# Robustness, Redundancy, and Validity of Copulas in Likelihood Models*

Artem Prokhorov Peter Schmidt

First draft: January 19, 2005
This draft: January 21, 2006

## Abstract

The paper considers likelihood-based estimation of multivariate models, in which only marginal distributions are correctly specified. The unknown joint distribution is modelled with a copula function, which may be misspecified. In a GMM framework, we study robustness and efficiency of resulting estimators, propose improvements to existing estimators and discuss tests of copula validity. It is shown that radially symmetric copulas are robust against misspecification in problems about sample means if the true joint density is also radially symmetric. Efficiency results suggest that knowledge of the true copula is redundant only if the covariance matrix for relevant moment conditions is singular. A simple simulation supports the theoretical result about robustness of the Frank, Farlie-Gumbel-Morgenstern and Ali-Mikhail-Haq copula families.

*JEL Classification*: C13
*Keywords*: Copula, (Q)MLE, GMM

---

[*] Helpful comments from J.Wooldridge, H.Koul, E.Iglesias and participants of MSU econometrics seminar and the 5th Villa Mondragone Workshop on Economic Theory and Econometrics are gratefully acknowledged.

# 1 Introduction

In multivariate economic models, one is often ready to assume marginal distributions but is reluctant to impose a joint distribution. For example, in a panel setting, economists often use a specific likelihood for each cross section separately (e.g., probit or logit) but avoid modelling the joint distribution of the cross-sections over time. Similarly, in selectivity models, it is often desired to allow for unrestricted dependence between the disturbances in the primary and the selection models, each of which has a well-defined likelihood.

The usual way to handle the indeterminacy of the joint distribution is to assume independence of the marginal distributions and employ quasi-MLE or to assume joint normality and employ pseudo-MLE (e.g. White, 1982; Gourieroux et al., 1984). In certain cases these approaches result in a consistent estimation while a "sandwich" covariance matrix may be used for valid inference.

However, these approaches suffer from major weaknesses. First, there are important cases when using a pseudo-likelihood does not result in consistent estimates. Green (2002, Section 17.9) and Wooldridge (2002, Chapter 13) discuss such cases. Second, as we show below, there are estimators that dominate traditional QMLE under non-independence.

The copula approach used here allows to replace normality or independence with an alternative assumption about the joint distribution. Clearly such a replacement is only warranted if the new distribution possesses some useful properties such as ease of computation, robustness to misspecification, and improved efficiency. Arguably, copulas (or at least some of their families) may have such properties in certain econometric models. The copula approach also incorporates multivariate normality and independence as special cases.

The copula approach is relatively new to econometrics. A note by Lee (1983) appears to be the earliest application of this approach in econometrics. Copulas have recently received a lot of attention in finance literature. They are used to model dependence in financial time series (e.g., Patton, 2001; Breymann et al., 2003) and in risk management applications (e.g., Embrechts et al., 2003, 2002). Bouyé et al. (2000) provide an extensive discussion of prospects for copula in finance. Use of copula in other subfields of econometrics still appears rather limited. Smith (2003) incorporates copula in selectivity models and provides applications to labor supply and duration of hospitalization; Cameron et al. (2004) use copula to develop a bivariate count data model with an application to the number of doctor visits.

We start by presenting some basics on copulas. This is done in Section 2. Section 3 introduces the GMM representation of the likelihood-based models used in the sequel. We show that imposing a joint distribution amounts to adding moment conditions.

Imposing moment constraints makes consistency of the resultant estimator conditional on the moment validity. Moreover, there are infinitely many alternative multivariate distributions that can be used. Section 4 shows that estimation of means remains robust against copula misspecification as long as the used copula and the true joint density share a symmetry property. A simple simulation employs most commonly used copula families to study their robustness properties.

It is well known that additional moment conditions cannot reduce asymptotic efficiency if properly used. However, sometimes the additional moments do not help even if properly used, i.e. are redundant in the sense of Breusch et al. (1999). In Section 5 we develop conditions for such redundancy.

Section 6 proposes tests of copula validity that can help deciding on the copula. Section 7 concludes.

## 2 Preliminaries

**Definition 2.1** *(Nelsen, 1999, p.40) An $M$-dimensional* ***copula*** *is a function $C : [0,1]^M \to [0,1]$ that has the following properties:*

*i.* $C(u_1, \dots, u_{m-1}, 0, u_{m+1}, \dots, u_M) = 0$, $m = 2, \dots, M-1$.

*ii.* $C(1, \dots, 1, u_m, 1, \dots, 1) = u_m$, $m = 1, \dots, M$.

*iii. $C$ is $M$-increasing: for every $M$-box $B = [a_1, b_1] \times [a_2, b_2] \times \dots \times [a_M, b_M]$, whose $2^M$ vertices $(c_1, \dots, c_M)$ are in $[0,1]^M$, the $C$-volume of $B$, defined by*

$$V_C(B) \equiv \sum_{i_1=1}^{2} \dots \sum_{i_M=1}^{2} (-1)^{i_1+\dots+i_M} C(c_{1i_1}, \dots, c_{Mi_M}),$$

*where $c_{j1} = a_j$ and $c_{j2} = b_j$ for all $j \in \{1, \dots, M\}$, satisfies*

$$V_C(B) \geq 0.$$

Property (iii) implies for $M = 2$ that $C(a_1, a_2) - C(a_1, b_2) - C(b_1, a_2) + C(b_1, b_2) \geq 0$ for any vectors $(a_1, a_2), (b_1, b_2) \in [0,1]^2$ such that $a_m \leq b_m$, $m = 1, 2$, i.e. $C(a, b)$ is non-decreasing in $(a, b)$.

It follows from the definition that an $M$-dimensional copula $C$ is an $M$-dimensional cdf whose $M$ marginals are uniform on $[0,1]$. One may also note that for any $M$-dimensional copula $C$, $M \geq 3$, each $m$-marginal of $C$, $2 \leq m < M$, is an $m$-dimensional copula.

The following well-known theorem establishes existence of such a function for any joint distribution function of random variables. We restate it with no proof.

**Theorem 2.1** *(Sklar, 1959, p.229-230) Let $H$ be an $M$-dimensional distribution function with margins $F_1, \ldots, F_M$. Then there exists an $M$-dimensional copula $C$ such that for all $x_m \in \mathbb{R}$, $m = 1, \ldots, M$*

$$H(x_1, \ldots, x_M) = C(F_1(x_1), \ldots, F_M(x_M)). \tag{1}$$

*If $F_1, \ldots, F_M$ are continuous, then $C$ is unique. Conversely, if $C$ is an $M$-dimensional copula and $F_1, \ldots, F_M$ are distribution functions, then the function $H$ in (1) is an $M$-dimensional distribution function with marginals $F_1, \ldots, F_M$.*

Thus, a copula is a multivariate distribution function that connects two or more marginal distributions so that to exactly form the joint distribution. A copula thus completely parameterizes the entire dependence structure between two or more random variables. It is important to note that a given joint distribution function $H$ defines only one set of marginal distribution functions $F_m$, $m = 1, \ldots, M$, whereas given marginal distributions do not determine a unique joint distribution (and the implied copula).

To connect copulas to likelihood-based models, let $h$ and $c$ be the derivatives of the distribution functions $H$ and $C$, respectively; let $f_m$ be the derivatives of the marginal distribution functions $F_m$, $m = 1, \ldots, M$. Then,

$$\begin{aligned} h(x_1, \ldots, x_M) &= \frac{\partial^M H(x_1, \ldots, x_M)}{\partial x_1 \ldots \partial x_M} \\ &= \frac{\partial^M C(F_1(x_1), \ldots, F_M(x_M))}{\partial x_1 \ldots \partial x_M} \\ &= \left.\frac{\partial^M C(u_1, \ldots, u_M)}{\partial u_1 \ldots \partial u_M}\right|_{u_m = F_m(x_m), m=1,\ldots,M} \prod_{m=1}^{M} f_m(x_m) \\ &= c(F_1(x_1), \ldots, F_M(x_M)) \prod_{m=1}^{M} f_m(x_m), \end{aligned}$$

i.e., the joint density is the product of the copula density and the marginal densities.

In what follows we restrict our attention to the bivariate case ($M = 2$). We let the marginal densities $f_1$ and $f_2$ be functions of an unknown parameter vector $\theta \in \mathbb{R}^p$ and the copula density $c$ and the joint density $h$ be functions of an additional parameter vector $\rho \in \mathbb{R}^q$. Then

$$\ln h(x_1, x_2; \theta, \rho) = \ln c(F_1(x_1; \theta), F_2(x_2; \theta); \rho) + \ln f_1(x_1; \theta) + \ln f_2(x_2; \theta). \tag{2}$$

Note that $\rho$ parameterizes the entire dependence between the two random variables. See Appendix A for selected copula families used in this paper.

For our discussion of copula misspecification, we let $K$ denote some copula other than the true copula $C$ and we let $k$ denote the corresponding copula density function.

# 3 GMM representation

MLE assumes a complete and correctly specified joint likelihood in (2). For the purposes of this paper, quasi-MLE (QMLE) assumes correctly specified marginal distributions and maintains their independence and thus only uses the last two terms in (2). In panel settings, what we call QMLE is often referred to as the *partial* likelihood method (see Wooldridge, 2002, Section 13.8). Pseudo-MLE (PMLE) assumes an incorrect joint distribution and thus uses an incorrectly specified copula term in (2). The (correct) copula term in (2) is therefore what distinguishes MLE from QMLE (and PMLE).

It is well known that likelihood-based models can be represented as GMM models based on likelihood equations (see Godambe, 1960, 1976). The expected value of the score function for the correctly specified joint log-likelihood (2) is zero. Furthermore, if the marginal densities are correctly specified, the same is true for the marginal log-likelihoods. Hence, under classical regularity conditions, the following four moment conditions hold at the true values of the parameters $(\theta_o, \rho_o)$:

$$\begin{array}{ll} \mathbb{E}\frac{\partial}{\partial\theta}\ln f_1(x_1;\theta_o)=0 & \text{(A)} \\ \mathbb{E}\frac{\partial}{\partial\theta}\ln f_2(x_2;\theta_o)=0 & \text{(B)} \\ \mathbb{E}\frac{\partial}{\partial\theta}\ln c(F_1(x_1;\theta_o),F_2(x_2;\theta_o);\rho_o)=0 & \text{(C)} \\ \mathbb{E}\frac{\partial}{\partial\rho}\ln c(F_1(x_1;\theta_o),F_2(x_2;\theta_o);\rho_o)=0. & \text{(D)} \end{array} \tag{3}$$

We call moment conditions (A) and (B) the **marginal moments** and (C) and (D) the **true copula moments**. Note that as stated in (3), the GMM problem is overidentified: it involves $p+q$ parameters and $3p+q$ moment conditions.

Here we will assume that the marginal distributions are correctly specified but the copula function may not be. If the copula is incorrectly specified, then copula moments (C-D) may not hold. If they do we will say that the copula is *robust*. In this case, (C) and (D) in (3) are replaced with

$$\begin{array}{ll} \mathbb{E}\frac{\partial}{\partial\theta}\ln k(F_1(x_1;\theta_o),F_2(x_2;\theta_o);\rho_o^k)=0 & \text{(C}'\text{)} \\ \mathbb{E}\frac{\partial}{\partial\rho}\ln k(F_1(x_1;\theta_o),F_2(x_2;\theta_o);\rho_o^k)=0, & \text{(D}'\text{)} \end{array} \tag{4}$$

where $\rho_o^k$ is the dependence parameter of copula $K$. We call (C$'$-D$'$) the **misspecified copula moments**. For the sense in which a parametric model for the density is correctly specified see, for example, Wooldridge (1994, p. 2672).

Our primary focus is estimation of $\theta_o$. GMM is an appropriate framework for our analysis because it allows studying robustness and efficiency of various likelihood-based estimators of $\theta_o$ (MLE, QMLE, PMLE) by considering misspecification and redundancy of copula moments.

Specifically, consider MLE versus QMLE in terms of efficiency. MLE of $(\theta_o, \rho_o)$ maximizes the joint likelihood in (2). This is equivalent to the optimal GMM estimation based on the expectation of the score of the joint likelihood, i.e.,

$$\left[\begin{array}{l} \mathbb{E}\frac{\partial}{\partial\theta}\ln f_1(x_1;\theta_o) + \mathbb{E}\frac{\partial}{\partial\theta}\ln f_2(x_2;\theta_o) + \mathbb{E}\frac{\partial}{\partial\theta}\ln c(F_1(x_1;\theta_o), F_2(x_2;\theta_o);\rho_o) \\ \mathbb{E}\frac{\partial}{\partial\rho}\ln c(F_1(x_1;\theta_o), F_2(x_2;\theta_o);\rho_o) \end{array}\right] = 0. \tag{5}$$

We show in Section 5.1 that this is equivalent to the optimal GMM based on (3).

At the same time, QMLE of $\theta_o$ maximizes the joint likelihood assuming independence of marginal distributions. This is equivalent to the optimal GMM based on

$$\mathbb{E}\frac{\partial}{\partial\theta}\ln f_1(x_1;\theta_o) + \mathbb{E}\frac{\partial}{\partial\theta}\ln f_2(x_2;\theta_o) = 0. \tag{6}$$

We will show in Section 5.1 that the optimal GMM based on (6) is no more efficient than the optimal GMM estimator based on

$$\left[\begin{array}{l} \mathbb{E}\frac{\partial}{\partial\theta}\ln f_1(x_1;\theta_o) \\ \mathbb{E}\frac{\partial}{\partial\theta}\ln f_2(x_2;\theta_o) \end{array}\right] = 0, \tag{7}$$

which we will call the Improved QMLE (IQMLE).

Thus MLE and (I)QMLE are equally efficient only if the extra copula moments in (3) do not help improve efficiency of estimation of $\theta_o$. In GMM literature this is known as partial redundancy of copula moment conditions given the marginal moment conditions in estimation of $\theta_o$ (see Breusch et al., 1999, Section 4).

Similarly, PMLE of $(\theta_o, \rho_o^k)$ maximizes the joint likelihood in (2) with a misspecified copula. This is equivalent to GMM based on

$$\left[\begin{array}{l} \mathbb{E}\frac{\partial}{\partial\theta}\ln f_1(x_1;\theta_o) + \mathbb{E}\frac{\partial}{\partial\theta}\ln f_2(x_2;\theta_o) + \mathbb{E}\frac{\partial}{\partial\theta}\ln k(F_1(x_1;\theta_o), F_2(x_2;\theta_o);\rho_o^k) \\ \mathbb{E}\frac{\partial}{\partial\rho}\ln k(F_1(x_1;\theta_o), F_2(x_2;\theta_o);\rho_o^k) \end{array}\right] = 0. \tag{8}$$

Since we assume correct specification of the marginal distributions, the moment conditions in (8) do not hold if and only if the moment conditions in (4) do not hold, i.e. if and only if the copula moments are not robust to misspecification.

Finally, for robust misspecified copula moments, we will show in Section 5.2 that PMLE is dominated by the optimal GMM estimator using (3A-B)-(4C′-D′), which we will call the Improved PMLE (IPMLE). Thus the question of relative efficiency of IQMLE versus IPMLE for $\theta_o$ is that of partial redundancy of the misspecified copula moments.

## 4 Robustness of copula terms

Redundancy applies to valid moment conditions. We therefore first discuss robustness of copula terms to misspecification. We seek to characterize an incorrect copula $K$, for which the copula moments in (4) hold in the population.

### 4.1 A Theoretical Result

Let $X_1$ and $X_2$ be random variables with joint distribution function $H$, marginal distribution functions $F_1$ and $F_2$, respectively, and copula $C$. Let $(\mu_1, \mu_2)$ be a point in $\mathbb{R}^2$.

**Definition 4.1** *$(X_1, X_2)$ is **radially symmetric (RS) about** $(\mu_1, \mu_2)$ if*

$$H(\mu_1 + x_1, \mu_2 + x_2) = 1 - F_1(\mu_1 - x_1) - F_2(\mu_2 - x_2) + H(\mu_1 - x_1, \mu_2 - x_2), \tag{9}$$

*for all $(x_1, x_2)$ in $\{\mathbb{R}^2 \bigcup \{\pm\infty\}\}$.*

Essentially, RS requires that any two points equally distant from $(\mu_1, \mu_2)$ that lie on the same line identify tail segments under the joint density function that have equal volume. It is clear from (9) that the true joint density $h(x_1, x_2)$ satisfies $h(\mu_1 + x_1, \mu_2 + x_2) = h(\mu_1 - x_1, \mu_2 - x_2)$ under RS. Moreover, if $x_1$ or $x_2$ in (9) is taken to be equal $\infty$, it follows that $F_i(\mu_i + x_i) = 1 - F_i(\mu_i - x_i)$, or $Prob(X_i - \mu_i \leq x_i) = Prob(\mu_i - X_i \leq x_i)$, i.e. $X_1$ and $X_2$ are **marginally symmetric (MS) about** $(\mu_1, \mu_2)$. RS is therefore a stronger symmetry concept than the usual (univariate) symmetry of random variables. It is however weaker than joint symmetry, which holds when $h(\mu_1 + x_1, \mu_2 + x_2) = h(\mu_1 + x_1, \mu_2 - x_2) = h(\mu_1 - x_1, \mu_2 + x_2) = h(\mu_1 - x_1, \mu_2 - x_2)$ (see Nelsen, 1993, for details). Many commonly used distributions are RS. For example, bivariate Normal, bivariate Student-t, bivariate Cauchy and other elliptically contoured distributions are RS. For a discussion of the elliptically contoured family of distributions, see Mardia et al. (1979, Section 2.7.2).

Now consider some copula $K \neq C$.

**Definition 4.2** *A copula $K$ is **radially symmetric (RS)** if*

$$K(1 - u, 1 - v) = 1 - u - v + K(u, v) \text{ for all } (u, v) \text{ in } \mathbb{I}^2. \tag{10}$$

Radial symmetry of copulas requires of the copula function what radial symmetry of random variables requires of the joint density function. Eq.(10) suggests that for the rectangles $[0, u] \times [0, v]$ and $[1 - u, 1] \times [1 - v, 1]$, the volume under the copula density function is the same for any $(u, v)$.

It can be shown that (marginally) symmetric random variables $X_1$ and $X_2$ are radially symmetric if and only if $C$ satisfies (10)(see Nelsen, 1999, p.33). So if $(X_1, X_2)$ is RS then (10) holds for the true copula $C$. However, (10) may hold for many other RS copulas.

It is sometimes easier to verify radial symmetry of a copula function $K$ by checking whether the copula density $k$ satisfies the equation $k(1-v, 1-u) = k(v, u), \forall u, v$. For example, for FGM family it is easier to verify that the density function satisfies this condition than to verify that the copula function satisfies (10). In contrast, for other families in Appendix A it is easier to check (10). Using one of the methods, one can establish that the independence, FGM, Normal, Plackett, and Frank families are RS, while the Logistic, AMH, Joe, Clayton and Gumbel families are not. Interestingly, Frank (1979) shows that the only Archimedean copula family (see Appendix A for the definition) that satisfies (10) is the Frank family. Joe, AMH, Clayton and Gumbel are all Archimedean copulas that are not RS.

**Theorem 4.1** *If $(X_1, X_2)$ are RS about $(\mu_1, \mu_2)$ then*

$$\mathbb{E}\frac{\partial}{\partial \mu} \ln k(F_1(\mu_1 + x_1), F_2(\mu_2 + x_2), \rho) = 0, \forall \rho \in \mathbb{R}^q,$$

*where $k$ is any RS copula density.*

**Proof:** See Appendix B for all proofs that are not given in the main text.

By Theorem 4.1, the misspecified copula moment condition in (C$'$) can be used to consistently estimate the symmetry point $(\mu_1, \mu_2)$ as long as the copula function and the true joint density share the property of radial symmetry.

Note that the theorem does not state anything about moment condition (D$'$) in (4) and the true copula dependence parameter $\rho$. Generally, under the regularity conditions, (D$'$) will hold in the population for some value of $(\theta, \rho)$ but not necessarily for $(\theta_o, \rho_o)$. However, (C$'$) holds under the conditions of the theorem no matter what value of $\rho$ is used in (C$'$).

## 4.2 An Illustrative Simulation

To illustrate the result of the theorem and to study the behavior of both the misspecified copula moments (C$'$) and (D$'$) in finite samples, this section presents results of a simple simulation concerning a sample mean problem.

For copula $K$, define the sample analogues of misspecified copula moments (C$'$) and (D$'$) in (4)

$$\bar{\delta}^{\theta}(\theta, \rho) \equiv T^{-1} \sum_{t=1}^{T} \frac{\partial}{\partial \theta} \ln k(F_1(x_{1t}; \theta), F_2(x_{2t}; \theta); \rho) \tag{11}$$

and

$$\bar{\delta}^{\rho}(\theta, \rho) \equiv T^{-1} \sum_{t=1}^{T} \frac{\partial}{\partial \rho} \ln k(F_1(x_{1t}; \theta), F_2(x_{2t}; \theta); \rho). \tag{12}$$

Clearly, if $K = C$ then $\bar{\delta}(\theta_o, \rho_o) \rightarrow_p 0$ since (C) and (D) hold in population. Moreover, by WLLN, for any misspecified copula for which (4) holds, $\bar{\delta}(\theta_o, \rho_o^k) \rightarrow_p 0$. However, for non-robust copulas, the probability limit may be non-zero.

In order to be able to compare copulas we define a common measure of dependence. There are very many such measures (see Nelsen, 1999, Section 5). We pick one that has a simple copula representation.

**Definition 4.3** *For any two continuous random variables $U$ and $V$ whose copula is $K$,* ***Kendall's $\tau$ measure of concordance*** *is given by*

$$\tau = 4 \int\int_{\mathbb{I}^2} K(u, v; \rho) dK(u, v; \rho) - 1. \tag{13}$$

It follows from (13) that

$$\tau = 4 \int\int_{\mathbb{I}^2} K(u, v; \rho) k(u, v; \rho) du dv - 1 = 4\mathbb{E}K(U, V; \rho) - 1. \tag{14}$$

For two random variables, Kendall's $\tau$ can be viewed as the probability that "large" ("small") values of one are associated with "large" ("small") values of the other (the probability of concordance) minus the probability that "large" ("small") values of one are associated with "small" ("large") values of the other (the probability of discordance). Importantly, various copulas cover unequal ranges of dependence as measured by Kendall's $\tau$ (see Appendix A). We therefore control for $\tau$ in all one-parameter copulas.

In the simulation, we use the fact (see, e.g., Kendall, 1949) that for the Normal copula with Normal margins, Pearson's correlation coefficient

$$\rho = \sin \frac{\pi}{2} \tau. \tag{15}$$

This allows us to derive the value of Kendall's $\tau$ that corresponds to the true value of Pearson's correlation coefficient $\rho$ employed in simulating the joint Normal distribution.

We employ the following procedure:

Step 1. Generate $T$ realizations of $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left( \begin{bmatrix} m \\ m \end{bmatrix}, \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix} \right)$ by

- generating $Z \sim N\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right)$;
- using the Cholesky decomposition

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = m + \begin{bmatrix} 1 & 0 \\ r & \sqrt{1 - r^2} \end{bmatrix} Z;$$

Step 2. For each realization $t$, calculate

- $u_{it}(\mu) = \Phi(X_{it} - \mu), i = 1, 2$, where $\Phi(\cdot)$ is the Standard Normal c.d.f.;
- $k_t(\mu, \rho) \equiv k(u_{1t}(\mu), u_{2t}(\mu); \rho)$;
- $\delta_t^\mu(\mu, \rho) \equiv \frac{\partial}{\partial \mu} \ln k_t(\mu, \rho)$ and $\delta_t^\rho(\mu, \rho) \equiv \frac{\partial}{\partial \rho} \ln k_t(\mu, \rho)$;

Step 3. Calculate sample averages

$$\bar{\delta}^\cdot(\mu, \rho) \equiv \sum_{t=1}^{T} \delta_t^\cdot(\mu, \rho).$$

Step 4. Plot the resultant functions $\bar{\delta}^\mu(\mu, \rho)$ and $\bar{\delta}^\rho(\mu, \rho)$ over a relevant range of $\mu$ and $\rho$.

Step 5. Evaluate the sample means $\bar{\delta}^\mu$ and $\bar{\delta}^\rho$ and the sample standard errors $se(\bar{\delta}^\mu) = s^\mu/\sqrt{T}$ and $se(\bar{\delta}^\rho) = s^\rho/\sqrt{T}$ at the true parameter values $\mu = m_o$ and $\rho = \rho_o^k$, where

$$s^\cdot = \sqrt{\frac{\sum_{t=1}^{T}(\delta_t^\cdot(m_o, \rho_o^k) - \bar{\delta}^\cdot(m_o, \rho_o^k))^2}{T-1}}.$$

The true parameter values in Step 1 are $m_o = 0$ and $r_o = 0.3$. We use (15) to calculate the true $\tau$ and then we use (14) to derive the value of $\rho$ corresponding to the true value of $\tau$ for each copula. We consider the independence, Logistic, Farlie-Gumbel-Morganstern, Joe, Ali-Mikhail-Haq, Clayton, Gumbel, Frank and Normal copulas. For some of these copulas it is possible to obtain an analytical solution for $\rho$ in terms of $\tau$ using (14) (see Appendix A), otherwise we use numerical methods to approximate the true value of $\rho$ with desired accuracy. Note that the independence, Farlie-Gumbel-Morganstern, Frank and Normal families are radially symmetric.

Table 1 contains the true values of $\tau$ and $\rho$ for the considered families of copulas. We choose $r_o = 0.3$ because it corresponds to a value of $\tau$ within the coverage of all the one-parameter copula families we consider. Note that the two no-parameter copulas, independence and Logistic, imply dependence measures that are different from the true.

Figures 1 through 8 of Appendix C contain the plots of $\bar{\delta}^\mu(\mu, \rho)$ and $\bar{\delta}^\rho(\mu, \rho)$ obtained in Step 4. The sample size used for the plots is 200. According to Figure 1, the independence copula is robust: the copula term is identically zero even though the marginal terms are not independent. The copula term for the Logistic copula is zero for a value of $\mu$ around 0.33.

Figures 2–8 illustrate how the one-parameter copulas compare in terms of robustness. Note that all the surfaces appear to intersect the zero plane at around the true values of the parameters, which suggests general robustness. As we show below, however, one cannot accept

Table 1: The true values for Kendall's $\tau$ and $\rho$ used in simulation

| Copula | $\rho_o^k$ | $\tau_o$ |
|---|---|---|
| Independence | – | 0 |
| Logistic | – | 1/3 |
| Farlie-Gumbel-Morganstern (FGM) | 0.872880 | 0.193973 |
| Joe | 1.426845 | 0.193973 |
| Ali-Mikhail-Haq (AMH) | 0.697058 | 0.193973 |
| Clayton | 0.481321 | 0.193973 |
| Gumbel | 1.240654 | 0.193973 |
| Frank | 1.801160 | 0.193973 |
| Normal with Normal margins | 0.3 | 0.193973 |

the hypothesis of zero $\bar{\delta}$ for all copula families. The benchmark for comparisons is the Normal copula – Figure 7.

Interestingly, the sample analogue of the Normal copula moment (C) is close to zero at the true value of $\mu$ for any value of $\rho$ and at $\rho = 0$ for any value of $\mu$ – panel (6a). The FGM, AMH and Frank families display a similar feature – panels (1a), (3a) and (7a). Clearly, when $\rho = 0$, these four families of copulas reduce to the independence copula, which is known to be robust. When $\rho \neq 0$, $\bar{\delta}^{\mu}$ is still close to zero at the true $m_o$. This observation suggests robustness of the FGM, AMH and Frank families. With these copulas, one can use the copula moment (C) with any assumed $\rho$ and obtain a consistent estimate of $\mu$. The other families do not exhibit this advantage.

Of course, the FGM and Frank families of copulas are RS. The observed robustness of these families is clearly a consequence of the theoretical result in the previous section. However, the AMH family is not RS. Why is the AMH copula robust? To answer this question, write the AMH copula as an infinite sum of a geometric sequence

$$\frac{uv}{1-\rho(1-u)(1-v)} = uv\sum_{k=0}^{\infty}[\rho(1-u)(1-v)]^k. \tag{16}$$

The FGM copula is then the first-order approximation to the AMH family, which explains similar robustness.

To test the features illustrated on the figures, in Step 5 we calculate $\bar{\delta}^{\mu}$ and $\bar{\delta}^{\rho}$ at the true parameter values $\mu = m_o = 0$ and $\rho = \rho_o$ and evaluate standard errors for these averages. Table 2 shows these values along with the estimated Pearson's correlation coefficient $\hat{r}_o$ as sample size grows from 200 to 30,000. The ratio of the sample average to the standard error in parenthesis is a test statistic. Under $H_o : \hat{\delta}^{\cdot} = 0$, it is asymptotically standard Normal.

Table 2: Relative robustness measures for selected copulas, their standard errors, and estimated Pearson's correlation coefficient $\hat{r}_o$ for three sample sizes

| | T=200 | | T=3,000 | | T=30,000 | |
|---|---|---|---|---|---|---|
| | $\bar{\delta}^{\mu}(m_o, \rho_o)$ | $\bar{\delta}^{\rho}(m_o, \rho_o)$ | $\bar{\delta}^{\mu}(m_o, \rho_o)$ | $\bar{\delta}^{\rho}(m_o, \rho_o)$ | $\bar{\delta}^{\mu}(m_o, \rho_o)$ | $\bar{\delta}^{\rho}(m_o, \rho_o)$ |
| Indep. | 0 | – | 0 | – | 0 | – |
| Logistic | -0.15036 | – | -0.17491 | – | -0.16269 | – |
| | (0.05027) | | (0.01143) | | (0.00365) | |
| FGM$^\sharp$ | 0.00243 | -0.00271 | -0.00684 | -0.01072 | -0.00399 | -0.00403 |
| | (0.02990) | (0.03026) | (0.00748) | (0.00747) | (0.00238) | (0.00237) |
| Joe | 0.07855 | -0.19118 | 0.03835 | -0.10287 | 0.03922 | -0.09700 |
| | (0.03015) | (0.05752) | (0.00857) | (0.01450) | (0.00260) | (0.00452) |
| AMH | 0.02062 | -0.00678 | -0.00308 | -0.01755 | 0.00218 | -0.01101 |
| | (0.02964) | (0.04951) | (0.00709) | (0.01218) | (0.00226) | (0.00384) |
| Clayton | -0.00061 | -0.08578 | -0.02670 | -0.08060 | -0.01942 | -0.06285 |
| | (0.03309) | (0.06315) | (0.00780) | (0.01404) | (0.00248) | (0.00418) |
| Gumbel | 0.04436 | -0.15441 | 0.01174 | -0.05754 | 0.01579 | -0.04967 |
| | (0.02682) | (0.08081) | (0.00739) | (0.02055) | (0.00223) | (0.00640) |
| Frank$^\sharp$ | -0.00039 | -0.00254 | -0.00453 | -0.00145 | -0.00390 | 0.00063 |
| | (0.02786) | (0.11547) | (0.00674) | (0.00297) | (0.00216) | (0.00932) |
| **Normal$^\sharp$** | **0.00984** | **-0.06024** | **-0.00599** | **-0.00481** | **-0.00348** | **0.00194** |
| | (0.02762) | (0.08817) | (0.00684) | (0.02151) | (0.00214) | (0.00666) |
| $\hat{r}_o$ | 0.3181 | | 0.3042 | | 0.3005 | |

Notes: $^\sharp$ denotes copulas for which $\bar{\delta}^{\mu}$ and $\bar{\delta}^{\rho}$ are insignificantly different from zero at the 5% level for every sample size.

The table entries for the Logistic copula are significantly different from zero. This copula is not RS and it implies a different measure of dependence ($\tau = 1/3$). This suggests running the same simulation with common $\tau = 1/3$ for all copulas. However, this value falls outside the coverage range for several one-parameter copula families (see Appendix A), making a general comparison infeasible.

As expected, the entries for the Normal copula are insignificantly different from zero for all sample sizes. For the two RS copula families, FGM and Frank, one cannot reject the null either. The AMH family is fairly robust, too. For the Joe, Clayton and Gumbel families, the sample averages are significantly different from zero for at least one sample size which confirms the observation that these non-RS copulas are not robust in this setting.

Among the one-parameter copula families, several entries in the table stand out. First, the

Frank family performs better than the Normal benchmark by a large margin for all sample sizes except for $\bar{\delta}^{\mu}$ for $T = 30,000$, when it is very close to Normal. Second, the FGM family is relatively more robust than the Normal copula for $T = 200$ and is comparable for the other sample sizes. Third, the AMH family also performs well in particular in terms of $\bar{\delta}^{\mu}$ for the larger samples. Finally, the Clayton family outperforms the true copula for the smaller sample size but not for the larger sizes.

In the previous section, it was noted that (D′) does not generally have to hold in the population for RS copulas. An interesting observation from Table 2 is that sample analogues of (D′) are insignificantly different from zero for RS copulas and significantly different from zero for others. This does not follow from Theorem 4.1.

# 5 Redundancy of copula terms

We now turn to the question of redundancy of copula moments. We assume that we either have the true copula moments (3C-D) or the robust misspecified copula moments (4C′-D′) that hold at the true value of $\theta$. We would like to study conditions under which using valid copula moments (either the true or misspecified ones) does not result in efficiency gains in estimation of $\theta$.

## 5.1 Redundancy with Correct Copula

We first prove a lemma that reveals the structure of the variance and derivative matrices of the moment functions in (3). Recall that correct specification of the copula is assumed in (3).

**Lemma 5.1** *Denote the covariance matrix of the moment functions in (3) by* $\mathbf{C}$*, their expected derivative matrix with respect to* $(\theta, \rho)$ *by* $\mathbf{D}$*. Then,*

$$\mathbf{C} = \left[\begin{array}{cc|cc} \mathbf{A} & \mathbf{G} & -\mathbf{G} & \mathbf{0} \\ \mathbf{G}' & \mathbf{B} & -\mathbf{G}' & \mathbf{0} \\ \hline -\mathbf{G}' & -\mathbf{G} & \mathbf{J} & \mathbf{E} \\ \mathbf{0} & \mathbf{0} & \mathbf{E}' & \mathbf{F} \end{array}\right] \tag{17}$$

*and*

$$\mathbf{D} = \left[\begin{array}{c|c} -\mathbf{A} & \mathbf{0} \\ -\mathbf{B} & \mathbf{0} \\ \hline \mathbf{G} + \mathbf{G}' - \mathbf{J} & -\mathbf{E} \\ -\mathbf{E}' & -\mathbf{F} \end{array}\right], \tag{18}$$

*where* $\mathbf{A}, \mathbf{B}, \mathbf{E}, \mathbf{F}, \mathbf{G}, \mathbf{J}$ *are matrix-functions of* $(\theta, \rho)$ *defined in Appendix B.*

Several important observations immediately follow from the lemma. First, (A) and (B) are uncorrelated with (C) if and only if (A) and (B) are uncorrelated with each other ($\mathbf{G} = 0$). Second, the optimal GMM based on (3) is identical to the ML estimation in (5), as claimed in Section 3. To see this explicitly, note that the optimal GMM on (3) does not change if (3) is pre-multiplied by a matrix $\mathbf{W}$ such that $\mathbf{W} = \mathbf{D}'\mathbf{C}^{-1}$, if $\mathbf{C}$ is nonsingular. But, by Lemma 5.1,

$$\mathbf{W} = \mathbf{D}'\mathbf{C}^{-1} = -\begin{bmatrix} \mathbb{I} & \mathbb{I} & \mathbb{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbb{I} \end{bmatrix} \mathbf{C}\mathbf{C}^{-1} = -\begin{bmatrix} \mathbb{I} & \mathbb{I} & \mathbb{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbb{I} \end{bmatrix},$$

where $\mathbb{I}$ denotes the identity matrix of the relevant dimension. Clearly, this reproduces the MLE first order conditions (5). Not surprisingly, estimators that use the same first order conditions yield the same asymptotic variance matrices. In particular, for non-singular $\mathbf{C}$, the asymptotic variance matrix of the optimal GMM estimator of $(\theta, \rho)$ based on (3) can be written as

$$\mathbb{V}_{\mathrm{GMM}} = (\mathbf{D}'\mathbf{C}^{-1}\mathbf{D})^{-1}. \tag{19}$$

(We use the standard notation according to which "$\mathbb{V}$ is the asymptotic variance of an estimator $\hat{\theta}$" means that "$\sqrt{N}(\hat{\theta} - \theta_o)$ converges in distribution to $N(\mathbf{0}, \mathbb{V})$.") By Lemma 5.1, this is identical to the asymptotic variance matrix of the MLE estimator of $(\theta, \rho)$

$$\mathbb{V}_{\mathrm{MLE}} = -\left(\begin{bmatrix} \mathbb{I} & \mathbb{I} & \mathbb{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbb{I} \end{bmatrix} \mathbf{D}\right)^{-1} = \left(\begin{bmatrix} \mathbb{I} & \mathbb{I} & \mathbb{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbb{I} \end{bmatrix} \mathbf{C} \begin{bmatrix} \mathbb{I} & \mathbf{0} \\ \mathbb{I} & \mathbf{0} \\ \mathbb{I} & \mathbf{0} \\ \mathbf{0} & \mathbb{I} \end{bmatrix}\right)^{-1}. \tag{20}$$

In contrast to $\mathbb{V}_{\mathrm{GMM}}$, $\mathbb{V}_{\mathrm{MLE}}$ is defined even if $\mathbf{C}$ is singular. In fact the last representation in (20) involves the outer-product-of-the-score form of the information matrix, while the one before the last involves the expected-Hessian form of the information matrix. Both are non-singular under regularity conditions.

By a similar argument, it follows from Lemma 5.1 that the marginal moments (7) are *not* equivalent to the QMLE first order conditions (6). To see this explicitly, partition $\mathbf{C}$ and $\mathbf{D}$ as follows:

$$\mathbf{C} = \begin{bmatrix} \mathbf{C_{11}} & \mathbf{C_{12}} \\ \mathbf{C_{21}} & \mathbf{C_{22}} \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} \mathbf{D_{11}} & \mathbf{0} \\ \mathbf{D_{21}} & \mathbf{D_{22}} \end{bmatrix}, \tag{21}$$

where $\mathbf{C_{11}}, \mathbf{C_{12}}, \mathbf{C_{21}}, \mathbf{C_{22}}, \mathbf{D_{11}}, \mathbf{D_{21}}, \mathbf{D_{22}}$ correspond to the blocks separated by the dotted lines in (17-18). The optimal GMM based on (7) does not change if the moment conditions

(7) are pre-multiplied by a matrix $\mathbf{W_{11}}$ such that $\mathbf{W_{11}} = \mathbf{D_{11}}'\mathbf{C_{11}^{-1}}$, if $\mathbf{C_{11}}$ is nonsingular. Now, using Lemma 5.1,

$$\mathbf{W_{11}} = \mathbf{D_{11}}'\mathbf{C_{11}^{-1}} = -\begin{bmatrix} \mathbb{I} & \mathbb{I} \end{bmatrix} - \begin{bmatrix} -\mathbf{G}' & -\mathbf{G} \end{bmatrix} \mathbf{C_{11}}^{-1}.$$

The last term is what distinguishes the optimal GMM based on the stacked marginal moments (7) from summation (6) employed by QMLE. Call the GMM estimator based on (7), the *Improved QML* estimator (IQMLE).

Schmidt (2004) shows that correlation between marginal scores used in the optimal weighting matrix results in efficiency gains over summation and that there are interesting cases when the two estimation methods are equally efficient. A trivial such case is when there is no correlation between the marginal scores, i.e. $\mathbf{G} = 0$. We provide a formal statement and a proof of this relative efficiency result in the following theorem. The logic of the proof will be used again when we compare PMLE and IPMLE.

**Theorem 5.1** *(Schmidt, 2004) Let* $\mathbb{V}_{\text{IQMLE}}$ *and* $\mathbb{V}_{\text{QMLE}}$ *denote the asymptotic variance matrices of the IQMLE and QMLE of* $\theta_o$*, respectively. Then,* $\mathbb{V}_{\text{QMLE}} - \mathbb{V}_{\text{IQMLE}}$ *is positive semi-definite.*

**Proof**. Define $\mathbb{A} = [\mathbb{I} \quad \mathbb{I}]$. Then, (6) can be rewritten as (7) pre-multiplied by $\mathbb{A}$. Correspondingly, the variance matrix of the moment functions in (6) can be expressed as $\mathbb{A}\mathbf{C_{11}}\mathbb{A}'$, where $\mathbf{C_{11}}$ is the variance matrix for the moment functions in (7), defined in (21). Similarly, the expected derivative matrix for the moment conditions in (6) can be expressed in terms of the relevant matrix for (7) as $\mathbb{A}\mathbf{D_{11}}$.

Then,

$$\mathbb{V}_{\text{QMLE}} = [(\mathbb{A}\mathbf{D_{11}})'(\mathbb{A}\mathbf{C_{11}}\mathbb{A}')^{-1}(\mathbb{A}\mathbf{D_{11}})]^{-1}, \tag{22}$$

while

$$\mathbb{V}_{\text{IQMLE}} = [\mathbf{D_{11}}'\mathbf{C_{11}^{-1}}\mathbf{D_{11}}]^{-1}. \tag{23}$$

But $\mathbb{V}_{\text{QMLE}} - \mathbb{V}_{\text{IQMLE}}$ is positive semi-definite (PSD) if and only if $\mathbb{V}_{\text{IQMLE}}^{-1} - \mathbb{V}_{\text{QMLE}}^{-1} = \mathbf{D_{11}}'\mathbf{C_{11}^{-1}}\mathbf{D_{11}} - \mathbf{D_{11}}'\mathbb{A}'(\mathbb{A}\mathbf{C_{11}}\mathbb{A}')^{-1}\mathbb{A}\mathbf{D_{11}}$ is PSD. The last expression can be rewritten as $\mathbf{D_{11}}'\mathbf{C_{11}^{-1/2}}[\mathbb{I} - \mathbf{C_{11}^{1/2}}\mathbb{A}'(\mathbb{A}\mathbf{C_{11}^{1/2}}\mathbf{C_{11}^{1/2}}\mathbb{A}')^{-1}\mathbb{A}\mathbf{C_{11}^{1/2}}]\mathbf{C_{11}^{-1/2}}\mathbf{D_{11}}$. This is PSD because the matrix in brackets is the PSD projection matrix orthogonal to $\mathbf{C_{11}^{1/2}}\mathbb{A}'$. □

Conditions under which the copula moments do not help in terms of efficiency for $\theta$ can be derived by comparing $\mathbb{V}_{\text{IQMLE}}$ with the upper left $p\times p$ block of $\mathbb{V}_{\text{MLE}}$. When $\mathbf{C}$ is non-singular, the comparisons can be equivalently made to the upper left $p \times p$ block of $\mathbb{V}_{\text{GMM}}$.

Breusch et al. (1999) (henceforth, BQSW) developed a very useful toolbox for analyzing redundancy of a set of moment conditions given another set of moment conditions. However, their analysis assumes nonsingular **C**. For this reason, we do not employ their results here but compare $\mathbb{V}_{\text{IQMLE}}$ with the relevant block of $\mathbb{V}_{\text{MLE}}$ directly.

**Theorem 5.2** *$\mathbb{V}_{\text{MLE}}$ for $\theta$ and $\mathbb{V}_{\text{IQMLE}}$ are equal if and only if*

$$\mathbf{J} - \mathbf{C}^{\theta}_{\mathbf{21}}\mathbf{C}^{-\mathbf{1}}_{\mathbf{11}}\mathbf{C}^{\theta}_{\mathbf{12}} - \mathbf{E}\mathbf{F}^{-\mathbf{1}}\mathbf{E} = \mathbf{0}, \tag{24}$$

*where* $\mathbf{C}^{\theta}_{\mathbf{21}} = {\mathbf{C}^{\theta}_{\mathbf{12}}}' = [-\mathbf{G}' \quad -\mathbf{G}]$.

The cumbersome expression in (24) has a simple interpretation in terms of singularity of **C**. It states that the linear projection of moment condition (C) on moment conditions (A), (B) and (D) is uncorrelated with moment condition (C). More specifically, (24) can be rewritten as follows

$$\mathbb{E}\left\{\left(\frac{\partial}{\partial\theta}\ln c - \boldsymbol{\Omega}_{\mathbf{21}}\boldsymbol{\Omega}^{-\mathbf{1}}_{\mathbf{11}}\begin{bmatrix}\frac{\partial}{\partial\theta}\ln f_1\\ \frac{\partial}{\partial\theta}\ln f_2\\ \frac{\partial}{\partial\rho}\ln c\end{bmatrix}\right)\frac{\partial}{\partial\theta'}\ln c\right\} = \mathbf{0},$$

where

$$\boldsymbol{\Omega}_{\mathbf{21}} = [-\mathbf{G}' \quad -\mathbf{G} \quad \mathbf{E}], \quad \boldsymbol{\Omega}_{\mathbf{11}} = \begin{bmatrix}\mathbf{A} & \mathbf{G} & \mathbf{0}\\ \mathbf{G}' & \mathbf{B} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{F}\end{bmatrix}$$

and the arguments of the moment functions have been suppressed for brevity. In other words, (C) has to be a linear combination of (A), (B) and (D) for the copula information to be redundant in terms of asymptotic efficiency of estimation of $\theta$. Thus **C** has to be singular.

Since $\mathbb{V}_{\text{MLE}} = \mathbb{V}_{\text{GMM}}$ for non-singular **C**, and $\mathbb{V}_{\text{IQMLE}}$ is equal to $\mathbb{V}_{\text{MLE}}$ for $\theta$ if and only if **C** is singular, thus equality of $\mathbb{V}_{\text{IQMLE}}$ and $\mathbb{V}_{\text{GMM}}$ for $\theta$ is impossible unless (C) is a linear combination of (A), (B) and (D).

**Corollary 5.1** *If* (C) *is a linear combination of* (A) *and* (B) *with $\rho$ known then*

1. $\mathbf{E} = \mathbf{0}$;
2. $\mathbf{J} - \mathbf{C}^{\theta}_{21}\mathbf{C}^{-1}_{11}\mathbf{C}^{\theta}_{12} = \mathbf{0}$;
3. *IQMLE is efficient.*

We therefore have redundancy of copula knowledge in (C) and (D) given the knowledge of the marginals in (A) and (B) if only the copula moment (C) is not informative given (A) and (B).

Examples at the end of this section illustrate how one can apply the redundancy results in practice.

### 5.2 Redundancy with Misspecified Copula

Now suppose incorrect but zero-mean copula terms in (4C′) and (4D′) are used in estimation. When is such knowledge redundant in terms of efficient estimation of $\theta$?

**Lemma 5.2** *Denote the covariance matrix of the moment functions in (3) that employ the copula moments (4C′) and (4D′) instead of (3C) and (3D), respectively, by* $\mathbf{C^k}$*, their expected derivative matrix with respect to* $(\theta, \rho^k)$ *by* $\mathbf{D^k}$*. Then,*

$$\mathbf{C^k} = \left[\begin{array}{cc|cc} \mathbf{A} & \mathbf{G} & -\mathbf{K} & -\mathbf{P} \\ \mathbf{G}' & \mathbf{B} & -\mathbf{L}' & -\mathbf{Q}' \\ \hline -\mathbf{K}' & -\mathbf{L} & \mathbf{N} & \mathbf{V} \\ -\mathbf{P}' & -\mathbf{Q} & \mathbf{V}' & \mathbf{W} \end{array}\right]$$

*and*

$$\mathbf{D^k} = \left[\begin{array}{c|c} -\mathbf{A} & \mathbf{0} \\ -\mathbf{B} & \mathbf{0} \\ \hline \mathbf{K}' + \mathbf{L} - \mathbf{M} & -\mathbf{S} \\ \mathbf{P}' + \mathbf{Q} - \mathbf{R} & -\mathbf{T} \end{array}\right],$$

*where* $\mathbf{A}, \mathbf{B}, \mathbf{G}$ *are as in Lemma 5.1,* $\mathbf{K}, \mathbf{L}, \mathbf{M}, \mathbf{N}, \mathbf{P}, \mathbf{Q}, \mathbf{R}, \mathbf{S}, \mathbf{T}, \mathbf{V}, \mathbf{W}$ *are matrix-functions of* $(\theta, \rho^k)$ *defined in Appendix B.*

Lemma 5.2 can be used to make the following observation. The optimal GMM estimator using (3A-B)-(4C′-D′) is *not* identical to the PML estimator. This is in contrast with Lemma 5.1, in which MLE coincided with GMM using (3A-D) because we had knowledge of the correct copula. More specifically, the optimal GMM estimator based on (3A-B)-(4C′-D′) is unchanged if (3A-B)-(4C′-D′) are pre-multiplied by matrix $\mathbf{W^k} = \mathbf{D^k}'(\mathbf{C^k})^{-1}$ if $\mathbf{C^k}$ is non-singular. Using Lemma 5.2, it can be shown that

$$\mathbf{D^k}'(\mathbf{C^k})^{-1} = -\begin{bmatrix} \mathbb{I} & \mathbb{I} & \mathbb{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbb{I} \end{bmatrix} + \mathbf{Z}(\mathbf{C^k})^{-1},$$

where $\mathbf{Z}$ contains $\mathbf{G}' - \mathbf{K}', \mathbf{G} - \mathbf{L}, \mathbf{N} - \mathbf{M}', \mathbf{R}' - \mathbf{V}, \mathbf{P}', \mathbf{Q}, \mathbf{V}' - \mathbf{S}', \mathbf{W} - \mathbf{T}'$. Clearly, Lemma 5.2 becomes Lemma 5.1 if $k = c$. In this case, $\mathbf{Z} = 0$, $\mathbf{W^k} = \mathbf{W}$, the optimal weighting retrieves (5), and PMLE is equivalent to MLE.

For $k \neq c$, correlation patterns impossible in Lemma 5.1 now provide potential efficiency gains over PMLE. We call the GMM estimator using (3A-B)-(4C′-D′) the *Improved PML* estimator (IPMLE).

**Theorem 5.3** *Let* $\mathbb{V}_{\text{IPMLE}}$ *and* $\mathbb{V}_{\text{PMLE}}$ *denote the asymptotic variance matrices of the IPMLE and PMLE of* $(\theta_o, \rho_o^k)$*, respectively. Then,* $\mathbb{V}_{\text{PMLE}} - \mathbb{V}_{\text{IPMLE}}$ *is positive semi-definite.*

**Proof.** Define

$$\mathbb{A} = \left[ \begin{array}{cccc} \mathbb{I} & \mathbb{I} & \mathbb{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbb{I} \end{array} \right].$$

Then, (8) can be rewritten as (3) pre-multiplied by $\mathbb{A}$. Correspondingly, the variance matrix of the moment functions in (8) can be expressed as $\mathbb{A}\mathbf{C^k_{11}}\mathbb{A}'$. Similarly, the expected derivative matrix for the moment conditions in (8) can be expressed as $\mathbb{A}\mathbf{D^k}$.

Then,

$$\mathbb{V}_{\text{PMLE}} = [(\mathbb{A}\mathbf{D^k})'(\mathbb{A}\mathbf{C^k}\mathbb{A}')^{-1}(\mathbb{A}\mathbf{D^k})]^{-1}, \tag{25}$$

while

$$\mathbb{V}_{\text{IPMLE}} = [\mathbf{D^k}'(\mathbf{C^k})^{\mathbf{-1}}\mathbf{D^k}]^{-1}. \tag{26}$$

$\mathbb{V}_{\text{PMLE}} - \mathbb{V}_{\text{IPMLE}}$ is PSD if and only if $\mathbb{V}_{\text{IPMLE}}^{-1} - \mathbb{V}_{\text{PMLE}}^{-1} = \mathbf{D^k}'\mathbf{C^k}^{\mathbf{-1}}\mathbf{D^k} - \mathbf{D^k}'\mathbb{A}'(\mathbb{A}\mathbf{C^k}\mathbb{A}')^{-1}\mathbb{A}\mathbf{D^k}$ is PSD. Rewrite the last expression as

$$\mathbf{D^k}'(\mathbf{C^k})^{\mathbf{-1/2}}[\mathbb{I} - (\mathbf{C^k})^{\mathbf{1/2}}\mathbb{A}'(\mathbb{A}(\mathbf{C^k})^{\mathbf{1/2}}(\mathbf{C^k})^{\mathbf{1/2}}\mathbb{A}')^{-1}\mathbb{A}(\mathbf{C^k})^{\mathbf{1/2}}](\mathbf{C^k})^{\mathbf{-1/2}}\mathbf{D^k}.$$

This is PSD because the matrix in brackets is the PSD projection matrix orthogonal to $(\mathbf{C^k})^{\mathbf{1/2}}\mathbb{A}'$. □

Clearly, (I)PMLE does not improve precision of estimation of $\theta$ over IQMLE if and only if the upper left $p \times p$ block of $\mathbb{V}_{\text{(I)PMLE}}$ is equal to $\mathbb{V}_{\text{IQMLE}}$. We focus on $\mathbb{V}_{\text{IPMLE}}$ because by Theorem 5.4, if IPMLE does not improve over precision of IQMLE for $\theta$, then neither does PMLE. $\mathbb{V}_{\text{IPMLE}}$ is only defined when $\mathbf{C^k}$ is non-singular, thus we can apply the redundancy toolbox of BQSW.

**Theorem 5.4** *$\mathbb{V}_{\mathrm{IPMLE}}$ for $\theta$ and $\mathbb{V}_{\mathrm{IQMLE}}$ are equal if and only if*

$$\mathbf{M} - \mathbf{C}_{\mathbf{21}}^{\theta\mathbf{k}}\mathbf{C}_{\mathbf{11}}^{-\mathbf{1}}\mathbf{C}_{\mathbf{12}}^{\theta} - \mathbf{S}\mathbf{T}^{-1}(\mathbf{R} - \mathbf{C}_{\mathbf{21}}^{\rho\mathbf{k}}\mathbf{C}_{\mathbf{11}}^{-\mathbf{1}}\mathbf{C}_{\mathbf{12}}^{\theta}) = \mathbf{0}, \tag{27}$$

*where* $\mathbf{C}_{\mathbf{21}}^{\theta\mathbf{k}} = [-\mathbf{K}' \quad -\mathbf{L}]$, $\quad \mathbf{C}_{\mathbf{21}}^{\rho\mathbf{k}} = [-\mathbf{P}' \quad -\mathbf{Q}]$.

In (27), $\mathbf{M} - \mathbf{C}_{\mathbf{21}}^{\theta\mathbf{k}}\mathbf{C}_{\mathbf{11}}^{-\mathbf{1}}\mathbf{C}_{\mathbf{12}}^{\theta}$ and $\mathbf{R} - \mathbf{C}_{\mathbf{21}}^{\rho\mathbf{k}}\mathbf{C}_{\mathbf{11}}^{-\mathbf{1}}\mathbf{C}_{\mathbf{12}}^{\theta}$ can be viewed as covariance matrices between copula moments (C′-D′) and the error in the linear projection of the true copula moment (C) on the marginal moments (A-B). More explicitly,

$$\mathbf{M} - \mathbf{C}_{\mathbf{21}}^{\theta\mathbf{k}}\mathbf{C}_{\mathbf{11}}^{-\mathbf{1}}\mathbf{C}_{\mathbf{12}}^{\theta} = \mathbb{E}\left\{\frac{\partial}{\partial\theta}\ln k\left(\frac{\partial}{\partial\theta}\ln c - \mathbf{C_{21}}\mathbf{C}_{\mathbf{11}}^{-\mathbf{1}}\begin{bmatrix}\frac{\partial}{\partial\theta}\ln f_1\\ \frac{\partial}{\partial\theta}\ln f_2\end{bmatrix}\right)'\right\}, \tag{28}$$

$$\mathbf{R} - \mathbf{C}_{\mathbf{21}}^{\rho\mathbf{k}}\mathbf{C}_{\mathbf{11}}^{-\mathbf{1}}\mathbf{C}_{\mathbf{12}}^{\theta} = \mathbb{E}\left\{\frac{\partial}{\partial\rho}\ln k\left(\frac{\partial}{\partial\theta}\ln c - \mathbf{C_{21}}\mathbf{C}_{\mathbf{11}}^{-\mathbf{1}}\begin{bmatrix}\frac{\partial}{\partial\theta}\ln f_1\\ \frac{\partial}{\partial\theta}\ln f_2\end{bmatrix}\right)'\right\}.$$

Clearly, when both of these matrices are zero, (27) holds for any $\mathbf{S}$. Also, if only (28) is zero and $\mathbf{S} = \mathbf{0}$, (27) holds for any $\mathbf{R}$ and $\mathbf{C}_{\mathbf{21}}^{\rho\mathbf{k}}$.

**Corollary 5.2** *If (C) is a linear combination of (A) and (B) with $\rho$ known then*

1. $\mathbf{M} - \mathbf{C}_{\mathbf{21}}^{\theta\mathbf{k}}\mathbf{C}_{\mathbf{11}}^{-\mathbf{1}}\mathbf{C}_{\mathbf{12}}^{\theta} = \mathbf{0}$*;*
2. $\mathbf{R} - \mathbf{C}_{\mathbf{21}}^{\rho\mathbf{k}}\mathbf{C}_{\mathbf{11}}^{-\mathbf{1}}\mathbf{C}_{\mathbf{12}}^{\theta} = \mathbf{0}$*;*
3. *IQMLE and IPMLE for $\theta$ are equally efficient.*

By the corollary, knowledge about robust but misspecified copulas is redundant in estimation of $\theta$ given (A) and (B) when the true copula moment (C) is not informative given (A) and (B).

## 5.3 Examples

The following four examples illustrate how the redundancy results can be used in practice. The first three examples show problems where the copula moment conditions are redundant and thus IQMLE is efficient. The last example considers a situation when copula moment conditions are not redundant in general and IQMLE is generally inefficient.

**Bivariate Normal with common mean.** Assume Normal marginal densities with $\sigma_1^2 = \sigma_2^2 = 1$ and $\mu_1 = \mu_2 = \mu$

$$f_1(x_1; \mu) = \frac{1}{\sqrt{2\pi}}e^{-\frac{(x_1-\mu)^2}{2}},$$

$$f_2(x_2;\mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_2-\mu)^2}{2}}.$$

Let the true joint density be Normal, i.e.,

$$h(x_1, x_2; \mu, \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{(x_1-\mu)^2+(x_2-\mu)^2-2\rho(x_1-\mu)(x_2-\mu)}{2(1-\rho^2)}}.$$

Then, the implied copula is the Normal copula

$$c(F_1(x_1;\mu), F_2(x_2;\mu); \rho) = \frac{1}{\sqrt{1-\rho^2}} e^{-\frac{\rho\left(\rho(x_1-\mu)^2+\rho(x_2-\mu)^2-2(x_1-\mu)(x_2-\mu)\right)}{2(1-\rho^2)}},$$

where $\rho$ is the copula dependence parameter (Pearson's correlation coefficient). Thus we have the setup of our simulation in Section 4.2.

The relevant moment conditions are

$$\mathbb{E}\left\{X_1 - \mu\right\} = 0 \qquad (A)$$

$$\mathbb{E}\left\{X_2 - \mu\right\} = 0 \qquad (B)$$

$$\mathbb{E}\left\{-\frac{((X_1-\mu)+(X_2-\mu))\rho}{\rho+1}\right\} = 0 \qquad (C)$$

$$\mathbb{E}\left\{-\frac{\rho(X_1^2+X_2^2)+\mu(1-\rho)^2(X_1+X_2)-\left(1+\rho^2\right)X_1X_2+\rho^3-\mu^2-\rho-\mu^2\rho^2+2\rho\mu^2}{(\rho-1)^2(\rho+1)^2}\right\} = 0. \qquad (D)$$

(C) is clearly a linear combination of (A) and (B) for known $\rho$. By Corollary 5.1, the true copula moments are redundant for estimation of $\mu$. Furthermore, by Corollary 5.2, any valid misspecified copula moments do not help improve precision of estimation over IQMLE of $\mu$. IQMLE of $\mu$ is efficient.

Section 4.2 provided evidence of robustness of independence, FGM, AMH and Frank copula families. None of them would allow to improve efficiency over IQMLE of $\mu$.

Note that using the Normal moment generating function, one can show that

$$\mathbf{C} = \begin{bmatrix} 1 & \rho & -\rho & 0 \\ \rho & 1 & -\rho & 0 \\ -\rho & -\rho & \frac{2\rho^2}{1+\rho} & 0 \\ 0 & 0 & 0 & \frac{1+\rho^2}{(\rho-1)^2(\rho+1)^2} \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} -1 & 0 \\ -1 & 0 \\ \frac{2\rho}{1+\rho} & 0 \\ 0 & -\frac{1+\rho^2}{(\rho-1)^2(\rho+1)^2} \end{bmatrix},$$

where $\det(\mathbf{C}) = 0$, and

$$\mathbb{V}_{\text{MLE}} = \mathbb{V}_{\text{IQMLE}} = \frac{1+\rho}{2}.$$

**Bivariate Normal regression.** Let $\mathbf{y} = \mathbf{x}\beta + \epsilon$, where $\mathbf{y} = (y_1, y_2)', \mathbf{x} = (x_1, x_2)'$. Suppose $\mathbf{x}$ is non-random. Let $\epsilon = (\epsilon_1, \epsilon_2)' \sim \mathbb{N}(\mathbf{0}, \mathbf{\Sigma})$, where

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{bmatrix}$$

and $\sigma_1^2, \sigma_2^2$ are known but $\rho$ is not.

Then,

$$f_1(y_1; x_1, \beta) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(y_1 - x_1\beta)^2}{2\sigma_1^2}},$$

$$f_2(y_2; x_2, \beta) = \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(y_2 - x_2\beta)^2}{2\sigma_2^2}},$$

$$h(\mathbf{y}; \mathbf{x}, \beta, \rho) = \frac{1}{2\pi\sqrt{|\mathbf{\Sigma}|}} e^{-\frac{1}{2}(\mathbf{y}-\mathbf{x}\beta)'\mathbf{\Sigma}^{-1}(\mathbf{y}-\mathbf{x}\beta)}.$$

Then, the implied copula is Normal,

$$c(F_1(y_1; x_1, \beta), F_2(y_2; x_2, \beta); \rho) = \frac{\sqrt{\sigma_1^2\sigma_2^2}}{\sigma_1^2\sigma_2^2 - \rho^2} e^{-\frac{1}{2}\left(\epsilon_1\left(\frac{\epsilon_1\sigma_2^2 - \epsilon_2\rho}{\sigma_1^2\sigma_2^2 - \rho^2}\right) + \epsilon_2\left(\frac{\epsilon_2\sigma_1^2 - \epsilon_1\rho}{\sigma_1^2\sigma_2^2 - \rho^2}\right) - \frac{\epsilon_1^2}{2\sigma_1^2} - \frac{\epsilon_2^2}{2\sigma_2^2}\right)},$$

where $\epsilon_i = y_i - x_i\beta$, $i = 1, 2$.

The relevant moment conditions are

$$\begin{aligned}
&\mathbb{E}\left\{\frac{x_1\epsilon_1}{\sigma_1^2}\right\} = 0 && (A)\\
&\mathbb{E}\left\{\frac{x_2\epsilon_2}{\sigma_2^2}\right\} = 0 && (B)\\
&\mathbb{E}\left\{-\frac{\rho(\sigma_1^2\sigma_2^2 x_1\epsilon_2 + \sigma_1^2\sigma_2^2 x_2\epsilon_1 - \sigma_1^2\rho x_2\epsilon_2 - \sigma_2^2\rho x_1\epsilon_1)}{\sigma_1^2\sigma_2^2(\sigma_1^2\sigma_2^2 - \rho^2)}\right\} = 0 && (C)\\
&\mathbb{E}\left\{\frac{\sigma_1^2\sigma_2^2\epsilon_1\epsilon_2 + \rho^2\epsilon_1\epsilon_2 - \sigma_2^2\rho\epsilon_1^2 - \sigma_1^2\rho\epsilon_2^2 + \rho\sigma_1^2\sigma_2^2 - \rho^3}{(\sigma_1^2\sigma_2^2 - \rho^2)^2}\right\} = 0. && (D)
\end{aligned}$$

Again, (C) is a linear combination of (A) and (B). The use of (C) and (D) or any other zero mean copula terms does not help estimate $\beta$ more precisely than IQMLE.

The covariance and expected derivative matrices are

$$\mathbf{C} = \begin{bmatrix}
\frac{x_1^2}{\sigma_1^2} & \frac{\rho x_1 x_2}{\sigma_1^2\sigma_2^2} & -\frac{\rho x_1 x_2}{\sigma_1^2\sigma_2^2} & 0\\
\frac{\rho x_1 x_2}{\sigma_1^2\sigma_2^2} & \frac{x_2^2}{\sigma_2^2} & -\frac{\rho x_1 x_2}{\sigma_1^2\sigma_2^2} & 0\\
-\frac{\rho x_1 x_2}{\sigma_1^2\sigma_2^2} & -\frac{\rho x_1 x_2}{\sigma_1^2\sigma_2^2} & \frac{\left(x_2^2\sigma_1^2 + x_1^2\sigma_2^2 - 2\rho x_1 x_2\right)\rho^2}{\sigma_1^2\sigma_2^2\left(\sigma_1^2\sigma_2^2 - \rho^2\right)} & 0\\
0 & 0 & 0 & \frac{\sigma_1^2\sigma_2^2 + \rho^2}{\left(\sigma_1^2\sigma_2^2 - \rho^2\right)^2}
\end{bmatrix},$$

$$\mathbf{D} = \begin{bmatrix}
-\frac{x_1^2}{\sigma_1^2} & 0\\
-\frac{x_2^2}{\sigma_2^2} & 0\\
\frac{\rho\left(2x_1x_2\sigma_1^2\sigma_2^2 - \rho x_1^2\sigma_2^2 - \rho\sigma_1^2 x_2^2\right)}{\left(\sigma_1^2\sigma_2^2 - \rho^2\right)\sigma_1^2\sigma_2^2} & 0\\
0 & -\frac{\sigma_1^2\sigma_2^2 + \rho^2}{\left(\sigma_1^2\sigma_2^2 - \rho^2\right)^2}
\end{bmatrix}.$$

**C** is singular. $\mathbb{V}_{\text{MLE}} = \mathbb{V}_{\text{IQMLE}} = \frac{\sigma_1^2\sigma_2^2-\rho^2}{x_2^2\sigma_1^2+x_1^2\sigma_2^2-2\rho x_1 x_2}$, which is also the variance of the GLS estimator of $\beta$, $(\mathbf{x'\Sigma^{-1}x})^{\mathbf{-1}}$.

**Bivariate Normal with common variance.** Assume Normal marginal densities with $\sigma_1^2 = \sigma_2^2 = \sigma^2$ and $\mu_1 = \mu_2 = 0$

$$f_1(x_1;\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x_1^2}{2\sigma^2}},$$

$$f_2(x_2;\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x_2^2}{2\sigma^2}}.$$

Again, let the true joint distribution be Normal, i.e.,

$$h(x_1, x_2; \sigma, \rho) = \frac{1}{2\pi\sqrt{\sigma^4-\rho^2}} e^{-\frac{x_1^2\sigma^2-2x_1x_2\rho+x_2^2\sigma^2}{2(\sigma^4-\rho^2)}}.$$

Then, the implied copula is Normal,

$$c(F_1(x_1;\sigma), F_2(x_2;\sigma);\rho) = \frac{\sigma^2}{\sqrt{\sigma^4-\rho^2}} e^{-\frac{\rho\left(x_1^2\rho+x_2^2\rho-2\sigma^2x_1x_2\right)}{2(\sigma^4-\rho^2)\sigma^2}}.$$

The relevant moment conditions are

$$\begin{array}{ll}
\mathbb{E}\left\{\frac{{X_1}^2-\sigma^2}{2\sigma^4}\right\} = 0 & (A)\\
\mathbb{E}\left\{\frac{{X_2}^2-\sigma^2}{2\sigma^4}\right\} = 0 & (B)\\
\mathbb{E}\left\{-\frac{\left(\left(3\rho\sigma^4-\rho^3\right)(X_1^2+X_2^2)-4\sigma^6X_1X_2-2\sigma^2\rho(\sigma^4-\rho^2)\right)\rho}{2(\sigma^2-\rho)^2(\sigma^2+\rho)^2\sigma^2}\right\} = 0. & (C)\\
\mathbb{E}\left\{-\frac{\rho\sigma^2(X_1^2+X_2^2)-(\rho^2+\sigma^4)X_1X_2-\rho(\sigma^4-\rho^2)}{(\sigma^2+\rho)^2(\sigma^2-\rho)^2}\right\} = 0. & (D)
\end{array}$$

(C) is not a linear combination of (A) and (B). But (C) is a linear combination of (A), (B), and (D). Thus (24) holds and **C** is singular.

$$\mathbf{C} = \left[\begin{array}{cccc}
\frac{1}{2\sigma^4} & \frac{\rho^2}{2\sigma^8} & -\frac{\rho^2}{2\sigma^8} & 0\\
\frac{\rho^2}{2\sigma^8} & \frac{1}{2\sigma^4} & -\frac{\rho^2}{2\sigma^8} & 0\\
-\frac{\rho^2}{2\sigma^8} & -\frac{\rho^2}{2\sigma^8} & \frac{\rho^2(4\sigma^8-3\sigma^4\rho^2+\rho^4)}{\sigma^8(\sigma^2-\rho)^2(\sigma^2+\rho)^2} & -\frac{2\sigma^2\rho}{(\sigma^2-\rho)^2(\sigma^2+\rho)^2}\\
0 & 0 & -\frac{2\sigma^2\rho}{(\sigma^2-\rho)^2(\sigma^2+\rho)^2} & \frac{\sigma^4+\rho^2}{(\sigma^2-\rho)^2(\sigma^2+\rho)^2}
\end{array}\right],$$

$$\mathbf{D} = \left[\begin{array}{cc}
-\frac{1}{2\sigma^4} & 0\\
-\frac{1}{2\sigma^4} & 0\\
\frac{\rho^2(\rho^2-3\sigma^4)}{\sigma^4(\sigma^2-\rho)^2(\sigma^2+\rho)^2} & \frac{2\sigma^2\rho}{(\sigma^2-\rho)^2(\sigma^2+\rho)^2}\\
\frac{2\sigma^2\rho}{(\sigma^2-\rho)^2(\sigma^2+\rho)^2} & -\frac{\sigma^4+\rho^2}{(\sigma^2-\rho)^2(\sigma^2+\rho)^2}
\end{array}\right].$$

By Theorem 5.2, IQMLE of $\sigma^2$ is efficient, in fact $\mathbb{V}_{\text{MLE}} = \mathbb{V}_{\text{IQMLE}} = \sigma^4 + \rho^2$.

**Farlie-Gumbel-Morganstern copula with general marginals.** For $i = \{1, 2\}$, denote the marginal p.d.f.'s and c.d.f.'s by

$$f_i \equiv f_i(x_i; \theta)$$

and

$$F_i \equiv F_i(x_i; \theta) = \int_{-\infty}^{x_i} f_i(z; \theta) dz$$

respectively.

Assume the FGM copula. Then

$$c(u, v; \rho) = 1 + \rho - 2\rho u - 2\rho v + 4\rho uv$$

Our moment conditions are now

$$\begin{array}{rr} \mathbb{E}\left\{ \frac{1}{f_1} \frac{\partial f_1}{\partial \theta} \right\} = 0 & (A) \\ \mathbb{E}\left\{ \frac{1}{f_2} \frac{\partial f_2}{\partial \theta} \right\} = 0 & (B) \\ \mathbb{E}\left\{ \frac{2\rho f_1 + 2\rho f_2 - 4\rho f_1 F_2 - 4\rho F_2 F_1}{1+\rho-2\rho F_1 - 2\rho F_2 + 4\rho F_1 F_2} \right\} = 0 & (C) \\ \mathbb{E}\left\{ \frac{1 - F_1 - F_2 + 4F_1 F_2}{1+\rho-2\rho F_1 - 2\rho F_2 + 4\rho F_1 F_2} \right\} = 0 & (D) \end{array}$$

In general, (C) is *not* a linear combination of (A), (B) or (A), (B) and (D). So the copula based terms are not redundant in general and IQMLE is generally inefficient.

# 6 Validity of copula terms

Suppose we are ready to assume the correctness of the marginal distributions (the marginal moments in (3)) but are doubtful about the correctness of the joint distribution (the copula moments in (3)). One may test the validity of a copula by testing the validity of the moment restrictions (C) and (D) in (3). There are at least two ways to do that.

It was noted earlier that the moment conditions in (3) are usually overidentified. There are at least as many marginal moments as marginal parameters (or more if the marginal distributions share parameters), plus there are as many copula moments as there are parameters in total. Since the parameters are overidentified, the moment conditions in (3) imply restrictions. Consequently, if the model that led to the moment conditions is incorrect (i.e., the assumed joint distribution is wrong) then at least some of the moment conditions will be systematically violated in the sample. This suggests the possibility for testing copula validity by a test of the overidentifying restrictions (see, e.g., Hansen, 1982; Newey and West, 1987).

We will need more notation. For $m = 1, 2$ and $i = 1, \ldots, N$, denote $f_{mi}(\theta) = f_m(x_{1i}; \theta)$, $c_i(\theta, \rho) = c(F_1(x_{1i}; \theta), F_2(x_{2i}; \theta); \rho)$,

$$\psi_i(\theta, \rho) = \begin{bmatrix} \frac{\partial}{\partial \theta} \ln f_{1i}(\theta) \\ \frac{\partial}{\partial \theta} \ln f_{2i}(\theta) \\ \frac{\partial}{\partial \theta} \ln c_i(\theta, \rho) \\ \frac{\partial}{\partial \rho} \ln c_i(\theta, \rho) \end{bmatrix}, \quad g_i(\theta) = \begin{bmatrix} \frac{\partial}{\partial \theta} \ln f_{1i}(\theta) \\ \frac{\partial}{\partial \rho} \ln f_{2i}(\theta) \end{bmatrix}, \quad r_i(\theta, \rho) = \begin{bmatrix} \frac{\partial}{\partial \theta} \ln c_i(\theta, \rho) \\ \frac{\partial}{\partial \rho} \ln c_i(\theta, \rho) \end{bmatrix}.$$

Note that $\psi_i$ is a (3p+q)-vector. Let

$$\bar{\psi}(\theta, \rho) \equiv \frac{1}{N} \sum_{i=1}^{N} \psi_i(\theta, \rho), \quad \bar{g}(\theta) \equiv \frac{1}{N} \sum_{i=1}^{N} g_i(\theta), \quad \bar{r}(\theta, \rho) \equiv \frac{1}{N} \sum_{i=1}^{N} r_i(\theta, \rho).$$

Following our previous notation, let $\mathbf{C_o} \equiv \mathbb{E}\psi(\theta_o, \rho_o)\psi(\theta_o, \rho_o)'$, $\mathbf{C^o_{11}} \equiv \mathbb{E}g(\theta_o)g(\theta_o)'$, $\mathbf{C^o_{22}} \equiv \mathbb{E}r(\theta_o, \rho_o)r(\theta_o, \rho_o)'$, $\mathbf{C^o_{12}} = \mathbf{C^o_{21}}' \equiv \mathbb{E}g(\theta_o)r(\theta_o, \rho_o)'$, and $\mathbf{D_o} \equiv \mathbb{E}\frac{\partial}{\partial(\theta', \rho')}\psi(\theta_o)$, $\mathbf{D^o_{11}} \equiv \mathbb{E}\frac{\partial}{\partial \theta'}g(\theta_o)$, $\mathbf{D^o_{21}} \equiv \mathbb{E}\frac{\partial}{\partial \theta'}r(\theta_o, \rho_o)$, $\mathbf{D^o_{22}} \equiv \mathbb{E}\frac{\partial}{\partial \rho'}r(\theta_o, \rho_o)$, where expectations are with respect to the joint density $h(x_1, x_2)$.

**Proposition 6.1** *Let $(\breve{\theta}, \breve{\rho})$ denote the optimal GMM estimate of $(\theta, \rho)$ based on (3). Then*

$$N\bar{\psi}(\breve{\theta}, \breve{\rho})'\mathbf{C_o^{-1}}\bar{\psi}(\breve{\theta}, \breve{\rho}) \overset{a}{\sim} \chi^2_{2p}. \tag{29}$$

This test is a specification test which, given that the marginal distributions are correct, should capture copula misspecification. A consistent estimator of $\mathbf{C_o}$ such as

$$\breve{\mathbf{C}}_{\mathbf{o}} = \frac{1}{N} \sum_{i=1}^{N} \psi_i(\breve{\theta}, \breve{\rho})\psi_i(\breve{\theta}, \breve{\rho})'$$

is usually used in (29). It is however important to note that the statistic in (29) can be used only if $\mathbf{C}$ in non-singular, i.e. if copula terms are not redundant.

The second way to test copula validity we propose is based on a two step procedure.

**Proposition 6.2** *Let $\hat{\theta}$ be the optimal GMM estimate based on $\mathbb{E}g(\theta) = 0$. Let $\hat{\rho}$ be obtained by minimizing $\bar{r}(\hat{\theta}, \rho)'\mathbf{B_o^{-1}}\bar{r}(\hat{\theta}, \rho)$, where*

$$\begin{aligned} \mathbf{B_o} &= \mathbf{C^o_{22}} \quad -\mathbf{D^o_{21}}(\mathbf{D^o_{11}}\mathbf{C^o_{11}}^{-1}\mathbf{D^o_{11}})^{-1}\mathbf{D^o_{11}}'\mathbf{C^o_{11}}^{-1}\mathbf{C^o_{12}} \\ &\qquad -\mathbf{C^o_{21}}\mathbf{C^o_{11}}^{-1}\mathbf{D^o_{11}}(\mathbf{D^o_{11}}\mathbf{C^o_{11}}^{-1}\mathbf{D^o_{11}})^{-1}\mathbf{D^o_{21}}' \\ &\qquad +\mathbf{D^o_{21}}(\mathbf{D^o_{11}}\mathbf{C^o_{11}}^{-1}\mathbf{D^o_{11}})^{-1}\mathbf{D^o_{21}}'. \end{aligned}$$

*Then,*

$$N\bar{r}(\hat{\theta}, \hat{\rho})'\mathbf{B_o^{-1}}\bar{r}(\hat{\theta}, \hat{\rho}) \overset{a}{\sim} \chi^2_p. \tag{30}$$

Similarly to Proposition 6.1, consistent estimates of the elements of $\mathbf{C_o}$ and $\mathbf{D_o}$ will be used in practice for calculating the test statistic in (30).

# 7 Concluding remarks

We have proposed considering likelihood-based models in a GMM setting, in which knowledge about the joint distribution can be represented as copula moment conditions and efficiency and robustness of estimators can be assessed in terms of redundancy and robustness of the copula moments.

In considering copula robustness, all of the copula families that we compared to the normal benchmark except the Frank family are not comprehensive, i.e., they do not cover all possible values of the dependence measure $\tau$. This makes such copula families relevant for modelling only certain degrees of dependence to which our robustness comparisons would apply.

For the Frank and Normal families, $\tau \in (-1, 1)$, so they are comprehensive. Given the simulation results, the Frank copula appears as useful in modelling any degree of dependence as the Normal family. It would be desirable to make comparisons with other comprehensive copulas such as the Plackett family. Similarly, comparisons of the Logistic copula to copulas with the same coverage should reveal its relative robustness.

The behavior of the AMH family of copulas was quite similar to that of the FGM family in our simulation. This was due to the small value of the dependence parameter $\rho$. The first order approximation in (16) is in this case quite accurate. It may not be so for larger $\rho$.

Finally, our results on copula robustness are problem-specific. For example, they are generally inapplicable to problems involving higher moments of a distribution. In similar simulations with problems other then sample-mean problems, radially symmetric copulas may not be robust to misspecification, but it should still be possible to compare robustness properties of copula families since the true copula is known.

# References

Bouyé, E., V. Durrleman, A. Nikeghbali, G. Riboulet, and T. Roncalli (2000): "Copulas for Finance: A Reading Guide and Some Applications," *Crédit Lyonnais Working Paper*, gro.creditlyonnais.fr/content/wp/copula-survey.pdf.

Breusch, T., H. Qian, P. Schmidt, and D. Wyhowski (1999): "Redundancy of moment conditions," *Journal of Econometrics*, 91, 89–111.

Breymann, W., A. Dias, and P. Embrechts (2003): "Dependence structures for multivariate high-frequency data in finance," *Quantitative Finance*, 3, 1–14, http://www.iop.org/EJ/abstract/1469-7688/3/1/301/.

Cameron, A. C., T. Li, P. K. Trivedi, and D. M. Zimmer (2004): "Modelling the Differences in Counted Outcomes Using Bivariate Copula Models with Application to Mismeasured Counts," *Econometrics Journal*, 7, 566–84.

EMBRECHTS, P., A. HÖING, AND A. JURI (2003): "Using copulae to bound the Value-at-Risk for functions of dependent risks," *Finance and Stochastics*, 7, 145 167.

EMBRECHTS, P., A. MCNEIL, AND D. STRAUMANN (2002): "Correlation and dependence in risk management: properties and pitfalls," in *Risk Management: Value at Risk and Beyond*, ed. by M. Dempster, Cambridge: Cambridge University Press, 176223.

FRANK, M. (1979): "On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$," *Aequationes Mathematicae*, 19, 194–226.

GODAMBE, V. P. (1960): "An optimum property of regular maximum likelihood estimation," *The Annals of Mathematical Statistics*, 31, 1208–1211.

——— (1976): "Conditional likelihood and unconditional optimum estimating equations," *Biometrika*, 63, 277–284.

GOURIEROUX, C., A. MONFORT, AND A. TROGNON (1984): "Pseudo maximum likelihood methods: theory," *Econometrica*, 52, 681–700.

GREEN, W. (2002): *Econometric Analysis*, Prentice Hall.

HANSEN, L. (1982): "Large sample properties of generalized method of moments estimators," *Econometrica*, 50, 1029–1054.

KENDALL, M. G. (1949): "Rank and product-moment correlation," *Biometrika*, 36, 177–193.

LEE, L.-F. (1983): "Generalized econometric models with selectivity," *Econometrica*, 51, 507–512.

MARDIA, K., J. KENT, AND J. BIBBY (1979): *Multivariate Analysis*, Probability and Mathematical Statistics, London: Academic Press.

NELSEN, R. B. (1993): "Some concepts of bivariate symmetry," *Journal of Nonparametric Statistics*, 3, 95–101.

——— (1999): *An Introduction to Copulas*, vol. 139 of *Lecture Notes in Statistics*, Springer.

NEWEY, W. AND K. WEST (1987): "Hypothesis testing with efficient method of moments estimation," *International Economic Review*, 28, 777–787.

PATTON, A. (2001): "Modelling time-varying exchange rate dependence using the conditional copula," *UC San Diego Department of Economics Discussion Paper 2001-09*.

SCHMIDT, P. (2004): "Likelihood-based estimation in a panel setting," *MSU Working Paper*.

SKLAR, A. (1959): "Fonctions de répartition à n dimensions et leurs marges," *Publications de l'Institut de Statistique de l'Université de Paris*, 8, 229–231.

SMITH, M. D. (2003): "Modelling sample selection using Archimedian copulas," *Econometrics Journal*, 6, 99–123.

WHITE, H. (1982): "Maximum likelihood estimation of misspecified models," *Econometrica*, 50, 1–26.

Wooldridge, J. (1994): "Estimation and inference for dependent processes," in *Handbook of Econometrics*, ed. by R. Engle and D. McFadden, vol. IV.

——— (2002): *Econometric analysis of cross section and panel data*, Cambridge, Mass.: MIT Press.

# A Selected copula families

1. Independence copula:

$$\begin{aligned} C(u,v) &= uv \\ c(u,v) &= 1 \\ \tau &= 0 \end{aligned}$$

2. Logistic copula:

$$\begin{aligned} C(u,v) &= \frac{uv}{u+v-uv} \\ c(u,v) &= \frac{2uv}{(u+v-uv)^3} \\ \tau &= 1/3 \end{aligned}$$

3. Farlie-Gumbel-Morgenstern family:

$$\begin{aligned} C(u,v,\rho) &= uv(1+\rho(1-u)(1-v)) \\ c(u,v,\rho) &= 1+\rho-2\rho u-2\rho v+4\rho uv \\ \rho &\in [-1;1] \\ \tau &= 2\rho/9 \in [-2/9, 2/9] \end{aligned}$$

4. Joe family$^*$:

$$\begin{aligned} C(u,v,\rho) &= 1-((1-u)^\rho+(1-v)^\rho-(1-u)^\rho(1-v)^\rho)^{1/\rho} \\ \rho &\in [1,\infty) \\ \varphi(t) &= -\log(1-(1-t)^\rho) \\ \tau &\in [0,1) \end{aligned}$$

5. Ali-Mikhail-Haq family$^*$:

$$\begin{aligned} C(u,v,\rho) &= \frac{uv}{1-\rho(1-u)(1-v)} \\ \rho &\in [-1,1) \\ \varphi(t) &= \log\frac{1-\rho(1-t)}{t} \\ \tau &\in [-0.182, 1/3) \end{aligned}$$

6. Clayton family$^*$:

$$\begin{aligned} C(u,v,\rho) &= \begin{cases} uv, & \rho = 0 \\ (u^{-\rho}+v^{-\rho}-1)^{-1/\rho}, & \rho \neq 0 \end{cases} \\ \rho &\in [0,\infty) \\ \varphi(t) &= \frac{1}{\rho}(t^{-\rho}-1) \\ \tau &= \frac{\rho}{\rho+2} \in [0,1) \end{aligned}$$

7. Gumbel family$^*$:

$$C(u,v,\rho) = \exp\left[-((-\ln u)^\rho + (-\ln v)^\rho)^{1/\rho}\right]$$
$$\rho \in [1,\infty)$$
$$\varphi(t) = (-\log t)^\rho$$
$$\tau = \frac{\rho - 1}{\rho} \in [0,1)$$

8. Frank family$^*$:

$$C(u,v,\rho) = \begin{cases} uv, & \rho = 0 \\ -\frac{1}{\rho}\ln\left[1 + \frac{(e^{-\rho u}-1)(e^{-\rho v}-1)}{e^{-\rho}-1}\right], & \rho \neq 0 \end{cases}$$
$$\rho \in (-\infty,\infty)$$
$$\varphi(t) = -\ln\frac{e^{-\rho t}-1}{e^{-\rho}-1}$$
$$\tau \in (-1,1)$$

9. Plackett family:

$$C(u,v,\rho) = \begin{cases} uv, & \rho = 1 \\ \frac{\left(1+(u+v)(\rho-1)-\sqrt{(1+(u+v)(\rho-1))^2-4uv\rho(\rho-1)}\right)}{2(\rho-1)}, & \rho \neq 1 \end{cases}$$
$$\rho \in (0,\infty)$$
$$\tau \in (-1,1)$$

10. Normal family:

$$C(u,v,\rho) = \Phi_2(\Phi^{-1}(u), \Phi^{-1}(v); \rho)$$
$$\rho \in (-1,1)$$
$$\tau = \frac{2}{\pi}\arcsin\rho \in (-1,1)$$

Note: $^*$ denotes Archimedean copulas, i.e. copulas generated as

$$C(u,v) = \varphi^{-1}(\varphi(u) + \varphi(v)),$$

where $\varphi : \mathbb{I} \to [0,\infty]$, continuous, $\varphi'(t) < 0$ and $\varphi''(t) > 0$ $\forall t \in (0,1)$ is called the generator function. It can be shown (see, e.g., Nelsen, 1999, p.130) that for Archimedean copulas, Kendall's

$$\tau = 1 + 4\int_0^1 \frac{\varphi(t)}{\varphi'(t)}dt.$$

# B Proofs

PROOF OF THEOREM 4.1:

We show that $\mathbb{E}\frac{\partial}{\partial\mu}\ln k(F_1(\mu_1+x_1), F_2(\mu_2+x_2); \rho^k) = 0$, where $\mu = (\mu_1,\mu_2)'$, holds for any RS $K$.

By the chain rule, $\frac{\partial}{\partial\mu}\ln k(F_1(\mu_1+x_1), F_2(\mu_2+x_2); \rho^k)$ contains terms of the form

$$\frac{1}{k(F_1(\mu_1+x_1), F_2(\mu_2+x_2); \rho^k)} \times \frac{\partial k(F_1(\mu_1+x_1), F_2(\mu_2+x_2); \rho^k)}{\partial F_i(\mu_i+x_i)} \times f_i(\mu_i+x_i), \quad i = 1,2. \tag{31}$$

Due to MS of $(X_1, X_2)$ and RS of $K$, $f_i(\mu_i + x_i) = f_i(\mu_i - x_i)$ and $k(F_1(\mu_1+x_1), F_2(\mu_2+x_2)) = k(1-F_1(\mu_1+x_1), 1-F_2(\mu_2+x_2)) = k(F_1(\mu_1-x_1), F_2(\mu_2-x_2))$. So the first term in (31) is the same whether evaluated at $(x_1,x_2)$ or $(-x_1,-x_2)$. Similarly, the last term is the same whether evaluated at $x_i$ or $-x_i$.

Furthermore,

$$\begin{array}{rcl}\frac{\partial k(F_1(\mu_1+x_1),F_2(\mu_2+x_2)'\rho^k)}{\partial F_i(\mu_i+x_i)} & = & \frac{\partial k(1-F_1(\mu_1+x_1),1-F_2(\mu_2+x_2);\rho^k)}{\partial(1-F_i(\mu_i-x_i))} \\ & = & -\frac{\partial k(F_1(\mu_1-x_1),F_2(\mu_2-x_2);\rho^k)}{\partial F_i(\mu_i-x_i)}.\end{array}$$

Thus, $\frac{\partial}{\partial\mu}\ln k(F_1(\mu_1+x_1),F_2(\mu_2+x_2);\rho^k) = -\frac{\partial}{\partial\mu}\ln k(F_1(\mu_1-x_1),F_2(\mu_2-x_2);\rho^k)$.

Denote $g(x_1,x_2)\equiv\frac{\partial}{\partial\mu}\ln k(F_1(\mu_1+x_1),F_2(\mu_2+x_2);\rho^k)\cdot h(\mu_1+x_1,\mu_2+x_2;\rho)$. From the above, it follows with RS that $g(-x_1,-x_2) = -g(x_1,x_2)$.

We thus have

$$\begin{array}{rcl}\mathbb{E}\frac{\partial}{\partial\mu}\ln k(F_1(\mu_1+x_1),F_2(\mu_2+x_2);\rho^k) & = & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}g(x_1,x_2)dx_1dx_2 \\ & = & \int_{-\infty}^{0}\int_{-\infty}^{0}g(x_1,x_2)dx_1dx_2 \\ & + & \int_{-\infty}^{0}\int_{0}^{\infty}g(x_1,x_2)dx_1dx_2 \\ & + & \int_{0}^{\infty}\int_{0}^{\infty}g(x_1,x_2)dx_1dx_2 \\ & + & \int_{0}^{\infty}\int_{-\infty}^{0}g(x_1,x_2)dx_1dx_2 \\ & = & \int_{0}^{\infty}\int_{0}^{\infty}g(-x_1,-x_2)dx_1dx_2 \\ & + & \int_{0}^{\infty}\int_{-\infty}^{0}g(-x_1,-x_2)dx_1dx_2 \\ & + & \int_{0}^{\infty}\int_{0}^{\infty}g(x_1,x_2)dx_1dx_2 \\ & + & \int_{0}^{\infty}\int_{-\infty}^{0}g(x_1,x_2)dx_1dx_2 \\ & = & 0.\end{array}$$

Proof of Lemma 5.1: By the information matrix equality (IME),

$$\mathbf{A}\equiv\mathbb{E}\left\{\frac{\partial}{\partial\theta}\ln f_1(x_1;\theta)\frac{\partial}{\partial\theta'}\ln f_1(x_1;\theta)\right\} = -\mathbb{E}\frac{\partial^2}{\partial\theta\partial\theta'}\ln f_1(x_1;\theta). \tag{32}$$

Similar for $\mathbf{B},\mathbf{F}$.

By the generalized IME (GIME),

$$\mathbf{E}\equiv\mathbb{E}\left\{\frac{\partial}{\partial\theta}\ln c(F_1(x_1;\theta),F_2(x_2;\theta);\rho)\frac{\partial}{\partial\rho'}\ln c(F_1(x_1;\theta),F_2(x_2;\theta);\rho)\right\} = -\mathbb{E}\frac{\partial^2}{\partial\theta\partial\rho'}\ln c(F_1(x_1;\theta),F_2(x_2;\theta);\rho) \tag{33}$$

and, for $i=1,2$,

$$\mathbb{E}\left\{\frac{\partial}{\partial\theta}\ln f_i(x_i;\theta)\frac{\partial}{\partial\rho'}\ln c(F_1(x_1;\theta),F_1(x_2;\theta);\rho)\right\} = -\mathbb{E}\frac{\partial^2}{\partial\theta\partial\rho'}\ln f_i(x_i;\theta) = \mathbf{0}.$$

Also by GIME and (2),

$$\mathbb{E}\left\{\frac{\partial}{\partial\theta}\ln f_i(x_i;\theta)\frac{\partial}{\partial\theta'}\left[\ln f_1(x_1;\theta)+\ln f_2(x_2;\theta)+\ln c(F_1(x_1;\theta),F_1(x_2;\theta);\rho)\right]\right\} = -\mathbb{E}\frac{\partial^2}{\partial\theta\partial\theta'}\ln f_i(x_i;\theta)$$

for $i=1,2$, which, along with (32), implies that

$$\mathbf{G}\equiv\mathbb{E}\left\{\frac{\partial}{\partial\theta}\ln f_1(x_1;\theta)\frac{\partial}{\partial\theta'}\ln f_2(x_2;\theta)\right\} = -\mathbb{E}\left\{\frac{\partial}{\partial\theta}\ln f_1(x_1;\theta)\frac{\partial}{\partial\theta'}\ln c(F_1(x_1;\theta),F_1(x_2;\theta);\rho)\right\}$$

and

$$\mathbb{E}\left\{\frac{\partial}{\partial\theta}\ln f_2(x_2;\theta)\frac{\partial}{\partial\theta'}\ln f_1(x_1;\theta)\right\} = -\mathbb{E}\left\{\frac{\partial}{\partial\theta}\ln f_2(x_2;\theta)\frac{\partial}{\partial\theta'}\ln c(F_1(x_1;\theta),F_1(x_2;\theta);\rho)\right\} = \mathbf{G}'.$$

Finally, by GIME and (2),

$$\mathbb{E}\left\{\tfrac{\partial}{\partial\theta}\ln c(F_1(x_1;\theta),F_1(x_2;\theta);\rho)\tfrac{\partial}{\partial\theta'}\left[\ln f_1(x_1;\theta)+\ln f_2(x_2;\theta)+\ln c(F_1(x_1;\theta),F_1(x_2;\theta);\rho)\right]\right\} = \\ = -\mathbb{E}\tfrac{\partial^2}{\partial\theta\partial\theta'}\ln c(F_1(x_1;\theta),F_1(x_2;\theta);\rho).$$

With $\mathbf{G}$ as defined above and

$$\mathbf{J}\equiv\mathbb{E}\left\{\frac{\partial}{\partial\theta}\ln c(F_1(x_1;\theta),F_1(x_2;\theta);\rho)\frac{\partial}{\partial\theta'}\ln c(F_1(x_1;\theta),F_1(x_2;\theta);\rho)\right\}$$

this implies that

$$\mathbb{E}\frac{\partial^2}{\partial\theta\partial\theta'}\ln c(F_1(x_1;\theta),F_1(x_2;\theta);\rho)=\mathbf{G}+\mathbf{G}'-\mathbf{J}.$$

Proof of Theorem 5.1: See text.

Proof of Theorem 5.2: By (20) and (23),

$$\mathbb{V}_{\text{MLE}}=\begin{bmatrix}\mathbf{A}+\mathbf{B}+\mathbf{J}-\mathbf{G}-\mathbf{G}' & \mathbf{E}'\\ \mathbf{E} & \mathbf{F}\end{bmatrix}^{-1},\quad \mathbb{V}_{\text{IQMLE}}=\left(\begin{bmatrix}-\mathbf{A} & -\mathbf{B}\end{bmatrix}\begin{bmatrix}\mathbf{A} & \mathbf{G}\\ \mathbf{G}' & \mathbf{B}\end{bmatrix}^{-1}\begin{bmatrix}-\mathbf{A}\\ -\mathbf{B}\end{bmatrix}\right)^{-1}.$$

Using partitioned inverse formulas, the upper left $p\times p$ block of $\mathbb{V}_{\text{MLE}}$ can be written as $\mathbf{\Sigma^{-1}}$, where $\mathbf{\Sigma}=\mathbf{A}+\mathbf{B}+\mathbf{J}-\mathbf{G}-\mathbf{G}'-\mathbf{E}'\mathbf{F^{-1}}\mathbf{E}$.

Also,

$$\begin{aligned}\mathbb{V}^{-1}_{\text{IQMLE}} &= \left([-\mathbf{G}'\quad -\mathbf{G}]+[\mathbb{I}\quad \mathbb{I}]\begin{bmatrix}\mathbf{A} & \mathbf{G}\\ \mathbf{G}' & \mathbf{B}\end{bmatrix}\right)\begin{bmatrix}\mathbf{A} & \mathbf{G}\\ \mathbf{G}' & \mathbf{B}\end{bmatrix}^{-1}\left(\begin{bmatrix}\mathbf{A} & \mathbf{G}\\ \mathbf{G}' & \mathbf{B}\end{bmatrix}\begin{bmatrix}\mathbb{I}\\ \mathbb{I}\end{bmatrix}+\begin{bmatrix}-\mathbf{G}\\ -\mathbf{G}'\end{bmatrix}\right)\\ &= [-\mathbf{G}'\quad -\mathbf{G}]\begin{bmatrix}\mathbf{A} & \mathbf{G}\\ \mathbf{G}' & \mathbf{B}\end{bmatrix}^{-1}\begin{bmatrix}-\mathbf{G}\\ -\mathbf{G}'\end{bmatrix}-\mathbf{G}'-\mathbf{G}+\mathbf{A}+\mathbf{G}+\mathbf{G}'+\mathbf{B}-\mathbf{G}-\mathbf{G}'. \qquad (34)\end{aligned}$$

Thus, $\mathbb{V}^{-1}_{\text{IQMLE}}=\mathbf{\Sigma}$ if and only if

$$\mathbf{J}-\mathbf{E}'\mathbf{F^{-1}}\mathbf{E}=[-\mathbf{G}'\quad -\mathbf{G}]\begin{bmatrix}\mathbf{A} & \mathbf{G}\\ \mathbf{G}' & \mathbf{B}\end{bmatrix}^{-1}\begin{bmatrix}-\mathbf{G}\\ -\mathbf{G}'\end{bmatrix}.$$

Proof of Corollary 5.1:

1. If (C) is a linear combination of (A) and (B) then covariances between moment functions in (C) and (D) are linear combinations of covariances between (D) and (A-B), which are all zero by Lemma 5.1.
2. Rewrite $\mathbf{J}-\mathbf{C}^{\theta}_{21}\mathbf{C}^{-1}_{11}\mathbf{C}^{\theta}_{12}$ as

   $$\mathbb{E}\left\{\left(\frac{\partial}{\partial\theta}\ln c-\mathbf{C}^{\theta}_{21}\mathbf{C}^{-1}_{11}\begin{bmatrix}\frac{\partial}{\partial\theta}\ln f_1\\ \frac{\partial}{\partial\theta}\ln f_2\end{bmatrix}\right)\frac{\partial}{\partial\theta'}\ln c\right\}.$$

   This is identically zero because, due to linearity of (C) in (A-B),

   $$\frac{\partial}{\partial\theta}\ln c-\mathbf{C}^{\theta}_{21}\mathbf{C}^{-1}_{11}\begin{bmatrix}\frac{\partial}{\partial\theta}\ln f_1\\ \frac{\partial}{\partial\theta}\ln f_2\end{bmatrix}=\mathbf{0}.$$

3. By Theorem 5.2.

Proof of Lemma 5.2: By construction, blocks $\mathbf{A}, \mathbf{B}, \mathbf{G}$ of matrices $\mathbf{C^k}$ and $\mathbf{D^k}$ are the same as in Lemma 5.1. However, GIME does not apply now. (For convenience we use $\rho$ instead of $\rho^k$.)

$$\begin{aligned}\mathbf{V} \equiv \mathbb{E}\left\{\tfrac{\partial}{\partial\theta} \ln k(F_1(x_1;\theta), F_2(x_2;\theta);\rho) \tfrac{\partial}{\partial(\rho)'} \ln k(F_1(x_1;\theta), F_2(x_2;\theta);\rho)\right\} \\ \neq -\mathbb{E}\left\{\tfrac{\partial^2}{\partial\theta\partial\rho'} \ln k(F_1(x_1;\theta), F_2(x_2;\theta);\rho)\right\} \equiv -\mathbf{S}.\end{aligned}$$

$$-\mathbf{P} \equiv \mathbb{E}\left\{\frac{\partial}{\partial\theta} \ln f_1(x_1;\theta) \frac{\partial}{\partial\rho'} \ln c(F_1(x_1;\theta), F_1(x_2;\theta);\rho)\right\} \neq -\mathbb{E}\frac{\partial^2}{\partial\theta\partial\rho'} \ln f_1(x_1;\theta) = \mathbf{0}$$

and

$$-\mathbf{Q}' \equiv \mathbb{E}\left\{\frac{\partial}{\partial\theta} \ln f_2(x_2;\theta) \frac{\partial}{\partial\rho'} \ln c(F_1(x_1;\theta), F_1(x_2;\theta);\rho)\right\} \neq -\mathbb{E}\frac{\partial^2}{\partial\theta\partial\rho'} \ln f_2(x_2;\theta) = \mathbf{0}.$$

$$\mathbf{G} \equiv \mathbb{E}\left\{\frac{\partial}{\partial\theta} \ln f_1(x_1;\theta) \frac{\partial}{\partial\theta'} \ln f_2(x_2;\theta)\right\} \neq -\mathbb{E}\left\{\frac{\partial}{\partial\theta} \ln f_1(x_1;\theta) \frac{\partial}{\partial\theta'} \ln k(F_1(x_1;\theta), F_1(x_2;\theta);\rho)\right\} \equiv \mathbf{K}$$

and

$$\mathbb{E}\left\{\frac{\partial}{\partial\theta} \ln f_2(x_2;\theta) \frac{\partial}{\partial\theta'} \ln f_1(x_1;\theta)\right\} \neq -\mathbb{E}\left\{\frac{\partial}{\partial\theta} \ln f_2(x_2;\theta) \frac{\partial}{\partial\theta'} \ln k(F_1(x_1;\theta), F_1(x_2;\theta);\rho)\right\} \equiv \mathbf{L}'.$$

However, by GIME and (2),

$$\begin{aligned}&\mathbb{E}\tfrac{\partial^2}{\partial\theta\partial\theta'} \ln k(F_1(x_1;\theta), F_1(x_2;\theta);\rho) = \\ &-\mathbb{E}\left\{\tfrac{\partial}{\partial\theta} \ln k(F_1(x_1;\theta), F_1(x_2;\theta);\rho) \tfrac{\partial}{\partial\theta'} \left[\ln f_1(x_1;\theta) + \ln f_2(x_2;\theta) + \ln c(F_1(x_1;\theta), F_1(x_2;\theta);\rho)\right]\right\} \equiv \mathbf{K}' + \mathbf{L} - \mathbf{M},\end{aligned}$$

and

$$\begin{aligned}&\mathbb{E}\tfrac{\partial^2}{\partial\rho\partial\theta'} \ln k(F_1(x_1;\theta), F_1(x_2;\theta);\rho) = \\ &-\mathbb{E}\left\{\tfrac{\partial}{\partial\rho} \ln k(F_1(x_1;\theta), F_1(x_2;\theta);\rho) \tfrac{\partial}{\partial\theta'} \left[\ln f_1(x_1;\theta) + \ln f_2(x_2;\theta) + \ln c(F_1(x_1;\theta), F_1(x_2;\theta);\rho)\right]\right\} \equiv \mathbf{P}' + \mathbf{Q} - \mathbf{R},\end{aligned}$$

$$-\mathbf{T} \equiv \mathbb{E}\frac{\partial^2}{\partial\rho\partial\rho'} \ln k(F_1(x_1;\theta), F_1(x_2;\theta);\rho) = -\mathbb{E}\left\{\frac{\partial}{\partial\rho} \ln k(F_1(x_1;\theta), F_1(x_2;\theta);\rho) \frac{\partial}{\partial\rho'} \ln c(F_1(x_1;\theta), F_1(x_2;\theta);\rho)\right\},$$

and

$$-\mathbf{S} \equiv \mathbb{E}\frac{\partial^2}{\partial\theta\partial\rho'} \ln k(F_1(x_1;\theta), F_1(x_2;\theta);\rho) = -\mathbb{E}\left\{\frac{\partial}{\partial\theta} \ln k(F_1(x_1;\theta), F_1(x_2;\theta);\rho) \frac{\partial}{\partial\rho'} \ln c(F_1(x_1;\theta), F_1(x_2;\theta);\rho)\right\}.$$

Finally,

$$\mathbf{N} \equiv \mathbb{E}\left\{\frac{\partial}{\partial\theta} \ln k(F_1(x_1;\theta), F_1(x_2;\theta);\rho) \frac{\partial}{\partial\theta'} \ln k(F_1(x_1;\theta), F_1(x_2;\theta);\rho)\right\} \neq \mathbf{M}$$

and

$$\mathbf{W} \equiv \mathbb{E}\left\{\frac{\partial}{\partial\rho} \ln k(F_1(x_1;\theta), F_1(x_2;\theta);\rho) \frac{\partial}{\partial\rho'} \ln k(F_1(x_1;\theta), F_1(x_2;\theta);\rho)\right\} \neq \mathbf{T}.$$

Proof of Theorem 5.3: See main text.

Proof of Theorem 5.4: By Theorem 8(C) of Breusch et al. (1999), (C$'$-D$'$) are redundant for $\theta$ given (A-B) if and only if

$$\begin{bmatrix} \mathbf{K}' + \mathbf{L} - \mathbf{M} \\ \mathbf{P}' + \mathbf{Q} - \mathbf{R} \end{bmatrix} - \begin{bmatrix} -\mathbf{K}' & -\mathbf{L} \\ -\mathbf{P}' & -\mathbf{Q} \end{bmatrix} \mathbf{C_{11}^{-1} D_{11}} = \begin{bmatrix} -\mathbf{S} \\ -\mathbf{T} \end{bmatrix} \mathbb{B},$$

for some matrix $\mathbb{B} : q \times p$.

This is equivalent to

$$-\mathbf{M} - [-\mathbf{K}' \quad -\mathbf{L}]\mathbf{C_{11}^{-1}} \begin{bmatrix} \mathbf{G} \\ \mathbf{G}' \end{bmatrix} = -\mathbf{S}\mathbb{B},$$

$$-\mathbf{R} - [-\mathbf{P}' \quad -\mathbf{Q}]\mathbf{C_{11}^{-1}} \begin{bmatrix} \mathbf{G} \\ \mathbf{G}' \end{bmatrix} = -\mathbf{T}\mathbb{B}.$$

$\mathbf{T}$ is symmetric and invertible, so we can substitute $\mathbb{B}$ from the latter equation into the former to obtain

$$\mathbf{M} - [-\mathbf{K}' \quad -\mathbf{L}]\mathbf{C_{11}^{-1} C_{12}^{\theta}} = \mathbf{S}\mathbf{T}^{-1}(\mathbf{R} - [-\mathbf{P}' \quad -\mathbf{Q}]\mathbf{C_{11}^{-1} C_{12}^{\theta}}),$$

which completes the proof.

Proof of Corollary 5.2:

1. By (28), $\mathbf{M} - \mathbf{C_{21}^{\theta k} C_{11}^{-1} C_{12}^{\theta}}$ is identically zero under linearity of (C) in (A-B).
2. As in 1.
3. By Theorem 5.4.

Proof of Proposition 6.1: See proof of Lemma 4.2 of Hansen (1982).

Proof of Proposition 6.2: First note that, by standard optimal GMM results, $\hat{\theta}$ satisfies

$$\sqrt{N}(\hat{\theta} - \theta_o) = -(\mathbf{D^o_{11}}'\mathbf{C^o_{11}}^{\mathbf{-1}}\mathbf{D^o_{11}})^{-1}\mathbf{D^o_{11}}'\mathbf{C^o_{11}}^{\mathbf{-1}}\sqrt{N}\bar{g}(\theta_o) + o_p(1). \tag{35}$$

The first order condition for $\hat{\rho}$ can equivalently be written as

$$\begin{aligned} [\frac{\partial}{\partial \rho'}\bar{r}(\hat{\theta}, \hat{\rho})]'\mathbf{B_o}^{-1}\bar{r}(\hat{\theta}, \hat{\rho}) &= 0 \\ \mathbf{D^o_{22}}'\mathbf{B_o}^{-1}\sqrt{N}\bar{r}(\hat{\theta}, \hat{\rho}) &= o_p(1) \end{aligned} \tag{36}$$

Now, by the mean-value theorem, we have

$$\sqrt{N}\bar{r}(\hat{\theta}, \hat{\rho}) = \sqrt{N}\bar{r}(\theta_o, \rho_o) + \mathbf{D^o_{21}}\sqrt{N}(\hat{\theta} - \theta_o) + \mathbf{D^o_{22}}\sqrt{N}(\hat{\rho} - \rho_o) + o_p(1). \tag{37}$$

Substituting (35) into (37), pre-multiplying by $\mathbf{D^o_{22}}'\mathbf{B_o}^{-1}$, and solving for $\sqrt{N}(\hat{\rho} - \rho_o)$ using (36) yields

$$\begin{aligned} \sqrt{N}(\hat{\rho} - \rho_o) = & -(\mathbf{D^o_{22}}'\mathbf{B_o^{-1}}\mathbf{D^o_{22}})^{-1}\mathbf{D^o_{22}}'\mathbf{B_o^{-1}}\sqrt{N}\bar{r}(\theta_o, \rho_o) \\ & +(\mathbf{D^o_{22}}'\mathbf{B_o^{-1}}\mathbf{D^o_{22}})^{-1}\mathbf{D^o_{22}}'\mathbf{B_o^{-1}}\mathbf{D^o_{21}}(\mathbf{D^o_{11}}'\mathbf{C^o_{11}}^{\mathbf{-1}}\mathbf{D^o_{11}})^{-1}\mathbf{D^o_{11}}'\mathbf{C^o_{11}}^{\mathbf{-1}}\sqrt{N}\bar{g}(\theta_o) \\ & +o_p(1) \end{aligned} \tag{38}$$

Substituting (38) and (35) into (37) and simplifying results in

$$\sqrt{N}\bar{r}(\hat{\theta},\hat{\rho}) \quad = \quad \mathbf{R_o}\sqrt{N}\bar{\phi}(\theta_o,\rho_o) + o_p(1), \tag{39}$$

where

$$\begin{aligned}\mathbf{R_o} \quad &= \quad \mathbb{I} - \mathbf{D^o_{22}}(\mathbf{D^o_{22}}'\mathbf{B_o^{-1}}\mathbf{D^o_{22}})^{-1}\mathbf{D^o_{22}}'\mathbf{B_o^{-1}} \\ \bar{\phi}(\theta_o,\rho_o) \quad &= \quad \bar{r}(\theta_o,\rho_o) - \mathbf{D^o_{21}}(\mathbf{D^o_{11}}'\mathbf{C^o_{11}}^{-1}\mathbf{D^o_{11}})^{-1}\mathbf{D^o_{11}}'\mathbf{C^o_{11}}^{-1}\bar{g}(\theta_o)\end{aligned}$$

Note that $\sqrt{N}\bar{\phi}(\theta_o,\rho_o) \sim \mathbb{N}(\mathbf{0},\mathbf{B_o})$, and thus $\mathbf{B_o}^{-1/2}\sqrt{N}\bar{\phi}(\theta_o,\rho_o) \sim \mathbb{N}(\mathbf{0},\mathbb{I})$. Also, note that $\mathbf{R_o'}\mathbf{B_o^{-1}}\mathbf{R_o} = \mathbf{B_o}^{-1/2}[\mathbf{I} - \mathbf{B_o}^{-1/2}\mathbf{D^o_{22}}(\mathbf{D^o_{22}}'\mathbf{B_o^{-1}}\mathbf{D^o_{22}})^{-1}\mathbf{D^o_{22}}'\mathbf{B_o}^{-1/2}]\mathbf{B_o}^{-1/2}$.

Thus, the test statistic in (30) can be written as

$$N\bar{h}(\hat{\theta},\hat{\rho})'\mathbf{B_o^{-1}}\bar{h}(\hat{\theta},\hat{\rho}), \tag{40}$$

i.e. as a quadratic form in standard normals with the coefficient matrix

$$\mathbb{P} = \mathbf{I}_{p+q} - \mathbf{B_o}^{-1/2}\mathbf{D^o_{22}}(\mathbf{D^o_{22}}'\mathbf{B_o^{-1}}\mathbf{D^o_{22}})^{-1}\mathbf{D^o_{22}}'\mathbf{B_o}^{-1/2}. \tag{41}$$

This matrix is idempotent (it is in fact the projection matrix orthogonal to $\mathbf{B_o}^{-1/2}\mathbf{D^o_{22}}$). The $\chi^2$-test in (30) follows immediately because $tr(\mathbb{P}) = p + q - rank(D^o_{22}) = p$.

# C Plots of simulated sample moments

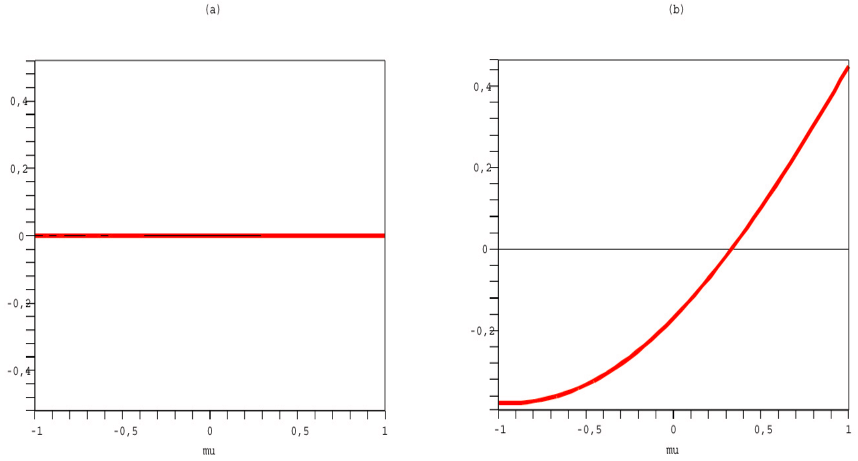

Figure 1: $\bar{\delta}^{\mu}(\mu)$ for no-parameter copulas: (a) Independence copula; (b) Logistic copula.

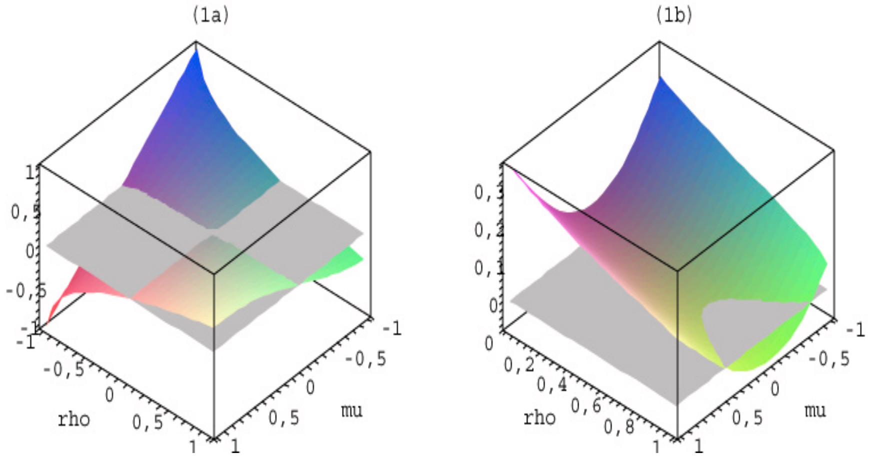

Figure 2: $\bar{\delta}^{\mu}(\mu, \rho)$ and $\bar{\delta}^{\rho}(\mu, \rho)$ for one-parameter copulas: (1) Farlie-Gumbel-Morganstern.

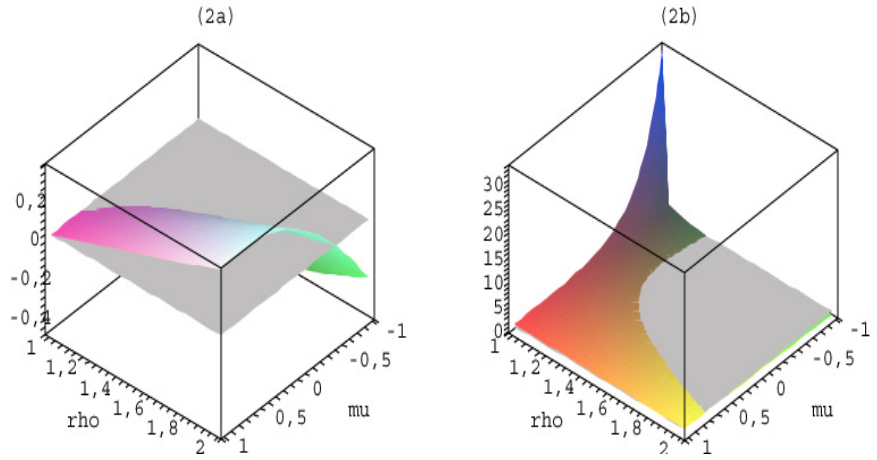

Figure 3: $\bar{\delta}^{\mu}(\mu,\rho)$ and $\bar{\delta}^{\rho}(\mu,\rho)$ for one-parameter copulas: (2) Joe.

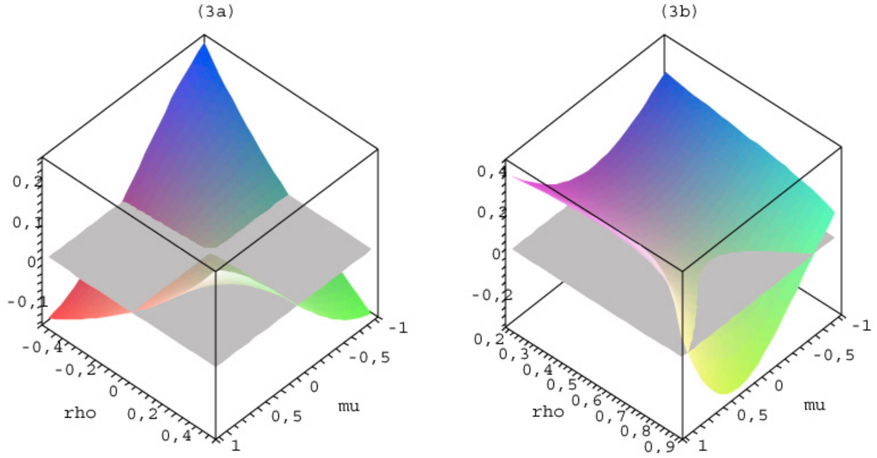

Figure 4: $\bar{\delta}^{\mu}(\mu,\rho)$ and $\bar{\delta}^{\rho}(\mu,\rho)$ for one-parameter copulas: (3) Ali-Mikhail-Haq.

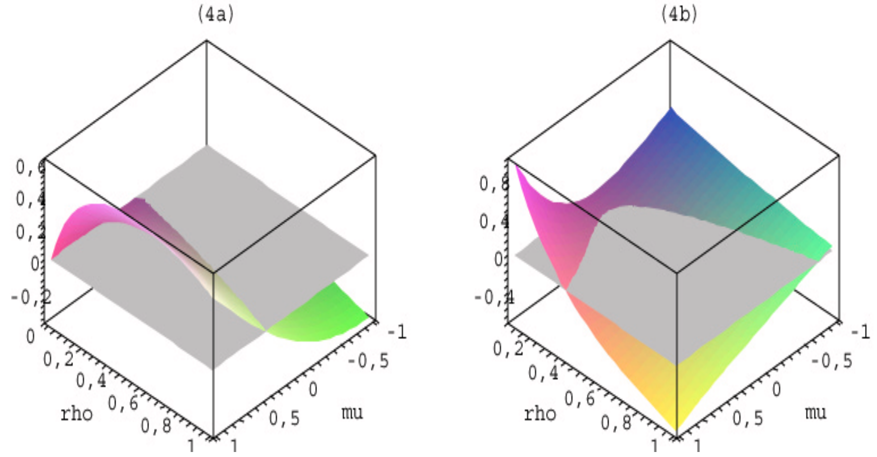

Figure 5: $\bar{\delta}^{\mu}(\mu,\rho)$ and $\bar{\delta}^{\rho}(\mu,\rho)$ for one-parameter copulas: (4) Clayton.

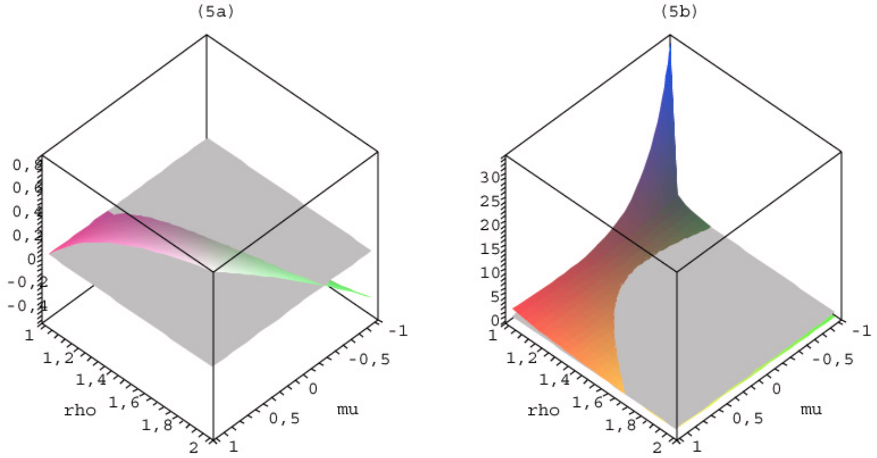

Figure 6: $\bar{\delta}^{\mu}(\mu,\rho)$ and $\bar{\delta}^{\rho}(\mu,\rho)$ for one-parameter copulas: (5) Gumbel.

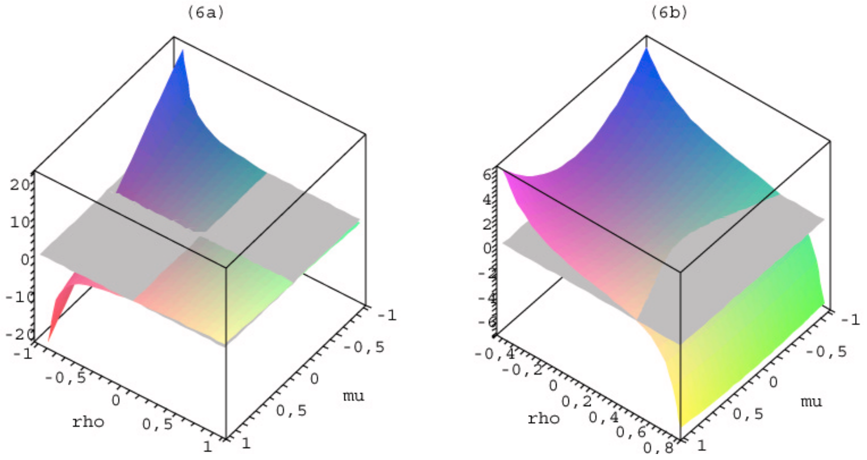

Figure 7: $\bar{\delta}^{\mu}(\mu, \rho)$ and $\bar{\delta}^{\rho}(\mu, \rho)$ for one-parameter copulas: (6) Normal.

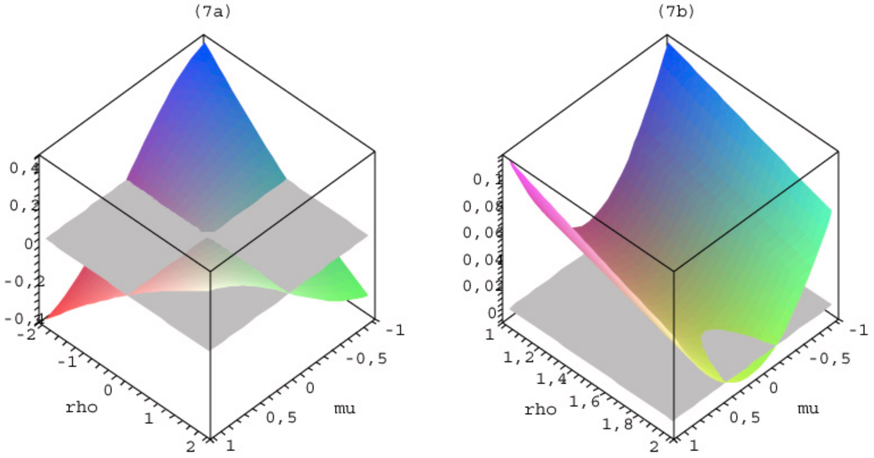

Figure 8: $\bar{\delta}^{\mu}(\mu, \rho)$ and $\bar{\delta}^{\rho}(\mu, \rho)$ for one-parameter copulas: (7) Frank.

Department of Economics
Michigan State University
East Lansing, MI 48824
*Email address:* prohorov@msu.edu