INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES
Série des Documents de Travail du CREST
(Centre de Recherche en Economie et Statistique)

n° 2002-20

Constrained Nonparametric Copulas

P. GAGLIARDINI[1]
C. GOURIEROUX[2]

July 2002

---

[1] Università della Svizzera Italiana.
[2] CREST and University of Toronto, Canada.

DURATION TIME SERIES MODELS WITH PROPORTIONAL HAZARD

P. GAGLIARDINI[1], C. GOURIEROUX[2]

May 27, 2002

[1]Università della Svizzera Italiana
[2]CREST and University of Toronto

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 such duration data have been introduced in the literature, the most famous being the ACD (Autoregressive Conditional Duration) model. However 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 potential forms of the stationary distributions are not satisfactorily known. The aim of this paper is to solve these difficulties by considering the properties of a duration time series satisfying the proportional hazard property. We describe in detail this class of dynamic models, discuss various representations, and give 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 they reach the parametric (respectively, nonparametric) efficiency bound.

Keywords: Duration, Copula, ACD model, Nonparametric Estimation, Proportional Hazard, Nonparametric Efficiency.
JEL classification: C14, C22, C41

Modèles dynamiques à hazard proportionnel

Résumé

L'analyse de la liquidité sur les marchés financiers est généralement menée par l'intermédiaire de la dynamique des durées observées entre transactions (éventuellement pondérées par les prix ou les volumes). Des modèles dynamiques pour de telles données de durée, comme le modèle ACD (Autoregressive Conditional Duration) on été introduits dans la littérature. Cependant ces modèles sont souvent trop contraints, conduisant par exemple pour le modèle ACD à lier espérance et variance conditionnelle des durées de façon déterministe. De plus leur propriétés de stationnarité sont mal connues, ainsi que les formes potentielles des distributions stationnaires. Le but de cet article est de résoudre ces difficultés en considérant les propriétés d'une série temporelle de durées satisfaisant l'hypothèse de hazard proportionnel. Nous décrivons en détail cette classe de modèles dynamiques, discutons diverses représentations, et donnons les conditions d'ergodicité. Le copule à hazard proportionnel peut être spécifié sous forme paramétrique ou non paramétrique. Nous discutons des méthodes d'estimation dans ces deux contextes et expliquons pourquoi elles sont efficaces, c'est à dire atteignent les bornes d'efficacité paramétrique et non paramétrique, respectivement.

Mots clés: Durée, Copula, modèle ACD, Estimation Non Paramétrique, Hazard Proportionnel, Efficacité Non Paramétrique.
Classification JEL: C14, C22, C41

1 Introduction.

Series of durations between consecutive trades of a given asset have been recently the object of a considerable body of research in financial econometrics (see e.g. Engle [2000], and Gourieroux and Jasiak [2001]a). The interest in this topic, supported by the increasing availability of (ultra-)high-frequency data, is motivated from a financial point of view along several lines. In addition to the links with microstructure theory and with the literature on stochastic time deformation[1], the dynamics of intertrade durations is an important aspect 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 class of dynamic models for intertrade durations which are 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]. Among the most significant ones are: 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 in some cases featuring long memory; underlying strong nonlinearities in the dynamics, as emerging from the analysis of nonlinear autocorrelograms; path dependent (under-)overdispersion in the conditional distribution; significant departures from exponentiality of the marginal 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 Time at Risk (TaR), that is the minimal time without a trade that may occur with a given probability (see Ghysels, Gourieroux, and Jasiak [1998]b). 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 presently 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, Gourieroux, and Jasiak [1998]b, one limitation of this specification

[1] see Clark [1973], Stock [1988], Ghysels and Jasiak [1994], Ghysels, Gourieroux and Jasiak [1998]a, Ané and Geman [2000], and references therein.

[2] See Engle, Russell [1998], Jasiak [1998], Ghysels, Gourieroux, and Jasiak [1998]b, Giot [2001], Gourieroux, Jasiak [2001]b.

[3] Various extensions of the basic specifications have been considered in the literature. As an example, Jasiak (1998) introduces fractionally integrated ACD (FIACD); Bauwens and Giot (2000) apply the GARCH dynamics on the log-durations and log expected durations; Zhang, Russell and Tsay (2001) introduce a nonlinear dynamics by means of a deterministic threshold autoregression.

is to impose quite restrictive assumptions on the conditional distribution of the duration process. The dynamics of conditional moments of any orders and of measures like TaR_t are all implicitly determined by the dynamics of the conditional mean. These restrictions are not supported by empirical evidence, since they imply for instance 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. As an example, Ghysels, Gourieroux, and Jasiak [1998]b propose the stochastic volatility duration (SVD) model, where conditional mean and conditional variance are allowed to follow independent dynamics due to the introduction of two underlying factors.

In this paper we introduce a Markov process for intertrade durations which is based on a proportional hazard specification. In this model, 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, up to identifiability conditions. This specification improves on the accelerated hazard specification of the ACD model in two directions. First, it provides a flexible specification for the conditional distribution of the duration process, which does not impose 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 in general 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. The second advantage of our specification is that it allows to separate marginal characteristics and dependence properties of the process. Specifically, we show that the bivariate copula between two consecutive durations X_t and X_{t-1} is completely characterized by a univariate functional parameter A (say) on $[0, 1]$. The copula is defined as the c.d.f. of the variables X_t and X_{t-1} after they have been transformed to get uniform marginal distributions on the interval $[0, 1]$. The copula summarizes the serial dependence between X_t and X_{t-1} which is invariant to monotonous transformations. This result implies that our model can be parameterized in terms of the marginal distribution of the process and the 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 serial dependence properties of the process are controlled, by letting its marginal distribution unaltered. We discuss how the shape of function A influences the patterns and the strength of serial dependence in the process, both in the whole distribution and in the tails, by introducing appropriate (functional)

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

concepts and measures of dependence. Specifically, it is shown that the duration process features positive dependence when the functional parameter A is decreasing, whereas its negative elasticity $-d \log A/dv$ can be used as an ordinal measure of serial dependence. In addition, the behaviour of A at $v = 1$ characterizes dependence in the tails of the process, which is responsible for clusterings of extreme large durations. We provide sufficient conditions on the behaviour of functional dependence parameter A in neighborhoods of the boundary points $v = 0$ and $v = 1$ ensuring ergodicity and mixing properties of the process.

The rest of the paper is organized as follows. In section 2 we define the first order Markov process with transition density satisfying the proportional hazard property. In section 3 the temporal 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 is concerned by statistical inference. Finally, section 7 concludes. The proofs are gathered in appendices.

2 Stationary Markov processes 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. The conditional hazard function is the product of a baseline hazard function λ_0 times a positive function a of the lagged duration:

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

Thus the effect of the lagged duration is a parallel 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: $\Lambda_0(x) = \int_0^x \lambda_0(u)du$, $x \geq 0$. Thus the distribution of the process is characterized by two functional parameters: the baseline cumulated hazard Λ_0, which corresponds (up to a multiplicative constant) to 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[5].

The proportional hazard specification satisfies an invariance property with respect to increasing transformations, that is 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.

In this section we are interested in transformations of process X_t, $t \in \mathbb{N}$, which follow autoregressive dynamics. In order to derive them, we consider the nonlinear autoregressive (NLAR) representation with exponential innovations of Markov process $X_t, t \in \mathbb{N}$, (see Tong [1990]), which is 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 $\gamma(1)$. Thus, the duration process $\underline{X_t}, t \in \mathbb{N}$, 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 if we consider another transformation of the duration variable X_t, $t \in \mathbb{N}$. Let us introduce the transformed process:

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

Then we have:

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

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

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

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

[5] The restriction on parameters a and Λ_0 implied by stationarity is derived later in this section.

where η_t, $t \in N$, is a white noise independent of $\underline{Y_{t-1}}$ with a Gompertz distribution.

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

$$\begin{aligned} h(x) &= \log \Lambda_0(x), & x \in [0, \infty), \quad (4) \\ \varphi(y) &= -\log a \left[\Lambda_0^{-1}(\exp y)\right], & y \in (-\infty, \infty). \quad (5) \end{aligned}$$

2.3 The copula representation.

We may also use the invariance property of the proportional hazard specification to obtain processes with given marginal distribution. Indeed, let F be a c.d.f. on $\mathbb{R}_+$ with strictly positive density, and let X_t, $t \in \mathbb{N}$, be a stationary Markov process with proportional hazard and a 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 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 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 given form of the marginal distribution of U_t. Indeed 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 thus the functional parameter A characterizes the distribution of the process $U_t, t \in \mathbb{N}$.

[6] The restriction on functional parameters h and φ, implied by stationarity, is considered later on in this section.

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 unique 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 ones (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 defined as the c.d.f. of the joint distribution of (U_t, U_{t-1}) (see Joe [1997], and 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). The copula summarizes all serial dependence between X_t and X_{t-1}, which is invariant to increasing transformations. Thus, in the proportional hazard model, the copula is characterized completely by a univariate functional parameter A on $[0, 1]$. Copula C_A is called proportional hazard copula.

From (8) and (9) the two 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 uniquely characterized by the two 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 $[0, 1]$, and characterizes the copula of (X_t, X_{t-1}), and hence the serial dependence of the process which is invariant to monotonous transformations[7]. This justifies the interpretation of A as a functional dependence parameter. It is identified up to a

[7] Equations (9) give in explicit form the restrictions on the parameters a and Λ_0 implied by the stationarity.

multiplicative constant. Indeed, from (8) and (10) two functions A which differ by a multiplicative constant define the same copula. The representation in terms of functional parameters (F, A) is called copula representation. It separates marginal characteristics from serial dependence properties of the process.

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

$$\begin{aligned}\varphi(y) &= -\log A\left[1-\int_0^1 \exp\left(-A(v)\exp y\right)dv\right], \; y \in (-\infty,\infty), && (11)\\ h(x) &= \log H_0\left[F(x)\right], \quad x \in [0,\infty). && (12)\end{aligned}$$

Note that 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 Gompertz innovations, where the autoregressive function is restricted by (11) to ensure stationarity.

2.4 Equivalent parameterizations of the copula.

When the functional dependence parameter A is monotonous, equivalent parameterizations of the copula C_A are available. We consider explicitly the case where A is decreasing[8]. Then copula C_A can also be characterized by $1 - A^{-1}$, which is the c.d.f. of the variable $A(U_{t-1})$, that is the transformation of the past transformed duration U_{t-1} having a proportional hazard effect on U_t [9]. In addition, 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)]. In this case it is equivalent to know A or H_0, and thus copula C_A is also characterized by the Laplace transform $1 - H_0^{-1}$, or the cumulated hazard H_0.

Proposition 3 *The copula of a proportional hazard process with monotonically decreasing functional dependence parameter A can be equivalently defined in*

[8] This corresponds to the case where process X_t, $t \in \mathbb{N}$, features positive serial dependence, as will be shown in the next section. The case where A is increasing is similar.

[9] The copula is invariant to scale transformations of the distribution $1 - A^{-1}$.

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 $S_0 \equiv \exp(-H_0)$.*

2.5 An example.

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

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, Stegun [1965]):

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

which has no closed form expression, but can be efficiently computed numerically. Then:

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

where inversion is with respect to the second argument. An analytic expression is available for H_0. Indeed:

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

and 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 modest positive serial dependence, with a tendency to clustering effects, which are stronger 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 $u = v = 1$. Intuitively, the rate of divergence is related to the strength of serial dependence in the tails, and thus to clustering. The asymmetry of the density reveals that the process is not time reversible. The autocorrelogram of duration process $X_t = F^{-1}(U_t)$, $t \in \mathbb{N}$, with Pareto marginal distribution $F(x) = 1 - (1+x)^{-\tau}$, $\tau = 1.05$, based on a simulation of length $S = 35000$ is reported in figure 3.

[insert figure 3: autocorrelogram for X, $\delta = 1/10$]

Let us increase the parameter δ to $\delta = 1$. A simulated trajectory of the process (see figure 4)

[insert figure 4: simulated path for U, $\delta = 1$]

features an increased positive serial dependence, with strong clustering effects, especially at upper boundary. The copula p.d.f. (see figure 5)

[insert figure 5: copula p.d.f., $\delta = 1$]

is more concentrated in a region close to the line $u = v$, and diverges more strongly at the corner points. Note the different limiting behaviour of the copula at the points $u = v = 0$ and $u = v = 1$. The autocorrelogram of corresponding process $X_t = F^{-1}(U_t)$, $t \in \mathbb{N}$, with the same marginal distribution as before, is reported in figure 6.

[insert figure 6: autocorrelogram for X, $\delta = 1$]

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

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[10]. We focus on notions of dependence, which are invariant by increasing transformations and thus involve only the copula.

We first recall two standard notions of positive dependence based on the conditional survivor function and conditional hazard function, respectively. They coincide for stationary processes with proportional hazard, and the condition is

[10] Beyond traditional methods based on autocorrelograms, considerable attention has been devoted in recent years to nonlinear autocorrelograms (see e.g. Gourieroux and Jasiak [2001]b), conditional Laplace transforms (see e.g. Darolles, Gourieroux and Jasiak [2000]) and copulas (see e.g. Bouyé, Gaussel and Salmon [2000], Rockinger and Jondeau [2001] and reference therein; see also chapter 8 in Joe [1997], and section 6.3 of Nelsen [1999]) .

easily 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 introduce functional measures of dependence. Then, we discuss tail dependence properties, and report a sufficient condition which 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)[11]. Moreover both dependence conditions are invariant by increasing transformation of process $(X_t, t \in \mathbb{N})$. In particular they 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. Then 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 the decrease of A (or a).*

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

$$\begin{aligned} \log S(u|v) &= -A(v)H_0(u), \\ \lambda(u|v) &= A(v)h_0(u), \end{aligned}$$

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.

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

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

Thus both notions of positive dependence coincide for proportional hazard models.

Finally, 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 iff the functional dependence parameter A is decreasing.

Corollary 5 *For a stationary Markov process with proportional hazard, the positive dependence (HI) or (SI) is satisfied iff 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 dependence between X_t and X_{t-1} with that between X_t^* and X_{t-1}^*, or equivalently between transformed processes (U_t, U_{t-1}) and $\left(U_t^*, U_{t-1}^*\right)$.

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 $v < v^{'}$, $v, v^{'} \in [0,1]$, 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 conditionally to $U_{t-1} = v$, 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^* is in X_{t-1}^* if for any $v, v^{'} \in [0,1]$, $v < v^{'}$:*

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

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

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

These pre-orderings are denoted by $\succeq_{(SI)}$, $\succeq_{(HI)}$, respectively[12]. They satisfy various axioms, desirable for dependence orderings (see Kimeldorf and

[12] The orderings $\succeq_{(SI)}$ and $\succeq_{(HI)}$ are derived from the (SI) and (HI) concepts of dependence: if X_t^* and X_{t-1}^* are independent, then $(X_t, X_{t-1}) \succeq_{(SI)} \left(X_t^*, X_{t-1}^*\right)$ iff X_t is SI in X_{t-1}, and similarly for $\succeq_{(HI)}$.

Sampson [1987,1989], and Capéràa and Genest [1990] for a discussion). Moreover, since $S_{v,v'}(1)/S^*_{v,v'}(1) = 1$, the ordering $\succeq_{(HI)}$ is stronger than $\succeq_{(SI)}$[13]. 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 for $\left(X_t^*, X_{t-1}^*\right)$. If in addition this is more and more true as we move towards the tail of the distribution, then $(X_t, X_{t-1}) \succeq_{(HI)} \left(X_t^*, X_{t-1}^*\right)$.

For two stationary processes with proportional hazard, $(X_t, t \in \mathbb{N})$ and $(X_t^*, t \in \mathbb{N})$, the following proposition characterizes the orderings in terms of functional dependence parameters A and A^*.

Proposition 6 *Let $(X_t, t \in \mathbb{N})$ and $(X_t^*, t \in \mathbb{N})$ be two stationary Markov processes with proportional hazard and dependence parameters A and A^*, respectively. Then 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*

$$A/A^* \text{ decreasing.}$$

Proof. *See Appendix 1.*

For the proportional hazard model, $\lambda(u \mid v)/\lambda\left(u \mid v'\right)$ is independent of u and is equal to $A(v)/A\left(v'\right)$. This implies that 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:

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

Finally, when the dependence parameters A and A^* are differentiable, the ordering conditions involve the elasticity of the 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 two stationary Markov processes with proportional hazard and differentiable dependence parameters A and A^*, respectively. Then 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:*

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

[13] $(X_t, X_{t-1}) \succeq_{(HI)} \left(X_t^*, X_{t-1}^*\right)$ 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:

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

for any monotonous transformation g such that the correlations exist.

or

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

As an illustration, the functions:

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

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

3.3 Measures of Dependence.

The previous discussion shows that, for the proportional hazard model, the appropriate functional dependence measure is not A itself, but preferably:

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

The properties above can be summarized as follows:

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

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

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

3.4 Tail dependence

In this section we provide sufficient conditions on the functional dependence parameter A that ensure that process X_t, $t\in\mathbb{N}$, features positive dependence in the tails. The coefficient of upper tail dependence is defined by (see Joe [1993], [1997]):

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

If $\lambda>0$, the process is said to have positive tail dependence. 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 *Assume the functional dependence parameter A is such that:*

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

for some $\delta > 0$ and $C > 0$. Then:

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

where P denotes the incomplete gamma function (see the example in section 2.5).

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 still be summarized by its copula, $C_{A,h}$, defined as the joint c.d.f. of U_t, U_{t-h}. By Chapman-Kolmogorov, the copula p.d.f. $c_{A,h}$ is given by:

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

The analytic expression of $c_{A,h}$ is not available in general. However, some dependence properties can be deduced from a theorem by Fang, Hu and Joe (1994). They show that, for a stationary Markov chain $(X_t, t \in \mathbb{N})$, 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 monotonous 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*

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

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 monotonous transformation g such that the correlations exist.

Thus, when A is decreasing, 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 said to be V-geometrically ergodic if there exists $\rho < 1$, a probability measure π and a finite function C such that:*

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

where $\|\mu\|_V = \sup_{f:|f|\leq V} \left| \int f d\mu \right|$.

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. In particular, conditions for geometric ergodicity will involve only either functional dependence parameter A, or functional autoregressive parameter φ. The NLAR representation with additive noise is the most appropriate to discuss geometric ergodicity, since the required drift conditions (see Meyn and Tweedie [1993]) are easy to derive, and have been already extensively investigated in the literature. Equivalent conditions can then 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 Gompertz distributed variable. Then 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 exists 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 exists constants $0 < R_1 < R_2 < \infty$, and $c < \exp(-\gamma) < C$, satisfying:

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

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

Proof. *See Appendix 4.*

Let us briefly discuss the ergodicity conditions[14]. 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 ergodicity condition is intuitive. Note that it is less stringent than the condition which is usually reported in the literature (see e.g. Doukhan [1994] and references therein): $|\varphi(y)+\gamma| \leq \rho\,|y|$, as $|y| \to +\infty$, for some $\rho < 1$. The weakening of the restriction on φ is possible since the 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 the conditions given in ii.[15]. They define restrictions on the dependence parameter A, and specifically on the behaviour of $A(v)$ as $v \to 0$ and $v \to 1$, respectively. They 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 the mass 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 to diverge to infinity or to be absorbed by 0. Let us now investigate these restrictions more precisely and focus on the restriction at $v = 1$[16], when $\lim_{v\to 1} A(v) = 0$. For simplicity, let us consider functions A which 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 $\quad \forall\delta > 0 : \lim_{v\to 1} \frac{A(v)}{(1-v)^\delta} = 0$.

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

[14] A geometric ergodicity condition for the functional parameters (a, Λ_0) in the transition density representation is immediately derived from condition ii. by noting that:

$$\frac{1}{A\left[1-\int_0^1 \exp\left(-A(v)y\right)dv\right]} = \frac{1}{a\left[\Lambda_0^{-1}(y)\right]},\quad y \geq 0.$$

[15] 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.

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

positive and finite[17]. Functions in class II (resp. III) dominate (resp. are dominated by) 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

By discussing mixing properties of a stochastic process we are concerned by 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 e.g. Bosq [1990]). 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.*

[17] In appendix 5 it is shown that functions A in class I imply autoregressive functions φ such that $\frac{\varphi(y)}{y} \to 1$ as $y \to +\infty$.

5 Examples.

In this section we discuss various examples of stationary Markov processes with proportional hazard. The associated dynamic models can be parametric or nonparametric. It is important to note that i) sufficient ergodicity conditions are easily written, ii) the invariant distribution (that 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 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:

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

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 nonzero, 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 $\frac{1}{\alpha w}$, $w \in \Omega$. The inverse of the baseline cumulated hazard H_0 is obtained by computing the Laplace transform of $1 - A^{-1}$:

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

5.2 Analytic examples.

Simple examples can be derived by considering standard distributions for which the Laplace transform admits an analytic expression (see Abramowitz, Stegun [1965] or Joe [1997], Appendix A.1, for an extensive list). In this section we consider only continuous distributions.

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:

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

Then:

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

and the baseline cumulated hazard is:

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

The corresponding copula is:

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

with density:

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

The associated proportional hazard process is geometrically ergodic. Indeed:

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

(see Proposition 11),

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

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 were derived in section 2:

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

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

Various qualitative features featured by the simulations provided in section 2.5 are consequences of the results derived in sections 3 and 4. These processes feature positive dependence since A is decreasing. The functional dependence measure is given by:

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

It is U-shaped and diverges at the boundaries $v = 0$ and $v = 1$ [see figure 7 where $\Delta(.;\delta)$ is plotted for $\delta = 1$ (dashed line) and $\delta = 0.1$ (solid line)].

[insert figure 7: functional dependence parameter]

Since $\Delta(.;1) \geq \Delta(.;0.1)$, the process corresponding to parameter $\delta = 1$ is more dependent.

For $w \sim 0$, we have:

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

and thus:

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

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

iii) Power distributions.

When:

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

with $\delta > 0$, we get:

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

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

The Laplace transform is:

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

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:

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

The functional measure of dependence is given by:

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

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

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

iv) α-stable distributions.

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

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

with $\alpha \geq 1$, and

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

This type of 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 $\alpha = \alpha_c/\alpha_m$. In particular, the larger is parameter α (that is the larger the mass of the distribution $1 - A^{-1}$ in a neighbourhood of 0), the larger is the duration dependence in the marginal distribution with respect to that 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 $J \in \mathbb{N} \cup \{+\infty\}$. Then the conditional distribution is characterized by the survivor function:

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

Thus 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:

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

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:

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

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

i) Uniform series.

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

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

Thus the function A is regularly decreasing and:

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

ii) Power Series.

When:

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

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):

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

with

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

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

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

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

iii) Logarithmic Series.

When:

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

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

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

and:

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

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

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

with

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

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

5.4 Proportional hazard in reversed time.

In this section we consider stationary Markov processes whose distribution features proportional hazard both in the initial and reversed time. The joint density of U_t and U_{t-1} will satisfy:

$$A(u)h_0(v;A)\exp\left[-H_0\left(v;A\right)A(u)\right] = A^*(v)h_0(u;A^*)\exp\left[-H_0\left(u;A^*\right)A^*(v)\right],$$

$u, v \in [0,1]$, for some functions A and A^*.

In Appendix 6 it is shown that the functional dependence parameter of a stationary Markov process with proportional hazard in both time directions is characterized by:

$$A(v) = \frac{\Psi^{-1}\left(\gamma v + \delta\right)}{\Psi^{-1}\left(\delta\right)}, \quad v \in [0,1], \tag{21}$$

where Ψ is a primitive on $\mathbb{R}_+$ of the function $y \mapsto \exp\left(-y\right)/y$, and γ and δ are constants such that:

$$\begin{aligned} \gamma &\geq 0, \\ \gamma + \delta &= \Psi\left(+\infty\right). \end{aligned}$$

In particular, these processes are either independent process ($\gamma = 0$), or processes with negative dependence ($\gamma > 0$). There exist no Markov process which features jointly proportional hazard in both time directions and positive serial dependence.

Since function A^* associated with functional dependence parameter A in (21) is $A^* = A$ (see Appendix 6), these processes satisfy the stronger condition of time reversibility: the density of the process is the same in both time directions, that is the copula is symmetric, $c_A(u,v) = c_A(v,u)$, $u, v \in [0,1]$. There exists no reversible process with proportional hazard and positive serial dependence.

6 Statistical Inference

In this section we assume available observations X_1, ..., X_T, and discuss efficient estimation of the dependence functional, when the marginal distribution F is unconstrained. The functional parameter A can be parametrically specified, or let unconstrained.

In practice it is generally proceeded in two steps. First the marginal c.d.f. can be estimated by its empirical counterpart $\widehat{F}$, say and $\widehat{U}_t = \widehat{F}(X_t)$, $t = 1$, ..., T provide approximations of the uniform variables U_t. $\widehat{U}_t$, $t = 1$, ..., T, are simply the ranks of the variables X_t, $t = 1$, ..., T. In a second step we can 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. Clearly this approach neglects the information on the copula, which 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 let unconstrained[18] [see Genest, Werker (2001), and Gagliardini, Gourieroux (2002) for more precise discussion].

However, since the aim of this section is just 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 the parametric framework, derive the expression of the score and of the efficiency bound. Then we consider the nonparametric estimation of functional parameter A. In section 6.2 we describe two nonparametric estimation methods, that are the minimum chi-square method and the Sieve method. These methods are nonparametrically efficient. We essentially provide the main ideas, which underlie the estimation approaches and the derivation of their asymptotic properties. Detailed proofs can be found in Gagliardini, Gourieroux (2002).

6.1 Parametric framework.

i) General results

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

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

The parameter θ can be estimated by maximum likelihood, that is by:

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

The score $\frac{\partial l_t}{\partial \theta}$ and the Cramer-Rao bound can be expressed in terms of backward conditional expectations. The results below are proved in Appendix 7.

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

[18] More precisely, the replacement of U_t by $\widehat{U}_t$ does not influence the pointwise asymptotic distribution of a nonparametric estimator of A, but it influences the asymptotic distribution of estimators of linear functionals of A.

ii. The Cramer-Rao bound is:

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

where

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

It is interesting to note that the 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:

$$\frac{\partial}{\partial \theta}\log H_0(U_t;\theta) = -E\left[\frac{\partial}{\partial \theta}\log A(U_{t-1};\theta) \mid U_t\right]. \tag{22}$$

ii) Stepwise functional parameter.

Let us consider the endogenous switching regime model (see section 5.3), with a regular grid. The dependence parameter is:

$$A(v;\theta) = \sum_{j=1}^{N} a_j \mathbb{I}_{(\frac{j-1}{N},\frac{j}{N}]}(v), \tag{23}$$

where $\theta = (a_1, a_2, ..., a_N)^{'}$. Let us introduce a vector of indicators $Z_t = (Z_{1t}, ..., Z_{Nt})^{'}$ such that $Z_{jt} = \mathbb{I}_{(\frac{j-1}{N},\frac{j}{N}]}(U_{t-1})$, $j = 1, ..., N$. Then the score is given by:

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

where $diag\,(a)$ is a diagonal matrix with the elements $a_1, a_2, ..., a_N$ on the diagonal. In addition, from equation (22), we deduce that:

$$\frac{\partial}{\partial \theta}\log H_0(U_t;\theta) = -diag(a)^{-1}E\left[Z_{t-1} \mid U_t\right],$$

that is:

$$\frac{\partial}{\partial a_j}\log H_0(U_t;\theta) = -\frac{1}{a_j}P\left[\frac{j-1}{N} < U_{t-1} < \frac{j}{N} \mid U_t\right],\ j = 1, ..., N.$$

Thus the score and the derivatives of the log baseline cumulated hazard are directly related to the backward predictions of the state variables.

In order to identify the model, we impose the following identification constraint on parameter θ [19]:

$$\frac{1}{N}\sum_{j=1}^{N} a_j=1. \tag{24}$$

Then the information matrix is given by:

$$\begin{aligned} I(\theta) &= \left(id_N - \frac{ee'}{N}\right) diag\,(a)^{-1}\, EV\left[\left(1 - A_{t-1}H_{0,t}\right)\left(Z_{t-1} - E\left[Z_{t-1} \mid U_t\right]\right) \mid U_t\right] \\ &\quad \cdot diag\,(a)^{-1}\left(id_N - \frac{ee'}{N}\right), \end{aligned}$$

where $e = (1, ..., 1)'$. In Appendix 10 it is shown that:

$$I(\theta_0) = \frac{1}{N} diag\,(a_0)^{-2} + O_N(\frac{1}{N^2}).$$

Thus, under regularity conditions, the maximum likelihood estimator $\widehat{\theta}_T = (\widehat{a}_{1T}, \widehat{a}_{2T}, ..., \widehat{a}_{NT})'$ under identification constraint (24) is asymptotically normal when T converges to infinity, with asymptotic variance-covariance matrix such that [20]:

$$Cov_{as}\left[\sqrt{T}\left(\widehat{a}_{k,T} - a_{k,0}\right), \sqrt{T}\left(\widehat{a}_{j,T} - a_{j,0}\right)\right] = N\left[a_{j,0}^2 \delta_{k,j} + O_N(1/N)\right]. \tag{25}$$

6.2 Nonparametric estimation methods.

We consider below two estimation methods for the functional A. The first approach considers the constrained nonparametric copula which is the closest to a kernel estimator of the copula for the chi-square proximity measure. The second one is based on a stepwise approximation of function A with the number of terms in the grid tending to infinity.

[19] It is necessary to impose an identification constraint since functions A and kA, where k is a constant, define the same copula (see section 2.3).

[20] In order to get intuition on these results, assume that function $H_0(.) = H_0(.; A_0)$ is known. Define the transformed variables: $W_t = H_0(U_t)$, $t = 1, ..., T$. Then the likelihood of W_t, $t = 1, ..., T$, is the sum of N independent exponential models:

$$\sum_{t=1}^{T} f\left(W_t | W_{t-1}\right) = \sum_{j=1}^{N}\left[\sum_{t=1}^{T}\left(\log a_j - a_j W_t\right)\mathbb{I}_{Z_{jt-1}=1}\right],$$

and $I(\theta_0) = (1/N)\, diag(a_0)^{-2}$ follows.

6.2.1 Minimum chi-square method.

i) Definition of the estimator.

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

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

where K is a kernel, $K_{h_T}(.) = (1/h_T)\, K(./h_T)$, and h_T is a bandwidth tending to 0. Under standard regularity conditions, including strict stationarity of (U_t):

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:

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

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 also asymptotically normal, but at higher nonparametric rate, and parametric rate, respectively:

$$V_{as}\left[\sqrt{Th_T}\int g(u,v)\widehat{c}_T(u,v)du\right] = E_0\left[g\left(U_t, U_{t-1}\right)^2 \mid U_{t-1} = v\right]\int K^2(w)dw, \tag{26}$$

$$V_{as}\left[\sqrt{T}\int\int g(u,v)\widehat{c}_T(u,v)dudv\right] = \sum_{h=-\infty}^{\infty} Cov\left[g\left(U_t, U_{t-1}\right), g\left(U_{t-h}, U_{t-h-1}\right)\right]. \tag{27}$$

The minimum chi-square estimator is defined by:

$$\widehat{A}_T = \min_A \int\int \frac{\left[\widehat{c}_T(u,v) - c(u,v;A)\right]^2}{\widehat{c}_T(u,v)} dudv, \tag{28}$$

where the optimization is performed under the identifying constraint:

$$\int A(v)dv = 1. \tag{29}$$

ii) Asymptotic properties of the estimator

The asymptotic properties of the minimum chi-square estimator $\widehat{A}_T$ defined in

(28) and (29) are reported in Proposition 14 below. In order to formulate this proposition we need some preliminary concepts [see Gagliardini, Gourieroux (2002)]. The derivation of the asymptotic properties of minimum chi-square estimators is based on the possibility to (Hadamard) differentiate 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 7):

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

where:

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

and γ_1 is given in Appendix 7, formula (a.13). Let ν be a measure on $[0,1]$ such that $D\log c\left(.,.;A\right)$ is a bounded linear operator from $L^2(\nu)$ to $L^2\left(P_A\right)$. Let us denote by H the tangent space of $\left\{A\in L^2\left(\nu\right):\int A(v)dv=1\right\}$ at A_0:

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

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:

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

for $g,h\in H$. For the proportional hazard copula the information operator I_H satisfies:

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

where:

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

and α_1 is defined in Appendix 8. The two components of the information operator I_H have different interpretations. The "local" component $\alpha_0(w)$ comes from differentiation of those parts of the density which depend from the value of A at point w, $w\in[0,1]$. The "functional" component α_1 derives from differentiation of those parts of the density which depend from continuous functionals of A.

We are now able to formulate the following Proposition (see Appendix 9).

Proposition 14 *Under standard regularity conditions:*

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

ii. We have the following asymptotic equivalence:

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

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 $(h, r_T)_{L^2(\nu)} \simeq 0$ for any $h \in H$.

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:

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

where P_H is the orthogonal projection on H.

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 $\int g(v)A(v)\nu(dv)$, g varying, which can be consistently estimated at rate $1/\sqrt{T}$ (see e.g. Severini, Tripathi [2001]). The nonparametric efficiency bound $B_A(g)$ is given by (see Gagliardini, Gourieroux [2002]):

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

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

iii) Estimation of H_0^{-1}.

Finally note that $H_0^{-1}(z,A) = 1 - \int_0^1 \exp[-A(v)z]\,dv$ 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:

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

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

Corollary 15 *Under regularity conditions:*

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

In Appendix 7 it is shown that H_0 and h_0 are both differentiable functionals of A. Therefore the corresponding pointwise estimators converge at parametric rate $1/\sqrt{T}$ [21]. 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 only consists in the local component α_0 . The asymptotic variance of $\widehat{A}_T(v)$ is (essentially) its inverse.

6.2.2 Sieve method.

Other nonparametric estimation methods can be considered. For instance it is possible to approximate the function A by a stepwise function: $A(v;\theta) = \sum_{j=1}^N a_j \mathbb{I}_{\left(\frac{j-1}{N},\frac{j}{N}\right]}(v)$, where $\theta=(a_1,...,a_N)$, and to estimate the parameter under the identifiability constraint:

$$\frac{1}{N}\sum_{j=1}^N a_j = 1,$$

which is the analog of (29). For any given N, we get maximum likelihood estimators $\widehat{a}_{j,N}$, $j=1$, ..., N, with properties described in section 6.1. This approach can be extended to a nonparametric framework, if we allow for a number N_T of intervals depending on the number T of observations. If N_T tends to infinity with T at an appropriate rate, this sieve method is expected to provide another nonparametrically efficient estimator of A, rather easy to implement[22].

[21] 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 on uniform margins of the copula, which implies that h_0 can be expressed as an integral of function A.

[22] Rewrite equation (25) with $N=N_T$ as:

$$Cov_{as}\left[\sqrt{T/N_T}\left(\widehat{a}_{k,T}-a_{k,0}\right), \sqrt{T/N_T}\left(\widehat{a}_{j,T}-a_{j,0}\right)\right] = a_{j,0}^2\delta_{k,j} + O_N(1/N_T),$$

and compare with Proposition 14 iii. See also Appendix 10, ii).

7 Conclusion.

In this paper we have introduced duration time series models with proportional hazard. These models allow 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. This has two important consequences from a modelling point of view. On the one hand, the marginal distribution of the process can be chosen freely, and we can then focus on serial dependence by considering function A. On the other hand, since parameter A is functional, this class of models is flexible enough for allowing various nonlinear and nongaussian 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 various characteristics of functional parameter A give rise to different forms of dependence, influence dependence in the tails, and imply ergodicity conditions.

Finally we have discussed the estimation of the dependence parameter A, both in parametric and nonparametric frameworks. A nonparametric 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 and asymptotically normal. In addition it reaches the nonparametric efficiency bound computed under the assumption that the uniform variables U_t are observed.

Appendix 1
Dependence Ordering

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

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

Appendix 2
Coefficient of upper tail dependence

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

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

Using $A\left(v\right) \simeq \left(1 - v\right)^{\delta}$, $v \to 1$, it follows:

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

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:

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

Appendix 3
Nonlinear Autoregressions

In this Appendix we report some probabilistic properties of nonlinear autoregressions 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 $E\left[\eta_t\right] = 0$.

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 [1993]) and aperiodic (see Proposition A1.2 of Tong [1990]).

We assume the autoregression function φ is continuous. Then Y_t, $t \in \mathbb{N}$, is a Feller chain (see Feigin and Tweedie [1993]). Indeed, if V is a bounded, continuous function defined on $\mathbb{R}$, by applying Lebesgue theorem it follows that:

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

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:*

$$|\varphi(y)| \leq |y| - \varepsilon, \quad |y| \geq R,$$

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 $k \in (0, r_0)$ define the functions:

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

We now show that for some k sufficiently small, the 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, Tweedie [1993]). Let us now prove the inequality (a.1). We have:

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

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

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

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

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

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:*

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

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

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

and the result follows. *Q.E.D.*

Finally, let us consider mixing properties. Using the results of Davydov (1973), it is seen that geometric ergodicity[23] implies β-mixing with geometric decay (see e.g. chapter 2.4 in Doukhan [1994]).

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

Appendix 4
Proof of Proposition 10

Condition i. implies geometric ergodicity

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 innovations are Gompertz distributed, 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, 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:

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

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.

[23] with integrable function C

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) depending on the behaviour of the functional dependence parameter at $v = 1$.

Case I: $\lim_{v\to 1} A(v) < \exp[\gamma]$

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 some $r_2 < \infty$, 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 $R_2 = \exp(r_2)$.

Case II: $\lim_{v\to 1} A(v) > \exp[\gamma]$

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 some $r_2 < \infty$, 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: $\lim_{v\to 1} A(v) = \exp[\gamma]$

In this case the 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 will be used in the proof.

Lemma A.3 *Let us assume that function A is strictly positive, 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:

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

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:

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

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

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

then:

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

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

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

Thus, Lebesgue theorem applies:

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

Therefore:

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

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

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

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

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

where γ is the expectation of a Gompertz distributed variable:

$$\gamma=\int_0^\infty (\ln\varepsilon)\exp(-\varepsilon)\,d\varepsilon.$$

The conclusion follows by using the following lemma.

Lemma A.4 *The function*

$$\delta\mapsto\Gamma(1+1/\delta)^\delta,\quad \delta\geq 0,$$

is decreasing, with:

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

Proof. *Define*

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

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

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

is the real LT of the negative of a Gompertz variable, ψ is convex and such that $\psi(0) = 0$ [24]*. It follows:*

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

and the first part of the Lemma is proved. Finally, we 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 assume now that A is in class II, and that there exists $C < \infty$ with:

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

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

$$\begin{aligned}
\int_0^1 \exp\left[-yA(v)\right] dv &\ge \int_{1-K}^1 \exp\left[-yA(v)\right] dv \\
&\ge K\exp\left(-\lambda y\right), \; y \ge 0.
\end{aligned}$$

Since A is decreasing near 1,

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

Then:

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

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

[24] We use that 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 $dQ_x(z) = \left\{\exp\left(-xz\right)/E\left[\exp\left(-xZ\right)\right]\right\} dF_Z(z)$.

Appendix 6
Proportional hazard in reversed time.

The condition for proportional hazard in both time directions is:
$\exists A^*$, H_0^* such that:

$$A(u)h_0(v)\exp\left[-A\left(u\right)H_0(v)\right]=A^*(v)h_0^*(u)\exp\left[-A^*\left(v\right)H_0^*(u)\right], \qquad \text{(a.4)}$$

for $u,v\in[0,1]$. By taking the logarithm of both sides, and deriving with respect to u and v we get:

$$\frac{\partial A(u)}{\partial u}\frac{\partial H_0(v)}{\partial v}=\frac{\partial A^*(v)}{\partial v}\frac{\partial H_0^*(u)}{\partial u}.$$

If U_t is not the independent process, we have:

$$\frac{\partial H_0^*(u)}{\partial u}/\frac{\partial A(u)}{\partial u}=\frac{\partial H_0(v)}{\partial v}/\frac{\partial A^*(v)}{\partial v},\ \ \forall u,v. \qquad \text{(a.5)}$$

Thus these ratios are constant, equal to α (say). It follows, by using $H_0(0)=H_0^*(0)=0$, and the normalizations $A(0)=A^*(0)=1$[25]:

$$\begin{aligned} H_0(v) &= \alpha\left[A^*(v)-1\right],\ \ v\in[0,1], \\ H_0^*(u) &= \alpha\left[A(u)-1\right],\ \ \ u\in[0,1]. \end{aligned}$$

Note in particular that A and A^* are monotonous. By replacing in equation (a.4), we get:

$$\begin{aligned} &\alpha A(u)a^*(v)\exp\left[-\alpha A\left(u\right)A^*(v)\right]\exp\left[\alpha A\left(u\right)\right] \\ =\ &\alpha A^*(v)a(u)\exp\left[-\alpha A^*\left(v\right)A(u)\right]\exp\left[\alpha A^*\left(v\right)\right], \end{aligned}$$

where $a(u)=dA(u)/du$ and $a^*(v)=dA^*(v)/dv$. Thus:

$$\frac{a(u)}{A(u)}\exp\left[-\alpha A\left(u\right)\right]=\frac{a^*(v)}{A^*(v)}\exp\left[-\alpha A^*\left(v\right)\right],\ \ \forall u,v.$$

In particular, function A is such that:

$$\frac{a(u)}{A(u)}\exp\left[-\alpha A\left(u\right)\right]=\gamma, \text{ where } \gamma \text{ is a constant.}$$

Let us denote $y(u)=\alpha A\left(u\right)$, $u\in[0,1]$. Then function y satisfies the separable differential equation:

$$\frac{\exp\left(-y\right)}{y}\frac{dy}{du}=\gamma.$$

[25] These normalizations are admissible, since $\frac{\partial H_0^*(u)}{\partial u}=\alpha\frac{\partial A(u)}{\partial u}$ implies that $\partial A/\partial u$ is integrable, and thus $A(0)<+\infty$, and similarly for $A^*(0)<+\infty$.

Let Ψ be a primitive of the function $y \mapsto \exp(-y)/y$ on $\mathbb{R}_+$. Ψ is continuous, strictly increasing such that $\Psi(+\infty) < +\infty$, and the solution is:

$$y(u) = \Psi^{-1}(\gamma u + \delta), \quad u \in [0,1],$$

where δ is such that:

$$\delta \leq \Psi(+\infty), \text{ if } \gamma \leq 0, \tag{a.6}$$

and

$$\gamma + \delta \leq \Psi(+\infty), \quad \text{if } \gamma > 0. \tag{a.7}$$

Therefore function A is such that:

$$A(u) = \frac{\Psi^{-1}(\gamma u + \delta)}{\Psi^{-1}(\delta)}, \; u \in [0,1],$$

and $\alpha = \Psi^{-1}(\delta)$. Since A^* satisfies the same differential equation as A, we have by symmetry:

$$A^* = A.$$

We now use restriction (8) of uniform margins. The function H_0 and its inverse are given by:

$$\begin{aligned} H_0(u) &= \alpha\left[A^*(u) - 1\right] = \alpha\left[\frac{\Psi^{-1}(\gamma u + \delta)}{\Psi^{-1}(\delta)} - 1\right] \\ &= \Psi^{-1}(\gamma u + \delta) - \Psi^{-1}(\delta), \; u \in [0,1], \end{aligned}$$

and:

$$H_0^{-1}(z) = \frac{1}{\gamma}\left\{\Psi\left[\Psi^{-1}(\delta) + z\right] - \delta\right\}, \; z \geq 0.$$

Thus the restriction is:

$$\frac{1}{\gamma}\left\{\Psi\left[\Psi^{-1}(\delta) + z\right] - \delta\right\} = 1 - \int_0^1 \exp\left[-z\frac{\Psi^{-1}(\gamma v + \delta)}{\Psi^{-1}(\delta)}\right] dv, \; z \geq 0. \tag{a.8}$$

After the change of variable:

$$\xi = \frac{\Psi^{-1}(\gamma v + \delta)}{\Psi^{-1}(\delta)},$$

the integral in the RHS becomes:

$$\begin{aligned} \int_0^1 \exp\left[-z\frac{\Psi^{-1}(\gamma v + \delta)}{\Psi^{-1}(\delta)}\right] dv &= \frac{1}{\gamma}\int_1^{\frac{\Psi^{-1}(\gamma+\delta)}{\Psi^{-1}(\delta)}} \exp\left[-\xi\left(z + \Psi^{-1}(\delta)\right)\right]\frac{d\xi}{\xi} \\ &= \frac{1}{\gamma}\int_{z+\Psi^{-1}(\delta)}^{\frac{\Psi^{-1}(\gamma+\delta)}{\Psi^{-1}(\delta)}\left(z+\Psi^{-1}(\delta)\right)} \frac{\exp(-\xi)}{\xi} d\xi \\ &= \frac{1}{\gamma}\left\{\Psi\left[\frac{\Psi^{-1}(\gamma+\delta)}{\Psi^{-1}(\delta)}\left(z + \Psi^{-1}(\delta)\right)\right] - \Psi\left[z + \Psi^{-1}(\delta)\right]\right\}. \end{aligned}$$

Thus restriction (a.8) becomes:

$$\gamma + \delta = \Psi \left[\Psi^{-1}(\gamma + \delta) + \frac{\Psi^{-1}(\gamma + \delta)}{\Psi^{-1}(\delta)} z \right], \forall z \geq 0.$$

This equation cannot be satisfied with values of δ and γ such that $\Psi^{-1}(\gamma + \delta) < +\infty$ and $\Psi^{-1}(\delta) < +\infty$, but is satisfied if either $\Psi^{-1}(\gamma + \delta) = +\infty$ or $\Psi^{-1}(\delta) = +\infty$ holds. The case $\Psi^{-1}(\delta) = +\infty$ is not admissible. When $\Psi^{-1}(\gamma + \delta) = +\infty$, condition (a.6) cannot be satisfied, whereas (a.7) is trivially satisfied. Thus, any pair of constants δ and γ such that:

$$\gamma \geq 0, \; \gamma + \delta = \Psi(+\infty),$$

satisfies the restriction.

Appendix 7
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, due to 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:

$$\log c(u, v; A) = \log A(v) + \log h_0(u, A) - A(v) H_0(u, A),$$

with respect to functional parameter A. We get:

$$\begin{aligned}
\log c(u, v; A + \delta A) &= \log[A(v) + \delta A(v)] + \log h_0(u, A + \delta A) \\
&\quad - [A(v) + \delta A(v)] H_0(u, A + \delta A) \\
&\simeq \log c(u, v; A) + \frac{\delta A(v)}{A(v)} + \langle D \log h_0(u, A), \delta A \rangle \\
&\quad - H_0(u, A) \delta A(v) - A(v) \langle D H_0(u, A), \delta A \rangle \\
&= \log c(u, v; A) + \frac{1 - A(v) H_0(u, A)}{A(v)} \delta A(v) \\
&\quad + \langle D \log h_0(u, A), \delta A \rangle - A(v) \langle D H_0(u, A), \delta A \rangle,
\end{aligned} \tag{a.9}$$

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

<u>Expression of $DH_0^{-1}(z,A)$</u>

We have:

$$\begin{aligned} H_0^{-1}(z, A+\delta A) &= 1-\int_0^1 \exp\left[-A(v)\,z-\delta A(v)\,z\right] dv \\ &\simeq 1-\int_0^1 \left[1-\delta A(v)\,z\right]\exp\left[-A(v)\,z\right] dv \\ &= H_0^{-1}(z,A)+\int_0^1 z\delta A(v)\exp\left[-A(v)\,z\right] dv, \end{aligned}$$

hence:

$$\left\langle DH_0^{-1}(z,A),\delta A\right\rangle = \int_0^1 z\exp\left[-A(v)\,z\right]\delta A(v)dv.$$

<u>Expression of $DH_0(u;A)$</u>

By applying the implicit function theorem we get:

$$\begin{aligned} \left\langle DH_0(u,A),\delta A\right\rangle &= -h_0(u,A)\left\langle DH_0^{-1}\left(H_0(u,A),A\right),\delta A\right\rangle \\ &= -h_0(u,A)\int_0^1 H_0(u,A)\exp\left[-A(v)\,H_0(u,A)\right]\delta A(v)dv \end{aligned} \tag{a.10}$$

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

We get:

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

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:

$$\begin{aligned}
q\left(u, A+\delta A\right) &= \int_0^1 \left[A(v)+\delta A(v)\right] \exp\left\{-\left[A(v)+\delta A(v)\right] H_0\left(u, A+\delta A\right)\right\} dv \\
&\simeq q(u, A) + \int_0^1 \delta A(v) \exp\left[-A(v) H_0\left(u, A\right)\right] dv \\
&\quad - H_0\left(u, A\right) \int_0^1 \delta A(v) A(v) \exp\left[-A(v) H_0\left(u, A\right)\right] dv \\
&\quad - \left\langle DH_0(u), \delta A\right\rangle \int_0^1 A(v)^2 \exp\left[-A(v) H_0\left(u, A\right)\right] dv \\
&\simeq q(u, A) + \int_0^1 \delta A(v) \left[1 - A(v) H_0\left(u, A\right)\right] \exp\left[-A(v) H_0\left(u, A\right)\right] dv \\
&\quad - \left\langle DH_0(u), \delta A\right\rangle \int_0^1 A(v)^2 \exp\left[-A(v) H_0\left(u, A\right)\right] dv.
\end{aligned}$$

It follows:

$$\begin{aligned}
&\left\langle D \log h_0\left(u, A\right), \delta A\right\rangle = -h_0\left(u, A\right) \left\langle Dq\left(u, A\right), \delta A\right\rangle \\
= \; & -h_0\left(u, A\right) \int_0^1 \delta A(v) \left[1 - A(v) H_0\left(u, A\right)\right] \exp\left[-A(v) H_0\left(u, A\right)\right] dv \\
& + h_0\left(u, A\right) \left(\int_0^1 A(v)^2 \exp\left[-A(v) H_0\left(u, A\right)\right] dv\right) \left\langle DH_0(u), \delta A\right\rangle \\
= \; & -h_0\left(u, A\right) \int_0^1 \delta A(v) \left[1 - A(v) H_0\left(u, A\right)\right] \exp\left[-A(v) H_0\left(u, A\right)\right] dv \\
& - h_0\left(u, A\right)^2 \left(\int_0^1 A(v)^2 \exp\left[-A(v) H_0\left(u, A\right)\right] dv\right) \\
& \cdot \int_0^1 H_0\left(u, A\right) \exp\left[-A\left(v\right) H_0\left(u, A\right)\right] \delta A(v) dv.
\end{aligned} \tag{a.11}$$

Explicit expression of the copula's derivative

By inserting (a.10) and (a.11) into (a.9), we see that the expansion of $\log c(u, v; A)$ is of the form:

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

where:

$$\gamma_0(u, v, A) = \frac{1 - A(v) H_0(u, A)}{A(v)}, \tag{a.12}$$

and:

$$
\begin{aligned}
& \gamma_1(u,v,w;A) \\
= \ & -h_0(u,A)\exp\left[-A(w)H_0(u,A)\right] \\
& \cdot\left\{1-H_0(u,A)\left[A(v)+A(w)-\int_0^1 A(z)^2 h_0(u,A)\exp\left[-A\left(z\right)H_0\left(u,A\right)\right]dz\right]\right\}.
\end{aligned} \tag{a.13}
$$

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

$$
\langle D\log c(u,v;A),\delta A\rangle = \gamma_0(u,v,A)\delta A(v) + \int \gamma_1(u,v,w;A)\delta A(w)dw. \tag{a.14}
$$

ii) Conditional expectations in reverse time.

Various functional derivatives with respect to A can be written as expectations in reverse time. From (a.10) we get:

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

or equivalently:

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

where $H_{0t}=H_0\left(U_t,A\right)$ and $A_{t-1}=A\left(U_{t-1}\right)$. Similarly, from (a.11) we get:

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

Then from (a.9) the score of the model can be written as an expectation error in reverse time:

$$
\begin{aligned}
& \langle D\log c\left(U_t,U_{t-1};A\right),\delta A\rangle \\
= \ & \left(1-A_{t-1}H_{0t}\right)\left(\delta A_{t-1}/A_{t-1}-E\left[\delta A_{t-1}/A_{t-1} \mid U_t\right]\right) \\
& -E\left\{\left(1-A_{t-1}H_{0t}\right)\left(\delta A_{t-1}/A_{t-1}-E\left[\delta A_{t-1}/A_{t-1} \mid U_t\right]\right) \mid U_t\right\}.
\end{aligned} \tag{a.15}
$$

iii) The parametric case.

When function A is parameterized:

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

the score of the model is obtained from (a.15) with:

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

We get:

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

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

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

and:

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

Appendix 8
The information operator

i) The expression of the information operator

Let us derive the information operator I_H. From (a.14) in Appendix 7, the differential $D \log c(.,.; A_0)$ admits a measure decomposition with both a discrete and a continuous part [see Gagliardini and Gourieroux (2002)]. Therefore:

$$(g, I_H h)_{L^2(\nu)} = \int g(v)\alpha_0(v; A_0)h(v)dv + \int g(w)\alpha_1(w, v; A_0)h(v)dwdv, \quad \text{(a.16)}$$

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

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

and:

$$\begin{aligned}
\alpha_1(w, v; A_0) &= \int \gamma_0(u, w; A_0)\gamma_1(u, w, v; A_0)du \\
&\quad + \frac{1}{2}\int \gamma_1(u, y, w; A_0)\gamma_1(u, y, v; A_0)dudy + (w \longleftrightarrow v).
\end{aligned}$$

Let us now derive an expression for $I_H h$, $h \in H$. From (a.16) we get:

$$\int g(w) \left[I_H h(w) \frac{d\nu}{d\lambda}(w) - \alpha_0(w; A_0) h(w) - \int \alpha_1(w, v; A_0) h(v) dv \right] dw = 0, \ \forall g \in H.$$

Thus there exists a constant k such that:

$$I_H h(w) \frac{d\nu}{d\lambda}(w) = \alpha_0(w; A_0) h(w) + \int \alpha_1(w, v; A_0) h(v) dv + k.$$

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

$$\begin{aligned} I_H h(w) &= \frac{\alpha_0(w; A_0)}{d\nu/d\lambda\,(w)} h(w) + \int \frac{\alpha_1(w, v; A_0)}{d\nu/d\lambda\,(w)} h(v) dv \\ &- \left(\int \frac{dw}{d\nu/d\lambda\,(w)} \right)^{-1} \left[\int \left(\frac{\alpha_0(w; A_0) h(w)}{d\nu/d\lambda\,(w)} + \int \frac{\alpha_1(w, v; A_0) h(v)}{d\nu/d\lambda\,(w)} dv \right) dw \right] \\ &\cdot \frac{1}{d\nu/d\lambda\,(w)}. \end{aligned} \tag{a.17}$$

Thus I_H admits the representation:

$$I_H h(w) = \frac{\alpha_{0,H}(w; A_0)}{d\nu/d\lambda\,(w)} h(w) + \int \frac{\alpha_{1,H}(w, v; A_0)}{d\nu/d\lambda\,(w)} h(v) dv, \quad \text{say},$$

with $\alpha_{0,H} = \alpha_0$.

ii) Boundedness and invertibility of I_H

We assume that, for any A, there exists a positive definite matrix $\alpha_H(.; A)$ such that:

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

Further let us introduce the measure ν such that:

$$\forall A : \exists C_A > 0 : C_A \frac{d\nu}{d\lambda}(v) \geq \max \left\{ \frac{1}{A(v)^2}, \alpha_H(v; A) \right\}, \ \forall v.$$

Then, from Proposition 22 in Gagliardini, Gourieroux (2002), I_H is a bounded operator from H in itself. Let us now consider the invertibility of I_H. We first show that the differential $D \log c\,(., .; A_0)$ has a zero null space. Indeed let us consider a function $h \in H$ such that:

$$\langle D \log c\,(U_t, U_{t-1}; a_0)\,, h \rangle = 0 \text{ a.s.}$$

Then by using the differential of the proportional hazard copula [see section 3.2 iv)], we deduce that:

$$
\begin{aligned}
& \left(1-a_{0t-1}H_{0t}\right)\left(h_{t-1}/a_{0t-1}-E\left[h_{t-1}/a_{0t-1}\mid U_t\right]\right) \\
= \ & \left[1-a_0\left(U_{t-1}\right)H_0\left(U_t\right)\right]\left\{h\left(U_{t-1}\right)/a_0\left(U_{t-1}\right)-E\left[h\left(U_{t-1}\right)/a_0\left(U_{t-1}\right)\mid U_t\right]\right\} \\
& \text{ is a function of } U_t \text{ only.}
\end{aligned}
$$

This implies that h/a_0 is a constant. Since $\int_0^1 h(v)dv = 0$, it follows that $h = 0$. Thus I_H has zero null space and it is positive.

Let us assume that ν is such that:

$$\forall A : \exists \widetilde{C}_A > 0 : \widetilde{C}_A \frac{d\nu}{d\lambda}(w) \leq \frac{1}{A\left(w\right)^2}, \ \forall w.$$

Then Proposition 22 in Gagliardini, Gourieroux (2002) implies that I_H is invertible.

Appendix 9
Asymptotic distributions

In this appendix we derive the asymptotic distribution of the minimum chi-square estimator reported in Proposition 14. To prove the result we use Proposition 23 in Gagliardini, Gourieroux (2002).

i) The efficient score ψ_T.

The efficient score $\psi_T \in L^2\left(\nu\right)$ is defined by:

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

From Gagliardini, Gourieroux (2002) we get:

$$\frac{d\nu}{d\lambda}(w)\psi_T(w) = \int \delta\widehat{c}_T(w,v)\gamma_0\left(w,v\right)dv + \int\int \delta\widehat{c}_T(u,v)\gamma_1\left(u,v,w\right)dudv. \tag{a.18}$$

ii) The first order condition.

From Gagliardini, Gourieroux (2002) the first order condition is given by:

$$I_H \delta\widehat{A}_T \simeq P_H\psi_T,$$

where P_H is the orthogonal projection on the tangent space H, which is given by:

$$P_H h(v) = h(v) - \left(\int \frac{dw}{d\nu/d\lambda\,(w)} \right)^{-1} \left(\int h(w) dw \right) \frac{1}{d\nu/d\lambda\,(v)}.$$

From (a.17) and (a.18) we get:

$$\begin{aligned}
& \alpha_0(w)\delta\widehat{A}_T(w) + \int \alpha_1(w,v)\delta\widehat{A}_T(v)dv \\
& - \left(\int \frac{dw}{d\nu/d\lambda\,(w)} \right)^{-1} \int \left(\frac{\alpha_0(w)}{d\nu/d\lambda\,(w)} \delta\widehat{A}_T(w) + \int \frac{\alpha_1(w,v)}{d\nu/d\lambda\,(w)} \delta\widehat{A}_T(v)dv \right) dw \\
\simeq \ & \int \delta\widehat{c}_T(w,v)\gamma_0\,(w,v)\,dv + \int\int \delta\widehat{c}_T(u,v)\gamma_1\,(u,v,w)\,dudv \\
& - \left(\int \frac{dw}{d\nu/d\lambda\,(w)} \right)^{-1} \int \frac{1}{d\nu/d\lambda\,(w)} \left(\int \delta\widehat{c}_T(w,v)\gamma_0\,(w,v)\,dv + \right. \\
& \left. \int\int \delta\widehat{c}_T(u,v)\gamma_1\,(u,v,w)\,dudv \right) dw,
\end{aligned} \tag{a.19}$$

which is the asymptotic expansion reported in Proposition 14 ii.

iii) Pointwise asymptotic distribution

Let us consider the pointwise asymptotic distribution of the minimum chi-square estimator $\widehat{A}_T$. Intuitively it can be derived from the asymptotic expansion (a.19), by noting that the second and third terms in the RHS are $O_p\left(1/\sqrt{T}\right)$ [see (27)], and similar orders are expected for the second and third terms in the LHS, leading to:

$$\sqrt{Th_T}\delta\widehat{A}_T\,(v) \simeq \alpha_0\,(v)^{-1} \int \delta\widehat{c}_T(w,v)\gamma_0\,(w,v)\,dv.$$

From (26) it follows:

$$\sqrt{Th_T}\delta\widehat{A}_T\,(v) \xrightarrow{d} N\left(0, \alpha_0\,(v)^{-1} \int K^2(w)dw\right), \ \ \lambda\text{-a.s. in } v \in [0,1]\,.$$

The complete proof of this result is given in Proposition 23 of Gagliardini, Gourieroux (2002).

iv) Asymptotic distribution of linear functionals of A

The asymptotic distribution of linear functionals $\int g(v)A(v)\nu\,(dv)$, $g \in L^2\,(\nu)$ is derived from Proposition 23 of Gagliardini, Gourieroux (2002).

Appendix 10
The efficiency bounds for the stepwise model

i) Determination of the parametric efficiency bound

We have:

$$\begin{aligned} I(\theta_0) &= \left(id_N - \frac{ee'}{N} \right) diag(a_0)^{-1} E_0 V_0 \left[\xi_t \left(Z_{t-1} - E\left[Z_{t-1} \mid U_t \right] \right) \mid U_t \right] \\ &\quad \cdot diag(a_0)^{-1} \left(id_N - \frac{ee'}{N} \right), \end{aligned}$$

where $\xi_t = 1 - A_{0t-1} H_{0,t} \sim iid(0,1)$, ξ_t independent of U_{t-1} [see equation (2)]. Let us transform the terms in the conditional variance. We have:

$$\begin{aligned} & E_0 V_0 \left[\xi_t \left(Z_{t-1} - E_0\left[Z_{t-1} \mid U_t \right] \right) \mid U_t \right] \\ = \; & E_0 E_0 \left[\xi_t^2 \left(Z_{t-1} - E_0\left[Z_{t-1} \mid U_t \right] \right) \left(Z_{t-1} - E_0\left[Z_{t-1} \mid U_t \right] \right)' \mid U_t \right] \\ & - E_0 \left\{ E_0 \left[\xi_t \left(Z_{t-1} - E_0\left[Z_{t-1} \mid U_t \right] \right) \mid U_t \right] E_0 \left[\xi_t \left(Z_{t-1} - E_0\left[Z_{t-1} \mid U_t \right] \right) \mid U_t \right]' \right\} \\ = \; & \underbrace{E_0\left[\xi_t^2\right]}_{=1} \underbrace{E_0\left[diag(Z_{t-1}) \right]}_{Id_N/N} - E_0 \left\{ E_0 \left[\xi_t^2 Z_{t-1} \mid U_t \right] E_0 \left[Z_{t-1} \mid U_t \right]' \right\} \\ & - E_0 \left\{ E_0 \left[Z_{t-1} \mid U_t \right] E_0 \left[\xi_t^2 Z_{t-1} \mid U_t \right]' \right\} \\ & + E_0 \left[E_0 \left[\xi_t^2 \mid U_t \right] E_0 \left[Z_{t-1} \mid U_t \right] E_0 \left[Z_{t-1} \mid U_t \right]' \right] \\ & - E_0 \left\{ \left(E_0 \left[\xi_t Z_{t-1} \mid U_t \right] - E_0 \left[\xi_t \mid U_t \right] E_0 \left[Z_{t-1} \mid U_t \right] \right) \right. \\ & \left. \left(E_0 \left[\xi_t Z_{t-1} \mid U_t \right] - E_0 \left[\xi_t \mid U_t \right] E_0 \left[Z_{t-1} \mid U_t \right] \right)' \right\}. \end{aligned}$$

An expression for the parametric efficiency bound $B(\theta_0) = I(\theta_0)^{-1}$ follows. Let us investigate its expansion for large N, and develop it in powers of $1/N$. By using:

$$\begin{aligned} \xi_t Z_{t-1} &= \left[Id_N - diag(a_0) H_{0t} \right] Z_{t-1}, \\ \xi_t^2 Z_{t-1} &= \left[Id_N - diag(a_0) H_{0t} \right]^2 Z_{t-1}, \\ E_0 \left[\xi_t \mid U_t \right] &= S' \left[Id_N - diag(a_0) H_{0t} \right] E_0 \left[Z_{t-1} \mid U_t \right], \\ E_0 \left[\xi_t^2 \mid U_t \right] &= S' \left[Id_N - diag(a_0) H_{0t} \right]^2 E_0 \left[Z_{t-1} \mid U_t \right], \end{aligned}$$

and the fact that:

$$E_0 \left[Z_{t-1} \mid U_t \right]' x = O(1/N),$$

for any vector x which is not a constant vector, we get:

$$I(\theta_0) = \frac{1}{N} diag(a_0)^{-2} + \frac{1}{N^2} M + o\left(\frac{1}{N^2}\right),$$

where M is a $N \times N$ matrix. Thus the parametric efficiency bound for the stepwise model is such that:

$$B(\theta_0) = N\left[diag(a_0)^2 + O(1/N)\right].$$

The asymptotic distribution for the maximum likelihood estimator $\widehat{\theta}_T = (\widehat{a}_{1,T}, ..., \widehat{a}_{N,T})$ follows:

$$Cov_{as}\left[\sqrt{T}\left(\widehat{a}_{k,T} - a_{k,0}\right), \sqrt{T}\left(\widehat{a}_{j,T} - a_{j,0}\right)\right] = N\left[a_{j,0}^2 \delta_{k,j} + O_N(1/N)\right]. \tag{a.20}$$

ii) Pointwise asymptotic distribution.

A pointwise estimator of A can be defined by:

$$\widehat{A}_T(v) = \sum_{j=1}^{N} \widehat{a}_{j,T} \mathbb{I}_{\left(\frac{j-1}{N}, \frac{j}{N}\right]}(v).$$

We deduce from (a.20) the asymptotic variance of the estimator $\widehat{A}_T(v)$, where T tends to infinity and $N = N_T$ tends to infinity at a much smaller rate:

$$\begin{aligned} & Cov_{as}\left[\sqrt{\frac{T}{N_T}}\left(\widehat{A}_T(v) - A_0(v)\right), \sqrt{\frac{T}{N_T}}\left(\widehat{A}_T(w) - A_0(w)\right)\right] \\ = & \sum_{i=1}^{N}\sum_{j=1}^{N} \mathbb{I}_{\left(\frac{i-1}{N}, \frac{i}{N}\right)}(v)\, \mathbb{I}_{\left(\frac{j-1}{N}, \frac{j}{N}\right)}(w)\left[a_i^2 \delta_{i,j} + o(1/N_T)\right] \\ \simeq & \begin{cases} A_0(v)^2, & v = w \\ 0 & v \neq w \end{cases}. \end{aligned}$$

This result can be directly compared with the pointwise asymptotic distribution of the minimum chi-square estimator given in Proposition 14 iii.

REFERENCES

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

Ané, T., and H., Geman, (2000): "Order Flow, Transaction Clock, and Normality of Asset Returns", Journal of Finance, 55, 2259-2284.

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

Bauwens, L., and P., Giot, (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 J. A., Wellner, (1993): *Efficient and Adaptive Estimation in Semiparametric Models,* John Hopkins University Press, Baltimore.

Bosq, D., (1990): *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction*, Lecture Notes in Statistics, 110, Springer.

Bouyé, E., Gaussel, N., and M., Salmon, (2000): "Investigating Dynamic Dependence using Copulae", Working Paper.

Capéraà, P., and C., Genest, (1990): "Concepts de dépendance et ordres stochastiques pour des lois bidimensionnelles", The Canadian Journal of Statistics, 18, 315-326.

Carrasco, M., and X., Chen, (1999): "β-mixing and Moment Properties of Various GARCH, Stochastic Volatility and ACD Models", CREST Working Paper Series.

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,* ed. by P. A. W. Lewis. New York, John Wiley, 55-66.

Darolles, S., Gourieroux, C., and J., Jasiak, (2000): "Compound Autoregressive Models", CREST DP.

Davydov, Y., (1973): "Mixing Conditions for Markov Chains", Stoch. Proc. Appl., 18, 312-328.

Doukhan, P., (1994): *Mixing. Properties and Examples,* Lecture Notes in Statistics, 85, Springer.

Engle, R., (2000): "The Econometrics of Ultra-High-Frequency Data", Econometrica, 68, 1-22.

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

Fang, Z., Hu, T., and H., Joe, (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 R. L., Tweedie, (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,* Second Edition, Vol. 2., Wiley.

Gagliardini, P., and C., Gourieroux, (2002): "Constrained Nonparametric Copulas", CREST Working Paper.

Genest, C., and B. J. M., Werker, (2001): "Conditions for the Asymptotic Semiparametric Efficiency of an Omnibus Estimator of Dependence Parameter in Copula Models", forthcoming, *Proceedings of the Conference on Distributions with Given Marginals and Statistical Modelling*, C. M. Cuadras and J. A. Rodríguez Lallena (eds).

Ghysels, E., Gourieroux, C., and J., Jasiak, (1998)a: "Market Time and Asset Price Movements. Theory and Estimation", in Hand, D., and S., Jacka, eds., *Statistics in Finance,* Arnold, London.

Ghysels, E., Gourieroux, C., and J., Jasiak, (1998)b: "Stochastic Volatility Duration Models", forthcoming Journal of Econometrics.

Ghysels, E., and J., Jasiak, (1994): "Stochastic Volatility and Time Deformation: an Application to Trading Volume and Leverage Effects", C.R.D.E., Université de Montreal, Unpublished Manuscript.

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

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

Gourieroux, C., and J., Jasiak, (2001)b: "Nonlinear Autocorrelograms; An Application to Inter-Trade Durations", Journal of Time Series Analysis, 23, 127-154.

Gourieroux, C., and J., Jasiak, (2001)c: "Count Data Models for Time Series", CREST DP.

Hautsch, N., (1999): "Analyzing the Time Between Trades with a Gamma Compounded Hazard Model. An Application to LIFFE Bund Future Transactions", Center of Finance and Econometrics, University of Konstanz, Unpublished Manuscript.

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*, Monograph on Statistics and Applied Probability, 73, Chapman & Hall.

Kimeldorf, G., and A. R., Sampson, (1987): "Positive Dependence Orderings", Ann. Inst. Statist. Math., 39, 113-128.

Kimeldorf, G., and A. R., Sampson, (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 R. L., Tweedie, (1993): *Markov Chains and Stochastic Stability*, Communications and Control Engineering Series, Springer.

Nelsen, R. B., (1999): *An Introduction to Copulas*, Lecture Notes in Statistics, 139, Springer.

Rockinger, M., and E., Jondeau, (2001): "Conditional Dependency of Financial Series: An Application of Copulas", Working paper, Banque de France.

Severini, T. A., and G., Tripathi, (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 University Press, Oxford.

Yanagimoto, T., and M., Okamoto, (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 R. S., Tsay, (2001): "A Nonlinear Autoregressive Conditional Duration Model with Applications to Financial Transaction Data", Journal of Econometrics, 104, 179-207.

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 $\delta = 0.1$.

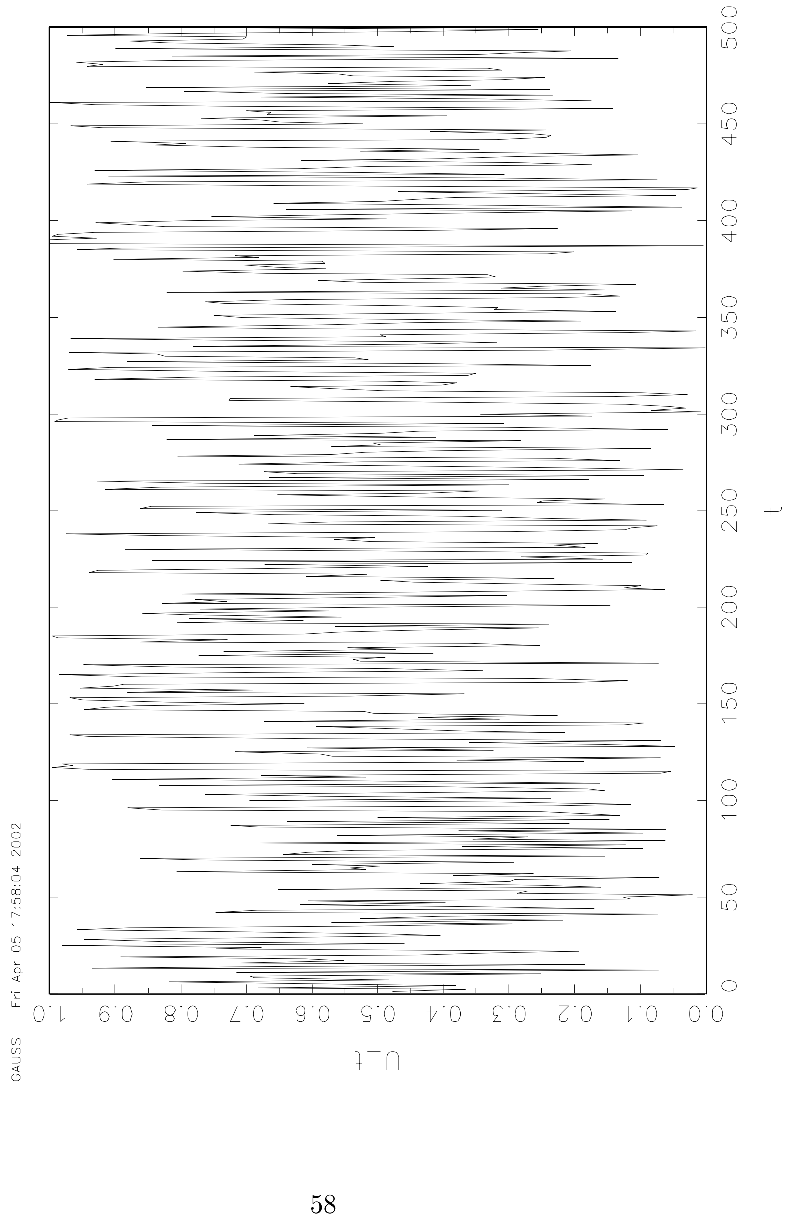

Figure 2: Copula p.d.f. for process U_t, $t \in \mathbb{N}$, with proportional hazard and functional dependence parameter A such that $1 - A^{-1}$ is a gamma distribution with parameter $\delta = 0.1$.

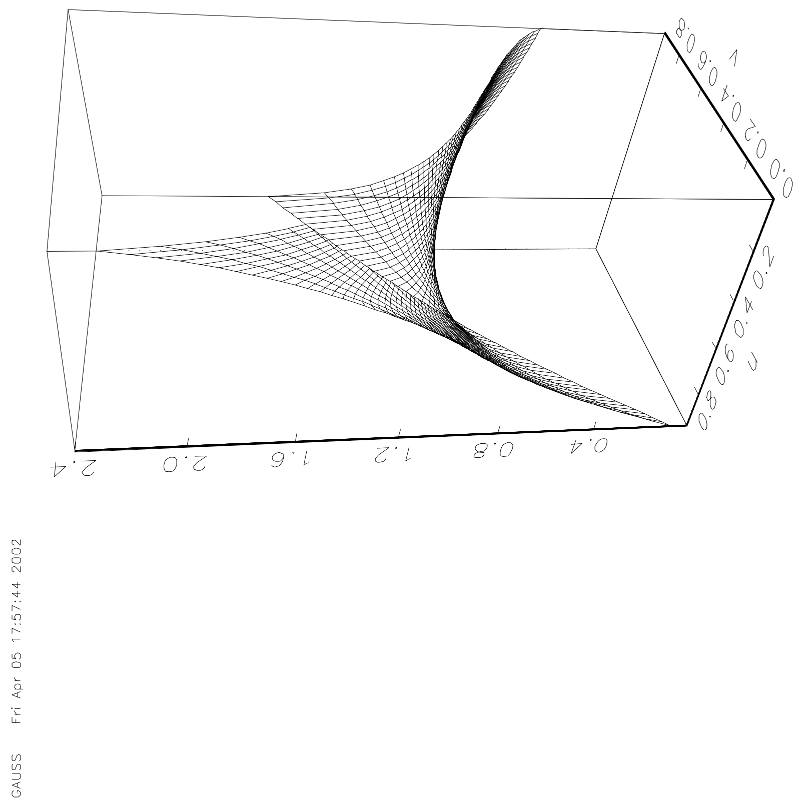

Figure 3: Autocorrelogram for process X_t, $t \in \mathbb{N}$, with functional dependence parameter A such that $1 - A^{-1} \sim \gamma(\delta)$, $\delta = 0.1$, and marginal distribution $F(x) = 1 - (1+x)^{\tau}$, $\tau = 1.05$.

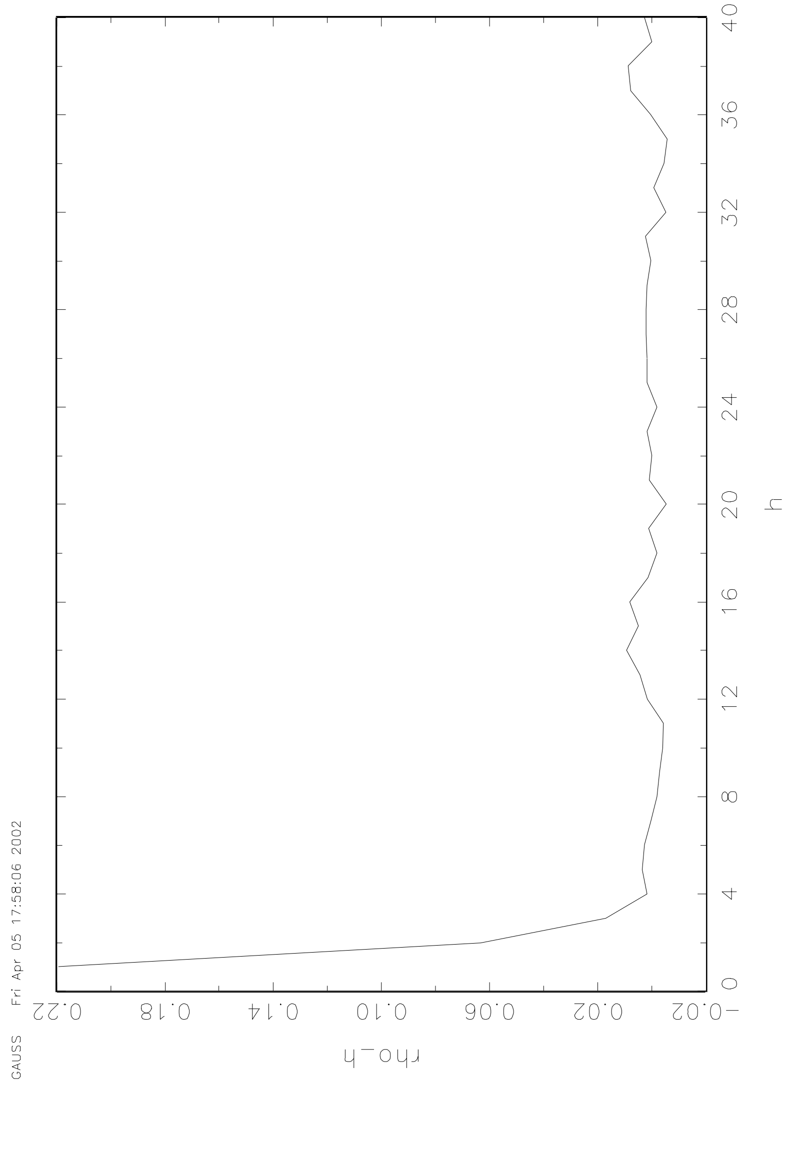

Figure 4: Simulated path for process U_t, $t \in \mathbb{N}$, with proportional hazard and functional dependence parameter A such that $1 - A^{-1}$ is a gamma distribution with parameter $\delta = 1$.

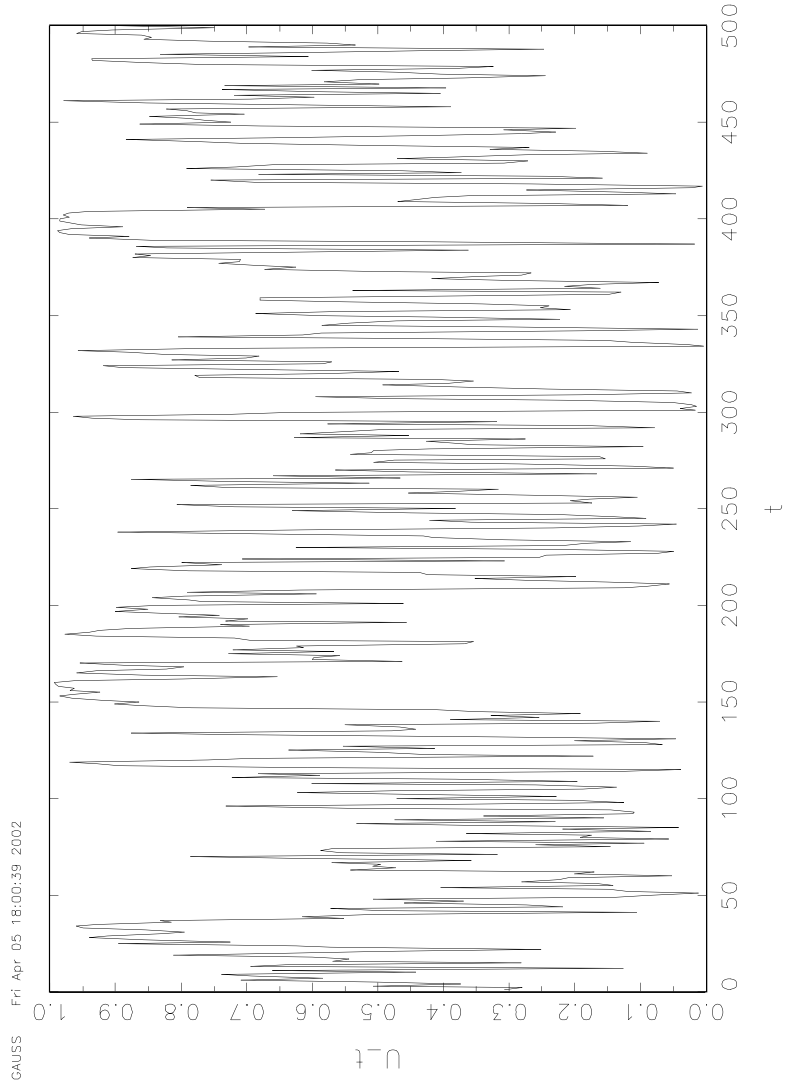

Figure 5: Copula p.d.f. for process U_t, $t \in \mathbb{N}$, with proportional hazard and functional dependence parameter A such that $1 - A^{-1}$ is a gamma distribution with parameter $\delta = 1$.

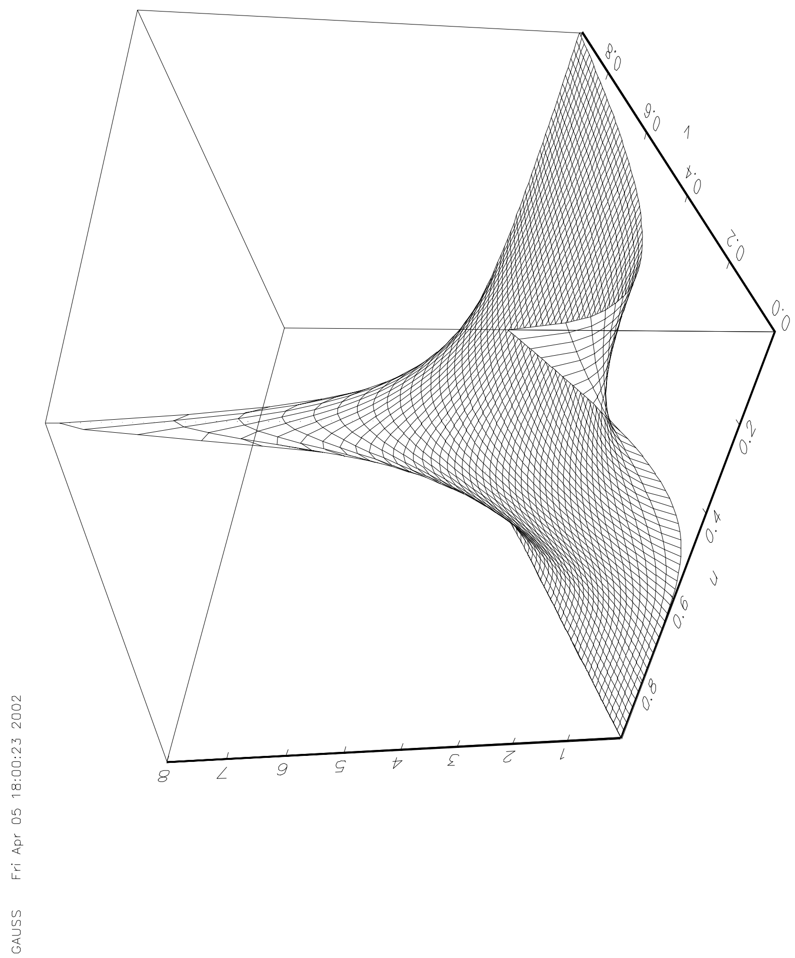

Figure 6: Autocorrelogram for process X_t, $t \in \mathbb{N}$, with functional dependence parameter A such that $1 - A^{-1} \sim \gamma(\delta)$, $\delta = 1$, and marginal distribution $F(x) = 1 - (1 + x)^{\tau}$, $\tau = 1.05$.

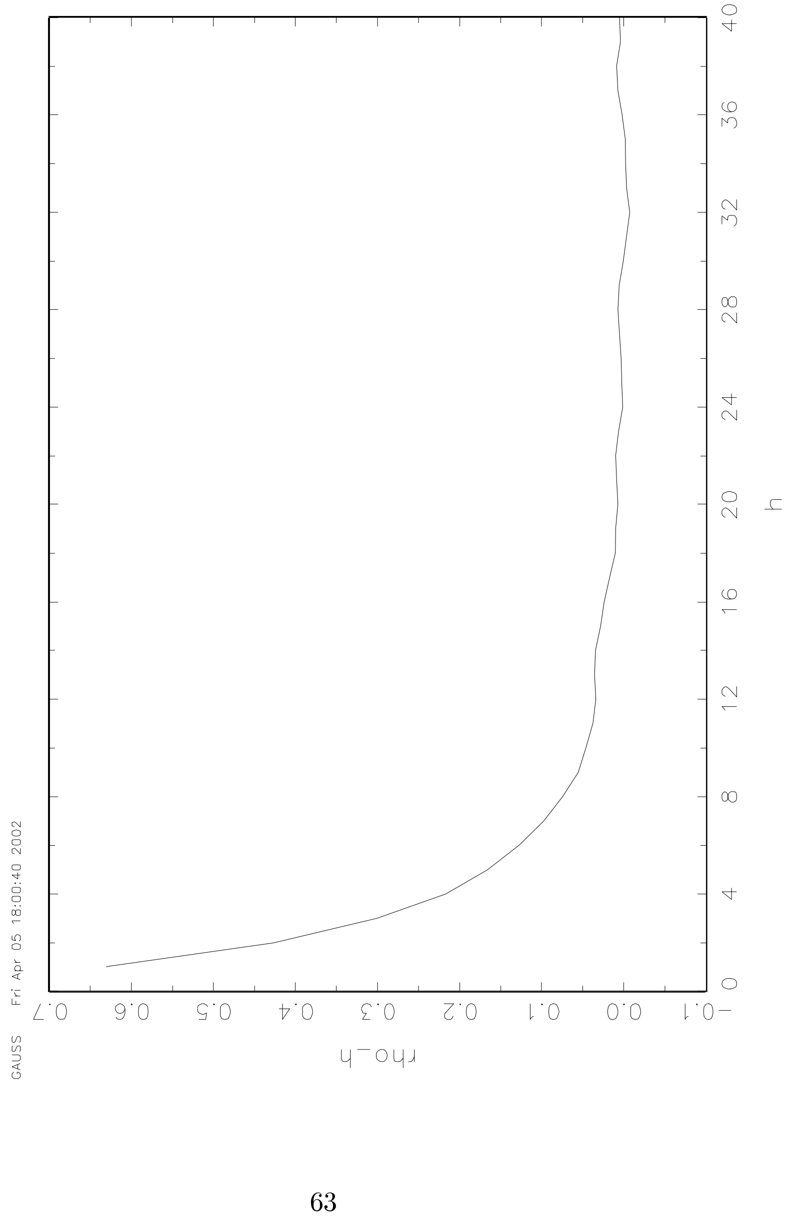

Figure 7: Functional dependence measure for process U_t, $t \in \mathbb{N}$, with $1 - A^{-1} \sim \gamma(\delta)$: $\delta = 0.1$ (solid line), $\delta = 1$ (dashed line).