it becomes that of the Gaussian copula,

$$
\frac{\partial \log \mathbf{c}_{Gaussian,t}\left(\mathbf{x}_t; \mathbf{R}_t\right)}{\partial \mathbf{R}_t} = -\frac{1}{2}\mathbf{R}_t^{-1} + \frac{1}{2}\mathbf{R}_t^{-1} \mathbf{x}_t \mathbf{x}_t' \mathbf{R}_t^{-1}.
$$

To derive the second and third factors in equation (19) we closely follow the logic and notations of Opschoor, *et al.* (2020). The second factor is re-written as

$$
\begin{aligned}
\frac{\partial vec\left(\mathbf{R}_t\right)}{\partial vec\left(\mathbf{L}_t\right)'} &= \frac{\partial vec\left(\widetilde{\mathbf{L}}_t' \widetilde{\mathbf{L}}_t\right)}{\partial vec\left(\widetilde{\mathbf{L}}_t\right)'} \cdot \frac{\partial vec\left(\widetilde{\mathbf{L}}_t\right)}{\partial vec\left(\mathbf{L}_t\right)'} + \frac{\partial vec\left(\mathbf{D}_t\right)}{\partial vec\left(\mathbf{L}_t\right)'} \\
&= \left(\mathbf{I}_{N^2} + J_{N,N}\right)\left(\mathbf{I}_N \otimes \widetilde{\mathbf{L}}_t'\right) \cdot \frac{\partial vec\left(\widetilde{\mathbf{L}}_t\right)}{\partial vec\left(\mathbf{L}_t\right)'} + \frac{\partial vec\left(\mathbf{D}_t\right)}{\partial vec\left(\mathbf{L}_t\right)'} \qquad (21)
\end{aligned}
$$

with another useful formula

$$
\frac{\partial vec\left(\mathbf{M}'\mathbf{M}\right)}{\partial vec\left(\mathbf{M}\right)'} = \left(\mathbf{I}_{n^2} + J_{n,n}\right)\left(\mathbf{I}_n \otimes \mathbf{M}'\right)
$$

where $\mathbf{M} \in \mathbb{R}^{m \times n}$, and $J_{m,n} \in \mathbb{R}^{mn \times mn}$ is the vectorized transpose matrix, i.e. $vec\left(\mathbf{M}'\right) =$

$J_{m,n} vec(\mathbf{M})$. Furthermore, we have

$$\frac{\partial vec\left(\widetilde{\mathbf{L}}_t\right)}{\partial vec\left(\mathbf{L}_t\right)'} = \begin{pmatrix} \mathbf{Q}_{1,t} & 0 & \cdots & 0 \\ 0 & \mathbf{Q}_{2,t} & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{Q}_{N,t} \end{pmatrix}, \quad \mathbf{Q}_{i,t} = \frac{\mathbf{I}_k}{\sqrt{1+\boldsymbol{\lambda}_{i,t}'\boldsymbol{\lambda}_{i,t}}} - \frac{\boldsymbol{\lambda}_{i,t}\boldsymbol{\lambda}_{i,t}'}{\left(1+\boldsymbol{\lambda}_{i,t}'\boldsymbol{\lambda}_{i,t}\right)^{3/2}}$$

for $i = 1, \ldots, N$. (The denominator on the final term in the above equation corrects a typo in Opschoor et al., 2020). The sparsity of $\left(\mathbf{I}_N \otimes \widetilde{\mathbf{L}}_t'\right)$ and $\partial vec\left(\widetilde{\mathbf{L}}_t\right) / \partial vec\left(\mathbf{L}_t\right)'$ simplifies product of those two factors to

$$\left(\mathbf{I}_N \otimes \widetilde{\mathbf{L}}_t'\right) \cdot \frac{\partial vec\left(\widetilde{\mathbf{L}}_t\right)}{\partial vec\left(\mathbf{L}_t\right)'} = \begin{pmatrix} \widetilde{\mathbf{L}}_t'\mathbf{Q}_{1,t} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \widetilde{\mathbf{L}}_t'\mathbf{Q}_{N,t} \end{pmatrix}. \tag{22}$$

We define T_{diag} as a $N^2 \times N$ transformation matrix such that $vec(A) = T_{diag} \cdot a$ where A is a $N \times N$ diagonal matrix with a $N \times 1$ vector a on the diagonal. Then,

$$\frac{\partial vec\left(\mathbf{D}_t\right)}{\partial vec\left(\mathbf{L}_t\right)'} = T_{diag} \cdot \frac{\partial \left[\sigma_{1,t}^2, \ldots, \sigma_{N,t}^2\right]'}{\partial vec\left(\mathbf{L}_t\right)'} = T_{diag} \cdot \begin{pmatrix} \frac{-2\boldsymbol{\lambda}_{1,t}'}{\left(1+\boldsymbol{\lambda}_{1,t}'\boldsymbol{\lambda}_{1,t}\right)^2} & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & \frac{-2\boldsymbol{\lambda}_{N,t}'}{\left(1+\boldsymbol{\lambda}_{N,t}'\boldsymbol{\lambda}_{N,t}\right)^2} \end{pmatrix}. \tag{23}$$

For the last factor in equation (19), recall that $\boldsymbol{\eta}_t$ in equation (9) is a vector of distinct factor loadings and $\eta_{t,i}$ denotes i-th element of $\boldsymbol{\eta}_t$. $\mathbf{L}_t$ is written as

$$\mathbf{L}_t = \sum_{i=1}^{2G} \eta_{t,i} \cdot S_i'$$

$$S_i = \begin{pmatrix} \delta_{1,i}\iota_{N_1} & \delta_{G+1,i}\iota_{N_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \delta_{2,i}\iota_{N_2} & \mathbf{0} & \delta_{G+2,i}\iota_{N_2} & & \vdots \\ \vdots & \vdots & & \ddots & \mathbf{0} \\ \delta_{G,i}\iota_{N_G} & \mathbf{0} & \cdots & \mathbf{0} & \delta_{2G,i}\iota_{N_G} \end{pmatrix} \in \mathbb{R}^{N\times(G+1)}$$

where ι_p is a $p \times 1$ vector filled with ones, $\delta_{i,j} = 1$ if $i = j$ and zero otherwise, N_g for $g = 1, \dots, G$ is the number of members in group g such that $N = \sum_{g=1}^{G} N_g$. Then

$$\frac{\partial vec\left(\mathbf{L}_t\right)}{\partial \boldsymbol{\eta}_t'} = \left(vec\left(S_1'\right), \quad \dots, \quad vec\left(S_G'\right), \quad vec\left(S_{G+1}'\right), \quad \dots, \quad vec\left(S_{2G}'\right) \right) \in \mathbb{R}^{(G+1)N\times 2G}. \quad (24)$$

Thus, the score expressed in equation (19) is obtained by combining equation (20), (21), (22), (23), and (24).

A.3 Variance targeting

The number of parameters to estimate in the proposed skewed t copula with GAS dynamics is $2G + 6$, and when G is large estimating all the parameters at once is not feasible. We adopt a two-step approach, so-called "variance targeting," to eliminate the need to numerically optimize over the intercept parameters $\left[\omega_1^M, \dots, \omega_G^M, \omega_1^C, \dots, \omega_G^C\right]$. Specifically, under the stationarity assumption, the unconditional expectation of all distinct factor loadings in equation (8), $\boldsymbol{\eta}_t = \left[\lambda_{M,1,t}, \dots, \lambda_{M,G,t}, \lambda_{1,1,t}, \dots, \lambda_{G,G,t}\right]'$ is

$$\bar{\boldsymbol{\eta}} = \boldsymbol{\omega} + \mathbf{B} \cdot \bar{\boldsymbol{\eta}}$$

where $\bar{\boldsymbol{\eta}} \equiv \mathbb{E}\left[\boldsymbol{\eta}_t\right]$, $\boldsymbol{\omega} \equiv \left[\omega_1^M, \dots, \omega_G^M, \omega_1^C, \dots, \omega_G^C\right]'$ and $\mathbf{B} = diag\left(\beta^M, \cdots, \beta^M, \beta^C, \cdots, \beta^C\right) \in \mathbb{R}^{2G\times 2G}$, so if $\bar{\boldsymbol{\eta}}$ is estimated from data in a first step, $\boldsymbol{\omega}$ can be replaced with $\left(I_{2G} - \mathbf{B}\right)\widehat{\bar{\boldsymbol{\eta}}}$ and only $\left[\alpha^M, \beta^M, \alpha^C, \beta^C, \nu, \zeta\right]$ are left to be estimated numerically in a second step.

Define a $(G+1)\times G$ matrix $\bar{\mathbf{L}}$ filled with elements of $\bar{\boldsymbol{\eta}}$ as

$$\bar{\mathbf{L}}' = \begin{pmatrix} \frac{\mathbb{E}[\lambda_{M,1,t}]}{v_1} & \frac{\mathbb{E}[\lambda_{1,1,t}]}{v_1} & 0 & \cdots & 0 \\ \frac{\mathbb{E}[\lambda_{M,2,t}]}{v_2} & 0 & \frac{\mathbb{E}[\lambda_{2,2,t}]}{v_2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\mathbb{E}[\lambda_{M,G,t}]}{v_G} & 0 & 0 & \cdots & \frac{\mathbb{E}[\lambda_{G,G,t}]}{v_G} \end{pmatrix}$$

where $v_i \equiv \sqrt{1+\mathbb{E}\left[\lambda_{M,i,t}\right]^2+\mathbb{E}\left[\lambda_{i,i,t}\right]^2}$, then the model implied correlation matrix of within- and across-group is obtained by $\bar{\mathbf{L}}'\bar{\mathbf{L}}$. The corresponding unconditional correlation matrix based on samples $\mathbf{x}_{it} = \Phi^{-1}\left(u_{it}\right)$ is denoted by $\hat{\Omega} \in \mathbb{R}^{G\times G}$ where the g-th diagonal element is the average correlation of any pair of variables belonging to group g and a (i,j) element is the average correlation of any pair of variables belonging to group i and group j. Then $\bar{\boldsymbol{\eta}}$ is estimated by minimizing the difference between the sample $(\hat{\Omega})$ and model-implied correlation matrices:

$$\hat{\bar{\boldsymbol{\eta}}} = \arg\min_{\bar{\boldsymbol{\eta}}} \left[\text{vech}\left(\hat{\Omega}-\bar{\mathbf{L}}'\bar{\mathbf{L}}\right)\right]' \text{vech}\left(\hat{\Omega}-\bar{\mathbf{L}}'\bar{\mathbf{L}}\right)$$

In most variance targeting applications, the estimation of the intercept can be done analytically. Here it requires numerical optimization, however it is extremely fast.

A.4 Proof of Theorem 1

Firstly, define the profile estimator

$$\tilde{\boldsymbol{\theta}}_T\left(\Gamma\right) = \arg\max_{\boldsymbol{\theta}} \frac{1}{T}\sum_{t=1}^{T} \log c\left(\mathbf{u}_t;\boldsymbol{\theta},\Gamma\right) \tag{25}$$

Assumptions 1, 2(a) and 3(a) are sufficient for $\tilde{\boldsymbol{\theta}}_T\left(\Gamma\right) \xrightarrow{p} \tilde{\boldsymbol{\theta}}^*\left(\Gamma\right)$ for each $\Gamma \in \mathcal{G}$, see White (1994, Theorem 3.5) for example. Next define the sample and population profile likelihoods as:

$$\begin{aligned} \bar{Q}_T\left(\Gamma\right) &= \frac{1}{T}\sum_{t=1}^{T} \log \mathbf{c}\left(\mathbf{u}_t;\tilde{\boldsymbol{\theta}}_T\left(\Gamma\right),\Gamma\right) \\ Q^*\left(\Gamma\right) &\equiv \mathbb{E}\left[\log \mathbf{c}\left(\mathbf{u}_t;\tilde{\boldsymbol{\theta}}^*\left(\Gamma\right),\Gamma\right)\right] \end{aligned}$$

Define the infeasible version of the sample likelihood using the population copula parameter as

$$\dot{Q}_T(\Gamma) \equiv \frac{1}{T}\sum_{t=1}^{T} \log \mathbf{c}\left(\mathbf{u}_t; \tilde{\boldsymbol{\theta}}^*(\Gamma), \Gamma\right)$$

Consider a mean-value expansion of the sample objective function:

$$\begin{aligned}
\log \mathbf{c}\left(\mathbf{u}_t; \tilde{\boldsymbol{\theta}}_T(\Gamma), \Gamma\right) &= \log \mathbf{c}\left(\mathbf{u}_t; \tilde{\boldsymbol{\theta}}^*(\Gamma), \Gamma\right) + \nabla_\theta \log \mathbf{c}\left(\mathbf{u}_t; \tilde{\boldsymbol{\theta}}^*(\Gamma), \Gamma\right)' \left(\tilde{\boldsymbol{\theta}}_T(\Gamma) - \tilde{\boldsymbol{\theta}}^*(\Gamma)\right) \quad (26) \\
&+ \frac{1}{2}\left(\tilde{\boldsymbol{\theta}}_T(\Gamma) - \tilde{\boldsymbol{\theta}}^*(\Gamma)\right)' \nabla_{\theta\theta} \log \mathbf{c}\left(\mathbf{u}_t; \ddot{\boldsymbol{\theta}}^*, \Gamma\right) \left(\tilde{\boldsymbol{\theta}}_T(\Gamma) - \tilde{\boldsymbol{\theta}}^*(\Gamma)\right) \\
\text{where} \quad \ddot{\boldsymbol{\theta}}^* &= \lambda \tilde{\boldsymbol{\theta}}_T(\Gamma) + (1-\lambda)\tilde{\boldsymbol{\theta}}^*(\Gamma) \quad \text{for some } \lambda \in [0,1]
\end{aligned}$$

Then summing the equation above over $t = 1, ..., T$ we have

$$\begin{aligned}
\bar{Q}_T(\Gamma) - \dot{Q}_T(\Gamma) &= \left(\frac{1}{T}\sum_{t=1}^{T} \nabla_\theta \log \mathbf{c}\left(\mathbf{u}_t; \tilde{\boldsymbol{\theta}}^*(\Gamma), \Gamma\right)\right)' \left(\tilde{\boldsymbol{\theta}}_T(\Gamma) - \tilde{\boldsymbol{\theta}}^*(\Gamma)\right) + \quad (27) \\
&+ \frac{1}{2}\left(\tilde{\boldsymbol{\theta}}_T(\Gamma) - \tilde{\boldsymbol{\theta}}^*(\Gamma)\right)' \frac{1}{T}\sum_{t=1}^{T} \nabla_{\theta\theta} \log \mathbf{c}\left(\mathbf{u}_t; \ddot{\boldsymbol{\theta}}^*, \Gamma\right) \left(\tilde{\boldsymbol{\theta}}_T(\Gamma) - \tilde{\boldsymbol{\theta}}^*(\Gamma)\right)
\end{aligned}$$

Assumption 2(b) and 3(a) imply that $\frac{1}{T}\sum_{t=1}^{T} \nabla_\theta \log c\left(\mathbf{u}_t; \tilde{\boldsymbol{\theta}}^*(\Gamma), \Gamma\right) \xrightarrow{p} 0$, as usual for M-estimation. Assumption 3(c) ensures the Hessian term has a finite limit. Thus we have $\bar{Q}_T(\Gamma) - \dot{Q}_T(\Gamma) = o_p(1)$. Further, by Assumptions 1 and 2(a) we also have $\dot{Q}_T(\Gamma) - Q^*(\Gamma) = o_p(1)$, and so we have $\bar{Q}_T(\Gamma) - Q^*(\Gamma) = o_p(1)$. Thus the sample objective function is pointwise (in Γ) consistent for the population objective function. Since the parameter space is discrete, uniform convergence simplifies to pointwise convergence (see, e.g., Choirat and Seri, 2012). This implies the estimator obtained by maximizing the sample objective function is consistent for the parameter that maximizes the population objective function, i.e., $\hat{\Gamma}_T \xrightarrow{p} \Gamma_0$ as $T \to \infty$, completing the proof.

References

[1] Amengual D. and E. Sentana, 2020, Is a Normal Copula the Right Copula?, *Journal of Business & Economic Statistics*, 38(2), 350-366.

[2] Ang, A. and J. Chen, 2002, Asymmetric Correlations of Equity Portfolios, *Journal of Financial Economics*, 63, 443-494

[3] Bester, A. and C. Hansen, 2016, Grouped Effects Estimators in Fixed Effects Models, *Journal of Econometrics*, 190(1), 197-208.

[4] Blasques, F., A. Lucas and S. J. Koopman, 2014, Stationarity and Ergodicity Conditions for Generalized Autoregressive Score Processes, *Electronic Journal of Statistics*, 8, 1088-1112.

[5] Bollerslev, T., R. F. Engle, and D. B. Nelson, 1994, Chapter 49 ARCH Models, *Handbook of Econometrics*, *Vol 4*, Elsevier.

[6] Bonhomme, S., and Manresa, E., 2015, Grouped patterns of heterogeneity in panel data, *Econometrica*, 83(3), 1147-1184.

[7] Carassco, M. and X. Chen, 2002, Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models, *Econometric Theory*, 18(1), 17-39.

[8] Choirat, C. and R. Seri, 2012, Estimation in Discrete Parameter Models, *Statistical Science*, 27(2), 278-293.

[9] Christoffersen, P., K. Jacobs, X. Jin and H. Langlois, 2018, Dynamic Dependence and Diversification in Corporate Credit, *Review of Finance*, 22(2), 521-560.

[10] Creal, D.D. and R. Tsay, 2015, High-dimensional Dynamic Stochastic Copula Models, *Journal of Econometrics*, 189(2), 335-345.

[11] Creal, D.D., S.J. Koopman, and A. Lucas, 2013, Generalized Autoregressive Score Models with Applications, *Journal of Applied Econometrics*, 28(5), 777-795.

[12] Demarta, S. and A. J. McNeil, 2005, The t Copula and Related Copulas. International Statistical Review, 73(1), 111-129.

[13] Diebold, F.X. and R.S. Mariano, 1995, Comparing Predictive Accuracy, *Journal of Business & Economic Statistics*, 13(3), 253–263.

[14] Engle R.F., 2002, Dynamic conditional correlation: A simple class of multivariate GARCH models, *Journal of Business & Economic Statistics*, 20, 339-350.

[15] Fan, J., Y. Fan and J. Lv, 2008, High dimensional covariance matrix estimation using a factor model, *Journal of Econometrics*, 147, 186-197.

[16] Fan, J., Y. Liao and M. Micheva, 2013, Large Covariance Estimation by Thresholding Principal Orthogonal Complements, *Journal of Royal Statistical Society Series B*, 75, 603-680.

[17] Francis, N., M.T. Owyang, and Ö. Savascin, 2017, An endogenously clustered factor approach to international business cycles, *Journal of Applied Econometrics*, 32, 1261-1276.

[18] Giacomini, R. and H. White, 2006, Tests of conditional predictive ability, *Econometrica*, 74, 1545-1578.

[19] Giesecke, K. B. and Kim, 2011, Systemic Risk: What Defaults are Telling Us, *Management Science*, 57, 1387-1405.

[20] Gneiting, T. and A.E. Raftery, 2007, Strictly Proper Scoring Rules, Prediction and Estimation, *Journal of the American Statistical Association*, 102, 358-378.

[21] Hafner, C.M. and A. Preminger, 2009, On Asymptotic Theory for Multivariate GARCH Models, *Journal of Multivariate Analysis*, 100, 2044-2054.

[22] Hahn, J. and R. Moon, 2010, Panel Data Models with Finite Number of Equilibria, *Econometric Theory*, 26(3), 863-881.

[23] Hansen, B. E., 1994, Autoregressive Conditional Density Estimation, *International Economic Review*, 35, 705–730.

[24] Harvey, A.C., 2013, *Dynamic Models for Volatility and Heavy Tails*, Econometric Society Monograph 52, Cambridge University Press, Cambridge.

[25] Hastie, T., R. Tibshirani and J. Friedman, 2009, *The Elements of Statistical Learning: Data Mining, Inference and Prediction, Second Edition*, Springer, New York.

[26] Hautsch, N., L.M. Kyj and R.C.A. Oomen, 2012, A blocking and regularization approach to high dimensional realized covariance estimation, *Journal of Applied Econometrics*, 27, 625-645, 2012.

[27] Hong, Y., J. Tu and G. Zhou, 2007, Asymmetries in Stock Returns: Statistical Tests and Economic Evaluation, *Review of Financial Studies*, 20, 1547-1581.

[28] Lin, C.-C. and S. Ng, 2012, Estimation of Panel Data Models with Parameter Heterogeneity when Group Membership is Unknown, *Journal of Econometric Methods*, 1(1), 42-55.

[29] Newey, W.K. and D. McFadden, 1993, Large Sample Estimation and Hypothesis Testing, in R.F. Engle and D. McFadden (eds.), *Handbook of Econometrics, Vol. IV*, Elsevier.

[30] Newey, W.K. and K.D. West, 1987, A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica*, 55(3), 703-708.

[31] Oh, D.H. and A.J. Patton, 2017, Modelling Dependence in High Dimensions with Factor Copulas, *Journal of Business & Economic Statistics*, 35(1), 139-154.

[32] Oh, D.H. and A.J. Patton, 2018, Time-Varying Systemic Risk: Evidence from a Dynamic Copula Model of CDS Spreads, *Journal of Business & Economic Statistics*, 36(2), 181-195.

[33] Opschoor, A., A. Lucas, I. Barra & D. van Dijk, 2020, Closed-form multi-factor copula models with observation-driven dynamic factor loadings, *Journal of Business & Economic Statistics*, forthcoming.

[34] Patton, A.J., 2004, On the Out-of-Sample Importance of Skewness and Asymmetric Dependence for Asset Allocation, 2004, *Journal of Financial Econometrics*, 2(1), 130-168.

[35] Patton, A.J., 2013, Copula Methods for Forecasting Multivariate Time Series, in G. Elliott and A. Timmermann (eds.) *Handbook of Economic Forecasting, Vol 2*, Springer Verlag.

[36] Patton, A.J. and B.M. Weller, 2019, Risk price variation: The missing half of empirical asset pricing, working paper, Duke University.

[37] Su, L., Z. Shi, and P.C.B. Phillips, 2016, Identifying Latent Structures in Panel Data, *Econometrica*, 6(84), 2215-2264.

[38] Su, L. X. Wang and S. Jin, 2019, Sieve Estimation of Time-Varying Panel Data Models With Latent Structures, *Journal of Business & Economic Statistics*, 37(2), 334-349.

[39] Tao, M., Y. Wang, Q. Yao and J. Zou, 2011, Large Volatility Matrix Inference via Combining Low-Frequency and High-Frequency Approaches, *Journal of the American Statistical Association*, 106(495), 1025-1040.

[40] Vogt, M. and O. Linton, 2020, Multiscale Clustering of Nonparametric Regression Curves, *Journal of Econometrics*, 216(1), 305-325.

[41] White, H., 1994, *Estimation, Inference and Specification Analysis*, Econometric Society Monographs No. 22, Cambridge University Press, USA.

Table 1: Simulation results for static copulas

Panel A: Parameter estimation accuracy

		Gaussian		***t***		**skew *t***	
	True	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
β_1^M	0.25	0.245	0.076	0.253	0.089	0.276	0.062
β_2^M	0.50	0.508	0.071	0.509	0.075	0.473	0.068
β_3^M	0.75	0.755	0.053	0.744	0.056	0.740	0.066
β_4^M	1.00	1.000	0.052	1.002	0.048	0.954	0.189
β_5^M	1.25	1.246	0.033	1.243	0.035	1.246	0.040
β_6^M	1.50	1.500	0.021	1.498	0.034	1.501	0.035
β_7^M	1.75	1.747	0.020	1.752	0.027	1.755	0.036
β_8^M	2.00	2.001	0.021	1.998	0.027	1.996	0.030
β_9^M	2.25	2.253	0.019	2.252	0.028	2.255	0.031
β_{10}^M	2.50	2.496	0.020	2.502	0.029	2.502	0.030
β_1^C	2.50	2.499	0.026	2.495	0.030	2.503	0.031
β_2^C	2.25	2.248	0.023	2.246	0.031	2.251	0.036
β_3^C	2.00	2.001	0.031	2.003	0.030	1.998	0.038
β_4^C	1.75	1.749	0.030	1.747	0.034	1.771	0.065
β_5^C	1.50	1.497	0.030	1.501	0.029	1.497	0.031
β_6^C	1.25	1.246	0.028	1.252	0.026	1.252	0.030
β_7^C	1.00	1.005	0.023	0.999	0.023	1.003	0.019
β_8^C	0.75	0.756	0.022	0.749	0.023	0.743	0.018
β_9^C	0.50	0.504	0.019	0.499	0.023	0.499	0.026
β_{10}^C	0.25	0.241	0.039	0.234	0.048	0.230	0.050
ν	5.00			5.007	0.069	4.858	0.369
ζ	-0.10					-0.095	0.015

Panel B: Group assignment estimation accuracy

	Gaussian	t	skew t
Number incorrect			
0	100	100	100
≥1	0	0	0

Panel C: Estimation details

	Gaussian	t	skew t
Time (hours)	4.74	4.74	4.72
EM iterations	87.61	87.72	87.76

Notes: This table presents results from 100 simulations from a static Gaussian, *t*, and skew *t* factor copula, with 10 groups. Panel A presents results on estimation accuracy of the copula parameters, Panel B presents results on estimation accuracy of the group assignments, and Panel C presents average estimation time based on a machine with 28 cores.

Table 2: Simulation results for time-varying copulas

Panel A: Parameter estimation accuracy

		Gaussian		***t***		**skew *t***	
	True	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
ω_1^M	0.04	0.042	0.007	0.042	0.007	0.044	0.008
ω_2^M	0.04	0.042	0.007	0.042	0.007	0.044	0.008
ω_3^M	0.04	0.042	0.007	0.042	0.007	0.043	0.007
ω_4^M	0.04	0.042	0.007	0.042	0.007	0.044	0.008
ω_5^M	0.04	0.042	0.007	0.042	0.007	0.043	0.007
ω_6^M	0.04	0.042	0.007	0.042	0.007	0.044	0.008
ω_7^M	0.04	0.042	0.006	0.042	0.007	0.043	0.008
ω_8^M	0.04	0.041	0.007	0.041	0.007	0.044	0.008
ω_9^M	0.04	0.042	0.007	0.042	0.007	0.044	0.008
ω_{10}^M	0.04	0.042	0.007	0.042	0.007	0.044	0.008
ω_1^C	0.04	0.043	0.007	0.043	0.007	0.042	0.007
ω_2^C	0.04	0.043	0.007	0.043	0.008	0.042	0.007
ω_3^C	0.04	0.042	0.007	0.043	0.007	0.042	0.007
ω_4^C	0.04	0.043	0.007	0.044	0.008	0.042	0.007
ω_5^C	0.04	0.043	0.007	0.043	0.008	0.041	0.007
ω_6^C	0.04	0.043	0.007	0.043	0.007	0.042	0.007
ω_7^C	0.04	0.043	0.008	0.043	0.007	0.041	0.006
ω_8^C	0.04	0.043	0.007	0.043	0.008	0.042	0.006
ω_9^C	0.04	0.043	0.008	0.043	0.007	0.042	0.007
ω_{10}^C	0.04	0.043	0.007	0.043	0.006	0.041	0.007
α^M	0.02	0.020	0.002	0.020	0.002	0.020	0.002
β^M	0.90	0.894	0.015	0.894	0.014	0.893	0.017
α^C	0.02	0.020	0.002	0.020	0.002	0.020	0.002
β^C	0.90	0.896	0.016	0.894	0.016	0.898	0.014
ν	5.00			5.014	0.071	5.016	0.108
ζ	-0.10					-0.100	0.007

Panel B: Group assignment estimation accuracy

	Gaussian	*t*	skew *t*
Number incorrect			
0	100	100	100
$\geq$1	0	0	0

Panel C: Estimation details

	Gaussian	*t*	skew *t*
Time (hours)	6.15	6.48	6.64
EM iterations	93.8	95.8	95.9

Notes: This table presents results from 100 simulations from a Gaussian, t, and skew t factor copula with GAS dynamics, each with 10 groups. Panel A presents results on estimation accuracy of the copula parameters, Panel B presents results on estimation accuracy of the group assignments, and Panel C presents average estimation time based on a machine with 28 cores.

Table 3: Summary statistics

	Cross-sectional distribution					
	Mean	5%	25%	Median	75%	95%
Panel A: Marginal moments						
Mean	0.001	0.000	0.001	0.001	0.001	0.001
Std	0.016	0.010	0.012	0.015	0.018	0.023
Skewness	-0.081	-0.748	-0.310	-0.091	0.092	0.648
Kurtosis	9.939	5.154	6.411	8.087	10.923	22.803
Panel B: Marginal model parameters						
Constant	0.001	0.000	0.001	0.001	0.001	0.001
AR(1)	-0.019	-0.068	-0.041	-0.017	0.000	0.031
$\varpi \times 10^4$	0.009	0.002	0.003	0.006	0.011	0.025
α	0.025	0.000	0.009	0.019	0.033	0.077
κ	0.099	0.029	0.064	0.095	0.131	0.179
β	0.885	0.756	0.864	0.904	0.932	0.958
ξ	5.089	3.401	4.234	4.846	5.798	7.256
ψ	-0.027	-0.087	-0.051	-0.025	-0.004	0.020
Panel C: Correlations of standardized residuals						
Pearson	0.322	0.170	0.256	0.314	0.378	0.492
Spearman	0.360	0.197	0.295	0.356	0.418	0.531

Notes: This table presents summary statistics on the 110 daily equity return series used in this paper. The sample period is January 2010 to December 2019. Panel A presents a summary of the cross-sectional distribution of the first four moments of these returns, Panel B presents a summary of the estimated AR(1)-GJR GARCH(1,1)-skew t model used for the marginal distributions, and Panel C presents a summary of the 5,995 pairwise correlations of the standardized residuals.

Table 4: Estimated group assignments

Group	Ticker	Name	SIC	Group	Ticker	Name	SIC
1	ABT	Abbott Lab.	28	7	CSCO	Cisco Sys	36
	AGN	Actavis	28		HPQ	Hewlett Pac	35
	AMGN	Amgen	28		INTC	Intel	36
	BAX	Baxter	38		MSFT	Microsoft	73
	BIIB	Biogen	28		NVDA	Nvidia	36
	BMY	Bristol-Myers	28		QCOM	Qualcomm	36
	GILD	Gilead	28		TXN	Texas Instru	36
	JNJ	Johnson & J	28				
	LLY	Lilly Eli	28	8	AIG	Ame Inter Group	63
	MDT	Medtronic	38		ALL	Allstate	63
	MRK	Merck	28		CMCSA	Comcast	48
	PFE	Pfizer	28		DIS	Disney Walt	48
	UNH	Unitedhealth	63		F	Ford	37
					GE	Gen Electric	35
2	BAC	Bank Of Am	60		XRX	Xerox	35
	BK	Bank Of NY	60				
	C	Citigroup Inc	60	9	AEP	Ame Elec Pow	49
	COF	Capital One	60		DUK	Duke Energy	49
	GS	Goldman Sachs	62		ETR	Entergy Corp	49
	JPM	Jpmorgan	60		EXC	Exelon	49
	MET	Metlife	63		NEE	Nextera Energy	49
	MS	Morgan Stanley	60		SO	Southern Co	49
	RF	Regions Fin	60				
	USB	U S Bancorp	60	10	COST	Costco	53
	WFC	Wells Fargo	60		CVS	C V S Health	59
					TGT	Target	53
3	APA	Apache	13		WBA	Walgreens	59
	BHI	Baker Hughes	35		WMT	Walmart	53
	COP	Conocophillips	13				
	CVX	Chevron	13	11	GD	Gen Dynamics	37
	DVN	Devon	13		LMT	Lockheed Martin	37
	HAL	Halliburton	13		RTN	Raytheon	38
	NOV	Nat. Oilwell	35				
	OXY	Occidental	13	12	AMT	American Tower	48
	SLB	Schlumberger	13		SPG	Simon Property	67
	WMB	Williams Co	49		WY	Weyerhaeuser	8
	XOM	Exxon Mobil	13				
				13	BA	Boeing	37
4	CAT	Caterpillar	35		FCX	Freeport Mcmo	10
	EMR	Emerson Ele	35		NKE	Nike	30
	FDX	Fedex	45				
	HON	Honeywell Int	37	14	ACN	Accenture	67
	MMM	3M	38		IBM	IBM	35
	NSC	Norfolk South	40		ORCL	Oracle	73
	UNP	Union Pacific	40				
	UPS	United Parcel	42	15	AXP	Amex	60
					BLK	Blackrock	62
5	AAPL	Apple	35				
	ADBE	Adobe	73	16	DHR	Danaher	38
	AMZN	Amazon	73		TMO	Thermo Fisher	38
	CRM	Salesforce	73				
	EBAY	Ebay	73	17	T	A T & T	48
	GOOGL	Google	73		VZ	Verizon	48
	NFLX	Netflix	78				
	PCLN	Priceline	73	18	AVP	Avon Products	28
					SNS	Steak N Shake	58
6	CL	Colgate Palmo	28				
	CPB	Campbell Soup	20	19	MA	Mastercard	73
	KO	Coca Cola	20		V	Visa	61
	MDLZ	Mondelez	20				
	MO	Altria	21	20	MCD	Mcdonalds	58
	PEP	Pepsico	20		SBUX	Starbucks	58
	PG	Procter Gamble	28				
	PM	Philip Morris	21	21	HD	Home Depot	52
					LOW	Lowes	52

Notes: This table presets the estimated group assignments based on the BIC-optimal number of groups, G=21. The groups are ordered by the number of members.

Table 5: Comparing different copula specifications

	Static vs. GAS			Copula shape		
	Gaussian	t	skew t	G vs. t	G vs. skew t	t vs. skew t
SIC 1 digit	7.861	12.067	11.552	9.288	8.596	-2.975
SIC 2 digit	9.887	15.730	16.501	8.938	8.465	-2.849
3 groups	6.528	6.636	6.761	8.339	7.711	-2.257
4 groups	7.681	10.059	9.412	9.121	8.381	-2.351
5 groups	7.548	10.804	10.913	9.236	9.088	-1.717
18 groups	10.571	16.052	15.711	9.295	7.945	-3.550
19 groups	9.806	15.457	15.259	9.553	8.847	-2.065
20 groups	10.908	16.193	14.741	9.426	7.995	-3.797
21 groups	10.916	16.626	17.321	9.474	8.670	-2.366
22 groups	10.732	16.891	16.802	9.688	8.962	-2.714
25 groups	11.817	19.001	19.140	9.475	8.569	-2.355
27 groups	10.725	15.836	15.794	9.591	8.951	-1.676
30 groups	10.917	17.169	15.917	9.697	8.717	-2.962

Notes: This table presents Diebold-Mariano t-statistics on pairwise comparisons of models using their out-of-sample log-likelihood. The left panel compares models assuming no dynamics with those using GAS dynamics, for three different copula shapes (Gaussian, t, and skew t) and for a variety of choices for the number of groups. The right panel compares the different copula shapes, using GAS dynamics in all cases, across a variety of choices for the number of groups. In a comparison labeled "A vs B," a positive t-statistic indicates that B is preferred; a negative t-statistic indicates that A is preferred. Note that there are 7 groups of firms using the 1-digit SIC, and 21 groups using the 2-digit SIC.

Table 6: Comparing different numbers of clusters

	SIC-1	SIC-2	3 groups	4 groups	5 groups	18 groups	19 groups	20 groups	21 groups	22 groups	30 groups
Panel A: t-statistics from pair-wise comparisons											
SIC-1		26.174	-4.783	14.180	21.898	33.495	32.074	33.127	33.292	32.973	29.975
SIC-2	-26.174		-21.579	-8.472	2.530	23.186	21.202	23.288	23.359	21.255	17.075
3 groups	4.783	21.579		18.347	24.175	31.398	30.848	30.926	31.247	31.529	28.722
4 groups	-14.180	8.472	-18.347		12.922	27.064	25.292	26.950	26.957	26.511	21.562
5 groups	-21.898	-2.530	-24.175	-12.922		23.175	20.562	22.535	22.689	20.069	15.187
18 groups	-33.495	-23.186	-31.398	-27.064	-23.175		-6.779	2.377	0.184	-6.754	-12.345
19 groups	-32.074	-21.202	-30.848	-25.292	-20.562	6.779		7.596	6.545	-3.270	-9.828
20 groups	-33.127	-23.288	-30.926	-26.950	-22.535	-2.377	-7.596		-2.915	-7.892	-13.062
21 groups	-33.292	-23.359	-31.247	-26.957	-22.689	-0.184	-6.545	2.915		-7.315	-12.798
22 groups	-32.973	-21.255	-31.529	-26.511	-20.069	6.754	3.270	7.892	7.315		-7.736
30 groups	-29.975	-17.075	-28.722	-21.562	-15.187	12.345	9.828	13.062	12.798	7.736	
Panel B: Out-of-sample log-likelihood values											
$\log \mathcal{L}$	34074.5	37887.6	33175.8	36353.3	38303.6	42041.2	41564.0	42145.5	42050.1	41276.9	40559.2

Notes: This table presents Diebold-Mariano t-statistics on pairwise comparisons of models using their out-of-sample log-likelihood. In all cases we use a t copula with GAS dynamics. A positive t-statistic indicates that the column model is preferred to the row model; a negative t-statistic indicates the opposite. Note that there are 7 groups of firms using the 1-digit SIC, and 21 groups using the 2-digit SIC.

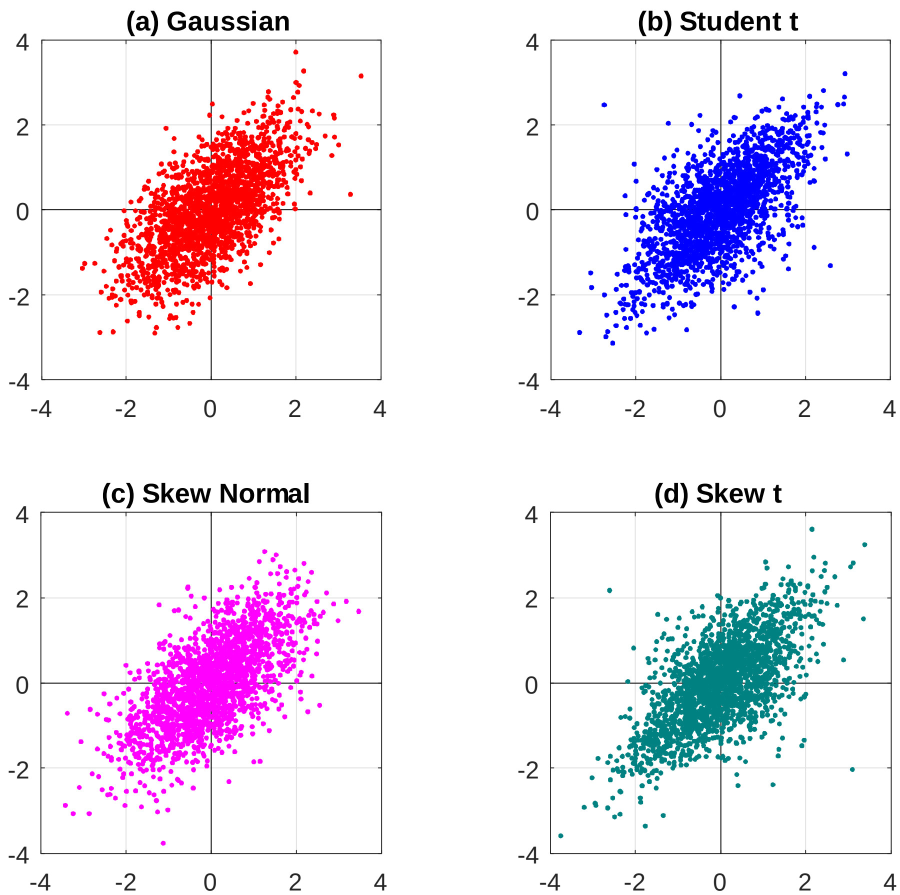

Figure 1: *This figure presents random draws from four joint distributions, all with standard Normal margins. Panel (a) uses a Gaussian copula, Panel (b) uses a Student's t copula, Panel (c) uses a skew Normal copula, Panel (d) uses a skew t copula. For all four copulas the correlation parameter is set to 0.5. For both t copulas the degrees of freedom parameter is set to 5. For both skewed copulas the skewness parameter is set to -0.1.*

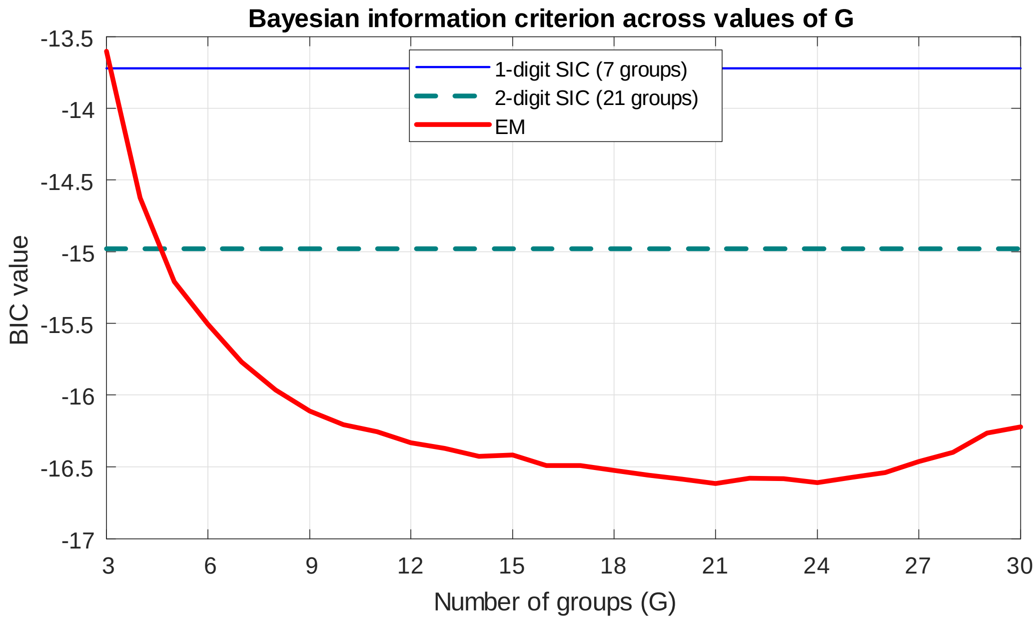

Figure 2: *Plot of BIC value as a function of the number of groups (G) for the EM-estimated model. The BIC values for the 1-digit and 2-digit SIC-based groups are also reported for comparison; these models have 7 and 21 groups respectively. As usual, lower BIC values are preferred. (Note the y-axis has been scaled by* 10^{-4} *for ease of presentation.)*

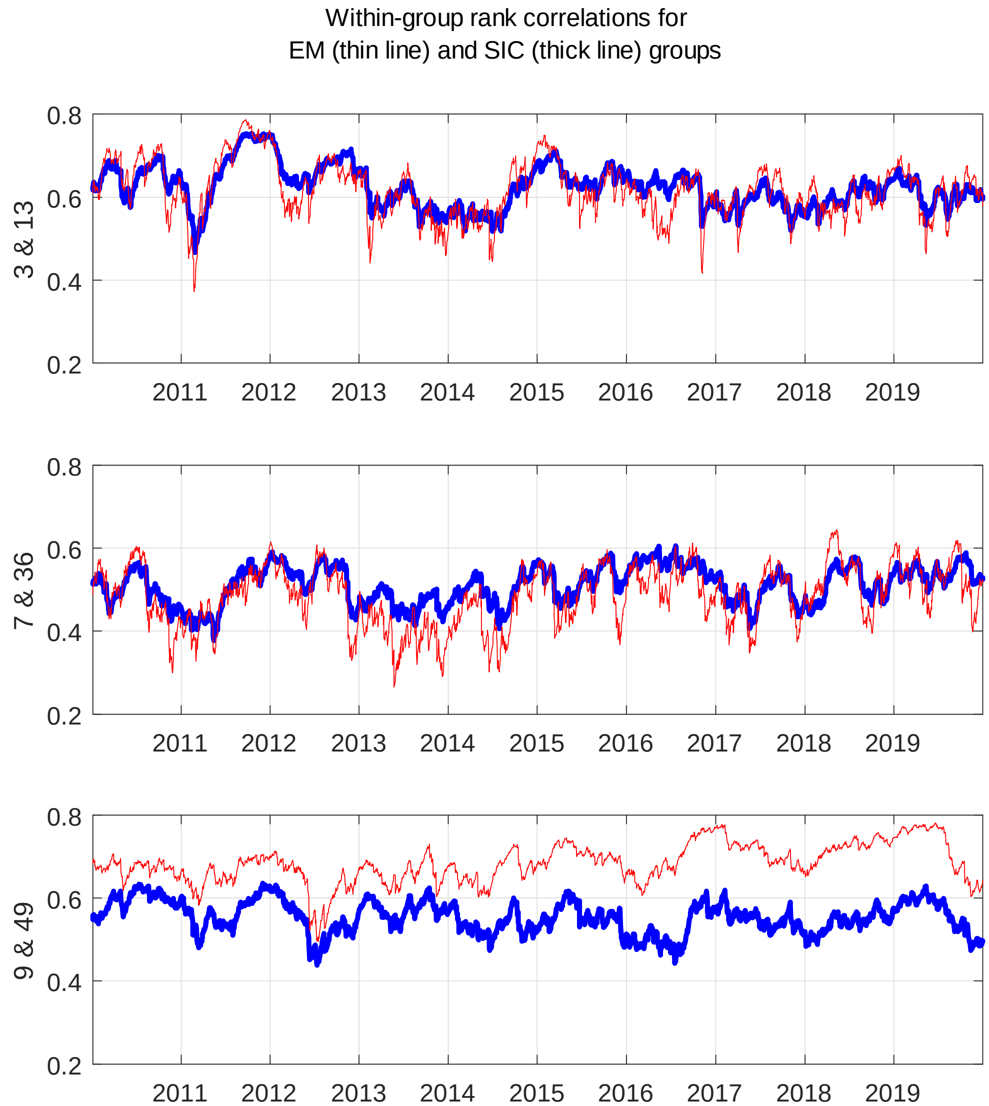

Figure 3: *Time series plots of model-implied within-group rank correlations. The upper panel presents estimated group 3 and SIC group 13; the middle panel presents estimated group 7 and SIC group 36; the lower panel presents estimated group 9 and SIC group 49.*

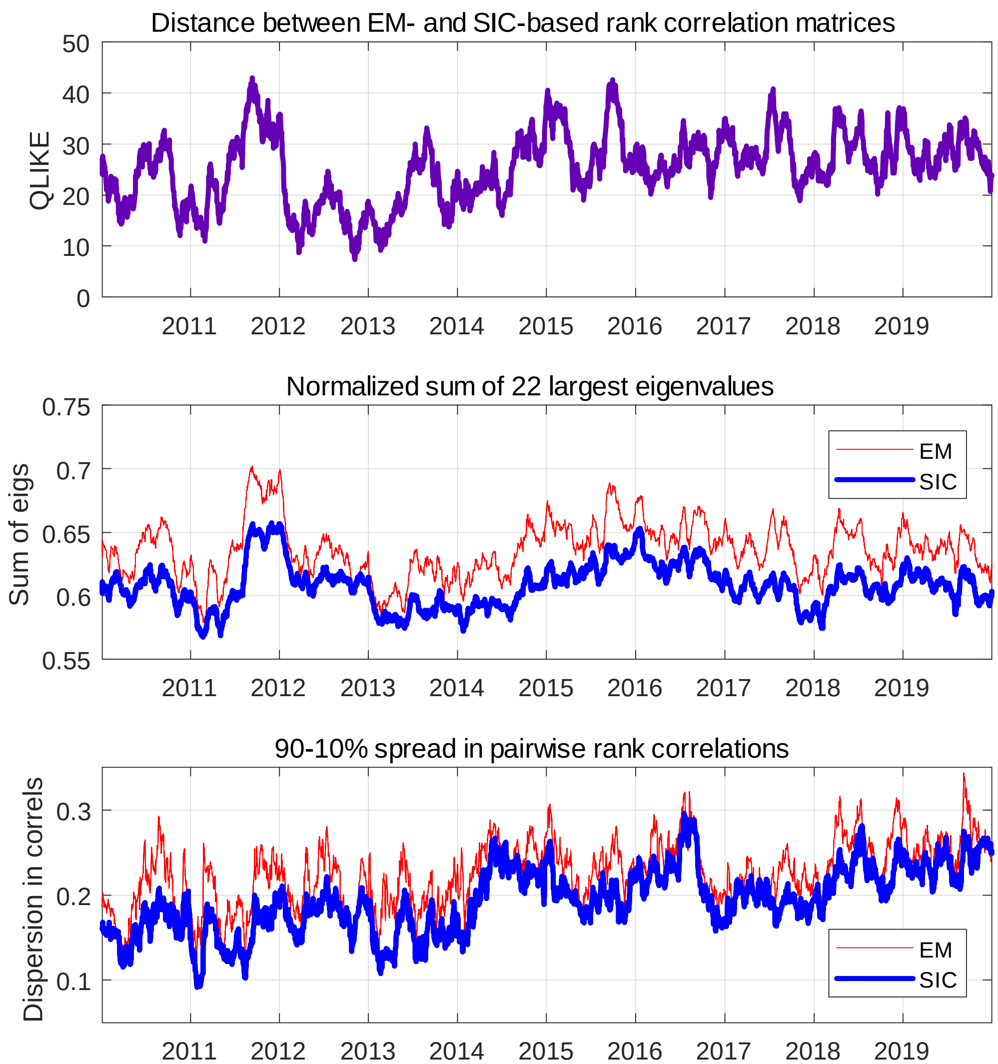

Figure 4: *The upper panel presents the QLIKE distance between the conditional rank correlation matrices implied by the 2-digit SIC-based model and the optimal EM-based factor copula model, both of which have a total of 22 factors. The middle panel presents the sum of the 22 largest eigenvalues of the conditional rank correlation matrices, divided by 110, the number of assets. The lower panel presents the difference between the 90% and 10% cross-sectional quantiles of all 5,995 pairwise rank correlations.*