Supplemental Appendices for

Is a normal copula the right copula?

Dante Amengual
CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain
<amengual@cemfi.es>

Enrique Sentana
CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain
<sentana@cemfi.es>

July 2018

B The score, Hessian and information matrix

Let $\boldsymbol{\varepsilon}$ denote a K-dimensional random vector with density function $f_K(\boldsymbol{\varepsilon};\boldsymbol{\rho},\boldsymbol{\varphi})$, where the $p+q$ parameters of interest are $\boldsymbol{\rho}$ (correlation) and $\boldsymbol{\varphi}$ (shape), whose true values are $(\boldsymbol{\rho}_0',\boldsymbol{\varphi}_0')'$. Similarly, let $f_{1k}(\varepsilon;\boldsymbol{\varphi})$ and $F_{1k}(\varepsilon;\boldsymbol{\varphi})$ denote the marginal density and distribution functions of the k^{th} element of this distribution, so that $\varepsilon_k(\boldsymbol{\varphi})$, which is implicitly defined by

$$\int_{-\infty}^{\varepsilon_k(\boldsymbol{\varphi})} f_{1k}(e;\boldsymbol{\varphi})de = F_{1k}[\varepsilon_k(\boldsymbol{\varphi});\boldsymbol{\varphi}] = u_K,$$

is the quantile with respect to the k^{th} marginal distribution of the assumed joint distribution evaluated at the probability integral transform of the k^{th} observation, $u_k = G_{1k}(x_k)$.

Assumption 2 $f_K(\boldsymbol{\varepsilon};\boldsymbol{\rho},\boldsymbol{\varphi})$ *is a well defined density function, strictly positive over its domain and twice continuously differentiable with respect to all its arguments.*

This assumption holds for the *GH* distribution, at least in the vicinity of the Gaussian null, as shown in the Supplemental Appendix of Mencía and Sentana (2012); see also Supplemental Appendix C.

Although we will relax it in Supplemental Appendix E.3, for clarity of exposition we also assume that:

Assumption 3 *The vectors of probability integral transforms of the observations,* $\mathbf{u}_n$*,* $n = 1,2,...N$*, are independent and identically distributed according to the assumed copula.*

Given our assumptions, the log-likelihood function of the copula derived from $\boldsymbol{\varepsilon}$ for a sample of size N will take the form $\sum_{n=1}^{N}\ln c(\mathbf{u}_n;\boldsymbol{\rho},\boldsymbol{\varphi})$, where

$$\begin{aligned} \ln c(\mathbf{u};\boldsymbol{\rho},\boldsymbol{\varphi}) &= \ln f_K[\boldsymbol{\varepsilon}(\boldsymbol{\varphi});\boldsymbol{\rho},\boldsymbol{\varphi}] - \sum_{k=1}^{K}\ln f_{1k}[\varepsilon_k(\boldsymbol{\varphi});\boldsymbol{\varphi}], \qquad \text{(B1)} \\ \boldsymbol{\varepsilon}(\boldsymbol{\varphi}) &= [\varepsilon_1(\boldsymbol{\varphi}),...,\varepsilon_K(\boldsymbol{\varphi})]' = \left[F_{11}^{-1}(u_1;\boldsymbol{\varphi}),...,F_{1K}^{-1}(u_K;\boldsymbol{\varphi})\right]'. \end{aligned}$$

Let $\mathbf{s}(\boldsymbol{\rho},\boldsymbol{\varphi})$ denote the score function, and partition it into $\mathbf{s}_{\boldsymbol{\rho}}(\boldsymbol{\rho},\boldsymbol{\varphi}) = \partial\ln c(\mathbf{u};\boldsymbol{\rho},\boldsymbol{\varphi})/\partial\boldsymbol{\rho}$ and $\mathbf{s}_{\boldsymbol{\varphi}}(\boldsymbol{\rho},\boldsymbol{\varphi}) = \partial\ln c(\mathbf{u};\boldsymbol{\rho},\boldsymbol{\varphi})/\partial\boldsymbol{\varphi}$, whose dimensions conform to those of $\boldsymbol{\rho}$ and $\boldsymbol{\varphi}$. Then

$$\mathbf{s}_{\boldsymbol{\rho}}(\boldsymbol{\rho},\boldsymbol{\varphi}) = \frac{d\ln f_K[\boldsymbol{\varepsilon}(\boldsymbol{\varphi});\boldsymbol{\rho},\boldsymbol{\varphi}]}{d\boldsymbol{\rho}} = -\mathbf{Z}_s(\boldsymbol{\rho})\mathbf{e}_s(\boldsymbol{\rho},\boldsymbol{\varphi}),$$

where

$$\begin{aligned} \mathbf{Z}_s(\boldsymbol{\rho}) &= \frac{\partial vec'[\mathbf{P}^{1/2}(\boldsymbol{\rho})]}{\partial\boldsymbol{\rho}}\cdot[\mathbf{I}_K\otimes\mathbf{P}^{-1/2\prime}(\boldsymbol{\rho})], \\ \mathbf{e}_s(\boldsymbol{\rho},\boldsymbol{\varphi}) &= vec\left\{\mathbf{I}_K + \frac{\partial\ln f[\boldsymbol{\varepsilon}^*(\boldsymbol{\rho},\boldsymbol{\varphi});\boldsymbol{\varphi}]}{\partial\boldsymbol{\varepsilon}^*}\cdot\boldsymbol{\varepsilon}^{*\prime}(\boldsymbol{\rho},\boldsymbol{\varphi})\right\} \end{aligned}$$

and $\boldsymbol{\varepsilon}^*(\boldsymbol{\rho},\boldsymbol{\varphi}) = \mathbf{P}^{-1/2}(\boldsymbol{\rho})\boldsymbol{\varepsilon}(\boldsymbol{\varphi})$, because $\boldsymbol{\rho}$ only enters through the joint distribution and not through the marginals or the quantile functions.

On the other hand,

$$\begin{aligned}
\mathbf{s}_{\boldsymbol{\varphi}}(\boldsymbol{\rho},\boldsymbol{\varphi}) &= \frac{d\ln f_K\left[\boldsymbol{\varepsilon}(\boldsymbol{\varphi});\boldsymbol{\rho},\boldsymbol{\varphi}\right]}{d\boldsymbol{\varphi}} - \sum_{k=1}^{K}\frac{d\ln f_{1k}\left[\varepsilon_k(\boldsymbol{\varphi});\boldsymbol{\varphi}\right]}{d\boldsymbol{\varphi}} \\
&= \frac{\partial\ln f_K\left[\boldsymbol{\varepsilon}(\boldsymbol{\varphi});\boldsymbol{\rho},\boldsymbol{\varphi}\right]}{\partial\boldsymbol{\varphi}} - \sum_{k=1}^{K}\frac{\partial f_{1k}\left[\varepsilon_k(\boldsymbol{\varphi});\boldsymbol{\varphi}\right]}{\partial\boldsymbol{\varphi}} \\
&\quad + \sum_{k=1}^{K}\left[\frac{\partial\ln f_K[\boldsymbol{\varepsilon}(\boldsymbol{\varphi});\boldsymbol{\rho},\boldsymbol{\varphi}]}{\partial\varepsilon_k} - \frac{\partial f_{1k}\left[\varepsilon_k(\boldsymbol{\varphi});\boldsymbol{\varphi}\right]}{\partial\varepsilon_k}\right]\frac{\partial\varepsilon_k(\boldsymbol{\varphi})}{\partial\boldsymbol{\varphi}}. \qquad \text{(B2)}
\end{aligned}$$

Expression (B2) decomposes the copula score into three easy to interpret components. The first one corresponds to the score of the joint distribution. The second one to the scores of the K marginal distributions. Finally, for the third component, we have to multiply the difference between the log-derivatives of the joint and marginal distributions with respect to each of their arguments by the derivatives of the corresponding marginal quantile functions with respect to the shape parameters, whose existence is guaranteed by our assumptions.

Let $\mathbf{h}(\boldsymbol{\rho},\boldsymbol{\varphi})$ denote the Hessian function $d\mathbf{s}(\boldsymbol{\rho},\boldsymbol{\varphi})/d(\boldsymbol{\rho}',\boldsymbol{\varphi}')$. We can then show that

$$\begin{aligned}
\mathbf{h}_{\boldsymbol{\varphi}\boldsymbol{\varphi}}(\boldsymbol{\rho},\boldsymbol{\varphi}) &= \frac{d\mathbf{s}_{\boldsymbol{\varphi}}\left(\boldsymbol{\rho},\boldsymbol{\varphi}\right)}{d\boldsymbol{\varphi}'} = \frac{d^2\ln f_K[\boldsymbol{\varepsilon}(\boldsymbol{\varphi});\boldsymbol{\rho},\boldsymbol{\varphi}]}{d\boldsymbol{\varphi}d\boldsymbol{\varphi}'} - \sum_{k=1}^{K}\frac{d^2\ln f_{1k}\left[\varepsilon_k(\boldsymbol{\varphi});\boldsymbol{\varphi}\right]}{d\boldsymbol{\varphi}d\boldsymbol{\varphi}'} \\
&= \frac{\partial^2\ln f_K[\boldsymbol{\varepsilon}(\boldsymbol{\varphi});\boldsymbol{\rho},\boldsymbol{\varphi}]}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\varphi}'} - \sum_{k=1}^{K}\frac{\partial^2\ln f_{1k}\left[\varepsilon_k(\boldsymbol{\varphi});\boldsymbol{\varphi}\right]}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\varphi}'} \\
&\quad +2\sum_{k=1}^{K}\frac{\partial\varepsilon_k(\boldsymbol{\varphi})}{\partial\boldsymbol{\varphi}}\left[\frac{\partial^2\ln f_K[\boldsymbol{\varepsilon}(\boldsymbol{\varphi});\boldsymbol{\rho},\boldsymbol{\varphi}]}{\partial\varepsilon_k\partial\boldsymbol{\varphi}'} - \frac{\partial^2\ln f_{1k}\left[\varepsilon_k(\boldsymbol{\varphi});\boldsymbol{\varphi}\right]}{\partial\varepsilon_k\partial\boldsymbol{\varphi}'}\right] \\
&\quad +\sum_{k=1}^{K}\sum_{j=1}^{K}\frac{\partial\varepsilon_k(\boldsymbol{\varphi})}{\partial\boldsymbol{\varphi}}\left[\frac{\partial^2\ln f_K[\boldsymbol{\varepsilon}(\boldsymbol{\varphi});\boldsymbol{\rho},\boldsymbol{\varphi}]}{\partial\varepsilon_k\partial\varepsilon_j} - \frac{\partial^2\ln f_{1k}\left[\varepsilon_k(\boldsymbol{\varphi});\boldsymbol{\varphi}\right]}{\partial\varepsilon_k\partial\varepsilon_j}\right]\frac{\partial\varepsilon_j(\boldsymbol{\varphi})}{\partial\boldsymbol{\varphi}'} \\
&\quad +\sum_{k=1}^{K}\frac{\partial^2\varepsilon_k(\boldsymbol{\varphi})}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\varphi}'}\left[\frac{\partial^2\ln f_K[\boldsymbol{\varepsilon}(\boldsymbol{\varphi});\boldsymbol{\rho},\boldsymbol{\varphi}]}{\partial\varepsilon_k} - \frac{\partial^2\ln f_{1k}\left[\varepsilon_k(\boldsymbol{\varphi});\boldsymbol{\varphi}\right]}{\partial\varepsilon_k}\right], \qquad \text{(B3)}
\end{aligned}$$

$$\mathbf{h}_{\boldsymbol{\rho}\boldsymbol{\rho}}(\boldsymbol{\rho},\boldsymbol{\varphi}) = \mathbf{Z}_s(\boldsymbol{\rho})\frac{\partial\mathbf{e}_s(\boldsymbol{\rho},\boldsymbol{\varphi})}{\partial\boldsymbol{\rho}'} + \left[\mathbf{e}_s'(\boldsymbol{\phi})\otimes\mathbf{I}_p\right]\frac{\partial vec[\mathbf{Z}_s(\boldsymbol{\rho})]}{\partial\boldsymbol{\rho}'},$$

and

$$\mathbf{h}_{\boldsymbol{\rho}\boldsymbol{\varphi}}(\boldsymbol{\rho},\boldsymbol{\varphi}) = \mathbf{Z}_s(\boldsymbol{\rho})\partial\mathbf{e}_s(\boldsymbol{\rho},\boldsymbol{\varphi})/\partial\boldsymbol{\varphi}'.$$

Importantly, while $\mathbf{Z}_s(\boldsymbol{\rho})$ and $\partial vec[\mathbf{Z}_s(\boldsymbol{\rho})]/\partial\boldsymbol{\rho}'$ depend on the specification of the correlation structure, the first and second derivatives of $\ln f_K(\varepsilon;\boldsymbol{\rho},\boldsymbol{\varphi})$ depend on the specific distributional assumption.

Finally, the (minus) expected value of $\mathbf{h}(\boldsymbol{\rho},\boldsymbol{\varphi})$ will give us the information matrix.

C A reparametrization of the *GH* distribution

To simplify the exposition, we focus on the symmetric case. In the vicinity of Gaussianity, Mencía and Sentana (2012) found that

$$s_\eta\left(\eta=0^+,\psi\right)=-s_\eta\left(\eta=0^-,\psi\right)=2\times s_\psi\left(\eta,\psi=0^+\right)=\sqrt{\frac{K(K+2)}{2}}\times L_2(\varsigma),$$

and $s_\eta\left(\eta,\psi=0^+\right)=s_\psi\left(\eta=0^-,\psi\right)=s_\psi\left(\eta=0^-,\psi\right)=0$. Since $s_\eta\left(\eta=0^+,\psi\right)$ and $s_\eta\left(\eta=0^-,\psi\right)$ have opposite signs, we consider each case separately.

Case I: $\eta\leq 0,\psi\geq 0$: We introduce the following reparametrization:

$$\tau_1=\eta\cdot\psi \quad \text{and} \quad \tau_2=\frac{2\eta+\psi}{\sqrt{5}}.$$

As a result,

$$\eta(\tau_1,\tau_2)=\frac{\sqrt{5}\tau_2-\sqrt{5\tau_2^2-8\tau_1}}{4} \quad \text{and} \quad \psi(\tau_1,\tau_2)=\frac{\sqrt{5}\tau_2+\sqrt{5\tau_2^2-8\tau_1}}{2}.$$

When evaluated at the Gaussian limit,

$$\eta(0,\tau_2)=\frac{1}{4}\sqrt{5}(\tau_2-|\tau_2|) \quad \text{and} \quad \psi(0,\tau_2)=\frac{1}{2}\sqrt{5}(\tau_2+|\tau_2|),$$

whence

$$\begin{aligned}
\left.\frac{\partial\eta}{\partial\tau_1}\right|_{\tau_1=0} &= \frac{1}{|\tau_2|\sqrt{5}}, \quad \left.\frac{\partial\eta}{\partial\tau_2}\right|_{\tau_1=0}=\frac{\sqrt{5}}{2}\times\mathbf{1}\{\tau_2<0\},\\
\left.\frac{\partial\psi}{\partial\tau_1}\right|_{\tau_1=0} &= -\frac{2}{|\tau_2|\sqrt{5}}, \text{ and } \left.\frac{\partial\psi}{\partial\tau_2}\right|_{\tau_1=0}=\sqrt{5}\times\mathbf{1}\{\tau_2>0\}.
\end{aligned}$$

When $\tau_1=0,\tau_2>0$, the chain rule implies

$$\begin{aligned}
s_{\tau_1}\left(\tau_1=0,\tau_2>0\right) &= -\frac{1}{|\tau_2|\sqrt{5}}\times\sqrt{\frac{K(K+2)}{2}}L_2(\varsigma),\\
s_{\tau_2}\left(\tau_1=0,\tau_2>0\right) &= 0.
\end{aligned}$$

Similarly, when $\tau_1=0,\tau_2<0$,

$$\begin{aligned}
s_{\tau_1}\left(\tau_1=0,\tau_2<0\right) &= -\frac{2}{|\tau_2|\sqrt{5}}\times\frac{1}{2}\sqrt{\frac{K(K+2)}{2}}L_2(\varsigma),\\
s_{\tau_2}\left(\tau_1=0,\tau_2<0\right) &= 0.
\end{aligned}$$

Notice that we have used the fact that $\tau_1=0,\tau_2>0$, which implies that in the limit $\eta=0$ and $\psi>0$, while for $\tau_1=0,\tau_2<0$ we have $\eta<0$ and $\psi=0$ in the limit.

Case II: $\eta\geq 0,\psi\geq 0$: We introduce the following reparametrization:

$$\tau_1=\eta\cdot\psi \quad \text{and} \quad \tau_2=\frac{2\eta-\psi}{\sqrt{5}}.$$

Analogous calculations deliver

$$
\begin{aligned}
s_{\tau_1}(\tau_1 = 0, \tau_2 > 0) &= \frac{1}{|\tau_2|\sqrt{5}} \times \sqrt{\frac{K(K+2)}{2}} L_2(\varsigma), \\
s_{\tau_2}(\tau_1 = 0, \tau_2 > 0) &= 0,
\end{aligned}
$$

and

$$
\begin{aligned}
s_{\tau_1}(\tau_1 = 0, \tau_2 < 0) &= \frac{2}{|\tau_2|\sqrt{5}} \times \frac{1}{2}\sqrt{\frac{K(K+2)}{2}} L_2(\varsigma), \\
s_{\tau_2}(\tau_1 = 0, \tau_2 < 0) &= 0.
\end{aligned}
$$

D Local power comparisons

We can assess the local power of the different score tests that we have proposed by computing the probability of rejecting the null hypothesis when it is false as a function of the shape parameters $\boldsymbol{\varphi}$ under the assumption that the asymptotic non-central chi-square distributions of the different LM and KT tests provide reliable rejection probabilities in finite samples. But given that the degrees of freedom are the same for copula and distributional tests, we can compare these two approaches against local alternatives by directly comparing their non-centrality parameters. In this regard, we explain in detail in Supplemental Appendix D.1 the way in which we compute the non-centrality parameters of our proposed tests, as well as the non-centrality parameters of distributional tests of Gaussian vs. Student t and Gaussian vs. asymmetric Student t, which ignore that the margins of the copula are Gaussian by construction.

Figures D1a-c depict the non-centrality parameters of symmetric Student t tests under asymmetric Student t local alternatives, while Figures D2a-c do the same for asymmetric Student t tests. In those plots, LM and LM^{NP} denote the LM versions of the copula tests applied to the Gaussian ranks when the marginal distributions of the observations are known and when they are estimated nonparametrically, respectively, while $Dist^{NP}$ indicates the LM version of the distributional test applied to the same ranks when the margins are estimated nonparametrically.

In Figures D1a and D2a we have represented η in the x-axis for fixed values of $\rho = .75$ and $b_k = 0$. As can be seen, the distributional tests have less power than the copula tests when the margins are estimated nonparametrically, which in turn have less power than the copula tests when they are known.

We then look at the non-centrality parameters for different values of ρ in the x-axis for fixed values of $\eta = .1$ and b_k =-.5 in Figures D1b and D2b. Interestingly, LM, LM^{NP} and $Dist^{NP}$ tend to have the same power as ρ approaches zero.

Finally, we plot the non-centrality parameters against asymmetric Student t alternatives with increasing skewness when $\eta = .1$ and $\rho = .75$. Not surprisingly, the Student t tests are not sensitive to the different values of b (Figure D1c), while the asymmetric Student t tests have more power as b moves away from zero.

D.1 Local power calculations

Let $\mathbf{m}_{\boldsymbol{\varphi}n}(\boldsymbol{\rho},\boldsymbol{\varphi})$ denote the h influence functions used to develop the following moment test of $H_0 : \boldsymbol{\varphi} = \mathbf{0}$:

$$M_N = N \times \bar{\mathbf{m}}'_{\boldsymbol{\varphi}N}(\boldsymbol{\rho}_0,\mathbf{0})\boldsymbol{\Psi}^{-1}\bar{\mathbf{m}}_{\boldsymbol{\varphi}N}(\boldsymbol{\rho}_0,\mathbf{0}), \tag{D1}$$

where $\bar{\mathbf{m}}_{\boldsymbol{\varphi}N}(\boldsymbol{\rho}_0,\mathbf{0})$ is the sample average of $\mathbf{m}_{\boldsymbol{\varphi}n}(\boldsymbol{\rho},\boldsymbol{\varphi})$ evaluated under the null, and $\boldsymbol{\Psi}$ is the corresponding asymptotic covariance matrix. In order to obtain the non-centrality parameter of this test under Pitman sequences of local alternatives of the form $H_l : \boldsymbol{\varphi}_N = \bar{\boldsymbol{\varphi}}/\sqrt{N}$, it is convenient to linearize $\mathbf{m}_{\boldsymbol{\varphi}n}(\boldsymbol{\rho}_0,\mathbf{0})$ with respect to $\boldsymbol{\varphi}$ around its true value $\boldsymbol{\varphi}_N$. This linearization yields

$$\sqrt{N}\bar{\mathbf{m}}_{\boldsymbol{\varphi}N}(\boldsymbol{\rho}_0,\mathbf{0}) = \sqrt{N}\bar{\mathbf{m}}_{\boldsymbol{\varphi}N}(\boldsymbol{\rho}_0,\boldsymbol{\varphi}_N) + \frac{1}{\sqrt{N}}\sum_{n=1}^{N}\frac{\partial \mathbf{m}_{\boldsymbol{\varphi}n}(\boldsymbol{\rho},\boldsymbol{\varphi}^*)}{\partial \boldsymbol{\varphi}'}\bar{\boldsymbol{\varphi}},$$

where $\boldsymbol{\varphi}^*$ is some "intermediate" value between $\boldsymbol{\varphi}_N$ and $\mathbf{0}$. As a result,

$$\sqrt{N}\bar{\mathbf{m}}_{\boldsymbol{\varphi}N}(\boldsymbol{\rho}_0,\mathbf{0}) \xrightarrow{d} N[\mathbf{M}(\boldsymbol{\rho}_0,\mathbf{0})\bar{\boldsymbol{\varphi}},\boldsymbol{\Psi}],$$

under standard regularity conditions, where

$$\mathbf{M}(\boldsymbol{\rho}_0,\mathbf{0}) = E\left[\frac{\partial \mathbf{m}_{\boldsymbol{\varphi}n}(\boldsymbol{\rho},\mathbf{0})}{\partial \boldsymbol{\varphi}'}\right],$$

so that the non-centrality parameter of the moment test (D1) will be

$$\bar{\boldsymbol{\varphi}}'\mathbf{M}'(\boldsymbol{\rho}_0,\mathbf{0})\boldsymbol{\Psi}^{-1}\mathbf{M}(\boldsymbol{\rho}_0,\mathbf{0})\bar{\boldsymbol{\varphi}}. \tag{D2}$$

On this basis, we can easily obtain the limiting probability of M_N exceeding some prespecified quantile of a central χ^2_h distribution from the cdf of a non-central χ^2 distribution with h degrees of freedom and non-centrality parameter (D2).

Finally, note that (D2) remains valid when we replace $\boldsymbol{\rho}_0$ by its ML estimator under the null if $\mathbf{m}_{\boldsymbol{\varphi}n}(\boldsymbol{\rho},\mathbf{0})$ and the scores corresponding to $\boldsymbol{\rho}$ are asymptotically uncorrelated when H_0 is true, as in all our tests. In addition, both $\mathbf{M}(\boldsymbol{\rho}_0,\mathbf{0})$ and $\boldsymbol{\Psi}$ coincide with the $(2,2)$ block of the information matrix when $\mathbf{m}_{\boldsymbol{\varphi}n}(\boldsymbol{\rho},\boldsymbol{\varphi})$ are the scores with respect to $\boldsymbol{\varphi}$.

To simplify the exposition, in what follows we focus on the bivariate case.

D.2 Student t alternatives

Propositions 1 and 5 contain expressions for $s_\eta(\rho)$ and $n_\eta(\rho)$, respectively, which allow us to compute

$$s^c_\eta(\rho) = s_\eta(\rho) - n_\eta(\rho).$$

Given that in the bivariate case both $V[m_{\eta n}(\rho)]$ and $E[\partial m_{\eta n}(\rho)/\partial\eta]$ coincide with the $(2,2)$ block of the information matrix, we only need to compute

$$V[s_\eta(\rho)] = 1 + \frac{3}{4}\rho^2$$

and

$$V\left[s_{\eta}^{c}\left(\rho\right)\right]=1+\frac{3}{4}\rho^{2}+\frac{3}{16}\left(\rho^{4}+\rho^{8}\right)$$

in order to obtain the corresponding non-centrality parameters. Similarly, for the distributional version of the test, we have that $m_{\eta n}(\rho)=d^{c}\left(\rho\right)$ with

$$d^{c}\left(\rho\right)=2L_{2}\left[\varsigma(\rho)\right]-\sqrt{\frac{3}{2}}\left[H_{4}(y_{1})+H_{4}(y_{2})\right].$$

Hence

$$V\left[d^{c}\left(\rho\right)\right]=1+3\rho^{4}$$

and

$$cov\left[d^{c}\left(\rho\right),s_{\eta}^{c}\left(\rho\right)\right]=1-\frac{3}{4}\rho^{6}.$$

D.3 Asymmetric Student t alternatives

The required quantities to compute the non-centrality parameters of the score test in the bivariate case are

$$V\left[s_{b_{k}}\left(\rho\right)\right]=26+24\rho^{2}+48\rho^{4},\ \text{for }k=1,2$$

$$cov\left[s_{b_{1}}\left(\rho\right),s_{b_{2}}\left(\rho\right)\right]=48\rho+26\rho^{3}+24\rho^{5},$$

$$V[s_{b_{k}}^{c}\left(\rho\right)]=2+\frac{2}{3}\left(\rho^{2}+\rho^{4}\right)+\frac{4}{3}\rho^{6},\ \text{for }k=1,2$$

and

$$cov[s_{b_{1}}^{c}\left(\rho\right),s_{b_{2}}^{c}\left(\rho\right)]=\frac{10}{3}\rho^{3}+\frac{2}{3}\left(\rho^{5}+\rho^{7}\right),$$

while $cov\left[s_{\eta}\left(\rho\right),s_{b_{k}}\left(\rho\right)\right]=cov[s_{\eta}^{c}\left(\rho\right),s_{b_{k}}^{c}\left(\rho\right)]=0$, for $k=1,2$. The same argument can be applied to the distributional test, yielding

$$d_{b_{1}}^{c}\left(\rho\right)=-2\left[\sqrt{\frac{3}{2}}H_{3}(y_{1})+\rho\sqrt{\frac{2}{3}}H_{3}(y_{2})\right]+y_{1}\left[\varsigma(\rho)-4\right]$$

and

$$d_{b_{2}}^{c}\left(\rho\right)=-2\left[\sqrt{\frac{3}{2}}H_{3}(y_{2})+\rho\sqrt{\frac{2}{3}}H_{3}(y_{1})\right]+y_{2}\left[\varsigma(\rho)-4\right].$$

As in the case of the score test, $d_{b_{k}}^{c}\left(\rho\right)$ for $k=1,2$ is orthogonal to $d_{\eta}^{c}\left(\rho\right)$. Therefore, the additional quantities required to compute the corresponding non-centrality parameters are

$$V\left[d_{b_{k}}^{c}\left(\rho\right)\right]=2-\frac{16}{3}\rho^{2}+8\rho^{4},\ \text{for }k=1,2$$

$$cov\left[d_{b_{1}}^{c}\left(\rho\right),m_{b_{2}}^{c}\left(\rho\right)\right]=-4\rho+6\rho^{3}+\frac{8}{3}\rho^{5},$$

$$cov\left[m_{b_{k}}^{c}\left(\rho\right),s_{b_{k}}^{c}\left(\rho\right)\right]=2-\frac{10}{3}\rho^{2}-2\rho^{4}-\frac{4}{3}\rho^{6},\ \text{for }k=1,2,$$

and

$$cov\left[m_{b_{1}}^{c}\left(\rho\right),s_{b_{2}}^{c}\left(\rho\right)\right]=-2\rho+\frac{2}{3}\rho^{3}-\frac{10}{3}\rho^{5}.$$

D.4 Interpretation of copula and distributional tests

D.4.1 When marginals are known

We can easily express both score copula tests as well as distributional LM tests in terms of Hermite polynomials of the marginal Gaussian ranks. Taking into account that $m_{b_2}(y_1, y_2; \rho) = m_{b_1}(y_2, y_1; \rho)$ and $d_{b_2}(y_1, y_2; \rho) = d_{b_1}(y_2, y_1; \rho)$, the relevant coefficients are in Table D1.

In order to characterize the loss of power of the distributional version of the test, for a given element φ of $\boldsymbol{\varphi}$ we could write

$$d_\varphi(\rho) = \beta_\varphi s_\varphi(\rho) + u_\varphi,$$

where

$$\beta_\varphi = \frac{cov\left[d_\varphi(\rho), s_\varphi(\rho)\right]}{V[s_\varphi(\rho)]},$$

so that the non-centrality parameter of $d_\varphi(\rho)$ under the sequence of local alternatives $H_l : \varphi_N = \bar{\varphi}/\sqrt{N}$ can be written as

$$\frac{\beta_\varphi^2 V[s_\varphi(\rho)]}{\beta_\varphi^2 V[s_\varphi(\rho)] + V(u_\varphi)}.$$

For instance, when $\varphi = \eta$ we have that

$$V[s_\eta(\rho)] = 1 + \frac{3}{4}\rho^2$$

and

$$cov\left[d_\eta(\rho), s_\eta(\rho)\right] = 1,$$

so that the power reduction of the distributional test relative to the copula one is captured by

$$V(u_\eta) = 4 - \frac{4}{4 + 3\rho^2},$$

where we have used the fact that $V[d_\eta(\rho)] = 4$. Similarly, doing the same calculations for $\varphi = b_i$, we find that

$$V\left[\begin{pmatrix} m_{b_1}(\rho) \\ m_{b_2}(\rho) \end{pmatrix}\right] = \begin{bmatrix} 2 & 2\rho^3 \\ 2\rho^3 & 2 \end{bmatrix}, \quad V\left[\begin{pmatrix} d_{b_1}(\rho) \\ d_{b_2}(\rho) \end{pmatrix}\right] = \begin{bmatrix} 8 & 8\rho \\ 8\rho & 8 \end{bmatrix}$$

and

$$cov\left[\begin{pmatrix} m_{b_1}(\rho) \\ m_{b_2}(\rho) \end{pmatrix}, \begin{pmatrix} d_{b_1}(\rho) \\ d_{b_2}(\rho) \end{pmatrix}'\right] = \begin{bmatrix} 2 - 4\rho^2 & -2\rho \\ -2\rho & 2 - 4\rho^2 \end{bmatrix}.$$

In this way, it is clear that for b_1,

$$d_{b_1}(\rho) = \frac{1 - \rho^2}{1 + \rho^2 + \rho^4} m_{b_1}(\rho) - \frac{\rho + 2\rho^3}{1 + \rho^2 + \rho^4} m_{b_2}(\rho) + u_{b_1},$$

so that the power reduction of the distributional test relative to the copula one is captured by

$$V(u_{b_k}) = \frac{6(1 + 2\rho^2)}{1 + \rho^2 + \rho^4}, \text{ for } k = 1, 2,$$

because $V[d_\eta(\rho)] = 4$.

D.4.2 Accounting for margins uncertainty

Direct application of Proposition 5 yields

$$n_\eta(\rho) = \frac{1}{4}\sqrt{\frac{3}{2}}\rho^2\left[H_4(y_1)+H_4(y_2)\right], \quad n_{b_k}(\rho) = \sqrt{\frac{2}{3}}\rho\left[\rho H_3(y_1)+H_3(y_2)\right],$$

for $k = 1, 2$ and $n_{b_2}(y_1, y_2; \rho) = n_{b_1}(y_2, y_1; \rho)$. Analogous calculations for the distributional test moments deliver

$$n^d_\eta(\rho) = \sqrt{\frac{3}{2}}\left[H_4(y_1)+H_4(y_2)\right], \quad n_{b_1}(\rho) = \sqrt{6}H_3(y_1) + 2\rho\sqrt{\frac{2}{3}}H_3(y_2),$$

and again $n_{b_2}(y_1, y_2; \rho) = n_{b_1}(y_2, y_1; \rho)$. In Table D2 we summarize the modified moments that account for nonparametric estimation of the marginals.

Again, in order to characterize the loss of power of the distributional version of the test we could write

$$d^{np}_\varphi(\rho) = \beta^{np}_\varphi s^{np}_\varphi(\rho) + u^{np}_\varphi,$$

where

$$\beta^{np}_\varphi = \frac{cov\left[d^{np}_\varphi(\rho), s_\varphi(\rho)\right]}{cov\left[s^{np}_\varphi(\rho), s_\varphi(\rho)\right]},$$

so that the non-centrality parameter of $d_\varphi(\rho)$ under the sequence of local alternatives $H_l : \varphi_N = \bar{\varphi}/\sqrt{N}$ can be written as

$$\frac{\beta^2_\varphi cov\left[s^{np}_\varphi(\rho), s_\varphi(\rho)\right]}{\beta^2_\varphi cov\left[s^{np}_\varphi(\rho), s_\varphi(\rho)\right] + V(u^{np}_\varphi)}$$

because $cov\left[s^{np}_\varphi(\rho), u^{np}_\varphi\right] = 0$. For instance, when $\varphi = \eta$ we have that

$$cov\left[s^{np}_\eta(\rho), s_\eta(\rho)\right] = 1 + \frac{3}{4}\rho^2$$

and

$$cov\left[d^{np}_\eta(\rho), s_\eta(\rho)\right] = 1,$$

so that the power reduction of the distributional test relative to the copula one is captured by

$$V(u_\eta) = \frac{12(1+\rho^2)(\rho+2\rho^3)^2}{(4+3\rho^2)^2},$$

where we have used the fact that $V[d^{np}_\eta(\rho)] = 1 + 3\rho^4$.

E Computational details

E.1 Simulation of random vectors

We simulate the distribution under the null and the symmetric Student t, as well as gamma and uniform random variables underlying the generation of the asymmetric Student t and discrete location-scale mixture of normals, using off-the-shelf MATLAB routines. Namely, we use `mvnrnd.m` for the bivariate normal, `mvtrnd.m` times $\sqrt{(\nu-2)/2}$, where ν denotes the degrees of freedom, for the bivariate symmetric Student t, `gamrnd.m` for the gamma distribution, and `rand.m` for the uniform. For the remaining ones, the procedure is as follows.

E.1.1 Generalized hyperbolic distributions

The simplest way of simulating a *GH* distribution exploits its interpretation as a location-scale mixture of normals in which the mixing variable is a Generalized Inverse Gaussian (*GIG*). Specifically, if $\boldsymbol{\varepsilon}$ is a *GH* vector, then it can be expressed as

$$\boldsymbol{\varepsilon} = \boldsymbol{\alpha} + \boldsymbol{\Upsilon}\boldsymbol{\beta}\xi^{-1} + \xi^{-\frac{1}{2}}\boldsymbol{\Upsilon}^{\frac{1}{2}}\boldsymbol{\varepsilon}^{\circ}, \tag{E1}$$

where $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^K$, $\boldsymbol{\Upsilon}$ is a symmetric positive definite matrix of order K, $\boldsymbol{\varepsilon}^{\circ} \sim iid\, N(\mathbf{0}, \mathbf{I}_K)$ and the positive mixing variable ξ is an independent *iid GIG* with parameters $-\nu$, γ and δ, or $\xi \sim GIG\,(-\nu, \gamma, \delta)$ for short, where $\nu \in \mathbb{R}$ and $\gamma, \delta \in \mathbb{R}^+$ (see Jørgensen (1982) and Johnson, Kotz and Balakrishnan (1994) for details). Since $\boldsymbol{\varepsilon}$ given ξ is Gaussian with conditional mean $\boldsymbol{\alpha} + \boldsymbol{\Upsilon}\boldsymbol{\beta}\xi^{-1}$ and covariance matrix $\boldsymbol{\Upsilon}\xi^{-1}$, it is clear that $\boldsymbol{\alpha}$ and $\boldsymbol{\Upsilon}$ play the roles of location vector and dispersion matrix, respectively. There is a further scale parameter, δ, two other scalars, ν and γ, to allow for flexible tail modelling, and the vector $\boldsymbol{\beta}$, which introduces skewness in this distribution, although for testing purposes it is more convenient to work with $\eta = -.5\nu^{-1}$ and $\psi = (1+\gamma)^{-1}$. The distribution of $\boldsymbol{\varepsilon}$ becomes a simple scale mixture of normals, and thereby spherical, when $\boldsymbol{\beta}$ is zero. In the symmetric and asymmetric Student t cases, ξ reduces to a gamma random variable with mean N and shape parameter ν, which is the most important special case of the *GIG*. In that case, the relevant expressions for $\boldsymbol{\alpha}$ and $\boldsymbol{\Upsilon}$ become

$$\boldsymbol{\alpha} = -c(\boldsymbol{\beta}, \eta)\boldsymbol{\beta} \quad \text{and} \quad \boldsymbol{\Upsilon} = \frac{1}{c(\boldsymbol{\beta}, \eta)}\left\{\mathbf{I}_K - \frac{[c(\boldsymbol{\beta}, \eta) - 1]}{\boldsymbol{\beta}'\boldsymbol{\beta}}\boldsymbol{\beta}\boldsymbol{\beta}'\right\},$$

where

$$c(\boldsymbol{\beta}, \eta) = \frac{1-4\eta}{2\eta}\,\frac{\sqrt{1 + 8\boldsymbol{\beta}'\boldsymbol{\beta}\eta/(1-4\eta)} - 1}{2\boldsymbol{\beta}'\boldsymbol{\beta}}.$$

E.1.2 Skew *t*-distributions

The family of multivariate Skew t distributions is an alternative extension of the multivariate Student t family via the introduction of another vector of parameters $\boldsymbol{\alpha} \in \mathbb{R}^K$ which regulates asymmetry. Specifically, when $\boldsymbol{\alpha} = \mathbf{0}$, the Skew t-distribution reduces to the symmetric multivariate Student t. As in the case of the *GH*, we choose its scale and location parameters so that the mean vector is 0 and the covariance matrix the identity. For additional information, see Section 6.2 of Azzalini and Capitanio (2014).

E.2 Monte Carlo details

The Monte Carlo analysis of the properties of our tests when we obtain the critical values through the parametric bootstrap can be divided in two main blocks:

1. Construction of the table with critical values.

2. Estimation of the correlation parameters and evaluation of the test size and power.

E.2.1 Construction of the table with critical values

To obtain the distribution of the test as a function of the estimated ρ's, the steps of the code are the following:

1. Create a grid of H points, $\mathcal{H} = \{\rho^{(1)}, ..., \rho^{(h)}, ..., \rho^{(H)}\}$, that covers $(-1, 1)$. In our design, we consider 199 equally spaced points between $-.99$ and $.99$.

2. Fix the seed to s_1.

3. For each point $h = 1, ..., H$:

   (a) Simulate data $\mathbf{X}_{N\times K}$ with exponential margins and Gaussian copula. Obviously, the choice of margins is inconsequential when we assume them known or when we estimate them nonparametrically. We use $N = 200$, 800 and 3,200, and $K = 2$ and 10.

      i. Simulate $\tilde{X}_i$ from $N(\mathbf{0}, \mathbf{P}_K^{(h)})$ iid across n.

      ii. $X_{nk} = F_k^{-1}[\tilde{F}_k(\tilde{X}_{nk}); \lambda_{k0}]$, with $F_k(x) = 1 - e^{-\lambda_{k0}x}$ and $\tilde{F}_k(x)$ is the true distribution of $\tilde{X}_{nk}$, i.e. under the null, $\tilde{F}_k(x) = \Phi(x)$. (The parameters we used are $\lambda_{10} = \lambda_{20} = 1$.)

   (b) Keep the copula and convert the marginal distributions to Gaussian to get the Gaussian ranks $\mathbf{Y}_{N\times K}$.

      i. For known margins, $Y_{nk}^k = \Phi^{-1}[F_k(X_{nk}; \lambda_{k0})] = \Phi^{-1}[\tilde{F}_k(\tilde{X_{nk}})]$. Under the null, we naturally use $Y_{nk}^k = \tilde{X}_{nk}$ directly.

      ii. For parametric margins, $Y_{nk}^p = \Phi^{-1}[F_k(X_{nk}; \hat{\lambda}_k)]$, with $\hat{\lambda}_k$ estimated by ML.

      iii. For non-parametric margins, $Y_{nk}^n = \Phi^{-1}[\hat{F}_k(X_{nk})]$, where $\hat{F}(x_{nk})$ denotes the empirical CDF of $\{x_{nk}\}_{n=1}^N$.

   (c) Estimate the correlation parameters $\hat{\rho}^k, \hat{\rho}^p, \hat{\rho}^n$ by ML using $\mathbf{Y}^k, \mathbf{Y}^p, \mathbf{Y}^n$.

   (d) Compute the tests evaluated at the parameter estimates in step c: $Test^k(s; h)$, $Test^p(s; h)$ and $Test^n(s; h)$, say.

Steps 3a–c are repeated 10,000 times, saving the test statistics for each $\rho^{(h)}$.

E.2.2 Estimation of the correlation parameters and evaluation of the test size and power

To obtain the size or power of the tests, the steps of the code are the following:

1. Load the results obtained in E.2.1.

2. For each test statistic, compute the relevant $(1 - \alpha)$ quantiles of the $Test(s, h)$ for each h: Q^α say.

3. Fix the seed to $s_2 \neq s_1$.

4. Simulate $\tilde{\mathbf{X}}$ from the relevant joint distribution $\tilde{F}$ for one of the two chosen correlation matrices (e.g. Gaussian with $\rho = .25$ for size).

5. Compute the $Y^{(k)}, Y^{(p)}, Y^{(n)}$ of these simulated observations following 3(a)ii and 3b and then estimate the parameter ρ by ML. For the asymmetric Student t and the Skew t Student alternatives, the calculation of $\tilde{F}_k$ is very time-consuming, so we did not calculate it for each sample. Instead, we first generated $N = 5,000,000$ draws from $\tilde{F}$ and calculated the empirical marginal cdf for $\tilde{X}_{nk}$. Given that N is very large, for all practical purposes $\tilde{F}_{nk} \approx \hat{F}_k(\tilde{X}_{nk})$. We save $\tilde{X}_{nk}$ and the approximate value of $\tilde{F}_{nk}$, and then draw samples of $\tilde{\mathbf{X}}$ using our bootstrap procedure.

6. Compute the test evaluated at $\hat{\rho}$: $Test(s)$, say.

7. Find the critical value (c^{α}) of the test at significance level α through a linear interpolation of the quantiles computed from the results in E.2.1.

Steps 4 to 7 are repeated 10,000 times and the number of times $Test(s) > c^{\alpha}$ is recorded for each test to compute size and power.

E.3 Pooled estimation and testing

For a given cross-section, we have $Y_t = \{(y^t_{11}, y^t_{21}), ..., (y^t_{1n}, y^t_{2n}), ..., (y^t_{1N_t}, y^t_{2N_t})\}$. The full sample would then consist of $\sum_{t=1}^{T} N_t$ bivariate observations $\mathbf{Y} = \{Y_1, ..., Y_T\}$. At each t, we can compute the average modified score, accounting for non-parametric estimation of the margins:

$$\bar{\mathbf{s}}^c_{\phi t}(Y_t; \rho) = \frac{1}{N_t} \sum_{n=1}^{N_t} \begin{bmatrix} s^c_{\rho}(Y_{tn}; \rho) \\ \mathbf{s}^c_{\boldsymbol{\varphi}}(Y_{tn}; \rho) \end{bmatrix},$$

which is the basis for the pooled average corrected score $\bar{\mathbf{s}}^c_{\phi}(Y_t; \rho) = T^{-1} \sum_{t=1}^{T} \bar{\mathbf{s}}^c_{\phi t}(Y_t; \rho)$.

As for Spearman's correlation coefficient, we can simplify our calculations by noticing that for large N, $\sum_{n=1}^{N_t} \Phi(y_{tn}) \approx 1/2$ and $\sum_{n=1}^{N_t} \Phi^2(y_{tn}) \approx 1/3$ so that

$$\frac{\sqrt{N_t}}{N_t} \frac{\sum_{n=1}^{N_t} \Phi(y_{1n})\Phi(y_{2n}) - 1/4}{1/12}$$

is the relevant moment function required to compute HAC robust standard errors.

Finally, to estimate Pearson correlation coefficient and its corresponding robust standard error, we can consider the following moment functions

$$\mathbf{m}(X_t) = \frac{1}{N_t} \sum_{n=1}^{N_t} \left[x^t_{1n},\ x^t_{2n},\ (x^t_{1n})^2,\ (x^t_{2n})^2,\ x^t_{1n}x^t_{2n}\right]'.$$

Specifically, if we introduce $g : \mathbb{R}^5 \to \mathbb{R}^3$,

$$\mathbf{g}\left[\mathbf{m}(X_t)\right] = \begin{bmatrix} \mathbf{m}_3(X_t) - \mathbf{m}_1^2(X_t) \\ \mathbf{m}_4(X_t) - \mathbf{m}_2^2(X_t) \\ \mathbf{m}_5(X_t) - \mathbf{m}_1(X_t)\mathbf{m}_2(X_t) \end{bmatrix} \text{ so that } \frac{\partial \mathbf{g}}{\partial \mathbf{m}} = \begin{bmatrix} -2m_1 & 0 & 1 & 0 & 0 \\ 0 & -2m_2 & 0 & 1 & 0 \\ -m_2 & -m_1 & 0 & 0 & 1 \end{bmatrix}$$

and $\ell : \mathbb{R}^3 \to [-1, 1]$,

$$\ell \{\mathbf{g} [\mathbf{m}(X_t)]\} = \frac{g_3}{\sqrt{g_1 g_2}} \text{ so that } \frac{\partial \ell}{\partial \mathbf{g}} = \left[\frac{-g_3}{2g_1\sqrt{g_1 g_2}}, \frac{-g_3}{2g_2\sqrt{g_1 g_2}}, \frac{1}{\sqrt{g_1 g_2}} \right]$$

we can apply the Delta method twice to obtain the corresponding asymptotic variance.

E.4 Variances of the moment functions

Below we present the relevant expressions for the bivariate copula testing procedures. See Amengual and Sentana (2015) for the corresponding expressions for the trivariate case.

E.4.1 Known marginals

The variances are

$$V [s_\eta (\rho)] = 1 + \frac{3}{4}\rho^2$$

and

$$V [m_{b_k} (\rho)] = 2, \quad \text{for} \quad k = 1, 2,$$

while the covariances are

$$cov [m_{b_1} (\rho), m_{b_2} (\rho)] = 2\rho^3$$

and

$$cov [s_\eta (\rho), m_{b_k} (\rho)] = 0, \text{ for } k = 1, 2.$$

E.4.2 Accounting for non-parametric estimation of the marginals

The variances are

$$V[s_\eta^{np} (\rho)] = 1 + \frac{3}{4}\rho^2 + \frac{3}{16} \left(\rho^4 + \rho^8\right)$$

and

$$V[m_{b_k}^{np} (\rho)] = 2 + \frac{2}{3} \left(\rho^2 + \rho^4 + 2\rho^6\right), \quad \text{for} \quad k = 1, 2,$$

while the covariances are

$$cov[m_{b_1}^{np} (\rho), m_{b_2}^{np} (\rho)] = 2\rho^3 + \frac{2}{3}\rho^3 \left(2 + \rho^2 + \rho^4\right)$$

and

$$cov[s_\eta^{np} (\rho), m_{b_k}^{np} (\rho)] = 0, \quad \text{for} \quad k = 1, 2.$$

F Additional Monte Carlo results

In this section we present the finite sample performance of the proposed tests for the same designs as in the main text when the correlation coefficient ρ is .75. Table F1 reports the parametric bootstrap rejection rates for all the different samples sizes and significance levels we consider. Specifically, Panel A reports rejection rates under the null at the 1%, 5% and 10% levels for the bivariate case while Panel B does the same for $K = 10$.

Similarly, Tables F2–4 report the Monte Carlo rejection rates at the 1%, 5% and 10% significance levels for the symmetric, asymmetric and Skew t, respectively. As in the case of $\rho = .25$, the behavior of the different test statistics is in accordance with expectations. In line with the evidence on local power in Supplemental Appendix D, the rejection rates are higher the higher the correlation.

G Additional empirical results

G.1 Industry level results

Industry definitions: Non Durables: Consumer NonDurables – Food, Tobacco, Textiles, Apparel, Leather, Toys; Durables :Consumer Durables – Cars, TV's, Furniture, Household Appliances; Manufacturing: Manufacturing – Machinery, Trucks, Planes, Off Furn, Paper, Com Printing; Energy: Oil, Gas, and Coal Extraction and Products; Chemicals: Chemicals and Allied Products; Business : Business Equipment – Computers, Software, and Electronic Equipment; Telecom: Telephone and Television Transmission; Utilities; Shops: Wholesale, Retail, and Some Services (Laundries, Repair Shops); Healthcare: Healthcare, Medical Equipment, and Drugs; Financials; and Other: Mines, Constr, BldMt, Trans, Hotels, Bus Serv, Entertainment. See Ken French's website for details.

As can be seen in Table G1, both Spearman and Gaussian rank correlations have the expected sign for all the industries when looking at momentum strategies, and the same is true for reversals with the exception of Telecommunications. In contrast, Pearson correlation estimates have the wrong sign for most of the industries, especially for short term reversals, which once again confirms their sensitivity to influential observations.

In Table G2 we report the Gaussian copula test statistics, with KT–t and KT–At denoting the Kuhn-Tucker versions of the tests against Student t and asymmetric Student t copulas, and Skew the Lagrange multiplier test based on the two moment conditions $m_{b_k}(\rho)$ in Proposition 3. We omit the Lagrange multiplier versions since they are numerically identical in our data. As can be seen, in all cases we reject the null hypothesis of a Gaussian copula for both short term reversals and momentum by a long margin.

Finally, in Table G3, we report the resulting pooled estimates of the correlation and shape parameters based on simulated sample paths of size 100,000. We find moderate negative tail dependence but quite substantive "leptokurtosis".

G.2 Trading implications of a non-Gaussian copula

The dependence between the (Gaussian) rank of a stock in period t and the rank of some of its characteristics in period $t-1$ we have found allows us to design sound trading strategies along the following lines:

1. We look at the rank of the chosen characteristic of an individual stock over the relevant

observation period.

2. Conditional on that rank, our estimated copula allows us to make probabilistic predictions about the rank of the return on that stock over the next month.

3. If the predicted probability of the rank being high is large, we buy the stock.

4. If the predicted probability of the rank being low is large, we sell it short.

5. Otherwise, we do not hold any position on it.

The Gaussian rank correlation is obviously very important in deriving probabilistic predictions about the rank of a stock over the next month given the current rank of its characteristic, but it is by no means the only determinant. In general, non-linear tail dependence also matters. To illustrate the importance of looking at the entire copula, we use the parameter estimates for the Gaussian, Student t and asymmetric Student t copulas in Table 5 to compute the probabilities that a stock will be in the bottom 30, middle 40 or top 30 percentiles during period t conditional on the same stock being in the bottom 5%, next 25%, middle 40%, next 25% and top 5% according to its short-term reversal or momentum characteristics at time $t-1$. A possible trading rule would be as follows: if the predicted probability of the rank being in the top/bottom 30% percentile is larger than the respective probabilities of being in the bottom/top 30% and middle 40%, we buy/short-sell the stock; otherwise, we do not hold any position on it (see Gagliardini, Gouriéroux and Rubin (2014) for a formal discussion of portfolio choice based on the maximization of the expected utility of the ranks).

Figure G1 presents the results for short-term reversals. As can be observed, the estimated negative correlation is not large enough for the Gaussian copula to suggest any position. In contrast, the non-linear dependence of both the symmetric and asymmetric Student t copulas results in long positions on recent losers (5%) and short positions on recent winners (95%).

Figure G2 contains the result of a similar exercise with momentum strategies. Once again, we find that the small positive correlation of the Gaussian copula is too weak to lead to any position. But the non-linear dependence of the symmetric Student t copula changes the probabilities enough to recommend taking short positions on past losers (5%) and long positions on past winners (95%). Somewhat surprisingly, though, the negative tail dependence of the asymmetric Student t in this case, which is higher than for short-term reversals, leads to the opposite trading strategy for the case of winners.

Table D1: Hermite polynomial coefficients for bivariate score copula tests and distributional LM tests when marginals are known

	Copula LM test		Distributional LM test	
Hermite polynomial	$s_\eta(\rho)$	$m_{b_1}(\rho)$	$d_\eta(\rho)$	$d_{b_1}(\rho)$
1	$\frac{2\rho^4+\rho^2}{(1-\rho^2)^2}$	0	$\frac{2\rho^4+\rho^2}{(1-\rho^2)^2}$	0
$H_1(y_1)$	0	$\frac{2\rho^2}{1-\rho^2}$	0	$\frac{4\rho^2}{1-\rho^2}$
$H_1(y_2)$	0	$-\frac{2(\rho^3+\rho)}{1-\rho^2}$	0	$-\frac{2\rho}{1-\rho^2}$
$H_2(y_1)$	$\frac{3\rho^2(\rho^2+3)}{2\sqrt{2}(1-\rho^2)^2}$	0	$\frac{3\sqrt{2}\rho^2}{(1-\rho^2)^2}$	0
$H_1(y_1)H_1(y_2)$	$-\frac{2(2\rho^3+\rho)}{(1-\rho^2)^2}$	0	$-\frac{2(2\rho^3+\rho)}{(1-\rho^2)^2}$	0
$H_2(y_2)$	$\frac{3\rho^2(\rho^2+3)}{2\sqrt{2}(1-\rho^2)^2}$	0	$\frac{3\sqrt{2}\rho^2}{(1-\rho^2)^2}$	0
$H_3(y_1)$	0	0	0	$\frac{\sqrt{6}}{1-\rho^2}$
$H_2(y_1)H_1(y_2)$	0	$-\frac{\sqrt{2}\rho}{1-\rho^2}$	0	$-\frac{2\sqrt{2}\rho}{1-\rho^2}$
$H_1(y_1)H_2(y_2)$	0	$\frac{\sqrt{2}(\rho^2+1)}{1-\rho^2}$	0	$\frac{\sqrt{2}}{1-\rho^2}$
$H_3(y_2)$	0	$-\frac{\sqrt{6}\rho}{1-\rho^2}$	0	0
$H_4(y_1)$	$\frac{\sqrt{\frac{3}{2}}\rho^2}{(1-\rho^2)^2}$	0	$\frac{\sqrt{\frac{3}{2}}}{(1-\rho^2)^2}$	0
$H_3(y_1)H_1(y_2)$	$-\frac{\sqrt{\frac{3}{2}}\rho(\rho^2+3)}{2(1-\rho^2)^2}$	0	$-\frac{\sqrt{6}\rho}{(1-\rho^2)^2}$	0
$H_2(y_1)H_2(y_2)$	$\frac{2\rho^2+1}{(1-\rho^2)^2}$	0	$\frac{2\rho^2+1}{(1-\rho^2)^2}$	0
$H_1(y_1)H_3(y_2)$	$-\frac{\sqrt{\frac{3}{2}}\rho(\rho^2+3)}{2(1-\rho^2)^2}$	0	$-\frac{\sqrt{6}\rho}{(1-\rho^2)^2}$	0
$H_4(y_2)$	$\frac{\sqrt{\frac{3}{2}}\rho^2}{(1-\rho^2)^2}$	0	$\frac{\sqrt{\frac{3}{2}}}{(1-\rho^2)^2}$	0

Table D2: Hermite polynomial coefficients for bivariate score copula tests and distributional LM tests when marginals are estimated nonparametrically

Hermite polynomial	Copula LM test		Distributional LM test	
	$s_{\eta}^{np}(\rho)$	$m_{b_1}^{np}(\rho)$	$d_{\eta}^{np}(\rho)$	$d_{b_1}^{np}(\rho)$
1	$\frac{2\rho^4+\rho^2}{(1-\rho^2)^2}$	0	$\frac{2\rho^4+\rho^2}{(1-\rho^2)^2}$	0
$H_1(y_1)$	0	$\frac{2\rho^2}{1-\rho^2}$	0	$\frac{4\rho^2}{1-\rho^2}$
$H_1(y_2)$	0	$-\frac{2(\rho^3+\rho)}{1-\rho^2}$	0	$-\frac{2\rho}{1-\rho^2}$
$H_2(y_1)$	$\frac{3\rho^2(\rho^2+3)}{2\sqrt{2}(1-\rho^2)^2}$	0	$\frac{3\sqrt{2}\rho^2}{(1-\rho^2)^2}$	0
$H_1(y_1)H_1(y_2)$	$-\frac{2(2\rho^3+\rho)}{(1-\rho^2)^2}$	0	$-\frac{2(2\rho^3+\rho)}{(1-\rho^2)^2}$	0
$H_2(y_2)$	$\frac{3\rho^2(\rho^2+3)}{2\sqrt{2}(1-\rho^2)^2}$	0	$\frac{3\sqrt{2}\rho^2}{(1-\rho^2)^2}$	0
$H_3(y_1)$	0	$\sqrt{\frac{2}{3}}\rho^2$	0	$\frac{\sqrt{6}\rho^2}{1-\rho^2}$
$H_2(y_1)H_1(y_2)$	0	$-\frac{\sqrt{2}\rho}{1-\rho^2}$	0	$-\frac{2\sqrt{2}\rho}{1-\rho^2}$
$H_1(y_1)H_2(y_2)$	0	$\frac{\sqrt{2}(\rho^2+1)}{1-\rho^2}$	0	$\frac{\sqrt{2}}{1-\rho^2}$
$H_3(y_2)$	0	$-\frac{\sqrt{\frac{2}{3}}\rho(\rho^2+2)}{1-\rho^2}$	0	$-2\sqrt{\frac{2}{3}}\rho$
$H_4(y_1)$	$\frac{\sqrt{\frac{3}{2}}\rho^2(\rho^4-2\rho^2+5)}{4(1-\rho^2)^2}$	0	$-\frac{\sqrt{\frac{3}{2}}\rho^2(\rho^2-2)}{(1-\rho^2)^2}$	0
$H_3(y_1)H_1(y_2)$	$-\frac{\sqrt{\frac{3}{2}}\rho(\rho^2+3)}{2(1-\rho^2)^2}$	0	$-\frac{\sqrt{6}\rho}{(1-\rho^2)^2}$	0
$H_2(y_1)H_2(y_2)$	$\frac{2\rho^2+1}{(1-\rho^2)^2}$	0	$\frac{2\rho^2+1}{(1-\rho^2)^2}$	0
$H_1(y_1)H_3(y_2)$	$-\frac{\sqrt{\frac{3}{2}}\rho(\rho^2+3)}{2(1-\rho^2)^2}$	0	$-\frac{\sqrt{6}\rho}{(1-\rho^2)^2}$	0
$H_4(y_2)$	$\frac{\sqrt{\frac{3}{2}}\rho^2(\rho^4-2\rho^2+5)}{4(1-\rho^2)^2}$	0	$-\frac{\sqrt{\frac{3}{2}}\rho^2(\rho^2-2)}{(1-\rho^2)^2}$	0

Table F1: Rejection rates under the null at 1%, 5%, and 10% significance levels

Margins		$N=200$			$N=800$			$N=3,200$		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A: $K=2$ and $\rho=0.75$										
Known	LM-t	9.5	4.6	1.0	10.0	5.1	0.9	10.2	5.1	0.8
	LM-At	9.4	4.8	1.0	10.4	5.2	1.0	9.8	5.0	1.0
	Skew	9.6	4.9	1.1	10.2	5.2	1.0	10.2	5.1	1.0
	KT-t	9.8	5.0	1.0	10.4	5.0	0.9	10.1	5.2	0.8
	KT-At	9.6	4.8	1.0	10.4	5.1	1.0	10.4	4.8	1.1
Parametric	LM-t	9.5	4.6	1.0	10.0	5.0	0.9	10.3	5.0	0.8
	LM-At	9.6	4.8	1.0	10.4	5.2	1.0	10.0	5.1	1.0
	Skew	9.8	4.9	1.2	10.3	5.2	1.0	10.1	5.3	1.0
	KT-t	9.9	4.8	1.0	10.1	5.1	0.8	10.2	5.1	0.9
	KT-At	9.5	4.8	1.0	10.1	5.0	0.9	10.4	4.8	1.1
Emp. CDF	LM-t	9.4	4.6	0.9	9.9	5.1	0.9	9.8	4.9	0.9
	LM-At	10.0	5.2	1.0	10.5	5.2	1.0	10.0	4.8	0.9
	Skew	10.3	5.1	1.1	10.2	5.2	0.9	9.8	5.1	1.0
	KT-t	9.8	4.6	0.9	10.1	4.9	0.9	9.9	5.0	1.0
	KT-At	10.0	5.2	1.0	10.2	5.2	1.0	10.2	4.9	1.1
	$S^{(C)}$	10.4	5.5	1.2	10.3	4.9	1.0			
	$S^{(B)}$	10.3	5.5	1.1	10.1	4.9	1.0			
	Q	10.5	5.3	0.9	10.9	5.3	1.1			
	KS	10.1	5.2	1.1	10.4	5.0	1.0			
	CvM	10.3	5.0	1.0	9.6	4.4	0.7			
Panel B: $K=10$ and $\rho_{kj}=0.75$										
Known	LM-t	9.5	4.9	0.9	10.1	5.0	1.1	10.1	5.3	0.9
	LM-At	10.3	5.4	1.1	10.3	4.9	1.0	9.8	4.9	1.1
	Skew	10.3	5.6	1.2	9.9	5.0	0.9	10.3	5.0	0.9
	KT-t	9.8	4.9	1.0	9.6	5.0	1.1	9.7	5.0	1.2
	KT-At	10.4	5.3	1.1	10.2	4.9	1.0	9.7	4.9	1.0
Emp. CDF	LM-t	9.6	4.5	0.8	9.1	4.7	0.8	9.9	4.8	1.0
	LM-At	9.8	5.0	1.0	9.6	4.7	1.1	10.1	4.8	1.0
	Skew	10.1	5.0	1.1	9.6	5.0	1.3	10.2	5.0	1.0
	KT-t	9.6	4.5	0.8	9.2	4.7	0.9	9.8	4.8	1.1
	KT-At	9.7	5.0	1.0	9.7	4.8	1.1	10.3	4.9	1.0

Notes: Critical values are computed using parametric bootstrap. LM–t and LM–At are the Lagrange multiplier tests based on the score of the symmetric and asymmetric Student t copula, respectively; while KT–t and KT–At are the corresponding Kuhn-Tucker versions (see Section 3 for details). Skew corresponds to the Lagrange multiplier test based on the moment conditions $m_{b_k}(\rho)$ of Proposition 3. $S^{(C)}$ and $S^{(B)}$ refer to procedures based on Rosenblatt's transform proposed by Genest et al. (2009), Q denotes the test statistic of Panchenko (2005), while KS and CvM denote the Kolmogorov–Smirnov and the Cramér–von Mises tests for copula models (see Rémillard (2017) for details). Parametric correspond to a DGP with exponential marginals whose parameters are estimated by maximum likelihood.

Table F2: Monte Carlo rejection rates at 1%, 5%, and 10% significance levels under the Student t alternative

Margins		$N=200$			$N=800$			$N=3,200$		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A: $K=2$ and $\rho=0.75$										
	LM-t	24.3	18.0	8.2	51.3	41.6	24.2	93.3	89.5	77.9
	LM-At	23.4	16.3	7.0	44.4	34.4	19.1	87.2	81.5	66.7
Known	Skew	15.8	9.2	3.0	17.2	10.4	3.2	16.4	9.8	3.0
	KT-t	31.9	21.7	8.2	61.2	48.9	25.8	96.4	92.9	80.9
	KT-At	25.3	17.0	7.1	47.4	36.9	19.5	89.5	83.7	68.5
	LM-t	24.5	18.1	8.1	51.0	41.4	24.4	93.3	89.4	78.2
	LM-At	23.6	16.3	6.9	44.2	34.5	19.0	87.3	81.6	66.7
Parametric	Skew	16.0	9.3	3.1	17.1	10.4	3.2	16.5	9.8	2.9
	KT-t	31.7	21.6	8.2	60.9	48.8	26.1	96.3	92.9	80.8
	KT-At	25.4	17.1	7.0	47.4	36.7	19.3	89.5	83.6	68.7
	LM-t	26.7	18.5	7.2	53.1	42.7	23.3	93.4	89.8	77.6
	LM-At	24.1	15.9	6.3	45.2	35.0	18.4	87.5	81.7	66.9
	Skew	15.9	9.0	2.9	16.8	10.2	3.1	16.2	9.8	2.9
	KT-t	29.3	19.1	7.2	58.2	45.5	23.7	95.5	91.6	78.9
	KT-At	24.9	16.1	6.3	47.2	36.0	18.6	89.0	83.2	68.2
Emp. CDF										
	$S^{(C)}$	11.8	6.1	1.3	16.3	8.9	2.3			
	$S^{(B)}$	11.9	6.1	1.2	16.1	9.1	2.3			
	Q	9.7	5.0	1.1	10.3	4.9	0.9			
	KS	10.3	5.3	1.1	11.5	5.8	1.3			
	CvM	10.2	5.0	1.0	11.0	5.3	1.0			
Panel B: $K=10$ and $\rho_{kj}=0.75$										
	LM-t	27.4	19.3	8.3	61.1	50.2	29.4	98.1	96.4	89.0
	LM-At	23.7	15.4	5.4	41.1	29.7	14.2	86.4	78.5	60.3
Known	Skew	17.4	9.8	2.7	17.9	10.7	2.9	18.5	10.3	2.7
	KT-t	38.5	26.0	9.7	73.2	60.5	34.7	99.3	98.2	93.1
	KT-At	24.4	15.8	5.5	42.8	31.4	14.4	87.5	80.5	61.8
	LM-t	30.9	19.9	6.5	65.7	53.0	28.5	98.3	96.6	89.2
	LM-At	24.7	15.7	4.8	45.5	34.0	16.2	87.9	80.6	62.2
Emp. CDF	Skew	16.4	9.3	2.3	18.0	10.1	3.1	17.9	10.6	3.1
	KT-t	31.1	20.0	6.5	67.3	53.9	28.6	98.7	97.1	90.3
	KT-At	24.7	15.7	4.8	46.2	34.1	16.5	88.5	81.5	62.8

Notes: DGP: Student t copula with 20 (100) degrees of freedom in Panel A (B). Critical values are computed using parametric bootstrap. LM–t and LM–At are the Lagrange multiplier tests based on the score of the symmetric and asymmetric Student t copula, respectively; while KT–t and KT–At are the corresponding Kuhn-Tucker versions (see Section 3 for details). Skew corresponds to the Lagrange multiplier test based on the moment conditions $m_{b_k}(\rho)$ of Proposition 3. $S^{(C)}$ and $S^{(B)}$ refer to procedures based on Rosenblatt's transform proposed by Genest et al. (2009), Q denotes the test statistic of Panchenko (2005), while KS and CvM denote the Kolmogorov–Smirnov and the Cramér–von Mises tests for copula models (see Rémillard (2017) for details). Parametric correspond to a DGP with exponential marginals whose parameters are estimated by maximum likelihood.

Table F3: Rejection rates at 1%, 5%, and 10% significance levels under the Asymmetric t alternative

		$N=200$			$N=800$			$N=3,200$		
Margins		10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A: $K=2$ and $\rho=0.75$										
	LM-t	27.2	20.4	9.7	56.9	47.4	29.5	96.1	93.5	84.5
	LM-At	37.4	27.3	12.6	80.6	72.3	52.5	100.0	99.9	99.7
Known	Skew	34.0	23.9	10.0	74.9	65.2	44.3	99.8	99.6	98.4
	KT-t	35.1	24.1	9.9	66.5	54.6	31.4	98.0	95.8	86.9
	KT-At	39.5	28.3	12.8	82.7	74.2	53.0	100.0	100.0	99.7
	LM-t	27.1	20.1	9.7	56.6	47.5	29.7	96.3	93.6	84.7
	LM-At	37.0	26.6	12.3	80.4	71.9	51.9	100.0	99.9	99.7
Parametric	Skew	33.4	23.4	9.8	74.7	64.9	43.6	99.8	99.6	98.4
	KT-t	34.7	23.7	9.8	66.5	54.2	31.3	98.0	96.0	87.1
	KT-At	39.0	27.6	12.4	82.7	74.1	52.5	100.0	100.0	99.7
	LM-t	28.9	20.1	8.1	58.6	48.6	27.9	96.2	93.7	83.9
	LM-At	34.0	24.2	9.8	75.9	66.9	45.0	99.9	99.8	99.1
	Skew	29.3	19.6	8.4	67.8	57.3	36.4	99.5	98.9	96.3
	KT-t	31.7	20.7	8.1	63.7	51.4	28.4	97.4	95.0	85.1
	KT-At	34.4	24.4	9.8	77.4	67.8	45.2	99.9	99.9	99.1
Emp. CDF										
	$S^{(C)}$	15.7	8.7	2.3	38.5	25.7	9.3			
	$S^{(B)}$	16.7	9.7	2.8	39.6	27.3	9.8			
	Q	9.6	4.6	0.8	13.9	7.2	1.7			
	KS	9.8	4.8	0.9	19.8	11.3	3.0			
	CvM	10.6	5.5	1.3	20.3	12.0	3.1			
Panel B: $K=10$ and $\rho_{kj}=0.75$										
	LM-t	27.05	19.12	8.54	61.12	49.6	29.66	97.97	96.4	88.03
	LM-At	24.77	16.53	5.99	46.33	34.32	16.78	93.08	87.81	74.04
Known	Skew	18.58	11.22	3.43	24.5	15.46	5.02	47.77	34.44	14.64
	KT-t	37.9	25.66	9.79	72.4	60.44	35.01	99.05	97.98	92.76
	KT-At	25.63	16.93	6.15	48.31	35.95	17.07	93.83	89.49	75.55
	LM-t	30.82	20.05	7.1	65.64	52.91	28.59	98.32	96.44	88.56
	LM-At	25.38	16.78	5.57	49.51	37.29	18.32	93.09	88.07	73.37
Emp. CDF	Skew	17.29	10.34	3.1	22.83	13.84	4.47	42.41	30.09	12.56
	KT-t	30.98	20.12	7.1	66.99	53.75	28.68	98.77	96.88	89.46
	KT-At	25.43	16.79	5.57	50.22	37.48	18.55	93.47	88.47	73.69

Notes: DGP: Asymmetric Student t copula with 20 (100) degrees of freedom and skewness vector $\mathbf{b}=-.75\ell$ ($\mathbf{b}=-.15\ell$) in Panel A (B). Critical values are computed using parametric bootstrap. LM–t and LM–At are the Lagrange multiplier tests based on the score of the symmetric and asymmetric Student t copula, respectively; while KT–t and KT–At are the corresponding Kuhn-Tucker versions (see Section 3 for details). Skew corresponds to the Lagrange multiplier test based on the moment conditions $m_{b_k}(\rho)$ of Proposition 3. $S^{(C)}$ and $S^{(B)}$ refer to procedures based on Rosenblatt's transform proposed by Genest et al. (2009), Q denotes the test statistic of Panchenko (2005), while KS and CvM denote the Kolmogorov–Smirnov and the Cramér–von Mises tests for copula models (see Rémillard (2017) for details). Parametric correspond to a DGP with exponential marginals whose parameters are estimated by maximum likelihood.

Table F4: Rejection rates at 1%, 5%, and 10% significance levels under the Skew t alternative

Margins		$N=200$			$N=800$			$N=3,200$		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A: $K=2$ and $\rho=0.75$										
	LM-t	24.0	17.6	7.8	51.0	41.5	23.8	93.7	89.9	77.7
	LM-At	23.7	16.4	6.7	45.4	35.4	19.5	90.3	84.8	70.4
Known	Skew	16.5	10.0	2.7	19.6	12.4	4.1	29.5	20.0	7.8
	KT-t	32.1	21.3	7.9	61.6	48.7	25.3	96.6	93.3	81.2
	KT-At	25.6	17.2	6.7	49.0	37.4	19.9	92.0	87.0	72.1
	LM-t	24.3	17.7	7.8	50.9	41.5	24.0	93.7	89.8	77.7
	LM-At	23.7	16.3	6.6	45.1	35.3	19.4	90.4	84.8	70.2
Parametric	Skew	16.2	10.2	2.8	19.4	12.2	4.1	29.2	20.1	7.6
	KT-t	31.9	21.4	7.9	61.4	48.5	25.7	96.5	93.2	81.1
	KT-At	25.5	17.3	6.6	48.8	37.7	19.7	92.1	86.8	72.1
	LM-t	26.9	18.7	6.9	53.7	43.1	22.4	93.9	90.3	77.4
	LM-At	25.0	16.8	6.4	46.5	35.7	18.8	89.7	83.9	69.3
	Skew	16.3	9.7	3.1	18.3	11.3	3.6	24.9	15.9	5.6
	KT-t	29.7	19.4	6.9	59.0	45.8	22.9	95.9	92.2	79.0
	KT-At	25.6	17.0	6.4	48.5	36.7	19.0	91.0	85.3	70.5
Emp. CDF										
	$S^{(C)}$	12.8	6.9	1.5	18.9	10.9	3.0			
	$S^{(B)}$	12.8	7.1	1.6	19.0	11.4	3.1			
	Q	9.3	4.6	0.9	9.8	4.6	0.8			
	KS	9.6	5.1	0.9	10.7	5.6	1.1			
	CvM	10.8	5.2	0.9	11.0	5.6	1.0			
Panel B: $K=10$ and $\rho_{kj}=0.75$										
	LM-t	27.87	19.28	8.64	61.11	50.1	28.75	98.47	97.07	90.22
	LM-At	24.99	16.21	5.55	41.15	30.4	15.58	89.37	82.5	66.58
Known	Skew	18.59	11.11	3.15	18.92	11.15	3.37	25.83	15.7	5.18
	KT-t	38.82	26.87	10.01	73.64	60.9	34.44	99.38	98.44	93.55
	KT-At	25.88	16.54	5.83	43.61	31.76	15.82	90.53	84.17	67.88
	LM-t	31.47	20.52	7.45	64.69	52.45	27.51	98.54	96.9	89.37
	LM-At	25.74	16.33	5.54	44.32	32.97	14.93	88.32	81.38	63.95
Emp. CDF	Skew	17.38	10.05	2.86	16.85	9.9	2.53	18.05	10.93	3.22
	KT-t	31.65	20.56	7.45	66.05	53.04	27.54	98.98	97.36	90.29
	KT-At	25.92	16.36	5.54	44.81	33.15	14.93	89.03	82.08	64.68

Notes: DGP: Skew t copula with 20 (100) degrees of freedom and skew parameter $\alpha=-.25$ ($\alpha=-.05$) in Panel A (B) (see Azzalini and Capitanio (2003) for details). Critical values are computed using parametric bootstrap. LM–t and LM–At are the Lagrange multiplier tests based on the score of the symmetric and asymmetric Student t copula, respectively; while KT–t and KT–At are the corresponding Kuhn-Tucker versions (see Section 3 for details). Skew corresponds to the Lagrange multiplier test based on the moment conditions $m_{b_k}(\rho)$ of Proposition 4. $S^{(C)}$ and $S^{(B)}$ refer to procedures based on Rosenblatt's transform proposed by Genest et al. (2009), Q denotes the test statistic of Panchenko (2005), while KS and CvM denote the Kolmogorov–Smirnov and the Cramér–von Mises tests for copula models (see Rémillard (2017) for details). Parametric correspond to a DGP with exponential marginals whose parameters are estimated by maximum likelihood.

Table G1: Correlation parameter estimates

Sector	N.Obs	Short-term reversal strategies				Momentum strategies			
			Correlation parameter				Correlation parameter		
		Beta OLS	Pearson	Spearman	Copula	Beta OLS	Pearson	Spearman	Copula
Non Durables	33,215	.013	.013	-.028	-.030	.000	.001	.029	.033
		(.010)	(.010)	(.004)	(.006)	(.002)	(.008)	(.004)	(.006)
Durables	6,816	.014	.015	-.038	-.044	.004	.022	.050	.048
		(.019)	(.020)	(.009)	(.011)	(.003)	(.015)	(.009)	(.011)
Manufacturing	62,609	.014	.015	-.040	-.038	-.001	-.003	.023	.024
		(.007)	(.007)	(.003)	(.004)	(.001)	(.006)	(.003)	(.005)
Energy	23,868	.005	.006	-.023	-.027	.000	-.001	.026	.025
		(.009)	(.009)	(.005)	(.007)	(.000)	(.002)	(.005)	(.007)
Chemicals	11,030	-.013	-.014	-.051	-.046	-.001	-.003	.004	.008
		(.016)	(.017)	(.007)	(.010)	(.004)	(.016)	(.007)	(.010)
Business	64,962	.010	.010	-.029	-.025	-.002	-.011	.017	.012
		(.008)	(.008)	(.003)	(.004)	(.001)	(.007)	(.003)	(.004)
Telecom	12,357	.041	.044	.018	.024	.003	.015	.062	.059
		(.017)	(.018)	(.007)	(.009)	(.002)	(.011)	(.007)	(.009)
Utilities	21,168	.019	.019	-.036	-.037	-.001	-.003	.004	.012
		(.014)	(.014)	(.005)	(.007)	(.003)	(.013)	(.005)	(.007)
Shops	65,864	.019	.019	-.025	-.025	-.001	-.004	.034	.033
		(.011)	(.011)	(.003)	(.004)	(.001)	(.004)	(.003)	(.004)
Healthcare	25,691	.018	.019	-.027	-.026	-.002	-.008	.032	.028
		(.012)	(.013)	(.005)	(.007)	(.002)	(.010)	(.005)	(.007)
Financials	123,850	-.008	-.009	-.062	-.055	.001	.016	.031	.034
		(.006)	(.006)	(.002)	(.003)	(.002)	(.018)	(.002)	(.003)
Others	155,624	.004	.004	-.024	-.024	.000	.002	.040	.037
		(.004)	(.004)	(.002)	(.003)	(.001)	(.004)	(.002)	(.003)
All									
	607,054	.008	.009	-.025	-.022	.000	.002	.037	.035
		(.003)	(.003)	(.001)	(.002)	(.000)	(.003)	(.001)	(.002)

Notes: The data is collected from CRSP and contains monthly series from January 1997 to December 2012. Industry definitions from Ken French's website. Beta OLS denotes the slope coefficient in a simple linear regression. Pearson and Spearman denote the Pearson linear correlation coefficient and the Spearman's rank correlation, respectively; while Copula denotes the Gaussian rank correlation (linear correlation coefficient of the Gaussian ranks). Numbers in parenthesis correspond to Newey and West (1987) standard errors; variances of ρ are corrected for heteroskedasticity and autocorrelation using 5 lags.

Table G2: Test statistics and p-values

		Short-term reversal strategies			Momentum strategies		
Sector	N.Obs	KT-t	Skew	KT-At	KT-t	Skew	KT-At
Non Durables	33,215	885.7	40.3	926.0	1,238.0	184.8	1,422.8
		(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Durables	6,816	61.5	11.4	72.9	82.7	22.0	104.7
		(.000)	(.003)	(.000)	(.000)	(.000)	(.000)
Manufacturing	62,609	1,437.5	23.4	1,461.0	2,245.6	133.3	2,378.9
		(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Energy	23,868	510.2	21.1	531.3	563.1	61.1	624.2
		(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Chemicals	11,030	168.4	2.8	171.1	209.7	15.5	225.2
		(.000)	(.247)	(.000)	(.000)	(.000)	(.000)
Business	64,962	810.0	34.3	844.3	1,443.7	120.6	1,564.2
		(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Telecom	12,357	236.0	25.9	261.9	211.9	98.6	310.6
		(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Utilities	21,168	500.6	34.6	535.3	559.5	34.1	593.7
		(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Shops	65,864	1,562.2	54.8	1,617.0	2,260.6	356.0	2,616.6
		(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Healthcare	25,691	429.8	19.7	449.5	564.5	120.0	684.5
		(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Financials	123,850	6,152.6	369.4	6,522.1	6,476.7	1,238.4	7,715.0
		(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Others	155,624	3,053.2	209.5	3,262.7	4,339.6	935.8	5,275.5
		(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
All							
	607,054	24,333.7	1,086.0	25,419.7	32,408.0	4,258.7	36,666.7
		(.000)	(.000)	(.000)	(.000)	(.000)	(.000)

Notes: The data is collected from CRSP and contains monthly series from January 1997 to December 2012. Industry definitions from Ken French's website. Numbers in parenthesis correspond to asymptotic p-values. Both, variances of the test moment functions are corrected for heteroskedasticity and autocorrelation using 5 lags. KT–t and KT–At are the Kuhn-Tucker tests based on the score of the symmetric and asymmetric Student t copula, respectively (see Section 3 for details). Skew corresponds to the Lagrange multiplier test based on the 2 moment conditions $m_{b_k}(u_1, u_2; \rho, 0)$ for $k = 1, 2$ of Proposition 4.

Table G3: Constrained indirect estimates of the shape parameters

Panel A: Short term reversals strategies

	Student t		Asymmetric Student t			
	$\hat{\rho}$	$\hat{\eta}$	$\hat{\rho}$	$\hat{\eta}$	$\hat{b}_1$	$\hat{b}_2$
Sector						
Non Durables	-.032	.154	-.029	.155	-.135	-.049
Durables	-.047	.093	-.045	.093	-.156	-.219
Manufacturing	-.040	.144	-.039	.144	-.065	-.045
Energy	-.029	.139	-.027	.139	-.076	-.091
Chemicals	-.049	.119	-.048	.120	-.103	-.014
Business	-.028	.108	-.027	.109	-.125	-.064
Telecom	.022	.131	.027	.130	-.154	-.142
Utilities	-.040	.146	-.036	.146	-.124	-.108
Shops	-.028	.146	-.026	.146	-.109	-.057
Healthcare	-.028	.124	-.026	.124	-.117	-.075
Financials	-.058	.209	-.042	.207	-.117	-.091
Others	-.027	.133	-.024	.134	-.163	-.069
All	-.025	.187	-.018	.187	-.112	-.069

Panel B: Momentum strategies

	Student t		Asymmetric Student t			
	$\hat{\rho}$	$\hat{\eta}$	$\hat{\rho}$	$\hat{\eta}$	$\hat{b}_1$	$\hat{b}_2$
Sector						
Non Durables	.032	.179	.051	.178	-.170	-.176
Durables	.046	.105	.051	.104	-.262	-.225
Manufacturing	.022	.176	.028	.176	-.087	-.114
Energy	.023	.144	.028	.143	-.131	-.159
Chemicals	.006	.131	.007	.130	-.054	-.140
Business	.010	.141	.014	.140	-.146	-.113
Telecom	.057	.123	.062	.118	-.113	-.467
Utilities	.010	.152	.014	.152	-.117	-.102
Shops	.032	.172	.046	.170	-.136	-.208
Healthcare	.026	.139	.034	.137	-.162	-.247
Financials	.033	.211	.080	.209	-.139	-.252
Others	.035	.155	.047	.153	-.150	-.259
All	.034	.213	.074	.212	-.124	-.190

Notes: The data is collected from CRSP and contains monthly series from January 1997 to December 2012. Estimates are obtained by generating sample paths of size 100,000 from this copula and matching in the simulated data the values in the original data of both the Gaussian rank correlation coefficients and the corresponding test statistics.

Figure D1: Power of Student t-based tests under asymmetric Student t local alternatives

Figure D1a: Non-centrality parameter for different kurtosis parameter values

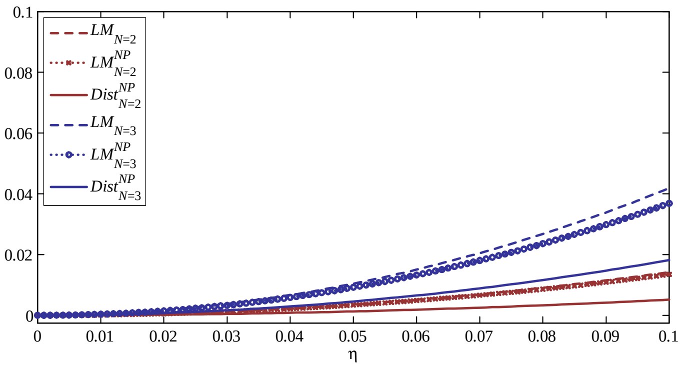


Figure D1b: Non-centrality parameter for different correlation parameter values

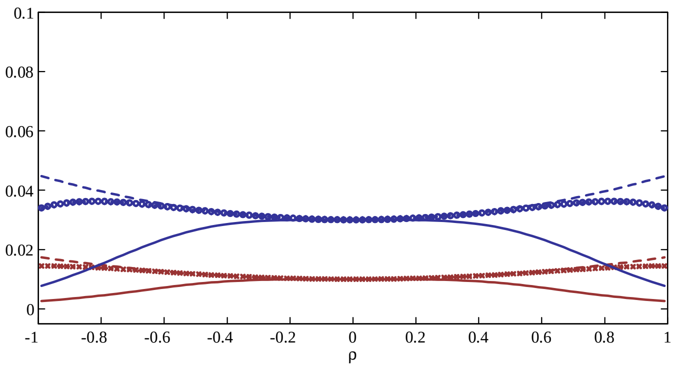
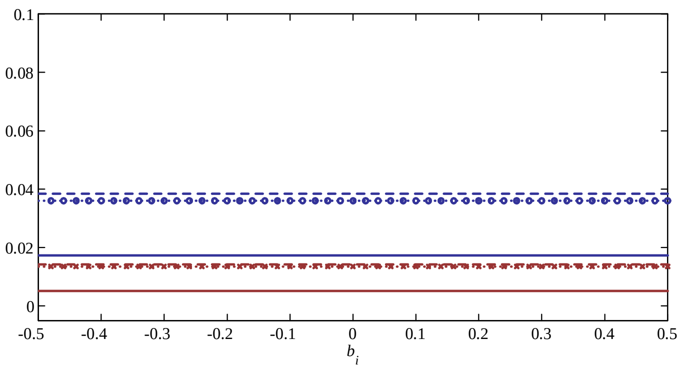


Figure D1c: Non-centrality parameter for different skewness parameter values



Notes: Non-centrality parameters of the Student t LM-copula and LM-distributional tests under asymmetric Student t alternatives. LM and LM^{NP} denote the LM-copula tests when marginals are known and when they are estimated nonparametrically, respectively; while $Dist^{NP}$ denotes the LM-distributional test when marginals are estimated nonparametrically. Figure D1a, power against asymmetric Student t alternatives with $\rho = .75$ and b_k =-.5 for $k = 1, 2$. Figure D1b, power against asymmetric Student t alternatives with different correlation parameter and $\eta = .1$, b_k =-.5 for $k = 1, 2$. Figure D1c, power against asymmetric Student t alternatives with increasing skewness and $\eta = .1$, $\rho = .75$. Figures D1b-c share the legend of Figure D1a.

Figure D2: Power of asymmetric Student t-based tests under asymmetric Student t local alternatives

Figure D2a: Non-centrality parameter for different kurtosis parameter values

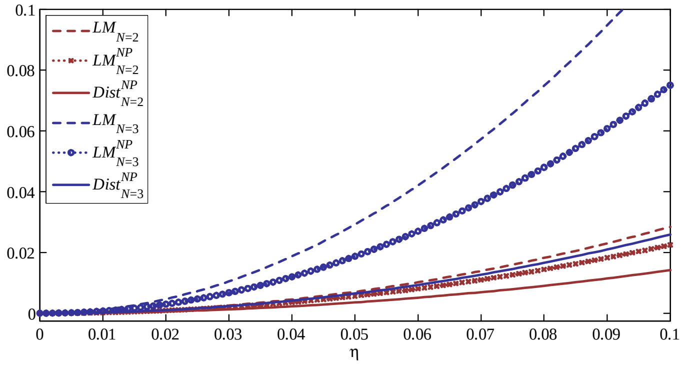


Figure D2b: Non-centrality parameter for different correlation parameter values

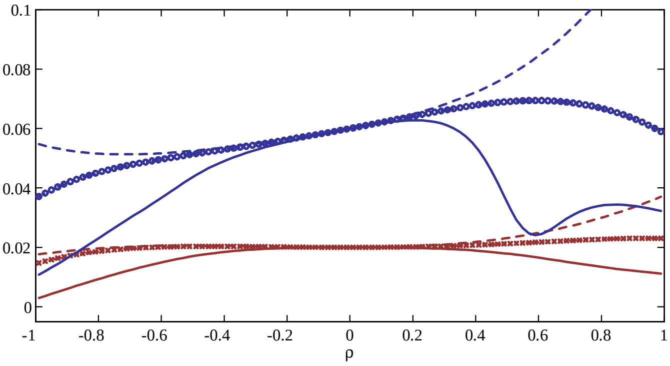
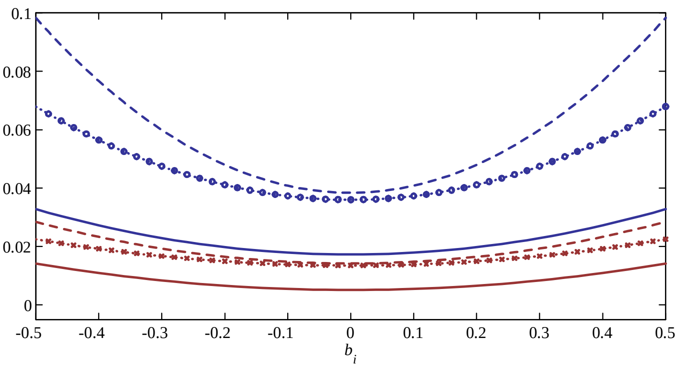


Figure D2c: Non-centrality parameter for different skewness parameter values



Notes: Non-centrality parameters of the Student t LM-copula and LM-distributional tests under asymmetric Student t alternatives. LM and LM^{NP} denote the LM-copula tests when marginals are known and when they are estimated nonparametrically, respectively; while $Dist^{NP}$ denotes the LM-distributional test when marginals are estimated nonparametrically. Figure D2a, power against asymmetric Student t alternatives with $\rho = .75$ and b_k =-.5 for $k = 1, 2$. Figure D2b, power against asymmetric Student t alternatives with different correlation parameter and $\eta = .1$, b_k =-.5 for $k = 1, 2$. Figure D2c, power against asymmetric Student t alternatives with increasing skewness and $\eta = .1$, $\rho = .75$. Figures D2b-c share the legend of Figure D2a.

Figure G1: Transition probabilities for short term reversals strategies

Figure G1a: Bottom 5%

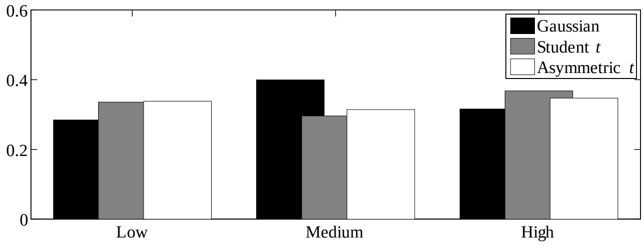


Figure G1b: Bottom-Middle 25%

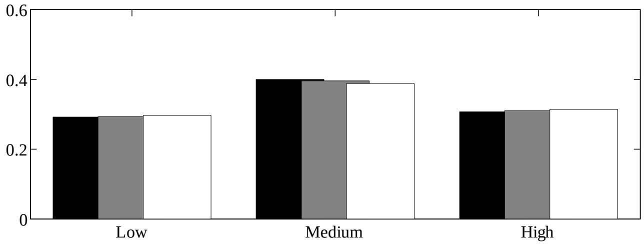


Figure G1c: Middle 40%

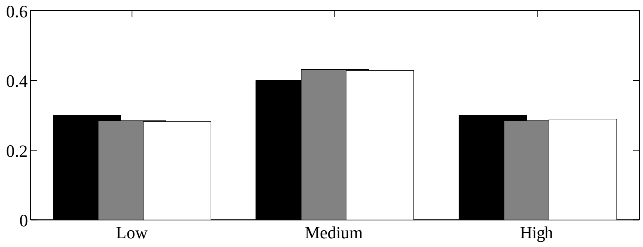


Figure G1d: Middle-Top 25%

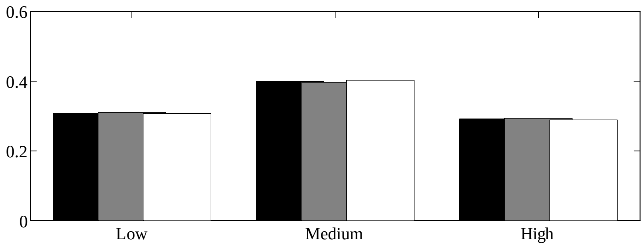
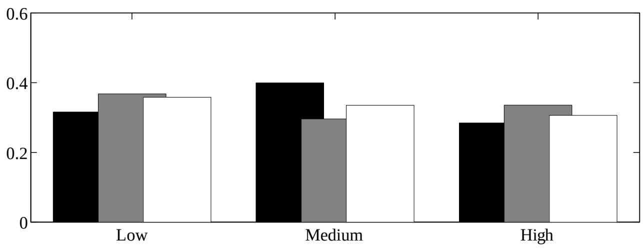


Figure G1e: Top 5%



Notes: The data is collected from CRSP and contains monthly series from January 1997 to December 2012. Gaussian copula with correlation coefficient ρ =-.022. For the Student t copula, ρ =-.025 and η = .187; while for the asymmetric Student t copula, ρ =-.018, η = .187, b_1 =-.112 and b_2 =-.069 (obtained by constrained indirect estimation).

Figure G2: Transition probabilities for momentum strategies

Figure G2a: Bottom 5%

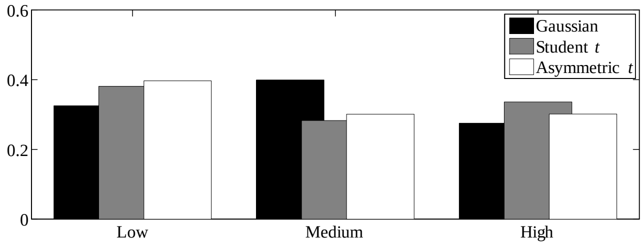


Figure G2b: Bottom-Middle 25%

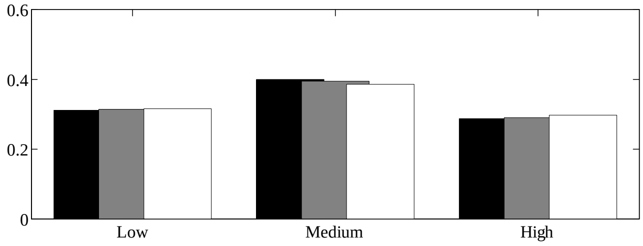


Figure G2c: Middle 40%

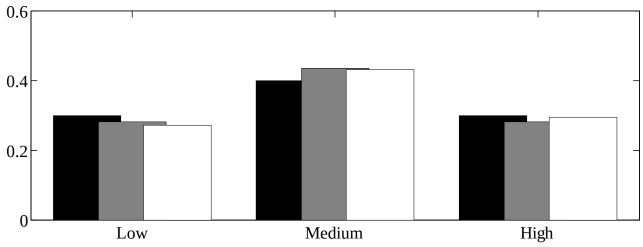


Figure G2d: Middle-Top 25%

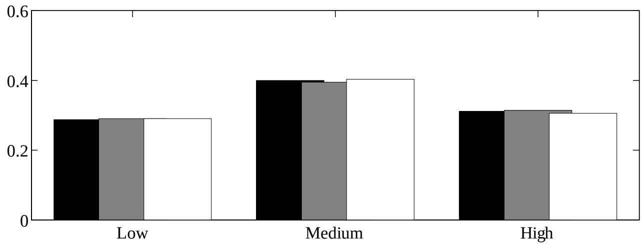
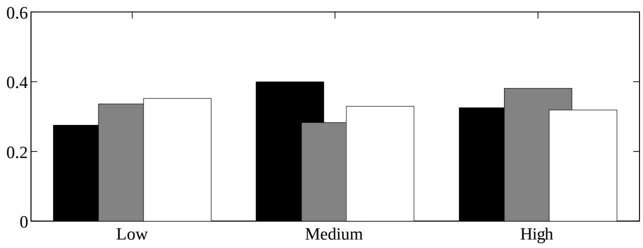


Figure G2e: Top 5%



Notes: The data is collected from CRSP and contains monthly series from January 1997 to December 2012. Gaussian copula with correlation coefficient $\rho = .035$. For the Student t copula, $\rho = .034$ and $\eta = .213$; while for the asymmetric Student t copula, $\rho = .074$, $\eta = .212$, b_1 =-.124 and b_2 =-.190 (obtained by constrained indirect estimation).

References

Amengual, D. and E. Sentana (2015): "Is a normal copula the right copula?" CEMFI Working Paper No. 1504.

Azzalini, A. and Capitanio, A. (2014). *The skew-normal and related families.* Cambridge University Press, IMS Monograph series.

Gagliardini, P., C. Gouriéroux and M. Rubin (2014): "Positional portfolio management", mimeo, Università della Svizzera Italiana.

Genest, C., B. Rémillard and D. Beaudoin (2009): "Goodness-of-fit tests for copulas: a review and a power study", *Insurance Mathematics and Economics*, 44, 199–213.

Johnson, N., S. Kotz and N. Balakrishnan (1994): *Continuous univariate distributions*, Wiley, New York.

Jørgensen, B. (1982): Statistical properties of the Generalized Inverse Gaussian distribution, Springer-Verlag, New York.

Mencía, J. and E. Sentana (2012): "Distributional tests in multivariate dynamic models with Normal and Student t innovations", *Review of Economics and Statistics*, 94, 133–152.

Newey, W.K and K.D. West (1987): "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix", *Econometrica* 55, 703–708.

Panchenko, V. (2005): "Goodness-of-fit test for copulas", *Physica A*, 355, 176–182.

Rémillard, B. (2017): "Goodness-of-fit tests for copulas of multivariate time series", *Econometrics*, 5, 13.