DURATION TIME SERIES MODELS

WITH PROPORTIONAL HAZARD

P. GAGLIARDINI[*], C. GOURIEROUX[†]

First version: May 2002

This version: March 2007

[*]University of Lugano and Swiss Finance Institute.

[†]CREST and University of Toronto.

Corresponding author: Patrick Gagliardini, Faculty of Economics, University of Lugano, Via Buffi 13, CH-6900 Lugano, Switzerland. Tel.: +41 58 666 4660. Fax: +41 58 666 4647. Email: gagliarp@lu.unisi.ch.

Duration Time Series Models with Proportional Hazard

Abstract

The analysis of liquidity in financial markets is generally performed by means of the dynamics of the observed intertrade durations (possibly weighted by price or volume). Various dynamic models for duration data have been considered in the literature, such as the ACD (Autoregressive Conditional Duration) model. These models are often excessively constrained, introducing for example a deterministic link between conditional expectation and variance in the case of the ACD model. Moreover, the stationarity properties and the patterns of the stationary distributions are often unknown. The aim of this paper is to solve these difficulties by considering a duration time series satisfying the proportional hazard property. We describe in detail this class of dynamic models, discuss its various representations, and provide the ergodicity conditions. The proportional hazard copula can be specified either parametrically, or nonparametrically. We discuss estimation methods in both contexts, and explain why they are efficient, that is, why they reach the parametric (respectively, nonparametric) efficiency bound.

Keywords: Duration, Copula, ACD Model, Nonparametric Estimation, Proportional Hazard, Nonparametric Efficiency.

JEL classification: C14, C22, C41

1 Introduction

Series of durations between consecutive trades of a given asset have been the object of a considerable body of research in financial econometrics (e.g. Engle, 2000; Gourieroux and Jasiak, 2001a). Interest in this topic, supported by the increasing availability of (ultra-)high-frequency data, has several financial motivations. In addition to its links with microstructure theory and with the literature on stochastic time deformation[1], the dynamics of intertrade durations is an important issue for the management of liquidity risk. Indeed, durations between consecutive trades are a natural measure of market liquidity and their variability is related to liquidity risk (risk on time). The aim of this paper is to introduce a new class of dynamic models for intertrade durations suitable for the analysis of liquidity risk.

Empirical investigations of series of intertrade durations report several stylized facts which must be taken into account in the specification of econometric models[2]. The most significant are the following ones: a positive serial dependence, in the form of positive autocorrelations and tendency of extremely large durations to come in clusters (clustering effects); persistency, with autocorrelations decreasing slowly with horizon, and possibly featuring long memory; strong nonlinearities in the dynamics, observed from the analysis of nonlinear autocorrelograms; path-dependent (under-)overdispersion in the conditional distribution; significant departures from unconditional exponential distribution, with negative duration dependence and fat tails. In addition to consistency with these stylized facts, flexible specifications for conditional mean and conditional variance are desirable for the management of liquidity risk. If extreme liquidity risks have to be taken into account, the first conditional moments may not be sufficient and measures based on the entire conditional distribution may be more appropriate. This is the case of the Time-at-Risk (TaR), which is the minimal time without a trade that may occur with a given probability (see Ghysels *et al.*, 2004). These measures require flexible specifications for the entire conditional distribution of the duration process.

The Autoregressive Conditional Duration (ACD) model introduced by Engle and Russell (1998) is the most successful dynamic model for intertrade durations. It is based on an accelerated hazard specification, where the conditional mean follows a deterministic autoregression[3]. The ACD is able to replicate various stylized effects observed in the data. However, as pointed out in Ghysels *et al.* (2004), a limitation of this specification is the quite restrictive set of assumptions on the conditional distribution of the duration process. The dynamics of conditional moments of any order and of liquidity risk measures like TaR_t are all determined by the dynamics of the conditional mean. These restrictions are not supported by empirical evidence, since they imply path-independent conditional dispersion and, more importantly, they are not desirable for management of liquidity risk. In order to overcome these difficulties, alternative specifications to accelerated hazard may be considered. For example, Ghysels *et al.* (2004) propose the Stochastic Volatility

Duration (SVD) model, where conditional mean and conditional variance follow independent dynamics as a result of the introduction of two underlying factors. For expository purposes and the sake of a thorough analytic study, we restrict our analysis to Markov duration processes.

In this paper, we introduce a Markov process for intertrade durations that is based on a proportional hazard specification. The conditional hazard function for duration X_t given the past durations $\underline{X_{t-1}}$, is the product of a baseline hazard function λ_0 times a positive function a of the lagged duration[4]:

$$\lambda\left(x \mid \underline{X_{t-1}}\right) = a\left(X_{t-1}\right)\lambda_0(x), \quad x \geq 0,$$

where a and λ_0 are unconstrained, except for identifiability conditions. This specification improves on the accelerated hazard specification of the ACD(0,1) model in two directions. First, it provides a flexible specification for the conditional distribution of the duration process, without restrictive assumptions on the joint dynamics of conditional moments. Since the past information scales the conditional hazard function instead of the duration variable itself, the effect of the lagged duration on the conditional moments and on the conditional distribution is not tied down by the specification of the conditional mean. On the contrary, the effect of the conditioning variable is determined by the interplay of the two functional parameters a and λ_0. Secondly, another advantage of the proportional hazard specification is to separate marginal and serial dependence characteristics of the process. More precisely, we show that the bivariate copula between two consecutive durations X_t and X_{t-1} is fully characterized by a univariate functional parameter A (say) on $[0,1]$. The copula is defined as the c.d.f. of variables X_t and X_{t-1} preliminarily transformed to get uniform marginal distributions on the interval $[0,1]$. The copula summarizes the serial dependence between X_t and X_{t-1} that is invariant to monotonic transformations. This result implies that the proportional hazard duration model can be parameterized in terms of the marginal distribution of the process and functional parameter A, which characterizes serial dependence. The marginal properties of the process are fixed by choosing the marginal distribution. By focusing on parameter A, the nonlinear serial dependence properties of the process are controlled, leaving its marginal distribution unaltered. We discuss how the shape of function A influences the pattern and the strength of serial dependence in the process, both in the whole distribution and in the tails, by introducing appropriate (functional) concepts and measures of dependence. In particular, the duration process features positive dependence when functional parameter A is decreasing, whereas its negative elasticity $-d\log A/dv$ is an ordinal functional measure of serial dependence. Moreover, the behaviour of A at $v=1$ characterizes dependence in the tails of the process and creates clusterings of extremely large durations. We provide sufficient conditions on the behaviour of functional dependence parameter A in a neighbourhood of the boundary points $v=0$ and $v=1$ ensuring ergodicity and mixing properties of the process [5].

The paper is organized as follows. In Section 2 we define the proportional hazard Markov process. In Section 3 the serial dependence properties of the Markov process with proportional hazard are discussed, and in Section 4 sufficient conditions for geometric ergodicity and mixing are provided. Section 5 reports several examples of Markov processes with proportional hazard. Section 6 considers statistical inference. Finally, Section 7 concludes. The proofs are gathered in appendices.

2 Markov process with proportional hazard

In this section we introduce the stationary Markov process with proportional hazard.

2.1 A Markov process of durations

Let X_t, $t \in \mathbb{N}$, denote the sequence of consecutive (intertrade) durations. We assume that X_t, $t \in \mathbb{N}$, is a stationary Markov process of order one and features proportional hazard. In the sequel, it is called Markov process with proportional hazard. The conditional hazard function is the product of a baseline hazard function λ_0 times a positive function a of the lagged duration:

$$\lambda\left(x \mid \underline{X_{t-1}}\right) \equiv \lim_{h \to 0} \frac{P\left[X_t \leq x+h \mid X_t \geq x, \underline{X_{t-1}}\right]}{h} = a\left(X_{t-1}\right)\lambda_0(x), \;\; x \geq 0.$$

The effect of the lagged duration is a proportional shift of the conditional hazard function. The transition density of the process is characterized by the conditional survivor function:

$$P\left[X_t \geq x_t \mid X_{t-1} = x_{t-1}\right] = \exp\left[-a(x_{t-1})\Lambda_0(x_t)\right], \, t \in \mathbb{N}, \tag{1}$$

where Λ_0 is the baseline cumulated hazard corresponding to λ_0, defined by: $\Lambda_0(x) = \int_0^x \lambda_0(u)du$, $x \geq 0$. Thus, the distribution of the process is characterized by two functional parameters, namely the baseline cumulated hazard Λ_0, which is (up to a multiplicative constant) the cumulated hazard of the conditional distribution of X_t given $X_{t-1} = x_{t-1}$, and the positive function a on $\mathbb{R}_+$, which describes the effect of the lagged duration X_{t-1} on the conditional distribution.

The proportional hazard specification satisfies an invariance property with respect to increasing transformations, since any increasing transformation $Y_t = h(X_t)$, $t \in \mathbb{N}$, of a Markov process X_t, $t \in \mathbb{N}$, with proportional hazard features proportional hazard. This suggests alternative representations of X_t, $t \in \mathbb{N}$, in which the distribution of the process features simpler characteristics. Two such representations are considered in the following sections.

2.2 The transformed nonlinear autoregressive representation

Let us consider the nonlinear autoregressive (NLAR) representation with exponential innovations of Markov process $X_t, t \in \mathbb{N}$, (see Tong, 1990) given by:

$$X_t = \Lambda_0^{-1}\left(\frac{1}{a(X_{t-1})}\varepsilon_t\right), \quad t \in \mathbb{N}, \tag{2}$$

where ε_t, $t \in \mathbb{N}$, is a white noise, independent of $\underline{X_{t-1}}$, with a standard exponential distribution $\gamma(1)$. The duration process $X_t, t \in \mathbb{N}$, can be represented (up to the transformation Λ_0^{-1}) as a stochastic time deformation of an i.i.d. series of exponential durations ε_t, $t \in \mathbb{N}$. The time deformation factor is function of past duration.

In the NLAR representation (2), the error term ε_t, $t \in \mathbb{N}$, does not enter in an additive way. An autoregressive representation with additive noise can be derived by considering another transformation of the duration variable X_t, $t \in \mathbb{N}$. Let us introduce the transformed process:

$$Y_t = \log\left[\Lambda_0(X_t)\right], t \in \mathbb{N}.$$

We have:

$$Y_t = -\log a(X_{t-1}) + \log \varepsilon_t = \varphi\left(Y_{t-1}\right) + \eta_t, \quad t \in \mathbb{N},$$

where $\varphi\left(y\right) = -\log a\left[\Lambda_0^{-1}\left(\exp y\right)\right]$, $y \in \mathbb{R}$, and $\eta_t = \log \varepsilon_t$ follows a type I extreme value distribution.

Proposition 1 *The stationary Markov process X_t, $t \in \mathbb{N}$, features proportional hazard if and only if there exists an increasing transformation of X_t: $Y_t = h(X_t)$, $t \in \mathbb{N}$, (say) such that:*

$$Y_t = \varphi\left(Y_{t-1}\right) + \eta_t, \quad t \in \mathbb{N}, \tag{3}$$

where η_t, $t \in \mathbb{N}$, is a white noise independent of $\underline{Y_{t-1}}$ with a type I extreme value distribution.

The additive NLAR representation (3) is characterized by two functional parameters, which are the autoregression function φ of the transformed process, and the transformation function h [6]. Representation (3) is equivalent to representation (1), since the functional parameters (a, Λ_0) and (h, φ) are in a one-to-one relationship:

$$h(x) = \log \Lambda_0(x), \qquad x \in [0, \infty), \tag{4}$$

$$\varphi(y) = -\log a\left[\Lambda_0^{-1}(\exp y)\right], \qquad y \in (-\infty, \infty). \tag{5}$$

2.3 The copula representation

We may also use the invariance property of the proportional hazard specification to obtain processes with given marginal distribution. Let F be a c.d.f. on $\mathbb{R}_+$ with strictly positive density, and X_t, $t \in \mathbb{N}$, be a stationary Markov process with proportional hazard and marginal c.d.f. F. Then $U_t = F(X_t)$, $t \in \mathbb{N}$, is a stationary Markov process with proportional hazard and uniform marginal distribution on $[0,1]$. Thus, the entire class of stationary Markov processes with proportional hazard can be obtained as a transformation of processes with uniform margins on $[0,1]$: $X_t = F^{-1}(U_t)$, $t \in \mathbb{N}$.

Functions A and H_0 in the conditional survivor function of process U_t, $t \in \mathbb{N}$:

$$P\left[U_t \geq u_t \mid U_{t-1} = u_{t-1}\right] = \exp\left[-A(u_{t-1})H_0(u_t)\right], \quad u_t,\, u_{t-1} \in [0,1],$$

are constrained by the form of the marginal distribution of U_t. We have:

$$P\left[U_t \geq u\right] = E\left[P\left[U_t \geq u \mid U_{t-1}\right]\right], \quad \forall u \in [0,1], \; t > 1,$$

or equivalently:

$$1 - u = \int_0^1 \exp\left(-A(v)H_0(u)\right) dv, \quad \forall u \in [0,1].$$

This condition identifies H_0 in terms of A:

$$H_0^{-1}(z) = 1 - \int_0^1 \exp\left(-A(v)z\right) dv, \; z \in [0, \infty),$$

and functional parameter A characterizes the distribution of the process $U_t, t \in \mathbb{N}$.

Proposition 2 *(i) Let F be a c.d.f. on $\mathbb{R}_+$ with strictly positive density. Stationary Markov processes $X_t, t \in \mathbb{N}$, with proportional hazard and marginal distribution F can be written as:*

$$X_t = F^{-1}(U_t), \; t \in \mathbb{N}, \tag{6}$$

where process $U_t, t \in \mathbb{N}$, is a stationary Markov process with proportional hazard and uniform marginal *distribution on* $[0,1]$.

(ii) The conditional survivor function of process $U_t, t \in \mathbb{N}$, *with uniform margins is given by:*

$$P\left[U_t \geq u_t \mid U_{t-1} = u_{t-1}\right] = \exp\left[-A(u_{t-1})H_0(u_t, A)\right], \; t \in \mathbb{N}, \tag{7}$$

where A *is a positive function on* $[0,1]$, *and :*

$$H_0^{-1}(z, A) = 1 - \int_0^1 \exp\left(-A(v)z\right) dv, \; z \in [0, \infty). \tag{8}$$

(iii) The parameters (a, Λ_0) *of process* $X_t, t \in \mathbb{N}$, *in (6) are obtained from the corresponding parameters* (A, H_0) *of process* $U_t, t \in \mathbb{N}$, *by compounding with* F*:*

$$a = A \circ F, \quad \Lambda_0 = H_0 \circ F. \tag{9}$$

Let $X_t, t \in \mathbb{N}$, be a stationary Markov process defined by (6), with transformed process $U_t, t \in \mathbb{N}$. The copula of (X_t, X_{t-1}) is the c.d.f. of the joint distribution of (U_t, U_{t-1}) (see Joe, 1997; Nelsen, 1999). It is given by:

$$C_A(u, v) = v - \int_0^v \exp\left(-A(y)H_0(u, A)\right) dy, \;\; u, v \in [0,1], \tag{10}$$

where $H_0(., A)$ is defined by (8). Copula C_A summarizes the serial dependence between X_t and X_{t-1} that is invariant to increasing transformations. In the proportional hazard model, the copula is fully characterized by a univariate functional parameter A on $[0,1]$. Copula C_A is called proportional hazard copula.

From (8) and (9), the sets of parameters (a, Λ_0) and (A, F) are in a one-to-one relationship. Thus, stationary Markov processes with proportional hazard and strictly positive marginal density can be characterized by the functional parameters F and A. F is the marginal distribution, and can be any c.d.f. on $\mathbb{R}_+$ with strictly positive density. A is any positive function on $[0,1]$; it characterizes the copula of (X_t, X_{t-1}) and the serial dependence of the process that is invariant to monotonic transformations. This justifies the interpretation of A as a functional dependence parameter. It is identified up to a multiplicative constant. The representation in terms of functional parameters (F, A) is called copula representation. It distinguishes marginal and serial dependence characteristics of the process.

Finally, let us relate the parameterization (F, A) involving the copula and the parameterization (φ, h) corresponding to the nonlinear autoregressive representation with additive noise. From (4), (5), (8) and (9), we get:

$$\varphi(y) = -\log A\left[1-\int_0^1 \exp\left(-A(v)\exp y\right)dv\right],\ y\in(-\infty,\infty), \tag{11}$$

$$h(x) = \log H_0\left[F(x)\right],\qquad x\in[0,\infty). \tag{12}$$

Function φ depends on A only. This is not surprising, since the copula of (X_t, X_{t-1}) is the same as that of (Y_t, Y_{t-1}), and the latter depends on the autoregression function φ only. Thus, C_A is the copula of a nonlinear autoregressive Markov process with type I extreme value innovations, where the autoregressive function is restricted by (11) to ensure stationarity.

2.4 Equivalent parameterizations of the copula

When functional dependence parameter A is monotonic, equivalent parameterizations of the copula C_A are available. We consider explicitly the case of positive serial dependence, which corresponds to a decreasing function A. Then copula C_A can also be characterized by $1-A^{-1}$, that is, the c.d.f. of the variable $A(U_{t-1})$, which is the transformation of the past transformed duration U_{t-1} with proportional hazard effect on U_t. Restriction (8) can be written as:

$$1-H_0^{-1}(z)=\int_\Omega \exp\left(-wz\right)d\left(1-A^{-1}\right)(w),\ z\in[0,\infty), \tag{13}$$

where Ω denotes the range of A. Thus, function $1-H_0^{-1}$ is the real Laplace transform (also called moment generating function) of the distribution with c.d.f. $1-A^{-1}$, and satisfies the property of complete monotonicity (see Feller, 1971). Knowing A is equivalent to knowing H_0, and thus copula C_A is also characterized by the Laplace transform $1-H_0^{-1}$, or by the cumulated hazard H_0.

Proposition 3 *A proportional hazard copula with monotonically decreasing functional dependence parameter A can be equivalently defined in terms of:*

i) *either the functional dependence parameter A itself, or*

ii) *the c.d.f. $1-A^{-1}$, with support $\Omega\subset\mathbb{R}_+$, or*

iii) *its Laplace transform $1-H_0^{-1}$, or*

iv) *the baseline cumulated hazard H_0, or*

v) *the baseline survivor function $S_0\equiv\exp\left(-H_0\right)$.*

2.5 An example

In this section, we consider an example of stationary Markov process with proportional hazard, and we plot simulated trajectories, autocorrelograms and the copula's p.d.f.. This allows us to have an initial qualitative idea of the serial dependence properties of these processes, which will be discussed extensively in Section 3.

Let us assume that $1 - A^{-1}$ is a gamma distribution with parameter $1/\delta$, $\delta > 0$. Thus, $1 - A^{-1}$ is given by the incomplete gamma function $P(1/\delta, .)$ (see Abramowitz and Stegun, 1970):

$$1 - A^{-1}(w) = P(1/\delta, w) = \frac{1}{\Gamma(1/\delta)} \int_0^w \exp(-u) u^{\frac{1}{\delta}-1} du, \quad w \in [0, +\infty). \tag{14}$$

Then:

$$A(v) = A(v; \delta) = P^{-1}(1/\delta, 1 - v), \ v \in [0, 1],$$

where inversion is defined with respect to the second argument. Since:

$$H_0^{-1}(z) = 1 - \frac{1}{(1+z)^{\frac{1}{\delta}}}, \quad z \in [0, +\infty),$$

the baseline cumulated hazard is:

$$H_0(u) = \frac{1}{(1-u)^{\delta}} - 1, \ \ u \in [0, 1].$$

Let us first consider the case $\delta = \frac{1}{10}$. A simulated trajectory of 500 observations of process U_t, $t \in \mathbb{N}$, (Figure 1),

[Insert Figure 1: Simulated path for U, $\delta = 1/10$]

features weak positive serial dependence, with a stronger tendency to clustering effects at the upper boundary (large durations). The associated copula p.d.f. (Figure 2)

[Insert Figure 2: Copula p.d.f., $\delta = 1/10$]

confirms the presence of positive dependence. The copula p.d.f. diverges at points $u = v = 0$ and $u = v = 1$. The rate of divergence is related with the strength of serial dependence in the tails, and thus with clustering. The asymmetry of the density shows that the process is not time reversible. These properties of the copula can be better seen in Figure 3, which displays a contour plot of the transition p.d.f. of process $X_t^* = \Phi^{-1}(U_t)$, $t \in \mathbb{N}$, with standard Gaussian marginal distribution. Positive serial dependence is revealed by the ellipsoidal contour, and it is stronger in the upper tail, where the peaks in the isodensity lines are more pronounced.

[Insert Figure 3: Contour plot, $\delta = 1/10$]

The autocorrelogram of duration process $X_t = F^{-1}(U_t)$, $t \in \mathbb{N}$, with Pareto marginal distribution $F(x) = 1 - (1+x)^{-\tau}$, $\tau = 5.5$, is reported in Figure 4.

[Insert Figure 4: Autocorrelogram for X, $\delta = 1/10$]

Let us now increase parameter δ to $\delta = 1$. A simulated trajectory of the process (see Figure 5)

[Insert Figure 5: Simulated path for U, $\delta = 1$]

features an increased positive serial dependence with strong clustering effects, especially at the upper boundary. The copula p.d.f. (see Figure 6)

[Insert Figure 6: Copula p.d.f., $\delta = 1$]

is more concentrated in a region close to line $u = v$, and diverges at the corner points. Note the different limiting behaviour of the copula at points $u = v = 0$ and $u = v = 1$. Similarly, the contour plot with standard Gaussian marginal distribution displayed in Figure 7 is more concentrated along the diagonal and more peaked in the upper tail.

[Insert Figure 7: Contour plot, $\delta = 1$]

The autocorrelogram of process $X_t = F^{-1}(U_t)$, $t \in \mathbb{N}$, with the same marginal distribution as before, is reported in Figure 8.

[Insert Figure 8: Autocorrelogram for X, $\delta = 1$]

In the next two sections, we introduce statistical tools that are useful to analyze the qualitative features observed above.

3 Positive dependence

The aim of this section is to discuss serial dependence for stationary Markov processes with proportional hazard. Several approaches have been proposed in the literature to analyse serial dependence in nonlinear time series[7]. We focus on notions of dependence that are invariant with respect to increasing transformations and involve the copula only.

We first recall two standard notions of positive dependence based on the conditional survivor function and hazard function, respectively. They coincide for stationary processes with proportional hazard, and the condition can be written in terms of either functional dependence parameter A, or autoregressive function φ. The notions of positive dependence are used to construct dependence orderings and to introduce functional measures of dependence. Then, we discuss tail dependence properties, and report a sufficient condition that ensures that the process features positive dependence in the tails. Finally, we discuss how the dependence between X_t and X_{t-h} varies with lag h, as an introduction to ergodicity properties of the process.

3.1 Notions of positive dependence

Different notions of positive bivariate dependence can be defined, which are invariant by increasing transformations of X_t and X_{t-1}. We describe below two standard definitions and discuss their interpretation.

Definition 1 *(Lehmann, 1966; Barlow and Proschan, 1975): X_t is stochastically increasing (SI) in X_{t-1} iff*

$$S(x \mid y) \equiv P\left[X_t \geq x \mid X_{t-1} = y\right] \text{ is increasing in } y \text{, for any } x \in \mathbb{R}_+.$$

Definition 2 *(Shaked, 1977): X_t is hazard increasing (HI) in X_{t-1} iff*

$$\lambda(x \mid y) \text{ is decreasing in } y \text{, for any } x \in \mathbb{R}_+,$$

where $\lambda(. \mid y)$ denotes the conditional hazard rate of X_t given $X_{t-1} = y$.

Since $S(x \mid y) = \exp\left(-\int_0^x \lambda\left(x^* \mid y\right) dx^*\right)$, the condition of increasing hazard (HI) is stronger than condition (SI)[8]. Both dependence conditions are invariant with respect to increasing transformations of process $(X_t, t \in \mathbb{N})$ and can be written in terms of the copula.

Proposition 4 *Let X_t, $t \in \mathbb{N}$, be a stationary Markov process with proportional hazard and dependence parameter A. X_t is hazard increasing in X_{t-1} if and only if it is stochastically increasing in X_{t-1}. This condition is equivalent to function A (or a) being decreasing.*

Proof. *It is a direct consequence of the relations:*

$$\log S(u|v) = -A(v)H_0(u), \quad \textit{and} \quad \lambda(u|v) = A(v)h_0(u),$$

for $u, v \in [0,1]$, where $S(u|v)$ [resp. $\lambda(u|v)$] denotes the conditional survivor function (resp. conditional hazard function) of (U_t, U_{t-1}). *Q.E.D.*

The condition can be written in terms of nonlinear autoregression with additive noise (see Proposition 1): $Y_t = \varphi(Y_{t-1}) + \eta_t$. Indeed from equation (11), the autoregressive function φ is increasing if and only if the functional dependence parameter A is decreasing.

Corollary 5 *A stationary Markov process with proportional hazard features (HI), or (SI), positive dependence if and only if the autoregressive function φ is increasing.*

3.2 Dependence orderings

Let $(X_t, t \in \mathbb{N})$ and $(X_t^*, t \in \mathbb{N})$ be two stationary processes with proportional hazard and dependence parameter A and A^*, respectively. The aim of this section is to introduce dependence orderings in order to compare the strength of the dependence between X_t and X_{t-1} with that between X_t^* and X_{t-1}^*, or equivalently between transformed variables U_t, U_{t-1}, and U_t^*, U_{t-1}^*.

Let us first recall two definitions proposed in the statistical literature (see Yanagimoto and Okamoto, 1969; Kimeldorf and Sampson, 1987, 1989; Capéràa and Genest, 1990). For $v < v^{'}$, with $v, v^{'} \in [0,1]$, let us denote:

$$S_{v,v'}(u) = S\left[S^{-1}(u \mid v) \middle| v^{'}\right], \;\; u \in [0,1],$$

where $S(. \mid v)$ is the survivor function of U_t conditional on $U_{t-1} = v$, and similarly for $S^*_{v,v'}(u)$, $u \in [0,1]$. Intuitively, $S_{v,v'}$ measures the effect on the conditional distribution of an increase of the conditioning variable from v to $v^{'}$.

Definition 3 *: X_t is more stochastically increasing in X_{t-1} than X_t^* in X_{t-1}^* if for any $v, v^{'} \in [0,1]$, $v < v^{'}$:*

$$S_{v,v'}(u) \Big/ S^*_{v,v'}(u) \geq 1, \textit{ for any } u \in [0,1].$$

Definition 4 *: X_t is more hazard increasing in X_{t-1} than X_t^* in X_{t-1}^* if for any $v, v^{'} \in [0,1]$, $v < v^{'}$:*

$$S_{v,v'}(u) \Big/ S^*_{v,v'}(u) \textit{ is decreasing in } u \in [0,1].$$

These pre-orderings, denoted by $\succeq_{(SI)}$ and $\succeq_{(HI)}$, respectively, satisfy various desirable axioms (see Kimeldorf and Sampson, 1987, 1989; Capéràa and Genest, 1990). Moreover, since $S_{v,v'}(1)/S^*_{v,v'}(1) = 1$, the ordering $\succeq_{(HI)}$ is stronger than $\succeq_{(SI)}$[9]. Intuitively, $(X_t, X_{t-1}) \succeq_{(SI)} \left(X_t^*, X_{t-1}^*\right)$ holds if the effect on the conditional distribution of an increase in the conditioning value is stronger for (X_t, X_{t-1}) than it is for $\left(X_t^*, X_{t-1}^*\right)$. Moreover, if this is more and more pronounced as we move towards the tail of the distribution, then $(X_t, X_{t-1}) \succeq_{(HI)} \left(X_t^*, X_{t-1}^*\right)$.

The following proposition characterizes the orderings in terms of the functional dependence parameter.

Proposition 6 *Let $(X_t, t \in \mathbb{N})$ and $(X_t^*, t \in \mathbb{N})$ be stationary Markov processes with proportional hazard and dependence parameters A and A^*, respectively. The conditions $(X_t, X_{t-1}) \succeq_{(SI)} \left(X_t^*, X_{t-1}^*\right)$ and $(X_t, X_{t-1}) \succeq_{(HI)} \left(X_t^*, X_{t-1}^*\right)$ are equivalent. They are also equivalent to the condition: A/A^* is decreasing.*

Proof. *See Appendix 1.*

For the proportional hazard model, $\lambda(u|v)/\lambda(u|v')$ is independent of u and is equal to $A(v)/A(v')$. Thus, the conditions $(X_t, X_{t-1}) \succeq_{(SI)} \left(X_t^*, X_{t-1}^*\right)$ and $(X_t, X_{t-1}) \succeq_{(HI)} \left(X_t^*, X_{t-1}^*\right)$ are also equivalent to:

$$\lambda(u \mid v)/\lambda^*(u \mid v) \text{ is decreasing in } v \text{, for any } u \in [0,1].$$

When dependence parameters A and A^* are differentiable, the ordering conditions involve the elasticity of dependence parameter A, or equivalently the elasticity of the hazard function with respect to the conditioning variable.

Corollary 7 *Let $(X_t, t \in \mathbb{N})$ and $(X_t^*, t \in \mathbb{N})$ be stationary Markov processes with proportional hazard and differentiable dependence parameters A and A^*, respectively. The conditions $(X_t, X_{t-1}) \succeq_{(SI)} \left(X_t^*, X_{t-1}^*\right)$ and $(X_t, X_{t-1}) \succeq_{(HI)} \left(X_t^*, X_{t-1}^*\right)$ are equivalent to:*

$$\frac{d}{dv}\log A(v) \leq \frac{d}{dv}\log A^*(v), \quad \forall v \in [0,1],$$

or

$$\frac{\partial}{\partial v}\log\lambda(u \mid v) \leq \frac{\partial}{\partial v}\log\lambda^*(u \mid v), \ \forall u, v \in [0,1].$$

For instance, the functions:

$$A(v;\alpha) = \exp(-\alpha v), \quad A(v;\alpha) = \frac{1}{(1+v)^\alpha}, \text{ and } A(v;\alpha) = (1-v)^\alpha,$$

induce three families of distributions such that serial dependence is increasing with respect to parameter α, in both the SI and HI sense.

3.3 Measures of dependence

The discussion above shows that the appropriate functional dependence measure is not A itself, but preferably:

$$\Delta_A(v) = -\frac{d}{dv} \log A(v),\ v \in [0,1]\,.$$

The properties above can be summarized as follows:

(i) $\Delta_A(v) = 0$, $\forall v \in [0,1] \Longleftrightarrow X_t$ and X_{t-1} are independent, $t \in \mathbb{N}$;

(ii) $\Delta_A(v) \geq 0$, $\forall v \in [0,1] \Longleftrightarrow X_t$ is SI and HI in X_{t-1}, $t \in \mathbb{N}$;

(iii) $\Delta_A(v) \geq \Delta_{A^*}(v)$, $\forall v \in [0,1] \Longleftrightarrow (X_t, X_{t-1}) \succeq \left(X_t^*, X_{t-1}^*\right)$, where $\succeq$ is any of the orderings $\succeq_{(SI)}$ or $\succeq_{(HI)}$.

3.4 Tail dependence

This section provides sufficient conditions on the functional dependence parameter A to get positive dependence in the tails. The coefficient of upper tail dependence is defined by (see Joe, 1993, 1997):

$$\lambda = \lim_{u \to 1} P\left[U_t \geq u \mid U_{t-1} \geq u\right].$$

The process features positive tail dependence if $\lambda > 0$. For a process with proportional hazard, the coefficient of upper tail dependence is given by:

$$\lambda = \lambda_A = \lim_{u \to 1} \frac{1}{1-u} \int_u^1 \exp\left[-A\left(v\right) H_0\left(u, A\right)\right] dv.$$

If $\lim_{v\to 1} A(v) > 0$, then $\lambda_A = 0$, and the process is independent in the tail. Hence tail dependence is possible only if $\lim_{v\to 1} A(v) = 0$, that is, if the conditional hazard function of U_t given $U_{t-1} = v$ converges to 0 as $v \to 1$.

Proposition 8 *If the functional dependence parameter A is such that:*

$$A(v) \sim C(1-v)^{\delta}, \quad v \sim 1,$$

for some constants $\delta > 0$ and $C > 0$, then:

$$\lambda_A = \lambda(\delta) = P\left(1/\delta, \Gamma\left(1 + 1/\delta\right)^{\delta}\right),$$

where $P(1/\delta, .)$ denotes the incomplete gamma function with parameter $1/\delta$.

Proof. *See Appendix 2.*

Function $\lambda(\delta)$, $\delta \geq 0$, is increasing, and ranges from 0 to 1.

3.5 Dependence at larger lag

Let $(X_t, t \in \mathbb{N})$ be a stationary Markov process with proportional hazard and dependence parameter A. Generally the pair (X_t, X_{t-h}) does not satisfy the property of proportional hazard. However, the dependence between X_t and X_{t-h}, $h \in \mathbb{N}$, can be summarized by its copula, $C_{A,h}$, defined as the joint c.d.f. of U_t, U_{t-h}. By the Chapman-Kolmogorov formula, the copula p.d.f. $c_{A,h}$ is given by (see also Darsow *et al.*, 1992):

$$c_{A,h}(u, v) = \int_0^1 \cdots \int_0^1 c_A(u, w_1) \cdots c_A(w_{i-1}, w_i) \cdots c_A(w_{h-1}, v) dw_1 \cdots dw_{h-1}.$$

The analytic expression of $c_{A,h}$ is not available in general, but some dependence properties can be deduced from a theorem by Fang *et al.* (1994). They show that, for a stationary Markov process $(X_t, t \in \mathbb{N})$, if X_t is stochastically increasing in X_{t-1}, then X_t is still stochastically increasing in X_{t-h}, $h \in \mathbb{N}$, and $corr\left[g(X_t), g(X_{t-h-1})\right] \leq corr\left[g(X_t), g(X_{t-h})\right]$, $h \in \mathbb{N}$, for any monotonic transformation g such that these correlations exist.

Proposition 9 *Let $(X_t, t \in \mathbb{N})$ be a stationary Markov process with proportional hazard and dependence parameter A. If A is decreasing, then*

$$X_t \text{ is stochastically increasing in } X_{t-h}, \text{ for any } h \in \mathbb{N},$$

and

$$corr\left[g(X_t), g(X_{t-h-1})\right] \leq corr\left[g(X_t), g(X_{t-h})\right], \text{ for any } h \in \mathbb{N},$$

for any monotonic transformation g such that the correlations exist.

When A is decreasing, serial dependence is positive at any lag, and decreases with the horizon.

4 Ergodicity properties

The aim of this section is to study the ergodicity properties of stationary Markov processes with proportional hazard.

4.1 Geometric ergodicity

Let us first recall the definition of geometric ergodicity.

Definition 5 *Let V be a function on $\mathbb{R}_+$, such that $V \geq 1$. The Markov process $(X_t, t \in \mathbb{N})$ is V-geometrically ergodic if there exist $\rho < 1$, a probability measure π and a finite function C such that:*

$$\left\|P^t(x,.) - \pi\right\|_V \leq \rho^t C(x), \text{ for any } x \in \mathbb{R}_+,\ t \in \mathbb{N},$$

where $P^t(x,.)$ is the probability measure of X_t given $X_0 = x$, and $\|\mu\|_V = \sup_{f:|f|\leq V} \left|\int f d\mu\right|$.

For a stationary Markov process with proportional hazard, geometric ergodicity can be equivalently discussed in any of the representations of the process introduced in Section 2. Conditions for geometric ergodicity will involve either functional dependence parameter A, or functional autoregressive parameter φ, only. The NLAR representation with additive noise is the most appropriate to discuss geometric ergodicity, since the required drift conditions are easy to derive, and have been extensively investigated in the literature (see Meyn and Tweedie, 1993). Equivalent conditions can be derived for the other representations.

Proposition 10 *Let $X_t, t \in \mathbb{N}$, be a stationary Markov process with proportional hazard, with dependence parameter A. Assume A is continuous on $(0,1)$. Denote by γ the expectation of a type I extreme value variable. The following conditions are equivalent and any of them implies geometric ergodicity of process $X_t, t \in \mathbb{N}$:*

(i) The autoregressive function φ is such that there exist constants $\varepsilon > 0$, $R < \infty$, satisfying:

$$|\varphi(y) + \gamma| \leq |y| - \varepsilon, \text{ for } |y| \geq R;$$

(ii) The functional dependence parameter A is such that there exist constants $0 < R_1 < R_2 < \infty$, and $c < \exp(-\gamma) < C$, satisfying:

$$Cy \leq \frac{1}{A\left[1 - \int_0^1 \exp(-A(v)y)\,dv\right]} \leq c\frac{1}{y}, \text{ for } 0 < y \leq R_1,$$

$$C\frac{1}{y} \leq \frac{1}{A\left[1-\int_0^1 \exp\left(-A(v)y\right)dv\right]} \leq cy, \; \textit{for } y \geq R_2.$$

Proof. *See Appendix 4.*

Let us briefly discuss the ergodicity conditions. Condition (i) restricts the absolute value of the autoregressive function (including the expectation of the innovation), $|\varphi(y)+\gamma|$, to be strictly bounded by $|y|$, as $|y| \to +\infty$. This condition is less stringent than the condition usually reported in the literature (see Doukhan, 1994), that is, $|\varphi(y)+\gamma| \leq \rho\,|y|$ as $|y| \to +\infty$, for some $\rho < 1$. The weakening of the restriction on φ is possible since innovation η_t in the additive NLAR representation has a distribution with sufficiently thin tails (see Proposition A.1 in Appendix 3).

Let us now consider condition (ii) [10]. It defines restrictions on dependence parameter A, and specifically on the behaviour of $A(v)$ as $v \to 0$ and $v \to 1$, respectively. These restrictions are not immediately satisfied only if $\lim_{v\to 0} A(v)$ or $\lim_{v\to 1} A(v)$ are either 0 or $+\infty$. The intuition beyond this condition is that when $A(v)$ approaches 0 (resp. $+\infty$), the distribution of duration U_t, conditionally on $U_{t-1}=v$, concentrates close to the upper (lower) boundary. Thus, geometric ergodicity imposes restrictions on the functional dependence parameter A in a neighborhood of $v=0$ and $v=1$ in order to prevent the process from diverging to infinity or being absorbed by 0. Let us now focus on the restriction at $v=1$, when $\lim_{v\to 1} A(v) = 0$ [11]. For simplicity's sake, let us consider functions A that are continuous on $(0,1)$, decreasing near $v=1$, and such that $\forall\delta>0 : \lim_{v\to 1}\frac{A(v)}{(1-v)^\delta}$ exists (in $[0,+\infty]$). Any such function belongs to one of the following categories:

I $\quad \exists\delta>0 : \lim_{v\to 1}\frac{A(v)}{(1-v)^\delta} \in]0,+\infty[$;

II $\quad \forall\delta>0 : \lim_{v\to 1}\frac{A(v)}{(1-v)^\delta} = +\infty$;

III $\quad \forall\delta>0 : \lim_{v\to 1}\frac{A(v)}{(1-v)^\delta} \in \{0,+\infty\}$ and $\exists\delta>0 : \lim_{v\to 1}\frac{A(v)}{(1-v)^\delta} = 0$.

A function A in class I converges to 0 as $(1-v)^\delta$, for some $\delta>0$, when $v\to 1$, that is, the elasticity δ of $A(1-v)$ with respect to v at $v=1$ is strictly positive and finite[12]. Functions in class II dominate any function in class I, when $v \to 1$.

Proposition 11 *When function A is either in class I, or in class II such that for some $C>0$: $A(v) \geq \frac{-C}{\log(1-v)}$, for v close to 1, then the second restriction in condition (ii) of Proposition 10 is satisfied.*

Proof. *See Appendix 5.*

4.2 Mixing properties

We are now concerned with the decay rate of the dependence between the σ-fields up to time s, $\sigma(X_t; t \leq s)$, and from time $s+h$ onward, $\sigma(X_t; t \geq s+h)$, as the horizon h goes to infinity (see Bosq, 1998). Let us recall the definition of β-mixing with geometric decay for a Markov process.

Definition 6 *A Markov process $X_t, t \in \mathbb{N}$, is β-mixing with geometric decay if the mixing coefficients β_h, defined by*

$$\beta_h = E\left[\sup_{C \in \sigma(X_t; t \geq h)} |P(C) - P(C \mid X_0)|\right], \; h \in \mathbb{N},$$

decay geometrically: $\beta_h \leq C\rho^h$, $h \in \mathbb{N}$, for some constants $\rho < 1$, $C < \infty$.

The next proposition provides sufficient conditions for β-mixing with geometric decay of a stationary Markov process $X_t, t \in \mathbb{N}$, with proportional hazard.

Proposition 12 *Under the ergodicity conditions of Proposition 10, a stationary Markov process $X_t, t \in \mathbb{N}$, with proportional hazard is β-mixing with geometric decay.*

Proof. *See Proposition A.2 in Appendix 3.*

5 Examples

Let us now discuss examples of stationary Markov processes with proportional hazard. The associated dynamic models can be parametric or nonparametric. In all cases, i) sufficient ergodicity conditions are easily written, ii) the invariant distribution (which is the uniform distribution) is known. This is an important advantage of these models compared to the dynamic duration models previously introduced in the literature (such as the ACD models) for which neither the ergodicity conditions, nor the stationary distribution are known.

5.1 Constant measure of dependence

When the measure of dependence Δ_A is constant, we get:

$$\Delta_A(v) = -\frac{d}{dv}\log A(v) = \alpha, \forall v \in [0,1] \Longrightarrow A(v) = \exp(-\alpha v + c), \; v \in [0,1],$$

and without loss of generality, we can set $c = 0$, to obtain:

$$A(v) = \exp(-\alpha v), \; v \in [0,1], \quad \alpha \in \mathbb{R}.$$

The distribution features (SI) and (HI) positive dependence when $\alpha \geq 0$, whereas the independence case corresponds to $\alpha = 0$. Moreover, since $A(0)$ and $A(1)$ are finite and non-zero, the process is geometrically ergodic.

When $\alpha > 0$, the c.d.f. $1 - A^{-1}$ is given by:

$$1 - A^{-1}(w) = 1 + \frac{1}{\alpha} \log w,\ w \in\ \Omega = \left[e^{-\alpha}, 1\right],$$

and admits the density $1/(\alpha w)$, $w \in \Omega$. The inverse of the baseline cumulated hazard H_0 is obtained by computing the Laplace transform of $1 - A^{-1}$:

$$H_0^{-1}(z) = 1 - \frac{1}{\alpha} \int_{\exp(-\alpha)}^{1} \frac{\exp(-zw)}{w} dw = 1 - \frac{1}{\alpha} \int_{z\exp(-\alpha)}^{z} \frac{\exp(-y)}{y} dy.$$

5.2 Analytic examples

Proposition 3 suggests that simple examples can be derived when the Laplace transform admits a closed form expression (see Abramowitz and Stegun, 1970; Joe, 1997, Appendix A.1, for an extensive list). In this section we consider continuous distributions only.

i) Exponential distribution

Let us assume an exponential distribution with parameter λ: $A^{-1}(w) = \exp(-\lambda w)$, $w \in \mathbb{R}_+$, $\lambda > 0$. Without loss of generality, we can set $\lambda = 1$, and get:

$$A(v) = -\log(v),\ \ v \in [0, 1]. \tag{15}$$

Then:

$$H_0^{-1}(z) \quad = \quad 1 - \int_0^{+\infty} \exp(-zw) \exp(-w)\, dw = 1 - \frac{1}{1+z} = \frac{z}{1+z},\quad z \in [0, +\infty),$$

and the baseline cumulated hazard is:

$$H_0(u) = \frac{u}{1-u},\ \ u \in [0, 1].$$

The corresponding copula is:

$$C_A(u, v) = v - (1-u)v^{\frac{1}{1-u}},\ \ u, v \in [0, 1],$$

with density:

$$c_A(u,v) = -\frac{1}{(1-u)^2}(\log v)\, v^{\frac{u}{1-u}}, \;\; u,v \in [0,1]\,.$$

The associated proportional hazard process is geometrically ergodic. Indeed:

$$A(v) = -\log v = -\log\left[1-(1-v)\right] \sim 1-v, \text{ for } v \sim 1,$$

(see Proposition 11),

$$A\left[H_0^{-1}(y)\right] = -\log\left(\frac{y}{1+y}\right) \sim -\log y, \;\; \text{as} \;\; y \to 0,$$

and $\lim_{y\to 0} yA\left[H_0^{-1}(y)\right] = 0$ (see Proposition 10).

ii) Gamma distribution

The exponential distribution is a special case of gamma distribution. In the general gamma case, the functional dependence parameter A and the baseline cumulated hazard H_0 have been derived in Section 2.5:

$$A(v) = A\left(v;\delta\right) = P^{-1}\left(1/\delta, 1-v\right),\, v \in [0,1]\,,$$

$$H_0(u) = \frac{1}{\left(1-u\right)^{\delta}} - 1, \;\; u \in [0,1]\,.$$

The process features positive dependence since A is decreasing. The functional dependence measure is given by:

$$\Delta_A(v) \equiv \Delta(v;\delta) = \frac{\Gamma\left(\frac{1}{\delta}\right)}{e^{-A(v;\delta)}A\left(v;\delta\right)^{\frac{1}{\delta}}}, \;\; v \in [0,1]\,.$$

It is U-shaped and diverges at the boundaries $v = 0$ and $v = 1$ [see Figure 9 where $\Delta(.;\delta)$ is plotted for $\delta = 0.1$ (dashed line) and $\delta = 1$ (solid line)].

[Insert Figure 9: Functional dependence measure]

Since $\Delta(.;1) \geq \Delta(.;0.1)$, serial dependence is stronger when $\delta = 1$.

For $w \sim 0$, we have:

$$P\left(1/\delta, w\right) \;=\; \frac{1}{\Gamma\left(1/\delta\right)}\int_0^w \exp\left(-u\right)u^{\frac{1}{\delta}-1}du \sim \frac{1}{\Gamma\left(1/\delta\right)}\int_0^w u^{\frac{1}{\delta}-1}du = \frac{w^{1/\delta}}{\Gamma\left(1+1/\delta\right)},$$

and:

$$A(v) = P^{-1}\left(1/\delta, 1-v\right) \sim \Gamma\left(1+1/\delta\right)^{\delta}\left(1-v\right)^{\delta}, \;\; v \sim 1.$$

It follows from Proposition 8 that the process features positive tail dependence.

iii) Power distributions

When:

$$1 - A^{-1}(w) = w^{\frac{1}{\delta}}, \;\; w \in [0,1]\,,$$

with $\delta > 0$, we get:

$$A(v) = (1-v)^{\delta}\,, \;\; v \in [0,1]\,. \tag{16}$$

For example, the Cox model (Cox, 1955, 1972) with $a(y) = \exp(-\alpha y)$, $y \geq 0$, and an exponential marginal distribution $F(x) = 1 - \exp(-\lambda x)$, $x \geq 0$, is in this class, with $\delta = \frac{\alpha}{\lambda}$.

The Laplace transform is:

$$\begin{aligned} 1 - H_0^{-1}(z) &= \int_0^1 \exp(-wz) \frac{w^{\frac{1}{\delta}-1}}{\delta} dw \\ &= \frac{1}{\delta z^{\frac{1}{\delta}}} \int_0^z \exp(-y)\, y^{\frac{1}{\delta}-1} dy = \frac{\Gamma(1/\delta + 1)}{z^{\frac{1}{\delta}}} P(1/\delta, z)\,, \;\; z \geq 0, \end{aligned}$$

and H_0 is derived by inversion. In the special case $\delta = 1$, which corresponds to the uniform distribution $\mathcal{U}_{[0,1]}$, we get:

$$H_0^{-1}(z) = 1 - \frac{1-\exp(-z)}{z}, \;\; z \geq 0.$$

The functional measure of dependence is given by:

$$\Delta_A(v) = \Delta(v;\delta) = \frac{\delta}{1-v}, \;\; v \in [0,1]\,.$$

It is increasing, and diverges at $v = 1$. Moreover, positive dependence is increasing in δ.

Since $A(0) = 1$, processes in this class are geometrically ergodic (see Propositions 10 and 11).

iv) α-stable distributions

For some distributions neither the density nor the c.d.f. is known explicitly, but an analytical expression of the Laplace transform can be available. For example, let us assume a positive α-stable distribution, where:

$$1 - H_0^{-1}(z) = \exp\left(-z^{\frac{1}{\alpha}}\right), \;\; z \geq 0,$$

with $\alpha \geq 1$, and

$$H_0(u) = \left[-\log(1-u)\right]^{\alpha}, \;\;\; u \in [0,1]\,.$$

This serial dependence is compatible with Weibull marginal and conditional distributions for process X_t, $t \in \mathbb{N}$. More precisely, let us assume:

$$\Lambda(x) \equiv -\log(1 - F(x)) = x^{\alpha_m}, \quad \Lambda_0(x) = x^{\alpha_c}, \quad x \geq 0,$$

where $\alpha_m < \alpha_c$, then:

$$H_0(u) = \Lambda_0\left[F^{-1}(u)\right] = \left[-\log(1-u)\right]^{\frac{\alpha_c}{\alpha_m}}, \quad u \in [0,1],$$

and $1 - A^{-1}$ corresponds to a positive α-stable distribution with parameter $\alpha = \alpha_c/\alpha_m$. The larger parameter α (that is, the larger the mass of the distribution $1 - A^{-1}$ in a neighbourhood of 0), the larger the duration dependence in the marginal distribution with respect to the one in the conditional distribution.

5.3 Endogenous switching regimes

Let us consider a stepwise functional dependence parameter:

$$A(v) = \sum_{j=0}^{J} a_j \mathbb{I}_{(u_j, u_{j+1}]}(v), \quad v \in [0,1], \tag{17}$$

where $0 = u_0 < u_1 < ... < u_j < ... < u_{J+1} = 1$, $a_j \geq 0$, $j = 0, ..., J$, and $J \in \mathbb{N} \cup \{+\infty\}$. The conditional distribution is characterized by the survivor function:

$$S(u_t|u_{t-1}) = P\left[U_t \geq u_t \mid U_{t-1} = u_{t-1}\right] = \sum_{j=0}^{J} \exp\left[-a_j H_0(u_t)\right] \mathbb{I}_{(u_j, u_{j+1}]}(u_{t-1}).$$

The proportional hazard process U_t, $t \in \mathbb{N}$, features endogenous regimes, induced by qualitative thresholds in lagged duration U_{t-1}, and characterized by hazard functions which differ by a scale factor.

The stationarity condition with uniform $\mathcal{U}_{[0,1]}$ margins is:

$$1 - u = \sum_{j=0}^{J} \exp\left[-a_j H_0(u)\right](u_{j+1} - u_j), \quad \forall u \in [0,1]. \tag{18}$$

When $a_j > 0$, for at least one $j \in \{0, ..., J\}$, condition (18) characterizes the baseline cumulated hazard H_0, whose inverse is given by:

$$H_0^{-1}(z) = 1 - \sum_{j=0}^{J} \exp(-a_j z)(u_{j+1} - u_j) = 1 - \sum_{j=0}^{J} \pi_j \exp(-a_j z), \quad z \geq 0, \tag{19}$$

where $\pi_j \equiv u_{j+1} - u_j$, $j = 0, ..., J$. Equation (19) is a discrete analogue of equation (8), and represents $1 - H_0^{-1}$

as the Laplace transform of a discrete distribution on $\mathbb{R}_+$, weighting a_j, $j = 0, ..., J$, with probabilities π_j, $j = 0, ..., J$.

Figure 10 displays a stepwise functional parameter and the corresponding autocorrelogram of process $X_t, t \in \mathbb{N}$, with the same Pareto marginal distribution as in Section 2.5.

[Insert Figure 10: Endogenous switching regimes]

The autocorrelation function in Figure 10 decays slowly with large horizon h and is significantly positive up to $h \sim 500$. This example shows that the proportional hazard Markov specification of order 1 is able to generate long memory patterns. Contrary to standard linear specifications, in which persistency is introduced by means of lagged variables of high order, in our setting persistency is induced by the nonlinearities in the serial dependence structure (see also Gourieroux and Robert, 2006).

Let us now discuss examples of endogenous switching regimes models with parametrically constrained functional dependence parameter.

i) Uniform series

Assume $J = N - 1 < +\infty$, and

$$a_j = N - j, \quad \pi_j = \frac{1}{N}, \; j = 0, 1, ..., N - 1.$$

Thus, function A is regularly decreasing and:

$$H_0^{-1}(z) = 1 - \frac{1}{N} \frac{1 - \exp(-Nz)}{\exp(z) - 1}, \; z \geq 0.$$

ii) Power series

When:

$$1 - H_0^{-1}(z) = 1 - [1 - \exp(-z)]^{\frac{1}{\theta}}, \; z \geq 0,$$

with $\theta \geq 1$, the corresponding baseline cumulated hazard is:

$$H_0(u) = -\log\left(1 - u^{\theta}\right), \; u \in [0, 1].$$

By using the binomial series expansion, we get (see Joe, 1997, Appendix A.1):

$$1 - H_0^{-1}(z) = \sum_{j=0}^{\infty} \pi_j \exp(-a_j z), \quad z \geq 0,$$

with

$$a_j = j + 1, \quad \pi_j = \frac{1}{\theta^{j+1}(j+1)!} \prod_{k=1}^{j} (k\theta - 1), \quad j = 0, 1, ...$$

This defines an increasing step function (17), with thresholds at:

$$u_{j+1} = \sum_{l=0}^{j} \pi_l, \quad j = 0, 1, ...$$

A decreasing step function, with the same baseline cumulated hazard, is obtained by considering $v \mapsto A(1-v)$.

iii) Logarithmic series

When:

$$1 - H_0^{-1}(z) = -\frac{1}{\theta} \log\left[1 - \left(1 - e^{-\theta}\right)\exp(-z)\right], \quad z \geq 0, \tag{20}$$

with $\theta > 0$, the corresponding baseline cumulated hazard and survivor function are:

$$H_0(u) = -\log\left(\frac{1 - e^{-\theta(1-u)}}{1 - e^{-\theta}}\right), \quad u \in [0, 1],$$

and:

$$S_0(u) = \frac{1 - e^{-\theta(1-u)}}{1 - e^{-\theta}}, \quad u \in [0, 1],$$

respectively. The corresponding discrete distribution is found by expanding the logarithmic series in (20) to get (see Joe, 1997, Appendix A.1):

$$1 - H_0^{-1}(z) = \sum_{j=0}^{\infty} \pi_j \exp(-a_j z), \quad z \geq 0,$$

with

$$a_j = j + 1, \quad \pi_j = \frac{1}{\theta(j+1)} \left(1 - e^{-\theta}\right)^{j+1}, \quad j = 0, 1, ...$$

Again, a decreasing step function, with the same baseline cumulated hazard, is obtained by considering $v \mapsto A(1-v)$.

6 Statistical inference

Let us assume available observations X_1, ..., X_T, and discuss the efficient estimation of the dependence functional parameter, when the marginal distribution F is unconstrained. The functional parameter A can be parametrically specified, or left unconstrained.

In practice, one generally proceeds in two steps. First, the marginal c.d.f. is estimated by its sample counterpart $\widehat{F}_T$, say, and the ranks $\widehat{U}_t = \widehat{F}_T(X_t)$, $t = 1, ..., T$ provide approximations of the uniform variables U_t. Secondly, we look for an estimator of the dependence functional A from the observed $\widehat{U}_t$ and study the asymptotic properties of the estimator as if $U_t = \widehat{U}_t$, $t = 1, ..., T$, were observed. This approach disregards the information on the copula that is contained in the level of the initial variables X_t. Firstly, a joint estimation of F and A can improve the accuracy of a copula estimator. Secondly, the asymptotic properties of the estimated copula can be influenced by the replacement of U_t by $\widehat{U}_t$, at least when the functional dependence parameter is left unconstrained[13] (see Genest and Werker, 2002; Gagliardini and Gourieroux, 2007, for a more precise discussion).

Since the aim of this section is merely to give a flavour of estimation on copula, we will assume that the transformed variables U_t, $t = 1, ..., T$, are observed. We first consider (Section 6.1) the parametric framework, and we derive the expression of the score and of the efficiency bound. Then, in Section 6.2, we consider the nonparametric estimation of functional parameter A. We describe a nonparametric estimation method based on the minimum chi-square principle. This estimation approach is nonparametrically efficient. We essentially provide the main ideas that underlie the estimator and the derivation of its asymptotic properties. Proofs are gathered in Appendices 6-9 and are based on the results of Gagliardini and Gourieroux (2007).

6.1 Parametric framework

When the dependence functional is parameterized, the conditional pdf is:

$$\begin{aligned} c\left[u_t, u_{t-1}; A\left(\theta\right)\right] &= A(u_{t-1};\theta)h_0(u_t;\theta)\exp[-H_0(u_t;\theta)A(u_{t-1};\theta)] \\ &= A_{t-1}(\theta)h_{0,t}(\theta)\exp[-H_{0,t}(\theta)A_{t-1}(\theta)]. \end{aligned}$$

Parameter θ can be estimated by maximum likelihood as:

$$\widehat{\theta}_T = \arg\max_{\theta} \sum_{t=1}^{T} \log c(u_t, u_{t-1}; \theta) = \sum_{t=1}^{T} l_t(\theta), \text{ say.}$$

The score $\frac{\partial l_t}{\partial \theta}$ and the Cramer-Rao bound can be written in terms of backward conditional expectations. The results below are proved in Appendix 6.

Proposition 13 *:*

(i) The score is given by:

$$\begin{aligned}\frac{\partial l_t}{\partial \theta} &= (1 - A_{t-1}H_{0,t})\left(\frac{\partial}{\partial \theta}\log A_{t-1} - E\left[\frac{\partial}{\partial \theta}\log A_{t-1} \mid U_t\right]\right) \\ &\quad -E\left\{(1 - A_{t-1}H_{0,t})\left(\frac{\partial}{\partial \theta}\log A_{t-1} - E\left[\frac{\partial}{\partial \theta}\log A_{t-1} \mid U_t\right]\right) \mid U_t\right\},\end{aligned}$$

where $A_{t-1} = A\left(U_{t-1};\theta\right)$*, and* $H_{0,t} = H_0\left(U_t;\theta\right)$*.*

(ii) The Cramer-Rao bound is:

$$B\left(\theta\right) = I\left(\theta\right)^{-1},$$

where

$$\begin{aligned}I\left(\theta\right) &= V\left(\frac{\partial l_t}{\partial \theta}\right) = E\left[V\left(\frac{\partial l_t}{\partial \theta} \mid U_t\right)\right] \\ &= E\ V\left[(1 - A_{t-1}H_{0,t})\left(\frac{\partial}{\partial \theta}\log A_{t-1} - E\left[\frac{\partial}{\partial \theta}\log A_{t-1} \mid U_t\right]\right) \mid U_t\right].\end{aligned}$$

Since process (U_t) is also a Markov process in reverse time, the expression of the score given in Proposition 13 has the form of an expectation error (martingale difference sequence) in reverse time.

The log-derivatives of functions A and H_0 are related by:

$$\frac{\partial}{\partial \theta}\log H_0(U_t;\theta) = -E\left[\frac{\partial}{\partial \theta}\log A(U_{t-1};\theta) \mid U_t\right]. \tag{21}$$

6.2 Nonparametric minimum chi-square estimation

The nonparametric estimation approach considers the constrained nonparametric copula that is the closest to a kernel estimator of the copula for the chi-square proximity measure.

i) The estimator

Let us introduce a kernel estimator of the copula density $\widehat{c}_T(u,v)$ (say), defined by:

$$\widehat{c}_T(u,v) = \frac{1}{T}\sum_{t=2}^{T} K_{h_T}\left(u - U_t\right)K_{h_T}\left(v - U_{t-1}\right),$$

where K is a kernel, $K_{h_T}(.) = (1/h_T)\, K(./h_T)$, and h_T is a bandwidth converging to 0. Under standard regularity conditions (see the set of Assumptions in Appendix 8):

(i) This estimator converges to the true copula p.d.f. $c(u,v) = c(u,v;A_0)$, and is $\sqrt{Th_T^2}$-asymptotically normal [14]:

$$\sqrt{Th_T^2}\left(\widehat{c}_T(u,v) - c(u,v)\right) \xrightarrow{d} N\left(0, c(u,v)\left(\int K^2(w)dw\right)^2\right).$$

(ii) The integrals of the type $\int g(u,v)\widehat{c}_T(u,v)du$ and $\int\int g(u,v)\widehat{c}_T(u,v)dudv$ are asymptotically normal, but at higher nonparametric rate, and parametric rate, respectively:

$$V_{as}\left[\sqrt{Th_T}\int g(u,v)\widehat{c}_T(u,v)du\right] = E_0\left[g\left(U_t,U_{t-1}\right)^2 \mid U_{t-1}=v\right]\int K^2(w)dw, \tag{22}$$

$$V_{as}\left[\sqrt{T}\int\int g(u,v)\widehat{c}_T(u,v)dudv\right] = \sum_{h=-\infty}^{\infty} Cov\left[g\left(U_t,U_{t-1}\right), g\left(U_{t-h},U_{t-h-1}\right)\right]. \tag{23}$$

The minimum chi-square estimator is defined as:

$$\widehat{A}_T = \min_{A\in\Theta}\int\int\frac{\left[\widehat{c}_T(u,v) - c(u,v;A)\right]^2}{\widehat{c}_T(u,v)}\omega_T(u,v)dudv, \tag{24}$$

where Θ is a set of functional parameters, ω_T is a smooth weighting function converging pointwise to the constant function 1 on $[0,1]^2$ as $T\to\infty$, and the optimization is performed under the identifying constraint:

$$\int A(v)dv = 1. \tag{25}$$

ii) Asymptotic properties of the estimator

The asymptotic properties of the minimum chi-square estimator $\widehat{A}_T$ defined in (24) and (25) are reported in Proposition 14 below. In order to formulate this proposition we need some preliminary concepts (see Gagliardini and Gourieroux, 2007, and references therein). The derivation of the asymptotic properties of the minimum chi-square estimator is based on the possibility of (Hadamard) differentiating the copula density with respect to the functional parameter. The differential of $\log c(.,.;A)$ with respect to A in direction h is given by (see Appendix 6):

$$\begin{aligned}\langle D\log c\left(U_t,U_{t-1};A\right),h\rangle &= \left(1-A_{t-1}H_{0t}\right)\left(h_{t-1}/A_{t-1}-E\left[h_{t-1}/A_{t-1}\mid U_t\right]\right)\\&\quad -E\left\{\left(1-A_{t-1}H_{0t}\right)\left(h_{t-1}/A_{t-1}-E\left[h_{t-1}/A_{t-1}\mid U_t\right]\right)\mid U_t\right\}\\&= \gamma_0(U_t,U_{t-1})h\left(U_{t-1}\right)+\int\gamma_1\left(U_t,U_{t-1},w\right)h(w)dw,\end{aligned}$$

where:

$$\gamma_0(u,v)=\left[1-A(v)H_0(u)\right]/A(v),$$

and γ_1 is given in Appendix 6, Formula (a.8). Let us denote by $L^2(\lambda)$ the space of square integrable functions w.r.t. the Lebesgue measure λ on $[0,1]$, equipped with the standard inner product $(.,.)_{L^2(\lambda)}$. We assume that $D\log c\left(U_t,U_{t-1};A\right)$ is a bounded linear operator from $L^2(\lambda)$ to $L^2(P_A)$, where $L^2(P_A)$ denotes the space of square integrable random variables w.r.t. the probability measure P_A associated with $c(u,v;A)$. Further, let us denote by H the tangent space of $\left\{A\in L^2(\lambda):\int A(v)dv=1\right\}$ at A_0:

$$H=\left\{h\in L^2(\lambda):\int h(x)dx=0\right\}.$$

The asymptotic distribution of the minimum chi-square estimator is characterized by the information operator I_H, which is the bounded linear operator from H into itself defined by:

$$\left(g,I_Hh\right)_{L^2(\lambda)}=E_0\left[\langle D\log c\left(U_t,U_{t-1};A_0\right),g\rangle\langle D\log c\left(U_t,U_{t-1};A_0\right),h\rangle\right],$$

for $g,h\in H$. For the proportional hazard copula, the information operator I_H satisfies (see Appendix 7):

$$\begin{aligned}\left(g,I_Hh\right)_{L^2(\lambda)} &= ECov_0\left\{\left(1-A_{t-1}H_{0t}\right)\left(g_{t-1}/A_{t-1}-E\left[g_{t-1}/A_{t-1}\mid U_t\right]\right),\right.\\&\quad \left.\left(1-A_{t-1}H_{0t}\right)\left(h_{t-1}/A_{t-1}-E\left[h_{t-1}/A_{t-1}\mid U_t\right]\right)\right\}\\&= \int_0^1 g(w)\alpha_0(w)h(w)dw+\int_0^1\int_0^1 g(w)\alpha_1(w,v)h(v)dwdv,\end{aligned}$$

where:

$$\alpha_0(w)=\frac{1}{A_0(w)^2}, \tag{26}$$

and α_1 is defined in Appendix 7. The two components of the information operator I_H have different interpretations. The "local" component $\alpha_0(w)$ comes from the differentiation of the terms in the copula density that depend on the value of A at a point w, $w\in[0,1]$. The "functional" component α_1 comes from the differentiation of the terms in the copula density that depend on continuous functionals of A.

Proposition 14 *Under the regularity conditions in Appendix 8,*

(i) The estimator $\widehat{A}_T$ is consistent in $L^2(\lambda)$-norm.

(ii) We have the following asymptotic equivalence:

$$\begin{aligned} &\alpha_0(v)\,\delta\widehat{A}_T(v) + \int \alpha_1(v,w)\,\delta\widehat{A}_T(w)\,dw \\ = &\int \delta\widehat{c}_T(u,v)\gamma_0(u,v)\,du + \int\int \delta\widehat{c}_T(u,w)\gamma_1(u,w,v)\,dudw + r_T(v), \end{aligned} \tag{27}$$

where $\delta\widehat{A}_T = \widehat{A}_T - A_0$, $\delta\widehat{c}_T = \widehat{c}_T - c$, and the residual term r_T is such that $(h, r_T)_{L^2(\lambda)} = o_p(1/\sqrt{T})$ for any $h \in H$ and $r_T(v) = o_p(1/\sqrt{Th_T})$ λ-a.s. in $v \in [0,1]$.

(iii) The estimator $\widehat{A}_T$ is pointwise asymptotically normal:

$$\sqrt{Th_T}\left(\widehat{A}_T(v) - A_0(v)\right) \xrightarrow{d} N\left(0, A_0(v)^2 \int K^2(w)dw\right), \quad \lambda\text{-a.s. in } v \in [0,1].$$

(iv) Continuous linear functionals of $\widehat{A}_T$ are asymptotically normal:

$$\sqrt{T}\left(g, \widehat{A}_T - A_0\right)_{L^2(\lambda)} \xrightarrow{d} N\left[0, \left(g, P_H I_H^{-1} P_H g\right)_{L^2(\lambda)}\right], \quad \text{for any } g \in L^2(\lambda),$$

where P_H is the orthogonal projection operator on H.

Proof. *See Appendix 8.*

Pointwise asymptotic normality of the minimum chi-square estimator follows from the asymptotic expansion in Proposition 14 (ii), since the second term in the RHS of (27) is $O_p(1/\sqrt{T})$ [see (23)], and the same order is expected for the second term in the LHS. Then, $\delta\widehat{A}_T(v) \simeq \alpha_0(v)^{-1} \int \delta\widehat{c}_T(u,v)\gamma_0(u,v)\,du$, and Proposition 14 (iii) follows from (22) and (26).

Let us now consider the nonparametric efficiency of the minimum chi-square estimator. The nonparametric efficiency bound for functional A is defined by the semiparametric efficiency bounds $B_A(g)$ for linear functional $\int g(v)A(v)dv$, g varying, which can be consistently estimated at rate $1/\sqrt{T}$ (e.g. Bickel *et al.*, 1993; Severini and Tripathi, 2001). The nonparametric efficiency bound $B_A(g)$ is given by (see Gagliardini and Gourieroux, 2007):

$$B_A(g) = \left(g, P_H I_H^{-1} P_H g\right)_{L^2(\lambda)}, \quad g \in L^2(\lambda).$$

From Proposition 14 (iv) the minimum chi-square estimator reaches the efficiency bound and is nonparametrically efficient.

iii) Estimation of H_0^{-1}

$H_0^{-1}(z, A) = 1 - \int_0^1 \exp[-A(v)z]\, dv$ is a differentiable functional of A. More precisely, we have:

$$H_0^{-1}(z, A + \delta A) = H_0^{-1}(z, A) - \int_0^1 z \exp[-A(v)z]\, \delta A(v) dv + o(\delta A).$$

Therefore,

$$H_0^{-1}\left(z, \widehat{A}_T\right) \simeq H_0^{-1}(z, A) - \int_0^1 z \exp[-A(v)z] \left(\widehat{A}_T(v) - A_0(v)\right) dv.$$

The estimator $\widehat{H_0^{-1}}(z) = H_0^{-1}\left(z, \widehat{A}_T\right)$ is asymptotically equivalent to a continuous linear functional of $\widehat{A}_T$, and thus converges at rate $1/\sqrt{T}$ [see Proposition 14 (iv)]:

Corollary 15 *Under the regularity conditions in Appendix 8:*

$$\sqrt{T}\left(\widehat{H_0^{-1}}(z) - H_0^{-1}(z, A_0)\right) \xrightarrow{d} N\left[0, z^2 \left(e^{-zA_0}, P_H I_H^{-1} P_H e^{-zA_0}\right)_{L^2(\lambda)}\right], \; z \in (0,1).$$

In Appendix 6, it is shown that H_0 and h_0 are both differentiable functionals of A. Therefore the corresponding pointwise estimators converge at parametric rate $1/\sqrt{T}$ [15]. The higher convergence rate of H_0 and h_0 sheds light on the pointwise asymptotic distribution of the minimum chi-square estimator given in Proposition 14 (iii). Indeed, for pointwise estimation of A, functions H_0 and h_0 can be assumed to be known, in which case the information operator I_H consists only of the local component α_0 . The asymptotic variance of $\widehat{A}_T(v)$ is (essentially) its inverse.

7 Conclusion

We have introduced duration time series models with proportional hazard. These models allow us to separate the marginal characteristics from the serial dependence properties. The latter are described by a copula with proportional hazard, characterized by a functional parameter A. The consequences from a modelling point of view are twofold. On the one hand, the marginal distribution of the process can be chosen freely, and we can focus on serial dependence by considering function A. On the other hand, since parameter A is functional, this class of models allows for various nonlinear and non-Gaussian dependence features, such as dependence in the extremes, serial persistence, nonreversibility, as confirmed in simulated examples.

We have related the pattern and strength of serial dependence to the shape of functional parameter A by using well-known concepts from copulas' theory. More precisely, we have shown how various characteristics of functional parameter A give rise to different forms of serial dependence, in particular, dependence in the tails. Furthermore, we have provided sufficient ergodicity conditions in terms of functional parameter A.

Finally, we have discussed parametric and nonparametric estimation of dependence parameter A. A non-parametric estimator of A can be obtained by minimizing a chi-square distance between the nonparametric constrained copula and an unconstrained kernel estimator of the copula density. This minimum chi-square estimator is consistent, asymptotically normal, and reaches the nonparametric efficiency bound computed under the assumption that the uniform variables U_t are observed.

Appendix 1

Dependence ordering: Proof of Proposition 6

For $u, v, v^{'} \in [0,1]$, we have:

$$S(u \mid v) = \exp\left(-A(v)H_0(u)\right),$$

$$S_{v,v'}(u) = S\left[S^{-1}(u \mid v) \mid v^{'}\right] = u^{A(v^{'})/A(v)},$$

and:

$$\frac{S_{v,v'}(u)}{S^{*}_{v,v'}(u)} = u^{A(v^{'})/A(v) - A^{*}(v^{'})/A^{*}(v)}.$$

Thus, for any $v < v^{'} \in [0,1]$:

$$\begin{aligned}
\frac{S_{v,v'}(u)}{S^{*}_{v,v'}(u)} \geq 1, \ \forall u \in [0,1] \quad &\Longleftrightarrow \quad \frac{S_{v,v'}(u)}{S^{*}_{v,v'}(u)} \text{ is decreasing in } u \in [0,1] \\
&\Longleftrightarrow \quad \frac{A(v^{'})}{A(v)} \leq \frac{A^{*}(v^{'})}{A^{*}(v)} \\
&\Longleftrightarrow \quad \frac{A(v^{'})}{A^{*}(v^{'})} \leq \frac{A(v)}{A^{*}(v)}.
\end{aligned}$$

Appendix 2

Coefficient of upper tail dependence: Proof of Proposition 8

Without loss of generality, we can set $C = 1$. It will be proved in Appendix 5 [equation (a.3)] that:

$$A\left[1 - \int_0^1 \exp\left[-yA(v)\right] dv\right] \simeq \frac{\Gamma\left(1 + 1/\delta\right)^{\delta}}{y}, \text{ as } y \to +\infty.$$

When $A\left(v\right) \simeq \left(1 - v\right)^{\delta}$, $v \to 1$, we get:

$$\int_0^1 \exp\left[-yA(v)\right] dv \simeq \frac{\Gamma\left(1 + 1/\delta\right)}{y^{1/\delta}}, \text{ as } y \to +\infty.$$

Thus:

$$H_0^{-1}(z, A) \simeq 1 - \frac{\Gamma\left(1 + 1/\delta\right)}{z^{1/\delta}}, \ z \to +\infty,$$

and

$$H_0(u, A) \simeq \frac{\Gamma\left(1 + 1/\delta\right)^{\delta}}{(1 - u)^{\delta}}, \ u \to 1.$$

It follows:

$$\begin{aligned}
\lambda_A &= \lim_{u\to 1}\frac{1}{1-u}\int_u^1 \exp\left[-A(v)\,H_0(u,A)\right]dv = \lim_{u\to 1}\frac{1}{1-u}\int_u^1 \exp\left[-(1-v)^{\delta}\,\frac{\Gamma(1+1/\delta)^{\delta}}{(1-u)^{\delta}}\right]dv \\
&= \frac{1}{\Gamma(1/\delta)}\int_0^{\Gamma(1+1/\delta)^{\delta}} \exp(-w)\,w^{1/\delta-1}dw = P\left(1/\delta, \Gamma(1+1/\delta)^{\delta}\right).
\end{aligned}$$

Appendix 3

Nonlinear Autoregression

Let us provide probabilistic properties of nonlinear autoregressive models with additive noise:

$$Y_t = \varphi(Y_{t-1}) + \eta_t,$$

where the innovation η_t is a white noise, independent of $\underline{Y_{t-1}}$, with strictly positive density g on $\mathbb{R}$, and $E[\eta_t] = 0$.

The conditional density of Y_t given $Y_{t-1} = y$ is given by:

$$f(x \mid y) = g\left(x - \varphi(y)\right),\ x, y \in \mathbb{R},$$

and is strictly positive. Thus $(Y_t, t \in \mathbb{N})$ is λ-irreducible, λ-Harris recurrent (see Feigin and Tweedie, 1985) and aperiodic (see Tong, 1990, Proposition A1.2).

We assume that the autoregression function φ is continuous. Then, $(Y_t, t \in \mathbb{N})$ is a Feller chain (see Feigin and Tweedie, 1985). Indeed, if V is a bounded, continuous function defined on $\mathbb{R}$, it follows by applying Lebesgue theorem that:

$$y \mapsto E\left[V(Y_t) \mid Y_{t-1} = y\right] = \int V\left(x + \varphi(y)\right) g(x)dx,$$

is continuous.

The following proposition provides a sufficient condition for geometric ergodicity.

Proposition A.1 *Assume that the real Laplace Transform (LT) of the innovation η_t is defined in an open neighbourhood of 0. Assume further that the autoregression function φ satisfies:*

$$|\varphi(y)| \le |y| - \varepsilon, \quad |y| \ge R,$$

for some constants $\varepsilon > 0$, $R < \infty$. Then, $(Y_t, t \in \mathbb{N})$ is geometrically ergodic.

Proof. *Let $r_0 > 0$ be such that the LT of η_t:*

$$\Psi(k) = E\left[\exp\left(-k\eta_t\right)\right],$$

is defined for $k \in (-r_0, r_0)$. For $k \in (0, r_0)$ let us introduce the functions:

$$V_k(y) = 1 + \exp\left(k\left|y\right|\right), \quad y \in \mathbb{R}.$$

We now show that for some k sufficiently small, function V_k satisfies the following drift condition:

$$\exists \gamma < 1 : E\left[V_k(Y_t) \mid Y_{t-1} = y\right] \leq \gamma V_k(y), \quad \text{for } |y| \text{ large enough.} \tag{a.1}$$

Since $(Y_t, t \in \mathbb{N})$ is an irreducible, aperiodic Feller chain, and V_k is continuous, condition (a.1) implies geometric ergodicity (see Theorem 1 of Feigin and Tweedie, 1985). Let us now prove the inequality (a.1). We have:

$$\begin{aligned}
E\left[V_k(Y_t) \mid Y_{t-1} = y\right] &= 1 + E\left[\exp\left(k\left|\varphi(y) + \eta_t\right|\right)\right] \\
&= 1 + \int_{-\infty}^{-\varphi(y)} \exp\left[-k\left(\varphi(y) + \eta\right)\right] g(\eta) d\eta + \int_{-\varphi(y)}^{+\infty} \exp\left[k\left(\varphi(y) + \eta\right)\right] g(\eta) d\eta \\
&= 1 + \exp\left(-k\varphi(y)\right) \int_{-\infty}^{-\varphi(y)} \exp\left(-k\eta\right) g(\eta) d\eta + \exp\left(k\varphi(y)\right) \int_{-\varphi(y)}^{+\infty} \exp\left(k\eta\right) g(\eta) d\eta.
\end{aligned}$$

It is sufficient to consider the case where $|\varphi(y)| \to +\infty$ as $|y| \to +\infty$. Then, we have:

$$E\left[V_k(Y_t) \mid Y_{t-1} = y\right] = 1 + o(1) + (1 + o(1))\, \Psi\left[-k \cdot sign\left(\varphi(y)\right)\right] \exp\left[k\left|\varphi(y)\right|\right],$$

where $o(1) \to 0$ as $|y| \to +\infty$. It follows:

$$E\left[V_k(Y_t) \mid Y_{t-1} = y\right] \leq O(1) + (1 + o(1)) \exp\left[k\left|y\right| - k\left(\varepsilon - \frac{\psi\left[-k \cdot sign\left(\varphi(y)\right)\right]}{k}\right)\right],$$

where $\psi(k) = \ln \Psi(k)$. Since:

$$\lim_{k \to 0} \left(\varepsilon - \frac{\psi\left[-k \cdot sign\left(\varphi(y)\right)\right]}{k}\right) = \varepsilon - sign\left(\varphi(y)\right) E\left[\eta_t\right] = \varepsilon > 0,$$

there exists $\delta > 0$ *such that for* k *small enough:*

$$E\left[V_k(Y_t) \mid Y_{t-1} = y\right] \le O(1) + (1 + o(1)) \exp\left[k\left|y\right| - \delta\right].$$

Therefore, there exists $\gamma < 1$ *such that for* k *small enough:*

$$E\left[V_k(Y_t) \mid Y_{t-1} = y\right] \le \gamma V_k(y), \quad |y| \textit{ large enough,}$$

and the result follows. Q.E.D.

Finally, by using the results of Davydov (1973), geometric ergodicity[16] implies β-mixing with geometric decay (see Doukhan, 1994, ch.2.4).

Proposition A.2 *Under the assumptions of Proposition A.1,* $(Y_t, t \in \mathbb{N})$ *is* β*-mixing with geometric decay.*

Appendix 4
Proof of Proposition 10

<u>Condition (i) implies geometric ergodicity</u>

Let us consider the transformed process $Y_t = h(X_t)$, $t \in \mathbb{N}$, which follows the nonlinear autoregression with additive noise in (3), where the innovations have a type I extreme value distribution, with density:

$$g(\eta) = \exp(\eta) \exp\left(-e^{\eta}\right), \quad \eta \in \mathbb{R}.$$

This density is strictly positive on $\mathbb{R}$. From Appendix 3, it follows that Y_t, $t \in \mathbb{N}$, (and hence X_t, $t \in \mathbb{N}$) is irreducible, Harris recurrent and aperiodic. Moreover, since the continuity of A on $(0,1)$ implies the continuity of the autoregressive function φ, the process Y_t, $t \in \mathbb{N}$, (and hence X_t, $t \in \mathbb{N}$) is a Feller chain. Finally, note that the density g of the innovation admits a real LT:

$$\Psi(k) = E\left[\exp\left(-k\eta_t\right)\right] = \int_0^\infty \frac{1}{\varepsilon^k} \exp\left(-\varepsilon\right) d\varepsilon,$$

defined for $k \in (-\infty, 1)$. From Proposition A.1 in Appendix 3, geometric ergodicity of Y_t, $t \in \mathbb{N}$, and hence of X_t, $t \in \mathbb{N}$, follows.

Conditions (i) and (ii) are equivalent

By using relation (11), condition (i) can be written as:

$$\left|\log\left(e^{-\gamma}A\left[1-\int_0^1 \exp\left(-A(v)\exp y\right)dv\right]\right)\right| \leq |y|-\varepsilon, \quad |y|\geq R. \tag{a.2}$$

Let us first consider the case $y \to +\infty$, and discuss the inequality (a.2) according to the behaviour of the functional dependence parameter at $v=1$.

Case I: $\lim_{v\to 1} A(v) = 0$

Condition (a.2) becomes:

$$-\log\left(e^{-\gamma}A\left[1-\int_0^1 \exp\left(-A(v)\exp y\right)dv\right]\right) \leq y-\varepsilon, \ \ y\geq r_2,$$

for a constant $r_2 < \infty$, that is:

$$\frac{1}{A\left[1-\int_0^1 \exp\left(-A(v)\exp y\right)dv\right]} \leq e^{-\varepsilon-\gamma}\exp\left(y\right), \ \ y\geq r_2,$$

which is equivalent to:

$$\frac{1}{A\left[1-\int_0^1 \exp\left(-A(v)y\right)dv\right]} \leq cy, \quad y\geq R_2,$$

for $c < e^{-\gamma}$, and $R_2 = \exp\left(r_2\right)$.

Case II: $\lim_{v\to 1} A(v) = +\infty$

Condition (a.2) becomes:

$$\log\left(e^{-\gamma}A\left[1-\int_0^1 \exp\left(-A(v)\exp y\right)dv\right]\right) \leq y-\varepsilon, \ \ y\geq r_2,$$

for a constant $r_2 < \infty$, that is:

$$A\left[1-\int_0^1 \exp\left(-A(v)\exp y\right)dv\right] \leq e^{-\varepsilon+\gamma}\exp\left(y\right), \ \ y\geq r_2,$$

which is equivalent to:

$$\frac{1}{A\left[1-\int_0^1 \exp\left(-A(v)y\right)dv\right]} \geq C\frac{1}{y}, \quad y\geq R_2,$$

for $C > e^{-\gamma}$, $R_2 = \exp(r_2)$.

Case III: $\lim_{v\to 1} A(v) \in]0, +\infty[$

In this case, inequality (a.2) implies no restrictions on the functional dependence parameter.

Case I and II give the second restriction in condition (ii). The case $y \to -\infty$ is similar, and provides the first restriction.

Appendix 5

Proof of Proposition 11

i) Let us first assume that A is in class I. The following lemma is used in the proof.

Lemma A.3 *Let us assume that function A is strictly positive on $[0,1[$, continuous on $(0,1)$, decreasing at $v = 1$, and satisfies $\lim_{v\to 1} A(v) = 0$. Then for any $\varepsilon > 0$ small enough:*

$$\lim_{y\to+\infty} \frac{\int_{1-\varepsilon}^{1} \exp[-yA(v)]\, dv}{\int_0^1 \exp[-yA(v)]\, dv} = 1.$$

Proof. *For any $\varepsilon > 0$ small enough, and $0 < \gamma < A(1-\varepsilon)$, there exists $\delta < \varepsilon$ such that:*

$$\begin{aligned} A(v) &\geq A(1-\varepsilon), \text{ on } [0, 1-\varepsilon], \\ A(v) &\leq A(1-\varepsilon) - \gamma, \text{ on } [1-\delta, 1]. \end{aligned}$$

Thus:

$$\begin{aligned} \frac{\int_0^{1-\varepsilon} \exp[-yA(v)]\, dv}{\int_{1-\varepsilon}^{1} \exp[-yA(v)]\, dv} &\leq \frac{\exp[-yA(1-\varepsilon)]}{\int_{1-\delta}^{1} \exp[-yA(v)]\, dv} \leq \frac{\exp[-yA(1-\varepsilon)]}{\int_{1-\delta}^{1} \exp[-y(A(1-\varepsilon)-\gamma)]\, dv} \\ &\leq \frac{1}{\delta \exp(y\gamma)} \to 0, \end{aligned}$$

as $y \to +\infty$. Q.E.D.

Without loss of generality, we can assume that for some $\delta > 0$

$$\lim_{v\to 1} \frac{A(v)}{(1-v)^{\delta}} = 1.$$

Let us now consider the function involved in the second restriction of (ii). For any $\varepsilon > 0$, we have:

$$\begin{aligned}
\lim_{y\to+\infty} yA\left[1-\int_0^1 \exp\left(-yA(v)\right)dv\right] &= \lim_{y\to+\infty} \frac{A\left[1-\int_0^1 \exp\left(-yA(v)\right)dv\right]}{\left(\int_0^1 \exp\left(-yA(v)\right)dv\right)^{\delta}} y\left(\int_0^1 \exp\left(-yA(v)\right)dv\right)^{\delta} \\
&= \lim_{y\to+\infty} y\left(\int_0^1 \exp\left(-yA(v)\right)dv\right)^{\delta} \\
&= \left(\lim_{y\to+\infty} y^{\frac{1}{\delta}} \int_{1-\varepsilon}^1 \exp\left(-yA(v)\right)dv\right)^{\delta} \\
&= \left(\lim_{y\to+\infty} \frac{1}{\delta}\int_0^{+\infty} 1_{z\le\varepsilon^{\delta}y} \exp\left[-yA\left(1-\left(\frac{z}{y}\right)^{\frac{1}{\delta}}\right)\right] z^{\frac{1}{\delta}-1}dz\right)^{\delta}.
\end{aligned}$$

Let us now check that the limit and integral can be commuted by using Lebesgue theorem. Since:

$$\lim_{y\to+\infty} yA\left(1-\left(\frac{z}{y}\right)^{\frac{1}{\delta}}\right) = \lim_{y\to+\infty} z\frac{A\left(1-\left(\frac{z}{y}\right)^{\frac{1}{\delta}}\right)}{\frac{z}{y}} = z,$$

we get:

$$\lim_{y\to+\infty} 1_{z\le\varepsilon^{\delta}y} \exp\left[-yA\left(1-\left(\frac{z}{y}\right)^{\frac{1}{\delta}}\right)\right] z^{\frac{1}{\delta}-1} = \exp\left(-z\right) z^{\frac{1}{\delta}-1}, \text{ for all } z>0.$$

Moreover, let $r<1$ be such that:

$$\frac{A(v)}{(1-v)^{\delta}} \ge \frac{1}{2}, \text{ for any } v\ge r,$$

then

$$yA\left(1-\left(\frac{z}{y}\right)^{\frac{1}{\delta}}\right) \ge \frac{1}{2}z, \quad \text{for any } z\le (1-r)^{\delta} y.$$

Therefore, by choosing $\varepsilon < 1-r$, we show that the integrand admits an integrable upper bound:

$$1_{z\le\varepsilon^{\delta}y} \exp\left[-yA\left(1-\left(\frac{z}{y}\right)^{\frac{1}{\delta}}\right)\right] z^{\frac{1}{\delta}-1} \le \exp\left(-\frac{1}{2}z\right) z^{\frac{1}{\delta}-1}, \text{ for any } z,y\ge 0.$$

Thus, Lebesgue theorem applies:

$$\lim_{y\to+\infty}\int_0^{+\infty} 1_{z\le\varepsilon^{\delta}y} \exp\left[-yA\left(1-\left(\frac{z}{y}\right)^{\frac{1}{\delta}}\right)\right] z^{\frac{1}{\delta}-1}dz = \int_0^{+\infty} \exp\left(-z\right) z^{\frac{1}{\delta}-1}dz = \Gamma\left(1/\delta\right),$$

and:

$$\lim_{y\to+\infty} yA\left[1-\int_0^1 \exp\left(-yA(v)\right)dv\right] = \left[(1/\delta)\,\Gamma\left(1/\delta\right)\right]^{\delta} = \Gamma\left(1+1/\delta\right)^{\delta}. \tag{a.3}$$

In particular, we deduce from (11) that the autoregressive function φ corresponding to A is such that:

$$\varphi(y) \sim y - \delta \log \Gamma\left(1 + 1/\delta\right), \quad y \to +\infty.$$

From (a.3), it follows that the second restriction in condition (ii) is satisfied iff:

$$\Gamma\left(1 + 1/\delta\right)^{\delta} > \exp\left(\gamma\right), \text{ for any } \delta > 0,$$

where γ is the expectation of a type I extreme value variable:

$$\gamma = \int_0^{\infty} \left(\ln \varepsilon\right) \exp\left(-\varepsilon\right) d\varepsilon.$$

The conclusion follows by using the following lemma.

Lemma A.4 *The function*

$$\delta \mapsto \Gamma\left(1 + 1/\delta\right)^{\delta}, \quad \delta > 0,$$

is decreasing, with:

$$\lim_{\delta \to +\infty} \Gamma\left(1 + 1/\delta\right)^{\delta} = \exp\left(\gamma\right).$$

Proof. *Define*

$$\psi(x) \equiv \log \Gamma(1 + x),\ x \geq 0.$$

Then $\delta \mapsto \Gamma\left(1 + 1/\delta\right)^{\delta}$, $\delta > 0$, *is decreasing iff* $x \mapsto \frac{\psi(x)}{x}$ *is increasing, that is, iff:* $x\psi^{'}(x) \geq \psi(x)$, $x \geq 0$. *Since*

$$\Gamma(1 + x) = \int_0^{+\infty} \exp\left(-z\right) \exp\left(x \log z\right) dz$$

is the real LT of the negative of a type I extreme value variable, ψ *is convex, such that* $\psi(0) = 0$ [17]. *We deduce:*

$$\psi(x) = \int_0^x \psi^{'}(z)dz \leq \int_0^x \psi^{'}(x)dz = x\psi^{'}(x),$$

and the first part of the Lemma is proved. Finally, let us show the second part:

$$\begin{aligned}
\lim_{\delta \to +\infty} \Gamma\left(1 + 1/\delta\right)^{\delta} &= \lim_{\delta \to +\infty} \left(\int_0^{\infty} \exp\left(-z\right) z^{\frac{1}{\delta}} dz\right)^{\delta} = \lim_{\delta \to +\infty} \left(\int_0^{\infty} \exp\left(-z\right)\left(1 + 1/\delta \ln z + o\left(1/\delta\right)\right) dz\right)^{\delta} \\
&= \lim_{\delta \to +\infty} \left(1 + \frac{1}{\delta} \int_0^{\infty} \exp\left(-z\right) \ln z dz\right)^{\delta} = \exp\left(\int_0^{\infty} \left(\ln z\right) \exp\left(-z\right) dz\right) = \exp\left(\gamma\right).
\end{aligned}$$

Q.E.D.

ii) Let us now assume that A is in class II, and that there exists $C < \infty$ with:

$$A(v) \geq -\frac{C}{\log(1-v)}, \text{ for } v \text{ close to } 1.$$

Since $\lim_{v\to 1} A(v) = 0$, for any $\lambda \in (0, +\infty)$ there exists $K = K(\lambda)$ such that $A(v) \leq \lambda$ for $v \geq 1-K$. Then:

$$\int_0^1 \exp\left[-yA(v)\right] dv \quad \geq \quad \int_{1-K}^1 \exp\left[-yA(v)\right] dv \geq K \exp\left(-\lambda y\right), \; y \geq 0.$$

Since A is decreasing near 1,

$$A\left[1 - \int_0^1 \exp\left(-yA(v)\right) dv\right] \geq A\left[1 - K\exp\left(-\lambda y\right)\right], \text{ for } y \text{ large}.$$

Then:

$$\begin{aligned}
yA\left[1 - \int_0^1 \exp\left(-yA(v)\right) dv\right] &\geq yA\left[1 - K\exp\left(-\lambda y\right)\right] \\
&= -\frac{1}{\lambda}\log\left(\frac{1 - \left[1 - K\exp\left(-\lambda y\right)\right]}{K}\right) A\left[1 - K\exp\left(-\lambda y\right)\right] \\
&= \frac{C}{\lambda} + o(1) > \exp\left(\gamma\right), \text{ for } y \text{ large enough,}
\end{aligned}$$

if we choose $\lambda < C\exp\left(-\gamma\right)$.

Appendix 6

Computation of the differential of $c\,(u, v; A)$ with respect to A

The aim of this appendix is to derive different expressions of the differential of the copula with respect to the functional parameter. In a first step, we derive the differential with respect to A, by taking into account that H_0 is a functional of A, as a result of the relationship implied by the condition of uniform marginal distribution. In a second step, we provide interpretations in terms of backward expectations. Finally, the results are particularized to the parametric framework.

i) The general expression

Let us derive the first-order expansion of the copula log-density:

$$\log c\left(u,v;A\right)=\log A\left(v\right)+\log h_{0}\left(u,A\right)-A\left(v\right)H_{0}\left(u,A\right),$$

with respect to functional parameter A. We get:

$$\begin{aligned}\log c\left(u,v;A+\delta A\right)&=\log\left[A\left(v\right)+\delta A\left(v\right)\right]+\log h_{0}\left(u,A+\delta A\right)-\left[A\left(v\right)+\delta A\left(v\right)\right]H_{0}\left(u,A+\delta A\right)\\&\simeq\log c\left(u,v;A\right)+\frac{\delta A\left(v\right)}{A\left(v\right)}+\left\langle D\log h_{0}\left(u,A\right),\delta A\right\rangle\\&-H_{0}\left(u,A\right)\delta A\left(v\right)-A\left(v\right)\left\langle DH_{0}\left(u,A\right),\delta A\right\rangle\\&=\log c\left(u,v;A\right)+\frac{1-A\left(v\right)H_{0}\left(u,A\right)}{A\left(v\right)}\delta A(v)\\&+\left\langle D\log h_{0}\left(u,A\right),\delta A\right\rangle-A\left(v\right)\left\langle DH_{0}\left(u,A\right),\delta A\right\rangle,\end{aligned}\tag{a.4}$$

where the expansions are in terms of Hadamard derivatives and the sign $\simeq$ means that the residual terms are negligible. Let us now derive the expressions of the derivative of H_0 and h_0 with respect to A.

Expression of $DH_0^{-1}\left(z,A\right)$

We have:

$$\begin{aligned}H_{0}^{-1}\left(z,A+\delta A\right)&=1-\int_{0}^{1}\exp\left[-A\left(v\right)z-\delta A\left(v\right)z\right]dv\simeq1-\int_{0}^{1}\left[1-\delta A\left(v\right)z\right]\exp\left[-A\left(v\right)z\right]dv\\&=H_{0}^{-1}\left(z,A\right)+\int_{0}^{1}z\delta A\left(v\right)\exp\left[-A\left(v\right)z\right]dv;\end{aligned}$$

hence:

$$\left\langle DH_{0}^{-1}\left(z,A\right),\delta A\right\rangle=\int_{0}^{1}z\exp\left[-A\left(v\right)z\right]\delta A(v)dv.$$

Expression of $DH_0(u;A)$

By applying the implicit function theorem, we get:

$$\begin{aligned}\left\langle DH_{0}\left(u,A\right),\delta A\right\rangle&=-h_{0}\left(u,A\right)\left\langle DH_{0}^{-1}\left(H_{0}\left(u,A\right),A\right),\delta A\right\rangle\\&=-h_{0}\left(u,A\right)\int_{0}^{1}H_{0}\left(u,A\right)\exp\left[-A\left(v\right)H_{0}\left(u,A\right)\right]\delta A(v)dv\end{aligned}\tag{a.5}$$

Expression of $D\log h_0(u;A)$

We get:

$$h_0(u,A) = \left(\frac{d}{dz} H_0^{-1}(z,A) \Big|_{z=H_0(u,A)} \right)^{-1} = \left(\int_0^1 A(v) \exp\left[-A(v) H_0(u,A)\right] dv \right)^{-1}.$$

Let us introduce the functional:

$$q(u,A) \equiv \frac{1}{h_0(u,A)} = \int_0^1 A(v) \exp\left[-A(v) H_0(u,A)\right] dv,$$

and derive its first-order expansion. We get:

$$\begin{aligned}
q(u, A+\delta A) &= \int_0^1 \left[A(v)+\delta A(v)\right] \exp\left\{-\left[A(v)+\delta A(v)\right] H_0(u, A+\delta A)\right\} dv \\
&\simeq q(u,A) + \int_0^1 \delta A(v) \exp\left[-A(v) H_0(u,A)\right] dv \\
&\quad - H_0(u,A) \int_0^1 \delta A(v) A(v) \exp\left[-A(v) H_0(u,A)\right] dv \\
&\quad - \left\langle D H_0(u), \delta A \right\rangle \int_0^1 A(v)^2 \exp\left[-A(v) H_0(u,A)\right] dv \\
&= q(u,A) + \int_0^1 \delta A(v) \left[1 - A(v) H_0(u,A)\right] \exp\left[-A(v) H_0(u,A)\right] dv \\
&\quad - \left\langle D H_0(u), \delta A \right\rangle \int_0^1 A(v)^2 \exp\left[-A(v) H_0(u,A)\right] dv.
\end{aligned}$$

It follows:

$$\begin{aligned}
&\left\langle D \log h_0(u,A), \delta A \right\rangle \\
&= -h_0(u,A) \left\langle D q(u,A), \delta A \right\rangle \\
&= -h_0(u,A) \int_0^1 \delta A(v) \left[1 - A(v) H_0(u,A)\right] \exp\left[-A(v) H_0(u,A)\right] dv \\
&+ h_0(u,A) \left(\int_0^1 A(v)^2 \exp\left[-A(v) H_0(u,A)\right] dv \right) \left\langle D H_0(u), \delta A \right\rangle \\
&= -h_0(u,A) \int_0^1 \delta A(v) \left[1 - A(v) H_0(u,A)\right] \exp\left[-A(v) H_0(u,A)\right] dv \\
&- h_0(u,A)^2 \left(\int_0^1 A(v)^2 \exp\left[-A(v) H_0(u,A)\right] dv \right) \int_0^1 H_0(u,A) \exp\left[-A(v) H_0(u,A)\right] \delta A(v) dv.
\end{aligned} \tag{a.6}$$

Explicit expression of the copula's derivative

By substituting (a.5) and (a.6) into (a.4), the expansion of $\log c(u,v;A)$ becomes:

$$\log c\,(u,v;A+\delta A) \simeq \log c(u,v;A) + \gamma_0(u,v,A)\delta A(v) + \int \gamma_1(u,v,w;A)\delta A(w)dw,$$

where:

$$\gamma_0(u,v,A) = \frac{1 - A(v)H_0(u,A)}{A(v)}, \tag{a.7}$$

and:

$$\begin{aligned}\gamma_1(u,v,w;A) = {} & -h_0(u,A)\exp\left[-A(w)H_0(u,A)\right] \\ & \cdot \left\{1 - H_0(u,A)\left[A(v) + A(w) - \int_0^1 A(z)^2 h_0(u,A)\exp\left[-A\,(z)\,H_0\,(u,A)\right]dz\right]\right\}.\end{aligned} \tag{a.8}$$

The expression of the differential of $\log c(u,v;A)$ follows:

$$\langle D\log c(u,v;A), \delta A\rangle = \gamma_0(u,v,A)\delta A(v) + \int \gamma_1(u,v,w;A)\delta A(w)dw. \tag{a.9}$$

ii) Conditional expectations in reverse time

Various functional derivatives with respect to A can be written as expectations in reverse time. From (a.5) we get:

$$\langle DH_0\,(u,A)\,,\delta A\rangle = -H_0\,(u,A)\,E\left[\delta A\,(U_{t-1})\,/A(U_{t-1}) \mid U_t = u\right],$$

or equivalently:

$$\langle D\log H_{0t}, \delta A\rangle = -E\left[\delta A_{t-1}/A_{t-1} \mid U_t\right],$$

where $H_{0t} = H_0\,(U_t, A)$ and $A_{t-1} = A\,(U_{t-1})$. Similarly, from (a.6) we get:

$$\langle D\log h_{0t}, \delta A\rangle \quad = \quad -E\left[(1 - A_{t-1}H_{0t})\,\delta A_{t-1}/A_{t-1} \mid U_t\right] - E\left[A_{t-1}H_{0t} \mid U_t\right]E\left[\delta A_{t-1}/A_{t-1} \mid U_t\right].$$

Then, from (a.4) the score of the model can be written as an expectation error in reverse time:

$$\begin{aligned}\langle D\log c\,(U_t, U_{t-1};A)\,,\delta A\rangle = {} & (1 - A_{t-1}H_{0t})\,(\delta A_{t-1}/A_{t-1} - E\left[\delta A_{t-1}/A_{t-1} \mid U_t\right]) \\ & - E\left\{(1 - A_{t-1}H_{0t})\,(\delta A_{t-1}/A_{t-1} - E\left[\delta A_{t-1}/A_{t-1} \mid U_t\right]) \mid U_t\right\}.\end{aligned} \tag{a.10}$$

iii) The parametric case

When function A is parameterized:

$$A(v) = A(v, \theta),$$

the score of the model is obtained from (a.10) with:

$$\delta A(v) = \frac{\partial A}{\partial \theta}(v, \theta)\delta\theta.$$

We get:

$$\begin{aligned}
\frac{\partial l_t}{\partial \theta}(\theta) &= \frac{\partial}{\partial \theta} \log c\left(U_t, U_{t-1}; A\left(\theta\right)\right) \\
&= \left(1 - A_{t-1}H_{0t}\right)\left(\frac{\partial}{\partial \theta}\log A_{t-1}\left(\theta\right) - E\left[\frac{\partial}{\partial \theta}\log A_{t-1}\left(\theta\right) \mid U_t\right]\right) \\
&\quad -E\left\{\left(1 - A_{t-1}H_{0t}\right)\left(\frac{\partial}{\partial \theta}\log A_{t-1}\left(\theta\right) - E\left[\frac{\partial}{\partial \theta}\log A_{t-1}\left(\theta\right) \mid U_t\right]\right) \mid U_t\right\}.
\end{aligned}$$

Similarly, the derivatives of $\log H_0\left(u, A\left(\theta\right)\right)$ and $\log h_0\left(u, A\left(\theta\right)\right)$ with respect to θ are given by:

$$\frac{\partial}{\partial \theta}\log H_{0t}\left(\theta\right) = -E\left[\frac{\partial}{\partial \theta}\log A_{t-1}\left(\theta\right) \mid U_t\right],$$

and:

$$\frac{\partial}{\partial \theta}\log h_{0t}\left(\theta\right) = -E\left[\left(1 - A_{t-1}H_{0t}\right)\frac{\partial}{\partial \theta}\log A_{t-1}\left(\theta\right) \mid U_t\right] - E\left[H_{0t}A_{t-1} \mid U_t\right]E\left[\frac{\partial}{\partial \theta}\log A_{t-1}\left(\theta\right) \mid U_t\right].$$

Appendix 7

The information operator

i) The expression of the information operator

Let us derive the information operator I_H. Using (a.9) in Appendix 6, we get

$$\begin{aligned}
\left(g, I_H h\right)_{L^2(\lambda)} &= E_0\left[\left\langle D\log c\left(U_t, U_{t-1}; A_0\right), g\right\rangle \left\langle D\log c\left(U_t, U_{t-1}; A_0\right), h\right\rangle\right] \\
&= \int g(v)\alpha_0(v)h(v)dv + \int g(w)\alpha_1(w, v)h(v)dwdv, \qquad \text{(a.11)}
\end{aligned}$$

for $g, h \in H$, where:

$$\alpha_0(v) = E_0 \left[\gamma_0 \left(U_t, U_{t-1} \right)^2 \mid U_{t-1} = v \right] = \frac{1}{A_0(v)^2},$$

and:

$$\alpha_1(w, v) \quad = \quad \int \gamma_0(u, w)\gamma_1(u, w, v)du + \int \gamma_0(u, v)\gamma_1(u, v, w)du + \int \gamma_1(u, y, w)\gamma_1(u, y, v)dudy.$$

Let us now derive an expression for $I_H h$, $h \in H$. From (a.11) we get:

$$\int g(w) \left[I_H h(w) - \alpha_0(w)h(w) - \int \alpha_1(w, v)h(v)dv \right] dw = 0, \; \forall g \in H.$$

Thus, there exists a constant k such that:

$$I_H h(w) = \alpha_0(w)h(w) + \int \alpha_1(w, v)h(v)dv + k.$$

Constant k is determined by the condition $I_H h \in H$, that is, $\int I_H h(w)dw = 0$. We get:

$$I_H h(w) = \alpha_0(w)h(w) + \int \alpha_1(w, v)h(v)dv - \int \alpha_0(w)h(w)dw - \int \alpha_1(w, v)h(v)dvdw. \tag{a.12}$$

Thus, I_H admits the representation:

$$I_H h(w) = \alpha_{0,H}(w)h(w) + \int \alpha_{1,H}(w, v)h(v)dv, \quad \text{say},$$

with $\alpha_{0,H} = \alpha_0$ (see equation (22) in Gagliardini and Gourieroux, 2007).

ii) Boundedness and invertibility of I_H

We assume that there exists a positive function $\alpha_H(.) > 0$ such that:

$$\int \int \frac{\alpha_{1,H}(w, v)^2}{\alpha_H(w)\alpha_H(v)} dwdv < +\infty.$$

Moreover, we assume that:

$$\sup_{v \in [0,1]} \max \left\{ \frac{1}{A_0(v)^2}, \alpha_H(v) \right\} < \infty.$$

Then, from Proposition B.1 in the technical Appendix of Gagliardini and Gourieroux (2007), $D \log c \left(U_t, U_{t-1}; A_0 \right)$ is a bounded operator from $L^2(\lambda)$ to $L^2(P_{A_0})$, and I_H is a bounded operator from H in itself.

Let us now consider the invertibility of I_H. In Gagliardini and Gourieroux (2007), Section 6.1, it is shown that $D \log c\,(U_t, U_{t-1}; A_0)$ has a zero null space on H. Further, let us assume that functional parameter A_0 is bounded. Then Proposition B.2 in the technical Appendix of Gagliardini and Gourieroux (2007) implies that I_H is invertible.

The above assumptions require functional parameter A to be bounded and bounded away from 0. It is possible to show that less restrictive conditions are sufficient, if we choose $\log A$ as functional parameter, instead of A itself. However, this would imply a nonlinear identification constraint. Similarly, less restrictive assumptions are sufficient if we allow for a general measure ν in the definition of the $L^2(\nu)$ scalar product in the functional parameter space.

Appendix 8

Asymptotic properties of the minimum chi-square estimator

In this Appendix we consider the asymptotic properties of the minimum chi-square estimator. We first introduce the set of regularity assumptions. Then, we prove Proposition 14.

A.8.1 Regularity assumptions

Assumption A.1: *Process $U_t, t \in \mathbb{N}$, is a Markov process with proportional hazard, with copula p.d.f. $c(u, v; A_0)$ and uniform stationary distribution. We denote by P_A the probability measure associated with $c(u, v; A)$, $A \in \mathcal{A}$, where $\mathcal{A}$ is an open subset of $L^2(\lambda)$ containing A_0.*

Assumption A.2: *The Hadamard derivative of $\log c(u, v; A)$ with respect to A, denoted by $D \log c(u, v; A)$, exists:*

$$\log c(u, v; A + h) - \log c(u, v; A) = \langle D \log c(u, v; A), h \rangle + R(u, v; A, h),$$

for $A, A + h \in \mathcal{A}$, where $D \log c(u, v; A)$ is a linear mapping from $L^2(\lambda)$ to $\mathbb{R}$ which associates to $h \in L^2(\lambda)$ the quantity $\langle D \log c(u, v; A), h \rangle \in \mathbb{R}$. When u, v are replaced by U_t, U_{t-1} with distribution P_A, the Hadamard derivative becomes stochastic and $D\log c(U_t, U_{t-1}; A)$ is a linear operator from $L^2(\lambda)$ to $L^2(P_A)$ which associates to $h \in L^2(\lambda)$ the random variable $\langle D \log c(U_t, U_{t-1}; A), h \rangle \in L^2(P_A)$. We also assume that:

(i) *the operator $D \log c(U_t, U_{t-1}; A) : L^2(\lambda) \to L^2(P_A)$ is bounded, $\forall\ A \in \mathcal{A}$,*

(ii) *the stochastic residual term $R(U_t, U_{t-1}; A, h)$ is such that $\forall A \in \mathcal{A}$, $K \subset \mathcal{A}$ compact: $\|R(U_t, U_{t-1}; A, h)\|_{L^2(P_A)} / \|h\|_{L^2(\lambda)} \to 0$, uniformly in $h \in K$.*

Assumption A.3: *The information operator I_H is invertible, with a continuous inverse I_H^{-1}.*

Assumption A.4: *The process U_t, $t \in \mathbb{N}$, is geometric β-mixing.*

Assumption A.5: *The copula density $c(u,v) = c(u,v;A_0)$ of (U_t, U_{t-1}) is of class $C^m\left([0,1]^2\right)$ and vanishes at the boundary.*

Assumption A.6: *There exist constants $\bar{C} > 0$, $\gamma > 0$, and an increasing sequence of sets $\Omega_T \subset (0,1)^2$, $T \in \mathbb{N}$, such that:*

$$\inf_{(u,v)\in\Omega_T} c(u,v) > \bar{C}(\log T)^{-\gamma}, \ \ \textit{for any } T.$$

Assumption A.7: *The conditional density $c_h(z,w \mid u,v)$* of (U_t, U_{t-1}) given $(U_{t-h}, U_{t-h-1}) = (u,v)$ *is such that:*

$$\sup_{h\in\mathbb{N}} \sup_{(z,w),(u,v)\in[0,1]^2 : c(u,v)>0} c_h(z,w \mid u,v) < +\infty.$$

Assumption A.8: *The kernel K is of class C^m, with derivatives in $L^2(\mathbb{R})$, and is Lipschitz. Moreover, the kernel K is of order $m \geq 2$, that is:*

$$\int u^s K(u)du = 0, \quad s = 1,...,m-1, \ \textit{and} \ \int |u|^m K(u)du < +\infty.$$

Assumption A.9: *The bandwidth h_T is such that $h_T = \bar{c}_T T^{-\alpha}$, $lim_{T\to\infty}\bar{c}_T = \bar{c} > 0$, with:*

$$\frac{1}{4m}\left(1 + \frac{2m-1}{4m^2+2m+1}\right) < \alpha < \frac{1}{4}\left(1 - \frac{1}{2}\frac{2m-1}{4m^2+2m+1}\right).$$

Assumption A.10: *There exist compact sets $\widetilde{\Omega}_T, \Omega_T$ such that $\widetilde{\Omega}_T \subset \Omega_T \subset [0,1]^2$, weighting function ω_T has support in Ω_T, is smaller than 1 with restriction $\omega_T|_{\widetilde{\Omega}_T} = 1$, $T \in \mathbb{N}$, and $\lambda_2\left(\widetilde{\Omega}_T\right) \to 1$, as $T \to \infty$, where λ_2 is the Lebesgue measure on $[0,1]^2$.*

Assumption A.11: *The set Ω_T in Assumption A.10 satisfies Assumption A.6.*

Assumption A.12: *For any $A, A_0 \in \Theta$: $c(U_t, U_{t-1}; A)/c(U_t, U_{t-1}) \in L^2(P_{A_0})$. Moreover, the first-order expansion of the copula density is such that:*

$$c(U_t, U_{t-1}; A+h) = c(U_t, U_{t-1}; A) + \langle Dc(U_t, U_{t-1}; A), h\rangle + R(U_t, U_{t-1}; A, h), \quad \forall A, A+h \in \Theta,$$

where:

i) $Dc(U_t, U_{t-1}; A)/c(U_t, U_{t-1})$ is a bounded operator from $L^2(\lambda)$ in $L^2(P_{A_0}), \forall A, A_0 \in \Theta$;

ii) the residual term satisfies: $\|R(U_t, U_{t-1}; A, h)/c(U_t, U_{t-1})\|_{L^2(P_{A_0})} = O\left(\|h\|^2_{L^2(\lambda)}\right), \quad A, A+h \in \Theta.$

iii) Moreover: $R(u, v; A_0, h) = O\left(\left(\|h(u)\| + \|h(v)\| + \|h\|_{L^2(\lambda)}\right)^2\right), \quad \lambda_2$*-a.s. in* $(u, v) \in (0,1)^2$.

Assumption A.13: *There exists* $p > 1$ *such that:*

$$\sup_{A \in \Theta} \left\| \frac{c(.,.;A)^2}{c(.,.)} \right\|_{L^p} < \infty.$$

Assumption A.14: *The set* Θ *is bounded and closed with respect to the norm* $\|.\|_{L^2(\lambda)}$.

Assumption A.15: *The set* $\{c(.,.;A), A \in \Theta\}$ *is bounded and weakly closed in* $L^2(\mu)$ *for any measure* μ *on* $(0,1)^2$ *with compact support and continuous density w.r.t* λ_2.

Assumption A.16: *The information operator* I_H *is such that:*

$$\inf_{h \in H: \|h\|_{L^2(\lambda)}=1} (h, I_H h)_{L^2(\lambda)} > 0.$$

Assumption A.17: *Parameter set* Θ *has a non-empty interior (w.r.t.* $\|.\|_{L^2(\lambda)}$*) containing the true functional parameter* A_0.

Assumption A.18: *The operator* $\frac{Dc(U_t, U_{t-1}; A)}{c(U_t, U_{t-1})}$ *is Lipschitz with respect to* A *at* A_0 :

$$\left\| \frac{Dc(U_t, U_{t-1}; A_0 + h)}{c(U_t, U_{t-1})} - \frac{Dc(U_t, U_{t-1}; A_0)}{c(U_t, U_{t-1})} \right\|_L \leq \bar{C} \|h\|_{L^2(\lambda)},$$

for a constant $\bar{C}$*, where* $\|.\|_L$ *denotes the* L^2*-norm on the space of bounded linear operators from* $L^2(\lambda)$ *into* $L^2(P_{A_0})$.

Assumption A.19: *There exists* $p > 1$ *such that:*

$$\|\langle D \log c(.,.;A_0), g \rangle \langle D \log c(.,.;A_0), h \rangle c(.,.)\|_{L^p} = O\left(\|g\|_{L^2(\lambda)} \|h\|_{L^2(\lambda)}\right).$$

Assumption A.20: *There exists* $\beta_2 > q/4$ *such that:*

$$\lambda_2(\widetilde{\Omega}_T^c) = O\left(T^{-\beta_2}\right),$$

where p *is the value given in Assumption A.19,* $1/p + 1/q = 1$*, and* $\widetilde{\Omega}_T$ *is defined in Assumption A.10.*

The boundedness of the differential operator (Assumption A.2) and the invertibility of the information operator (Assumption A.3) can be verified using primitive conditions on functions α_0 and α_1 (see Appendix

7). Moreover, the decomposition of the information operator I_H in Assumption A.3 of Gagliardini and Gourieroux (2007) is satisfied (see Appendix 7). Sufficient conditions on the functional parameter A_0 to ensure geometric β-mixing (Assumption A.4) are given in Proposition 12. Assumptions A.5-A.9 are standard conditions on the copula p.d.f., the kernel and the bandwidth for kernel density estimation. In particular, Assumption A.9 allows for optimal bandwidth choice $h_T = O\left(T^{-1/(2m+1)}\right)$. The minimum chi-square estimator is pointwise asymptotically unbiased if, in addition to A.9, we have $\alpha > 1/(2m+1)$ (see Gagliardini and Gourieroux, 2007). To control the boundary bias, less restrictive assumptions on the copula p.d.f. can be introduced, if generalized kernels are used (e.g. Rice, 1984; Jones, 1993). Assumptions A.10, A.11 and A.20 explain how the sequence of weighting functions ω_T with compact support Ω_T converges to the constant function 1 on $[0,1]^2$. Assumptions A.14 and A.17 describe the functional parameter set Θ. Assumption A.15 is useful to prove the existence of the minimum chi-square estimator $\widehat{A}_T$, since the criterion defining $\widehat{A}_T$ is a distance in an Hilbert space. Finally, Assumptions A.12, A.13, A.16, A.18 and A.19 are technical restrictions on functional parameter A to ensure integrability conditions for the copula p.d.f. and its differential. They are used to bound the residual terms in the asymptotic expansion of the minimum chi-square estimator.

A.8.2 Proof of Proposition 14

Let us now prove Proposition 14. Point (i) follows from Proposition 12 in Gagliardini and Gourieroux (2007). To prove point (ii), let us first derive the efficient score $\psi_T \in L^2(\lambda)$, which is defined by:

$$(h, \psi_T)_{L^2(\lambda)} = \int\int \delta\widehat{c}_T(u,v) \left\langle D \log c\,(u,v;A_0)\,, h\right\rangle dudv, \quad \forall h \in L^2(\lambda)\,.$$

From Gagliardini and Gourieroux (2007), equation (16), we get:

$$\psi_T(w) = \int \delta\widehat{c}_T(w,v)\gamma_0\,(w,v)\,dv + \int\int \delta\widehat{c}_T(u,v)\gamma_1\,(u,v,w)\,dudv. \tag{a.13}$$

The first-order condition of the minimum chi-square estimator is given by (see Gagliardini and Gourieroux, 2007, Proposition 4, Sections 4.5 and 6.2):

$$I_H \delta\widehat{A}_T = P_H \psi_T + \widetilde{r}_T,$$

where P_H is the orthogonal projection operator on the tangent space H, defined by $P_H h(v) = h(v) - \left(\int h(w)dw\right)$, and the residual term $\widetilde{r}_T$ is such that $\|\widetilde{r}_T\|_{L^2(\lambda)} = o_p(1/\sqrt{T})$ and $\widetilde{r}_T(v) = o_p(1/\sqrt{Th_T})$ λ-a.s.

in $v \in [0,1]$. From (a.12) and (a.13), we get:

$$\begin{aligned}
&\alpha_0(w)\delta\widehat{A}_T(w) + \int \alpha_1(w,v)\delta\widehat{A}_T(v)dv - \int\left(\alpha_0(w)\delta\widehat{A}_T(w) + \int \alpha_1(w,v)\delta\widehat{A}_T(v)dv\right)dw \\
&= \int \delta\widehat{c}_T(w,v)\gamma_0\left(w,v\right)dv + \int\int \delta\widehat{c}_T(u,v)\gamma_1\left(u,v,w\right)dudv \\
&- \left(\int\int \delta\widehat{c}_T(w,v)\gamma_0\left(w,v\right)dv + \int\int \delta\widehat{c}_T(u,v)\gamma_1\left(u,v,w\right)dudv\right)dw + \widetilde{r}_T(w). \qquad \text{(a.14)}
\end{aligned}$$

This gives the asymptotic expansion reported in Proposition 14 (ii) with:

$$\begin{aligned}
r_T(v) &= \widetilde{r}_T(v) - \left(\int\int \delta\widehat{c}_T(w,v)\gamma_0\left(w,v\right)dv + \int\int \delta\widehat{c}_T(u,v)\gamma_1\left(u,v,w\right)dudv\right)dw \\
&\quad + \int\left(\alpha_0(w)\delta\widehat{A}_T(w) + \int \alpha_1(w,v)\delta\widehat{A}_T(v)dv\right)dw \equiv \widetilde{r}_T(v) + k_T.
\end{aligned}$$

Then, $(r_T, h)_{L^2(\lambda)} = (\widetilde{r}_T, h)_{L^2(\lambda)} \leq \|\widetilde{r}_T\|_{L^2(\lambda)} \|h\|_{L^2(\lambda)} = o_p(1/\sqrt{T})$ for any $h \in H$. Moreover, from equation (23) and point (iv), we have $k_T = O_p(1/\sqrt{T})$. Thus, $r_T(v) = o_p(1/\sqrt{Th_T})$ λ-a.s. in $v \in [0,1]$ and point (ii) is proved.

Finally, points (iii) and (iv) follow from Propositions 13 and 16 in Gagliardini and Gourieroux (2007).

Acknowledgements

We thank P. Balestra, S. Cain and an anonymous referee for helpful comments. Financial support by the National Center of Competence in Research "Financial Valuation and Risk Management" (NCCR FINRISK) is gratefully acknowledged by the first author. The second author has benefitted from the support of the Chair AXA: "Large Risks in Insurance".

References

ABRAMOWITZ, M. and STEGUN, I. (1970) *Handbook of Mathematical Functions.* New-York: Dover.

ANÉ, T. and GEMAN, H. (2000) Order Flow, Transaction Clock, and Normality of Asset Returns. *Journal of Finance* 55, 2259-2284.

BARLOW, R. E. and PROSCHAN, F. (1975) *Statistical Theory of Reliability and Life Testing Probability Models.* New York: Holt, Reinhart and Winston.

BAUWENS, L. and GIOT, P. (2000) The Logarithmic ACD Model: An Application to the Bid-Ask Quote Process of Three NYSE Stocks. *Annales d'Economie et Statistique* 60, 117-149.

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1993) *Efficient and Adaptive Estimation in Semiparametric Models.* Baltimore: John Hopkins University Press.

BOSQ, D. (1998) *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction.* New York: Springer.

BOUYÉ, E., GAUSSEL, N. and SALMON, M. (2002) Investigating Dynamic Dependence using Copulae. Working Paper, Financial Econometrics Research Center.

CAPÉRAÀ, P. and GENEST, C. (1990) Concepts de dépendance et ordres stochastiques pour des lois bidimensionnelles. *The Canadian Journal of Statistics* 18, 315-326.

CARRASCO, M. and CHEN, X. (2002) Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models. *Econometric Theory* 18, 17-39.

CLARK, P. (1973) A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices. *Econometrica* 41, 135-156.

COX, D. R. (1955) Some Statistical Models Connected with Series of Events (with discussion). *Journal of the Royal Statistical Society Series B* 17, 129-164.

COX, D. R. (1972) The Statistical Analysis of Dependences in Point Processes. In *Symposium on Point Processes*, P. A. W. Lewis (ed), 55-66. New York: John Wiley.

DAROLLES, S., GOURIEROUX, C. and JASIAK, J. (2006) Structural Laplace Transform and Compound Autoregressive Models. *Journal of Time Series Analysis* 27, 477-504.

DARSOW, W., NGUYEN, B. and OLSEN, E. (1992) Copulas and Markov Processes. *Ill. J. Maths* 36, 600-642.

DAVYDOV, Y. (1973) Mixing Conditions for Markov Chains. *Stoch. Proc. Appl.* 18, 312-328.

DOUKHAN, P. (1994) *Mixing. Properties and Examples.* New-York: Springer.

ENGLE, R. (2000) The Econometrics of Ultra-High-Frequency Data. *Econometrica* 68, 1-22.

ENGLE, R. and RUSSELL, J. R. (1998) Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. *Econometrica* 66, 1127-1162.

FANG, Z., HU, T. and JOE, H. (1994) On the Decrease in Dependence with Lag for Stationary Markov Chains. *Probability in the Engineering and Informational Sciences* 8, 385-401.

FEIGIN, P. D. and TWEEDIE, R. L. (1985) Random Coefficients Autoregressive Processes: a Markov Chain Analysis of Stationarity and Finiteness of Moments. *Journal of Time Series Analysis* 6, 1-14.

FELLER, W. (1971) *An Introduction to Probability Theory an Its Applications* (2nd edn). New York: Wiley.

GAGLIARDINI, P. and GOURIEROUX, C. (2007) An Efficient Nonparametric Estimator for Models with Nonlinear Dependence. *Journal of Econometrics* 137, 189-229.

GENEST, C. and WERKER, B. J. M. (2002) Conditions for the Asymptotic Semiparametric Efficiency of an Omnibus Estimator of Dependence Parameter in Copula Models. In *Proceedings of the Conference on Distributions with Given Marginals and Statistical Modelling*, C. M. Cuadras and J. A. Rodríguez Lallena (eds). Dordrecht: Kluwer.

GHYSELS, E., GOURIEROUX, C. and JASIAK, J. (1998) Market Time and Asset Price Movements. Theory and Estimation. In *Statistics in Finance*, D. Hand and S. Jacka (eds). London: Arnold.

GHYSELS, E., GOURIEROUX, C. and JASIAK, J. (2004) Stochastic Volatility Duration Models. *Journal of Econometrics* 119, 413-433.

GHYSELS, E. and JASIAK, J. (1994) Stochastic Volatility and Time Deformation: an Application to Trading Volume and Leverage Effects. Working Paper, C.R.D.E. and Université de Montreal.

GIOT, P. (2001) Time Deformations, Intra-Day Data and Volatility Models. *Journal of Computational Finance* 4, 31-62.

GOURIEROUX, C. and JASIAK, J. (2001a) *Financial Econometrics.* Princeton: Princeton University Press.

GOURIEROUX, C., and JASIAK, J. (2001b) Nonlinear Autocorrelograms: An Application to Inter-Trade Durations. *Journal of Time Series Analysis* 23, 127-154.

GOURIEROUX, C. and ROBERT, C. Y. (2006) Stochastic Unit Root Models. *Econometric Theory* 22, 1052-1090.

HAUTSCH, N. (1999) Analyzing the Time Between Trades with a Gamma Compounded Hazard Model. An Application to LIFFE Bund Future Transactions. Working Paper 99.03, Center of Finance and Econometrics, University of Konstanz.

JASIAK, J. (1998) Persistence in Intertrade Durations. *Finance* 19, 166-195.

JOE, H. (1993) Parametric Family of Multivariate Distributions with Given Margins. *Journal of Multivariate Analysis* 46, 262-282.

JOE, H. (1997) *Multivariate Models and Dependence Concepts.* London: Chapman & Hall.

JONDEAU, E. and ROCKINGER, M. (2006) The Copula-GARCH Model of Conditional Dependencies: An International Stock Market Application. *Journal of International Money and Finance* 25, 827-853.

JONES, M. C. (1993) Simple Boundary Correction for Kernel Density Estimation. *Statistics and Computing* 3, 135-146.

KIMELDORF, G. and SAMPSON, A. R. (1987) Positive Dependence Orderings. *Ann. Inst. Statist. Math.* 39, 113-128.

KIMELDORF, G. and SAMPSON, A. R. (1989) A Framework for Positive Dependence. *Ann. Inst. Statist. Math.* 341, 31-45.

LEHMANN, E. L. (1966) Some Concepts of Dependence. *Ann. Math. Statist.* 37, 1137-1153.

MEYN, S. P. and TWEEDIE, R. L. (1993) *Markov Chains and Stochastic Stability.* New York: Springer.

NELSEN, R. B. (1999) *An Introduction to Copulas.* New-York: Springer.

RICE, J. (1984) Boundary Modification for Kernel Regression. *Communications in Statistics. Theory and Methods* 13, 893-900.

SEVERINI, T. A. and TRIPATHI, G. (2001) A Simplified Approach to Computing Efficiency Bounds in Semiparametric Models. *Journal of Econometrics* 102, 23-66.

SHAKED, M. (1977) A Family of Concepts of Dependence for Bivariate Distributions. *J. Amer. Statist. Assoc.* 72, 642-650.

STOCK, J. (1988) Estimating Continuous Time Processes Subject to Time Deformation. *J. Amer. Statist. Assoc.* 83, 77-85.

TONG, H. (1990) *Non-linear Time Series: A Dynamical System Approach.* Oxford: Oxford University Press.

YANAGIMOTO, T. and OKAMOTO, M. (1969) Partial Orderings of Permutations and Monotonicity of a Rank Correlation Statistic. *Ann. Inst. Statist. Math.* 21, 489-506.

ZHANG, M. Y., RUSSELL, J. R. and TSAY, R. S. (2001) A Nonlinear Autoregressive Conditional Duration Model with Applications to Financial Transaction Data. *Journal of Econometrics* 104, 179-207.

Notes

[1]See Clark (1973), Stock (1988), Ghysels and Jasiak (1994), Ghysels *et al.* (1998), Ané and Geman (2000).

[2]See Engle and Russell (1998), Jasiak (1998), Giot (2001), Gourieroux and Jasiak (2001b), Ghysels *et al.* (2004).

[3]Various extensions of the basic specifications have been considered in the literature. For instance, Jasiak (1998) introduces fractionally integrated ACD (FIACD); Bauwens and Giot (2000) apply the GARCH dynamics to the log-durations and log-expected durations; Zhang *et al.* (2001) introduce a nonlinear dynamics by means of a deterministic threshold autoregression.

[4]In the Cox (1972) model, function a is exponential linear. See Hautsch (1999) for an application to intertrade durations.

[5]See Carrasco and Chen (2002) for mixing properties of ACD models.

[6]The restrictions on functional parameters h and φ to get stationarity are considered later on in this section.

[7]Beyond traditional methods based on autocorrelograms, considerable attention has been devoted in recent years to nonlinear autocorrelograms (e.g. Gourieroux and Jasiak, 2001b), conditional Laplace transforms (e.g. Darolles *et al.*, 2006) and copulas (see Bouyé *et al.*, 2002; Jondeau and Rockinger, 2006, and references therein; Joe, 1997, ch.8; Nelsen, 1999, section 6.3).

[8]A link with the literature on nonlinear autocorrelograms is provided by the fact that condition (SI) implies that any monotonic transformation $h(X_t)$, $t \in \mathbb{N}$, of the process features positive correlation:

$$corr\left[h(X_t), h(X_{t-1})\right] \geq 0.$$

[9]$(X_t, X_{t-1}) \succeq_{(HI)} \left(X_t^*, X_{t-1}^*\right)$, or $(X_t, X_{t-1}) \succeq_{(SI)} \left(X_t^*, X_{t-1}^*\right)$, implies that the Kendall's tau of (X_t, X_{t-1}) is larger than that of $\left(X_t^*, X_{t-1}^*\right)$; moreover, if $(X_t, t \in \mathbb{N})$ and $(X_t^*, t \in \mathbb{N})$ have the same margins, then:

$$corr\left[g\left(X_t\right), g\left(X_{t-1}\right)\right] \geq corr\left[g\left(X_t^*\right), g\left(X_{t-1}^*\right)\right],$$

for any monotonic transformation g such that the correlations exist.

[10]Note that:

$$\frac{1}{A\left[1-\int_0^1 \exp\left(-A(v)y\right) dv\right]} = \frac{1}{A\left[H_0^{-1}(y)\right]},\ y \geq 0,$$

is the conditional expectation of the transformed process $Z_t = H_0(U_t)$, $t \in \mathbb{N}$, with constant conditional hazard.

[11]The symmetric case $v = 0$ is analogous.

[12]Functions A in class I imply autoregressive functions φ such that $\frac{\varphi(y)}{y} \to 1$ as $y \to +\infty$ (see Appendix 5).

[13]The replacement of U_t by $\widehat{U}_t$ does not influence the pointwise asymptotic distribution of a nonparametric estimator of A, but can influence the asymptotic distribution of estimators of linear functionals of A.

[14]For expository purposes, we assume that the bandwidth h_T is such that the bias term can be neglected (see Gagliardini and Gourieroux, 2007, for the analysis of the bias term of the minimum chi-square estimator).

[15]The fact that the pointwise estimator for the baseline hazard function h_0 converges at a parametric rate may seem unusual. This result is due to the restriction of uniform margins for the copula, which implies that h_0 can be expressed as an integral of function A.

[16]With integrable function C.

[17]Indeed, if $\psi(x) = \log E\left[\exp\left(-xZ\right)\right]$, then $\psi^{''}(x) = V_{Q_x}\left[Z\right]$, where distribution Q_x is defined by $dQ_x(z) = \left\{\exp\left(-xz\right)/E\left[\exp\left(-xZ\right)\right]\right\} dF_Z(z)$.

Figure 1: Simulated path for process U_t, $t \in \mathbb{N}$, with proportional hazard and functional dependence parameter A such that $1 - A^{-1}$ is a gamma distribution with parameter $1/\delta$, $\delta = 0.1$.

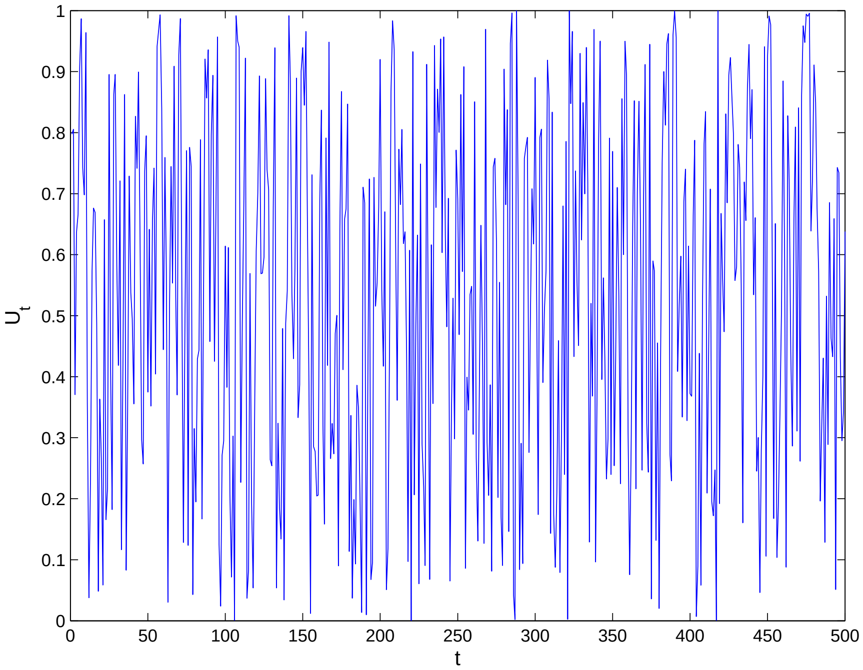

Figure 2: Copula p.d.f. for process U_t, $t \in \mathbb{N}$, with proportional hazard and functional dependence parameter A such that $1 - A^{-1}$ is a gamma distribution with parameter $1/\delta$, $\delta = 0.1$.

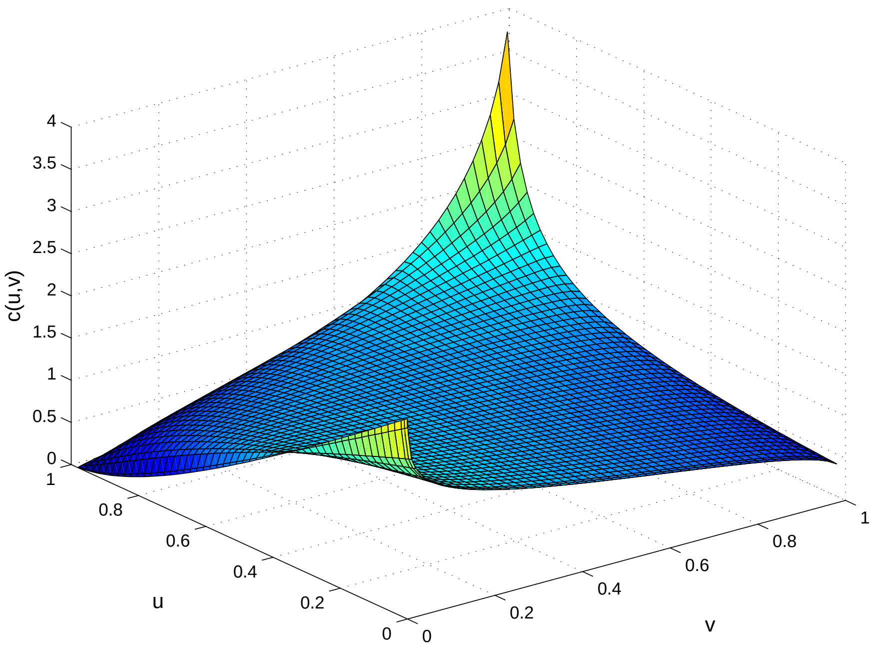

Figure 3: Contour plot of transition p.d.f. for process X_t^*, $t \in \mathbb{N}$, with proportional hazard, functional dependence parameter A such that $1 - A^{-1}$ is a gamma distribution with parameter $1/\delta$, $\delta = 0.1$, and standard Gaussian marginal distribution.

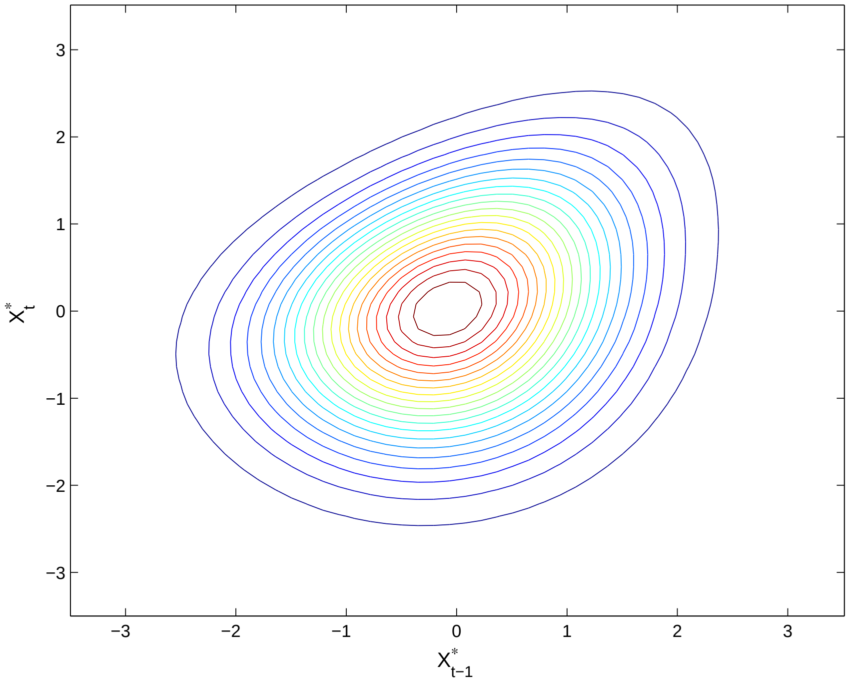

Figure 4: Autocorrelogram for process X_t, $t \in \mathbb{N}$, with functional dependence parameter A such that $1 - A^{-1} \sim \gamma(1/\delta)$, $\delta = 0.1$, and marginal distribution $F(x) = 1 - 1/(1+x)^{\tau}$, $\tau = 5.5$.

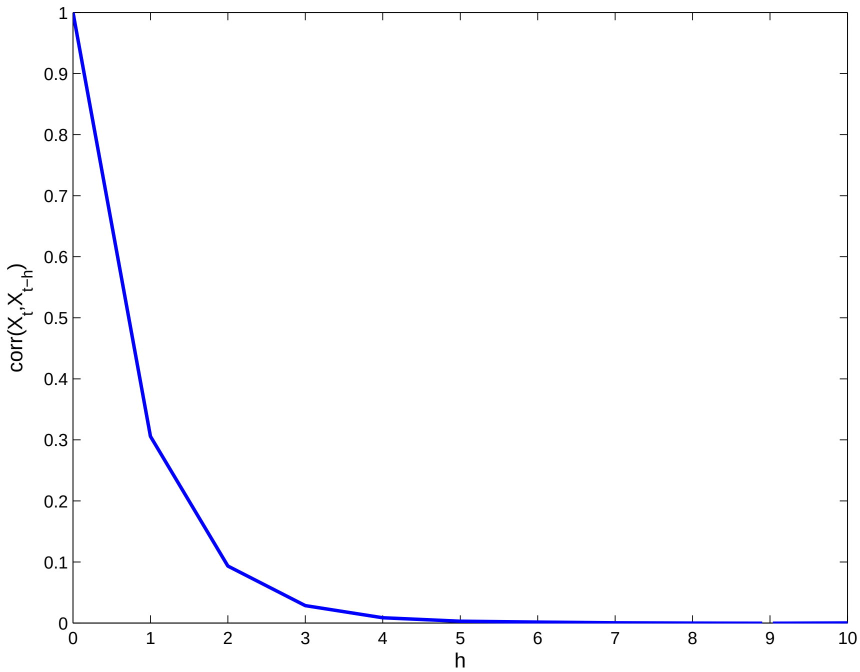

Figure 5: Simulated path for process U_t, $t \in \mathbb{N}$, with proportional hazard and functional dependence parameter A such that $1 - A^{-1}$ is a gamma distribution with parameter $1/\delta$, $\delta = 1$.

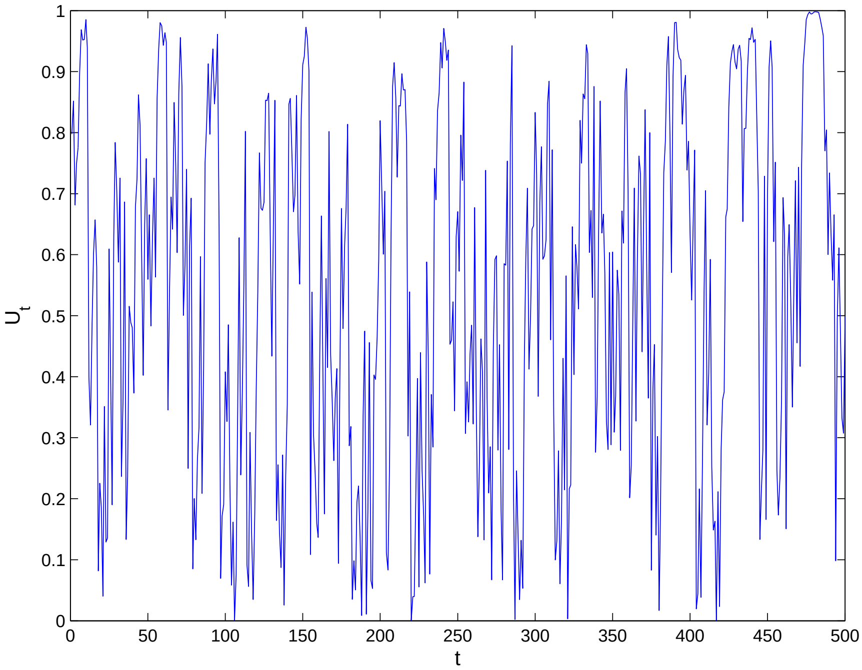

Figure 6: Copula p.d.f. for process U_t, $t \in \mathbb{N}$, with proportional hazard and functional dependence parameter A such that $1 - A^{-1}$ is a gamma distribution with parameter $1/\delta$, $\delta = 1$.

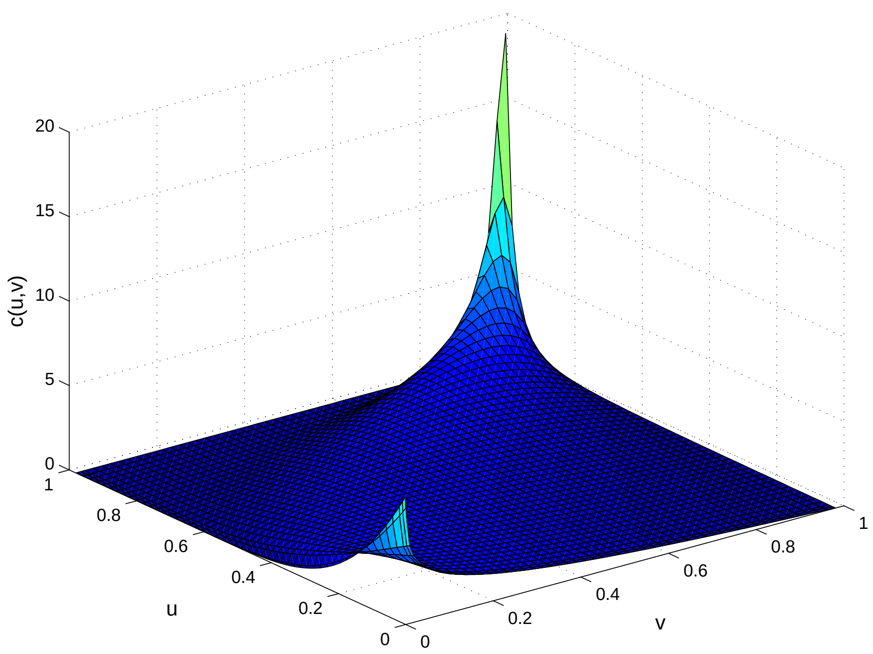

Figure 7: Contour plot of transition p.d.f. for process X_t^*, $t \in \mathbb{N}$, with proportional hazard, functional dependence parameter A such that $1 - A^{-1}$ is a gamma distribution with parameter $1/\delta$, $\delta = 1$, and standard Gaussian marginal distribution.

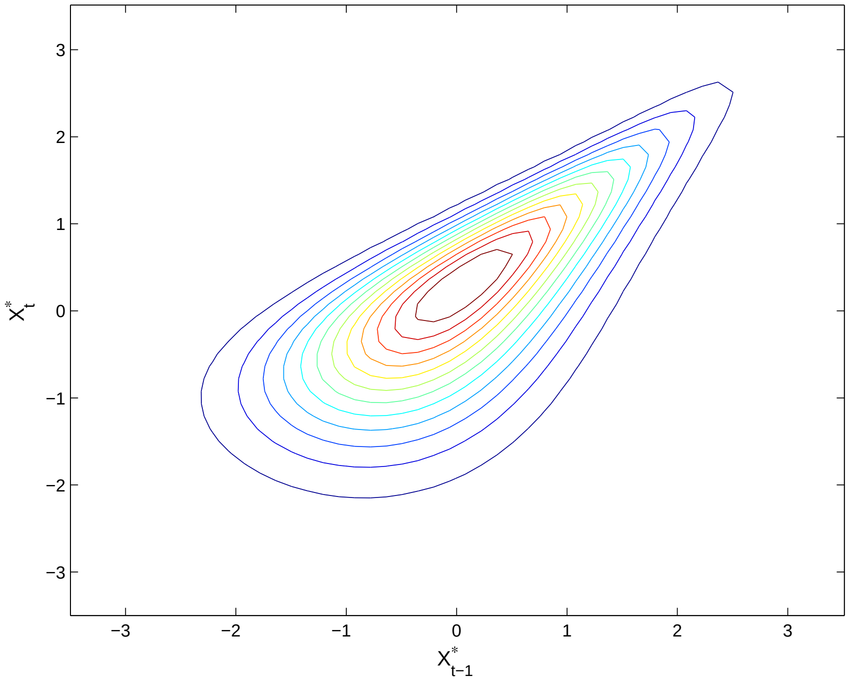

Figure 8: Autocorrelogram for process X_t, $t \in \mathbb{N}$, with functional dependence parameter A such that $1 - A^{-1} \sim \gamma(1/\delta)$, $\delta = 1$, and marginal distribution $F(x) = 1 - 1/(1+x)^{\tau}$, $\tau = 5.5$.

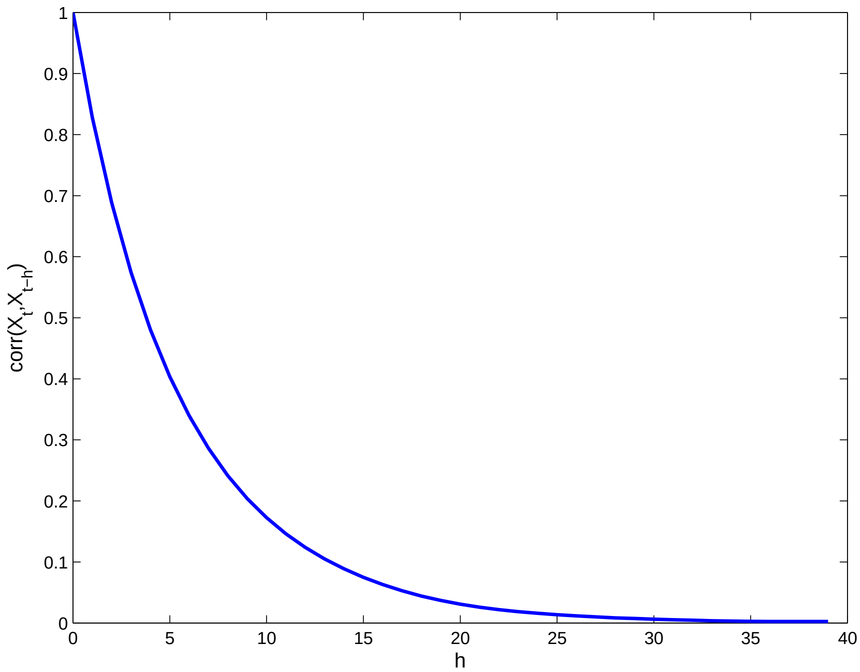

Figure 9: Functional dependence measure for process U_t, $t \in \mathbb{N}$, with $1 - A^{-1} \sim \gamma(1/\delta)$: $\delta = 0.1$ (dashed line), $\delta = 1$ (solid line).

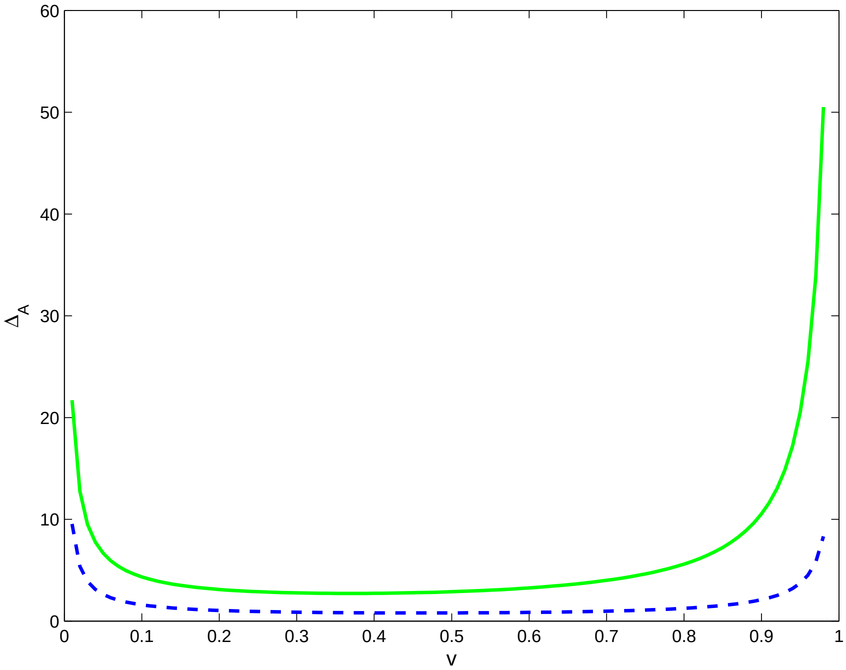

Figure 10: Stepwise functional parameter A (upper Panel) of a Markov process with proportional hazard and endogenous switching regimes, and the corresponding autocorrelogram (lower Panel) for marginal distribution $F(x) = 1 - 1/(1+x)^{\tau}$, $\tau = 5.5$.