Gaussian rank correlation and regression*

Dante Amengual
CEMFI
<amengual@cemfi.es>

Enrique Sentana
CEMFI
<sentana@cemfi.es>

Zhanyuan Tian
Boston University
<zhanyuan@bu.edu>

September 2018

Revised: July 2020

Abstract

We study the statistical properties of Pearson correlation coefficients of Gaussian ranks, and Gaussian rank regressions – OLS applied to those ranks. We show that these procedures are fully efficient when the true copula is Gaussian and the margins are non-parametrically estimated, and remain consistent for their population analogues otherwise. We compare them to Spearman and Pearson correlations and their regression counterparts theoretically and in extensive Monte Carlo simulations. Empirical applications to migration and growth across US states, the augmented Solow growth model, and momentum and reversal effects in individual stock returns confirm that Gaussian rank procedures are insensitive to outliers.

Keywords: Copula, Growth Regressions, Migration, Misspecification, Momentum, Robustness, Short-term reversals.

JEL: C13, C46, G14, O47

*Some of the material in this paper originally appeared in sections 2.3, 2.5 and 2.7 of Amengual and Sentana (2015). We are grateful to Xinyue Bei and participants at Encuentro de la SEU (Montevideo, 2018), 8^{th} ICEEE conference (Lecce, 2019), AMES (Xiamen, 2019), CMES (Guangzhou, 2019), IAAE (Nicosia, 2019), ESEM (Manchester, 2019) and SAEe (Alicante, 2019) for their helpful comments, discussions and suggestions. An anonymous referee has also helped us greatly improve the paper. Of course, the usual caveat applies. Financial support from the Spanish Ministry of Economy, Industry & Competitiveness through grant ECO 2017-89689 and the Santander CEMFI Research Chair in Banking and Finance is gratefully acknowledged.

1 Introduction

As is well known, short term reversal strategies regularly invest in those stocks that have underperformed in the past month, while momentum strategies typically invest in stocks that outperformed over the previous months of the past year. However, most of the literature has focused on assessing the excess returns obtained by various trading strategies rather than on precisely analyzing the cross-sectional dependence between stock returns this month and those characteristics at the time investment decisions are made. Such an analysis is important not only because it can potentially lead to better decisions, but also because it can shed some light on the sources of the dependence.

There are several different ways of characterizing dependence. The most straightforward one is by means of linear relationships, as it is often done in the extensive growth convergence literature in macroeconomics, which we will revisit in section 7. Specifically, a researcher could cross-sectionally regress individual stock returns this month on a constant and returns over previous months, and look at the size and significance of the Pearson correlation coefficient. However, a few high-leverage observations can unduly affect the value and sign of the estimated coefficients. As a case in point, Figures 1a (reversals) and 1b (momentum) contain the results of cross-sectionally regressing individual stock returns in the CRSP database in August 2007 on a constant and returns over previous months. The problem with this linear approach is that the OLS coefficient estimates may be extremely sensitive to a few outliers, as illustrated in Figure 1b, in which the slightly negative slope is largely driven by the southeasternmost stocks. In fact, if we trim the sample of 2,463 observations by simply excluding those five stocks whose cumulative return over the period September 2006 - June 2007 exceeded 300%, we obtain a positive correlation.

A procedure which is far less sensitive to outliers involves rank regressions, whereby one regresses the cross-sectional rank of stocks this month on a constant and their cross-sectional rank over the relevant period in the past. Figures 1c (reversals) and 1d (momentum) contain the scatterplots of the corresponding normalized ranks for the same month, the associated regression lines and Spearman correlation coefficients. This procedure is closely related to the concept of "copula", which allows us to separate joint distributions from marginal ones by fixing the latter. In the case of rank regressions, in particular, the empirical marginal distributions are discrete uniform by construction.

But this is not the only possibility. A closely related approach is to look at the dependence between the so-called Gaussian ranks, which are simple monotonic transformations of the usual ranks obtained by applying the standard normal quantile function. In fact, one may convincingly argue that scatterplots of Gaussian ranks are easier to interpret than scatterplots of uniform ranks, if only because empirical researchers are more used to analyzing real data with approximately Gaussian marginals than uniform ones (see Joe (2015) for a more formal justification). Figures 1e (reversals) and 1f (momentum) show the scatterplots of the Gaussian ranks, the corresponding regression lines, and the Gaussian rank correlation coefficients. As can be seen, both the Spearman and Gaussian rank correlation coefficients confirm the presence of momentum and short term reversals in individual stock returns.[1]

Boudt et al (2012) study the numerical sensitivity of different correlation coefficients with respect to observations with unusually large magnitudes. While those results are very useful, the purpose of our paper is to study the usual statistical properties –namely consistency and asymptotic efficiency– of Gaussian rank correlations, which are the Pearson correlation coefficients of the Gaussian ranks. We also consider Gaussian rank regressions, which coincide with OLS applied to those ranks. We show that these procedures are as efficient as maximum likelihood when the true copula is Gaussian and the margins are non-parametrically estimated, and remain consistent for their population analogues otherwise, thereby inheriting the properties of the Gaussian pseudo maximum likelihood estimators of first and second moments. Therefore, Gaussian rank correlations and regressions are robust in both the statistical and econometric senses of the word: they are not too sensitive to outliers and they remain consistent under misspecification of the copula.

We also compare these estimators to Spearman and Pearson correlations based on the original data. In addition, we compare the regression counterparts to the Gaussian rank correlations with both standard OLS and some of its robust versions, specifically the least trimmed squares and least median of squares estimators proposed by Rousseeuw (1984, 1985).

Finally, we apply the aforementioned procedures to study two important empirical issues: (i) the relationship between migration and growth rates across US states over the twentieth century using the data set in chapter 11 of Barro and Xala-i-Martin (2003); and (ii) the augmented Solow growth model in Mankiw, Romer and Weil (1992) (MRW, henceforth), which Temple (1998) re-assessed using alternative robust regression techniques. Our results confirm that Gaussian rank procedures are insensitive to outliers, unlike Pearson correlations and OLS regressions. Thus, they are indeed doubly robust.

The rest of the paper is organised as follows. In Section 2, we introduce Gaussian copulas, and derive the first and second derivatives of the associated log-likelihood function. Then, in Section 3 we obtain the asymptotic variance of the maximum likelihood estimators and compare them to some closely related

[1]In Amengual, Sentana and Tian (2020) we also study the combined effect of short term reversals and momentum by running a multiple regression of individual stock returns $r_{t,t-1}$ on a constant, $r_{t-1,t-2}$, and $r_{t-2,t-12}$. Given the very low dependence between the two regressors, the multiple regression coefficients are very close to the pairwise correlations, which in turn implies that the conclusions derived from Figure 1 are by and large preserved.

moment estimators, both when the marginal distributions are known, and when they are replaced by their (re-scaled) empirical cumulative distribution function (cdf) counterparts. In Section 4, we extend those results to realistic situations in which the true copula is not Gaussian. Next, Section 5 compares the theoretical properties of Gaussian rank correlations and regression to those of the well known Pearson and Spearman counterparts, while Section 6 looks at their behaviour in finite samples by means of an extensive Monte Carlo exercise. The results of our empirical applications can be found in Section 7, followed by our conclusions and directions for futher research. Finally, some practical considerations of interest for practitioners are discussed in the appendix, while proofs and auxiliary results are relegated to the Online Appendix.

2 Theoretical background

2.1 Econometric model

Let $\mathbf{x}$ denote a vector of K continuous random variables. The traditional way of modelling the dependence between the elements of $\mathbf{x}$ is through the joint cdf $F_K(\mathbf{x})$ 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 $x_{k-1}, x_{k-2}, \ldots, x_1$ $(k = 2, \ldots, K)$ times the marginal distribution of x_1.

In contrast, the standard copula approach first transforms each of the elements of $\mathbf{x}$ into a uniform random variable by means of the probability integral transform $u_k = G_k(x_k)$, where $G_k(.)$ is the marginal cdf of x_k, which we assume known until Section 3.2, and then models the dependence of the random vector $\mathbf{u} = (u_1, \ldots, u_K)'$ through a joint distribution function $C_K(\mathbf{u})$ 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 it name suggests, it is the copula function that corresponds to the multivariate Gaussian distribution, which remains dominant in multivariate statistical analysis.

More formally, define $\mathbf{y} = (y_1, \ldots, y_K)'$, where $y_k = \Phi^{-1}(u_k)$, $\Phi(.)$ denotes the univariate standard normal cumulative distribution function and $\Phi^{-1}(.)$ the corresponding quantile function. The Gaussian copula with correlation matrix $\mathbf{P}(\boldsymbol{\rho})$ is derived from the cdf 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 potentially restricted positive definite correlation matrix which contains* $K(K-1)/2$ *twice continuously differentiable functions of the* $p \leq \frac{1}{2}K(K-1)$ *free parameters in* $\boldsymbol{\rho}$*, such that* $\mathbf{P}(\mathbf{0}) = \mathbf{I}_K$*.*

In the unrestricted case in which $\boldsymbol{\rho} = vecl[\mathbf{P}(\boldsymbol{\rho})]$, where $vecl(\mathbf{P})$ is the vec-type operator which stacks by columns the elements in the strict lower triangle of the matrix $\mathbf{P}$, $\mathbf{P}(\boldsymbol{\rho})$ is trivially twice differentiable. The same is true for many popular restricted parametrizations, such as an equicorrelated one-factor structure. In turn, the requirement that $\boldsymbol{\rho} = \mathbf{0}$ yields the independent copula is just a convenient normalization.

Under Assumption 1, the Gaussian copula density function will be given by

$$c_K(\mathbf{u}; \boldsymbol{\rho}) = |\mathbf{P}(\boldsymbol{\rho})|^{-1/2} \exp\left\{-\frac{1}{2}\mathbf{y}'[\mathbf{P}^{-1}(\boldsymbol{\rho}) - \mathbf{I}_K]\mathbf{y}\right\} = |\mathbf{P}(\boldsymbol{\rho})|^{-1/2} \exp\left\{-\frac{1}{2}[\varsigma(\boldsymbol{\rho}) - \varsigma(\mathbf{0})]\right\}, \quad (1)$$

where $\varsigma(\boldsymbol{\rho}) = \mathbf{y}'\mathbf{P}^{-1}(\boldsymbol{\rho})\mathbf{y}$ and $\varsigma(\mathbf{0}) = \mathbf{y}'\mathbf{y}$. Figures 2a-b display a bivariate Gaussian copula density with Gaussian rank correlation .25 and Gaussian margins.

We can directly use the Gaussian ranks to write the likelihood function as

$$\phi_K(\mathbf{y}; \boldsymbol{\rho}) = (2\pi)^{-K/2}|\mathbf{P}(\boldsymbol{\rho})|^{-1/2} \exp\left[-\frac{1}{2}\mathbf{y}'\mathbf{P}^{-1}(\boldsymbol{\rho})\mathbf{y}\right] = (2\pi)^{-K/2}|\mathbf{P}(\boldsymbol{\rho})|^{-1/2} \exp\left[-\frac{1}{2}\varsigma(\boldsymbol{\rho})\right], \quad (2)$$

which we use to obtain ML estimators of $\boldsymbol{\rho}$.

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 distributions without conditioning variables.

2.2 The score vector and the Hessian matrix

Before studying the asymptotic properties of our proposed estimators, it is convenient to obtain generic expressions for the score and Hessian of a K-dimensional Gaussian copula or equivalently, of a multivariate Gaussian distribution for a vector of random variables $\mathbf{y}$.

If we assume *i.i.d.* observations, the relevant log-likelihood function for a sample of uniform ranks of size N will take the form $\sum_{i=1}^{N} \ln c_K(\mathbf{u}_i; \boldsymbol{\rho})$, where $c_K(\mathbf{u}_i; \boldsymbol{\rho})$ is given by (1), or analogously, for a sample of N Gaussian ranks, $\sum_{i=1}^{N} \ln \phi(\mathbf{y}_i; \boldsymbol{\rho})$, where $\phi_K(\mathbf{y}_i; \boldsymbol{\rho})$ is given by (2). In the unrestricted case, letting $\tilde{\mathbf{L}}$ and $\mathbf{K}$ denote the (strictly lower triangular) elimination matrix of order K and the commutation matrix of orders K, K, respectively (see Magnus (1988) for details), we can prove the following result:

Proposition 1 *Let* $\mathbf{s}_i(\boldsymbol{\rho})$ *denote the score function* $\partial \ln \phi_K(\mathbf{y}_i;\boldsymbol{\rho})/\partial\boldsymbol{\rho}$ *with* $\boldsymbol{\rho} = vecl(\mathbf{P})$*. Then*

$$\mathbf{s}_i(\boldsymbol{\rho}) = \frac{1}{2}(\tilde{\mathbf{L}} + \tilde{\mathbf{L}}\mathbf{K})[\mathbf{P}^{-1}(\boldsymbol{\rho}) \otimes \mathbf{P}^{-1}(\boldsymbol{\rho})]vec[\mathbf{y}_i\mathbf{y}_i' - \mathbf{P}(\boldsymbol{\rho})]. \tag{3}$$

Maximum likelihood (ML) estimation of $\boldsymbol{\rho}$ usually requires a numerical optimization procedure. However, given that in the bivariate case the score (3) takes the form

$$s_{\rho i}(\rho) = \frac{(1+\rho^2)y_{1i}y_{2i} - \rho(y_{1i}^2 + y_{2i}^2 - 1) - \rho^3}{(1-\rho^2)^2}, \tag{4}$$

(cf eq. (2.54) in Martin, Hurn and Harris (2013)), it is possible to obtain a numerically convenient closed-form expression for the ML estimator of ρ. Specifically,

Corollary 1 *In the bivariate case, the ML estimator of* ρ*,* $\hat{\rho}$*, will be the real root to the cubic equation*

$$\sum_{i=1}^{N} s_{\rho i}(\rho) = 0 \tag{5}$$

that leads to the largest log-likelihood value, where $s_{\rho i}(\rho)$ *is given by (4).*

Algebraic solutions to any cubic equation have been available since at least the early 16th century even though they remain relatively unknown. What is well known, though, is that every cubic equation with real coefficients has at least one real solution, while the other two solutions can be either real or a pair of complex conjugates. In Section 3.2 we revisit the uniqueness of the real root to equation (5).

Another example of practical interest that we will consider in the next sections arises when imposing an equicorrelated structure on $\mathbf{P}(\boldsymbol{\rho})$. Using the general expression (3), we can get the corresponding score by using the fact that

$$s_{eqi}(\rho) = \ell_{K(K-1)/2}'\mathbf{s}_i(\rho\ell_{K(K-1)/2}),$$

where $\ell_{K(K-1)/2}$ denotes a column vector of $K(K-1)/2$ ones. If we exploit the equicorrelated structure of

$$\mathbf{P}(\rho) = \rho\ell_K'\ell_K + (1-\rho)\mathbf{I}_K,$$

then we can write

$$\mathbf{P}^{-1}(\rho) = \frac{1}{1-\rho}\mathbf{I}_K - \frac{\rho\ell_K\ell_K'}{(1-\rho)^2 + \rho(1-\rho)K}.$$

As a result,

$$\ell_{K(K-1)/2}'(\tilde{\mathbf{L}} + \tilde{\mathbf{L}}\mathbf{K})[\mathbf{P}^{-1}(\rho) \otimes \mathbf{P}^{-1}(\rho)] = vec[-a^2\mathbf{I}_K + (b^2K^2 - b^2K - 2ab + 2abK)\ell_K\ell_K']'$$

with

$$a = \frac{1}{1-\rho} \text{ and } b = -\frac{\rho}{(1-\rho)^2 + \rho(1-\rho)K}.$$

Hence, we obtain:

Corollary 2 *Assume that* $\mathbf{P}(\rho)$ *has an equicorrelated structure with scalar parameter* ρ*. Then, the ML estimator of* ρ, $\hat{\rho}$, *will solve*

$$0 = \sum_{i=1}^{N} s_{eqi}(\rho),$$

where

$$s_{eqi}(\rho) = \frac{1}{2} tr\{[-a^2\mathbf{I}_K + (-2ab - b^2K + 2abK + b^2K^2)\ell_K\ell_K'][\mathbf{y}_i\mathbf{y}_i' - \mathbf{P}(\rho)]\}.$$

In any case, and regardless of the restrictions, $\mathbf{P}(\hat{\boldsymbol{\rho}})$ will be positive definite because of the penalty term induced by the Jacobian in the log-likelihood function.

In practice, the log-likelihood score is often used not only as the input to a steepest ascent, Berndt, Hall, Hall, and Hausman (1974) or quasi-Newton numerical optimization routine, but also to estimate the asymptotic covariance matrix of the ML parameter estimators. Nevertheless, both of these uses could be problematic. First, the results of Fiorentini et al. (1996) and many others suggest that alternative gradient methods, such as scoring or Newton–Raphson, usually show much better convergence properties, particularly when the parameters are close to the optimum, which in this case could be obtained by using the closed-form sample correlation coefficients of the Gaussian ranks as starting values.

Similarly, it is well known that the outer-product-of-the-score standard errors and test statistics can be very badly behaved in finite samples (Davidson and MacKinnon, 1993).

For both these reasons, we derive analytical expressions for the elements of the Hessian matrix. In the unrestricted case, in particular:

Proposition 2 *Let* $\mathbf{h}_i(\boldsymbol{\rho})$ *denote the Hessian function* $\partial\mathbf{s}_i(\boldsymbol{\rho})/\partial\boldsymbol{\rho}' = \partial^2 \ln\phi_K(\mathbf{y}_i;\boldsymbol{\rho})/\partial\boldsymbol{\rho}\partial\boldsymbol{\rho}'$, *with* $\boldsymbol{\rho} = vecl(\mathbf{P})$*. Then*

$$\begin{aligned} \mathbf{h}_i(\boldsymbol{\rho}) \quad = \quad & \frac{1}{2}(\tilde{\mathbf{L}} + \tilde{\mathbf{L}}\mathbf{K})\{-[\mathbf{P}^{-1}(\boldsymbol{\rho})\mathbf{y}_i\mathbf{y}_i'\mathbf{P}^{-1}(\boldsymbol{\rho})] \otimes \mathbf{P}^{-1}(\boldsymbol{\rho}) \\ & -\mathbf{P}^{-1}(\boldsymbol{\rho}) \otimes [\mathbf{P}^{-1}(\boldsymbol{\rho})\mathbf{y}_i\mathbf{y}_i'\mathbf{P}^{-1}(\boldsymbol{\rho})] + \mathbf{P}^{-1}(\boldsymbol{\rho}) \otimes \mathbf{P}^{-1}(\boldsymbol{\rho})\}(\tilde{\mathbf{L}}' + \mathbf{K}\tilde{\mathbf{L}}'). \end{aligned} \tag{6}$$

On this basis, we can easily show that:

Corollary 3 *In the bivariate case, the Hessian of* ρ *is given by*

$$h_{\rho i}(\rho) = \frac{1 - \rho^4 + 2y_{1i}y_{2i}\rho(3+\rho^2) - (y_{1i}^2 + y_{2i}^2)(1+3\rho^2)}{(1-\rho^2)^3}.$$

In turn:

Corollary 4 *Assume that* $\mathbf{P}(\rho)$ *has an equicorrelated structure with scalar parameter* ρ*. Then, the Hessian of* ρ *is given by*

$$h_{eqi}(\rho) = \ell_{K(K-1)/2}'\mathbf{h}_i(\rho\ell_{K(K-1)/2})\ell_{K(K-1)/2}.$$

2.3 Partial correlation and regression

When there are three or more variables involved, the interest of the researcher may lie in the correlation between two of them after partialling out the effect of the others. For example, in the case of three variables, y_1, y_2 and y_3, the partial correlation of the first two given the third is given by the well known expression:

$$\rho_{12.3} = \frac{\rho_{12} - \rho_{13}\rho_{23}}{\sqrt{1-\rho_{13}^2}\sqrt{1-\rho_{23}^2}}. \tag{7}$$

In this case, the ML estimator of $\rho_{12.3}$ can be directly obtained by plugging the ML estimators of ρ_{12}, ρ_{13} and ρ_{23} in (7) by virtue of the invariance property of ML estimation. In that regard, it is straightforward to prove that the Jacobian of $\rho_{12.3}$ with respect to ρ_{12}, ρ_{13} and ρ_{23} will be given by

$$\begin{aligned}
\frac{\partial \rho_{12.3}}{\partial \rho_{12}} &= \frac{1}{\sqrt{1-\rho_{13}^2}\sqrt{1-\rho_{23}^2}}, \\
\frac{\partial \rho_{12.3}}{\partial \rho_{13}} &= -\frac{\rho_{23} - \rho_{12}\rho_{13}}{\left(1-\rho_{13}^2\right)^{\frac{3}{2}}\sqrt{1-\rho_{23}^2}}, \\
\frac{\partial \rho_{12.3}}{\partial \rho_{23}} &= -\frac{\rho_{13} - \rho_{12}\rho_{23}}{\sqrt{1-\rho_{13}^2}\left(1-\rho_{23}^2\right)^{\frac{3}{2}}}.
\end{aligned}$$

The concept of partial correlation is intimately related to the idea of linear regression. Specifically, given that the Gaussian ranks have zero mean and unit variance by construction, the coefficients of the least squares projection of some y_k onto the remaining elements of $\mathbf{y}$, $\mathbf{y}_{(k)}$, which we denote by $\boldsymbol{\beta}^{(k)}$ henceforth, will coincide with the partial correlation of y_k with each of the elements of $\mathbf{y}_{(k)}$ given the other $K-2$. In matrix notation, we can write

$$\boldsymbol{\beta}^{(k)} = \mathbf{P}_{(kk)}^{-1}\mathbf{P}_{(k)k}, \tag{8}$$

where $\mathbf{P}_{(kk)}$ is the block of the correlation matrix that excludes the k^{th} row and column, while the vector $\mathbf{P}_{(k)k}$ contains its k^{th} column except the 1.

Once again, the ML estimators of $\boldsymbol{\beta}^{(k)}$ will be the result of plugging the ML estimators of $\mathbf{P}$ in (8). In that regard, the Jacobian of $\boldsymbol{\beta}^{(k)}$ with respect to $\boldsymbol{\rho}$ will be given by

$$\boldsymbol{\nabla}\boldsymbol{\beta}^{(k)}(\boldsymbol{\rho}) = \frac{\partial \boldsymbol{\beta}^{(k)}}{\partial \boldsymbol{\rho}'} = -\mathbf{P}_{(k)k}(\mathbf{P}_{(kk)}^{-1} \otimes \mathbf{P}_{(kk)}^{-1})\frac{\partial vec(\mathbf{P}_{(kk)})}{\partial \boldsymbol{\rho}'} + \mathbf{P}_{(kk)}^{-1}\frac{\partial \mathbf{P}_{(k)k}}{\partial \boldsymbol{\rho}'} \tag{9}$$

because

$$\mathrm{d}\boldsymbol{\beta}^{(k)} = \mathrm{d}\mathbf{P}_{(kk)}^{-1}\mathbf{P}_{(k)k} + \mathbf{P}_{(kk)}^{-1}\mathrm{d}\mathbf{P}_{(k)k} = -\mathbf{P}_{(kk)}^{-1}\mathrm{d}\mathbf{P}_{(kk)}\mathbf{P}_{(kk)}^{-1} + \mathbf{P}_{(kk)}^{-1}\mathrm{d}\mathbf{P}_{(k)k}.$$

In some cases, the magnitude of a coefficient may be important on its own. The growth regressions in MRW are one such example. Given that by construction Gaussian ranks lose the information on the original scale of the variables, it would be necessary to re-scale the coefficients appropriately. Relying on sample standard deviations to do so, however, seems unwise in the presence of high leverage observations. In Appendix A we show how to take this into consideration by adjusting regression slopes and intercept using interquartile ranges.

3 Asymptotic properties under correct specification

3.1 When margins are known

3.1.1 Information matrix equality

Under the maintained assumption of a Gaussian copula, computing the (minus) expectation of the Hessian is straightforward since (6) only involves squares and second-order cross products of correlated Gaussian variables. Specifically, letting $\boldsymbol{\rho}_0$ denote the true vector of the correlation parameters, which are such that $\mathbf{P}(\boldsymbol{\rho}_0) = E(\mathbf{y}\mathbf{y}')$, we can show that in the general unrestricted case

$$-E[\mathbf{h}_i(\boldsymbol{\rho}_0)] = \frac{1}{2}(\tilde{\mathbf{L}} + \tilde{\mathbf{L}}\mathbf{K})[\mathbf{P}^{-1}(\boldsymbol{\rho}_0) \otimes \mathbf{P}^{-1}(\boldsymbol{\rho}_0)](\tilde{\mathbf{L}}' + \mathbf{K}\tilde{\mathbf{L}}'), \tag{10}$$

which in the bivariate case reduces to

$$-E[h_{\rho i}(\rho_0)] = \frac{1 + \rho_0^2}{(1 - \rho_0^2)^2}.$$

Similarly, in the restricted equicorrelated case

$$-E[h_{eqi}(\rho_0)] = \frac{K(K + \rho_0^2 - 2K\rho_0^2 + K^2\rho_0^2 - 1)}{2(1 - \rho_0)^2[1 + (K - 1)\rho_0]^2},$$

which trivially reduces to the previous expression for $K = 2$.

Similarly, computing the variance of the score vector under the maintained assumption that the copula is Gaussian is also straightforward because (3) involves fourth powers and fourth-order cross products of correlated Gaussian variables. Indeed, we can easily show that in the unrestricted case

$$Var[\mathbf{s}_i(\boldsymbol{\rho}_0)] = \frac{1}{4}(\tilde{\mathbf{L}} + \tilde{\mathbf{L}}\mathbf{K})[\mathbf{P}^{-1}(\boldsymbol{\rho}_0) \otimes \mathbf{P}^{-1}(\boldsymbol{\rho}_0)]Var[vec(\mathbf{y}_i\mathbf{y}_i')][\mathbf{P}^{-1}(\boldsymbol{\rho}_0) \otimes \mathbf{P}^{-1}(\boldsymbol{\rho}_0)](\tilde{\mathbf{L}}' + \mathbf{K}\tilde{\mathbf{L}}'). \tag{11}$$

Not surprisingly, we can combine the usual formulas for

$$Var[vec(\mathbf{y}_i\mathbf{y}_i')] = E[vec(\mathbf{y}_i\mathbf{y}_i')vec(\mathbf{y}_i\mathbf{y}_i')'] - vec[\mathbf{P}(\boldsymbol{\rho}_0)]vec[\mathbf{P}(\boldsymbol{\rho}_0)]'$$

with the properties of the commutation and elimination matrices to show that (10) and (11) coincide.

3.1.2 Asymptotic distribution of the ML estimators

Under standard regularity conditions, we can exploit the expressions in the previous subsection, together with the delta method, to obtain the asymptotic variance of the ML estimators of $\boldsymbol{\rho}$ and $\boldsymbol{\beta}$, which in turn characterize their asymptotic distributions:

Proposition 3 *a) Let $\hat{\boldsymbol{\rho}}_N$ denote the ML estimator of $\boldsymbol{\rho}$ in model (2). Then,*

$$\sqrt{N}(\hat{\boldsymbol{\rho}}_N - \boldsymbol{\rho}_0) \xrightarrow{d} N[\mathbf{0}, \mathcal{I}^{-1}(\boldsymbol{\rho}_0)]$$

where $\mathcal{I}(\boldsymbol{\rho}_0) = Var[\mathbf{s}_i(\boldsymbol{\rho}_0)] = -E[\mathbf{h}_i(\boldsymbol{\rho}_0)]$.
b) Let $\hat{\boldsymbol{\beta}}_N^{(k)}$ denote the ML estimator of $\boldsymbol{\beta}^{(k)}$ in (8). Then,

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_N^{(k)} - \boldsymbol{\beta}_0^{(k)}) \xrightarrow{d} N[\mathbf{0}, \boldsymbol{\nabla}\boldsymbol{\beta}^{(k)}(\boldsymbol{\rho}_0)\mathcal{I}^{-1}(\boldsymbol{\rho}_0)\boldsymbol{\nabla}\boldsymbol{\beta}^{(k)}(\boldsymbol{\rho}_0)'],$$

where $\boldsymbol{\nabla}\boldsymbol{\beta}^{(k)}(\boldsymbol{\rho}) = \partial\boldsymbol{\beta}^{(k)}(\boldsymbol{\rho})/\partial\boldsymbol{\rho}'$ is given in (9).

3.2 Replacing margins with empirical cdf's

The marginal distributions of the K variables in the observed vector $\mathbf{x}$ are rarely known in practice. The most common solution is a two-step estimation procedure, whereby the margins $G_k(x_k)$ are replaced by their (re-scaled) empirical cdf counterparts $\hat{G}_k(x_k)$. Thus, the proposed estimators can be viewed as functions of the Gaussian ranks obtained from the (uniform) sample ranks, where the scaling factor $N/(N+1)$ is simply introduced to avoid potential problems with the copula density blowing up at the boundary of $[0,1]^K$. Smoothed versions of the empirical cdf can also be used, but the asymptotic effects are 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. Among other things, this in turn implies that the real solution to the cubic equation in (5), which defines the unrestricted ML estimator of ρ in the bivariate case, will be unique for $N > 6$.

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)).

Following Chen and Fan (2006), we can obtain the variance of a generic influence function $m_\phi(\boldsymbol{\rho})$ adjusted for non-parametric estimation of the margins by computing the variance of the adjusted function

$$m_\phi^c(\boldsymbol{\rho}, \mathbf{0}) = m_\phi(\boldsymbol{\rho}, \mathbf{0}) + n_\phi(\boldsymbol{\rho}),$$

where

$$n_{\phi}=\sum_{j=1}^{K}\int_{0}^{1}[1\{U_j\leq u_j\}-u_j]W_{\phi}^{j}du_j,$$

with

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

In the case of the correlation parameters, one can capture the resulting inflation in variance by adding linear combinations of second order Hermite polynomials of each of the variables (in Gaussian form) to the original scores $\mathbf{s}_{\boldsymbol{\rho}i}(\boldsymbol{\rho},\mathbf{0})$, as the following result shows:

Lemma 1 *a) In the bivariate case, the correction to $s_{\rho i}(\rho)$ is given by*

$$n_{\rho i}(\rho)=\frac{\rho}{1-\rho^2}\left[\frac{H_2(y_{1i})+H_2(y_{2i})}{\sqrt{2}}\right].$$

b) In the trivariate case, the correction to $s_{\rho_{12}i}(\rho_{12},\rho_{13},\rho_{23})$ is given by

$$n_{\rho_{12}i}(\rho_{12},\rho_{13},\rho_{23})=\frac{\rho_{12}-\rho_{13}\rho_{23}}{1-\rho_{12}^2-\rho_{13}^2-\rho_{23}^2+2\rho_{12}\rho_{13}\rho_{23}}\left[\frac{H_2(y_{1i})+H_2(y_{2i})}{\sqrt{2}}\right],$$

and the same applies to $s_{\rho_{13}i}$ and $s_{\rho_{23}i}$ by suitably changing the subscripts.

Analogous expressions apply for general K.

3.3 Efficiency comparison with other moment estimators

3.3.1 Correlation measures

Obvious consistent moment-based estimators of ρ in the bivariate case are $\check{\rho}=N^{-1}\sum_{i=1}^{N}y_{1i}y_{2i}$ and $\tilde{\rho}=\check{\rho}/\sqrt{\tilde{\sigma}_1^2\tilde{\sigma}_2^2}$, with $\tilde{\sigma}_k^2=N^{-1}\sum_{i=1}^{N}y_{ki}^2$. These are the sample covariance and correlation coefficients of the Gaussian ranks y_1 and y_2, respectively, which impose that their population mean is 0. Nevertheless, these estimators are generally inefficient relative to the ML estimator when the margins are known:

Proposition 4 *When the bivariate copula is Gaussian and the margins are known, the asymptotic variances of $\hat{\rho}$, $\tilde{\rho}$ and $\check{\rho}$, which are the ML estimator of ρ, and the sample correlation and covariance coefficients of the Gaussian ranks, respectively, are*

$$AVar(\hat{\rho})=\frac{(1-\rho^2)^2}{1+\rho^2},\quad AVar(\tilde{\rho})=(1-\rho^2)^2\quad\text{and}\quad AVar(\check{\rho})=\frac{1}{1+\rho^2},$$

so that

$$AVar(\hat{\rho})\leq AVar(\tilde{\rho})\leq AVar(\check{\rho}),$$

with equality if and only if $\rho=0$.

Although when $K=3$ the asymptotic variances of these three estimators, which we omit for the sake of brevity, are different, exactly the same ranking applies, as we show in the proof of Proposition 6.

Interestingly, though, when the margins are non-parametrically estimated, we obtain the following modified version of Proposition 4:

Proposition 5 *When the bivariate copula is Gaussian and the margins are replaced by their empirical cdfs, the asymptotic variances of $\hat{\rho}$, $\tilde{\rho}$ and $\check{\rho}$, which are the ML estimator of ρ, and the sample correlation and covariance coefficients of the Gaussian ranks, respectively, are given by*

$$AVar(\hat{\rho}) = AVar(\tilde{\rho}) = AVar(\check{\rho}) = (1-\rho^2)^2. \tag{12}$$

In other words, all the alternative moment estimators of ρ are equally efficient and their asymptotic variance coincide with the one corresponding to the sample Pearson correlation of the Gaussian ranks with known margins.

Importantly, we show in the proof of Proposition 7 that exactly the same result applies to the unrestricted trivariate case. As a result, with non-parametric margins no efficiency gains accrue from maximizing the log-likelihood function (2) by implicitly solving the non-linear equations in (3), at least up to the usual first-order terms.

Figure 3a reports the asymptotic variances for the unconstrained ML estimators in both bivariate and equicorrelated trivariate contexts assuming known margins, which we denote by $ML_{K=2}$ and $ML_{K=3}$, respectively. It also includes plots for the same estimators in the more realistic situation in which margins are estimated non-parametrically, which we denote by ML^{NP}. As expected, the relative rankings coincide with the statements of Propositions 4 and 5. In particular, the gains from increasing K disappear when the margins are non-parametrically estimated, as the asymptotic variance of ML^{NP} is the same regardless of the cross-sectional dimension.

3.3.2 Partial correlation coefficients and regression

As mentioned earlier, the concept of partial correlation is intimately related to the idea of linear regression. Therefore a natural alternative estimator to the ML ones described in Section 2.3 is given by simply applying OLS to the Gaussian ranks of the original data.

We will restrict attention to the trivariate case for the rest of the subsection. Specifically, in the case of three variables, y_1, y_2 and y_3, without loss of generality we consider the regression of y_1 onto y_2 and y_3:

$$y_1 = \alpha + \beta_2^{(1)} y_2 + \beta_3^{(1)} y_3 + \epsilon. \tag{13}$$

Let $\hat{\boldsymbol{\beta}}^{(1)}$, $\check{\boldsymbol{\beta}}^{(1)}$ and $\tilde{\boldsymbol{\beta}}^{(1)}$ denote the ML, the OLS without intercept, and the OLS with intercept estimators of $\boldsymbol{\beta}^{(1)} = (\beta_2, \beta_3)'$ in (13), respectively. In view of the discussion following Proposition 4, it is not surprising that both OLS estimators are generally inefficient relative to the ML estimator when the margins are known, as the following proposition shows:

Proposition 6 *When the copula is Gaussian and the margins are known, the asymptotic variances of* $\hat{\boldsymbol{\beta}}^{(1)}$, $\check{\boldsymbol{\beta}}^{(1)}$ *and* $\tilde{\boldsymbol{\beta}}^{(1)}$ *in (13), are such that*

$$AVar(\hat{\beta}_j^{(1)}) \leq AVar(\check{\beta}_j^{(1)}) = AVar(\tilde{\beta}_j^{(1)}), \quad j = 2, 3.$$

Nevertheless, these gains disappear when the margins are non-parametrically estimated by the empirical cdf:

Proposition 7 *When the copula is Gaussian and the margins are replaced by their empirical cdfs, the asymptotic variances of* $\hat{\boldsymbol{\beta}}^{(1)}$, $\check{\boldsymbol{\beta}}^{(1)}$ *and* $\tilde{\boldsymbol{\beta}}^{(1)}$ *in (13), are such that*

$$AVar(\hat{\beta}_j^{(1)}) = AVar(\check{\beta}_j^{(1)}) = AVar(\tilde{\beta}_j^{(1)}) \quad j = 2, 3.$$

In fact, it is easy to prove that $\check{\boldsymbol{\beta}}^{(1)}$ and $\tilde{\boldsymbol{\beta}}^{(1)}$ numerically coincide because the sample means of the estimated Gaussian ranks are identically 0. Therefore, with non-parametric margins researchers can use standard OLS routines to efficiently estimate the Gaussian rank regression coefficients without the need to numerically maximise the log-likelihood function (2).

Figure 3b is the counterpart to Figure 3a for regression coefficients instead of correlations. As can be seen, the general patterns are in line with the results in Propositions 6 and 7. Specifically, when the margins are non-parametrically estimated, OLS is as efficient as ML, as stated in Proposition 7.

4 Misspecification analysis

4.1 Pseudo-true values

In the context of multivariate location-scale models with non-normal observations, many empirical researchers continue to use the Gaussian pseudo-maximum likelihood estimators advocated by Bollerslev and Wooldridge (1992) among others because they remain consistent for the (conditional) mean and variance parameters as long as those moments are correctly specified. However, no such result seems to be available for copulas. The following result characterizes the analogous property for the Gaussian rank correlations:

Proposition 8 *Assume there exists* $\boldsymbol{\rho}_\infty$ *that solves* $\mathbf{P}(\boldsymbol{\rho}_\infty) = E(\mathbf{y}\mathbf{y}')$*, where* $\mathbf{P}(\boldsymbol{\rho})$ *is the potentially restricted, but correctly specified correlation matrix of the Gaussian ranks* $\mathbf{y}$ *in Assumption 1. Then, the Gaussian pseudo-ML estimator of the* $p \times 1$ *vector of free parameters* $\boldsymbol{\rho}$*, with* $p \leq \frac{1}{2}K(K-1)$*, remains consistent even when the true copula is not Gaussian.*

The same is true of the $K(K-1)/2$ sample Gaussian rank correlation coefficients $\tilde{\rho}_{kj}$ if $\mathbf{P}(\boldsymbol{\rho})$ were unrestricted. In fact, it is easy to see that in the unrestricted case, the pseudo-true values of the ML estimators coincide with the population values of the usual Pearson correlation coefficients of the Gaussian ranks.

An interesting question worth investigating is the behavior of these correlation coefficients for some well-known non-Gaussian copulas. In particular, we consider the Clayton copula and the Student t copula. The first one is a member of the Archimidean family, whose copula function admits an explicit formula, a popular feature when modeling dependence. Figures 2c-d display a bivariate Clayton copula density with Gaussian rank correlation .25 and Gaussian margins. Figures 2e-f does the same but for the Student t copula. As is well known, the Student t copula, which nests the Gaussian copula, is a very popular example of elliptical copula; see Amengual and Sentana (2020) for tests of one versus the other.

Figure 4a presents the population Gaussian rank correlation ρ_∞ as a function of the dependence parameter θ of a Clayton copula. Similarly, Figure 4b presents the analogous functions for several Student t copulas that differ in the number of degrees of freedom, with θ denoting the value of the correlation of the bivariate Student t distributions underlying those copulas.

4.2 Asymptotic distribution

In Section 3.1.2, we derived the asymptotic distribution of the ML estimator of the Gaussian rank correlations when the true copula is Gaussian. In this section, in contrast, we find the asymptotic variance when the true copula is not Gaussian. To do so, we simply need to combine the expected value of the Hessian in (10) with the asymptotic variance of the average score in (11), which depends on the true copula through the fourth moments in $Var[vec(\mathbf{y}_i\mathbf{y}_i')]$. Given that the Gaussian ranks are a non-linear transformation of the uniform ranks, we are forced to resort to numerical quadrature for the calculation of the fourth moments of the Gaussian ranks.

Figure 5a shows the asymptotic variance of the pseudo-ML estimator of the Gaussian rank correlation as a function of the dependence parameter of the Clayton copula. Similarly, Figure 5b contains analogous results for the Student t copulas. Not surprisingly, the asymptotic variance converges to the values in Proposition 4 as the degrees of freedom of the t copula increase without bound.

5 Comparison with alternative estimators

For the sake of brevity, in this section we restrict the analysis to the bivariate case.

5.1 Spearman correlation

Consider the following moment conditions for the uniform ranks

$$E[\mathbf{m}_i^I(\boldsymbol{\theta})] = E\left[\begin{pmatrix} u_{1i}u_{2i} - \mu_1\mu_2 - \sqrt{\sigma_1^2\sigma_2^2}\rho \\ u_{1i} - \mu_1 \\ u_{2i} - \mu_2 \\ u_{1i}^2 - (\mu_1^2 + \sigma_1^2) \\ u_{2i}^2 - (\mu_2^2 + \sigma_2^2) \end{pmatrix}\right] = \mathbf{0}, \tag{14}$$

where $\boldsymbol{\theta} = (\rho, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2)'$. Let $\tilde{\rho}_S^I$ denote method of moments estimator of ρ based on (14). Given that u_1 and u_2 are uniform ranks, we can interpret this estimator as a sample version of the Spearman correlation coefficient.[2]

As expected, the value of the Spearman correlation coefficient is zero for the independence copula. In addition, given that its asymptotic standard error is 1 in that case, tests of indepedence between two random variables based on their Spearman correlation coefficient will have exactly same power against identical sequences of local alternatives as independence tests based on their Gaussian rank correlation.

5.2 Pearson correlation

Pearson correlations are applied directly to the original data $\mathbf{x}$. If the data were Gaussian (uniform) then we would end up with the same figures as for the Gaussian (Spearman) rank correlation. To make the comparisons more interesting, in this section we consider four alternative marginal distributions for the raw data:

1. Weibull,
2. Asymmetric Laplace,
3. Tukey (symmetric), and
4. Mixture of two Weibulls.

Figure 6 displays the densities of these distributions, whose descriptions appear in Online Appendix E.

Once more, we resort to numerical integration to obtain the relevant cross-moments involved in computing both the pseudo-true values and the asymptotic variances of the Pearson correlation coefficients. A convenient feature of the four marginal distributions above is that there are closed-form expressions for the corresponding quantile functions, which speeds up the calculations.

[2] In Online Appendix D.1, we consider three alternative estimators of the Spearman correlation, while in Online Appendix D.2 we also study the asymptotic properties of all of them under the assumption that the data is generated from a Gaussian copula.

5.3 Comparison

In Figure 7a we report the pseudo-true values of the Pearson and Spearman correlation coefficients when the true copula is Gaussian for the four marginal distributions in the previous section. As can be seen, the bias is more pronounced for the Pearson correlations, especially for the mixture of Weibulls.

In turn, Figure 7b presents the asymptotic variances for the Pearson, Spearman and Gaussian rank correlations for the same data generating processes (DGP). Not surprisingly, the Gaussian rank correlation estimator has smaller variance than the Spearman correlation coefficent for all values of $\rho \neq 0$, which in turn is more precise than its Pearson counterpart.

Once again, however, the pseudo-true value of the Pearson correlation coefficients are zero for the independence copula. In addition, given that its asymptotic standard error is 1 under the same circumstances, indepedence tests based on the Pearson, Spearman or Gaussian rank correlation coefficients will have exactly the same power against identical sequences of local alternatives.

6 Monte Carlo Evidence

6.1 Design and estimation details

In this section, we study the finite sample performance of the different estimators discussed in previous sections by means of an extensive Monte Carlo exercise, with several experimental designs aimed to assess the estimators under both correct specification and misspecification. In all cases, we consider 10,000 replications.

We first simulate and estimate bivariate and trivariate –equicorrelated– copula models for correlation parameters $-.25, -.1, -.05, .05, .1$ and $.25$ when the true copula is Gaussian. Then, we study the effects of misspecification by simulating from a Student t copula with 8 degrees of freedom and the same correlation parameters for the underlying multivariate distribution. Importantly, we also consider a third DGP which consists of a Gaussian copula contaminated with five atypical observations that we keep fixed across samples. As we shall see, the impact of those five ouliers is more dramatic the smaller the sample size. In that respect, in all our designs we consider four samples sizes: N=50, 200, 800 and 3,200. As for the margins, we use the asymmetric Laplace distributions in Section 5.2 with location, scale, and shape parameter values equal to 0, 10 and 0.9, respectively (see Online Appendix E for details).[3]

[3]We have also repeated the entire Monte Carlo exercise using log-normal marginal distributions instead. The results, which are available upon request, indicate that the behavior of the different estimators is qualitatively very similar to the reported in this section.

6.2 Sampling distribution of the different estimators

Table 1 contains means and standard deviations of the Monte Carlo sampling distributions for the Pearson, Spearman and Gaussian rank correlation estimators in the bivariate case for both Gaussian and Student t copulas.

By and large, the behavior of these different estimators when the copula is Gaussian, which is reported in the first six columns, is in accordance with the asymptotic results in Section 3.3. In particular, the bias arising in the Pearson correlation coefficients is in line with the pseudo-true values reported in Figure 7 for asymmetric Laplace margins. In turn, the last six columns show that when the copula is Student t, the bias seems to be systematically smaller for the Gaussian rank correlations despite the misspecification.

In turn, Table 2 looks at the trivariate regression case, in which we consider not only Gaussian rank regressions and OLS but also *Least trimmed squares* (LTS). This last estimator is such that a fraction κ of the observations corresponding to the largest κN OLS residuals is considered unrepresentative and subsequently omitted from the calculations; see Rousseeuw (1984, 1985) for further details.

Once again, the sampling distributions of the Gaussian rank-based betas present lower biases than the corresponding OLS estimates based on the raw data. Remarkably, the performance of the LTS estimator with $\kappa = .5$ is not very good, as it shows considerable biases. Moreover, when the copula is Student t, the standard deviations for the Gaussian rank-based betas are about 10% smaller than the ones for the OLS and LTS coefficients based on the raw data.

6.3 Finite sample inference

In order to gauge the extent to which our proposed asymptotic corrections for non-parametric estimation of the marginal cdfs work in finite samples, we also look at the t tests associated to the estimated Pearson and Gaussian rank correlation coefficients in the bivariate case, as well as the t and F tests in the trivariate case. We do so under both correct specification and misspecification of the copula.

Specifically, the first two columns of Table 3 report the finite sample sizes at the 5% level of the F tests of $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}(\rho_0)$ for both OLS and Gaussian ranks in the trivariate case when a Gaussian copula is used to generate the data, while the next two columns do the same for the two-sided t tests of the same null hypothesis.[4] In turn, the last four columns report analogous rejection rates of $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_\infty(\theta)$, where $\boldsymbol{\beta}_\infty(\theta)$ is the pseudo-true value of the Gaussian rank regression coefficient vector corresponding to a Student t copula with 8 degrees of freedom and equicorrelation parameter θ for the underlying

[4] Tables 4 and 5 in Amengual, Sentana and Tian (2020) contain the rejection rates at the 1% and 10% significance levels too.

multivariate distribution. Given the number of Monte Carlo replications, the 95% asymptotic confidence interval for the rejection probabilities for all those tests is (4.57,5.43).

For small samples of $N = 50$, none of the test statistics seem to follow their asymptotic distributions. However, the size distortions become much smaller in samples of size 800. Therefore, the correction for non-parametric first-stage estimation of the margins does not seem to work well unless the sample size is large. The same pattern is present for the t tests in the bivariate case, which for the sake of brevity we omit here (see Table 3 in Amengual, Sentana and Tian (2020)).

In addition, in the case of the regression tests, the univariate t tests present smaller size distortions than the joint F tests irrespective of whether we look at OLS and LTS applied to raw data or Gaussian rank-based coefficients.

6.4 The effect of outliers

The tougher DGP we consider is the one in which the original Gaussian copula is contaminated with five extreme observations taken as

$$(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = \mathbf{x}_0 \times (-.5, 1, 0) \times IQR_0 + \ell_5 \ell_3' \times MED_0$$

where $\mathbf{x}_0 = (5, 6, 7, 8, 9)'$, while IQR_0 and MED_0 denote the interquartile range and median of the standardized asymmetric Laplace distribution, respectively. Those observations aim to induce additional negative dependence between x_1 and x_2 while reducing dependence between both x_1 and x_3 and x_2 and x_3.

Results for the bivariate case are reported in Table 4. In addition to means and standard deviations, we also report the frequency of estimates bigger than zero. According to this last statistic, the Spearman correlation coefficient is the winner and Pearson's the worst by far, with the Gaussian rank correlation close to Spearman's.

Tables 7 and 8 in Amengual, Sentana and Tian (2020) do the same as Table 4 here but for the regression coefficients $\beta_2^{(1)}$ and $\beta_3^{(1)}$, respectively. Importantly, the Gaussian rank regression coefficients do not only outperform OLS but also the LTS robust estimator. Nevertheless, we should emphasize that the effect of the contaminated sample is quite strong in terms of the biases of correlations and regression coefficients, particularly for sample sizes of 50 and 200 observations.

7 Empirical applications

7.1 Migration and growth rates

As mentioned in the introduction, we apply the different estimators discussed in previous sections to study the relationship between migration and growth rates across US states over the twentieth century using the dataset in chapter 11 of Barro and Xala-i-Martin (2003). Specifically, we look at the relationship between the annual rate of net migration into region i between years $t-T$ and t, $m_{i,t}$, and (log) per capita income at the beginning of the period, $\ln(Y_{i,T-t})$, to assess whether there exists a positive effect of per capita income on migration across US states.

We first consider OLS (black solid line), LTS (red dotted line) and also *Least median of squares* (LMS, red dashed line) applied to the original data. As is well known, LMS minimizes the median of the square residuals instead of the mean square residual; see Rousseeuw (1984, 1985) for further details. Figure 8a replicates the scatter plot in Figure 11.10 in Barro and Xala-i-Martin (2003), with log of 1900 per capita income on the horizontal axis and the average net migration rate for 48 U.S. states or territories from 1900 to 1990 on the vertical axis. As can be seen, the three procedures deliver a positive relationship between those varables as the theory predicts. Nevertheless, it can be easily noticed that the OLS slope is more pronounced than the LTS and LMS, which are very close to each other. This discrepancy is mostly driven by Florida, Arizona, California, and Nevada (the four points with $m_{i,t} > .025$), which have notably higher net migration rates than the values predicted by their initial levels of income.

In Figure 8b we transform the original data into Gaussian ranks and then we compute their Pearson correlation coefficient, which coincides with the Gaussian rank correlation of the raw observations. We also report LTS and LMS applied to the transformed data. Interestingly, now the three lines look very much alike, confirming that Gaussian rank procedures are insensitive to outliers, unlike Pearson correlations.

7.2 The augmented Solow growth model

In an influential paper, MRW proposed an augmented version of the Solow growth model which takes not only physical capital but also human capital into account. They also showed that their augmented model improves the performance of the textbook Solow model in two important respects: (i) the OLS regression R^2 increases from .59 to .78, and (ii) the implied Cobb-Douglas coefficients are much closer to their predicted values.

Nevertheless, Temple (1998) highlighted that an important characteristic of the cross-section growth data used by MRW is that it contains many influential observations which could substantially alter the validity of their empirical conclusions. For that reason, he used reweighted least squares (RWLS) –a particular case of Rousseeuw (1984) LTS estimator– to deal with outliers. His results showed that the augmented Solow growth model continues to perform well not only in the full sample but also in several alternative subsamples.

We apply our proposed Gaussian regression procedures to the same data set.[5] Using the same country groups as MRW, the results reported in Panel A of Table 5 also support the augmented Solow growth model because the R^2 are high for all groups of countries except the OECD, and the signs of the coefficients coincide with the theoretical predictions. At the same time, we also find that our proposed Gaussian rank regression prodedure shrinks considerably the coefficient of $\ln(n + g + \delta)$, where n is population growth, g is physical capital growth and δ is the depreciation rate. In addition, the R^2 are also smaller for LTS and Gaussian rank regressions when we use the country classification by Temple (1998). By and large, we can conclude that although the original MRW results are not very accurate because of the presence of outliers, their main conclusions are not severely influenced by them.

8 Conclusions and directions for futher research

In this paper we study the asymptotic properties of both Pearson correlation coefficients of Gaussian ranks, and Gaussian rank regressions, namely OLS applied to those ranks. We show that these procedures are as efficient as maximum likelihood when the true copula is Gaussian and the margins are non-parametrically estimated, and remain consistent for their population analogues otherwise. We compare them to Spearman and Pearson correlations, and their regression counterparts based on raw data. Empirical applications to migration and growth rates across US states, the augmented Solow growth model, and individual stocks momentum and reversals during the global financial crisis confirm that Gaussian rank procedures are insensitive to outliers, unlike Pearson correlations and OLS regressions. Thus, they are doubly robust.

Several important topics deserve further investigation. From the theoretical point of view, we would like to extend our study of the properties of the Gaussian rank correlation and regression procedures under misspecification of the Gaussian copula to a situation in which the margins are non-parametrically estimated by means of the empirical cumulative distribution function. The study of the statistical properties of Spearman correlations and uniform rank regressions in those circumstances is also worth exploring.

In addition, we could compare the finite sample size and power of (conditional) independence tests

[5] Unfortunately, the data reported by MRW only contains two decimal figures, which prevents us from exactly replicating their empirical results.

based on the different correlation and regression procedures that we have considered, which is especially relevant given their markedly different sensitivity to outliers.

Finally, the modification of our procedures to deal with instrumental variables and panel data would substantially widen their scope. In that respect, it is important to remember that in their comment to Islam (1995), Lee, Pesaran and Smith (1998) highlighted that the conclusions of the cross-sectional growth empirics literature might be altered in the context of dynamic panel data models.

All these extensions constitute very interesting avenues for further research.

References

Amengual, D. and E. Sentana (2015): "Is a normal copula the right copula?" CEMFI Working Paper 1504.

Amengual, D. and E. Sentana (2020): "Is a normal copula the right copula?", *Journal of Business and Economic Statistics*, 38, 350–366.

Amengual, D., E. Sentana and Z. Tian (2020): "Gaussian rank correlation and regression" CEMFI Working Paper 2004.

Babu, G. J., and C.R. Rao (1989): "Joint asymptotic distribution of marginal quantiles and quantile functions in samples from a multivariate population", in C.R. Rao and M.M. Rao (eds.) *Multivariate Statistics and Probability, Essays in Memory of Paruchuri R. Krishnaiah*, Academic Press.

Barro, R.J. and X. Sala-i-Martin (2003): *Economic Growth,* 2nd edition, MIT Press.

Berndt, E.R., B.H. Hall, R.E. Hall and J.A. Hausman, (1974): "Estimation and inference in nonlinear structural models", *Annals of Economic and Social Measurement*, 3, 653–665.

Boudt, K., J. Cornelissen and C. Croux (2012): "The Gaussian rank correlation estimator: robustness properties", *Statistics and Computing*, 22, 471–483.

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

Davidson R., and J.G. MacKinnon (1993): *Estimation and inference in econometrics*, Oxford University Press.

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

Fiorentini, G., G. Calzolari and L. Panattoni (1996): "Analytical derivatives and the computation of GARCH models", *Journal of Applied Econometrics*, 11, 399–417.

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

Islam, N. (1995), "Growth empirics: a panel data approach", *Quarterly Journal of Economics*, 110, 1127-1170.

Joe, H. (2015): *Dependence modeling with copulas*, CRC Press.

Koenker, R., and G. Bassett Jr (1978): "Regression quantiles", *Econometrica*, 46, 33–50.

Lee, K., M.H. Pesaran and R. Smith (1998): "Growth empirics: a panel data approach – a comment", *Quarterly Journal of Economics*, 113, 319-323.

Magnus, J.R. (1988): *Linear structures*, Oxford University Press.

Mankiw, N.G., D. Romer and D.N. Weil (1992): "A contribution to the empirics of economic growth", *Quarterly Journal of Economics*, 107, 407-437.

Martin, V., S. Hurn and D. Harris (2013): *Econometric modelling with time series: specification, estimation and testing*, Cambridge University Press.

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, Elsevier.

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

Rousseeuw, P.J. (1984): "Least median of squares regression", *Journal of the American Statistical Association*, 79, 871–880.

Rousseeuw, P.J. (1985): "Multivariate estimation with high breakdown point", in W. Grossman, G.Pflug, I. Vincze and W.Wertz (eds.) *Mathematical Statistics and Applications* vol. B, Reidel Publishing Company.

Temple, J.R.W. (1998): "Robustness tests of the augmented Solow model, *Journal of Applied Econometrics*, 13, 361–375.

Var der Vaart, A.W. (1998): *Asymptotic statistics*, Cambridge University Press.

Appendix

A Some practical considerations

Uniform or Gaussian ranks make sense when the original variables are continuous. But if one or more of the explanatory variables is a dummy, they should be left unchanged, although strictly speaking, rank transformations would only change the two values that a dummy can take, which will depend on the original fraction of 0's and 1's.

Correlation coefficients are invariant to the scale of the variables involved, so the comparison between Pearson, Spearman and Gaussian rank correlations is straightforward. In some important contexts, such as autoregressions, the same is true of regression coefficients. Similarly, if the main objective of the empirical analysis is to assess whether or not a specific regression coefficient is zero, the scale of the variables is also irrelevant.

Nevertheless, in some cases the magnitude of a coefficient may be important on its own. Given that by construction Gaussian ranks lose the information on the original scale of the variables, it would be necessary to re-scale the regression slope coefficients appropriately. Relying on sample standard deviations to do so, however, seems unwise in the presence of high leverage observations. For that reason, our suggestion is to use the following simple transformation:

$$\beta^*_{1|2} = \beta_{1|2} \frac{IQR(x_1)}{IQR(x_2)},$$

where $\beta_{1|2}$ is the Gaussian rank regression coefficient in the regression of y_1 on y_2, while $IQR(x_j)$ is the theoretical interquartile range of the relevant raw variable.

However, since those interquartile ranges are usually unknown and must be replaced by their sample counterparts, it becomes necessary to adjust the standard errors of the different estimators of $\beta^*_{1|2}$ to take into account both the sampling variability of the sample interquartile ranges and their covariability with the Gaussian rank correlation coefficients. We can do so by relying on standard GMM methods. Specifically, if we write the estimators as the solution to the exactly identified system of moment conditions

$$E\begin{bmatrix} y_1y_2 - \mu_1\mu_2 - \sqrt{\sigma_1^2\sigma_2^2}\rho \\ y_1 - \mu_1 \\ y_2 - \mu_2 \\ y_1^2 - \mu_1^2\sigma_1^2 \\ y_2^2 - \mu_2^2\sigma_2^2 \\ 1\{x_1 \leq q_{1,0.25}\} - 0.25 \\ 1\{x_1 \leq q_{1,0.75}\} - 0.75 \\ 1\{x_2 \leq q_{2,0.25}\} - 0.25 \\ 1\{x_2 \leq q_{2,0.75}\} - 0.75 \end{bmatrix} = E[\mathbf{m}_i^q(\boldsymbol{\theta})] = \mathbf{0},$$

then the only thing we need is the expected Jacobian matrix of the above moment conditions and the variance of the associated influence functions. The non-differentiability of the influence functions corresponding to the quartiles may appear problematic at first sight, but it can be easily dealt with by using the procedures discussed in Koenker and Bassett (1978). As for the covariance matrix of the influence functions, the additional terms we need would be $cov(1\{x_i \leq q_{i,l}\}, y_j)$, $cov(1\{x_i \leq q_{i,l}\}, y_j^2)$, $cov(1\{x_i \leq q_{i,l}\}, y_iy_j)$ and $cov(1\{x_i \leq q_{i,l}\}, 1\{x_j \leq q_{j,k}\})$ for $i, j \in \{1, 2\}$. In this regard, it is well known that if $q_{i,l} \leq q_{i,k}$, then

$$cov(1\{x_i \leq q_{i,l}\}, 1\{x_i \leq q_{i,k}\}) = F_i(q_{i,l})[1 - F_i(q_{i,k})],$$

Similarly, we can use the results in Babu and Rao (1989) to show that

$$
\begin{aligned}
cov(1\{x_i \leq q_{i,l}\}, y_i) &= E(1\{x_i \leq q_{i,l}\}y_i) - E(1\{x_i \leq q_{i,l}\})E(y_i) \\
&= E\{y_i | y_i \leq \Phi^{-1}[F_i(q_{i,l})]\}F_i(q_{i,l}).
\end{aligned}
$$

Not surprisingly, we obtain the same asymptotic variance for the quantiles of x_i if we rely on the information matrix equality

$$
\frac{1}{N}\sum_{i=1}^{N} \frac{\partial 1\{x_i \leq q_{i,l}\}}{\partial q_{i,l}} \approx \frac{\partial E(1\{x_i \leq q_{i,l}\})}{\partial q_{i,l}} = f_i(q_{i,l}),
$$

where the approximation error is $o_p(1)$ (see e.g. Van der Vaart (1998)).

Either way, we can finally show that

$$
\begin{aligned}
cov(1\{x_i \leq q_{i,l}\}, y_i^2) &= E\{y_i^2 | y_i \leq \Phi^{-1}[F_i(q_{i,l})]\}F_i(q_{i,l}) - E(y_i^2)F_i(q_{i,l}) \\
&= [E\{y_i^2 | y_i \leq \Phi^{-1}[F_i(q_{i,l})]\} - 1]F_i(q_{i,l})
\end{aligned}
$$

and

$$
cov(1\{x_i \leq q_{i,l}\}, y_i y_j) = E\{y_i y_j | y_i \leq \Phi^{-1}[F_i(q_{i,l})]\}F_i(q_{i,l}) - E(y_i y_j)F_i(q_{i,l}).
$$

Obviously, these adjustments only make sense when the regressors are continuous variables. If some of the regressors are dummy variables, we would only need to scale the regression coefficient by $IQR(x_1)$ to get to the desired scale.

Similar issues arise with the intercept. In many empirical regressions, either the fitted line is restricted to go through the origin, or the only parameters of interest are the slopes. In some cases, though, the magnitude of the intercept itself may be relevant. Given that Gaussian rank regressions based on non-parametric marginal cdfs will have a zero intercept even if we added a constant to the regressions, it is also convenient to have a robust estimator of the coefficient of the constant. By analogy with the usual OLS intercept estimator, in the bivariate case we could consider

$$
med(x_1) - \beta_{12}\frac{IQR(x_2)}{IQR(x_1)}med(x_2),
$$

where $med(x_j)$ denotes the (population) median of the corresponding observed variable. The asymptotic distribution of this estimator can be easily obtained by adapting the GMM procedures for the adjusted slope coefficients that we have described above to an extended set of influence functions that also includes

$$
1\{x_1 \leq q_{1,0.5}\} - 0.5 \quad \text{and} \quad 1\{x_2 \leq q_{2,0.5}\} - 0.5.
$$

Once again, if x_2 or any of the other regressors were a dummy variable, no adjustment for scale would be necessary because it would usually be sufficient to compare the median of x_1 when the dummy is 0 with its median when it is one.

Table 1: Correlation parameter estimators

	Gaussian copula						Student t copula					
	Pearson		*Spearman*		*Gaussian*		*Pearson*		*Spearman*		*Gaussian*	
ρ_0/θ	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.
Panel A: $N = 50$												
-.25	-.239	.136	-.235	.136	-.242	.135	-.242	.166	-.230	.142	-.237	.147
-.10	-.097	.143	-.094	.142	-.098	.142	-.095	.177	-.091	.150	-.094	.155
-.05	-.047	.144	-.047	.143	-.048	.143	-.050	.175	-.046	.149	-.049	.154
.05	.050	.141	.048	.141	.050	.141	.050	.175	.046	.150	.048	.156
.10	.094	.142	.092	.141	.095	.141	.099	.173	.093	.146	.096	.152
.25	.241	.137	.237	.136	.244	.135	.248	.165	.234	.140	.242	.145
Panel B: $N = 200$												
-.25	-.240	.068	-.238	.068	-.248	.067	-.244	.085	-.234	.070	-.245	.074
-.10	-.096	.071	-.096	.071	-.100	.070	-.099	.090	-.095	.072	-.099	.077
-.05	-.048	.071	-.047	.071	-.049	.071	-.048	.089	-.047	.073	-.049	.077
.05	.048	.071	.047	.071	.049	.071	.050	.091	.047	.074	.049	.078
.10	.096	.071	.095	.071	.099	.070	.100	.091	.094	.073	.099	.078
.25	.240	.068	.238	.067	.247	.067	.248	.086	.235	.070	.246	.074
Panel C: $N = 800$												
-.25	-.239	.034	-.239	.034	-.249	.033	-.245	.043	-.235	.035	-.247	.038
-.10	-.096	.035	-.096	.035	-.100	.035	-.098	.045	-.094	.036	-.099	.039
-.05	-.048	.035	-.047	.036	-.049	.036	-.048	.046	-.047	.037	-.049	.039
.05	.048	.035	.048	.035	.050	.035	.051	.046	.047	.036	.050	.039
.10	.096	.035	.095	.035	.099	.035	.099	.046	.094	.036	.099	.039
.25	.241	.034	.240	.034	.250	.034	.248	.043	.236	.035	.248	.037
Panel D: $N = 3,200$												
-.25	-.240	.017	-.239	.017	-.250	.017	-.246	.021	-.236	.017	-.248	.019
-.10	-.096	.017	-.096	.018	-.100	.018	-.098	.023	-.094	.018	-.100	.020
-.05	-.048	.018	-.048	.018	-.050	.018	-.049	.023	-.047	.018	-.050	.020
.05	.048	.018	.048	.018	.050	.018	.050	.023	.047	.018	.050	.020
.10	.096	.018	.095	.018	.100	.018	.099	.023	.094	.018	.099	.020
.25	.240	.017	.239	.017	.250	.016	.248	.022	.236	.017	.248	.019

Notes: Results based on 10,000 replications. DGP: Asymmetric Laplace margins with location, scale, and shape parameter values 0, 10 and 0.9, respectively (see Online Appendix E for details); Gaussian copula with parameter ρ_0 (left) and Student t copula with 8 degrees of freedom and correlation parameter θ (right). We report the mean and standard deviation of the sampling distribution of the following estimators: *Pearson* denotes the usual Pearson correlation applied to the simulated raw data, *Spearman* refers to $\tilde{\rho}_S^I$ in section 5.1 applied to the empirical cdf of the simulated data, while *Gaussian* refers to the Pearson correlation applied to the Gaussian ranks of the simulated data.

Table 2: Regression parameter estimators

		Gaussian copula						Student t copula					
		OLS		*LTS*		*Gaussian*		*OLS*		*LTS*		*Gaussian*	
ρ_0/θ	β	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.
Panel A: $N = 50$													
-.25	*-.33*	-.321	.152	-.232	.167	-.320	.131	-.329	.176	-.270	.192	-.317	.143
-.10	*-.11*	-.107	.152	-.075	.154	-.106	.143	-.108	.183	-.085	.184	-.105	.155
-.05	*-.05*	-.052	.152	-.034	.152	-.051	.144	-.051	.182	-.041	.182	-.050	.155
.05	*.05*	.048	.152	.031	.152	.047	.144	.048	.183	.038	.184	.046	.155
.10	*.09*	.087	.154	.061	.155	.086	.144	.091	.182	.073	.181	.088	.153
.25	*.20*	.198	.155	.142	.163	.195	.140	.200	.178	.164	.190	.194	.149
Panel B: $N = 200$													
-.25	*-.33*	-.318	.074	-.203	.079	-.330	.064	-.326	.088	-.244	.097	-.326	.071
-.10	*-.11*	-.107	.072	-.065	.071	-.110	.070	-.108	.091	-.077	.089	-.108	.078
-.05	*-.05*	-.051	.073	-.031	.070	-.052	.072	-.051	.091	-.036	.088	-.052	.078
.05	*.05*	.046	.073	.028	.071	.047	.072	.049	.091	.033	.089	.048	.078
.10	*.09*	.088	.073	.054	.072	.090	.071	.090	.092	.064	.090	.089	.078
.25	*.20*	.195	.074	.124	.077	.199	.069	.199	.090	.146	.096	.197	.075
Panel C: $N = 800$													
-.25	*-.33*	-.316	.036	-.195	.037	-.332	.032	-.326	.045	-.234	.048	-.329	.036
-.10	*-.11*	-.107	.036	-.063	.035	-.111	.035	-.109	.046	-.075	.045	-.110	.039
-.05	*-.05*	-.050	.036	-.030	.035	-.052	.035	-.051	.046	-.035	.044	-.052	.039
.05	*.05*	.046	.036	.027	.035	.048	.035	.048	.046	.031	.044	.048	.039
.10	*.09*	.087	.035	.052	.035	.090	.035	.091	.046	.060	.044	.090	.039
.25	*.20*	.194	.036	.118	.037	.200	.034	.199	.046	.140	.047	.199	.038
Panel D: $N = 3,200$													
-.25	*-.33*	-.316	.018	-.194	.018	-.333	.016	-.326	.022	-.233	.024	-.330	.018
-.10	*-.11*	-.106	.018	-.063	.017	-.111	.017	-.109	.023	-.074	.022	-.111	.020
-.05	*-.05*	-.050	.018	-.030	.017	-.052	.018	-.051	.023	-.035	.022	-.052	.020
.05	*.05*	.046	.018	.027	.017	.048	.018	.048	.023	.031	.022	.047	.020
.10	*.09*	.088	.018	.052	.017	.091	.018	.090	.023	.060	.022	.090	.020
.25	*.20*	.194	.018	.118	.018	.200	.017	.198	.023	.138	.023	.199	.019

Notes: Results based on 10,000 replications. DGP: Asymmetric Laplace margins with location, scale, and shape parameter values 0, 10 and 0.9, respectively (see Online Appendix E for details); Gaussian copula with parameter ρ_0 (left) and Student t copula with 8 degrees of freedom and correlation parameter θ (right). We report the mean and standard deviation of the sampling distribution of the following estimators: *OLS* denotes the usual OLS regression applied to the simulated raw data, *LTS* refers to the Least Trimmed Squares, which classifies some observations as unrepresentative and subsequently omits them from the sample (see Rousseeuw 1984, 1985 for details), while *Gaussian* refers to OLS regression applied to the Gaussian ranks of the simulated data.

Table 3: 5% finite sample sizes of F and t tests

		F tests				t tests (two-sided)			
		Gaussian copula		Student t copula		Gaussian copula		Student t copula	
ρ_0/θ	β_0	*OLS*	*Gaussian*	*OLS*	*Gaussian*	*OLS*	*Gaussian*	*OLS*	*Gaussian*
		Panel A: $N = 50$							
-.25	*-.33*	13.8	17.0	16.6	18.7	10.2	11.8	11.8	12.9
-.10	*-.11*	12.2	17.7	15.1	20.5	9.4	12.7	11.3	13.6
-.05	*-.05*	11.5	17.8	15.4	20.0	9.1	12.6	11.2	13.5
.05	*.05*	11.5	18.0	15.3	19.8	9.3	12.5	11.2	13.7
.10	*.09*	11.8	18.3	16.3	19.2	9.4	12.9	11.7	13.1
.25	*.20*	13.5	18.0	16.4	18.8	10.0	12.6	11.4	12.7
		Panel B: $N = 200$							
-.25	*-.33*	7.6	8.4	8.7	9.4	6.6	7.5	7.2	7.7
-.10	*-.11*	6.7	8.8	8.3	9.6	5.9	7.1	7.2	7.6
-.05	*-.05*	6.5	8.9	8.1	8.9	5.9	7.8	7.0	7.3
.05	*.05*	6.6	9.2	8.0	9.2	6.1	7.6	6.9	7.6
.10	*.09*	6.5	8.9	8.1	9.1	6.1	7.3	6.8	7.7
.25	*.20*	7.2	8.7	8.2	8.9	6.4	7.4	7.0	7.5
		Panel C: $N = 800$							
-.25	*-.33*	5.8	5.7	6.2	6.4	5.4	5.7	5.6	6.0
-.10	*-.11*	5.2	6.0	6.0	6.2	5.3	5.6	5.8	5.9
-.05	*-.05*	5.3	6.4	6.1	6.4	5.2	5.9	5.7	5.8
.05	*.05*	5.8	6.3	6.1	6.5	5.6	5.9	5.5	5.7
.10	*.09*	5.3	6.0	6.0	6.0	5.3	5.7	5.8	5.7
.25	*.20*	5.5	6.3	5.9	6.4	5.5	5.9	5.6	5.9
		Panel D: $N = 3,200$							
-.25	*-.33*	4.9	5.3	5.1	5.6	4.9	5.0	5.1	5.4
-.10	*-.11*	4.9	5.4	5.1	5.2	4.9	5.4	5.0	5.3
-.05	*-.05*	5.3	5.3	5.1	5.3	5.1	5.3	5.1	5.2
.05	*.05*	5.2	5.3	5.1	5.3	5.2	5.3	5.2	5.3
.10	*.09*	5.1	5.7	5.4	5.7	5.2	5.5	5.1	5.6
.25	*.20*	4.8	4.9	5.1	5.0	5.0	4.9	5.1	5.2

Notes: Results based on 10,000 replications. DGP: Asymmetric Laplace margins with location, scale, and shape parameter values 0, 10 and 0.9, respectively (see Online Appendix E for details); Gaussian copula with parameter ρ_0 (left) and Student t copula with 8 degrees of freedom and correlation parameter θ (right). The first four columns report the finite sample sizes of the F tests of $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}(\rho_0)$ for the trivariate case when a Gaussian copula is used to generate the data, and the analogous rejection rates of $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_\infty(\theta)$, where $\boldsymbol{\beta}_\infty(\theta)$ is the pseudo-true value of the Gaussian rank regression coefficient vector corresponding to a Student t copula with 8 degrees of freedom and equicorrelation parameter θ for the underlying multivariate distribution. The last four coulumns do the same but for the two sided t tests. For each dependence measure, we substract the pseudo-true value and then standardize using feasible standard error estimates. *OLS* denotes the usual OLS regression applied to the simulated raw data; while *Gaussian* refers to OLS regression applied to the Gaussian ranks of the simulated data.

Table 4: Correlation parameter estimators, contaminated sample

	Pearson			*Spearman*			*Gaussian*		
	Mean	Std.Dev.	$(\hat{\rho}>0)\%$	Mean	Std.Dev.	$(\hat{\rho}>0)\%$	Mean	Std.Dev.	$(\hat{\rho}>0)\%$
ρ_0									
				Panel A: $N = 50$					
-.25	-.758	.052	.000	-.441	.106	.000	-.498	.095	.000
-.10	-.718	.060	.000	-.336	.110	.001	-.401	.100	.000
-.05	-.706	.062	.000	-.302	.109	.002	-.369	.099	.000
.05	-.680	.067	.000	-.234	.109	.014	-.308	.099	.000
.10	-.668	.070	.000	-.200	.110	.033	-.276	.100	.002
.25	-.631	.078	.000	-.099	.105	.178	-.182	.094	.022
				Panel B: $N = 200$					
-.25	-.508	.042	.000	-.293	.064	.000	-.344	.059	.000
-.10	-.424	.046	.000	-.160	.065	.007	-.215	.061	.000
-.05	-.396	.048	.000	-.116	.067	.043	-.172	.063	.003
.05	-.340	.051	.000	-.028	.067	.339	-.086	.063	.086
.10	-.312	.052	.000	.016	.066	.601	-.043	.062	.239
.25	-.228	.057	.000	.149	.063	.992	.085	.059	.921
				Panel C: $N = 800$					
-.25	-.334	.028	.000	-.253	.033	.000	-.282	.032	.000
-.10	-.213	.030	.000	-.112	.034	.001	-.139	.034	.000
-.05	-.172	.030	.000	-.066	.035	.029	-.092	.034	.004
.05	-.092	.031	.002	.028	.035	.785	.003	.034	.537
.10	-.050	.031	.055	.076	.035	.985	.052	.034	.936
.25	.071	.032	.987	.216	.033	1.00	.194	.032	1.00
				Panel D: $N = 3,200$					
-.25	-.267	.016	.000	-.243	.017	.000	-.260	.017	.000
-.10	-.129	.017	.000	-.100	.018	.000	-.112	.017	.000
-.05	-.083	.017	.000	-.052	.018	.002	-.063	.017	.000
.05	.009	.017	.700	.043	.018	.992	.036	.017	.981
.10	.054	.017	.999	.090	.018	1.00	.085	.018	1.00
.25	.193	.016	1.00	.233	.017	1.00	.233	.016	1.00

Notes: Results based on 10,000 replications. DGP: Gaussian copula with parameter ρ_0 and asymmetric Laplace margins with location, scale, and shape parameter values 0, 10 and 0.9, respectively (see Online Appendix E for details) with 5 outliers given by $X_1^{outlier} = (9, 8, 7, 6, 5)' IQR_1$ and $X_2^{outlier} = -.5X_1^{outlier}$. We report the mean, standard deviation and the fraction of parameter estimates that are positive of the sampling distribution of the following estimators: *Pearson* denotes the usual Pearson correlation applied to the simulated raw data, *Spearman* refers to the $\tilde{\rho}_S^I$ in section 5.1 applied to the empirical cdf of the simulated data, while *Gaussian* refers to the Pearson correlation applied to the Gaussian ranks of the simulated data.

Table 5: Growth regressions

Panel A: Mankiw, Romer and Weil (1992) classification

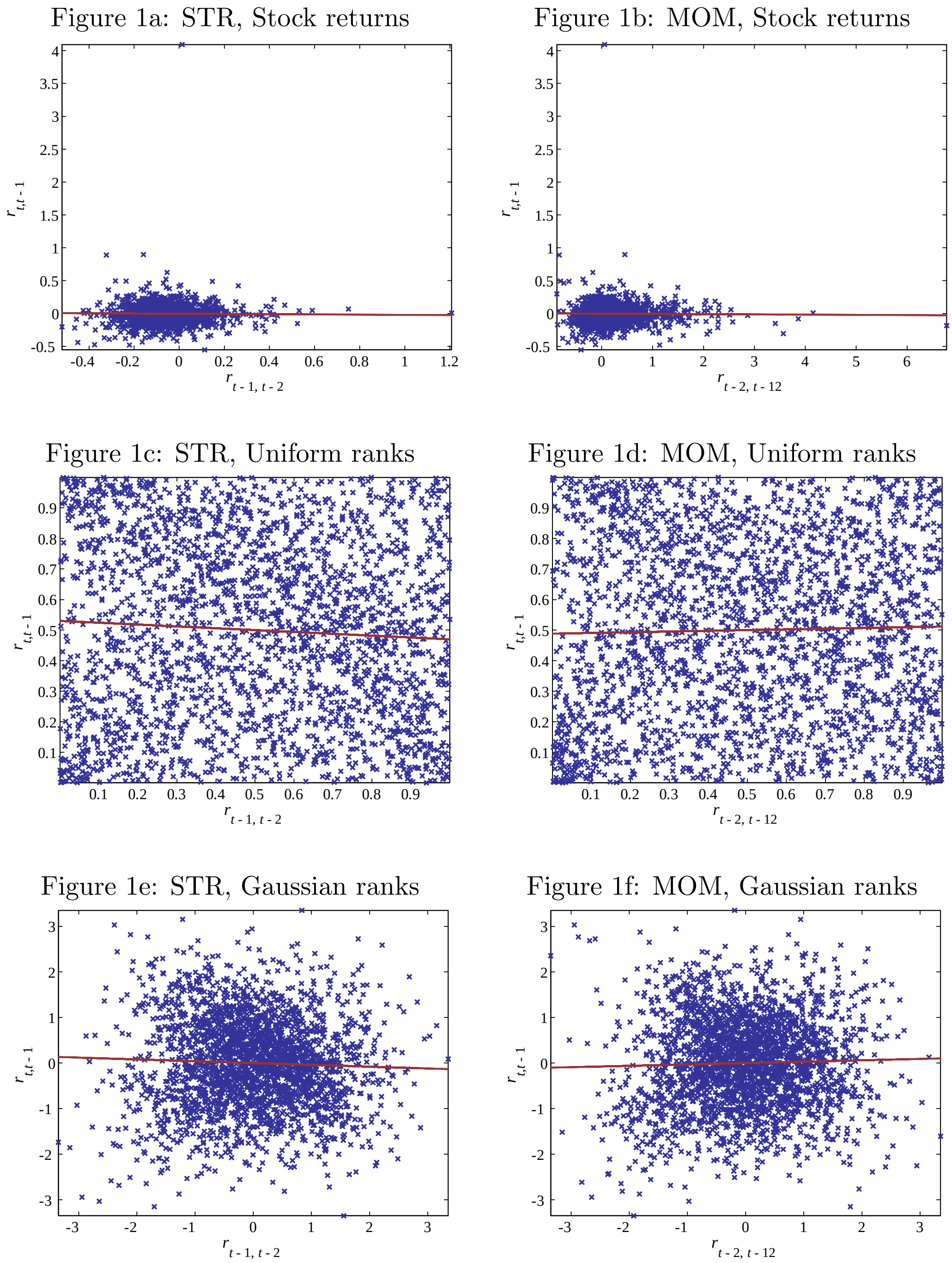
	OLS				LTS				Gaussian			
	Non-oil	Interm.	OECD	All	Non-oil	Interm.	OECD	All	Non-oil	Interm.	OECD	All
Constant	6.84	7.79	8.64	10.10	6.76	7.92	9.40	6.95	8.94	9.34	9.32	12.38
	(1.18)	(1.19)	(2.21)	(1.28)	(1.92)	(2.03)	(6.64)	(2.18)	(1.35)	(1.28)	(1.14)	(2.08)
$\ln(n+g+\delta)$	-1.75	-1.50	-1.08	-0.58	-1.78	-1.52	-1.36	-1.76	-1.24	-1.20	-0.51	0.04
	(0.42)	(0.40)	(0.76)	(0.45)	(0.19)	(0.19)	(1.04)	(0.23)	(0.48)	(0.42)	(0.40)	(0.75)
$\ln(I/GDP)$	0.70	0.70	0.28	0.70	0.60	0.73	0.23	0.63	0.93	0.93	0.31	0.86
	(0.13)	(0.15)	(0.39)	(0.16)	(0.03)	(0.03)	(0.25)	(0.04)	(0.22)	(0.27)	(0.21)	(0.23)
$\ln(SCHOOL)$	0.65	0.73	0.77	0.69	0.69	0.76	1.42	0.71	0.81	0.93	0.35	0.85
	(0.07)	(0.10)	(0.29)	(0.08)	(0.02)	(0.03)	(0.51)	(0.03)	(0.15)	(0.22)	(0.25)	(0.16)
R-squared	0.79	0.78	0.35	0.69	0.78	0.78	0.16	0.67	0.71	0.68	0.24	0.63

Panel B: Temple (1998) classification

	OLS				LTS				Gaussian			
	Poorest	Second	Third	Richest	Poorest	Second	Third	Richest	Poorest	Second	Third	Richest
Constant	8.41	11.12	9.63	8.58	7.64	9.87	9.03	8.78	6.94	12.29	13.77	9.77
	(2.32)	(2.13)	(2.61)	(2.15)	(10.54)	(17.26)	(14.40)	(4.48)	(1.06)	(3.09)	(4.23)	(1.01)
$\ln(n+g+\delta)$	-0.25	0.68	-0.20	-0.95	-0.15	0.15	-0.12	-0.87	-0.34	1.15	1.42	-0.60
	(0.83)	(0.85)	(0.96)	(0.58)	(1.39)	(2.89)	(1.98)	(0.41)	(0.38)	(1.24)	(1.56)	(0.30)
$\ln(I/GDP)$	0.46	0.41	0.86	0.42	0.18	0.46	0.57	0.46	0.30	0.63	0.89	0.43
	(0.19)	(0.34)	(0.37)	(0.44)	(0.04)	(0.24)	(0.23)	(0.22)	(0.12)	(0.36)	(0.45)	(0.29)
$\ln(SCHOOL)$	0.28	0.30	0.06	0.50	0.20	0.29	0.00	0.42	0.10	0.16	0.10	0.56
	(0.10)	(0.20)	(0.31)	(0.28)	(0.02)	(0.07)	(0.18)	(0.21)	(0.05)	(0.14)	(0.25)	(0.30)
R-squared	0.55	0.44	0.46	0.48	0.41	0.43	0.28	0.37	0.34	0.40	0.11	0.44

Notes: We report parameter estimates and standard errors (in parantheses) of the following estimators: *OLS* denotes the usual OLS regression applied to the raw data, *LTS* refers to the Least Trimmed Squares, which classifies some observations as unrepresentative and subsequently omits them from the sample (see Rousseeuw 1984, 1985 for details), while *Gaussian* refers to OLS regression applied to the Gaussian ranks of the data.

Figure 1: Short term reversal and momentum, August 2007



Notes: The data is collected from CRSP. STR refers to short term reversal and MOM to momentum. Red lines in the top panels represent the regression lines of the original data, with beta coefficients: -.019 in Figure 1a and -.004 in Figure 1b; red lines in the middle panels correspond to the Spearman rank correlation: -.062 and .023 in Figures 1c and 1d, respectively; and red lines in the bottom panels represent the Gaussian rank correlation: -.040 in Figure 1e and .030 in Figure 1f.

Figure 2: Gaussian, Clayton and Student copulas with Gaussian margins

Figure 2a: Bivariate Gaussian copula with Gaussian margins

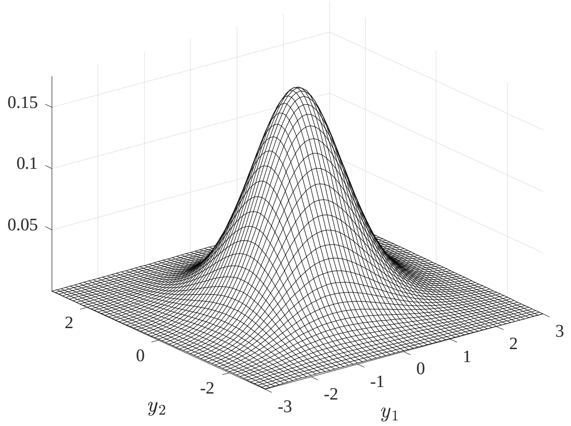


Figure 2b: Contours of a bivariate Gaussian copula with Gaussian margins

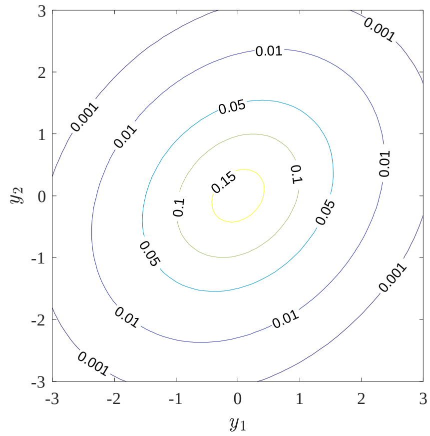


Figure 2c: Bivariate Clayton copula with Gaussian margins

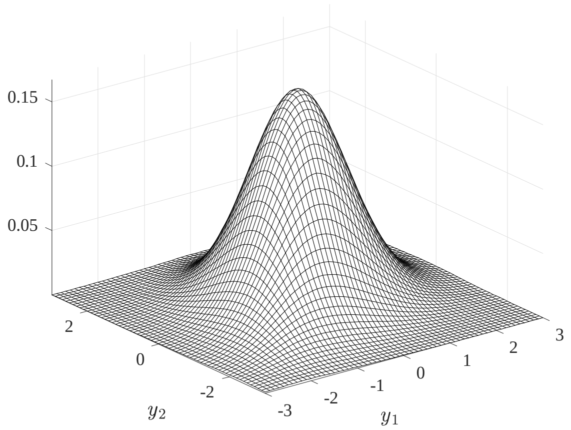


Figure 2d: Contours of a bivariate Clayton copula with Gaussian margins

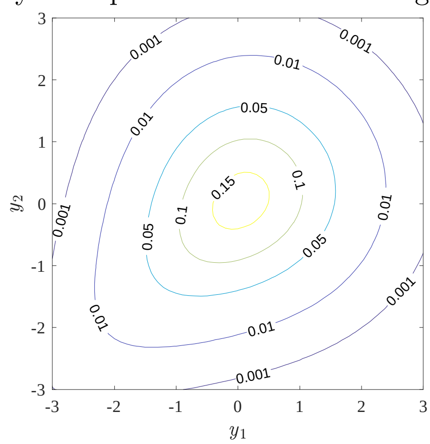


Figure 2e: Bivariate Gaussian copula with Student t margins

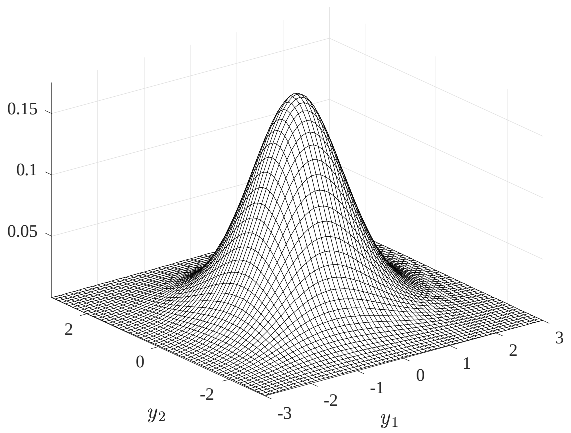
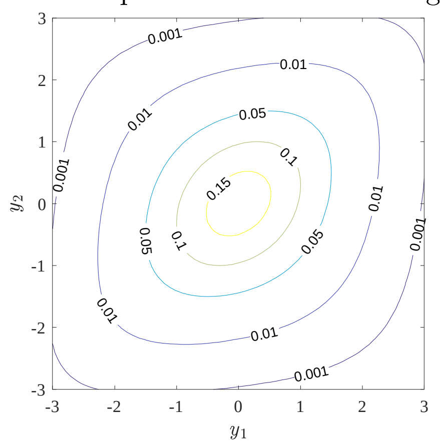


Figure 2f: Contours of a bivariate Student t copula with Gaussian margins



Notes: Figures 2a.b report a Gaussian copula, Figures 2c.d a Clayton copula, and Figures 2e.f a Student t copula with 8 degrees of freedom. All of them are represented with standard normal margins and have been calibrated so that their Gaussian rank correlation coefficient is .25.

Figure 3: Asymptotic variance of Gaussian rank correlation and regression coefficients

Figure 3a: Correlation coefficients

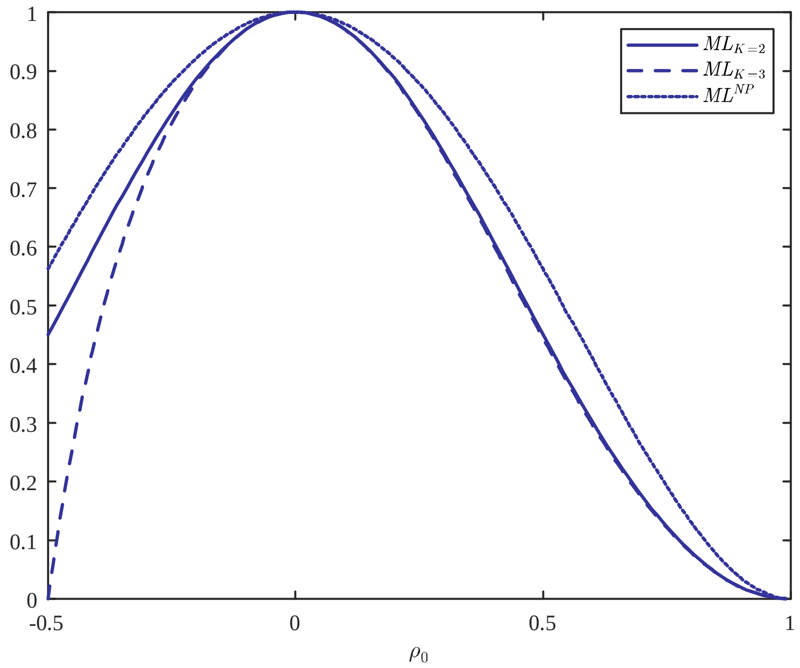
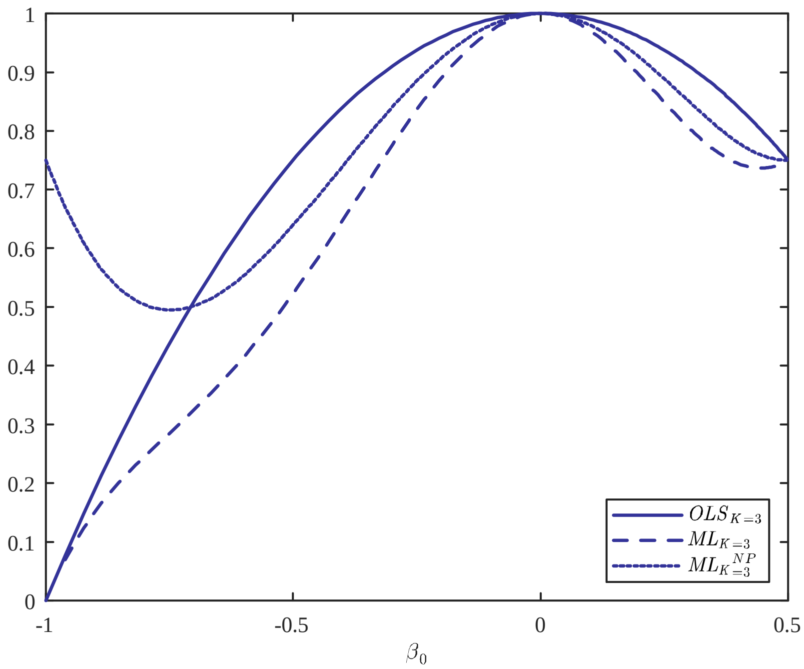


Figure 3b: Regression coefficients



Notes: $ML_{K=2}$ and $ML_{K=3}$ denote the unconstrained ML estimators in both bivariate and trivariate cases assuming known margins; ML^{NP} denotes the unconstrained ML estimators in both bivariate and trivariate contexts when margins are estimated non-parametrically; while OLS denotes the slope coefficients in a multiple linear regression.

Figure 4: Pseudo-true values of the Gaussian rank correlation coefficient

Figure 4a: Clayton copula

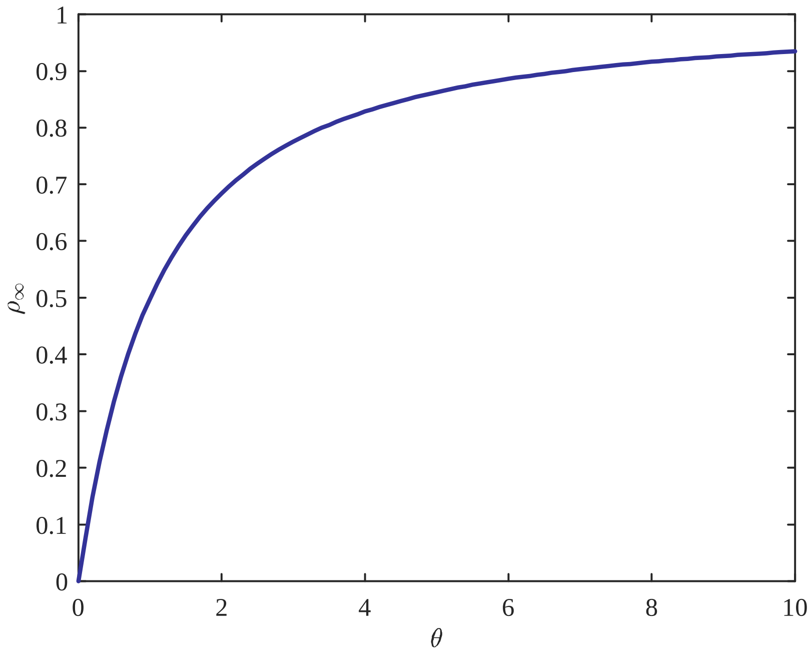
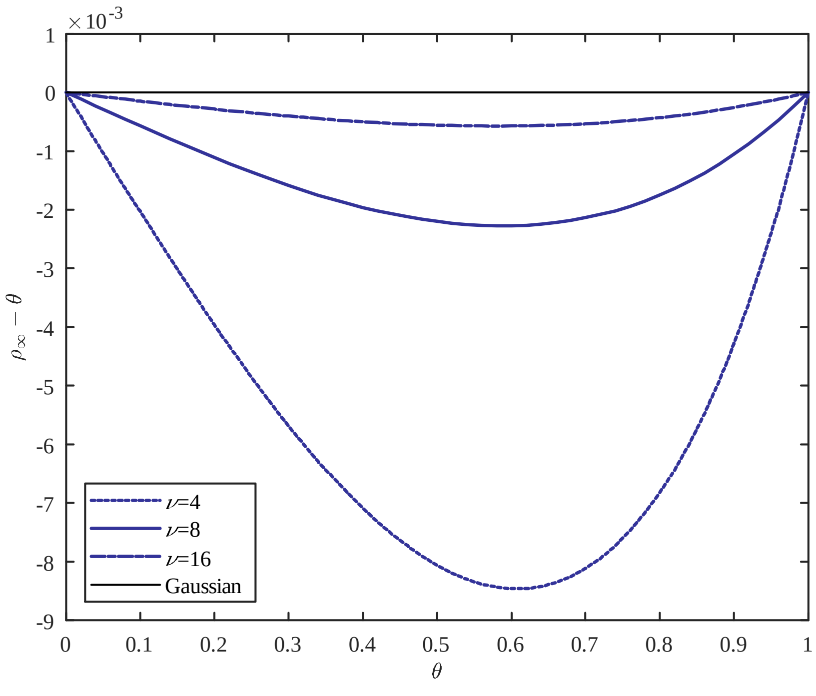


Figure 4b: Student t copula



Notes: ρ_∞ is the pseudo-true value of the Gaussian rank correlation. In Figure 4a, θ denotes the dependence parameter of the Clayton copula, while in Figure 4b it represents the value of the correlation of the bivariate Student t distributions underlying the copulas. We use numerical integration to obtain the relevant cross-moments involved in the Gaussian rank expressions when the true copula is either Clayton or Student t. Computations involving the Clayton copula are done in Mathematica (Cartesian rule) while those for the Student t copula in Matlab (Simpson rule).

Figure 5: Asymptotic variance of the ML estimator of the Gaussian rank correlation

Figure 5a: Clayton copula

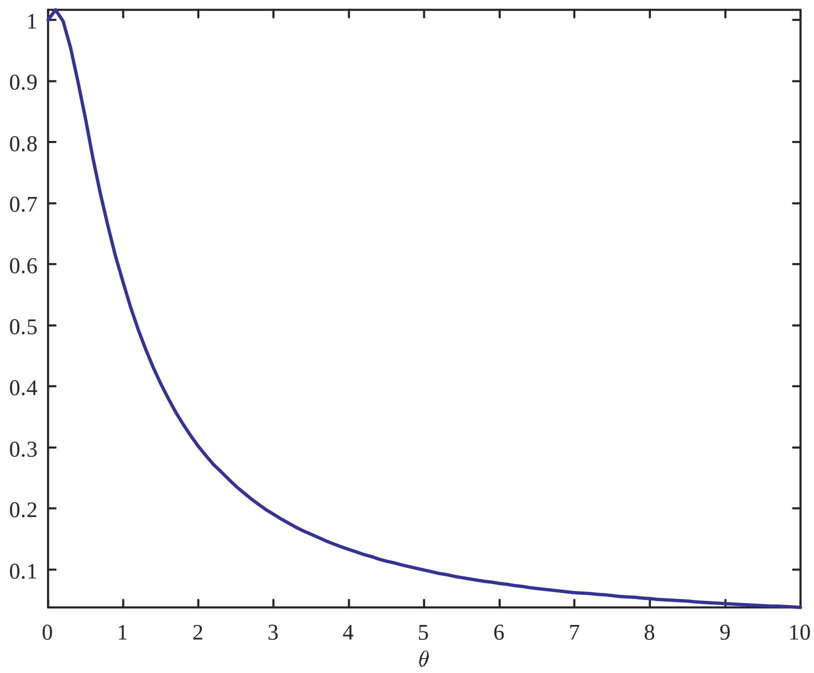
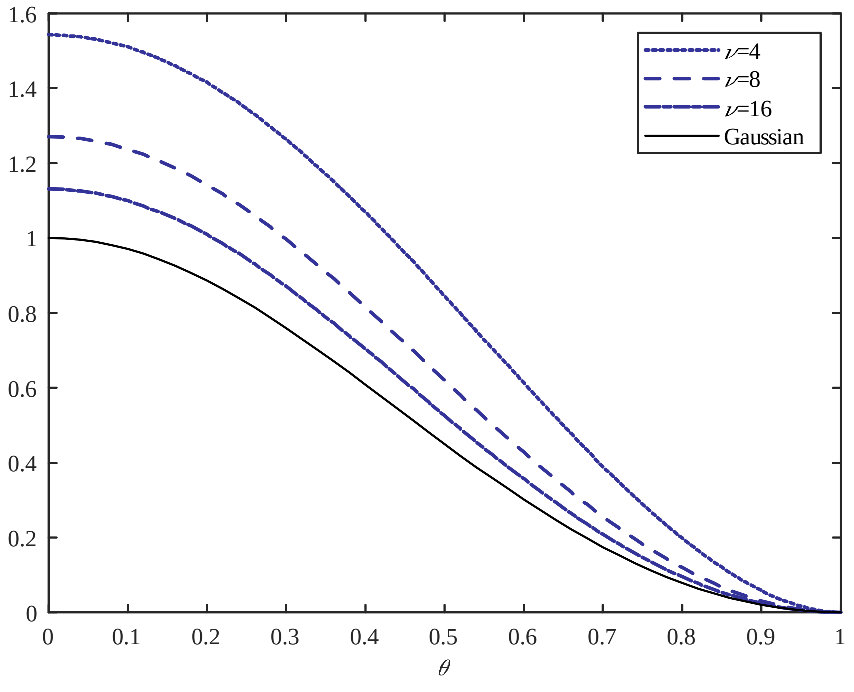


Figure 5b: Student t copula



Notes: In Figure 5a, θ denotes the dependence parameter of the Clayton copula, while in Figure 5b it represents the value of the correlation of the bivariate Student t distributions underlying the copulas. We use numerical integration to obtain the relevant cross-moments involved in the Gaussian rank expressions when the true copula is either Clayton or Student t. Computations involving the Clayton copula are done in Mathematica (Cartesian rule) while those for the Student t copula in Matlab (Simpson rule).

Figure 6: Alternative marginal distributions

Figure 6a: Tukey density

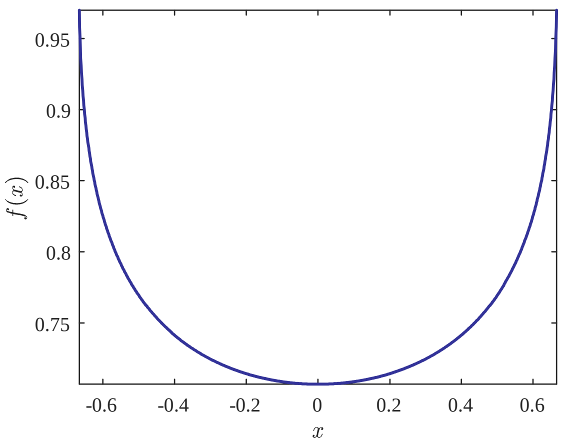


Figure 6b: Asymmetric Laplace density

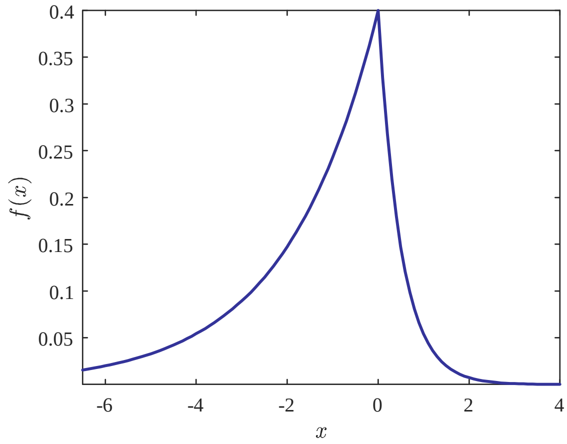


Figure 6c: Weibull density

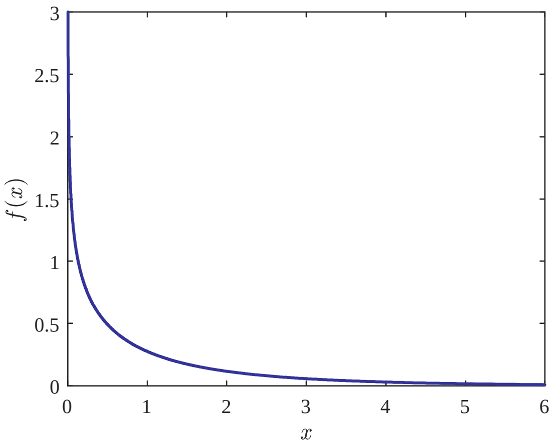
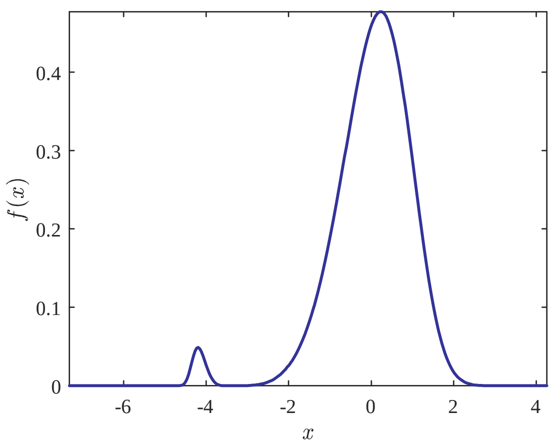


Figure 6d: Mixture of Weibulls density



Notes: Weibull random variable with parameters $k = 0.75$ and $\lambda = 1$. Asymmetric Laplace random variable with parameters $m = 0$, $k = 2$ and $\lambda = 1$. Tukey random variable with parameter $\lambda = 1.5$. Mixture of Weibulls random variable with parameters $k_1 = 5$, $\lambda_1 = 10$, $k_2 = 5$, , $\lambda_1 = 2$ and mixing probability $\alpha = .98$. See Online Appendix E for a description of the marginal distributions.

Figure 7: Pseudo-true values and asymptotic variances of Pearson, Spearman and Gaussian rank correlations

Figure 7a: Pseudo-true values (dicrepancy)

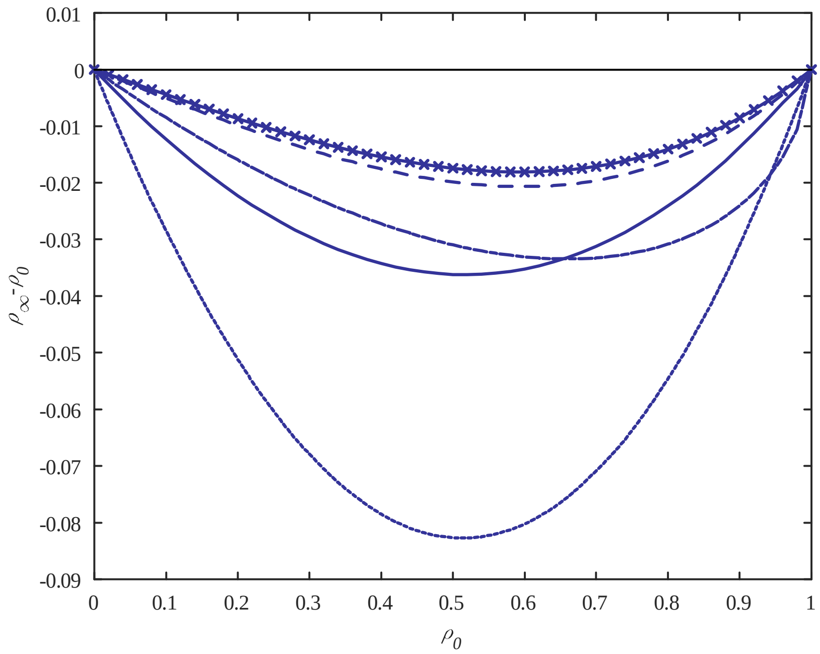
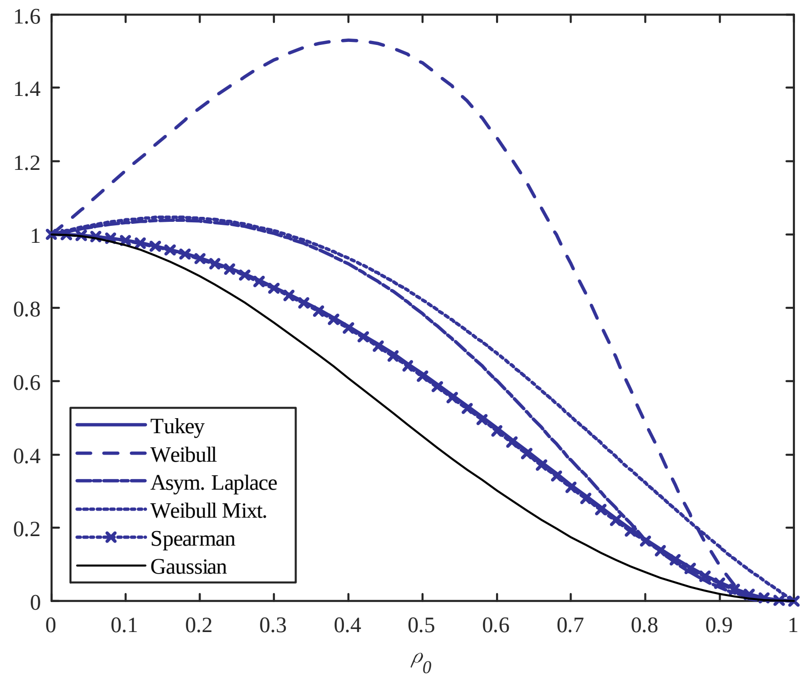


Figure 7b: Asymptotic variance



Notes: ρ_0 denotes the Gaussian copula correlation while ρ_∞ denotes the pseudo-true value of the Spearman and Pearson correlation coefficients. Tukey random variable with parameter $\lambda = 1.5$; Weibull random variable with parameters $k = 0.75$ and $\lambda = 1$; asymmetric Laplace random variable with parameters $m = 0$, $k = 2$ and $\lambda = 1$; mixture of Weibulls random variable with parameters $k_1 = 5$, $\lambda_1 = 10$, $k_2 = 5$, , $\lambda_1 = 2$ and mixing probability $\alpha = .98$. See Online Appendix E for a description of the marginal distributions.

Figure 8: Determinants of migration across US states

Figure 8a: Annual Migration Rate, 1900-1987

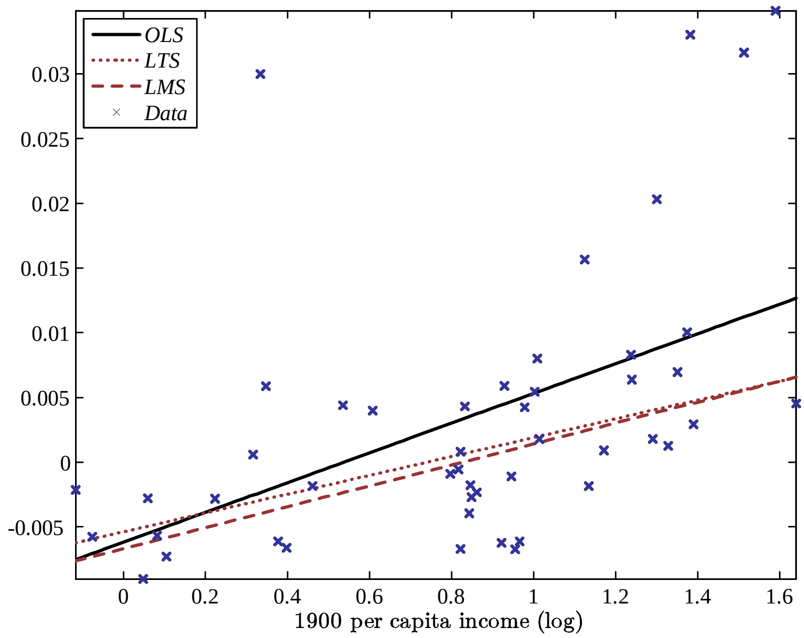
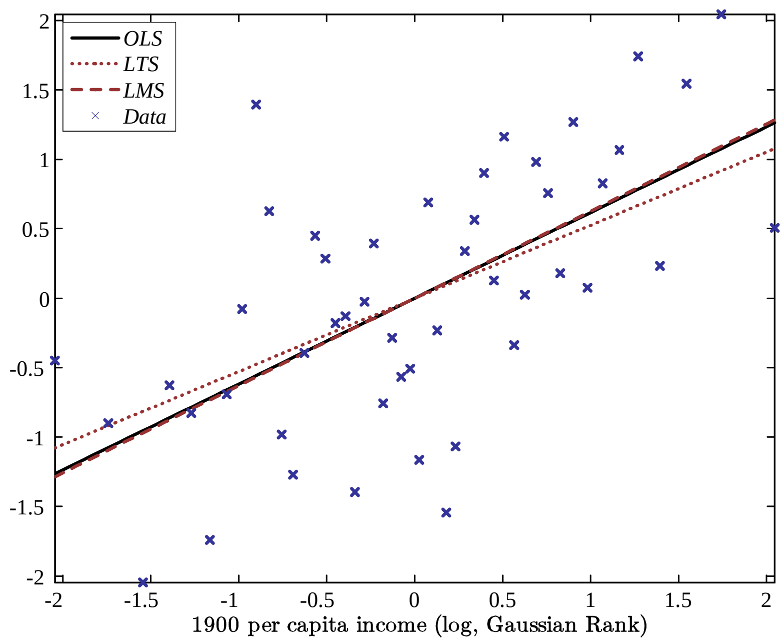


Figure 8b: Annual Migration Rate, 1900-1987 (Gaussian ranks)



Notes: Data set in chapter 11 of Barro and Xala-i-Martin (2003): log of 1900 per capita income on the horizontal axis and the average net migration rate for 48 US states or territories from 1900 to 1990 on the vertical axis. The top panel contains the original raw data while the corresponding Gaussian ranks are plotted in the bottom panel. Estimators: *OLS* denotes the usual OLS regression, *LTS* refers to the Least Trimmed Squares, which classifies some observations as unrepresentative and subsequently omits them from the sample, while *LMS* refers to the Least Median of Squares, which minimizes the median square residuals instead of the mean square residuals (see Rousseeuw 1984, 1985 for details).

Online Appendix for

Gaussian rank correlation and regression

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>

Zhanyuan Tian
Boston University
<zhanyuan@bu.edu>

July 2020

B Proofs

For the sake of brevity, Assumption 1 is maintained throughout.

Proposition 1

Consider the differential of (2),

$$
\begin{aligned}
\mathrm{d}\phi_K(\mathbf{y}_i, \boldsymbol{\rho}) &= -\frac{1}{2} tr\{\mathbf{P}^{-1}(\boldsymbol{\rho})\mathrm{d}[\mathbf{P}(\boldsymbol{\rho})]\} - \frac{1}{2}\mathbf{y}_i'\mathrm{d}[\mathbf{P}^{-1}(\boldsymbol{\rho})]\mathbf{y}_i \\
&= -\frac{1}{2} vec[\mathbf{P}^{-1}(\boldsymbol{\rho})]' vec\{\mathrm{d}[\mathbf{P}(\boldsymbol{\rho})]\} + \frac{1}{2}\mathbf{y}_i'\mathbf{P}^{-1}(\boldsymbol{\rho})\mathrm{d}[\mathbf{P}(\boldsymbol{\rho})]\mathbf{P}^{-1}(\boldsymbol{\rho})\mathbf{y}_i \\
&= -\frac{1}{2} vec[\mathbf{P}^{-1}(\boldsymbol{\rho})]' vec\{\mathrm{d}[\mathbf{P}(\boldsymbol{\rho})]\} + \frac{1}{2}[\mathbf{y}_i'\mathbf{P}^{-1}(\boldsymbol{\rho}) \otimes \mathbf{y}_i'\mathbf{P}^{-1}(\boldsymbol{\rho})] vec\{\mathrm{d}[\mathbf{P}(\boldsymbol{\rho})]\} \\
&= -\frac{1}{2}\{vec[\mathbf{P}^{-1}(\boldsymbol{\rho})]' - \mathbf{y}_i'\mathbf{P}^{-1}(\boldsymbol{\rho}) \otimes \mathbf{y}_i'\mathbf{P}^{-1}(\boldsymbol{\rho})\} vec\{\mathrm{d}[\mathbf{P}(\boldsymbol{\rho})]\} \\
&= -\frac{1}{2}\{vec[\mathbf{P}^{-1}(\boldsymbol{\rho})\mathbf{P}(\boldsymbol{\rho})\mathbf{P}^{-1}(\boldsymbol{\rho})]' - vec[\mathbf{P}^{-1}(\boldsymbol{\rho})\mathbf{y}_i\mathbf{y}_i'\mathbf{P}^{-1}(\boldsymbol{\rho})]'\} vec\{\mathrm{d}[\mathbf{P}(\boldsymbol{\rho})]\} \\
&= \frac{1}{2}\{[\mathbf{P}^{-1}(\boldsymbol{\rho}) \otimes \mathbf{P}^{-1}(\boldsymbol{\rho})] vec[\mathbf{y}_i\mathbf{y}_i' - \mathbf{P}(\boldsymbol{\rho})]\}' vec\{\mathrm{d}[\mathbf{P}(\boldsymbol{\rho})]\}
\end{aligned}
$$

Transposing and simplifying terms, we can get the fist order condition:

$$
\frac{1}{2}\frac{\partial vec'[\mathbf{P}(\boldsymbol{\rho})]}{\partial \boldsymbol{\rho}}[\mathbf{P}^{-1}(\boldsymbol{\rho}) \otimes \mathbf{P}^{-1}(\boldsymbol{\rho})] vec[\mathbf{y}_i\mathbf{y}_i' - \mathbf{P}(\boldsymbol{\rho})] = \mathbf{0}.
$$

Then, using the fact that

$$
vec[\mathbf{P}(\boldsymbol{\rho})] = vec(\mathbf{I}_K) + \tilde{\mathbf{L}}'\boldsymbol{\rho} + \mathbf{K}\tilde{\mathbf{L}}'\boldsymbol{\rho},
$$

so that

$$
\frac{\partial \mathbf{P}(\boldsymbol{\rho})}{\partial \boldsymbol{\rho}'} = \tilde{\mathbf{L}}' + \mathbf{K}\tilde{\mathbf{L}}',
$$

the result follows. □

Proposition 2

Starting from the expression for the score in (3), we can derive Hessian matrix by differencing once again, namely

$$
\begin{aligned}
\mathrm{d}\mathbf{s}_{Ki}(\boldsymbol{\rho}) &= \frac{1}{2}(\tilde{\mathbf{L}} + \tilde{\mathbf{L}}\mathbf{K})[\mathrm{d}\{vec[\mathbf{P}^{-1}(\boldsymbol{\rho})\mathbf{y}_i\mathbf{y}_i'\mathbf{P}^{-1}(\boldsymbol{\rho})]\} - \mathrm{d}\{vec[\mathbf{P}(\boldsymbol{\rho})]\}] \\
&= \frac{1}{2}(\tilde{\mathbf{L}} + \tilde{\mathbf{L}}\mathbf{K})[vec\{\mathrm{d}[\mathbf{P}^{-1}(\boldsymbol{\rho})]\mathbf{y}_i\mathbf{y}_i'\mathbf{P}^{-1}(\boldsymbol{\rho})\} \\
&\quad + vec\{\mathbf{P}^{-1}(\boldsymbol{\rho})\mathbf{y}_i\mathbf{y}_i'\mathrm{d}[\mathbf{P}^{-1}(\boldsymbol{\rho})]\} - \mathrm{d}\{vec[\mathbf{P}^{-1}(\boldsymbol{\rho})]\}] \\
&= \frac{1}{2}(\tilde{\mathbf{L}} + \tilde{\mathbf{L}}\mathbf{K})[-vec\{\mathbf{P}^{-1}(\boldsymbol{\rho})\mathrm{d}[\mathbf{P}(\boldsymbol{\rho})]\mathbf{P}^{-1}(\boldsymbol{\rho})\mathbf{y}_i\mathbf{y}_i'\mathbf{P}^{-1}(\boldsymbol{\rho})\} \\
&\quad - vec\{\mathbf{P}^{-1}(\boldsymbol{\rho})\mathbf{y}_i\mathbf{y}_i'\mathbf{P}^{-1}(\boldsymbol{\rho})\mathrm{d}[\mathbf{P}(\boldsymbol{\rho})]\mathbf{P}^{-1}(\boldsymbol{\rho})\} + vec\{\mathbf{P}^{-1}(\boldsymbol{\rho})\mathrm{d}[\mathbf{P}(\boldsymbol{\rho})]\mathbf{P}^{-1}(\boldsymbol{\rho})\}] \\
&= \frac{1}{2}(\tilde{\mathbf{L}} + \tilde{\mathbf{L}}\mathbf{K})\{-[\mathbf{P}^{-1}(\boldsymbol{\rho})\mathbf{y}_i\mathbf{y}_i'\mathbf{P}^{-1}(\boldsymbol{\rho})] \otimes \mathbf{P}^{-1}(\boldsymbol{\rho}) \\
&\quad - \mathbf{P}^{-1}(\boldsymbol{\rho}) \otimes [\mathbf{P}^{-1}(\boldsymbol{\rho})\mathbf{y}_i\mathbf{y}_i'\mathbf{P}^{-1}(\boldsymbol{\rho})] + \mathbf{P}^{-1}(\boldsymbol{\rho}) \otimes \mathbf{P}^{-1}(\boldsymbol{\rho})\} vec\{\mathrm{d}[\mathbf{P}(\boldsymbol{\rho})]\}.
\end{aligned}
$$

Hence, $\mathrm{d}\mathbf{s}_{Ki}(\boldsymbol{\rho})$ can be written as in (6) after noticing that $vec\{\mathrm{d}[\mathbf{P}(\boldsymbol{\rho})]\} = (\tilde{\mathbf{L}} + \tilde{\mathbf{L}}\mathbf{K})$. □

Proposition 3

Part (a) follows from the *i.i.d.* assumption on $\{\mathbf{y}_i\}$ together with either expression (10) or (11), while part (b) is a direct application of the Delta method to the result in part (a). $\square$

Lemma 1

As in Chen and Fan (2006), we need to compute

$$n_2 = \int_0^1 [1\{u_1 \leq U_1\} - u_1] W^1_{\rho_{12}} dU_1 + \int_0^1 [1\{u_2 \leq U_2\} - u_2] W^2_{\rho_{12}} dU_2,$$

with $W^j_{\rho_{12}} = \int [\partial s_{\rho_{12}}(u_1, u_2; \rho_{12})/\partial u_j] c(u_1, u_2; \rho_{12}) du_j$ for $j = 1, 2$. Then, the result follows from

$$W^j_{\rho_{12}} = \int \left[\frac{1+\rho_{12}^2}{(1-\rho_{12}^2)^2} y_{-j} - \frac{2\rho_{12}}{(1-\rho_{12}^2)^2} y_j \right] \phi(y_j) dy_{-j} = \frac{1+\rho_{12}^2}{(1-\rho_{12}^2)^2} y_j$$

and the fact that

$$\int_{-\infty}^{y} H_1(x)\Phi(x)dx = \frac{H_1(y)}{2}\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) + \frac{1}{2\sqrt{2}} H_2(y) \left[1 + \mathrm{erf}\left(\frac{y}{\sqrt{2}}\right)\right]$$

and

$$\int_{y}^{\infty} H_1(x)[1-\Phi(x)]dx = \frac{H_1(y)}{2}\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) - \frac{1}{2\sqrt{2}} H_2(y) \left[\mathrm{erfc}\left(\frac{y}{\sqrt{2}}\right)\right],$$

where $\mathrm{erf}(z) = 2\pi^{-1/2}\int_0^z e^{-t^2} dt$ and $\mathrm{erfc}(z) = 1 - \mathrm{erf}(z)$. Analogous calculations in the trivariate case yield the corresponding correction. $\square$

Proposition 4

Under the maintained assumption of a Gaussian copula, it is straightforward to obtain the variance of the score in (4) using the moments of the bivariate normal, whose reciprocal is $AVar(\hat{\rho})$.

To obtain the asymptotic variance of $\check{\rho} = \sum_i y_{1i}y_{2i}/\sum_i y_{2i}^2$, consider the following vector of influence functions:

$$\check{\mathbf{m}}_{2i}(\boldsymbol{\theta}) = (y_{1i}y_{2i} - \sigma_2^2\rho, y_{2i}^2 - \sigma_2^2)'$$

where $\boldsymbol{\theta} = (\rho, \sigma_2^2)'$. Then, we can easily compute

$$\check{\mathbf{A}}_2 = E\left[\frac{\partial \check{\mathbf{m}}_{2i}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right] = -\begin{pmatrix} \sigma_2^2 & \rho \\ 0 & 1 \end{pmatrix} \text{ and } \check{\mathbf{B}}_2 = Var[\check{\mathbf{m}}_{2i}(\boldsymbol{\theta})] = \begin{pmatrix} 1+\rho^2 & 2\rho \\ 2\rho & 2 \end{pmatrix},$$

so that imposing $\sigma^2 = 1$ and applying the sandwich formula yields $Var(\check{\rho}) = (1-\rho^2)^2/(1+\rho^2)$ as the $(1,1)$ element of $\check{\mathbf{A}}_2^{-1}\check{\mathbf{B}}_2\check{\mathbf{A}}_2^{-1\prime}$.

As for $\tilde{\rho} = \sum_i (y_{1i} - \bar{y}_1)(y_{2i} - \bar{y}_2)/\sum_i y_{2i}^2$, where $\bar{y}_j = N^{-1}\sum_i y_{ji}$, we consider the following alternative vector of influence functions:

$$\tilde{\mathbf{m}}_{2i}(\boldsymbol{\theta}) = [y_{1i}y_{2i} - (\mu_1\mu_2 + \sigma_2^2\rho), y_{2i}^2 - (\mu_2^2 + \sigma_2^2), y_{1i} - \mu_1, y_{2i} - \mu_2]'$$

where $\boldsymbol{\theta} = (\rho, \sigma_2^2, \mu_1, \mu_2)'$. Then, we can compute

$$\tilde{\mathbf{A}}_2 = E\left[\frac{\partial \tilde{\mathbf{m}}_{2i}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right] = -\begin{pmatrix} \sigma_2^2 & \rho & \mu_2 & \mu_1 \\ 0 & 1 & 0 & 2\mu_2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{and } \tilde{\mathbf{B}}_2 = Var\left[\tilde{\mathbf{m}}_{2i}(\boldsymbol{\theta})\right] = \begin{pmatrix} \check{\mathbf{B}}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{P}(\rho) \end{pmatrix},$$

so that imposing the Gaussian copula assumption and applying the sandwich formula yields $Var(\tilde{\rho}) = (1-\rho^2)^2$ as the $(1,1)$ element of $\tilde{\mathbf{A}}_2^{-1}\tilde{\mathbf{B}}_2\tilde{\mathbf{A}}_2^{-1\prime}$. □

Proposition 5

Analogous calculations to the ones used in the proof of Lemma 1 allow us to obtain

$$n_{\mu_j i}(\rho) = H_1(y_{ji}) \quad \text{and} \quad n_{\sigma_j^2 i}(\rho) = \sqrt{2}H_2(y_{ji}) \quad \text{for} \quad j = 1, 2.$$

Hence, the asymptotic variance of the ML estimator of ρ can be obtained as

$$AVar(\hat{\rho}^{np}) = \frac{Var[s_{\rho i}^{np}(\rho)]}{\{Var[s_{\rho i}(\rho)]\}^2} = (1-\rho^2)^2.$$

As for the other estimators, letting $\mathbf{B}_2^{np} = Var[\mathbf{m}_{2i}(\boldsymbol{\theta}) + \mathbf{n}_{2i}(\boldsymbol{\theta})]$, we can show that $AVar(\hat{\rho}^{np})$ coincides with the common asymptotic variance of $\check{\rho}^{np}$ and $\tilde{\rho}^{np}$, which is given by

$$AVar(\tilde{\rho}^{np}) = AVar(\check{\rho}^{np}) = (1-\rho^2)^2$$

because $\check{\mathbf{B}}_2^{np}$ and $\tilde{\mathbf{B}}_2^{np}$ have all the elements equal to zero except the $(1,1)$ one, which is equal to $(1-\rho^2)$. □

Proposition 6

First, we can obtain the asymptotic variance of $\hat{\boldsymbol{\rho}}$ as $AVar(\hat{\boldsymbol{\rho}}) = \mathcal{A}^{-1}(\boldsymbol{\rho})$, where the expressions for the expected (minus) Hessian $\mathcal{A}(\boldsymbol{\rho})$ are reported in Online Appendix C. Then, regarding the ML estimator $\hat{\boldsymbol{\beta}}^{(1)}$, we can exploit

$$\hat{\boldsymbol{\beta}}^{(1)} = \begin{bmatrix} 1 & \hat{\rho}_{23} \\ \hat{\rho}_{23} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\rho}_{12} \\ \hat{\rho}_{13} \end{bmatrix}$$

to obtain the asymptotic variance of say $\hat{\beta}_2^{(1)}$ by applying the Delta method, namely

$$AVar(\hat{\beta}_2^{(1)}) = \nabla\hat{\beta}_2^{(1)\prime}(\boldsymbol{\rho})AVar(\hat{\boldsymbol{\rho}})\nabla\hat{\beta}_2^{(1)}(\boldsymbol{\rho})$$

where

$$\nabla\hat{\beta}_2^{(1)}(\boldsymbol{\rho}) = \left[\frac{1}{1-\rho_{23}^2}, -\frac{\rho_{23}}{1-\rho_{23}^2}, \frac{2\rho_{12}\rho_{23} - \rho_{13}(1+\rho_{23}^2)}{(1-\rho_{23}^2)^2}\right]'.$$

This yields

$$\begin{aligned} AVar(\hat{\beta}_2^{(1)}) &= -(((-1+\rho_{23}^2)^3(1+\rho_{23}^2)+\rho_{12}^6(1+3\rho_{23}^2)-2\rho_{12}^5\rho_{13}\rho_{23}(5+7\rho_{23}^2) \\ &\quad +\rho_{13}^6(-1+3\rho_{23}^2+2\rho_{23}^4)+\rho_{13}^2(-1+\rho_{23}^2)^2(1+3\rho_{23}^2+2\rho_{23}^4) \\ &\quad +\rho_{13}^4(1-6\rho_{23}^2+\rho_{23}^4+4\rho_{23}^6)-2\rho_{12}\rho_{13}\rho_{23}(3(-1+\rho_{23}^2)^2(1+\rho_{23}^2) \\ &\quad +4\rho_{13}^2(-1+\rho_{23}^2)(1+\rho_{23}^2)^2+\rho_{13}^4(1+7\rho_{23}^2+4\rho_{23}^4)) \\ &\quad -4\rho_{12}^3\rho_{13}\rho_{23}(4(-1+\rho_{23}^4)+\rho_{13}^2(4+11\rho_{23}^2+5\rho_{23}^4)) \\ &\quad +\rho_{12}^4(-3-2\rho_{23}^2+5\rho_{23}^4+3\rho_{13}^2(1+11\rho_{23}^2+8\rho_{23}^4)) \\ &\quad +\rho_{12}^2(3(-1+\rho_{23}^2)^2(1+\rho_{23}^2)+\rho_{13}^4(1+25\rho_{23}^2+26\rho_{23}^4+8\rho_{23}^6)+ \\ &\quad 2\rho_{13}^2(-2-11\rho_{23}^2+4\rho_{23}^4+9\rho_{23}^6))) \\ &\quad /((-1+\rho_{23}^2)^3(-1+\rho_{12}^4+\rho_{13}^4-2\rho_{12}^2\rho_{13}^2\rho_{23}^2+\rho_{23}^4))). \end{aligned}$$

In turn, to obtain the asymptotic variance of

$$\breve{\boldsymbol{\beta}}^{(1)} = \begin{pmatrix} \sum_i y_{2i}^2 & \sum_i y_{2i}y_{3i} \\ \sum_i y_{2i}y_{3i} & \sum_i y_{3i}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_i y_{1i}y_{2i} \\ \sum_i y_{1i}y_{3i} \end{pmatrix}$$

we consider the following vector of influence functions:

$$\breve{\mathbf{m}}_{3i}(\boldsymbol{\theta}) = (y_{1i}y_{2i}-\sqrt{\sigma_1^2\sigma_2^2}\rho_{12}, y_{1i}y_{3i}-\sqrt{\sigma_1^2\sigma_3^2}\rho_{13}, y_{2i}y_{3i}-\sqrt{\sigma_2^2\sigma_3^2}\rho_{23}, y_{1i}^2-\sigma_1^2, y_{2i}^2-\sigma_2^2, y_{3i}^2-\sigma_3^2)'$$

where $\boldsymbol{\theta} = (\rho_{12}, \rho_{13}, \rho_{23}, \sigma_1^2, \sigma_2^2, \sigma_2^3)'$. Then, under the assumption of a Gaussian copula we will have

$$\breve{\mathbf{A}}_3 = E\left[\frac{\partial \breve{\mathbf{m}}_{3i}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right] = -\begin{pmatrix} 1 & 0 & 0 & \rho_{12}/2 & \rho_{12}/2 & 0 \\ 0 & 1 & 0 & \rho_{13}/2 & 0 & \rho_{13}/2 \\ 0 & 0 & 1 & 0 & \rho_{23}/2 & \rho_{23}/2 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

and

$$\breve{\mathbf{B}}_3 = Var[\breve{\mathbf{m}}_{3i}(\boldsymbol{\theta})] = \begin{pmatrix} 1+\rho_{12}^2 & \rho_{12}\rho_{13}+\rho_{23} & \rho_{12}\rho_{23}+\rho_{13} & 2\rho_{12} & 2\rho_{12} & 2\rho_{13}\rho_{23} \\ \rho_{12}\rho_{13}+\rho_{23} & 1+\rho_{13}^2 & \rho_{13}\rho_{23}+\rho_{12} & 2\rho_{13} & 2\rho_{12}\rho_{23} & 2\rho_{13} \\ \rho_{12}\rho_{23}+\rho_{13} & \rho_{13}\rho_{23}+\rho_{12} & 1+\rho_{23}^2 & 2\rho_{12}\rho_{13} & 2\rho_{23} & 2\rho_{23} \\ 2\rho_{12} & 2\rho_{13} & 2\rho_{12}\rho_{13} & 3 & 2\rho_{12}^2 & 2\rho_{13}^2 \\ 2\rho_{12} & 2\rho_{12}\rho_{23} & 2\rho_{23} & 2\rho_{12}^2 & 3 & 2\rho_{23}^2 \\ 2\rho_{13}\rho_{23} & 2\rho_{13} & 2\rho_{23} & 2\rho_{13}^2 & 2\rho_{23}^2 & 3 \end{pmatrix},$$

which allow us to obtain $AVar(\breve{\boldsymbol{\theta}})$ as $\breve{\mathbf{A}}_3^{-1}\breve{\mathbf{B}}_3\breve{\mathbf{A}}_3^{-1\prime}$. We can then use the Delta method to obtain the asymptotic variance of $\breve{\beta}_2^{(1)}$. Specifically, we have

$$\nabla\breve{\beta}_2^{(1)}(\boldsymbol{\theta}) = \left[\frac{1}{1-\rho_{23}^2}, \frac{-\rho_{23}}{1-\rho_{23}^2}, \frac{2\rho_{12}\rho_{23}-\rho_{13}-\rho_{13}\rho_{23}^2}{(1-\rho_{23}^2)^2}, \frac{\rho_{12}-\rho_{13}\rho_{23}}{2(1-\rho_{23}^2)}, -\frac{\rho_{12}-\rho_{13}\rho_{23}}{2(1-\rho_{23}^2)}, 0\right]',$$

and therefore

$$AVar(\breve{\beta}_2^{(1)}) = \frac{1-\rho_{12}^2-\rho_{13}^2+2\rho_{12}\rho_{13}\rho_{23}-\rho_{23}^2}{(1-\rho_{23}^2)^2}.$$

Finally, to obtain the asymptotic variance of

$$\tilde{\boldsymbol{\beta}}^{(1)} = \begin{pmatrix} N^{-1}\sum_i y_{2i}^2 - \bar{y}_2^2 & N^{-1}\sum_i y_{2i}y_{3i} - \bar{y}_2\bar{y}_3 \\ N^{-1}\sum_i y_{2i}y_{3i} - \bar{y}_2\bar{y}_3 & N^{-1}\sum_i y_{3i}^2 - \bar{y}_3^2 \end{pmatrix}^{-1} \begin{pmatrix} N^{-1}\sum_i y_{1i}y_{2i} - \bar{y}_1\bar{y}_2 \\ N^{-1}\sum_i y_{1i}y_{3i} - \bar{y}_1\bar{y}_3 \end{pmatrix}$$

we consider the following vector of influence functions:

$$\tilde{\mathbf{m}}_{3i}(\boldsymbol{\theta}) = \begin{pmatrix} y_{1i}y_{2i} - \mu_1\mu_2 - \sqrt{\sigma_1^2\sigma_2^2}\rho_{12} \\ y_{1i}y_{3i} - \mu_1\mu_3 - \sqrt{\sigma_1^2\sigma_3^2}\rho_{13} \\ y_{2i}y_{3i} - \mu_2\mu_3 - \sqrt{\sigma_2^2\sigma_3^2}\rho_{23} \\ y_{1i} - \mu_1 \\ y_{2i} - \mu_2 \\ y_{3i} - \mu_3 \\ y_{1i}^2 - (\mu_1^2 + \sigma_1^2) \\ y_{2i}^2 - (\mu_2^2 + \sigma_2^2) \\ y_{3i}^2 - (\mu_3^2 + \sigma_3^2) \end{pmatrix}$$

where $\boldsymbol{\theta} = (\rho_{12}, \rho_{13}, \rho_{23}, \sigma_1^2, \sigma_2^2, \sigma_2^3, \mu_1, \mu_2, \mu_3)'$. Then, under the assumption of a Gaussian copula we will have

$$\tilde{\mathbf{A}}_3 = E\left[\frac{\partial \tilde{\mathbf{m}}_{3i}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right] = \begin{pmatrix} \check{\mathbf{A}}_3 & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_3 \end{pmatrix} \text{ and } \tilde{\mathbf{B}}_3 = Var[\tilde{\mathbf{m}}_{3i}(\boldsymbol{\theta})] = \begin{pmatrix} \check{\mathbf{A}}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{P}(\boldsymbol{\rho}) \end{pmatrix},$$

which allow us to obtain $AVar(\tilde{\boldsymbol{\theta}})$ as $\tilde{\mathbf{A}}_3^{-1}\tilde{\mathbf{B}}_3\tilde{\mathbf{A}}_3^{-1\prime}$. We can then use the Delta method to obtain the asymptotic variance of $\tilde{\beta}_2^{(1)}$. Specifically, we have

$$\nabla\tilde{\beta}_2^{(1)}(\boldsymbol{\theta}) = \left[\frac{1}{1-\rho_{23}^2}, \frac{-\rho_{23}}{1-\rho_{23}^2}, \frac{2\rho_{12}\rho_{23} - \rho_{13} - \rho_{13}\rho_{23}^2}{(1-\rho_{23}^2)^2}, \frac{\rho_{12} - \rho_{13}\rho_{23}}{2(1-\rho_{23}^2)}, -\frac{\rho_{12} - \rho_{13}\rho_{23}}{2(1-\rho_{23}^2)}, 0, 0, 0, 0\right]',$$

and therefore

$$AVar(\tilde{\beta}_2^{(1)}) = \frac{1 - \rho_{12}^2 - \rho_{13}^2 + 2\rho_{12}\rho_{13}\rho_{23} - \rho_{23}^2}{(1-\rho_{23}^2)^2},$$

as desired. □

Proposition 7

We first compute the variance of the ML correlation estimator by using the correction for the trivariate case given in Lemma 1. Specifically, the resulting diagonal elements for the variance of the corrected scores are

$$V_{11} = \mathcal{V}_i^c(\rho_{12}, \rho_{13}, \rho_{23}),\ V_{22} = \mathcal{V}_i^c(\rho_{13}, \rho_{12}, \rho_{23}) \text{ and } V_{33} = \mathcal{V}_i^c(\rho_{23}, \rho_{12}, \rho_{13}),$$

where

$$\mathcal{V}_i^c(\rho_{12}, \rho_{13}, \rho_{23}) = \frac{1 + 2\rho_{12}^2 + \rho_{12}^4 - \rho_{13}^2 - 4\rho_{12}\rho_{13}\rho_{23} - 2\rho_{12}^3\rho_{13}\rho_{23} - \rho_{23}^2 + 3\rho_{13}^2\rho_{23}^2 + \rho_{12}^2\rho_{13}^2\rho_{23}^2}{(1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23})^2}.$$

In turn, the corresponding off-diagonal elements are

$$V_{12} = \mathcal{V}_{ij}^c(\rho_{12}, \rho_{13}, \rho_{23}),\ V_{13} = \mathcal{V}_{ij}^c(\rho_{13}, \rho_{23}, \rho_{12}) \text{ and } V_{23} = \mathcal{V}_{ij}^c(\rho_{23}, \rho_{12}, \rho_{13}),$$

with

$$\begin{aligned} V_{ij}^c(\rho_{12}, \rho_{13}, \rho_{23}) = {} & [5\rho_{12}\rho_{13} + \rho_{12}^3\rho_{13} + \rho_{12}\rho_{13}^3 - 2\rho_{23} - 3\rho_{12}^2\rho_{23} - \rho_{12}^4\rho_{23} - 3\rho_{13}^2\rho_{23} \\ & -2\rho_{12}^2\rho_{13}^2\rho_{23} - \rho_{13}^4\rho_{23} + 2\rho_{12}\rho_{13}\rho_{23}^2 + \rho_{12}^3\rho_{13}\rho_{23}^2 + \rho_{12}\rho_{13}^3\rho_{23}^2 + 2\rho_{23}^3 \\ & -\rho_{12}^2\rho_{23}^3 - \rho_{13}^2\rho_{23}^3 + \rho_{12}\rho_{13}\rho_{23}^4]/[2(1 - \rho_{12}^2 - \rho_{13}^2 + 2\rho_{12}\rho_{13}\rho_{23} - \rho_{23}^2)^2]. \end{aligned}$$

These quantities, together with the expressions for the expected (minus) Hessian in Online Appendix C, allow us to compute the corrected asymptotic variance of the ML estimators via the usual sandwich formula $\mathcal{H}(\boldsymbol{\rho})^{-1}\mathcal{V}^c(\boldsymbol{\rho})\mathcal{H}(\boldsymbol{\rho})^{-1}$.

As for the moment-based estimators, we can also correct the corresponding moment conditions using the following terms:

$$n_{\mu_j i}(\boldsymbol{\theta}) = -H_1(y_{ji}) \;\; \text{and} \;\; n_{\sigma_j^2 i}(\boldsymbol{\theta}) = -\sqrt{2}H_2(y_{ji}) \;\; \text{for} \;\; j = 1, 2, 3$$

and

$$n_{\sigma_{jh} i}(\boldsymbol{\theta}) = -\frac{1}{2}(y_{ji}^2 + y_{hi}^2 - 2)\rho_{hj}, \;\text{for}\; h = 1, 2, 3, \;\text{and}\; h \neq j.$$

As in the bivariate case, if we define $\mathbf{B}_3^{np} = Var[\mathbf{m}_{3i}(\boldsymbol{\theta}) + \mathbf{n}_{3i}(\boldsymbol{\theta})]$, then we will have

$$\check{\mathbf{B}}_3^{np} = \begin{pmatrix} r_{12} & r_{123} & r_{132} & \mathbf{0} \\ r_{123} & r_{13} & r_{231} & \mathbf{0} \\ r_{132} & r_{231} & r_{23} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \;\; \text{and} \;\; \tilde{\mathbf{B}}_3^{np} = \begin{pmatrix} \check{\mathbf{B}}_3^{np} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

where

$$r_{jh} = (1 - \rho_{jh}^2)^2 \;\; \text{and} \;\; r_{jhk} = \frac{1}{2}[\rho_{jh}^3\rho_{jk} - 2\rho_{jh}^2\rho_{hk} + 2(1 - \rho_{jk}^2)\rho_{hk} + \rho_{jh}\rho_{jk}(\rho_{jk}^2 + \rho_{hk}^2 - 1)].$$

Finally, the corrected variance of both moment estimators of the regression coefficients $\boldsymbol{\beta}$ can be obtained by combining the Delta method with the sandwich formula, and it turns out to be the same as the corrected variance of the ML estimators. □

Proposition 8

The combination of *i.i.d.* data with Assumption 1 implies that under standard regularity conditions we can effectively prove consistency by showing that the expected value of the score in (3) is zero. Let us start by considering the case in which $\mathbf{P}(\boldsymbol{\rho})$ is unrestricted, so that $\boldsymbol{\rho}$ contains the $K(K-1)/2$ off-diagonal elements of the correlation matrix. But since

$$E(y_i^2) = 1 \;\; \text{and} \;\; \rho_{ij} = E(y_i y_j),$$

then $\mathbf{P}(\boldsymbol{\rho}_\infty) = E(\mathbf{yy})$. More generally, consider $\mathbf{P}(\boldsymbol{\rho})$, where $\boldsymbol{\rho}$ is a $p \times 1$ vector with $p < K(K-1)/2$. In this case,

$$E[\mathbf{s}_{\boldsymbol{\rho} i}(\mathbf{y}; \boldsymbol{\rho})] = \frac{\partial vecl'[\mathbf{P}(\boldsymbol{\rho})]}{\partial \boldsymbol{\rho}} E[\mathbf{s}_{Ki}\{vecl[\mathbf{P}(\boldsymbol{\rho})]\}] = \mathbf{0},$$

where the first equality follows from the chain rule and the last one from the fact that $\mathbf{P}(\boldsymbol{\rho})$ is correctly specified. □

C Trivariate copula expressions

C.1 Score

Applying the general formula in (3) to the trivariate case yields

$$
\begin{aligned}
s_{\rho_{12}}(y_1,y_2,y_3,\rho_{12},\rho_{13},\rho_{23}) \quad = \quad & \frac{1}{(1-\rho_{12}^2-\rho_{13}^2+2\rho_{12}\rho_{13}\rho_{23}-\rho_{23}^2)^2} \\
& \times \left[y_1^2(\rho_{12}-\rho_{13}\rho_{23})(\rho_{23}^2-1)-\rho_{12}^3+\rho_{12}^2\rho_{13}[(3+y_3^2)\rho_{23}-y_2y_3]\right. \\
& +\rho_{13}[y_2^2(\rho_{23}-\rho_{13}^2\rho_{23})+y_2y_3(\rho_{13}^2-\rho_{23}^2-1)+\rho_{23}(y_3^2+\rho_{13}^2+\rho_{23}^2-1)] \\
& -\rho_{12}[-1-y_2^2(-1+\rho_{13}^2)-2y_2y_3\rho_{23}+\rho_{23}^2+y_3^2\rho_{23}^2+\rho_{13}^2(1+y_3^2+2\rho_{23}^2)] \\
& +y_1\{-y_3(\rho_{23}+\rho_{12}^2\rho_{23}+\rho_{13}^2\rho_{23}-2\rho_{12}\rho_{13}-\rho_{23}^3) \\
& \left.+y_2[1+\rho_{12}^2-2\rho_{12}\rho_{13}\rho_{23}-\rho_{23}^2-\rho_{13}^2(1-2\rho_{23}^2)]\}\right]
\end{aligned}
$$

$$s_{\rho_{13}}(y_1,y_2,y_3,\rho_{12},\rho_{13},\rho_{23}) = s_{\rho_{12}}(y_1,y_3,y_2,\rho_{13},\rho_{12},\rho_{23}),$$

and

$$s_{\rho_{23}}(y_1,y_2,y_3,\rho_{12},\rho_{13},\rho_{23}) = s_{\rho_{12}}(y_2,y_3,y_1,\rho_{23},\rho_{12},\rho_{13}).$$

C.2 Hessian

The expected value of the (minus) Hessian under correct of specification of the correlation matrix is given by

$$E[-\mathbf{h}_i(\boldsymbol{\rho}_\infty)] = \begin{bmatrix} h_{11}(\rho_{12\infty},\rho_{13\infty},\rho_{23\infty}) & h_{12}(\rho_{12\infty},\rho_{13\infty},\rho_{23\infty}) & h_{13}(\rho_{12\infty},\rho_{13\infty},\rho_{23\infty}) \\ & h_{22}(\rho_{12\infty},\rho_{13\infty},\rho_{23\infty}) & h_{23}(\rho_{12\infty},\rho_{13\infty},\rho_{23\infty}) \\ & & h_{33}(\rho_{12\infty},\rho_{13\infty},\rho_{23\infty}) \end{bmatrix}$$

where

$$
\begin{aligned}
h_{11}(\rho_{12},\rho_{13},\rho_{23}) &= \frac{1+\rho_{12}^2-2\rho_{12}\rho_{13}\rho_{23}-\rho_{23}^2-\rho_{13}^2(1-2\rho_{23}^2)}{(1-\rho_{12}^2-\rho_{13}^2-\rho_{23}^2+2\rho_{12}\rho_{13}\rho_{23})^2} \\
h_{22}(\rho_{12},\rho_{13},\rho_{23}) &= h_{11}(\rho_{13},\rho_{12},\rho_{23}), \\
h_{33}(\rho_{12},\rho_{13},\rho_{23}) &= h_{11}(\rho_{23},\rho_{12},\rho_{13}), \\
h_{12}(\rho_{12},\rho_{13},\rho_{23}) &= \frac{\rho_{23}^3+2\rho_{12}\rho_{13}-\rho_{23}(1+\rho_{12}^2+\rho_{13}^2)}{(1-\rho_{12}^2-\rho_{13}^2-\rho_{23}^2+2\rho_{12}\rho_{13}\rho_{23})^2}, \\
h_{13}(\rho_{12},\rho_{13},\rho_{23}) &= h_{12}(\rho_{12},\rho_{23},\rho_{13})
\end{aligned}
$$

and

$$h_{23}(\rho_{12},\rho_{13},\rho_{23}) = h_{12}(\rho_{13},\rho_{23},\rho_{12}).$$

D Spearman's calculations

D.1 Alternative estimators

Alternative estimators to $\tilde{\rho}_S^I$, which is based on the moment conditions (14), can be obtained as follows.

Given that u_{1i} and u_{2i} are uniform by definition, one could exploit the fact that $E(u_{ji}) = 1/2$ and $Var(u_{ji}) = 1/12$ to estimate ρ based on the single moment condition

$$E\left(u_{1i}u_{2i} - \frac{1}{4} - \frac{1}{12}\rho\right) = 0,$$

whence

$$\tilde{\rho}_S^{II} = 12\left(\frac{1}{N}\sum_{i=1}^{N} u_{1i}u_{2i} - \frac{1}{4}\right). \tag{D1}$$

A third estimator in which the mean of each component is substracted before computing the cross-moment is given by

$$\tilde{\rho}_S^{III} = \frac{1}{1/12}\sum_{i=1}^{N}\left(u_{1i} - \frac{1}{2}\right)\left(u_{2i} - \frac{1}{2}\right). \tag{D2}$$

Finally, the fourth estimator we could consider, which is the closest to the one Matlab implements, is

$$\tilde{\rho}_S^{IV} = 1 - \frac{6(N+1)^2}{N(N^2-1)}\sum_{i=1}^{N}(u_{1i} - u_{2i})^2,$$

which in large samples can be interpreted in terms of the following moment conditions

$$E[\mathbf{m}_i^{IV}(\boldsymbol{\theta})] = E\left[\begin{pmatrix} u_{1i}u_{2i} - \frac{1}{2}(\mu_1^2+\sigma_1^2) - \frac{1}{2}(\mu_2^2+\sigma_2^2) - \frac{1}{12}(\rho-1) \\ u_{1i} - \mu_1 \\ u_{2i} - \mu_2 \\ u_{1i}^2 - (\mu_1^2+\sigma_1^2) \\ u_{2i}^2 - (\mu_2^2+\sigma_2^2) \end{pmatrix}\right] = \mathbf{0}. \tag{D3}$$

D.2 Asymptotic variances

Regarding $\tilde{\rho}_S^{I}$, we can easily compute the expected value of the Jacobian and variance of the moment conditions to obtain the asymptotic variance for $\boldsymbol{\theta}$ in (14). In particular,

$$\mathcal{A}^I(\boldsymbol{\theta}) = E\left[\frac{\partial \mathbf{m}_i^I(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right] = -\begin{pmatrix} \sqrt{\sigma_1^2\sigma_2^2} & \mu_2 & \mu_1 & \frac{1}{2}\rho\sqrt{\sigma_2^2/\sigma_1^2} & \frac{1}{2}\rho\sqrt{\sigma_1^2/\sigma_2^2} \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 2\mu_1 & 0 & 1 & 0 \\ 0 & 0 & 2\mu_2 & 0 & 1 \end{pmatrix},$$

and

$$\begin{aligned} \mathcal{B}^I(\boldsymbol{\theta}) &= Var[\mathbf{m}_i^I(\boldsymbol{\theta})] \\ &= \begin{pmatrix} E_{22} - E_{11}^2 & E_{21} - E_{11}E_{10} & E_{12} - E_{11}E_{01} & E_{31} - E_{11}E_{20} & E_{13} - E_{11}E_{02} \\ & Var(u_{1i}) & cov(u_{1i}, u_{2i}) & E_{30} - E_{20}E_{10} & cov(u_{1i}, u_{2i}^2) \\ & & Var(u_{2i}) & cov(u_{1i}^2, u_{2i}) & E_{03} - E_{02}E_{01} \\ & & & E_{40} - E_{20}^2 & cov(u_{1i}^2, u_{2i}^2) \\ & & & & E_{04} - E_{02}^2 \end{pmatrix}, \end{aligned}$$

where $E_{h,j}$ denotes $E(u_{1i}^h u_{2i}^j)$.

As for $\tilde{\rho}_S^{II}$, it is straightforward to prove that (D1) implies $AVar(\hat{\rho}) = 144 \times Var(u_{1i}u_{2i})$.

To obtain the asymptotic variance of $\tilde{\rho}_S^{III}$ from (D2), it is convenient to use the following moment conditions

$$E\begin{bmatrix} u_{1i}u_{2i} - \frac{1}{2}\mu_1 - \frac{1}{2}\mu_2 - \frac{1}{12}\rho + \frac{1}{4} \\ u_{1i} - \mu_1 \\ u_{2i} - \mu_2 \\ u_{1i}^2 - (\mu_1^2 + \sigma_1^2) \\ u_{2i}^2 - (\mu_2^2 + \sigma_2^2) \end{bmatrix} = E[\mathbf{m}_i^{III}(\boldsymbol{\theta})] = \mathbf{0},$$

whence

$$\mathcal{A}^{III}(\boldsymbol{\theta}) = E\left[\frac{\partial \mathbf{m}_i^{III}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right] = -\begin{pmatrix} 1/12 & 1/2 & 1/2 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 2\mu_1 & 0 & 1 & 0 \\ 0 & 0 & 2\mu_2 & 0 & 1 \end{pmatrix}.$$

In addition, it is easy to see that $\mathcal{B}^{III}(\boldsymbol{\theta}) = V[\mathbf{m}_i^{III}(\boldsymbol{\theta})]$ coincides with $\mathcal{B}^{I}(\boldsymbol{\theta})$.

Finally, we can use (D3) to show that $\mathcal{B}^{IV}(\boldsymbol{\theta}) = V[\mathbf{m}_i^{IV}(\boldsymbol{\theta})]$ is equal to $\mathcal{B}^{I}(\boldsymbol{\theta})$ and

$$\mathcal{A}^{IV}(\boldsymbol{\theta}) = E\left[\frac{\partial \mathbf{m}_i^{IV}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right] = -\begin{pmatrix} 1/12 & \mu_1 & \mu_2 & 1/2 & 1/2 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 2\mu_1 & 0 & 1 & 0 \\ 0 & 0 & 2\mu_2 & 0 & 1 \end{pmatrix},$$

whence we can obtain the asymptotic variance of $\tilde{\rho}_S^{IV}$.

E Description of the marginal distributions used in Section 5

E.1 Tukey distribution

The Tukey lambda distribution is a continuous, symmetric probability distribution defined in terms of its quantile function

$$Q(p, \lambda) = \begin{cases} \frac{1}{\lambda}[p^\lambda - (1-p)^\lambda], & \text{if } \lambda \neq 0 \\ \ln[p/(1-p)], & \text{if } \lambda = 0, \end{cases}$$

where λ is its single shape parameter. It nests the logistic distribution for $\lambda = 0$ and the uniform distribution for both $\lambda = 1$ and $\lambda = 2$. In Figure 6a, we plot the density of a Tukey random variable with parameter $\lambda = 1.5$.

E.2 Asymmetric Laplace distribution

The Asymmetric Laplace distribution is a continuous probability distribution consisting of two exponential distributions of unequal scale, adjusted to ensure continuity and normalization. Its density is

$$f(x; m, \kappa, \lambda) = \frac{\lambda}{\kappa + 1/\kappa} \begin{cases} \exp[(\lambda/k)(x - m)], & x \leq m \\ \exp[-\lambda\kappa(x - m)], & x > m \end{cases}$$

The quantiles for this distribution can be easily obtained from those of the two underlying exponential distributions. In Figure 6b, we plot the density of an Asymmetric Laplace random variable with parameters $m = 0$, $k = 2$ and $\lambda = 1$.

E.3 Weibull distribution

The probability density function of the Weibull distribution is

$$f(x; k, \lambda) = \begin{cases} \frac{k}{\lambda}(\frac{x}{\lambda})^{k-1} \exp[-(x/\lambda)^k], & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{E1}$$

where $k > 0$ is the shape parameter and $\lambda > 0$ is the scale parameter. Its quantile function is $F^{-1}(p; k, \lambda) = \lambda[-\ln(1-p)]^k$. When $k = 0$, it particularizes to the exponential distribution with parameter λ^{-1}. We plot the density of a Weibull random variable with parameters $k = 0.75$ and $\lambda = 1$ in Figure 6c.

E.4 Mixture of Weibull distributions

This distribution is generated by mixing a regular Weibull distribution and a mirror image of another Weibull distribution whose support is the negative real line. Suppose that x_1 follows a Weibull distribution with shape and scale parameters k_1 and λ_1, and that $-x_2$ follows a Weibull distribution with shape and scale parameters k_2 and λ_2. Further, let α denote the mixing probability associated to the first component. Then, the nonstandardized mixture x has density given by

$$f(x; k_1, k_2, \lambda_1, \lambda_2, \alpha) = \alpha f(x; k_1, \lambda_1) + (1-\alpha) f(x; k_2, \lambda_2),$$

where $f(x; k, \lambda)$ is given in (E1). We standardize x to achieve zero mean and unit variance. The quantiles for this distribution can be easily obtained from those of the two underlying Weibull distributions. In Figure 6d we plot the density of a mixture of Weibull random variables with parameters $k_1 = 5$, $\lambda_1 = 10$, $k_2 = 5$, $\lambda_1 = 2$ and mixing probability $\alpha = .98$.