UC San Diego
Recent Work

Title
Copulas and Temporal Dependence

Permalink
https://escholarship.org/uc/item/87p829d4

Author
Beare, Brendan K.

Publication Date
2009-03-23
Copulas and Temporal Dependence

Brendan K. Beare∗†

University of California at San Diego

March 23, 2009

Abstract: An emerging literature in time series econometrics concerns the modeling of potentially nonlinear temporal dependence in stationary Markov chains using copula functions. We obtain sufficient conditions for a geometric rate of mixing in models of this kind. Geometric \( \beta \)-mixing is established under a rather strong sufficient condition that rules out asymmetry and tail dependence in the copula function. Geometric \( \rho \)-mixing is obtained under a weaker condition that permits both asymmetry and tail dependence. We verify one or both of these conditions for a range of parametric copula functions that are popular in applied work.

Keywords and phrases: copula; Markov chain; maximal correlation; mean square contingency; mixing; canonical correlation; tail dependence.

JEL classifications: primary C22; secondary C63.

∗ Address: Department of Economics, UCSD, 9500 Gilman Drive, La Jolla CA 92093-0508, USA. Email: bbeare@ucsd.edu.
† Parts of this paper derive from my doctoral research at Yale University. I thank my advisor Peter Phillips and committee members Donald Andrews and Yuichi Kitamura for their support and advice. I also thank Xiaohong Chen, Rustam Ibragimov, Bent Nielsen, Andres Santos, and three anonymous referees for helpful comments. Financial support from the Cowles Foundation under a Carl Arvid Anderson Prize Fellowship is gratefully acknowledged.
1 Introduction

In recent years, a number of authors have considered the possibility of modeling a univariate stationary discrete-time Markov chain by specifying (a) the invariant distribution $F$, and (b) a bivariate copula $C$ characterizing the dependence between consecutive realizations. All finite dimensional distributions of the chain are uniquely determined by $C$ and $F$. Some interesting patterns of temporal dependence can be generated using certain copula functions. A copula exhibiting tail dependence may generate a Markov chain which appears to become substantially more serially dependent as it draws towards the extremes of the state space. For instance, one could think of the phenomenon of unemployment hysteresis, in which unemployment appears to become “stuck” at relatively high levels, as corresponding to a copula exhibiting upper tail dependence. Asymmetric copulas can be used to model nonreversible temporal behaviour, such as the tendency for many economic and financial time series to exhibit relatively long periods of fairly steady growth, interspersed with shorter periods of sharp decline.

Various procedures for estimating models of this kind have been proposed, ranging from fully parametric (Joe, 1997, ch. 8) to semiparametric (Chen and Fan, 2006; Chen et. al., 2009) and quasi-nonparametric (Gagliardini and Gouriéroux, 2007). Fentaw and Naik-Nimbalkar (2008) develop methods applicable to Markov chains generated by copula functions with time-varying parameters. Applications of these procedures have considered air quality measurements (Joe, 1997, ch. 8), stochastic volatility in stock returns (Ibragimov and Lentzas, 2008), and coffee prices in Ethiopia (Fentaw and Naik-Nimbalkar, 2008). Gagliardini and Gouriéroux (2008) discuss the application of their methods to intertrade durations in financial markets. In a panel data context, related methods have been applied by Bonhomme and Robin (2006) and Dearden et. al. (2008) to model earnings dynamics. Patton (2008) provides a brief review of a number of these papers, among others.

An important issue to have arisen in much of this literature is the following: how does the form of the copula $C$ relate to the strength of temporal dependence in the Markov chain? More specifically, what conditions on $C$ will ensure that weak dependence conditions sufficient for the application of invariance principles are satisfied? As suggested by Chen and Fan (2006), verification of the stability conditions of Meyn and Tweedie (1993) guaranteeing geometric ergodicity provides one possible approach to demonstrating weak dependence properties for specific copula functions. Gagliardini and Gouriéroux (2008) use this approach to identify a condition under which proportional hazard copulas generate a geometric rate of $\beta$-mixing. Ibragimov and Lentzas (2008) claim to provide numerical evidence that is suggestive of long memory in chains generated by Clayton copulas. Ibragimov (2008) proposes a class of Fourier copulas which generate $m$-dependent Markov chains under suitable conditions. $m$-dependence implies the satisfaction of mixing...
In this paper we identify conditions on $C$ that suffice for geometrically fast mixing rates. Geometric $\beta$-mixing, equivalent to geometric ergodicity for stationary Markov chains, is established under a rather strong condition that excludes copulas exhibiting tail dependence or asymmetry. Geometric $\rho$-mixing, which implies geometric $\alpha$-mixing, is obtained under a much weaker condition. We verify this condition for various parametric copula functions popular in applied work. $\rho$-, $\beta$- and $\alpha$-mixing conditions may be used as the basis for a range of inequalities and limit theorems that are useful in demonstrating the asymptotic validity of statistical methods; see e.g. the monograph by Doukhan (1994) or the recent three volume series by Bradley (2007).

Contemporaneous work by Chen et. al. (2009) relates closely to the results reported here. Specifically, Chen et. al. establish that the Clayton, Gumbel and $t$-copulas generate Markov chains with a geometric rate of $\beta$-mixing. Since these three copula functions exhibit positive tail dependence, they are not covered by our results concerning geometric $\beta$-mixing. Nevertheless, we do provide results pertaining to the $\rho$-mixing properties of Markov chains generated by these copula functions. Since the $\beta$-mixing property does not imply $\rho$-mixing, or vice-versa, our results are complementary to those of Chen et. al.

The structure of the paper is as follows. In Section 2 we introduce our notation and some basic definitions and results. In Section 3 we consider $\beta$-mixing conditions, while in Section 4 we consider $\rho$-mixing conditions. Section 5 concludes. Mathematical proofs are collected in the Appendix.

2 Basic setup

We begin with the following rather basic definition of a bivariate copula function.

Definition 2.1. A bivariate copula function is a bivariate probability distribution function on $[0, 1]^2$ for which the two univariate marginal distribution functions are uniform on $[0, 1]$.

Our concern in this paper is with bivariate dependence; when we refer to a copula function or copula, we mean a bivariate copula function. Suppose that $X$ and $Y$ are real valued random variables with joint distribution function $F_{X,Y}$ and marginal distribution functions $F_X$ and $F_Y$. We will say that $X$ and $Y$ admit the copula $C$ if $C(F_X(x), F_Y(y)) = F_{X,Y}(x, y)$ for all $x, y \in \mathbb{R}$. A fundamental result concerning copulas is Sklar’s theorem (1959), proved in Schweizer and Sklar (1974). A useful, more recent treatment of Sklar’s theorem may be found in Nelsen (1999). Sklar’s theorem ensures that for any random variables $X$ and $Y$, there exists a copula $C$ such that $X$ and $Y$ admit $C$. Moreover, $C$ is uniquely defined on the product of the ranges of the marginal distribution functions of $X$ and $Y$. Hence, $C$ is unique if $X$ and $Y$ are continuous random
variables. If $X$ or $Y$ are not continuous, $C$ may nevertheless be uniquely defined by bilinear interpolation between uniquely defined values; see e.g. the proof of Lemma 2.3.5 in Nelsen (1999). Following Darsow et. al. (1992), we will refer to this unique copula as the copula of $X$ and $Y$.

Let $\{Z_t : t \in \mathbb{Z}\}$ be a stationary sequence of real valued random variables defined on a probability space $(\Omega, \mathcal{F}, P)$, and for $s, t \in \mathbb{Z} \cup \{-\infty, \infty\}$, let $\mathcal{F}_t^s \subseteq \mathcal{F}$ denote the $\sigma$-field generated by the random variables $\{Z_r : s \leq r \leq t\}$. We assume that $\{Z_t\}$ is a Markov chain. That is, for any finite collection of integers $\{t_1, \ldots, t_n\}$ satisfying $t_1 < t_2 < \cdots < t_n$, and any $z \in \mathbb{R}$, we have

$$E\left(1_{\{Z_{t_n} \leq z\}}|Z_{t_1}, \ldots, Z_{t_{n-1}}\right) = E\left(1_{\{Z_{t_n} \leq z\}}|Z_{t_{n-1}}\right)$$

almost surely. Let $F$ denote the marginal distribution function of $Z_0$, and for $k \in \mathbb{N}$ let $C_k$ be the copula of $Z_0$ and $Z_k$. We will often write $C$ in place of $C_1$. Theorem 3.2 of Darsow et. al. (1992) asserts that the copulas $C_k$ satisfy

$$C_{k+1}(x, y) = \int_0^1 \frac{\partial C_k(x, z)}{\partial z} \frac{\partial C(z, y)}{\partial z} dz$$

for all $k \in \mathbb{N}$ and all $x, y \in [0, 1]^2$. Existence of the partial derivatives in (2.1) for almost all $z \in [0, 1]$ is a well-known property of copula functions; see e.g. Theorem 2.2.7 in Nelsen (1999).

(2.1) constitutes a version of the Chapman-Kolmogorov equations for Markov chains, expressed in terms of copula functions. It implies that all bivariate copulas $C_k$ may be expressed in terms of the copula $C$. In the next two sections we will identify conditions on $C$ such that mixing coefficients determined by the copulas $C_k$ decay to zero at a geometric rate as $k$ increases.

### 3 Sufficient conditions for geometric $\beta$-mixing

Sequences of $\beta$-mixing coefficients provide one way to characterize the serial persistence of time series. We shall employ the following definition.

**Definition 3.1.** The $\beta$-mixing coefficients $\{\beta_k : k \in \mathbb{N}\}$ corresponding to the sequence of random variables $\{Z_t\}$ are given by

$$\beta_k = \frac{1}{2} \sup_{m \in \mathbb{Z}} \sup_{A_1, \ldots, A_I, B_1, \ldots, B_J} \sum_{i=1}^I \sum_{j=1}^J |P(A_i \cap B_j) - P(A_i)P(B_j)|,$$

where the second supremum is taken over all finite partitions $\{A_1, \ldots, A_I\}$ and $\{B_1, \ldots, B_J\}$ of $\Omega$ such that $A_i \in \mathcal{F}_{-\infty}^m$ for each $i$ and $B_j \in \mathcal{F}_{m+k}^\infty$ for each $j$.

An equivalent form of Definition 3.1 was originally stated by Volkonskii and Rozanov (1959), though they attribute it to Kolmogorov. The condition $\beta_k \to 0$ as $k \to \infty$ is variously referred to as strong regularity.
complete regularity, absolute regularity, or simply $\beta$-mixing. Several equivalent formulations of Definition 3.1 have been used in the literature, often involving conditional probabilities or total variation norms. Numerous results concerning these equivalencies may be found in Chapter 3 of Bradley (2007).

Before stating our first theorem, we require an additional definition.

**Definition 3.2.** The maximal correlation $\rho_C$ of the copula $C$ is given by

$$\sup_{f,g} \left| \int_0^1 \int_0^1 f(x)g(y)C(dx,dy) \right|,$$

where the supremum is taken over all $f,g \in L^2[0,1]$ such that $\int f = \int g = 0$ and $\int f^2 = \int g^2 = 1$. The integral in (3.2) is defined in the usual Lebesgue-Stieltjes sense. Maximal correlation is an old concept; Rényi (1959) provides an early discussion of its properties as a measure of dependence between random variables.

The first result of this section is as follows.

**Theorem 3.1.** Suppose that $C$ is symmetric and absolutely continuous with square integrable density $c$, and that $\rho_C < 1$. Then there exists $A < \infty$ and $\gamma > 0$ such that $\beta_k \leq Ae^{-\gamma k}$ for all $k$.

**Remark 3.1.** Here and elsewhere, we say that a copula $C$ is symmetric if $C(x,y) = C(y,x)$ for all $x,y \in [0,1]$.

**Remark 3.2.** Theorem 3.1 is proved in two main steps. First, one establishes the bound $\beta_k \leq \frac{1}{2}\|c_k - 1\|_2$, where $c_k$ is the density of $C_k$. Next, a spectral decomposition of $c_k$ is used to show $\|c_k - 1\|_2 \leq \rho_C^{-1}\|c - 1\|_2$. The validity of this decomposition depends crucially on $c$ being symmetric and square integrable.

**Remark 3.3.** The quantity $\|c - 1\|_2^2$ is referred to as the mean square contingency of the joint distribution of $Z_0$ and $Z_1$. Mean square contingency was defined formally by Lancaster (1958) without reference to copula functions, but its origins can be traced back to work by Pearson in the early twentieth century. Lancaster (1958), Rényi (1959) and Sarmanov (1958ab, 1961) used spectral or singular value decompositions to study the structure of bivariate distributions exhibiting finite mean square contingency. The work of Sarmanov (1961) is of special relevance here, as he was concerned in particular with the bivariate distributions associated with stationary Markov chains.

**Remark 3.4.** The proof of Theorem 3.1 establishes geometric decay of $\|c_k - 1\|_2$, which is stronger than the statement of the theorem. Geometric decay of $\|c_k - 1\|_2$ implies geometric decay of a range of dependence measures between $Z_0$ and $Z_k$ other than $\beta$-mixing coefficients. For instance, Proposition 1 of Ibragimov and Lentzas (2008) establishes $\|c_k - 1\|_2$ or $\|c_k - 1\|_2^2$ as an upper bound on the relative entropy, Hellinger distance, linear correlation, and Schweizer-Wolff (1981) distances between $Z_0$ and $Z_k$. These quantities must therefore decay geometrically as $k \to \infty$, under the assumptions of Theorem 3.1.
Remark 3.5. For stationary Markov chains, a geometric rate of $\beta$-mixing holds if and only if the chain satisfies geometric ergodicity. See Definition 21.18 and Theorem 21.19 in Bradley (2007) for a definition of geometric ergodicity and statement of this result.

Remark 3.6. The following result may prove useful in verifying the condition $\rho_C < 1$ that appears in the statement of Theorem 3.1.

**Theorem 3.2.** Suppose that $C$ is absolutely continuous with square integrable density $c$. Then $\rho_C = 1$ if and only if there exist measurable sets $A, B \subset [0, 1]$ with measure strictly between zero and one such that

$$c = 0 \text{ a.e. on } (A \times B) \cup (A^c \times B^c).$$

In particular, if $c > 0$ a.e. on $[0,1]^2$, then $\rho_C < 1$.

Theorem 3.2 indicates that, for absolutely continuous copulas with square integrable densities, the condition $\rho_C < 1$ is rather easy to satisfy. Copula functions used in applied work typically have a density that is positive almost everywhere.

Remark 3.7. The assumption that $C$ is symmetric implies that the Markov chain $\{Z_t\}$ is time reversible.

There is substantial evidence that many economic and financial series exhibit irreversible behavior; see e.g. McCausland (2007). The proportional hazard copulas considered by Gagliardini and Gouriéroux (2008) are asymmetric in general. Nevertheless, many parametric families of copulas used in applied work do satisfy the symmetry condition. In particular, all Archimedean copulas are symmetric.

Remark 3.8. Commonly used parametric copula functions satisfying the conditions of Theorem 3.1 include the Farlie-Gumbel-Morgenstern, Frank, and Gaussian copulas. See Nelsen (1999) for definitions and origins. For the latter two copulas, this assumes that the copula parameter is in the interior of the parameter space, so that the copula does not degenerate to the Fréchet-Hööfdding upper or lower bound. Square integrability of the Farlie-Gumbel-Morgenstern and Frank copula densities follows from the fact that they are bounded. For the Gaussian copula density, square integrability follows from Mehler’s identity, which provides a mean square convergent expansion of the bivariate Gaussian density in terms of Hermite polynomials. That is,

$$f_\rho(x,y) = f(x)f(y)\sum_{k=0}^{\infty} \rho^k H_k(x)H_k(y),$$

where $f_\rho$ is the standard bivariate Gaussian density with correlation coefficient $\rho \in (0,1)$, $f$ is the standard Gaussian density, and $\{H_k\}$ is a sequence of functions that is orthonormal with respect to the density $f$, with each $H_k$ being a polynomial of order $k$. (3.3) implies that the Gaussian copula density $c_\rho$ satisfies

$$c_\rho(x,y) = \sum_{k=0}^{\infty} \rho^k \phi_k(x)\phi_k(y),$$

6
with \( \{ \phi_k \} \) being an orthonormal sequence of functions in \( L_2[0,1] \). It follows that \( \|c_\rho\|_2 = 1/\sqrt{1 - \rho^2} < \infty \).

Other bivariate distributions admitting expansions analogous to (3.3) include the bivariate gamma, Poisson, binomial, and hypergeometric distributions, and the compound correlated bivariate Poisson distribution. See Hamdan and Al-Bayyati (1971) for references and further discussion. The copula functions corresponding to these distributions satisfy the conditions of Theorem 3.1.

Remark 3.9. Not all copulas of interest satisfy the conditions of Theorem 3.1. The Marshall-Olkin copula (see e.g. Nelsen, 1999, p. 46) is not absolutely continuous, nor is it symmetric in general. Further, any copula exhibiting upper or lower tail dependence will not admit a square integrable density. Following e.g. McNeil et. al. (2005, p. 209), we define tail dependence as follows.

**Definition 3.3.** The coefficient of lower tail dependence \( \mu_L \) corresponding to a copula \( C \) is given by

\[
\mu_L = \lim_{x \to 0^+} \frac{C(x,x)}{x},
\]

if the limit exists. The coefficient of upper tail dependence is given by

\[
\mu_U = \lim_{x \to 1^-} \frac{1 - 2x + C(x,x)}{1 - x},
\]

if the limit exists.

If \( \mu_L \) exists and is positive we say that \( C \) exhibits lower tail dependence, while if \( \mu_U \) exists and is positive we say that \( C \) exhibits upper tail dependence.

The following result states that tail dependence is ruled out by our requirement in Theorem 3.1 that \( C \) have square integrable density \( c \).

**Theorem 3.3.** Let \( C \) be an absolutely continuous copula with square integrable density \( c \). Then \( C \) exhibits neither upper nor lower tail dependence.

Theorem 3.3 implies that several parametric classes of copula used frequently in applied work have a density that is not square integrable. In particular, the Gumbel, Clayton, and \( t \)-copulas all exhibit upper or lower tail dependence; see e.g. Examples 5.3.1 and 5.3.3 in McNeil et. al. (2005). The connection between tail dependence and the square integrability of \( c \) does not seem to have been noted in previous literature.

Remark 3.10. Although Theorem 3.3 implies that the conditions of Theorem 3.1 are not satisfied by copulas exhibiting upper or lower tail dependence, we do not assert that such copulas necessarily generate stationary Markov chains for which the decay rate of \( \beta \)-mixing coefficients is sub-geometric. Indeed, Chen et. al.
(2009) establish that the Gumbel, Clayton and t-copulas all generate a geometric rate of β-mixing. This result indicates that the conditions of Theorem 3.1 are sufficient but not necessary for the stated conclusion.

Remark 3.11. Theorem 3.3 bears an additional implication that is not related to time series applications in particular. Gagliardini and Gourieroux (2007) propose a quasi-nonparametric approach to copula estimation that involves minimizing a weighted chi-square distance between an unconstrained kernel estimate of \( c \), and a constrained estimate. The constrained estimate restricts \( c \) to be in a class of copula densities that is specified up to a one-dimensional functional parameter; that is, a vector-valued function of one variable. Archimedean copulas are an example of a class of copulas that may be parameterized in this way, with the additive generator of each copula being its functional parameter; see e.g. p. 92 in Nelsen (1999). Another example is the class of proportional hazard copulas, as shown by Gagliardini and Gourieroux (2008).

Suppose we have a sample of \( n \) pairs of observations, assumed for simplicity to have marginal distributions that are uniform on \([0, 1]\). Given a set of functional parameters \( \Theta \), the estimated functional parameter \( \hat{\theta}_n \) minimizes the criterion function

\[
M_n(\theta) = \int_0^1 \int_0^1 \left[ \frac{c(x, y; \theta) - \hat{c}_n(x, y)}{\hat{c}_n(x, y)} \right]^2 w_n(x, y) \, dx \, dy,
\]

where \( \hat{c}_n(x, y) \) is the unconstrained estimate of \( c(x, y) \), \( c(x, y; \theta) \) is the copula corresponding to the functional parameter \( \theta \), and \( w_n(x, y) \) is a smooth weighting function that converges pointwise to one as \( n \to \infty \). Gagliardini and Gourieroux (2007) do not discuss the choice of the weighting function \( w_n \). Theorem 3.3 implies that, if \( c(x, y; \theta) \) exhibits tail dependence for values of \( \theta \) in a neighborhood of the true value, then the choice of \( w_n \) may be rather important. Specifically, for each \( n \), we will need to have \( w_n(x, x) \to 0 \) at some rate as \( x \to 0 \) and as \( x \to 1 \), in order for \( M_n(\theta) \) to be finite in a neighborhood of the true value. This issue merits further investigation, but goes beyond the scope of this paper.

4 Sufficient conditions for geometric \( \rho \)-mixing

\( \beta \)-mixing coefficients provide one way to characterize the serial persistence of \( \{Z_t\} \). One may also characterize this persistence using \( \rho \)-mixing coefficients. We shall employ the following definition.

Definition 4.1. The \( \rho \)-mixing coefficients \( \{\rho_k : k \in \mathbb{N}\} \) corresponding to the sequence of random variables \( \{Z_t\} \) are given by

\[
\rho_k = \sup_{m \in \mathbb{Z}} \sup_{f, g} |\text{Corr}(f, g)|,
\]

where the second supremum is taken over all square integrable random variables \( f \) and \( g \) measurable with
respect to $\mathcal{F}_{m}$ and $\mathcal{F}_{m+k}$ respectively, with positive and finite variance, and where $\text{Corr}(f, g)$ denotes the correlation between $f$ and $g$.

Definition 4.1 was originally stated by Kolmogorov and Rozanov (1960). Rosenblatt (1971, ch. 7) provides an important discussion of $\rho$-mixing conditions in the context of stationary Markov chains. $\rho$-mixing conditions appear to have been largely ignored in the econometrics literature. A recent exception is Chen et. al. (2008), who study the $\rho$- and $\beta$-mixing properties of nonlinear diffusion processes.

A definition of mixing that has been much more heavily employed in econometrics is that of $\alpha$-mixing.

**Definition 4.2.** The $\alpha$-mixing coefficients $\{\alpha_k : k \in \mathbb{N}\}$ corresponding to the sequence of random variables $\{Z_t\}$ are given by

$$
\alpha_k = \sup_{m \in \mathbb{Z}} \sup_{A \in \mathcal{F}_{m}, B \in \mathcal{F}_{m+k}} |P(A \cap B) - P(A)P(B)|.
$$

Definition 4.2 is commonly attributed to Rosenblatt (1956). However, Beare (2007) observes that this definition is in fact different to that proposed by Rosenblatt, and appears to have first been stated by Volkonskii and Rozanov (1959), who refer to their condition as being ‘analogous’ to that of Rosenblatt. The inequalities $2\alpha_k \leq \beta_k$ and $4\alpha_k \leq \rho_k$ (see e.g. Proposition 3.11 in Bradley, 2007) ensure that $\alpha$-mixing is a weaker dependence condition than both $\beta$- and $\rho$-mixing. Neither of the $\beta$- and $\rho$-mixing conditions is implied by the other: there exist processes for which $\beta_k \to 0$ but $\rho_k \not\to 0$ as $k \to \infty$, and vice-versa. Central limit theorems are available for $\rho$-mixing processes that involve weaker assumptions on moments and mixing rates than do those available for $\beta$-mixing processes. See, for instance, Remark 10.11(8) and Theorems 10.7 and 11.4 in Bradley (2007). On the other hand, $\beta$-mixing conditions have been used to establish a variety of other useful results, such as, for instance, the central limit theorem for $U$-statistics in Arcones (1995).

The following result identifies a simple condition on $C$ such that $\rho_k$ decays at a geometric rate.

**Theorem 4.1.** Suppose $\rho_C < 1$. Then there exists $A < \infty$ and $\gamma > 0$ such that $\rho_k \leq Ae^{-\gamma k}$ for all $k$.

**Remark 4.1.** Theorem 4.1 is little more than a reformulation of Theorem 7.5(I)(a) of Bradley (2007) in terms of copula functions. Though well understood in the probability literature, the connection between mixing conditions and maximal correlation appears to have gone unmentioned in the recent literature in statistics and econometrics on copula-based time series.

**Remark 4.2.** The following example of a stationary autoregressive process that is not $\alpha$-mixing may be familiar to econometricians. It was studied in detail by Andrews (1984); see also Example 2.15 in Bradley (2007) and the references given there. Let $\{\varepsilon_t : t \in \mathbb{Z}\}$ be an independent sequence of random variables that
each take the value 0 with probability 1/2 and the value 1/2 with probability 1/2, and for $t \in \mathbb{Z}$, let $Z_t$ be the limit in mean square of the series $\sum_{k=0}^{\infty} 2^{-k} \varepsilon_{t-k}$. Andrews (1984) showed that the $\alpha$-mixing coefficients corresponding to $\{Z_t\}$ satisfy $\alpha_k = 1/4$ for all $k$. Since $\{Z_t\}$ is a stationary Markov chain, it follows from Theorem 4.1 and the inequality $4\alpha_k \leq \rho_k$ that the copula $C_A$ of $Z_0$ and $Z_1$, which we will refer to as the Andrews copula, must have maximal correlation coefficient $\rho_{C_A} = 1$.

Let us consider the form of the Andrews copula in some more detail. It is known that the marginal distribution of $Z_0$ and of $Z_1$ is uniform on $[0,1]$ (Bradley, 2007, vol. 1, p. 58). Consequently, $C_A$ is simply the joint distribution function of $Z_0$ and $Z_1$. It takes the form

$$C_A(x,y) = \min\left\{x, y, \frac{1}{2} x - \frac{1}{2} + \max\left\{y, \frac{1}{2}\right\}\right\}.$$  \hspace{1cm} (4.2)

(4.2) can be verified using elementary methods. Viewed as a probability distribution on $[0,1]^2$, the Andrews copula is absolutely singular with respect to Lebesgue measure, assigning half its mass uniformly along the line $y = x/2$, and half its mass uniformly along the line $y = x/2 + 1/2$. Letting $f(x) = x$ and $g(y) = 2y1(y \leq 1/2) + (2y - 1)1(y > 1/2)$, it is straightforward to see that $f(Z_0) = g(Z_1)$ a.s., confirming that the Andrews copula does indeed satisfy $\rho_{C_A} = 1$.

**Remark 4.3.** From Theorem 3.2 and Remark 3.8, we know that the Farlie-Gumbel-Morgenstern, Frank, and Gaussian copulas satisfy $\rho_C < 1$, provided in the case of the latter two copulas that the copula parameters are in the interior of the respective parameter spaces. Consequently, stationary Markov chains generated using such copula functions satisfy $\rho$-mixing conditions at a geometric rate.

**Remark 4.4.** The following result may prove useful in verifying the assumption $\rho_C < 1$ for specific copula functions.

**Theorem 4.2.** If the density of the absolutely continuous part of $C$ is bounded away from zero on a set of measure one, then $\rho_C < 1$.

The proof of Theorem 4.2 uses only elementary methods. Compared to Theorem 3.2, Theorem 4.2 relaxes the requirement that $C$ be absolutely continuous with square integrable density, but rules out copulas whose densities become arbitrarily close to zero. The same tradeoff may be found in Corollary 2.7 and Proposition 2.8 of Bryc (1996), which concern the solubility of a certain inverse problem involving conditional expectation operators. The proof of Theorem 4.2 resembles the proof of Lemma 3.4 in that paper.

The $t$-copula and Marshall-Olkin copula (with parameters in the interior of the respective parameter spaces) provide two examples of copula functions whose densities are bounded away from zero. Consequently, Theorem 4.1 and Theorem 4.2 jointly imply that stationary Markov chains constructed using $t$-copulas or
Marshall-Olkin copulas will satisfy $\rho$- and $\alpha$-mixing conditions at a geometric rate. Unfortunately, not all copula functions used in applied work satisfy the condition of Theorem 4.2. In particular, the Clayton and Gumbel copula densities tend to zero at the off-diagonal corners $(0,1)$ and $(1,0)$.

**Remark 4.5.** One might conjecture that Theorem 4.2 would remain true if one required that the density of the absolutely continuous component of $C$ merely be positive on a set of measure one, rather than bounded away from zero. In fact, this conjecture is false. We now provide an example of a pair of random variables $X$ and $Y$ whose unique copula function $C$ satisfies $\rho_C = 1$ and is absolutely continuous with density $c > 0$ a.e. This example is inspired by Figure 1 in Rényi (1959). Let $X,Y$ have joint pdf $h : [0,1]^2 \to \mathbb{R}$ given by

$$h(x,y) = x^3y^3 + A \cdot 1(\log(1+x) \leq y \leq e^x - 1),$$

where $A > 0$ is such that $\int_0^1 \int_0^1 h(x,y) dx dy = 1$. Note that $h(x,y) = h(y,x)$ for all $x,y \in [0,1]$. Since the joint and marginal densities of $X$ and $Y$ are positive a.e., $X$ and $Y$ admit a unique copula $C$ that is absolutely continuous with density $c > 0$ a.e. (see e.g. p. 197 in McNeil et. al., 2005). As $z \to 0^+$, one can show using Taylor approximations that

$$P(X \leq z, Y \leq z) = \frac{1}{16} z^8 + A(z^2 + 2z - 2(1 + z) \log(1 + z)) = \frac{1}{3} Az^3 + O(z^4)$$

and that

$$P(X \leq z) = \frac{1}{16} z^4 + A(e^z - 1 - (1 + z) \log(1 + z)) = \frac{1}{3} Az^3 + O(z^4),$$

with the first equality in the second line holding for $z \in [0,\log 2]$. Therefore, the coefficient of lower tail dependence corresponding to $C$ satisfies

$$\mu_L = \lim_{z \to 0^+} \frac{C(z,z)}{z} = \lim_{z \to 0^+} \frac{C(F(z),F(z))}{F(z)} = \lim_{z \to 0^+} \frac{P(X \leq z, Y \leq z)}{P(X \leq z)} = 1,$$

where $F$ is the common cdf of $X$ and $Y$. For $n \in \mathbb{N}$, define the function $f_n \in L_2[0,1]$ by

$$f_n(x) = \frac{n}{\sqrt{n-1}} \left( 1 \left( x \leq \frac{1}{n} \right) - \frac{1}{n} \right).$$

Note that $\int f_n = 0$ and $\int f_n^2 = 1$. It is simple to show that

$$\int_0^1 \int_0^1 f_n(x)f_n(y)C(dx,dy) = \frac{n^2}{n-1} C \left( \frac{1}{n}, \frac{1}{n} \right) - \frac{1}{n-1} = nC \left( \frac{1}{n}, \frac{1}{n} \right) + o \left( \frac{1}{n} \right).$$
Letting \( n \to \infty \), we obtain

\[
\lim_{n \to \infty} \int_0^1 \int_0^1 f_n(x)f_n(y)C(dx, dy) = \lim_{z \to 0^+} \frac{C(z, z)}{z} = \mu_L = 1,
\]

implying that \( \rho_C = 1 \).

Remark 4.6. For copulas not satisfying the condition of Theorem 4.2, and for which \( \rho_C \) is unknown, one may nevertheless seek to verify the condition \( \rho_C < 1 \) numerically.

**Theorem 4.3.** Given a copula \( C \), for \( n \in \mathbb{N} \) let \( K_n \) be the \( n \times n \) matrix with \( (i,j) \)th element given by

\[
K_n(i,j) = C \left( \frac{i}{n}, \frac{j}{n} \right) - C \left( \frac{i-1}{n}, \frac{j}{n} \right) - C \left( \frac{i}{n}, \frac{j-1}{n} \right) + C \left( \frac{i-1}{n}, \frac{j-1}{n} \right) - \frac{1}{n^2}.
\]

Let \( \varrho_n \) denote the maximum eigenvalue of \( K_n \). Then \( n\varrho_n \to \rho_C \) as \( n \to \infty \).

Theorem 4.3 shows that \( \rho_C \) may be approximated arbitrarily well by computing the maximum eigenvalue of a matrix of sufficiently large dimension. No assumptions whatsoever are placed on the copula \( C \). In Figure 4.1 we plot approximate values of \( \rho_C \) for the Clayton and Gumbel copulas. These copulas are given by

\[
C_{\text{Clayton}}(x, y; \theta) = \left( x^{-\theta} + y^{-\theta} - 1 \right)^{-1/\theta},
\]

\[
C_{\text{Gumbel}}(x, y; \theta) = \exp \left( - \left( (-\ln x)^\theta + (-\ln y)^\theta \right)^{1/\theta} \right),
\]

where we have \( \theta \in (0, \infty) \) for the Clayton copula and \( \theta \in (1, \infty) \) for the Gumbel copula. The horizontal axis measures lower tail dependence for the Clayton copula, and upper tail dependence for the Gumbel copula. In the former case we have \( \mu_L = 2^{-1/\theta} \), while in the latter case we have \( \mu_U = 2 - 2^{1/\theta} \); see e.g. McNeil et. al. (2005). At each increment of 0.01 on the horizontal axis, an approximate maximal correlation coefficient for each copula was calculated by computing the maximum eigenvalue of a \( 200 \times 200 \) matrix. Maximal correlation values between these increments were obtained by linear interpolation. Computations were implemented using Ox version 5.10.

Figure 4.1 confirms what basic intuition would suggest: the maximal correlation coefficient for both the Clayton and Gumbel copulas increases smoothly from zero to one as the tail dependence coefficient increases from zero to one. Consequently, Theorem 4.1 implies that stationary Markov chains constructed using a Clayton or Gumbel copula with \( \theta < \infty \) will satisfy \( \rho \)- and \( \alpha \)-mixing conditions at a geometric rate. Note that the curves in Figure 4.1 plotting maximal correlation never fall below the 45 degree line. This is because the upper and lower tail dependence coefficients can be written as the limit of correlations between indicator functions, and are therefore bounded by the maximal correlation coefficient.
Figure 4.1: Maximal correlation coefficients for Clayton and Gumbel copulas

We are unable to give precise bounds on the accuracy of the numerical approximations used to generate Figure 4.1. In the case of the Gaussian copula with linear correlation parameter \( \rho \) ranging between 0.01 and 0.99 in intervals of 0.01, we found that the approximate maximal correlation obtained using a 200 \( \times \) 200 matrix differed from the true maximal correlation (known to be equal to \( \rho \)) by no more than 0.001.

Our conclusion that Clayton and Gumbel copulas generate a geometric rate of \( \rho \)-mixing is at odds with a claim made by Ibragimov and Lentzas (2008), who argue that Clayton copulas generate a polynomial decay rate of the \( L_1 \) Schweizer-Wolff distance \( \kappa_k = \| C_k(x,y) - xy \|_1 \). Since \( |C_k(x,y) - xy| \leq \alpha_k \leq (1/4) \rho_k \), a geometric rate of \( \rho \)-mixing implies that \( \kappa_k \) must decay at least geometrically fast, and not at a polynomial rate. Ibragimov and Lentzas arrive at their conclusion by approximating \( \kappa_k \) numerically using a discretization method, and then regressing the log of the approximate \( \kappa_k \) on \( \log k \) and a constant. They find that the coefficient of \( \log k \) is significantly negative, and argue that this implies a polynomial decay rate of \( \kappa_k \). However, in view of the slowly varying nature of \( \log k \) as \( k \to \infty \), one may expect the regression estimates reported by Ibragimov and Lentzas to be largely determined by the regression fit at small values of \( k \) - yet our interest lies with the behavior of \( \kappa_k \) as \( k \to \infty \). A further concern is that positive bias in the approximation of \( \kappa_k \) when \( \kappa_k \) is close to zero may cause the rate of decay of \( \kappa_k \) to appear slower than is actually the case.

5 Conclusion

In this paper we have identified conditions under which a copula function generates a stationary Markov chain that satisfies mixing conditions at a geometric rate. In particular, for non-extreme parameter values, we have demonstrated that the Farlie-Gumbel-Morgenstern, Frank and Gaussian copulas generate geometric
rates of $\beta$- and $\rho$-mixing, and that the Marshall-Olkin, Clayton, Gumbel and $t$-copulas generate a geometric rate of $\rho$-mixing. The conditions under which we obtain geometric $\rho$-mixing are substantially weaker than those under which we obtain geometric $\beta$-mixing, as well as being easily verifiable. Our results complement those of Chen et. al. (2009), who establish a geometric rate of $\beta$-mixing for Clayton, Gumbel and $t$-copulas. Taken together, the results here and in that paper establish that, for a wide class of copula functions, copula-based Markov models exhibit dependence properties typical of short memory time series.

A Mathematical Appendix

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

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

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 (1958ab, 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

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

As observed by Lancaster (1958), Rényi (1959) and Sarmanov (1958ab, 1961), $\lambda_1$ is equal to the maximal correlation of $C$. Since this quantity is assumed less than one, the proof is complete. 

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

$$\iint_0^1 f(x)g(y)c(x,y)dxdy = \left(\iint_0^1 f^2c(x,y)dxdy\right)^{1/2} \left(\iint_0^1 g^2c(x,y)dxdy\right)^{1/2},$$

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 $c = 0$ a.e. on $(A \times B) \cup (A^c \times B^c)$.

Suppose next that $c = 0$ a.e. 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 $[0, 1]$, over which $c$ integrates to one. Since neither $f$ nor $g$ is constant a.e., it follows that $\rho_C = 1$. 

**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

$$x^{-1} \left(\int_0^x \phi_i(z)dz\right)^2 \leq x^{-1} \left(\int_0^x dz\right) \left(\int_0^x \phi_i(z)^2dz\right) = \left(\int_0^x \phi_i(z)^2dz\right).$$
Square integrability of \( \phi_i \) therefore implies that \( \lim_{x \to 0^+} x^{-1/2} \int_0^x \phi_i(z) dz = 0 \). 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-Schwartz, we have

\[
\|\xi_n\|_{\infty} = \left\| x^{-1} \int_0^x \int_0^x \left( c(u, v) - 1 - \sum_{i=1}^n \lambda_i \phi_i(u) \phi_i(v) \right) du \right\|_{\infty} \\
\leq \left\| \left( \int_0^x \int_0^x \left( c(u, v) - 1 - \sum_{i=1}^n \lambda_i \phi_i(u) \phi_i(v) \right)^2 du \right) \right\|_{\infty}^{1/2} \\
= \left( \int_0^1 \int_0^1 \left( c(u, v) - 1 - \sum_{i=1}^n \lambda_i \phi_i(u) \phi_i(v) \right)^2 du \right)^{1/2}.
\]

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

**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 \( c(x, y) \geq \varepsilon \) a.e. 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

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

Since \( (f(x) - g(y))^2 \geq 0 \) and \( c(x, y) \geq \varepsilon \) a.e., we have

\[
\iint (f(x) - g(y))^2C(dx, dy) \geq \iint (f(x) - g(y))^2c(x, y)dx dy \\
\geq \varepsilon \iint (f(x) - g(y))^2dx dy \\
= 2\varepsilon.
\]
Since it is also the case that \( \iint (f(x)^2 + g(y)^2)C(dx, dy) = 2 \), we obtain \( \iint f(x)g(y)C(dx, dy) \leq 1 - \varepsilon \), 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

\[
f(x) = \sum_{i=1}^{n} f_i 1(\frac{i-1}{n}, \frac{i}{n}) (x),
\]

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

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

Consequently, \( n\varrho_n \) 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 \( n\varrho_n \) is the maximum of \( \iint f(x)g(y)C(dx, dy) \) over \( f, g \in S_n \) 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 \( L_2[0, 1] \).

**References**


