Statistica Sinica 20(2010): Supplement

Goodness-of-fit Tests for Archimedean Copula Models

Antai Wang

Georgetown University

Supplementary Material

In this note, we will prove Theorem 3, Corollaries 1 and 2, Theorems 4 and 5.

Proof of Theorem 3(1): we need to determine the form of

$$H_1(u,v) = \Pr(U \le u, V \le v | T_1 > c_1, T_2 > c_2) = \Pr(U \le u, V \le v, T_1 > c_1, T_2 > c_2)/S(c_1,c_2)$$

for $0 \le v \le S(c_1,c_2)$ and $0 \le u \le 1$. From this we know that we only need to work on the probability:

$$\Pr(U \le u, V \le v, T_1 > c_1, T_2 > c_2).$$

Assuming that the marginal distributions are continuous, using the monotonicity properties of the survivor function and the function q, we can see the probability equals to

$$\Pr(U \le u, V \le v, q\{S_1(T_1)\} > q\{S_1(c_1)\}, q\{S_2(T_2)\} > q\{S_2(c_2)\}).$$

From the definition of U and V, we know that $q\{S_1(T_1)\} = q(V)U$ and $q\{S_2(T_2)\} = q(V)(1-U)$. Hence the probability can be simplified as:

$$\Pr(U \le u, V \le v, q(V)U > q\{S_1(c_1)\}, q(V)(1-U) > q\{S_2(c_2)\})$$

$$= \Pr\left(V \le v, \frac{q\{S_1(c_1)\}}{q(V)} \le U \le \min\{u, 1 - \frac{q\{S_2(c_2)\}}{q(V)}\}\right)$$

It turns out that there are five situations we need to consider to derive this probability:

1. when $u < 1 - q\{S_2(c_2)\}/q(v)$, $u < q\{S_1(c_1)\}/q\{S(c_1, c_2)\}$ and $uq(v) < q\{S_1(c_1)\}$ (actually, because $uq(v) < q\{S_1(c_1)\}$ implies $u < q\{S_1(c_1)\}/q\{S(c_1, c_2)\}$, only the condition $uq(v) < q\{S_1(c_1)\}$ is needed here, we include both conditions here for clarity) : in this situation, we have

$$\frac{q\{S_1(c_1)\}}{q(V)} \le U \le u \le 1 - \frac{q\{S_2(c_2)\}}{q(V)}$$

and

$$0 \le V < p[q\{S_1(c_1)/u\}] < \min\{v, p[q\{S_2(c_2)/(1-u)\}]\}.$$

Therefore the probability

$$H_1(u, v) = \Pr\left(0 \le V \le p[q\{S_1(c_1)/u\}], \frac{q\{S_1(c_1)\}}{q(V)} \le U \le u\right) = \int_0^{p[q\{S_1(c_1)/u\}]} \int_{\frac{q\{S_1(c_1)\}}{q(V)}}^{u} k(v) du dv$$

$$= \int_0^{p[q\{S_1(c_1)/u\}]} \int_{\frac{q\{S_1(c_1)\}}{q(V)}}^{u} \frac{q''(v)q(v)}{q'(v)^2} du dv = up[\frac{q\{S_1(c_1)\}}{u}]$$

when $uq(v) < q\{S_1(c_1)\}$ for $0 \le v \le S(c_1, c_2)$.

2. when $u < 1 - q\{S_2(c_2)\}/q(v)$, $u < q\{S_1(c_1)\}/q\{S(c_1, c_2)\}$ but $uq(v) > q\{S_1(c_1)\}$: in this situation, we have

$$0 \le V \le v < p[q\{S_1(c_1)/u\}] < p[q\{S_2(c_2)/(1-u)\}],$$

and

$$q\{S_1(c_1)\}/q(V) \le U \le u.$$

Hence

$$H_1(u, v) = \Pr\left(0 \le V \le v, \frac{q\{S_1(c_1)\}}{q(V)} \le U \le u\right) = \int_0^{v} \int_{\frac{q\{S_1(c_1)\}}{q(V)}}^{u} k(v) du dv$$

$$= uv + \frac{q\{S_1(c_1)\} - uq(v)}{q'(v)}$$

for $q\{S_1(c_1)\}/q(v) < u < q\{S_1(c_1)\}/q\{S(c_1, c_2)\}$ and $0 \leq v \leq S(c_1, c_2)$.

3. when $q\{S_1(c_1)\}/q\{S(c_1, c_2)\} < u < 1 - q\{S_2(c_2)\}/q(v)$ but $q\{S_2(c_2)\} < (1-u)q(v)$: in this situation, we have

$$0 \leq V \leq v < p[q\{S_2(c_2)/(1-u)\}] < p[q\{S_1(c_1)/u\}],$$

and

$$q\{S_1(c_1)\}/q(V) \leq U \leq u.$$

Hence

$$H_1(u, v) = \Pr\left(0 \leq V \leq v, \frac{q\{S_1(c_1)\}}{q(V)} \leq U \leq u\right) = \int_0^v \int_{\frac{q\{S_1(c_1)\}}{q(V)}}^u k(v) du dv$$
$$= uv + \frac{q\{S_1(c_1)\} - uq(v)}{q'(v)}$$

for $q\{S_1(c_1)\}/q\{S(c_1, c_2)\} < u < 1 - q\{S_2(c_2)\}/q\{S(v)\}$ and $0 \leq v \leq S(c_1, c_2)$.

4. when $q\{S_1(c_1)\}/q\{S(c_1, c_2)\} < u < 1 - q\{S_2(c_2)\}/q(v)$ but $q\{S_2(c_2)\} > (1-u)q(v)$: in this situation, we have

$$0 \leq V \leq p[q\{S_2(c_2)/(1-u)\}] < v,$$

and

$$q\{S_1(c_1)\}/q(V) \leq U \leq u.$$

Hence

$$H_1(u, v) = \Pr\left(0 \leq V \leq p[q\{S_2(c_2)/(1-u)\}], \frac{q\{S_1(c_1)\}}{q(V)} \leq U \leq u\right)$$
$$= \int_0^{p[q\{S_2(c_2)/(1-u)\}]} \int_{\frac{q\{S_1(c_1)\}}{q(V)}}^u k(v) du dv$$

$$= up[q\{S_2(c_2)/(1-u)\}] - uq\{S_2(c_2)\}/(1-u)/q'[p\{q[S_2(c_2)/(1-u)]\}]$$

$$+q\{S_1(c_1)\}/q'[p\{q[S_2(c_2)/(1-u)]\}]$$

for $1 - q\{S_2(c_2)\}/q\{S(v)\} < u$ and $0 \le v \le S(c_1, c_2)$.

5. when $u > 1 - q\{S_2(c_2)\}/q(v)$: in this situation, we have

$$p[\frac{q\{S_2(c_2)\}}{(1-u)}] \le V \le v$$

and

$$q\{S_1(c_1)\}/q(V) \le U \le 1 - q\{S_2(c_2)\}/q(V).$$

Hence

$$\begin{aligned} H_1(u,v) &= \Pr\left(p[\frac{q\{S_2(c_2)\}}{(1-u)}] \le V \le v, q\{S_1(c_1)\}/q(V) \le U \le 1 - q\{S_2(c_2)\}/q(V)\right) \\ &= \int_{p[\frac{q\{S_2(c_2)\}}{(1-u)}]}^{v} \int_{q\{S_1(c_1)\}/q(v)}^{1-q\{S_2(c_2)\}/q(v)} k(v)dudv \\ &= v - q(v)/q'(v) - p[\frac{q\{S_2(c_2)\}}{(1-u)}] + q(p[\frac{q\{S_2(c_2)\}}{(1-u)}])/q'(p[\frac{q\{S_2(c_2)\}}{(1-u)}]) \\ &+q\{S(c_1,c_2)\}\left[1/q'(v) - 1/q'(p[\frac{q\{S_2(c_2)\}}{(1-u)}])\right] \end{aligned}$$

for $1 - q\{S_2(c_2)\}/q\{S(v)\} < u$ and $0 \le v \le S(c_1, c_2)$. Combining all five results and after some simple algebra, we can reach the conclusion for the joint distribution of (U, V) given $T_1 > c_1, T_2 > c_2$.

Proof of Theorem 3(2): we need to determine the form of

$$H_2(u,v) = \Pr(U \le u, V \le v | T_1 = t_1, T_2 > c_2) = \Pr(U \le u, V \le v, T_1 = t_1, T_2 > c_2)/\Pr(T_1 = t_1, T_2 > c_2)$$

for $0 \le v \le S(t_1, c_2)$ and $0 \le u \le 1$. It is easily seen that because $S(t_1,t_2) = p[q\{S_1(t_1)\} + q\{S_2(t_2)\}]$, we have

$$\Pr(T_1 = t_1, T_2 > c_2) = -p'[q\{S_1(t_1)\} + q\{S_2(c_2)\}]q'\{S_1(t_1)\}S_1'(t_1) = -p'[q\{S(t_1,c_2)\}]q'\{S_1(t_1)\}S_1'(t_1).$$

Therefore, we only need to work on

$$\Pr(U \leq u, V \leq v, T_1 = t_1, T_2 > c_2).$$

Using the same technique as presented in the proof of previous Theorem, we can express the above probability as

$$\Pr(q(v) \leq q(V), \frac{q\{S_1(T_1)\}}{u} \leq q(V), T_1 = t_1, T_2 > c_2).$$

Based on the fact that $q(V) = q\{S(T_1, T_2)\} = q\{S_1(T_1)\} + q\{S_2(T_2)\}$ and $T_1 = t_1$, the above probability can be further simplified as:

$$\Pr\left(q\{S_2(T_2)\} \geq \min\left[q(v) - q\{S_1(t_1)\}, q\{S_1(t_1)\}(\frac{1}{u} - 1)\right], T_1 = t_1, T_2 > c_2\right).$$

Again, we need to consider several situations:

1. $q(v) - q\{S_1(t_1)\} > q\{S_1(t_1)\}(1/u - 1)$, i.e. $u > q\{S_1(t_1)\}/q(v)$. In this case, after some simple algebra, we can show that the above probability equals to $-p'\{q(v)\}q'\{S_1(t_1)\}S_1'(t_1)$.

2. $q(v) - q\{S_1(t_1)\} \leq q\{S_1(t_1)\}(1/u - 1)$, i.e. $u \leq q\{S_1(t_1)\}/q(v)$. In this case, after some simple algebra, we can show that the above probability equals to $-p'[q\{S_1(t_1)\}/u]q'\{S_1(t_1)\}S_1'(t_1)$.

Combining these two results and after some algebra, we can reach the conclusion for the joint distribution of (U, V) given $T_1 = t_1, T_2 > c_2$.

Proof of Theorem 3(3): As in the proof of Theorem 3(2), we need to determine the form of

$$H_3(u, v) = \Pr(U \leq u, V \leq v | T_1 > c_1, T_2 = t_2) = \Pr(U \leq u, V \leq v, T_1 > c_1, T_2 = t_2)/\Pr(T_1 > c_1, T_2 = t_2)$$

for $1 - q\{S_2(t_2)\}/q(v) \leq u \leq 1$ and $0 \leq v \leq S(c_1, t_2)$. It is easily seen as before that

$$\Pr(T_1 > c_1, T_2 = t_2) = -p'[q\{S_1(c_1)\} + q\{S_2(t_2)\}]q'\{S_2(t_2)\}S_2'(t_2) = -p'[q\{S(c_1, t_2)\}]q'\{S_2(t_2)\}S_2'(t_2).$$

Therefore, we only need to work on

$$\Pr(U \leq u, V \leq v, T_1 > c_1, T_2 = t_2).$$

Using the same technique as presented in the proof of previous result, we can express the above probability as

$$\Pr\left(q\{S_2(t_2)\}/(1/u-1) \geq q\{S_1(T_1)\} \geq q(v) - q\{S_2(t_2)\}, T_2 = t_2\right).$$

$$= p'\{q(v)\} - p'[q\{S_2(t_2)\}/(1-u)]q'\{S_2(t_2)\}S_2'(t_2) = -p'[q\{S(c_1,t_2)\}]q'\{S_2(t_2)\}S_2'(t_2).$$

for $1 - q\{S_2(t_2)\}/q(v) \leq u \leq 1$. After some algebra, we can reach the conclusion for the joint distribution of (U, V) given $T_1 > c_1, T_2 = t_2$.

Proof of Corollary 1 *and* 2: Let $v = S(c_1, c_2)$, $v = S(t_1, c_2)$ and $v = S(c_1, t_2)$ in H_1, H_2 and H_3 respectively, we can prove Corollary 1. The same idea applies in the proof of Corollary 2. After plugging $u = 1$ into the expression of H_1, H_2 and H_3, we can reach the desired conclusions.

Proof of Theorem 4: From Corollary 1(a), we know the density function of $(V|T_1 > c_1, T_2 > c_2)$ is $f(v|T_1 > c_1, T_2 > c_2) = q''(v)[q(v) - q\{S(c_1,c_2)\}]/\{q'(v)^2 S(c_1,c_2)\}$. Hence we only need to determine

$$\Pr(U \leq u, V = v, T_1 > c_1, T_2 > c_2)$$

$$= \Pr(U \leq u, V = v, q(V)U > q\{S_1(c_1)\}, q(V)(1-U) > q\{S_2(c_2)\})$$

$$= \Pr\left(\frac{q\{S_1(c_1)\}}{q(v)} \leq U \leq \min\left\{u, 1 - \frac{q\{S_2(c_2)\}}{q(v)}\right\}, V = v\right)$$

Because $T_2 > c_2$, we have $q\{S_2(T_2)\} \geq q\{S_2(c_2)\}$. Therefore

$$U = q\{S_2(T_2)\}/q(v) = [q(v) - q\{S_2(T_2)\}]/q(v) \leq 1 - q\{S_2(c_2)\}/q(v).$$

On the other hand, $T_1 > c_1$, we have $U = q\{S_1(T_1)\}/q(v) \geq q\{S_1(c_1)\}/q(v)$. Hence

$$\Pr(U \leq u, V = v, T_1 > c_1, T_2 > c_2) = \Pr\left(\frac{q\{S_1(c_1)\}}{q(v)} \leq U \leq u, V = v\right)$$

$$= k(v)(u - \frac{q\{S_1(c_1)\}}{q(v)}) = \frac{q''(v)}{q'(v)^2}(uq(v) - q\{S_1(c_1)\}).$$

Hence the conditional distribution of $(U|V = v, T_1 > c_1, T_2 > c_2)$ is $[uq(v)-q\{S_1(c_1)\}]/[q(v)-q\{S(c_1, c_2)\}]$ for $q\{S_1(c_1)\}/q(v) \leq u \leq 1 - q\{S_2(c_2)\}/q(v)$. The conclusion follows.

Proof of Theorem 5: we only need to show that the covariance matrix

$$\text{cov}\{(\hat{U}_i, \hat{V}_i)^T, (\hat{U}_j, \hat{V}_j)^T\} = E\{(\hat{U}_i - E(\hat{U}_i), \hat{V}_i - E(\hat{V}_i))^T(\hat{U}_j - E(\hat{U}_j), \hat{V}_j - E(\hat{V}_j))\}\} \rightarrow \mathbf{0}_{2\times 2}$$

for $i \neq j$ when $n \rightarrow \infty$, where $\mathbf{0}_{2\times 2}$ represents a 2×2 zero matrix. Therefore, we need to prove this conclusion is correct for each entry of the matrix. There are several cases we need to consider:

1. when $(X_{1i}, X_{2i}) = (T_{1i}, T_{2i})$ and $(X_{1j}, X_{2j}) = (T_{1j}, T_{2j})$ i.e., both components of (T_{1i}, T_{2i}) and (T_{1j}, T_{2j}) are uncensored. Starting with the last entry, we must show that $E(\hat{V}_i - E(\hat{V}_i))(\hat{V}_j - E(\hat{V}_j)) \rightarrow 0$ when $n \rightarrow \infty$. In this situation, $\hat{V}_i$ and $\hat{V}_j$ are both Dabrowska's estimates of survivor functions at (X_{1i}, X_{2i}) and (X_{1j}, X_{2j}) respectively. By the almost sure consistency of Dabrowska's estimator (Dabrowska 1988), we can conclude that $\hat{V}_i = V_i + (\hat{V}_i - V_i) = V_i + Z_i$ and $\hat{V}_j = V_j + (\hat{V}_j - V_j) = V_j + Z_j$, where Z_i and Z_j converge to zero a.s. on $[0, \tau_1]\times[0, \tau_2]$ where $\Pr(X_1 > \tau_1, X_2 > \tau_2) > 0$. Notice the fact that $|\text{cov}(Z_i, V_j)| \leq \text{var}(Z_i)\text{var}(V_j) \leq E(Z_i^2)E(V_j^2) \leq E(Z_i^2) \rightarrow 0$ when $n \rightarrow \infty$ by the bounded convergence Theorem because Z_i^2 converges almost surely to 0 and $Z_i^2 \leq 4$. Similarly, one can show $\text{cov}(V_i, Z_j)$ and $\text{cov}(Z_i, Z_j)$ converge to zero when $n \rightarrow \infty$ as well. Therefore, we can conclude that $\text{cov}(\hat{V}_i, \hat{V}_j) \rightarrow 0$ because

$$\text{cov}(\hat{V}_i, \hat{V}_j) = \text{cov}(V_i + Z_i, V_j + Z_j) = \text{cov}(V_i, V_j) + \text{cov}(Z_i, V_j) + \text{cov}(V_i, Z_j) + \text{cov}(Z_i, Z_j).$$

$$= \text{cov}(Z_i, V_j) + \text{cov}(V_i, Z_j) + \text{cov}(Z_i, Z_j).$$

Now we need to show $\text{cov}(\hat{U}_i, \hat{U}_j) \to 0$ when $n \to \infty$, where

$$\hat{U}_i = \frac{q_{\hat{\theta}}(\hat{S}_1(t_{1i}))}{q_{\hat{\theta}}\{\hat{S}(t_{1i}, t_{2i})\}} = f\{\hat{\theta}, \hat{S}_1(t_{1i}), \hat{S}(t_{1i}, t_{2i})\},$$

where $f = f(\theta, w_1, w_2) = q_\theta(w_1)/q_\theta\{w_2\}$. Using the Taylor expansion, we have

$$\hat{U}_i = U_i + f_\theta\{\theta, S_1(t_{1i}), S(t_{1i}, t_{2i})\}(\hat{\theta} - \theta)$$

$$+f_1\{\theta, S_1(t_{1i}), S(t_{1i}, t_{2i})\}(\hat{S}_1(t_{1i}) - S_1(t_{1i})) + f_2\{\theta, S_1(t_{1i}), S(t_{1i}, t_{2i})\}(\hat{S}(t_{1i}, t_{2i}) - S(t_{1i}, t_{2i})),$$

where f_θ, f_1 and f_2 denotes the derivatives of function f with respective to θ, w_1 and w_2 respectively. Under appropriate regularity conditions on f such as the boundedness of the second derivatives of f, considering the fact that $\hat{\theta}$, $\hat{S}_1$ and $\hat{S}$ are consistent estimators of θ, S_1 and S respectively, one can conclude that $\hat{U}_i = U_i + o_p(1)$ and $\hat{U}_j = U_j + o_p(1)$. Following the similar arguments as before, we have $\text{cov}(\hat{U}_i, \hat{U}_j) \to 0$ when $n \to \infty$ because U_i and U_j are independent. Exactly the same arguments can be applied to show that $\text{cov}(\hat{U}_i, \hat{V}_j) \to 0$ and $\text{cov}(\hat{V}_i, \hat{U}_j) \to 0$. We have therefore proved the conclusion of Theorem 5 when both components of (T_{1i}, T_{2i}) are uncensored.

2. When both components of (T_{1i}, T_{2i}) and (T_{1j}, T_{2j}) are censored, i.e., $(X_{1i}, X_{2i}) = (C_{1i}, C_{2i})$ and also $(X_{1j}, X_{2j}) = (C_{1j}, C_{2j})$. In this situation, from the MI step we have described,

$$\hat{V}_i = F_1^{-1}(Q_{1i}, \hat{\theta}, \hat{S}(C_{1i}, C_{2i})), \hat{V}_j = F_1^{-1}(Q_{1j}, \hat{\theta}, \hat{S}(C_{1j}, C_{2j})),$$

$$\hat{U}_i = Q_{2i}\{1 - q(\hat{S}(C_{1i}, C_{2i}))/q(\hat{V}_i)\} + q(\hat{S}_1(C_{1i}))/q(\hat{V}_i) \text{ and}$$

$$\hat{U}_j = Q_{2j}\{1 - q(\hat{S}(C_{1j}, C_{2j}))/q(\hat{V}_j)\} + q(\hat{S}_1(C_{1j}))/q(\hat{V}_j),$$

where Q_{1i}, Q_{1j}, Q_{2i} and Q_{2j} are independently uniformly distributed random variables on $[0, 1]$. Applying the Taylor expansion again, one can show that $\hat{V}_i = F_1^{-1}(Q_{1i}, \theta, S(C_{1i}, C_{2i})) + o_p(1)$,

$\hat{V}_j = F_1^{-1}(Q_{1j}, \theta, S(C_{1j}, C_{2j})) + o_p(1)$,

$$\hat{U}_i = Q_{2i}\{1 - q(S(C_{1i}, C_{2i}))/q(V_i)\} + q(S_1(C_{1i}))/q(V_i) + o_p(1)$$

and

$$\hat{U}_j = Q_{2j}\{1 - q(S(C_{1j}, C_{2j}))/q(V_j)\} + q(S_1(C_{1j}))/q(V_j) + o_p(1)$$

where $V_i = F_1^{-1}(Q_{1i}, \theta, S(C_{1i}, C_{2i}))$ and $V_j = F_1^{-1}(Q_{1j}, \theta, S(C_{1j}, C_{2j}))$ under suitable regularity conditions on F_1 and q. Because (Q_{1i}, Q_{1j}), (Q_{2i}, Q_{2j}), (C_{1i}, C_{2i}) and also (C_{1j}, C_{2j}) are independent, one can show that $(\hat{U}_i, \hat{V}_i)$ and $(\hat{U}_j, \hat{V}_j)$ are asymptotically independent for $i \neq j$, $i, j \in \{1, 2, \ldots, n\}$ following the similar arguments as before.

3. When at least one component of (T_{1i}, T_{2i}) or (T_{1j}, T_{2j}) is censored and the other component in the same pair is uncensored, the proof is essentially the same as before (we only need to replace F_1 by F_2 or F_3 accordingly). This completes our proof.