...

Supplemental Appendix for

Modelling Dependence in High Dimensions with Factor Copulas

Dong Hwan Oh and Andrew J. Patton
Federal Reserve Board Duke University

18 May 2015

Appendix S.A.1: Proofs of propositions

Proof of Proposition 1. Consider a simple case first: $\beta_1 = \beta_2 = \beta > 0$. This implies that X_1 and X_2 have the same distribution function G, and so we can use the same threshold for both X_1 and X_2. Then the upper tail dependence coefficient is:

$$\tau^U = \lim_{s\to\infty} \frac{\Pr\left[X_1 > s, X_2 > s\right]}{\Pr\left[X_1 > s\right]}.$$

From standard extreme value theory, see Feller (1970) and Embrechts *et al.* (1997) for example, we have the probability of an exceedence by the sum as the sum of the probabilities of an exceedence by each component of the sum, as the exceedence threshold diverges:

$$\Pr\left[X_i > s\right] = \Pr\left[\beta Z + \varepsilon_i > s\right] \sim s^{-\alpha}\left(A_z^U \beta^\alpha + A_\varepsilon^U\right).$$

Further, we have the probability of two sums of variables both exceeding some diverging threshold being driven completely be the common component of the sums:

$$\Pr\left[X_1 > s, X_2 > s\right] = \Pr\left[\beta Z + \varepsilon_1 > s, \beta Z + \varepsilon_2 > s\right] \sim s^{-\alpha} A_z^U \beta^\alpha.$$

And so we obtain:

$$\tau^U = \lim_{s\to\infty} \frac{s^{-\alpha} A_z^U \beta^\alpha}{s^{-\alpha}\left(A_z^U \beta^\alpha + A_\varepsilon^U\right)} = \frac{A_z^U \beta^\alpha}{A_z^U \beta^\alpha + A_\varepsilon^U}.$$

(a) Now we consider the case that $\beta_1 \neq \beta_2$, and *wlog* assume $\beta_2 > \beta_1 > 0$. This complicates the problem as the thresholds, s_1 and s_2, must be set such that $G_1(s_1) = G_2(s_2) = q \to 1$, and when $\beta_1 \neq \beta_2$ we have $G_1 \neq G_2$ and so $s_1 \neq s_2$. We can find the link between the thresholds as follows:

$$\Pr[X_i > s] \sim s_i^{-\alpha}\left(A_z^U \beta_i^\alpha + A_\varepsilon^U\right),$$

and we require (s_1, s_2) such that $s_1^{-\alpha}\left(A_z^U \beta_1^\alpha + A_\varepsilon^U\right) = s_2^{-\alpha}\left(A_z^U \beta_2^\alpha + A_\varepsilon^U\right)$. This implies:

$$s_2 = s_1 \left(\frac{A_z^U \beta_2^\alpha + A_\varepsilon^U}{A_z^U \beta_1^\alpha + A_\varepsilon^U}\right)^{1/\alpha}$$

Note that s_1 and s_2 diverge at the same rate. Further note that since $\beta_2 > \beta_1$, straightforward calculations imply that $s_1/\beta_1 > s_2/\beta_2$, which is used below. The numerator of the tail dependence coefficient is:

$$\Pr[X_1 > s_1, X_2 > s_2] \sim \Pr[Z > \max\{s_1/\beta_1, s_2/\beta_2\}] = \Pr[Z > s_1/\beta_1] = s_1^{-\alpha} A_z^U \beta_1^\alpha.$$

Using either $\Pr[X_1 > s_1]$ or $\Pr[X_2 > s_2]$ in the denominator we obtain:

$$\tau^U = \frac{\beta_1^\alpha A_z^U}{\beta_1^\alpha A_z^U + A_\varepsilon^U}.$$

(b) Say $\beta_2 < \beta_1 < 0$. Then similar to part (a) we obtain:

$$\Pr[X_i > s] = \Pr[\beta_i Z + \varepsilon_i > s] \sim s^{-\alpha}\left(A_z^L |\beta_i|^\alpha + A_\varepsilon^U\right)$$

Next we find the thresholds (s_1, s_2) such that $\Pr[X_1 > s_1] = \Pr[X_2 > s_2]$, which yields

$$s_2 = s_1 \left(\frac{A_z^L |\beta_2|^\alpha + A_\varepsilon^U}{A_z^L |\beta_1|^\alpha + A_\varepsilon^U}\right)^{1/\alpha}$$

Using the same steps as for part (a), we find that $s_1/|\beta_1| > s_2/|\beta_2|$. Thus the numerator becomes:

$$\Pr[X_1 > s_1, X_2 > s_2] \sim \Pr[(-Z) > \max\{s_1/|\beta_1|, s_2/|\beta_2|\}] = \Pr[(-Z) > s_1/|\beta_1|] = s_1^{-\alpha} A_z^L |\beta_1|^\alpha$$

and so

$$\tau^U = \frac{|\beta_1|^\alpha A_z^L}{|\beta_1|^\alpha A_z^L + A_\varepsilon^U}$$

(c) Consider $\beta_2 > \beta_1 = 0$:

$$\begin{aligned}
\Pr[X_1 > s_1] &\sim s_1^{-\alpha} A_\varepsilon^U \\
\text{and} \quad \Pr[X_2 > s_2] &\sim s_2^{-\alpha}\left(A_z^U \beta_2^\alpha + A_\varepsilon^U\right) \\
\text{but} \quad \Pr[X_1 > s_1, X_2 > s_2] &= \Pr[\varepsilon_1 > s_1] \Pr[\beta_2 Z + \varepsilon_2 > s_2] \\
&\sim s_1^{-\alpha} s_2^{-\alpha} A_\varepsilon^U \left(A_z^U \beta_2^\alpha + A_\varepsilon^U\right) \\
&= s_1^{-2\alpha}\left(A_\varepsilon^U\right)^2
\end{aligned}$$

using $s_2 = s_1 \left(\frac{A_z^U \beta_2^\alpha + A_\varepsilon^U}{A_z^U \beta_1^\alpha + A_\varepsilon^U}\right)^{1/\alpha}$ from part (a). Thus the denominator of the tail dependence coefficientis of order $\mathcal{O}\left(s^{-\alpha}\right)$ while the numerator is of order $\mathcal{O}\left(s^{-2\alpha}\right)$, so the coefficient will be zero.

(d) Consider $\beta_1 < 0 < \beta_2$. Then the denominator of the tail dependence coefficient will again be of order $\mathcal{O}\left(s_1^{-\alpha}\right)$, but the numerator will be of a lower order:

$$\begin{aligned}
\Pr\left[X_1 > s_1, X_2 > s_2\right] &= \Pr\left[\beta_1 Z + \varepsilon_1 > s_1, \beta_2 Z + \varepsilon_2 > s_2\right] \\
&= \Pr\left[\beta_1 Z > s_1, \beta_2 Z > s_2\right] + o\left(s^{-\alpha}\right) \text{ as } s \to \infty
\end{aligned}$$

using the same results as above. But note that $\Pr\left[\beta_1 Z > s_1, \beta_2 Z > s_2\right] = 0$ since $s_1, s_2 > 0$ and $sgn\left(\beta_1 Z\right) = -sgn\left(\beta_2 Z\right)$, and so the numerator will be of order $o\left(s^{-\alpha}\right)$, implying the tail dependence coefficient will be zero. All of the results for parts (a) through (d) apply for lower tail dependence, *mutatis mutandis.* ■

Proof of Proposition 2. It is more convenient to work with the density than the distribution function for skew t random variables. Note that if F_z has a regularly varying tails with tail index $\alpha > 0$, and has corresponding density function f_z that is monotone decreasing in the tails (i.e., it satisfies the Monotone Density Theorem, see Bingham *et al.* (1987)), then f_z satisfies:

$$f_z(s) \sim \alpha A_z^U s^{-\alpha-1}$$

and we can thus obtain:

$$A_z^U = \lim_{s\to\infty} \frac{f_z(s)}{\alpha s^{-\alpha-1}}$$

The skew t distribution of Hansen (1994) has a unique mode and is monotone decreasing on each side of this, thus satisfying the monotonicity condition.

For $\nu \in (2,\infty)$ and $\lambda \in (-1,1)$, the skew t distribution of Hansen (1994) has density:

$$\begin{aligned}
f_z(s;\nu,\lambda) &= \begin{cases} bc\left(1 + \frac{1}{\nu-2}\left(\frac{bs+a}{1-\lambda}\right)^2\right)^{-(\nu+1)/2}, & s < -a/b \\ bc\left(1 + \frac{1}{\nu-2}\left(\frac{bs+a}{1+\lambda}\right)^2\right)^{-(\nu+1)/2}, & s \geq -a/b \end{cases} \\
\text{where } a &= 4\lambda c\left(\frac{\nu-2}{\nu-1}\right), \; b = \sqrt{1+3\lambda^2-a^2}, \; c = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi(\nu-2)}}
\end{aligned}$$

and its tail index is equal to the degrees of freedom parameter, so $\alpha = \nu$. Straightforward algebra yields

$$A_z^U = \lim_{s\to\infty} \frac{f_z(s)}{\nu s^{-\nu-1}} = \frac{bc}{\nu}\left(\frac{b^2}{(\nu-2)(1+\lambda)^2}\right)^{-(\nu+1)/2}$$

For the left tail we similarly obtain:

$$f_z(s) \sim \alpha A_z^L (-s)^{-\alpha-1}$$

and

$$A_z^L = \lim_{s\to-\infty} \frac{f_z(s)}{\alpha(-s)^{-\alpha-1}} = \frac{bc}{\nu}\left(\frac{b^2}{(\nu-2)(1-\lambda)^2}\right)^{-(\nu+1)/2}$$

When $\lambda = 0$ we recover the non-skewed, standardized Student's t distribution. In that case we have $a = 0$, $b = 1$ (and c unchanged), and so $A_\varepsilon^U = A_\varepsilon^L = \frac{c}{\nu}\left(\frac{1}{\nu-2}\right)^{-(\nu+1)/2}$. ■

Proof of Proposition 3. First consider the denominator of the upper tail dependence coefficient:

$$\begin{aligned}
\Pr[X_i > s_i] &= \Pr\left[\sum_{k=1}^K \beta_{ik} Z_k + \varepsilon_i > s_i\right] \\
&\sim \Pr[\varepsilon_i > s_i] + \sum_{k=1}^K \Pr[\beta_{ik} Z_k > s_i] \\
&= s_i^{-\alpha}\left(A_\varepsilon^U + \sum_{k=1}^K A_k^U \beta_{ik}^\alpha\right)
\end{aligned}$$

We need to choose (s_i, s_j) such that $\Pr[X_i > s_i] = \Pr[X_j > s_j]$, which implies

$$s_j = s_i \left(\frac{A_\varepsilon^U + \sum_{k=1}^K A_k^U \beta_{jk}^\alpha}{A_\varepsilon^U + \sum_{k=1}^K A_k^U \beta_{ik}^\alpha}\right)^{1/\alpha} \equiv s_i \gamma_{U,ij}$$

As in Proposition 1, note that s_i and s_j diverge at the same rate.

When $\beta_{ik}\beta_{jk} = 0$, the factor Z_k does not contribute to the numerator of the tail dependence coefficient, as it appears in at most one of X_i and X_j. Thus we need only keep track of factors such that $\beta_{ik}\beta_{jk} > 0$. In this case, we again need to determine the larger of s_i/β_{ik} and s_j/β_{jk} for each $k = 1, 2, ..., K$. Unlike the one-factor model, a general ranking cannot be obtained. To keep notation compact we introduce δ_{ijk}. Note

$$\begin{aligned}
\max\left\{\frac{s_i}{\beta_{ik}}, \frac{s_j}{\beta_{jk}}\right\} &= \max\left\{\frac{s_i}{\beta_{ik}}, \frac{s_i}{\beta_{jk}}\gamma_{U,ij}\right\} = \frac{s_i}{\beta_{ik}} \max\left\{1, \frac{\beta_{ik}}{\beta_{jk}}\gamma_{U,ij}\right\} \equiv \frac{s_i}{\beta_{ik}\delta_{ijk}} \\
\text{where } \delta_{ijk}^{-1} &\equiv \max\left\{1, \frac{\beta_{ik}}{\beta_{jk}}\gamma_{U,ij}\right\}
\end{aligned}$$

To cover the case that $\beta_{ik}\beta_{jk} = 0$, we generalize the definition of δ_{ijk} so that it is well defined in that case. The use of any finite number here will work (as it will be multiplied by zero in this case) and here we set it to one:

$$\delta_{ijk}^{-1} \equiv \begin{cases} \max\left\{1, \gamma_{U,ij}\beta_{ik}/\beta_{jk}\right\}, & \text{if } \beta_{ik}\beta_{jk} > 0 \\ 1, & \text{if } \beta_{ik}\beta_{jk} = 0 \end{cases}$$

Now we can consider the numerator

$$\begin{aligned} \Pr\left[X_i > s_i, X_j > s_j\right] &= \Pr\left[\sum_{k=1}^{K} \beta_{ik} Z_k + \varepsilon_i > s_i, \sum_{k=1}^{K} \beta_{jk} Z_k + \varepsilon_j > s_j\right] \\ &\sim \sum_{k=1}^{K} \Pr\left[\beta_{ik} Z_k > s_i, \beta_{jk} Z_k > s_j\right] \end{aligned}$$

$$\begin{aligned} &= \sum_{k=1}^{K} \mathbf{1}\left\{\beta_{ik}\beta_{jk} > 0\right\} \Pr\left[\beta_{ik} Z_k > s_i, \beta_{jk} Z_k > s_j\right] \\ &= \sum_{k=1}^{K} \mathbf{1}\left\{\beta_{ik}\beta_{jk} > 0\right\} \Pr\left[Z_k > \max\left\{\frac{s_i}{\beta_{ik}}, \frac{s_j}{\beta_{jk}}\right\}\right] \\ &\equiv \sum_{k=1}^{K} \mathbf{1}\left\{\beta_{ik}\beta_{jk} > 0\right\} \Pr\left[Z_k > \frac{s_i}{\beta_{ik}\delta_{ijk}}\right] \\ &= s_i^{-\alpha} \sum_{k=1}^{K} \mathbf{1}\left\{\beta_{ik}\beta_{jk} > 0\right\} A_k^U \beta_{ik}^{\alpha} \delta_{ijk}^{\alpha} \end{aligned}$$

And so we obtain

$$\tau_{ij}^U = \lim_{s \to \infty} \frac{\Pr\left[X_i > s_i, X_j > s_j\right]}{\Pr\left[X_i > s_i\right]} = \frac{\sum_{k=1}^{K} \mathbf{1}\left\{\beta_{ik}\beta_{jk} > 0\right\} A_k^U \beta_{ik}^{\alpha} \delta_{ijk}^{\alpha}}{A_\varepsilon^U + \sum_{k=1}^{K} A_k^U \beta_{ik}^{\alpha}}$$

The results for lower tail dependence can be obtained using similar derivations to those above. ■

Proof of Proposition 4. First we derive some initial results. By the results of Fermanian *et al.* (2004), assumption 1 implies that sample rank dependence measures converge in probability to their population counterparts. Thus $\hat{\mathbf{R}}_T \xrightarrow{p} \mathbf{R}$, and $g_k(\hat{\mathbf{R}}_T) \xrightarrow{p} g_k(\mathbf{R})$ for $k = 1, 2, ..., N$. By assumption, the copula of $\mathbf{Y}$ is the same as that of $\mathbf{X}$, and so $RankCorr\left[\mathbf{Y}\right] = RankCorr\left[\mathbf{X}\right]$, since rank correlations are solely functions of the copula, see Nelsen (2006, Chapter 5).

Re-writing the factor model in equation (2) in matrix form, $\mathbf{X} = \mathbf{BZ} + \boldsymbol{\varepsilon}$, we can easily obtain its covariance and correlation matrix:

$$\begin{aligned} V\left[\mathbf{X}\right] &= \mathbf{BB}' + I \\ \mathbf{R}^L &= \mathbf{\Gamma BB}'\mathbf{\Gamma} + \mathbf{\Gamma}^2 \\ \text{where } \mathbf{\Gamma} &\equiv diag\left\{\left[\gamma_1, \gamma_2, ..., \gamma_N\right]\right\} \\ \gamma_i &\equiv \left(1 + \boldsymbol{\beta}_i'\boldsymbol{\beta}_i\right)^{-1/2}, \;\; i = 1, 2, ..., N \end{aligned}$$

Following Proposition 4 of Chamberlain and Rothschild (1983) we obtain a bound on the eigenvalues of $\mathbf{R}^L$. Specifically, we use the result that if A and B are symmetric matrices, then $g_i(A+B) \leq g_j(A) + g_k(B)$, for $j+k \leq i+1$. Thus we find

$$g_{K+1}\left(\mathbf{R}^L\right) \leq g_{K+1}\left(\mathbf{\Gamma B B'\Gamma}\right) + g_1\left(\mathbf{\Gamma}^2\right) = g_1\left(\mathbf{\Gamma}^2\right) \leq 1$$

where the first inequality follows from the previously-mentioned bound, the equality follows from $rank\left(\mathbf{\Gamma B B'\Gamma}\right) = K$, and the second inequality follows from the fact that $g_1\left(\mathbf{\Gamma}^2\right) = \max_i\ \gamma_i^2$, and $\gamma_i^2 \leq 1$ since $\boldsymbol{\beta}_i'\boldsymbol{\beta}_i \geq 0$.

(i) $\Pr\left[\hat{K}_T > K\right] = \Pr\left[g_{K+1}(\hat{\mathbf{R}}_T^y) > 1\right] \to 0$ as $T \to \infty$, since $g_{K+1}(\hat{\mathbf{R}}_T^y) \xrightarrow{p} g_{K+1}(\mathbf{R}) = g_{K+1}(\mathbf{R}^L) \leq 1$. Thus $\Pr\left[\hat{K}_T \leq K\right] \to 1$ as $T \to \infty$.

(ii) $\Pr\left[\hat{K}_T < K\right] = \Pr\left[g_K(\hat{\mathbf{R}}_T^y) < 1\right] \to 0$ as $T \to \infty$, since $g_K(\hat{\mathbf{R}}_T^y) \xrightarrow{p} g_K(\mathbf{R}) = g_K(\mathbf{R}^L) > 1$ under assumption 3. Thus, combining with part (i) we have $\Pr\left[\hat{K}_T = K\right] \to 1$ as $T \to \infty$. ■

Appendix S.A.2: A Monte Carlo study of SMM estimation of high dimension factor copulas

In this section we present a study of the finite sample properties of the simulated method of moments (SMM) estimator defined in equation (15) of the main paper. In the case where a likelihood for the copula model is available in closed form we contrast the properties of the SMM estimator with those of the maximum likelihood estimator.

We initially consider three different factor copulas, all of them of the form:

$$\begin{aligned} X_i &= \beta Z + \varepsilon_i, \quad i = 1, 2, ..., N \\ Z &\sim Skew\ t\,(\nu, \lambda) \\ \varepsilon_i &\sim iid\ t\,(\nu)\,, \text{ and } \varepsilon_i \perp\!\!\!\perp Z\ \forall\ i \\ [X_1, ..., X_N]' &\sim \mathbf{F}_x = \mathbf{C}\,(G_x, ..., G_x) \end{aligned} \tag{27}$$

and we use the skewed t distribution of Hansen (1994) for the common factor. In all cases we set $\beta = 1$, implying that the common factor (Z) accounts for one-half of the variance of each X_i, implying rank correlation of around 0.5. In the first model we set $\nu \to \infty$ and $\lambda = 0$, which implies that the resulting factor copula is simply the Gaussian copula, with equicorrelation parameter $\rho = 0.5$. In this case we can estimate the model by SMM and also by GMM and MLE, and we use this case to study the loss of efficiency in moving from MLE to GMM to SMM. In the second model we set $\nu = 4$ and $\lambda = 0$, yielding a symmetric factor copula that generates tail dependence. In the third case we set $\nu = 4$ and $\lambda = -0.5$ yielding a factor copula that generates tail dependence as well as "asymmetric dependence", in that the lower tails of the copula are more dependent than the upper tails. We estimate the inverse degrees of freedom parameter, ν_z^{-1}, so that its parameter space is $[0, 0.5)$ rather than $(2, \infty]$.

We also consider an extension of the above equidependence model which allow each X_i to have a different coefficient on Z. For $N = 3$ we set $[\beta_1, \beta_2, \beta_3] = [0.5, 1, 1.5]$. For $N = 10$ we set $[\beta_1, \beta_2, ..., \beta_{10}] = [0.25, 0.50, ..., 2.5]$, which corresponds to pair-wise rank correlations ranging from approximately 0.1 to 0.8. Motivated by our empirical application below, for the $N = 100$ case we consider a "block equidependence" model, where we assume that the 100 variables can be grouped *ex ante* into 10 groups, and that all variables within each group have the same β_i. We use the same set of values for β_i as in the $N = 10$ case.

We consider two different scenarios for the marginal distributions of the variables of interest. In the first case we assume that the data are *iid* with standard Normal marginal distributions, meaning that the only parameters that need to be estimated are those of the factor copula. This simplified case is contrasted with a second scenario where the marginal distributions of the variables are assumed to follow an AR(1)-GARCH(1,1) process:

$$\begin{aligned} Y_{it} &= \phi_0 + \phi_1 Y_{i,t-1} + \sigma_{it}\eta_{it}, \quad t = 1, 2, ..., T \\ \sigma_{it}^2 &= \omega + \gamma\sigma_{i,t-1}^2 + \alpha\sigma_{i,t-1}^2\eta_{i,t-1}^2 \\ \boldsymbol{\eta}_t &\equiv [\eta_{1t}, ..., \eta_{Nt}] \sim iid \;\; \mathbf{F}_\eta = \mathbf{C}(\Phi, \Phi, ..., \Phi) \end{aligned} \tag{28}$$

where Φ is the standard Normal distribution function and $\mathbf{C}$ is the factor copula implied by equation (27). We set the parameters of the marginal distributions as $[\phi_0, \phi_1, \omega, \gamma, \alpha] = [0.01, 0.05, 0.05, 0.85, 0.10]$, which broadly matches the values of these parameters when estimated using daily equity return data. In this scenario the parameters of the marginal distribution are estimated via QML in a separate first stage, following which the estimated standardized residuals, $\hat{\eta}_{it}$, are obtained and used in a second stage to estimate the factor copula parameters. In all cases we consider a time series of length $T = 1000$, corresponding to approximately 4 years of daily return data, and we use $S = 25 \times T$ simulations in the computation of the dependence measures to be matched in the SMM optimization. We repeat each scenario 100 times. In all results below we use the identity weight matrix for estimation.[12] We use the same dependence measures in the SMM estimation as in our empirical analysis, described in detail in the appendix to the main paper.

Table A.1 reveals that for all three dimensions ($N = 3,\ 10$ and 100) and for all three copula models the estimated parameters are centered on the true values, with the average estimated bias being small relative to the standard deviation, and with the median of the simulated distribution centered on the true values. The measures of estimator accuracy (the standard deviation and the 90-10 percentile difference) reveal that adding more parameters to the model, *ceteris paribus*, leads to greater estimation error, as expected; β, for example, is more accurately estimated when it is the only unknown parameter compared with when it is one of three unknown parameters. Looking across the dimension size, we see that the copula model parameters are almost always more precisely estimated as the dimension grows. This is intuitive, given the equidependence nature of all three models: increasing the dimension of the model does not increase the number of parameters to be estimated but it does increase the amount of information available on the unknown parameters.

Comparing the SMM estimator with the ML estimator, which is only feasible for the Normal copula (as the other two factor copulas do not have a copula likelihood in closed form) we see that the SMM estimator performs quite well. As predicted by theory, the ML estimator is always more efficient than the SMM estimator, however the loss in efficiency is moderate, ranging from around 25% for $N = 3$ to around 10% for $N = 100$. This provides some confidence that our move to SMM, prompted by the lack of a closed-form likelihood, does not come at a cost of a large loss in efficiency. Comparing the SMM estimator to the GMM estimator provides us with a measure of the loss in accuracy from having to estimate the population moment function via simulation. We find that this loss is at most 3% and in some cases ($N = 100$) is slightly negative. Thus little is lost from using SMM rather than GMM when we set $S = 25 \times T$.

Table A.2 shows results for the block equidependence model for the $N = 100$ case, which can be compared to the results in the lower panel of Table A.1. This table shows that the parameters of these models are well estimated using the proposed dependence measures described in the appendix. The accuracy of the "shape" parameters, ν^{-1} and λ, is slightly lower in the more general model, consistent with the estimation error from having to estimate ten factor loadings (β_i) being greater than from having to estimate just a single loading parameter, however this loss is not great.

[12] Corresponding results based on the efficient weight matrix generally comparable to those based on the identity weight matrix, however the coverage rates are worse than those based on the identity weight matrix.

[INSERT TABLES A.1 AND A.2 ABOUT HERE]

In Tables A.3 and A.4 we present the finite-sample coverage probabilities of 95% confidence intervals based on the estimated asymptotic covariance matrix. As discussed in Oh and Patton (2013), a critical input to the asymptotic covariance matrix estimator is the step size used in computing the numerical derivative matrix $\hat{G}$. This step size, ε_T, must go to zero, but at a slower rate than $T^{-1/2}$. Ignoring constants, our simulation sample size of $T = 1000$ suggests setting $\varepsilon_T > 0.03$, which is much larger than standard step sizes used in computing numerical derivatives.[13] We consider a range of values from 0.0001 to 0.1. Table A.4 shows that when the step size is set to 0.01, 0.03 or 0.1 the finite-sample coverage rates are close to their nominal levels. However if the step size is chosen too small (0.003 or smaller) then the coverage rates are much lower than nominal levels. For example, setting $\varepsilon_T = 0.0001$ (which is still 16 times larger than the default setting in Matlab) we find coverage rates as low as 38% for a nominal 95% confidence interval. Thus this table shows that the asymptotic theory provides a reliable means for obtaining confidence intervals, so long as care is taken not to set the step size too small.

[INSERT TABLES A.3 AND A.4 ABOUT HERE]

Finally in Table A.5 we present the results of a study of the rejection rates for the J test of over-identifying restrictions. Given that we consider $W = I$ in this table, the test statistic has a non-standard distribution (see Proposition 4 of Oh and Patton, 2013), and we use $10,000$ simulations to obtain critical values. In this case, the limiting distribution also depends on $\hat{G}$, and we present the rejection rates for various choices of step size ε_T. Table A.5 reveals that the rejection rates are close to their nominal levels, for both the equidependence models and the "different loading" models (which is a block equidependence model for the $N = 100$ case). The J test rejection rates are less sensitive to the choice of step size than the coverage probabilities of confidence intervals, however the best results are again generally obtained when ε_T is 0.01 or greater.

[INSERT TABLE A.5 ABOUT HERE]

[13] For example, the default in many *Matlab* functions is a step size of $\varepsilon^{1/3} \approx 6 \times 10^{-6} \approx 1/(165{,}000)$, where $\varepsilon = 2.22 \times 10^{-16}$ is machine epsilon. This choice is optimal in certain applications, see Judd (1998) for example.

Appendix S.A.3: A Monte Carlo study of the use of "scree" plots for factor copulas

In this section we study the usefulness of "scree" plots based on rank correlation matrices for identifying the number of factors in a factor copula. We consider factor copulas with the number of factors $K \in \{1, 2, 4, 8\}$, and in all cases we set $N = 100$ and $T = 1000$. We firstly consider the following DGP:

$$\begin{aligned} X_i &= \boldsymbol{\beta}_i' \mathbf{Z} + \varepsilon_i, \quad i = 1, 2, ..., N \\ Z_k &\sim Skew\ t(\nu_k, \lambda_k), \text{ where } \nu_k \sim Unif[3, 33], \ \lambda_k \sim Unif[-1, 1] \\ \varepsilon_i &\sim t(\nu_i), \text{ where } \nu_i \sim Unif[3, 33] \end{aligned} \tag{29}$$

and all variables and parameters are *iid* across variables and across simulations. We choose the factor loadings, β_{ik}, to imply cross-sectional average correlations that are around 0.5, consistent with our empirical application. To implement this, we need to adjust the factor loadings as K varies (else the common factors get too strong as K grows). We assume $\beta_{ik} \sim N\left(\mu_\beta, \sigma_\beta^2\right)$, where:

	$K=1$	$K=2$	$K=4$	$K=8$
μ_β	1.00	0.75	0.50	0.40
σ_β^2	0.20^2	0.20^2	0.20^2	0.20^2

In the figure below we plot all $N(N-1)/2$ linear and rank correlations for one simulation from each of these cases. These scatterplots reveal that these two measures of association are not identical, but are indeed very close, suggesting that assumption 2 of Proposition 4 of the paper is reasonable for the types of factor copulas we consider.

In a second simulation design we attempt to match more closely the model we find works well in our empirical results. This model is an eight-factor model, but with any given variable only having non-zero loading on a common "market" factor, and one of seven "industry" factors; the loadings on six of the factors are imposed to be zero. We use the following design:

$$\begin{aligned} X_i &= \boldsymbol{\beta}_i' \mathbf{Z} + \varepsilon_i \\ Z_0 &\sim Skew\ t(4, -0.5) \quad \text{and } \beta_{i0} \sim N\left(1, 0.1^2\right) \\ Z_k &\sim t(4) \quad \text{and} \quad \beta_{ik} \sim N\left(0.45, 0.3^2\right), \quad k = 1, 2, ..., 7 \\ \varepsilon_i &\sim t(4), \quad i = 1, 2, .., N \end{aligned} \tag{30}$$

We group the $N = 100$ variables into seven "industries," each containing 14 variables, except for the last group which contains 16. The results for the "no groups" and the "industry groups" DGPs are presented in Table A.6. We find that $\hat{K}_T$ correctly estimates the number of factors for 90% to 99% of simulations.

[INSERT TABLE A.6 ABOUT HERE]

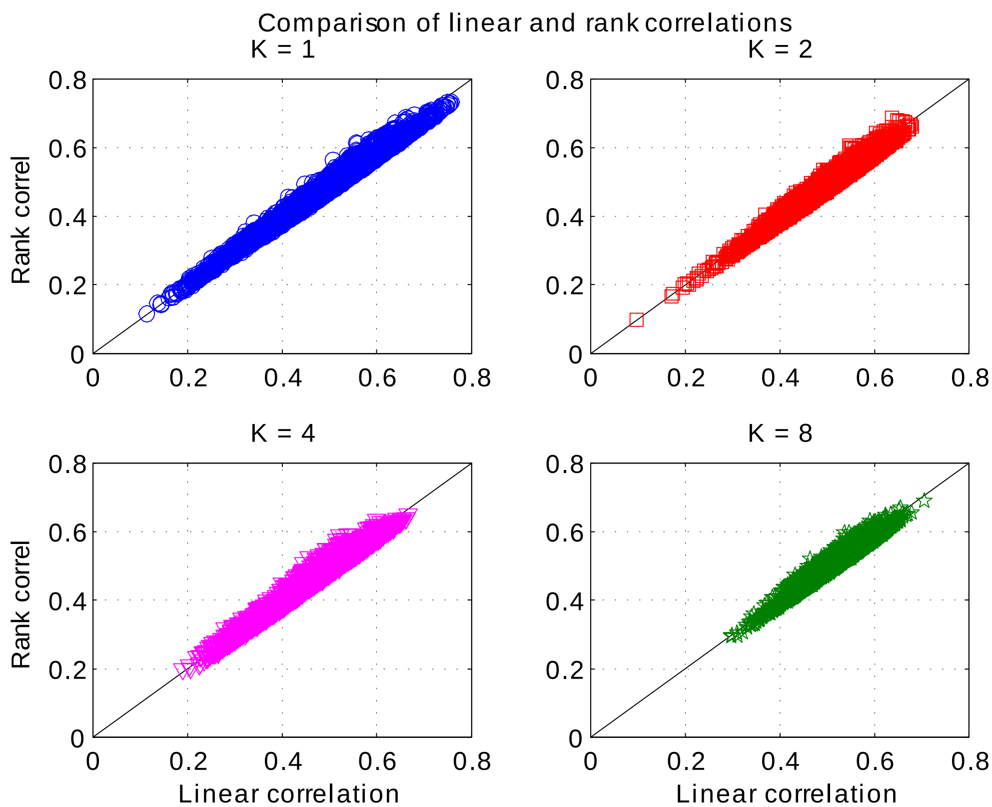

Figure 6: *Comparison of linear and rank correlations for variables generated by the factor copulas in equation (29).*

References

[1] Bingham, N.H., C.M. Goldie and J.L. Teugels, 1987, *Regular Variation*, Cambridge University Press, Cambridge.

[2] Bollerslev, T., 1990, Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH approach, *Review of Economics and Statistics*, 72, 498-505.

[3] Chamberlain, G. and M. Rothschild, 1983, Arbitrage, factor structure, and mean-variance analysis on large asset markets, *Econometrica*, 51(5), 1281-1304.

[4] Chen, X. and Y. Fan, 2006, Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification, *Journal of Econometrics*, 135, 125-154.

[5] Embrechts, P., C. Klüppelberg and T. Mikosch, 1997, *Modelling Extremal Events*, Springer-Verlag, Berlin.

[6] Feller, W., 1970, *An Introduction to Probability Theory and Its Applications*, Volume II, John Wiley & Sons, USA.

[7] Fermanian, J., D. Radulovic and M. Wegkamp, M., 2004, Weak convergence of empirical copula processes, *Bernoulli*, 10, 847-860.

[8] Hansen, B.E., 1994, Autoregressive conditional density estimation, *International Economic Review*, 35(3), 705-730.

[9] Oh, D.H. and A.J. Patton, 2013, Simulated method of moments estimation for copula-based multivariate models, *Journal of the American Statistical Association*, 108, 689-700

[10] Nelsen, R.B., 2006, *An Introduction to Copulas*, Second Edition, Springer, U.S.A.

[11] Rémillard, B., 2010, Goodness-of-fit tests for copulas of multivariate time series, working paper.

Table A.1: Simulation results for factor copula models

	Normal MLE	Normal GMM	Normal SMM	Factor $t-t$		Factor $Skew\ t-t$		
	β	β	β	β	ν^{-1}	β	ν^{-1}	λ
True value	1.00	1.00	1.00	1.00	0.25	1.00	0.25	-0.50
	$N=3$							
Bias	0.0141	-0.0143	-0.0164	-0.0016	-0.0185	0.0126	-0.0199	-0.0517
Std	0.0803	0.1014	0.1033	0.1094	0.0960	0.1205	0.1057	0.1477
Median	1.0095	0.9880	0.9949	0.9956	0.2302	1.0050	0.2380	-0.5213
90%	1.1180	1.1103	1.1062	1.1448	0.3699	1.1772	0.3636	-0.3973
10%	0.9172	0.8552	0.8434	0.8721	0.0982	0.8662	0.0670	-0.7538
90-10 Diff	0.2008	0.2551	0.2628	0.2727	0.2716	0.3110	0.2966	0.3565
	$N=10$							
Bias	0.0113	-0.0099	-0.0119	-0.0025	-0.0137	-0.0039	-0.0161	-0.0119
Std	0.0559	0.0651	0.0666	0.0724	0.0611	0.0851	0.0790	0.0713
Median	1.0125	0.9874	0.9898	0.9926	0.2360	0.9897	0.2376	-0.5084
90%	1.0789	1.0644	1.0706	1.0967	0.3102	1.1095	0.3420	-0.4318
10%	0.9406	0.9027	0.8946	0.9062	0.1704	0.8996	0.1331	-0.5964
90-10 Diff	0.1383	0.1617	0.1761	0.1905	0.1398	0.2100	0.2089	0.1645
	$N=100$							
Bias	0.0167	-0.0068	-0.0080	-0.0011	-0.0138	0.0015	-0.0134	-0.0099
Std	0.0500	0.0554	0.0546	0.0659	0.0549	0.0841	0.0736	0.0493
Median	1.0164	0.9912	0.9956	1.0011	0.2346	0.9943	0.2402	-0.5101
90%	1.0805	1.0625	1.0696	1.0886	0.3127	1.1060	0.3344	-0.4465
10%	0.9534	0.9235	0.9279	0.9112	0.1685	0.8970	0.1482	-0.5734
90-10 Diff	0.1270	0.1390	0.1418	0.1773	0.1442	0.2089	0.1861	0.1270

Notes: This table presents the results from 100 simulations of three different factor copulas, the Normal copula, the $t-t$ factor copula and the $Skew\ t-t$ factor copula. The Normal copula is estimated by ML, GMM, and SMM, and the other two copulas are estimated by SMM. The marginal distributions of the data are assumed to follow AR(1)-GARCH(1,1) processes, as described in Section 3. Problems of dimension $N=3$, 10 and 100 are considered, the sample size is $T=1000$ and the number of simulations used is $S=25\times T$. The first row of each panel presents the average difference between the estimated parameter and its true value. The second row presents the standard deviation in the estimated parameters. The third, fourth and fifth rows present the 50^{th}, 90^{th} and 10^{th} percentiles of the distribution of estimated parameters, and the final row presents the difference between the 90^{th} and 10^{th} percentiles.

Table A.2: Simulation results for block equidependence factor copula model, N=100

	ν^{-1}	λ_z	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
True value	0.25	-0.5	0.25	0.5	0.75	1	1.25	1.5	1.75	2	2.25	2.5
						Normal						
Bias	-	-	-0.0010	-0.0038	-0.0040	-0.0072	-0.0071	-0.0140	-0.0178	-0.0119	-0.0194	-0.0208
Std	-	-	0.0128	0.0182	0.0248	0.0322	0.0377	0.0475	0.0651	0.0784	0.1022	0.1291
Median	-	-	0.2489	0.4970	0.7440	0.9942	1.2421	1.4868	1.7279	1.9918	2.2256	2.4832
90%	-	-	0.2645	0.5204	0.7787	1.0291	1.2970	1.5470	1.8226	2.0874	2.3609	2.6458
10%	-	-	0.2304	0.4701	0.7158	0.9502	1.1982	1.4197	1.6526	1.8825	2.0921	2.3090
90-10 diff	-	-	0.0341	0.0503	0.0629	0.0788	0.0987	0.1273	0.1700	0.2049	0.2689	0.3368
						Factor $t-t$						
Bias	-0.0120	-	0.0000	0.0009	0.0018	-0.0045	0.0011	-0.0073	-0.0080	-0.0122	-0.0061	-0.0065
Std	0.0574	-	0.0149	0.0236	0.0300	0.0343	0.0443	0.0580	0.0694	0.0867	0.1058	0.1332
Median	0.2384	-	0.2503	0.5056	0.7528	0.9985	1.2550	1.4881	1.7409	1.9820	2.2234	2.4737
90%	0.3056	-	0.2678	0.5255	0.7896	1.0348	1.3052	1.5697	1.8270	2.1012	2.4089	2.6597
10%	0.1683	-	0.2348	0.4689	0.7187	0.9462	1.1965	1.4282	1.6517	1.8744	2.1303	2.3196
90-10 diff	0.1373	-	0.0330	0.0566	0.0709	0.0886	0.1086	0.1416	0.1754	0.2268	0.2786	0.3401
						Factor *skew* $t-t$						
Bias	-0.0119	-0.0019	0.0008	0.0001	0.0028	-0.0029	-0.0036	-0.0096	-0.0114	-0.0232	-0.0178	-0.0194
Std	0.0633	0.0451	0.0134	0.0246	0.0320	0.0443	0.0588	0.0806	0.0902	0.1111	0.1373	0.1635
Median	0.2434	-0.5051	0.2477	0.5001	0.7520	0.9986	1.2468	1.4826	1.7417	1.9803	2.2107	2.4786
90%	0.3265	-0.4392	0.2680	0.5309	0.7961	1.0613	1.3028	1.5856	1.8378	2.1094	2.4430	2.7034
10%	0.1550	-0.5527	0.2358	0.4660	0.7155	0.9505	1.1756	1.4042	1.6230	1.8395	2.0494	2.2739
90-10 diff	0.1714	0.1134	0.0321	0.0648	0.0807	0.1107	0.1272	0.1814	0.2148	0.2699	0.3936	0.4294

Notes: This table presents the results from 100 simulations of three different factor copulas: the Normal copula, the $t-t$ factor copula and the *Skew* $t-t$ factor copula. We divide the $N = 100$ variables into ten groups and assume that all variables in the same group have the same loading on the common factor. The marginal distributions of the data are assumed to follow AR(1)-GARCH(1,1) processes, as described in Section 3. The sample size is $T = 1000$ and the number of simulations used is $S = 25 \times T$. The first row of each panel presents the average difference between the estimated parameter and its true value. The second row presents the standard deviation in the estimated parameters. The third, fourth and fifth rows present the 50^{th}, 90^{th} and 10^{th} percentiles of the distribution of estimated parameters, and the final row presents the difference between the 90^{th} and 10^{th} percentiles.

Table A.3: Simulation results on coverage rates

	Normal	Factor $t-t$		Factor *Skew* $t-t$		
	β	β	ν^{-1}	β	ν^{-1}	λ
			$N=3$			
ε_T						
0.1	89	93	97	99	100	96
0.03	90	94	98	99	98	96
0.01	88	92	98	99	96	95
0.003	85	95	95	96	89	95
0.001	83	89	89	92	84	93
0.0003	58	69	69	74	74	74
0.0001	38	49	53	57	70	61
			$N=10$			
ε_T						
0.1	87	93	99	97	98	99
0.03	87	95	99	97	98	97
0.01	87	94	96	97	98	95
0.003	87	95	95	98	95	96
0.001	87	95	93	96	90	95
0.0003	86	94	87	91	77	93
0.0001	71	87	81	71	81	85
			$N=100$			
ε_T						
0.1	95	93	95	94	95	94
0.03	95	94	94	94	94	94
0.01	95	93	93	94	94	94
0.003	94	95	93	94	94	94
0.001	94	94	92	94	93	95
0.0003	92	94	92	94	92	93
0.0001	84	94	89	94	88	95

Notes: This table presents the results from 100 simulations of three different factor copulas, the Normal copula, the $t-t$ factor copula and the *Skew* $t-t$ factor copula, all estimated by SMM. The marginal distributions of the data are assumed to follow AR(1)-GARCH(1,1) processes, as described in Section 3. Problems of dimension $N = 3$, 10 and 100 are considered, the sample size is $T = 1000$ and the number of simulations used is $S = 25 \times T$. The rows of each panel contain the step size, ε_T, used in computing the matrix of numerical derivatives, $\hat{G}_{T,S}$. The numbers in the table present the percentage of simulations for which the 95% confidence interval based on the estimated covariance matrix contained the true parameter.

Table A.4: Coverage rate for block equidependence factor copula model, N=100

	ν^{-1}	λ	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
						Normal						
ε_T												
0.1	-	-	97	91	92	89	95	93	94	95	95	90
0.03	-	-	97	91	92	90	95	95	94	95	95	90
0.01	-	-	97	91	92	90	95	94	94	96	94	91
0.003	-	-	97	90	93	90	95	94	95	96	95	90
0.001	-	-	97	90	94	93	94	94	94	96	94	92
0.0003	-	-	97	92	93	92	95	94	91	93	92	94
0.0001	-	-	94	94	91	88	90	92	94	91	88	86
						Factor $t-t$						
ε_T												
0.1	95	-	94	93	96	96	98	91	93	92	95	93
0.03	94	-	94	91	96	96	98	92	93	92	97	93
0.01	95	-	94	94	97	96	97	93	93	92	98	93
0.003	94	-	94	94	97	96	97	94	94	95	98	95
0.001	94	-	93	93	97	97	97	92	96	94	100	94
0.0003	90	-	94	95	98	97	99	94	95	95	99	93
0.0001	65	-	95	96	96	98	98	92	96	94	97	91
						Factor *Skew* $t-t$						
ε_T												
0.1	93	95	98	95	96	94	94	92	91	91	90	92
0.03	93	95	98	95	95	94	95	92	91	91	89	90
0.01	93	95	97	96	95	94	94	92	92	91	91	91
0.003	93	95	97	96	96	94	95	92	92	92	90	89
0.001	93	94	97	96	95	94	94	91	91	93	89	88
0.0003	84	93	98	95	95	95	95	90	90	88	83	85
0.0001	69	86	98	97	94	91	90	88	87	84	83	80

Notes: This table presents the results from 100 simulations of three different factor copulas: the Normal copula, the $t-t$ factor copula and the *Skew* $t-t$ factor copula. We divide the $N = 100$ variables into ten groups and assume that all variables in the same group have the same loading on the common factor. The marginal distributions of the data are assumed to follow AR(1)-GARCH(1,1) processes, as described in Section 3. The sample size is $T = 1000$ and the number of simulations used is $S = 25 \times T$. The rows of each panel contain the step size, ε_T, used in computing the matrix of numerical derivatives, $\hat{G}_{T,S}$. The numbers in the table present the percentage of simulations for which the 95% confidence interval based on the estimated covariance matrix contained the true parameter.

Table A.5: Rejection frequencies for the test of overidentifying restrictions

	Equidependence			Different loadings		
	Normal	Factor $t-t$	Factor $Skew\ t-t$	Normal	Factor $t-t$	Factor $Skew\ t-t$
			$N=3$			
ε_T						
0.1	97	97	99	95	97	97
0.03	97	98	99	95	95	96
0.01	97	97	100	93	95	95
0.003	97	98	100	92	95	96
0.001	98	96	100	93	93	97
0.0003	99	97	100	91	92	97
0.0001	99	97	99	92	94	98
			$N=10$			
ε_T						
0.1	97	97	98	98	95	98
0.03	98	97	97	98	95	99
0.01	96	97	97	97	94	98
0.003	97	96	97	98	92	99
0.001	98	95	97	96	89	100
0.0003	97	94	97	97	93	100
0.0001	97	94	98	98	95	100
			$N=100$			
ε_T						
0.1	97	95	99	95	95	99
0.03	97	95	98	96	94	99
0.01	97	95	98	96	93	99
0.003	97	95	97	95	94	99
0.001	97	94	99	95	91	100
0.0003	97	94	99	95	89	100
0.0001	98	92	98	93	90	100

Notes: This table presents the results from 100 simulations of three different factor copulas, the Normal copula, the $t-t$ factor copula and the $Skew\ t-t$ factor copula, all estimated by SMM. The marginal distributions of the data are assumed to follow AR(1)-GARCH(1,1) processes, as described in Section 3. Problems of dimension $N=3$, 10 and 100 are considered, the sample size is $T=1000$ and the number of simulations used is $S=25\times T$. The rows of each panel contain the step size, ε_T, used in computing the matrix of numerical derivatives, $\hat{G}_{T,S}$, needed for the critical value. The confidence level for the test of over-identifying restrictions is 0.95, and the numbers in the table present the percentage of simulations for which the test statistic was less than its computed critical value.

Table A.6: Properties of the estimator of the number of factors

	No "industry" groups				"Industry" groups
	$K = 1$	$K = 2$	$K = 4$	$K = 8$	$K = 8$
Mean	1.267	2.078	4.040	7.999	7.896
Std dev	1.154	0.293	0.201	0.045	0.305
$\Pr[\hat{K}_T < K]$	0.000	0.000	0.000	0.002	0.104
$\Pr[\hat{K}_T = K]$	0.903	0.928	0.961	0.998	0.896
$\Pr[\hat{K}_T > K]$	0.097	0.072	0.039	0.000	0.000

Notes: This table presents results from 1000 simulations of five different factor copulas, with the true number of factors denoted K. In all simulations we set $N = 100$ and $T = 1000$. The estimator for the number of factors, $\hat{K}_T$, is presented in Proposition 4 of the main paper.