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>

May 2013
Revised: July 2018

Abstract

We derive computationally simple and intuitive expressions for score tests of Gaussian copulas against Generalized Hyperbolic alternatives, including symmetric and asymmetric Student t, and many other examples. We decompose our tests into third and fourth moment components, and obtain one-sided Likelihood Ratio analogues, whose standard asymptotic distribution we provide. Our Monte Carlo exercises confirm the reliable size of parametric bootstrap versions of our tests, and their substantial power gains over alternative procedures. In an empirical application to CRSP stocks, we find that short-term reversals and momentum effects are better captured by non-Gaussian copulas, whose parameters we estimate by indirect inference.

Keywords: Cokurtosis, Coskewness, Indirect inference, Kuhn-Tucker test, Momentum strategies, Non-linear dependence, Short-term reversals, Supremum test.

JEL: C46, C52, G11, G14

1 Introduction

Nowadays copulas are extensively used in many economic and finance applications, with the Gaussian copula being very popular despite ruling out non-linear dependence, particularly in the lower tail. Nevertheless, the validity of this copula in finance has been the subject of considerable public debate, to the extent that the media declared it "the formula that felled Wall Street" (see the provocative article by Salmon (2009), the more nuanced analysis by MacKenzie and Spears (2012), and the academic response by Donnelly and Embrechts (2010)). To be fair, the statistics and econometric literatures were well aware of the possibility of misspecification of the assumed copula, and several more or less formal diagnostics had already been proposed (see e.g. Malevergne and Sornette (2003), Panchenko (2005), Berg and Quessy (2009) and Genest, Rémillard and Beaudoin (2009)).

The first objective of our paper is precisely to derive simple to implement and interpret score-based specification tests that can detect the non-normality of a copula. As a flexible nesting alternative, we consider the family of copulas associated to the Generalized Hyperbolic distribution, which includes the symmetric and asymmetric Student t, normal-gamma mixtures, hyperbolic, normal inverse Gaussian and symmetric and asymmetric Laplace distributions. Aside from computational considerations, the advantage of score tests is that rejections provide a clear indication of the specific directions along which modelling efforts should focus. In addition, they often coincide with tests of easy to understand moment conditions (see Newey (1985) and Tauchen (1985)). In our case, in particular, we decompose our tests into third and fourth moment analogues for the Gaussian ranks, which will continue to have non-trivial power even in situations for which they are not optimal. Further, we obtain more powerful one-sided Kuhn-Tucker versions that are asymptotically equivalent to the Likelihood Ratio test under the null and sequences of local alternatives, and therefore share its optimality properties. In all cases, we derive closed-form expressions for the asymptotic covariance matrices of the influence functions we use for testing, which should improve the finite sample reliability of our tests.

For pedagogical reasons, we initially assume known margins, but since this rarely happens in practice, we also consider popular two-step estimation procedures in which the marginal distributions are either replaced by their (re-scaled) empirical cumulative distribution function (cdf) counterparts or estimated by maximum likelihood. In the first case, we show that it is possible to capture the variance modification in the scores of the shape parameters of the distributions that we consider as alternatives by adding linear combinations of third and fourth Hermite polynomials in the Gaussian ranks. In the second case, on the other hand, we exploit Joe (2005)'s results to obtain the asymptotic variance of the influence functions that account for parameter uncertainty in the margins.

We also study the finite sample properties of parametric bootstrap versions of our proposed tests with an extensive Monte Carlo exercise. The rejection rates of our proposals are by and large close to being perfect for all samples sizes. In addition, their finite sample power agrees with what an asymptotic local power analysis suggests, showing substantial gains over alternative non-parametric procedures.

Finally, we employ our proposed tests to assess the suitability of the Gaussian copula for capturing the short-term reversals and momentum effects observed in the cross-section of individual stock returns in the CRSP data base. In both cases, we reject the null hypothesis of a Gaussian copula by a long margin, the source of the rejection being not only the "cokurtosis" between the Gaussian ranks, but also their "coskewness", especially for momentum strategies. For that reason, we estimate the parameters of some non-Gaussian copulas by means of a constrained indirect inference approach that uses the Gaussian rank correlation coefficients and our score tests as sample statistics to match, as suggested by Calzolari, Fiorentini and Sentana (2004).

The rest of the paper is divided as follows. In Section 2, we discuss the relevant theoretical background to the problem, and develop our proposed tests in Section 3. Next, we report the results from an extensive Monte Carlo exercise in Section 4. We then analyze the cross-sectional dependence between monthly returns on individual U.S. stocks in the CRSP database and some of their observable characteristics in Section 5, followed by our conclusions. Proofs and additional results are relegated to appendices.

2 Theoretical background

2.1 The model under the null

Let $\mathbf{x}$ denote a vector of K observed continuous random variables. The traditional way of modelling the dependence between the elements of $\mathbf{x}$ is through the joint distribution function $F_K(\mathbf{x})$ or the associated density function $f_K(\mathbf{x})$ when it is well defined. These functions are often recursively factorized for a predetermined ordering as the sequence of conditional distributions of x_k given $x_{k-1}, x_{k-2}, \dots x_1$ $(k = 2, \dots, K)$ times the marginal distribution of x_1. In contrast, the standard copula approach first instantaneously transforms each of the elements of $\mathbf{x}$ into a uniform random variable by means of the probability integral transform $u_k = G_{1k}(x_k)$, where $G_{1k}(.)$ is the marginal cumulative distribution function of x_k, which we assume known until Section 3.5.2, and then models the dependence of the random vector $\mathbf{u} = (u_1, \ldots, u_K)'$ through a joint distribution function $C(\mathbf{u})$ with uniform marginals defined over the unit hypercube in $\mathbb{R}^K$. This distribution function is known as the copula distribution function, and the associated density as the copula density function.

Although there are many well known examples of bivariate copulas, some of them are popular simply because they are mathematically convenient, as opposed to being motivated by empirical observations on real life phenomena. More importantly, they are difficult to generalize to multiple dimensions. On the other hand, the Gaussian copula is a popular choice both in bivariate and multivariate contexts since it is easily scalable. Moreover, as its name suggests, it is the copula function that corresponds to the multivariate Gaussian distribution, which remains dominant in multivariate statistical analysis.

More formally, define $\mathbf{y} = (y_1, \ldots, y_K)'$ as the vector of Gaussian ranks of the observed variables $\mathbf{x}$, so that $y_k = \Phi_1^{-1}(u_k)$, $\Phi_1(.)$ denotes the univariate standard normal cdf and $\Phi_1^{-1}(.)$ the corresponding quantile function. The Gaussian copula with correlation matrix $\mathbf{P}(\boldsymbol{\rho})$ is derived from the cumulative distribution function of a multivariate random vector $\mathbf{y} \sim N[\mathbf{0}, \mathbf{P}(\boldsymbol{\rho})]$. In what follows, we assume that:

Assumption 1 $\mathbf{P}(\boldsymbol{\rho})$ *is a positive definite matrix which contains $K(K-1)/2$ possibly distinct, twice continuously differentiable functions of the $p \times 1$ vector of correlation parameters $\boldsymbol{\rho}$, such that $\mathbf{P}(\mathbf{0}) = \mathbf{I}_K$.*

It should be noticed that, in the unrestricted case, $\mathbf{P}(\boldsymbol{\rho})$ is trivially twice differentiable and the same is true for many popular restricted parametrizations, such as an equicorrelated single factor structure. In turn, the requirement that $\boldsymbol{\rho} = \mathbf{0}$ yields the independent copula is just a convenient normalization.

Under this assumption, the Gaussian copula density function will be given by

$$c(\mathbf{u}; \boldsymbol{\rho}) = |\mathbf{P}(\boldsymbol{\rho})|^{-1/2} \exp\left\{-\frac{1}{2}\mathbf{y}'[\mathbf{P}^{-1}(\boldsymbol{\rho}) - \mathbf{I}_K]\mathbf{y}\right\}.$$

Figure 1a-b displays a bivariate Gaussian copula density with Spearman correlation of .115 ($\rho = .12$) with Gaussian margins.

In principle, we could consider more complex models by conditioning on past values of $\mathbf{x}$ or present and past values of some exogenous variables $\mathbf{z}$ (see e.g. Patton (2012) and Fan and Patton (2014) for detailed reviews), but for the sake of clarity we will only explicitly cover unconditional contexts without conditioning variables.

2.2 Some parametric models that nest the Gaussian copula

As we mentioned in the introduction, the Gaussian copula rules out any type of non-linear dependence between the elements of $\mathbf{y}$. For that reason, empirical researchers have considered more flexible copulas that nest the normal copula as a special case, which are often generated from a multivariate distribution that in turn nests the multivariate normal distribution. Some important examples are the symmetric and asymmetric versions of the Student t, and the more flexible Generalized Hyperbolic family.

2.2.1 Student t copulas

The Student t distribution generalizes the multivariate normal distribution through a single additional parameter, usually known as the degrees of freedom. To construct the Student t copula, we use the log density of a standardized multivariate Student t distribution with mean $\mathbf{0}$, correlation matrix $\mathbf{P}(\boldsymbol{\rho})$ and $\nu > 2$ degrees of freedom. This distribution is such that its marginal components are also univariate Student t's with mean 0, unit variance and ν degrees of freedom. Let $\eta = 1/\nu$ denote the reciprocal of the degrees of freedom parameter and $\boldsymbol{\varepsilon}(\eta) = [F_1^{-1}(u_1;\eta), ..., F_1^{-1}(u_K;\eta)]'$ the vector of Student t ranks of the observed variables $\mathbf{x}$, where $F_1^{-1}(u_k;\eta)$ denotes the common quantile function of a univariate standardized Student t. Importantly, the standardized version that we use differs from the textbook multivariate Student t distribution in that the kernel is $\ln[1+\eta(1-2\eta)^{-1}\boldsymbol{\varepsilon}'\mathbf{P}^{-1}\boldsymbol{\varepsilon}]$ instead of $\ln[1+\eta\boldsymbol{\varepsilon}'\mathbf{P}^{-1}\boldsymbol{\varepsilon}]$ to guarantee that $\mathbf{P}$ coincides with the correlation matrix (see Fiorentini, Sentana and Calzolari (2003)). Nevertheless, the difference is inconsequential in the neighborhood of the null hypothesis $H_0 : \eta = 0$.

Therefore, the Student t copula will be given by the expression

$$c(\mathbf{u};\boldsymbol{\rho},\eta) = |\mathbf{P}(\boldsymbol{\rho})|^{-1/2}\exp[h_K(\eta) - Kh_1(\eta)]\frac{\left[1+\eta\boldsymbol{\varepsilon}'(\eta)\mathbf{P}^{-1}(\boldsymbol{\rho})\boldsymbol{\varepsilon}(\eta)/(1-2\eta)\right]^{(K\eta+1)/(2\eta)}}{\left\{\Pi_{k=1}^{K}\left[1+\eta\varepsilon_k^2(\eta)/(1-2\eta)\right]\right\}^{(\eta+1)/(2\eta)}},$$

where

$$h_k(\eta) = \ln\left[\Gamma\left(\frac{k\eta+1}{2\eta}\right)\right] - \ln\left[\Gamma\left(\frac{1}{2\eta}\right)\right] - \frac{k}{2}\ln\left(\frac{1-2\eta}{\eta}\right) - \frac{k}{2}\ln\pi.$$

As expected, this copula converges to the Gaussian one as $\eta \to 0^+$ but otherwise it induces tail dependence even when $\boldsymbol{\rho} = \mathbf{0}$. Figures 1c-d display a bivariate Student t copula density with the same Spearman correlation as in Figure 1a-b ($\rho = .122$) and with Gaussian margins.

2.2.2 Generalized hyperbolic copulas

Barndorff-Nielsen (1977) introduced a rather flexible family of multivariate densities that he called Generalized Hyperbolic (GH), which nests not only the normal and Student t but also many other examples such as the asymmetric Student t, the hyperbolic and normal inverse Gaussian distributions, as well as symmetric and asymmetric versions of the normal-gamma mixture and Laplace (see also Blæsild (1981)).

Mencía and Sentana (2012) derive a standardized version of the GH distribution with zero mean and identity covariance matrix, which depends exclusively on three shape parameters that we can set to zero under normality: $\boldsymbol{\beta}$, which introduces asymmetries, and η and ψ, whose product $\tau = \eta\psi$ effectively controls excess kurtosis in the vicinity of the Gaussian null. Then, one can easily obtain a correlated GH by means of the linear transformation given by the matrix $\mathbf{P}^{\frac{1}{2}\prime}(\boldsymbol{\rho})$, where $\mathbf{P}^{1/2}(\boldsymbol{\rho})$ denotes some particular "square root" matrix such that $\mathbf{P}^{1/2}(\boldsymbol{\rho})\mathbf{P}^{1/2\prime}(\boldsymbol{\rho}) = \mathbf{P}(\boldsymbol{\rho})$. They also parametrize $\boldsymbol{\beta}$ as a function of a new vector of parameters $\mathbf{b}$ in the following way:

$$\boldsymbol{\beta}(\boldsymbol{\rho}, \mathbf{b}) = \mathbf{P}^{\frac{1}{2}\prime}(\boldsymbol{\rho})\mathbf{b}, \tag{1}$$

so that the resulting distribution does not depend on the choice of square root matrix.

A rather useful property of the GH distributions is that the marginal distributions of linear combinations (including the individual components) also follow univariate GH distributions (see Blæsild (1981)). As a result, we can once again construct the GH copula by combining the joint distribution and its marginals.

Figures 1e-f display a bivariate asymmetric Student t copula density with negative tail dependence but the same Spearman correlation as in Figure 1a-b ($\rho = .186$) and with Gaussian margins. The evidence in Mencía and Sentana (2009) shows that this member of the GH family is empirically very relevant.

3 Score tests

3.1 Test against Student t copulas

Under the null hypothesis, $\mathbf{y}'\mathbf{P}^{-1}(\boldsymbol{\rho}_0)\mathbf{y}$ will be distributed as a χ^2 random variable with K degrees of freedom (see Malevergne and Sornette (2003)). Let

$$L_2\left[\mathbf{y}'\mathbf{P}^{-1}(\boldsymbol{\rho})\mathbf{y}\right] = \frac{1}{4}[\mathbf{y}'\mathbf{P}^{-1}(\boldsymbol{\rho})\mathbf{y}]^2 - \frac{K+2}{2}\mathbf{y}'\mathbf{P}^{-1}(\boldsymbol{\rho})\mathbf{y} + \frac{K(K+2)}{4}$$

denote the second-order Laguerre polynomial associated to this particular example of a gamma random variable. Further, let $\mathbf{y}'_{(k)} = (y_1, \ldots, y_{k-1}, y_{k+1}, \ldots, y_K)$, $\mathbf{P}_{(kj)}(\boldsymbol{\rho})$ the $(K-1)\times(K-1)$ matrix obtained from $\mathbf{P}(\boldsymbol{\rho})$ after erasing row k and column j, and $\mathbf{p}_{(k)}(\boldsymbol{\rho})$ the coefficients in the theoretical least squares projection of y_k on to (the linear span of) $\mathbf{y}_{(k)}$.

Proposition 1 *Let $H_j(.)$ denote the (standardized) j^{th}-order Hermite polynomial. The score of the Student t copula with respect to the reciprocal of the degrees of freedom parameter η when $\eta = 0$ is given by*

$$\begin{aligned} s_\eta(\boldsymbol{\rho}) \quad = \quad & \sqrt{\frac{K(K+2)}{2}} L_2\left[\mathbf{y}'\mathbf{P}^{-1}(\boldsymbol{\rho})\mathbf{y}\right] - \sqrt{\frac{3}{2}}\sum_{k=1}^{K} H_4(y_k) \\ & + \frac{1}{2}\sqrt{\frac{3}{2}}\sum_{k=1}^{K}\left[\frac{\mathbf{p}'_{(k)}(\boldsymbol{\rho})\mathbf{P}^{-1}_{(kk)}(\boldsymbol{\rho})[\mathbf{y}_{(k)} - \mathbf{p}_{(k)}(\boldsymbol{\rho})y_k]}{1-\mathbf{p}'_{(k)}(\boldsymbol{\rho})\mathbf{P}^{-1}_{(kk)}(\boldsymbol{\rho})\mathbf{p}_{(k)}(\boldsymbol{\rho})}\right] H_3(y_k). \end{aligned} \tag{2}$$

Therefore, the LM test will simply be given by N times the square of the sample average of (2) evaluated at some consistent estimator of $\boldsymbol{\rho}$ divided by the variance of this score. [Supplemental Appendix E.4 contains the detailed expression for the asymptotic variance of (2) in the bivariate case; see Amengual and Sentana (2015) for the trivariate case]

The fact that $\eta = 0$ lies at the boundary of the admissible parameter space invalidates the usual distribution of the LR and Wald tests, which under the null will be a 50:50 mixture of χ^2_0 (=0 with probability 1) and χ^2_1. Although the distribution of the LM test statistic remains valid, intuition suggests that the one-sided nature of the alternative hypothesis should be taken into account to obtain a more powerful test. For that reason, we follow Fiorentini, Sentana and Calzolari (2003) in using the Kuhn-Tucker (KT) multiplier test introduced by Gouriéroux, Holly and Monfort (1980) instead, which is equivalent in large samples to the LR and Wald tests, and therefore, implicitly one-sided. Thus, we will reject H_0 at the $100\varkappa\%$ significance level if the average score with respect to η evaluated under the Gaussian null is strictly positive *and* the LM statistic exceeds the $100(1-2\varkappa)$ percentile of a χ^2_1 distribution. In this regard, the KT test is entirely analogous to the Wald test, which coincides with the one-sided t-test based on the ML estimator of η, except that it is based on the average score in (2). Further, it is important to remember that when there is a single restriction, such as in our case, those one-sided tests would be asymptotically locally more powerful (see e.g. Andrews (2001)).

Finally, given that (2) is computed in terms of Gaussian ranks, copula tests are invariant to strictly increasing univariate transformations of the observed series but not to affine multivariate transformations. Intuitively, the reason is that in the case of a copula the original variables are of direct interest.

3.2 Tests against asymmetric Student t copulas

As we mentioned before, the asymmetric Student t distribution is an important special case of the *GH* distribution. The derivation of the LM test for a multivariate normal copula versus an asymmetric one is complicated by the fact that $\mathbf{b}$ drops out from both the joint and marginal distributions when $\eta \to 0$ (see Mencía and Sentana (2012)). One standard solution in the literature to deal with testing situations with underidentified parameters under the null involves fixing those parameters to some arbitrary values, and then computing the appropriate test statistic for the chosen values.

Proposition 2 *The score of the asymmetric Student t copula with respect to the reciprocal of the degrees of freedom parameter η when $\eta = 0$ for fixed values of the skewness parameters $\mathbf{b}$ is given by*

$$\begin{aligned} s_{\eta}(\boldsymbol{\rho}, \mathbf{b}) &= s_{\eta}(\boldsymbol{\rho}) + \mathbf{b}'\mathbf{y}\left[\mathbf{y}'\mathbf{P}^{-1}(\boldsymbol{\rho})\mathbf{y} - (K+2)\right] - \sqrt{6}\sum_{k=1}^{K}\left[\beta_k(\boldsymbol{\rho}, \mathbf{b})H_3(y_k)\right] \\ &\quad + \sqrt{2}\sum_{k=1}^{K}\frac{\mathbf{p}'_{(k)}(\boldsymbol{\rho})\mathbf{P}^{-1}_{(kk)}(\boldsymbol{\rho})[\mathbf{y}_{(k)} - \mathbf{p}_{(k)}(\boldsymbol{\rho})y_k]}{1 - \mathbf{p}'_{(k)}(\boldsymbol{\rho})\mathbf{P}^{-1}_{(kk)}(\boldsymbol{\rho})\mathbf{p}_{(k)}(\boldsymbol{\rho})}\beta_k(\boldsymbol{\rho}, \mathbf{b})H_2(y_k). \end{aligned} \tag{3}$$

where $\beta(\boldsymbol{\rho}, \mathbf{b}) = \mathbf{P}(\boldsymbol{\rho})\mathbf{b}$.

On this basis, it would be straightforward to develop the associated test statistic, $LM_N(\mathbf{b})$. Unfortunately, it is not generally clear a priori what values of $\mathbf{b}$ are likely to prevail under the alternative of *GH* innovations. For that reason, we consider instead a second approach, which consists in computing the LM test for all possible values of $\mathbf{b}$, and then take the supremum over those parameter values.

It turns out that we can maximize $LM_N(\mathbf{b})$ with respect to $\mathbf{b}$ in closed form, and also obtain the asymptotic distribution of the resulting test statistic:

Proposition 3 *The supremum with respect to* $\mathbf{b}$ *of the LM tests based on (3) is equal to the sum of two asymptotically independent components: the symmetric Student t LM test based on (2), and a moment test based on the following K influence functions*

$$\begin{aligned} m_{b_k}(\boldsymbol{\rho}) &= y_k\left[\varsigma(\boldsymbol{\rho})-(K+2)\right]-\sqrt{6}\sum_{j=1}^{K}\mathbf{P}_{kj}(\boldsymbol{\rho})H_3(y_j) \\ &+\sqrt{2}\sum_{j=1}^{K}\frac{\mathbf{p}'_{(j)}(\boldsymbol{\rho})\mathbf{P}^{-1}_{(jj)}(\boldsymbol{\rho})[\mathbf{y}_{(j)}-\mathbf{p}_{(j)}(\boldsymbol{\rho})y_j]}{1-\mathbf{p}'_{(j)}(\boldsymbol{\rho})\mathbf{P}^{-1}_{(jj)}(\boldsymbol{\rho})\mathbf{p}_{(j)}(\boldsymbol{\rho})}\mathbf{P}_{kj}(\boldsymbol{\rho})H_2(y_j). \end{aligned} \tag{4}$$

This second moment test is asymptotically distributed as a χ^2 *distribution with* K *degrees of freedom when the true copula is Gaussian.*

Given that $s_\eta(\boldsymbol{\rho})$ is orthogonal to the K moment conditions in (4), we can conduct a partially one-sided test by combining the KT one-sided version of the Student t test and the moment test based on $m_{b_k}(\boldsymbol{\rho})$. This one-sided version should be equivalent in large samples to the corresponding LR test. The asymptotic distribution of the joint test under the null will be a 50:50 mixture of χ^2_K and χ^2_{K+1}, whose p-values are the equally weighted average of those χ^2 p-values.

Interestingly, the K moment conditions $E[m_{b_k}(\boldsymbol{\rho})]=0$ can also be used to consistently test a symmetric Student copula against an asymmetric one because we can show that the expected values of those influence functions would remain zero under this new null. But the test will be incorrectly sized if we used the covariance matrix of (4) under Gaussianity. To avoid size distortions, we can either compute the correct covariance expression by numerical quadrature or Monte Carlo integration for a given value of η, or else run the univariate regression of 1 on $m_{b_1}(\hat{\boldsymbol{\rho}}_T),\ldots,m_{b_K}(\hat{\boldsymbol{\rho}}_T)$. We use a heteroskedasticity and autocorrelation consistent (HAC) version of this second approach in the empirical application.

3.3 Tests against Generalized Hyperbolic copulas

As discussed by Mencía and Sentana (2012), there are three different paths along which a symmetric *GH* distribution converges to a Gaussian distribution. Specifically, the normal distribution can be achieved when (i) $\eta \to 0^-$ or (ii) $\eta \to 0^+$, regardless of the value of ψ; and (iii) $\psi \to 0$ irrespective of the value of η. In addition, one of the shape parameters becomes increasingly underidentified when the other one is on a normality path. Nevertheless, we can show that the score of the remaining identified parameter evaluated under the null of normality is (proportional to) (2) along those three paths. As a result, the LM/KT tests of Gaussian copula against a "symmetric" *GH* copula are numerically identical to the LM/KT tests against symmetric Student t that use the score in Proposition 1. In other words, all the symmetric versions of the *GH* copulas are asymptotically locally equivalent hypotheses (see Godfrey (1988)), to the extent that if a researcher decided to consider a normal-gamma mixture copula as the alternative to the normal copula, or indeed any other special case of the *GH* copula, she would end up with exactly the same statistic as in the case of the Student t.

The non-linear ARCH(1) model $y_t = (1+\gamma y_{t-1}^2)^{1/2} e^{\frac{1}{2}\delta y_{t-1}^2}\sqrt{\omega}\varepsilon_t^*$, $\varepsilon_t^*|I_{t-1} \sim iid\ N(0,1)$, which nests several popular specifications, provides a useful analogy. Conditionally homoskedasticity will happen *if and only if* both $\gamma = 0$ and $\delta = 0$. Under suitable conditions related to the magnitudes of those parameters, this model has a bounded full rank information matrix under the alternative. Under the null, though, its score contains two identical elements, so its information matrix is singular. As a result, the score test will asymptotically follow a χ_1^2 despite the fact that the null hypothesis $H_0 : \gamma = \delta = 0$ involves two parametric restrictions. As expected, the LM statistic for homoskedasticity against heteroskedasticity in this combined additive-multiplicative model also coincides with both Engle's (1982) test for additive heteroskedasticity and a test against multiplicative heteroskedasticity (see Godfrey (1988)). Therefore, it would be inappropriate to regard it as a test of only a single one of those alternatives.

In Supplemental Appendix C, we provide some further intuition for this result by introducing an alternative parametrization in terms of $\tau_1 = \eta\psi$ and a second parameter τ_2. We show that when $\tau_1 \to 0$, τ_2 drops out of the log-likelihood function along the aforementioned three paths to normality. As a result, the flexibility that the *GH* copula adds over and above the Student t copula provides no extra power to detect local deviations from normality. Obviously, the same would happen with Wald and LR tests because they are asymptotically equivalent to the KT tests both under the null and sequences of local alternatives.

Since the same is true for asymmetric *GH* alternatives, the LM/KT tests of Gaussian copula against an "asymmetric" *GH* copula will also be numerically identical to the LM/KT tests against asymmetric Student t in Propositions 2 and 3.

3.4 Comparison with distributional tests

It is interesting to compare our score test of the Student t copula to the corresponding test of the multivariate Student t distribution. Following Fiorentini, Sentana and Calzolari (2003), the score for η under the null is proportional to $L_2\left[\mathbf{y}'\mathbf{P}^{-1}(\boldsymbol{\rho})\mathbf{y}\right]$. Therefore, in the bivariate case with $\rho = 0$ their test becomes a moment test of

$$E\left(y_1^4 - 6y_1^2 + 3\right) + E\left(y_2^4 - 6y_2^2 + 3\right) + 2E[\left(y_1^2 - 1\right)\left(y_2^2 - 1\right)] = 0,$$

while our proposed copula score test reduces to

$$E[\left(y_1^2 - 1\right)\left(y_2^2 - 1\right)] = 0. \tag{5}$$

Given that the Gaussian ranks satisfy $E\left(y_1^4 - 6y_1^2 + 3\right) = E\left(y_2^4 - 6y_2^2 + 3\right) = 0$ by construction irrespective of the copula being Gaussian, and that $\left(y_1^2 - 1\right)\left(y_2^2 - 1\right)$ is orthogonal to those marginal fourth-order polynomials under the null, including them necessarily reduces (local) power.

In turn, the moments checked by the asymmetric component of the Mencía and Sentana (2012) test in exactly the same set up are

$$\begin{aligned} E\left(y_1^3 - 3y_1\right) + E[y_1(y_2^2 - 1)] &= 0, \\ E\left(y_2^3 - 3y_2\right) + E[(y_1^2 - 1)y_2] &= 0, \end{aligned}$$

while our proposed copula test reduces to a moment test based on

$$E[y_1(y_2^2 - 1)] = 0, \quad E[(y_1^2 - 1)y_2] = 0. \tag{6}$$

Given that the Gaussian ranks satisfy $E(y_1^3 - 3y_1) = E(y_2^3 - 3y_2) = 0$ by construction irrespective of the copula being Gaussian, and that $y_1(y_2^2 - 1)$ and $(y_1^2 - 1)y_2$ are orthogonal to those marginal third-order polynomials under the null, including them necessarily reduces (local) power, as in the symmetric component case.

Similar derivations in Supplemental Appendix D.4 for the case of $\rho \neq 0$ lead to analogous but longer expressions involving additional third and fourth cross-moments (of Hermite polynomials) of the Gaussian ranks involved.

3.5 Estimation uncertainty

So far we have assumed that the marginal distributions are known and the correlation matrix of the Gaussian copula is also known. Next, we study the implications of the fact that they will often have to be replaced by estimated counterparts.

3.5.1 Gaussian rank correlations

It turns out that under the Gaussian null, the information matrix is block diagonal between the correlation and shape parameters $\boldsymbol{\rho}$ and $\boldsymbol{\varphi}$:

Proposition 4 *The scores* $\mathbf{s}_{\boldsymbol{\rho}}(\boldsymbol{\rho}, \boldsymbol{\varphi})$ *and* $\mathbf{s}_{\boldsymbol{\varphi}}(\boldsymbol{\rho}, \boldsymbol{\varphi})$ *evaluated at* $\boldsymbol{\varphi} = \mathbf{0}$ *are orthogonal when the true copula is Gaussian.*

This result, which is the analogue for copulas of Proposition 3 in Fiorentini and Sentana (2007), is particularly convenient for our purposes because it will allow us to evaluate our score tests at any root-N consistent estimator of $\boldsymbol{\rho}$ without having to adjust the asymptotic variance of $\mathbf{s}_{\boldsymbol{\varphi}}(\boldsymbol{\phi})$ for parameter uncertainty, where $\boldsymbol{\phi} = (\boldsymbol{\rho}', \boldsymbol{\varphi}')'$.

3.5.2 Replacing margins with empirical cdf's

The most common solution to the fact that the marginal distributions of the K variables in the observed vector x are rarely known in practice is a two-step estimation procedure, whereby the margins $G_{1k}(x_k)$ are replaced by their (re-scaled) empirical cdf counterparts $\hat{G}_{1k}(x_k)$, where the scaling factor $N/(N+1)$ is only introduced to avoid potential problems with the copula density blowing up at the boundary of $[0,1]^K$. In this manner, the proposed tests can be viewed as functions of the Gaussian ranks obtained from the (uniform) sample ranks. Smoothed versions of the empirical cdf can also be used, but the effects should be the same (up to first-order).

The use of sample ranks has two implications. First, the exact discrete uniform nature of their distribution simplifies some of the previous expressions. Specifically, the sample averages of all the odd-order Hermite polynomials of the Gaussian ranks will be identically zero, while the sample averages of the even-order ones will converge to zero at faster than square root N rates. Second, it effectively transforms the Gaussian ML estimation procedure we have considered so far into a sequential semiparametric procedure, which requires us to take into account the sample uncertainty resulting from its non-parametric first-stage (see Newey and McFadden (1994)). Otherwise, our test statistics will have size distortions even in large samples.

Following Chen and Fan (2006a), for the *GH* copulas that we consider as alternatives, we show that it is possible to capture the variance modification in the scores of their shape parameters, $\mathbf{s}_{\boldsymbol{\varphi}}(\boldsymbol{\rho}, \mathbf{0})$, by adding linear combinations of third and fourth Hermite polynomials in those variables. The following result provides the expressions for the corrections of the influence functions in Propositions 1, 2 and 3:

Proposition 5 *The correction of* $s_\eta(\boldsymbol{\rho})$ *is given by*

$$\begin{aligned} n_\eta(\boldsymbol{\rho}) &= -\frac{1}{4}\sqrt{\frac{3}{2}}\sum_{k=1}^{K}\frac{\mathbf{p}'_{(k)}(\boldsymbol{\rho})\mathbf{P}^{-1}_{(kk)}(\boldsymbol{\rho})\mathbf{p}_{(k)}(\boldsymbol{\rho})}{1-\mathbf{p}'_{(k)}(\boldsymbol{\rho})\mathbf{P}^{-1}_{(kk)}(\boldsymbol{\rho})\mathbf{p}_{(k)}(\boldsymbol{\rho})}H_4(y_k) \\ &\quad +\frac{1}{4}\sqrt{\frac{3}{2}}\sum_{k=1}^{K}\sum_{h\neq k}\frac{\mathbf{p}'_{(h)}(\boldsymbol{\rho})\mathbf{P}^{-1}_{(h)}(\boldsymbol{\rho})}{1-\mathbf{p}'_{(h)}(\boldsymbol{\rho})\mathbf{P}^{-1}_{(hh)}(\boldsymbol{\rho})\mathbf{p}_{(h)}(\boldsymbol{\rho})}\mathbf{P}^3_{kh}(\boldsymbol{\rho})H_4(y_k), \end{aligned}$$

the correction of $m_{b_k}(\boldsymbol{\rho})$ *by*

$$\begin{aligned} n_{b_k}(\boldsymbol{\rho}) &= -\sqrt{\frac{2}{3}}\frac{\mathbf{p}'_{(k)}(\boldsymbol{\rho})\mathbf{P}^{-1}_{(kk)}(\boldsymbol{\rho})\mathbf{p}_{(k)}(\boldsymbol{\rho})}{1-\mathbf{p}'_{(k)}(\boldsymbol{\rho})\mathbf{P}^{-1}_{(kk)}(\boldsymbol{\rho})\mathbf{p}_{(k)}(\boldsymbol{\rho})}H_3(y_k) \\ &\quad +\sqrt{\frac{2}{3}}\sum_{j\neq k}\frac{\mathbf{p}'_{(j)}(\boldsymbol{\rho})\mathbf{P}^{-1}_{(jj)}(\boldsymbol{\rho})}{1-\mathbf{p}'_{(j)}(\boldsymbol{\rho})\mathbf{P}^{-1}_{(jj)}(\boldsymbol{\rho})\mathbf{p}_{(j)}(\boldsymbol{\rho})}\mathbf{P}^2_{kj}(\boldsymbol{\rho})H_3(y_k) \\ &\quad -\sqrt{\frac{2}{3}}\sum_{j\neq k}\frac{\mathbf{p}'_{(j)}(\boldsymbol{\rho})\mathbf{P}^{-1}_{(jj)}(\boldsymbol{\rho})\mathbf{p}_{(j)}(\boldsymbol{\rho})}{1-\mathbf{p}'_{(j)}(\boldsymbol{\rho})\mathbf{P}^{-1}_{(jj)}(\boldsymbol{\rho})\mathbf{p}_{(j)}(\boldsymbol{\rho})}\mathbf{P}_{kj}(\boldsymbol{\rho})H_3(y_j) \\ &\quad +\sqrt{\frac{2}{3}}\sum_{j\neq k}\sum_{h\neq j}\frac{\mathbf{p}'_{(h)}(\boldsymbol{\rho})\mathbf{P}^{-1}_{(hh)}(\boldsymbol{\rho})}{1-\mathbf{p}'_{(h)}(\boldsymbol{\rho})\mathbf{P}^{-1}_{(hh)}(\boldsymbol{\rho})\mathbf{p}_{(h)}(\boldsymbol{\rho})}\mathbf{P}_{kh}(\boldsymbol{\rho})\mathbf{P}^2_{jh}(\boldsymbol{\rho})H_3(y_j), \end{aligned}$$

and the correction of $s_\eta(\boldsymbol{\rho},\mathbf{b})$ *by* $n_\eta(\boldsymbol{\rho})+\sum_{k=1}^{K}b_k n_{b_k}(\boldsymbol{\rho})$.

For example, for the bivariate asymmetric Student t, the score corrections will be

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

$$n_{b_1}(\rho) = \sqrt{\frac{2}{3}}\rho^2\left[(6-\rho^2)H_3(y_1)+5\rho H_3(y_2)\right]$$

and

$$n_{b_2}(\rho) = \sqrt{\frac{2}{3}}\rho^2\left[(6-\rho^2)H_3(y_2)+5\rho H_3(y_1)\right].$$

Importantly, given that Amengual and Sentana (2017) show that the adjustment to the scores of the correlation coefficients resulting from the nonparametric estimation of the margins only involves linear combinations of second-order polynomials in each of the variables, the orthogonality between the original scores for correlation and shape parameters stated in Proposition 4 is preserved in the modified scores because Hermite polynomials form an orthonormal basis under Gaussianity. For the same reason, the modified scores for η and the elements of $\mathbf{b}$ remain uncorrelated, as in Proposition 3.

3.5.3 Replacing margins with parametric cdf's

Sometimes, researchers prefer to specify parametric marginal distributions, estimating their parameters by univariate ML. Let $\boldsymbol{\lambda}=(\boldsymbol{\lambda}'_1,\ldots,\boldsymbol{\lambda}'_K)'$ denote the parameters of the K marginal distributions and $L_1(\boldsymbol{\lambda}_1), \ldots, L_K(\boldsymbol{\lambda}_K)$ the corresponding log-likelihood functions. Similarly, let $m_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}, \boldsymbol{\rho}) = [\partial L_1(\boldsymbol{\lambda}_1)/\partial \boldsymbol{\lambda}_1', \ldots, \partial L_K(\boldsymbol{\lambda}_K)/\partial \boldsymbol{\lambda}_K']'$ denote the associated scores, with $m_{\boldsymbol{\rho}}(\boldsymbol{\lambda}, \boldsymbol{\rho}) = [\partial L(\boldsymbol{\rho}, \mathbf{0})/\partial \boldsymbol{\rho}']'$. On the basis of the results in Joe (2005), we can prove the following result:

Proposition 6 *When the Gaussian ranks are obtained after estimating the parametric marginal distributions by ML, the asymptotic variance of the influence functions*

$$\mathbf{m}_{\boldsymbol{\varphi}}(\boldsymbol{\lambda}, \boldsymbol{\rho}) = [s_{\eta}(\boldsymbol{\rho}), s_{b_1}(\boldsymbol{\rho}), \ldots, s_{b_K}(\boldsymbol{\rho})]'$$

under the null is given by

$$\mathcal{V} = \mathcal{V}_{\boldsymbol{\varphi}} + \mathcal{A}\mathcal{D}^{-1}\mathcal{M}\mathcal{D}'^{-1}\mathcal{A}',$$

where

$$\mathcal{V}_{\boldsymbol{\varphi}} = V[\mathbf{m}_{\boldsymbol{\varphi}}(\boldsymbol{\lambda}, \boldsymbol{\rho})],$$

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

$$\mathcal{M} = V\left[\begin{pmatrix} \mathbf{m}_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}_0, \boldsymbol{\rho}_0) \\ \mathbf{m}_{\boldsymbol{\rho}}(\boldsymbol{\lambda}_0, \boldsymbol{\rho}_0) \end{pmatrix}\right] = E\begin{pmatrix} \frac{\partial L_1(\boldsymbol{\lambda}_1)}{\partial \boldsymbol{\lambda}_1}\frac{\partial L_1(\boldsymbol{\lambda}_1)}{\partial \boldsymbol{\lambda}_1'} & \cdots & \frac{\partial L_1(\boldsymbol{\lambda}_1)}{\partial \boldsymbol{\lambda}_1}\frac{\partial L_K(\boldsymbol{\lambda}_K)}{\partial \boldsymbol{\lambda}_K'} & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots \\ \frac{\partial L_K(\boldsymbol{\lambda}_K)}{\partial \boldsymbol{\lambda}_K}\frac{\partial L_1(\boldsymbol{\lambda}_1)}{\partial \boldsymbol{\lambda}_1'} & \cdots & \frac{\partial L_K(\boldsymbol{\lambda}_K)}{\partial \boldsymbol{\lambda}_K}\frac{\partial L_K(\boldsymbol{\lambda}_K)}{\partial \boldsymbol{\lambda}_K'} & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \frac{\partial L(\boldsymbol{\rho},\mathbf{0})}{\partial \boldsymbol{\rho}}\frac{\partial L(\boldsymbol{\rho},\mathbf{0})}{\partial \boldsymbol{\rho}'} \end{pmatrix},$$

and

$$\mathcal{D} = E\begin{pmatrix} \partial \mathbf{m}_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}_0, \boldsymbol{\rho}_0)/\partial \boldsymbol{\lambda}' & \partial \mathbf{m}_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}_0, \boldsymbol{\rho}_0)/\partial \boldsymbol{\rho}' \\ \partial \mathbf{m}_{\boldsymbol{\rho}}(\boldsymbol{\lambda}_0, \boldsymbol{\rho}_0)/\partial \boldsymbol{\lambda}' & \partial \mathbf{m}_{\boldsymbol{\rho}}(\boldsymbol{\lambda}_0, \boldsymbol{\rho}_0)/\partial \boldsymbol{\rho}' \end{pmatrix} = E\begin{pmatrix} \frac{\partial^2 L_1(\boldsymbol{\lambda}_1)}{\partial \boldsymbol{\lambda}_1 \partial \boldsymbol{\lambda}_1'} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \cdots & \frac{\partial^2 L_K(\boldsymbol{\lambda}_K)}{\partial \boldsymbol{\lambda}_K \partial \boldsymbol{\lambda}_K'} & \mathbf{0} \\ \frac{\partial^2 L(\boldsymbol{\rho},\mathbf{0})}{\partial \boldsymbol{\rho} \partial \boldsymbol{\lambda}_1'} & \cdots & \frac{\partial^2 L(\boldsymbol{\rho},\mathbf{0})}{\partial \boldsymbol{\rho} \partial \boldsymbol{\lambda}_K'} & \frac{\partial^2 L(\boldsymbol{\rho},\mathbf{0})}{\partial \boldsymbol{\rho} \partial \boldsymbol{\rho}'} \end{pmatrix}.$$

One can then modify our proposed tests by simply replacing $\mathcal{V}_{\boldsymbol{\varphi}}$ by the adjusted asymptotic covariance matrix $\mathcal{V}$, as we illustrate in the Monte Carlo section below.

3.6 Constrained indirect estimation

If a researcher who uses our proposed tests does not reject the null hypothesis, she can rely on the Gaussian copula evaluated at the Gaussian rank correlation coefficients. However, if she rejects, she might be interested in estimating the parameters of the alternative distributions that we have considered.

Conceptually, the most straightforward procedure for estimating the parameters of those non-Gaussian copulas would be ML using the analytical expressions for the scores that we have derived. Unfortunately, this is easier said than done because the scores in (B2) are computationally involved under the alternative (see the Supplemental Appendix to Mencía and Sentana (2012) for some of the required expressions), especially taking into account that the evaluation of those scores also requires the derivatives of the quantile function of the univariate GH distribution with respect to its shape parameters.

Nevertheless, it is possible to come up with much simpler consistent estimators of $\boldsymbol{\rho}$ and $\boldsymbol{\varphi}$ along the lines of Calzolari, Fiorentini and Sentana (2004). Specifically, we can estimate those coefficients for a specific parametric copula by generating data from this copula and matching in the simulated data the values in the original data of both the Gaussian rank correlation coefficients and the test statistics we have proposed. The fact that the scores of the two shape parameters of the GH copula are proportional under the Gaussian null implies that researchers have to decide which non-Gaussian copula within the GH family (asymmetric t, normal-gamma mixture, hyperbolic, normal inverse Gaussian or Laplace) they would like to estimate. Proposition 4 in Calzolari, Fiorentini and Sentana (2004) guarantees the consistency and asymptotic normality of these constrained indirect estimators. Similarly, their Proposition 7 characterizes their efficiency loss relative to MLE.

4 Monte Carlo evidence

In this section, we assess the finite sample size and power properties of the testing procedures discussed above by means of several extensive Monte Carlo exercises.

4.1 Design and estimation details

As in our empirical application, we look at the case $K = 2$, but we also consider $K = 10$ to assess the performance of our testing procedures in moderately large dimensions. To simplify the bootstrap procedure, though, we impose an equicorrelated structure on $\mathbf{P}(\boldsymbol{\rho})$. We consider three different sample sizes: 200, 800 and 3,200, and two different values of the correlation parameter ρ: .25 and .75 (the latter relegated to the Supplemental Appendix). Regarding the marginal distributions, we consider both the benchmark case in which they are known and when they are replaced by the empirical CDFs. In the bivariate case, we also consider parametrically estimated margins. Specifically, we assume exponential margins, whose parameters we estimate individually by ML. As for the relevant quantities involved in Proposition 6, they reduce to:

$$
\begin{aligned}
-E\left[\frac{\partial^2 L_k(\lambda_k)}{\partial \lambda_k^2}\right] &= E\left\{\left[\frac{\partial L_k(\lambda_k)}{\partial \lambda_k}\right]^2\right\} = \frac{1}{\lambda_k^2} \text{ for } k=1,2, \text{ and} \\
-E\left[\frac{\partial^2 L(\rho,\mathbf{0})}{\partial \rho^2}\right] &= E\left\{\left[\frac{\partial L(\rho,\mathbf{0})}{\partial \rho}\right]^2\right\} = \frac{1+\rho^2}{(1-\rho^2)^2},
\end{aligned}
$$

while then non-zero elements in $\mathcal{A}$, together with $E(\partial L_1/\partial\lambda_1 \times \partial L_2/\partial\lambda_2)$ and $E[\partial^2 L/(\partial\rho\partial\lambda_k)]$ for $k=1,2$, can be easily computed by numerical integration. Details on how we simulate those multivariate distributions and their parametrizations can be found in Supplemental Appendix E.1.

Given that the asymptotic distributions that we have derived in Section 3 turn out to be unreliable in small samples (see Amengual and Sentana (2015)), we compute bootstrap critical values. Specifically, we employ a parametric bootstrap procedure with 10,000 simulated samples for all tests. In this way, we can automatically compute size-adjusted rejection rates, as forcefully argued by Horowitz and Savin (2000). Despite the asymptotic orthogonality of the scores corresponding to the correlation and shape parameters, our bootstrap procedure takes into account the sensitivity of the critical values to the values of ρ to avoid ruling out higher-order refinements. We achieve this by computing simulated critical values for a fine grid of values of ρ between -1 and 1 (see Supplemental Appendix E.2 for further details).

Importantly, we compare our proposed score tests to the Kolmogorov–Smirnov (KS) and Cramér–von Mises (CvM) tests for copula models, which are often reported in empirical work (see Rémillard (2017) for details), as well as to the recent consistent tests proposed by Panchenko (2005) and Genest et al. (2009). Since the asymptotic distributions of these tests in copula models with non-parametric margins are unknown, we rely on their bootstrap values.

4.2 Size properties

Table 1 shows that the parametric bootstrap rejection rates are close to being perfect for all the different samples sizes and significance levels we consider. Specifically, Panel A contains rejection rates under the null at the 1%, 5% and 10% levels for the bivariate case while Panel B does the same for $K=10$. The row labels LM–t and LM–At correspond to the Lagrange multiplier tests based on the score of the symmetric and asymmetric Student t copulas, respectively, while KT–t and KT–At are the corresponding Kuhn-Tucker versions. In turn, Skew denotes the Lagrange multiplier test based on the K moment conditions $m_{b_k}(\rho)$ in Proposition 4. As for the non-parametric tests, $S^{(C)}$ 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 correspond to the Kolmogorov–Smirnov and Cramér–von Mises tests for copula models.

4.3 Power properties

For our first alternative, we draw from (symmetric) Student t copulas with 20 (100) degrees of freedom when $K = 2$ ($K = 10$) but the same correlation matrices as in the Gaussian case. Table 2 reports the Monte Carlo rejection rates at the 1%, 5% and 10% significance levels. By and large, the behavior of the different test statistics is in accordance with expectations. In particular, the Student t tests are the most powerful, with the Kuhn-Tucker versions being more powerful than the Lagrange multiplier's ones. In contrast, all the non-parametric competitors have close to trivial power in samples of 200 observations. In line with the evidence on local power in Supplemental Appendix D (see Figures D1 and D2), the rejection rates are higher the higher the correlation. Not surprisingly, power is lower when the margins are estimated non-parametrically than when they are known, with the parametric case usually in between.

Next, we generate observations from Asymmetric Student t copulas with 20 (100) degrees of freedom $K = 2$ ($K = 10$) and identical correlation matrices but negative tail dependence characterized by $b_k = -.75$ ($b_k = -.15$). As can be seen from Table 3, in this case the asymmetric Student t tests are the most powerful, with the Kuhn-Tucker version being more powerful than the Lagrange multiplier's one.

Therefore, our Monte Carlo results confirm the local power analysis in Supplemental Appendix D, so that our proposed tests provide a very strong indication of the directions along which the efforts to improve the specification of the model should focus.

Arguably, though, this conclusion is not entirely surprising. For that reason, we also assess the power of our tests against certain nonnormal alternatives outside the GH family by considering the Skew t copula, which is obtained from the Skew t distribution proposed by Azzalini and Capitanio (2003) (see also Kotz and Nadarajah (2004)). The results reported in Table 4 clearly show that our proposed tests continue to have good power in situations in which the true copula does not correspond with the alternative they are designed for. Moreover, they beat the competition by a long margin.

5 Momentum and reversals in stock returns

In this section we apply our Gaussian copula tests to formally analyze the cross-sectional dependence between monthly returns on individual U.S. stocks in the CRSP database and some of their observable characteristics. Given their prominence in the empirical finance literature, we focus on (i) short term reversals, in which the observable characteristic are the individual stock returns over the previous month; and (ii) momentum, where the relevant variable are the individual stock returns from month $t-2$ to month $t-12$. In line with most previous studies (see e.g. Asness, Moskowitz and Pedersen (2013)), we only consider common equities (CRSP sharecodes 10 and 11) and exclude those stocks with share prices less than \$1 at the beginning of the holding period. We also restrict our analysis to those firms with at least 60 months of return history, so that we focus on liquid stocks with low transaction costs and high tradability.

An important advantage of working with either uniform or Gaussian ranks is that we obtain exactly the same numerical results whether we work with the original returns or with their deviations from the returns on an aggregate stock market index or the level of the risk free rate. Nevertheless, the presence of other time-varying effects that may potentially affect different firms differently could alter the cross-sectional dependence. For that reason, we carry out our analysis both at the aggregate level, i.e. using all individual stocks, and at the industry level (see Supplemental Appendix G.1 for the latter).

Our dataset spans the period from January 1997 to December 2012. For each and every month, we first transform the observed variables for each individual stock into their cross-sectional (uniform) ranks and then into their Gaussian ranks. Importantly, though, the empirical cdfs used to transform the original observations into Gaussian ranks are re-estimated for every single month so as to allow for complete flexibility in the time-variation of the marginal distributions. Thus, we end up with a cross-section of the form $Y_t = \{(y_{11}^t, y_{21}^t), ..., (y_{1N_t}^t, y_{2N_t}^t)\}$, where N_t is the number of individual stocks for which we have data on both their return and the relevant observable characteristic for month t. Although we could apply our tests on a monthly basis, from the point of view of devising trading strategies, a period by period cross-sectional analysis is of little interest. For that reason, we pool all $\sum_{t=1}^{T} N_t$ bivariate observations $\mathbf{Y} = \{Y_1, ..., Y_T\}$ and exploit the moment-based interpretation of our tests as follows. First, for each t we compute the cross-sectional average of the log-likelihood scores with respect to all the parameters $\boldsymbol{\phi}$, say $\bar{\mathbf{s}}_{\boldsymbol{\phi}t}(Y_t; \rho) = N_t^{-1} \sum_{n=1}^{N_t} \mathbf{s}_{\boldsymbol{\phi}}(y_{1n}^t, y_{2n}^t; \rho)$. Then, we time-average those scores, thereby creating a pooled average score $\bar{\mathbf{s}}_{\phi}(Y_t;\rho) = T^{-1}\sum_{t=1}^{T}\bar{\mathbf{s}}_{\phi t}(Y_t;\rho)$, on the basis of which we can estimate the correlation coefficient ρ and construct our tests. The only complication is that our pooled procedure requires the computation of robust standard errors to capture the potential time-series dependence in $\bar{\mathbf{s}}_{\phi t}(Y_t;\rho)$ for different t's (see Supplemental Appendix E.3 for details).

Before characterizing dependence through the copula, though, it is convenient to look at correlations. Panel A of Table 5 presents the parameter estimates and their corresponding asymptotically robust standard errors for the Pearson, Spearman and Gaussian rank correlation coefficients (again, for a detailed description see Supplemental Appendix E.3). All correlation parameters have the expected sign with the exception of the Pearson correlation estimate for short term reversals. In addition, the Pearson correlation coefficient for momentum has the right sign but it is statistically insignificant, in sharp contrast to the Spearman and Gaussian rank correlations. This confirms the sensitivity of the estimators of the Pearson coefficient and the associated slopes to the presence of outliers.

In Panel B of Table 5 we report the Gaussian copula test statistics, with KT–t and KT–At denoting the KT versions of the tests against Student t and asymmetric Student t copulas (LM versions are numerically identical in our data), and Skew the LM test based on the two moment conditions $m_{b_k}(\boldsymbol{\rho})$ in Proposition 4. 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. Importantly, the source of the rejection is not only the "cokurtosis" between the Gaussian ranks, but also their "coskewness", specially for momentum strategies. In this regard, it is worth emphasizing that the use of the HAC procedure ensures that the asymmetric component of the test is correctly sized under the null of a symmetric Student t copula too, as argued at the end of section 3.2. We discuss the trading implications of these empirical results in Supplemental Appendix G.2.

Given those rejections, the natural next step is to gauge the parameters of the alternative distributions that we have considered. As explained in Section 3.6, we can consistently do so by means of an equality constrained indirect estimation procedure which matches the observed tests statistics and the estimated Gaussian rank correlations. In Panel C of Table 5 we report the resulting pooled estimates of the correlation and shape parameters of both symmetric and asymmetric Student t copulas based on simulated sample paths of size 100,000. We find moderate negative tail dependence but quite substantive "leptokurtosis", with estimated degrees of freedom in the neighborhood of 5. Importantly, Supplemental Appendix G.1 shows that our empirical results are robust to estimating the model at the industry level.

6 Conclusions

We derive computationally simple and intuitive expressions for score tests of Gaussian copulas against GH alternatives. In this regard, we show that all the GH copulas are asymptotically locally equivalent hypothesis to the Gaussian null, so that if a researcher decided to consider a normal-gamma mixture copula, or indeed any other special case of the GH copula as the alternative, she would end up with exactly the same statistic as in the case of the Student t. Asymptotically, the same is true of the Likelihood Ratio and Wald tests.

We decompose our score tests into simple moment tests based on linear combinations of cross products of Hermite polynomials of the Gaussian ranks up to order four. By taking into account the partial one-sided nature of some of the alternative hypotheses, we also obtain more powerful one-sided Kuhn-Tucker versions that are equivalent to the Likelihood Ratio test, whose standard asymptotic distribution under the null we derive. This equivalence implies that our approach has a likelihood interpretation, which provides a formal justification for focusing on the specific moments that we test. In turn, this likelihood interpretation confirms that we can learn from our tests in which directions the model is really worth extending because the score vector gives the direction of steepest ascent. Finally, the expression of our score tests as moment tests also allows us to show that they are more powerful than multivariate distributional tests applied to the Gaussian ranks because they do not waste power in checking the normality of the marginal distributions, which are Gaussian by construction.

We conduct detailed Monte Carlo exercises with a range of correlation matrices to study our proposed tests in finite samples. We find that the parametric bootstrap rejection rates are almost perfect for all samples sizes. Moreover, the finite sample power of the different test statistics agrees with what the asymptotic results would suggest. Importantly, our findings indicate that our parametric tests have substantially more power than the existing non-parametric ones even for departures for which our procedures are not optimal.

In an empirical application to CRSP data, we assess the widely held view that stocks that underperformed in the past month (short term reversals) and those that outperformed in previous months (momentum) show superior performance. Our tests indicate that those effects are better captured by non-Gaussian copulas, whose parameters we estimate by an indirect inference procedure that matches our test statistics in the simulated and real data.

As a valuable extension, we could explicitly consider more complex models by conditioning on past values of $\mathbf{x}$ or present and past values of some exogenous variables $\mathbf{z}$ (see e.g. Patton (2006) or Chen and Fan (2006a) for some interesting examples of dynamic copula models).

Another interesting extension would be to develop testing procedures that direct power over the third quadrant, say, instead of the entire distribution. One possible approach would be to use a *GH*-based test in which we fix both asymmetric parameters to be big and negative.

It would also be interesting to compare our score tests to information criteria approaches (see e.g. Chen and Fan (2005 and 2006b)), as well as to tests based on non-parametric estimates of the copula density (see Fermanian (2005) and Scaillet (2007)). In addition, we could develop tests that take as their null hypothesis other special cases of the *GH* copula, such as the popular (symmetric) Student t, which is nested in the *GH* family when $\eta > 0$ and $\psi = 1$. This poses two technical complications relative to the Gaussian tests. First, the information matrix will no longer be block diagonal between the correlation and shape parameters. Second, the score with respect to ψ will be identically 0 under the null, which means that we will have to rely on what Lee and Chesher (1986) called an extremum test. All these extensions constitute promising avenues for further research.

Acknowledgements

The authors would like to thank Manuel Arellano, Yanqin Fan, Pedro García Ares, Pascal Lavergne, Javier Mencía, Andrew Patton and Javier Perote for their comments, advice and suggestions. We have also benefitted from the feedback provided by seminar participants at Alicante, CEMFI, Duke, Erasmus, ESSEC, Exeter, Indiana, Liverpool, Murcia, Surrey, Torcuato Di Tella, UCL, UPF and Yale, as well as audiences at the TSE Financial Econometrics Conference (May 2013), the XXI Finance Forum (Segovia, November 2013), the XXXVIII Symposium on Economic Analysis (Santander, December 2013), the LATAM Econometrics Workshop (São Paulo, December 2013), the IEEM Meeting (Montevideo, December 2013), the 68^{th} ESEM (Toulouse, August 2014), the 11^{th} World Congress of the Econometric Society (Montreal, August 2015, the New Methods for the Empirical Analysis of Financial Markets SanFI conference (Comillas, June 2017) and the EcoSta (Hong Kong, June 2017). We thank Difang Huang and Zaici Li for their help in collecting the data and especially Xinyue Bei and Julio Crego for their excellent research assistance. The input of a co-editor and two anonymous referees has also greatly improved the paper. Of course, the usual caveat applies. Financial support from the Spanish Ministries of Science & Innovation and Economy & Competitiveness through grants ECO 2011-26342 and 2014-59262, respectively, and the Santander - CEMFI Research Chair on Finance is gratefully acknowledged.

A Proofs

Henceforth, we maintain Assumption 1 in the text, and Assumptions 2-3 in Supplemental Appendix B.

A.1 Proposition 1

We can easily compute the first two terms of (B2) in Supplemental Appendix B, whose limits are

$$\begin{aligned}\lim_{\eta\to 0}\frac{\partial \ln f_K\left(y_1,...,y_K;\rho,\eta\right)}{\partial\eta} &= \sqrt{\frac{K(K+2)}{2}}\times L_2(\varsigma),\\ \lim_{\eta\to 0}\sum\nolimits_{k=1}^{K}\frac{\partial \ln f_{1k}\left(y_k;\eta\right)}{\partial\eta} &= \sqrt{\frac{3}{2}}\times\sum_{k=1}^{K}H_4(y_k),\end{aligned}$$

where $L_j(\cdot)$ and $H_j(\cdot)$ are the normalized Laguerre and Hermite polynomials of order j, respectively. As for the remaining terms in (B2), we can use the fact that for a generic copula density such as the one in (B1),

$$\lim_{\boldsymbol{\varphi}\to\mathbf{0}}\frac{\partial \ln f_K\left(y_1,...,y_K;\boldsymbol{\rho},\boldsymbol{\varphi}\right)}{\partial\varsigma}=-\frac{1}{2}\quad\text{and}\quad\lim_{\boldsymbol{\varphi}\to\mathbf{0}}\frac{\partial \ln f_{1k}\left(y_k;\boldsymbol{\varphi}\right)}{\partial y_k}=-y_k,$$

so that

$$\lim_{\boldsymbol{\varphi}\to\mathbf{0}}\left[\frac{\partial \ln f_K\left(\varsigma;\boldsymbol{\rho},\boldsymbol{\varphi}\right)}{\partial\varsigma}\frac{\partial\varsigma}{\partial y_k}-\frac{\partial \ln f_{1k}\left(y_k;\boldsymbol{\varphi}\right)}{\partial y_k}\right]=\frac{\mathbf{p}'_{(k)}(\boldsymbol{\rho})\mathbf{P}^{-1}_{(kk)}(\boldsymbol{\rho})[\mathbf{y}-\mathbf{p}_{(k)}(\boldsymbol{\rho})y_i]}{1-\mathbf{p}'_{(k)}(\boldsymbol{\rho})\mathbf{P}^{-1}_{(kk)}(\boldsymbol{\rho})\mathbf{p}_{(k)}(\boldsymbol{\rho})},\tag{A1}$$

where $\mathbf{y}-\mathbf{p}_{(k)}y_k$ are residuals of univariate simple regressions of $\mathbf{y}_{-k}$ onto y_k because $\mathbf{p}_{(k)}$ are the corresponding OLS coefficients, while the denominator is the residual variance in a regression of y_k onto the remaining components of $\mathbf{y}$. As for $\partial\varepsilon_k(\eta)/\partial\eta=\partial F^{-1}_{1k}(u_k;\eta)/\partial\eta$, differentiating $\int_{-\infty}^{F^{-1}_{1k}(u_k,\eta)}f_{1k}(y_k;\eta)dy_k=u_k$ with respect to η yields

$$\frac{\partial F^{-1}_{1k}(u_k;\eta)}{\partial\eta}=\frac{-1}{f_{1k}[F^{-1}_{1k}(u_k;\eta);\eta]}\int_{-\infty}^{F^{-1}_{1k}(u_k,\eta)}\frac{\partial f_{1k}(y_k;\eta)}{\partial\eta}dy_k.$$

But then, noticing that $\lim_{\eta\to 0}f_{1k}[F^{-1}_{1k}(u_k;\eta);\eta]=\phi[\Phi^{-1}(u_k)]$ and that

$$\lim_{\eta\to 0}\frac{\partial f_{1k}(y_k;\eta)}{\partial\eta}=\phi(y_k)\times\sqrt{\frac{3}{2}}\times H_4(y_k),$$

we obtain

$$\lim_{\eta\to 0}\frac{\partial F^{-1}_{1k}(u_k;\eta)}{\partial\eta}=\frac{1}{2}\sqrt{\frac{3}{2}}\times H_3(y_k).$$

Collecting terms finally yields (2). □

A.2 Proposition 2

For fixed $\mathbf{b}$, the LM test is based on the average score with respect to η evaluated at the limit of $\eta \to 0$. In this regard, we first obtain the parameters of the corresponding marginal distributions appearing in (B1). Specifically, if $\mathbf{y} \sim At(\mathbf{0}, \mathbf{P}(\boldsymbol{\rho}), \eta, \mathbf{b})$ with $\boldsymbol{\beta} = \mathbf{P}^{1/2\prime}(\boldsymbol{\rho})\mathbf{b}$, then $y_k \sim At(0, 1, \eta, ß_k(\eta, \mathbf{b}))$ where

$$ß_k(\eta, \mathbf{b}) = \frac{c(\mathbf{b}'\mathbf{P}(\boldsymbol{\rho})\mathbf{b}, \eta)\imath_k'\mathbf{P}(\boldsymbol{\rho})\mathbf{b}}{1 + [c(\mathbf{b}'\mathbf{P}(\boldsymbol{\rho})\mathbf{b}, \eta) - 1]\imath_k'\mathbf{P}(\boldsymbol{\rho})\mathbf{b}\mathbf{b}'\mathbf{P}(\boldsymbol{\rho})\imath_k/\mathbf{b}'\mathbf{P}(\boldsymbol{\rho})\mathbf{b}},$$

with $\imath_k$ denoting a $K \times 1$ vector with 1 in its k'th position and 0's otherwise, and

$$c(\mathbf{b}'\mathbf{P}(\boldsymbol{\rho})\mathbf{b}, \eta) = \frac{-(1-4\eta) + \sqrt{(1-4\eta)^2 + 8\eta(1-4\eta)\mathbf{b}'\mathbf{P}(\boldsymbol{\rho})\mathbf{b}}}{4\eta\mathbf{b}'\mathbf{P}(\boldsymbol{\rho})\mathbf{b}}.$$

In this context, we can write $s_\eta(\boldsymbol{\rho}, \mathbf{b})$ as

$$\begin{aligned}
&\frac{\partial \ln f_K(y_1, ..., y_K; \boldsymbol{\rho}, \eta, \mathbf{b})}{\partial \eta} - \sum_{k=1}^{K} \frac{\partial \ln f_{1k}[y_k; \eta, ß_k(\eta, \mathbf{b})]}{\partial \eta} - \sum_{k=1}^{K} \frac{\partial \ln f_{1k}[y_k; \eta, ß_k(\eta, \mathbf{b})]}{\partial ß_k} \frac{\partial ß_k(\eta, \mathbf{b})}{\partial \eta} \\
&+ \sum_{k=1}^{K} \left\{ \frac{\partial \ln f_K(y_1, ..., y_K; \boldsymbol{\rho}, \eta, \mathbf{b})}{\partial y_k} - \frac{\partial \ln f_{1k}[y_k; \eta, ß_k(\eta, \mathbf{b})]}{\partial y_k} \right\} \frac{\partial F_{1k}^{-1}[u_k; \eta, ß_k(\eta, \mathbf{b})]}{\partial \eta}. \qquad \text{(A2)}
\end{aligned}$$

As for the first two terms of (A2), Mencía and Sentana (2012) provide the corresponding expressions, which reduce to

$$\begin{aligned}
\lim_{\eta \to 0} \frac{\partial \ln f_K(y_1, ..., y_K, \boldsymbol{\rho}, \eta, \mathbf{b})}{\partial \eta} &= \sqrt{\frac{K(K+2)}{2}} \times L_2(\varsigma) + \mathbf{b}'\mathbf{y}\left[\varsigma - (K+2)\right], \\
\lim_{\eta \to 0} \frac{\partial \ln f_{1k}[y_k; \eta, ß_k(\eta, \mathbf{b})]}{\partial \eta} &= \sqrt{\frac{3}{2}} H_4(y_k) + \mathbf{P}_{[k]}(\boldsymbol{\rho})\mathbf{b}H_3(y_k),
\end{aligned}$$

where $\mathbf{P}_{[k]}(\boldsymbol{\rho})$ denotes the k'th row of $\mathbf{P}(\boldsymbol{\rho})$. Regarding the third term of (A2), they show that

$$\lim_{\eta \to 0} \frac{\partial \ln f_{1k}[y_k; \eta, ß_k(\eta, \mathbf{b})]}{\partial ß_k} = 0.$$

As for $\partial F_{1k}^{-1}[u_k; \eta, ß_k(\eta, \mathbf{b})]/\partial \eta$, differentiating $\int_{-\infty}^{F_{1k}^{-1}(u_k, \eta, ß_k(\eta, \mathbf{b}))} f_{1k}[y_k; \eta, ß_k(\eta, \mathbf{b})]dy_k = u_k$ with respect to η yields

$$\frac{\partial F_{1k}^{-1}[u_k; \eta, ß_k(\eta, \mathbf{b})]}{\partial \eta} = \frac{-1}{f_{1k}(F_{1k}^{-1}[u_k; \eta, ß_k(\eta, \mathbf{b})]; \eta)} \int_{-\infty}^{F_{1k}^{-1}(u_k, \eta, ß_k(\eta, \mathbf{b}))} \frac{\partial f_{1k}[y_k; \eta, ß_k(\eta, \mathbf{b})]}{\partial \eta} dy_k.$$

Then, noticing that $\lim_{\eta \to 0} f_{1k}(F_{1k}^{-1}[u_k; \eta, ß_k(\eta, \mathbf{b})]; \eta) = \phi[\Phi^{-1}(u_k)]$ and that

$$\lim_{\eta \to 0} \frac{\partial f_{1k}[y_k; \eta, ß_k(\eta, \mathbf{b})]}{\partial \eta} = \phi(y_k) \left[\sqrt{\frac{3}{2}} H_4(y_k) + ß_k(\eta, \mathbf{b})\sqrt{6}H_3(y_k) \right],$$

we obtain

$$\lim_{\eta \to 0} \frac{\partial F_{1k}^{-1}[u_k; \eta, ß_k(\eta, \mathbf{b})]}{\partial \eta} = ß_k(\eta, \mathbf{b})\sqrt{2}H_2(y_k) + \frac{1}{2}\sqrt{\frac{3}{2}} H_3(y_k).$$

Collecting terms and using (A1) yields (3). □

A.3 Proposition 3

Under normality, the score with respect to $\mathbf{b}$ is 0, while the score with respect to η for fixed values of $\mathbf{b}$ is given in Proposition 2. Now consider a reparametrization in terms of $\eta^{\ddagger}$ and $\mathbf{b}^{\ddagger}$, where $\eta^{\ddagger} = \eta$ and $\mathbf{b}^{\ddagger} = \mathbf{b}\eta$. This reparametrization is such that under normality, both $\eta^{\ddagger}$ and $\mathbf{b}^{\ddagger}$ will be zero, while under local alternatives of the form $\eta_T^{\ddagger} = T^{-1/2}\bar{\eta}^{\ddagger}$ and $\mathbf{b}_T^{\ddagger} = T^{-1/2}\bar{\mathbf{b}}^{\ddagger}$, we will have an asymmetric Student t distribution with parameters $\eta_T = T^{-1/2}\bar{\eta}$ and $\mathbf{b}_T = \bar{\mathbf{b}}$. As for the score test, we start by defining

$$c^{\ddagger}(\mathbf{u};\boldsymbol{\rho},\eta^{\ddagger},\mathbf{b}^{\ddagger}) = c\left(\mathbf{u};\boldsymbol{\rho},\eta,\frac{\mathbf{b}^{\ddagger}}{\eta^{\ddagger}}\right).$$

We can then expand $\ln c\left(\mathbf{u};\boldsymbol{\rho},\eta,\mathbf{b}\right)$ around $\eta = 0$ as follows

$$\ln c\left(\mathbf{u};\boldsymbol{\rho},\eta,\mathbf{b}\right) = \ln c\left(\mathbf{u};\boldsymbol{\rho},0,\mathbf{b}\right) + s_{\eta}\left(\boldsymbol{\rho},\mathbf{b}\right)\eta + O(\eta^2),$$

and similarly, we can also expand $\ln c^{\ddagger}\left(\mathbf{u};\boldsymbol{\rho},\eta^{\ddagger},\mathbf{b}^{\ddagger}\right)$ as

$$\ln c^{\ddagger}(\mathbf{u};\boldsymbol{\rho},\eta^{\ddagger},\mathbf{b}^{\ddagger}) = \ln c^{\ddagger}\left(\mathbf{u};\boldsymbol{\rho},0,\mathbf{0}\right) + s_{\eta}^{\ddagger}\left(\boldsymbol{\rho},\mathbf{0}\right)\eta + \mathbf{s}_{\mathbf{b}^{\ddagger}}^{\ddagger}\left(\boldsymbol{\rho},\mathbf{0}\right)\mathbf{b}^{\ddagger} + O(\eta^2) + O(\mathbf{b}^{\ddagger\prime}\mathbf{b}^{\ddagger}) + O(\mathbf{b}^{\ddagger}\eta^{\ddagger}).$$

Since $\ln c\left(\mathbf{u};\boldsymbol{\rho},0,\mathbf{b}\right)$ does not depend on $\mathbf{b}$ and

$$s_{\eta}\left(\boldsymbol{\rho},\mathbf{b}\right)\eta = J_0\left(\boldsymbol{\rho}\right)\eta + \sum_{k=1}^{K} J_k\left(\boldsymbol{\rho}\right) b_i \eta$$

in light of Proposition 2, we can identify $J_0\left(\boldsymbol{\rho}\right)$ with $s_{\eta}^{\ddagger}\left(\boldsymbol{\rho},\mathbf{0}\right)$ and $J_k\left(\boldsymbol{\rho}\right)$ with $\mathbf{s}_{\mathbf{b}^{\ddagger}}^{\ddagger}\left(\boldsymbol{\rho},\mathbf{0}\right)$ for $k = 1,...,K$ because $\mathbf{b}^{\ddagger} = \mathbf{b}\eta$. □

A.4 Proposition 4

We can use the generalized information matrix equality (see e.g. Newey and McFadden (1994)) to show that

$$E\left\{\mathbf{s}_{\boldsymbol{\rho}}(\boldsymbol{\rho},\mathbf{0})\mathbf{s}_{\boldsymbol{\varphi}}'(\boldsymbol{\rho},\boldsymbol{\varphi})\middle|\,\boldsymbol{\rho},\boldsymbol{\varphi}\right\} = -E\left\{\left[\frac{\partial \mathbf{s}_{\boldsymbol{\rho}}(\boldsymbol{\rho},\mathbf{0})}{\partial \boldsymbol{\varphi}'}\right]\middle|\,\boldsymbol{\rho},\boldsymbol{\varphi}\right\} = \mathbf{0},$$

irrespective of the assumed copula, where we have used the fact that $\mathbf{s}_{\boldsymbol{\rho}}(\boldsymbol{\rho},\mathbf{0})$ does not vary with $\boldsymbol{\varphi}$ when regarded as the influence function for the Gaussian rank correlation. □

A.5 Proposition 5

Following Chen and Fan (2006a), to obtain the correction for non-parametric estimation of the marginals for a generic score s_{ϕ}, we need to compute

$$n_{\phi} = \sum_{j=1}^{K}\int_0^1 [1\{U_j \leq u_j\} - u_j] W_{\phi_k}^{j} du_j$$

with

$$W_\phi^j = \int ... \int \frac{\partial s_\phi}{\partial u_j} c(u_1, ..., u_K; \boldsymbol{\phi}) du_1 ... du_{k-1} du_{k+1} ... du_K.$$

To do so, we can exploit the fact that $\mathbf{z}_k = \mathbf{y}_{(k)} - \mathbf{p}_{(k)}(\boldsymbol{\rho}) y_k \sim N(\mathbf{0}, \boldsymbol{\Upsilon}_k)$ with $\boldsymbol{\Upsilon}_k = \mathbf{P}_{(kk)}(\boldsymbol{\rho}) - \mathbf{p}_{(k)}(\boldsymbol{\rho}) \mathbf{p}'_{(k)}(\boldsymbol{\rho})$, and that

$$\varsigma(\boldsymbol{\rho}) = y_k^2 + \mathbf{z}'_k [\mathbf{P}_{(kk)}(\boldsymbol{\rho}) - \mathbf{p}_{(k)}(\boldsymbol{\rho}) \mathbf{p}'_{(k)}(\boldsymbol{\rho})]^{-1} \mathbf{z}_k, \tag{A3}$$

$$y_k - \mathbf{p}'_{(k)}(\boldsymbol{\rho}) \mathbf{P}^{-1}_{(kk)}(\boldsymbol{\rho}) \mathbf{y}_{(k)} = [1 - \mathbf{p}'_{(k)}(\boldsymbol{\rho}) \mathbf{P}^{-1}_{(kk)}(\boldsymbol{\rho}) \mathbf{p}_{(k)}(\boldsymbol{\rho})] y_k + \mathbf{p}'_{(k)}(\boldsymbol{\rho}) \mathbf{P}^{-1}_{(kk)}(\boldsymbol{\rho}) \mathbf{z}_k. \tag{A4}$$

We can also make use of the fact that

$$c(\mathbf{u}_{(k)}; \boldsymbol{\phi}) d\mathbf{u}_{(k)} = c(u_1, ..., u_K; \boldsymbol{\phi}) du_1 ... du_{k-1} du_{k+1} ... du_K$$

involves integrating with respect to

$$f(\mathbf{z}_k; \boldsymbol{\phi}) = \frac{(2\pi)^{-(K-1)/2}}{|\boldsymbol{\Upsilon}_k|^{1/2}} \times \exp\left(-\frac{1}{2} \mathbf{z}'_k \boldsymbol{\Upsilon}_k^{-1} \mathbf{z}_k\right).$$

Specifically, for the first term of (2), using the fact that

$$\frac{\partial}{\partial \varsigma}\left[\sqrt{\frac{K(K+2)}{2}} L_2(\varsigma)\right] = \frac{\varsigma - (K+2)}{2} \quad \text{and} \quad \frac{\partial \varsigma(\boldsymbol{\rho})}{\partial y_k} = 2 \times \frac{y_k - \mathbf{p}'_{(k)}(\boldsymbol{\rho}) \mathbf{P}^{-1}_{(kk)}(\boldsymbol{\rho}) \mathbf{y}_{(k)}}{1 - \mathbf{p}'_{(k)}(\boldsymbol{\rho}) \mathbf{P}^{-1}_{(kk)}(\boldsymbol{\rho}) \mathbf{p}_{(k)}(\boldsymbol{\rho})},$$

and substituting (A3) and (A4), we obtain

$$\int \frac{\partial}{\partial y_k}\left[\sqrt{\frac{K(K+2)}{2}} L_2(\varsigma)\right] c(\mathbf{u}_{(k)}; \boldsymbol{\phi}) d\mathbf{u}_{(k)} = \sqrt{6} H_3(y_k), \tag{A5}$$

where the last equality follows from

$$\int \mathbf{z}_k \mathbf{z}'_k \boldsymbol{\Upsilon}_k^{-1} \mathbf{z}_k f(\mathbf{z}_k; \boldsymbol{\phi}) d\mathbf{z}_k = \mathbf{0} \quad \text{and} \quad \int \mathbf{z}'_k \boldsymbol{\Upsilon}_k^{-1} \mathbf{z}_k f(\mathbf{z}_k; \boldsymbol{\phi}) d\mathbf{z}_k = K - 1.$$

Similarly, for the second term of (2), using $\partial H_j(y)/\partial y = \sqrt{j} H_{j-1}(y)$ we notice that

$$\int \frac{\partial}{\partial y_k}\left[\sqrt{\frac{3}{2}} \sum_{h=1}^K H_4(y_h)\right] f(\mathbf{z}_k; \boldsymbol{\phi}) d\mathbf{z}_k = \sqrt{6} H_3(y_k),$$

which cancels with (A5). Regarding the final term of (2), given that $\int \mathbf{z}_k f(\mathbf{z}_k; \boldsymbol{\phi}) d\mathbf{z}_k = \mathbf{0}$ and

$$\int H_3(y_j) f(\mathbf{z}_k; \boldsymbol{\phi}) d\mathbf{z}_k = \int H_3(y_j) \frac{1}{\sqrt{1 - \rho_{jk}^2}} \phi\left(\frac{y_j - \rho_{jk} y_k}{\sqrt{1 - \rho_{jk}^2}}\right) f(\mathbf{z}_k; \boldsymbol{\phi}) dy_j = \rho_{jk}^3 H_3(y_k),$$

we can show that

$$\begin{aligned} &\int \frac{\partial}{\partial y_k}\left\{\sqrt{\frac{3}{8}} \sum_{h=1}^K \left[\frac{\mathbf{p}'_{(h)}(\boldsymbol{\rho}) \mathbf{P}^{-1}_{(h)}(\boldsymbol{\rho}) \mathbf{z}_h}{1 - \mathbf{p}'_{(h)}(\boldsymbol{\rho}) \mathbf{P}^{-1}_{(hh)}(\boldsymbol{\rho}) \mathbf{p}_{(h)}(\boldsymbol{\rho})}\right] H_3(y_h)\right\} f(\mathbf{z}_k; \boldsymbol{\phi}) d\mathbf{z}_k \\ = \; & -\sqrt{\frac{3}{8}} \frac{\mathbf{p}'_{(k)}(\boldsymbol{\rho}) \mathbf{P}^{-1}_{(kk)}(\boldsymbol{\rho}) \mathbf{p}_{(k)}(\boldsymbol{\rho})}{1 - \mathbf{p}'_{(k)}(\boldsymbol{\rho}) \mathbf{P}^{-1}_{(kk)}(\boldsymbol{\rho}) \mathbf{p}_{(k)}(\boldsymbol{\rho})} H_3(y_k) + \sqrt{\frac{3}{8}} \sum_{h \neq k} \frac{\mathbf{p}'_{(h)}(\boldsymbol{\rho}) \mathbf{P}^{-1}_{(h)}(\boldsymbol{\rho})}{1 - \mathbf{p}'_{(h)}(\boldsymbol{\rho}) \mathbf{P}^{-1}_{(hh)}(\boldsymbol{\rho}) \mathbf{p}_{(h)}(\boldsymbol{\rho})} \mathbf{P}^3_{kj}(\boldsymbol{\rho}) H_3(y_k). \end{aligned}$$

As for $m_{b_k}(\boldsymbol{\rho})$, we can use (A3) to rewrite its first term as

$$\frac{\partial\left\{y_k\left[\varsigma(\boldsymbol{\rho})-(K+2)\right]\right\}}{\partial y_k}=3y_k^2+\mathbf{z}_k'\boldsymbol{\Upsilon}_k^{-1}\mathbf{z}_k-(K+2)-2\times\frac{\mathbf{p}_{(k)}'(\boldsymbol{\rho})\mathbf{P}_{(kk)}^{-1}(\boldsymbol{\rho})\mathbf{z}_k}{1-\mathbf{p}_{(k)}'(\boldsymbol{\rho})\mathbf{P}_{(kk)}^{-1}(\boldsymbol{\rho})\mathbf{p}_{(k)}(\boldsymbol{\rho})},$$

so that

$$\int\frac{\partial\left\{y_k\left[\varsigma(\boldsymbol{\rho})-(K+2)\right]\right\}}{\partial y_k}f(\mathbf{z}_k;\boldsymbol{\phi})d\mathbf{z}_k=3(y_k^2-1)=3\sqrt{2}H_2(y_k),$$

where we have used the fact that $\int\mathbf{z}_k'\boldsymbol{\Upsilon}_k^{-1}\mathbf{z}_kf(\mathbf{z}_k;\boldsymbol{\phi})d\mathbf{z}_k=K-1$ and $\int\mathbf{z}_kf(\mathbf{z}_k;\boldsymbol{\phi})d\mathbf{z}_k=\mathbf{0}$, which again cancels with the correction corresponding to the second term because

$$\int\frac{\partial}{\partial y_k}\left[\sqrt{6}\sum_{j=1}^{K}\mathbf{P}_{kj}(\boldsymbol{\rho})H_3(y_j)\right]f(\mathbf{z}_k;\boldsymbol{\phi})d\mathbf{z}_k=3\sqrt{2}H_2(y_k)$$

in view of $\mathbf{P}_{kk}(\boldsymbol{\rho})=1$ and $\int f(\mathbf{z}_k;\boldsymbol{\phi})d\mathbf{z}_k=1$. Finally, to find

$$\frac{\partial}{\partial y_k}\left\{\sqrt{2}\sum_{j=1}^{K}\frac{\mathbf{p}_{(k)}'(\boldsymbol{\rho})\mathbf{P}_{(kk)}^{-1}(\boldsymbol{\rho})[\mathbf{y}_{(k)}-\mathbf{p}_{(k)}(\boldsymbol{\rho})y_k]}{1-\mathbf{p}_{(k)}'(\boldsymbol{\rho})\mathbf{P}_{(kk)}^{-1}(\boldsymbol{\rho})\mathbf{p}_{(k)}(\boldsymbol{\rho})}\mathbf{P}_{kj}(\boldsymbol{\rho})H_2(y_j)\right\},$$

we have to deal with the following two terms:

$$-\sqrt{2}\frac{\mathbf{p}_{(k)}'(\boldsymbol{\rho})\mathbf{P}_{(kk)}^{-1}(\boldsymbol{\rho})\mathbf{p}_{(k)}(\boldsymbol{\rho})}{1-\mathbf{p}_{(k)}'(\boldsymbol{\rho})\mathbf{P}_{(kk)}^{-1}(\boldsymbol{\rho})\mathbf{p}_{(k)}(\boldsymbol{\rho})}H_2(y_k)+\sqrt{2}\sum_{j\neq k}\frac{\mathbf{p}_{(j)}'(\boldsymbol{\rho})\mathbf{P}_{(jj)}^{-1}(\boldsymbol{\rho})}{1-\mathbf{p}_{(j)}'(\boldsymbol{\rho})\mathbf{P}_{(jj)}^{-1}(\boldsymbol{\rho})\mathbf{p}_{(j)}(\boldsymbol{\rho})}\mathbf{P}_{kj}(\boldsymbol{\rho})H_2(y_j)$$

and

$$2\frac{\mathbf{p}_{(k)}'(\boldsymbol{\rho})\mathbf{P}_{(kk)}^{-1}(\boldsymbol{\rho})\mathbf{z}_k}{1-\mathbf{p}_{(k)}'(\boldsymbol{\rho})\mathbf{P}_{(kk)}^{-1}(\boldsymbol{\rho})\mathbf{p}_{(k)}(\boldsymbol{\rho})}\mathbf{P}_{kk}(\boldsymbol{\rho})H_1(y_k).$$

The integral of the last term is zero since $\int\mathbf{z}_kf(\mathbf{z}_k;\boldsymbol{\phi})d\mathbf{z}_k=\mathbf{0}$. As for the first one, given

$$\int H_2(y_j)f(\mathbf{z}_k;\boldsymbol{\phi})d\mathbf{z}_k=\int H_2(y_j)\frac{1}{\sqrt{1-\rho_{jk}^2}}\phi\left(\frac{y_j-\rho_{jk}y_k}{\sqrt{1-\rho_{jk}^2}}\right)f(\mathbf{z}_k;\boldsymbol{\phi})dy_j=\rho_{jk}^2H_2(y_k),$$

we obtain that

$$\begin{aligned}&-\int\sqrt{2}\sum_{j=1}^{K}\frac{\mathbf{p}_{(j)}'(\boldsymbol{\rho})\mathbf{P}_{(jj)}^{-1}(\boldsymbol{\rho})\mathbf{p}_{(j)}(\boldsymbol{\rho})}{1-\mathbf{p}_{(j)}'(\boldsymbol{\rho})\mathbf{P}_{(jj)}^{-1}(\boldsymbol{\rho})\mathbf{p}_{(j)}(\boldsymbol{\rho})}\mathbf{P}_{kj}(\boldsymbol{\rho})H_2(y_j)f(\mathbf{z}_k;\boldsymbol{\phi})d\mathbf{z}_k\\ =\;&-\sqrt{2}\left[\frac{\mathbf{p}_{(k)}'(\boldsymbol{\rho})\mathbf{P}_{(kk)}^{-1}(\boldsymbol{\rho})\mathbf{p}_{(k)}(\boldsymbol{\rho})}{1-\mathbf{p}_{(k)}'(\boldsymbol{\rho})\mathbf{P}_{(kk)}^{-1}(\boldsymbol{\rho})\mathbf{p}_{(k)}(\boldsymbol{\rho})}+\sum_{j\neq k}\frac{\mathbf{p}_{(j)}'(\boldsymbol{\rho})\mathbf{P}_{(jj)}^{-1}(\boldsymbol{\rho})}{1-\mathbf{p}_{(j)}'(\boldsymbol{\rho})\mathbf{P}_{(jj)}^{-1}(\boldsymbol{\rho})\mathbf{p}_{(j)}(\boldsymbol{\rho})}\mathbf{P}_{kj}^2(\boldsymbol{\rho})\right]H_2(y_k).\end{aligned}$$

Analogous calculations allow us to obtain the relevant quantities for $\partial m_{b_k}(\boldsymbol{\rho})/\partial y_j$. Finally, the results stated in the proposition are obtained by collecting terms and integrating y_k out using the fact that

$$\int_{-\infty}^{y}H_3(x)\Phi(x)dx=\frac{H_3(y)}{4}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{y^2}{2}\right)+\frac{1}{4}H_4(y)\left[1+\operatorname{erfc}\left(\frac{y}{\sqrt{2}}\right)\right],$$

$$\int_y^{\infty} H_3(x)[1-\Phi(x)]dx = \frac{H_3(y)}{4}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{y^2}{2}\right) - \frac{1}{4}H_4(y)\,\mathrm{erf}\left(\frac{y}{\sqrt{2}}\right),$$

so that

$$\int_{-\infty}^{y} H_3(x)\Phi(x)dx - \int_y^{\infty} H_3(x)[1-\Phi(x)]dx = \frac{1}{2}H_4(y),$$

and

$$\int_{-\infty}^{y} H_2(x)\Phi(x)dx = \frac{H_2(y)}{3}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{y^2}{2}\right) + \frac{1}{2\sqrt{3}}H_3(y)\left[1+\mathrm{erfc}\left(\frac{y}{\sqrt{2}}\right)\right],$$

$$\int_y^{\infty} H_3(x)[1-\Phi(x)]dx = \frac{H_2(y)}{3}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{y^2}{2}\right) - \frac{1}{2\sqrt{3}}H_3(y)\,\mathrm{erf}\left(\frac{y}{\sqrt{2}}\right),$$

so that

$$\int_{-\infty}^{y} H_2(x)\Phi(x)dx - \int_y^{\infty} H_2(x)[1-\Phi(x)]dx = \frac{1}{\sqrt{3}}H_3(y).$$

□

A.6 Proposition 6

We start by simplifying the expressions for $\mathcal{A}$, $\mathcal{M}$, and $\mathcal{D}$. Regarding $\mathcal{A}$, the fact that

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

follows directly from Proposition 4. As for $\mathcal{M}$, Joe (2005) proves in the appendix that

$$cov\left[\mathbf{m}_{\boldsymbol{\lambda}_k}(\boldsymbol{\lambda}_0,\boldsymbol{\rho}_0), \mathbf{m}_{\boldsymbol{\rho}}(\boldsymbol{\lambda}_0,\boldsymbol{\rho}_0)\right] = \mathbf{0}, \text{ for } k = 1,...,K.$$

In regards to $\mathcal{D}$,

$$E\left(\frac{\partial^2 L_k(\boldsymbol{\lambda}_k)}{\partial \boldsymbol{\lambda}_k \partial \boldsymbol{\lambda}_h'}\right) = \mathbf{0}, \text{ whenever } h \neq k \text{ and } E\left(\frac{\partial^2 L_k(\boldsymbol{\lambda}_k)}{\partial \boldsymbol{\lambda}_k \partial \boldsymbol{\rho}'}\right) = \mathbf{0} \text{ for } k = 1,...,K$$

follows trivially from the fact that $L_k(\boldsymbol{\lambda}_k)$ does not depend on $\boldsymbol{\lambda}_h$ or $\boldsymbol{\rho}$. Next, to keep the notation simple, we group the correlation parameters $\boldsymbol{\rho}$ with those characterizing the marginals, $\boldsymbol{\lambda}$, into $\boldsymbol{\theta} = (\boldsymbol{\lambda}', \boldsymbol{\rho}')'$ and the associated score vectors as $\mathbf{m}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = [\mathbf{m}_{\boldsymbol{\lambda}}'(\boldsymbol{\lambda},\boldsymbol{\rho}), \mathbf{m}_{\boldsymbol{\rho}}'(\boldsymbol{\lambda},\boldsymbol{\rho})]'$. Then, it is easy to see that the first-order expansion

$$\mathbf{0} = \frac{1}{\sqrt{N}}\sum_{n=1}^{N}\mathbf{m}_{\boldsymbol{\theta}n}(\hat{\boldsymbol{\theta}}) = \frac{1}{\sqrt{N}}\sum_{n=1}^{N}\mathbf{m}_{\boldsymbol{\theta}n}(\boldsymbol{\theta}_0) + \frac{1}{N}\sum_{n=1}^{N}\frac{\partial \mathbf{m}_{\boldsymbol{\theta}n}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\sqrt{N}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) + o_p(1)$$

yields $\sqrt{N}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) \to N(\mathbf{0}, \mathcal{D}^{-1}\mathcal{M}\mathcal{D}'^{-1})$ (see e.g. Newey and McFadden, 1994). If we then exploit the fact that

$$cov[\mathbf{m}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0), \mathbf{m}_{\boldsymbol{\varphi}}(\boldsymbol{\theta}_0)] = \mathbf{0},$$

which is again proved in the appendix of Joe (2005), we can write

$$\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\mathbf{m}_{\boldsymbol{\varphi}n}(\hat{\boldsymbol{\theta}},\mathbf{0}) = \frac{1}{\sqrt{N}}\sum_{n=1}^{N}\mathbf{m}_{\boldsymbol{\varphi}n}(\boldsymbol{\theta}_0,\mathbf{0}) + \frac{1}{N}\sum_{n=1}^{N}\frac{\partial \mathbf{m}_{\boldsymbol{\varphi}n}(\boldsymbol{\theta}_0,\mathbf{0})}{\partial \boldsymbol{\theta}'}\sqrt{N}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) + o_p(1),$$

whence the result follows. □

References

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

Amengual, D. and E. Sentana (2017): “Gaussian rank correlation: a coefficient with attractive statistical properties”, mimeo, CEMFI.

Andrews, D.W.K. (2001): “Testing when a parameter is on the boundary of the maintained hypothesis”, *Econometrica*, 69, 683–734.

Asness, C.S., T.J. Moskowitz and L.H Pedersen (2013): “Value and momentum everywhere”, *Journal of Finance*, 68, 929–985.

Azzalini, A. and A. Capitanio (2003): “Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution”, *Journal of the Royal Statistical Society. Series B*, 65, 367–389.

Barndorff-Nielsen, O. (1977): “Exponentially decreasing distributions for the logarithm of particle size”, *Proceedings of the Royal Society*, 353, 401–419.

Berg, D. and J.-F. Quessy (2009): “Local power analyses of goodness-of-fit tests for copulas”, *Scandinavian Journal of Statistics*, 36, 389–412.

Blæsild, P. (1981): “The two-dimensional hyperbolic distribution and related distributions, with an application to Johannsen’s bean data”, *Biometrika*, 68, 251–263.

Calzolari, G., G. Fiorentini and E. Sentana (2004): “Constrained indirect estimation”, *Review of Economic Studies*, 71, 945–973.

Chen, X. and Y. Fan (2005): “Pseudo-likelihood ratio tests for semiparametric multivariate copula model selection”, *Canadian Journal of Statistics*, 33, 389–414.

Chen, X. and Y. Fan (2006a): “Estimation of copula-based semiparametric time series models”, *Journal of Econometrics,* 130, 307–335.

Chen, X. and Y. Fan (2006b): “Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification”, *Journal of Econometrics*, 135, 125–154.

Donnelly, C. and P. Embrechts (2010): “The devil is in the tails: actuarial mathematics and the subprime mortgage crisis”, *ASTIN Bulletin*, 40, 1–33.

Engle, R.F. (1982): "Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation", *Econometrica*, 50, 987–1007.

Fan, Y. and A.J. Patton (2014): "Copulas in econometrics", *Annual Review of Economics*, 6, 179–200.

Fermanian, J.-D. (2005): "Goodness-of-fit tests for copulas", *Journal of Multivariate Analysis*, 95, 119–152.

Fiorentini, G. and E. Sentana (2007): "On the efficiency and consistency of likelihood estimation in multivariate conditionally heteroskedastic dynamic regression models", CEMFI Working Paper 0713.

Fiorentini, G., E. Sentana and G. Calzolari (2003): "Maximum likelihood estimation and inference in multivariate conditionally heteroskedastic dynamic regression models with Student t innovations", *Journal of Business and Economic Statistics,* 21, 532–546.

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.

Godfrey, L.G. (1988): *Misspecification tests in econometrics: the Lagrange multiplier principle and other approaches*, Econometric Society Monographs.

Gouriéroux C., A. Holly and A. Monfort (1980): "Kuhn–Tucker, likelihood ratio and Wald tests for nonlinear models with inequality constraints on the parameters", Discussion Paper 770, Harvard Institute of Economic Research.

Horowitz, J. and N.E. Savin (2000): "Empirically relevant critical values for hypothesis tests: a bootstrap approach", *Journal of Econometrics*, 95, 375–389.

Joe, H. (2005): "Asymptotic efficiency of the two-stage estimation method for copula-based models", *Journal of Multivariate Analysis*, 94, 401–419.

Kotz, S. and S. Nadarajah (2004): *Multivariate t distributions and their applications*, Cambridge University Press, Cambridge.

Lee, L. F. and A. Chesher (1986): "Specification testing when score test statistics are identically zero", *Journal of Econometrics*, 31, 121–149.

MacKenzie, D. and T. Spears (2012): "The formula that killed Wall Street? The Gaussian copula and the material cultures of modelling", mimeo, School of Social & Political Science, University of Edinburgh.

Malevergne, Y. and D. Sornette (2003): "Testing the Gaussian copula hypothesis for financial assets dependencies", *Quantitative Finance*, 3, 231–250.

Mencía, J. and E. Sentana (2009): "Multivariate location-scale mixture of normals and mean-variance-skewness portfolio allocation", *Journal of Econometrics*, 153, 105–121.

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. (1985): "Maximum likelihood specification testing and conditional moment tests", *Econometrica*, 53, 1047–70.

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

Newey, W.K. and D.L. McFadden (1994): "Large sample estimation and hypothesis testing", in R.F. Engle and D.L. McFadden (eds.) *Handbook of Econometrics* vol. IV, 2111–2245, Elsevier.

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

Patton, A.J. (2006): "Modelling asymmetric exchange rate dependence", *International Economic Review*, 47, 527–56.

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

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

Salmon, F. (2009): "Recipe for disaster: the formula that killed Wall Street", *Wired*, 23 February, http://www.wired.com/print/techbiz/it/magazine/17-03/wp_quant

Scaillet, O. (2007): "Kernel based goodness-of-fit tests for copulas with fixed smoothing parameters", *Journal of Multivariate Analysis*, 98, 533–543.

Tauchen, G. (1985): "Diagnostic testing and evaluation of maximum likelihood models", *Journal of Econometrics*, 30, 415–443.

Table 1: 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 $\rho = 0.25$										
Known	LM-t	10.3	5.2	1.0	10.2	5.4	1.1	9.8	4.9	1.0
	LM-At	10.3	5.2	1.0	10.2	5.2	0.9	9.7	5.0	1.0
	Skew	10.0	4.9	1.2	9.8	4.9	1.0	10.2	4.7	1.1
	KT-t	10.4	5.1	0.9	10.2	5.3	1.2	9.7	4.9	0.9
	KT-At	10.1	5.1	0.9	10.1	5.2	0.9	9.6	4.9	1.0
Parametric	LM-t	10.3	5.2	0.9	10.2	5.5	1.1	9.8	4.9	1.0
	LM-At	10.2	5.0	0.9	10.4	5.2	0.9	9.8	4.9	1.0
	Skew	10.0	4.9	1.1	9.9	4.9	1.0	10.2	4.7	1.1
	KT-t	10.4	5.1	1.0	10.3	5.2	1.2	9.7	4.9	0.9
	KT-At	10.1	5.0	0.9	10.1	5.2	0.9	9.6	4.9	1.0
Emp. CDF	LM-t	10.2	4.9	0.9	10.3	5.3	1.2	9.7	4.8	1.0
	LM-At	10.0	5.0	0.8	10.0	4.8	1.0	9.6	4.9	0.9
	Skew	9.6	4.8	1.0	9.8	4.6	1.0	9.9	4.7	1.0
	KT-t	10.4	5.2	0.9	10.2	5.3	1.2	9.9	5.0	0.8
	KT-At	10.1	5.1	0.8	10.0	4.9	1.0	9.6	5.0	1.0
	$S^{(C)}$	9.9	4.7	0.9	9.9	5.2	1.1			
	$S^{(B)}$	9.4	4.5	0.9	9.9	5.2	1.1			
	Q	10.1	4.9	0.9	10.1	5.1	1.0			
	KS	10.7	5.2	1.3	10.2	5.2	1.0			
	CvM	9.0	4.4	0.7	8.9	4.4	0.9			
Panel B: $K = 10$ and $\rho_{kj} = 0.25$										
Known	LM-t	10.0	4.9	1.0	10.2	5.1	1.1	10.1	5.1	1.2
	LM-At	10.3	5.3	1.1	9.7	4.7	0.9	9.6	4.9	1.0
	Skew	10.2	5.2	1.2	9.8	4.9	0.9	9.6	4.9	1.0
	KT-t	10.1	5.0	1.1	9.5	5.0	1.0	10.3	4.9	1.1
	KT-At	10.4	5.3	1.1	9.7	4.5	0.8	9.7	4.9	1.0
Emp. CDF	LM-t	10.2	4.9	0.9	9.8	5.0	1.1	10.3	5.1	1.2
	LM-At	10.1	5.3	0.9	9.8	4.8	1.1	9.7	4.9	1.1
	Skew	10.3	5.0	0.9	10.1	4.8	1.0	9.7	4.7	1.0
	KT-t	9.9	4.8	0.9	9.4	4.8	1.2	10.4	5.2	1.1
	KT-At	10.1	5.4	0.9	9.7	4.8	1.0	9.7	4.7	1.1

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 2: Mote Carlo rejection rates at 1%, 5%, and 10% significance levels under the Student t alternative

		$N=200$			$N=800$			$N=3,200$		
Margins		10%	5%	1%	10%	5%	1%	10%	5%	1%
		Panel A: $K=2$ and $\rho=0.25$								
	LM-t	20.2	14.2	5.8	41.8	32.2	16.7	86.7	80.7	63.2
	LM-At	20.6	13.3	5.0	35.4	25.9	12.4	77.4	69.0	49.0
Known	Skew	15.1	8.5	2.6	15.0	8.5	2.5	15.0	8.3	2.2
	KT-t	29.2	18.0	6.2	53.5	39.8	19.0	92.3	86.5	67.6
	KT-At	22.6	14.2	5.0	39.2	28.1	12.7	81.0	72.3	51.0
	LM-t	20.2	13.9	5.6	41.6	32.1	16.5	86.7	80.8	63.2
	LM-At	20.6	13.3	4.8	35.7	25.8	12.3	77.5	69.1	48.8
Parametric	Skew	14.9	8.5	2.6	15.0	8.5	2.5	14.9	8.4	2.2
	KT-t	28.7	18.0	5.9	53.3	39.9	18.8	92.3	86.4	67.6
	KT-At	22.3	14.2	4.9	38.9	28.3	12.6	81.1	72.2	51.2
	LM-t	20.2	13.8	5.1	41.3	32.2	16.5	87.0	80.5	63.2
	LM-At	19.6	12.0	4.1	34.7	25.3	11.2	77.8	68.8	48.2
	Skew	13.9	7.7	2.1	14.5	7.8	2.1	14.6	8.1	2.1
	KT-t	27.0	16.8	5.3	52.6	38.8	18.4	92.3	86.2	67.3
	KT-At	20.9	12.5	4.2	38.0	27.3	11.5	81.0	71.9	50.7
Emp. CDF										
	$S^{(C)}$	10.77	5.4	1.1	12.4	6.5	1.4			
	$S^{(B)}$	10.66	5.3	1.1	12.0	6.3	1.3			
	Q	10.62	5.2	1.2	10.7	5.5	1.1			
	KS	11.86	6.0	1.4	12.4	6.7	1.4			
	CvM	9.12	4.5	0.7	11.1	5.4	1.1			
		Panel B: $K=10$ and $\rho_{kj}=0.25$								
	LM-t	26.7	18.5	7.6	58.4	47.3	26.9	97.7	95.4	87.3
	LM-At	22.0	14.1	4.9	37.6	26.8	11.8	83.0	74.6	54.3
Known	Skew	16.2	9.3	2.3	17.2	9.8	2.5	16.7	8.9	2.2
	KT-t	37.1	24.8	9.3	71.0	57.2	31.3	99.0	97.7	91.4
	KT-At	23.1	14.7	5.0	39.6	28.5	12.2	84.7	76.5	55.8
	LM-t	28.6	19.2	7.3	60.9	49.0	27.4	97.8	95.7	87.9
	LM-At	21.1	12.7	4.0	37.7	26.6	11.2	83.2	75.0	54.5
Emp. CDF	Skew	15.3	8.3	1.8	16.9	9.3	2.3	16.4	8.9	2.4
	KT-t	33.5	22.0	7.5	68.7	55.3	30.3	98.8	97.5	90.7
	KT-At	21.3	12.9	4.0	39.1	27.6	11.5	84.6	76.4	55.8

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

Margins		$N=200$			$N=800$			$N=3,200$		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A: $K=2$ and $\rho=0.25$										
	LM-t	20.6	14.1	5.9	43.5	33.8	17.9	88.1	82.2	65.6
	LM-At	28.3	19.9	7.7	66.2	55.5	32.7	99.5	99.0	96.1
Known	Skew	26.6	17.4	6.3	59.4	47.0	25.6	98.4	96.5	89.6
	KT-t	28.4	18.4	6.2	56.4	42.0	20.1	93.5	87.8	70.0
	KT-At	30.9	20.7	7.8	70.1	58.3	33.5	99.7	99.3	96.6
	LM-t	20.4	14.2	5.9	43.5	33.9	17.7	88.1	82.1	65.6
	LM-At	28.3	19.6	7.6	66.3	55.7	32.8	99.5	99.0	96.2
Parametric	Skew	26.4	17.2	6.2	59.5	47.1	25.5	98.4	96.6	89.7
	KT-t	28.4	18.3	6.1	56.3	41.9	20.0	93.2	87.7	69.9
	KT-At	30.6	20.4	7.7	69.8	58.4	33.5	99.7	99.3	96.6
	LM-t	20.2	13.2	4.5	43.2	33.2	16.0	88.5	81.9	65.1
	LM-At	27.7	17.7	5.3	66.1	54.5	31.6	99.5	99.0	95.8
	Skew	24.5	14.7	4.3	57.6	44.8	23.4	98.0	96.3	88.6
	KT-t	26.8	16.7	4.7	54.6	40.6	18.3	93.3	87.8	69.6
	KT-At	29.1	18.3	5.3	69.1	57.4	32.1	99.7	99.2	96.4
Emp. CDF										
	$S^{(C)}$	11.0	5.5	1.0	23.4	13.9	3.9			
	$S^{(B)}$	11.6	5.7	1.2	24.0	14.5	4.1			
	Q	12.0	6.4	1.2	19.9	11.2	2.4			
	KS	13.6	7.1	1.7	32.6	22.1	8.0			
	CvM	10.0	5.1	1.0	13.8	7.3	1.9			
Panel B: $K=10$ and $\rho_{kj}=0.25$										
	LM-t	26.6	18.4	7.3	58.2	46.7	26.8	97.2	95.1	86.3
	LM-At	22.5	14.5	5.4	40.9	30.0	13.1	88.6	82.3	64.0
Known	Skew	17.1	9.9	2.9	21.1	12.6	3.5	35.9	23.5	8.7
	KT-t	36.4	25.1	9.0	70.0	57.0	31.1	99.0	97.3	90.5
	KT-At	23.5	14.8	5.6	43.0	31.5	13.6	89.9	83.7	65.3
	LM-t	28.3	19.6	7.1	59.6	48.5	27.0	97.6	95.4	86.6
	LM-At	21.7	13.8	4.3	40.4	29.2	12.5	88.1	81.7	63.1
Emp. CDF	Skew	15.9	9.2	2.4	20.0	11.4	3.5	33.5	22.2	8.0
	KT-t	33.3	22.1	7.4	68.1	54.7	29.7	98.9	97.3	89.8
	KT-At	22.1	14.1	4.3	41.8	30.2	12.9	89.3	82.7	64.5

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 4: 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 $\rho = 0.25$										
Known	LM-t	19.9	13.9	5.7	41.9	32.4	16.7	86.0	79.6	62.0
	LM-At	20.0	13.1	4.8	37.2	27.5	13.1	79.7	71.4	52.3
	Skew	14.7	8.1	2.1	17.1	9.8	2.9	23.6	14.3	4.6
	KT-t	27.3	17.6	6.0	54.2	40.1	19.0	92.1	85.6	66.8
	KT-At	21.9	13.9	4.9	40.8	30.0	13.6	82.7	74.5	54.3
Parametric	LM-t	19.7	13.7	5.9	41.6	32.4	16.6	85.8	79.6	62.1
	LM-At	20.0	13.1	4.7	37.2	27.5	13.0	79.7	71.3	52.1
	Skew	14.7	8.4	2.3	17.0	10.0	2.9	23.5	14.5	4.6
	KT-t	27.4	17.5	6.1	54.3	39.9	18.9	92.0	85.4	66.8
	KT-At	21.7	13.8	4.8	40.7	30.1	13.5	82.8	74.5	54.6
Emp. CDF	LM-t	19.0	12.7	5.0	42.0	32.1	16.3	86.5	79.9	62.2
	LM-At	18.8	11.8	3.9	36.2	26.5	11.8	79.6	71.4	51.5
	Skew	13.4	7.4	1.7	16.4	9.2	2.5	23.0	13.8	4.2
	KT-t	25.5	15.7	5.2	53.5	39.6	18.4	92.3	85.8	66.6
	KT-At	20.0	12.2	3.9	39.6	28.3	12.2	82.5	74.2	53.7
	$S^{(C)}$	10.2	4.9	0.9	12.7	6.6	1.6			
	$S^{(B)}$	9.7	4.9	0.8	12.7	6.6	1.5			
	Q	10.8	5.6	1.1	11.3	6.2	1.4			
	KS	11.5	5.6	1.1	13.3	7.2	1.4			
	CvM	9.1	4.5	0.8	10.8	5.4	1.0			
Panel B: $K = 10$ and $\rho_{kj} = 0.25$										
Known	LM-t	25.6	17.6	6.9	58.7	47.5	26.2	97.9	95.9	87.7
	LM-At	22.0	14.3	4.8	37.5	26.6	10.9	83.4	74.9	54.8
	Skew	16.4	9.3	2.6	16.7	9.6	2.5	17.2	9.7	2.6
	KT-t	35.7	24.2	8.3	71.2	57.7	31.8	99.1	97.9	91.6
	KT-At	23.1	14.6	4.8	39.5	28.0	11.3	85.2	76.9	57.3
Emp. CDF	LM-t	19.5	11.7	3.9	37.0	25.9	10.7	83.5	75.2	54.7
	LM-At	28.4	18.4	5.8	60.7	48.3	26.9	98.0	96.3	87.8
	Skew	14.4	7.8	2.0	16.6	9.8	2.5	16.8	9.7	2.5
	KT-t	32.5	21.1	6.1	69.2	55.7	30.6	99.2	97.9	91.1
	KT-At	19.8	11.9	3.9	38.5	27.1	10.9	84.9	76.8	56.5

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 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 5: Momentum and reversals in stock returns

Panel A: Correlation parameter estimates

Strategy	Beta OLS	Correlation parameter: Pearson	Spearman	Copula
Short-term reversals	.008	.009	-.025	-.022
	(.003)	(.003)	(.001)	(.002)
Momentum	2.94×10^{-4}	.002	.037	.035
	(4.21×10^{-4})	(.003)	(.001)	(.002)

Panel B: Test statistics and p-values

Strategy	KT-t	Skew	KT-At
Short-term reversals	24,333.7	1,086.0	25,419.7
	(.000)	(.000)	(.000)
Momentum	32,408.0	4,258.7	36,666.7
	(.000)	(.000)	(.000)

Panel C: Constrained indirect estimates of the shape parameters

Strategy	Student t: $\hat{\rho}$	$\hat{\eta}$
Short-term reversals	-.025	.187
Momentum	.034	.213

Strategy	Asymmetric Student t: $\hat{\rho}$	$\hat{\eta}$	$\hat{b}_1$	$\hat{b}_2$
Short-term reversals	-.018	.187	-.112	-.069
Momentum	.074	.212	-.124	-.190

Notes: The data is collected from CRSP and contains monthly series from January 1997 to December 2012. Total number of observations is 607,054. Panel A: Beta OLS denotes the slope coefficient in a simple linear regression. Pearson and Spearman denote the Pearson linear correlation coefficient and Spearman 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. Panel B: 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 $m_{b_k}(\rho)$ for $k = 1, 2$ of Proposition 3. Panel C: 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 1a: Bivariate Gaussian copula with Gaussian margins

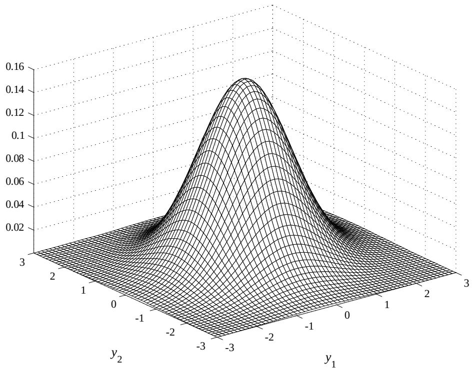


Figure 1b: Contours of a bivariate Gaussian copula with Gaussian margins

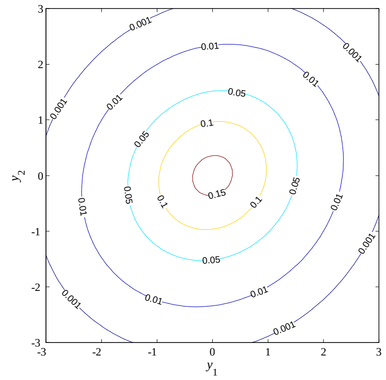


Figure 1c: Bivariate Student t copula density with Gaussian margins

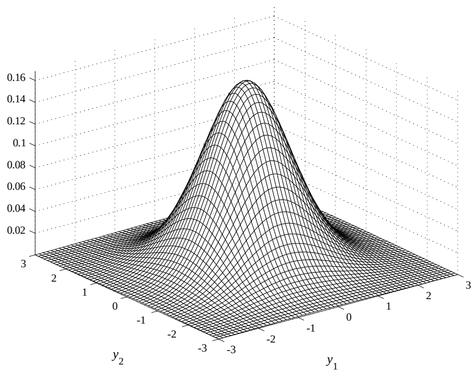


Figure 1d: Contours of a bivariate Student t copula with Gaussian margins

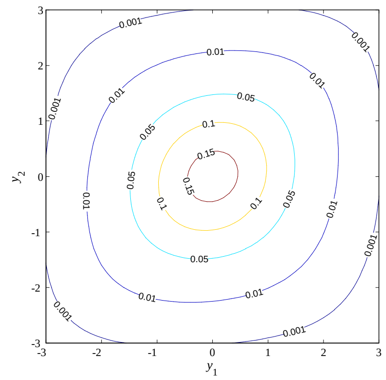


Figure 1e: Bivariate asymmetric Student t copula density with Gaussian margins

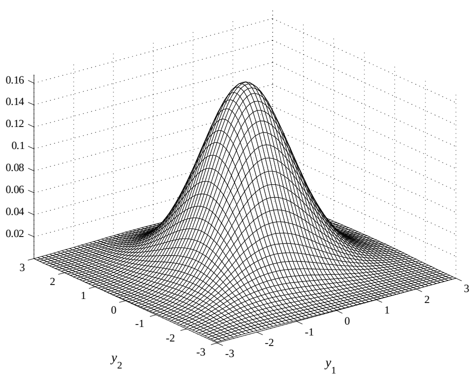
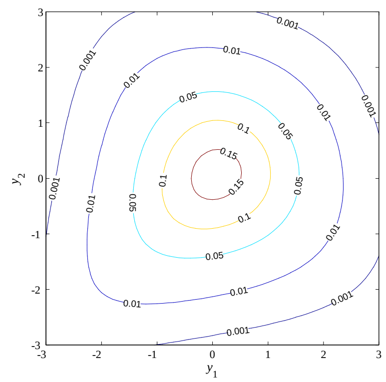


Figure 1f: Contours of a bivariate asymmetric Student t copula with Gaussian margins



Notes: Panels a-b: Gaussian copula with correlation coefficient $\rho = .120$ (Spearman correlation $\rho_S = .115$). Panels c-d: Student t copula with 10 degrees of freedom and correlation coefficient $\rho = .122$ (Spearman correlation $\rho_S = .115$). Panels e-f: Asymmetric Student t copula with 10 degrees of freedom, skewness parameters b_i =-.5 and correlation coefficient $\rho = .186$ (Spearman correlation $\rho_S = .115$). In all panels the marginals are standard normal.

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

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

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

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

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,

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

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\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}$$

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

and

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

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

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

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

$$d_{\varphi}^{np}(\rho) = \beta_{\varphi}^{np} s_{\varphi}^{np}(\rho) + u_{\varphi}^{np},$$

where

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

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

$$\frac{\beta_{\varphi}^2 cov\left[s_{\varphi}^{np}(\rho), s_{\varphi}(\rho)\right]}{\beta_{\varphi}^2 cov\left[s_{\varphi}^{np}(\rho), s_{\varphi}(\rho)\right] + V(u_{\varphi}^{np})}$$

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

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

and

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

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 $V[d_{\eta}^{np}(\rho)] = 1 + 3\rho^4$.

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 $Y_t = \{(y_{11}^t, y_{21}^t), ..., (y_{1n}^t, y_{2n}^t), ..., (y_{1N_t}^t, y_{2N_t}^t)\}$. 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:

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

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

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

$$\mathbf{m}(X_t) = \frac{1}{N_t}\sum_{n=1}^{N_t}\left[x_{1n}^t,\ x_{2n}^t,\ (x_{1n}^t)^2,\ (x_{2n}^t)^2,\ x_{1n}^t x_{2n}^t\right]'.$$

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

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

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	$-\frac{2\left(\rho^3+\rho\right)}{1-\rho^2}$	0	$-\frac{2\rho}{1-\rho^2}$
$H_2(y_1)$	$\frac{3\rho^2\left(\rho^2+3\right)}{2\sqrt{2}(1-\rho^2)^2}$	0	$\frac{3\sqrt{2}\rho^2}{(1-\rho^2)^2}$	0
$H_1(y_1)H_1(y_2)$	$-\frac{2\left(2\rho^3+\rho\right)}{(1-\rho^2)^2}$	0	$-\frac{2\left(2\rho^3+\rho\right)}{(1-\rho^2)^2}$	0
$H_2(y_2)$	$\frac{3\rho^2\left(\rho^2+3\right)}{2\sqrt{2}(1-\rho^2)^2}$	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	$\frac{\sqrt{2}\left(\rho^2+1\right)}{1-\rho^2}$	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
$H_3(y_1)H_1(y_2)$	$-\frac{\sqrt{\frac{3}{2}}\rho\left(\rho^2+3\right)}{2(1-\rho^2)^2}$	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
$H_1(y_1)H_3(y_2)$	$-\frac{\sqrt{\frac{3}{2}}\rho\left(\rho^2+3\right)}{2(1-\rho^2)^2}$	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

	Copula LM test		Distributional LM test	
Hermite polynomial	$s_{\eta}^{np}(\rho)$	$m_{b_1}^{np}(\rho)$	$d_{\eta}^{np}(\rho)$	$d_{b_1}^{np}(\rho)$
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 $\rho = 0.75$										
	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
Known	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
	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
Parametric	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
	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
Emp. CDF										
	$S^{(C)}$	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 $\rho_{kj} = 0.75$										
	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
Known	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
	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
Emp. CDF	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

		$N=200$			$N=800$			$N=3,200$		
Margins		10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A: $K=2$ and $\rho=0.75$										
	LM-t	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

		Short-term reversal strategies				Momentum strategies			
			Correlation parameter				Correlation parameter		
	N.Obs	Beta OLS	Pearson	Spearman	Copula	Beta OLS	Pearson	Spearman	Copula
Sector									
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		
	N.Obs	KT-t	Skew	KT-At	KT-t	Skew	KT-At
Sector							
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

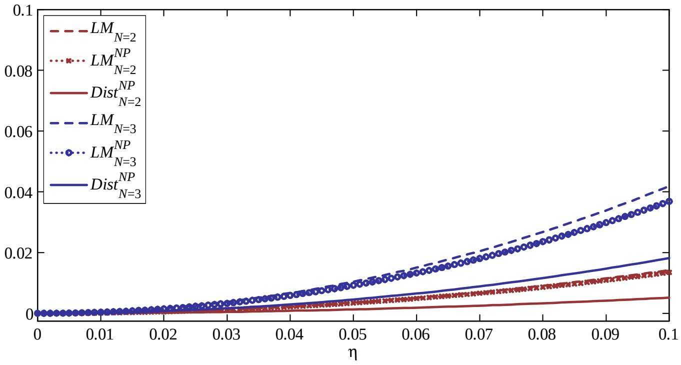


Figure D1b: Non-centrality parameter for different correlation parameter values

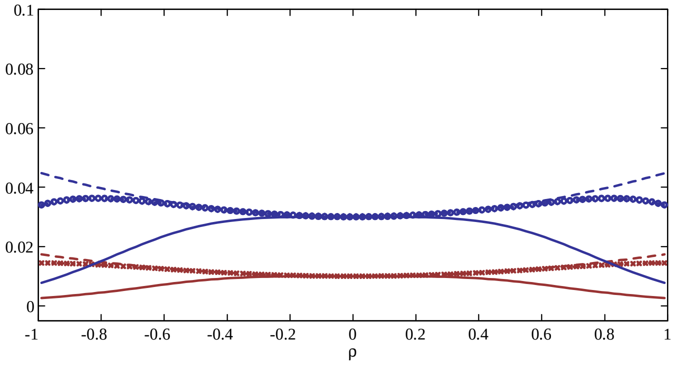
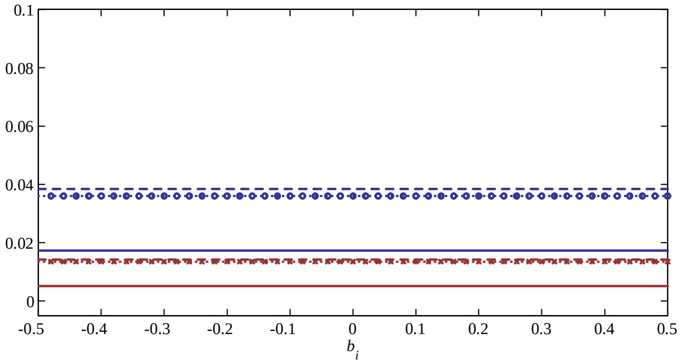


Figure D1c: Non-centrality parameter for different skewness parameter values



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 ρ = .75 and b_k =-.5 for $k = 1, 2$. Figure D1b, power against asymmetric Student t alternatives with different correlation parameter and η = .1, b_k =-.5 for $k = 1, 2$. Figure D1c, power against asymmetric Student t alternatives with increasing skewness and η = .1, ρ = .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

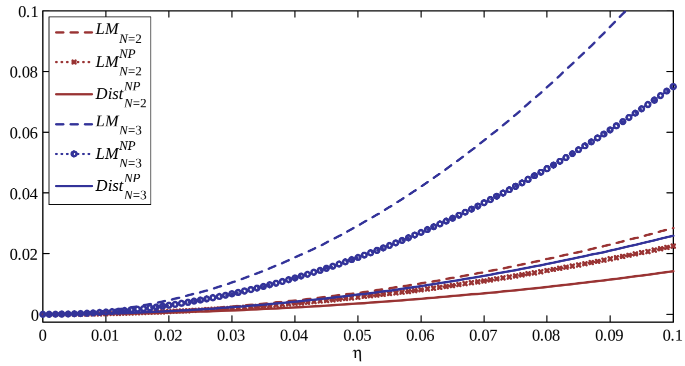


Figure D2b: Non-centrality parameter for different correlation parameter values

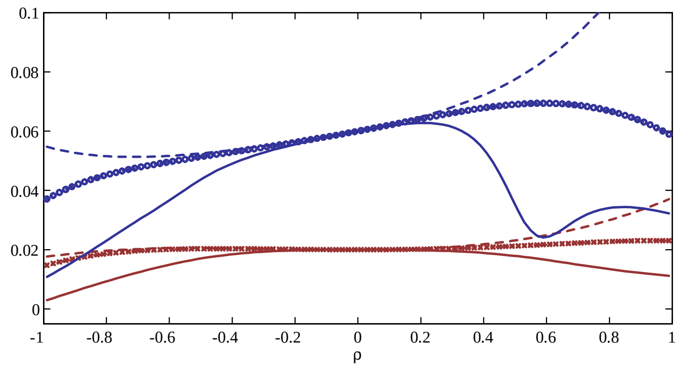
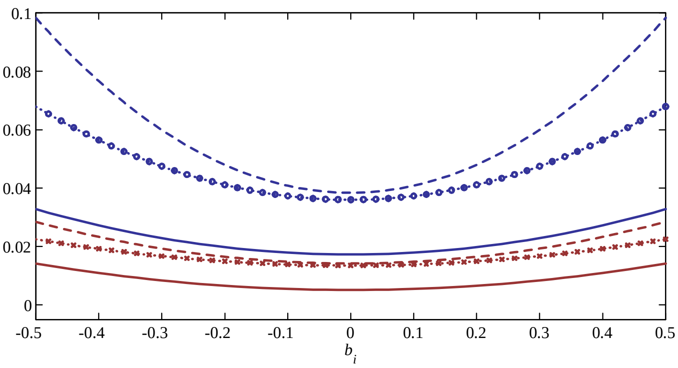


Figure D2c: Non-centrality parameter for different skewness parameter values



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%

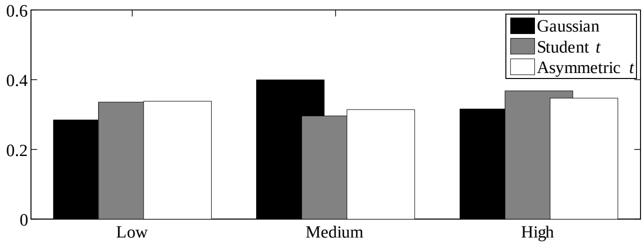


Figure G1b: Bottom-Middle 25%

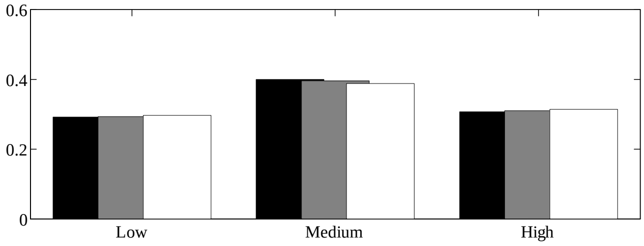


Figure G1c: Middle 40%

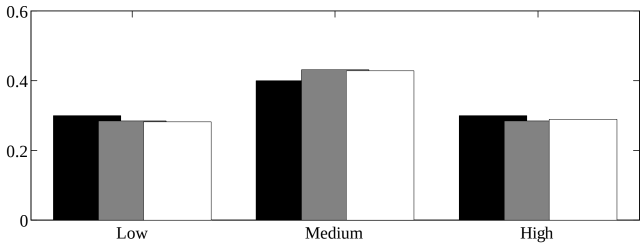


Figure G1d: Middle-Top 25%

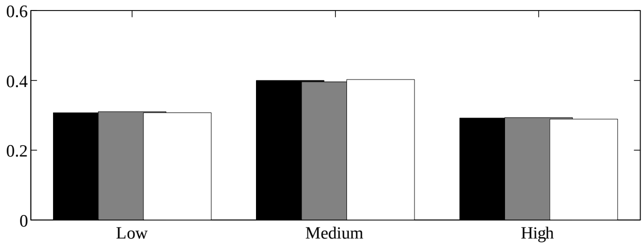
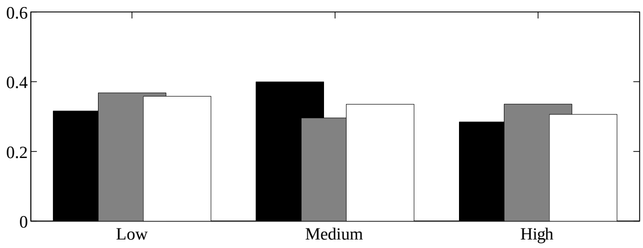


Figure G1e: Top 5%



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%

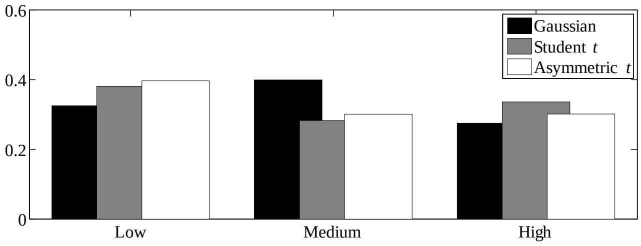


Figure G2b: Bottom-Middle 25%

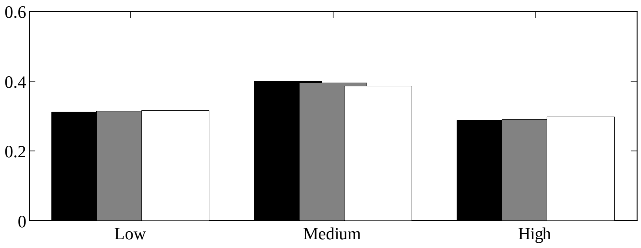


Figure G2c: Middle 40%

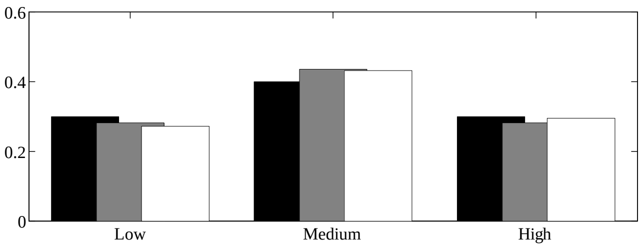


Figure G2d: Middle-Top 25%

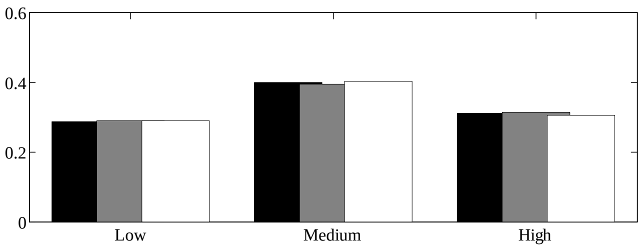
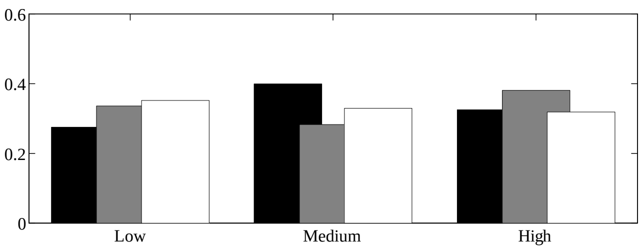


Figure G2e: Top 5%



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.