SUPPLEMENT TO “COPULAS AND TEMPORAL DEPENDENCE”: APPENDIX

BY BRENDAN K. BEARE

(Econometrica, Vol. 78, No. 1, January 2010, 395–410)

This supplementary appendix contains proofs of the theorems given in the main paper.

**Proof of Theorem 3.1:** Since \( \{Z_t\} \) is a stationary Markov chain, it is known (see Theorems 7.3(b) and 3.29(II) in Bradley (2007)) that its \( \beta \)-mixing coefficients satisfy

\[
\beta_k = \frac{1}{2} \| F_{0,k}(x, y) - F(x)F(y) \|_{TV},
\]

where \( F_{0,k} \) is the joint distribution function of \( Z_0 \) and \( Z_k \), and \( \| \cdot \|_{TV} \) is total variation (in the Vitali sense).

From Sklar’s theorem, we thus have

\[
\beta_k = \frac{1}{2} \| C_k(F(x), F(y)) - F(x)F(y) \|_{TV} \leq \frac{1}{2} \| C_k(x, y) - xy \|_{TV}.\]

Equation (2.1) implies that \( C_k \) inherits the property of absolute continuity from \( C \). Letting \( c_k \) denote the density of \( C_k \), we now have that \( \beta_k \leq \frac{1}{2} \| c_k - 1 \|_1 \) and hence \( \beta_k \leq \frac{1}{2} \| c_k - 1 \|_2 \).

As a symmetric square-integrable joint density function with uniform marginals, \( c \) admits the mean square convergent expansion

\[
(A.1) \quad c(x, y) = 1 + \sum_{i=1}^{\infty} \lambda_i \phi_i(x)\phi_i(y),
\]

where the eigenvalues \( \{ \lambda_i \} \) form a nonincreasing square-summable sequence of nonnegative real numbers and the eigenfunctions \( \{ \phi_i \} \) form a complete orthonormal sequence in \( L^2[0,1] \). Expansions of this form were studied by Lancaster (1958), Rényi (1959), and Sarmanov (1958a, 1958b, 1961). Using (2.1), we deduce that the densities \( c_k \) satisfy

\[
c_k(x, y) = 1 + \sum_{i=1}^{\infty} \lambda_i^k \phi_i(x)\phi_i(y),
\]
which is simply a restatement of a result due to Sarmanov (1961) in terms of copula functions. We now have

\[ \|c_k - 1\|_2 = \left\| \sum_{i=1}^{\infty} \lambda_i^k \phi_i(x) \phi_i(y) \right\|_2, \]

and so with two applications of Parseval’s equality, we obtain

\[ \|c_k - 1\|_2 = \left( \sum_{i=1}^{\infty} \lambda_i^{2k} \right)^{1/2} \leq \lambda_1^{k-1} \left( \sum_{i=1}^{\infty} \lambda_i^2 \right)^{1/2} = \lambda_1^{k-1}\|c - 1\|_2. \]

As observed by Lancaster (1958), Rényi (1959), and Sarmanov (1958a, 1958b, 1961), \( \lambda_1 \) is equal to the maximal correlation of \( C \). Since this quantity is assumed to be less than 1, the proof is complete. \( Q.E.D. \)

**PROOF OF THEOREM 3.2:** Suppose first that \( \rho_C = 1 \). As observed by Lancaster (1958), Rényi (1959), and Sarmanov (1958a, 1958b), the supremum in (3.1) is achieved by a specific pair of functions \( f, g \) when \( c \) is square integrable. Consequently, for such \( f, g \), we have \( \iint f(x)g(y)c(x,y) \, dx \, dy = 1 \). Further, since \( \int f^2 = \int g^2 = 1 \) and the density \( c \) has uniform marginals, we have \( \iint f(x)^2c(x,y) \, dx \, dy = \iint g(y)^2c(x,y) \, dx \, dy = 1 \). It follows that

\[
\int_0^1 \int_0^1 f(x)g(y)c(x,y) \, dx \, dy = \left( \int_0^1 \int_0^1 f(x)^2c(x,y) \, dx \, dy \right)^{1/2} \left( \int_0^1 \int_0^1 g(y)^2c(x,y) \, dx \, dy \right)^{1/2},
\]

and so the Cauchy–Schwarz inequality holds with equality. This can be true only if the set \( D = \{(x, y) : f(x) \neq g(y)\} \) satisfies \( \iint_D c = 0 \). Let \( A = \{x : f(x) \geq 0\} \) and \( B = \{y : g(y) < 0\} \). The conditions \( \int f = \int g = 0 \) and \( \int f^2 = \int g^2 = 1 \) ensure that \( A \) and \( B \) have measure strictly between zero and one. Since \( (A \times B) \cup (A^c \times B^c) \subseteq D \), we have \( \iint_{(A \times B)\cup(A^c \times B^c)} c = 0 \), and hence \( c = 0 \) almost everywhere on \( (A \times B) \cup (A^c \times B^c) \).

Suppose next that \( c = 0 \) almost everywhere on \( (A \times B) \cup (A^c \times B^c) \), where \( A, B \) have measure strictly between zero and one. Let \( f(x) = 1 \ (x \in A) \) and \( g(y) = 1 \ (y \notin B) \). It is easily verified that \( f(x) = g(y) \) on a subset of \([0, 1]^2\) over which \( c \) integrates to 1. Since neither \( f \) nor \( g \) is constant almost everywhere, it follows that \( \rho_C = 1 \). \( Q.E.D. \)

**PROOF OF THEOREM 3.3:** We will show that \( C \) cannot exhibit lower tail dependence when \( c \) is square integrable and \( \mu_L \) exists. The corresponding result
for upper tail dependence can be shown in essentially the same way. For any \( n \in \mathbb{N} \) and any \( x \in (0, 1] \), we may write

\[
\frac{C(x, x)}{x} = x + \sum_{i=1}^{n} \lambda_i x^{-1} \left( \int_{0}^{x} \phi_i(z) \, dz \right)^2 + \xi_n(x),
\]

where \( \xi_n \) is defined by this equation. The Cauchy–Schwarz inequality implies that

\[
x^{-1} \left( \int_{0}^{x} \phi_i(z) \, dz \right)^2 \leq x^{-1} \left( \int_{0}^{x} \phi_i(z)^2 \, dz \right) = \left( \int_{0}^{x} \phi_i(z)^2 \, dz \right).
\]

Square integrability of \( \phi_i \) therefore implies that \( \lim_{x \to 0^+} \frac{x^{-1/2} \int_{0}^{x} \phi_i(z) \, dz}{x} = 0 \). We thus obtain

\[
\lim_{x \to 0^+} \frac{C(x, x)}{x} = \lim_{x \to 0^+} \xi_n(x) \leq \|\xi_n\|_{\infty}
\]

for each \( n \in \mathbb{N} \). It thus suffices to show that \( \|\xi_n\|_{\infty} \to 0 \) as \( n \to \infty \). Using Cauchy–Schwarz, we have

\[
\|\xi_n\|_{\infty} = \left\| x^{-1} \int_{0}^{x} \int_{0}^{x} \left( c(u, v) - 1 - \sum_{i=1}^{n} \lambda_i \phi_i(u) \phi_i(v) \right) \, du \, dv \right\|_{\infty}
\]

\[
\leq \left\| \left( \int_{0}^{x} \int_{0}^{x} \left( c(u, v) - 1 - \sum_{i=1}^{n} \lambda_i \phi_i(u) \phi_i(v) \right)^2 \, du \, dv \right)^{1/2} \right\|_{\infty}
\]

\[
= \left( \int_{0}^{1} \int_{0}^{1} \left( c(u, v) - 1 - \sum_{i=1}^{n} \lambda_i \phi_i(u) \phi_i(v) \right)^2 \, du \, dv \right)^{1/2}.
\]

Convergence of this last term to zero as \( n \to \infty \) is the content of our series expansion (A.1).

Q.E.D.

**Proof of Theorem 4.1:** Since \( \{Z_t\} \) is a Markov chain, Theorem 7.5(I)(a) of Bradley (2007) implies that \( \rho_k \) decays geometrically fast if \( \rho_1 < 1 \). We thus need only show that \( \rho_1 \leq \rho_C \). Given \( \sigma \)-fields \( \mathcal{A}, \mathcal{B} \subseteq \mathcal{F} \), let \( \rho(\mathcal{A}, \mathcal{B}) = \sup_{f, g} |\text{Corr}(f, g)| \), where the supremum is taken over all random variables \( f \) and \( g \) measurable with respect to \( \mathcal{A} \) and \( \mathcal{B} \), respectively, with positive and finite variance. Since \( \{Z_t\} \) is a stationary Markov chain, Theorem 7.3(c) in Bradley (2007) implies that \( \rho_1 = \rho(\sigma(Z_0), \sigma(Z_1)) \). Let \( U, V \) be random variables with joint distribution function \( C \), and let \( F^{-1} \) denote the quasi-inverse
distribution function given by $F^{-1}(z) = \inf_x \{ F(x) \geq z \}$. Then $Z_0^* = F^{-1}(U)$ and $Z_1^* = F^{-1}(V)$ have the same joint distribution as $Z_0$ and $Z_1$, and so Proposition 3.6(I)(c) of Bradley (2007) implies that $\rho_1 = \rho(\sigma(Z_0^*), \sigma(Z_1^*))$. Since $\sigma(Z_0^*) \subseteq \sigma(U)$ and $\sigma(Z_1^*) \subseteq \sigma(V)$, it follows that $\rho_1 \leq \rho(\sigma(U), \sigma(V))$. We conclude by noting that $\rho(\sigma(U), \sigma(V)) = \rho_C$.

**PROOF OF THEOREM 4.2:** Let $\varepsilon > 0$ be such that $c(x, y) \geq \varepsilon$ almost everywhere on $[0, 1]^2$. Consider $f, g \in L_2[0, 1]$ with $\int f = \int g = 0$ and $\int f^2 = \int g^2 = 1$. Begin by writing

$$
\int \int f(x)g(y)C(dx, dy) = \frac{1}{2} \int \int (f(x)^2 + g(y)^2)C(dx, dy) - \frac{1}{2} \int \int (f(x) - g(y))^2C(dx, dy).
$$

Since $(f(x) - g(y))^2 \geq 0$ and $c(x, y) \geq \varepsilon$ almost everywhere, we have

$$
\int \int (f(x) - g(y))^2C(dx, dy) \geq \int \int (f(x) - g(y))^2c(x, y)dx dy \geq \varepsilon \int \int (f(x) - g(y))^2 dx dy = 2\varepsilon.
$$

Since it is also the case that $\int \int (f(x)^2 + g(y)^2)C(dx, dy) = 2$, we obtain $\int \int f(x)g(y)C(dx, dy) \leq 1 - \varepsilon$, implying that the maximal correlation of $C$ cannot exceed $1 - \varepsilon$.

**PROOF OF THEOREM 4.3:** Let $S_n$ denote the class of real-valued functions $f$ on $[0, 1]$ that can be written in the form

$$
f(x) = \sum_{i=1}^{n} f_i 1_{((i-1)/n, i/n]}(x),
$$

where $f_1, \ldots, f_n$ are real numbers. If $f, g \in S_n$, then

$$
(A.2) \quad \int_0^1 \int_0^1 f(x)g(y)C(dx, dy) = - \left( \int_0^1 f(x) dx \right) \left( \int_0^1 g(y) dy \right) = \sum_{i=1}^{n} \sum_{j=1}^{n} f_i g_j K_n(i, j).
$$

Consequently, $n_0$ is the maximum of the left-hand side of $(A.2)$ over $f, g \in S_n$ such that $\int f^2 = \int g^2 = 1$. It follows that $n_0$ is the maximum of
\[ \int \int f(x)g(y)C(dx, dy) \] over \( f, g \in S \), such that \( \int f = \int g = 0 \) and \( \int f^2 = \int g^2 = 1 \). Our desired result now follows from the definition of \( \rho_C \) and the fact that \( \bigcup_{n \in \mathbb{N}} S_n \) is a dense subset of \( L_2[0,1] \). Q.E.D.

**REFERENCES**


Dept. of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508, U.S.A.; bbeare@ucsd.edu.

Manuscript received September, 2008; final revision received June, 2009.