

# Copulas and long memory\*

Rustam Ibragimov<sup>†</sup>

*Imperial College Business School and Innopolis University*

*e-mail: [irustam@imperial.ac.uk](mailto:irustam@imperial.ac.uk)*

and

George Lentzas

*Columbia Business School*

**Abstract:** This paper focuses on the analysis of persistence properties of copula-based time series. We obtain theoretical results that demonstrate that Gaussian and Eyraud-Farlie-Gumbel-Morgenstern copulas always produce short memory stationary Markov processes. We further show via simulations that, in finite samples, stationary Markov processes, such as those generated by Clayton copulas, may exhibit a spurious long memory-like behavior on the level of copulas, as indicated by standard methods of inference and estimation for long memory time series. We also discuss applications of copula-based Markov processes to volatility modeling and the analysis of nonlinear dependence properties of returns in real financial markets that provide attractive generalizations of GARCH models. Among other conclusions, the results in the paper indicate non-robustness of the copula-level analogues of standard procedures for detecting long memory on the level of copulas and emphasize the necessity of developing alternative inference methods.

**Keywords and phrases:** Long memory processes, short memory processes, copulas, measures of dependence, autocorrelations, persistence, volatility, GARCH.

Received March 2014.

---

\*A working paper version of the manuscript was previously circulated as Ibragimov and Lentzas (2008). The authors are grateful to an anonymous referee for many helpful suggestions. We thank Brendan Beare, Xiaohong Chen, Umberto Cherubini, Yanqin Fan, Jerry Hausman, Alexander McNeil, Anna Mikusheva, Andrew Patton, Artem Prokhorov, Neil Shephard and Murad Taqqu for useful comments. We also thank the participants at seminars at the Departments of Economics at Harvard University and MIT, the exploratory seminar on “Stochastics and Dependence in Finance, Risk and Insurance” at the Radcliffe Institute for Advanced Study, Young Economists’ Jamboree in Econometrics at Duke, the Symposium on “Modeling Multivariate Dependence” at the Oxford-Man Institute of Quantitative Finance and the conference on “Dynamic copula methods in finance” at the University of Bologna for discussion of the results. The research of Rustam Ibragimov for this paper was supported by a grant from the Russian Science Foundation (Project No. 16-18-10432).

<sup>†</sup>Address for manuscript correspondence: Rustam Ibragimov, Imperial College Business School, Imperial College London, South Kensington Campus, London SW7 2AZ, United Kingdom, e-mail: [irustam@imperial.ac.uk](mailto:irustam@imperial.ac.uk), Phone: +44 (0)20 7594 9344, Fax: +44 (0)20 7594 9184

## 1. Introduction

### 1.1. *Dependence and memory in time series*

Dependence is one of the most fundamental concepts in econometrics, statistics and probability. Numerous notions of dependence are available in the literature (see, among others, the reviews in Joe, 1997; Nze and Doukhan, 2004). Motivated, in part, by empirical applications in economics and finance, many works provide examples of time series and stochastic processes that exhibit dependence and autocorrelation properties ranging from short to long memory (see, among others, Cont, 2001, and references therein). Naturally, stationary Markov processes are regarded as canonical examples of short memory processes. The value of such processes at a given time depends only on their value at the previous period. In addition, the autocorrelation functions of the most commonly used stationary Markov processes, stationary autoregressive time series, exhibit fast exponential decline to zero as lags increase. There are several definitions of long memory and persistence in a time series  $\{X_t\}$  available in the literature. These definitions differ in measures of dependence between the variables  $X_t$  and  $X_{t+h}$  they are based upon. The most commonly used notions of long memory employ the standard autocovariance or autocorrelation functions and take their slow decay to be the defining property of long memory processes (see, among others, Lo 1991, Baillie 1996, Hosking 1996, Doukhan, Oppenheim and Taqqu 2003 and Appendix A1 for a review of the commonly used definitions of long memory processes and their properties).

### 1.2. *Copulas in economics and finance*

In recent years, a number of studies in economics, finance and econometrics have argued that the use of (auto)correlations and (auto)covariances is problematic in many settings, including the departure from Gaussianity and elliptic distributions that is common in economic, financial and insurance market data (see, among others, Embrechts, Klüppelberg and Mikosch, 1997, Embrechts, McNeil and Straumann 2002, McNeil, Frey and Embrechts 2015, Ibragimov, Ibragimov and Walden 2015, and Ibragimov and Prokhorov, 2017). As discussed in Granger (2003), naturally, since (auto)correlations capture only linear relationships, they may not be appropriate in describing persistence in unconditional distributions of time series, that is, persistent relations that are invariant under increasing transformations of data. In addition, (auto)correlations and (auto)covariances are defined only in the case of data with finite second moments. Furthermore, their reliable estimation is problematic in the case of infinite fourth moments (see Davis and Mikosch 1998, Mikosch and Stărică 2000 and the discussion in Cont 2001). At the same time, as discussed in a number of studies (see Loretan and Phillips, 1994; Cont, 2001; Ibragimov, 2009b; Ibragimov, Jaffee and Walden, 2009, and references therein) heavy-tailed behavior with infinite fourth moments is present in many financial, insurance and economic market data sets and even first moments or variances

are infinite for certain time series in finance and economics. Several approaches have been proposed recently to deal with the above problems. One of these approaches, which is becoming increasingly popular in dependence modeling and analysis is the one based on copulas. Copulas are (dependence) functions that allow one, by a famous theorem due to Sklar, to represent a joint distribution of random variables (r.v.'s) as a function of marginal distributions (see Joe 1997, Nelsen 1999 and Appendix A2 for the definition of copulas and a review of their main properties). Copulas, therefore, allow one to separate the analysis of dependence from the properties of marginals (for instance, heavy-tailedness and skewness) and to quantify their relative contributions to a model in consideration. In recent years, copulas and related concepts have been applied to a wide range of problems in economics, finance, econometrics, statistics and probability (see Cherubini, Luciano and Vecchiato 2004, and references therein, Ibragimov 2009a, de la Peña, Ibragimov and Sharakhmetov 2006, Granger, Teräsvirta and Patton 2006, Hu 2006, Patton 2006 and Lowin 2007, McNeil, Frey and Embrechts 2015, and Ibragimov and Prokhorov, 2017). Chen and Fan (2004, 2006) consider copula estimation procedures for time-series based on bivariate copulas and apply the results in the problems of evaluating density forecasts. Fermanian, Radulović and Wegkamp (2004) establish weak convergence of empirical copula processes. Doukhan, Fermanian and Lang (2004) focus on the analysis of the asymptotics of empirical copula processes for weakly dependent sequences of random vectors. A number of works have also focused on dependence characterizations for time series and joint cdf's (see the review in Nze and Doukhan 2004, the monographs by Joe 1997 and Nelsen 1999 and the references therein). Darsow, Nguyen and Olsen (1992) obtain characterizations of first-order Markov chains in terms of copula functions corresponding to their bivariate distributions. Ibragimov (2009a) provides extensions of these characterizations to the case of Markov processes of an arbitrary order and establishes necessary and sufficient copula-based conditions for (possibly higher-order) Markov processes to exhibit  $m$ -dependence,  $r$ -independence or conditional symmetry. The results in Ibragimov (2009a) are based on general  $U$ -statistics-based representations for joint distributions and copulas of dependent r.v.'s obtained in de la Peña, Ibragimov and Sharakhmetov (2006) (see Section 1.4).

### 1.3. Copula-based approaches to long memory

Granger (2003) proposes definitions of long memory and short memory processes  $\{X_t\}_{t=-\infty}^{\infty}$  using the (copula-linked) Hellinger measure of dependence  $H(t, h) = H_{X_t, X_{t+h}}$  between the r.v.'s  $X_t$  and  $X_{t+h}$  given by relations (33) and (36) in Appendix A2. He suggests calling a process  $\{X_t\}$  long or short memory depending on whether, for some constant  $A > 0$ ,

$$H(t, h) \sim Ah^{-p}, \quad h \rightarrow \infty, \quad (1)$$

where  $p > 0$ , or  $H(t, h) = O[\exp(-Ah)]$ ,  $h \rightarrow \infty$  (as in relations (19) and (21) in Appendix A1 for the case of autocovariances).<sup>1</sup>

Granger (2003) further indicates that the difficulty with the above definition is that it depends on a particular measure of dependence and it has to be shown that some general rule applies. He also indicates the possibility of using other measures of dependence and remarks that there seems to be no single measure that provides a dominant alternative to autocovariances and autocorrelations, while the measure  $H$  has the advantage of simplicity and of having the link to copulas via relation (36) in Appendix A2.

#### 1.4. Objectives and key results

This paper focuses on the analysis of persistence properties of copula-based time series. We introduce and compare various definitions of long memory on the level of copulas (Section 2). In particular, we derive relationships between the commonly used measures of dependence and the implied long memory concepts (Proposition 1). We further present theoretical results that generalize the well-known autocovariance-based short memory properties of stationary autoregressive processes and show that several widely used copula families, such as Gaussian and Eyraud-Farlie-Gumbel-Morgenstern (EFGM) copulas, always produce short memory stationary Markov processes (Proposition 2). This is the case in terms of an exponential decay of both the autocorrelation functions and copula-based dependence measures. In Section 3 we show via simulations that, in finite samples, standard methods of inference and estimation for long memory time series may indicate a spurious long memory-like behavior on the level of copulas for stationary Markov processes. In particular, the standard estimation methods applied to Clayton copula-based stationary Markov processes produce the point estimates ranging from 0.08 to 0.14 for the (spurious) long memory parameter  $p$  in an analogue of characteristic property (1) for the dependence measure based on the  $L^1$ -distance to the independence copula (Section 3.5). We further discuss applications copula-based Markov processes to volatility modeling and the analysis of nonlinear dependence properties of returns in real financial markets that provide attractive generalizations of GARCH families of models (Section 4). The constructions and the results obtained in the paper overcome several technical problems due to complexity of copulas and computations (see Sections 3.1 and 3.2). The methods proposed can also be used in the analysis of a number of related problems in econometrics and stochastic processes (see the discussion in Section 5).

The focus of the theoretical part of the paper on the Gaussian and EFGM copula families is motivated by importance of these classes of dependence functions and corresponding and related distributions (Gaussian and EFGM copulas

---

<sup>1</sup>As usual, throughout the paper, for functions  $g$  and  $h$ ,  $g(x) \sim h(x)$  as  $x \rightarrow \infty$  denotes that  $\lim_{x \rightarrow \infty} \frac{g(x)}{h(x)} = 1$ , and  $g(x) = O(h(x))$  denotes that  $|g(x)| \leq A|h(x)|$  for some constant  $A > 0$  and all  $x > x_0$ .

and distributions, their mixtures and approximations based on them) in copula theory and many other areas of probability and statistics.

Importance of Gaussian distribution and copula families cannot be overemphasized. In particular, as discussed in many works in the literature (see, among others, Chs. 3 and 10 in [McLachlan and Peel, 2000](#) and Chs. 6 and 8 in [Fruhwrth-Schnatter, 2006](#)), their mixtures can be used to represent a variety of heterogeneity and dependence patterns arising in many applications.<sup>2</sup> Naturally, Gaussian copulas and distributions provide a building block for defining tail-dependent Student- $t$  copula and distribution families and their mixtures (*Op. cit.*). Many works in the copula literature focus on the analysis of tail-dependent copula mixtures, including those that involve Gaussian and related copulas, and their applications (see, among others, [Patton, 2006](#), and [Hu, 2006](#)).

Further, as is shown in [Sharakhmetov and Ibragimov \(2002\)](#) (see also [de la Peña et al., 2006](#), p. 190), EFGM families can be used to represent any joint distribution of two-valued random variables. [de la Peña et al. \(2006\)](#) obtain  $U$ -statistic based representations for general multivariate copulas and distributions and apply them to define new wide classes of copulas, including power copula families. From the results in [de la Peña et al. \(2006\)](#) it follows, essentially, that extensions of EFGM families in the form of expansions by degenerate  $U$ -statistics kernels can be used to represent and approximate any multivariate copulas and joint distributions, including those that exhibit tail dependence properties important in financial and economic applications. The  $U$ -statistic based characterizations of multivariate distributions and copulas are used in [Ibragimov \(2009a\)](#) to obtain copula-based representations for general Markov processes and other time series, characterize their dependence properties and introduce some new flexible copula classes, such as Fourier copulas.<sup>3</sup>

[Ibragimov and Prokhorov \(2016\)](#) focus on the analysis of diversification optimality for heavy-tailed risks with dependence in the form of EFGM copulas and their power copula family extensions (see also [Ibragimov et al., 2015](#), and Ch. 4 in [Ibragimov and Prokhorov, 2017](#)). As discussed in [Ibragimov and Prokhorov \(2016\)](#), EFGM copulas provide first order approximations to Ali-Mikhail-Haq (AMH), Plackett and Frank copula families (see, e.g., [Nelsen, 1999](#), p. 100, 133). From the approximation results in [Nelsen et al. \(1997\)](#); [Cuadras \(2009\)](#); [Cuadras and Diaz \(2012\)](#) it follows that power generalizations of the bivariate EFGM copulas with cubic terms can be used to approximate some well-known families of copulas such as the copulas of [Kimeldorf and Sampson \(1975\)](#) and [Lin \(1987\)](#). From the results it also follows that such extensions of the EFGM copulas are second-degree Maclaurin approximations to members of the Frank and Plackett copula families. [Cuadras \(2009\)](#) studies the power series class of copulas, obtained as weighted geometric means of the EFGM and AMH copulas, and shows that it provides first-order approximations to Gumbel-Barnett and Cuadras-Auge copulas. [Cuadras and Diaz \(2012\)](#) provide approximations of tail-dependent Clayton-Oakes copulas, which also have the form of power-type

---

<sup>2</sup>We are grateful to an anonymous referee for these references.

<sup>3</sup>See also [Lowin \(2007\)](#) for the analysis of properties and applications of Fourier copulas in finance and economies.

generalizations of EFGM copulas.

In papers that appeared after an earlier version of this work referred to in them was prepared for publication, Chen, Wu and Yi (2009) and Beare (2010) obtain several results on short memory and mixing properties of stationary copula-based Markov processes that are closely related to those presented in this paper (see the discussion in Section 5). In particular, among many other results, Beare (2010) establishes sufficient conditions for short memory properties of copula-based stationary Markov processes that generalize the results for stationary Markov time series generated by Gaussian and EFGM copula families obtained, for the first time, in this work. Beare (2010) also provides numerical results that suggest exponential decay in mixing coefficients and, thus, also in copula-based dependence measures between  $X_t$  and  $X_{t+h}$  for Clayton copula-based stationary Markov processes. Chen, Wu and Yi (2009) obtain theoretical results that show that tail-dependent Clayton, survival Clayton, Gumbel and  $t$ -copulas always generate Markov processes that are geometric ergodic and hence geometric  $\beta$ -mixing and short memory in any meaningful sense of the world: In particular, the processes are short memory on the level of copulas. Given the above subsequent recent results in the literature on short memory properties of many copula-based time series, the conclusions on a spurious long memory-like behavior of copula-based stationary Markov processes in finite samples in this paper further indicate non-robustness of copula-level analogues of standard procedures for detecting long memory on the level of copulas and emphasize the necessity of developing alternative inference methods (see further discussion in Section 5).

### 1.5. Organization of the paper

The paper is organized as follows. Section 2 discusses several copula-based definitions of long memory processes and relations among them. It also presents the results on short memory properties of stationary Markov processes based on widely used Gaussian and EFGM copulas. Section 3 provides numerical results that indicate that stationary Markov processes, such as those generated by Clayton copulas, may exhibit a spurious long memory-like behavior on the level of copulas in finite samples. Section 4 presents empirical applications of copula-based time series to volatility modeling. Section 5 makes some concluding remarks. Section 6 contains the proofs of the results derived in the paper. Appendix A1 reviews definitions of long memory processes based on autocorrelation functions. Appendix A2 discusses the definition and main properties of copula functions, together with their examples. Appendix A3 reviews copula-based characterizations of Markov processes.

## 2. Copulas and long and short memory: Definitions and main properties

The concepts related to copulas and dependence measures dealt with in this section and throughout the paper are defined in Appendices A2 and A3.

In order to highlight the main concepts and ideas discussed, we assume throughout the paper that all copulas, r.v.'s and their distributions considered are absolutely continuous, if not stated otherwise. These assumptions imply, in particular, that the copulas corresponding to finite-dimensional distributions of the processes in consideration are unique (see Proposition 4 in Appendix A2). However, most of the results discussed in the paper can be extended to the case of not necessarily absolutely continuous distributions, copulas and processes.

Similar to Granger (2003), it is natural to consider definitions of long and short memory that involve the speed of decay of dependence measures different from the Hellinger distance discussed in Section 1.3. In particular, one can define long memory in a way analogous to Granger (2003) using the following measures of dependence between the r.v.'s  $X_t$  and  $X_{t+h}$  (see Appendix A2 for definition and a review): Pearson's  $\phi^2$  coefficient  $\phi_X^2(t, h) = \phi_{X_t, X_{t+h}}^2$ , relative entropy  $\delta_X(t, h) = \delta_{X_t, X_{t+h}}$ , general divergence measures  $D_X^\psi(t, h) = D_{X_t, X_{t+h}}^\psi$ , or measures of dependence based on the distances to the product (independence) copula  $\kappa_X(t, h) = \kappa_{X_t, X_{t+h}}$  and  $\lambda_X^2(t, h) = \lambda_{X_t, X_{t+h}}^2$ . We will say that a process  $\{X_t\}_{t=-\infty}^\infty$  exhibits  $\phi^2$ -long memory (resp.,  $\phi^2$ -short memory) if, for some constant  $A > 0$ ,  $\phi_X^2(t, h) = \phi_{X_t, X_{t+h}}^2 \sim Ah^{-p}$  with  $p > 0$  (resp.,  $\phi_X^2(t, h) = O[\exp(-Ah)]$ ). The notions of  $\delta$ -,  $D^\psi$ -,  $\kappa$ - and  $\lambda^2$ -long memory and short memory processes are defined in a similar way. We will refer to time series  $\{X_t\}$  exhibiting long (short) memory in the sense of the definition in Granger (2003) as  $H$ -long memory (resp.,  $H$ -short memory) processes.

The following proposition provides several relationships between the measures of dependence  $\phi_{X,Y}^2$ ,  $\delta_{X,Y}$ ,  $H_{X,Y}$ ,  $D_{X,Y}^\psi$ ,  $\kappa_{X,Y}$ ,  $\lambda_{X,Y}^2$ ,  $\nu_{X,Y}$  and the correlation coefficient  $Corr(X, Y)$ . This proposition holds for r.v.'s  $X, Y$  with arbitrary dependence.

**Proposition 1.** *The following inequalities hold:*

$$\delta_{X,Y} \leq \ln(1 + \phi_{X,Y}^2) \leq \phi_{X,Y}^2, \tag{2}$$

$$\kappa_{X,Y} \leq \lambda_{X,Y} \leq \nu_{X,Y} \leq \phi_{X,Y}, \tag{3}$$

$$H_{X,Y} \leq (1/2)\phi_{X,Y}. \tag{4}$$

If  $EX^2, EY^2 < \infty$ , then

$$Corr(X, Y) \leq \phi_{X,Y}. \tag{5}$$

From Proposition 1 it follows that  $\phi$ -long memory is not shorter than long memory in the usual sense (see (19)) or than  $\delta$ -,  $\kappa$ -,  $\lambda$ -,  $\nu$ - and  $H$ -long memory. More precisely, if a process  $\{X_t\}_{t=-\infty}^\infty$  exhibits  $\phi$ -short memory, then it is also a  $\delta$ -,  $\kappa$ -,  $\lambda$ -,  $\nu$ - and  $H$ -short memory process. By (5), provided  $EX_t^2 < \infty$ , such process also has short memory in the sense of the usual definition with autocovariances exhibiting at most exponential decline in (21). In addition, if a process  $\{X_t\}_{t=-\infty}^\infty$  exhibits long memory with hyperbolic decay for autocovariances in (19) or for one of the measures of dependence  $\delta_X(t, h)$ ,  $\kappa_X(t, h)$ ,  $\lambda_X(t, h)$ ,  $\nu_X(t, h)$  or  $H_X(t, h)$  as  $h \rightarrow \infty$ , then the decay of  $\phi_X^2(t, h)$  is not slower than hyperbolic.

**Remark 1.** Evidently, the inequalities in (1) hold as equalities in the case of independent r.v.'s  $X, Y$ , as, under independence, all the dependence measures  $\delta, \phi, \kappa, \lambda, \nu, H$  and the correlation  $\text{Corr}$  (for finite second moments) equal zero. As follows from Example 1 in the Appendix, for r.v.'s  $X, Y$  that have a Gaussian copula  $C_\rho(u, v)$  with the correlation parameter  $\rho$ , one has  $\delta_{X,Y} = -0.5 \ln(1 - \rho^2)$ ,  $\phi_{X,Y}^2 = \rho^2/(1 - \rho^2)$ . Therefore, interestingly, inequalities (2) also hold as equalities in the limiting case  $\rho \rightarrow \pm 1$ , that is, in the case of comonotonic/countermonotonic r.v.'s  $X, Y$  with perfect positive/negative dependence.

In what follows, we refer to the processes  $\{X_t\}_{t=1}^\infty$  constructed via (53) in Appendix A3 as stationary Markov processes based on the copula  $C$  or as  $C$ -based stationary Markov processes for short.<sup>4</sup> Under stationarity, the measures of dependence between the r.v.'s  $X_t$  and  $X_{t+h}$  considered above are independent of  $t$  and will be denoted, in what follows, by  $\phi_X^2(h) = \phi_{X_t, X_{t+h}}^2$ ,  $\delta_X(h) = \delta_{X_t, X_{t+h}}$ ,  $\kappa_X(h) = \kappa_{X_t, X_{t+h}}$ ,  $\lambda_X^2(h) = \lambda_{X_t, X_{t+h}}^2$ ,  $H_X(h) = H_{X_t, X_{t+h}}$ ,  $\nu_X(h) = \nu_{X_t, X_{t+h}}$  and  $\gamma(h) = \text{Cov}(X_t, X_{t+h})$  (as usual, we assume finiteness of the second moments whenever the covariance or correlation are used as a measure of dependence). In what follows, for two r.v.'s  $X$  and  $Y$ , the notation  $X \stackrel{d}{=} Y$  means that their (one-dimensional) distributions are the same.

The following proposition justifies, in part, the definitions of short memory processes proposed in Granger (2003) and in the above discussion and shows that all stationary Markov processes based on Gaussian or EFGM copulas  $C_\rho^G(u, v)$ ,  $C_\alpha^{EFGM}(u, v)$  in (29), (31) exhibit short memory regardless of what dependence measure ( $\phi^2, \delta, \kappa, \lambda^2, H, \nu$  or the covariance  $\gamma$ ) is used in their definition.

**Proposition 2.** Let  $\{X_t\}_{t=1}^\infty$  be a stationary Markov process based on a Gaussian copula  $C_\rho^G(u, v)$  with the correlation coefficient  $\rho$  or on an EFGM copula  $C_\alpha^{EFGM}(u, v)$  with the parameter  $\alpha$ . Then the process exhibits short memory in the sense that its measures of dependence satisfy, for some constant  $A > 0$ ,

$$\phi_X^2(h), \delta_X(h), \kappa_X(h), \lambda_X^2(h), H_X(h), \nu_X(h), \gamma_X(h) = O[\exp(-Ah)] \quad (6)$$

as  $h \rightarrow \infty$ .

**Remark 2.** According to Proposition 2, Gaussian and EFGM copulas cannot be used to construct long memory copula-based stationary Markov processes. This complements the impossibility/reduction results in Ibragimov (2009a) who shows that stationary  $k$ -th order Markov processes cannot exhibit  $m$ -dependence or  $k$ -independence if they are based on Gaussian, Student- $t$  (see relation (30)) or EFGM-type copulas that involve products of functions of their arguments. The proposition also complements the results in Cambanis (1991) that demonstrate that constant, exponential and  $m$ -dependence cannot be exhibited by stationary processes  $\{X_t\}$  whose finite-dimensional copulas are multivariate analogues of bivariate EFGM copulas (see the discussion in Ibragimov, 2009a).

<sup>4</sup>Throughout the paper, stationarity refers to strict stationarity.



Let, for a copula  $C(u, v)$ ,  $\tilde{C}(u, v) = u + v - 1 + C(1 - u, 1 - v)$  denote its survival copula (see Nelsen, 1999, and McNeil, Frey and Embrechts, 2015). From the following proposition it follows that the short (long) memory properties of  $C$ - and  $\tilde{C}$ -based stationary Markov processes are the same. Similar to Proposition 1, this proposition holds for any copula and dependence structures.

**Proposition 3.** *Let  $\{X_t\}$  be a  $C$ -based stationary Markov process and let  $\{\tilde{X}_t\}$  be a stationary Markov process based on the survival copula  $\tilde{C}(u, v)$  of  $C$ . The measures of dependence of the processes  $\{X_t\}$  and  $\{\tilde{X}_t\}$  are the same:  $\phi_X^2(h) = \phi_{\tilde{X}}^2(h)$ ,  $\delta_X(h) = \delta_{\tilde{X}}(h)$ ,  $\kappa_X(h) = \kappa_{\tilde{X}}(h)$ ,  $\lambda_X^2(h) = \lambda_{\tilde{X}}^2(h)$ ,  $H_X(h) = H_{\tilde{X}}(h)$  and  $\nu_X(h) = \nu_{\tilde{X}}(h)$ . In particular, the process  $\{X_t\}$  exhibits long (short) memory in the sense of hyperbolic (resp., at most exponential) decay of the measures of dependence  $\phi_X^2(h)$ ,  $\delta_X(h)$ ,  $\kappa_X(h)$ ,  $\lambda_X^2(h)$ ,  $H_X(h)$  or  $\nu_X(h)$  to zero as  $h \rightarrow \infty$  if and only if the same holds for the process  $\{\tilde{X}_t\}$ .*

The next section shows that stationary Markov processes, such as those generated by Clayton copulas, may exhibit a spurious long memory-like behavior on copula level in finite samples, as indicated by standard inference and estimation methods for long memory time series. In what follows, we choose to use  $\kappa(h) = \kappa_{X_t, X_{t+h}}$  as the measure of copula dependence for two reasons. First, it is more intuitive than divergence measures (where the choice of specific functional form is not trivial). Second, compared to  $\lambda^2(h)$  and to  $\nu(h)$ , the measure  $\kappa(h)$  introduces less numerical error when using the grid approximation. Last and most important, it follows from Proposition 1 that, for example,  $\phi$ -long memory is not shorter than  $\kappa$ -long memory and similar conclusions hold for  $\delta$ -,  $\lambda$ - and  $\nu$ - long memory.

### 3. Copula-based persistent long memory-like processes: Construction

#### 3.1. Empirical problems

Despite their theoretical appeal, applications of copulas to construction of Markov time series context run into serious problems. In particular, given  $C(u, v) := C_{t,t+1}(u, v)$ , the copula between  $X_t$  and  $X_{t+1}$ , evaluation of  $C_{t,t+h}(u, v)$  is extremely difficult due to the presence of nested integrals in the expression for  $C_{t,t+h}(u, v)$  implied by iterations on relation (52), with the complexity increasing as the time lag  $h$  increases. For example,  $C_{t,t+2}(u, v)$  and  $C_{t,t+3}(u, v)$  are given by

$$C_{t,t+2}(u, v) = \int_0^1 \frac{\partial C_{t,t+1}(u, t)}{\partial t} \frac{\partial C_{t+1,t+2}(t, v)}{\partial t} dt \tag{7}$$

$$= \int_0^1 \frac{\partial C(u, t)}{\partial t} \frac{\partial C(t, v)}{\partial t} dt, \tag{8}$$

$$C_{t,t+3}(u, v) = \int_0^1 \frac{\partial C_{t,t+2}(s, v)}{\partial s} \frac{\partial C_{t+2,t+3}(u, s)}{\partial s} ds \tag{9}$$

$$\begin{aligned}
&= \int_0^1 \frac{\partial C(u, s)}{\partial s} \frac{\partial C_{t,t+2}(s, v)}{\partial s} ds \\
&= \int_0^1 \int_0^1 \frac{\partial C(u, s)}{\partial s} \frac{\partial^2 C(s, t)}{\partial s \partial t} \frac{\partial C(t, v)}{\partial t} dt ds. \quad (10)
\end{aligned}$$

Analytical evaluation of these integrals when  $C_{t,t+1}(u, v)$  is, for example, the Clayton, or any other non-product copula, is usually impossible while numerical approximation of these integrals will also fail after two or three lags, even with very low accuracy. Moreover, the use of higher (than one) order Markov models will make the evaluation of  $C_{t,t+h}(u, v)$  even more complicated, making the use of copulas in a Markov time series context problematic in any real world application.

### 3.2. Discretization method

We propose to overcome the problems discussed in Section 3.1 by introducing a copula discretization method where the copula is redefined as a grid which is then used to approximate the partial derivatives and definite integrals in equation (8). This method has three important advantages. First, it allows for the reasonably fast evaluation of  $C_{t,t+h}(u, v)$ , thus opening the way for meaningful empirical applications of copula models to time series. Below, we provide the results for calibrated copulas with  $h$  up to  $50^{th}$  lag. We also propose a method to increase the estimation speed in applications with large  $h$  ( $h$  greater than 50). Second, under the proposed discretization method, the numerical complexity of  $C_{t,t+h}(u, v)$  no longer increases with  $h$  but remains constant at the level implied by the chosen accuracy. This allows one to evaluate the time necessary for the evaluation of  $C_{t,t+h}(u, v)$  so that the accuracy can be chosen given the time constraints of the application. Moreover, the discretization makes it possible to evaluate numerically functionals of the copula, including dependence measures like  $\phi^2(C)$ ,  $\kappa(C)$ ,  $\lambda(C)$ ,  $\delta(C)$ ,  $H(C)$  and  $\nu(C)$ . Last but not least, the discretization method can be directly applied to any parametric and non-parametric copulas and easily generalized to higher-order Markov copula models.

The discretization method works as follows. We start with the copula  $C(u, v) := C_{t,t+1}(u, v)$ , describing the dependence between the r.v.'s  $X_t$  and  $X_{t+1}$ . This may be a copula with a simple parametric closed form (an ‘‘explicit’’ copula), for example, the Clayton copula  $C^{Clayton, \theta}(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}}$  in (26) or a copula with a more complicated form obtained via inversion method discussed in Appendix A2 (an ‘‘implicit’’ copula), like a Gaussian or a  $t$ -copula in (29) and (30).<sup>5</sup> We evaluate the two partial derivatives

$$\frac{\partial C_{t,t+1}(u, v)}{\partial u}$$

---

<sup>5</sup>The use of the terms ‘‘explicit’’ and ‘‘implicit’’ for the considered classes of copulas follows that in Ch. 5 in McNeil, Frey and Embrechts (2015).

$$\frac{\partial C_{t,t+1}(u, v)}{\partial v},$$

analytically in the case of an explicit copula, or by the grid method presented below for an implicit copula. Then we use

$$C_{t,t+2}(u, v) = \int_0^1 \frac{\partial C(u, w)}{\partial w} \frac{\partial C_{t,t+1}(w, v)}{\partial w} dw. \tag{11}$$

to define  $C_{t,t+2}^{Grid}$  by calculating  $C_{t,t+2}(u, v)$  on a  $M \times M$  grid, where  $M$  is the number of grid points and where the integral in equation (11) is evaluated numerically given  $(u, v)$ . The grid  $C_{t,t+2}^{Grid}$  is an  $M \times M$  matrix whose  $(i, j)$  element gives the probability

$$P\left(F_{X_t}(x_t) \leq \frac{i}{M}, F_{X_{t+2}}(x_{t+2}) \leq \frac{j}{M}\right) = P\left(U \leq \frac{i}{M}, V \leq \frac{j}{M}\right),$$

where  $(\frac{i}{M}, \frac{j}{M}) \in [0, 1]^2$ . Then, given  $C_{t,t+2}^{Grid}$ , we use

$$C_{t,t+3}(u, v) = \int_0^1 \frac{\partial C(u, w)}{\partial w} \frac{\partial C_{t,t+2}(w, v)}{\partial w} dw, \tag{12}$$

to create  $C_{t,t+3}^{Grid}$ . The partial derivative in equation (12) is approximated using the  $C_{t,t+2}^{Grid}$  grid, namely by

$$\begin{aligned} \frac{\partial C_{t,t+2}(w, v)}{\partial w} &= \lim_{\delta \rightarrow 0} \frac{C_{t,t+2}(w + \delta, v) - C_{t,t+2}(w, v)}{\delta} \\ &\approx \frac{C_{t,t+2}^{Grid}(i + \frac{1}{M}, j) - C_{t,t+2}^{Grid}(i, j)}{\frac{1}{M}}. \end{aligned}$$

We then iterate up to the required time lag  $h$  to get  $C_{t,t+h}$ . Finally, in order to calculate, for example,  $\kappa(h)$ , the distance of  $C_{t,t+h}$  from the independence copula we numerically approximate the double integral

$$\kappa(h) = \int_0^1 \int_0^1 |C_{t,t+h}(u, v) - uv| dudv,$$

using the  $C_{t,t+h}^{Grid}$ , namely by

$$\begin{aligned} \kappa(h) &= \frac{1}{M^2} \sum_{i=1}^M \sum_{j=1}^M \left| C_{t,t+h}\left(\frac{i}{M}, \frac{j}{M}\right) - \frac{i \times j}{M^2} \right| \\ &= \frac{1}{M^2} \sum_{i=1}^M \sum_{j=1}^M \left| C_{t,t+h}^{Grid}(i, j) - \frac{i \times j}{M^2} \right|. \end{aligned}$$

### 3.3. Clayton copula-based stationary Markov process

Using the discretization method we start with the Clayton copula  $C_{t,t+1}(u, v)$  and we then calculate the implied copulas  $C_{t,t+s}(u, v)$ ,  $s = 1, \dots, h$ , copulas (as captured by the  $C_{t,t+s}^{Grid}$  grids) for a number of different values of  $\theta$ . We numerically evaluate the distance  $\kappa(h)$  for  $h = 1, \dots, 50$  and report the results below. The numerical results indicate spurious slow decrease in  $\kappa(h)$  as the time lag  $h$  increases, pointing out to a spurious  $\kappa$ -long memory-like behavior of the copula-based time series considered in finite samples.

Since we are approximating each copula using a finite  $M \times M$  grid, error is inevitably introduced in our estimates. This error cumulates as the time lag increases, hence a larger number of grid points is required the larger the number of time lags  $h$ . Figure 1 shows how the cumulated error affects  $\kappa(h)$ , the distance from the independence copula for the first ten time lags. We see that the error can cause  $\kappa(h)$  to spuriously increase with  $h$ , as for  $M = 10$  to  $M = 50$ , although in the latter case the increase happens at a higher lag. We also see that  $M \approx 100$  produces reasonably accurate results for the first 10 – 20 lags. Similar results hold for  $\theta = 100$  and  $\theta = 130$  and a higher number of lags, summarized in Figure 2. Moreover, since higher  $\theta$  means that  $C_{t,t+1}$  is closer to the upper Fréchet-Hoeffding bound  $\bar{C}(u, v)$  in (25) giving the measure  $\kappa(h)$  bound of  $\frac{1}{12}$  and thus the error is comparatively smaller, the grid approximation works better for higher  $h$ . This is evident in Figure 3 where we keep the number of grid points constant and alter  $h$  for the first ten time lags.

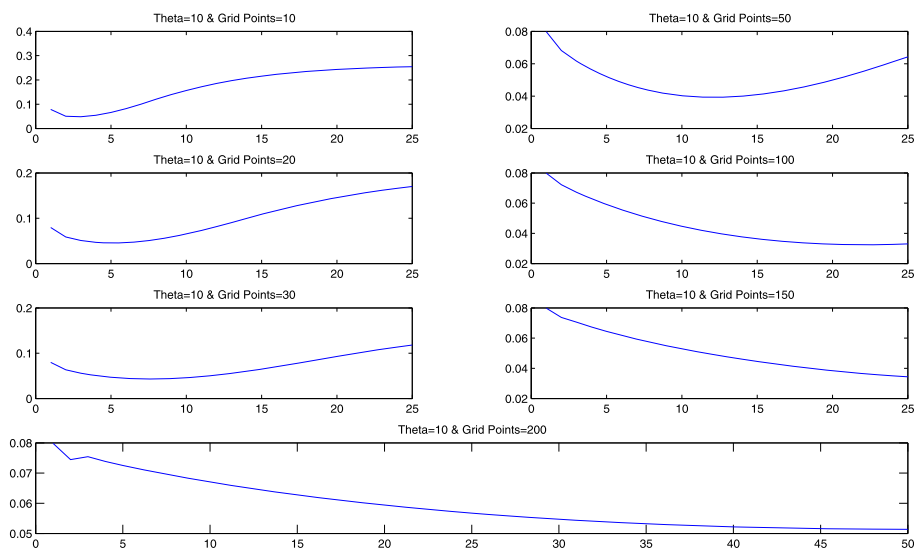


FIG 1. The measure  $\kappa(h)$  (OY axis) as function of lag  $h$  (OX axis): fixed  $\theta = 10$  and the increasing number of grid points  $M = 10, 20, 30, 50, 100, 150, 200$ .

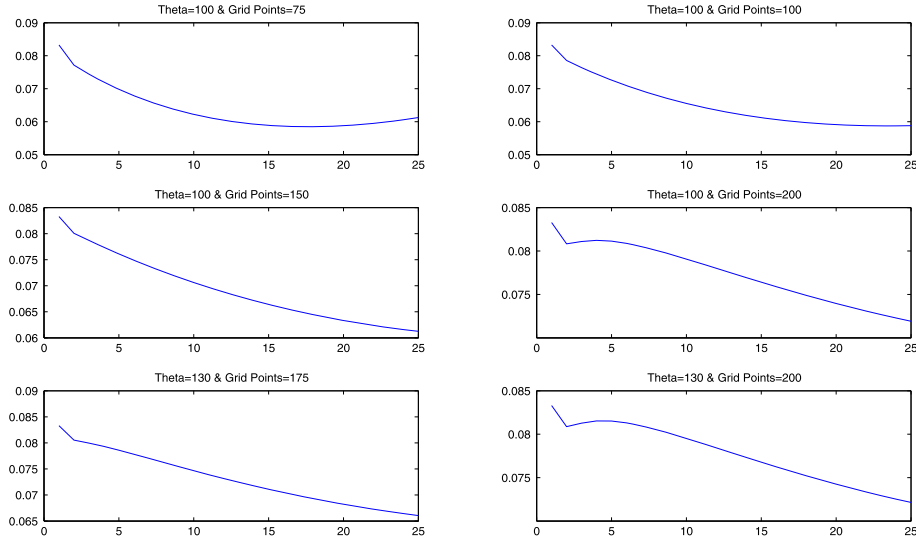


FIG 2. The measure  $\kappa(h)$  (OY axis) as a function of lag  $h$  (OX axis):  $\theta = 100, 130$  and the number of grid points  $M = 75, 100, 150, 175, 200$ .

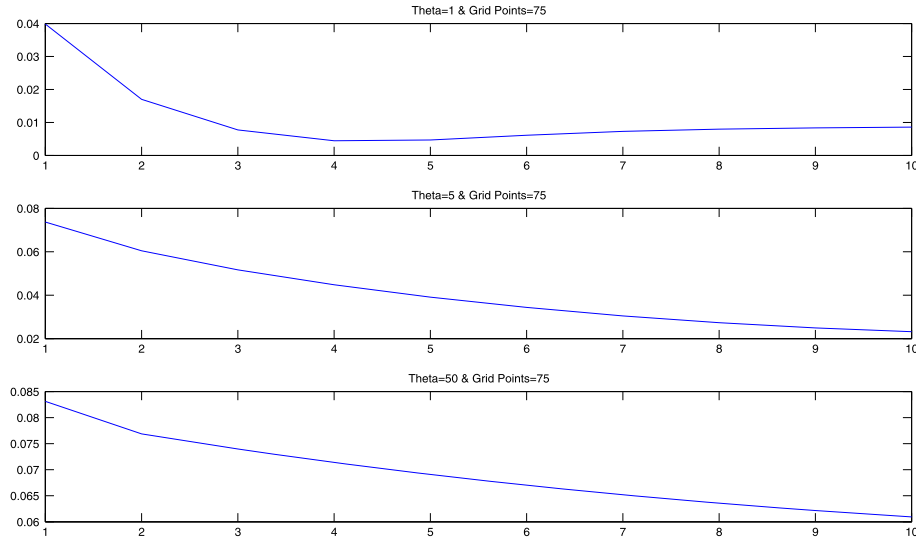


FIG 3. The measure  $\kappa(h)$  (OY axis) as a function of lag  $h$  (OX axis):  $\theta = 1, 5, 50$  and the fixed number of grid points  $M = 75$ .

### 3.4. Tail dependence coefficient

The grid approximation for  $C_{t,t+s}(u, v)$ ,  $s = 1, \dots, h$ , has the advantage that we can also use the estimated grids to fit parametric copulas at each time lag. We

can thus approximate the grids by fitting the Clayton copulas at each lag and then infer the parameters  $\theta$  and the implied coefficients of lower tail dependence for the copulas given by  $\lambda_L = 2^{-1/\theta}$  (see McNeil, Frey and Embrechts, 2015).<sup>6 7</sup> This method can be applied to increase the calibration speed in applications with large  $h$ , by having an increasing number of grid points  $M$  and periodically refitting the Clayton at each increase of  $M$ . Actually, the tail dependence coefficients can be more accurately estimated for a given copula  $C$  directly from the grid, namely by approximating

$$\lambda_L = \lim_{q \rightarrow 0^+} \frac{C(q, q)}{q},$$

for the coefficient of lower tail dependence and

$$\lambda_U = \lim_{q \rightarrow 1^-} \frac{\tilde{C}(q, q)}{q},$$

for the coefficient of upper tail dependence, where, as in Section 2,  $\tilde{C}$  is the survival copula of  $C$ . Using this method, we present in Figure 4 the decay of the estimated coefficient of lower tail dependence  $\lambda_L$  for the first 20 lags. We see that tail dependence decreases fast and essentially disappears after the 10<sup>th</sup> lag.

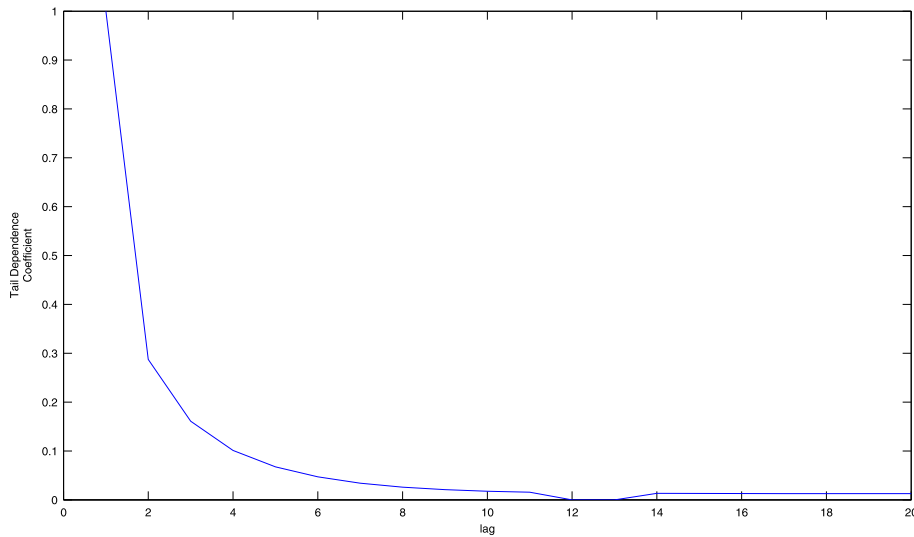


FIG 4. The coefficient of lower tail dependence  $\lambda(h)$  (OY axis) as a function of lag  $h$  (OX axis).

<sup>6</sup>The coefficient of upper tail dependence is zero for Clayton family:  $\lambda_U = 0$ .

<sup>7</sup>An alternative method would be to fit flexible mixture copulas at each lag.

### 3.5. Spurious long memory-like behavior in finite samples

Similar to the preceding sections, we start with the Clayton copula  $C_{t,t+1}(u, v)$ , calculate the implied copulas  $C_{t,t+h}(u, v)$  and numerically evaluate the corresponding distances  $\kappa(h)$  for  $h = 1, \dots, 54$  using the discretization method as described in Sections 3.2-3.4. To obtain estimates of the parameter  $p$  in the following characteristic long memory relation given by an analogue of (1):

$$\kappa(h) \sim Ah^{-p}, \quad (13)$$

we use a grid of  $M = 200$  points and Clayton copula parameter  $\theta = 10$  and use the standard estimation approach where the logarithm of the left-hand side of (13) is regressed on the logarithm of its right-hand side for  $h = 1, \dots, 59$ . The results are as follows (the OLS standard errors are in parentheses):

$$\ln[k(h)] = -2.428 - 0.136 \ln(h), \quad R^2 = 0.96. \\ (0.013) \quad (0.004) \quad (14)$$

Thus, the estimate of the parameter  $p \approx 0.136$  indicates (spurious) slow decay of the distances  $\kappa(h)$  to zero at commonly used lag numbers. Since we are particularly interested in the behavior of the distances  $\kappa(h)$  as the lag length  $h$  increases we also calculate three more regressions, this time using truncated samples, in which we drop the first 10, 20 and 30 observations respectively in order to decrease the effect of the first few lags. The resulting estimates for the parameter  $p$  are as follows (standard errors and  $R^2$  in parentheses): 0.159 (0.005, 0.96), 0.129 (0.007, 0.90), 0.077 (0.008, 0.79).

As discussed in the introduction and Section 5, Chen, Wu and Yi (2009) (see also Beare, 2010) show that Clayton copulas always generate stationary Markov processes that are geometrically ergodic and thus, geometrically  $\beta$ -mixing and short memory on the level of copulas (in particular,  $\kappa$ -short memory). These properties imply that the long memory-like behavior of Clayton copula-based stationary Markov processes in finite samples in the above numerical results is indeed spurious. This, in particular, points out to potential non-robustness of copula-level analogues of standard procedures for detecting long memory and emphasize the necessity of developing alternative inference methods. In particular, the estimates of the parameter  $p$  obtained using (14) imply that such procedures applied for commonly used lag numbers may lead to a conclusion on long memory on the level of copulas despite the series in consideration being weakly dependent, as follows from the results in Beare (2010) and Chen, Wu and Yi (2009).

## 4. An empirical application: Volatility modeling using copulas

We apply the copula-based first-order Markov model to capturing the dependence structure of squared returns. This is motivated by the autoregressive

representation of the GARCH family of models, for which the squared residual process is ARMA with a martingale difference series innovation. Other authors have already drawn on such squared residuals structures to extend the usual GARCH models; Drost and Nijman (1993) introduced the class of weak GARCH models, captured by a weak white noise innovation ARMA structure for squared residuals, and showed that (unlike usual GARCH) are closed under temporal dependence. Similarly, Meddahi and Renault (2004) built on this idea to introduce the class of square-root stochastic autoregressive volatility models, characterized by the autoregressive structure of the residuals and overcoming important empirical limitations, inference difficulties and limiting symmetry assumptions of the weak GARCH class approach.

Here we focus on a simple, zero mean, ARCH(1) model, for which

$$\begin{aligned} r_t &= \varepsilon_t, \sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 \\ \varepsilon_t &= \sigma_t e_t, e_t \stackrel{i.i.d.}{\sim} N(0, 1), \end{aligned}$$

where  $r_t$  is the asset return,  $\sigma_t^2$  is the  $t - 1$  conditional variance,  $\omega > 0$  and  $0 < \alpha < 1$ . This can be expressed as

$$r_t^2 = \omega + \alpha r_{t-1}^2 + v_t,$$

where  $v_t$ , the volatility surprise, is a martingale difference series. Taking expectations<sup>8</sup> conditional on  $t - 1$  we obtain  $\sigma_t^2$ , the  $t - 1$  conditional variance.

Abstracting from any time variation in the mean, the building block of our model is the copula between  $r_t^2$  and  $r_{t-1}^2$ . For instance, using the Clayton copula we assume that

$$\begin{aligned} C_{r_t^2, r_{t-1}^2}(u, v) &= \text{Clayton}(\theta) \\ \sigma_t^2 &= E(r_t^2 | r_{t-1}^2), \end{aligned}$$

plus some appropriate marginals  $r_t^2, r_{t-1}^2$ . This setup, is fundamentally different from a standard ARCH specification (where the marginal distributions are a byproduct of the conditionally normal distributions) and has several important advantages. It generalizes the ARCH model by keeping the Markov property for the stationary process of squared returns, but allowing for a non additive-linear dependence structure between  $r_t^2$  and  $r_{t-1}^2, \dots, r_{t-k}^2$ , as captured by any copula. Since we only consider here first-order Markov processes the results are directly comparable to an ARCH(1), yet easily extendable to an ARCH(p) setting using generalizations of copula-based characterizations of higher order Markov processes in Ibragimov (2009a). Standard GARCH processes exhibit short memory, regardless of their order. Short memory behavior of GARCH is also obvious from their ARMA characterization, which contradicts the empirical stylized fact of long memory of squared returns. In fact, a number of authors, including Bollerslev and Mikkelsen (1996), Comte and Renault (1998) and Robinson and Zaffaroni (1998), among others, have discussed methods to introduce

<sup>8</sup>These are assumed to exist.



long memory in a GARCH or a stochastic volatility option pricing setting. The approach to volatility modeling based on copulas introduces another potential way to model nonlinear dependence in financial returns. The proposed models, in particular, seamlessly allows for the calculation of descriptive quantities other than the conditional variance, for example higher conditional moments, tail dependence coefficients, quantiles or other descriptive statistics of interest.

We use demeaned Microsoft daily log-returns for the period 1997-2000 which is the data set used by McNeil et. al. (2005) to fit a GARCH(1,1) model<sup>9</sup>. McNeil et al. (2005) report that although the raw returns show no evidence of autocorrelation their absolute values and squares show significant serial correlation until the 19th lag, as illustrated in Figure 5.

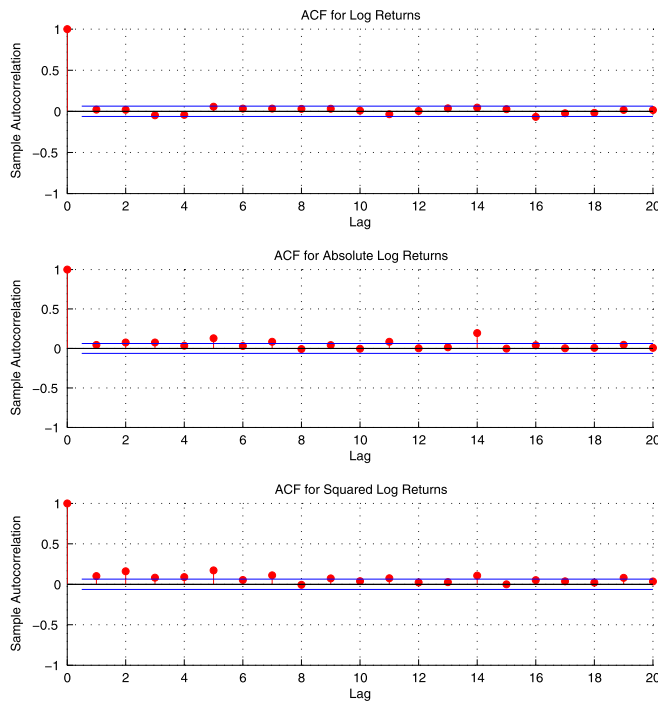


FIG 5. Microsoft log returns 1997-2000; autocorrelation functions of raw, absolute and squared log-returns.

Consistency and asymptotic normality of the semiparametric estimators of the parameters of Clayton copulas and Clayton survival copulas in the models considered below follows from Propositions 4.2 and 4.3 in Chen and Fan (2006), the results in Chen, Wu and Yi (2009) that show that stationary Markov processes based on Clayton and Clayton survival copulas are  $\beta$ -mixing with exponential decay rates (see Theorem 2.1 and Remark 2.2 in Chen, Wu and Yi

<sup>9</sup>We thank Alexander McNeil for providing the data. During this time the stock splits twice and for these dates we use CRSP simple return instead of logreturn.

2009 and the discussion in Section 5 in this paper) and Section 5.3 in Chen and Fan (2006) that verifies the remaining conditions in Propositions 4.2 and 4.3 for Clayton copulas (verification of these conditions for Clayton survival copulas is similar).

The GARCH model assumes that the distribution of  $r_t$  conditional on  $r_{t-1}$

$$r_t|r_{t-1} \sim N(0, \sigma_t^2),$$

and hence that  $r_t$  follows a fat tailed mixture of normals. Here we instead explicitly describe the marginal distributions of the squared returns  $r_t^2, r_{t-1}^2$ , using the Weibull distribution. The Weibull is one of many candidate distributions such as, for example, the lognormal or the gamma laws as was chosen using a likelihood ratio test. (Choice of alternative distributions for  $r_t^2$  does not alter our results).

We consider mainly two copulas as candidates for the dependence of squared returns, the Clayton copula and the Clayton survival copula and fit the non-parametrically derived quantile sample to estimate their dependence parameters. We expect that the Clayton survival copula will be more appropriate for capturing the dependence of squared returns than the Clayton copula because the former allows for upper tail dependence (and hence volatility clustering) while the latter for lower tail dependence. Upper tail dependence is also possible using the Gumbel copula so estimates for the Gumbel copula are also provided as a benchmark case. Figure 6 shows the densities of the three copulas.

To calculate the conditional volatility  $\sigma_t$  in copula models we use

$$\begin{aligned} \sigma_{t+1}^2 &= E(r_{t+1}^2|r_t^2) \\ &= \int_0^\infty r_{t+1}^2 \frac{f(r_{t+1}^2, r_t^2)}{f(r_t^2)} dr_{t+1}^2 \\ &= \int_0^\infty r_{t+1}^2 c(F(r_{t+1}^2), F(r_t^2)) f(r_{t+1}^2) dr_{t+1}^2, \end{aligned} \quad (15)$$

where  $c(u, v)$  is the copula density and  $F(r_{t+1}^2)$  and  $f(r_{t+1}^2)$  are the kernel or parametrically estimated cumulative and probability density functions respectively. Since Weibull densities are used the asymptotic theory of Chen and Fan (2006) for conditional moments cannot be applied but the volatility estimates are still consistent.

**Example 3.** (Independence) If  $C_{r_t^2, r_{t-1}^2}(u, v) = uv$  then

$$\begin{aligned} \sigma_{t+1}^2 &= E(r_{t+1}^2|r_t^2) \\ &= \int_0^\infty r_{t+1}^2 \frac{f(r_{t+1}^2, r_t^2)}{f(r_t^2)} dr_{t+1}^2 \\ &= \int_0^\infty r_{t+1}^2 f(r_{t+1}^2) dr_{t+1}^2 \\ &= \bar{\sigma}^2, \end{aligned}$$

no matter what the value of  $r_t^2$  is.

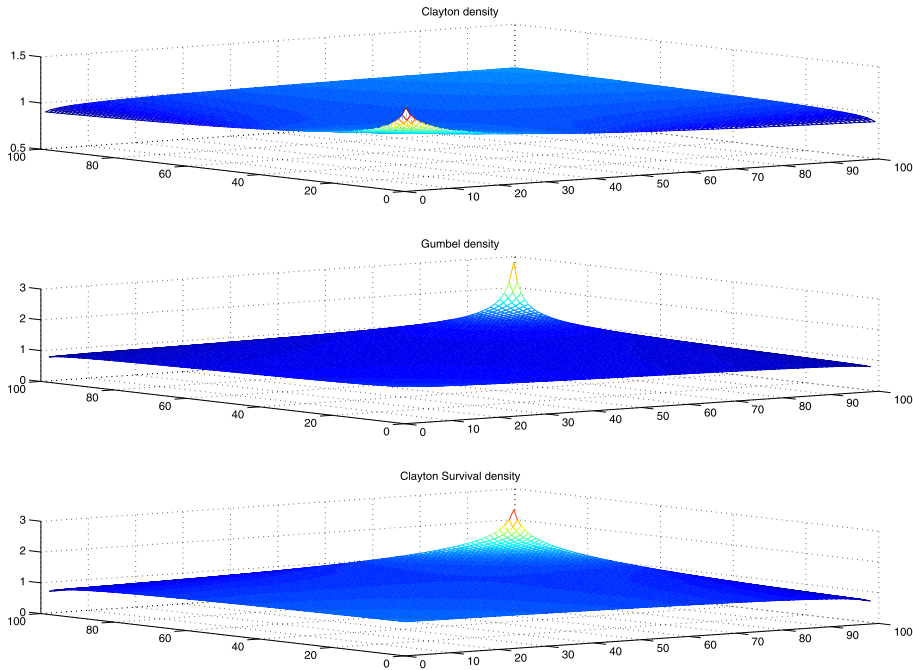


FIG 6. Copula Densities.

Therefore, the intuitive explanation of (15) is that the copula reweighs the values of  $r_{t+1}^2$  according to the realized  $r_t^2$ . We can actually calculate and draw the distribution of weights given  $r_t^2$ , an example of which is shown in Figure 7. We see that a higher theta (i.e. higher dependence) makes the distribution of weights more centered around the  $r_t^2$  value on which we have conditioned

We estimate the conditional volatility using the Clayton, Clayton survival and Gumbel copulas whose parameters have in turn been estimated using maximum likelihood. The results show that, as expected, the Clayton copula performs poorly in capturing the dependence structure of square returns (see Figure 8) since it results in an estimated volatility time series that looks like white noise around the unconditional level. Moreover, as Figure 9 shows, the ACF of the devolatilized data using the Clayton copula shows little improvement over that of the original data. Intuitively, this is because -due to lack of upper tail dependence- high volatility today does not mean high volatility tomorrow for the Clayton copula model.

As a benchmark against which to compare first order Markov copula models, Figures 10 and 11 show the ACF of the devolatilized returns and the estimated volatility time series for an ARCH(1) model. Compared to the Clayton copula model we can see that the simple ARCH(1) creates a much more realistic time series of estimated volatilities and an improvement in the ACF of devolatilized returns.

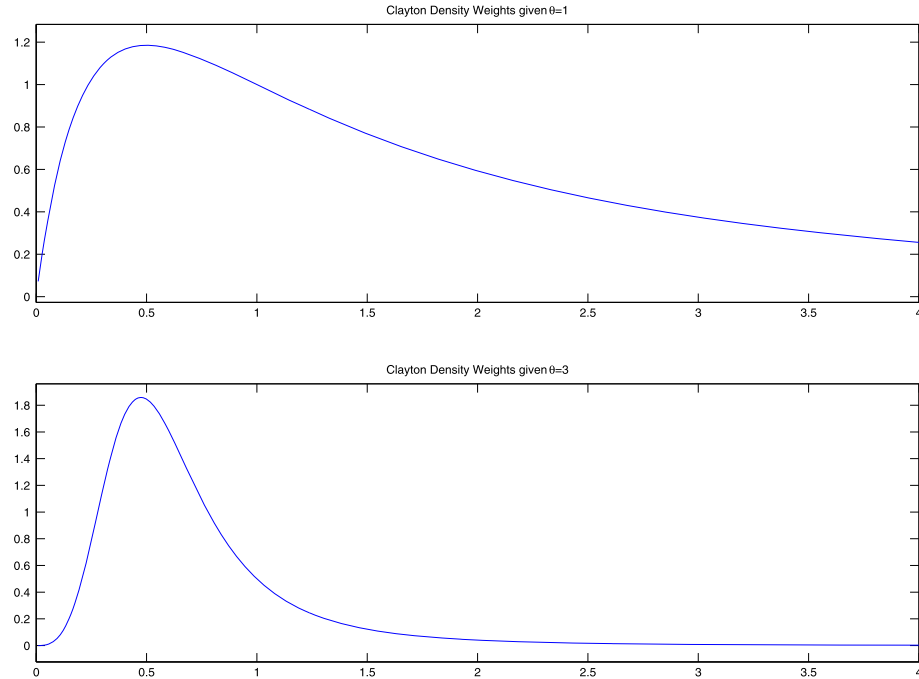


FIG 7. *Conditional Clayton distribution of weights.*

The time series of estimated conditional standard deviations for the Clayton survival copula and the corresponding ACF for devolatilized data are shown in Figures 12 and 13. Comparing with the autocorrelation functions of the original data and of the devolatilized data using ARCH(1) model we see that the Clayton survival copula performs very well. One of the main conclusion is that the Clayton copula almost reduces to zero the first order autocorrelation of the devolatilized data that provides the benchmark for comparison of first order Markov processes (see Figure 13): the magnitude of the first order autocorrelation for the fitted Clayton survival copula-based Markov process in Figure 13 is much smaller than that for the ARCH(1) case in Figure 10. In addition, as is seen in Figure 12, the fitted survival Clayton copula-based Markov process creates a convincing estimated volatility time series.

However, the need for a model with a longer lag is evident by the significant autocorrelations of lags higher than one in the ARCH case and all copula models considered. Similar results follow from estimating the Gumbel copula, shown in Figures 14 and 15. According to the results in this section and the paper, despite these shortcomings, one of the main advantages of the copula-based approaches to volatility modeling is that such copula-based models, in contrast to GARCH-type processes, allow one to separate the analysis of marginal and dependence properties of time series dealt with.

In lack of a formal statistical test it is hard to compare the Clayton survival model to the ARCH(1) and, therefore, developing such a test is a definite direc-

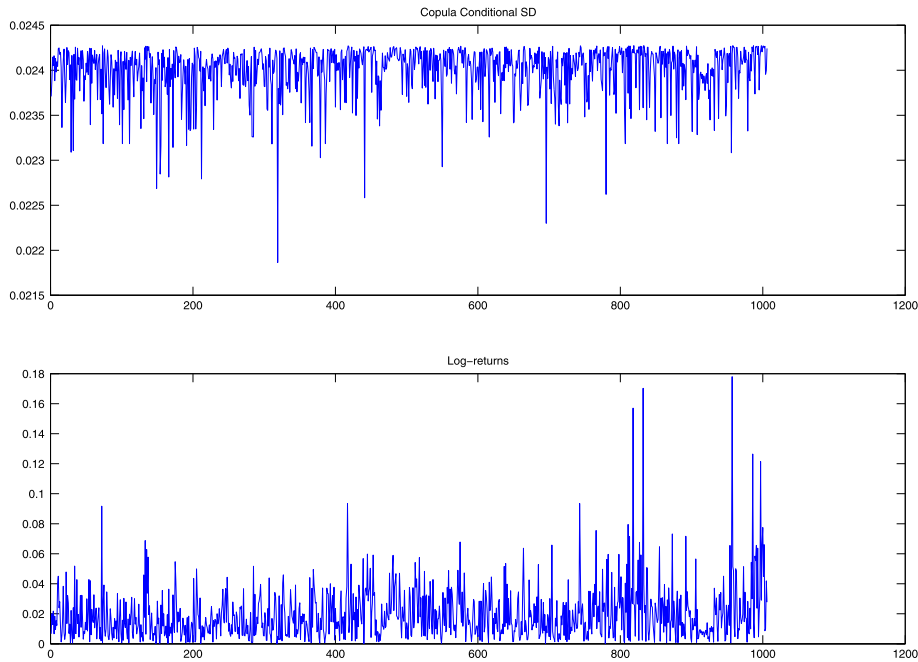


FIG 8. Microsoft log returns 1997-2000; Conditional standard deviation estimated from the fitted Clayton copula-based Markov process

tion for future work. Moreover, a straightforward extension of the copula model is to estimate time varying higher moments, such as for example the conditional kurtosis of the returns in order to model potential “hetero-kurtosis” of asset returns. Since squaring the square returns is an increasing transformation the copula parameter does not change and only the margins need to be estimated again. The estimated conditional kurtosis is shown in Figure 16 and exhibits a structure similar to the estimated conditional volatility.

## 5. Conclusion

Theory and applications of copulas and long range dependence are now well-developed in economics, econometrics, statistics and probability. However, only recently these concepts started to intersect in the literature, with copula-based definitions of long memory introduced by Granger (2003). In this paper, we provide an analysis of relations between different concepts of long memory, including those based on autocorrelations and copula functions. The theoretical results obtained in the paper demonstrate that Gaussian and Eyraud-Farlie-Gumbel-Morgenstern copulas always produce short memory stationary Markov processes. We further show via simulations that, on the other hand, Clayton copula-based stationary Markov processes may exhibit a spurious long memory-

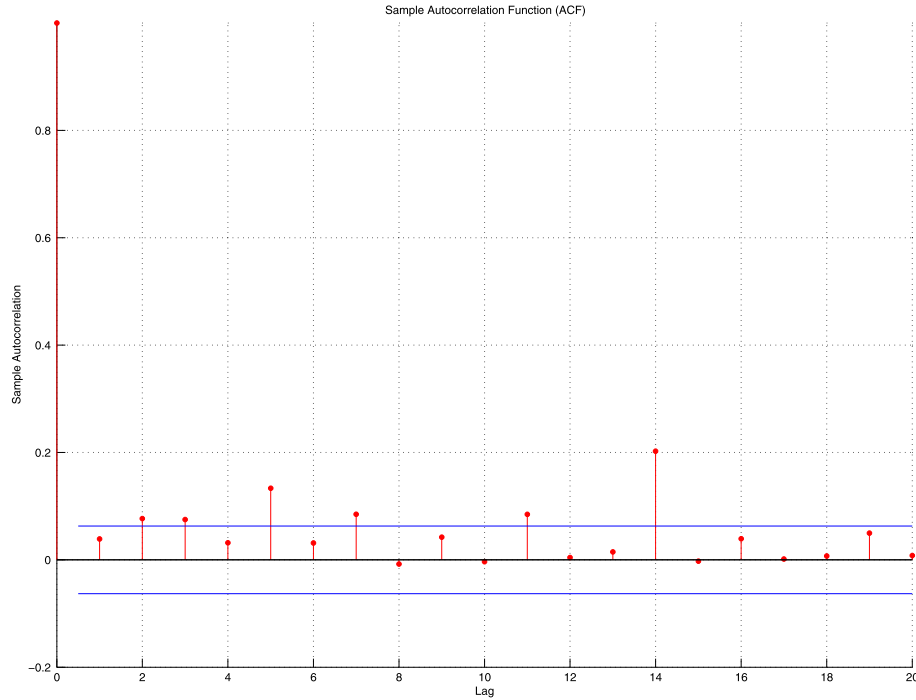


FIG 9. Microsoft log returns 1997-2000; autocorrelation function of devolatilized data using the Clayton copula

like behavior on the level of copulas in finite samples, as indicated by standard inference and estimation methods for long memory time series. We also discuss applications of copula-based Markov processes to volatility modeling and the analysis of nonlinear dependence properties of returns in real financial markets that provide attractive generalizations of GARCH models.

The results presented in the paper overcome several technical and computational difficulties that have so far hindered empirical applications of copula models to time series modeling. These constructions and the methods proposed can be used in the study of several related problems in econometrics, statistics and probability. This includes the analysis of properties of copula-based higher-order Markov processes, applications of various copula families in volatility modeling using copula-based processes and computations involving copula-based time series. In addition, the copula-based modeling may be perspective in empirical applications by providing parsimonious alternatives with a simple low-order Markovian structure and fewer parameters to higher-order GARCH models and their analogues.

As discussed in the introduction, in recent papers that appeared after an earlier version of this work referred to in them was prepared for publication, Beare (2010) and Chen, Wu and Yi (2009) obtained several results on short

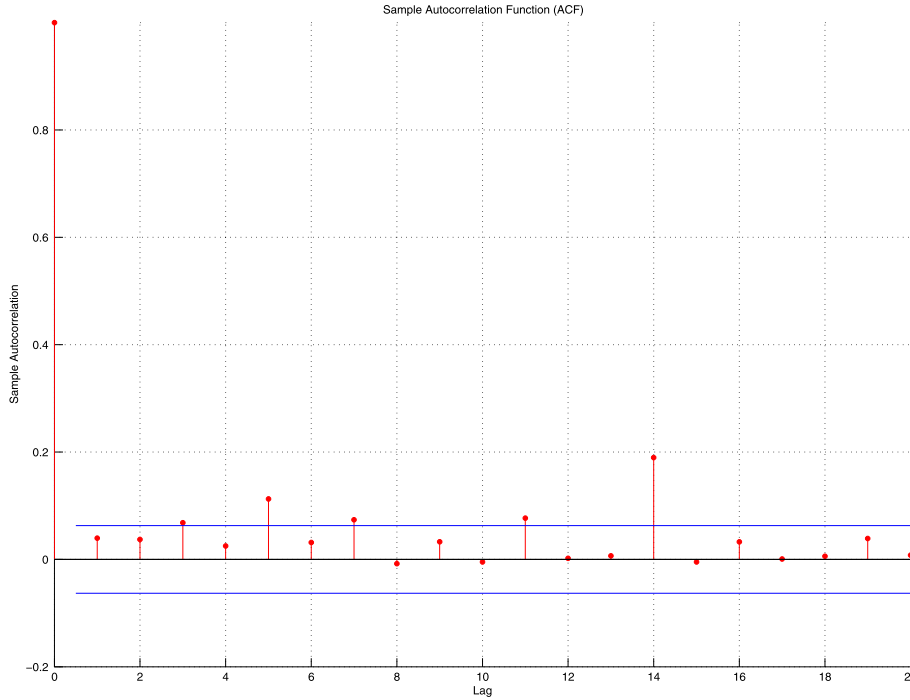


FIG 10. Microsoft log returns 1997-2000; autocorrelation function of devolatilized data using an ARCH(1)

memory and mixing properties of stationary copula-based Markov processes that are closely related to those presented in this paper. For a copula  $C : [0, 1]^2 \rightarrow [0, 1]$  denote by  $\bar{\rho}(C)$  the maximal correlation coefficient of  $C : \bar{\rho}(C) = \sup_{g,h} \text{Corr}(g(U), h(V))$ , where the supremum is taken over all square integrable functions  $g, h : [0, 1] \rightarrow [0, 1]$ , and  $U, V$  are r.v.'s with the joint cdf  $C$  (see Definition 1). The results in Beare (2010) provide generalizations of Proposition 2 that imply that (6) holds for the  $C$ -based stationary Markov processes  $\{X_t\}$  if  $C$  is symmetric and absolutely continuous with square integrable density  $c$  and  $\bar{\rho}(C) < 1$  (as is shown in Beare 2010, square integrability of  $c$  rules out copulas  $C$  that exhibit lower or upper tail dependence, such as Clayton, Clayton survival or Gumbel copulas). Since, as is discussed in Beare (2010),  $\alpha$ - and  $\beta$ -mixing coefficients  $\alpha_X(h)$  and  $\beta_X(h)$  corresponding to  $\{X_t\}$  satisfy  $\kappa_X(h) \leq \alpha_X(h) \leq \beta_X(h) \leq 1/2\phi_X(h)$ , this implies, in particular, that, under the same conditions,  $\{X_t\}$  exhibits  $\alpha$ - and  $\beta$ -mixing with exponential decay rates. As follows from Lancaster (1957),  $\bar{\rho}(C) = \rho$  for the Gaussian copula  $C_\rho^G(u, v)$  (see also Kendall and Stuart, 1973, pp. 599-600). Inequality (16) with  $G(x, y) = g(x)h(y)$  implies that, for any copula  $C : [0, 1]^2 \rightarrow [0, 1]$ ,  $\bar{\rho}(C) \leq \phi(C)$ . Thus, for the EFGM copula  $C_\alpha^{EFGM}(u, v)$  one has, by (47),  $\bar{\rho}(C) \leq 1/3$ . These inequalities for Gaussian and EFGM copulas imply (6) (see also Remark 3.8

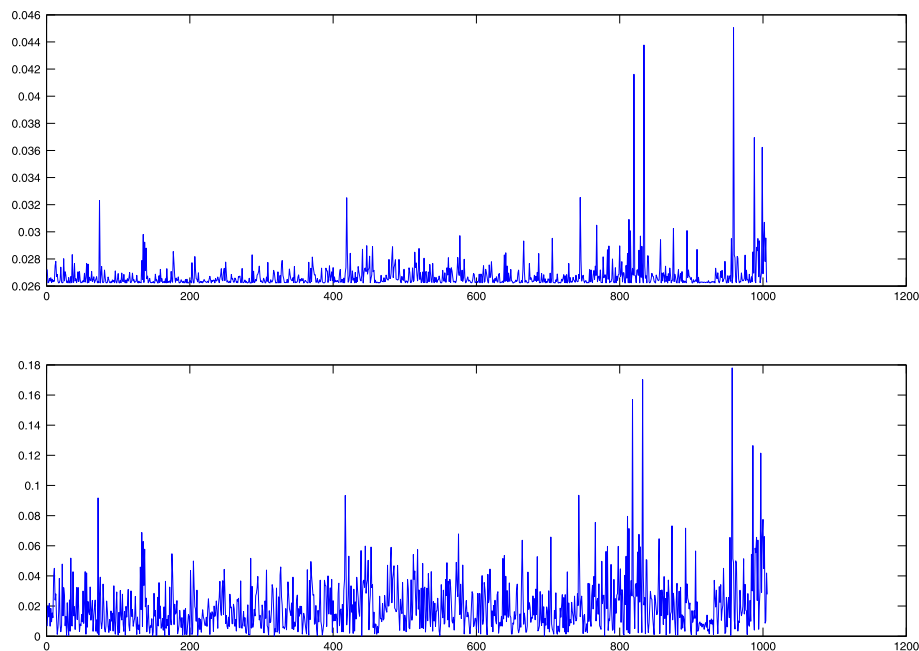


FIG 11. *Microsoft log returns 1997-2000; Conditional standard deviation estimated from an ARCH(1)*

in [Beare 2010](#) on implications for mixing properties of stationary Markov processes generated by Gaussian and EFGM copulas). [Beare \(2010\)](#) further shows that a stationary  $C$ -based Markov process  $\{X_t\}$  with  $\bar{\varrho}(C) < 1$  has exponential decay rate for  $\rho$ -mixing coefficients  $\rho_X(h)$  that implies, together with the inequalities  $\kappa_X(h) \leq \alpha_X(h) \leq 1/4\rho_X(h)$ , the exponential decay in the coefficients  $\kappa_X(h)$  and  $\alpha_X(h)$ . [Beare \(2010\)](#) also provides numerical results that suggest exponential decay in  $\beta$ -mixing coefficients and, thus, also in  $\alpha$ -mixing and  $\kappa(h)$  coefficients of Clayton copula-based stationary Markov processes. [Chen, Wu and Yi \(2009\)](#) obtain theoretical results that show that tail-dependent Clayton, survival Clayton, Gumbel and  $t$ -copulas always generate Markov processes that are geometric ergodic and hence geometric  $\beta$ -mixing and short memory on the level of copulas (in particular,  $\kappa$ -short memory).

From the theoretical results in this paper and in [Chen, Wu and Yi \(2009\)](#) it thus follows that short-memory obtains both for Markov processes generated by both tail-independent (Gaussian or EFGM) copulas as in this work and also those generated by tail-dependent (e.g., Clayton, Gumbel and  $t$ -) copulas. The numerical results in this paper demonstrate that, on the other hand, standard inference and estimation methods for long memory time series may indicate a spurious long memory-like behavior in stationary copula-based Markov processes in finite samples. The conclusions in [Chen, Wu and Yi \(2009\)](#), together with the results presented in this paper, thus further indicate non-robustness



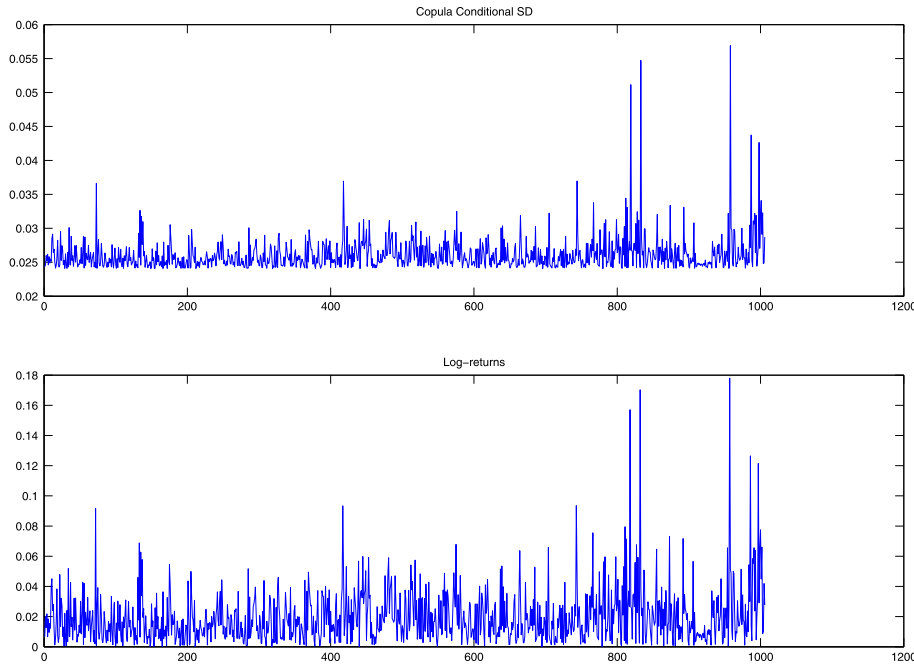


FIG 12. Microsoft log returns 1997-2000; Conditional standard deviation estimated from the fitted Clayton survival copula-based Markov process

of the copula-level analogues of standard procedures for detecting long memory: The inference approaches may lead to a conclusion on persistence on the level of copulas despite the series in consideration being weakly dependent. As discussed throughout the paper, this emphasizes importance of development of robust methods for differentiating short and long memory in copula-based time series. This important problem and other questions discussed in this section are left for further research.

The numerical results in this paper demonstrate that spurious long memory-like behavior (indicated by standard inference methods) may be exhibited by (short-memory) Markov processes generated by tail-dependent (e.g., Clayton) copulas. In subsequent research, it may be of interest to explore whether a similar spurious long memory-like behavior is also indicated by standard inference approaches even in the case of Markov processes generated by tail-independent (e.g., EFGM or Gaussian) copula classes.

Typically, applications of volatility dynamics models in risk management and finance mostly focus on downfalls in financial markets and negative extreme movements of financial returns. Such applications may thus be naturally based on Clayton, survival Clayton and Gumbel copulas that exhibit one (e.g., lower) type of tail dependence. At the same time, motivated by the presence of clusters of both upward and lower extreme movements of financial returns and other time series in economic and financial markets, it would also be of interest to

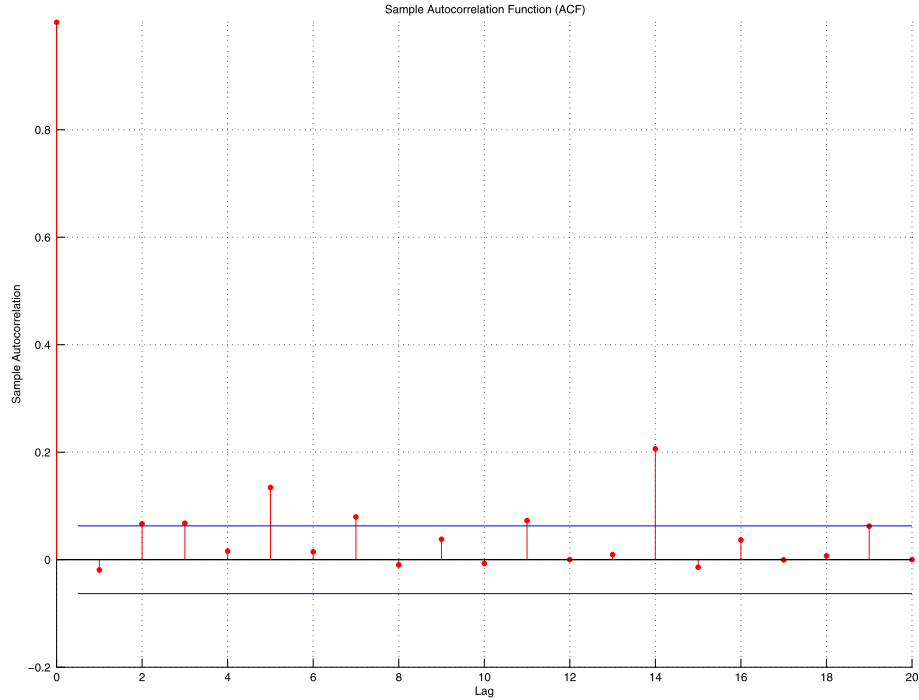


FIG 13. *Microsoft log returns 1997-2000; autocorrelation function of devolatilized data using the Clayton survival copula*

consider extensions of copula-based models for volatility dynamics in the paper to the case of copulas that exhibit both upper- and lower tail dependence such as symmetrized Joe-Clayton copulas dealt with in Patton (2006) or  $t$ -copulas. These and other extensions of the results in the paper are currently under way by the authors and co-authors.

## 6. Proofs

*Proof of Proposition 1.* Relation (2) follows from more general results in Theorem 7.1 in de la Peña, Ibragimov and Sharakhmetov (2006). According to inequality (7.7) in de la Peña, Ibragimov and Sharakhmetov (2006), for any Borel measurable function  $G : \mathbf{R}^2 \rightarrow \mathbf{R}$ , provided the expectations are finite,

$$EG(X, Y) \leq EG(\xi, \eta) + \phi_{X, Y}(EG^2(\xi, \eta))^{1/2}, \quad (16)$$

where, as in Proposition 6,  $\xi, \eta$  are independent copies of  $X, Y : \xi \stackrel{d}{=} X, \eta \stackrel{d}{=} Y$ . Taking in (16)  $G(x, y) = (x - EX)(y - EY)$ , we obtain (5).

The first inequality in (3) follows from Hölder's inequality, and the second inequality in (3) is evident. Using Hölder's inequality, we further have

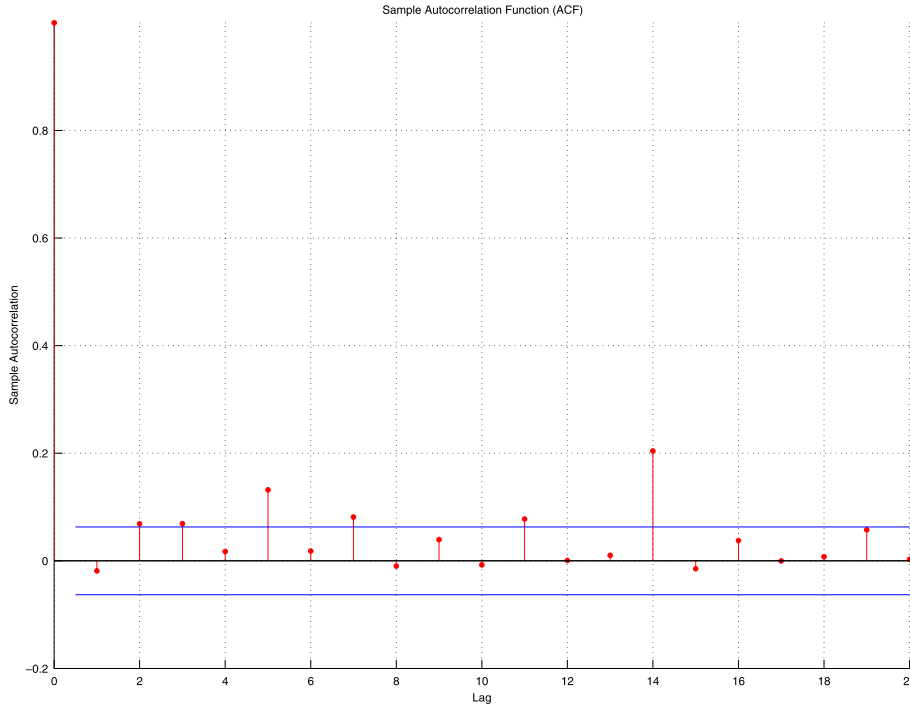


FIG 14. Microsoft log returns 1997-2000; autocorrelation function of devolatilized data using the Gumbel copula

$$\nu_{X,Y} \leq \sup_{u,v \in [0,1]} \int_0^u \int_0^v |c(u,v) - 1| dudv \leq \int_0^1 \int_0^1 |c(u,v) - 1| dudv = E|c(U,V) - 1| \leq [E(c(u,v) - 1)^2]^{1/2} = \phi_{X,Y}.$$

This proves the third inequality in (3). Using Hölder’s inequality again, we get

$$\begin{aligned} 2H_{X,Y} &= E[c^{1/2}(U,V) - 1]^2 \leq (E[c^{1/2}(U,V) - 1]^4)^{1/2} \\ &= \{E[c(U,V) - 1]^2 - 4E[c^{1/2}(U,V)(c^{1/2}(U,V) - 1)^2]\}^{1/2} \\ &\leq \{E[c(U,V) - 1]^2\}^{1/2} = \phi_{X,Y}. \end{aligned}$$

Therefore, (4) holds. □

*Proof of Proposition 2.* Let  $\{X_t\}_{t=1}^\infty$  be a stationary Markov process based on an EFGM copula  $C_\alpha^{EFGM}(u,v)$  with the parameter  $\alpha$ . Since, by Example 4.4 in Darsow, Nguyen and Olsen (1992) the class of EFGM copulas is closed under  $*$ -operator, we then have that the copula  $C_{t,t+h}$  between the r.v.’s  $X_t$  and  $X_{t+h}$  is also an EFGM copula with the parameter  $\alpha^h : C_{t,t+h} = C_{\alpha^h}^{EFGM}$ . By relation (47), this implies  $\phi(h) = O(|\alpha|^h)$ . Thus, by Proposition 1, relations (6) indeed

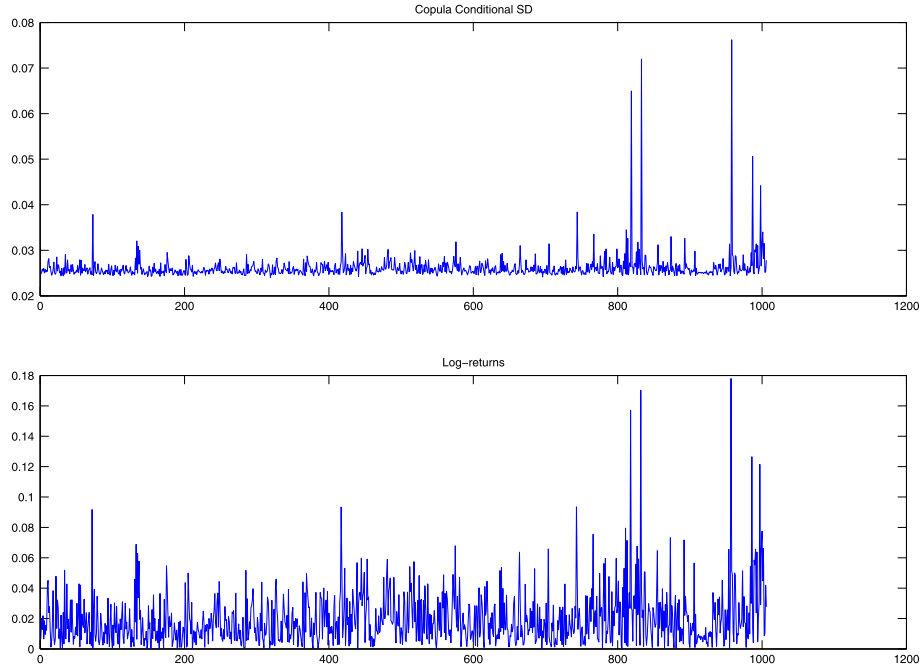


FIG 15. Microsoft log returns 1997-2000; Conditional standard deviation estimated from the fitted Gumbel copula-based Markov process

hold (these relations for  $\lambda(h)$ ,  $\kappa(h)$  and  $\nu(h)$  can also be obtained from Example 2 in Appendix A2).

Suppose now that  $\{X_t\}_{t=1}^{\infty}$  is a stationary Markov process based on a Gaussian copula  $C_{\rho}^G(u, v)$  with the correlation coefficient  $\rho$ . Then (see Example 1 in Chen and Fan 2004 and Section 4 in Ibragimov 2009a) we conclude that  $\{Y_t\}_{t=1}^{\infty}$ , where  $Y_t = \Phi^{-1}[F(X_t)]$ ,  $t \geq 1$ , is a Gaussian process and, thus,

$$Y_t = \rho Y_{t-1} + \epsilon_t = \rho^h Y_{t-h} + \sum_{k=0}^{h-1} \rho^k \epsilon_{t-k}, \quad (17)$$

where  $\epsilon_t$  has a normal distribution:  $\epsilon_t \sim \mathcal{N}(0, 1 - \rho^2)$ . Representation (17) implies that the joint distribution of  $Y_t$  and  $Y_{t-h}$  is normal with the correlation coefficient  $\rho^h$  and, thus, the copula of the r.v.'s  $X_t$  and  $X_{t-h}$  is Gaussian with the correlation coefficient  $\rho^h$ :  $C_{t,t-h} = C_{\rho^h}^G$  (these arguments thus show that Gaussian copulas are closed under  $*$ -operator, similar to the case of EFGM copulas). Using (42), we conclude that  $\phi(h) = O(|\rho|^h)$ . Consequently, by Proposition 1, relations (6) indeed hold for  $\{X_t\}_{t=1}^{\infty}$  (these relations for  $\delta(h)$ ,  $H(h)$ ,  $\kappa(h)$  and  $\nu(h)$  can also be obtained from Example 1 in Appendix A2).  $\square$

*Proof of Proposition 3.* Consider a  $C$ -based stationary Markov process  $\{X_t\}$  and a stationary Markov process  $\{\tilde{X}_t\}$  based on the survival copula  $S(u, v) =$

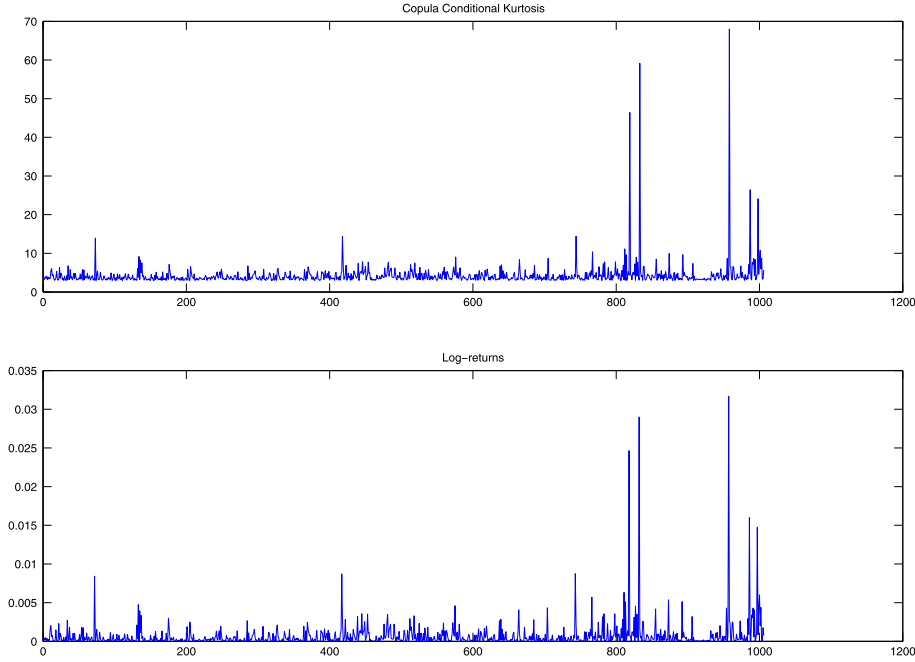


FIG 16. Microsoft log returns 1997-2000; Conditional kurtosis estimated from the fitted Clayton survival copula-based Markov process

$\tilde{C}(u, v)$  of  $C$ . As before, let  $C_{t,t+h}(u, v) = C^h(u, v)$  denote the copula of the r.v.'s  $X_t$  and  $X_{t+h}$ , and, similarly, let  $S_{t,t+h}(u, v) = S^h(u, v)$  denote the copula of the r.v.'s  $X_t$  and  $X_{t+h}$ . It is not difficult to see, using the definition of the  $*$ -operator in Appendix A3, that, for any two copulas  $A, B : [0, 1]^2 \rightarrow [0, 1]$ , the  $*$ -product of the survival copulas  $\tilde{A}$  and  $\tilde{B}$  is the survival copula of  $A * B$ :  $\tilde{A} * \tilde{B}(u, v) = \widetilde{A * B}(u, v)$ . Consequently, the copula  $S_{t,t+h}(u, v) = S^h(u, v)$  is the survival copula of  $C_{t,t+h}(u, v) = C^h(u, v)$ :  $S_{t,t+h}(u, v) = \tilde{C}_{t,t+h}(u, v)$ . Since the densities  $c(u, v)$  and  $\tilde{c}_{t,t+h}(u, v)$  of  $C_{t,t+h}(u, v)$  and  $\tilde{C}_{t,t+h}(u, v)$  satisfy  $\tilde{c}_{t,t+h}(u, v) = c_{t,t+h}(1 - u, 1 - v)$ , from representations (34)-(36) we thus obtain that the processes  $\{X_t\}$  and  $\{\tilde{X}_t\}$  have the same measures of dependence  $\phi^2, \delta, H$ :  $\phi_X^2(h) = \phi_{\tilde{X}}^2(h)$ ,  $\delta_X(h) = \delta_{\tilde{X}}(h)$ ,  $H_X(h) = H_{\tilde{X}}(h)$ .<sup>10</sup> Furthermore, since, evidently,  $\tilde{C}_{t,t+h}(u, v) - uv = C_{t,t+h}(1 - u, 1 - v) - (1 - u)(1 - v)$ , it is easy to obtain from representations (38)-(40), that the measures of dependence  $\kappa, \lambda, \nu$  are also the same for the processes  $\{X_t\}$  and  $\{\tilde{X}_t\}$ :  $\kappa_X(h) = \kappa_{\tilde{X}}(h)$ ,  $\lambda_X(h) = \lambda_{\tilde{X}}(h)$ ,  $\nu_X(h) = \nu_{\tilde{X}}(h)$ . From these conclusions it follows, in particular, that the process  $\{X_t\}$  exhibits short (long) memory in the sense of exponential (resp., hyperbolic) decay of the measures of dependence  $\phi_X^2(h), \delta_X(h), \kappa_X(h)$ ,

<sup>10</sup>From representation (37) it follows that, for all functions  $\psi$  in (32), the measure  $D^\psi$  is the same for the processes  $\{X_t\}$  and  $\{\tilde{X}_t\}$  as well:  $D_X^\psi(h) = D_{\tilde{X}}^\psi(h)$ .

$\lambda_X^2(h)$ ,  $H_X(h)$  or  $\nu_X(h)$  to zero as  $h \rightarrow \infty$  if and only if the same holds for the process  $\{\tilde{X}_t\}$ .  $\square$

## Appendix

### Appendix A1: Long memory and autocorrelation functions

According to one of the commonly employed definitions, a weakly stationary process  $\{X_t\}_{t=-\infty}^{\infty}$  with autocovariance function  $\gamma(h) = Cov(X_t, X_{t+h})$  is said to have long memory if

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| \quad (18)$$

is divergent, and to have short memory otherwise.

Another widely applied definition is based on the hyperbolic decay of autocovariances  $\gamma(h)$  as  $h \rightarrow \infty$ . More precisely, a weakly stationary process  $\{X_t\}$  is said to exhibit long memory or long-range dependence if

$$\gamma(h) \sim \begin{cases} h^\beta l(h), & \text{for } \beta \in (-1, 0), \text{ or} \\ -h^\beta l(h), & \text{for } \beta \in (-2, -1), \end{cases} \quad \text{as } h \rightarrow \infty, \quad (19)$$

where  $l(h)$  is a slowly varying function at infinity:  $l(\lambda h)/l(h) \rightarrow 1$ , as  $h \rightarrow \infty$ , for all  $\lambda > 0$  (see Lo 1991 and Section 2.6 in Lo 1997).

Well-known examples of long memory time series are given by fractional white noise sequences and, more generally, by autoregressive fractionally integrated moving average (ARFIMA) processes. A fractional white noise process  $\{X_t\}$  of order  $d \in \mathbf{R}$  is defined using the difference equation

$$(1 - L)^d X_t = \epsilon_t, \quad (20)$$

where  $L$  is the lag operator and  $\{\epsilon_t\}$  is a white noise process with  $E(\epsilon_t) = 0$ ,  $Var(\epsilon_t) = \sigma_\epsilon^2$  and  $Cov(\epsilon_s, \epsilon_t) = 0$  for  $s \neq t$ . In (20),  $(1 - L)^d$  is defined as

$$(1 - L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k - d)L^k}{\Gamma(-d)\Gamma(k + 1)},$$

where  $\Gamma(\cdot)$  denotes the gamma function:  $\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$ . ARFIMA processes are obtained by replacing the white noise process  $\epsilon_t$  in (20) by stationary and invertible autoregressive moving average (ARMA) processes. Long memory properties of ARFIMA processes are similar to those of fractional white noise.

The process  $\{X_t\}$  in (20) is both stationary and invertible if  $|d| < 0.5$ . For  $0 < d < 0.5$ , the process  $\{X_t\}$  exhibits long memory in the sense of definition (18). For  $-0.5 < d < 0$ , the process is short memory in the sense of definition

(18). In this case,  $\{X_t\}$  is said to exhibit intermediate memory (anti-persistence) or long-range negative dependence.

A fractional white noise process with  $d \in (-0.5, 0.5)$  exhibits long memory in the sense of definition (19) with  $\beta = 2d - 1$  and  $l(h) \equiv A$ , where  $A$  is a positive constant.

The decay of autocovariances in (19) is qualitatively different from and is significantly slower than that of autocovariances of stationary and invertible ARMA processes. Autocovariance functions  $\gamma(h)$  of such processes satisfy

$$\gamma(h) = O[\exp(-Ah)], \quad A > 0, \tag{21}$$

as  $h \rightarrow \infty$ . For instance, in the case of a stationary AR(1) process

$$X_t = \rho X_{t-1} + \epsilon_t, \tag{22}$$

where  $|\rho| < 1$ , and  $\{\epsilon_t\}$  is a white noise process with  $E(\epsilon_t) = 0$ ,  $Var(\epsilon_t) = \sigma_\epsilon^2$  and  $Cov(\epsilon_s, \epsilon_t) = 0$  for  $s \neq t$ , one has  $\gamma(h) = \rho^h \sigma_\epsilon^2 / (1 - \rho^2)$ . Thus, autocorrelations of stationary and invertible ARMA processes exhibit at most exponential decline to zero and, thus, such processes are short memory in the sense of both definitions (18) and (19).

### Appendix A2: Copulas and measures of dependence

We begin with the definition of copulas and formulation of Sklar’s theorem mentioned in the introduction (see Nelsen, 1999; Embrechts, McNeil and Straumann, 2002; McNeil, Frey and Embrechts, 2015).

**Definition 1.** A function  $C : [0, 1]^2 \rightarrow [0, 1]$  is called a (bivariate) copula if it satisfies the following conditions:

1.  $C(u, v)$  is increasing in each component  $u$  and  $v$ .
2.  $C(u, 0) = C(0, v) = 0$  for all  $u, v \in [0, 1]$ .
3.  $C(u, 1) = u$  and  $C(1, v) = v$  for all  $u, v \in [0, 1]$ .
4. For all  $(u_1, u_2), (v_1, v_2) \in [0, 1]^2$  with  $u_1 \leq v_1$  and  $u_2 \leq v_2$ ,  $C(v_1, v_2) + C(u_1, u_2) - C(u_1, v_2) - C(u_2, v_1) \geq 0$ .

Equivalently,  $C$  is a copula if it is a joint cdf of two r.v.’s  $U, V$  each of which is uniformly distributed on  $[0, 1]$ .

**Definition 2.** A copula  $C : [0, 1]^2 \rightarrow [0, 1]$  is called absolutely continuous if, when considered as a joint cdf, it has a joint density given by  $c(u, v) = \partial^2 C(u, v) / \partial u \partial v$ .

**Proposition 4.** (Sklar’s theorem). If  $X, Y$  are r.v.’s defined on a common probability space, with the one-dimensional cdf’s  $F_X(x) = P(X \leq x)$  and  $F_Y(y) = P(Y \leq y)$  and the joint cdf  $F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$ , then there exists a copula  $C_{X,Y}(u, v)$  such that  $F_{X,Y}(x, y) = C_{X,Y}(F_X(x), F_Y(y))$  for all  $x, y \in \mathbf{R}$ . If the univariate marginal cdf’s  $F_X, F_Y$  are both continuous, then the copula is unique and can be obtained via inversion method:

$$C_{X,Y}(u, v) = F_{X,Y}(F_X^{-1}(u), F_Y^{-1}(v)), \tag{23}$$

where  $F_X^{-1}(u) = \inf\{x : F_X(x) \geq u\}$  and  $F_Y^{-1}(v) = \inf\{y : F_Y(y) \geq v\}$ . Otherwise, the copula is uniquely determined at points  $(u, v)$ , where  $u$  is in the range of  $F_X$  and  $v$  is in the range of  $F_Y$ .

As is well-known, copulas are invariant under strictly increasing transformations of r.v.'s with continuous univariate cdf's.

**Proposition 5.** *Let  $X, Y$  be r.v.'s with continuous univariate marginal cdf's  $F_X$  and  $F_Y$  and a copula  $C_{X,Y}$ . If  $g, h : \mathbf{R} \rightarrow \mathbf{R}$  are strictly increasing functions, then the r.v.'s  $g(X)$  and  $h(Y)$  have the same copula  $C_{X,Y}$ .*

The following proposition is an immediate corollary of Sklar's theorem given by Proposition 4.

**Proposition 6.** *Let the r.v.'s  $X, Y$  have the univariate marginal cdf's  $F_X$  and  $F_Y$  and a copula  $C_{X,Y}$  with the density  $c_{X,Y}(u, v)$ . Further, let  $\xi, \eta$  be independent copies of  $X, Y$  (that is, independent r.v.'s with the same one-dimensional distributions as those of  $X, Y$ , respectively:  $\xi \stackrel{d}{=} X$ ,  $\eta \stackrel{d}{=} Y$ ). Then, for any Borel measurable function  $G : \mathbf{R}^2 \rightarrow \mathbf{R}$ , provided the expectations are finite,  $EG(X, Y) = EG(\xi, \eta)c(F_X(\xi), F_Y(\eta))$ .*

de la Peña, Ibragimov and Sharakhmetov (2006) develop  $U$ -statistics-based representations for multivariate joint distributions and copulas that generalize Proposition 6. As a corollary of the results, de la Peña, Ibragimov and Sharakhmetov (2006) derive similar representations for multivariate dependence measures and obtain sharp complete decoupling moment and probability inequalities for dependent r.v.'s in terms of their dependence characteristics (see also the review in Ibragimov, 2009a)

R.v.'s  $X, Y$  with the copula  $C_{X,Y}(u, v)$  are independent if and only if  $C_{X,Y}$  is the product (or independence) copula:

$$C_{X,Y}(u, v) = uv. \quad (24)$$

Any bivariate copula  $C(u, v)$  satisfies the following Fréchet-Hoeffding bounds:

$$\underline{C}(u, v) \leq C(u, v) \leq \overline{C}(u, v), \quad (25)$$

where  $\underline{C}(u, v) = \max(u + v - 1, 0)$  and  $\overline{C}(u, v) = \min(u, v)$ .

Well-known examples of copulas with simple closed forms ("explicit" copulas) are given by Clayton and Gumbel copulas (see Nelsen, 1999; Joe, 1997; McNeil, Frey and Embrechts, 2015):

$$C^{Clayton, \theta}(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}}, \quad \theta > 0, \quad (26)$$

$$C^{Gumbel, \theta}(u, v) = \exp \left[ - \left( (-\ln u)^\theta + (-\ln v)^\theta \right)^{\frac{1}{\theta}} \right], \quad \theta \geq 1. \quad (27)$$

The survival copula  $\tilde{C}(u, v) = u + v - 1 + C(1 - u, 1 - v)$  for the Clayton family also has a simple explicit form

$$\tilde{C}^{Clayton, \theta}(u, v) = u + v - 1 + \left( (1 - u)^{-\theta} + (1 - v)^{-\theta} - 1 \right)^{-\frac{1}{\theta}}, \quad \theta > 0. \quad (28)$$



Taking in (23)  $F_{X,Y}(x, y) = \Phi_\rho(x, y)$  to be the bivariate normal cdf with the linear correlation coefficient  $\rho$ , one obtains the well-known normal, or Gaussian, copula  $C_\rho^G(u, v)$ :

$$C_\rho^G(u, v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{u^2 - 2\rho uv + v^2}{2(1-\rho^2)}\right) dudv, \quad (29)$$

where  $\Phi(x)$  denotes the standard normal univariate cdf. Similarly, let  $q > 0$  and let  $F$  be the bivariate Student- $t$  cdf  $t_{q,\rho}$  with  $q$  degrees of freedom, the linear correlation coefficient  $\rho$  and the location parameter  $0 \in \mathbf{R}^2$ . That is,  $F(x, y) = t_{q,\rho}(x, y)$  is the joint cdf of the random vector  $(\sqrt{q}X/\sqrt{S}, \sqrt{q}Y/\sqrt{S})$  where  $Z = (X, Y) \sim \mathcal{N}(0, \rho)$  has the bivariate normal distribution with the correlation coefficient  $\rho$  and  $S \sim \chi^2(q)$  is a chi-square r.v. with  $q$  degrees of freedom that is independent of  $Z$ . Formula (23) then gives  $t$ -copulas with the correlation coefficient  $\rho$ :

$$C_{q,\rho}^t(u, v) = t_{q,\rho}(t_q^{-1}(u), t_q^{-1}(v)) = \int_{-\infty}^{t_q^{-1}(u)} \int_{-\infty}^{t_q^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \left(1 + \frac{u^2 - 2\rho uv + v^2}{q(1-\rho^2)}\right)^{-(q+2)/2} dudv, \quad (30)$$

where  $t_q(x)$  denotes the cdf of the univariate Student- $t$  distribution with  $q$  degrees of freedom.

A simple class of dependence functions is given by EFGM copulas that have the form

$$C_\alpha^{EFGM}(u, v) = uv(1 + \alpha(1-u)(1-v)), \quad (31)$$

where  $|\alpha| < 1$ .

A number of dependence measures for r.v.'s can be expressed or defined using their copulas. Let  $X, Y$  be r.v.'s with the one-dimensional pdf's  $f_X(x)$  and  $f_Y(y)$ , and the joint pdf  $f_{X,Y}(x, y)$ . Consider the following measures of dependence (see Joe, 1989):

$$\phi_{X,Y}^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{f_{X,Y}^2(x, y)}{f_X(x)f_Y(y)} dx dy - 1$$

(Pearson's  $\phi^2$  coefficient), and

$$\delta_{X,Y} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \ln\left(\frac{f_{X,Y}(x, y)}{f_X(x)f_Y(y)}\right) f_{X,Y}(x, y) dx dy$$

(relative entropy or Kullback-Leibler mutual information).

The Pearson's  $\phi^2$  coefficient and the relative entropy are particular cases of divergence measures

$$D_{X,Y}^\psi = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi\left(\frac{f_{X,Y}(x, y)}{f_X(x)f_Y(y)}\right) f_X(x)f_Y(y) dx dy, \quad (32)$$

where  $\psi$  is a strictly convex function on  $\mathbf{R}$  satisfying  $\psi(1) = 0$ . The multivariate Pearson's  $\phi^2$  corresponds to  $\psi(x) = x^2 - 1$  and the relative entropy is obtained with  $\psi(x) = x \ln x$ .

The choice  $\psi(x) = (1/2)(1 - x^{1/2})^2$  in (32) leads to the Hellinger measure of dependence

$$H_{X,Y} = (1/2) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left( f_{X,Y}^{1/2}(x,y) - f_X^{1/2}(x)f_Y^{1/2}(y) \right)^2 dx dy = 1 - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}^{1/2}(x,y)f_X^{1/2}(x)f_Y^{1/2}(y) dx dy \quad (33)$$

considered in Granger (2003); Granger, Maasoumi and Racine (2004). As discussed in Granger, Maasoumi and Racine (2004), the measure  $H_{X,Y}$  and its scaled versions are rather unique among divergence measures since they satisfy the triangular inequality and are, thus, proper measures of distance (see below).

Let  $C(u, v) = C_{X,Y}(u, v)$  be the copula of  $X, Y$  and let  $U, V$  denote i.i.d. r.v.'s uniformly distributed on  $[0, 1]$ . It is easy to see that the measures  $\phi_{X,Y}^2, \delta_{X,Y}, H_{X,Y}, D_{X,Y}^\psi$  can be written in terms of the copula density  $c(u, v) = \partial C^2(u, v) / \partial u \partial v$  as follows:

$$\phi_{X,Y}^2 = \phi^2(C) = \int_0^1 \int_0^1 c^2(u, v) dudv - 1 = Ec^2(U, V) - 1 = E(c(U, V) - 1)^2, \quad (34)$$

$$\delta_{X,Y} = \delta(C) = \int_0^1 \int_0^1 c(u, v) \ln[c(u, v)] dudv = Ec(U, V) \ln[c(U, V)], \quad (35)$$

$$H_{X,Y} = H(C) = (1/2) \int_0^1 \int_0^1 [c^{1/2}(u, v) - 1]^2 dudv = (1/2)E[c^{1/2}(U, V) - 1]^2 = 1 - Ec^{1/2}(U, V). \quad (36)$$

$$D_{X,Y}^\psi = D^\psi(C) = \int_0^1 \int_0^1 \psi[c(u, v)] dudv = E\psi[c(U, V)]. \quad (37)$$

The following measures of dependence between two r.v.'s  $X$  and  $Y$  with the copula  $C(u, v) = C_{X,Y}(u, v)$  are also of importance (see Schweizer and Wolff, 1981):

$$\kappa_{X,Y} = \kappa(C) = \int_0^1 \int_0^1 |C(u, v) - uv| dudv = E|C(U, V) - UV|, \quad (38)$$

$$\lambda_{X,Y}^2 = \lambda^2(C) = \int_0^1 \int_0^1 (C(u, v) - uv)^2 dudv = E(C(U, V) - UV)^2, \quad (39)$$

$$\nu_{X,Y} = \nu(C) = \sup_{u,v \in [0,1]} |C(u, v) - uv|. \quad (40)$$

Using concavity of the function  $f(x) = \sqrt{x}$  it is easy to see that the following triangular inequality holds for the Hellinger measure of dependence in the case

of copula mixtures. Suppose that the copula  $C(u, v) = C_{X,Y}(u, v)$  of the r.v.'s  $X, Y$  is given by

$$C(u, v) = w_1 C_1(u, v) + w_2 C_2(u, v), \quad w_1, w_2 \in [0, 1]. \quad (41)$$

Then  $H(C) \leq w_1 H(C_1) + w_2 H(C_2)$ . It is not difficult to see, using Minkowski inequality (see [Marshall and Olkin 1979](#), Section 16.D) in the case of  $\lambda_{X,Y}$ , that the measures of dependence  $\kappa_{X,Y}$ ,  $\lambda_{X,Y}$  and  $\nu_{X,Y}$  satisfy similar triangular inequalities as well: for the above copula mixtures,  $\kappa(C) \leq w_1 \kappa(C_1) + w_2 \kappa(C_2)$ ,  $\lambda(C) \leq w_1 \lambda(C_1) + w_2 \lambda(C_2)$  and  $\nu(C) \leq w_1 \nu(C_1) + w_2 \nu(C_2)$ .

**Example 1.** For a Gaussian copula  $C_\rho^G(u, v)$  with the correlation coefficient  $\rho$  in (29), one has

$$\phi^2(C_\rho^G) = \rho^2 / (1 - \rho^2), \quad (42)$$

$$\delta(C_\rho^G) = -0.5 \ln(1 - \rho^2) \quad (43)$$

(see [Joe, 1989](#)),

$$H(C_\rho^G) = 1 - \frac{(1 - \rho^2)^{5/4}}{(1 - \rho^2/2)^{3/2}} \quad (44)$$

(see [Granger, Maasoumi and Racine, 2004](#)),

$$\kappa(C_\rho^G) = \frac{1}{2\pi} \arcsin(|\rho|/2), \quad (45)$$

$$\nu(C_\rho^G) = \frac{1}{2\pi} \arcsin(|\rho|) \quad (46)$$

(see Section 4 in [Schweizer and Wolff, 1981](#)).

**Example 2.** It is easy to see that the following relations hold for the EFGM copula  $C_\alpha^{EFGM}(u, v)$  with the parameter  $\alpha$  in (31):

$$\phi^2(C_\alpha^{EFGM}) = \alpha^2 / 9, \quad (47)$$

$$\lambda(C_\alpha^{EFGM}) = |\alpha| / 30, \quad (48)$$

$$\kappa(C_\alpha^{EFGM}) = |\alpha| / 36, \quad (49)$$

$$\nu(C_\alpha^{EFGM}) = |\alpha| / 16. \quad (50)$$

### Appendix A3: Markov processes and copulas

[Darsow, Nguyen and Olsen \(1992\)](#) obtain the following necessary and sufficient conditions for a time series process based on bivariate copulas to be first-order Markov. For copulas  $A, B : [0, 1]^2 \rightarrow [0, 1]$ , set

$$(A * B)(x, y) = \int_0^1 \frac{\partial A(x, t)}{\partial t} \cdot \frac{\partial B(t, y)}{\partial t} dt.$$

Further, for copulas  $A : [0, 1]^m \rightarrow [0, 1]$  and  $B : [0, 1] \rightarrow [0, 1]$ , define their  $\star$ -product  $A \star B : [0, 1]^{m+n-1} \rightarrow [0, 1]$  via

$$A \star B(x_1, \dots, x_{m+n-1}) = \int_0^{x_m} \frac{\partial A(x_1, \dots, x_{m-1}, \xi)}{\partial \xi} \cdot \frac{\partial B(\xi, x_{m+1}, \dots, x_{m+n-1})}{\partial \xi} d\xi.$$

As shown in Darsow, Nguyen and Olsen (1992), the operators  $*$  and  $\star$  on the class of copulas are distributive over convex combinations, associative and continuous in each place, but not jointly continuous. For a copula  $C$  denote by  $C^s$  the  $s$ -fold product  $\star^k$  of  $C$  with itself.

Let  $C_{t_1, \dots, t_k}$ ,  $t_i \in T$ ,  $i = 1, \dots, k$ ,  $t_1 < \dots < t_k$ , stand for copulas corresponding to the joint distribution of the r.v.'s  $X_{t_1}, \dots, X_{t_k}$  in the process  $\{X_t\}_{t=-\infty}^{\infty}$  in consideration. Darsow, Nguyen and Olsen (1992) prove that the transition probabilities  $P(s, x, t, A) = P(X_t \in A | X_s = x)$  of a real-valued stochastic process  $\{X_t\}_{t=-\infty}^{\infty}$ , satisfy the Chapman-Kolmogorov equations

$$P(s, x, t, A) = \int_{-\infty}^{\infty} P(u, \xi, t, A) P(s, x, u, d\xi) \quad (51)$$

for all Borel sets  $A$ , all  $s < r < t$  and for almost all  $x \in \mathbf{R}$  if and only if the copulas corresponding to the bivariate distributions of  $X_t$  are such that

$$C_{st} = C_{sr} * C_{rt} \quad (52)$$

for all  $s < r < t$ . Darsow, Nguyen and Olsen (1992) also show that a real-valued stochastic process  $\{X_t\}_{t=-\infty}^{\infty}$  is a stationary first-order Markov process if and only if the copulas corresponding to the finite-dimensional distributions of  $\{X_t\}$  satisfy the conditions

$$C_{1, \dots, n}(u_1, \dots, u_n) = C \star^k C \star^k \dots \star^k C(u_1, \dots, u_n) = C^{n-k+1}(u_1, \dots, u_n) \quad (53)$$

for all  $n \geq 2$ . Ibragimov (2009a) provides copula-based characterizations of higher-order Markov processes. The results are applied to establish necessary and sufficient conditions for Markov processes of a given order to exhibit  $m$ -dependence,  $r$ -independence or conditional symmetry. Ibragimov (2009a) also presents a study of applicability and limitations of different copula families in constructing higher-order Markov processes with the above dependence properties.

## References

- Baillie, R. T. (1996), 'Long memory processes and fractional integration in econometrics', *Journal of Econometrics* **73**, 5–59. [MR1410000](#)
- Beare, B. (2010), 'Copulas and temporal dependence', *Econometrica* **78**, 395–410. [MR2642867](#)
- Bollerslev, T. and Mikkelsen, H. O. (1996), 'Modeling and pricing long memory in stock market volatility', *Journal of Econometrics* **73**, 151–184.

- Cambanis, S. (1991), On Eyrraud-Farlie-Gumbel-Morgenstern random processes, in ‘Advances in probability distributions with given marginals (Rome, 1990)’, Vol. 67 of *Mathematics and its Applications*, Kluwer Academic Publishing, Dordrecht, pp. 207–222. [MR1215953](#)
- Campbell, J. Y., Lo, A. W. and MacKinlay, A. (1997), *The econometrics of financial markets*, Princeton University Press, Princeton.
- Chen, X. and Fan, Y. (2004), ‘Evaluating density forecasts via copula approach’, *Finance Research Letters* **1**, 74–84.
- Chen, X. and Fan, Y. (2006), ‘Estimation of copula-based semiparametric time series models’, *Journal of Econometrics* **130**, 307–335. [MR2211797](#)
- Chen, X., Wu, W. B. and Yi, Y. (2009), ‘Efficient estimation of copula-based semiparametric Markov models’, *Annals of Statistics* **37**, 4214–4253. [MR2572458](#)
- Cherubini, U., Luciano, E. and Vecchiato, W., eds (2004), *Copula methods in finance*, Wiley, Chichester. [MR2250804](#)
- Comte, F. and Renault, E. (1998), ‘Long memory in continuous-time stochastic volatility models’, *Mathematical Finance* **8**, 291–323. [MR1645101](#)
- Cont, R. (2001), ‘Empirical properties of asset returns: Stylized facts and statistical issues’, *Quantitative Finance* **1**, 223–236.
- Cuadras, C. (2009), ‘Constructing copula functions with weighted geometric means’, *Journal of Statistical Planning and Inference* **139**, 3766–3772. [MR2553761](#)
- Cuadras, C. and Diaz, W. (2012), ‘Another generalization of the bivariate fgm distribution with two-dimensional extensions’, *Acta et Commentationes Universitatis Tartuensis de Mathematica* **16**, 3–12. [MR2988796](#)
- Darsow, W. F., Nguyen, B. and Olsen, E. T. (1992), ‘Copulas and Markov processes’, *Illinois Journal of Mathematics* **36**, 600–642. [MR1215798](#)
- Davis, R. A. and Mikosch, T. (1998), ‘The sample autocorrelations of heavy-tailed processes with applications to ARCH’, *Annals of Statistics* **26**, 2049–2080. [MR1673289](#)
- de la Peña, V. H., Ibragimov, R. and Sharakhmetov, S. (2006), Characterizations of joint distributions, copulas, information, dependence and decoupling, with applications to time series, in J. Rojo, ed., ‘2nd Erich L. Lehmann Symposium – Optimality, IMS Lecture Notes – Monograph Series’, Vol. 49, Institute of Mathematical Statistics, Beachwood, Ohio, USA, pp. 183–209. Available at <http://dx.doi.org/10.1214/074921706000000455>.
- Doukhan, P., Fermanian, J.-D. and Lang, G. (2004), ‘Copulas of a vector-valued stationary weakly dependent process’, *Working paper, CREST*.
- Doukhan, P., Oppenheim, G. and Taqqu, M. S., eds (2003), *Theory and applications of long-range dependence*, Birkhäuser, Boston. [MR1957509](#)
- Drost, F. C. and Nijman, T. E. (1993), ‘Temporal aggregation of garch processes’, *Econometrica* **61**, 909–927. [MR1231681](#)
- Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997), *Modelling extremal events for insurance and finance*, Springer, New York. [MR1458613](#)
- Embrechts, P., McNeil, A. J. and Straumann, D. (2002), Correlation and dependence in risk management: properties and pitfalls, in ‘Risk management:

- value at risk and beyond (Cambridge, 1998)', Cambridge University Press, Cambridge, pp. 176–223. [MR1892190](#)
- Fermanian, J.-D., Radulović, D. and Wegkamp, M. (2004), 'Weak convergence of empirical copula processes', *Bernoulli* **10**, 847–860. [MR2093613](#)
- Fruhworth-Schnatter, S. (2006), *Finite mixture models and Markov switching models*, Wiley Series in Probability and Statistics, Springer, New York. [MR2265601](#)
- Granger, C. W. J. (2003), 'Time series concepts for conditional distributions', *Oxford Bulletin of Economics and Statistics* **65**, 689–701 Suppl. S.
- Granger, C. W. J., Teräsvirta, T. and Patton, A. J. (2006), 'Common factors in conditional distributions for bivariate time series', *Journal of Econometrics* **132**, 43–57. [MR2271390](#)
- Granger, C. W., Maasoumi, E. and Racine, J. (2004), 'A dependence metric for possibly nonlinear processes', *Journal of Time Series Analysis* **25**, 649–669. [MR2086354](#)
- Hosking, J. R. M. (1996), 'Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series', *Journal of Econometrics* **73**, 261–284. [MR1410007](#)
- Hu, L. (2006), 'Dependence patterns across financial markets: methods and evidence', *Applied Financial Economics* **16**, 717–729.
- Ibragimov, M., Ibragimov, R. and Walden, J. (2015), *Heavy-Tailed Distributions and Robustness in Economics and Finance*, Vol. 214 of *Lecture Notes in Statistics*, Springer, New York. [MR3331281](#)
- Ibragimov, R. (2009a), 'Copula-based characterizations for higher-order Markov processes', *Econometric Theory* **25**, 819–846. [MR2507535](#)
- Ibragimov, R. (2009b), Heavy tailed densities, in S. N. Durlauf and L. E. Blume, eds, 'The New Palgrave Dictionary of Economics Online', Palgrave Macmillan, New York.
- Ibragimov, R., Jaffee, D. and Walden, J. (2009), 'Nondiversification traps in catastrophe insurance markets', *Review of Financial Studies* **22**, 959–993.
- Ibragimov, R. and Lentzas, G. (2008), 'Copulas and long memory', *Harvard University Research Discussion Paper 2160*.
- Ibragimov, R. and Prokhorov, A. (2016), 'Heavy tails and copulas: Limits of diversification revisited', *Economics Letters* **149**, 102–107. [MR3576394](#)
- Ibragimov, R. and Prokhorov, A. (2017), *Heavy Tails and Copulas: Topics in Dependence Modelling in Economics and Finance*, World Scientific, New York.
- Joe, H. (1989), 'Relative entropy measures of multivariate dependence', *Journal of the American Statistical Association* **84**(405), 157–164. [MR0999674](#)
- Joe, H. (1997), *Multivariate models and dependence concepts*, Vol. 73 of *Monographs on Statistics and Applied Probability*, Chapman & Hall, London. [MR1462613](#)
- Kendall, M. G. and Stuart, A. (1973), *The advanced theory of statistics, Vol. 2*, third edn, Hafner Publishing Co., New York. [MR0474561](#)
- Kimeldorf, G. and Sampson, A. (1975), 'Uniform representations of bivariate distributions', *Communications in Statistics* **4**, 617–627. [MR0397989](#)

- Lancaster, H. O. (1957), ‘Some properties of the bivariate normal distribution considered in the form of contingency table’, *Biometrika* **44**, 289–292.
- Lin, G. (1987), ‘Relationships between two extensions of farlie–gumbel–morgenstern distribution’, *Annals of the Institute of Statistical Mathematics* **39**, 129–140. [MR0886511](#)
- Lo, A. W. (1991), ‘Long-term memory in stock market prices’, *Econometrica* **59**, 1279–1313.
- Loretan, M. and Phillips, P. C. B. (1994), ‘Testing the covariance stationarity of heavy-tailed time series’, *Journal of Empirical Finance* **1**, 211–248. [MR2687497](#)
- Lowin, J. (2007), *The Fourier copula: Theory and applications*, Senior thesis, Harvard University.
- Marshall, A. W. and Olkin, I. (1979), *Inequalities: theory of majorization and its applications*, Vol. 143 of *Mathematics in Science and Engineering*, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York. [MR0552278](#)
- McLachlan, G. J. and Peel, D. (2000), *Finite mixture models*, Wiley Series in Probability and Statistics, John Wiley & Sons, New York. [MR1789474](#)
- McNeil, A. J., Frey, R. and Embrechts, P. (2015), *Quantitative risk management: Concepts, Techniques and Tools*, Princeton Series in Finance, Princeton University Press, Princeton, NJ. [MR3445371](#)
- Meddahi, N. and Renault, E. (2004), ‘Temporal aggregation of volatility models’, *Journal of Econometrics* **119**, 355–379. [MR2057104](#)
- Mikosch, T. and Stărică, C. (2000), ‘Limit theory for the sample autocorrelations and extremes of a GARCH (1, 1) process’, *Annals of Statistics* **28**, 1427–1451. [MR1805791](#)
- Nelsen, R. B. (1999), *An introduction to copulas*, Vol. 139 of *Lecture Notes in Statistics*, Springer-Verlag, New York. [MR1653203](#)
- Nelsen, R. B., Quesada-Molina, J. and Rodriguez-Lallena, J. (1997), ‘Bivariate copulas with cubic sections’, *Nonparametric Statistics* **7**, 205–220. [MR1443354](#)
- Nze, P. A. and Doukhan, P. (2004), ‘Weak dependence: models and applications to econometrics’, *Econometric Theory* **20**, 995–1045. [MR2101950](#)
- Patton, A. (2006), ‘Modelling asymmetric exchange rate dependence’, *International Economic Review* **47**, 527–556. [MR2216591](#)
- Robinson, P. M. and Zaffaroni, P. (1998), ‘Nonlinear time series with long memory: a model for stochastic volatility’, *Journal of Statistical Planning and Inference* **68**, 359–371. [MR1629599](#)
- Schweizer, B. and Wolff, E. F. (1981), ‘On nonparametric measures of dependence for random variables’, *Annals of Statistics* **9**, 879–885. [MR0619291](#)
- Sharakhmetov, S. and Ibragimov, R. (2002), ‘A characterization of joint distribution of two-valued random variables and its applications’, *Journal of Multivariate Analysis* **83**, 389–408. [MR1945960](#)