

UNIVERSITÀ DELLA SVIZZERA ITALIANA
FACOLTÀ DI SCIENZE ECONOMICHE

CONSTRAINED NONPARAMETRIC
DEPENDENCE WITH APPLICATIONS IN
FINANCE

A thesis submitted by

PATRICK GAGLIARDINI

for the degree of

DOTTORE IN SCIENZE ECONOMICHE

January 2003

Members of the jury:

Prof. G. Barone Adesi, President of the jury

Prof. C. Gouriéroux, External member and thesis codirector

Prof. E. Ronchetti

Prof. P. Balestra, thesis codirector

Alla mia famiglia

Acknowledgements

I wish to express my deep gratitude to my thesis directors, Prof. Pietro Balestra and Prof. Christian Gouriéroux, for their invaluable help and for all they have given to me, both at a scientific and at a human level. Without their constant support and unfailing guidance, I would not have overcome most difficulties or indeed enjoyed many exciting moments.

It is a pleasure and an honour for me to thank Prof. Giovanni Barone Adesi and Prof. Elvezio Ronchetti for accepting to sit on the jury, and for their precious comments and suggestions.

I wish to thank the teaching and research staff of the Faculty of Economics, Lugano, especially Fabio Trojani and all the colleagues and friends at the Institute of Finance. I very much appreciated all the stimulating discussions we had in countless occasions.

I also wish to remember the researchers of the Laboratoire de Finance et Assurance at CREST, for their very kind hospitality during my visiting periods in Paris.

Finally, I would like to express my gratitude to all persons at the Università della Svizzera Italiana who have been close to me during these years, for their support and all the good moments we have shared.

Contents

Introduction	1
1 Constrained Nonparametric Dependence with Application in Finance and Insurance	5
1.1 Introduction	5
1.2 Copulas	8
1.2.1 Definition and Sklar's Theorem	8
1.2.2 Concepts and measures of dependence	10
1.2.3 Examples	14
1.3 Other functional dependence measures.	20
1.3.1 Canonical analysis	21
1.3.2 Laplace transforms	24
1.4 Applications: static framework.	26
1.4.1 Scoring comparison	27
1.4.2 Competing default risks	27
1.4.3 Age structure of default correlation	29
1.4.4 Income and wealth inequality	30
1.4.5 Moment based problems	31
1.4.6 Control of extreme risk	37
1.5 Nonlinear time series models	40
1.5.1 Characterizations of serial dependence	40
1.5.2 Transitions at any horizon	41
1.5.3 Examples	45
1.6 Applications: dynamic framework	51
1.6.1 Trend correction	51
1.6.2 Dynamic duration model with proportional hazard	52
1.6.3 Derivative pricing	53
1.6.4 Prediction and pricing of default risk	57
1.6.5 Continuous time models	60
2 Duration Time Series Models with Proportional Hazard	105
2.1 Introduction.	106
2.2 Stationary Markov processes with proportional hazard.	108
2.2.1 A Markov Process of Durations.	108

2.2.2	The transformed nonlinear autoregressive representation.	109
2.2.3	The copula representation.	110
2.2.4	Equivalent parameterizations of the copula.	111
2.2.5	An example.	112
2.3	Positive Dependence.	114
2.3.1	Notions of positive dependence.	114
2.3.2	Dependence Orderings.	115
2.3.3	Measures of Dependence.	117
2.3.4	Tail dependence	117
2.3.5	Dependence at larger lag	118
2.4	Ergodicity Properties.	118
2.4.1	Geometric ergodicity.	119
2.4.2	Mixing properties	121
2.5	Examples.	121
2.5.1	Constant measure of dependence.	121
2.5.2	Analytic examples.	122
2.5.3	Endogenous switching regimes.	125
2.5.4	Proportional hazard in reversed time.	127
2.6	Statistical Inference	128
2.6.1	Parametric framework.	128
2.6.2	Nonparametric estimation methods.	131
2.7	Conclusion.	134
3	Efficient Nonparametric Estimation of Models with Nonlinear Dependence	169
3.1	Introduction	170
3.2	The information operator	173
3.2.1	Differentiability condition	174
3.2.2	Identification and Information.	174
3.2.3	Decompositions of the information operator	176
3.3	Examples	179
3.3.1	Differentials of the copula and of the conditional and marginal densities.	179
3.3.2	Examples.	182
3.4	Minimum chi-square estimators.	191
3.4.1	Definition of the estimator.	191
3.4.2	Consistency of the estimators	193
3.4.3	Asymptotic expansion of the minimum chi-square estimator	194
3.4.4	The asymptotic distribution of the minimum chi-square estimator	196
3.4.5	Time series framework.	198
3.5	Nonparametric efficiency.	199
3.5.1	Nonparametric efficiency bound	199
3.5.2	Nonparametric efficiency of the minimum chi-square estimator	200
3.6	Constrained estimation. Identifying restrictions.	201
3.6.1	Restricted information operator.	201
3.6.2	The minimum chi-square estimator.	203

3.6.3	The nonparametric efficiency bound.	204
3.7	Concluding remarks.	204
General Bibliography		251

Introduction

General motivation of the thesis

The developments of financial theory in the last decades have shown that one of the most fundamental topics in Finance is the specification of dependence between different risk variables. For instance, in portfolio analysis, it is well-known at least since Markowitz (1971) that the optimal diversification of risk in a portfolio is related to the contemporaneous dependence between the different assets included in the portfolio. Similarly, in a multiperiod framework, optimal dynamic portfolio allocations are based on the predictability of assets' returns and state variables, and therefore involve the serial dependence of the latter.

Empirical evidence on financial series such as returns, interest rates, or exchange rates suggest that dependence between financial variables exhibits strong features of nonlinearity. Beyond these by now well-documented empirical stylized facts, the logical need for nonlinear modeling has been emphasized by recent developments in risk management, such as the analysis of dependence between default risks of different borrowers, or dependence between extreme risks. Indeed, the analysis of default risk, for instance, typically involves either positive variables, such as times to default, or qualitative variables, such as binary indicators for default in a given period. In these cases, due to the nature of the risk variables, linear specifications are not appropriate. Moreover, in risk management applications nonlinear dependence relates mainly to the whole joint distribution of the variables (and not only to conditional moments, which typically underlie linear models), and the focus is often placed on the tails of the distribution.

Supported by these motivations, a large interest has arisen in financial econometrics for modeling nonlinear dependence. Beyond traditional specifications such as ARCH or switching regimes models, a considerable amount of research has been recently devoted to methodologies based on the joint distribution of the risk variables such as copulas, which are standardized versions of the joint cdf where marginal features have been eliminated. Nonlinear models have been successfully adopted in several financial applications, a tangible proof of their relevance for a careful specification of dependence between risk variables. Despite these very promising achievements, more work is still needed. Indeed, on the one hand, traditional parametric specifications are typically excessively constrained, and are not appropriate for performing a separate analysis of dependence between medium and high risk, as required for instance in Value at Risk applications. On the other hand, in a pure nonparametric approach the interpretation of the patterns of nonlinear dependence may

be difficult due to the complete lack of structure, and moreover the statistical estimates are inaccurate when a large number of risk variables or the tails of the distribution are concerned.

In this thesis we consider an approach to nonlinear dependence which is intermediate between pure parametric and pure nonparametric specifications, combining desirable features of the two. In this approach, called constrained nonparametric dependence, the joint density is constrained and depends on a small number of one-dimensional functional parameters, that are functions of one variable. Constrained nonparametric dependence presents several advantages for modeling nonlinear dependence in financial applications. Firstly the presence of functional components in the model provides flexibility to the specifications and richness to the admissible patterns of nonlinear dependence, while affording clear structural interpretations of nonlinear dependence. Indeed, the nonparametric constraints are typically introduced by means of latent factors, omitted heterogeneities, proportional hazard specifications, or exponential affine restrictions, which are appropriate for financial analysis. Secondly, this approach allows to interpret and visualize the patterns of nonlinear dependence by relating them to specific features of the one-dimensional functional parameters, such as their shape, monotonicity, elasticity, or divergence behaviour at the boundary points of the support. Thirdly, the nonparametric dimensionality is controlled, and the rate of convergence of the estimators is the standard one-dimensional nonparametric rate, independent of the number of underlying risk variables. This leads to more accurate and robust results for estimation and inference on the functional parameters characterizing nonlinear dependence.

The purpose of this thesis is to develop new econometric methodologies involving constrained nonparametric dependence, aimed at providing valuable tools for financial applications, both from a modeling and a statistical inference point of view. The thesis is organized in three chapters¹. The first chapter introduces constrained nonparametric specifications, and motivates their application in Finance. The core of the thesis consists of Chapter 2 and 3. Chapter 2 is devoted to modeling methodologies, and presents the analysis of nonlinear serial dependence in a dynamic constrained nonparametric specification by treating the case of dynamic duration models with proportional hazard. Statistical inference is considered in Chapter 3, where we provide efficient nonparametric estimators for the functional parameters characterizing nonlinear dependence. The content of the three chapters is detailed below.

Detailed outline of the chapters

In chapter 1 we consider the specification of models with constrained nonparametric dependence, discuss the analysis of nonlinear dependence in this framework, and present several financial applications. In constrained nonparametric models nonlinear dependence between risk variables is specified by introducing one-dimensional functional parameters in an appropriate representation of the joint density. Different approaches are possible according to which functional representation of dependence is selected. Three alternative representations are: copulas, nonlinear canonical decompositions of the joint density, and conditional Laplace transforms (also called conditional moment generating functions). We review the main related definitions and results, and illustrate their relationship with traditional econo-

¹They correspond to three papers written with Christian GOURIEROUX.

metric specifications such as factor models. Constrained nonparametric specifications are implemented by introducing restrictions in that characterization of dependence, which is best-suited for the financial problem of interest. The analysis of the patterns of nonlinear dependence is performed by relating them to the shape of the functional parameters, using appropriate dependence concepts and measures. We present several financial and economic applications involving either contemporaneous or dynamic nonlinear dependence between two (or more) risk variables, and discuss the adequacy of various constrained nonparametric specifications. The applications in the cross-sectional framework include: i) the study of the age structure of default correlation, ii) problems based on nonlinear cross-moments, such as expected utility maximization, and iii) the control of dependence between extreme risks. In the dynamic framework we discuss: i) duration models with proportional hazard for the analysis of liquidity risk, ii) derivative pricing and term structure of interest rates, and iii) applications of nonlinear canonical analysis for continuous time models.

Chapter 2 illustrates the use of constrained nonparametric specifications for nonlinear dynamic modeling by introducing duration time series with proportional hazard. These models are useful for the analysis of the liquidity risk, and assume lagged durations as regressors with proportional hazard effect. Due to invariance by increasing transformation, the proportional hazard constraint only concerns the copula of the process, and any stationary distribution F may be imposed by appropriate marginal transformations. The copula of the Markov process with proportional hazard is characterized by a one-dimensional functional parameter a defined on $[0, 1]$. In this specification marginal features, included in the stationary cdf, and serial dependence features, characterized by function a , are completely separated. Markov processes with proportional hazard provide a dynamic duration model with rich dependence features and flexible marginal specifications. The patterns and strength of serial dependence (namely various forms of dependence, dependence in the tails, ergodicity) are related to the elasticity of parameter a and to its behaviour at the boundary points of the support.

Chapter 3 is devoted to statistical inference in constrained nonparametric families of densities, which depend on one-dimensional functional parameters. The functional parameter may be defined up to one-to-one transformations, and the choice of the appropriate normalization requires a nondegenerate differential and information operator. Various examples of constrained nonparametric families are discussed, and closed forms expressions for the differential and the information operator are provided. A natural nonparametric estimator for the functional parameter characterizing nonlinear dependence can be defined by minimizing a chi-square distance between the constrained densities in the family and an unconstrained kernel estimator of the density. We derive the asymptotic properties of the estimator and of its linear functionals. In particular, the pointwise estimator for the one-dimensional parameter is shown to have the expected one-dimensional nonparametric rate. Finally, nonparametric efficiency bounds are derived, and the nonparametric efficiency of the minimum chi-square estimator is shown.

Concluding remarks and future developments

Through this thesis we hope to provide valuable econometric methodologies for the analysis of nonlinear dependence in financial applications. We have shown that the approach of constrained nonparametric dependence presents several advantages in terms of flexibility, structural interpretations of the patterns of nonlinear dependence, and control of the nonparametric dimension. Further we have provided efficient nonparametric estimators for the functional parameters characterizing nonlinear dependence.

Different future developments seem promising. On the one hand, Chapter 1 presents several interesting economic and financial applications, which deserve a deeper theoretical and empirical analysis. On the other hand, further methodological developments for statistical inference are natural, for instance the introduction of specification tests based on the minimum chi-square criterion.

Chapter 1

Constrained Nonparametric Dependence with Application in Finance and Insurance

1.1 Introduction

Dependence between risk variables, either contemporaneous or dynamic, is at the heart of many topics in Finance and Insurance, including asset pricing, portfolio analysis and risk management. Based on empirical investigations, there is by now a convincing evidence that dependence between financial variables involves strong departures from the gaussian assumption, such as fat tails or asymmetries¹, and important nonlinearities, such as autoregressive conditional heteroschedasticity or switching regimes². These features are generally ascribed to different causes, for instance complexity of the economic agents' behaviour, such as attitude towards risk, or peculiarities of financial contracts, which may involve nonlinear payoffs, such as options. Beyond these empirical stylized facts, nonlinearity and non-normality features are implied in some cases simply by the nature of the financial variables, which may be qualitative, such as price variations at a tick by tick frequency, or positive, such as durations until default in credit risk models³.

Several approaches have been proposed in the econometric literature to model nonlinear dependence. On the one hand, parametric specifications have been typically adopted for their simplicity. They include for instance extensions of the traditional linear regression framework, such as ARCH, ACD, or Markov switching regimes models⁴, and parametric

¹See e.g. Mandelbrot (1963), Clark (1973) for original references on fat tails in financial returns, and e.g. Ang, Chen (2002) for a recent discussion of asymmetric correlations in equity portfolios.

²See Bollerslev, Engle, Nelson (1990) and Bollerslev, Chou, Kroner (1992) for reviews on ARCH effects, Hamilton (1993) for a survey on regime switches, and e.g. Ang, Bakaert (2002b) for a recent discussion of regime switches in interest rates. See further Tong (1990), and Teräsvirta, Tjøstheim, Granger (1994) for general reviews on nonlinear time series with economic and financial applications.

³See Campbell, Lo, MacKinlay (1997), chapter 12, and Gouriéroux, Jasiak (2001a) for general reviews on nonlinear models in finance.

⁴See Engle (1982), Engle, Russell (1998), and Hamilton (1989, 1999).

families of standardized joint cdf's called copulas⁵. However, parametric specifications are often excessively constrained, resulting in a poor fit to the data, with serious consequences for pricing and risk management. In addition, in a nonlinear setting, it may be difficult to find specifications whose parameters have a clear economic or financial interpretation, diminishing the attractiveness of this approach. On the other hand, fully nonparametric approaches⁶ suffer from the curse of dimensionality, which causes inaccurate estimates when the number of variables is larger than 3, or when we are interested in dependence between extremes (since the number of informative observations becomes too small). Moreover, the complete lack of structure complicates the interpretation of the models and diminishes the robustness of the results, and therefore may be undesirable in economic or financial applications.

Intermediate approaches, which combine desirable features of parametric and nonparametric specifications, are also possible. For instance nonlinear dependence between risk variables can be summarized by one-dimensional functional parameters, that are functions of one variable, characterizing the joint distribution. This approach is called constrained nonparametric dependence. A typical example of this approach are transformation models, where an unknown transformation of the endogenous variable satisfies a linear regression model⁷. A further example is provided by regression models where the mean and volatility are unrestricted functions of a set of regressors⁸. Indeed, in order to avoid the curse of dimensionality when the number of regressors is large, these models typically assume that the mean (and the volatility) is either the sum (the product, respectively) of one-dimensional functions, or depends on a scalar transformation of the regressors, called index⁹. Recently, different specifications with constrained nonparametric dependence have been introduced, which involve more general restrictions and are more appropriate for financial applications, such as the analysis of nonlinear dependence between durations until default for several borrowers, or modeling the term structure of interest rates¹⁰. These constrained nonparametric specifications presents several advantages. Firstly, as in a nonparametric approach, the presence of functional components contributes to the flexibility of the specification and to the richness of the admissible patterns of dependence. Secondly, by means of appropriate dependence concepts, nonlinear dependence may be related to the shape of these functional parameters, which becomes the focus of interest. Thirdly, the nonparametric dimensionality is controlled, which allows for more accurate and robust results.

Another distinction between the approaches proposed to study nonlinear dependence concerns the selection of an appropriate representation of the joint distribution summarizing its characteristics. Typically the distribution of a one-dimensional continuous variable can be characterized by either its probability density function, or its cumulative distribution function, or its survivor function, or its hazard function, or its characteristic function. It

⁵See e.g. Rockinger, Jondeau (2001), Patton (2002).

⁶See e.g. Aït-Sahalia, Lo (1998), Brandt (1999), Fermanian, Scaillet (2002).

⁷See e.g. Han (1987a,b), Horowitz (1996), Gorgens, Horowitz (1999).

⁸Haerdle, Tsybakov (1999).

⁹See Hastie, Tibshirani (1990) for generalized additive models, and Härdle, Stocker (1989), Powel, Stock (1989), Ichimura (1993), Horowitz, Härdle (1996) for index models.

¹⁰See e.g. Darolles, Gourioux, Jasiak (2002), Gagliardini, Gourioux, Robert (2002), Gourioux, Monfort, Polimenis (2002).

is well known that some characteristics are more appropriate for given applications. For instance the hazard function is easily interpreted for duration variables, and the constrained duration models are usually defined by means of restrictions on the hazard function. The same type of remark holds when we consider two variables and focus on nonlinear dependence. Intuitively, we have to separate the marginal features and the dependence features. However the latter one can be characterized in various ways, for instance by means of either a conditional distribution, or a copula, or the canonical decomposition of the joint pdf, or a conditional Laplace transform, and so on. In practice we have to select the representation of nonlinear dependence, which is appropriate for the problem of interest. For instance, it is known that the analysis of the term structure of interest rates and of default risk naturally involves conditional Laplace transforms¹¹, which are appropriate characterizations of nonlinear dependence for these applications. The parametric or nonparametric specifications are then implemented by introducing constraints on the well-chosen representation of nonlinear dependence.

In section 2 we recall the definition of a copula, which is a measure of nonlinear dependence widely used for joint analysis of financial risks¹². A copula function summarizes dependence which is invariant to increasing transformations of the variables. We consider associated functional dependence concepts, measures and orderings, which may be used to interpret the pattern and strength of nonlinear dependence. Finally, we introduce several examples of parametric and nonparametric copula families, and discuss their dependence properties.

In section 3 we consider two other functional characterizations of nonlinear dependence, namely nonlinear canonical analysis and Laplace transforms. In nonlinear canonical analysis the joint pdf is decomposed into orthonormal functional directions of dependence (called canonical directions), and associated canonical correlations, in decreasing dependence order¹³. We introduce a constrained nonparametric specification featuring finite dimensional dependence by imposing a finite number of non-zero canonical correlations¹⁴. Finally in the approach based on Laplace transforms, the joint density is characterized by the Laplace transforms of a marginal and of a conditional distribution. We consider especially the compound model, where the conditional Laplace transform is assumed to be exponential affine in the conditioning variable¹⁵.

In section 4 we present several financial and economic applications involving contemporaneous nonlinear dependence between two (or more) risk variables, and discuss the appropriateness of the characterizations of nonlinear dependence introduced in section 2 and 3. The applications include: i) the comparison of two scores (for default, for instance) attributed to the same individuals by means of different criteria; ii) the analysis of dependence between

¹¹Duffie, Pan, Singleton (2000)], Gouriéroux, Monfort, Polimenis (2002).

¹²See e.g. Embrechts, McNeil, Straumann (1999) for a general discussion of the usefulness of copulas in Finance, Jouanin et al. (2001) for an application to credit risk, Rockinger, Jondeau (2001) and Patton (2002) for applications to dynamic portfolio selection.

¹³See e.g. Dunford, Schwarz (1968) and Lancaster (1968). See also Gouriéroux, Jasiak (2001a) for an application to intertrade durations, or Chen, Hansen, Scheinkman (2002) for a macroeconomic application.

¹⁴Gouriéroux, Jasiak (2001b).

¹⁵Darolles, Gouriéroux, Jasiak (2002).

competing default risks; iii) the study of the age structure of default correlation; iv) the role of dependence between income and wealth for inequality theory; v) problems based on nonlinear cross-moments, such as expected utility maximization and pricing of derivatives written on more than one underlying asset, as well as vi) the control of dependence between extreme risks. In section 5 we discuss nonlinear dependence in dynamic models, and characterize serial dependence in nonlinear time series by means of copulas, nonlinear canonical analysis, and Laplace transforms. Different constrained nonparametric specifications are introduced, including Markov processes with finite-dimensional dependence, and Compound Autoregressive processes. We provide simulated trajectories, autocorrelation functions, and isodensity curves at several horizons for different examples, in order to analyse how the nonlinear dependence pattern affects the dynamics of the process. Several financial applications in the dynamic framework are presented in section 6, including: i) trend correction in nonlinear time series to study the ranking dynamics of different firms, ii) dynamic duration models with proportional hazard for the analysis of liquidity risk, iii) derivative pricing and term structure of interest rates, iv) prediction and pricing of default risk, as well as v) applications to continuous time models. These applications require the specification of dynamic nonlinear dependence either for an observed time series (such as a series of intertrade durations), or for latent factors (such as the intensity processes in credit risk models, or the stochastic discount factor in pricing models). Section 7 concludes.

For ease of exposition the theoretical results are generally presented in a continuous bivariate framework, but most of them can be directly extended to a multivariate case.

1.2 Copulas

In this section we consider a functional measure of dependence called copula, which is invariant with respect to increasing transformations of the variables. We first review the main definitions and results in copulas' theory [see Joe (1997) and Nelsen (1999) for surveys on copulas]. Then we give several examples of copula families, including the gaussian copula and copula families characterized by one-dimensional functional parameters.

1.2.1 Definition and Sklar's Theorem

Copula functions have been introduced to specify and analyse multivariate distributions covering various types of dependence structures. A copula function couples marginal distributions to get a joint distribution, and summarizes the dependence which is invariant to increasing transformations of the variables. Let us first introduce the definition of a copula (for the bivariate case).

Definition 1.1 *A joint cumulative distribution function (c.d.f.) C on $[0, 1]^2$, with uniform marginal distributions on $[0, 1]$, is called a copula.*

Thus a function $C : [0, 1]^2 \longrightarrow [0, 1]$ is a copula if:

- i) $C(0, v) = C(u, 0) = 0, \quad \forall u, v \in [0, 1];$
- ii) $C(u, 1) = u, \quad C(1, v) = v, \quad \forall u, v \in [0, 1];$
- iii) for any rectangle $R = [u_1, u_2] \times [v_1, v_2] \subset [0, 1]^2$:

$$\int \int_R C(du, dv) = C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) + C(u_1, v_1) \geq 0. \quad (1.1)$$

When the distribution C is continuous, the associated density:

$$c(u, v) = \frac{\partial^2 C}{\partial u \partial v}(u, v), \quad u, v \in [0, 1], \quad (1.2)$$

is called the copula density.

The main theorem in copulas' theory is Sklar's Theorem [Sklar (1959)]. In order to introduce it, let F_X and F_Y denote univariate c.d.f. and let C be a copula. Then the function F defined by:

$$F(x, y) = C[F_X(x), F_Y(y)], \quad x, y \in \mathbb{R}, \quad (1.3)$$

is a bivariate c.d.f., with marginal cdf F_X and F_Y . Sklar's Theorem shows that the reverse is also true.

Theorem 1.1 (Sklar) *Let F be a bivariate c.d.f., with marginal cdf F_X and F_Y . Then there exists a copula C such that:*

$$F(x, y) = C[F_X(x), F_Y(y)], \quad \forall x, y.$$

This copula is unique, if F is a continuous distribution.

The copula C of continuous variables (X, Y) with joint distribution F and marginal cdf F_X, F_Y is defined as the unique copula satisfying (1.3). Let us sketch a proof of Sklar's Theorem, which is instructive to understand how the copula C is derived. Let X, Y be continuous variables with joint cdf F and marginal distributions F_X, F_Y . Then the transformed variables:

$$U = F_X(X) \quad \text{and} \quad V = F_Y(Y), \quad (1.4)$$

have uniform marginal distributions, and their joint c.d.f. is given by:

$$\begin{aligned} P[U \leq u, V \leq v] &= P[X \leq F_X^{-1}(u), Y \leq F_Y^{-1}(v)] \\ &= F[F_X^{-1}(u), F_Y^{-1}(v)], \quad \forall u, v \in [0, 1]. \end{aligned}$$

We deduce that the c.d.f. of (U, V) is a copula C satisfying condition (1.3). The uniqueness follows from the continuity of the marginal distributions F_X and F_Y .

Corollary 1.1: *The copula C of (X, Y) is the c.d.f. of the transformed variables (U, V) . The variables (U, V) are called standardized variables (or quantile transformations).*

Corollary 1.2: *The copula of (X, Y) is the same as the copula of $(g(X), h(Y))$, where g, h are any increasing transformations.*

Thus Sklar's Theorem allows to separate the information contained in a joint c.d.f. F into: i) the marginal features, described by F_X and F_Y , and ii) some dependence characteristics, described by the copula C . The dependence features characterized by the copula C are invariant with respect to increasing transformations of the variables X and Y [see Corollary 2].

Sklar's Theorem states the equivalence between the joint distribution on the one hand, the copula and marginal distributions on the other hand. This equivalence explains the usefulness of copulas. Firstly, given a joint c.d.f. F , we can describe the dependence which is invariant to increasing transformations by recovering the associated copula C :

$$C(u, v) = F[F_X^{-1}(u), F_Y^{-1}(v)], \quad u, v \in [0, 1]. \quad (1.5)$$

Symmetrically, if variables X and Y have marginal distributions F_X, F_Y , we can specify a joint distribution for (X, Y) by specifying a copula C and defining the joint c.d.f. as $F(x, y) = C[F_X(x), F_Y(y)], \quad \forall x, y$.

1.2.2 Concepts and measures of dependence

Various dependence concepts, measures and orderings based on copulas have been introduced in the statistical literature to describe, measure and compare dependence in joint distributions. Contrary to standard correlation, they are mainly concerned with dependence which is invariant to increasing transformations¹⁶. Moreover, whereas correlation measures linear affine dependence, they will summarize nonlinear dependence features.

These concepts and measures can be either global in nature, when they summarize the dependence in the whole distribution, or local, when they focus on some regions of the sample space $[0, 1]^2$. Moreover, they may involve either scalar, or functional summaries of dependence.

i) Kendall's tau

Kendall's tau [Kendall (1938)] is one of the most well-known and frequently used global measures of dependence, which is invariant by increasing transformations. Let F be a bivariate c.d.f., and $(X, Y), (X', Y')$ denote independent pairs with identical joint c.d.f. F . Then Kendall's tau τ_F of distribution F measures the probability of concordant pairs:

$$\tau_F = P\left[\left(X - X'\right)\left(Y - Y'\right) > 0\right] - P\left[\left(X - X'\right)\left(Y - Y'\right) < 0\right]. \quad (1.6)$$

¹⁶The correlation $\text{corr}(X, Y)$ is not invariant to nonlinear increasing transformations of the variables X, Y , and it is not characterized by the copula of (X, Y) .

Kendall's tau τ_F can be written in terms of the copula C of F as:

$$\tau_F = 4 \int \int C(u, v) C(du, dv) - 1 = \tau_C. \quad (1.7)$$

The range of Kendall's tau is $[-1, 1]$. Kendall's tau reaches its maximal (minimal) value $\tau_F = 1$ ($\tau_F = -1$, respectively) if and only if variables X and Y are in an increasing (decreasing) deterministic relationship: $Y = g(X)$, (say), where g is an increasing (decreasing) function. This result is the counterpart of the following standard property of correlation: $\text{corr}(X, Y) = 1$ [$\text{corr}(X, Y) = -1$] iff X and Y are in increasing (decreasing) affine relationship.

When X and Y are independent, then $\tau_F = 0$. However, the reverse is not true, and a zero Kendall's tau does not imply independence. For instance, the Kendall's tau of (X, Y) is zero when the distribution of X given Y is symmetric. The analogous in terms of linear correlation is that $\text{corr}(X, Y) = 0$, when $E[X | Y] = 0$.

ii) Positive Quadrant Dependence

Two random variables $(X, Y) \sim F$ are Positive Quadrant Dependent (PQD) if they are more likely both small under F , than it would be under the independence hypothesis:

$$P[X \leq x, Y \leq y] \geq P[X \leq x] P[Y \leq y], \quad \forall x, y. \quad (1.8)$$

Positive Quadrant Dependence is a property of the copula, since X and Y are PQD iff their copula C is such that:

$$C(u, v) \geq uv, \quad \forall u, v \in [0, 1], \quad (1.9)$$

where $C_{ind}(u, v) = uv$ is the copula of independent variables. Moreover, since any decreasing function of X , $g(x)$ (say), is the limit of linear combinations of indicator functions $\mathbb{I}_{(-\infty, x_i]}$:

$$g(x) = \lim_n \searrow \left\{ \alpha_0 + \sum_{i=1}^n \alpha_i \mathbb{I}_{(-\infty, x_i]}(x) \right\},$$

with positive components α_i , $i = 1, \dots, n$, it is immediately deduced from the PQD condition (1.8) that:

$$E[g(X)h(Y)] \geq E[g(X)] E[h(Y)], \quad \text{for any decreasing functions } g, h,$$

or equivalently that:

$$\text{Cov}[g(X), h(Y)] \geq 0, \quad \text{for any increasing functions } g, h, \quad (1.10)$$

(such that the covariance exists). In fact property (1.10) is equivalent to the PQD condition (1.8) [see Dhaene, Goovaerts (1996)] and provides the interpretation of PQD in terms of correlations.

Finally, an equivalent characterization of PQD in terms of survivor probabilities can be derived. Variables X and Y are PQD iff they are more likely simultaneously large than their independent copies:

$$P[X \geq x, Y \geq y] \geq P[X \geq x] P[Y \geq y], \quad \forall x, y.$$

To summarize, we have the following equivalent characterizations of PQD:

Property 1.1: *Variables (X, Y) are PQD if and only if any of the following conditions is satisfied:*

- i) $P[X \leq x, Y \leq y] \geq P[X \leq x]P[Y \leq y], \quad \forall x, y;$
- ii) $C(u, v) \geq uv, \quad \forall u, v \in [0, 1];$
- iii) $Cov[g(X), h(Y)] \geq 0$, for any increasing functions g, h , such that the covariance exists;
- iv) $P[X \geq x, Y \geq y] \geq P[X \geq x]P[Y \geq y], \quad \forall x, y.$

Positive Quadrant Dependence between two variables can be introduced by means of common factors.

Example 1: Let variables X and Y be such that:

$$X = a(Z, \varepsilon), \quad Y = b(Z, \eta),$$

where Z, ε and η are independent variables, and functions a and b are increasing in the argument Z . In particular, the distribution of X given Z (and of Y given Z) is increasing in Z for the first order stochastic dominance. We get:

$$Cov[g(X), h(Y)] = Cov\{E(g[a(Z, \varepsilon)] | Z), E(h[b(Z, \eta)] | Z)\} \geq 0,$$

for any increasing functions g, h , since $E(g[a(Z, \varepsilon)] | Z)$ and $E(h[b(Z, \eta)] | Z)$ are both increasing with respect to Z . From Property 1 iii) we deduce that X and Y are PQD. The analogous result in terms of linear correlation is that the variables:

$$X = \alpha Z + \varepsilon, \quad Y = \beta Z + \eta,$$

where $\alpha, \beta \geq 0$, have a positive correlation.

PQD can be used to define a dependence ordering. Copula C_1 is said to be more PQD than copula C_2 , noted $C_1 \succeq_{PQD} C_2$, if:

$$C_1(u, v) \geq C_2(u, v), \quad \forall u, v \in [0, 1]. \quad (1.11)$$

The maximal and minimal elements with respect to the PQD ordering are the upper and lower Frechet bounds, respectively. They are characterized by the copulas:

$$C_u(u, v) = \min(u, v), \quad C_l(u, v) = \max\{u + v - 1, 0\},$$

respectively. They correspond to deterministic linear dependence between the quantile variables, $U = V$ and $U = 1 - V$, respectively, and to increasing, and decreasing, respectively, deterministic nonlinear dependence between X and Y .

Kendall's tau is compatible with the PQD ordering: if copula C_1 is more PQD than copula C_2 , then Kendall's tau of C_1 is larger than that of C_2 .

The PQD dependence ordering can also be defined for the pair of initial variables (X_1, Y_1)

and (X_2, Y_2) . In particular if X_1 and X_2 (resp. Y_1 and Y_2) have the same distribution, the PQD ordering can be characterized in terms of covariance:

$$\begin{aligned} \text{Cov}[g(X_1), h(Y_1)] &\geq \text{Cov}[g(X_2), h(Y_2)], \\ &\text{for any increasing functions } g, h, \end{aligned} \quad (1.12)$$

(such that the covariances exist)¹⁷, or in terms of survivor functions:

$$S_1(x, y) \geq S_2(x, y), \quad \forall x, y, \quad (1.13)$$

where S_1 and S_2 are the joint survivor functions of (X_1, Y_1) and (X_2, Y_2) , respectively.

The PQD ordering is useful to investigate the effect of nonlinear dependence between X and Y on cross-moments such as $E[g(X, Y)]$. For instance if function g is the cumulative function of a positive measure, by Fubini theorem we get:

$$E[g(X, Y)] = \overset{g}{E}[S(X, Y)], \quad (1.14)$$

where S is the survivor function of (X, Y) , and $\overset{g}{E}$ denotes the expectation with respect to measure g . Then, by characterization (1.13) of PQD, we deduce the following proposition [see Tchen (1980), and Müller, Scarsini (2000) for equivalent formulations].

Proposition 1.2: *Let g be the cumulative function of a positive bivariate measure. Let F_1 and F_2 be bivariate cdf with the same marginal distributions, and copulas C_1 and C_2 , such that $C_1 \succeq_{PQD} C_2$. Then:*

$$\overset{F_1}{E}[g(X, Y)] \geq \overset{F_2}{E}[g(X, Y)].$$

Thus the expectation $E[g(X, Y)]$ is monotone with respect to PQD between X and Y , for given marginal distributions.

iii) Tail dependence

The observations drawn from the tails of a distribution are called extremes, or extreme values. The dependence between extremes of several variables is important for many applications in Finance and Insurance. For instance, the management of extreme risk in a financial portfolio and the determination of the capital required to hedge these risks are based on a careful analysis of the dependence between extreme returns of the assets included in the portfolio. Therefore it is important to introduce measures which focus on dependence in the joint tails. For this purpose the statistical literature has introduced the tail dependence coefficients [Joe (1993)]. The lower tail dependence coefficient λ_L is defined by:

$$\lambda_L = \lim_{\alpha \rightarrow 0} P[X \leq F_X^{-1}(\alpha) \mid Y \leq F_Y^{-1}(\alpha)],$$

¹⁷Note that the same inequalities hold for the correlations if X_1 and X_2 , Y_1 and Y_2 have the same marginal distributions. Otherwise the inequalities are generally not valid for the correlations.

when this limit exists¹⁸. Thus λ_L is given by the limiting probability that X is extreme given that Y is extreme. When $\lambda_L > 0$, X and Y feature positive lower tail dependence; otherwise they are lower tail independent. The coefficient of lower tail dependence λ_L is symmetric in variables X and Y , and is invariant by increasing transformations. It can be written in terms of the copula C :

$$\begin{aligned}\lambda_L &= \lim_{u \rightarrow 0} P[U \leq u \mid V \leq u] \\ &= \lim_{u \rightarrow 0} \frac{C(u, u)}{u}.\end{aligned}\tag{1.15}$$

iv) Tail conditional copulas

However the lower tail dependence coefficient is a scalar measure considering only the extreme behaviour when α tends to 0. For practical purposes, such as the determination of the required capital (Value at Risk), it is usual to study the extreme when α is small, but not infinitely small. For instance we can be interested in the dependence when α is of order 5% or 1%, without tending to zero. In order to get a richer description of dependence in the tails of a bivariate distribution, we have to introduce functional measures. Charpentier (2002) and Juri, Wuethrich (2002 a,b) consider tail conditional copulas. The lower tail conditional copula C_α of X and Y at threshold α is the copula of (X, Y) given $X \leq F_X^{-1}(\alpha)$, $Y \leq F_Y^{-1}(\alpha)$, that is the copula of (U, V) given $U \leq \alpha$, $V \leq \alpha$. It is given by:

$$C_\alpha(u, v) = \frac{C[F_{U,\alpha}^{-1}(u), F_{V,\alpha}^{-1}(v)]}{C(\alpha, \alpha)},$$

where C is the copula of (X, Y) ,

$F_{U,\alpha}(u) = P[U \leq u \mid U \leq \alpha, V \leq \alpha] = C(\min\{u, \alpha\}, \alpha) / C(\alpha, \alpha)$, and similarly $F_{V,\alpha}(v) = P[V \leq v \mid U \leq \alpha, V \leq \alpha] = C(\alpha, \min\{v, \alpha\}) / C(\alpha, \alpha)$. The conditional copula can be considered for a given level of α , such as $\alpha = 5\%$ or $\alpha = 1\%$. Its asymptotic behaviour when $\alpha \rightarrow 0$ can also be investigated [see Juri, Wuethrich (2002 a,b)].

1.2.3 Examples

In this section we introduce examples of copula families and discuss their dependence properties. The most famous copula is the Gaussian copula, which is characterized by a linear correlation parameter ρ . Constrained nonparametric specifications may be introduced by considering copula families characterized by one-dimensional functional parameters.

¹⁸Dependence between the upper tails can be measured by the upper tail dependence coefficient λ_U , which is defined by: $\lambda_U = \lim_{\alpha \rightarrow 1} P[X \geq F_X^{-1}(\alpha) \mid Y \geq F_Y^{-1}(\alpha)]$

1.2.3.1 Gaussian copula.

i) Definition

A gaussian copula is associated with a bivariate gaussian distribution:

$$N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right]. \quad (1.16)$$

Thus the copula family is indexed by the correlation parameter $\rho \in [-1, 1]$ and is given by:

$$C_\rho(u, v) = \Phi_\rho [\Phi^{-1}(u), \Phi^{-1}(v)],$$

where $\Phi_\rho(x, y)$ denotes the c.d.f. of the gaussian distribution (1.16) with zero mean, unitary variances and correlation coefficient ρ , and $\Phi(x)$ denotes the c.d.f. of a standard univariate Gaussian distribution. Two variables X, Y with marginal distributions F_X, F_Y and gaussian copula C_ρ admit the joint c.d.f.:

$$F(x, y) = C_\rho[F_X(x), F_Y(y)] = \Phi_\rho[\Phi^{-1}(F_X(x)), \Phi^{-1}(F_Y(y))].$$

ii) Dependence

The gaussian family C_ρ is positively PQD ordered with respect to the dependence parameter ρ . The case $\rho = 0$ corresponds to independence. When $\rho = 1$, we have $U = V$, and variables X and Y are in increasing deterministic dependence: $Y = F_Y^{-1}[F_X(X)]$ ¹⁹. In particular, when F_X and F_Y are not identical, the deterministic relationship between X and Y is in general nonlinear, and differs from the affine relationship between U and V . It is important to note that the correlation between X and Y is an increasing function of ρ ; it differs from ρ , as soon as $\Phi^{-1} \circ F_X$ and $\Phi^{-1} \circ F_Y$ are not affine functions. Finally, gaussian copulas feature both upper and lower tail independence: $\lambda_U = \lambda_L = 0$.

iii) Scatterplots and isodensity curves

Gaussian copula may be used to define random variables (X, Y) with a gaussian dependence structure, and nonnormal marginal distributions, featuring for instance fat tails. In Figure 1, Panels A and B, we report scatterplots of simulated pairs (X, Y) with identical student marginal distributions t_4 and gaussian copula C_ρ for different values of ρ : $\rho = 0.2$ in Panel A, and $\rho = 0.8$ in Panel B.

[insert Figure 1A: scatterplot X, Y , marginal student, $\rho = 0.2$]

[insert Figure 1B: scatterplot X, Y , marginal student, $\rho = 0.8$]

In Panels C and D we provide the scatterplots of the corresponding standardized variables (U, V) .

[insert Figure 1C: scatterplot U, V , $\rho = 0.2$]

¹⁹Similarly for $\rho = -1$.

[insert Figure 1D: scatterplot U, V , $\rho = 0.8$]

As expected from the PQD order, when the dependence parameter ρ increases, the distribution gets more concentrated along the 45 degree line. Copula densities are reported in Figure 2, Panels A and B.

[insert Figure 2A: copula p.d.f., $\rho = 0.2$]

[insert Figure 2B: copula p.d.f., $\rho = 0.8$]

The density is more concentrated along the line $u = v$, when ρ increases. For sake of visual interpretation and comparison with other families, it is useful to consider the density of the transformations of X and Y having standard normal margins:

$$X^* = \Phi^{-1} [F_X(X)], \quad Y^* = \Phi^{-1} [F_Y(Y)].$$

If X, Y have a gaussian copula C_ρ , X^*, Y^* have joint standard normal distribution, with correlation ρ . The isodensity curves of X^*, Y^* are ellipsoids represented in Figure 2, Panels C and D.

[insert Figure 2C: isodensity curves, $\rho = 0.2$]

[insert Figure 2D: isodensity curves, $\rho = 0.8$]

1.2.3.2 Archimedean Copula

i) Definition

Let ϕ be a convex, decreasing, positive function on $(0, 1]$, such that $\phi(1) = 0$ and $\phi(0) = +\infty$. An Archimedean copula with (strict) generator ϕ is defined by [Genest, Mc Kay (1986)]:

$$C_\phi(u, v) = \phi^{-1} [\phi(u) + \phi(v)].$$

ii) Factor representation

A wide class of Archimedean copulas, including the most usual ones, admit a factor representation. More precisely, let us consider a positive random variable Z such that X and Y are independent conditionally on Z , and:

$$P[X \leq x \mid Z] = G_X(x)^Z, \quad P[Y \leq y \mid Z] = G_Y(y)^Z, \quad (1.17)$$

where G_X and G_Y are c.d.f.'s. The variable Z is a latent factor with a common effect on X and Y , which admits an interpretation in terms of proportional hazard. Indeed, let us recall that variable W features proportional hazard in variable Z if the conditional hazard function of W given Z is proportional to Z , that is: $P[W \geq w \mid Z] = \exp[-Z\Lambda_0(w)]$, where Λ_0 is an increasing function called baseline integrated hazard. For Archimedean copulas, we deduce from (1.17) that $P[-X \geq x \mid Z] = \exp\{-Z[-\log G_X(-x)]\} = \exp\{-Z\Lambda_X(x)\}$,

(say), and similarly for Y . Thus, if for instance variables X and Y are two duration variables associated with competing risks, factor Z has a proportional hazard effect on X and Y , and is often interpreted as omitted heterogeneity²⁰.

The joint c.d.f. of variables X and Y is given by:

$$\begin{aligned} F(x, y) &= E \left[G_X(x)^Z G_Y(y)^Z \right] = E \exp \left[-Z (-\log G_X(x) - \log G_Y(y)) \right] \\ &= \psi \left[-\log G_X(x) - \log G_Y(y) \right], \end{aligned}$$

where ψ is the real Laplace transform (moment generating function)²¹ of the distribution of Z , and the marginal cdf are given by:

$$F_X(x) = \psi \left[-\log G_X(x) \right], \quad F_Y(y) = \psi \left[-\log G_Y(y) \right].$$

Thus the copula of the pair (X, Y) is:

$$C(u, v) = F \left[F_X^{-1}(u), F_Y^{-1}(v) \right] = \psi \left[\psi^{-1}(u) + \psi^{-1}(v) \right].$$

This is an Archimedean copula with generator $\phi = \psi^{-1}$.

The variables X and Y admit the stochastic representations: $X = G_X^{-1}(\varepsilon^{1/Z})$, $Y = G_Y^{-1}(\eta^{1/Z})$, with ε, η, Z independent, and ε, η uniformly distributed on $[0, 1]$. Variables X and Y are both increasing in the common factor Z for first order stochastic dominance. Thus from Example 1 Archimedean copulas with factor structure (1.17) feature positive dependence.

iii) PQD ordering

The PQD ordering in the Archimedean family can be described in terms of the generator ϕ . Let C_{ϕ_i} , $i = 1, 2$, be two Archimedean copulas with generators ϕ_i , $i = 1, 2$. Different characterizations of the Archimedean PQD ordering have been considered in the literature [see e.g. Joe (1997), chapter 4, and Nelsen (1999), chapter 4]. For instance, a sufficient condition for PQD ordering is the following [see Corollary 4.4.4 in Nelsen (1999)]: C_{ϕ_1} is more PQD than C_{ϕ_2} if:

$$\nu = \phi_2 \circ \phi_1^{-1} \text{ is concave.} \quad (1.18)$$

This condition has an interesting interpretation for Archimedean copulas derived from factor models (1.17), in terms of the dispersion of the latent factor. Let Z_1 and Z_2 denote the latent factors corresponding to Archimedean copula C_{ϕ_1} and C_{ϕ_2} , with real Laplace transforms ψ_1 and ψ_2 , respectively. Let us assume for simplicity that Z_1 and Z_2 have finite expectation. Since they are defined up to a multiplicative constant, we may assume without loss of generality that $E[Z_1] = E[Z_2] = 1$. Then, since $\psi_1'(0) = \psi_2'(0) = -1$, we deduce from (1.18) that

²⁰This proportional hazard interpretation explains why these models are called frailty models [see e.g. Joe (1997)].

²¹The real Laplace transform of variable Z is the function defined by:

$$\psi(u) = E \left[\exp(-uZ) \right].$$

When Z is positive, its domain of definition includes \mathbb{R}_+ .

$\nu'(0) = 1$ and: $V(Z_1) = \psi_1''(0) - \psi_1'(0)^2 = \psi_2''(0) - \psi_2'(0)^2 - \nu''(0) = V(Z_2) - \nu''(0) > V(Z_2)$. Thus, when condition (1.18) is satisfied, factor Z_1 is more dispersed than factor Z_2 , and the Archimedean copula C_{ϕ_1} is more PQD than C_{ϕ_2} .

iv) Tail dependence

Let us now consider tail dependence in Archimedean copulas. From section 2.2 iii) the coefficient of upper tail dependence is given by [see Joe (1997), Theorem 4.12]:

$$\lambda_U = 2 - 2 \lim_{s \rightarrow 0} \frac{\psi'(2s)}{\psi'(s)},$$

where $\psi = \phi^{-1}$. When the Archimedean copula is derived from a latent factor model, tail dependence is characterized by the magnitude of the factor tail. If the factor Z has a finite mean, $-\psi'(0) = E[Z] < \infty$, then $\lambda_U = 0$. Thus Archimedean copulas may feature upper tail dependence only if $E[Z] = \infty$. Let us assume for instance a factor Z with Pareto tails: $P[Z \geq z] \sim l(z)/z^\delta$, where l is a slowly varying function, and $0 \leq \delta \leq 1$ [which implies $E[Z] = +\infty$]. Then, by the Tauberian Theorem [see Feller (1971)], the Laplace transform ψ of Z near $s = 0$ is such that: $\psi(s) \sim 1 - s^\delta l(1/s)$, and thus: $\psi'(2s)/\psi'(s) \sim 2^{\delta-1}$, $s \rightarrow 0$. Therefore the tail dependence parameter is given by: $\lambda_U = 2 - 2^\delta$. The smaller δ , the fatter the tail of Z , and the stronger tail dependence in the copula.

v) Parametric Archimedean copula

Classical Archimedean copulas are the Gumbel copula and the Clayton copula.

The Gumbel copula [Gumbel (1960)] is defined by:

$$C_\alpha(u, v) = \exp \left\{ - [(-\log u)^\alpha + (-\log v)^\alpha]^{1/\alpha} \right\},$$

where $1 \leq \alpha \leq \infty$. The Gumbel family C_α corresponds to a positive α -stable factor Z , with a real Laplace transform $\psi_\alpha(s) = \exp[-s^{1/\alpha}]$, $s \geq 0$. Its generator is given by: $\phi_\alpha(u) = (-\log u)^\alpha$, $u \in (0, 1]$. The Gumbel family C_α is positively PQD ordered with respect to α , since $\phi_{\alpha_2} \circ \phi_{\alpha_1}^{-1}(s) = s^{\alpha_2/\alpha_1}$, $s \geq 0$, is sign alternating if $\alpha_1 \geq \alpha_2$. In accordance with the general result, the PQD ordering is related to the dispersion of the factor Z : the smaller is α , the faster is the decay of the Laplace transform ψ_α , and thus the more concentrated the distribution of Z close to 0. Finally, since $\psi_\alpha(s) \sim 1 - s^{1/\alpha}$, when $s \rightarrow 0$, the Gumbel copula features upper tail dependence, with $\lambda_U = 2 - 2^{1/\alpha}$. The upper tail dependence coefficient is increasing with respect to α . This is consistent with the interpretation of α as the tail parameter of the distribution of Z : $P[Z \geq z] \sim z^{-1/\alpha}$, $z \rightarrow \infty$.

The Kimeldorf and Sampson copula, also called Clayton copula, [Kimeldorf, Sampson (1975), Clayton (1978)] is given by:

$$C_\delta(u, v) = (u^{-\delta} + v^{-\delta} - 1)^{-1/\delta}, \quad u, v \in [0, 1],$$

where $\delta \geq 0$, with generator: $\phi_\delta(u) = 1 - u^{-\delta}$, $u \in (0, 1]$. Clayton copula corresponds to a common factor Z following a gamma distribution with parameter δ , and Laplace transform

$\psi_\delta(s) = 1/(1+s)^{1/\delta}$, $s \geq 0$. This family is positively PQD ordered with respect to δ , and features lower tail dependence.

vi) Scatterplots and isodensity curves

Scatterplots of (U, V) and isodensity curves of random variables (X, Y) with standard normal marginal distributions for Gumbel (resp. Clayton) copula are reported in Figure 3, Panels A and B (Panels C and D, respectively). The parameters are chosen to get Kendall's tau identical to that of a Gaussian copula with $\rho = 0.8$.

[insert Figure 3A, B: scatterplot and isodensity curves, Gumbel copula]

[insert Figure 3C, D: scatterplot and isodensity curves, Clayton copula]

Two important differences emerge with respect to the Gaussian copula (see Figure 1, Panel D, and Figure 2, Panel D). First, isodensity curves are wedge-shaped in the upper (resp. lower) quadrant for the Gumbel (resp. Clayton) copula. Intuitively the dependence between large observations is stronger than dependence between small ones for the Gumbel copula, and the reverse holds for the Clayton copula. Such an asymmetry in dependence is not possible with gaussian copulas. The second, and related, important difference evidenced by the scatterplots is that Gumbel (Clayton) copula features upper (resp. lower) tail dependence, whereas Gaussian copula features tail independence in both tails.

1.2.3.3 Extreme value copulas

i) Definition

Let (X_i^*, Y_i^*) , $i = 1, \dots, n$, be independent pairs of random variables. Extreme value copulas are associated with the limiting joint distribution of $X = \max_i X_i^*$, $Y = \max_i Y_i^*$, as n tends to infinity. Extreme value copulas are of the form [see e.g. Joe (1997)]:

$$C_\chi(u, v) = \exp \left\{ (\log u + \log v) \chi \left(\frac{\log u}{\log u + \log v} \right) \right\},$$

where function χ is defined on $[0, 1]$, convex, and satisfies: $\max(v, 1-v) \leq \chi(v) \leq 1$, $\forall v \in [0, 1]$.

ii) PQD dependence

Positive Quadrant Dependence is easily characterized in terms of the functional parameter χ . If C_1 and C_2 are two extreme value copulas, with generators χ_1 and χ_2 , respectively, then C_1 is more PQD than C_2 iff $\chi_1 \leq \chi_2$. The limiting generators $\chi_1(v) = \max(v, 1-v)$ and $\chi_2(v) = 1$ correspond to positive deterministic dependence, and independence, respectively. In particular, only positive dependence is allowed.

iii) Parametric extreme value copulas

The Gumbel copula is an example of extreme value copula, with generator $\chi(v) = \left(v^\delta + (1-v)^\delta\right)^{1/\delta}$. Another example is provided by the asymmetric logistic copula:

$$C_{\alpha, \beta_1, \beta_2}(u, v) = \exp \left\{ - [(\beta_1 z)^\alpha + (\beta_2 w)^\alpha]^{1/\alpha} - (1 - \beta_1) z - (1 - \beta_2) w \right\}, \quad (1.19)$$

where $z = -\log u$, $w = -\log v$, and $\beta_1, \beta_2 \in [0, 1]$, $\alpha \geq 1$ ²². The generator is given by: $\chi(v) = [(\beta_1 v)^\alpha + (\beta_2 (1-v))^\alpha]^{1/\alpha} + (1 - \beta_1) v + (1 - \beta_2) (1 - v)$.

iv) Scatterplots and isodensity curves

Scatterplots of variables (U, V) with asymmetric logistic copula and parameters $\alpha = 2.62$, $\beta_1 = 1.3$, $\delta = 0.8$ ²³, and isodensity curves of corresponding variables (X, Y) with standard normal marginal distributions are reported in Figure 4, Panels A and B, respectively.

[insert Figure 4A: scatterplot, asymmetric logistic copula]

[insert Figure 4B : isodensity curves, gaussian marginal distribution, asymmetric logistic copula]

The most important difference with respect to previous examples is that variables U and V do not play a symmetric role: the copula is not exchangeable, that is the distribution of (U, V) is not invariant for permutation of the components: $C(u, v) \neq C(v, u)$, $u, v \in [0, 1]$. Realizations with U small and V large are more likely than the opposite ones. In the asymmetric logistic copula family asymmetry is described by parameters β_1, β_2 . The copula is symmetric when $\beta_1 = \beta_2$. More generally, an extreme value copula is symmetric if and only if its generator χ is such that: $\chi(v) = \chi(1 - v)$, $\forall v \in [0, 1]$.

1.3 Other functional dependence measures.

As seen in section 2 copulas summarize the dependence properties which are invariant with respect to increasing transformations. Thus they do not take into account the levels of both variables. Other characterizations of dependence can be introduced and be more appropriate for some applications. In section 3.1, we consider the so-called nonlinear canonical analysis. We first define nonlinear canonical correlations and canonical directions, and provide their interpretations. We then describe the nonlinear canonical decomposition of a bivariate distribution, and of the associated copula density. Finally we introduce constrained non-parametric specifications based on a finite-dimensional canonical decomposition. In section 3.2 the distribution is characterized by means of a marginal and a conditional Laplace transform. We review the basic properties of Laplace transforms and introduce a nonparametric constrained model for the conditional Laplace transform.

²²This copula is obtained by the asymmetrization technique of Genest, Ghoudi, Rivest (1998). In this case it is derived by mixing a Gumbel copula and an independent copula.

²³These parameters are such that the upper tail dependence coefficient is equal to that of the Gumbel copula considered in ii).

1.3.1 Canonical analysis

1.3.1.1 Definition of canonical correlations and canonical directions

Let X, Y be a pair of random variables with joint distribution F and marginal distributions F_X, F_Y . The canonical directions of order j , $\varphi_j \in L^2(F_X)$, $\psi_j \in L^2(F_Y)$, and the associated canonical correlations λ_j , $j \in \mathbb{N}$, are defined recursively by:

$$\lambda_j = \text{corr} [\varphi_j(X), \psi_j(Y)] = \max_{\varphi \in L^2(F_X), \psi \in L^2(F_Y)} \text{corr} [\varphi(X), \psi(Y)],$$

where the maximization is subject to the constraints²⁴:

$$\varphi \perp \text{span} \{\varphi_1, \dots, \varphi_{j-1}\}, \quad \psi \perp \text{span} \{\psi_1, \dots, \psi_{j-1}\},$$

and to the normalizations:

$$E[\varphi(X)] = E[\psi(Y)] = 0, \quad V[\varphi(X)] = V[\psi(Y)] = 1.$$

More explicitly, the first canonical directions φ_1 and ψ_1 are the transformations of X and Y with maximal correlation (equal to λ_1), given zero mean and unitary variance restrictions; the second canonical directions φ_2 and ψ_2 are the transformations uncorrelated with φ_1 and ψ_1 , with the maximal correlation (equal to λ_2); and so on. Thus nonlinear canonical analysis provides a sequence of orthonormal functional directions of nonlinear dependence [the canonical directions φ_j, ψ_j , j varying] and associated canonical correlations $\lambda_j \geq 0$, j varying, in decreasing dependence order: $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$.

Example 2: Let (X, Y) follow a bivariate standard gaussian distribution with correlation $\rho > 0$. Thus the canonical directions are given by [Barrett, Lampard (1955), Wiener (1958), lecture 5, Neveu (1968)]:

$$\varphi_j = \psi_j = \frac{1}{\sqrt{j!}} H_j, \quad j \in \mathbb{N},$$

where H_j , j varying, are the Hermite polynomials defined by:

$$H_j(x) = \sum_{0 \leq m \leq \lfloor j/2 \rfloor} \frac{j!}{(j-2m)!m!2^m} (-1)^m x^{j-2m}.$$

The associated canonical correlations are:

$$\lambda_j = \rho^j, \quad j \in \mathbb{N}.$$

For a bivariate gaussian distribution the directions of stronger dependence are affine, and the largest canonical correlation coincides with the linear correlation ρ . The canonical correlations λ_j decrease geometrically with the order $j \in \mathbb{N}$, and the associated directions correspond to polynomial transformations of increasing degree.

²⁴ \perp denotes orthogonality with respect to the standard inner product in $L^2(F_X)$ and $L^2(F_Y)$, respectively. Thus $\varphi \perp \varphi_1$ means $E[\varphi(X)\varphi_1(X)] = 0$, and $\psi \perp \psi_1$ means $E[\psi(Y)\psi_1(Y)] = 0$.

1.3.1.2 The decomposition theorem

The canonical directions can be used to get a decomposition of the joint distribution separating the effect of the marginal distributions and the nonlinear dependence. Let φ_j, ψ_j and λ_j, j varying, be the canonical directions and canonical correlations of variables X and Y , respectively. Let us denote by f [resp. f_X, f_Y] the joint pdf of (X, Y) [the marginal pdf of X and Y , respectively]. Under weak conditions²⁵, the joint p.d.f. of X and Y admits the decomposition [see Lancaster (1968), Dunford, Schwartz (1968)]:

$$f(x, y) = f_X(x)f_Y(y) \left[1 + \sum_{j=1}^{\infty} \lambda_j \varphi_j(x) \psi_j(y) \right]. \quad (1.20)$$

The canonical decomposition (1.20) provides another characterization of the joint p.d.f. of (X, Y) , where marginal effects f_X, f_Y are clearly distinguished.

1.3.1.3 Invariance by increasing transformation

By definition the canonical correlations $\lambda_j, j \in \mathbb{N}$, are invariant to increasing transformations of the variables X and Y , whereas the canonical directions $\varphi_j, \psi_j, j \in \mathbb{N}$, are transformed by compounding: the canonical directions of $g(X), h(Y)$, where g, h are increasing transformations, are $\varphi_j \circ g^{-1}, \psi_j \circ h^{-1}, j \in \mathbb{N}$. Thus the canonical decomposition of the distribution of X and Y is characterized, up to increasing transformations of the canonical directions, by the canonical decomposition of the copula p.d.f.:

$$c(u, v) = 1 + \sum_{j=1}^{\infty} \lambda_j a_j(u) b_j(v), \text{ say,}$$

where $\varphi_j = a_j \circ F_X, \psi_j = b_j \circ F_Y, j \in \mathbb{N}$. Canonical directions $a_j, b_j, j \in \mathbb{N}$, satisfy the following normalization and orthogonality conditions with respect to the uniform distribution on $[0, 1]$:

$$\begin{aligned} \int_0^1 a_j(u) du &= \int_0^1 b_j(v) dv = 0, \forall j, \\ \int_0^1 a_i(u) a_j(u) du &= \int_0^1 b_i(v) b_j(v) dv = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}, \forall i, j. \end{aligned}$$

²⁵Such as $I = \int \int f(x, y)^2 / [f_X(x) f_Y(y)] dx dy < \infty$. Since $I = \int \int \lambda(x, y)^2 f_X(x) f_Y(y) dx dy$ is an average of nonlinear dependence measure $\lambda(x, y) = f(x, y) / f_X(x) f_Y(y)$, the condition means that the dependence is not too strong. When $I < \infty$, we deduce from (1.20) that I is related to the canonical correlations by $I = \sum_{j=1}^{\infty} \lambda_j^2 < +\infty$. Thus canonical correlations λ_j, j varying, cannot decrease too slowly.

1.3.1.4 Finite dimensional dependence

From the decomposition theorem, the joint p.d.f. of (X, Y) may be equivalently represented by a nonlinear canonical decomposition as in the RHS of equation (1.20). This suggests model specifications with constrained nonparametric dependence by imposing a finite number of non-zero canonical correlations. We say that these models feature finite dimensional dependence [see Gouriéroux, Jasiak (2001)]. Let us consider one-dimensional dependence. The copula admits the representation:

$$c(u, v) = 1 + \lambda a(u)b(v),$$

where the canonical directions a and b satisfy the normalizations:

$$\begin{aligned} \int_0^1 a(u)du &= \int_0^1 b(v)dv = 0, \\ \int_0^1 a(u)^2 du &= \int_0^1 b(v)^2 dv = 1. \end{aligned} \quad (1.21)$$

By the positivity condition of the copula density, the canonical directions a and b are bounded: $\underline{a} \leq a \leq \bar{a}$, $\underline{b} \leq b \leq \bar{b}$, for some constants $\underline{a}, \underline{b} \leq 0 \leq \bar{a}, \bar{b}$, and the canonical correlation λ is constrained by: $\lambda \leq -1/\min\{\underline{a}\bar{b}, \bar{a}\underline{b}\}$.

Let us study the positive quadrant dependence between X and Y . The copula c.d.f. is given by:

$$C(u, v) = uv + \lambda A(u)B(v),$$

where $A(u) = \int_0^u a(w)dw$, and $B(v) = \int_0^v b(w)dw$. Thus (X, Y) features PQD iff:

$$\text{either } \begin{cases} A \geq 0 \\ B \geq 0 \end{cases}, \text{ or } \begin{cases} A \leq 0 \\ B \leq 0 \end{cases},$$

that is A and B have the same constant sign. Since functions A and B vanish at the boundaries $u = 0$ and $u = 1$, this condition is satisfied for instance when functions a and b are either both monotonically decreasing, or both monotonically increasing.

In a model with one-dimensional dependence, PQD is increasing with respect to the canonical correlation λ . Let us now investigate how the patterns of canonical directions a and b affect the strength of PQD. Let (X, Y) and (X^*, Y^*) feature one-dimensional dependence, with canonical directions a, b and a^*, b^* , and canonical correlations λ and λ^* , respectively. Let us assume for simplicity $A, B, A^*, B^* \geq 0$. Then (X, Y) is more PQD than (X^*, Y^*) iff:

$$\lambda A(u)B(v) \geq \lambda^* A^*(u)B^*(v), \quad \forall u, v,$$

or equivalently:

$$\begin{aligned} \frac{\lambda}{\lambda^*} \frac{A(u)}{A^*(u)} \frac{B(v)}{B^*(v)} &\geq 1, \quad \forall u, v \\ \iff \frac{\lambda}{\lambda^*} \min \frac{A}{A^*} \min \frac{B}{B^*} &\geq 1. \end{aligned}$$

Example 3: An example of copula with one-dimensional dependence is Morgenstern copula [Morgenstern (1956)]:

$$c(u, v) = 1 + 3\lambda(1 - 2u)(1 - 2v),$$

where $|\lambda| \leq 1/3$. The canonical directions $a(w) = b(w) = \sqrt{3}(1 - 2w)$ are affine. In particular the linear correlation between U and V is equal to λ and is necessarily smaller than $1/3$ in absolute value. When $\lambda > 0$ this copula features PQD.

1.3.2 Laplace transforms

In the nonlinear canonical decomposition the joint pdf plays an important role as the appropriate characterization of the joint distribution. In this section we consider the characterization of a distribution by means of the Laplace transform (also called moment generating function). More precisely, it is well-known that the joint distribution is characterized by a marginal distribution and a conditional one. We will use the equivalent decomposition in terms of Laplace transform.

1.3.2.1 Definition

The (real) Laplace transform of a variable Y is the function:

$$u \mapsto \Psi_Y(u) = E[\exp(-uY)], \quad u \in \mathbb{D} \subset \mathbb{R}.$$

The domain of definition \mathbb{D} is an interval including 0. The expectation does not necessarily exist for any real value of u , but the larger the domain \mathbb{D} , the thinner are the tails of the distribution of Y .

Important properties of the Laplace transform are known for positive variables. The domain \mathbb{D} contains the positive half-line \mathbb{R}_+ . In addition, function Ψ is completely monotone²⁶: $(-1)^{(m)} d^m \Psi_Y(u) / du^m = E[Y^m \exp(-uY)] \geq 0$, $u \geq 0$, $\forall m \in \mathbb{N}$. Finally, the real Laplace transform Ψ on \mathbb{R}_+ characterizes completely the distribution of Y . Thus, in this case, the knowledge of Ψ is equivalent to the knowledge of the distribution of Y ²⁷.

Let us now consider two random variables X and Y . Their nonlinear dependence may be described by the conditional distribution of X given Y . Its Laplace transform is given by:

$$u \mapsto \Psi(u, Y) = E[\exp(-uX) | Y].$$

Under weak conditions²⁸, the conditional Laplace transform $\Psi(u, Y)$ characterizes the conditional distribution, and provides an equivalent description of nonlinear dependence, which is appropriate in many applications.

²⁶Together with $\Psi(0) = 1$, this is an equivalent characterization of Laplace transforms for positive variables [see Feller (1971)].

²⁷In the general case, we have to consider the Laplace transform defined on the complex domain: $\Psi_Y(w) = E[\exp(-wY)]$, $w \in \mathbb{D}^{\mathbb{C}} \subset \mathbb{C}$.

²⁸For instance X is positive, or $w \mapsto \Psi(w, Y)$ is analytic in a neighbourhood of $0 \in \mathbb{C}$.

Finally, the joint Laplace transform of X and Y , which characterizes the joint distribution of (X, Y) , is defined by:

$$(u, v) \mapsto \Psi(u, v) = E[\exp(-uX - vY)].$$

It can be written in terms of the conditional Laplace transform of X given Y and the marginal distribution of Y as:

$$\Psi(u, v) = E[\Psi(u, Y) \exp(-vY)]. \quad (1.22)$$

1.3.2.2 The compound model

i) Definition

Constrained nonparametric specifications of the nonlinear dependence may be introduced by imposing functional restrictions on the conditional Laplace transform $\Psi(u, Y)$. In the so-called compound model, $\Psi(u, Y)$ is an exponential affine function of the conditioning variable:

$$\Psi(u, Y) = E[\exp(-uX) | Y] = \exp[-a(u)Y - b(u)], \quad (1.23)$$

where a and b are one dimensional functions defined on a domain $\mathbb{D} \subset \mathbb{R}$. Notice that this implies $\Psi_X(u) = \Psi_Y[a(u)] \exp[-b(u)]$. From (1.22) the joint Laplace transform of (X, Y) becomes:

$$\begin{aligned} \Psi(u, v) &= E[\Psi(u, Y) \exp(-vY)] \\ &= E[\exp(-[a(u) + v]Y - b(u))] \\ &= \Psi_Y[a(u) + v] \exp[-b(u)] \\ &= \Psi_X(u) \frac{\Psi_Y[a(u) + v]}{\Psi_Y[a(u)]}. \end{aligned} \quad (1.24)$$

Example 4: Let (X, Y) follow a bivariate standard gaussian distribution with linear correlation ρ . Then the conditional distribution of X given Y is gaussian $N(\rho Y, 1 - \rho^2)$, and we get:

$$E[\exp(-uX) | Y] = \exp\left[-u\rho Y + \frac{1}{2}(1 - \rho^2)u^2\right].$$

Thus (X, Y) follows a compound model with $a(u) = \rho u$, and $b(u) = -\frac{1}{2}(1 - \rho^2)u^2$, $u \in \mathbb{R}$.

Example 5: Exponential affine conditional Laplace transforms are typical in compound risk aggregation, which explains the name of the model. As an illustration let us denote by Y the number of car accidents during a year. Moreover let us assume that the costs of the claims (W_i) of the different accidents are independent of Y and i.i.d., with Laplace

transform $E[\exp(-uW)] = \exp[-a(u)]$. Further let V denote an additional loss (the usual "fixed" cost), which is not due to car accidents, and is independent from Y and (W_i) with Laplace transform $\exp[-b(u)]$. Then the nonlinear dependence between the total loss $X = \sum_{i=1}^Y W_i + V$ and the number of accidents Y is such that:

$$E[\exp(-uX) | Y] = \{\exp[-a(u)]\}^Y \exp[-b(u)] = \exp[-a(u)Y - b(u)].$$

ii) Nonlinear dependence

Let us now discuss nonlinear dependence in the compound model. For simplicity, we consider the case of positive variables, satisfying the condition of Example 5: $X = \sum_{i=1}^Y W_i + V = Z(Y) + V$, (say), where V , W_i , i varying, are independent positive variables. In particular the values $a(u)$, $b(u)$ are nonnegative for positive argument u .

It is rather difficult to characterize the PQD ordering in terms of the functional parameter a summarizing the dependence. However it is possible to study a slightly weaker ordering. Let us consider two pairs of variables (X_1, Y_1) , (X_2, Y_2) following the compound model of Example 5, with functional parameters a_1 , a_2 , respectively, and identical marginal distributions. Then we get:

$$\text{Cov}(g(X_1), h(Y_1)) \geq \text{Cov}(g(X_2), h(Y_2))$$

for any increasing functions g , h which are limits of positive linear combinations of affine and increasing exponential functions, if and only if: $a_1(u) \geq a_2(u)$, $\forall u \geq 0$ [see Appendix 1]. In particular any pair (X, Y) satisfying the conditions of example 5 features PQD, since $a(u) \geq 0$, $\forall u \geq 0$ [see Appendix 1].

iii) Link with canonical analysis

Finally, it is possible to relate the approach based on Laplace transforms with the approach based on nonlinear canonical analysis. For instance, for a compound model with symmetric joint distribution, the nonlinear canonical decomposition has been derived in Darolles, Gouriéroux, Jasiak (2002). For positive variables, the canonical correlations are geometrically decreasing and related to function a by $\lambda_j = [da/du(0)]^j$, $j \in \mathbb{N}$, whereas the canonical directions $\varphi_j = \psi_j$ are appropriate polynomials with increasing degrees.

1.4 Applications: static framework.

In this section we consider economic and financial applications involving contemporaneous dependence between two risk variables X and Y . According to the problem, different notions of nonlinear dependence are concerned. We discuss the appropriateness of the different specifications of nonlinear dependence introduced in sections 2 and 3.

1.4.1 Scoring comparison

A typical example of the usefulness of copula is the comparison of grades (or scores, or ratings) attributed to individuals by two different auditors. For instance these grades can represent the results at two intelligence quotient (IQ) tests, the results in mathematics and literature of a set of students, or two scoring for default. The latter example is used in this section for illustration.

Let us consider a set of borrowers $i = 1, \dots, n$ and the indicator variable representing default during a given period: $Y_i = 1$ if default occurs, 0, otherwise. The prediction of default is usually performed by means of a dichotomous qualitative model, which assumes:

$$P[Y_i = 1 | X_i] = G(z_i' b),$$

where X_i are observed explanatory characteristics, z_i transformations of the basic explanatory characteristics introduced to include nonlinear or cross effects, b unknown parameters, and G a cdf [see e.g. Gourioux, Jasiak (2002b)]. A logit (resp. probit) specification is selected when G is the cdf of the logistic distribution (resp. standard normal distribution). The dichotomous specification can be used in two different ways: we can use either the structural model to approximate the probability of default by $G(z_i' \hat{b})$, or simply the score $S_i = z_i' \hat{b}$ to rank the individuals by increasing grade. For the latter application the score is defined up to an increasing transformation and the same individual ranking can be derived by introducing the standardized score $U = F_S(S)$, where F_S is the marginal distribution of S . Note that, while the endogenous variable Y is qualitative, the set of exogenous variables z includes generally quantitative covariates, which will imply a continuous score.

In practice it is usual to compare different scores S and S^* , say, corresponding to an old score and an updated score. They differ by the choice of the set of transformed variables z and by the selected cdf G . The compatibility of the scores S and S^* can be analyzed by means of the copula. Intuitively, the larger is the dependence in their copula, the more compatible are the two scores. In this example Kendall's tau measures the probability that two borrowers are ranked in the same way by the two scores S and S^* . In particular S and S^* provide identical ranking if and only if the Kendall's tau is equal to one.

1.4.2 Competing default risks

The standard competing risk models assume independent competing events and focus on the distribution of the date of arrival of the first event (or second, third ... one). Mathematically they assume independent duration variables and study the distribution of the corresponding order statistics. It is interesting to see if the usual results are modified when the event occurrences are linked.

To understand the interest of the question, we can consider a credit portfolio, including homogeneous credits, that are credits with the same design (initial balance, interest rate, maturity, pattern of monthly payments) granted to similar borrowers. The borrowers can default and the defaults can be characterized by the duration variables giving the time before default for each individual. Since there exist different credit derivatives providing for

instance 1 money unit if the first (resp. the second ...) default occurs before maturity²⁹, it is necessary to know the distribution of the order statistics to price these derivatives. By assuming the independence between the competing risks (the individual defaults in the example), we neglect the possibility of default correlation and as a consequence the derivative assets are likely mispriced.

Let us discuss the distribution of the order statistics for two homogeneous competing risks (say), when the risks are dependent. The two duration variables are denoted by X and Y , and have the same marginal distribution F_X , because of the homogeneity assumption. We note that:

$$\begin{aligned} X &= F_X^{-1}(U), Y = F_X^{-1}(V), \\ \min(X, Y) &= F_X^{-1}(\min(U, V)), \quad \max(X, Y) = F_X^{-1}(\max(U, V)). \end{aligned}$$

Thus the dependence properties can be studied with respect to the standardized durations U, V only, since the X, Y variables and the order statistics $\min(X, Y)$ and $\max(X, Y)$ are deduced from the U, V variables and the corresponding order statistics $\min(U, V)$ and $\max(U, V)$ by some transformation F_X^{-1} , which can be interpreted as a nonlinear change of time unit (time deformation). Moreover, the copula of the order statistics depends only on the copula of (X, Y) , and not on the marginal distribution F_X .

The magnitude of the dependence between U and V has both an effect on the marginal distributions of $\min(U, V)$ and $\max(U, V)$ and on their dependence. These effects can be illustrated by considering the limiting cases of deterministic dependence. The highest positive dependence is obtained for $V = U$. Then $\min(U, V) = \max(U, V) = U$, and the two order statistics coincide. The smallest dependence is obtained for $V = 1 - U$. Then $\min(U, V) = \min(U, 1 - U) = 1 - \max(U, 1 - U) = 1 - \max(U, V)$, and the order statistics are in negative deterministic dependence. The difference $\max(U, 1 - U) - \min(U, 1 - U) = |2U - 1|$ is likely the largest possible one.

These results will be clarified by considering the marginal distributions of $\min(U, V)$ and $\max(U, V)$, and their copula.

i) Marginal distributions

The marginal distributions of $\min(U, V)$ and $\max(U, V)$ are given by:

$$\begin{aligned} P[\min(U, V) \leq x^*] &= 1 - P[U \geq x^*, V \geq x^*] \\ &= 2x^* - C(x^*, x^*) = \varphi_C(x^*), \text{ say,} \end{aligned} \tag{1.25}$$

and:

$$P[\max(U, V) \leq y^*] = P[U \leq y^*, V \leq y^*] = C(y^*, y^*) = \psi_C(y^*), \text{ say.}$$

Both marginal distributions depend on the value of the copula on the diagonal. In particular when the copula increases, that is when the standardized durations U, V are more PQD, the survivor function of $\min(U, V)$ [resp. $\max(U, V)$] increases [resp. decreases]. Therefore $\min(U, V)$ [resp. $\max(U, V)$] is larger [resp. smaller] for first order stochastic dominance.

²⁹These derivatives are called first (second, ...) to default baskets.

ii) Copula

Let us now compute the copula of $(\min(U, V), \max(U, V))$. It is given by [see Appendix 2]:

$$C^*(u, v) = C(\varphi_C^{-1}(u), \psi_C^{-1}(v)) + C(\psi_C^{-1}(v), \varphi_C^{-1}(u)) \\ + u - 2\varphi_C^{-1}(u),$$

for u, v satisfying $C[(u+v)/2, (u+v)/2] \leq v$, and

$$C^*(u, v) = v,$$

otherwise.

1.4.3 Age structure of default correlation

Let us still consider the example of a credit portfolio including two similar credits. It is interesting to analyse the seasoning effect, that is the evolution of the risk during the life of the portfolio. Let us consider a given age h at which the two borrowers are still alive. It is natural to study the dependence between the residual durations $X - h, Y - h$. Their copula at age h coincides with the copula of (X, Y) conditional to $X > h, Y > h$, and is directly related to the tail conditional copulas C_α introduced in section 2.2 iv). Thus, after an appropriate time deformation, the term structure of default correlation, that is the way the copula depends on the age, corresponds to the application $\alpha \rightarrow C_\alpha$. In particular, the limiting copula $\lim_{\alpha \rightarrow 0} C_\alpha$ measures the default dependence in the long run.

Example 6: Let us assume that the two durations X, Y are independent conditionally to a factor Z with constant hazard function Z . Then the conditional survivor function of X, Y is:

$$P[X \geq x, Y \geq y | Z] = \exp[-Z(x + y)],$$

and by integrating out factor Z we get the unconditional survivor function: $P[X \geq x, Y \geq y] = \Psi(x + y)$, where Ψ denotes the Laplace transform of factor Z . Thus the survivor copula corresponding to age $h = 0$ is Archimedean (see section 2.3.2):

$$C^s(u, v) = \Psi[\Psi^{-1}(u) + \Psi^{-1}(v)].$$

At age h the survivor function of residual times becomes:

$$P[X \geq x + h, Y \geq y + h | X > h, Y > h] = \frac{P[X \geq x + h, Y \geq y + h]}{P[X > h, Y > h]} \\ = \frac{\Psi(x + y + 2h)}{\Psi(2h)}.$$

Thus the survivor copula at age h is:

$$C_h^s(u, v) = \frac{\Psi[\Psi^{-1}(u\Psi(2h)) + \Psi^{-1}(v\Psi(2h)) - 2h]}{\Psi(2h)},$$

which is an Archimedean copula with inverse generator $\Psi_h(s) = \Psi(s + 2h) / \Psi(2h)$. Thus for the Archimedean family the pattern of the term structure of default correlation is characterized by the Laplace transform of the factor Z . Let us for instance assume that the factor Z has a gamma distribution with parameter δ , corresponding to a Clayton survivor copula with $\Psi(s) = (1 + s)^{-1/\delta}$. Then the survivor copula at age h has inverse generator $\Psi_h(s) = \Psi[s/(1 + 2h)]$. Since the Archimedean copula is invariant to scale transformations of the factor Z , the survivor copula at age h is still a Clayton copula with parameter δ . Thus the age structure of default correlation is flat in this case. In general the pattern of the term structure of default correlation is related to the derivatives of Laplace transform Ψ . For instance, by means of characterization (1.18) of the PQD ordering in the Archimedean family, it can be shown that default correlation increases with the age h , that is $C_{h_1}^s$ is more PQD than $C_{h_2}^s$ for $h_1 \geq h_2$, if the Laplace transform Ψ satisfies the condition:

$$\frac{\Psi''(s)\Psi(s)}{\Psi'(s)^2} \text{ is increasing in } s.$$

This condition is equivalent to the fact that $Var_s(Z)/E_s(Z)^2$ is increasing in s , where E_s and Var_s denote expectation and variance with respect to the distribution $\exp(-sz)G(dz)$, where G is the cdf of Z . When s increases, the moments E_s and Var_s downweight progressively large values of Z . Intuitively this condition requires that the variability of the distribution of the factor Z is located sufficiently close to the origin.

Finally, the long term behaviour of default correlation depends on the asymptotic behaviour of the Laplace transform of Z at infinity. For instance, from Juri, Wüthrich (2002a,b), if the Laplace transform of Z is asymptotically equivalent to the Laplace transform of a gamma distribution:

$$\Psi(s) \sim s^{-1/\delta}, \quad s \rightarrow \infty, \quad (1.26)$$

with $\delta > 0$, then the survivor copula C_h^s converges to a Clayton copula with parameter δ when the age h goes to infinity:

$$\lim_{h \rightarrow \infty} C_h^s(u, v) = (u^{-\delta} + v^{-\delta} - 1)^{-1/\delta},$$

By the Tauberian theorem, the tail of the Laplace transform of Z is related to the distribution of Z at the origin. In particular, condition (1.26) is equivalent to [see Feller (1971), chapter 13]:

$$P[Z \leq z] \sim \frac{1}{\Gamma(1 + 1/\delta)} z^{1/\delta}, \quad z \rightarrow 0.$$

The larger is δ , the more concentrated the distribution of Z close to the origin, and the stronger is PQD in the long term default copula.

1.4.4 Income and wealth inequality

An important question in inequality theory is how to combine income and wealth. Indeed, whereas an inequality ordering is clearly defined for a scalar variable by means of the Lorenz curve, the analogous does not exist for a pair of variables. A natural idea to circumvent this

difficulty is to transform any wealth into a regular income. For instance let us assume an (indirect) intertemporal utility function of the type $\sum_{k=0}^{\infty} \delta^k u(\pi_k)$, where δ is the discounting factor and (π_k) an income pattern. We can apply an equivalent utility principle to define the regular income flow which is equivalent to the wealth by solving:

$$u(W) = \sum_{k=0}^{\infty} \delta^k u(\pi_W),$$

where W is the wealth level and π_W the implied regular income. Note that π_W is an increasing function of the wealth. Then the inequality could be measured on the aggregate income $\pi + \pi_W$, defined as the sum of the current income and the wealth equivalent income. But an alternative approach can be based on wealth by considering the inequality measured on the aggregate wealth $W + W_\pi$, where W is the current wealth and W_π is the implied wealth defined by:

$$W_\pi = u^{-1} \left[\sum_{k=0}^{\infty} \delta^k u(\pi) \right] = u^{-1} \left[\frac{1}{1-\delta} u(\pi) \right].$$

Intuitively a larger PQD between income and wealth, with fixed marginal distributions, will increase the inequality measures defined in either equivalent incomes, or equivalent wealths. More precisely, we have to show that equivalent income and equivalent wealth are decreasing with respect to PQD between income and wealth (with fix marginal distributions), in the sense of second order stochastic dominance: the more PQD are income and wealth, the riskier are equivalent income and equivalent wealth. Since equivalent income [resp. equivalent wealth] is of the form $\pi + h(W)$ [$W + h^*(\pi)$, respectively], with h (and h^*) a positive increasing function, we have to show that for any pairs of random variables (X_1, Y_1) and (X_2, Y_2) with identical marginal distributions, such that $(X_1, Y_1) \succeq_{PQD} (X_2, Y_2)$:

$$E[u(X_1 + h(Y_1))] \leq E[u(X_2 + h(Y_2))], \quad (1.27)$$

for any concave, increasing utility function u , and any positive increasing function h . The result follows from Proposition 2, since the function $g(X, Y) = -\{u[X + h(Y)] - u(X) - u[h(Y)] + u(0)\}$ is the cumulative function of the positive measure³⁰ with density $-u''[X + h(Y)]h'(Y) \geq 0$.

1.4.5 Moment based problems

Different applications require the computation of a nonlinear cross moment between two variables X and Y , such as:

$$E[g(X, Y; A)],$$

where $g(\cdot, \cdot; A)$ is a parametric family of functions. According to the problem, the parametric family is different and involves a different notion of nonlinearity.

³⁰Without loss of generality we can assume that $h(0) = 0$.

1.4.5.1 Expected Utility

Let us consider an investor allocating wealth between two risky assets, with prices $p_{X,t}$ and $p_{Y,t}$ at date t , respectively, and a riskfree asset over an horizon of h periods. We assume that the investor has a Constant Absolute Risk Aversion (CARA) utility function $u(W; A) = -\exp(-AW)$, $W \geq 0$, where $A > 0$ denotes the absolute risk aversion parameter. Let us denote by α_0 and $\alpha = (\alpha_X, \alpha_Y)$ the allocations in the riskfree asset and in the risky assets, respectively. The optimal portfolio is determined by:

$$\arg \max_{\alpha_0, \alpha} E[-\exp(-A[\alpha_X p_{X,h} + \alpha_Y p_{Y,h} + \alpha_0(1 + r_{0,h})])],$$

subject to the budget constraint: $\alpha_X p_{X,0} + \alpha_Y p_{Y,0} + \alpha_0 = W_0$, where $r_{0,h}$ is the interest rate for period $[0, h]$, and W_0 is the initial wealth. After eliminating the quantity invested in the riskfree asset α_0 , the optimization problem becomes:

$$\alpha^* = \arg \max_{\alpha} E(-\exp[-A(\alpha_X X + \alpha_Y Y)]) = -\Psi(A\alpha), \quad (1.28)$$

where $X = p_{X,h} - (1 + r_{0,h})p_{X,0}$, $Y = p_{Y,h} - (1 + r_{0,h})p_{Y,0}$ denote the excess gains, and Ψ is the joint Laplace transform of X and Y ³¹. Therefore for portfolio management with CARA utility function it is more convenient to specify the nonlinear dependence by means of the joint Laplace transform (instead of using the joint density, the copula, or the canonical decomposition). This dependence can be either let unconstrained [see Brandt (1999), and Gouriéroux, Monfort (2002b)], or based on a parametric or semi-nonparametric specification.

Different constrained specifications have been considered in the literature on portfolio management.

Example 7: The standard model assumes gaussian returns which leads to the usual mean-variance framework [Markowitz (1967)]. Let us assume that (X, Y) is jointly normal $N(\mu, \Omega)$, where μ is the vector of expected excess returns and Ω is the variance-covariance matrix. Then the joint Laplace transform of (X, Y) is given by:

$$\Psi(u, v) = \exp\left[-\mu'w + \frac{1}{2}w'\Omega w\right],$$

where $w = (u, v)'$, and the optimization problem becomes:

$$\arg \max_{\alpha} \alpha' \mu - \frac{A}{2} \alpha' \Omega \alpha.$$

Its solution provides the standard mean-variance efficient allocation:

$$\alpha^* = \frac{1}{A} \Omega^{-1} \mu.$$

³¹Note that the joint Laplace transform Ψ of excess gains X, Y is immediately deduced from that of the prices $p_{X,h}, p_{Y,h}$, which are positive variables.

Example 8: A stochastic volatility model.

However it is well-known that the gaussian specification is ill-specified since it neglects fat tails phenomenon. Fat tails can be introduced by including a stochastic volatility. It can be checked that the Laplace transform is still appropriate in this framework. Indeed let us assume that (X, Y) follows a normal distribution $N(\mu, \eta^2 \Omega)$ conditionally to factor η^2 :

$$E \left[\exp(-uX - vY) \mid \eta^2 \right] = \exp \left[-\mu' w + \frac{\eta^2}{2} w' \Omega w \right]. \quad (1.29)$$

When the stochastic volatility is integrated out, the joint Laplace transform is given by:

$$\Psi(u, v) = \exp \left\{ -\mu' w - \phi \left[-\frac{1}{2} w' \Omega w \right] \right\},$$

where $\exp(-\phi)$ is the real Laplace transform of the factor distribution³². This is a semi-nonparametric specification characterized by vector μ , matrix Ω , and functional parameter ϕ ³³.

A natural question in this framework is whether stochastic volatility increases PQD dependence between X and Y . The answer is negative. Indeed, let us for convenience consider the case $\mu = 0$, $\Omega = Id$. Then if η^2 were constant, variables X and Y would be independent. Let us verify whether stochastic volatility induces PQD dependence between X and Y . Let us consider the increasing transformations $g(X) = (X - K)_+$ and $h(Y) = -(L - Y)_+$, which correspond to the payoffs of a call on X with strike K , and of a short position in a put on Y with strike L , respectively. Then:

$$\begin{aligned} Cov[g(X), h(Y)] &= -Cov \left(E[(X - K)_+ \mid \eta^2], E[(L - Y)_+ \mid \eta^2] \right) \\ &< 0, \end{aligned}$$

since call and put prices $E[(X - K)_+ \mid \eta^2]$, $E[(L - Y)_+ \mid \eta^2]$ are increasing functions of the volatility η^2 . Since there exist increasing transformations of X and Y with negative correlation, X and Y are not PQD³⁴.

The first order condition to the optimization of expected utility is:

$$\mu = -A\phi' \left[-\frac{A^2}{2} \alpha^{*'} \Omega \alpha^* \right] \Omega \alpha^*.$$

Thus the ratio of optimal allocations α_X^*/α_Y^* do not depend on functional parameter ϕ . This is due to the fact that the conditional distribution of (X, Y) given η^2 is gaussian, with a single factor η^2 scaling the variance-covariance matrix of the assets.

³²We assume that the Laplace transform of η^2 is defined on $]-\lambda, +\infty[$, where $\lambda > 0$. Then Ψ is defined for $w = (u, v)$ such that $w' \Omega w / 2 < \lambda$.

³³In the literature on portfolio analysis, such joint distributions are also known as elliptical distributions [see e.g. Ingersoll (1987)]. Normal and student distributions are members of the family. In the literature typically ϕ is not treated as a parameter, but it is specified a priori.

³⁴Note the difference with Example 1. In the case of stochastic volatility, the common factor has not a positive effect on X and Y in the sense of first order stochastic dominance, but instead in the sense of second order stochastic dominance.

1.4.5.2 Derivative pricing

Joint distributions of prices are also involved when we consider derivatives written on two underlying assets, with risk neutral distribution replacing the standard historical one. Let us denote by S_1 and S_2 the prices at maturity of two underlying assets, and $\tilde{g}(S_1, S_2; K)$ the payoff at maturity of an European derivative, where K is a parameter characterizing the derivative design. There exist various examples of derivative families written on two underlying assets, for instance:

- i) the quanto derivative with payoff $\tilde{g}(S_1, S_2; K) = (S_1 - K)^+ S_2$. A quanto derivative is a call option, denominated in domestic currency and written on a foreign asset with price S_1 . S_2 corresponds to the exchange rate at maturity.
- ii) The spread derivative, with payoff $\tilde{g}(S_1, S_2; K) = (S_1 - S_2 - K)^+$, is a call option on the difference of prices.
- iii) The basket derivative, with payoff $\tilde{g}(S_1, S_2; K) = (\alpha S_1 + \beta S_2 - K)^+$, is a call option on a portfolio formed by the two assets.
- iv) The exchange (or chooser) derivative has a payoff equal to the maximum (or the minimum) of the two prices at maturity, $\tilde{g}(S_1, S_2) = \max\{S_1, S_2\}$ [$\tilde{g}(S_1, S_2) = \min\{S_1, S_2\}$, respectively].
- v) The Max and Min derivatives (also called under- and over-performance derivatives, respectively, or Rainbow derivatives), with payoffs $\tilde{g}_1(S_1, S_2; K) = (\max\{S_1, S_2\} - K)^+$, $\tilde{g}_2(S_1, S_2; K) = (\min\{S_1, S_2\} - K)^+$, are call derivatives on the maximum and the minimum of the two prices, respectively.

Let us denote by X and Y the returns of the two underlying assets over the holding period: $S_1 = \exp(X)$, $S_2 = \exp(Y)$ ³⁵. The payoff can be written as $g(X, Y; K) = \tilde{g}(\exp(X), \exp(Y); K)$. Due to the no arbitrage conditions, and assuming a zero riskfree interest rate for convenience, the derivative price is equal to:

$$C(g) = \overset{Q}{E}[g(X, Y; K)],$$

where $\overset{Q}{E}$ denotes the expectation with respect to a risk neutral distribution Q . Thus it is natural to select a risk neutral distribution for (X, Y) [or (S_1, S_2)] which allows for tractable computation of such derivative prices.

i) Truncated Laplace transforms

Let us for instance consider the case of a quanto derivative:

$$\begin{aligned} C(K) &= \overset{Q}{E}[(\exp(X) - K)_+ \exp(Y)] \\ &= \overset{Q}{E}[\exp(X + Y) \mathbb{I}(X \geq \log K)] \\ &\quad - K \overset{Q}{E}[\exp(Y) \mathbb{I}(X \geq \log K)]. \end{aligned}$$

³⁵Both current prices are normalized to 1 without loss of generality.

Clearly the derivative price has no simple expression in terms of either the pdf, the copula, or the Laplace transform. However, by introducing returns instead of prices, we point out the importance of exponential payoffs [see e.g. Bakshi, Madan (2000), Duffie, Pan, Singleton (2000), Gouriéroux, Monfort (2001a)]. Moreover it is seen that the derivative prices are easily derived when the dependence is summarized by the risk neutral truncated Laplace transform [Duffie, Pan, Singleton (2000), Gouriéroux, Monfort, Polimenis (2002)]:

$$\Phi(u, v; k) = \overset{Q}{E} [\exp(-uX - vY) \mathbb{I}(X \geq k)].$$

Indeed the price of a quanto option is equal to:

$$C(K) = \Phi(-1, -1, \log K) - K\Phi(0, -1, \log K).$$

The example of quanto options shows that any family of derivatives written on two underlying assets requires an appropriate summary of the joint distribution, and that the summary has no reason to correspond to either the pdf, or the copula ...

ii) Bounds based on copula.

However copula theory can be used to find bounds on derivative prices for some payoff functions [Rapuch (2001), Embrechts, Höing, Juri (2001)]. Let us assume that the derivative payoff \tilde{g} is the cumulative distribution function of a positive measure. Then from Proposition 2 [section 2], the derivative price $C(g) = \overset{Q}{E} [\tilde{g}(S_1, S_2)]$ is monotone increasing with respect to PQD between S_1 and S_2 (or equivalently between X and Y) in the risk neutral distribution, for given risk neutral marginal distributions. Indeed, by partial integration we know that [see (1.14)]:

$$\overset{Q}{E} [\tilde{g}(S_1, S_2)] = \overset{\tilde{g}}{E} [S^Q(S_1, S_2)], \quad (1.30)$$

where S^Q is the joint risk neutral survivor function of (S_1, S_2) , and the result follows by characterization (1.13) of PQD.

In particular, if Q and Q^* are two risk neutral distributions with the same pair of marginal distributions, and copulas C and C^* such that $C \succeq_{PQD} C^*$, we have:

$$\overset{Q}{C}(g) = \overset{Q}{E} [\tilde{g}(S_1, S_2)] \geq \overset{Q^*}{E} [g(S_1, S_2)] = \overset{Q^*}{C}(g). \quad (1.31)$$

Monotonicity of the derivative price in the reversed direction holds when \tilde{g} is a negative measure.

Since $S^Q(s, r)$ is the price of the digital option with payoff $\mathbb{I}(S_1 \leq s, S_2 \leq r)$, equation (1.30) corresponds to a decomposition of the original derivative as a portfolio of digital options with different thresholds (s, r) , distributed according to measure \tilde{g} . Since the price of a bivariate digital option $S^Q(s, r)$ is increasing with respect to risk neutral PQD (with fixed risk neutral marginal distributions), the monotonicity of the derivative price follows when the measure associated to g is positive (or negative).

Let us now consider the examples above of derivative families, and compute the measure associated with their payoffs.

i) For a quanto derivative, it is easily verified that the payoff $\tilde{g}(S_1, S_2) = (S_1 - K)_+ S_2$ is the cumulative function of the Lebesgue measure restricted to the set $\{(S_1, S_2) : S_1 \geq K\}$. Thus the price of a quanto derivative is monotone increasing with respect to risk neutral PQD.

ii) Spread derivatives. The payoff $\tilde{g}(S_1, S_2) = (S_1 - S_2 - K)_+$ of a spread option is neither the cumulative function of a positive, nor a negative measure. However, it is possible to introduce the payoff $g^*(S_1, S_2) = \sqrt{2} [-(S_1 - S_2 - K)_+ + (S_1 - K)_+]$, which is the cumulative function of the Lebesgue measure restricted to the set $\{(S_1, S_2) : S_1 - S_2 = K\}$. Thus the price of the spread derivative is monotone decreasing with respect to risk neutral PQD between S_1 and S_2 , for given risk neutral marginal distributions.

iii) Let us consider a basket derivative with payoff $\tilde{g}(S_1, S_2) = (\alpha S_1 + \beta S_2 - K)_+$. Since either α , or β has to be positive (otherwise the payoff is identical to zero), we may assume that $\alpha > 0$. When $\beta > 0$ (resp. $\beta < 0$) the payoff $(\alpha\beta)^{-1} \sqrt{2} [(\alpha S_1 + \beta S_2 - K)_+ - (\alpha S_1 - K)_+ - (\beta S_2 - K)_+]$ (resp. $(\alpha|\beta|)^{-1} \sqrt{2} [-(\alpha S_1 + \beta S_2 - K)_+ + (\alpha S_1 - K)_+]$) is the cumulative function of the Lebesgue measure restricted to the set $\{(S_1, S_2) : \alpha S_1 + \beta S_2 = K\}$. Thus the price of a basket derivative is monotone increasing (resp. decreasing) with respect to risk neutral PQD between S_1 and S_2 when $\alpha\beta > 0$ ($\alpha\beta < 0$, respectively).

iv) Exchange derivative. The payoff $\tilde{g}(S_1, S_2) = \min(S_1, S_2)$ is the cumulative function of the Lebesgue measure restricted to the 45 degree line $S_1 = S_2$. The price of the corresponding exchange derivative is monotone increasing with respect to risk neutral PQD. The payoff of the chooser derivative $\tilde{g}(S_1, S_2) = \max(S_1, S_2)$ is not the cumulative function of a positive measure. However, since $\max(S_1, S_2) = S_1 + S_2 - \min(S_1, S_2)$, we deduce that the price of the chooser derivative is monotone decreasing with respect to risk neutral PQD.

v) The payoff of a Min derivative $\tilde{g}(S_1, S_2) = (\min(S_1, S_2) - K)_+$ is the cumulative function of the Lebesgue measure restricted to the set $\{(S_1, S_2) : S_1 = S_2 \geq K\}$. The price of the corresponding derivative is monotone increasing with respect to risk neutral PQD.

In the examples above the measure \tilde{g} is a Lebesgue measure restricted to a subset of \mathbb{R}_+^2 , and the designs differ by the selected subset. This subset can admit a two- or one-dimensional support.

The monotonicity of the derivative price with respect to PQD between S_1 and S_2 in the risk neutral distribution can be used to derive bounds on the derivative price. Indeed, if the copula C of the risk neutral distribution Q is such that $C' \succeq_{PQD} C \succeq_{PQD} C^*$, for

two copulas C' and C^* , then: $\overset{Q'}{E} [\tilde{g}(S_1, S_2)] \geq \overset{Q}{E} [\tilde{g}(S_1, S_2)] \geq \overset{Q^*}{E} [\tilde{g}(S_1, S_2)]$, where Q' and Q^* are risk neutral distributions with the same pair of marginal distributions as Q , and copula C' and C^* , respectively. In particular, for given risk neutral marginal distributions, the derivative price is contained between the bounds corresponding to the upper and lower Frechet risk neutral copulas. We illustrate these lower and upper bounds in the example of a spread option. Let us assume that the gross returns S_1/S_1^0 and S_2/S_2^0 of the two assets have marginal lognormal distributions with parameters $(\mu_1 T, \sigma_1^2 T)$ and $(\mu_2 T, \sigma_2^2 T)$, respectively, corresponding to marginal Black Scholes models, where S_1^0 and S_2^0 are the prices at time 0,

and T is the maturity of the spread option. In Figure 5 we plot the upper and lower bounds for the spread option price $C(g)$ as functions of the strike price K , for different values of the initial asset prices S_1^0, S_2^0 : in Panel A we have $S_1^0 = 90, S_2^0 = 100$, in Panel B $S_2^0 = 100 = S_1^0$, in Panel C $S_1^0 = 110, S_2^0 = 100$, in Panel D $S_1^0 = 120, S_2^0 = 100$.

[insert Figure 5: upper and lower bounds for the spread option price]

The parameter values are $T = 1/12$ (one month), and $\mu_1 = 0.06, \sigma_1 = 0.25, \mu_2 = 0.05, \sigma_2 = 0.2$. The lower and the upper bounds are close to each other, identifying a narrow interval for admissible option prices, when S_1^0 is larger than S_2^0 and the strike price K is small.

1.4.6 Control of extreme risk

In order to control the extreme risks included in financial investments the regulators are defining new rules for computing the reserves. For a payoff $g(X, Y)$, say, written on two risk variables, the required capital is defined from a quantile of the payoff distribution, called Value at Risk (VaR). More precisely the Value at Risk $VaR(g, \alpha)$ at level α of the payoff g is defined as the α -quantile:

$$P[g(X, Y) \leq VaR(g, \alpha)] = \alpha.$$

The payoff can admit different patterns in finance or insurance applications. The VaR of a portfolio including the quantities a, b of two assets corresponds to the quantile of a linear payoff $g(X, Y) = aX + bY$, where X and Y denote the asset returns. Nonlinear payoffs are involved in insurance problems. For instance let X denote a loss, and Y a stochastic reimbursement that a reinsurance company has to pay when the loss exceeds a threshold K . The reinsurance company is interested in the VaR of the payoff $g(X, Y) = Y \cdot \mathbb{I}(X \geq K)$.

The VaR of the payoff $g(X, Y)$ is influenced by the dependence structure of (X, Y) . The dependence measures introduced in sections 2 and 3 are not appropriate for an analytical computation of the VaR. However they can be used to derive exact bounds for the VaR, or approximations of the VaR for small α .

i) Bounds for the VaR

As for derivative prices it is possible to derive bounds for the admissible Value at Risk $VaR(g, \alpha)$ of the payoff $g(X, Y)$, for given marginal distributions F_X, F_Y of X, Y . Let us denote by G the c.d.f. of the payoff $g(X, Y)$, and assume that the copula C of (X, Y) is known to be larger than a lower bound C_0 : $C \succeq_{PQD} C_0$ [in absence of additional information this bound is the lower Frechet bound, see section 2.2.ii)]. Durrleman, Nikeghbali, Roncalli (2000) and Embrechts, Höing, Juri (2001), following the original works of Makarov (1981) and Williams (1987), derive functions H and K which provide lower and upper bounds for the cdf of the payoff: $H \leq G \leq K$. The lower and upper bounds for the VaR of the payoff $g(X, Y)$, for given marginal distributions of X, Y , are deduced immediately:

$$K^{-1}(\alpha) \leq VaR(g, \alpha) \leq H^{-1}(\alpha).$$

Let us briefly sketch the derivation of bounds H and K . For expositional purpose, let us assume that the payoff $g(X, Y)$ is strictly increasing in variable Y . Then, for any x , function $y \mapsto g(x, y)$ is invertible, and let us denote by g_x^{-1} its inverse.

To derive the upper bound, let us remark that, for any x, s , the condition $X > x$ and $Y > g_x^{-1}(s)$ implies $g(X, Y) > s$. Thus we get:

$$\begin{aligned} G(s) &= P[g(X, Y) \leq s] \\ &\leq P[\{X \leq x\} \cup \{Y \leq g_x^{-1}(s)\}] \\ &= P[X \leq x] + P[Y \leq g_x^{-1}(s)] - P[X \leq x, Y \leq g_x^{-1}(s)] \\ &= F_X(x) + F_Y[g_x^{-1}(s)] - C(F_X(x), F_Y[g_x^{-1}(s)]). \end{aligned}$$

Since this inequality holds for any x , we get:

$$G(s) \leq \inf_x \{F_X(x) + F_Y[g_x^{-1}(s)] - C_0(F_X(x), F_Y[g_x^{-1}(s)])\} \equiv K(s).$$

Similarly, to derive the lower bound, we use the property that for any x, s , the condition $X \leq x$ and $Y \leq g_x^{-1}(s)$ implies $g(X, Y) \leq s$. We get:

$$\begin{aligned} G(s) &= 1 - P[g(X, Y) > s] \\ &\geq 1 - P[\{X > x\} \cup \{Y > g_x^{-1}(s)\}] \\ &= P[X \leq x, Y \leq g_x^{-1}(s)] \\ &= C(F_X(x), F_Y[g_x^{-1}(s)]), \end{aligned}$$

and thus:

$$G(s) \geq \sup_x C_0(F_X(x), F_Y[g_x^{-1}(s)]) \equiv H(s).$$

For instance, if (X, Y) are PQD, the lower and upper bounds for the cdf of a portfolio value $g(X, Y) = aX + bY$ reduce to: $H(s) = \sup_x F_X(x)F_Y[s/b - ax/b]$, $K(s) = \inf_x \{F_X(x) + F_Y[s/b - ax/b] - F_X(x)F_Y[s/b - ax/b]\}$.

Functions H and G do not correspond to the cdf of the payoff $g(X, Y)$ under the upper or lower Frechet copulas, nor to the cdf of the payoff under any particular copula. For instance, in the case of a portfolio $g(X, Y) = aX + bY$, the worst scenario for the VaR does not correspond to positive deterministic dependence between X and Y , that is the upper Frechet bound. This point distinguishes the present result from that in Proposition 2.

ii) Tail approximation of the VaR

Whereas the PQD ordering associated with copulas seems appropriate to derive bounds for the VaR, closer approximations can be derived from the Laplace transform when the critical level α is small [Darolles, Gouriou, Jasiak (2002)]. This possibility follows from Karamata's Tauberian theorem, which explains how the behaviour of a cdf at infinity is related to the behaviour of the real Laplace transform at the origin. Typically, if W is a random variable such that:

$$E[\exp(-uW)] \sim 1 - u^\delta l(1/u), \quad \text{for } u \approx 0,$$

where $0 \leq \delta \leq 1$, and l is a slowly varying function, then:

$$P[W \leq w] \sim 1 - \frac{l(w)}{w^\delta \Gamma(1-\delta)}, \quad \text{for } w \rightarrow +\infty,$$

where Γ is the gamma function [see e.g. Feller (1971) chapter 13, Bingham, Goldie, Teugels (1987) Corollary 8.17].

This property can be applied to a portfolio value $W = aX + bY$, say. Let us assume that the joint Laplace transform of (X, Y) satisfies:

$$\Psi(u, v) \sim 1 - c_1 u^{\delta_1} - c_2 v^{\delta_2}, \quad u, v \approx 0,$$

where $0 \leq \delta_1, \delta_2 \leq 1$. Then the real Laplace transform of the portfolio value W is such that:

$$\begin{aligned} E[\exp(-uW)] &= E[\exp(-auX - buY)] \\ &= \Psi(au, bu) \sim 1 - (c_1 a^{\delta_1}) u^{\delta_1} - (c_2 b^{\delta_2}) u^{\delta_2}, \quad \text{for } u \approx 0. \end{aligned}$$

The behaviour at the origin of the real Laplace transform of the portfolio value W depends on the relative magnitude of the exponents δ_1, δ_2 , which are related to the tails of the joint distribution of X, Y . Let us for instance consider the case where $\delta_1 = \delta_2 = \delta$. Then:

$$E[\exp(-uW)] \sim 1 - (c_1 a^\delta + c_2 b^\delta) u^\delta, \quad \text{for } u \approx 0,$$

and the Tauberian theorem implies that the cdf at infinity of the portfolio value W is such that:

$$P[W \leq w] \sim 1 - \frac{c_1 a^\delta + c_2 b^\delta}{w^\delta \Gamma(1-\delta)}, \quad \text{for } w \rightarrow +\infty.$$

Thus the VaR for the portfolio value W at a small confidence level α is approximated by:

$$VaR(\alpha) \sim \left[\frac{c_1 a^\delta + c_2 b^\delta}{(1-\alpha) \Gamma(1-\delta)} \right]^{1/\delta}.$$

Example 9: Let us assume that the returns X, Y follow compounds models in a latent factor Z such that: $E[\exp(-uX) | Z] = \exp[-a_1(u)Z]$, $E[\exp(-vY) | Z] = \exp[-a_2(v)Z]$, where a_1 and a_2 are positive functions, and moreover that X and Y are independent conditionally to factor Z :

$$E[\exp(-uX - vY) | Z] = \exp\{-[a_1(u) + a_2(v)]Z\}.$$

By integrating out factor Z , the joint Laplace transform of X and Y is given by:

$$\begin{aligned} \Psi(u, v) &= E[\exp\{-[a_1(u) + a_2(v)]Z\}] \\ &= \psi[a_1(u) + a_2(v)], \end{aligned}$$

where ψ is the Laplace transform of factor Z . Let us further assume that functions a_1, a_2 are such that: $a_1(u) \sim c_1 u^{\gamma_1}$, $a_2(v) \sim c_2 v^{\gamma_2}$, for $u, v \approx 0$.

The behaviour of the joint Laplace transform Ψ in a neighbourhood of the origin depends on the tails of factor Z .

i) If factor Z has gamma tails:

$$\psi(s) \sim \left(\frac{\lambda}{\lambda + s} \right)^\delta, \quad s \approx 0,$$

for some $\lambda, \delta > 0$, we get:

$$\Psi(u, v) \sim 1 - \frac{\delta}{\lambda} (c_1 u^{\gamma_1} + c_2 v^{\gamma_2}), \quad u, v \approx 0.$$

If $\gamma_1 = \gamma_2 = \gamma$, the VaR at a small critical level α is approximated by:

$$VaR(\alpha) \sim \left[\frac{\delta}{\lambda} \frac{c_1 a^\gamma + c_2 b^\gamma}{(1 - \alpha) \Gamma(1 - \gamma)} \right]^{1/\gamma}.$$

In particular, the exponent of $(1 - \alpha)$ is affected only by the parameter γ characterising dependence functions a_1, a_2 , and the distribution of the factor Z only affects the scale of $VaR(\alpha)$.

ii) Let us now assume that factor Z has Pareto tails:

$$\psi(s) \approx 1 - s^{1/\delta}, \quad s \approx 0.$$

Then if $\gamma_1 = \gamma_2 = \gamma$ we get:

$$\Psi(au, bu) \approx 1 - (c_1 a^\gamma + c_2 b^\gamma)^{1/\delta} u^{\delta/\gamma}, \quad u \approx 0,$$

and the VaR is approximated by:

$$VaR(\alpha) \sim \frac{(c_1 a^\gamma + c_2 b^\gamma)^{1/\gamma}}{[(1 - \alpha) \Gamma(1 - \gamma/\delta)]^{\delta/\gamma}}.$$

In this case both parameter γ and parameter δ characterising the tails of Z affect the exponent of $(1 - \alpha)$.

1.5 Nonlinear time series models

1.5.1 Characterizations of serial dependence

In this section we discuss the serial dependence in nonlinear time series X_t , $t \in \mathbb{N}$. For expository purpose, we consider a one-dimensional stationary Markov process. Thus the distribution of the process is fully characterized by the joint distribution of X_t and X_{t-1} . Moreover due to the stationarity assumption, this bivariate distribution admits identical marginal distributions, which coincide with the stationary one. The serial dependence can

be characterized in different ways.

i) Copulas

First the joint distribution of (X_t, X_{t-1}) can be defined by its copula C and a stationary distribution F . The Markov process X_t , $t \in \mathbb{N}$, can be represented as an increasing (nonlinear) transformation of a standardized Markov process U_t , $t \in \mathbb{N}$:

$$X_t = F^{-1}(U_t),$$

with uniform marginal distribution, and a transition density given by:

$$f_{U_t|U_{t-1}}(u | v) = c(u, v), \quad u, v \in [0, 1],$$

where c is the copula density. Serial dependence properties of process X_t , $t \in \mathbb{N}$, which are invariant by stationary increasing transformations $X_t \longrightarrow g(X_t)$, $\forall t$, of all components of the process, are fully characterized by the copula C . Indeed, such serial dependence properties are valid for the standardized process $U_t = F(X_t)$, $t \in \mathbb{N}$, whose distribution is characterized by the copula C .

ii) Nonlinear canonical analysis

This approach is based on a decomposition of the joint density f_1 of X_t and X_{t-1} as:

$$f_1(x_t, x_{t-1}) = f(x_t)f(x_{t-1}) \left[1 + \sum_{j=1}^{\infty} \lambda_j \varphi_j(x_t) \psi_j(x_{t-1}) \right], \quad (1.32)$$

where φ_j , ψ_j , j varying, are the current and lagged canonical directions, λ_j , j varying, are the canonical correlations, and f is the marginal density [see section 3.1].

iii) Laplace transforms

Finally, the distribution of (X_t, X_{t-1}) may be characterized by the joint Laplace transform:

$$\Psi(u, v) = E[\exp(-uX_t - vX_{t-1})].$$

When the Markov process admits a unique invariant distribution, the distribution of (X_t, X_{t-1}) is also characterized by the Laplace transform of the conditional distribution of X_t given X_{t-1} :

$$\Psi(u, X_{t-1}) = E[\exp(-uX_t) | X_{t-1}].$$

1.5.2 Transitions at any horizon

The investigation of persistency properties, forecasting, and the computation of conditional expectations at any horizon of the process X_t , $t \in \mathbb{N}$, require the analysis of the dependence between X_t and X_{t-h} , at any lag $h \in \mathbb{N}$. We discuss below the convenience of the different representations of serial dependence for the analysis at larger horizon.

i) Copulas

The copula density c_h of (X_t, X_{t-h}) may be deduced from the copula at horizon 1 by Chapman-Kolmogorov equation. The copula c_h is given by:

$$c_h(u_t, u_{t-h}) = \int \dots \int c(u_t, u_{t-1}) \dots c(u_{t-h+1}, u_{t-h}) du_{t-1} \dots du_{t-h+1}. \quad (1.33)$$

Example 10: Markov process with finite dimensional dependence.

Markov processes with finite dimensional dependence have been introduced by Gouriéroux, Jasiak (2001), and are characterized by a nonlinear canonical decomposition of the density of (X_t, X_{t-1}) with only a finite number of non-zero canonical correlations. For a Markov process with one-dimensional dependence, the copula of (X_t, X_{t-1}) is given by:

$$c(u, v) = 1 + \lambda a(u) b(v),$$

where a and b are the first canonical directions, and λ the first canonical correlation (see section 3.1.4). By Chapman-Kolmogorov equation (1.33), and the normalizations of canonical directions, the copula p.d.f. c_h of (X_t, X_{t-h}) is given by:

$$c_h(u, v) = 1 + \lambda_h a(u) b(v), \quad (1.34)$$

where the canonical correlation λ_h at horizon h is given by³⁶:

$$\begin{aligned} \lambda_h &= \lambda \left(\lambda \int_0^1 a(u) b(u) du \right)^{h-1} \\ &= \lambda \rho^{h-1}, \text{ say.} \end{aligned}$$

In the general framework c_h does not admit a tractable form. However interesting properties of the sequence c_h , $h \in \mathbb{N}$, that is of the age structure of copula, can be derived in special cases. For instance, some properties can be deduced from a theorem by Fang, Hu and Joe (1994), which involves the concept of stochastic increasing dependence. Let us recall that X_t is said to be stochastic increasing (SI) in X_{t-1} , if X_t is increasing with respect to X_{t-1} for first order stochastic dominance³⁷. The SI dependence only involves the copula of (X_t, X_{t-1}) , and is stronger than PQD [see e.g. Joe (1997)].

Theorem 1.2 (Fang, Hu, Joe) *Let X_t , $t \in N$, be a stationary Markov process, such that X_t is SI in X_{t-1} . Then the sequence C_h , $h \in N$, is decreasing, such that:*

$$uv \leq C_{h+1}(u, v) \leq C_h(u, v), \quad \forall u, v, \quad \forall h \in \mathbb{N}.$$

Thus, under conditions of Theorem 2, the process features PQD at any lag, and this dependence decreases with the horizon³⁸. In particular, the sequence C_h admits a limit:

³⁶We assume that $\int_0^1 a(u) b(u) du \geq 0$. Otherwise, the minus sign of λ_h has to be assigned to one of the canonical directions.

³⁷That is the conditional survivor function of X_t given $X_{t-1} = y$, $S(x|y)$ say, is increasing in y , for any x .

³⁸More precisely, Fang, Hu, Joe (1994) show that X_t is SI in X_{t-h} for any h .

$C_\infty = \lim_{h \rightarrow \infty} C_h$ (for pointwise convergence). It is easily checked that function C_∞ is a copula such that $C_\infty(u, v) \geq uv$, which can be called the long-term copula.

Example 10 (cont.) For the stationary Markov process with one-dimensional dependence we have: $P(U_t \leq u \mid U_{t-1} = v) = u + \lambda A(u)b(v)$, where $A(u) = \int_0^u a(w)dw$. Thus X_t is SI in X_{t-1} if and only if functions a and b are both increasing, or both decreasing. The copula cdf at horizon h is given by $C_h(u, v) = uv + \lambda \rho^{h-1} A(u)B(v)$, where $B(v) = \int_0^v b(w)dw$. When $\rho < 1$, the sequence C_h decreases geometrically to the long-term copula $C_\infty(u, v) = uv$, which is the independent copula.

ii) Canonical decomposition

The density f_h of (X_t, X_{t-h}) admits a canonical decomposition:

$$f_h(x_t, x_{t-h}) = f(x_t)f(x_{t-h}) \left[1 + \sum_{j=1}^{\infty} \lambda_{j,h} \varphi_{j,h}(x_t) \psi_{j,h}(x_{t-h}) \right], \quad (1.35)$$

where $\varphi_{j,h}, \psi_{j,h}, j$ varying, are the current and lagged canonical directions, and $\lambda_{j,h}, j$ varying, are the canonical correlations at horizon $h \in \mathbb{N}$. For instance, for the Markov process with one-dimensional dependence, the canonical decomposition at horizon h is given in (1.34). Generally the canonical directions and correlations at horizon h have no tractable expression in terms of the canonical directions and correlations at horizon 1. However a simple relation can be derived for reversible Markov processes. A Markov process is said to be reversible if its distribution is the same in direct and reversed time, that is if (X_t, X_{t-1}) and (X_{t-1}, X_t) have the same distribution. This is the case if the current and lagged canonical directions at horizon one are equal up to sign: $\varphi_j = \pm \psi_j$. By applying Chapman-Kolmogorov and the orthogonality properties of the canonical directions φ_j , it is easily checked that the canonical decomposition of f_h for a reversible process is given by:

$$f_h(x_t, x_{t-h}) = f(x_t)f(x_{t-h}) \left[1 + \sum_{j=1}^{\infty} \lambda_j^h \varphi_j(x_t) \varphi_j(x_{t-h}) \right].$$

Since $\lambda_j \leq \lambda_1 < 1$, we deduce that a reversible stationary Markov process admits an asymptotic copula corresponding to independence, and that the convergence of c_h to the long term copula is at a geometric rate.

iii) Laplace transforms

Finally, the distribution of (X_t, X_{t-h}) may be characterized by the joint Laplace transform:

$$\Psi_h(u, v) = E[\exp(-uX_t - vX_{t-h})],$$

or equivalently by the conditional Laplace transform at horizon h :

$$\Psi_h(u, X_{t-h}) = E[\exp(-uX_t) \mid X_{t-h}].$$

In general, no explicit expressions for $\Psi_h(u, v)$ or $\Psi_h(u, X_{t-h})$ can be given. However, an important special case where this is possible is the Compound Autoregressive (CAR) process introduced by Darolles, Gourioux, Jasiak (2002). The CAR process is a Markov process with an exponentially affine Laplace transform [see section 3.2.2]:

$$E \left[\exp(-uX_t) \mid \underline{X_{t-1}} \right] = \exp[-a(u)X_{t-1} - b(u)],$$

where a et b are one-dimensional functions. Let us now compute the conditional Laplace transform at horizon h . We have:

$$\begin{aligned} E \left[\exp(-uX_t) \mid X_{t-h} \right] &= E \left\{ E \left[\exp(-uX_t) \mid \underline{X_{t-1}} \right] \mid X_{t-h} \right\} \\ &= \exp[-b(u)] E \left\{ \exp[-a(u)X_{t-1}] \mid X_{t-h} \right\}. \end{aligned}$$

By iteration we get:

$$E \left[\exp(-uX_t) \mid X_{t-h} \right] = \exp \left\{ -a^{oh}(u)X_{t-h} - b(u) - b[a(u)] - \dots - b[a^{o(h-1)}(u)] \right\},$$

where a^{oh} denotes function a compounded h times with itself. From this formula we deduce that the Markov process X_t , $t \in \mathbb{N}$, is stationary when $\lim_{h \rightarrow \infty} a^{oh}(u) = 0$, $\forall u$, since the transition at large horizon no longer depends on the initial value of the process. The Laplace transform $\exp(-c)$ of the marginal distribution is related to functions a and b by:

$$c(u) = c[a(u)] + b(u).$$

Thus two different parameterizations are possible for stationary CAR processes: either in terms of functions a and b defining the conditional Laplace transform, or in terms of function a and function c , which is related to the marginal distribution. In particular the conditional Laplace transform at horizon h can be written in terms of a and c as:

$$E \left[\exp(-uX_t) \mid X_{t-h} \right] = \exp \left\{ -a^{oh}(u)X_{t-h} - c(u) + c[a^{oh}(u)] \right\}.$$

By using similar arguments we can also derive the joint conditional Laplace transform at horizon h :

$$E \left[\exp \left(- \sum_{s=1}^h u_s X_{t+s} \right) \mid X_t \right] = \exp[-A(u, h)X_t - B(u, h)], \quad (1.36)$$

where $u = (u_1, \dots, u_h)$ and:

$$\begin{aligned} A(u, h) &= a \{ u_1 + a[u_2 + \dots + a(u_{h-1} + a(u_h))] \}, \\ B(u, h) &= b(u_h) + b[u_{h-1} + a(u_h)] + b \{ u_1 + a[u_2 + \dots + a(u_{h-1} + a(u_h))] \}. \end{aligned}$$

In many applications (see section 6), we need to evaluate conditional expectations such as: $E[\exp(-\gamma X_{t+h} - \delta X_{t+h-1} - \dots - \delta X_{t+1}) \mid X_t]$. The vector u is given by: $u = (\delta, \dots, \delta, \gamma)$, and the coefficients $A_h = A(u, h)$ and $B_h = B(u, h)$, $h \in \mathbb{N}$, satisfy the recursion formulae:

$$\begin{aligned} A_h &= a(\delta + A_{h-1}), \quad A_1 = a(\gamma), \\ B_h &= B_{h-1} + b(\delta + A_{h-1}), \quad B_1 = b(\gamma). \end{aligned} \quad (1.37)$$

Finally, unconditional joint Laplace transforms are easily obtained from (1.36):

$$E \left[\exp \left(- \sum_{s=1}^h u_s X_s \right) \right] = \exp \{ -c [A(u, h)] - B(u, h) \}. \quad (1.38)$$

1.5.3 Examples

Different examples are provided in this section to see how the nonlinear dependence pattern affects the dynamics of the process.

1.5.3.1 Markov processes with gaussian copula

Let X_t , $t \in \mathbb{N}$, be a stationary Markov process, whose joint density is characterized by a gaussian copula C_ρ , $|\rho| < 1$, [see section 2.3.i)] and a marginal distribution F . Then process X_t , $t \in \mathbb{N}$, is a nonlinear transformation of a gaussian autoregressive process X_t^* , $t \in \mathbb{N}$:

$$\begin{aligned} X_t &= F^{-1} [\Phi(X_t^*)], \\ X_t^* &= \rho X_{t-1}^* + \varepsilon_t, \quad \varepsilon_t \sim IIN(0, 1 - \rho^2). \end{aligned}$$

i) Simulated trajectories

We report in Figure 6, Figure 7, and Figure 8 simulated trajectories of length $T = 200$ of Markov processes with Gaussian copula, correlation parameter $\rho = 0$, $\rho = 0.5$, and $\rho = 0.95$, respectively, and different marginal distributions.

[insert Figure 6: simulated trajectory, Gaussian copula, $\rho = 0$]

[insert Figure 7: simulated trajectory, Gaussian copula, $\rho = 0.5$]

[insert Figure 8: simulated trajectory, Gaussian copula, $\rho = 0.95$]

For each figure, Panel A reports the standardized process U_t with uniform marginal distribution, Panel B the gaussian process X_t^* , whereas Panel C and D report Markov processes with Gaussian copula and Pareto³⁹ or Cauchy⁴⁰ marginal distribution, respectively. The Pareto and Cauchy distributions have fat tails, and Markov processes with these marginal distributions may represent for instance the durations between consecutive trades, and the returns, respectively, of a financial asset.

When $\rho = 0$ the process is iid, and positive dependence increases with ρ . Moreover different behaviour of extremes can be seen according to the tail and skewness of the marginal distribution.

³⁹The marginal cdf is given by: $F(x) = 1 - 1/(1 + ax)^\tau$, with $a = 4$, $\tau = 2$.

⁴⁰The marginal cdf is given by: $F(x) = (1/\pi) \arctan(ax) + 1/2$, with $a = 10$.

ii) Autocorrelograms

Figure 9, Figure 10 and Figure 11 provide the autocorrelograms for Markov processes with Gaussian copula considered above⁴¹.

[insert Figure 9: ACF, Gaussian copula, $\rho = 0$]

[insert Figure 10: ACF, Gaussian copula, $\rho = 0.5$]

[insert Figure 11: ACF, Gaussian copula, $\rho = 0.95$]

Since the canonical correlations at horizon h are $\lambda_{j,h} = \rho^{hj}$, j varying, the autocorrelograms decrease geometrically with lag h , for any marginal distribution. However, although all Markov processes with Gaussian copula and same ρ parameter have the same copula, their autocorrelograms strongly differ. Indeed the correlation is not invariant by nonlinear transformation and the magnitude of the autocorrelation depends on the pattern of the marginal cdf. Since the first canonical directions of the gaussian distribution are affine, the transformed process with gaussian marginal distribution has the larger autocorrelogram [see Example 2].

iii) Isodensity curves

Since (X_t^*, X_{t-h}^*) has a gaussian distribution with linear correlation ρ^h , the copula of (X_t, X_{t-h}) at horizon h is gaussian, with correlation parameter ρ^h . When $\rho = 0$, C_h is the independent copula at any horizon. In Figure 12 and Figure 13 we plot isodensity curves for the distribution of (X_t^*, X_{t-h}^*) for correlation parameters $\rho = 0.5$ and $\rho = 0.95$, respectively, at different horizons.

[insert Figure 12: isodensity curves, Gaussian copula, $\rho = 0.5$]

[insert Figure 13: isodensity curves, Gaussian copula, $\rho = 0.95$]

The sequence C_h is monotone decreasing, and converges geometrically to the long-term independent copula.

iv) Clustering of extremes

In Figure 14 we plot a time series of indicator variable $I_t = \mathbb{I}(U_t \geq \lambda)$ and of the counting process $N_t = \sum_{s=1}^t \mathbb{I}(U_s \geq \lambda)$ for observations above the λ -quantile in a simulated trajectory of length $T = 2000$ of a Markov process with Gaussian copula and $\rho = 0.5$. Panels A and B correspond to 1% upper quantile $\lambda = 0.99$, and Panels C and D to 0.5% upper quantile $\lambda = 0.995$.

[insert Figure 14: Extremes, Gaussian copula, $\rho = 0.5$]

Most of the large observations do not cluster, but instead are isolated in time. This is consistent with the tail independence featured by Gaussian copula.

⁴¹ Autocovariances are computed by Monte Carlo simulation whenever necessary.

1.5.3.2 Markov process with Archimedean copula

Let us now consider a stationary Markov process (X_t) with a copula in the Archimedean family [see Bouyé, Gaussel, Salmon (2000)]. For instance, let us assume a Gumbel copula C_α , $\alpha \geq 1$ [see section 2.3.2 v)].

i) Simulated trajectories

Simulated trajectories of length $T = 200$ of a Markov process with Gumbel Copula C_α are reported in Figure 15 and Figure 16 for $\alpha = 1.5$, and $\alpha = 4.946$, respectively. The choices of α parameter ensure that the Kendall's tau of (X_t, X_{t-1}) is equal to that of Gaussian copulas with $\rho = 0.5$ and $\rho = 0.95$, respectively⁴². Moreover the marginal distributions are the same as in the previous section.

[insert Figure 15: simulated trajectory, Gumbel copula, $\alpha = 1.5$]

[insert Figure 16: simulated trajectory, Gumbel copula, $\rho = 0.95$]

Serial dependence is stronger for the process with the larger α parameter [$\alpha = 4.946$]; this is consistent with the fact that the Gumbel family is positively PQD ordered with respect to α . Moreover, compared to Markov processes with Gaussian copula, trajectories of Markov processes with Gumbel copula feature clusters (that are patches) of large observations.

ii) Autocorrelogram

The autocorrelograms of Markov processes with Gumbel copula are reported in Figure 17 and Figure 18.

[insert Figure 17: ACF, Gumbel copula, $\alpha = 1.5$]

[insert Figure 18: ACF, Gumbel copula, $\alpha = 4.946$]

These autocorrelograms differ in several aspects compared with the ones derived for Markov process with Gaussian copula. Firstly, since the first canonical directions of a Markov process with Archimedean copula and standard gaussian marginal distribution are not affine, the transformed process with standard gaussian marginal distribution has no longer the largest autocorrelation. Among the transformed processes considered here, the Markov process with Pareto marginal distribution has the largest autocorrelation function.

Secondly, the Markov process with Gumbel copula has more persistent autocorrelograms. For instance, let us compare the autocorrelation function of a Markov process with Gaussian copula with $\rho = 0.95$ and Pareto marginal distribution (Figure 11, Panel C) with that of a Markov process with Gumbel copula with $\alpha = 4.946$ and Pareto marginal distribution (Figure 18, Panel C). Both processes, X_t and X_{t-1} have the same Kendall's tau, but different copula patterns. The autocorrelation function of the Markov process with Gumbel copula is larger at any lag, and decays more slowly with the horizon. However, since Archimedean copulas are symmetric, Markov process with Gumbel copula is time reversible, and thus

⁴²The choice $\alpha = 1$ would correspond to an iid process.

canonical correlations and autocorrelograms decrease asymptotically with the lag at a geometric rate, as in the Gaussian case.

iii) Clustering of extremes

Figure 19 provides the time series of indicator variable $I_t = \mathbb{I}(U_t \geq \lambda)$ and of the counting process $N_t = \sum_{s=1}^t \mathbb{I}(U_s \geq \lambda)$ for observations above the λ -quantile in a simulated trajectory of length $T = 2000$ of a Markov process with Gumbel copula and $\alpha = 1.5$. Panels A and B correspond to $\lambda = 0.99$, and Panels C and D to $\lambda = 0.995$.

[insert Figure 19: Extremes, Gumbel copula, $\alpha = 1.5$]

Comparing with the Markov process with Gaussian copula (Figure 14), we see that large observations for the Markov process with Gumbel copula have a tendency to cluster, that is to come in patches. This is consistent with the evidence from the simulated trajectories discussed in i), and with the upper tail dependence featured by Gumbel copula.

1.5.3.3 Markov processes with finite dimensional dependence

Let us consider a stationary Markov process with one-dimensional dependence and canonical directions $a(u) = (2\alpha + 1)^{1/2} |2u - 1|^\alpha \text{sign}(u - 1/2)$, $\alpha = 2$, $b(v) = (2\beta + 1)^{1/2} |2v - 1|^\beta \text{sign}(v - 1/2)$, $\beta = 0.5$, and canonical correlation $\lambda = 0.3$.

[insert Figure 20: canonical directions]

Current and lagged canonical directions are plotted in Figure 20 for different marginal distributions, that are the uniform, gaussian, Pareto and Cauchy distribution. Since current and lagged canonical directions a and b are both monotonically increasing, this Markov process with one-dimensional dependence features PQD (see section 3.1.4 ⁴³).

i) Simulated trajectories

In Figure 21 we plot simulated time series of length $T = 200$ of Markov process with one-dimensional dependence. As usual, the standardized Markov process U_t , the Markov process X_t^* with standard gaussian marginal distribution, and Markov process with Pareto and Cauchy marginal distributions are reported.

[insert Figure 21: simulated trajectories]

The processes feature weak positive serial dependence and absence of clustering of extremes.

ii) Autocorrelograms

The autocorrelograms of the Markov processes with one-dimensional dependence are reported in Figure 22.

[insert Figure 22: autocorrelograms]

⁴³This Markov process features even SI dependence, see Example 9 in section 5.2.

Since the canonical correlation at horizon h is given by $\lambda_h = \lambda \rho^{h-1}$, where $\rho = \lambda \int_0^1 a(w)b(w)dw \simeq 0.27$, the autocorrelation function converges quickly to zero with the lag. The standardized Markov process U_t and the Markov process with standard gaussian marginal distribution have the largest and most persistent autocorrelograms. Indeed, for these processes the canonical directions (especially the current canonical direction a and φ) are approximately linear on a relevant part of the sample space. For instance, let us compare the present canonical direction of the Markov process with standard gaussian marginal distribution (Figure 20, Panel B) with that of a Markov process with Pareto marginal distribution (Figure 20, Panel C). For the latter one, the canonical directions involve a stronger downweighting of the upper half of observations⁴⁴. Consequently, the autocorrelation function of the process with Pareto marginal distribution is smaller and converges more quickly to zero.

iii) Reversed time autocorrelograms

Since the current and lagged canonical directions are different, $a \neq b$, the Markov process with one-dimensional dependence is not time reversible. Therefore the correlations are different in calendar and reversed time. For instance, let us consider the correlation between X_t^* with standard gaussian marginal distribution and the lagged transformation $g(X_{t-h}^*) = \text{sign}(X_{t-h}^*) |X_{t-h}^*|^{1/10}$, for $h = 0, 1, 2, \dots$. When process X_t^* is observed in reversed time, the corresponding correlations are: $\text{corr}(X_t^*, g(X_{t+h}^*)) = \text{corr}(g(X_t^*), X_{t-h}^*)$, $h = 0, 1, 2, \dots$. Calendar and reversed time autocorrelations are reported in Figure 23, solid and dashed line, respectively.

[insert Figure 23: reversed time autocorrelograms]

The autocorrelations in reversed time are smaller. This is easy to understand from the patterns of the canonical directions, since transformation g is more correlated with the lagged canonical direction than with the current canonical direction.

iv) Isodensity curves

Isodensity curves of (X_t^*, X_{t-h}^*) for Markov process with one-dimensional dependence and standard Gaussian marginal distribution are reported in Figure 24, for different horizons.

[insert Figure 24: isodensity curves]

The copula $C_h(u, v)$ of X_t and X_{t-h} is not symmetric in u and v , which is consistent with the time irreversibility of this Markov process. More precisely, if we compare with Markov processes with Gaussian or Gumbel copula (Figure 12 and Figure 3, Panel B, respectively), the isodensity curves are distorted, for X_t and X_{t-h} large in absolute value, in directions of smaller $|X_{t-h}|$. This is due to the fact that the lagged canonical direction downweights large observations (in absolute value) more than the present canonical direction doe.

⁴⁴The 0.95-quantile of the marginal distribution is 0.86.

1.5.3.4 The Compound Autoregressive process

Let us consider the Autoregressive Gamma (ARG) process introduced by Gouriéroux and Jasiak (2000), which is a discrete time counterpart of the Cox, Ingersoll, Ross diffusion process [Cox, Ingersoll, Ross (1985)]. The ARG process is a positively valued Markov process, whose conditional distribution of X_t given X_{t-1} is a noncentral gamma distribution $\gamma(\delta, \beta X_{t-1})$, with $\delta, \beta > 0$ ⁴⁵. The ARG process is a CAR process characterized by functions $a(u) = \beta u / (1 + u)$ and $b(u) = -\delta \log(1 + u)$. It is stationary when $|\beta| < 1$, and the marginal distribution is such that $(1 - \beta) X_t$ follows a gamma distribution $\gamma(\delta)$. Parameters β and δ have different interpretations. Indeed, parameter β characterizes function a , and thus the dynamics of the process. Parameter δ is related to the under- or overdispersion of the marginal and conditional distribution of the process [$\delta > 1$ or $\delta < 1$, respectively].

i) Simulated trajectories

In Figure 25 we report simulated trajectories of length $T = 200$ of the ARG processes with $\delta = 0.5, \beta = 0.5$ (Panel A), $\delta = 1.5, \beta = 0.5$ (Panel B), $\delta = 0.5, \beta = 0.95$ (Panel C), and $\delta = 1.5, \beta = 0.95$ (Panel D).

[insert Figure 25: ARG process, simulated trajectories]

The pdf of the marginal distribution is reported in Figure 26 for the different values of the parameters β and δ .

[insert Figure 26: ARG process, marginal pdf]

The process with the largest β parameter [$\beta = 0.95$] features stronger positive dependence. This is consistent with the fact that function a is proportional to β [see discussion in section 3.2.2 ii)]. The marginal distributions of the processes with $\delta > 1$ [$\delta = 1.5$] are hump-shaped, and their trajectories rarely come close to zero.

ii) Autocorrelograms

The autocorrelogram of the ARG process is given by $\text{corr}(X_t, X_{t-h}) = \beta^h$, $h \in \mathbb{N}$, and decays geometrically [Gouriéroux, Jasiak (2000)]. Therefore the autocorrelograms of the ARG processes with $\beta = 0.5$ and $\beta = 0.95$ are the same as those of Gaussian autoregressive processes with $\rho = 0.5$ and $\rho = 0.95$, respectively (Figure 10, Panel B, and Figure 11, Panel B). Parameter δ does not affect the autocorrelogram.

iii) Isodensity curves

Isodensity curves of the ARG process with standard gaussian marginal distribution are reported in Figure 27, for the different values of the parameters.

[insert Figure 27: ARG process, copula isodensity lines, $h = 1$]

⁴⁵Let us recall that X_t follows a noncentral gamma distribution $\gamma(\delta, \beta X_{t-1})$ given X_{t-1} if there exists an intermediary factor Z_t such that the conditional distribution of X_t given Z_t is a gamma distribution $\gamma(\delta + Z_t)$, and the conditional distribution of Z_t given X_{t-1} is a Poisson distribution $\mathcal{P}(\beta X_{t-1})$.

Both parameters β and δ affect the density. In particular, the density is more concentrated along the 45 degree line when β increases, and it is more wedge shaped in the upper quadrant for smaller δ . The density is symmetric since the ARG process is time reversible.

The copula at larger horizons may be easily derived since the conditional distribution of X_t given X_{t-h} is such that $X_t/\frac{1-\beta^h}{1-\beta}$ follows a noncentral gamma distribution $\gamma\left(\delta, \beta^h X_{t-h}/\frac{1-\beta^h}{1-\beta}\right)$. Isodensity curves of the copula at horizon h of the ARG process with $\beta = 0.95$ (with standard gaussian marginal distribution) are reported in Figure 28 ($h = 5, h = 10$) and Figure 29 ($h = 20, h = 50$).

[insert Figure 28: ARG process, copula isodensity lines, $h = 5, 10$]

[insert Figure 29: ARG process, copula isodensity lines, $h = 20, 50$]

1.6 Applications: dynamic framework

These applications involve time series (X_t) , which can be observable, or introduced as latent factors. For expository purpose, we consider Markov processes and the functional dependence parameters characterize the transition distribution. We first explain how copulas can be used for trend correction, or for specifying a dynamic proportional hazard model, useful for liquidity analysis.

The conditional Laplace transform seems appropriate for a number of different problems concerning duration models with stochastic intensity, derivative pricing, as well as the term structure of interest rates.

Finally we explain how the basis of eigenfunctions appearing in nonlinear canonical analysis can be used to identify the drift and volatility function of a diffusion equation, or to approximate option pricing, in particular to define the price of skewness and kurtosis, as a complement to the price of volatility.

1.6.1 Trend correction

Let us consider the standard gaussian autoregressive model with drift:

$$y_t = \alpha_t + \rho y_{t-1} + u_t,$$

where the error terms are iid. Nonstationary features can be introduced by means of either the drift (such as a deterministic linear trend), or the autoregressive coefficient (by a unit root $\rho = 1$). These types of trend effects are of different nature. Indeed we can write:

$$y_t = \beta_t + \tilde{y}_t,$$

where β_t is the solution of the deterministic equation:

$$\beta_t = \alpha_t + \rho \beta_{t-1},$$

and \tilde{y}_t is the zero mean autoregressive process satisfying:

$$\tilde{y}_t = \rho \tilde{y}_{t-1} + u_t, \quad u_t \sim IIN(0, \sigma^2).$$

Equivalently we get:

$$F_t(y_t) = \rho F_{t-1}(y_{t-1}) + u_t, \quad (1.39)$$

where $F_t(y_t) = y_t - \beta_t$. If $\alpha_t = at + b$, say, and $|\rho| < 1$, the variable of interest is transformed by a nonstationary transformation F_t , whereas the copula coincides with the gaussian copula corresponding to $|\rho| < 1$, which is stationary. Symmetrically if $\alpha_t = \alpha$, say, and $\rho = 1$, the transformation F_t becomes stationary, whereas the gaussian copula corresponds to a nonstationary standardized process. This example shows that it can be useful to distinguish marginal nonstationarity (through the transformation F_t) and serial nonstationarity (through the copula).

The principle can be applied to the comparative analysis of the results of firms of a given industrial sector, or of the performances of fund managers. Indeed, it can be interesting to distinguish the nonstationary evolution of the industrial sector (resp. the financial market) and the more stationary comparative rankings of the firms (resp. fund managers). A panel model distinguishing these features can be:

$$X_{it} = G(Z_t, U_{it}), \quad i = 1, \dots, n, \quad (1.40)$$

where the processes (Z_t) , (U_{1t}) , ..., (U_{nt}) , are independent, and:

(Z_t) is nonstationary,

(U_{1t}) , ..., (U_{nt}) , have the same stationary distribution on $[0, 1]$,
associated to the same stationary copula $C(u_t, u_{t-1})$, say,

G is increasing with respect to U for any Z .

In panel model (1.40) X_{it} denotes the profit of firm i in period t (resp. a performance measure for fund manager i in period t), Markov process (U_{it}) represents a latent quality of the firm (resp. of the fund manager), which may be interpreted as a comparative ranking of the firm (resp. of the fund manager), and (Z_t) denotes a nonstationary common factor, representing the trend of the industrial sector (resp. of the financial market).

1.6.2 Dynamic duration model with proportional hazard

Liquidity phenomena can be analyzed from intertrade duration data. The idea is to study the dynamics of the successive intertrade durations, for instance to detect clustering effects, that is the existence of subperiods with large durations, corresponding to weak liquidity [resp. small durations, corresponding to high liquidity]. Different dynamic specifications have been introduced in the literature on high frequency data, such as the autoregressive conditional duration (ACD) models [see Engle, Russell (1998)], which extends the ARCH approach to

durations analysis, or the stochastic volatility duration (SVD) model [see Ghysels, Gouriéroux, Jasiak (2002)], which is the analogue of the stochastic volatility model. However these specifications are introduced by analogy with models usually considered for returns and are not very easy to reconcile with standard models used for durations in the microeconomic framework. The later line can also be followed. Let us denote (X_t) the series of intertrade durations and assume that (X_t) is a Markov process. The dynamics is defined by the transition probability, which explains how the lagged durations X_{t-1} will influence the current one X_t . Thus we get a standard econometric duration model, in which the explanatory variable is the lagged duration. For instance let us consider a specification with proportional hazard in which the conditional hazard function of X_t given X_{t-1} is written as:

$$\lambda(x_t | x_{t-1}) = a(x_{t-1})h_0(x_t),$$

where h_0 is the baseline hazard function, and a is a positive function describing the effect of the lagged duration [see Gagliardini, Gouriéroux (2002a) ⁴⁶]. From the invariance property of the proportional hazard specification by increasing transformation of the variables, the proportional hazard constraint concerns the copula only. In particular, the standardized Markov process $U_t = F(X_t)$, where F is the stationary distribution of (X_t) , features proportional hazard:

$$P[U_t \geq u_t | U_{t-1} = u_{t-1}] = \exp[-A(u_{t-1})H_0(u_t)], \text{ say,}$$

and the condition of uniform marginal distribution identifies function H_0 in terms of A . Thus the functional parameter A characterizes the copula of the Markov process with proportional hazard. Gagliardini, Gouriéroux (2002a) relate the serial dependence patterns of Markov process with proportional hazard (X_t) to different properties of function A . For instance, the strength of positive serial dependence is related to the elasticity of functional parameter A .

1.6.3 Derivative pricing

i) The principle

In this section we consider the pricing of derivative assets within a discrete time stochastic discount factor approach [Harrison, Kreps (1979), Garman, Ohlson (1980), Hansen, Richard (1987)]. It is known that, if agents make their investment decisions at date t ($t \in \mathbb{N}$) based on the available information set, then the prices of actively traded assets are conditional expectations of the payoffs, discounted by means of an appropriate stochastic factor. More precisely, if there are no arbitrage opportunities in the market, there exists a positive random variable $M_{t,t+1}$ such that the price at time t of a derivative asset with payoff g_{t+1} at time $t + 1$ is given by:

$$C(t, g) = E_t[M_{t,t+1}g_{t+1}],$$

⁴⁶In the Cox (1972) model function a is exponential linear. See Hautsch (1999) for an application to intertrade durations.

where E_t denotes expectation conditional to the information at time t . Variable $M_{t,t+1}$ is known as the stochastic discount factor (sdf, or pricing kernel) between t and $t+1$; it accounts for both risk adjustment and time discount. Moreover, the price at time t of a derivative with maturity h and payoff g_{t+h} is given by recursive discounting:

$$C(t, h, g) = E_t [M_{t,t+1} M_{t+1,t+2} \dots M_{t+h-1,t+h} g_{t+h}] = E_t [M_{t,t+h} g_{t+h}], \quad (\text{say}).$$

In discrete time, markets are incomplete and the no-arbitrage opportunity condition does not identify a unique stochastic discount factor. The latter may be selected by imposing additional restrictions. A natural specification ensuring the positivity of the sdf is that of exponential affine sdf [Stutzer (1995,1996), Gouriou, Monfort (2001a,b), Gouriou, Monfort, Polimenis (2002)]:

$$M_{t,t+1} = \exp [-\alpha_t X_{t+1} - \beta_t], \quad (1.41)$$

where (X_t) is a state variable, and α_t and β_t are functions of the information at time t ⁴⁷. An exponential affine sdf is consistent with several econometric specifications and financial equilibrium models considered in the literature [see Gouriou, Monfort (2001a,b), and Polimenis (2002)]. It has the advantage to allow departures from time independence and conditional normality, which underlie usual models, such as the Black-Scholes model.

ii) Stock derivatives

The importance of the conditional Laplace transform of the state variable to summarize serial dependence is due to the exponential form of the sdf. Indeed let us assume that the state variable X_t is the return of a stock with price S_t , $X_t = \log (S_t/S_{t-1})$, and assume for simplicity a zero riskfree rate. Then the no-arbitrage restrictions are:

$$\begin{aligned} 1 &= E_t [M_{t,t+1}], \\ 1 &= E_t [M_{t,t+1} \exp (X_{t+1})]. \end{aligned}$$

From (1.41) they can be written in terms of the conditional Laplace transform $\Psi_t(u) = \Psi(u, X_t) = E[\exp(-uX_{t+1}) | X_t]$ of the state variable:

$$\begin{aligned} \beta_t &= \log \Psi_t(\alpha_t), \\ \Psi_t(\alpha_t) &= \Psi_t(\alpha_t + 1). \end{aligned} \quad (1.42)$$

The no-arbitrage restrictions (1.42) have in general a unique solution, which identifies a unique sdf. The corresponding risk neutral distribution Q of the returns X_t is characterized by the conditional Laplace transform:

$$\frac{Q}{E_t} [\exp(-uX_{t+1})] = \frac{E_t [M_{t,t+1} \exp(-uX_{t+1})]}{E_t [M_{t,t+1}]} = \frac{\Psi_t(\alpha_t + u)}{\Psi_t(\alpha_t)}. \quad (1.43)$$

⁴⁷The exponential affine specification corresponds to the approach based on the Esscher transform [see Esscher (1932), Buhlman et al. (1996), Shyriaev (1999)].

Thus the risk neutral distribution is deduced from the historical distribution by a shift in the Laplace transform.

Example 11: Black-Scholes model in discrete time.

Let us assume that returns X_t are $IIN(\mu - \sigma^2/2, \sigma^2)$. Then the conditional Laplace transform is independent of X_t , and given by $\Psi(u) = \exp[-(\mu - \sigma^2/2)u + \sigma^2 u^2/2]$. The no-arbitrage restrictions imply:

$$\begin{aligned}\alpha &= \frac{\mu - \sigma^2}{\sigma^2}, \\ \beta &= -\frac{\mu}{2} \left(\frac{\mu - \sigma^2}{\sigma^2} \right).\end{aligned}$$

In particular, the coefficient α depends on the ratio of expected returns to variance, and the sdf does not depend on the information at time t , due to time independence of the returns. From (1.43), the dynamics of the returns (X_t) under the risk neutral distribution Q is characterized by the conditional Laplace transform:

$${}^Q E_t [\exp(-uX_{t+1})] = \exp(-\sigma^2 u/2 + \sigma^2 u^2/2).$$

Thus under the risk neutral distribution Q the returns (X_t) are $IIN(-\sigma^2/2, \sigma^2)$, and, as well-known, the correction for risk corresponds to a change of drift, the historical drift μ being replaced by the riskfree rate $r = 0$.

iii) Interest rate derivatives

Let us now consider the pricing of interest rate derivatives. Let the state variable X_t be the short rate between time t and $t + 1$, denoted by r_{t+1} , and assumed to be predetermined at t . The no-arbitrage condition is given by:

$$\exp(-r_{t+1}) = E_t[M_{t,t+1}].$$

If the sdf is exponential affine in the future short rate:

$$M_{t,t+1} = \exp(-\alpha r_{t+2} - \beta_t),$$

and the short rate follows a CAR process: $\Psi_t(u) = E_t[\exp(-ur_{t+2})] = \exp[-a(u)r_{t+1} - b(u)]$, then the restricted sdf becomes:

$$M_{t,t+1} = \exp\{-\alpha r_{t+2} - [1 - a(\alpha)]r_{t+1} + b(\alpha)\}.$$

In particular, any $\alpha \in \mathbb{R}$ is admissible, and the restriction of exponential affine sdf does not select a unique pricing kernel. From the analogue to (1.43) we deduce that the short rate follows a CAR process also under the risk neutral distribution Q . The CAR dynamics is characterized by functions $a^Q(u) = a(u + \alpha) - a(\alpha)$, $b^Q(u) = b(u + \alpha) - b(\alpha)$, deduced from functions a and b by an appropriate drift.

Example 12: The Cox, Ingersoll, Ross model in discrete time.

Let us assume that the short rate (r_{t+1}) follows an Autoregressive Gamma (ARG) process with parameters β and δ [see section 5.3.4]. Under the risk neutral distribution Q the functional parameters characterizing the CAR dynamics of (r_{t+1}) are given by:

$$a^Q(u) = \left(\frac{\beta}{1+\alpha} \right) \frac{u/(1+\alpha)}{1+u/(1+\alpha)}, \quad b^Q(u) = -\delta \log [1 + u/(1+\alpha)].$$

Thus, under the risk neutral distribution Q , the rescaled short rate $(1+\alpha)r_{t+1}$ follows an ARG process with parameters $\beta^Q = \beta/(1+\alpha)^2$, $\delta^Q = \delta$. In particular, the coefficient α of the sdf is such that $\alpha > -1$. The short rate (r_{t+1}) is stationary under Q if $\alpha > -1 + \sqrt{\beta}$, and for $\alpha \in (-1 + \sqrt{\beta}, 0)$ serial dependence is larger under the risk neutral distribution Q than under the original historical probability.

Let us now consider the pricing of interest rate derivatives with exponential payoff⁴⁸. The price at time t of a derivative with residual maturity h and payoff $g(r_{t+h+1}) = \exp(-ur_{t+h+1})$ is given by:

$$\begin{aligned} C(t, h, u) &= E_t [M_{t,t+1} \dots M_{t+h-1,t+h} \exp(-ur_{t+h+1})] \\ &= E_t^Q \{ \exp[-(r_{t+1} + \dots + r_{t+h})] \exp(-ur_{t+h+1}) \}. \end{aligned}$$

Using results (1.36), (1.37) under the risk neutral dynamics, we get:

$$C(t, h, u) = \exp[-A_h(u)r_{t+1} - B_h(u)], \quad h \in \mathbb{N},$$

where $A_h(u)$, $B_h(u)$, $h \in \mathbb{N}$, satisfy the recursion formulae:

$$\begin{aligned} A_h(u) &= a[A_{h-1}(u) + \alpha] - a(\alpha) + 1, & A_1(u) &= a(u + \alpha) - a(\alpha) + 1, \\ B_h(u) &= B_{h-1}(u) + b[A_{h-1}(u) + \alpha] - b(\alpha), & B_1(u) &= b(u + \alpha) - b(\alpha). \end{aligned} \tag{1.44}$$

iv) Term structure of interest rate

Let us finally derive the term structure of interest rates. Let $B(t, t+h)$ denote the price at time t of a zero-coupon bond with residual maturity h . The payoff $g(r_{t+h+1}) = 1$ of the bond is exponential with $u = 0$, and its price $B(t, t+h)$ may be deduced from the results of the previous section. Thus the term structure of interest rates $(r_{t,t+h}, h \in \mathbb{N})$ at time t is given by:

$$\begin{aligned} r_{t,t+h} &= -\frac{1}{h} \log B(t, t+h) \\ &= \frac{A_h}{h} r_{t+1} + \frac{B_h}{h}, \quad h \in \mathbb{N}, \end{aligned}$$

⁴⁸They can be used as a basis for pricing more general derivatives [see e.g. Bakshi, Madan (2000), Duffie, Pan, Singleton (2000), Gouriéroux, Monfort, Polimenis (2002)]

where $A_h, B_h, h \in \mathbb{N}$, satisfy the recursion formulae:

$$\begin{aligned} A_h &= a(A_{h-1} + \alpha) - a(\alpha) + 1, \quad A_1 = 1, \\ B_h &= B_{h-1} + b(A_{h-1} + \alpha) - b(\alpha), \quad B_1 = 0. \end{aligned} \quad (1.45)$$

For any given value of the state variable r_{t+1} , the shape of the term structure $(r_{t,t+h}, h \in \mathbb{N})$ at time t is determined by the patterns of the sequences $A_h/h, B_h/h, h \in \mathbb{N}$. Under mild conditions on function a , we can prove that:

- i) A_h/h is decreasing in $h \in \mathbb{N}$, B_h/h , is increasing in $h \in \mathbb{N}$;
- ii) the smoother function b , the smoother the sequence $B_h/h, h \in \mathbb{N}$; the smoother function a , the smoother $B_h/h, h \in \mathbb{N}$, and the steeper $A_h/h, h \in \mathbb{N}$.

Thus the patterns of functions a and b allow for a large variety of term structures.

Finally, the long term interest rate is given by [see Gouriou, Monfort, Polimenis (2002)]:

$$\lim_{h \rightarrow \infty} r_{t,t+h} = b(A^* + \alpha) - b(\alpha) \equiv r_\infty,$$

where A^* is defined by: $A^* = a(A^* + \alpha) - a(\alpha) + 1$. The flatter b , the lower r_∞ . Similarly, the flatter a , the lower A^* and thus the lower r_∞ .

1.6.4 Prediction and pricing of default risk

In this section we consider prediction and pricing of default risk in intensity based models [see e.g. Duffie, Singleton (1999) in continuous time]. In this approach the probability of default of the borrowers is affected by a latent factor, called intensity, which changes randomly in time. We analyse how serial dependence of the intensity process characterizes the distributions of times to default, the patterns of default correlation, and the way the latter change with the age of the credits.

i) One-dimensional survivor functions

Let τ be a discrete variable ($\tau \in \mathbb{N}$) denoting the time to default of a firm created at time 0, and assume a survivor intensity process (λ_t) [Brémaud (1981)]:

$$P[\tau > t \mid \tau > t-1, (\lambda_t), (X_t)] = \exp(-\lambda_t),$$

where (X_t) denotes the whole path of the state variable. The survivor function of time to default, conditionally to $(\lambda_t), (X_t)$, is given by:

$$\begin{aligned} P[\tau > t \mid (\lambda_t), (X_t)] &= P[\tau > 0 \mid (\lambda_t), (X_t)] \prod_{s=1}^t P[\tau > s \mid \tau > s-1, (\lambda_t), (X_t)] \\ &= \exp\left(-\sum_{s=1}^t \lambda_s\right). \end{aligned}$$

The unconditional survivor function of time to default is deduced by integrating out the intensity (λ_t) . The survivor function of τ at age h is given by:

$$\begin{aligned} S_h(x) &= P[\tau > h+x \mid \tau > h] \\ &= \frac{P[\tau > h+x]}{P[\tau > h]} = \frac{E\left[\exp\left(-\sum_{s=1}^{h+x} \lambda_s\right)\right]}{E\left[\exp\left(-\sum_{s=1}^h \lambda_s\right)\right]}. \end{aligned}$$

The patterns of the survivor function $x \rightarrow S_h(x)$ at age h , and the way it changes with age h , depend on the serial dependence of the intensity process (λ_t) . For instance, if (λ_t) is an iid process, the survivor function corresponds to a geometric distribution: $S_h(x) = \Psi(1)^x$, $x = 0, 1, \dots$, where Ψ is the marginal real Laplace transform of λ_t , and is age independent. At the opposite, when the intensity (λ_t) is perfectly persistent, $\lambda_t = Z$, $\forall t$ (Z a random variable), we get: $S_h(x) = \Psi(x+h)/\Psi(h)$, and the pattern of S_h and its age dependence are related to the Laplace transform Ψ of Z .

More generally, if the intensity process (λ_t) follows a CAR process with parameters a, b , the survivor function $S_h(x)$ at age h admits an analytical expression [see results (1.37), (1.38)], and its patterns and age dependence are related to the functional parameters a, b characterizing the CAR dynamics of the intensity process [see Gouriéroux, Monfort (2002)c].

ii) The term structure of defaultable bonds

The survivor function $S_h(x)$ is also related to the prices of defaultable bonds. Indeed, let us consider a zero-coupon bond with residual maturity h issued by the firm. The payoff at time $t+h$ is $\mathbb{I}(\tau > t+h)$ [where for simplicity we assume a zero recovery rate]. If at time t the firm has not yet defaulted, the price of the bond at time t is given by:

$$\begin{aligned} B^*(t, t+h) &= E\left[M_{t,t+h} \mathbb{I}(\tau > t+h) \mid \underline{X}_t, \tau > t\right] \\ &= \frac{E\left[M_{t,t+h} \mathbb{I}(\tau > t+h) \mid \underline{X}_t\right]}{E\left[\mathbb{I}(\tau > t) \mid \underline{X}_t\right]} \end{aligned}$$

Let us assume that the intensity process (λ_t) is independent of the state variable (X_t) . Then we get:

$$B^*(t, t+h) = B(t, t+h) S_t(h).$$

Thus the price at time t of a defaultable bond with maturity h is equal to the price of a riskfree bond with the same maturity times the survivor probability for h periods at time t . In particular the term structure of the spread between defaultable bonds and riskfree bonds is given by:

$$s_{t,t+h} = -\left(\frac{1}{h} \log B^*(t, t+h) - \frac{1}{h} \log B(t, t+h)\right) = -\frac{1}{h} \log S_t(h).$$

Under the assumption of independence between intensity and state variables (the so-called actuarial assumption), the default spread is a deterministic function of the historical survivor

function. For instance, when the intensity process (λ_t) is iid, the term structure of the spread is flat, and constant during the life of the firm: $s_{t,t+h} = -\log \Psi(1) > 0$. In general, the patterns and dynamics of the spread are related to serial dependence of the intensity process.

iii) Term structure of default dependence

Let us now consider two firms and study the dependence between their default events. Let τ_1 and τ_2 be discrete variables denoting times to default of the two firms. We can generalize the specification in section 4.3 and assume that times to default τ_1 and τ_2 are conditionally independent given a common intensity process (λ_t) :

$$P[\tau_1 > t_1, \tau_2 > t_2 \mid (\lambda_t), (X_t)] = \exp \left[- \left(\sum_{s=1}^{t_1} \lambda_s + \sum_{s=1}^{t_2} \lambda_s \right) \right].$$

Then the joint survivor function at age h when both firms are still alive is given by:

$$\begin{aligned} S_h(x, y) &= P[\tau_1 > h+x, \tau_2 > h+y \mid \tau_1 > h, \tau_2 > h] \\ &= \frac{E \left\{ \exp \left[- \left(2 \sum_{s=1}^{x^*+h} \lambda_s + \sum_{s=x^*+h+1}^{y^*+h} \lambda_s \right) \right] \right\}}{E \left[\exp \left(-2 \sum_{s=1}^h \lambda_s \right) \right]}, \end{aligned}$$

where $x^* = \min(x, y)$, $y^* = \max(x, y)$. In particular the marginal survivor function of time to default τ_1 of firm 1, say, at age h when both firms are still alive is given by:

$$S_h(x, 0) = \frac{E \left\{ \exp \left[- \left(2 \sum_{s=1}^h \lambda_s + \sum_{s=h+1}^{x+h} \lambda_s \right) \right] \right\}}{E \left[\exp \left(-2 \sum_{s=1}^h \lambda_s \right) \right]},$$

it differs in general from the one-dimensional survivor function $S_h(x)$ computed in i) since it takes into account information on both firms.

When the intensity process (λ_t) is perfectly persistent, $\lambda_t = Z \forall t$, we get the discrete time analogue of the results in section 4.3:

$$S_h(x, y) = E \{ \exp [-(x+y+2h)Z] \} / E [\exp (-2hZ)] = \Psi(x+y+2h) / \Psi(2h).$$

iv) Distribution of the first-to-default time

Several credit derivatives involve the first (resp. the second, ...) to default time in a basket of borrowers [see section 4.2 and the next section]. Therefore it is important to study the distribution of the order statistics. Let us for instance consider the first-to-default time $\tau = \min(\tau_1, \tau_2)$. Its survivor distribution at age h is given by:

$$\begin{aligned} P[\tau > h+x \mid \tau > h] &= P[\tau_1 > h+x, \tau_2 > h+x \mid \tau_1 > h, \tau_2 > h] \\ &= S_h(x, x) = \frac{E \left\{ \exp \left[- \left(2 \sum_{s=1}^{x+h} \lambda_s \right) \right] \right\}}{E \left[\exp \left(-2 \sum_{s=1}^h \lambda_s \right) \right]}. \end{aligned}$$

It corresponds to the one-dimensional survivor function $S_h(x)$ of paragraph i), with a doubled intensity⁴⁹, and is written in terms of conditional Laplace transform of the intensity process.

v) Derivative pricing

Let us consider the pricing of credit derivatives when information about default events of both firms is available. Let us first consider a bond with residual maturity h issued by firm 1, say. If at time t both firms have not yet defaulted, the price $B_1^*(t, t+h)$ of this bond is given by:

$$\begin{aligned} B_1^*(t, t+h) &= E \left[M_{t,t+h} \mathbb{I}(\tau_1 > t+h) \mid \underline{X}_t, \tau_1 > t, \tau_2 > t \right] \\ &= \frac{E \left[M_{t,t+h} \mathbb{I}(\tau_1 > t+h) \mathbb{I}(\tau_2 > t) \mid \underline{X}_t \right]}{E \left[\mathbb{I}(\tau_1 > t) \mathbb{I}(\tau_2 > t) \mid \underline{X}_t \right]} \\ &= B(t, t+h) S_t(h, 0). \end{aligned}$$

Since the payoff depends on default of firm 1 only, the marginal survivor distribution of duration τ_1 is involved. However, since the default intensity of the two firms are linked, the conditional information at time t on both firms matters.

Let us now consider a first-to-default basket, that is a derivative which pays 1 \$ if the first default $\tau = \min(\tau_1, \tau_2)$ occurs after maturity $t+h$. The payoff at maturity is given by: $\mathbb{I}(\tau > t+h)$. If no default has occurred at time t , the price at time t of such a derivative with residual maturity h is given by:

$$\begin{aligned} C(t, t+h) &= E \left[M_{t,t+h} \mathbb{I}(\tau > t+h) \mid \underline{X}_t, \tau > t \right] \\ &= \frac{E \left[M_{t,t+h} \mathbb{I}(\tau > t+h) \mid \underline{X}_t \right]}{E \left[\mathbb{I}(\tau > t) \mid \underline{X}_t \right]} \\ &= B(t, t+h) S_t(h, h), \end{aligned}$$

and involves the survivor probability of the first-to-default time.

1.6.5 Continuous time models

Nonlinear canonical analysis can be useful, when a dynamics is specified in continuous time and assumed to satisfy a one-dimensional diffusion equation:

$$dy_t = \mu(y_t) dt + \sigma(y_t) dW_t, \quad (1.46)$$

where μ and σ are the drift and volatility function, respectively. Indeed it is known that the transition at horizon 1 admits a reversible nonlinear canonical decomposition:

$$f(y_t \mid y_{t-1}) = f(y_t) \left[1 + \sum_{j=1}^{\infty} \exp(-\lambda_j) \varphi_j(y_t) \varphi_j(y_{t-1}) \right], \quad (1.47)$$

⁴⁹Indeed in this conditional independence framework the intensity process corresponding to $\tau = \min(\tau_1, \tau_2)$ is the sum of the intensities of the two durations, that is $(2\lambda_t)$.

where $\lambda_j, j = 1, 2, \dots$, $\varphi_j, j = 1, 2, \dots$ are eigenvalues and eigenfunctions of the infinitesimal generator associated with the diffusion equation⁵⁰. The generator A is defined by:

$$A\varphi(y) = \mu(y) \frac{d\varphi}{dy}(y) + \frac{1}{2}\sigma(y)^2 \frac{d^2\varphi}{dy^2}(y), \quad (1.48)$$

and underlies the well-known Ito's formula.

1.6.5.1 Nonparametric identification of the drift and volatility functions

In practice the volatility function is useful to compute the price of a derivative written on y , whereas both the drift and volatility functions have to be known for predicting the future value of a portfolio of derivatives, in particular for computing the Value at Risk of this portfolio. The spectral properties of the infinitesimal generator can be used as the basis for identifying nonparametrically the drift and volatility function. Indeed let us assume known the nonlinear canonical decomposition (1.47). Then we get:

$$\begin{aligned} \mu(y) \frac{d\varphi_1}{dy}(y) + \frac{1}{2}\sigma(y)^2 \frac{d^2\varphi_1}{dy^2}(y) &= \lambda_1\varphi_1(y), \quad \forall y, \\ \mu(y) \frac{d\varphi_2}{dy}(y) + \frac{1}{2}\sigma(y)^2 \frac{d^2\varphi_2}{dy^2}(y) &= \lambda_2\varphi_2(y), \quad \forall y. \end{aligned} \quad (1.49)$$

For any value y , the system can be solved to get $\mu(y)$ and $\sigma^2(y)$ as functions of $\lambda_1, \lambda_2, \varphi_1, \varphi_2$. This interpretation has been used to construct a nonparametric estimation method of μ, σ^2 [Darolles, Florens, Gouriéroux (2001)]. In a first step the joint density of (y_t, y_{t-1}) is estimated nonparametrically by a kernel estimator, and nonlinear canonical analysis is performed on the estimated joint density to deduce estimates of the first two canonical directions φ_1, φ_2 and canonical correlations $\exp(-\lambda_1), \exp(-\lambda_2)$. In a second step equations (1.49) are used to derive a non parametric estimator for the drift and volatility functions μ, σ .

1.6.5.2 Derivative pricing

Nonlinear canonical analysis can also be used for approximate derivative pricing. Indeed let us assume that the stochastic differential equation (1.46) provides the dynamics of (y_t) under the risk neutral probability and that the riskfree interest rate is zero. The price at t of a payoff $g(y_{t+h})$ is:

$$C_t(g, h) = E_t[g(y_{t+h})].$$

The derivative price is easily computed for the payoff corresponding to the eigenfunction of the infinitesimal operator, since:

$$C_t(\varphi_j, h) = \exp(-\lambda_j h) \varphi_j(y_t), \quad \text{for any } j, h, t.$$

⁵⁰See Demoura (1993), Hansen, Scheinkman, Touzi (1998), Chen, Hansen, Scheinkman (1999), Darolles, Florens, Gouriéroux (2001).

Approximated pricing formulas can be deduced for a general payoff g by approximating the payoff by its projection on the first eigenfunctions $\varphi_1, \varphi_2, \dots$. In particular several continuous time models used in applied finance, such as the Black-Scholes model, the Vasicek model, the Cox-Ingersoll-Ross model are such that the eigenfunction φ_j is a polynomial of degree j . Thus the idea is to price φ_1 that is the expected value, φ_2 that is the volatility, φ_3 that is the skewness, φ_4 that is the kurtosis, and to deduce an approximate price for g from the valuation of the first moments [Madan, Milne (1994)]. Loosely speaking the usual basis of payoffs $(y - k)^+$ proposed on the market by means of European call is not necessarily the most appropriate for pricing and hedging with a limited number of basis derivatives.

REFERENCES

- Aït-Sahalia, Y., and A., Lo, (1998): "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices", *Journal of Finance*, 53, 499-547.
- Ang, A., and G., Bekaert, (2002a): "Short Rate Nonlinearities and Regime Switches", *Journal of Economic Dynamics and Control*, 26, 2343-1274.
- Ang, A., and G., Bekaert, (2002b): "Regime Switches in Interest Rates", *Journal of Business and Economics Statistics*, forthcoming.
- Bakshi, G., and D., Madan, (2000): "Spanning and Derivative-Security Valuation", *Journal of Financial Economics*, 55, 205-238.
- Barrett, J., and D., Lampard, (1955): "An Expansion for Some Second Order Probability Distributions and its Application to Noise Problems", *IRE Trans. PGIT*, 10-15.
- Bingham, N., Goldie, C., and J., Teugels, (1987): *Regular Variation*, Cambridge University Press.
- Bollerslev, T., Chou, R., and K., Kroner, (1992) "ARCH Modelling in Finance: a Review of the Theory and Empirical Evidence", *Journal of Econometrics*, 52, 5-59.
- Bollerslev, T., Engle, R., and D., Nelson, (1994) "ARCH Models", in *Handbook of Econometrics, Vol. IV*, R. F. Engle and D. L. McFadden eds., North Holland, Amsterdam.
- Bouyé, E., Gaussel, N., and M., Salmon, (2000): "Investigating Dynamic Dependence using Copulae", Working Paper.
- Brandt, M. W., (1999): "Estimating Portfolio and Consumption Choice: A Conditional Euler Equation Approach", *Journal of Finance*, Vol. LIV, 1609-1645.
- Brémaud, P., (1981): *Point Processes and Queues, Martingale Dynamics*. Springer Verlag.
- Buhlman, H., Delbaen, F., Embrechts, P., and A., Shyraev, (1996): "No Arbitrage, Change of Measure and Conditional Esscher Transforms in a Semi-Martingale Model of Stock Processes", *CWI Quarterly*, 9, 291-317.
- Campbell, J., Lo, A., and C., MacKinlay, (1997): *The Econometrics of Financial Markets*, Princeton University Press, New Jersey.
- Chen, X., Hansen, L., and J., Scheinkman, (1999): "Shape Preserving Estimation of Diffusion", University of Chicago Working Paper.

Cherubini, U., and E., Luciano, (2001): "Pricing Vulnerable Options With Copulas", University of Turin, Working Paper.

Clark, P., (1973): "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices", *Econometrica*, 41, 135-155.

Clayton, D. G., (1978): "A Model for Association in Bivariate Life Tables and its Application in Epidemiological Studies of Family Tendency in Chronic Disease Incidence", *Biometrika*, 65, 141-151.

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.

Cox, J., Ingersoll, J., and S., Ross (1985): "A Theory of the Term Structure of Interest Rates", *Econometrica*, 53, 385-408.

Darolles, S., Florens, J.-P., and C., Gouriéroux, (2001): "Kernel Based Nonlinear Canonical Analysis", forthcoming *Journal of Econometrics*.

Darolles, S., Gouriéroux, C., and J., Jasiak, (2002): "Structural Laplace Transform and Compound Autoregressive Models", CREST Working Paper.

Demoura, S., (1993): "Theory and Application of Transition Operator to the Analysis of Economic Time Series. Necessary and Sufficient Conditions for Nonlinearities in Economic Dynamics; Aliasing Problem, Manuscript.

Dhaene, J., and M. J., Goovaerts, (1996): "Dependency of Risks and Stop-Loss Order", *Astin Bull.*, 26, 201-212.

Duffie, D., (1998): "First-To-Default Valuation", University of Paris, Dauphine, and Graduate School of Business, Stanford University, Working Paper.

Duffie, D., Pan, J., and K., Singleton, (2000): "Transform Analysis and Asset Pricing for Affine Jump Diffusions", *Econometrica*, 68, 1343-1376.

Duffie, D., and K., Singleton, (1999): "Modelling Term Structures of Defaultable Bonds", *The Review of Financial Studies*, 12, 687-720.

Dunford, N. and J., Schwartz, (1968): *Linear Operators: Part I*, Wiley, New York.

Durrleman, V., Nikeghbali, A., and T., Roncalli, (2000): "How to Get Bounds for Distribution Convolution? A Simulation Study and an Application to Risk Management", GRO Crédit Lyonnais, Working Paper.

Embrechts, P., Höing, A., and A., Juri, (2001): "Using Copulae to Bound the Value-at-Risk for Functions of Dependent Risks", Working Paper.

Embrechts, P., Mc Neil, A. J., and D., Straumann, (1999): "Correlation: Pitfalls and Alternatives", *Risk*, 12, 93-113.

Engle, R. (1982): "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of the U. K. Inflation", *Econometrica*, 50, 987-1008.

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

Esscher, F., (1932): "On the Probability Function in the Collective Theory of Risk", *Skandinavisk Aktuarietidskrift*, 15, 165-195.

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.

Feller, W., (1971): *An Introduction to Probability Theory and Its Applications*, Second Edition, Vol. 2., Wiley.

Fermanian, J., and O., Scaillet, (2002): "Nonparametric Estimation of Copulas for Time Series", Working Paper.

Frey, R., and A. J., McNeil (2001): "Modelling Dependent Defaults", ETH Zurich working paper.

Gagliardini, P., and C., Gouriéroux, (2002a), "Duration Time Series Models with Proportional Hazard", CREST and USI Working Paper.

Gagliardini, P., and C., Gouriéroux, (2002b), "Constrained Nonparametric Copulas", CREST and USI Working Paper.

Garman, M., and J., Ohlson, (1980): "Information and the Sequential Valuation of Assets in Arbitrage Free Economics", *Journal of Accounting Research*, 18, 420-440.

Genest, C., (1987): "Frank's Family of Bivariate Distributions", *Biometrika*, 74, 549-555.

Genest, C., Ghoudi, K., and L.-P., Rivest, (1998): Discussion of "Understanding Relationships Using Copulas", by E. Frees and E. Valdez, *North American Actuarial Journal*, 3, 143-149.

Genest, C., Ghoudi, K., and L.-P., Rivest, (1998): "A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions", *Biometrika*, 82, 543-552.

Genest, C., and R. J., Mc Kay, (1986): "Copules Archimédiennes et familles de lois bidimensionnelles dont les marges sont données", *Can. J. Statistics*, 14, 145-159.

Genest, C., and L.-P., Rivest, (1993): "Statistical Inference Procedures for Bivariate Archimedean Copulas", *Journal of the American Statistical Association*, 88, 1034-1043.

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., Gouriéroux, C., and J., Jasiak, (2002): "Stochastic Volatility Duration Models", forthcoming *Journal of Econometrics*.

Gorgens, T., and J., Horowitz, (1999): "Semi-parametric Estimation of a Censored Regression Model with an Unknown Transformation of the Dependent Variable", *Journal of Econometrics*, 90, 155-191.

Gouriéroux, C., and J., Jasiak, (1999): "Nonlinear Persistence and Copersistence", CREST and University of Toronto Working Paper.

Gouriéroux, C., and J., Jasiak, (2000): "Autoregressive Gamma Processes", CREST Working Paper.

Gouriéroux, C., and J., Jasiak, (2001a): "State-space Models with Finite Dimensional Dependence", *Journal of Time Series Analysis*, 22, 665-678.

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

Gouriéroux, C., and J., Jasiak, (2002a): "Nonlinear Autocorrelograms: An Application to Intertrade Durations", *Journal of Time Series Analysis*, 33, 127-154.

Gouriéroux, C., and J., Jasiak, (2002b): *Econometric Analysis of Individual Risks*, www.dept.econ.yorku.ca/~jasiakj.

Gouriéroux, C., and A., Monfort, (2001a): "Econometric Specifications of Stochastic Discount Factor Models", CREST Working Paper.

Gouriéroux, C., and A., Monfort, (2001b): "Pricing with Splines", CREST Working Paper.

Gouriéroux, C., and A., Monfort, (2002a): "Equidependence in Qualitative and Duration Models", CREST Working Paper.

Gouriéroux, C., and A., Monfort, (2002b): "Econometrics of Efficient Portfolios", CREST Working Paper.

Gouriéroux, C., and A., Monfort, (2002c): "Age and Term Structure in Duration Models", CREST Working Paper.

Gourieroux, C., Monfort, A., and V., Polimenis, (2002): "Affine Term Structure Models", CREST Working Paper.

Gourieroux, C., and C. Y., Robert, (2001): "Tails and Extremal Behaviour of Stochastic Unit Root Models", CREST Working Paper.

Gumbel, E. J., (1960): "Distributions des valeurs extrêmes en plusieurs dimensions", Publ. Inst. Statist. Univ. Paris, 9, 171-173.

Hamilton, J., (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle", *Econometrica*, 57, 357-384.

Hamilton, J., (1990): "Analysis of Time Series Subject to Changes in Regime", *Journal of Econometrics*, 45, 39-70.

Hamilton, J., (1993): "Estimation, Inference, and Forecasting of Time Series Subject to Changes in Regime", in *Handbook of Statistics, Vol 11*, G. Maddala, C. Rao, and H. Vinod, eds., North Holland, New York.

Han, A., (1987)a: "Nonparametric Analysis of a Generalized Regression Model", *Journal of Econometrics*, 35, 305-316.

Han, A. (1987)b: "A Nonparametric Analysis of Transformations", *Journal of Econometrics*, 35, 191-209.

Hansen, L., and S., Richard, (1987): "The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models", *Econometrica*, 55, 587-613.

Hansen, L., Scheinkman, J., and N., Touzi, (1998): "Spectral Methods for Identifying Scalar Diffusions", *Journal of Econometrics*, 86, 1-32.

Härdle, W., and T., Stocker, (1989): "Investigating Smooth Multiple Regression by the Method of Average Derivatives", *JASA*, 84, 986-995.

Härdle, W., and A., Tsybakov, (1997): "Local Polynomial Estimation of the Volatility Function", *Journal of Econometrics*, 81, 223-242.

Harrison, M., and D., Kreps, (1979): "Martingales and Arbitrage in Multiperiod Securities Markets", *Journal of Economic Theory*, 20, 381-408.

Hastie, T., and R., Tibshirani, (1990): *Generalized Additive Models*, in Monographs on Statistics and Applied Probability, 43, Chapman and Hall, London.

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.

Horowitz, J., (1996): "Semiparametric Estimation of a Regression Model with an Unknown Transformation of the Dependent Variable", *Econometrica*, 64, 103-137.

Horowitz, J., and W., Härdle (1996): "Direct Semiparametric Estimation of Single Index Models with Discrete Covariates", *JASA*, 91, 1632-1640.

Ichimura, M., (1993): "Semi-parametric Least-Squares and Weighted SLS Estimation of Single Index Models", *Journal of Econometrics*, 58, 71-120.

Ingersoll, J., (1987): *Theory of Financial Decision Making*, Rowman & Littlefield, Totoway NJ.

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

Jouanin, J.-F., Rapuch, G., Riboulet, G., and T., Roncalli, (2001): "Modelling Dependence for Credit Derivatives with Copulas", GRO Crédit Lyonnais, Working Paper.

Juri, A., and M. V., Wuethrich, (2002a): "Copula Convergence Theorems for Tail Events", *Insurance: Mathematics and Economics*, 30, 411-427.

Juri, A., and M. V., Wuethrich, (2002b): "Tail Dependence from a Distributional Point of View", ETHZ Working Paper.

Kendall, M. G., (1938): "A New Measure of Rank Correlation", *Biometrika*, 30, 81-93.

Kimeldorf, G., and A. R., Sampson (1975): "Uniform Representations of Bivariate Distributions", *Comm. Statist.*, 4, 617-627.

Lancaster, H., (1968): "The Structure of Bivariate Distributions", *Annals of Mathematical Statistics*, 29, 719-736.

Li, D. X., (2000): "On Default Correlation: a Copula Function Approach", *Journal of Fixed Income*, 9, 43-54.

Madan, D., and F., Milne (1994): "Contingent Claims Valued and Hedged by Pricing and Investing in a Basis", *Mathematical Finance*, 4, 223-245.

Mandelbrot, R. (1963): "The Variation of Certain Speculative Prices", *The Journal of Business*, 36, 394-411.

Markowitz, H. (1976): *Portfolio Selection*, Yale University Press, New Heaven.

Marshall, A. W., and I., Olkin, (1967): "A Multivariate Exponential Distribution", *J. Amer. Statist. Assoc.*, 62, 30-44.

Merton, R. C., (1974): "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates", *Journal of Finance*, 29, 449-470.

Meyn, S. P., and R. L., Tweedie, (1993): *Markov Chains and Stochastic Stability*, Communications and Control Engineering Series, Springer.

Morgenstern, D., (1956): "Einfache Beispiele Zweidimensionaler Verteilungen", *Mitt. Math. Statist.*, 8, 234-235.

Müller, A., and M., Scarsini, (2000): "Some Remarks on the Supermodular Order", *Journal of Multivariate Analysis*, 73, 107-119.

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

Neveu, J., (1968): "Processus aléatoires gaussiens", Presses de l'Université de Montréal.

Oakes, D., (1986): "Semiparametric Inference in a Model for Association in Bivariate Survival Data", *Biometrika*, 73, 353-361.

Oakes, D., (1989): "Bivariate Survival Models Induced by Frailties", *Journal of the American Statistical Association*, 84, 487-493.

Oakes, D., (1994): "Multivariate Survival Distributions", *J. Nonparametric Stat.*, 3, 343-354.

Patton, A., (2002): "On the Importance of Skewness and Asymmetric Dependence in Stock Returns for Asset Allocation", UCSD Working Paper.

Polimenis, V., (2002): "Essays in Discrete Time Asset Pricing", PhD Thesis, The Wharton School, University of Pennsylvania.

Powell, J., Stock, J., and T., Stocker, (1989): "Semi-parametric Estimation of Index Coefficients", *Econometrica*, 57, 474-523.

Rapuch, G., (2001): "Dépendance stochastique et pricing des actifs contingents", Mémoire de stage de DEA de Finance et Probabilités de Paris 6.

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

Rosenberg, J. V., (1999): "Semiparametric Pricing of Multivariate Contingent Claims", Stern School of Business Working Paper.

Rosenberg, J. V., (2001): "Nonparametric Pricing of Multivariate Contingent Claims", Stern School of Business Working Paper.

Schönbucher, P. J., and D., Schubert, (2001): "Copula-Dependent Default Risk in Intensity Models", University of Bonn, Working Paper.

Shyraev, A., (1999): *Essentials of Stochastic Finance: Facts, Models, Theory*. World Scientific Publishing, London.

Sklar, A., (1959): "Fonctions de répartition à n dimensions et leurs marges", Publ. Inst. Statist. Univ. Paris, 8, 229-231.

Stutzer, M., (1995): "A Bayesian Approach to Diagnosis of Asset Pricing Models", Journal of Econometrics, 68, 367-397.

Stutzer, M., (1996): "A Simple Nonparametric Approach to Derivative Security Valuation", Journal of Finance, 51, 1633-1652.

Tchen, A. H., (1980): "Inequalities for Distributions with Given Marginals", Annals of Statistics, 8, 814-827.

Teräsvirta, T., Tjøstheim, D., and C., Granger, (1994): "Aspects of Modelling Nonlinear Time Series", in R. Engle and D. McFadden (eds.), *Handbook of Econometrics*, Vol. IV, Elsevier, Amsterdam.

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

Wiener, N., (1958): "Nonlinear Problems in Random Theory", MIT Press, Cambridge.

Appendix 1

Dependence ordering in the compound model

i) Dependence ordering

Let us introduce a dependence ordering in the compound model. We have the following Lemma.

Lemma A.1: Let (X_1, Y_1) and (X_2, Y_2) be pairs of positive variables satisfying compound models, with functions a_1 and a_2 respectively, and same pair of marginal distributions. Then:

$$\begin{aligned} \text{cov} [-\exp(-uX_1), -\exp(-vY_1)] &\geq \text{cov} [-\exp(-uX_2), -\exp(-vY_2)], \\ \forall u, v &\geq 0, \end{aligned} \tag{a.1}$$

if and only if:

$$a_1(u) \geq a_2(u), \quad \forall u \geq 0.$$

Proof: By the restriction of identical marginal distributions, condition (a.1) is equivalent to:

$$E [\exp(-uX_1 - vY_1)] \geq E [\exp(-uX_2 - vY_2)], \quad \forall u, v \geq 0.$$

Using the expression (1.24) for the joint Laplace transform of a compound model, we get:

$$\frac{\Psi_Y [a_1(u) + v]}{\Psi_Y [a_1(u)]} \geq \frac{\Psi_Y [a_2(u) + v]}{\Psi_Y [a_2(u)]}, \quad \forall u, v \geq 0.$$

Since $-\log \Psi_Y$ is concave, function $s \mapsto \log \Psi_Y [s + v] - \log \Psi_Y [s]$ is increasing, for any $v \geq 0$. Therefore (a.1) is equivalent to:

$$a_1(u) \geq a_2(u), \quad \forall u \geq 0.$$

Q.E.D

Condition (a.1) is equivalent to a condition involving more general transformations of X, Y . Indeed, let us assume that $\text{cov} [g(X_1), h(Y_1)] \geq \text{cov} [g(X_2), h(Y_2)]$, for any increasing exponential transformations $g(X) = -\exp(-uX)$, $h(Y) = -\exp(-vY)$. Then, by considering the limit $u, v \rightarrow 0$, the inequality is valid for increasing affine transformations g, h of X, Y . Moreover, by a continuity argument, the inequality applies to any increasing transformations g, h , which are limit of positive combinations of increasing affine or exponential functions.

ii) PQD ordering

The ordering introduced in the previous section is weaker than PQD ordering. However any

pair (X, Y) following a compound model and satisfying the conditions of Example 5 features PQD. Indeed, for any increasing transformations g, h we get:

$$Cov[g(x), h(Y)] = Cov(E[g(X) | Y], h(Y)) \geq 0,$$

since the conditional distribution of X given Y is increasing in Y for first order stochastic dominance, and thus $E[g(X) | Y]$ is an increasing function of Y .

Appendix 2

The copula of $(\min(U, V), \max(U, V))$

i) The cdf of $(\min(U, V), \max(U, V))$

Let us compute $P[\min(U, V) \leq x^*, \max(U, V) \leq y^*]$. Two cases can be distinguished. If $x^* \leq y^*$, we get:

$$\begin{aligned}
 & P[\min(U, V) \leq x^*, \max(U, V) \leq y^*] \\
 = & P[\min(U, V) \leq x^*] - P[\min(U, V) \leq x^*, \max(U, V) \geq y^*] \\
 = & 1 - P[\min(U, V) \geq x^*] - P[U \leq x^*, V \geq y^*] - P[V \leq x^*, U \geq y^*] \\
 = & \{1 - P[U \geq x^*, V \geq x^*] - P[U \leq x^*] - P[V \leq x^*]\} \\
 & + P[U \leq x^*, V \leq y^*] + P[U \leq y^*, V \leq x^*] \\
 = & C(x^*, y^*) + C(y^*, x^*) - C(x^*, x^*), \text{ by Poincaré formula.}
 \end{aligned}$$

If $x^* > y^*$, we get:

$$\begin{aligned}
 P[\min(U, V) \leq x^*, \max(U, V) \leq y^*] &= P[\max(U, V) \leq y^*] \\
 &= C(y^*, y^*).
 \end{aligned}$$

ii) Marginal distributions

The marginal distributions of $\min(U, V)$ and $\max(U, V)$ are given by:

$$\begin{aligned}
 P[\min(U, V) \leq x^*] &= 2x^* - C(x^*, x^*) =: \varphi_C(x^*), \\
 P[\max(U, V) \leq y^*] &= C(y^*, y^*) =: \psi_C(y^*).
 \end{aligned}$$

Since $\min(U, V) \leq \max(U, V)$, we have $\varphi_C \geq \psi_C$.

iii) The copula

Let us denote by $C^*(u, v)$ the copula of $(\min(U, V), \max(U, V))$. Two cases have to be distinguished, according to whether $\varphi_C^{-1}(u) \leq \psi_C^{-1}(v)$ or $\varphi_C^{-1}(u) \geq \psi_C^{-1}(v)$. The first condition is equivalent to:

$$\begin{aligned}
 u &\leq \varphi_C[\psi_C^{-1}(v)] = 2\psi_C^{-1}(v) - C[\psi_C^{-1}(v), \psi_C^{-1}(v)] \\
 &= 2\psi_C^{-1}(v) - v,
 \end{aligned}$$

that is $(u + v)/2 \leq \psi_C^{-1}(v)$, or $C[(u + v)/2, (u + v)/2] \leq v$. Thus for $u, v \in [0, 1]$ such that $C[(u + v)/2, (u + v)/2] \leq v$ we get:

$$\begin{aligned}
 C^*(u, v) &= C(\varphi_C^{-1}(u), \psi_C^{-1}(v)) + C(\psi_C^{-1}(v), \varphi_C^{-1}(u)) \\
 &\quad - C(\varphi_C^{-1}(u), \varphi_C^{-1}(u)) \\
 &= C(\varphi_C^{-1}(u), \psi_C^{-1}(v)) + C(\psi_C^{-1}(v), \varphi_C^{-1}(u)) \\
 &\quad + u - 2\varphi_C^{-1}(u).
 \end{aligned}$$

For $u, v \in [0, 1]$ such that $C[(u + v)/2, (u + v)/2] \geq v$ we get:

$$C^*(u, v) = v.$$

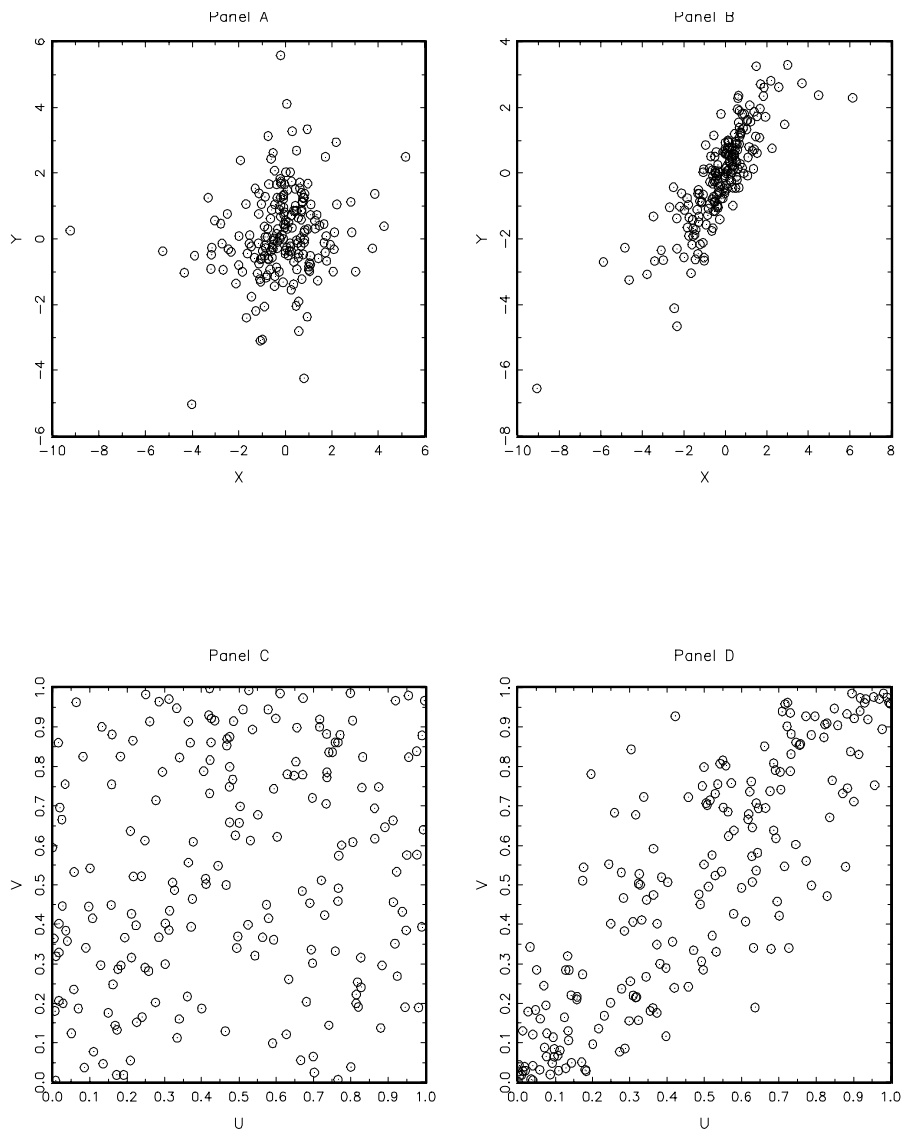


Figure 1.1: Gaussian copula with t_4 -margins. Scatterplot of (X, Y) and (U, V) with correlation parameter $\rho = 0.2$ in Panel A and Panel B, respectively. Scatterplot of (X, Y) and (U, V) with correlation parameter $\rho = 0.8$ in Panel C and Panel D, respectively.

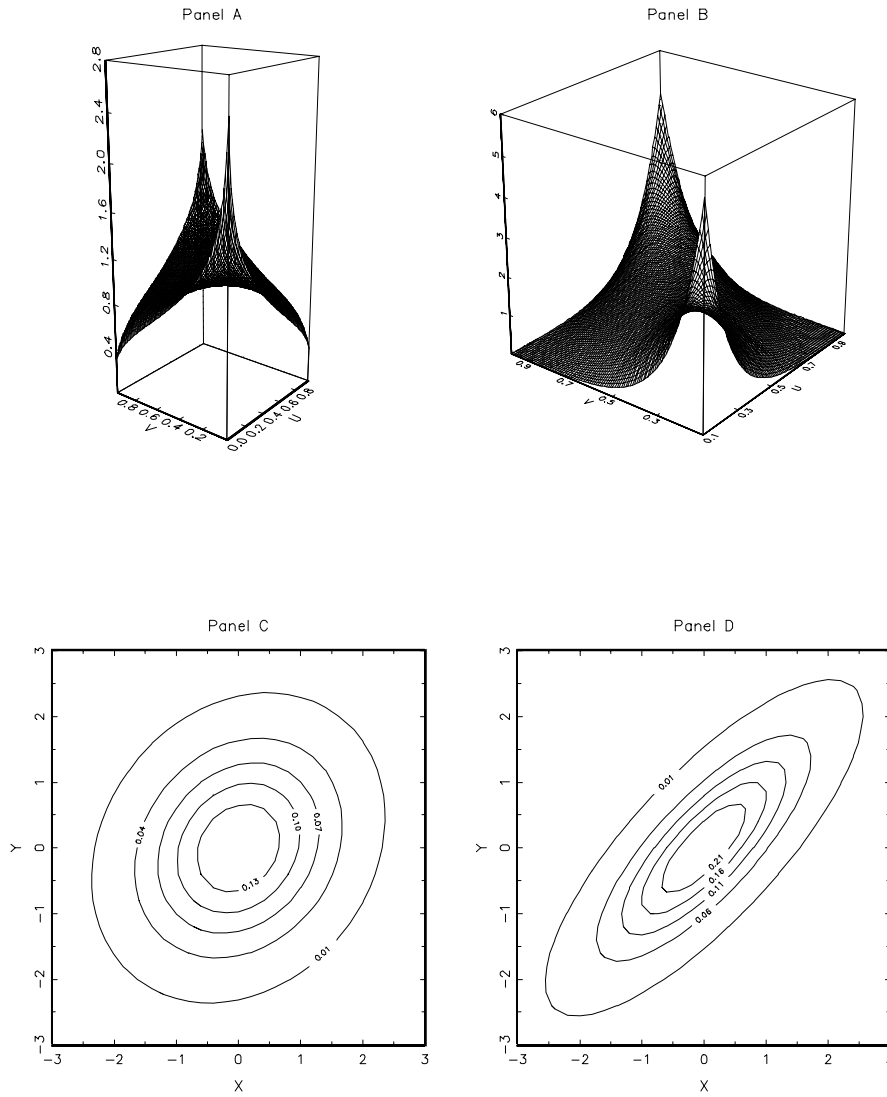


Figure 1.2: Gaussian copula. In Panel A and B we report density plots for correlation parameters $\rho = 0.2$ and $\rho = 0.8$, respectively. Isodensity curves of variables (X, Y) with standard gaussian margins are reported in Panel C and D, for $\rho = 0.2$ and $\rho = 0.8$, respectively.

Figure 1.3: Archimedean Copulas. Scatterplot of variables (U, V) with Gumbel (Clayton) copula in Panel A (Panel C, respectively), and isodensity curves of (X, Y) with Gumbel (Clayton) copula and standard normal marginal distributions in Panel B (Panel D, respectively). Copula parameters are chosen such that Kendall's tau is equal to that of a Gaussian copula with $\rho = 0.8$.

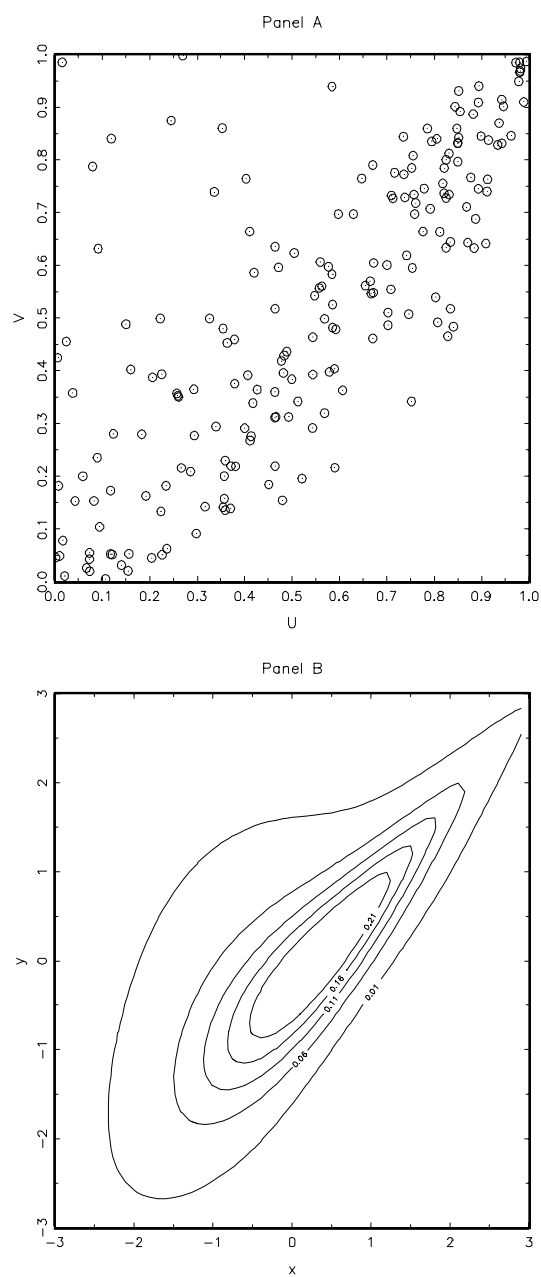


Figure 1.4: Asymmetric logistic copula. Scatterplot of variables (U, V) (Panel A), and isodensity curves of variables (X, Y) with standard normal marginal distributions (Panel B).

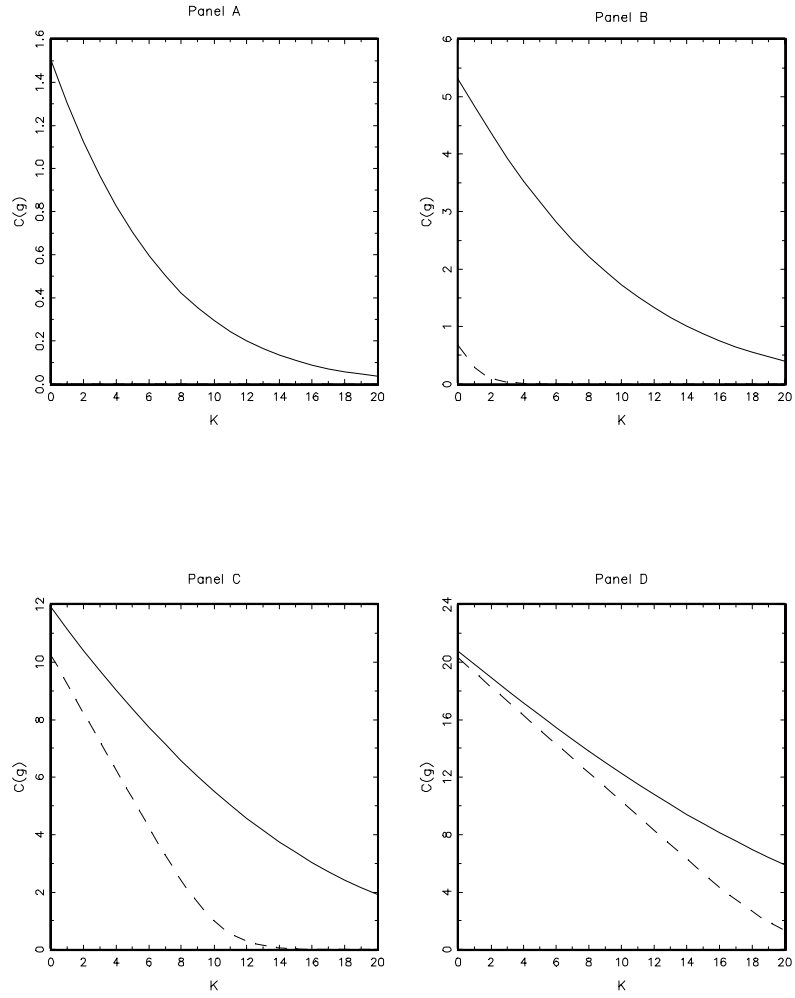


Figure 1.5: Upper and lower bounds for the spread option price $C(g)$ as a function of the strike price K , for marginally lognormal distributed assets. In Panel A the initial prices are $S_1^0 = 90$, $S_2^0 = 100$, in Panel B $S_1^0 = 100 = S_2^0$, in Panel C $S_1^0 = 110$, $S_2^0 = 100$, in Panel D $S_1^0 = 120$, $S_2^0 = 100$.

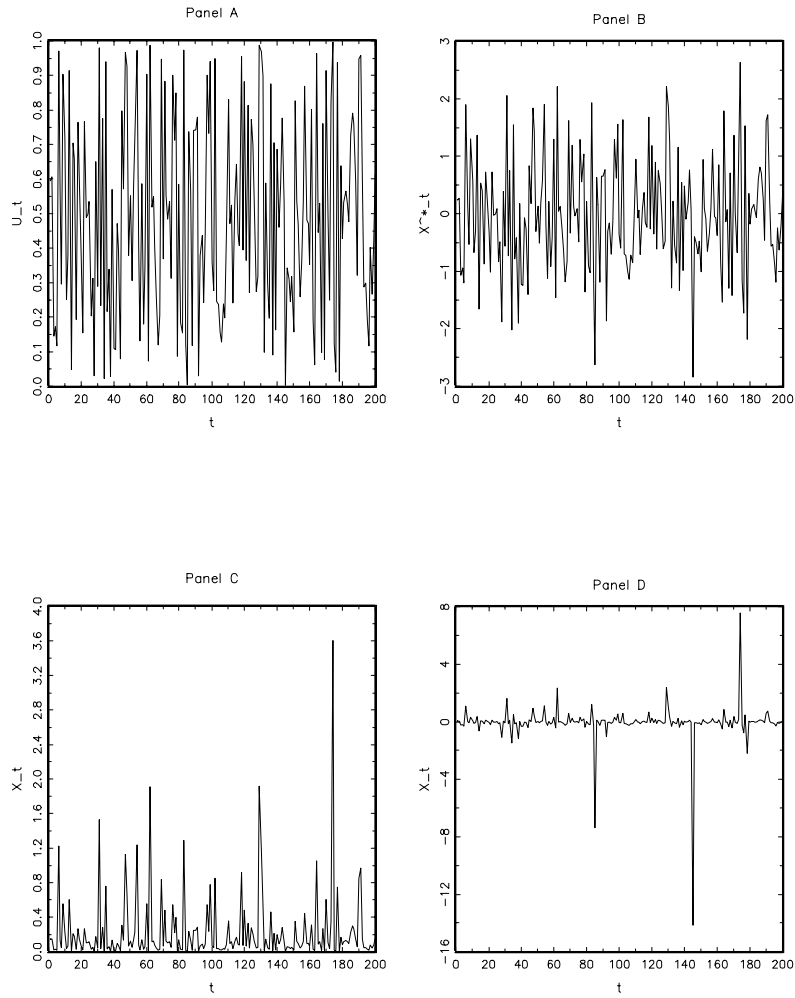


Figure 1.6: Simulated trajectories of Markov process with Gaussian copula and $\rho = 0$. Panel A corresponds to the standardized process U_t , Panel B to the gaussian process X_t^* , Panel C to a Pareto marginal process, and Panel D to a Cauchy marginal process.

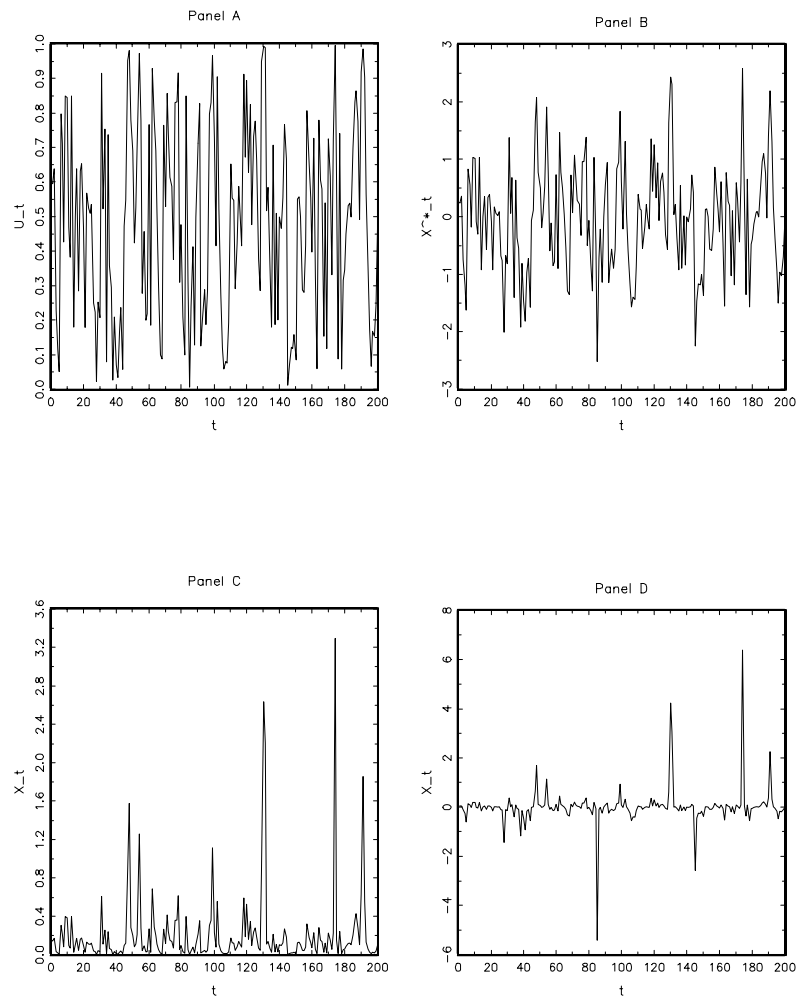


Figure 1.7: Simulated trajectories of Markov process with Gaussian copula and $\rho = 0.5$. Panel A corresponds to the standardized process U_t , Panel B to the gaussian process X_t^* , Panel C to a Pareto marginal process, and Panel D to a Cauchy marginal process.

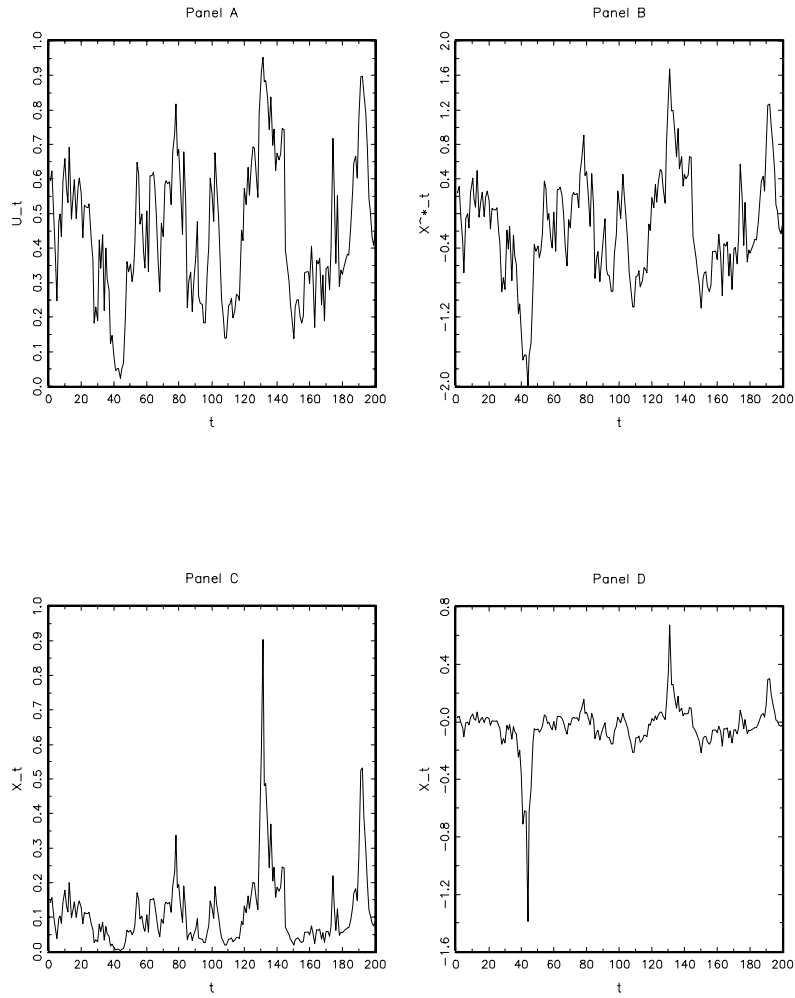


Figure 1.8: Simulated trajectories of Markov process with Gaussian copula and $\rho = 0.95$. Panel A corresponds to the standardized process U_t , Panel B to the gaussian process X_t^* , Panel C to a Pareto marginal process, and Panel D to a Cauchy marginal process.

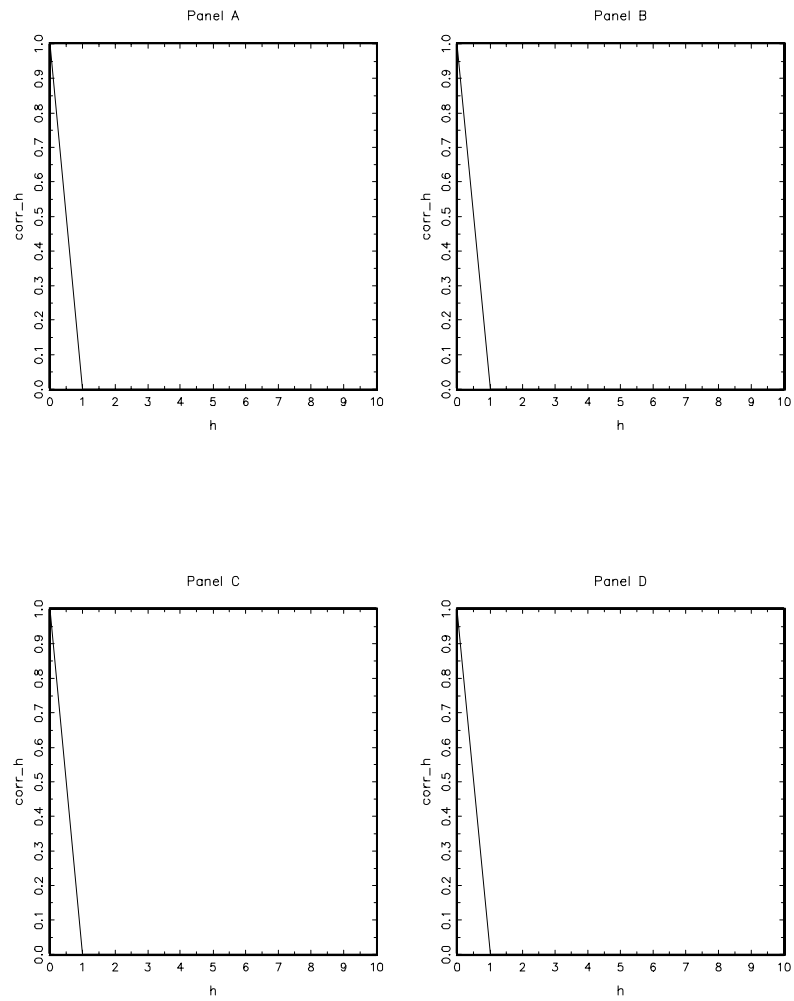


Figure 1.9: ACF of Markov process with Gaussian copula and $\rho = 0$. Panel A refers to the standardized process U_t , Panel B to the gaussian process X_t^* , Panel C to a Pareto marginal process, and Panel D to a Cauchy marginal process.

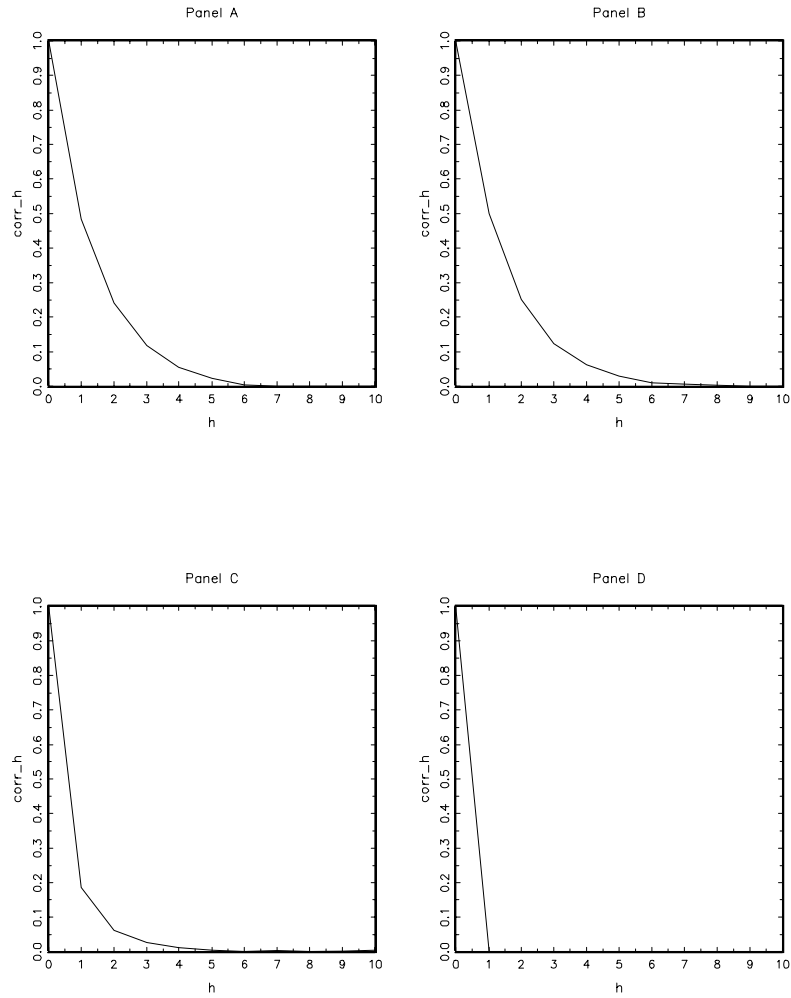


Figure 1.10: ACF of Markov process with Gaussian copula and $\rho = 0.5$. Panel A refers to the standardized process U_t , Panel B to the gaussian process X_t^* , Panel C to a Pareto marginal process, and Panel D to a Cauchy marginal process.

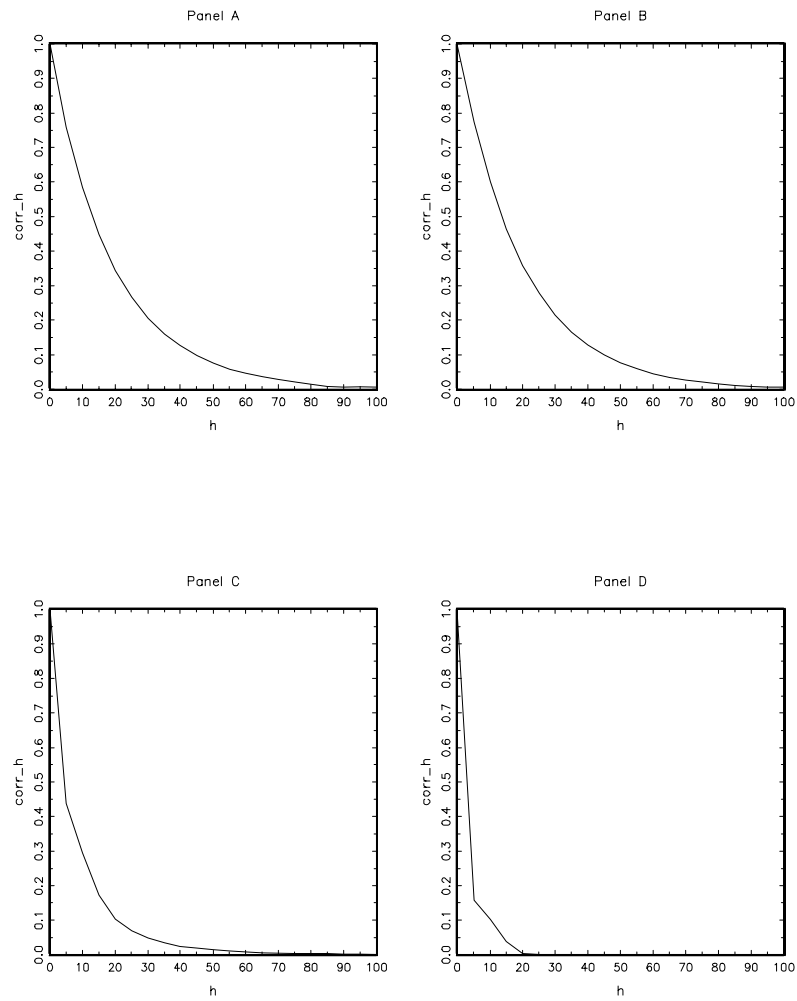


Figure 1.11: ACF of Markov process with Gaussian copula and $\rho = 0.95$. Panel A refers to the standardized process U_t , Panel B to the gaussian process X_t^* , Panel C to a Pareto marginal process, and Panel D to a Cauchy marginal process.

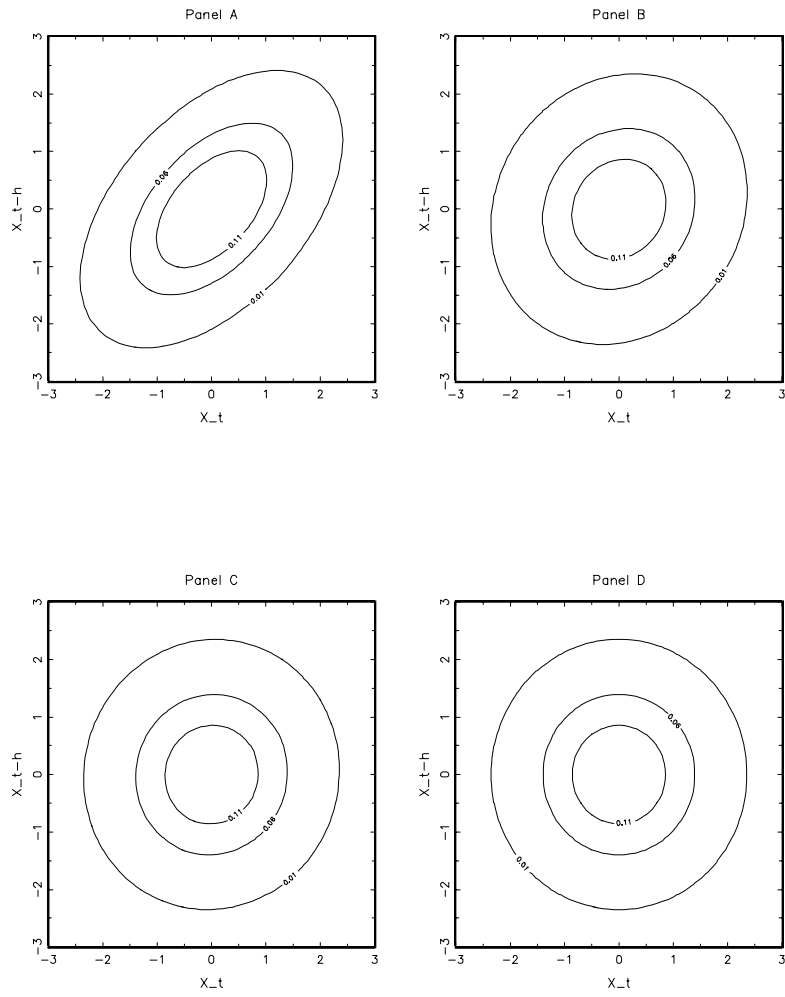


Figure 1.12: Isodensity curves for the distribution of (X_t^*, X_{t-h}^*) with $\rho = 0.5$. Panels A, B, C, and D correspond to horizon $h = 1$, $h = 3$, $h = 5$ and $h = 10$, respectively.

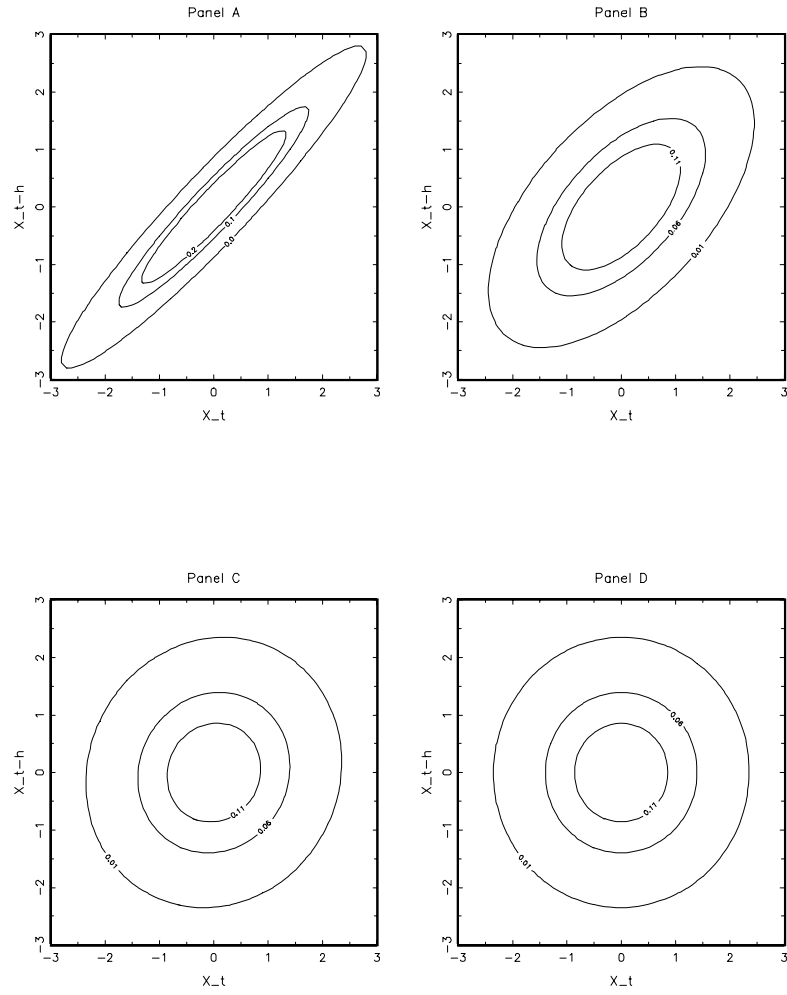


Figure 1.13: Isodensity curves for the distribution of (X_t^*, X_{t-h}^*) with $\rho = 0.95$. Panels A, B, C, and D correspond to horizon $h = 1$, $h = 10$, $h = 50$ and $h = 100$, respectively.

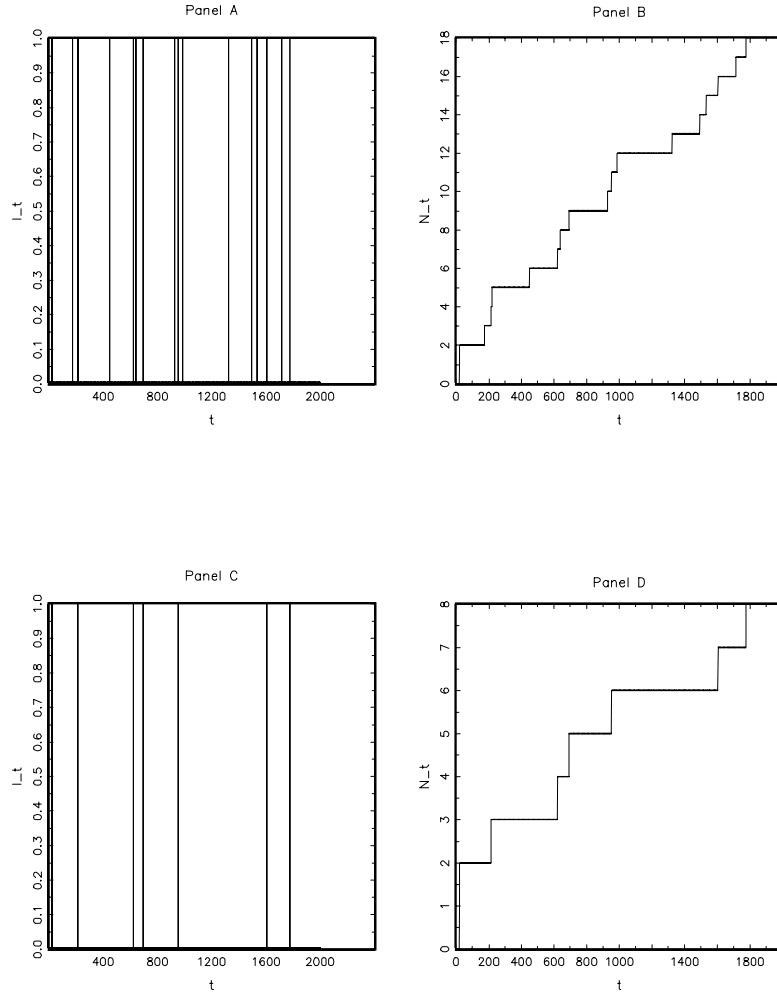


Figure 1.14: Extremes for Markov process with Gaussian copula with $\rho = 0.5$. Panels A and B correspond to the indicator variable $I_t = \mathbb{I}(U_t \geq 0.99)$ and the counting process $N_t = \sum_{s=1}^t \mathbb{I}(U_s \geq 0.99)$, respectively, for the observations above the 0.99-quantile in a simulation of length $T = 2000$. Panels C and D correspond to observations above the 0.995-quantile in the same simulated trajectory.

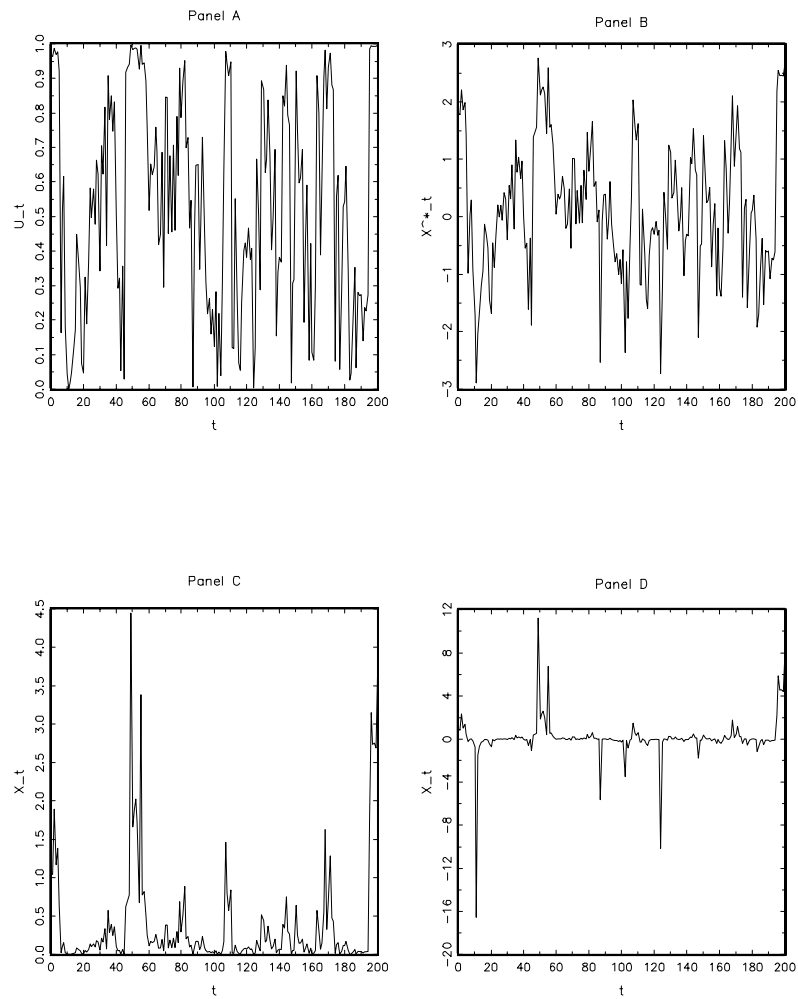


Figure 1.15: Simulated trajectories of Markov process with Gumbel copula and $\alpha = 1.5$. Panel A corresponds to the standardized process U_t , Panel B to the gaussian process X_t^* , Panel C to a Pareto marginal process, and Panel D to a Cauchy marginal process.

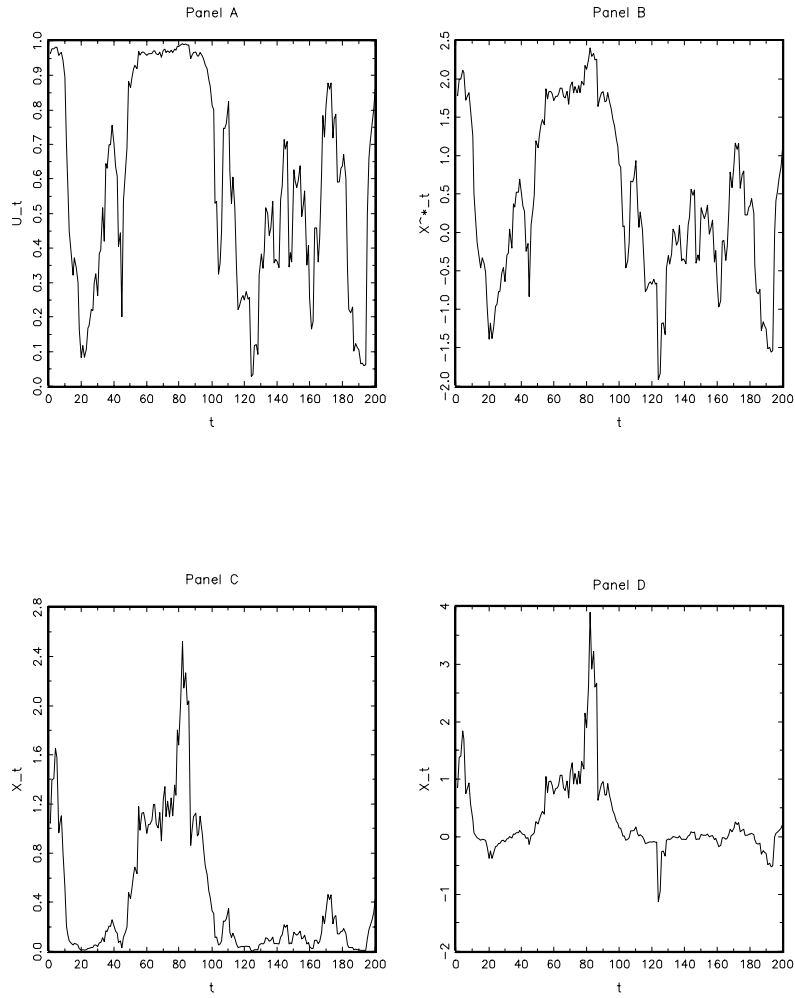


Figure 1.16: Simulated trajectories of Markov process with Gumbel copula and $\alpha = 4.946$. Panel A corresponds to the standardized process U_t , Panel B to the gaussian process X_t^* , Panel C to a Pareto marginal process, and Panel D to a Cauchy marginal process.

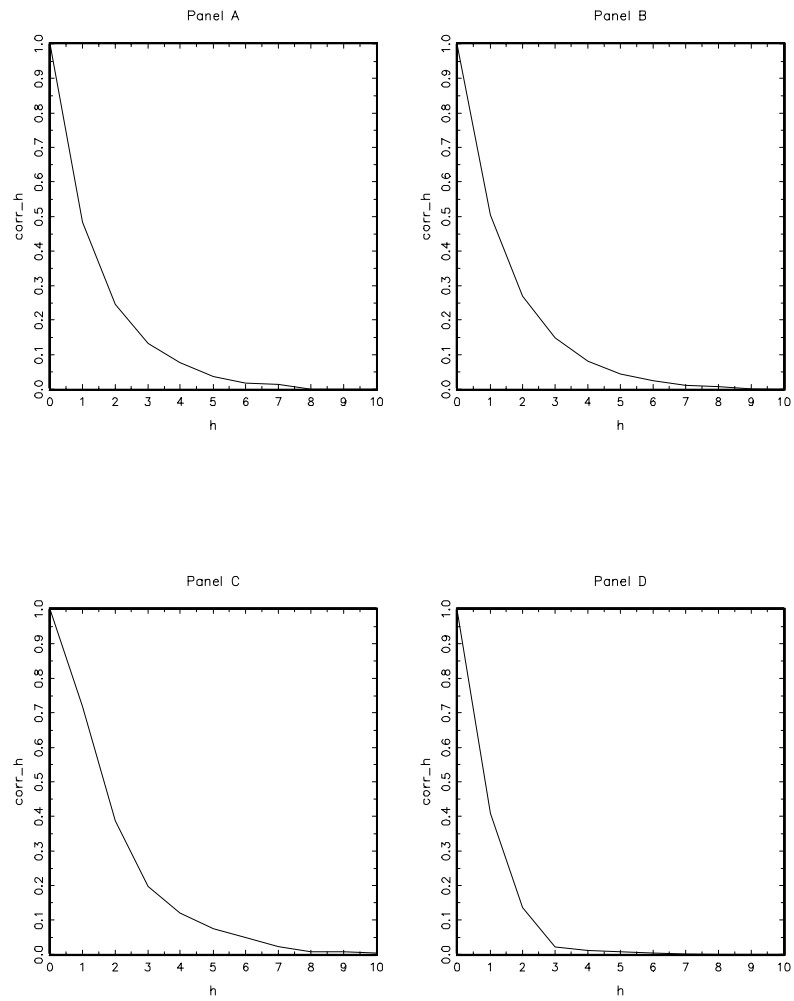


Figure 1.17: ACF of Markov process with Gumbel copula and $\alpha = 1.5$. Panel A refers to the standardized process U_t , Panel B to the process with standard gaussian marginal distribution, Panel C to a Pareto marginal process, and Panel D to a Cauchy marginal process.

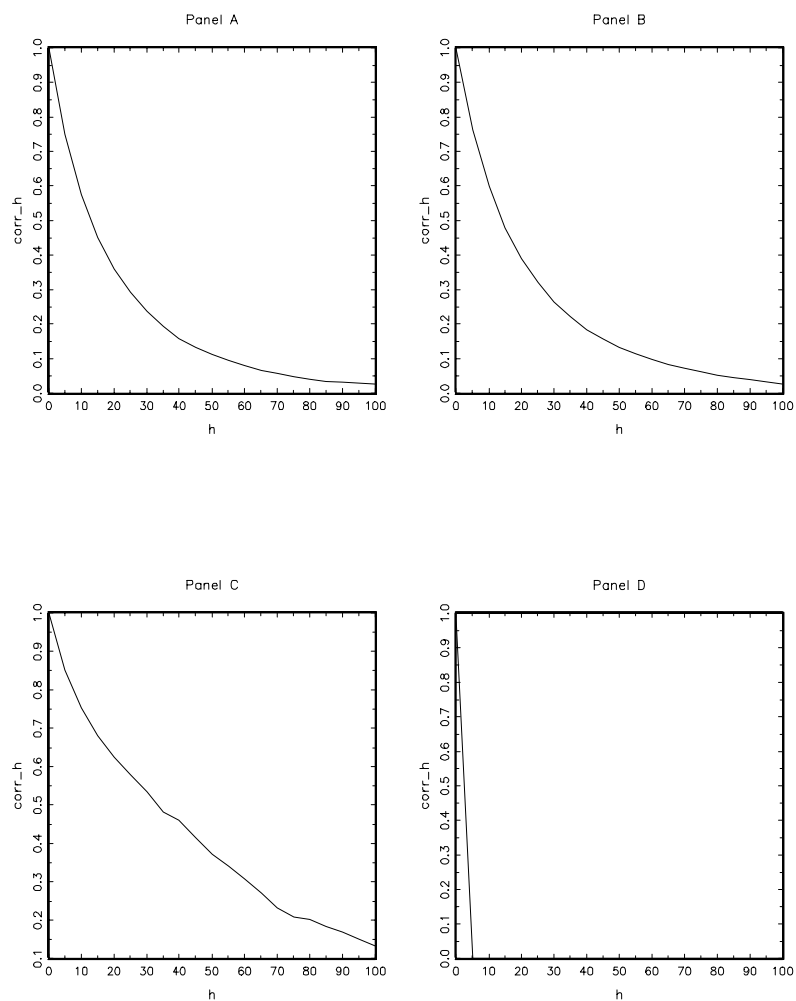


Figure 1.18: ACF of Markov process with Gumbel copula and $\alpha = 4.946$. Panel A refers to the standardized process U_t , Panel B to the process with standard gaussian marginal process, Panel C to a Pareto marginal process, and Panel D to a Cauchy marginal process.

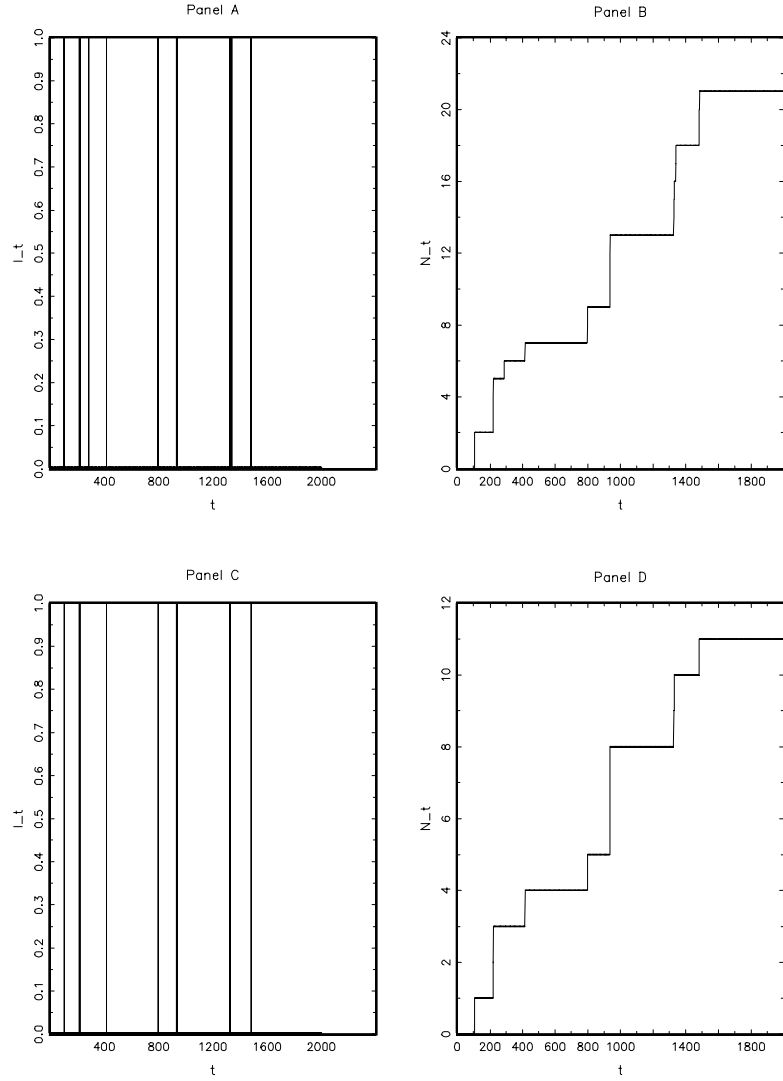


Figure 1.19: Extremes for Markov process with Gumbel copula and $\alpha = 1.5$. Panels A and B correspond to the indicator variable $I_t = \mathbb{I}(U_t \geq 0.99)$ and the counting process $N_t = \sum_{s=1}^t \mathbb{I}(U_s \geq 0.99)$, respectively, for the observations above the 0.99-quantile in a simulation of length $T = 2000$. Panels C and D correspond to observations above the 0.995-quantile with the same simulated trajectory.

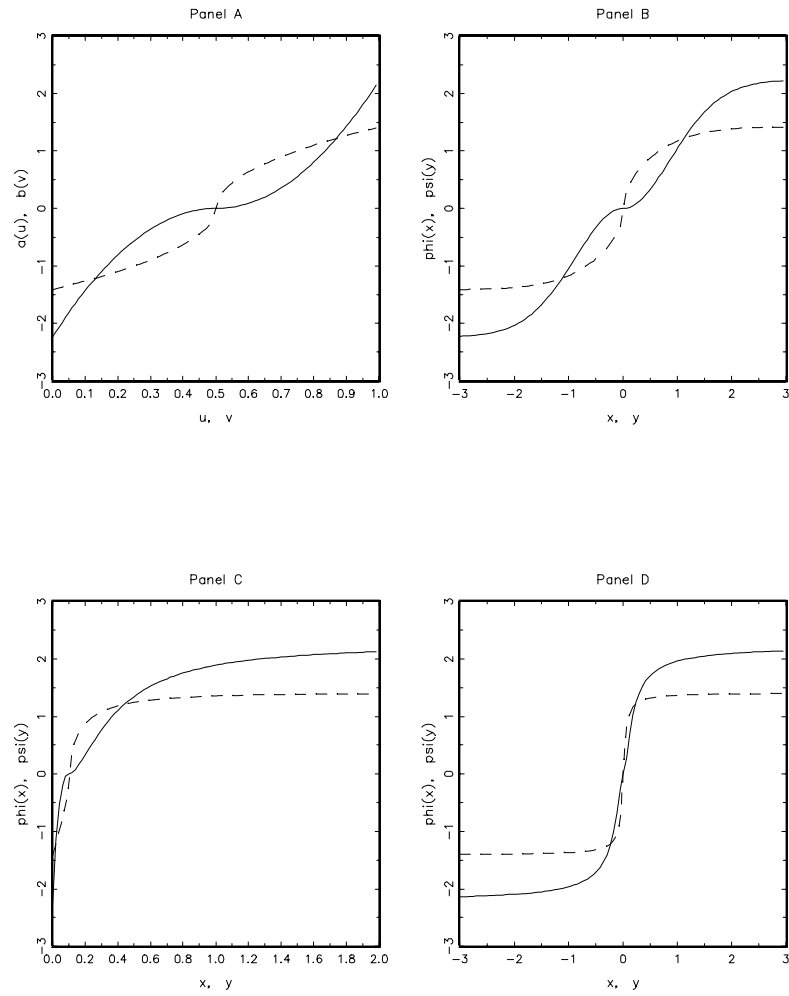


Figure 1.20: Current and lagged canonical directions of Markov process with one-dimensional dependence (solid and dashed line, respectively). Panel A corresponds to the standardized Markov process, Panel B to the Markov process with standard gaussian distribution, Panels C and D to the Markov process with Pareto and Cauchy marginal distributions, respectively.

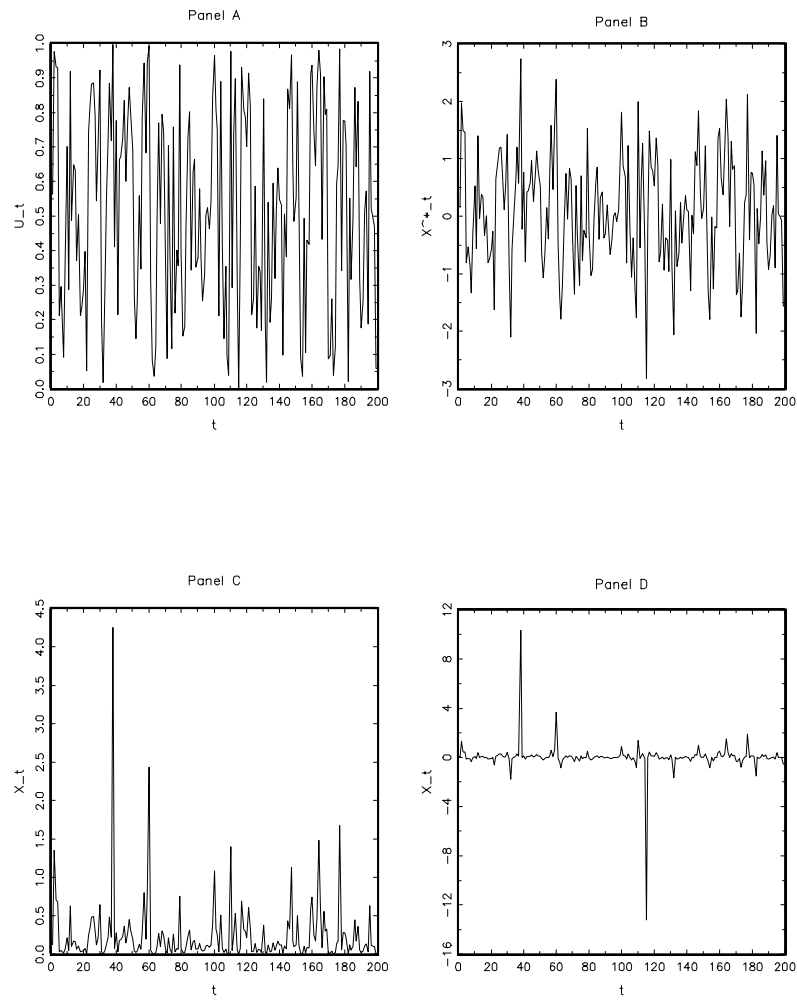


Figure 1.21: Simulated series of length $T = 200$ of Markov process with one-dimensional dependence. Panel A corresponds to the standardized Markov process, Panel B to the Markov with standard gaussian marginal distribution, Panels C and D to the Markov process with Pareto and Cauchy marginal distributions, respectively.

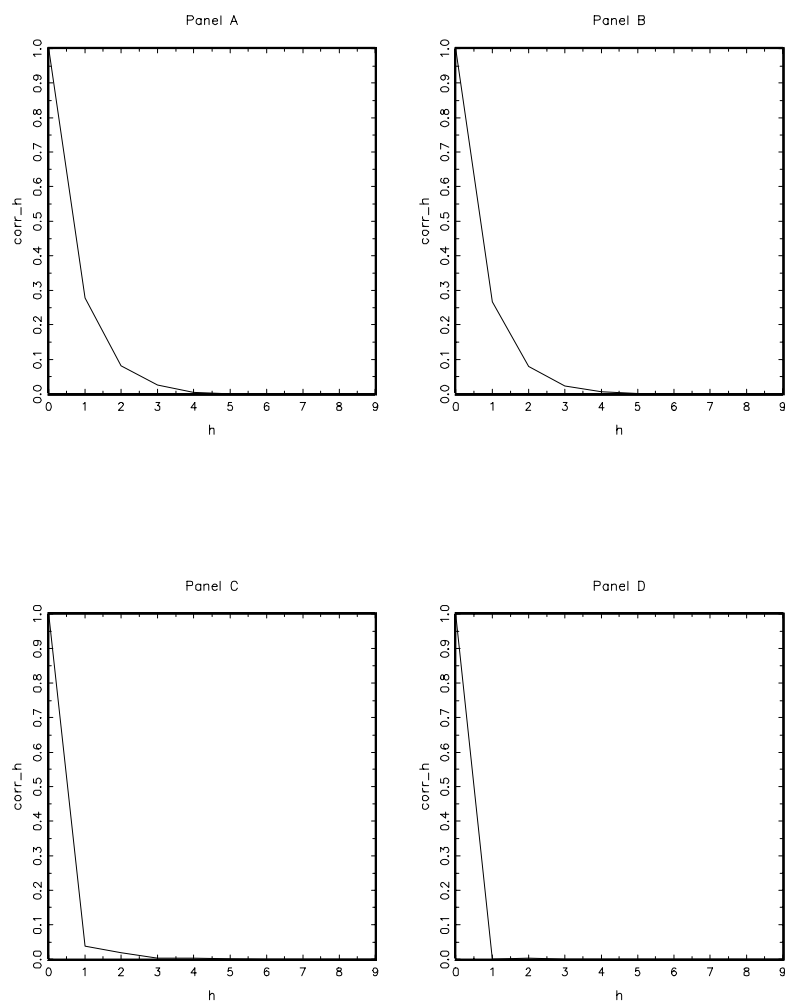


Figure 1.22: Autocorrelograms of Markov process with one-dimensional dependence. Panel A corresponds to the standardized Markov process, Panel B to the Markov process with standard gaussian marginal distribution, Panels C and D to the Markov process with Pareto and Cauchy marginal distributions, respectively.

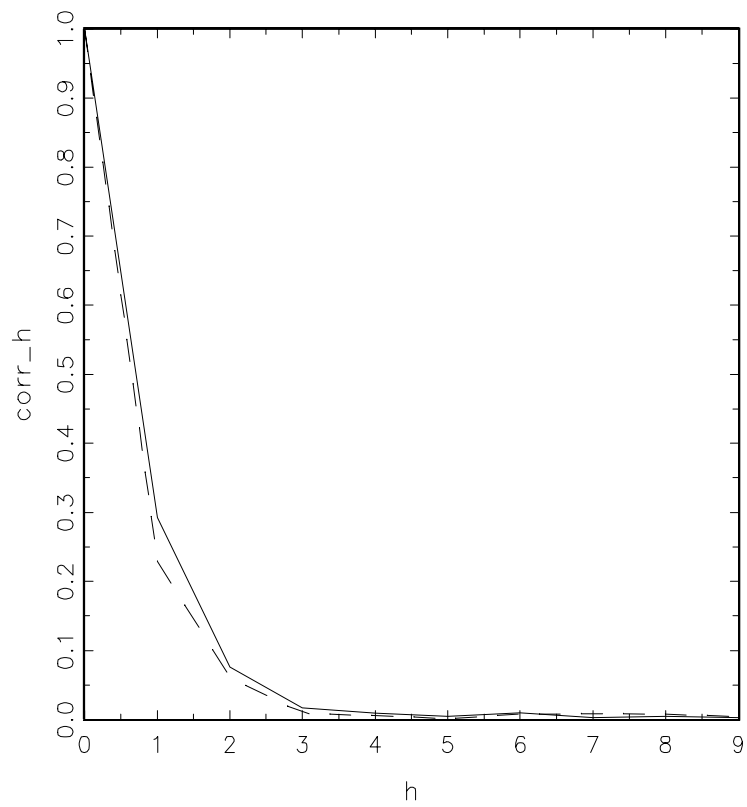


Figure 1.23: Cross-correlations between X_t^* and $\text{sign}(X_t^*)|X_t^*|^{1/10}$: in calendar time (solid line) and in reversed time (dashed line).

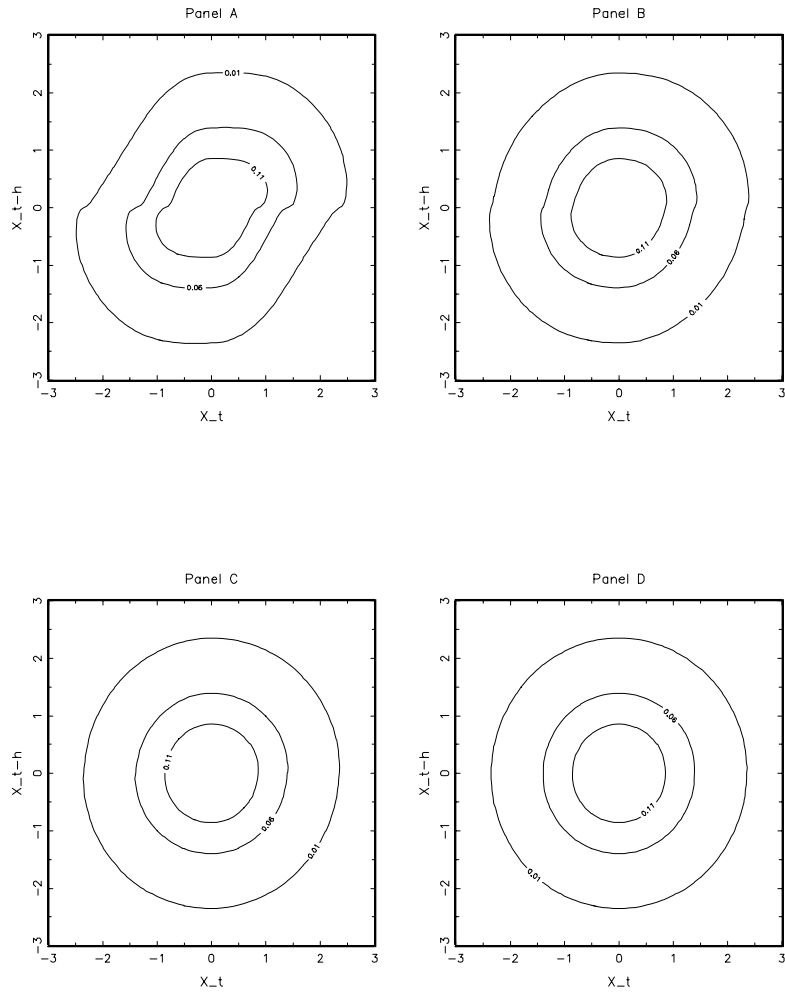


Figure 1.24: Isodensity curves of (X_t^*, X_{t-h}^*) for Markov process with one-dimensional dependence and standard gaussian marginal distribution, at different horizons. In Panel A $h = 1$, in Panel B $h = 2$, in Panel C $h = 3$, and in Panel D $h = 4$.

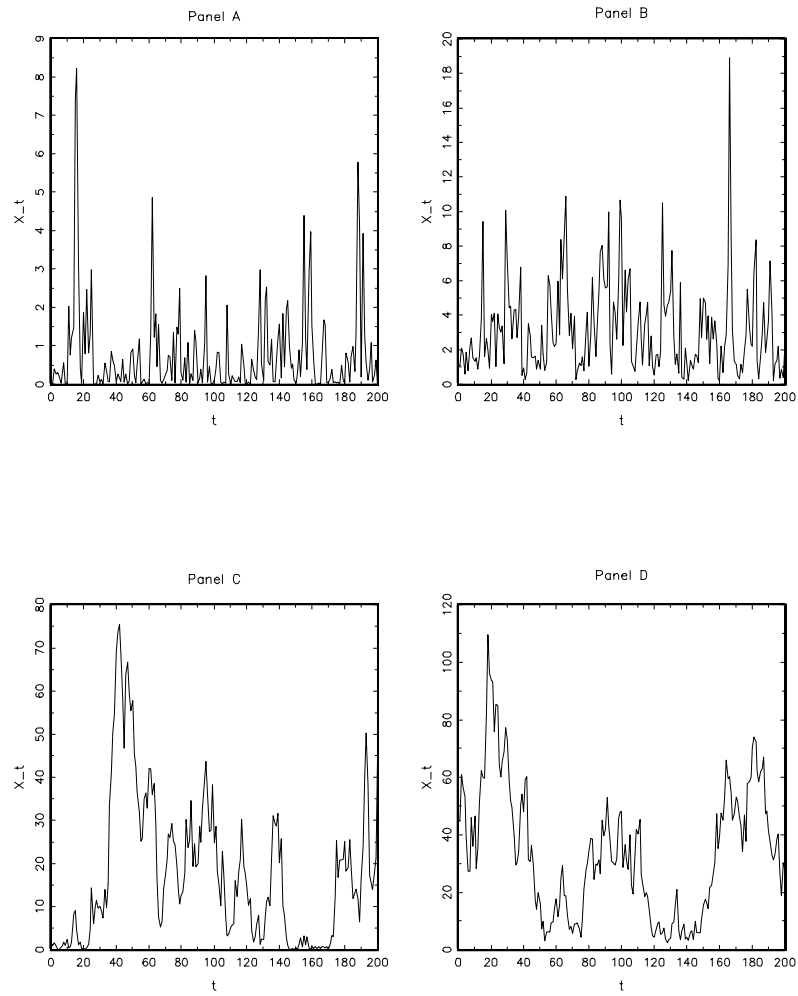


Figure 1.25: Simulated time series of length $T = 200$ of the ARG process. In Panel A $\delta = 0.5, \beta = 0.5$, in Panel B $\delta = 1.5, \beta = 0.5$, in Panel C $\delta = 0.5, \beta = 0.95$, and in Panel D $\delta = 1.5, \beta = 0.95$.

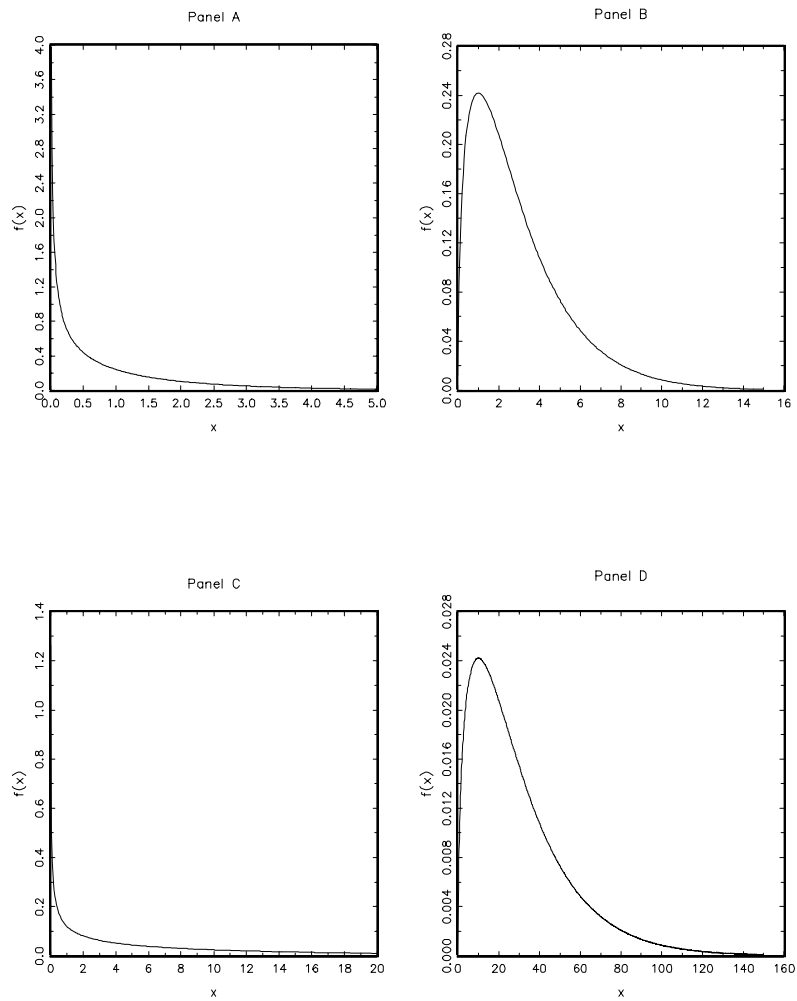


Figure 1.26: Marginal pdf of the ARG process. In Panel A $\delta = 0.5, \beta = 0.5$, in Panel B $\delta = 1.5, \beta = 0.5$, in Panel C $\delta = 0.5, \beta = 0.95$, and in Panel D $\delta = 1.5, \beta = 0.95$.

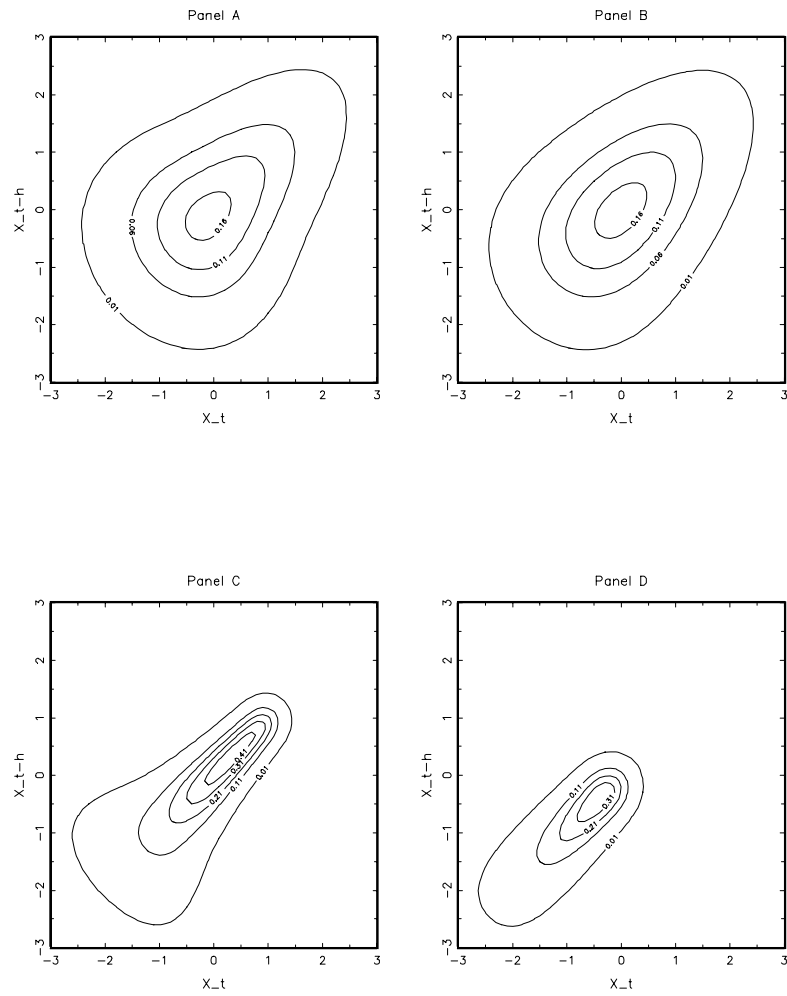
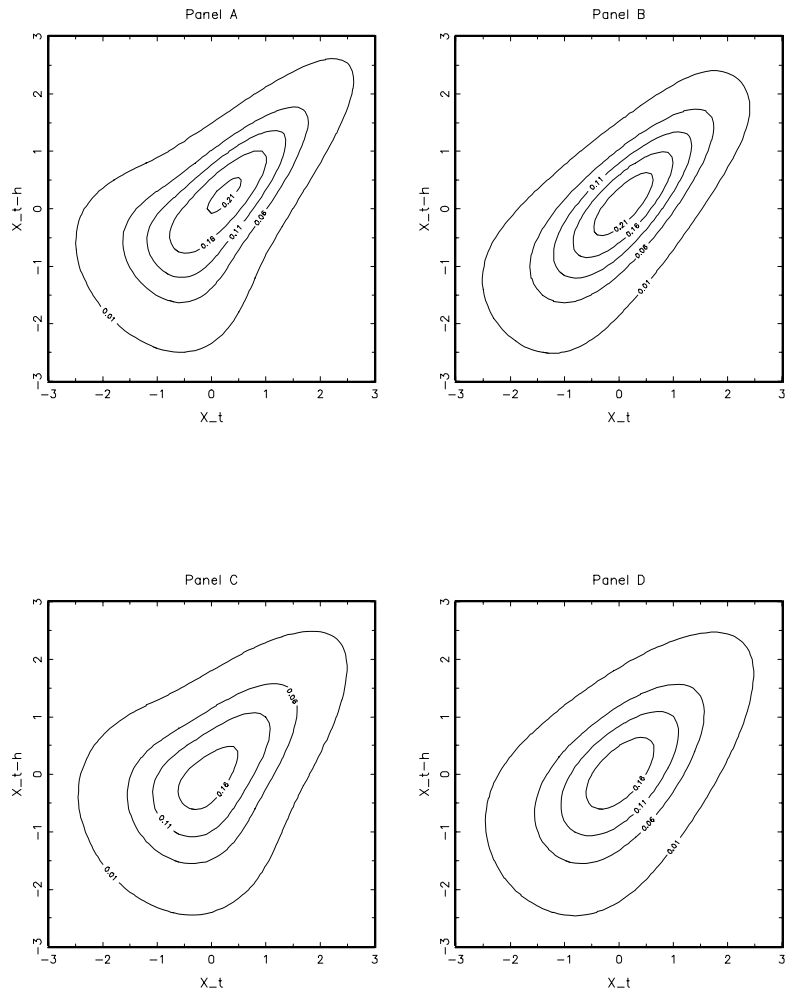


Figure 1.27: Isodensity curves at lag $h = 1$ for the ARG process with standard marginal distribution. In Panel A $\delta = 0.5$, $\beta = 0.5$, in Panel B $\delta = 1.5$, $\beta = 0.5$, in Panel C $\delta = 0.5$, $\beta = 0.95$, and in Panel D $\delta = 1.5$, $\beta = 0.95$.



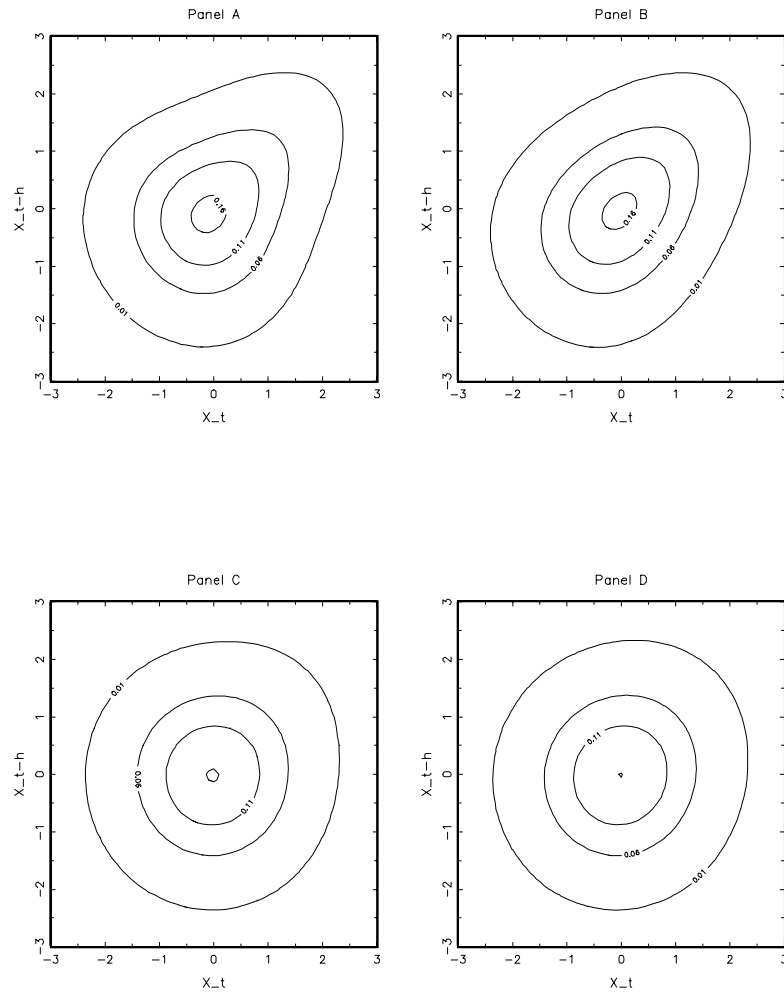


Figure 1.29: Isodensity curves at horizon h for the ARG process with $\beta = 0.95$ and standard gaussian marginal distribution. Horizon $h = 20$ in Panel A and B ($\delta = 0.5$ and $\delta = 1.5$, respectively), horizon $h = 50$ in Panels C and D ($\delta = 0.5$ and $\delta = 1.5$, respectively).

Chapter 2

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

2.1 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 Gouriéroux 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 deformation¹, 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 models². 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, Gouriéroux, 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 autoregression³. The ACD is able to replicate various stylized effects observed in the data. However, as pointed out in Ghysels, Gouriéroux, and Jasiak [1998]b, one limitation of this specification 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

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

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

³Various 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.

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, Gouriéroux, 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 duration⁴:

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

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) 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.

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

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.

2.2 Stationary Markov processes with proportional hazard.

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

2.2.1 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:

$$\lambda(x | \underline{X_{t-1}}) \equiv \lim_{h \rightarrow 0} \frac{P[X_t \leq x + h | X_t \geq x, X_{t-1}]}{h} = a(X_{t-1}) \lambda_0(x), \quad x \geq 0.$$

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:

$$P[X_t \geq x_t | X_{t-1} = x_{t-1}] = \exp[-a(x_{t-1})\Lambda_0(x_t)], \quad t \in \mathbb{N}, \quad (2.1)$$

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 distribution⁵.

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.

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

2.2.2 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:

$$X_t = \Lambda_0^{-1} \left(\frac{1}{a(X_{t-1})} \varepsilon_t \right), \quad t \in \mathbb{N}, \quad (2.2)$$

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

In the NLAR representation (2.2) 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:

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

Then we have:

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

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

Proposition 2.1 *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:*

$$Y_t = \varphi(Y_{t-1}) + \eta_t, \quad t \in \mathbb{N}, \quad (2.3)$$

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

The additive NLAR representation (2.3) 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 (2.1), since the functional parameters (a, Λ_0) and (h, φ) are in a one to one relationship:

$$h(x) = \log \Lambda_0(x), \quad x \in [0, \infty), \quad (2.4)$$

$$\varphi(y) = -\log a[\Lambda_0^{-1}(\exp y)], \quad y \in (-\infty, \infty). \quad (2.5)$$

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

2.2.3 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 $U_t, t \in \mathbb{N}$:

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

are constrained by the given form of the marginal distribution of U_t . Indeed we have:

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

or equivalently:

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

This condition identifies H_0 in terms of A :

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

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

Proposition 2.2 *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:*

$$X_t = F^{-1}(U_t), \quad t \in \mathbb{N}, \quad (2.6)$$

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:

$$P[U_t \geq u_t \mid U_{t-1} = u_{t-1}] = \exp(-A(u_{t-1})H_0(u_t, A)), \quad t \in \mathbb{N}, \quad (2.7)$$

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

$$H_0^{-1}(z, A) = 1 - \int_0^1 \exp(-A(v)z) dv, \quad z \in [0, \infty). \quad (2.8)$$

iii. The parameters (a, Λ_0) of process $X_t, t \in \mathbb{N}$, in (2.6) are obtained from the corresponding ones (A, H_0) of process $U_t, t \in \mathbb{N}$, by compounding with F :

$$a = A \circ F, \quad \Lambda_0 = H_0 \circ F. \quad (2.9)$$

Let $X_t, t \in \mathbb{N}$, be a stationary Markov process defined by (2.6), 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:

$$C_A(u, v) = v - \int_0^v \exp(-A(y)H_0(u, A)) dy, \quad u, v \in [0, 1], \quad (2.10)$$

where $H_0(., A)$ is defined by (2.8). 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 (2.8) and (2.9) 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 transformations⁷. This justifies the interpretation of A as a functional dependence parameter. It is identified up to a multiplicative constant. Indeed, from (2.8) and (2.10) 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 (2.4), (2.5), (2.8) and (2.9) we get:

$$\varphi(y) = -\log A \left[1 - \int_0^1 \exp(-A(v) \exp y) dv \right], \quad y \in (-\infty, \infty), \quad (2.11)$$

$$h(x) = \log H_0[F(x)], \quad x \in [0, \infty). \quad (2.12)$$

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 (2.11) to ensure stationarity.

2.2.4 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 decreasing⁸. Then copula C_A can also be characterized by $1 - A^{-1}$, which is the c.d.f. of the variable $A(U_{t-1})$,

⁷Equations (2.9) give in explicit form the restrictions on the parameters a and Λ_0 implied by the stationarity.

⁸This 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.

that is the transformation of the past transformed duration U_{t-1} having a proportional hazard effect on U_t ⁹. In addition, restriction (2.8) can be written as:

$$1 - H_0^{-1}(z) = \int_{\Omega} \exp(-wz) d(1 - A^{-1})(w), \quad z \in [0, \infty), \quad (2.13)$$

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

- i) *either the functional dependence parameter A itself, or*
- ii) *the c.d.f. $1 - A^{-1}$, with support $\Omega \subset \mathbb{R}_+$, or*
- iii) *its Laplace transform $1 - H_0^{-1}$, or*
- iv) *the baseline cumulated hazard H_0 , or*
- v) *the baseline survivor function $S_0 \equiv \exp(-H_0)$.*

2.2.5 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 $P(1/\delta, \cdot)$ (see Abramowitz, Stegun [1965]):

$$1 - A^{-1}(w) = P(1/\delta, w) = \frac{1}{\Gamma(1/\delta)} \int_0^w \exp(-u) u^{\frac{1}{\delta}-1} du, \quad w \in [0, +\infty), \quad (2.14)$$

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

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

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

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

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

and the baseline cumulated hazard is:

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

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.

2.3 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 series¹⁰. 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 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.

2.3.1 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 2.1 (*Lehmann [1966], Barlow and Proschan [1975]*): X_t is stochastically increasing (SI) in X_{t-1} iff

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

Definition 2.2 (*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 $\lambda(\cdot \mid y)$ denotes the conditional hazard rate of X_t given $X_{t-1} = y$.

Since $S(x \mid y) = \exp\left(-\int_0^x \lambda(x^* \mid y) dx^*\right)$, 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.

¹⁰Beyond traditional methods based on autocorrelograms, considerable attention has been devoted in recent years to nonlinear autocorrelograms (see e.g. Gouriéroux and Jasiak [2001]b), conditional Laplace transforms (see e.g. Darolles, Gouriéroux 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]).

¹¹A 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):

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

Proposition 2.4 *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.

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 (2.11), the autoregressive function φ is increasing iff the functional dependence parameter A is decreasing.

Corollary 2.5 *For a stationary Markov process with proportional hazard, the positive dependence (HI) or (SI) is satisfied iff the autoregressive function φ is increasing.*

2.3.2 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 (U_t^*, U_{t-1}^*) .

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

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

where $S(\cdot | v)$ 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 2.3 : 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 2.4 : 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 $\succeq_{(SI)}$, $\succeq_{(HI)}$, respectively¹². They satisfy various axioms, desirable for dependence orderings (see Kimeldorf and 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, $(X_t, X_{t-1}) \succeq_{(SI)} (X_t^*, X_{t-1}^*)$ 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 (X_t^*, X_{t-1}^*) . If in addition this is more and more true as we move towards the tail of the distribution, then $(X_t, X_{t-1}) \succeq_{(HI)} (X_t^*, X_{t-1}^*)$.

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 2.6 *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 $(X_t, X_{t-1}) \succeq_{(SI)} (X_t^*, X_{t-1}^*)$, and $(X_t, X_{t-1}) \succeq_{(HI)} (X_t^*, X_{t-1}^*)$ are equivalent. They are also equivalent to the condition*

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

Proof. See Appendix 1.

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

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

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.

¹²The 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 $(X_t, X_{t-1}) \succeq_{(SI)} (X_t^*, X_{t-1}^*)$ iff X_t is SI in X_{t-1} , and similarly for $\succeq_{(HI)}$.

¹³ $(X_t, X_{t-1}) \succeq_{(HI)} (X_t^*, X_{t-1}^*)$ or $(X_t, X_{t-1}) \succeq_{(SI)} (X_t^*, X_{t-1}^*)$ implies that the Kendall's tau of (X_t, X_{t-1}) is larger than that of (X_t^*, X_{t-1}^*) ; moreover, if $(X_t, t \in \mathbb{N})$ and $(X_t^*, t \in \mathbb{N})$ have the same margins, then:

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

for any monotonous transformation g such that the correlations exist.

Corollary 2.7 *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 $(X_t, X_{t-1}) \succeq_{(SI)} (X_t^*, X_{t-1}^*)$ and $(X_t, X_{t-1}) \succeq_{(HI)} (X_t^*, X_{t-1}^*)$ are equivalent to:*

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

or

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

As an illustration, the functions:

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

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

2.3.3 Measures of Dependence.

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

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

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

2.3.4 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]):

$$\lambda = \lim_{u \rightarrow 1} P[U_t \geq u | U_{t-1} \geq u].$$

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:

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

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

Proposition 2.8 *Assume the functional dependence parameter A is such that:*

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

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

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

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.

2.3.5 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:

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

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 $\text{corr}[g(X_t), g(X_{t-h-1})] \leq \text{corr}[g(X_t), g(X_{t-h})]$, $h \in \mathbb{N}$, for any monotonous transformation g such that these correlations exist.

Proposition 2.9 *Let $(X_t, t \in \mathbb{N})$ be a stationary Markov process with proportional hazard and dependence parameter A . If A is decreasing, then*

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

and

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

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.

2.4 Ergodicity Properties.

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

2.4.1 Geometric ergodicity.

Let us first recall the definition of geometric ergodicity.

Definition 2.5 *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:*

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

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 2.10 *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:*

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

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

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

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

Proof. See Appendix 4.

Let us briefly discuss the ergodicity conditions¹⁴. Condition i. restricts the absolute value of the autoregressive function (including the expectation of the innovation), $|\varphi(y) + \gamma|$, to

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

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

be strictly bounded by $|y|$, as $|y| \rightarrow +\infty$. 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 $|y| \rightarrow +\infty$, 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 $v \rightarrow 0$ and $v \rightarrow 1$, respectively. They are not immediately satisfied only if $\lim_{v \rightarrow 0} A(v)$ or $\lim_{v \rightarrow 1} A(v)$ 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 $\lim_{v \rightarrow 1} A(v) = 0$. For simplicity, let us consider functions A which are continuous on $(0, 1)$, decreasing near $v = 1$, and such that $\forall \delta > 0 : \lim_{v \rightarrow 1} \frac{A(v)}{(1-v)^\delta}$ exists (in $[0, +\infty]$). Any such function belongs to one of the following categories:

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

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

$$\text{III} \quad \forall \delta > 0 : \lim_{v \rightarrow 1} \frac{A(v)}{(1-v)^\delta} = 0.$$

A function A in class I converges to 0 as $(1-v)^\delta$, for some $\delta > 0$, when $v \rightarrow 1$, that is the elasticity δ of $A(1-v)$ with respect to v at $v = 1$ is strictly positive and finite¹⁷. Functions in class II (resp. III) dominate (resp. are dominated by) any function in class I, when $v \rightarrow 1$.

Proposition 2.11 *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.

¹⁵Note that:

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

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

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

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

2.4.2 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 2.6 *A Markov process $X_t, t \in \mathbb{N}$, is β -mixing with geometric decay if the mixing coefficients β_h , defined by:*

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

decay geometrically:

$$\beta_h \leq C\rho^h, \quad h \in \mathbb{N},$$

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 2.12 *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.*

2.5 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.

2.5.1 Constant measure of dependence.

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

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

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

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

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

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

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

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 $1 - A^{-1}$:

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

2.5.2 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:

$$A(v) = -\log(v), \quad v \in [0, 1]. \quad (2.15)$$

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),

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

and $\lim_{y \rightarrow 0} y A [H_0^{-1}(y)] = 0$, (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:

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

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

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:

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

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

[insert Figure 7: functional dependence parameter]

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

For $w \sim 0$, we have:

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

and thus:

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

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:

$$A(v) = (1 - v)^{\delta}, \quad v \in [0, 1]. \quad (2.16)$$

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

The Laplace transform is:

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

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

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

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:

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

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.

2.5.3 Endogenous switching regimes.

Let us consider a stepwise functional dependence parameter:

$$A(v) = \sum_{j=0}^J a_j \mathbb{I}_{(u_j, u_{j+1}]}(v), \quad v \in [0, 1], \quad (2.17)$$

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

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

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:

$$1 - u = \sum_{j=0}^J \exp[-a_j H_0(u)] (u_{j+1} - u_j), \quad \forall u \in [0, 1]. \quad (2.18)$$

When $a_j > 0$, for at least one $j \in \{0, \dots, J\}$, condition (2.18) characterizes the baseline cumulated hazard H_0 , whose inverse is given by:

$$\begin{aligned} H_0^{-1}(z) &= 1 - \sum_{j=0}^J \exp(-a_j z) (u_{j+1} - u_j) \\ &= 1 - \sum_{j=0}^J \pi_j \exp(-a_j z), \quad z \geq 0, \end{aligned} \quad (2.19)$$

where $\pi_j \equiv u_{j+1} - u_j$, $j = 0, \dots, J$. Equation (2.19) is a discrete analogue of equation (2.8), and it represents $1 - H_0^{-1}$ as the Laplace transform of a discrete distribution on \mathbb{R}_+ , weighting a_j , $j = 0, \dots, J$, with probabilities π_j , $j = 0, \dots, J$.

i) Uniform series.

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

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

Thus the function A is regularly decreasing and:

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

ii) Power Series.

When:

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

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

$$H_0(u) = -\log(1 - u^\theta), \quad u \in [0, 1].$$

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

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

with

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

This defines an increasing step function (2.17), with thresholds at:

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

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

iii) Logarithmic Series.

When:

$$1 - H_0^{-1}(z) = -\frac{1}{\theta} \log [1 - (1 - e^{-\theta}) \exp(-z)], \quad z \geq 0, \quad (2.20)$$

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

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

and:

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

respectively. The corresponding discrete distribution is found by expanding the logarithmic series in (2.20) to get (see Joe [1997], Appendix A.1):

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

with

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

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

2.5.4 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:

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

$u, v \in [0, 1]$, 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:

$$A(v) = \frac{\Psi^{-1}(\gamma v + \delta)}{\Psi^{-1}(\delta)}, \quad v \in [0, 1], \quad (2.21)$$

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

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

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 (2.21) 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.

2.6 Statistical Inference

In this section we assume available observations X_1, \dots, 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 \hat{F} , say and $\hat{U}_t = \hat{F}(X_t)$, $t = 1, \dots, T$ provide approximations of the uniform variables U_t . \hat{U}_t , $t = 1, \dots, T$, are simply the ranks of the variables X_t , $t = 1, \dots, T$. In a second step we can look for an estimator of the dependence functional A from the observed \hat{U}_t and study the asymptotic properties of the estimator as if $U_t = \hat{U}_t$, $t = 1, \dots, T$, 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 \hat{U}_t , at least when the functional dependence parameter is let unconstrained¹⁸ [see Genest, Werker (2001), and Gagliardini, Gouriéroux (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 , $t = 1, \dots, T$, 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, Gouriéroux (2002).

2.6.1 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:

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

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.

¹⁸More precisely, the replacement of U_t by \hat{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 .

Proposition 2.13 :

i. The score is given by:

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

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

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:

$$\frac{\partial}{\partial \theta} \log H_0(U_t; \theta) = -E \left[\frac{\partial}{\partial \theta} \log A(U_{t-1}; \theta) \mid U_t \right]. \quad (2.22)$$

ii) Stepwise functional parameter.

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

$$A(v; \theta) = \sum_{j=1}^N a_j \mathbb{I}_{(\frac{j-1}{N}, \frac{j}{N}]}(v), \quad (2.23)$$

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

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

where $\text{diag}(a)$ is a diagonal matrix with the elements a_1, a_2, \dots, a_N on the diagonal. In addition, from equation (2.22), we deduce that:

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

that is:

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

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:

$$\frac{1}{N} \sum_{j=1}^N a_j = 1. \quad (2.24)$$

Then the information matrix is given by:

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

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

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

Thus, under regularity conditions, the maximum likelihood estimator $\hat{\theta}_T = (\hat{a}_{1T}, \hat{a}_{2T}, \dots, \hat{a}_{NT})'$ under identification constraint (2.24) is asymptotically normal when T converges to infinity, with asymptotic variance-covariance matrix such that²⁰:

$$Cov_{as} \left[\sqrt{T}(\hat{a}_{k,T} - a_{k,0}), \sqrt{T}(\hat{a}_{j,T} - a_{j,0}) \right] = N [a_{j,0}^2 \delta_{k,j} + O_N(1/N)]. \quad (2.25)$$

¹⁹It 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).

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

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

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

2.6.2 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.

2.6.2.1 Minimum chi-square method.

i) Definition of the estimator.

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

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

where K is a kernel, $K_{h_T}(\cdot) = (1/h_T) K(\cdot/h_T)$, 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:

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

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

$$V_{as} \left[\sqrt{Th_T} \int g(u, v) \hat{c}_T(u, v) du \right] = E_0 [g(U_t, U_{t-1})^2 | U_{t-1} = v] \int K^2(w) dw, \quad (2.26)$$

$$V_{as} \left[\sqrt{T} \int \int g(u, v) \hat{c}_T(u, v) dudv \right] = \sum_{h=-\infty}^{\infty} Cov [g(U_t, U_{t-1}), g(U_{t-h}, U_{t-h-1})]. \quad (2.27)$$

The minimum chi-square estimator is defined by:

$$\hat{A}_T = \min_A \int \int \frac{[\hat{c}_T(u, v) - c(u, v; A)]^2}{\hat{c}_T(u, v)} dudv, \quad (2.28)$$

where the optimization is performed under the identifying constraint:

$$\int A(v) dv = 1. \quad (2.29)$$

ii) Asymptotic properties of the estimator

The asymptotic properties of the minimum chi-square estimator \hat{A}_T defined in (2.28) and (2.29) are reported in Proposition 14 below. In order to formulate this proposition we need some preliminary concepts [see Gagliardini, Gouriou (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 $\log c(\cdot, \cdot; A)$ with respect to A in direction h is given by (see Appendix 7):

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

where:

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

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

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

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

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

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 2.14 *Under standard regularity conditions:*

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

ii. *We have the following asymptotic equivalence:*

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

where $\delta \hat{A}_T = \hat{A}_T - A_0$, $\delta \hat{c}_T = \hat{c}_T - c$, 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 \hat{A}_T is pointwise asymptotically normal:*

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

iv. *Continuous linear functionals of \hat{A}_T are asymptotically normal:*

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

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, Gouriéroux [2002]):

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

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:

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

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

Corollary 2.15 *Under regularity conditions:*

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

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.

2.6.2.2 Sieve method.

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

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

which is the analog of (2.29). For any given N , we get maximum likelihood estimators $\widehat{a}_{j,N}$, $j = 1, \dots, N$, 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 implement²².

2.7 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,

²¹The 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 .

²²Rewrite equation (2.25) with $N = N_T$ as:

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

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

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.

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., Gouriéroux, 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 and Its Applications*, Second Edition, Vol. 2., Wiley.

Gagliardini, P., and C., Gouriéroux, (2002): "Efficient Nonparametric Estimation of Models with Nonlinear Dependence", 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., Gouriéroux, 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., Gouriéroux, 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.

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

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

Gouriéroux, 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.

Appendix 1 Dependence Ordering

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

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

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

and:

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

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

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

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:

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

Using $A(v) \simeq (1 - v)^\delta$, $v \rightarrow 1$, it follows:

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

Thus:

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

and

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

It follows:

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

Appendix 3

Nonlinear 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 $E[\eta_t] = 0$.

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

$$f(x | y) = g(x - \varphi(y)), \quad x, y \in \mathbb{R},$$

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:

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

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 :*

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

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

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

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

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

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:

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

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

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

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

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

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

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

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

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

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

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

and the result follows.

Q.E.D.

Finally, let us consider mixing properties. Using the results of Davydov (1973), it is seen that geometric ergodicity²³ 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.*

²³with integrable function C

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 (2.3), where innovations are Gompertz distributed, with density:

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

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:

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

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

Conditions i. and ii. are equivalent

By using relation (2.11), condition i. can be written as:

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

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

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

Condition (a.2) becomes:

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

for some $r_2 < \infty$, that is:

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

which is equivalent to:

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

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

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

Condition (a.2) becomes:

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

for some $r_2 < \infty$, that is:

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

which is equivalent to:

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

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

Case III: $\lim_{v \rightarrow 1} A(v) = \exp[\gamma]$

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 $y \rightarrow -\infty$ 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 $\lim_{v \rightarrow 1} A(v) = 0$. Then for any $\varepsilon > 0$ small enough:*

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

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:

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

as $y \rightarrow +\infty$. *Q.E.D.*

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

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

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

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

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

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

we get:

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

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

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

then:

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

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:

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

Therefore:

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

In particular, we deduce from (2.11) that the autoregressive function φ corresponding to A is such that:

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

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:

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

Proof. *Define*

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

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:

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

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

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

Q.E.D.

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

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

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

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

²⁴We use that if $\psi(x) = \log E[\exp(-xZ)]$, then $\psi''(x) = V_{Q_x}[Z]$, where distribution Q_x is defined by $dQ_x(z) = \{\exp(-xz)/E[\exp(-xZ)]\} dF_Z(z)$.

Since A is decreasing near 1,

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

Then:

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

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

Appendix 6

Proportional hazard in reversed time.

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

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

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:

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

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:

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

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

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

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

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

In particular, function A is such that:

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

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

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

²⁵These 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:

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

and

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

Therefore function A is such that:

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

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

$$A^* = A.$$

We now use restriction (2.8) of uniform margins. The function H_0 and its inverse are given by:

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

and:

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

Thus the restriction is:

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

After the change of variable:

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

the integral in the RHS becomes:

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

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:

$$\begin{aligned} \log c(u, v; A + \delta A) &= \log [A(v) + \delta A(v)] + \log h_0(u, A + \delta A) \\ &\quad - [A(v) + \delta A(v)] H_0(u, A + \delta A) \\ &\simeq \log c(u, v; A) + \frac{\delta A(v)}{A(v)} + \langle D \log h_0(u, A), \delta A \rangle \\ &\quad - H_0(u, A) \delta A(v) - A(v) \langle DH_0(u, A), \delta A \rangle \\ &= \log c(u, v; A) + \frac{1 - A(v) H_0(u, A)}{A(v)} \delta A(v) \\ &\quad + \langle D \log h_0(u, A), \delta A \rangle - A(v) \langle DH_0(u, A), \delta A \rangle, \end{aligned} \tag{a.9}$$

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 .

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

We have:

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

hence:

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

Expression of $DH_0(u; A)$

By applying the implicit function theorem we get:

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

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

We get:

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

Let us introduce the functional:

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

and derive its first order expansion. We get:

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

It follows:

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

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:

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

where:

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

and:

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

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:

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

or equivalently:

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

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

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

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

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

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:

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

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:

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

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 $D \log c(\cdot, \cdot; A_0)$ admits a measure decomposition with both a discrete and a continuous part [see Gagliardini and Gouriéroux (2002)]. Therefore:

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

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

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

and:

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

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

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

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:

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

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 $\alpha_H(\cdot; A)$ 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 $D \log c(\cdot, \cdot; A_0)$ has a zero null space. Indeed:

$$\langle D \log c(U_t, U_{t-1}; A_0), h \rangle = 0 \text{ a.s., } h \in H,$$

implies that:

$$(1 - A_{0t-1}H_{0t})(h_{t-1}/A_{0t-1} - E[h_{t-1}/A_{0t-1} | U_t]) \text{ is a function of } U_t, h \in H,$$

that is:

$$h_{t-1}/A_{0t-1} \text{ is a constant, and } \int h(v) dv = 0,$$

which can only be the case if $h = 0$. Thus I_H has zero null space and it is positive.

Let us assume that ν is such that:

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

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 $\psi_T \in L^2(\nu)$ is defined by:

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

From Gagliardini, Gourieroux (2002) we get:

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

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:

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

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 \hat{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 (2.27)], and similar orders are expected for the second and third terms in the LHS, leading to:

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

From (2.26) it follows:

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

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

Appendix 10

The efficiency bounds for the stepwise model

i) Determination of the parametric efficiency bound

We have:

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

where $\xi_t = 1 - A_{0t-1}H_{0,t} \sim iid(0, 1)$, ξ_t independent of U_{t-1} [see equation (2.2)]. Let us transform the terms in the conditional variance. We have:

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

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:

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

and the fact that:

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

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:

$$B(\theta_0) = N [\text{diag}(a_0)^2 + O(1/N)].$$

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

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

ii) Pointwise asymptotic distribution.

A pointwise estimator of A can be defined by:

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

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

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

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

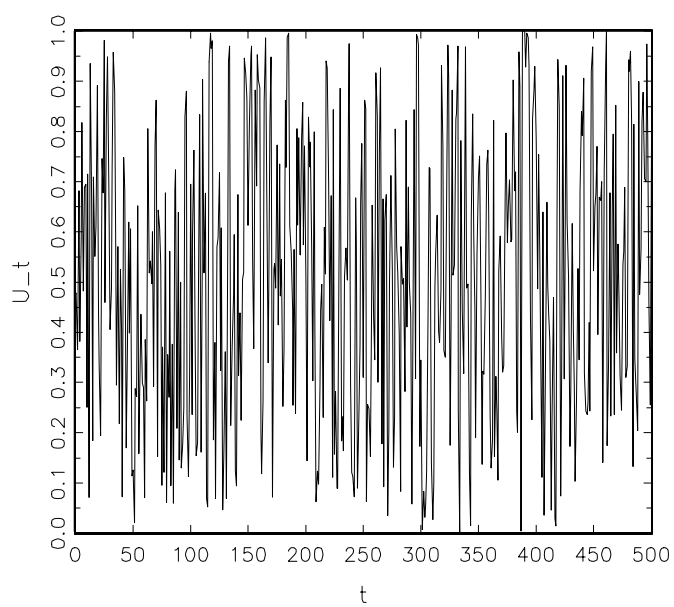


Figure 2.1: Simulated path for process U_t , $t \in \mathbb{N}$, with proportional hazard and functional dependence parameter A such that $1 - A^{-1}$ is a gamma distribution with parameter $\delta = 0.1$.

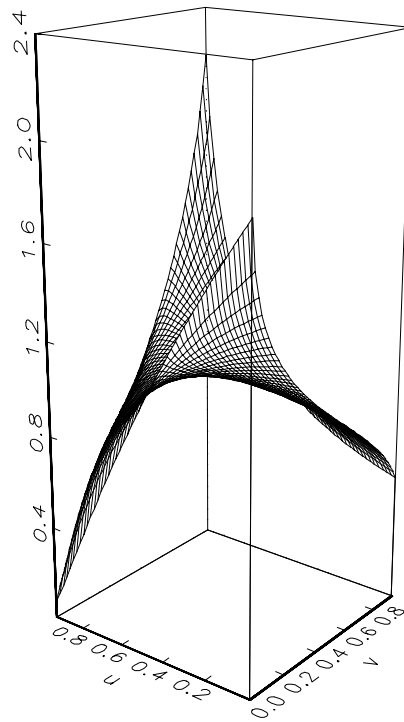


Figure 2.2: Copula p.d.f. for process U_t , $t \in \mathbb{N}$, with proportional hazard and functional dependence parameter A such that $1 - A^{-1}$ is a gamma distribution with parameter $\delta = 0.1$.

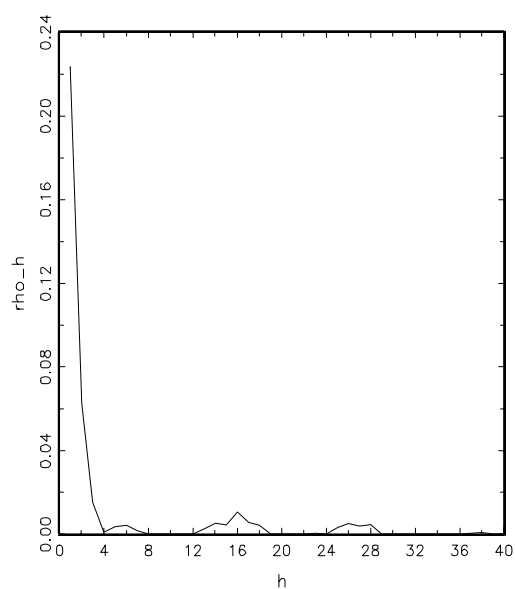


Figure 2.3: Autocorrelogram for process $X_t, t \in \mathbb{N}$, with functional dependence parameter A such that $1 - A^{-1} \sim \gamma(\delta)$, $\delta = 0.1$, and marginal distribution $F(x) = 1 - (1 + x)^\tau$, $\tau = 1.05$.

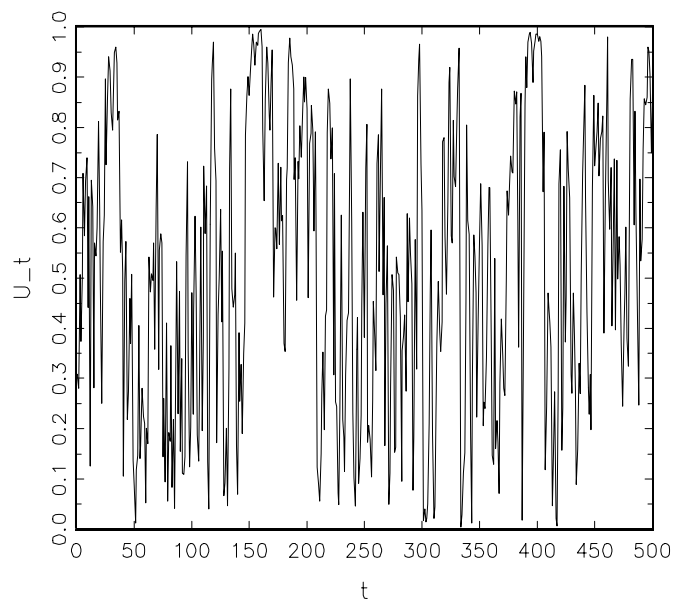


Figure 2.4: Simulated path for process U_t , $t \in \mathbb{N}$, with proportional hazard and functional dependence parameter A such that $1 - A^{-1}$ is a gamma distribution with parameter $\delta = 1$.

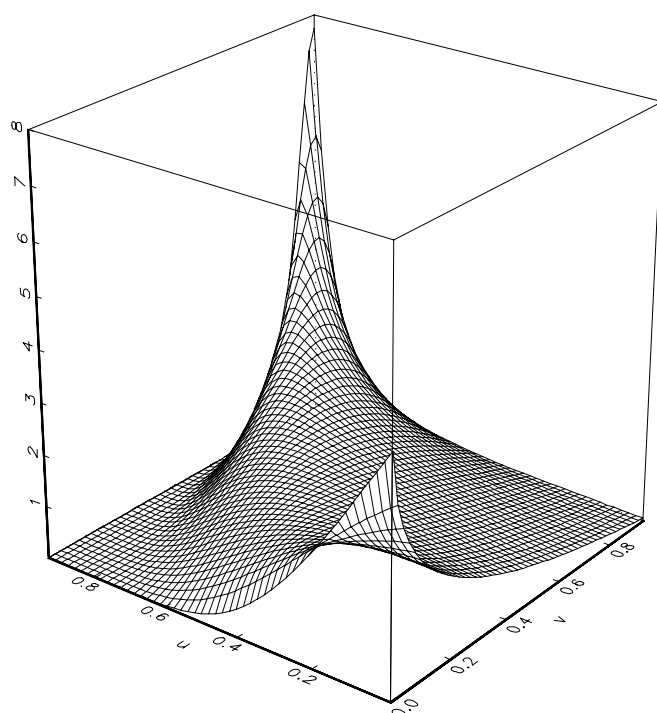


Figure 2.5: Copula p.d.f. for process U_t , $t \in \mathbb{N}$, with proportional hazard and functional dependence parameter A such that $1 - A^{-1}$ is a gamma distribution with parameter $\delta = 1$.

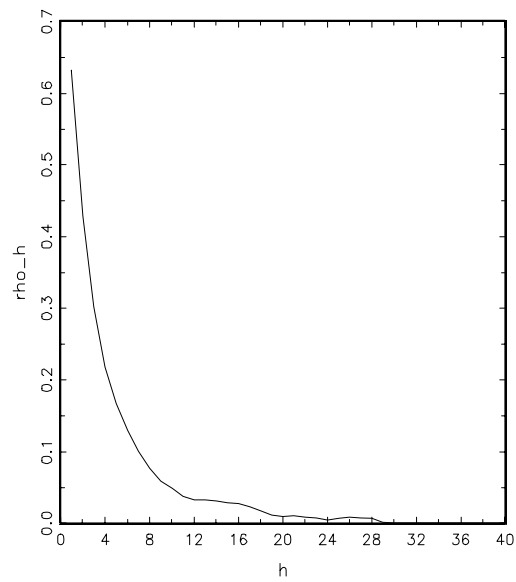


Figure 2.6: Autocorrelogram for process X_t , $t \in \mathbb{N}$, with functional dependence parameter A such that $1 - A^{-1} \sim \gamma(\delta)$, $\delta = 1$, and marginal distribution $F(x) = 1 - (1 + x)^\tau$, $\tau = 1.05$.

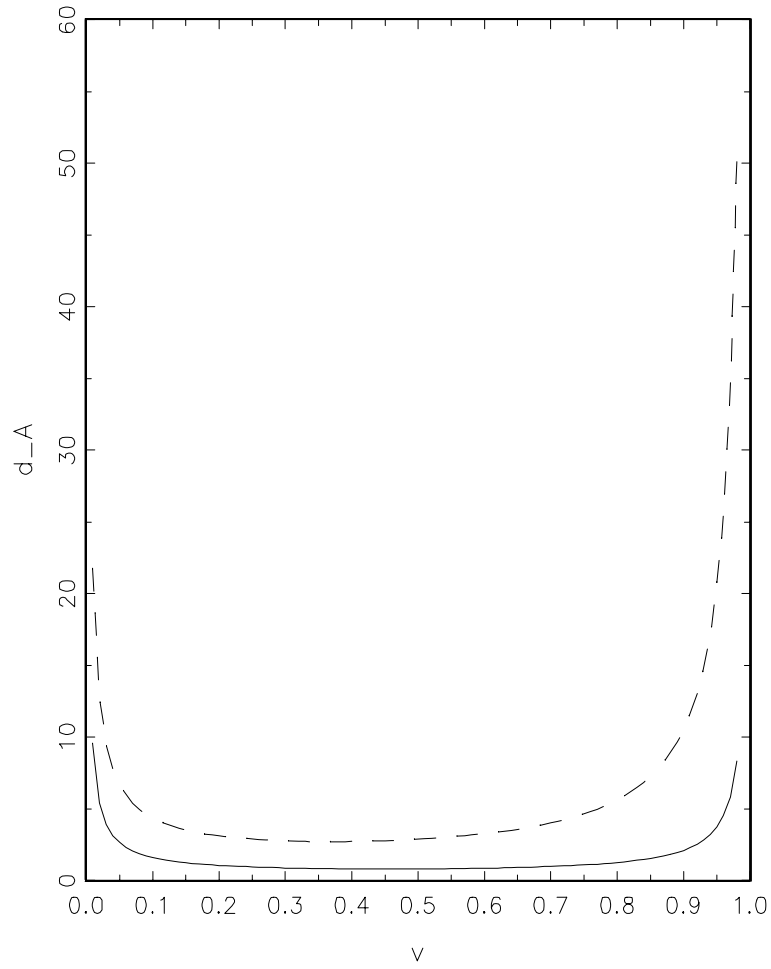


Figure 2.7: Functional dependence measure for process U_t , $t \in \mathbb{N}$, with $1 - A^{-1} \sim \gamma(\delta)$: $\delta = 0.1$ (solid line), $\delta = 1$ (dashed line).

Chapter 3

Efficient Nonparametric Estimation of Models with Nonlinear Dependence

Abstract

In this paper we introduce models with constrained nonparametric dependence, which specify the conditional distribution or the copula in terms of a one-dimensional functional parameter. They provide a convenient framework for the analysis of nonlinear dependence in financial applications. As such they can be viewed as an approach that lies in between standard parametric specifications (which are in general too restrictive) and the fully unrestricted approach (which is not well-suited for the interpretation of the patterns of nonlinear dependence and suffers from the curse of dimensionality). A natural nonparametric estimator is defined by minimizing a chi-square distance between the constrained densities in the family and an unconstrained kernel estimator of the density. We derive the asymptotic properties of this estimator and of its linear functionals. We show that, under an appropriate choice of the functional parameter, the expected nonparametric one-dimensional rate of convergence of the estimator is obtained. Finally we derive the nonparametric efficiency bound and show that the minimum chi-square estimator is nonparametrically efficient.

Keywords: Nonlinear Dependence, Copula, Nonparametric Estimation, Efficiency.

JEL classification: C14, C51

3.1 Introduction

The modeling of nonlinear dependence is a topic of crucial importance in applied finance. In addition to traditional problems, such as the dependence between returns of different assets for portfolio analysis, recent developments in risk management in finance emphasize the need to carefully assess the nonlinear dependence between risks. Typical examples are the study of dependence:

- i) between the default risk of different firms to capture the so-called default correlation, that is some clustering in corporate failure [see e.g. Duffie, Singleton (1999), Li (2000), Schönbucher, Schubert (2001), and Gouriéroux, Monfort (2002a)],
- ii) between the risk on interest rate and the default risk to analyze the term structure of the spread between T-bonds and corporate bonds,
- iii) between the extreme risks in different budget lines of a bank's balance sheet, in order to aggregate the Value at Risk (VaR), and the required capital, computed per line [see e.g. Durrleman, Nikeghbali, Roncalli (2000), and Embrechts, Höing, Juri (2001)],
- iv) between intraday durations (durations between consecutive trades of an asset) to detect clustering effects in trade activity and analyse the liquidity risk.

In most of these problems the nonlinear dependence relates to the whole joint distribution of the variables (not only the first conditional moments) and the main concern is often about the tail of the joint distribution, as when the required capital is introduced to hedge extreme risks. Moreover, these problems generally involve a rather large number of variables. Indeed in example i) above the number of firms may run well over hundred, and in example iii) the number of budget lines is typically between ten and twenty.

Different approaches have been proposed in the econometric and statistical literature to describe nonlinear dependence. They can conveniently be classified in two broad groups: parametric and nonparametric approaches.

Among the class of parametric specifications, beyond the traditional approaches such as ARCH or switching regimes models, a considerable attention has been recently devoted to methods based on the joint distribution of the risk variables, such as copulas¹, especially in the framework of financial risk management. Let us briefly recall the definition of a copula, focusing for expository purpose on a pair of continuous variables X and Y , even if the definition can be extended to any multidimensional framework [see Joe (1997), and Nelsen (1999) for general presentations and the references therein]. Let $F(x, y)$ denote the bivariate cumulative distribution function (c.d.f.), $F_X(x)$ and $F_Y(y)$ the two marginal c.d.f. The joint c.d.f. can always be written as [Sklar (1959)]:

$$F(x, y) = C[F_X(x), F_Y(y)],$$

where C is the c.d.f. of a distribution on $[0, 1]^2$, with uniform marginal distributions. C is called the copula c.d.f., and is the c.d.f. of the standardized variables $U = F_X(X)$, $V = F_Y(Y)$ which are uniformly distributed on $[0, 1]$. The associated density

$$c(u, v) = \frac{\partial^2 C}{\partial u \partial v}(u, v),$$

¹Other methodologies involve for instance real Laplace transforms, or nonlinear canonical analysis [see Gagliardini, Gouriéroux (2002a) for a survey].

is the copula p.d.f. (simply called copula in the rest of the paper). Copulas present several advantages for modeling nonlinear dependence. First, they allow to specify the joint distribution by separating the marginal features (included in F_X and F_Y) and some dependence features (included in the copula). The dependence features are those which are invariant by increasing transformations of either X or Y . Second, they are a functional representation of dependence, providing a rich description of the patterns of nonlinear dependence in the different regions of the bivariate distribution.

There exists a large literature on copulas, which focuses on the analysis of positive dependence and on the search for parametric families of copulas [Joe (1997), Nelsen (1999)]. However, the dependence between financial variables such as returns, volatility or times to default is not well-captured by standard parametric copulas. Indeed the standard parametric copulas are often excessively constrained, which causes a poor fit to the data. Moreover, they are not appropriate for separate analysis of the dependence between low, medium and high risk [as required in example iii)], since copula parameters typically affect dependence in the whole sample space. In addition, they are not well suited for describing the dependence between quantitative and qualitative risks [as in example ii)]. Finally, it is rare that a standard parametric family of copulas admits a clear structural interpretation for financial applications.

The alternative approach to the modeling of nonlinear dependence consists in estimating nonparametrically the unrestricted joint density [see e.g. Silverman (1986), Härdle (1990), Scott (1992) for surveys on density estimation]. The method has been used by Deheuvels (1980) and Fermanian, Scaillet (2002) to deduce a nonparametric estimate of the associated unrestricted copula. The advantage of this approach is that it does not require any additional assumption on the nonlinear dependence. However it also has some drawbacks. Indeed, the absence of any structure complicates the interpretation of the patterns of nonlinear dependence, especially when more than 2 variables are considered, since the joint density is hard to visualize. Moreover this approach suffers from the curse of dimensionality when the number of variables of interest is larger than 4 or 5. Even in the bivariate case, it can provide inaccurate and erratic results for the VaR [see example iii)]. Indeed the Value at Risk is evaluated by considering rather extreme observations; the number of extreme observations is typically small even if the total number of observations is large and sufficient to apply bivariate nonparametric techniques for the estimation of the central part of the density function.

In this paper we explore the intermediate approach in which the joint density is constrained and depends on a small number of one-dimensional functional parameters (that are functions of one variable). Our aim is to provide efficient nonparametric estimators for the one-dimensional functional parameters that characterize nonlinear dependence. Our approach has several advantages. First, by using functional parameters instead of scalar parameters, we achieve high flexibility, while maintaining a clear structural interpretation of nonlinear dependence in terms of latent factors, or omitted heterogeneity. Second, the graphical representation of the one-dimensional functional parameters highlights the patterns of nonlinear dependence. For instance, in the example of dynamic proportional hazard models used for the analysis of liquidity risk [see example iv) above, and example iv) in section 3.2], the serial dependence in the whole sample depends on the elasticity of the functional

parameter, while tail dependence is revealed by its behaviour close to the boundary points of its support. Third, we show in the paper that the rate of convergence of the estimators, both for the functional parameters and for the joint density, is the standard one-dimensional nonparametric rate, and is independent of the number of underlying variables of interest.

Constrained nonparametric densities have already been analyzed in the literature, under various restrictions. A typical example is the transformation model, in which an unknown transformation of the endogenous variable satisfies a linear regression model with iid errors [see Han (1987a,b), and Horowitz (1996); see Gorgens, Horowitz (1999) for the case with censoring], or the location-scale model in which the mean and the volatility are unrestricted functions of a set of regressors [see e.g. Härdle, Tsybakov (1997)]. To avoid the curse of dimensionality when the number of regressors is high, these models typically adopt additivity assumptions [see e.g. Hastie, Tibshirani (1990)], or assume an index structure [see e.g. Härdle, Stocker (1989), Powel, Stock, Stocker (1989), Ichimura (1993), Horowitz, Härdle (1996)], that is the endogenous variable is explained by the set of regressors only by means of an unknown scalar transformation (called index). These nonparametric constrained regressions are suitable for describing dependence between an endogenous variable and a set of regressors, but not for instance for modeling dependence between several endogenous variables, such as times to default for several borrowers [as in example i)]. Moreover, they have been introduced as extensions of the standard linear model, which explains the form of the index function, which is often linear and therefore neglects cross effects. Moreover they assume that the same index matters for the extreme and standard values of the endogenous variable. Our purpose is to consider other types of nonparametric constraints better suited for financial or duration analysis, and admitting structural interpretations, for instance in terms of factors, or omitted heterogeneity.

For expository purpose, we discuss the case of bivariate distributions, even if the results of the paper can be extended to any multidimensional framework. We consider two alternative ways to specify the nonlinear dependence between two variables by introducing a one-dimensional functional parameter a into the conditional distribution, or the copula. In the latter case the parameterized copula is denoted by $c(u, v; a)$, where a is a function defined on $[0, 1]$. Such a constrained copula can be used for different purposes. In cross-sectional studies, it will be used to specify the joint distribution $F(x, y)$ of two risk variables, such as corporate lifetimes in the joint analysis of default. This amounts to parameterize the bivariate density $f(x, y; A)$ by three one-dimensional functional parameters:

$$A = (f_X, f_Y, a),$$

where f_X and f_Y are the unconstrained marginal densities and a the one-dimensional parameter, which characterizes the copula.

In a time series framework, it can be used to study the risk dynamics. If (X_t) is a stationary Markov process, the dynamics is fully characterized by the joint bivariate distribution $F(x_t, x_{t-1})$, whose marginal distributions are identical because of stationarity. In this case the bivariate distribution $f(x_t, x_{t-1}; A)$ is parameterized by two one-dimensional functional parameters: $A = (f, a)$, where f is the p.d.f. of the stationary distribution and a the functional parameter which characterizes the copula. Such a dynamic specification is relevant for liquidity analysis [see example iv) above], where (X_t) measures intertrade durations, or in

term structure models where the variables correspond to underlying factor processes which influence both the dynamics and patterns of the term structure.

Since the functional parameters are one-dimensional, we can expect that the estimators converge at the one-dimensional nonparametric rate $\sqrt{Th_T}$, where h_T is a bandwidth². However it is well-known that the rate of convergence is not invariant by one to one changes of the functional parameter. For instance a nonparametric estimator of a marginal p.d.f. converges generally at rate $\sqrt{Th_T}$, whereas the corresponding estimator of the c.d.f. converges at a parametric rate \sqrt{T} . To ensure the expected rate, it is necessary to assume that the joint density $f(x, y; A)$ is first order differentiable with respect to functional parameter A , and the differential is nondegenerate.

The paper is organized as follows. In section 2 we introduce the differentiability assumption, define the information operator and discuss identifiability. Various representations of the information operator are introduced, and its invertibility is discussed. In section 3 we consider several examples of constrained nonparametric families of bivariate densities, for which the joint p.d.f. is specified either by means of the conditional density and a marginal distribution, or by the copula and the two marginal distributions. For each example we discuss the structural interpretations, the parameter choice, and provide the closed form expression of the first order differential and of the information operator. In section 4, we consider a natural nonparametric estimator of functional parameter A . In the cross-sectional framework the idea is to minimize a chi-square distance between the constrained density $f(x, y; A)$ and an unconstrained kernel estimator of the density, whereas in the time series framework the conditional densities are used. We derive the asymptotic properties of the estimator and of its linear functionals. Intuitively the estimator will take account of the whole information included in the observations, since the unconstrained kernel estimator of the joint density provides semi-parametric efficient estimators for any marginal or cross-moment of (X, Y) . Thus we can expect some efficiency property of the chi-square estimator. The nonparametric efficiency of the minimum chi-square estimator is proved in section 5, where the nonparametric efficiency bounds are also derived for the cross-sectional and time series framework. In many examples the functional parameter A is subject to restrictions, which are due either to the natural constraint on the marginal density to sum up to 1, or to identification restrictions. The extension of the results to these cases is considered in section 6. Proofs are gathered in Appendices.

3.2 The information operator

In this section we discuss the local identification of the functional parameter by introducing the information operator I . The main ideas are similar to those in the usual parametric framework, but they are generalized to take into account the functional nature of the parameters. As mentioned before the basic notions are presented in dimension 2, but their extension to any dimension is straightforward.

²Note that this argument is independent of the initial number of variables.

3.2.1 Differentiability condition

Let $f(x, y; A)$ be a nonparametric family of bivariate densities, where the functional parameter A belongs to an open set \mathcal{A} of \mathbb{R}^q -valued one-dimensional functions, with a norm $\|\cdot\|_{L^2(\nu)}$, where the measure ν will be specified later on in the text [see section 2.3 ii)]. The family $f(x, y; A)$ can be parameterized in different ways. For instance, if A is differentiable, we can replace the initial function A by its derivative dA/dw , which provides the same information (up to a scalar parameter). However it is well-known that nonparametric estimators of A and dA/dw can have very different rates of convergence [see e.g. Silverman (1978), Stone (1983)]. This explains why it is necessary to standardize the functional parameter A . This standardization is introduced by means of the derivative of the density with respect to A .

Assumption A.1 *The distributions of interest are continuous with respect to the Lebesgue measure λ_2 , with p.d.f. $f(x, y; A)$. We denote by P_A the distribution associated to $f(x, y; A)$.*

Assumption A.2 *The Hadamard derivative of $\log f(x, y; A)$ with respect to A exists:*

$$\log f(x, y; A + h) - \log f(x, y; A) = \langle D \log f(x, y; A), h \rangle + R(x, y; A, h),$$

for $A, A + h \in \mathcal{A}$, where:

- i. $D \log f(\cdot, \cdot; A) : L^2(\nu) \rightarrow L^2(P_A)$ is a bounded linear operator, for any $A \in \mathcal{A}$;
- ii. the residual term $R(x, y; A, h)$ is such that $\|R(X, Y; A, h)\|_{L^2(P_A)} = o(\|h\|_{L^2(\nu)})$, uniformly on h in the class of compact sets, for any $A \in \mathcal{A}$ ³.

3.2.2 Identification and Information.

Let $A_0 \in \mathcal{A}$ denote the true value of the functional parameter, and $f(\cdot, \cdot) = f(\cdot, \cdot; A_0)$ the corresponding true p.d.f.. In this section we discuss the identification of A_0 as a minimizer of the chi-square proximity measure:

$$Q(A) = \int \int \frac{[f(x, y) - f(x, y; A)]^2}{f(x, y)} dx dy, \quad A \in \mathcal{A}.$$

Under Assumption A.2 and an additional technical condition⁴, Q is well-defined in a neighborhood of A_0 (w.r.t $\|\cdot\|_{L^2(\nu)}$) and it is locally equivalent to the Kullback proximity measure $\mathcal{K}(A) = E_0 \log [f(X, Y; A)/f(X, Y)]$ (see Appendix 2).

i) Global identification

³Precisely: $\forall A \in \mathcal{A}, K \subset \mathcal{A}$ compact: $\|R(X, Y; A, h)\|_{L^2(P_A)} / \|h\|_{L^2(\nu)} \rightarrow 0$, uniformly in $h \in K$ [see Aït-Sahalia (1993), Van der Vaart, Wellner (1996)].

⁴See Assumption A.2.bis in Appendix 2.

Under the global identification condition:

$$f(x, y; A) = f(x, y; A_0) \quad \lambda_2\text{-a.s.}, \quad A \in \mathcal{A} \implies A = A_0,$$

A_0 is the unique minimizer of Q over \mathcal{A} .

ii) Local identification.

Under Assumption A.2 we can introduce the information operator I defined by⁵:

$$(g, Ih)_{L^2(\nu)} = E_0 [\langle D \log f(X, Y; A_0), g \rangle \langle D \log f(X, Y; A_0), h \rangle], \quad (3.1)$$

for $g, h \in L^2(\nu)$. Under Assumption A.2 the information operator I is a bounded, nonnegative, self-adjoint operator from $L^2(\nu)$ in itself.

Let us consider the following assumption:

Assumption A.3. *i. Local identification: the differential operator has zero null space:*

$$\langle D \log f(X, Y; A_0), h \rangle = 0 \quad P_0\text{-a.s.}, \quad h \in L^2(\nu) \implies h = 0.$$

Assumption A.3 i. is equivalent to any of the following conditions on the information operator (see Appendix 2):

i. the information operator I has a zero null space:

$$Ih = 0, h \in L^2(\nu) \implies h = 0;$$

ii. I is a positive operator:

$$(h, Ih)_{L^2(\nu)} = 0, h \in L^2(\nu) \implies h = 0.$$

Under Assumption A.3. i. and an additional technical condition⁶, A_0 is locally identified in the following sense (see Appendix 2): A_0 is the unique minimizer of Q over any sufficiently small compact set $\Theta \subset \mathcal{A}$ that contains A_0 , and:

$$\forall \varepsilon > 0 : \inf_{A \in \Theta \setminus B_\varepsilon(A_0)} Q(A) > Q(A_0) = 0,$$

where $B_\varepsilon(A_0)$ is a $L^2(\nu)$ -ball of radius ε centered at A_0 . Assumption A.3 i. is weaker than invertibility of the information operator I . In the next section we show that, if the information operator admits a specific representation, then Assumption A.3 i. is sufficient for invertibility of I .

The identification of A_0 over noncompact subsets requires a stronger assumption:

⁵See e.g. Begun, Hall, Huang, Wellner (1983), Bickel, Klaassen, Ritov, Wellner (1993), Gill, Van der Vaart (1993), Holly (1995). In Appendix 1 we relate definition (3.1) to those adopted in the literature.

⁶See Assumption A.3.* in Appendix 2.

Assumption A.3. *ii. Local identification:*

$$\inf_{h: \|h\|_{L^2(\nu)}=1} (h, Ih)_{L^2(\nu)} > 0.$$

Under Assumption A.3. ii. A_0 is the unique minimizer of Q over any sufficiently small subset $\Theta \subset \mathcal{A}$ containing A_0 , and:

$$\forall \varepsilon > 0 : \inf_{A \in \Theta \setminus B_\varepsilon(A_0)} Q(A) > Q(A_0) = 0.$$

Assumption A.3 ii. implies in particular that operator I is invertible⁷.

3.2.3 Decompositions of the information operator

The differential operator and the information operator admit different forms in the applications. We consider in this section a particular decomposition of the information operator which is common in applied examples [see section 3.2], and discuss in this framework the choice of the measure ν and the invertibility of the information operator.

i) A decomposition of the information operator.

A case of particular importance for the applications is when the information operator I admits the representation:

$$(g, Ih)_{L^2(\nu)} = \int g(w)' \alpha_0(w; A_0) h(w) dw + \int \int g(w)' \alpha_1(w, v; A_0) h(v) dv dw, \quad (3.2)$$

where α_0 and α_1 are matrix-valued functions, such that $\alpha_0(w; A_0) = \alpha_0(w; A_0)'$, $\alpha_1(v, w; A_0) = \alpha_1(w, v; A_0)'$, $\forall v, w$. Thus the information operator I is given by:

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

and admits a decomposition in two components, corresponding to functions α_0 and α_1 , respectively. This decomposition is valid for the applied examples presented in section 3.2. To provide some insights, let us consider the case where the joint density $f(x, y; A)$ depends both on the value of function A at points x, y and on functionals of A . Then the differential operator admits the form:

$$\begin{aligned} \langle D \log f(x, y; A), h \rangle &= \gamma_0(x, y; A)' h(x) + \gamma_1(x, y; A)' h(y) \\ &\quad + \int \gamma_2(x, y, w; A)' h(w) dw, \end{aligned} \quad (3.3)$$

⁷Since I is a bounded self-adjoint operator, we have: $\inf_{h: \|h\|_{L^2(\nu)}=1} (h, Ih)_{L^2(\nu)} = \inf_{\lambda \in \sigma(I)} \lambda$, where $\sigma(I) \subset \mathbb{R}_+$ is the spectrum of I [see Yosida (1995), Theorem 2, p. 320]. Thus Assumption A.3. ii. is equivalent to $\inf_{\lambda \in \sigma(I)} \lambda > 0$, whereas the invertibility of I just requires $0 \notin \sigma(I)$.

where $\gamma_0, \gamma_1, \gamma_2$ are \mathbb{R}^q -valued functions. The information operator admits decomposition (3.2) with:

$$\begin{aligned}\alpha_0(w; A) &= \int \gamma_0(w, y; A) \gamma_0(w, y; A)' f(w, y) dy \\ &\quad + \int \gamma_1(x, w; A) \gamma_1(x, w; A)' f(x, w) dx \\ &\equiv E \left[\gamma_{0,t} \gamma_{0,t}' \mid X_t = w \right] f_X(w) + E \left[\gamma_{1,t} \gamma_{1,t}' \mid Y_t = w \right] f_Y(w),\end{aligned}\tag{3.4}$$

$$\begin{aligned}\alpha_1(w, v; A) &= \gamma_0(w, v; A) \gamma_1(w, v; A)' f(w, v) \\ &\quad + \int \gamma_0(w, z; A) \gamma_2(w, z, v; A)' f(w, z) dz \\ &\quad + \int \gamma_1(z, w; A) \gamma_2(z, w, v; A)' f(z, w) dz \\ &\quad + \frac{1}{2} \int \int \gamma_2(z, y, w; A) \gamma_2(z, y, v; A)' f(z, y) dz dy + \text{sym}(w \leftrightarrow v)'\end{aligned}\tag{3.5}$$

where $\gamma_{0,t} = \gamma_0(X_t, Y_t; A)$, $\gamma_{01,t} = \gamma_1(X_t, Y_t; A)$. The component α_0 of the information operator arises from differentiation of those parts of the joint density $f(x, y; A)$ which depend on the value of the parameter A at some point. α_0 is called local component. The components of the density which depend on functionals of A contribute only to term α_1 ⁸.

ii) Choice of the measure ν

Let us assume that the information operator satisfies decomposition (3.2), and discuss the choice of the measure ν to ensure that the differential operator $D \log f(x, y; A)$ is a bounded operator from $L^2(\nu)$ to $L^2(P_A)$.

Proposition 3.1 : Assume that the information operator satisfies the decomposition (3.2). For any $A \in \mathcal{A}$, let $\alpha(\cdot; A)$ be a positive definite matrix function such that:

$$\int \int \left\| \alpha(x; A)^{-1/2} \alpha_1(x, y; A) \alpha(y; A)^{-1/2} \right\|^2 dx dy < \infty, \forall A, \tag{3.6}$$

where $\|\cdot\|$ is a matrix norm on $\mathbb{R}^{q \times q}$. Let the measure ν be such that:

$$\forall A : \exists C_A > 0 : C_A \frac{d\nu}{d\lambda}(v) Id_q \geq \max \{ \alpha_0(v; A), \alpha(v; A) \}, \quad \forall v. \tag{3.7}$$

Then $D \log f(\cdot, \cdot; A)$ is a bounded operator from $L^2(\nu)$ to $L^2(P_A)$, for any $A \in \mathcal{A}$.

Proof. See Appendix 1.

⁸A more complete discussion of the link between representations of the differential operator and representations of the information operator is provided in Appendix 1.

The choice of a measure ν which satisfies the conditions in Proposition 1 depends in general on the parameterization. In order to illustrate this point, let us consider an independent family: $f(x, y; A) = f_X(x; A)f_Y(y; A)$.

i) If parameter A consists of the marginals themselves, $A = (f_X, f_Y)$, we get:

$$\langle D \log f(x, y; A), h \rangle = \frac{h_X(x)}{f_X(x; A)} + \frac{h_Y(y)}{f_Y(y; A)}, \quad h = (h_X, h_Y)',$$

and:

$$\begin{aligned} E_0 [\langle D \log f(X, Y; A_0), g \rangle \langle D \log f(X, Y; A_0), h \rangle] &= \int \frac{g_X(x)h_X(x)}{f_X(x; A_0)} dx \\ &+ \int \frac{g_Y(y)h_Y(y)}{f_Y(y; A_0)} dy + \int \int g_X(x)h_Y(y) dx dy + \int \int h_X(x)g_Y(y) dx dy. \end{aligned}$$

Thus:

$$\alpha_0(w; A_0) = \begin{pmatrix} 1/f_X(w; A_0) & 0 \\ 0 & 1/f_Y(w; A_0) \end{pmatrix}, \quad \alpha_1(w, v; A_0) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The choice $\alpha = \alpha_0$ satisfies condition (3.6) in Proposition 1. Condition (3.7) becomes:

$$\forall A : \exists C_A : C_A \frac{d\nu}{d\lambda}(v) \geq \max \left\{ \frac{1}{f_X(v; A)}, \frac{1}{f_Y(v; A)} \right\}, \quad \forall v.$$

ii) If instead we choose $A = (f_X^{1/2}, f_Y^{1/2})$, we get:

$$\begin{aligned} E_0 [\langle D \log f(X, Y; A_0), g \rangle \langle D \log f(X, Y; A_0), h \rangle] &= 4 \left\{ \int g_X(x)h_X(x) dx \right. \\ &+ \left. \int g_Y(y)h_Y(y) dy + \int \int [g_X(x)h_Y(y) + h_X(x)g_Y(y)] f_X(x; A)^{1/2} f_Y(y; A)^{1/2} dx dy \right\}, \end{aligned}$$

that is $\alpha_0(v; A_0) = 4Id_2$ and:

$$\alpha_1(w, v; A_0) = 4 \begin{pmatrix} 0 & f_X(w; A)^{1/2} f_Y(v; A)^{1/2} \\ f_X(v; A)^{1/2} f_Y(w; A)^{1/2} & 0 \end{pmatrix}.$$

Conditions (3.6) and (3.7) are satisfied by $\alpha = Id_2$, $\nu = \lambda$.

iii) Finally, if $A = (\log f_X, \log f_Y)$, we get:

$$\begin{aligned} E_0 [\langle D \log f(X, Y; A_0), g \rangle \langle D \log f(X, Y; A_0), h \rangle] &= \int g_X(x)h_X(x)f_X(x; A) dx \\ &+ \int g_Y(y)h_Y(y)f_Y(y; A) dy + \int \int [g_X(x)h_Y(y) + h_X(x)g_Y(y)] f_X(x; A)f_Y(y; A) dx dy, \end{aligned}$$

that is:

$$\begin{aligned}\alpha_0(v; A_0) &= \begin{pmatrix} f_X(v; A_0) & 0 \\ 0 & f_Y(w; A_0) \end{pmatrix}, \\ \alpha_1(w, v; A_0) &= \begin{pmatrix} 0 & f_X(w; A)f_Y(v; A) \\ f_X(v; A)f_Y(w; A) & 0 \end{pmatrix}.\end{aligned}$$

The choice $\alpha = \alpha_0$ satisfies condition (3.6) in Proposition 1. Condition (3.7) is equivalent to:

$$\forall A : \exists C_A : C_A \frac{d\nu}{d\lambda}(v) \geq \max \{f_X(v; A), f_Y(v; A)\}, \quad \forall v,$$

that is the measure ν dominates both marginal distributions in the family.

iii) Invertibility of the information operator

When the information operator satisfies decomposition (3.2) with additional restrictions on α_0 , a zero null space of the information operator I is sufficient for its invertibility⁹.

Proposition 3.2 : Assume the conditions of Proposition 1, and in addition let $\alpha_0(v; A)$ be invertible, $\forall v, \forall A$, such that:

$$\forall A : \exists \tilde{C}_A > 0 : \tilde{C}_A \frac{d\nu}{d\lambda}(v) Id_q \leq \alpha_0(v; A), \quad \forall v.$$

Assume further that the information operator I has a zero null space. Then the information operator is continuously invertible.

Proof. See Appendix 1.

3.3 Examples

3.3.1 Differentials of the copula and of the conditional and marginal densities.

A family of bivariate joint densities can be specified in various ways. One possibility is to parameterize the conditional density and one marginal distribution. Alternatively we can parameterize the copula and the marginal distributions. In both cases, the differential of the joint density can be recovered from the differentials of its components.

i) Conditional density and marginal density.

Assume $f_{X|Y}(x | y; A)$ [resp. $f_Y(y; A)$] is a differentiable family of conditional distributions of X given Y [resp. of marginal distributions of Y], parameterized by function A . Let

⁹The following proposition uses the theory of Fredholm operators [see Van der Vaart (1994) for similar results].

$D \log f_{X|Y}$, and $D \log f_Y$ denote their differentials with respect to A . A family of bivariate densities is defined by:

$$f(x, y; A) = f_{X|Y}(x | y; A) f_Y(y; A).$$

We have (see Appendix 3) ¹⁰:

Proposition 3.3 : *The differential of $\log f(x, y; A)$ is given by:*

$$D \log f(x, y; A) = D \log f_{X|Y}(x | y; A) + D \log f_Y(y; A).$$

Moreover:

$$\begin{aligned} D \log f_Y(y; A) &= E_A [D \log f(X, Y; A) | Y = y] \\ &= \int D \log f(x, y; A) f_{X|Y}(x | y; A) dx. \end{aligned}$$

Thus $D \log f_{X|Y}(x | y; A)$ is the residual in the projection of $D \log f(x, y; A)$ on Y ; in particular it is orthogonal to $D \log f_Y(y; A)$:

$$E_A [\langle D \log f_{X|Y}(X | Y; A), h \rangle \langle D \log f_Y(Y; A), g \rangle] = 0, \quad \forall h, g \in L^2(\nu). \quad (3.8)$$

As a consequence the information operator I is the sum of a conditional and a marginal information operator:

$$I = I_{X|Y} + I_Y,$$

where $I_{X|Y}$ and I_Y are defined by:

$$\begin{aligned} (g, I_{X|Y} h)_{L^2(\nu)} &= E_0 [\langle D \log f_{X|Y}(X | Y; A_0), g \rangle \langle D \log f_{X|Y}(X | Y; A_0), h \rangle], \\ (g, I_Y h)_{L^2(\nu)} &= E_0 [\langle D \log f_Y(Y; A_0), g \rangle \langle D \log f_Y(Y; A_0), h \rangle], \end{aligned}$$

for $h, g \in L^2(\nu)$.

An interesting special case occurs in the stationary time-series framework when there exists a unique stationary distribution. Then the conditional and marginal distributions are linked by the Chapman-Kolmogorov equation:

$$f(x; A) = \int f_{X|Y}(x | y; A) f(y; A) dy. \quad (3.9)$$

By differentiating this equation, we get the relationship satisfied by the associated differentials.

¹⁰The differential $D \log f_Y(y; A)$ is an operator indexed by y . Its integral $\int \varphi(y) D \log f_Y(y; A) dy$ with respect to a function φ is defined in the usual distributional sense as:

$$\left\langle \int \varphi(y) D \log f_Y(y; A) dy, h \right\rangle := \int \varphi(y) \langle D \log f_Y(y; A), h \rangle dy.$$

Similarly for the other differentials.

Proposition 3.4 : *If the marginal distribution satisfies the Chapman-Kolmogorov condition, the differential Df satisfies the integral equation:*

$$Df(x; A) = \int Df_{X|Y}(x|y; A)f(y; A)dy + \int f_{X|Y}(x|y; A)Df(y; A)dy.$$

ii) Copula and marginal distributions

A family of bivariate densities for (X, Y) can also be defined by specifying the copula $c(u, v; A)$, and the marginal distributions $f_X(x; A)$, $f_Y(y; A)$:

$$f(x, y; A) = c[F_X(x; A), F_Y(y; A); A] f_X(x; A) f_Y(y; A).$$

Proposition 3.5 : *The differential of the density $f(x, y; A)$ is given by:*

$$\begin{aligned} D \log f(x, y; A) &= D \log c[F_X(x; A), F_Y(y; A); A] \\ &+ D \log f_X(x; A) + D \log f_Y(y; A) \\ &+ \frac{\partial \log c}{\partial u}[F_X(x; A), F_Y(y; A); A] \int_{-\infty}^x f_X(z; A) D \log f_X(z; A) dz \\ &+ \frac{\partial \log c}{\partial v}[F_X(x; A), F_Y(y; A); A] \int_{-\infty}^y f_Y(z; A) D \log f_Y(z; A) dz. \end{aligned} \quad (3.10)$$

Proof. See Appendix 3.

In a cross-sectional framework the functional parameter A is often chosen as:

$$A = (f_X, f_Y, a),$$

where a characterizes the copula. The differential of $\log f(x, y; A)$ is given in the following corollary, where the effects of the different functional parameters are distinguished.

Corollary 3.6 : *The differential of the density $f(x, y; A)$ is given by:*

$$\begin{aligned} D_a \log f(x, y; A) &= D \log c[F_X(x), F_Y(y); a], \\ \langle D_{f_X} \log f(x, y; A), h \rangle &= \frac{\partial \log c}{\partial u}[F_X(x), F_Y(y); a] \int_{-\infty}^x h(z) dz + \frac{h(x)}{f_X(x)}, \\ \langle D_{f_Y} \log f(x, y; A), h \rangle &= \frac{\partial \log c}{\partial v}[F_X(x), F_Y(y); a] \int_{-\infty}^y h(z) dz + \frac{h(y)}{f_Y(y)}. \end{aligned}$$

Let us define the information operator I_{cop} associated with the copula density:

$$(g, I_{cop} h)_{L^2(\nu)} = E_0 [\langle D \log c(U, V; A_0), g \rangle \langle D \log c(U, V; A_0), h \rangle],$$

for $h, g \in L^2(\nu)$. Since:

$$E_A [\langle D \log c(U, V; A_0), h \rangle | U] = E_A [\langle D \log c(U, V; A), h \rangle | V] = 0,$$

$\forall h \in L^2(\nu)$, the first term in the decomposition of the differential [see equation (3.10)] is orthogonal to the second and the third ones. Let I_X and I_Y be the marginal information operators [defined in i)], and I_{XY} , I_{YX} the cross operators, defined by $(g, I_{XY}h)_{L^2(\nu)} = E_0 [\langle D \log f_X(X; A_0), g \rangle \langle D \log f_Y(Y; A_0), h \rangle]$, and similarly for I_{YX} . Then the information operator I can be decomposed as:

$$I = I_{cop} + I_X + I_Y + I_{XY} + I_{YX} + J,$$

where the term J comes from the last two terms in (3.10).

In particular when the parameter is $A = (f_X, f_Y, a)$ [see Corollary 6], the information operator I has a block decomposition, with univariate versions of I_X , I_Y , and I_{cop} on the diagonal. The elements out of the diagonal corresponding to (f_X, a) and (f_Y, a) are not zero due to the first terms in the differentials with respect to the marginal distributions given in Corollary 6. These terms arise since the efficient copula estimator provides information on the marginal distributions (see Genest, Werker [2001]).

3.3.2 Examples.

We consider below different constrained nonparametric families of bivariate densities, and give the expressions of the differential of the copula or of the conditional density (see Appendix 4 for some derivations). We provide an appropriate choice of the functional parameter in each example, in order to ensure that Assumption A.2 is satisfied and the information operator admits the representation (3.2). As seen from the examples, this choice is the difficult part when specifying nonlinear dependence.

i) Truncated model

Let us consider a latent variable X^* with p.d.f. f^* , $f^* > 0$, and assume that, for any value of $Y = y$, the value of X is drawn in the conditional distribution of X^* given $X^* < y$. This situation occurs in models with truncation, where the truncation variable Y is independent of the latent variable X^* of interest. The parameter of interest is the pdf f^* of the latent variable. The conditional p.d.f. of X given Y is:

$$f(x | y) = \frac{f^*(x)}{\int_{-\infty}^y f^*(z) dz} \mathbb{I}_{x \leq y}.$$

By choosing the parametrization $A = \log f^*$, the differential of $\log f(x | y; A)$, for $x \leq y$, is given by:

$$\begin{aligned} \langle D \log f(x | y; A), h \rangle &= h(x) - \int f(z | y; A) h(z) dz \\ &= h(x) - E_A [h(X) | Y = y]. \end{aligned}$$

Let us now consider the conditional information operator $I_{X|Y}$. By definition we have:

$$\begin{aligned} (g, I_{X|Y}h)_{L^2(\nu)} &= E_0 \{ (g(X) - E_0[g(X) | Y]) (h(X) - E_0[h(X) | Y]) \} \\ &= E_0 \text{Cov}_0(g(X), h(X) | Y). \end{aligned}$$

It admits the decomposition (3.2) with:

$$\begin{aligned} \alpha_0(x; A) &= f_X(x; A), \\ \alpha_1(x, y; A) &= - \int f(x | z; A) f(y | z; A) f_Y(z; A) dz. \end{aligned}$$

Let us finally discuss the boundedness of the differential operator (Proposition 1). If we choose $\alpha(x; A) = f_X(x; A)$ we get:

$$\int \int \frac{\alpha_1(x, y; A)^2}{\alpha(x; A)\alpha(y; A)} dx dy = \int \int \frac{[\int f(x | z; A) f(y | z; A) f_Y(z; A) dz]^2}{f_X(x; A)f_X(y; A)} dx dy.$$

Thus condition (3.6) of Proposition 1 requires¹¹:

$$\int \int \frac{[\int f(x | z; A) f(y | z; A) f_Y(z; A) dz]^2}{f_X(x; A)f_X(y; A)} dx dy < \infty. \quad (3.11)$$

Moreover the measure ν has to satisfy:

$$\forall A : \exists C_A > 0 : C_A \frac{d\nu}{d\lambda}(x) \geq f_X(x; A), \quad \forall x. \quad (3.12)$$

The measure ν must dominate the marginal density of X , for any distribution in the family.

ii) Truncated dynamic models.

Let S be a differentiable survivor function on \mathbb{R}_+ and let a be a positive function on \mathbb{R}_+ . The positive valued Markov process (X_t) follows a truncated dynamic model if its transition survivor function satisfies:

$$P(X_t \geq x_t | X_{t-1} = x_{t-1}) = \frac{S[x_t + a(x_{t-1})]}{S[a(x_{t-1})]}.$$

Thus the distribution of X_t given $X_{t-1} = x_{t-1}$ is the distribution of the excess $X^* - a(x_{t-1})$, where X^* is truncated at $a(x_{t-1})$, $X^* \geq a(x_{t-1})$, and $X^* \sim S$. The patterns of serial dependence in Markov process (X_t) are characterized by functional parameter a .

¹¹Note that $\int f(x | z; A) f(y | z; A) f_Y(z; A) dz$ is the joint density of two observations of X having the same (unknown) conditioning value Y . This distribution has marginals equal to $f_X(\cdot; A)$, and the expression in the LHS of (3.11) is the sum of its squared canonical correlations, see e.g. Dunford, Schwartz (1968), and Lancaster (1968).

Let us denote by g (resp. λ) the density (resp. the hazard function) corresponding to S . The conditional distribution is given by:

$$f(x_t | x_{t-1}; A) = \frac{g[x_t + a(x_{t-1})]}{\int_{a(x_{t-1})}^{+\infty} g(z) dz}, \quad A = (a, \log g)'.$$

The differential is:

$$\begin{aligned} \langle D_a \log f(x | y; A), h \rangle &= \left(\frac{d \log g}{dz} [x + a(y)] + \lambda[a(y)] \right) h(y), \\ \langle D_{\log g} \log f(x | y; A), h \rangle &= h(x + a(y)) - E_A[h(X_t + a(X_{t-1})) | X_{t-1} = y]. \end{aligned}$$

The information operator admits the representation (3.2), with local component:

$$\alpha_0(w; A) = \begin{pmatrix} E_A \left[\left(\frac{d \log g}{dz} [X_t + a_{t-1}] + \lambda[a_{t-1}] \right)^2 | X_{t-1} = w \right] f(w; A) & 0 \\ 0 & f_{X_t + a_{t-1}}(w; A) \end{pmatrix},$$

where $a_{t-1} = a(X_{t-1})$ and f [resp. $f_{X_t + a_{t-1}}$] is the stationary density of X_t [resp. $X_t + a(X_{t-1})$].

iii) Stochastic unit root.

The stochastic unit root model has been introduced by Gouriéroux and Robert (2001) to study the links between long memory, endogenous switching regimes and heavy tails, often encountered in financial time series. The process is defined by:

$$X_t = \begin{cases} X_{t-1} + \varepsilon_t & , \text{ with prob. } \pi(X_{t-1}), \\ \varepsilon_t & , \text{ with prob. } 1 - \pi(X_{t-1}), \end{cases}$$

where the ε_t are i.i.d. errors independent from X_{t-1} , with density g , $g > 0$, and π is a function with values in $]0, 1]$. This is a Markov process with two possible stochastic regimes, corresponding to either a random walk, or a white noise¹². A latent binary variable Z_t can be introduced, with $Z_t = 1$ (resp. $Z_t = 0$) when the process is in the random walk (resp. white noise) regime. Such a specification underlies the analysis of purchasing power parity (PPP), when it is assumed that unit roots can exist inside a band for the PPP equilibrium, whereas mean-reverting effects exist outside the band [see e.g. Bec, Salem, Carrasco, (2001, 2002), and Rahbek, Shephard, (2002)]. Function π characterizes nonlinear serial dependence properties of Markov process (X_t) [see Gouriéroux, Robert (2001)]. For instance, the tail behaviour of $\pi(y)$ when $y \rightarrow \infty$ characterizes the durations of (X_t) in the random walk regime.

The conditional density is given by:

$$f(x | y) = \pi(y)g(x - y) + [1 - \pi(y)]g(x).$$

¹²The specification is easily extended to a second regime which is a stationary autoregression.

For parameterization $A = (\log \pi, \log g)'$, the differential is given by:

$$\begin{aligned}\langle D_{\log \pi} \log f(x | y; A), h \rangle &= r(x, y; A) h(y), \\ \langle D_{\log g} \log f(x | y; A), h \rangle &= p_1(x, y; A) h(x - y) + p_0(x, y; A) h(x),\end{aligned}$$

where $r(x, y; A) = [f(x | y; A) - g(x; A)] / f(x | y; A)$, and p_0, p_1 are the filtering probabilities:

$$\begin{aligned}p_1(x_t, x_{t-1}; A) &= P_A[Z_t = 1 | \underline{x}_t] = P_A[Z_t = 1 | x_t, x_{t-1}] \\ &= \pi(x_{t-1}) g(x_t - x_{t-1}) / [\pi(x_{t-1}) g(x_t - x_{t-1}) + (1 - \pi(x_{t-1})) g(x_t)],\end{aligned}$$

and:

$$p_0(x_t, x_{t-1}; A) = P_A[Z_t = 0 | \underline{x}_t] = 1 - p_1(x_t, x_{t-1}; A).$$

The information operator admits representation (3.2) with:

$$\alpha_0(z; A_0) = \begin{pmatrix} E_0[r_t^2 | X_{t-1} = z] f(z) & 0 \\ 0 & E_0[p_{1,t}^2 | X_t - X_{t-1} = z] f_{X_t - X_{t-1}}(z) \\ & + E_0[p_{0,t}^2 | X_t = z] f(z) \end{pmatrix}$$

and α_1 given in Appendix 4, where $r_t = r(X_t, X_{t-1}; A_0)$, $p_{0,t} = p_0(X_t, X_{t-1}; A_0)$, $p_{1,t} = p_1(X_t, X_{t-1}; A_0)$, f (resp. $f_{X_t - X_{t-1}}$) is the stationary density of X_t (resp. $X_t - X_{t-1}$), and all functions are evaluated at A_0 . The component of $\alpha_0(z; A_0)$ relative to $\log \pi$ depends on $E_0[r_t^2 | X_{t-1} = z]$, which is the conditional chi-square distance between the conditional distribution and the density of the innovation. The component relative to $\log g$ depends on conditional expectations of the squared filtering probabilities, $p_{1,t}^2$ and $p_{0,t}^2$, given $X_t - X_{t-1} = z$ and $X_t = z$ respectively. The filtering probabilities are conditional to the innovation, since the innovation ε_t is either equal to $X_t - X_{t-1}$, when the process is in the random walk regime, or to X_t when it is in the white noise regime.

iv) Dynamic models with proportional hazard.

This specification describes time series (X_t) of duration variables, where the lagged values are explanatory variables with proportional hazard effect. Such models are used for measuring liquidity risk from intertrade duration data [see Gagliardini, Gouriou, (2002b)]. Since the proportional hazard condition is invariant by increasing transformation, it only concerns the copula of the process, and any stationary distribution can be imposed by an appropriate marginal transformation. The distribution of the Markov process (U_t) with proportional hazard and uniform marginal distribution can be written as:

$$P[U_t \geq u | U_{t-1} = v] = \exp[-a(v) H_0(u)],$$

where a is a positive function on $[0, 1]$, and H_0 is a baseline cumulated hazard on $[0, 1]$. Functions a and H_0 are restricted by the condition of uniform margins:

$$1 - u = E[P[U_t \geq u | U_{t-1}]], \forall u \in [0, 1],$$

that is:

$$H_0^{-1}(z) = 1 - \int_0^1 \exp[-za(v)] dv, \quad z \geq 0. \quad (3.13)$$

Thus the proportional hazard copula of (U_t, U_{t-1}) is characterized by the functional parameter a and it is given by:

$$c(u, v; a) = a(v)h_0(u; a) \exp[-a(v)H_0(u; a)],$$

where $H_0(u; a)$ is defined by (3.13), and $h_0 = dH_0/du$. The distribution of Markov process $X_t = F^{-1}(U_t)$ with proportional hazard and marginal cdf F is characterized by the two one-dimensional functional parameters (f, a) . In Gagliardini, Gouriéroux (2002b) it is shown that the strength of serial dependence in Markov process (X_t) is related to the elasticity of function a , whereas the behaviour of the latter close to the boundary points $v = 0$, $v = 1$ characterizes the tail dependence properties of the process. Note finally that two functional parameters differing by a multiplicative constant, a and ka (say), define the same proportional hazard copula.

The differential of the copula density is given by [see Gagliardini, Gouriéroux (2002b)]:

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

where $a_{t-1} = a(U_{t-1})$, $H_{0t} = H_0(U_t, a)$,

$$\gamma_0(u, v; a) = \frac{1 - a(v)H_0(u; a)}{a(v)},$$

and γ_1 is given in Gagliardini, Gouriéroux (2002b), formula (a.13), Appendix 7. From (3.4), (3.5) the copula information operator admits the form (3.2) with local component:

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

and α_1 given in Appendix 8 of Gagliardini, Gouriéroux (2002b) ^{13 14}.

¹³It is possible to consider the example of general distributions (X, Y) with proportional hazard:

$$P[X \geq x | Y = y] = \exp[-a(y)\Lambda(x)],$$

where a is a positive function, and Λ is the baseline cumulated hazard.

¹⁴The results on proportional hazard copula can be extended to more general transformation copulas, corresponding to the c.d.f. of variables (U_t, V_t) with uniform margins, satisfying:

$$H_0(U_t) = \frac{\varepsilon_t}{a(V_t)},$$

where a is a positive function, H_0 is increasing, and the innovation ε_t is independent from V_t , with a distribution with support in \mathbb{R}_+ . The case where ε_t has an exponential distribution corresponds to proportional hazard.

v) Archimedean copula.

The family is usually defined by [see Genest and Mc Kay (1986)]:

$$C(u, v) = \phi [\phi^{-1}(u) + \phi^{-1}(v)], \quad (3.14)$$

where the (strict) generator ϕ^{-1} is a convex, decreasing function defined on $(0, 1]$, such that $\phi^{-1}(1) = 0$, and $\phi^{-1}(0) = +\infty$. Many of the most well-known archimedean copulas are derived from factor models. Typically they correspond to duration models, where the duration variables X and Y are independent identically distributed conditionally to an omitted factor Z , and the factor Z has identical proportional hazard effects on the duration distributions [see e.g. Van der Berg (2001)]¹⁵. In this case ϕ is the Laplace transform of the positive random variable Z :

$$\phi(s) = E[\exp(-sZ)], \quad s \geq 0. \quad (3.15)$$

This specification is useful for modeling default correlation, with heterogeneity Z being a latent economic factor with a common proportional hazard effect on the durations until default X, Y of two firms. The patterns of the nonlinear dependence between X and Y are characterized by the Laplace transform ϕ of the omitted factor Z . For instance, the strength of the dependence is related to the dispersion of factor Z , whereas the tails of Z characterize tail dependence and the age structure of default correlation in the distribution of X and Y ¹⁶.

Assume ϕ is twice continuously differentiable. The copula p.d.f. is:

$$c(u, v) = \frac{\phi'' [\phi^{-1}(u) + \phi^{-1}(v)]}{\phi' [\phi^{-1}(u)] \phi' [\phi^{-1}(v)]}.$$

Even if the generator ϕ (or ϕ^{-1}) is a natural functional parameter for the Archimedean copula, it does not satisfy the differentiability condition given in Assumption A.2. The proposition below introduces an equivalent functional parameter in one-to-one relationship with ϕ . Let us consider the transformed variables:

$$\begin{aligned} W &= C(U, V) \\ Z &= V. \end{aligned}$$

Proposition 3.7 : *The joint p.d.f. of W and Z is given by:*

$$f(w, z) = \frac{f^*(w)}{\int_0^z f^*(v) dv} 1_{w \leq z}, \quad w, z \in (0, 1),$$

¹⁵The Archimedean copula admits a direct extension to multidimensional framework as $\phi [\sum_{i=1}^n \phi^{-1}(u_i)]$. This is a typical example of symmetric copula with large dimension depending on a single one-dimensional functional parameter. Clearly such symmetric copulas, useful in default correlation analysis, do not belong to the class of index models.

¹⁶See Gagliardini, Gouriéroux (2002a) and Gouriéroux, Monfort (2002b).

where the latent measure density f^* is given by:

$$f^*(w) = -\frac{\phi''[\phi^{-1}(w)]}{\phi'[\phi^{-1}(w)]}, \quad w \in (0, 1). \quad (3.16)$$

Moreover we have a one-to-one relationship between the measure F^* and the generator ϕ^{-1} since:

$$F^*(w) = -\phi'[\phi^{-1}(w)] \iff \phi^{-1}(y) = \int_y^1 \frac{dw}{F^*(w)},$$

under the condition $\int_0^1 1/F^*(w)dw = \infty$ ¹⁷.

Proof. See Appendix 4.

The generator ϕ^{-1} and the function f^* are identifiable up to a multiplicative constant. This identification problem can be solved by imposing that f^* is a p.d.f., as we will do in the following. Then variables W and Z follow a truncation model [see example i)], with latent density f^* in (3.16) and $Z \sim U(0, 1)$.

We choose to parameterize the copula density by means of function $a = f^*$. Thus the copula density is:

$$c(u, v; a) = a[C(u, v; a)] \frac{F^*[C(u, v; a); a]}{F^*(u; a) F^*(v; a)},$$

where functional parameter a is a positive function defined on $[0, 1]$ and such that:

$$\int_0^1 a(v)dv = 1.$$

The differential is given by:

$$\langle D \log c(u, v; a), h \rangle = \frac{h[C(u, v; a)]}{a[C(u, v; a)]} + \int_0^1 \gamma(u, v, w; a) h(w)dw,$$

where function γ is given in Appendix 4. The information operator is of the form (3.2), where the local component is given by:

$$\alpha_0(w, a) = \frac{f_W(w; a)}{a(w)^2} = \frac{\phi^{-1}(w; a)}{a(w)},$$

¹⁷By the mean value theorem: $F^*(v) \simeq f^*(0)v$, for $v \simeq 0$, and thus condition $\int_0^1 1/F^*(v)dv = \infty$ is satisfied if $f^*(0) < \infty$. Since:

$$f^*(0) = \lim_{w \rightarrow 0} f^*(w) = \lim_{w \rightarrow \infty} -\frac{\phi''(w)}{\phi'(w)},$$

this condition is equivalent to:

$$\lim_{w \rightarrow \infty} -\frac{\phi''(w)}{\phi'(w)} < \infty.$$

where $f_W(\cdot; a)$ is the p.d.f. of variable W , and α_1 is reported in Appendix 4.

vi) Extreme value copula

Let (Z_i, W_i) , $i = 1, \dots, n$ be independent pairs of random variables. Extreme value bivariate copulas are associated with the limiting joint distribution of marginal maxima $(\max_i Z_i, \max_i W_i)$, as n tends to infinity. Extreme value copulas are of the form [see e.g. Joe (1997)]:

$$C_\chi(u, v) = \exp \left\{ (\log u + \log v) \chi \left(\frac{\log u}{\log u + \log v} \right) \right\},$$

where the generator χ is a function defined on $[0, 1]$, is convex, and satisfies:

$$\max(v, 1 - v) \leq \chi(v) \leq 1.$$

Assume function χ is differentiable. The copula p.d.f. is given by:

$$\begin{aligned} c_\chi(u, v) = & \frac{C(u, v)}{uv} \left\{ -\frac{\tilde{u}\tilde{v}}{\log u + \log v} \chi''(\tilde{u}) \right. \\ & \left. + \left[\chi(\tilde{u}) + \tilde{v}\chi'(\tilde{u}) \right] \left[\chi(\tilde{u}) - \tilde{u}\chi'(\tilde{u}) \right] \right\}, \end{aligned}$$

where $\tilde{u} = \log u / (\log u + \log v)$, $\tilde{v} = \log v / (\log u + \log v)$. The functional parameter χ does not satisfy Assumption A.2. As in the example of the archimedean family, we look for a parameter which is related to χ'' . In order to get intuition, let us consider an alternative characterization of function χ . The generator χ of an extreme value copula can be written as (see e.g. Joe [1997], and Appendix 4):

$$\chi(v) = 2 \int_0^1 \max\{(1-z)v, z(1-v)\} dF^*(z),$$

where F^* is a c.d.f. on $[0, 1]$ such that: $\int_0^1 z dF^*(z) = 1/2$. When F^* admits a density f^* , we get:

$$\chi'' = 2f^*.$$

Thus, an extreme value copula can be parameterized by the functional parameter $a = f^* = \chi''/2$:

$$\begin{aligned} c(u, v; a) = & \frac{C(u, v; a)}{uv} \left\{ -\frac{2\tilde{u}\tilde{v}}{\log u + \log v} a(\tilde{u}) \right. \\ & \left. + \left[\chi(\tilde{u}; a) + \tilde{v}\chi'(\tilde{u}; a) \right] \left[\chi(\tilde{u}; a) - \tilde{u}\chi'(\tilde{u}; a) \right] \right\}, \end{aligned}$$

and the functional parameter a is a positive function defined on $[0, 1]$ satisfying the constraints:

$$\int_0^1 a(v) dv = 1, \quad \int_0^1 va(v) dv = 1/2.$$

The differential of the copula density is of the form:

$$\langle D \log c(u, v; a), h \rangle = \gamma_0(u, v; a) h(\tilde{u}) + \int_0^1 \gamma_1(u, v, w; a) h(w) dw,$$

where:

$$\gamma_0(u, v; a) = \left\{ a(\tilde{u}) - \frac{\log u + \log v}{2\tilde{u}\tilde{v}} \left[1 - \int_0^{\tilde{u}} w a(w) dw \right] \left[\int_0^{\tilde{u}} a(w) dw - \int_0^{\tilde{u}} w a(w) dw \right] \right\}^{-1}.$$

The copula information operator admits representation (3.2) with local component:

$$\alpha_0(w; a) = E_a \left[\gamma_0(U, V; a)^2 \mid \tilde{U} = w \right] f_{\tilde{U}}(w; a),$$

where $\tilde{U} = \log U / (\log U + \log V)$, and $f_{\tilde{U}}$ is the density of \tilde{U} .

vii) Markov processes with finite dimensional canonical decomposition

Nonlinear canonical analysis provides a decomposition of a stationary Markov process X_t , $t \in \mathbb{N}$, in orthogonal functional directions $\varphi_j(X_t)$, $\psi_j(X_{t-1})$, $j \in \mathbb{N}$ varying, of decreasing nonlinear dependence¹⁸. Functions φ_j , ψ_j , j varying, are called canonical directions, and $\lambda_j = \text{corr}[\varphi_j(X_t), \psi_j(X_{t-1})]$, j varying, are the associated canonical correlations. The canonical decomposition of Markov process (X_t) is characterized, up to increasing transformations of the canonical directions, by the canonical decomposition of the copula.

A stationary Markov process with one dimensional canonical decomposition is obtained when $\lambda_j = 0$, $j \geq 2$, and $\lambda_1 = \lambda > 0$ [see Gouriéroux, Jasiak (2001)]. Its copula is given by:

$$c(u, v) = 1 + \lambda \varphi(u) \psi(v),$$

where the canonical directions φ and ψ satisfy the conditions:

$$\int_0^1 \varphi(u) du = \int_0^1 \psi(v) dv = 0,$$

with the normalization:

$$\int_0^1 \varphi(u)^2 du = \int_0^1 \psi(v)^2 dv = 1,$$

and are such that the copula density is positive. Let us for simplicity consider the case of reversible Markov processes, that is $\varphi = \psi$. Then the copula density can be parameterized by $a = \sqrt{\lambda}\varphi$, and we get:

$$c(u, v) = 1 + a(u)a(v),$$

where the functional parameter a satisfies the constraint:

$$\int_0^1 a(v) dv = 0.$$

¹⁸See Lancaster (1968), and Dunford, Schwartz (1968); see also Gouriéroux, Jasiak (2002) for an application to intertrade durations.

The canonical correlation λ and the canonical direction φ are deduced from a by the equations:

$$\lambda = \int_0^1 a(v)^2 dv, \quad \varphi(u) = \frac{1}{\sqrt{\lambda}} a(u).$$

The differential of the copula is given by:

$$D \log c(u, v; a) = \frac{a(v)}{1 + a(u)a(v)} h(u) + \frac{a(u)}{1 + a(u)a(v)} h(v),$$

and from (3.4), (3.5) the information operator admits representation (3.2), with local component:

$$\alpha_0(w; a) = \frac{2}{a(w)^2} E_a \left[\left(\frac{c(U, V; a) - 1}{c(U, V; a)} \right)^2 \mid V = w \right],$$

and:

$$\alpha_1(w, v; a) = 2 \frac{a(w)a(v)}{1 + a(w)a(v)}.$$

Thus the local component α_0 involves the conditional chi-square distance between the copula $c(., .; a)$ and the independent copula.

3.4 Minimum chi-square estimators.

In this section we study the properties of minimum chi-square estimators. We first consider the cross-sectional framework, where the observations (X_t, Y_t) , t varying, are i.i.d., define the estimator, prove its consistency and derive its asymptotic distribution. Then we provide similar results in the time series framework.

3.4.1 Definition of the estimator.

Let us consider the cross-sectional framework:

Assumption A.4: *The variables (X_t, Y_t) , t varying, are i.i.d., with distribution $f(x, y; A)$. The support of the p.d.f. is $[0, 1]^2$.*

It is always possible to transform variables (X_t^*, Y_t^*) with values in \mathbb{R} into variables with values in $[0, 1]$ for instance by applying the logit transformation. Therefore the assumption of compact support $[0, 1]^2$ is not restrictive.

Let us introduce a kernel estimator of the unconstrained bivariate density function [Rosenblatt (1956), Parzen (1962)]:

$$\hat{f}_T(x, y) = \frac{1}{Th_T^2} \sum_{t=1}^T K\left(\frac{x - X_t}{h_T}\right) K\left(\frac{y - Y_t}{h_T}\right), \quad (3.17)$$

where K is a kernel and h_T is a bandwidth. Under standard regularity properties (see Appendix 5, Assumptions B.1-B.4), the estimator is consistent and asymptotically normal:

$$\sqrt{Th_T^2} \left[\hat{f}_T(x, y) - f(x, y; A) \right] \xrightarrow{d} N(0, \sigma^2(x, y; A)), \quad (3.18)$$

where $\sigma^2(x, y; A) = f(x, y; A) \left(\int K^2(w) dw \right)^2$. Moreover, we have also the consistency and asymptotic normality of linear functionals of f , that are conditional and cross-moments, at rates depending on the number of integrations:

$$\sqrt{Th_T} \left[\int g(x) \hat{f}_T(x, y) dx - \int g(x) f(x, y; A) dx \right] \xrightarrow{d} N(0, \sigma^2(y, g; A)), \quad (3.19)$$

where $\sigma^2(y, g; A) = E_A[g(X_t)^2 | Y_t = y] f_Y(y) \int K^2(w) dw$, and

$$\sqrt{T} \left[\int \int g(x, y) \hat{f}_T(x, y) dx dy - \int \int g(x, y) f(x, y; A) dx dy \right] \xrightarrow{d} N(0, \sigma^2(g; A)), \quad (3.20)$$

where $\sigma^2(g) = V_A[g(X_t, Y_t)]$.

The unconstrained estimator of the bivariate density can be used to derive a minimum chi-square estimator of A :

$$\hat{A}_T = \arg \min_{A \in \Theta} Q_T(A) = \int_0^1 \int_0^1 \frac{\left[\hat{f}_T(x, y) - f(x, y; A) \right]^2}{\hat{f}_T(x, y)} \omega_T(x, y) dx dy, \quad (3.21)$$

where Θ is a subset of \mathcal{A} , ω_T is a smooth weighting function, converging pointwise to the identity function on $(0, 1)^2$, when T tends to infinity. Estimator \hat{A}_T is well defined under the assumption:

Assumption A.5 *Either:*

- i. *the criterion Q_T is continuous and the set Θ is compact with respect to the norm $\|\cdot\|_{L^2(\nu)}$; or*
- ii. *the criterion Q_T is weakly lower semicontinuous and the set Θ is bounded and closed with respect to the norm $\|\cdot\|_{L^2(\nu)}$.*¹⁹

¹⁹Let $(X, \|\cdot\|)$ be a normed linear space. A sequence $(x_n) \subset X$ converges weakly to $x \in X$, noted $x_n \xrightarrow{w} x$, if for every linear functional l in the dual space X^* : $l(x_n) \rightarrow l(x)$. A function Φ on X is weakly lower semicontinuous (w.l.s.c.) if: $x_n \xrightarrow{w} x$ implies $\Phi(x) \leq \liminf \Phi(x_n)$. Assume that the space X is reflexive, that is the bidual space X^{**} is in one-to-one relationship with X under the canonical isomorphism (this is the case if X is an Hilbert space). Let function Φ be w.l.s.c., and let $M \subset X$ be closed and bounded. Then function Φ reaches a minimum over M [see Theorem S.6 of Reed, Simon (1980), p. 356].

The constrained estimator of the bivariate density is given by:

$$\hat{f}_T^0(x, y) = f(x, y; \hat{A}_T). \quad (3.22)$$

The aim of this paper is not to discuss the practical implementation of a nonparametric minimum chi-square estimator, but to prove the existence of nonparametrically efficient estimators. However, even if the optimization problem involves functionals, it is interesting to note that in practice two general approaches can be followed.

- i) We can optimize over a finite dimensional vector space of functions A and then use the usual optimization software. When the dimension of the space tends to infinity sufficiently fast with T the asymptotic properties of the estimator will be the same.
- ii) Since the derivative of the chi-square criterion are related to the information operator, and explicit expressions of the derivative are available for the examples, we can compute recursively the solution by a Newton-Raphson type algorithm, or apply a step of the algorithm from a consistent, but inefficient functional estimator.

Finally, some concentration with respect to a part of the functional parameters is possible on some examples.

Remark: *The chi-square measure is invariant by one to one transformation Φ of the basic variables X^*, Y^* . Thus it is equivalent to minimize a chi-square distance between f and \hat{f} or a distance between f^* and the transformation of \hat{f} by Φ . Similarly, the information operators corresponding to the families induced by f and f^* are the same. However it can be noted that the transformation of the kernel estimator of f is not a kernel estimator of f^* .*

3.4.2 Consistency of the estimators

Let us consider the consistency of the minimum chi-square estimator \hat{A}_T . In Appendix 6 it is shown that under the following two assumptions and additional regularity conditions (see Assumptions A.8 - A.11 in Appendix 6), Q_T converges to the chi-square proximity measure Q , uniformly in $A \in \Theta$, and that Q is continuous.

Assumption A.6 *There exists compact sets $\tilde{\Omega}_T, \Omega_T$ such that $\tilde{\Omega}_T \subset \Omega_T \subset [0, 1]^2$, ω_T has support in Ω_T , is smaller than 1 with restriction $\omega_T|_{\tilde{\Omega}_T} = 1$, $T \in \mathbb{N}$, and $\lambda_2(\tilde{\Omega}_T) \rightarrow 1$, as $T \rightarrow \infty$, where λ_2 is the Lebesgue measure.*

Assumption A.7 *$D \log f(X, Y; A)$ is a bounded operator from $L^2(\nu)$ in $L^2(P_0)$, for any $A, A_0 \in \Theta$.*

In particular, under Assumption A.7, the information operator I_A at A , defined by:

$$E_0 [\langle D \log f(X, Y; A), g \rangle \langle D \log f(X, Y; A), h \rangle] = (g, I_A h)_{L^2(\nu)},$$

for $h, g \in L^2(\nu)$, is a bounded operator from $L^2(\nu)$ in itself, for any $A, A_0 \in \Theta$.

We have the following proposition.

Proposition 3.8 : Under Assumptions A.1-A.11 ²⁰ the chi-square estimator \hat{A}_T is consistent in norm:

$$\left\| \hat{A}_T - A_0 \right\|_{L^2(\nu)} \xrightarrow{p} 0.$$

Let us now consider the constrained density estimator \hat{f}_T^0 , and show its consistency in L^1 -norm²¹. Convergence of \hat{A}_T to A_0 and continuity of Q implies convergence of $Q(\hat{A}_T)$ to $Q(A_0) = 0$. By using the Cauchy-Schwarz inequality:

$$\left\| f(\cdot, \cdot; \hat{A}_T) - f(\cdot, \cdot) \right\|_{L^1} \leq \left\| \frac{f(\cdot, \cdot; \hat{A}_T) - f(\cdot, \cdot)}{\sqrt{f(\cdot, \cdot)}} \right\|_{L^2} \left\| \sqrt{f(\cdot, \cdot)} \right\|_{L^2} = Q(\hat{A}_T)^{1/2},$$

we deduce the following proposition.

Proposition 3.9 : Under Assumptions of Proposition 8, the constrained density estimator \hat{f}_T^0 is consistent in L^1 norm:

$$\left\| \hat{f}_T^0 - f \right\|_{L^1} \xrightarrow{p} 0.$$

3.4.3 Asymptotic expansion of the minimum chi-square estimator

In this section we derive asymptotic expansions of the minimum chi-square estimator. We assume that the minimum chi-square estimator satisfies the first order condition in the following sense²².

Assumption A.12 For any $g \in L^2(\nu)$: $\hat{A}_T + tg \in \Theta$ with probability approaching to 1, for t in a neighborhood of 0 small enough.

Then it is possible to derive a set of first order conditions along the one-dimensional paths defined in Assumption A.12. The expansion of the first order condition satisfied by the minimum chi-square estimator is performed in Appendix 8 under additional regularity conditions (Assumptions A.13-A.15) described in this Appendix.

²⁰We assume that either A.3 i. and A.5 i., or A.3 ii. and A.5 ii. hold.

²¹Let $\Omega \subset [0, 1]^2$ be λ_2 -measurable. We denote by $L^p(\Omega)$, $p \geq 1$, the space of p -integrable functions with respect to the Lebesgue measure restricted on Ω , and $L^p \equiv L^p([0, 1]^2)$.

²²This assumption is immediately satisfied when A_0 is an interior point of Θ , in the sense that a $L^2(\nu)$ -ball $B_r(A_0)$ centered at A_0 is contained in Θ . This is typically the case under Assumption A.5 ii.

Proposition 3.10 : *Under Assumptions A.1-A.15 the minimum chi-square estimator \hat{A}_T is such that:*

$$I\left(\hat{A}_T - A_0\right) \simeq \psi_T, \quad (3.23)$$

where the efficient score $\psi_T \in L^2(\nu)$ is defined by²³:

$$(\psi_T, h)_{L^2(\nu)} = \int \int \delta \hat{f}_T(x, y) \omega_T(x, y) \langle D \log f(x, y; A_0), h \rangle dx dy, \quad h \in L^2(\nu),$$

where $\delta \hat{f}_T = \hat{f}_T - f$.

As an example, when the differential operator is of the form (3.3), function ψ_T is given by:

$$\begin{aligned} \frac{d\nu}{d\lambda}(w) \psi_T(w) &= \int \delta \hat{f}_T(w, y) \omega_T(w, y) \gamma_0(w, y) dy + \int \delta \hat{f}_T(x, w) \omega_T(x, w) \gamma_1(x, w) dx \\ &+ \int \int \delta \hat{f}_T(x, y) \omega_T(x, y) \gamma_2(x, y, w) dx dy. \end{aligned} \quad (3.24)$$

Moreover when the information operator admits the representation (3.2), the first order condition is equivalent to:

$$\alpha_0(w) \delta \hat{A}_T(w) + \int \alpha_1(w, v) \delta \hat{A}_T(v) dv \simeq \frac{d\nu}{d\lambda}(w) \psi_T(w), \quad (3.25)$$

where $\delta \hat{A}_T = \hat{A}_T - A_0$. To deduce the asymptotic expansion of the estimator itself, we have to assume that the information operator is invertible and that its inverse is continuous at zero [see section 2.3 iii) for sufficient conditions].

Corollary 3.11 : *When I is invertible and continuous at zero:*

$$\hat{A}_T - A_0 \simeq I^{-1} \psi_T. \quad (3.26)$$

Since $I = D \log f_0^* D \log f_0$ and $\psi_T = D \log f_0^* \left(\omega_T \delta \hat{f}_T / f \right)$, where $D \log f_0^*$ denotes the adjoint of the differential operator $D \log f_0 \equiv D \log f(\cdot, \cdot; A_0)$, the asymptotic expansion in (3.26) can be written as:

$$\hat{A}_T - A_0 \simeq [D \log f_0^* D \log f_0]^{-1} D \log f_0^* \left(\omega_T \delta \hat{f}_T / f \right),$$

that is a regression of the "errors" $\delta \hat{f}_T / f$ on the score $D \log f_0$.

Let us finally consider the expansion of the constrained estimator of the density [see Appendix 8, v)]:

Proposition 3.12 : *The constrained estimator is such that:*

$$\hat{f}_T^0(x, y) - f(x, y) \simeq \left\langle D f(x, y; A_0), \delta \hat{A}_T \right\rangle.$$

²³The differential operator $D \log f(x, y; A_0)$ smoothed by the kernel density estimator, that is $\int \int \delta \hat{f}_T(x, y) \omega_T(x, y) D \log f(x, y; A_0) dx dy$ becomes a linear functional on $L^2(\nu)$. Function $\psi_T \in L^2(\nu)$ corresponds to the Riesz representation of this functional. See Appendix 7.

3.4.4 The asymptotic distribution of the minimum chi-square estimator

The asymptotic distribution of the minimum chi-square estimator \hat{A}_T is derived from the asymptotic expansion given in Corollary 11. To simplify the presentation we assume decomposition of both differential and information operators [see section 2.3 i)]. We distinguish the pointwise estimation of A and the estimation of linear functionals of A , such as $\int_0^1 g(w)' \hat{A}_T(w) \nu(dw)$, for which different orders are expected $1/\sqrt{Th_T}$ and $1/\sqrt{T}$, respectively.

i) Pointwise estimation

To give some intuition on the asymptotic distribution let us consider equation (3.25), and assume $\alpha_0(w)$ is invertible for any w . For pointwise estimation, the second term of order $1/\sqrt{T}$ can be neglected leading to [see Appendix 8 iv)]:

$$\sqrt{Th_T} \delta \hat{A}_T(w) \simeq \alpha_0(w)^{-1} \sqrt{Th_T} \frac{d\nu}{d\lambda}(w) \psi_T(w).$$

When the differential operator admits the representation (3.3) we directly deduce from (3.24), (3.19), and (3.20) that $\sqrt{Th_T} \psi_T(w)$ is pointwise asymptotically normal (see Appendix 9²⁴).

Lemma 3.13 : *When the differential admits the decomposition (3.3):*

$$\sqrt{Th_T} \frac{d\nu}{d\lambda}(w) \psi_T(w) \xrightarrow{d} N \left[0, \left(\int K^2(x) dx \right) \alpha_0(w) \right], \quad \lambda\text{-a.s. in } w.$$

The asymptotic distribution of \hat{A}_T follows.

Proposition 3.14 : *Under Assumptions A.1-A.15 the estimator \hat{A}_T is λ -a.s. pointwise asymptotically normal:*

$$\sqrt{Th_T} \left(\hat{A}_T(w) - A_0(w) \right) \xrightarrow{d} N \left(0, \left(\int K^2(x) dx \right) \alpha_0(w)^{-1} \right),$$

λ -a.s. in w .

The intuition beyond this result is the following: since functionals of A converge at a parametric rate $1/\sqrt{T}$ (see below), for pointwise estimation we can neglect differentiation of those parts of the density which depend on functionals of A . The relevant component of the information operator is the local component α_0 , and the asymptotic variance of the estimator is essentially its inverse.

When the differential operator admits the representation (3.3), the asymptotic variance is given by:

$$\left(\int K^2(x) dx \right) \left(E \left[\gamma_{0,t} \gamma'_{0,t} \mid X_t = w \right] f_X(w) + E \left[\gamma_{1,t} \gamma'_{1,t} \mid Y_t = w \right] f_Y(w) \right)^{-1}.$$

Finally we get from Proposition 12 the asymptotic distribution of the constrained estimator.

²⁴Representation (3.3) is valid in example i), iv) and vii) in section 3.2. It is possible to extend the result to more general cases including the other examples.

Corollary 3.15 : *The constrained estimator $\sqrt{Th_T} \left(\hat{f}_T^0(x, y) - f(x, y) \right)$ is asymptotically normal, with asymptotic variance:*

$$\left(\int K^2(x) dx \right) f(x, y)^2 \left[\gamma_0(x, y) \alpha_0(x)^{-1} \gamma_0(x, y)' + \gamma_1(x, y) \alpha_0(y)^{-1} \gamma_1(x, y)' \right].$$

In particular the constrained estimator has a one-dimensional nonparametric convergence rate, and:

$$\begin{aligned} \sqrt{Th_T^2} \left[\hat{f}_T(x, y) - \hat{f}_T^0(x, y) \right] &\simeq \sqrt{Th_T^2} \left[\hat{f}_T(x, y) - f(x, y) \right] \\ &\xrightarrow{d} N \left[0, f(x, y) \left(\int K^2(w) dw \right)^2 \right]. \end{aligned}$$

The discrepancy $\sqrt{Th_T^2} \left[\hat{f}_T(x, y) - \hat{f}_T^0(x, y) \right]$, x, y varying, between the unconstrained and the constrained estimators can be used as a basis for a (pointwise) misspecification test.

ii) Estimation of linear functional

Let us now consider the estimation of a linear functional $G = \int g(v)' A_0(v) \nu(dv)$, with $g \in L^2(\nu)$. We expect the estimator $\hat{G}_T = \int g(v)' \hat{A}_T(v) \nu(dv)$ to have a parametric rate, so that the second term of equation (3.25), which is of order $1/\sqrt{T}$, can no longer be neglected. We deduce from Corollary 11 [see also Appendix 8 iii):

$$\begin{aligned} \sqrt{T} \left(\hat{G}_T - G \right) &= \sqrt{T} \int g(v)' \delta \hat{A}_T(v) \nu(dv) = \sqrt{T} \left(g, \delta \hat{A}_T \right)_{L^2(\nu)} \\ &\simeq \sqrt{T} \left(g, I^{-1} \psi_T \right)_{L^2(\nu)}, \text{ from (3.26)} \\ &= \sqrt{T} \left(I^{-1} g, \psi_T \right)_{L^2(\nu)}, \text{ since } I^{-1} \text{ is self-adjoint on } L^2(\nu). \end{aligned}$$

The following Lemma provides the asymptotic distribution of $\sqrt{T} (g, \psi_T)_{L^2(\nu)}$, $g \in L^2(\nu)$.

Lemma 3.16 *For $g \in L^2(\nu)$:*

$$\sqrt{T} (g, \psi_T)_{L^2(\nu)} \xrightarrow{d} N \left[0, (g, Ig)_{L^2(\nu)} \right].$$

Proof. *We have:*

$$\begin{aligned} \sqrt{T} (g, \psi_T)_{L^2(\nu)} &= \sqrt{T} \int \int \delta \hat{f}_T(x, y) \omega_T(x, y) \langle D \log f(x, y; A_0), g \rangle dx dy \\ &\simeq \sqrt{T} \int \int \delta \hat{f}_T(x, y) \langle D \log f(x, y; A_0), g \rangle dx dy. \end{aligned}$$

By using (3.20), the latter expression is asymptotically normal. Its variance is given by:

$$\begin{aligned} \sigma^2(g) &= V_0 [\langle D \log f(X_t, Y_t; A), g \rangle] \\ &= E_0 [\langle D \log f(X_t, Y_t; A), g \rangle^2] \\ &= (g, Ig)_{L^2(\nu)}. \end{aligned}$$

■

The asymptotic distribution of a linear functional follows.

Proposition 3.17 *Under Assumptions A.1-A.15 the estimator $\hat{G}_T = \int g(v)' \hat{A}_T(v) \nu(dv)$ of a linear functional of A is asymptotically normal, with parametric rate of convergence:*

$$\sqrt{T} \left(\hat{G}_T - G \right) \xrightarrow{d} N \left(0, (g, I^{-1}g)_{L^2(\nu)} \right).$$

3.4.5 Time series framework.

The previous results are easily extended to the time series framework. We need some mixing condition.

Assumption A.4.TS *Process X_t , t varying, is strictly stationary, Markov, with transition distribution $f(x|y; A)$, and β -mixing coefficients such that: $\beta_k = O(k^{-\delta})$, $\delta > 1$. The support of the marginal p.d.f. is $[0, 1]$.*

Moreover the minimum chi-square estimator is now defined by minimizing a chi-square divergence between the conditional distribution in the family and its unconstrained kernel estimator:

$$\hat{A}_T = \arg \min_{A \in \Theta} Q_T(A) = \int_0^1 \int_0^1 \frac{\left[\hat{f}_T(x|y) - f(x|y; A) \right]^2}{\hat{f}_T(x|y)} \omega_T(x, y) \hat{f}_{Y,T}(y) dx dy. \quad (3.27)$$

We also need some assumptions similar to A.1-A.3, A.5-A.15, valid for the time series framework. They are deduced by considering the conditional distribution $f(x|y; A)$, instead of the joint one, and the conditional differential operator $D \log f(x|y; A)$. They are denoted by adding TS.

Proposition 3.18 : *Under Assumptions A.1.TS-A.11.TS the minimum chi-square estimator \hat{A}_T is consistent.*

The asymptotic expansion of the chi-square estimator in the time series framework is given by [see Appendix 10]:

$$I_{X|Y} \left(\hat{A}_T - A_0 \right) \simeq \tilde{\psi}_T,$$

where the function $\tilde{\psi}_T \in L^2(\nu)$ is defined by:

$$\left(\tilde{\psi}_T, h \right)_{L^2(\nu)} = E_0 \left[\frac{\delta \hat{f}_T(X|Y)}{f(X|Y)} \omega_T(X, Y) \langle D \log f(X|Y; A_0), h \rangle \right], \quad h \in L^2(\nu).$$

In particular, when the conditional information operator $I_{X|Y}$ admits a representation with $\tilde{\alpha}_0, \tilde{\alpha}_1$, say, the asymptotic expansion becomes:

$$\tilde{\alpha}_0(w) \delta \hat{A}_T(w) + \int \tilde{\alpha}_1(w, v) \delta \hat{A}_T(v) dv \simeq \frac{d\nu}{d\lambda}(w) \tilde{\psi}_T(w).$$

The asymptotic distribution of \widehat{A}_T is immediately deduced from that of $\widetilde{\psi}_T$:

$$\begin{aligned} \sqrt{Th_T} \frac{d\nu}{d\lambda}(w) \widetilde{\psi}_T(w) &\xrightarrow{d} N \left[0, \left(\int K^2(x) dx \right) \widetilde{\alpha}_0(w) \right], \quad \lambda\text{-a.s. in } w, \\ \sqrt{T} \left(g, \widetilde{\psi}_T \right)_{L^2(\nu)} &\xrightarrow{d} N \left[0, (g, I_{X|Y} g)_{L^2(\nu)} \right], \quad \text{for } g \text{ in } L^2(\nu). \end{aligned}$$

Note that the asymptotic variance $(g, I_{X|Y} g)_{L^2(\nu)} = V_0 [\langle D \log f(X_t | X_{t-1}; A_0), g \rangle]$ includes no cross-term, since $\langle D \log f(X_t | X_{t-1}; A_0), g \rangle$ is a martingale difference sequence.

We deduce:

Proposition 3.19 : *Under Assumptions A.1.TS-A.15.TS we have:*

$$\sqrt{Th_T} \left(\widehat{A}_T(v) - A_0(v) \right) \xrightarrow{d} N \left(0, \left(\int K^2(x) dx \right) \widetilde{\alpha}_0(v)^{-1} \right), \quad \lambda\text{-a.s. in } v,$$

and:

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

3.5 Nonparametric efficiency.

The aim of this section is to show that a minimum chi-square estimator is nonparametrically efficient. We first review the approach to derive the nonparametric efficiency bound.

3.5.1 Nonparametric efficiency bound

The nonparametric "efficiency bound" for functional A is defined in the usual way from the parametric efficiency bound. The idea is to consider continuous linear functionals of function A , such as $\int A(v)' g(v) \nu(dv)$, which can be consistently estimated at rate $1/\sqrt{T}$, and to construct the semi-parametric bound $B(g)$, say, for this parameter [see e.g. Severini, Tripathi (2001)].

More precisely the approach consists in two steps:

- i. First introduce a one dimensional parametric model $A(., \theta)$, and derive the Cramer-Rao lower bound $B_A(g, \theta)$ for $\int_0^1 A(v; \theta)' g(v) \nu(dv)$ in this model.
- ii. Then the nonparametric efficiency bound is defined by:

$$B_A(g) = \max B_A(g, \theta), \quad g \text{ varying},$$

where the maximization is performed on all possible parametric specifications $A(., \theta)$.

Since a parameter is defined up to an invertible transformation, for any parametric specification we can select the parameter θ such that:

$$\int A(v; \theta)' g(v) \nu(dv) = \theta.$$

In a neighbourhood of θ_0 , this condition is equivalent to:

$$\int g(v)' \frac{\partial A}{\partial \theta}(v; \theta_0) \nu(dv) = 1.$$

Then the nonparametric efficiency bound is given by:

$$\begin{aligned} B_A(g) &= \max B_A(g, \theta), \\ \text{s.t.} \quad & \int g(v)' \frac{\partial A}{\partial \theta}(v; \theta_0) \nu(dv) = 1, \end{aligned} \tag{3.28}$$

g varying, where maximization is performed over all parameterizations $A(., \theta)$.

Proposition 3.20 : *i) In the cross-sectional framework the nonparametric efficiency bound is given by:*

$$B_A(g) = (g, I^{-1}g)_{L^2(\nu)},$$

where:

$$(g, Ih)_{L^2(\nu)} = E_0 [\langle D \log f(X, Y; A_0), g \rangle \langle D \log f(X, Y; A_0), h \rangle].$$

ii) In the time series framework the nonparametric efficiency bound is given by:

$$B_A(g) = (g, I_{X|Y}^{-1}g)_{L^2(\nu)},$$

where:

$$(g, I_{X|Y}h)_{L^2(\nu)} = E_0 [\langle D \log f(X_t | X_{t-1}; A_0), g \rangle \langle D \log f(X_t | X_{t-1}; A_0), h \rangle].$$

Proof. : See Appendix 11. ■

3.5.2 Nonparametric efficiency of the minimum chi-square estimator

From Propositions 17 and 19, we immediately deduce that the estimator $\hat{G}_T = \int g(v)' \hat{A}_T(v) \nu(dv)$ reaches the nonparametric efficiency bound.

Corollary 3.21 : *The minimum chi-square estimator \hat{A}_T is nonparametrically efficient.*

The efficiency property of the minimum chi-square estimator is important in practice. Indeed a number of inefficient nonparametric estimation methods have been introduced for some specific copulas (see e.g. Genest, Rivest [1993] for archimedean copulas, Abdous, Ghoudi, Khoudraji [2000] and references therein for extreme value copulas). Similarly the usual estimator of the transformation in transformed regression model, based on the ratio of partial derivatives of the conditional distribution due to the nonparametric identification constraint suggested by Ridder [Ridder (1990)], is consistent, asymptotically normal [Horowitz (1996), Gorgens, Horowitz (1999)], but in general inefficient. However these inefficient nonparametric estimators can be used as a first step of a Newton-Raphson type algorithm.

3.6 Constrained estimation. Identifying restrictions.

Until now we have assumed that the functional parameter A is free to vary over an open ball of $L^2(\nu)$. However this condition is not met in some examples described in section 3. We consider therefore in this section the case of a constrained functional parameter. From the examples, two sources of constraints can be distinguished. First, when one component of A is a marginal distribution, f_Y say, this component satisfies the unit mass restriction $\int f_Y(y)dy = 1$. Second, some parameters may be not identified unless additional restrictions are imposed. This is the case for the copula parameter a in the proportional hazard and archimedean copulas [examples iv) and v)], since a and ka , where k is a positive constant, define the same copula. A possible identifying restriction is: $\int a(v)dv = 1$.

3.6.1 Restricted information operator.

Let us assume that functional parameter A satisfies n linear constraints:

$$\int A(v)' g_i(v) \nu(dv) = (A, g_i)_{L^2(\nu)} = k_i, \quad i = 1, \dots, n,$$

where $g_i \in L^2(\nu)$, $k_i \in \mathbb{R}$, $i = 1, \dots, n$. Let us denote by $\tilde{\mathcal{A}} \subset \mathcal{A}$ the subset of functional parameters satisfying these restrictions. The tangent space H of $\tilde{\mathcal{A}}$ at $A_0 \in \tilde{\mathcal{A}}$ does not depend on A_0 , has a finite codimension, and it is given by:

$$H = \{h \in L^2(\nu) : (h, g_i)_{L^2(\nu)} = 0, \quad i = 1, \dots, n\}.$$

The differential operator $D \log f(., .; A_0)$ can be restricted to the linear space $H \subset L^2(\nu)$, and we assume that $D \log f(., .; A_0) : H \rightarrow L^2(P_0)$ is a bounded operator. The information operator I_H is the bounded linear operator from H in itself defined by:

$$(g, I_H h)_{L^2(\nu)} = E_0 [\langle D \log f(X, Y; A_0), g \rangle \langle D \log f(X, Y; A_0), h \rangle], \quad h, g \in H.$$

Then the definitions of identification and decomposition of the information operator can be extended to the constrained framework.

i) Local identification

Let us introduce the following local identification condition:

Assumption A.3. *i. Local identification:*

$$\langle D \log f(X, Y; A_0), h \rangle = 0 \text{ } P_0\text{-a.s.}, \quad h \in H \implies h = 0.$$

Assumption A.3. i. is equivalent to the assumption that I_H has a zero null space or that I_H is positive, and implies that A_0 is locally identified over any sufficiently small compact subset $\tilde{\Theta} \subset \tilde{\mathcal{A}}$ containing A_0 .

Local identification over non-compact subsets requires a stronger assumption:

Assumption A.3. *ii. Local identification:*

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

ii) Decomposition of the information operator

When the information operator I_H admits a decomposition:

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

it is possible to characterize boundedness and invertibility of I_H in terms of $\alpha_{0,H}$ and $\alpha_{1,H}$ ²⁵ [see Appendix 12].

Proposition 3.22 :

i. Assume that for any A there exists a positive definite matrix $\alpha_H(., A)$ such that:

$$\int \int \left\| \alpha_H(x; A)^{-1/2} \alpha_{1,H}(x, y; A) \alpha_H(y; A)^{-1/2} \right\|^2 dx dy < \infty, \forall A, \quad (3.29)$$

where $\|\cdot\|$ is a matrix norm on $\mathbb{R}^{q \times q}$. Let the measure ν be such that:

$$\forall A : \exists C_A > 0 : C_A \frac{d\nu}{d\lambda}(v) Id_q \geq \max \{ \alpha_{0,H}(v; A), \alpha_H(v; A) \}, \quad \forall v. \quad (3.30)$$

Then I_H is a bounded operator from H in itself.

ii. Assume further that $\alpha_{0,H}(v; A)$ is invertible, $\forall v, \forall A$, and such that:

$$\forall A : \exists \tilde{C}_A > 0 : \tilde{C}_A \frac{d\nu}{d\lambda}(v) Id_q \leq \alpha_{0,H}(v; A), \quad \forall v. \quad (3.31)$$

Assume finally that I_H has a zero null space. Then I_H is invertible, with bounded inverse.

Let us consider the example of the proportional hazard copula [example iv) in section 3.2]. The functional parameter a of the copula is subject to the identifying constraint: $\int_0^1 a(v) dv = 1$. The corresponding tangent space H is given by:

$$H = \left\{ h \in L^2(\nu) : \int_0^1 h(v) dv = 0 \right\}.$$

²⁵Let $I : L^2(\nu) \rightarrow L^2(\nu)$ denote the unrestricted information operator defined by the differential $D \log f(., .; A_0)$ with domain $L^2(\nu)$. Since $I_H = P_H I P_H = I - P_{H^\perp} I - I P_{H^\perp} - P_{H^\perp} I P_{H^\perp}$, where P_H (resp. P_{H^\perp}) denotes the orthogonal projector on H (resp. H^\perp), and H^\perp has finite dimension, it follows that I_H has a decomposition of the form (3.2) if I has such a decomposition. Moreover, both decompositions have identical local component: $\alpha_{0,H} = \alpha_0$.

Boundedness and invertibility of the copula information operator I_H^c on H is discussed in Gagliardini, Gourieroux (2002b) using Proposition 22. Let us for instance show that I_H^c has a zero null space on H . 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:

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

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^c has a zero null space and is a positive operator. The copula information operator is not invertible when defined on the entire space $L^2(\nu)$, since the differential $D \log c(., .; a_0)$ has a non zero null space, consisting in functions ka_0 , where k is a constant.

3.6.2 The minimum chi-square estimator.

Let $\tilde{\Theta}$ be a subset of $\tilde{\mathcal{A}}$. The minimum chi-square estimator is obtained by minimizing the chi-square divergence under the constraints:

$$\hat{A}_T = \arg \min_{A \in \tilde{\Theta}} Q_T(A) = \int_0^1 \int_0^1 \frac{[\hat{f}_T(x, y) - f(x, y; A)]^2}{\hat{f}_T(x, y)} \omega_T(x, y) dx dy. \quad (3.32)$$

The consistency of the constrained estimator is proved in complete analogy with section 4. Here we focus on the asymptotic expansion. We modify Assumption A.12 and assume that \hat{A}_T satisfies the first order condition in the sense that $\hat{A}_T + th \in \tilde{\Theta}$ with probability approaching to 1, for t small enough, for any $h \in H$. The first order condition is equivalent to (see Appendix 12):

$$(h, I_H \delta \hat{A}_T - \psi_T) \simeq 0, \forall h \in H,$$

that is:

$$I_H \delta \hat{A}_T \simeq P_H \psi_T.$$

If the operator I_H is invertible, the asymptotic expansion of \hat{A}_T is:

$$\hat{A}_T - A_0 \simeq I_H^{-1} P_H \psi_T.$$

By using:

$$\sqrt{T} (h, \psi_T)_{L^2(\nu)} \xrightarrow{d} N \left[0, (h, I_H h)_{L^2(\nu)} \right], \quad h \in H,$$

we get (see Appendix 12):

Proposition 3.23 : *Under Assumptions A.1-A.15:*

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

When the differential operator admits a decomposition (3.3):

$$\sqrt{Th_T} \left(\hat{A}_T(v) - A_0(v) \right) \xrightarrow{d} N \left(0, \left(\int K^2(x) dx \right) \alpha_{0,H}(v)^{-1} \right),$$

λ -a.s in v .

3.6.3 The nonparametric efficiency bound.

The following proposition reports the efficiency bound $B_A(g)$ for linear functionals $(g, A)_{L^2(\nu)}$, $g \in L^2(\nu)$, under the constraint $A \in \tilde{\mathcal{A}}$.

Proposition 3.24 : *The nonparametric efficiency bound is given by:*

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

The constrained minimum chi-square estimator is therefore nonparametrically efficient.

3.7 Concluding remarks.

The analysis of nonlinear dependence is crucial for financial applications and requires an appropriate specification of the joint density for often a rather large dimension. To avoid the curse of dimensionality and to select models with structural interpretations the density cannot be let unconstrained. At the opposite a standard parametric specification is generally too restrictive to get the expected fit. In this paper we have considered the intermediate case in which the conditional distribution or the copula depends on one-dimensional functional parameters. The functional parameter is defined up to a one to one transformation. We have explained what representation of the functional parameter has to be selected to get results on the information operator, efficiency bound, and efficient estimators similar to the standard results of the pure parametric framework. The approach has been illustrated by discussing different families of constrained nonparametric densities, useful for financial and duration analysis.

REFERENCES

- Abdous, B., Ghoudi, K., and A., Khoudraji, (2000): "Nonparametric Estimation of the Limit Dependence Function of Multivariate Extremes", *Extremes*, 2, 245-268.
- Aït-Sahalia, Y., (1993): "The Delta and Bootstrap Methods for Nonparametric Kernel Functionals", MIT Discussion Paper.
- Back, K., and D., Brown, (1992): "GMM, Maximum Likelihood, and Nonparametric Efficiency", *Economic Letters*, 39, 23-28.
- Back, K., and D., Brown, (1993): "Implied Probabilities in GMM Estimators", *Econometrica*, 61, 971-975.
- Bec, F., Salem, M. B., and M., Carrasco, (2001): "Tests for Unit-Root versus Threshold Specifications with an Application to the PPP", CREST, Université de Paris 1 and University of Rochester, Working Paper.
- Bec, F., Salem, M. B., and M., Carrasco, (2002): "The Purchasing Power Parity Puzzle: Evidence from a LSTAR Specification", CREST, Université de Paris 1 and University of Rochester, Working Paper.
- Begun, J. M., Hall, W. J., Huang, W.-M., and J. A., Wellner, (1983): "Information and Asymptotic Efficiency in Parametric-Nonparametric Models", *Annals of Statistics*, 11, 432-452.
- Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and J. A., Wellner, (1993): *Efficient and Adaptive Estimation in Semiparametric Models*, Johns Hopkins University Press, Baltimore.
- Bickel, P. J., and M., Rosenblatt, (1973): "On Some Global Measures of Deviations of Density Function Estimate", *Annals of Statistics*, 1, 1071-1095.
- Billingsley, P., (1968): *Convergence of Probability Measures*, Wiley, New-York.
- Chamberlain, G., (1987): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions", *Journal of Econometrics*, 34, 305-334.
- Debnath, L., and P., Mikusinski, (1998): *Introduction to Hilbert Spaces with Applications*, 2nd ed., Academic Press.
- Deheuvels, P., (1981): "A Nonparametric Test for Independence", *Pub. Instit. Stat. Univ. Paris*, 26, 29-50.
- Duffie, D., and K., Singleton, (1999): "Simulating Correlated Defaults", Stanford University, Working Paper.

Dunford, N., and J., Schwartz, (1968): *Linear Operators: Part I*, Wiley, New York.

Durrleman, V., Nikeghbali, A., and T., Roncalli, (2000): "How to Get Bounds for Distribution Convolution? A Simulation Study and an Application to Risk Management", GRO Crédit Lyonnais, Working Paper.

Embrechts, P., Höing, A., and A., Juri, (2001): "Using Copulae to Bound the Value-at-Risk for Functions of Dependent Risks", forthcoming in *Finance and Stochastics*.

Fermanian, J. D., and Scaillet, O., (2002): "Nonparametric Estimation of Copulas for Time Series", Working Paper.

Gagliardini, P., and C., Gouriéroux, (2002a): "Constrained Nonparametric Dependence with Applications to Finance and Insurance", CREST Working Paper.

Gagliardini, P., and C., Gouriéroux, (2002b): "Duration Time Series Models with Proportional Hazard", CREST Working Paper.

Genest, C., (1987): "Frank's Family of Bivariate Distributions", *Biometrika*, 74, 549-555.

Genest, C., Ghoudi, K., and L.-P., Rivest, (1995): "A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions", *Biometrika*, 82, 543-552.

Genest, C., and R. J., Mc Kay, (1986): "Copules Archimédiennes et familles de lois bidimensionnelles dont les marges sont données", *Can. J. Statistics*, 14, 145-159.

Genest, C., and L.-P., Rivest, (1993): "Statistical Inference Procedures for Bivariate Archimedean Copulas", *Journal of the American Statistical Association*, 88, 1034-1043.

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

Gill, R. D., and A. W., van der Vaart, (1993): "Non- and Semi-parametric Maximum Likelihood Estimators and the von Mises Method-II", *Scandinavian Journal of Statistics*, 20, 271-288.

Gorgens, T., and J., Horowitz, (1999): "Semi-parametric Estimation of a Censored Regression Model with an Unknown Transformation of the Dependent Variable", *Journal of Econometrics*, 90, 155-191.

Gouriéroux, C., and J., Jasiak, (2001): "State-space Models with Finite Dimensional Dependence", *Journal of Time Series Analysis*, 22, 665-678.

Gourieroux, C., and J., Jasiak, (2002): "Nonlinear Autocorrelograms: An Application to Intertrade Durations", *J. of Time Series Analysis*, 33, 127-154.

Gourieroux, C., and A., Monfort, (2002a): "Equidependence in Qualitative and Duration Models with Application to Credit Risk", CREST Working Paper.

Gourieroux, C., and A., Monfort, (2002b): "Term Structure in Duration Models", CREST Working Paper.

Gourieroux, C., and C. Y., Robert, (2001): "Tails and Extremal Behaviour of Stochastic Unit Root Models", CREST Working Paper.

Han, A., (1987)a: "Nonparametric Analysis of a Generalized Regression Model", *Journal of Econometrics*, 35, 305-316.

Han, A. (1987)b: "A Nonparametric Analysis of Transformations", *Journal of Econometrics*, 35, 191-209.

Härdle, W., (1990): *Applied Nonparametric Regression*, Econometric Society Monographs, Cambridge University Press.

Härdle, W., and T., Stocker, (1989): "Investigating Smooth Multiple Regression by the Method of Average Derivatives", *JASA*, 84, 986-995.

Härdle, W., and A., Tsybakov, (1997): "Local Polynomial Estimation of the Volatility Function", *Journal of Econometrics*, 81, 223-242.

Hastie, T.J., and R.J., Tibshirani, (1990): *Generalized Additive Models*, Chapman and Hall, New York.

Holly, A., (1995): "A Random Linear Functional Approach to Efficiency Bounds", *Journal of Econometrics*, 65, 235-261.

Horowitz, J., (1996): "Semiparametric Estimation of a Regression Model with an Unknown Transformation of the Dependent Variable", *Econometrica*, 64, 103-137.

Horowitz, J., and W., Härdle (1996): "Direct Semiparametric Estimation of Single Index Models with Discrete Covariates", *JASA*, 91, 1632-1640.

Ichimura, M., (1993): "Semi-parametric Least-Squares and Weighted SLS Estimation of Single Index Models", *Journal of Econometrics*, 58, 71-120.

Joe, H., (1997): *Multivariate Models and Dependence Concepts*, Monographs on Statistics and Applied Probability, 73, Chapman & Hall.

Kitamura, Y., (1997): "Empirical Likelihood Methods with Weakly Dependent Processes", *Annals of Statistics*, 25, 2084-2103.

Koshevnik, Y., and B., Levit, (1976): "On a Nonparametric Analogue of the Information Matrix", *Theory of Probability and its Applications*, 21, 738-753.

Lancaster, H., (1968): "The Structure of Bivariate Distributions", *Annals of Mathematical Statistics*, 29, 719-736.

Li, D. X., (2000): "On Default Correlation: a Copula Function Approach", *Journal of Fixed Income*, 9, 43-54.

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

Oakes, D., (1986): "Semiparametric Inference in a Model for Association in Bivariate Survival Data", *Biometrika*, 73, 353-361.

Oakes, D., (1994): "Multivariate Survival Distributions", *J. Nonparametric Stat.*, 3, 343-354.

Owen, A., (1990): "Empirical Likelihood Ratio Confidence Regions", *Annals of Statistics*, 18, 90-120.

Owen, A., (2001): *Empirical Likelihood*, Monographs on Statistics and Applied Probability, Chapman & Hall.

Parzen, E., (1962): "On Estimation of a Probability Density Function and the Mode", *Annals of Mathematical Statistics*, 33, 1065-1076.

Powell, J., Stock, J., and T., Stocker, (1989): "Semi-parametric Estimation of Index Coefficients", *Econometrica*, 57, 474-523.

Quin, J., and J., Lawless, (1994): "Empirical Likelihood and General Estimating Equations", *Annals of Statistics*, 22, 300-325.

Rahbek, A., and N., Shephard, (2002), "Autoregressive Conditional Root Model", University of Copenhagen and Nuffield College, Working Paper.

Reed, M., and B., Simon, (1980): *Methods of Modern Mathematical Physics. I: Functional Analysis*, Academic Press.

Ridder, G., (1990): "The Nonparametric Identification of Generalized Accelerated Failure Time Models", *Review of Economic Studies*, 57, 167-182.

Rosenblatt, M., (1956): "Remarks on Some Nonparametric Estimates of a Density Function", *Annals of Mathematical Statistics*, 27, 832-835.

Rudin, W., (1973): *Functional Analysis*, McGraw-Hill, New York.

Schönbucher, P. J., and D., Schubert, (2001): "Copula-Dependent Default Risk in Intensity Models", Working Paper, University of Bonn.

Scott, D., (1992): *Multivariate Density Estimation. Theory, Practice and Visualization*, Wiley.

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

Silverman, B., (1978): "Weak and Strong Consistency of the Kernel Estimate of a Density and its Derivatives", *Annals of Statistics*, 6, 177-184.

Silverman, B., (1986): *Density Estimation for Statistics and Data Analysis*, Monographs on Statistics and Applied Probability, 26, Chapman and Hall.

Sklar, A., (1959): "Fonctions de répartition à n dimensions et leurs marges", *Publ. Inst. Statist. Univ. Paris*, 8, 229-231.

Stone, C., (1983): "Optimal Uniform Rates of Convergence of Nonparametric Estimators of a Density Function or its Derivatives", in *Recent Advances in Statistics, Essays in Honor of Herman Chernoff's Sixtieth Birthday*, Academic Press, New-York.

Vaart, A. W. Van der, (1994): "Infinite Dimensional M-Estimation", *Prob. Theory and Math. Stat.*, 1, 1-13.

Vaart, A. W. Van der, and J. A., Wellner, (1996): *Weak Convergence and Empirical Processes*, Springer.

Van den Berg, G., (2001): "Duration Models: Specification, Identification, and Multiple Durations", in *Handbook of Econometrics, Volume V*, J. J. Heckman and E. Leamer eds., North-Holland, Amsterdam.

Von Mises, R., (1947): "On the Asymptotic Distribution of Differentiable Statistical Functions", *Annals of Mathematical Statistics*, 18, 309-348.

Yosida, K., (1995): *Functional Analysis*, Sixth Edition, Springer.

Appendix 1

The information operator

i) Definition.

Let us relate the definition of the information operator given in (3.1) with those normally adopted in the literature. For functions h such that $A_0(1+h)^2 \in \mathcal{A}$, denote by $f^{1/2}(h)$ the square root density:

$$f^{1/2}(h) = \left[\frac{f(., .; A_0(1+h)^2)}{f(., .; A_0)} \right]^{1/2} \in L^2(P_0).$$

Assume there exists a measure ν such that the mapping $f^{1/2} : L^2(\nu) \rightarrow L^2(P_0)$ is differentiable at $h = 0$, with continuous derivative:

$$df_0^{1/2} : L^2(\nu) \rightarrow L^2(P_0).$$

Then, following Begun, Hall, Huang, Wellner [1983], and Gill, Van der Vaart [1993], the information operator can be defined as:

$$I = df_0^{1/2*} df_0^{1/2} : L^2(\nu) \rightarrow L^2(\nu).$$

Operator I is bounded, nonnegative, self-adjoint, and satisfies:

$$E_0 \left[\left\langle df_0^{1/2}, g \right\rangle \left\langle df_0^{1/2}, h \right\rangle \right] = (g, Ih)_{L^2(\nu)}, \quad h, g \in L^2(\nu).$$

Under the differentiability Assumption A.2, $df_0^{1/2}$ is equal to the differential operator $D \log f(., .; A_0)$. Indeed:

$$\begin{aligned} f^{1/2}(th) &\simeq \left[\frac{f(., .; A_0(1+2th))}{f(., .; A_0)} \right]^{1/2} \simeq \left[1 + \frac{2t \langle Df(., .; A_0), h \rangle}{f(., .; A_0)} \right]^{1/2} \\ &\simeq 1 + t \langle D \log f(., .; A_0), h \rangle, \quad t \text{ small,} \end{aligned}$$

and the mapping $f^{1/2} : L^2(\nu) \rightarrow L^2(P_0)$ is differentiable at $h = 0$, with continuous derivative $df_0^{1/2} = D \log f(., .; A_0)$. The information operator reduces to $I = D \log f(., .; A_0)^* D \log f(., .; A_0)$, and satisfies:

$$E_0 [\langle D \log f(X, Y; A_0), g \rangle \langle D \log f(X, Y; A_0), h \rangle] = (g, Ih)_{L^2(\nu)}, \quad h, g \in L^2(\nu).$$

This is the definition adopted in our paper, and in Holly (1995).

ii) Representations by measures

The differential operator and the information operator can often be represented in terms of measure. We discuss here the link between the two representations.

The differential operator can generally be written in terms of a measure:

$$\langle D \log f(x, y, A), h \rangle = \int h(w)' \mu(x, y, A; dw), \quad (\text{a.1})$$

where $\mu(x, y, A; \cdot)$ is a q -vector of measures, $\forall x, y$. When this measure $\mu(x, y, A; \cdot)$ has both a discrete and a continuous part, we get for instance:

$$\begin{aligned} \langle D \log f(x, y, A), h \rangle &= \gamma_0(x, y, A)' h(x) + \gamma_1(x, y, A)' h(y) \\ &\quad + \int \gamma_2(x, y, w, A)' h(w) dw, \end{aligned} \quad (\text{a.2})$$

where γ_0, γ_1 and γ_2 are \mathbb{R}^q -valued functions, that is:

$$\mu(x, y, A; dw) = \gamma_0(x, y, A) \delta_x(dw) + \gamma_1(x, y, A) \delta_y(dw) + \gamma_2(x, y, w, A) \lambda(dw).$$

This corresponds to representation (3.3) in the paper. We can deduce the form of the information operator I when the differential operator $D \log f(X, Y; A_0)$ admits representation (a.1). We get:

$$(g, Ih)_{L^2(\nu)} = \int g(w)' Ih(w) \nu(dw),$$

where²⁶:

$$Ih(w) \nu(dw) = \int E_0 \left[\mu(X, Y; A_0; dw) \mu(X, Y; A_0; dv)' \right] h(v).$$

Ih is an \mathbb{R}^q -valued function in $L^2(\nu)$, and ν is a scalar measure.

A case of particular importance is when the measure μ is such that:

$$\begin{aligned} E_0 \left[\mu(X, Y; A_0; dw) \mu(X, Y; A_0; dv)' \right] &= \alpha_0(w, A_0) \delta_w(dv) \lambda(dw) \\ &\quad + \alpha_1(w, v, A_0) \lambda(dv, dw), \end{aligned}$$

where α_0 and α_1 are matrix-valued functions, such that $\alpha_0(w, A_0) = \alpha_0(w, A_0)'$, $\alpha_1(v, w, A_0) = \alpha_1(w, v, A_0)'$, $\forall v, w$. In this case the information operator is such that:

$$(g, Ih)_{L^2(\nu)} = \int g(w)' \alpha_0(w, A_0) h(w) dw + \int \int g(w)' \alpha_1(w, v, A_0) h(v) dv dw,$$

corresponding to representation (3.2) in the paper. This decomposition is valid for instance when the measure μ admits both a discrete and a continuous part as in (a.2) [resp. (3.3)], and the corresponding function α_0, α_1 are given in (3.4), (3.5) in the text.

iii) Choice of the measure ν

Let us prove Proposition 1. We have:

$$\begin{aligned} \|\langle D \log f(X, Y; A), h \rangle\|_{L^2(P_A)}^2 &= E_A [\langle D \log f(X, Y; A), h \rangle^2] \\ &= \int h(v)' \alpha_0(v, A) h(v) dv + \int \int h(v)' \alpha_1(v, w, A) h(w) dv dw. \end{aligned}$$

²⁶We assume that the integrals with respect to μ and P_0 can be commuted.

Both terms are easily bounded. For the first one we get:

$$\int h(v)' \alpha_0(v; A) h(v) dv \leq C_A \int h(v)' h(v) \nu(dv) = C_A \|h\|_{L^2(\nu)}^2.$$

Let us now consider the second one, and denote:

$$k_A = \left(\int \int \left\| \alpha(v; A)^{-1/2} \alpha_1(v, w; A) \alpha(w; A)^{-1/2} \right\|^2 dv dw \right)^{1/2} < \infty.$$

We get:

$$\begin{aligned} & \int \int h(v)' \alpha_1(v, w; A) h(w) dv dw \\ &= \int \int \left(\alpha(v; A)^{1/2} h(v) \right)' \left[\alpha(v; A)^{-1/2} \alpha_1(v, w; A) \alpha(w; A)^{-1/2} \right] \alpha(v; A)^{1/2} h(w) dv dw \\ &\leq \int \int \left\| \alpha(v; A)^{1/2} h(v) \right\| \left\| \alpha(v; A)^{-1/2} \alpha_1(v, w; A) \alpha(w; A)^{-1/2} \right\| \left\| \alpha(w; A)^{1/2} h(w) \right\| dv dw \\ &\leq \left(\int \int \left\| \alpha(v; A)^{-1/2} \alpha_1(v, w; A) \alpha(w; A)^{-1/2} \right\|^2 dv dw \right)^{1/2} \\ &\quad \left(\int \left\| \alpha(v; A)^{1/2} h(v) \right\|^2 dv \right), \text{ by applying twice Cauchy-Schwarz inequality,} \\ &= k_A \int h(v)' \alpha(v; A) h(v) dv \\ &\leq k_A C_A \int h(v)' h(v) \nu(dv) \\ &= k_A C_A \|h\|_{L^2(\nu)}^2. \end{aligned}$$

Thus:

$$\|\langle D \log f(X, Y; A), h \rangle\|_{L^2(P_0)}^2 \leq C_A (1 + k_A) \|h\|_{L^2(\nu)}^2,$$

and Proposition 1 is proved.

iv) Invertibility

Let us prove Proposition 2. The information operator can be decomposed in two components:

$$\begin{aligned} Ih(w) &= \frac{1}{d\nu/d\lambda(w)} \alpha_0(w; A_0) h(w) + \int \frac{1}{d\nu/d\lambda(w)} \alpha_1(w, v; A_0) h(v) dv \\ &\equiv I^0 h(w) + I^1 h(w). \end{aligned}$$

The invertibility of I is proved by using results on Fredholm operators, as in Van der Vaart (1994). In particular, let us consider the following Lemma [see e.g. Rudin (1973), p. 99-103].

Lemma A.1. *Let H be a Banach space. Let $I^0 : H \rightarrow H$ be a continuously invertible*

operator, and let $I^1 : H \rightarrow H$ be a compact operator. Assume that $I = I^0 + I^1$ has a zero null space. Then I is continuously invertible.

Let us verify that the conditions of this Lemma are satisfied by operators I^0 and I^1 defined above. In the previous paragraph it has been shown that they are both bounded operators of $L^2(\nu)$ into itself. In addition:

$$\begin{aligned} \|I^{0^{-1}}h\|_{L^2(\nu)}^2 &= \int \left(\frac{d\nu}{d\lambda}(v) \right)^2 h(v)' \alpha_0(v)^{-2} h(v) \nu(dv) \\ &\leq \tilde{C}_A^{-2} \int h(v)' h(v) \nu(dv) = \tilde{C}_A^{-2} \|h\|_{L^2(\nu)}^2; \end{aligned}$$

thus I^0 is continuously invertible. Let us now consider the operator I^1 . We have:

$$I^1 h(w) = \int K(w, v; A_0) h(v) \nu(dv),$$

where

$$K(w, v; A_0) = \frac{1}{d\nu/d\lambda(w)} \frac{1}{d\nu/d\lambda(v)} \alpha_1(w, v; A_0).$$

We have:

$$\begin{aligned} &\int \int \|K(w, v; A_0)\|^2 \nu(dw) \nu(dv) \\ &\leq \int \int \frac{\|\alpha(x; A)\| \left\| \alpha(x; A)^{-1/2} \alpha_1(x, y; A) \alpha(y; A)^{-1/2} \right\|^2 \|\alpha(x; A)\|}{(d\nu/d\lambda(x) d\nu/d\lambda(y))^2} \nu(dx) \nu(dy) \\ &\leq C_A^2 \int \int \frac{\left\| \alpha(x; A)^{-1/2} \alpha_1(x, y; A) \alpha(y; A)^{-1/2} \right\|^2}{d\nu/d\lambda(x) d\nu/d\lambda(y)} \nu(dx) \nu(dy) \\ &= C_A^2 \int \int \left\| \alpha(x; A)^{-1/2} \alpha_1(x, y; A) \alpha(y; A)^{-1/2} \right\|^2 dx dy < \infty. \end{aligned}$$

It then follows from Hilbert-Schmidt theorem [see e.g. a generalization of Theorem VI.23 in Reed, Simon (1980)] that I^1 is a compact operator. Then all conditions of Lemma A.1 are satisfied, and Proposition 2 is proved.

Appendix 2

Local Identification

i) Local equivalence of the minimum chi-square and Kullback proximity measures.

The Kullback proximity measure between $f(x, y; A)$ and $f(x, y)$ is defined by:

$$\mathcal{K}(A) = E_0 \log \left[\frac{f(X, Y)}{f(X, Y; A)} \right].$$

Its second order expansion in a neighbourhood of $A = A_0$ is:

$$\begin{aligned} \mathcal{K}(A) &= -E_0 \log \left[1 + \frac{f(X, Y; A) - f(X, Y)}{f(X, Y)} \right] \\ &\simeq -E_0 \left[\frac{f(X, Y; A) - f(X, Y)}{f(X, Y)} \right] + \frac{1}{2} E_0 \left[\left(\frac{f(X, Y; A) - f(X, Y)}{f(X, Y)} \right)^2 \right] \\ &= \frac{1}{2} Q(A). \end{aligned}$$

ii) Local expansion of the minimum chi-square proximity measure.

In Appendix 6 we will derive expansions of the minimum chi-square proximity measure. In particular, it will be shown that the expansion around A_0 is given by [see equation (a.8) in Appendix 6 ii)]:

$$\begin{aligned} Q(A_0 + h) &= (h, Ih)_{L^2(\nu)} + \int \int \frac{R(x, y; A_0, h)^2}{f(x, y)} dx dy \\ &\quad + O \left[\left(\int \int \frac{R(x, y; A_0, h)^2}{f(x, y)} dx dy \right)^{\frac{1}{2}} (h, Ih)_{L^2(\nu)}^{\frac{1}{2}} \right], \end{aligned}$$

where $R(x, y; A_0, h)$ denotes the residual term in the first-order expansion of the density: $f(x, y; A_0 + h) = f(x, y; A_0) + \langle Df(x, y; A_0), h \rangle + R(x, y; A_0, h)$. Let us assume:

Assumption A.2.bis. *For any $A_0 \in \mathcal{A}$, there exists a neighborhood \mathcal{N}_0 of A_0 such that:*

$$\int \int \frac{R(x, y; A_0, h)^2}{f(x, y)} dx dy = O \left(\|h\|_{L^2(\nu)}^4 \right), \quad A_0 + h \in \mathcal{N}_0.$$

We get:

$$Q(A_0 + h) = (h, Ih)_{L^2(\nu)} + r(h), \quad A_0 + h \in \mathcal{N}_0,$$

where $r(h) = O\left(\|h\|_{L^2(\nu)}^2 (h, Ih)_{L^2(\nu)}^{1/2}\right) = O\left(\|h\|_{L^2(\nu)}^3\right)$. In particular, Q is well-defined on \mathcal{N}_0 .

iii) Local identification over compact sets.

Let $\Theta \subset \mathcal{N}_0$ be a compact set containing A_0 . Let us first give an upper bound for the residual term $r(h)$. For h such that $A_0 + h \in \Theta$ we have:

$$\frac{|r(h)|}{(h, Ih)_{L^2(\nu)}} \leq \frac{C \|h\|_{L^2(\nu)}^2}{(h, Ih)_{L^2(\nu)}^{1/2}},$$

for some constant C .

Assumption A.3*:

$$\inf_{\substack{h \in (\Theta - A_0) \\ h \neq 0}} \frac{1}{\|h\|_{L^2(\nu)}^2} \frac{(h, Ih)_{L^2(\nu)}}{\|h\|_{L^2(\nu)}^2} > 4C^2.$$

Thus: $|r(h)| \leq \frac{1}{2} (h, Ih)_{L^2(\nu)}$, $h \in \Theta - A_0$, and we get:

$$Q(A_0 + h) = (h, Ih)_{L^2(\nu)} + r(h) \geq \frac{1}{2} (h, Ih)_{L^2(\nu)}, \quad h \in \Theta - A_0.$$

Let us now show that A_0 is locally identified. We get:

$$Q(A_0 + h) \geq \frac{1}{2} (h, Ih)_{L^2(\nu)} > 0, \text{ for any } h \in \Theta - A_0, h \neq 0,$$

since I is positive, and

$$\begin{aligned} \inf_{A \in \Theta \setminus B_\varepsilon(A_0)} Q(A) &= \inf_{h \in (\Theta - A_0) \setminus B_\varepsilon(0)} Q(A_0 + h) \\ &\geq \frac{1}{2} \inf_{h \in (\Theta - A_0) \setminus B_\varepsilon(0)} (h, Ih)_{L^2(\nu)} \\ &= \frac{1}{2} (h^*, Ih^*)_{L^2(\nu)} > 0, \text{ say,} \end{aligned}$$

since $(\Theta - A_0) \setminus B_\varepsilon(0)$ is compact.

iv) Local identification over non-compact sets.

Let $\Theta \subset \mathcal{N}_0$ contain A_0 . Under Assumption A.3 ii., Assumption A.3* is immediately satisfied if Θ is small enough. Thus $r(h)$ can be bounded and for any $h \in \Theta - A_0$ we get:

$$Q(A_0 + h) \geq \frac{1}{2} (h, Ih)_{L^2(\nu)}.$$

Let us now show that A_0 is locally identified. We get:

$$Q(A_0 + h) \geq \frac{1}{2} (h, Ih)_{L^2(\nu)} > 0, \text{ for any } h \in \Theta - A_0, h \neq 0,$$

since I is positive, and:

$$\begin{aligned} \inf_{A \in \Theta \setminus B_\varepsilon(A_0)} Q(A) &= \inf_{h \in (\Theta - A_0) \setminus B_\varepsilon(0)} Q(A_0 + h) \\ &\geq \frac{1}{2} \inf_{h \in (\Theta - A_0) \setminus B_\varepsilon(0)} \|h\|_{L^2(\nu)}^2 \frac{(h, Ih)_{L^2(\nu)}}{\|h\|_{L^2(\nu)}^2} \\ &\geq \frac{1}{2} \varepsilon^2 \inf_{h \neq 0} \frac{(h, Ih)_{L^2(\nu)}}{\|h\|_{L^2(\nu)}^2} > 0. \end{aligned}$$

v) Equivalence of Assumption A.3 i. and the conditions on the information operator.

$ii) \implies i)$: Let $h \in L^2(\nu)$ such that $Ih = 0$. It follows $(h, Ih)_{L^2(\nu)} = 0$ and thus $h = 0$.

$i) \implies \text{A.3 i.}$: Let $h \in L^2(\nu)$ such that $\langle D \log f(X, Y; A_0), h \rangle = 0$ P_0 -a.s. Then for any $g \in L^2(\nu)$: $(g, Ih)_{L^2(\nu)} = 0$. It follows $Ih = 0$, and thus $h = 0$.

$\text{A.3 i.} \implies ii)$: Let $h \in L^2(\nu)$ such that $(h, Ih)_{L^2(\nu)} = 0$. Then $E_0 [\langle D \log f(X, Y; A_0), h \rangle^2] = 0$. Therefore $\langle D \log f(X, Y; A_0), h \rangle = 0$ P_0 -a.s., and thus $h = 0$.

Appendix 3

Differential of the copula and of the conditional and marginal densities

i) Proof of Proposition 3.

The first equation is clear. To derive the second one, let us differentiate the relationship:

$$f_Y(y; A) = \int f(x, y; A) dx.$$

We get:

$$\begin{aligned} f_Y(y; A + h) &= \int f(x, y; A + h) dx \simeq \int f(x, y; A) dx + \int \langle Df(x, y; A), h \rangle dx \\ &= f_Y(y; A) + \int \langle D \log f(x, y; A), h \rangle f(x, y; A) dx. \end{aligned}$$

Thus:

$$\langle D \log f_Y(y; A), h \rangle = \int \langle D \log f(x, y; A), h \rangle f_{X|Y}(x|y; A) dx.$$

ii) Proof of Proposition 5.

By taking the logarithm of the joint density we get:

$$\log f(x, y; A) = \log c[F_X(x; A), F_Y(y; A); A] + \log f_X(x; A) + \log f_Y(y; A).$$

Let us derive the expansion of the first term with respect to A . We have:

$$\begin{aligned} & \log c[F_X(x; A+h), F_Y(y; A+h); A+h] \\ \simeq & \log c[F_X(x; A) + \langle DF_X(x; A), h \rangle, F_Y(y; A) + \langle DF_Y(y; A), h \rangle; A+h] \\ \simeq & \log c[F_X(x; A), F_Y(y; A); A] \\ & + \frac{\partial \log c}{\partial u}[F_X(x; A), F_Y(y; A); A] \langle DF_X(x; A), h \rangle \\ & + \frac{\partial \log c}{\partial v}[F_X(x; A), F_Y(y; A); A] \langle DF_Y(y; A), h \rangle \\ & + \langle D \log c[F_X(x; A), F_Y(y; A); A], h \rangle. \end{aligned}$$

Thus the differential of $\log f(x, y; A)$ with respect to A is:

$$\begin{aligned} & \frac{\partial \log c}{\partial u}[F_X(x; A), F_Y(y; A); A] \langle DF_X(x; A), h \rangle \\ & + \frac{\partial \log c}{\partial v}[F_X(x; A), F_Y(y; A); A] \langle DF_Y(y; A), h \rangle \\ & + \langle D \log c[F_X(x; A), F_Y(y; A); A], h \rangle \\ & + \langle D \log f_X(x; A), h \rangle + \langle D \log f_Y(y; A), h \rangle. \end{aligned}$$

Moreover the differentials $DF_X(x; A)$ and $DF_Y(y; A)$ can be expressed by means of $D \log f_X(x; A)$ and $D \log f_Y(y; A)$, respectively. For instance:

$$\begin{aligned} F_X(x; A+h) &= \int_{-\infty}^x f_X(z; A+h) dz \\ &\simeq \int_{-\infty}^x [f_X(z; A) + \langle Df_X(z; A), h \rangle] dz \\ &= F_X(x; A) + \int_{-\infty}^x f_X(z; A) \langle D \log f_X(z; A), h \rangle dz, \end{aligned}$$

and thus:

$$\langle DF_X(x; A), h \rangle = \int_{-\infty}^x f_X(z; A) \langle D \log f_X(z; A), h \rangle dz.$$

The proposition follows.

Appendix 4 Examples

ii) Truncated dynamic models

Let us first derive the differential with respect to a . The first order expansion is given by:

$$\begin{aligned}\log f(x|y; a+h, \log g) &= \log g[x+a(y)+h(y)] - \log \left(\int_{a(y)+h(y)}^{+\infty} g(z) dz \right) \\ &\simeq \log f(x|y; a, \log g) + \frac{d \log g}{dz} [x+a(y)] h(y) \\ &\quad + \frac{g[a(y)]}{\int_{a(y)}^{+\infty} g(z) dz} h(y).\end{aligned}$$

Thus we get:

$$\langle D_a \log f(x|y; A) \rangle = \left(\frac{d \log g}{dz} [x+a(y)] + \lambda[a(y)] \right) h(y).$$

Let us now derive the differential with respect to $\log g$. The first order expansion is given by:

$$\begin{aligned}\log f(x|y; a, \log g+h) &= \log g[x+a(y)] + h[x+a(y)] - \log \left(\int_{a(y)}^{+\infty} g(z) e^{h(z)} dz \right) \\ &\simeq \log f(x|y; a, \log g) + h[x+a(y)] - \frac{\int_{a(y)}^{+\infty} g(z) h(z) dz}{\int_{a(y)}^{+\infty} g(z) dz},\end{aligned}$$

and thus:

$$\langle D_{\log g} \log f(x|y; A) \rangle = h[x+a(y)] - E_A[h(X_t + a(X_{t-1})) | X_{t-1} = y].$$

iii) Stochastic unit roots

Let us first compute the differential of $f(x|y; \pi, g)$ with respect to π and g . By linearity, we have:

$$\begin{aligned}\langle D_\pi f(x|y; \pi, g), h \rangle &= [g(x-y) - g(x)] h(y), \\ \langle D_g f(x|y; \pi, g), h \rangle &= \pi(y) h(x-y) + [1 - \pi(y)] h(x).\end{aligned}$$

Thus the differential of $\log f(x|y; A)$ with respect to $A = (\log \pi, \log g)$ is given by:

$$\begin{aligned}\langle D_{\log \pi} \log f(x|y; A), h \rangle &= \frac{[g(x-y) - g(x)] \pi(y)}{f(x|y; A)} h(y) \\ &= \frac{f(x|y; A) - g(x)}{f(x|y; A)} h(y) \\ &= r(x, y; A) h(y), \\ \langle D_{\log g} \log f(x|y; A), h \rangle &= \frac{\pi(y) g(x-y)}{f(x|y; A)} h(x-y) + \frac{[1 - \pi(y)] g(x)}{f(x|y; A)} h(x) \\ &= p_1(x, y; A) h(x-y) + p_0(x, y; A) h(x).\end{aligned}$$

Let us now derive the information operator. We compute separately each block. We get:

$$\begin{aligned}
\left(\tilde{h}, I_{\log \pi, \log \pi} h \right)_{L^2(\nu)} &= E_A \left[\left\langle D_{\log \pi} \log f(X | Y; A), \tilde{h} \right\rangle \left\langle D_{\log \pi} \log f(X | Y; A), h \right\rangle \right] \\
&= E_A \left[r(X, Y; A)^2 \tilde{h}(Y) h(Y) \right] \\
&= E_A \left[E_A \left[r(X, Y; A)^2 | Y \right] \tilde{h}(Y) h(Y) \right] \\
&= \int E_A \left[r(X, Y; A)^2 | Y = z \right] f_Y(z; A) \tilde{h}(z) h(z) dz, \\
\\
\left(\tilde{h}, I_{\log g, \log g} h \right)_{L^2(\nu)} &= E_A \left[\left\langle D_{\log g} \log f(X | Y; A), \tilde{h} \right\rangle \left\langle D_{\log g} \log f(X | Y; A), h \right\rangle \right] \\
&= E_A \left[E_A \left[p_1(X, Y; A)^2 | X - Y \right] \tilde{h}(X - Y) h(X - Y) \right] \\
&\quad + E_A \left[p_1(X, Y; A) p_0(X, Y; A) \tilde{h}(X - Y) h(X) \right] \\
&\quad + E_A \left[p_1(X, Y; A) p_0(X, Y; A) \tilde{h}(X) h(X - Y) \right] \\
&\quad + E_A \left[E_A \left[p_0(X, Y; A)^2 | X \right] \tilde{h}(X) h(X) \right] \\
&= \int E_A \left[p_1(X, Y; A)^2 | X - Y = z \right] f_{X-Y}(z) \tilde{h}(z) h(z) dz \\
&\quad + \int \tilde{h}(x) \left(\int p_0(z, z - x; A) p_1(z, z - x; A) f(z, z - x; A) h(z) dz \right) dx \\
&\quad + \int \tilde{h}(x) \left(\int p_0(x, x - z; A) p_1(x, x - z; A) f(x, x - z; A) h(z) dz \right) dx \\
&\quad + \int E_A \left[p_0(X, Y; A)^2 | X = z \right] f_X(z) \tilde{h}(z) h(z) dz,
\end{aligned}$$

and finally:

$$\begin{aligned}
\left(\tilde{h}, I_{\log \pi, \log g} h \right)_{L^2(\nu)} &= E_A \left[\left\langle D_{\log \pi} \log f(X | Y; A), \tilde{h} \right\rangle \left\langle D_{\log g} \log f(X | Y; A), h \right\rangle \right] \\
&= E_A \left[r(X, Y; A) \tilde{h}(Y) p_1(X, Y; A) h(X - Y) \right] \\
&\quad + E_A \left[r(X, Y; A) \tilde{h}(Y) p_0(X, Y; A) h(X) \right] \\
&= \int \tilde{h}(x) \left(\int [f(z + x | x; A) - g(z + x)] p_1(z + x, x; A) h(z) dz \right) f_Y(x) dx \\
&\quad + \int \tilde{h}(x) \left(\int [f(z | x; A) - g(z)] p_0(z, x; A) h(z) dz \right) f_Y(x) dx.
\end{aligned}$$

Thus the information operator admits a measure decomposition with:

$$\alpha_0(z; A_0) = \begin{pmatrix} E_0[r_t^2 | X_{t-1} = z] f(z) & 0 \\ 0 & E_0[p_{1,t}^2 | X_t - X_{t-1} = z] f_{X_t - X_{t-1}}(z) \\ & + E_0[p_{0,t}^2 | X_t = z] f(z) \end{pmatrix}$$

and:

$$\alpha_1(x, z; A_0) = \begin{pmatrix} 0 & f(x) \{ [f(x+z|x) - g(x+z)] p_1(x+z, x) \\ & + [f(z|x) - g(z)] p_0(z, x) \} \\ 0 & p_0(x, x-z) p_1(x, x-z) f(x, x-z) \end{pmatrix} \\ + (x \longleftrightarrow z)'.$$

v) Archimedean Copulas

a) Proof of Proposition 7.

The Jacobian of the transformation is:

$$\det \frac{\partial(w, z)}{\partial(u, v)} = \frac{\phi' [\phi^{-1}(u) + \phi^{-1}(v)]}{\phi' [\phi^{-1}(u)]} \equiv J(u, v).$$

Thus:

$$\frac{c(u, v)}{J(u, v)} = \frac{\phi'' \{ \phi^{-1} [C(u, v)] \}}{\phi' \{ \phi^{-1} [C(u, v)] \} \phi' [\phi^{-1}(v)]},$$

and the joint p.d.f. of W and Z is given by:

$$f(w, z) = \frac{\phi'' [\phi^{-1}(w)]}{\phi' [\phi^{-1}(w)] \phi' [\phi^{-1}(z)]} \mathbb{I}_{w \leq z}.$$

Let us define the function:

$$f^*(w) = -\frac{\phi'' [\phi^{-1}(w)]}{\phi' [\phi^{-1}(w)]} = -\frac{d}{dw} \phi' [\phi^{-1}(w)], \quad w \in [0, 1].$$

Since $\phi' [\phi^{-1}(0)] = \phi' [+\infty] = 0$, we have:

$$\phi' [\phi^{-1}(z)] = -\int_0^z f^*(v) dv = -F^*(z), \text{ say.}$$

Thus the joint p.d.f. of W and Z can also be written as:

$$f(w, z) = \frac{f^*(w)}{\int_0^z f^*(v) dv} \mathbb{I}_{w \leq z}.$$

Let us now show that ϕ and f^* are in one-to-one relationship. We have:

$$F^*(w) = -\phi' [\phi^{-1}(w)],$$

or equivalently:

$$-\frac{1}{F^*(w)} = \frac{d\phi^{-1}(w)}{dw}.$$

By integration, with $\phi^{-1}(1) = 0$:

$$\phi^{-1}(y) = \int_y^1 \frac{dv}{\int_0^v f^*(w)dw}, \quad y \in (0, 1).$$

Let us finally check that this function satisfies the properties of a (strict) archimedean generator. The properties $\phi^{-1}(1) = 0$ and $\phi^{-1}(0) = \infty$ are evident. Moreover:

$$\begin{aligned} \frac{d}{dy}\phi^{-1}(y) &= -\frac{1}{\int_0^y f^*(w)dw} \leq 0, \\ \frac{d^2}{dy^2}\phi^{-1}(y) &= \frac{f^*(y)}{(\int_0^y f^*(w)dw)^2} \geq 0, \end{aligned}$$

and thus ϕ^{-1} is decreasing and convex.

b) Differential of the copula.

The log copula density is given by:

$$\begin{aligned} \log c(u, v; a) &= \log a [C(u, v; a)] + \log F^* [C(u, v; a); a] \\ &\quad - \log F^* (u; a) - \log F^* (v; a), \end{aligned}$$

where:

$$a = f^*.$$

The general expression

Let us derive the differential with respect to a . We get:

$$\begin{aligned} \langle D \log c(u, v; a), h \rangle &= \frac{h [C(u, v; a)]}{a [C(u, v; a)]} + \frac{d \log a}{dw} [C(u, v; a)] \langle DC(u, v; a), h \rangle \\ &\quad + \langle D \log F^* [C(u, v; a); a], h \rangle + \frac{a [C(u, v; a)]}{F^* [C(u, v; a); a]} \langle DC(u, v; a), h \rangle \\ &\quad - \langle D \log F^* (u; a), h \rangle - \langle D \log F^* (v; a), h \rangle \\ &= \frac{h [C(u, v; a)]}{a [C(u, v; a)]} + \langle D \log F^* [C(u, v; a); a], h \rangle \\ &\quad - \langle D \log F^* (u; a), h \rangle - \langle D \log F^* (v; a), h \rangle \\ &\quad + \left(\frac{d \log a}{dw} [C(u, v; a)] + \frac{a [C(u, v; a)]}{F^* [C(u, v; a); a]} \right) \langle DC(u, v; a), h \rangle. \end{aligned} \tag{a.3}$$

Let us now derive the differentials of $C(u, v; a)$ and $F^*(u, v; a)$ with respect to a .

Differential of $C(u, v; a)$

We get:

$$\begin{aligned}
\langle DC(u, v; a), h \rangle &= \langle D\phi [\phi^{-1}(u; a) + \phi^{-1}(v; a); a], h \rangle \\
&\quad + \phi' [\phi^{-1}(u; a) + \phi^{-1}(v; a); a] \{ \langle D\phi^{-1}(u; a), h \rangle + \langle D\phi^{-1}(v; a), h \rangle \} \\
&= \langle D\phi (\phi^{-1}[C(u, v; a); a]; a), h \rangle \\
&\quad + \phi' (\phi^{-1}[C(u, v; a); a]; a) \{ \langle D\phi^{-1}(u; a), h \rangle + \langle D\phi^{-1}(v; a), h \rangle \}.
\end{aligned}$$

By the implicit function theorem we have:

$$\langle D\phi [\phi^{-1}(y; a); a], h \rangle = -\phi' [\phi^{-1}(y; a); a] \langle D\phi^{-1}(y; a), h \rangle,$$

and thus we get:

$$\begin{aligned}
\langle DC(u, v; a), h \rangle &= F^*[C(u, v; a); a] \{ \langle D\phi^{-1}[C(u, v; a); a], h \rangle \\
&\quad - \langle D\phi^{-1}(u; a), h \rangle - \langle D\phi^{-1}(v; a), h \rangle \}.
\end{aligned} \tag{a.4}$$

Differential of $F^*(y; a)$

Let us now derive the differential of $F^*(y; a)$. We get:

$$\begin{aligned}
\langle D \log F^*(y; a), h \rangle &= \frac{1}{F^*(y; a)} \int_0^y h(v) dv \\
&= E_a [h(W)/a(W) \mid Z = y].
\end{aligned} \tag{a.5}$$

By inserting (a.4) and (a.5) in (a.3) we get:

$$\begin{aligned}
\langle D \log c(u, v; a), h \rangle &= \frac{h[C(u, v; a)]}{a[C(u, v; a)]} + E_a [h(W)/a(W) \mid Z = C(u, v; a)] \\
&\quad - E_a [h(W)/a(W) \mid Z = u] - E_a [h(W)/a(W) \mid Z = v] \\
&\quad + \left\{ a[C(u, v; a)] + \frac{d \log a}{dw} [C(u, v; a)] F^*[C(u, v; a); a] \right\} \\
&\quad \cdot \{ \langle D\phi^{-1}[C(u, v; a); a], h \rangle - \langle D\phi^{-1}(u; a), h \rangle - \langle D\phi^{-1}(v; a), h \rangle \}.
\end{aligned} \tag{a.6}$$

Let us finally compute the differential of $\phi^{-1}(y; a)$ with respect to a .

Differential of $\phi^{-1}(y; a)$

We have:

$$\phi^{-1}(y; a) = \int_y^1 \frac{dw}{\int_0^w a(v) dv}.$$

Let us consider the first order expansion:

$$\begin{aligned}
 \phi^{-1}(y; a + h) &= \int_y^1 \frac{dw}{\int_0^w a(v) dv + \int_0^w h(v) dv} \\
 &\simeq \int_y^1 \frac{1}{\int_0^w a(v) dv} \left[1 - \frac{\int_0^w h(v) dv}{\int_0^w a(v) dv} \right] dw \\
 &\simeq \phi^{-1}(y; a) - \int_y^1 \frac{\int_0^w h(v) dv}{F^*(w; a)^2} dw.
 \end{aligned}$$

Thus:

$$\begin{aligned}
 \langle D\phi^{-1}(y; a), h \rangle &= - \int_y^1 \frac{1}{F^*(w; a)^2} \left(\int_0^w h(v) dv \right) dw \\
 &= \left(\int_y^1 \frac{dw}{F^*(w; a)^2} \right) \left(\int_0^w h(v) dv \right) \Big|_y^1 - \int_y^1 \left(\int_w^1 \frac{dw}{F^*(w; a)^2} \right) h(w) dw \\
 &= - \left(\int_y^1 \frac{dw}{F^*(w; a)^2} \right) \int_0^y h(v) dv - \int_y^1 \left(\int_w^1 \frac{dw}{F^*(w; a)^2} \right) h(w) dw. \\
 &= k(y; a) \int_0^y h(v) dv + \int_y^1 k(w; a) h(w) dw,
 \end{aligned} \tag{a.7}$$

where $k(y; a) = - \int_y^1 (1/F^*(v)^2) dv$.

By inserting (a.7) in (a.6), we get the differential of the copula density, which is of the form:

$$\langle D \log c(u, v; a), h \rangle = \frac{h[C(u, v; a)]}{a[C(u, v; a)]} + \int_0^1 \gamma(u, v, w; a) h(w) dw, \text{ say.}$$

c) The information operator.

Let us now compute the information operator I_c of the copula. We get:

$$\begin{aligned}
 &E_0 [\langle D \log c(U, V; a_0), g \rangle \langle D \log c(U, V; a_0), h \rangle] \\
 &= E_0 \{ g[C_0(U, V)] h[C_0(U, V)] / a_0[C_0(U, V)]^2 \} \\
 &\quad + \int E_0 \{ g[C_0(U, V)] \gamma(U, V, y) / a_0[C_0(U, V)] \} h(y) dy \\
 &\quad + \int E_0 \{ \gamma(U, V, y) h[C_0(U, V)] / a_0[C_0(U, V)] \} g(y) dy \\
 &\quad + \int \int E_0 \{ \gamma(U, V, x) \gamma(U, V, y) \} g(x) h(y) dx dy.
 \end{aligned}$$

Let us consider the four terms separately. The first one is:

$$E_0 \{ g[C_0(U, V)] h[C_0(U, V)] / a_0[C_0(U, V)]^2 \} = \int g(w) h(w) \frac{f_W(w; a_0)}{a_0(w)^2} dw,$$

where $f_W(., a_0)$ is the density of W . The second term is:

$$\begin{aligned} & \int E_0 \{g[C_0(U, V)] \gamma(U, V, y) / a_0[C_0(U, V)]\} h(y) dy \\ &= \int E_0 \{g(W) \tilde{\gamma}(W, Z, y) / a_0(W)\} h(y) dy, \text{ say,} \\ &= \int \int g(w) \frac{E_0 \{\tilde{\gamma}(W, Z, y) \mid W = w\} f_W(w, a_0)}{a_0(w)} h(y) dy. \end{aligned}$$

Similarly we get for the third and fourth terms:

$$\begin{aligned} & \int E_0 \{\gamma(U, V, y) h[C_0(U, V)] / a_0[C_0(U, V)]\} g(y) dy \\ &= \int \int g(y) \frac{E_0 \{\tilde{\gamma}(W, Z, y) \mid W = w\} f_W(w, a_0)}{a_0(w)} h(w) dw, \end{aligned}$$

and:

$$\begin{aligned} & \int \int E_0 \{\gamma(U, V, x) \gamma(U, V, y)\} g(x) h(y) dx dy \\ &= \int \int g(x) E_0 \{\tilde{\gamma}(W, Z, x) \tilde{\gamma}(W, Z, y)\} h(y) dx dy. \end{aligned}$$

Thus the information operator I_c admits representation (3.2), with local component:

$$\alpha_0(w; a) = \frac{f_W(w; a)}{a_0(w)^2},$$

and:

$$\begin{aligned} \alpha_1(x, y; a) &= E_a \{\tilde{\gamma}(W, Z, y) \mid W = x\} f_W(x, a) / a(x) \\ &\quad + E_a \{\tilde{\gamma}(W, Z, x) \mid W = y\} f_W(y, a) / a(y) \\ &\quad + E_0 \{\tilde{\gamma}(W, Z, x) \tilde{\gamma}(W, Z, y)\}. \end{aligned}$$

d) The density of the variable W

The c.d.f. of W is given by [see Genest, Rivest (1993)]:

$$\begin{aligned} F_W(w) &= P[C(U, V) \leq w] = w - \frac{\phi^{-1}(w)}{d\phi^{-1}(w)/dw} \\ &= w - \phi^{-1}(w) \phi'[\phi^{-1}(w)] \\ &= w + \phi^{-1}(w) F^*(w). \end{aligned}$$

Thus the density of W is given by:

$$\begin{aligned} f_W(w) &= 1 + \frac{1}{\phi'[\phi^{-1}(w)]} F^*(w) + \phi^{-1}(w) f^*(w) \\ &= \phi^{-1}(w) f^*(w). \end{aligned}$$

vi) Extreme value copula

a) Copula p.d.f.

Let us introduce the variables $x = \log u$, $y = \log v$, and the function:

$$D(x, y) = (x + y)\chi\left(\frac{x}{x + y}\right).$$

Then we have:

$$C(u, v) = \exp[D(x, y)],$$

and thus:

$$\frac{\partial C(u, v)}{\partial u} = \frac{C(u, v)}{u} \frac{\partial D(x, y)}{\partial x},$$

and:

$$\frac{\partial^2 C(u, v)}{\partial u \partial v} = \frac{C(u, v)}{uv} \left\{ \frac{\partial D(x, y)}{\partial x} \frac{\partial D(x, y)}{\partial y} + \frac{\partial^2 D(x, y)}{\partial x \partial y} \right\}.$$

The derivatives of function D are:

$$\begin{aligned} \frac{\partial D(x, y)}{\partial x} &= \chi\left(\frac{x}{x + y}\right) + \frac{y}{x + y} \chi'\left(\frac{x}{x + y}\right), \\ \frac{\partial D(x, y)}{\partial y} &= \chi\left(\frac{x}{x + y}\right) - \frac{x}{x + y} \chi'\left(\frac{x}{x + y}\right), \\ \frac{\partial^2 D(x, y)}{\partial x \partial y} &= -\frac{xy}{(x + y)^3} \chi''\left(\frac{x}{x + y}\right). \end{aligned}$$

By substitution, the expression of the copula p.d.f. follows.

b) Characterization of the generator χ

By the Pickands representation (see e.g. Joe [1997], Theorem 6.3), a c.d.f. C with uniform margins is an extreme value copula iff function $A(x, y) = -\log C(e^{-x}, e^{-y})$ admits the representation:

$$A(x, y) = \int_{S^1} \max\{q_1 x, q_2 y\} \sigma(dq),$$

where σ is a finite measure on the one-dimensional simplex $S^1 = \{q = (q_1, q_2) \in \mathbb{R}_+^2 : q_1 + q_2 = 1\}$. Thus the generator χ of an extreme value copula is such that there exists a measure F^* on $[0, 1]$ with:

$$\begin{aligned} \chi(v) &= 2 \int_0^1 \max\{(1 - z)v, z(1 - v)\} dF^*(z), \\ \chi(0) &= \chi(1) = 1. \end{aligned}$$

The boundary conditions on χ are equivalent to:

$$\int_0^1 (1 - z) dF^*(z) = \int_0^1 z dF^*(z) = \frac{1}{2},$$

that is F^* is a c.d.f. such that $\int_0^1 z dF^*(z) = 1/2$.

c) Expression of the generator and of its derivatives

When F^* admits a density f^* , we get:

$$\chi(v) = 2v \int_0^v (1-z) f^*(z) dz + 2(1-v) \int_v^1 z f^*(z) dz.$$

Let us now compute the derivatives of χ . We get:

$$\begin{aligned} \chi'(v) &= 2 \int_0^v (1-z) f^*(z) dz - 2 \int_v^1 z f^*(z) dz \\ &= 2 \int_0^v f^*(z) dz - 1, \end{aligned}$$

and:

$$\chi''(v) = 2f^*(v).$$

Let us introduce functional parameter $a = f^*$. Using the restrictions on f^* , we deduce the expressions of χ and χ' in terms of functional parameter a :

$$\begin{aligned} \chi(v) &= v \int_0^v a(w) dw - \int_0^v w a(w) dw + 1 - v, \\ \chi'(v) &= \int_0^v a(w) dw - 1. \end{aligned}$$

Appendix 5

Kernel estimators

Let us consider the following assumptions.

Assumption B.1: $Z_t = (X_t, Y_t)$, t varying, is a strictly stationary process, with β -mixing coefficients $\beta(k)$ such that: $\beta(k) = O(k^{-\delta})$, $\delta > 1$.

Assumption B.2: The stationary density f has compact support $[0, 1]^2$, vanishes at its boundary, and is of class C^s .

Assumption B.3: The kernel K is of class C^r , with derivatives in $L^2(\mathbb{R})$. Further K is of order $m = s$.

We have the following theorem [see Theorem 3 of Aït-Sahalia (1993)].

Theorem. Consider a functional Φ from an open subset of C^s to \mathbb{R} . Suppose that Φ is Hadamard differentiable at the true c.d.f. F with Hadamard derivative $\langle D\Phi(F), H \rangle = \int \varphi[F](x, y) dH(x, y)$:

$$\Phi(F + H) = \Phi(F) + \int \varphi[F](x, y) dH(x, y) + R[F, H],$$

with $|R[F, H]| = O(\|H\|_\infty^2)$, uniformly on H in the class of compact set. Assume the bandwidth h_T is such that $h_T \rightarrow \infty$, $Th_T^2 \rightarrow \infty$. Then under Assumptions B.1-B.3:

- i. if $\varphi[F]$ is a cadlag function, and $Th_T^{2m} \rightarrow 0$:

$$\sqrt{T} \left[\Phi(\hat{F}_T) - \Phi(F) \right] \xrightarrow{d} N[0, V_\Phi(F)],$$

where:

$$V_\Phi(F) = \sum_{k=-\infty}^{\infty} \text{cov}(\varphi[F](Z_t), \varphi[F](Z_{t-k})).$$

- ii. If $\varphi[F]$ is of the form $\varphi[F](x, y) = \gamma_0(x, y) \delta_{x_0}(x) + \gamma_1(x, y) \delta_{y_0}(y) + \gamma_2(x, y)$, where $\gamma_0, \gamma_1 \in C^0$, $\gamma_2 \in C^1$, and $Th_T^{2m+1} \rightarrow 0$:

$$\sqrt{Th_T} \left[\Phi(\hat{F}_T) - \Phi(F) \right] \xrightarrow{d} N[0, V_\Phi(F)],$$

where:

$$\begin{aligned} V_\Phi(F) = & \left(\int K(z)^2 dz \right) (E[\gamma_0(Z_t)^2 | X_t = x_0] f_X(x_0) \\ & + E[\gamma_1(Z_t)^2 | Y_t = y_0] f_Y(y_0)). \end{aligned}$$

- iii. If $\varphi[F]$ is of the form $\varphi[F](x, y) = \alpha(x, y) \delta_{(x_0, y_0)}(x, y)$, and $Th_T^{2m+2} \rightarrow 0$:

$$\sqrt{Th_T^2} \left[\Phi(\hat{F}_T) - \Phi(F) \right] \xrightarrow{d} N[0, V_\Phi(F)],$$

where:

$$V_\Phi(F) = \left(\int K(z)^2 dz \right)^2 \alpha(x_0, y_0) f(x_0, y_0).$$

Let us introduce the last assumption:

Assumption B.4: The bandwidth h_T is such that $h_T \rightarrow \infty$, $Th_T^2 \rightarrow \infty$, $Th_T^{2m} \rightarrow 0$.

i) Density estimators.

Let us consider the kernel estimator for the density at (x_0, y_0) , $\hat{f}_T(x_0, y_0)$. The functional $\Phi(F) = f(x_0, y_0)$ is Hadamard differentiable, with $\varphi[F](x, y) = \delta_{(x_0, y_0)}(x, y)$, and $R[F, H] = 0$. Thus, under Assumptions B.1-B.4:

$$\sqrt{Th_T^2} \left(\hat{f}_T(x_0, y_0) - f(x_0, y_0) \right) \xrightarrow{d} N \left[0, f(x_0, y_0) \left(\int K(z)^2 dz \right)^2 \right].$$

ii) Conditional moment estimators.

Let us consider a conditional moment of the type:

$$\begin{aligned} g(x_0, y_0) &= \int \gamma_0(x_0, y) f(x_0, y) dy + \int \gamma_1(x, y_0) f(x, y_0) dx \\ &\quad + \int \int \gamma_2(x, y) f(x, y) dx dy, \end{aligned}$$

where $\gamma_0, \gamma_1 \in C^0, \gamma_2 \in C^1$, and $x_0, y_0 \in \mathbb{R}$. The functional $\Phi(F) = g(x_0, y_0)$ is Hadamard differentiable, with $\varphi[F](x, y) = \gamma_0(x, y) \delta_{x_0}(x) + \gamma_1(x, y) \delta_{y_0}(y) + \gamma_2(x, y)$, and $R(F, H) = 0$. Then the conditional moment estimator:

$$\begin{aligned} g_T(x_0, y_0) &= \int \gamma_0(x_0, y) \hat{f}_T(x_0, y) dy + \int \gamma_1(x, y_0) \hat{f}_T(x, y_0) dx \\ &\quad + \int \int \gamma_2(x, y) \hat{f}_T(x, y) dx dy, \end{aligned}$$

is asymptotically normal, with:

$$\sqrt{Th_T} [g_T(x_0, y_0) - g(x_0, y_0)] \xrightarrow{d} N(0, V_\Phi(F))$$

where:

$$\begin{aligned} V_\Phi(F) &= \left(\int K(z)^2 dz \right) (E[\gamma_0(Z_t)^2 | X_t = x_0] f_X(x_0) \\ &\quad + E[\gamma_1(Z_t)^2 | Y_t = y_0] f_Y(y_0)). \end{aligned}$$

Formula (3.19) is a special case.

iii) Moment estimators.

Finally let us consider a moment estimator $\int \int g(x, y) \hat{f}_T(x, y) dx dy$, where g is cadlag. The functional $\Phi(F) = \int \int g(x, y) f(x, y) dx dy$ is Hadamard differentiable, with $\varphi[F](x, y) = g(x, y)$ and $R[F, H] = 0$. Thus, under Assumptions B.1-B.4:

$$\sqrt{T} \left(\int \int g(x, y) \hat{f}_T(x, y) dx dy - \int \int g(x, y) f(x, y) dx dy \right) \xrightarrow{d} N(0, V_\Phi(F)),$$

where:

$$V_{\Phi}(F) = \sum_{k=-\infty}^{\infty} \text{cov}[g(Z_t), g(Z_{t-k})].$$

Appendix 6 Consistency

It is well-known that the estimator is consistent under the following assumptions:

- i) Q_T converges in probability to a deterministic limit Q_{∞} , uniformly in A ;
- ii) Q_{∞} is continuous with respect to A ;
- iii) $\forall \varepsilon > 0$: $\inf_{A \in B_{\varepsilon}(A_0)^c \cap \Theta} Q_{\infty}(A) > Q_{\infty}(A_0)$, where $B_{\varepsilon}(A_0)$ denotes a ball of radius ε around A_0 , w.r.t. the norm $\|\cdot\|_{L^2(\nu)}$.

In the proof of these three points we use the following technical assumptions.

Assumption A.8: *There exist $p > 1$ such that:*

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

Assumption A.9: *For $q > 1$ such that $1/p + 1/q = 1$:*

$$\left\| \frac{\widehat{f}_T(\cdot, \cdot) - f(\cdot, \cdot)}{\widehat{f}_T(\cdot, \cdot)} \right\|_{L^q(\Omega_T)} \xrightarrow{p} 0.$$

Assumption A.10:

$$\int_0^1 \int_0^1 \frac{[\widehat{f}_T(x, y) - f(x, y)]^2}{f(x, y)} dx dy = O_p(1).$$

Assumption A.11: *Let $R(x, y; A, h)$ be the residual term in the first order expansion of the density with respect to A :*

$$f(x, y; A + h) = f(x, y; A) + \langle Df(x, y; A), h \rangle + R(x, y; A, h).$$

For any $A \in \Theta$:

$$\int \int \frac{R(x, y; A, h)^2}{f(x, y)} dx dy = O\left(\|h\|_{L^2(\nu)}^4\right), \quad h \in L^2(\nu).$$

i) Uniform Convergence.

We have:

$$\begin{aligned} Q_T(A) &= \int \int \widehat{f}_T(x, y) \omega_T(x, y) dx dy \\ &\quad - 2 \int \int f(x, y; A) \omega_T(x, y) dx dy \\ &\quad + \int \int \frac{f(x, y; A)^2}{\widehat{f}_T(x, y)} \omega_T(x, y) dx dy, \end{aligned}$$

and:

$$Q_\infty(A) = Q(A) = \int \int \frac{f(x, y; A)^2}{f(x, y)} dx dy - 1.$$

Thus:

$$\begin{aligned} Q_T(A) - Q_\infty(A) &= \int \int \widehat{f}_T(x, y) \omega_T(x, y) dx dy - 1 \\ &\quad - 2 \left(\int \int f(x, y; A) \omega_T(x, y) dx dy - 1 \right) \\ &\quad + \int_0^1 \int_0^1 f(x, y; A)^2 \left(\frac{1}{\widehat{f}_T(x, y)} - \frac{1}{f(x, y)} \right) \omega_T(x, y) dx dy \\ &\quad + \int_0^1 \int_0^1 \frac{f(x, y; A)^2}{f(x, y)} (\omega_T(x, y) - 1) dx dy \\ &\equiv S_{1,T} + S_{2,T} + S_{3,T} + S_{4,T}, \text{ say.} \end{aligned}$$

Let us now check that each term converges in probability to 0, uniformly in $A \in \Theta$. We have:

$$\begin{aligned} |S_{1,T}| &= \left| \int \int \widehat{f}_T(x, y) (\omega_T(x, y) - 1) dx dy \right| \\ &\leq \int \int \widehat{f}_T(x, y) |\omega_T(x, y) - 1| dx dy \\ &\leq \int \int \widehat{f}_T(x, y) \mathbb{I}_{\widetilde{\Omega}_T^c}(x, y) dx dy \\ &\leq \left\| \frac{\widehat{f}_T}{\sqrt{f}} \right\|_{L^2} \left\| \sqrt{f} \mathbb{I}_{\widetilde{\Omega}_T^c} \right\|_{L^2} \\ &\leq \left(\int_0^1 \int_0^1 \frac{\widehat{f}_T(x, y)^2}{f(x, y)} dx dy \right)^{\frac{1}{2}} \left(\int_0^1 \int_0^1 \mathbb{I}_{\widetilde{\Omega}_T^c}(x, y) f(x, y) dx dy \right)^{\frac{1}{2}} \\ &\leq \left(\int \int \frac{[\widehat{f}_T(x, y) - f(x, y)]^2}{f(x, y)} dx dy + 1 \right)^{\frac{1}{2}} P_0 \left[(X_t, Y_t) \in \widetilde{\Omega}_T^c \right]^{1/2} \xrightarrow{P} 0, \end{aligned}$$

due to Assumptions A.6 and A.10.

The proof is similar for $S_{2,T}$:

$$\begin{aligned}
|S_{2,T}| &\leq \int \int f(x, y; A) \mathbb{I}_{\tilde{\Omega}_T^c}(x, y) dx dy \\
&\leq \left(\int \int \frac{f(x, y; A)^2}{f(x, y)} dx dy \right)^{\frac{1}{2}} \left(\int \int \mathbb{I}_{\tilde{\Omega}_T^c}(x, y) f(x, y) dx dy \right)^{\frac{1}{2}} \\
&\leq \left(\sup_{A \in \Theta} Q(A) + 1 \right)^{\frac{1}{2}} P_0 \left[(X_t, Y_t) \in \tilde{\Omega}_T^c \right]^{1/2} \\
&\rightarrow 0, \text{ in probability uniformly in } A \in \Theta,
\end{aligned}$$

due to Assumption A.6, whenever $\sup_{A \in \Theta} Q(A) < \infty$. Under Assumption A.5. i. Θ is compact, and $\sup_{A \in \Theta} Q(A) < \infty$ since Q is continuous [see ii) below]. Under Assumption A.5. ii. Θ is bounded, and $\sup_{A \in \Theta} Q(A) < \infty$ since:

$$Q(A_0 + h) = C_1 \|h\|_{L^2(\nu)}^2 + C_2 \|h\|_{L^2(\nu)}^3 + C_3 \|h\|_{L^2(\nu)}^4,$$

for some constants C_1, C_2, C_3 [see ii) below].

Let us now consider $S_{3,T}$:

$$\begin{aligned}
|S_{3,T}| &\leq \int_0^1 \int_0^1 \frac{f(x, y; A)^2}{f(x, y)} \left| \frac{\hat{f}_T(x, y) - f(x, y)}{\hat{f}_T(x, y)} \right| \omega_T(x, y) dx dy \\
&\leq \sup_{A \in \Theta} \left\| \frac{f(\cdot, \cdot; A)^2}{f(\cdot, \cdot)} \right\|_{L^p} \left\| \frac{\hat{f}_T - f}{\hat{f}_T} \right\|_{L^q(\Omega_T)} \\
&\rightarrow 0, \text{ in probability uniformly in } A \in \Theta,
\end{aligned}$$

due to Assumptions A.8 and A.9.

Finally, the last term $S_{4,T}$ is such that:

$$\begin{aligned}
|S_{4,T}| &\leq \int \int \frac{f(x, y; A)^2}{f(x, y)} |\omega_T(x, y) - 1| dx dy \\
&\leq \int \int \frac{f(x, y; A)^2}{f(x, y)} \mathbb{I}_{\tilde{\Omega}_T^c}(x, y) dx dy \\
&\leq \left\| \frac{f(\cdot, \cdot; A)^2}{f(\cdot, \cdot)} \right\|_{L^p} \left\| \mathbb{I}_{\tilde{\Omega}_T^c} \right\|_{L^q}, \\
&\leq \sup_{A \in \Theta} \left\| \frac{f(\cdot, \cdot; A)^2}{f(\cdot, \cdot)} \right\|_{L^p} \cdot \lambda_2 \left(\tilde{\Omega}_T^c \right)^{1/q}, \\
&\rightarrow 0, \text{ in probability uniformly in } A \in \Theta,
\end{aligned}$$

due to Assumptions A.6 and A.8.

ii) Continuity of the chi-square criterion.

To show the continuity of the limit criterion $Q_\infty = Q$, we have to prove:

$$\lim_{h \rightarrow 0} Q(A + h) = Q(A), \quad \forall A \in \Theta,$$

where $h \rightarrow 0$ denotes convergence in norm $\|\cdot\|_{L^2(\nu)}$. For this purpose let us consider the expansion of the chi-square criterion:

$$\begin{aligned} Q(A + h) &= \int \int \frac{[f(x, y) - f(x, y; A + h)]^2}{f(x, y)} dx dy \\ &= \int \int \frac{[f(x, y) - f(x, y; A) - \langle Df(x, y; A), h \rangle - R(x, y; A, h)]^2}{f(x, y)} dx dy \\ &= Q(A) + \int \int \langle D \log f(x, y; A), h \rangle^2 f(x, y) dx dy \\ &\quad + \int \int \frac{R(x, y; A, h)^2}{f(x, y)} dx dy \\ &\quad - 2 \int \int [f(x, y) - f(x, y; A)] \langle D \log f(x, y; A), h \rangle dx dy \\ &\quad + 2 \int \int \langle D \log f(x, y; A), h \rangle R(x, y; A, h) dx dy \\ &\quad - 2 \int \int \frac{f(x, y) - f(x, y; A)}{f(x, y)} R(x, y; A, h) dx dy. \end{aligned}$$

Let us now bound the terms in the last three lines. For the first one we have:

$$\begin{aligned} &\left| \int \int [f(x, y) - f(x, y; A)] \langle D \log f(x, y; A), h \rangle dx dy \right| \\ &= \left| E_0 \left[\frac{f(X, Y) - f(X, Y; A)}{f(X, Y)} \langle D \log f(X, Y; A), h \rangle \right] \right| \\ &\leq E_0 \left[\left(\frac{f(X, Y) - f(X, Y; A)}{f(X, Y)} \right)^2 \right]^{1/2} E_0 [\langle D \log f(X, Y; A), h \rangle^2]^{1/2} \\ &= Q(A)^{1/2} (h, I_A h)_{L^2(\nu)}^{1/2}. \end{aligned}$$

Similar upper bounds can be obtained for the last two terms. Thus the expansion of Q is:

$$\begin{aligned} Q(A + h) &= Q(A) + (h, I_A h)_{L^2(\nu)} + \int \int \frac{R(x, y; A, h)^2}{f(x, y)} dx dy \\ &\quad + O \left[(h, I_A h)_{L^2(\nu)}^{1/2} Q(A)^{1/2} \right] \\ &\quad + O \left[\left(\int \int \frac{R(x, y; A, h)^2}{f(x, y)} dx dy \right)^{1/2} (h, I_A h)_{L^2(\nu)}^{1/2} \right] \\ &\quad + O \left[\left(\int \int \frac{R(x, y; A, h)^2}{f(x, y)} dx dy \right)^{1/2} Q(A)^{1/2} \right]. \end{aligned} \tag{a.8}$$

Under Assumptions A.7 and A.11 we get:

$$Q(A + h) = Q(A) + O\left(\|h\|_{L^2(\nu)}^2\right),$$

and the continuity follows.

iii) Identification.

Under Assumption A.3 i. or ii. we have (see Appendix 2):

$$\inf_{A \in \Theta \setminus B_\varepsilon(A_0)} Q(A) > 0 = Q(A_0).$$

iv) Sufficient conditions for compactness.

In Assumption A.5 i. the set Θ is supposed to be compact in $L^2(\nu)$. We report here a theorem providing sufficient conditions for compactness in L^p spaces [see e.g. Yosida (1995)].

Theorem. (*Fréchet-Kolmogorov*). *Let Θ be a subset of the Banach space L^p of p -integrable functions with respect to the Lebesgue measure on \mathbb{R} . Assume:*

- i. Θ is bounded: $\sup_{A \in \Theta} \|A\|_{L^p} < \infty$;
- ii. $\sup_{A \in \Theta} \|A(\cdot + u) - A(\cdot)\|_{L^p} \longrightarrow 0$, as $u \rightarrow 0$;
- iii. $\lim_{\alpha \rightarrow \infty} \sup_{A \in \Theta} \int_{|x| > \alpha} A(x)^p dx = 0$.

Then Θ is precompact, that is its closure is compact.

Generalizations of this theorem when the L^p -space is defined with respect to a general measure are possible.

Appendix 7

The efficient score ψ_T

Let g be a function on $[0, 1]^2$, such that $g(\cdot, \cdot)/f(\cdot, \cdot; A_0) \in L^2(P_0)$. By Riesz representation theorem, there exists $\psi(g) \in L^2(\nu)$ such that:

$$(\psi(g), h)_{L^2(\nu)} = E_0 \left[\frac{g(X, Y)}{f(X, Y)} \langle D \log f(X, Y; A_0), h \rangle \right], \quad \forall h \in L^2(\nu).$$

It is given by $\psi(g) = \langle D \log f(., .; A_0)^*, g/f \rangle$. When the differential admits a decomposition (3.3), function $\psi(g)$ is given by:

$$\begin{aligned} \psi(g)(z) &= \frac{1}{d\nu/d\lambda(z)} \left[\int g(z, y) \gamma_0(z, y) dy + \int g(x, z) \gamma_1(x, z) dx \right. \\ &\quad \left. + \int \int g(x, y) \gamma_2(x, y, z) dx dy \right]. \end{aligned}$$

Let us now apply these results to function $g_T = \left[(\hat{f}_T - f) / f \right] \omega_T = (\delta \hat{f}_T / f) \omega_T$. For any $T \in \mathbb{N}$, $(\delta \hat{f}_T / f) \omega_T \in L^2(P_0)$ with probability 1. Thus there exists $\psi_T \in L^2(\nu)$ such that:

$$(\psi_T, h)_{L^2(\nu)} = E_0 \left[\frac{\delta \hat{f}_T(X, Y)}{f(X, Y)} \omega_T(X, Y) \langle D \log f(X, Y; A_0), h \rangle \right], \quad \forall h \in L^2(\nu).$$

When the differential admits a decomposition (3.3), function ψ_T is given by:

$$\begin{aligned} \psi_T(z) &= \frac{1}{d\nu/d\lambda(z)} \left[\int \delta \hat{f}_T(z, y) \omega_T(z, y) \gamma_0(z, y) dy + \int \delta \hat{f}_T(x, z) \omega_T(x, z) \gamma_1(x, z) dx \right. \\ &\quad \left. + \int \int \delta \hat{f}_T(x, y) \omega_T(x, y) \gamma_2(x, y, z) dx dy \right]. \end{aligned}$$

Appendix 8

Asymptotic expansion of first order conditions

i) Expansion of the first order condition

From Assumption A.12, \hat{A}_T satisfies the set of first order conditions:

$$\int \int \frac{\hat{f}_T(x, y) - f(x, y; \hat{A}_T)}{\hat{f}_T(x, y)} \langle Df(x, y; \hat{A}_T), g \rangle \omega_T(x, y) dx dy = 0, \quad \forall g \in L^2(\nu).$$

Let us denote $\delta \hat{A}_T = \hat{A}_T - A_0$. We can expand the functions involved in the first order condition. We get:

$$\begin{aligned} f(x, y; \hat{A}_T) &= f(x, y) + \langle Df(x, y; A_0), \delta \hat{A}_T \rangle + R(x, y; \delta \hat{A}_T), \\ \langle Df(x, y; \hat{A}_T), g \rangle &= \langle Df(x, y; A_0), g \rangle + \tilde{R}(x, y; \delta \hat{A}_T, g). \end{aligned}$$

The behaviour of the residual terms R and \tilde{R} has to be constrained to ensure that they are negligible for small $\delta\hat{A}_T$. This is achieved for R by Assumption A.2.bis. For \tilde{R} we assume:

Assumption A.13: *The residual term \tilde{R} is such that:*

$$\int \int \frac{\tilde{R}(x, y; h, g)^2}{f(x, y)} dx dy = O\left(\|h\|_{L^2(\nu)}^2 \|g\|_{L^2(\nu)}^2\right).$$

By writing:

$$\frac{1}{\hat{f}_T(x, y)} = \frac{1}{f(x, y)} \left(1 - \frac{\delta\hat{f}_T(x, y)}{\hat{f}_T(x, y)}\right),$$

where $\delta\hat{f}_T = \hat{f}_T - f$, the first order condition becomes:

$$\begin{aligned} 0 = & \int \int \delta\hat{f}_T(x, y) \omega_T(x, y) \langle D \log f(x, y; A_0), g \rangle \left(1 - \frac{\delta\hat{f}_T(x, y)}{\hat{f}_T(x, y)}\right) dx dy \\ & - \int \int \frac{\langle Df(x, y; A_0), \delta\hat{A}_T \rangle \langle Df(x, y; A_0), g \rangle}{f(x, y)} \left(1 - \frac{\delta\hat{f}_T(x, y)}{\hat{f}_T(x, y)}\right) \omega_T(x, y) dx dy \\ & - \int \int R(x, y; \delta\hat{A}_T) \frac{\langle Df(x, y; A_0), g \rangle}{f(x, y)} \left(1 - \frac{\delta\hat{f}_T(x, y)}{\hat{f}_T(x, y)}\right) \omega_T(x, y) dx dy \\ & + \int \int \frac{\delta\hat{f}_T(x, y)}{f(x, y)} \tilde{R}(x, y; \delta\hat{A}_T, g) \left(1 - \frac{\delta\hat{f}_T(x, y)}{\hat{f}_T(x, y)}\right) \omega_T(x, y) dx dy \\ & - \int \int \frac{\langle Df(x, y; A_0), \delta\hat{A}_T \rangle}{f(x, y)} \tilde{R}(x, y; \delta\hat{A}_T, g) \left(1 - \frac{\delta\hat{f}_T(x, y)}{\hat{f}_T(x, y)}\right) \omega_T(x, y) dx dy \\ & - \int \int \frac{R(x, y; \delta\hat{A}_T)}{f(x, y)} \tilde{R}(x, y; \delta\hat{A}_T, g) \left(1 - \frac{\delta\hat{f}_T(x, y)}{\hat{f}_T(x, y)}\right) \omega_T(x, y) dx dy. \end{aligned}$$

The leading terms are the first one [where we recognize $(g, \psi_T)_{L^2(\nu)}$, see Appendix 7] and the second one [with $(g, I\delta\hat{A}_T)_{L^2(\nu)}$]. Thus the first order condition can be rewritten as:

$$(g, \psi_T - I\delta\hat{A}_T)_{L^2(\nu)} + R(\delta\hat{A}_T, g) = 0, \forall g \in L^2(\nu),$$

where the residual term $R\left(\delta\hat{A}_T, g\right)$ is:

$$\begin{aligned}
R\left(\delta\hat{A}_T, g\right) &= - \int \int \delta\hat{f}_T(x, y) \langle D \log f(x, y; A_0), g \rangle \frac{\delta\hat{f}_T(x, y)}{\hat{f}_T(x, y)} \omega_T(x, y) dx dy \\
&- \int \int \left\langle D \log f(x, y; A_0), \delta\hat{A}_T \right\rangle \langle D \log f(x, y; A_0), g \rangle f(x, y) \\
&\cdot \left[\left(1 - \frac{\delta\hat{f}_T(x, y)}{\hat{f}_T(x, y)} \right) \omega_T(x, y) - 1 \right] dx dy \\
&- \int \int R\left(x, y; \delta\hat{A}_T\right) \langle D \log f(x, y; A_0), g \rangle \left(1 - \frac{\delta\hat{f}_T(x, y)}{\hat{f}_T(x, y)} \right) \omega_T(x, y) dx dy \\
&+ \int \int \frac{\delta\hat{f}_T(x, y)}{f(x, y)} \tilde{R}\left(x, y; \delta\hat{A}_T, g\right) \left(1 - \frac{\delta\hat{f}_T(x, y)}{\hat{f}_T(x, y)} \right) \omega_T(x, y) dx dy \\
&- \int \int \frac{\left\langle D f(x, y; A_0), \delta\hat{A}_T \right\rangle}{f(x, y)} \tilde{R}\left(x, y; \delta\hat{A}_T, g\right) \left(1 - \frac{\delta\hat{f}_T(x, y)}{\hat{f}_T(x, y)} \right) \omega_T(x, y) dx dy \\
&- \int \int \frac{R\left(x, y; \delta\hat{A}_T\right)}{f(x, y)} \tilde{R}\left(x, y; \delta\hat{A}_T, g\right) \left(1 - \frac{\delta\hat{f}_T(x, y)}{\hat{f}_T(x, y)} \right) \omega_T(x, y) dx dy \\
&\equiv R_1\left(\delta\hat{A}_T, g\right) + R_2\left(\delta\hat{A}_T, g\right) + R_3\left(\delta\hat{A}_T, g\right) + R_4\left(\delta\hat{A}_T, g\right) + R_5\left(\delta\hat{A}_T, g\right) \\
&+ R_6\left(\delta\hat{A}_T, g\right).
\end{aligned}$$

ii) A bound for the residual term

The following Lemma provides a bound for the residual term $R\left(\delta\hat{A}_T, g\right)$ under the additional assumption:

Assumption A.14: *There exists $p > 1$ such that:*

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

Lemma A.2: *Under Assumptions A.13 and A.14 the residual term $R\left(\delta\hat{A}_T, g\right)$ is such that:*

$$R\left(\delta\hat{A}_T, g\right) = \|g\|_{L^2(\nu)} O_p\left[\tau_{T,1}^2 + (\tau_{T,1} + \tau_{T,2}) \left\|\delta\hat{A}_T\right\|_{L^2(\nu)} + \left\|\delta\hat{A}_T\right\|_{L^2(\nu)}^2\right],$$

where

$$\tau_{T,1} = \left\|\frac{\delta\hat{f}_T}{\hat{f}_T}\right\|_{L^\infty(\Omega_T)}, \quad \tau_{T,2} = \lambda_2\left(\tilde{\Omega}_T^c\right)^{1/q}, \quad 1/p + 1/q = 1,$$

and p is defined as in Assumption A.14.

Proof. We bound each of the six terms in the expression of $R(\delta\hat{A}_T, g)$.

i) The first term is such that:

$$\begin{aligned}
\left| R_1(\delta\hat{A}_T, g) \right| &\leq \left\| \frac{\delta\hat{f}_T}{\hat{f}_T} \right\|_{L^\infty(\Omega_T)}^2 \int |\langle D \log f(x, y; A_0), g \rangle| \sqrt{f(x, y)} \frac{\hat{f}_T(x, y)}{\sqrt{f(x, y)}} dx dy \\
&\leq \left\| \frac{\delta\hat{f}_T}{\hat{f}_T} \right\|_{L^\infty(\Omega_T)}^2 E_0 [\langle D \log f(X, Y; A_0), g \rangle^2]^{1/2} \left(\int \frac{\hat{f}_T(x, y)^2}{f(x, y)} dx dy \right)^{1/2} \\
&= \left\| \frac{\delta\hat{f}_T}{\hat{f}_T} \right\|_{L^\infty(\Omega_T)}^2 (g, Ig)_{L^2(\nu)}^{1/2} \left(\int \int \frac{[\hat{f}_T(x, y) - f(x, y)]^2}{f(x, y)} dx dy + 1 \right)^{1/2} \\
&= O_p \left[\|g\|_{L^2(\nu)} \tau_{1,T}^2 \right],
\end{aligned}$$

by continuity of the information operator I and Assumption A.10.

ii) The second term is such that:

$$\begin{aligned}
\left| R_2(\delta\hat{A}_T, g) \right| &\leq \int \int \left| \left\langle D \log f(x, y; A_0), \delta\hat{A}_T \right\rangle \left\langle D \log f(x, y; A_0), g \right\rangle \right. \\
&\quad \left. f(x, y) |\omega_T(x, y) - 1| dx dy \right. \\
&\quad \left. + \int \int \left| \left\langle D \log f(x, y; A_0), \delta\hat{A}_T \right\rangle \left\langle D \log f(x, y; A_0), g \right\rangle \right| \right. \\
&\quad \left. f(x, y) \frac{|\delta\hat{f}_T(x, y)|}{\hat{f}_T(x, y)} \omega_T(x, y) dx dy \right. \\
&\leq \left\| \left\langle D \log f(\cdot, \cdot; A_0), g \right\rangle \left\langle D \log f(\cdot, \cdot; A_0), \delta\hat{A}_T \right\rangle f(\cdot, \cdot) \right\|_{L^p} \|\omega_T - 1\|_{L^q} \\
&\quad + \left\| \frac{\delta\hat{f}_T}{\hat{f}_T} \right\|_{L^\infty(\Omega_T)} \left\| \left\langle D \log f(\cdot, \cdot; A_0), g \right\rangle \left\langle D \log f(\cdot, \cdot; A_0), \delta\hat{A}_T \right\rangle f(\cdot, \cdot) \right\|_{L^1} \\
&= O_p \left[\|g\|_{L^2(\nu)} \|\delta\hat{A}_T\|_{L^2(\nu)} \right] \left(\lambda \left(\tilde{\Omega}_T^c \right)^{1/q} + \left\| \frac{\delta\hat{f}_T}{\hat{f}_T} \right\|_{L^\infty(\Omega_T)} \right) \\
&= O_p \left[\|g\|_{L^2(\nu)} \|\delta\hat{A}_T\|_{L^2(\nu)} (\tau_{1,T} + \tau_{2,T}) \right],
\end{aligned}$$

by Assumption A.14, where we used that $\|\varphi\|_{L^1} \leq \|\varphi\|_{L^p}$, $p > 1$, for a function φ defined on $[0, 1]^2$, by Hölder inequality.

iii) The third term satisfies:

$$\begin{aligned}
\left| R_3 \left(\delta \hat{A}_T, g \right) \right| &\leq (1 + \tau_{1,T}) \int \int |\langle D \log f(x, y; A_0), g \rangle| \sqrt{f(x, y)} \frac{|R(x, y; \delta \hat{A}_T)|}{\sqrt{f(x, y)}} dx dy \\
&\leq (1 + \tau_{1,T}) (g, Ig)_{L^2}^{1/2} \left(\int \int \frac{R(x, y; \delta \hat{A}_T)^2}{f(x, y)} dx dy \right)^{1/2} \\
&= O_p \left(\|g\|_{L^2(\nu)} \left\| \delta \hat{A}_T \right\|_{L^2(\nu)}^2 \right),
\end{aligned}$$

by Assumption A.2.bis.

iv) The term R_4 is such that:

$$\begin{aligned}
\left| R_4 \left(\delta \hat{A}_T, g \right) \right| &\leq (1 + \tau_{1,T}) \tau_{1,T} \int \int \frac{|\tilde{R}(x, y; \delta \hat{A}_T, g)|}{\sqrt{f(x, y)}} \frac{\hat{f}_T(x, y)}{\sqrt{f(x, y)}} dx dy \\
&\leq (1 + \tau_{1,T}) \tau_{1,T} \left(\int \int \frac{\tilde{R}(x, y; \delta \hat{A}_T, g)^2}{f(x, y)} dx dy \right)^{1/2} \left(\int \frac{\hat{f}_T(x, y)^2}{f(x, y)} dx dy \right) \\
&= O_p \left(\tau_{1,T} \|g\|_{L^2(\nu)} \left\| \delta \hat{A}_T \right\|_{L^2(\nu)} \right),
\end{aligned}$$

by Assumptions A.10, A.13.

v) The fifth term is bounded by:

$$\begin{aligned}
\left| R_5 \left(\delta \hat{A}_T, g \right) \right| &\leq (1 + \tau_{1,T}) \left(\delta \hat{A}_T, I \delta \hat{A}_T \right)_{L^2(\nu)}^{1/2} \left(\int \int \frac{\tilde{R}(x, y; \delta \hat{A}_T, g)^2}{f(x, y)} dx dy \right)^{1/2} \\
&= O_p \left(\|g\|_{L^2(\nu)} \left\| \delta \hat{A}_T \right\|_{L^2(\nu)}^2 \right).
\end{aligned}$$

vi) Finally, the last term:

$$\begin{aligned}
\left| R_6 \left(\delta \hat{A}_T, g \right) \right| &\leq (1 + \tau_T) \left(\int \int \frac{R(x, y; \delta \hat{A}_T)^2}{f(x, y)} dx dy \right)^{1/2} \\
&\quad \cdot \left(\int \int \frac{\tilde{R}(x, y; \delta \hat{A}_T, g)^2}{f(x, y)} dx dy \right)^{1/2} \\
&= O_p \left(\|g\|_{L^2(\nu)} \left\| \delta \hat{A}_T \right\|_{L^2(\nu)}^3 \right).
\end{aligned}$$

By gathering the dominant terms, the bound for $R\left(\delta\hat{A}_T, g\right)$ is proved.
Q.E.D.

iii) Negligibility of the residual term.

Finally we have to introduce an additional assumption to ensure that the residual term is negligible with respect to the other terms.

Assumption A.15:

$$\tau_{T,1} = \left\| \frac{\delta\hat{f}_T}{\hat{f}_T} \right\|_{L^\infty(\Omega_T)} = o_p(T^{-1/4}).$$

Lemma A.3: *Under Assumptions A.1-A.15:*

i.

$$\left\| \delta\hat{A}_T \right\|_{L^2} = O_p\left(1/\sqrt{T}\right).$$

ii.

$$\sqrt{T} \left(g, \delta\hat{A}_T \right)_{L^2(\nu)} = \sqrt{T} \left(g, I^{-1}\psi_T \right)_{L^2(\nu)} + o_p(1), \quad g \in L^2(\nu).$$

Proof. *From Lemma A.2, Assumptions A.6 and A.15 we get:*

$$R\left(\delta\hat{A}_T, g\right) = o_p\left(\|g\|_{L^2}/\sqrt{T}\right) + o_p\left(\|g\|_{L^2} \left\| \delta\hat{A}_T \right\|_{L^2(\nu)}\right).$$

Then the first order condition is such that:

$$\left(g, I\delta\hat{A}_T \right)_{L^2} = (g, \psi_T)_{L^2} + o_p\left(\|g\|_{L^2}/\sqrt{T}\right) + o_p\left(\|g\|_{L^2} \left\| \delta\hat{A}_T \right\|_{L^2(\nu)}\right),$$

for any $g \in L^2(\nu)$, and since I^{-1} is bounded we get:

$$\begin{aligned} \left(g, \delta\hat{A}_T \right)_{L^2(\nu)} &= (g, I^{-1}\psi_T)_{L^2(\nu)} + o_p\left(\|g\|_{L^2}/\sqrt{T}\right) \\ &\quad + o_p\left(\|g\|_{L^2(\nu)} \left\| \delta\hat{A}_T \right\|_{L^2(\nu)}\right), \end{aligned}$$

(a.9)

for any $g \in L^2(\nu)$.

Let us now deduce a bound for $\|\delta\hat{A}_T\|_{L^2(\nu)}$. Since $\sqrt{T}(I^{-1}g, \psi_T)_{L^2(\nu)} \xrightarrow{d} N[0, (g, I^{-1}g)_{L^2(\nu)}]$ (see Lemma 16 in the text) and I^{-1} is bounded:

$$(g, I^{-1}\psi_T)_{L^2(\nu)} = O_p\left(\|g\|_{L^2}/\sqrt{T}\right).$$

Thus:

$$(g, \delta\hat{A}_T)_{L^2(\nu)} = O_p\left(\|g\|_{L^2}/\sqrt{T}\right) + o_p\left(\|g\|_{L^2(\nu)}\|\delta\hat{A}_T\|_{L^2(\nu)}\right), \quad g \in L^2(\nu).$$

We get:

$$\begin{aligned} \|\delta\hat{A}_T\|_{L^2(\nu)} &= \sup_{g \in L^2(\nu): \|g\|_{L^2(\nu)}=1} (g, \delta\hat{A}_T)_{L^2(\nu)} \\ &= O_p\left(1/\sqrt{T}\right) + o_p\left(\|\delta\hat{A}_T\|_{L^2(\nu)}\right), \end{aligned}$$

that is $\|\delta\hat{A}_T\|_{L^2(\nu)} = O_p\left(1/\sqrt{T}\right)$. From (a.9) we deduce ii.

Q.E.D.

iv) Pointwise expansion.

Let us now focus on pointwise expansions. Intuitively, these are derived from the first order condition corresponding to a variation g of the functional parameter A which involves only a point $x_0 \in [0, 1]$. We use an approach by localization, and consider variations g which are more and more concentrated around x_0 as $T \rightarrow \infty$, at an higher speed than kernel localization. For simplicity we consider the case where A has one component.

Let $\varphi \in C_0^\infty$ be a symmetric kernel with compact support, and let \tilde{h}_T be a bandwidth converging to 0. For any $x_0 \in [0, 1]$, define the function:

$$g_{T,x_0}(x) = \frac{1}{\sqrt{\tilde{h}_T}} \varphi\left(\frac{x - x_0}{\tilde{h}_T}\right), \quad x \in [0, 1].$$

Then:

$$\begin{aligned} \|g_{T,x_0}\|_{L^2(\nu)}^2 &= \int \frac{1}{\tilde{h}_T} \varphi\left(\frac{x - x_0}{\tilde{h}_T}\right)^2 \frac{d\nu}{d\lambda}(x) dx \\ &= \int \varphi(u)^2 \frac{d\nu}{d\lambda}(x_0 + \tilde{h}_T u) du \simeq \left(\int \varphi(u)^2 du\right) \frac{d\nu}{d\lambda}(x_0). \end{aligned}$$

Thus $g_{T,x_0} \in L^2(\nu)$ λ -a.s. in x_0 , and $\|g_{T,x_0}\|_{L^2(\nu)}$ converges to a constant as $T \rightarrow \infty$. In addition, for any $h \in L^2(\nu)$:

$$\begin{aligned} (g_{T,x_0}, h)_{L^2(\nu)} &= \int \frac{1}{\sqrt{\tilde{h}_T}} \varphi\left(\frac{x-x_0}{\tilde{h}_T}\right) h(x) \frac{d\nu}{d\lambda}(x) dx \\ &= \sqrt{\tilde{h}_T} \int \varphi(u) h(x_0 + \tilde{h}_T u) \frac{d\nu}{d\lambda}(x_0 + \tilde{h}_T u) du \\ &= \sqrt{\tilde{h}_T} h(x_0) \frac{d\nu}{d\lambda}(x_0) \\ &\quad + \sqrt{\tilde{h}_T} \int \varphi(u) \left[\left(h \frac{d\nu}{d\lambda} \right)(x_0 + \tilde{h}_T u) - \left(h \frac{d\nu}{d\lambda} \right)(x_0) \right] du. \end{aligned}$$

The idea is to apply Lemma A.3 ii. to $g = g_{T,x_0}$. Since function g_{T,x_0} depends on T , it is important to know the rate of the residual term in Lemma A.3 ii) and for this purpose we have to strength Assumption A.15.

Assumption A.15':

$$\tau_{T,1} = O_p(T^{-1/4-\beta_1}), \quad \tau_{T,2} = O_p(T^{-\beta_2}), \quad \beta_1, \beta_2 > 0.$$

Lemma A.4: Let $g_T \in L^2(\nu)$ for any T , such that $\|g_T\|_{L^2(\nu)} \leq \text{const}$, independent of T , for T sufficiently large. Then under Assumption A.15':

$$\sqrt{T} \left(g_T, I\delta\hat{A}_T \right)_{L^2(\nu)} = \sqrt{T} (g_T, \psi_T)_{L^2(\nu)} + O_p(T^{-\beta}),$$

where $\beta = \min\{2\beta_1, 1/4 + \beta_1, \beta_2, 1/2\} > 0$.

Proof. Since the first order condition holds for any given T , and $g_T \in L^2(\nu)$:

$$\left(g_T, I\delta\hat{A}_T \right)_{L^2(\nu)} = (g_T, \psi_T)_{L^2(\nu)} + R(\delta\hat{A}_T, g_T).$$

From Lemma A.2, Lemma A.3 i., and using A.15', we get:

$$\begin{aligned} R(\delta\hat{A}_T, g_T) &= \|g_T\|_{L^2(\nu)} O_p[T^{-1/2-2\beta_1} + (T^{-1/4-\beta_1} + T^{-\beta_2}) T^{-1/2} + T^{-1}] \\ &= O_p(T^{-1/2-\beta}). \end{aligned}$$

Q.E.D.

Let us apply Lemma A.4 to $g_T = g_{T,x_0}$, where the bandwidth for localization \tilde{h}_T is selected such that:

$$\tilde{h}_T = o(h_T), \quad h_T = o(\tilde{h}_T T^{2\beta}).$$

We get:

$$\sqrt{Th_T/\tilde{h}_T} \left(g_{T,x_0}, I\delta\hat{A}_T \right)_{L^2(\nu)} = \sqrt{Th_T/\tilde{h}_T} (g_{T,x_0}, \psi_T)_{L^2(\nu)} + O_p \left(T^{-\beta} \sqrt{h_T/\tilde{h}_T} \right). \quad (\text{a.10})$$

Let us consider the RHS of (a.10). We get:

$$\begin{aligned} & \sqrt{Th_T/\tilde{h}_T} (g_{T,x_0}, \psi_T)_{L^2(\nu)} + O_p \left(T^{-\beta} \sqrt{h_T/\tilde{h}_T} \right) \\ & \simeq \sqrt{Th_T} \frac{d\nu}{d\lambda}(x_0) \psi_T(x_0) \\ & \quad + \sqrt{Th_T} \int \varphi(u) \left[\left(\frac{d\nu}{d\lambda} \psi_T \right) (x_0 + \tilde{h}_T u) - \left(\frac{d\nu}{d\lambda} \psi_T \right) (x_0) \right] du. \end{aligned}$$

Let us now consider the LHS of (a.10). We get:

$$\begin{aligned} & \sqrt{Th_T/\tilde{h}_T} \left(g_{T,x_0}, I\delta\hat{A}_T \right)_{L^2(\nu)} \\ & = \sqrt{Th_T} \frac{d\nu}{d\lambda}(x_0) I\delta\hat{A}_T(x_0) \\ & \quad + \sqrt{Th_T} \int \varphi(u) \left[\left(\frac{d\nu}{d\lambda} I\delta\hat{A}_T \right) (x_0 + \tilde{h}_T u) - \left(\frac{d\nu}{d\lambda} I\delta\hat{A}_T \right) (x_0) \right] du \\ & \simeq \sqrt{Th_T} \alpha_0(x_0) \delta\hat{A}_T(x_0) \\ & \quad + \sqrt{Th_T} \int \varphi(u) \left[\left(\alpha_0 \delta\hat{A}_T \right) (x_0 + \tilde{h}_T u) - \left(\alpha_0 \delta\hat{A}_T \right) (x_0) \right] du. \end{aligned}$$

Thus, from (a.10) we get:

$$\begin{aligned} & \sqrt{Th_T} \alpha_0(x_0) \delta\hat{A}_T(x_0) \\ & \quad + \sqrt{Th_T} \int \varphi(u) \left[\left(\alpha_0 \delta\hat{A}_T \right) (x_0 + \tilde{h}_T u) - \left(\alpha_0 \delta\hat{A}_T \right) (x_0) \right] du \\ & \simeq \sqrt{Th_T} \frac{d\nu}{d\lambda}(x_0) \psi_T(x_0) \\ & \quad + \sqrt{Th_T} \int \varphi(u) \left[\left(\frac{d\nu}{d\lambda} \psi_T \right) (x_0 + \tilde{h}_T u) - \left(\frac{d\nu}{d\lambda} \psi_T \right) (x_0) \right] du. \end{aligned}$$

Let us now show that the second term on the RHS is negligible, since $\tilde{h}_T = o(h_T)$. We have:

$$\begin{aligned} & \sqrt{Th_T} \int \varphi(u) \left[\left(\frac{d\nu}{d\lambda} \psi_T \right) (x_0 + \tilde{h}_T u) - \left(\frac{d\nu}{d\lambda} \psi_T \right) (x_0) \right] du \\ & \simeq \sqrt{Th_T} \frac{\tilde{h}_T^2}{2} \frac{d^2}{dx^2} \left(\frac{d\nu}{d\lambda} \psi_T \right) (x_0) \int u^2 \varphi(u) du, \text{ (since the kernel } \varphi \text{ is symmetric),} \\ & = o_p(1). \end{aligned} \quad (\text{a.11})$$

Indeed, since $\frac{d\nu}{d\lambda}\psi_T(x_0)$ involves conditional moments of kernel estimators of a density [see (3.24)], we have $\frac{d\nu}{d\lambda}\psi_T(x_0) = O_p\left[(Th_T)^{-1/2}\right]$ (see Lemma 13), and since each differentiation diminishes the rate of convergence of a kernel estimator by the factor h_T (see Theorem 3 in Aït-Sahalia [1993]), we deduce $\frac{d^2}{dx^2}\left(\frac{d\nu}{d\lambda}\psi_T\right)(x_0) = O_p\left[(Th_T)^{-1/2}h_T^{-2}\right]$. Thus we get:

$$\begin{aligned} & \sqrt{Th_T}\alpha_0(x_0)\delta\hat{A}_T(x_0) \\ \simeq & -\sqrt{Th_T}\int\varphi(u)\left[\left(\alpha_0\delta\hat{A}_T\right)\left(x_0+\tilde{h}_Tu\right)-\left(\alpha_0\delta\hat{A}_T\right)(x_0)\right]du \\ & +\sqrt{Th_T}\frac{d\nu}{d\lambda}(x_0)\psi_T(x_0), \quad \lambda\text{-a.s. in } x_0 \in [0, 1]. \end{aligned}$$

This is an integral equation for $\sqrt{Th_T}\alpha_0\delta\hat{A}_T$ which has a unique solution [see e.g. Theorem 5.2.3 in Debnath, Mikusinski (1998)]. By substitution and using (a.11), we see that the solution is of the form $\sqrt{Th_T}\alpha_0\delta\hat{A}_T = \sqrt{Th_T}\frac{d\nu}{d\lambda}\psi_T + o_p(1)$. We conclude:

$$\sqrt{Th_T}\alpha_0(x_0)\delta\hat{A}_T(x_0) \simeq \sqrt{Th_T}\frac{d\nu}{d\lambda}(x_0)\psi_T(x_0), \quad \lambda\text{-a.s. in } x_0 \in [0, 1].$$

v) Expansion of the constrained estimator.

Let us now consider the asymptotic expansion of the constrained estimator $\hat{f}_T^0(x, y)$. We get:

$$\begin{aligned} \hat{f}_T^0(x, y) - f(x, y) &= f(x, y; \hat{A}_T) - f(x, y; A_0) \\ &= \left\langle Df(x, y; A_0), \delta\hat{A}_T \right\rangle + R(x, y; \delta\hat{A}_T). \end{aligned}$$

Let us now derive a bound for $R(x, y; \delta\hat{A}_T)$. By Assumption A.2.bis and Lemma A.3.i. we get:

$$\left(\int\int\frac{R(x, y; \delta\hat{A}_T)^2}{f(x, y)}dxdy\right)^{1/2} = O_p\left(\left\|\delta\hat{A}_T\right\|_{L^2(\nu)}^2\right) = O_p(1/T). \quad (\text{a.12})$$

For any $x_0, y_0 \in [0, 1]$ let us introduce the function:

$$g_{T, x_0, y_0}(x, y) = \frac{1}{\tilde{h}_T}\varphi\left(\frac{x-x_0}{\tilde{h}_T}\right)\varphi\left(\frac{y-y_0}{\tilde{h}_T}\right),$$

where $\varphi \in C_0^\infty$ is a symmetric kernel with compact support and the localization bandwidth \tilde{h}_T is selected such that $\tilde{h}_T = o(h_T)$ and $\sqrt{h_T/T} = o(\tilde{h}_T)$ ²⁷. Then:

$$\begin{aligned} \|g_{T, x_0, y_0}\|_{L^2(P_0)}^2 &= \int\frac{1}{\tilde{h}_T^2}\varphi\left(\frac{x-x_0}{\tilde{h}_T}\right)^2\varphi\left(\frac{y-y_0}{\tilde{h}_T}\right)^2f(x, y)dxdy \\ &\simeq \left(\int\varphi(u)^2du\right)^2f(x_0, y_0). \end{aligned}$$

²⁷This is possible since $Th_T \rightarrow \infty$ by Assumption B.4 in Appendix 5.

Thus $g_{T,x_0,y_0} \in L^2(P_0)$ with $\|g_{T,x_0,y_0}\|_{L^2(P_0)} \leq C$ independent of T , for T sufficiently large. Then by Cauchy-Schwarz inequality:

$$\begin{aligned} \int \int g_{T,x_0,y_0}(x,y) R(x,y; \delta \hat{A}_T) dx dy &\leq \|g_{T,x_0,y_0}\|_{L^2(P_0)} \left(\int \int \frac{R(x,y; \delta \hat{A}_T)^2}{f(x,y)} dx dy \right)^{1/2} \\ &= O_p(1/T), \text{ [from (a.12)]}. \end{aligned} \quad (\text{a.13})$$

On the other hand:

$$\begin{aligned} \int \int g_{T,x_0,y_0}(x,y) R(x,y; \delta \hat{A}_T) dx dy &= \int \int \frac{1}{\tilde{h}_T} \varphi\left(\frac{x-x_0}{\tilde{h}_T}\right) \varphi\left(\frac{y-y_0}{\tilde{h}_T}\right) R(x,y; \delta \hat{A}_T) dx dy \\ &= \tilde{h}_T \int \int \varphi(u) \varphi(v) R(x_0 + \tilde{h}_T u, y_0 + \tilde{h}_T v; \delta \hat{A}_T) du dv \\ &\simeq \tilde{h}_T R(x_0, y_0; \delta \hat{A}_T), \end{aligned} \quad (\text{a.14})$$

since $\tilde{h}_T = o(h_T)$. In particular, from (a.13) and (a.14) we get:

$$R(x_0, y_0; \delta \hat{A}_T) = O_p(1/T\tilde{h}_T).$$

Since $1/\tilde{h}_T = o(\sqrt{T/h_T})$ it follows:

$$R(x,y; \delta \hat{A}_T) = O_p(1/T\tilde{h}_T) = o_p\left(1/\sqrt{Th_T}\right), \quad \lambda\text{-a.s. in } x, y.$$

Thus:

$$\hat{f}_T^0(x,y) - f(x,y) = \left\langle Df(x,y; A_0), \delta \hat{A}_T \right\rangle + o_p\left(1/\sqrt{Th_T}\right), \quad \lambda\text{-a.s. in } x, y.$$

Appendix 9

Asymptotic distribution of ψ_T

Let us consider the case where the differential operator admits the decomposition (3.3). From Appendix 7, function ψ_T is such that:

$$\begin{aligned} \frac{d\nu}{d\lambda}(w) \psi_T(w) &= \int \delta \hat{f}_T(w,y) \omega_T(w,y) \gamma_0(w,y) dy + \int \delta \hat{f}_T(x,w) \omega_T(x,w) \gamma_1(x,w) dx \\ &\quad + \int \int \delta \hat{f}_T(x,y) \omega_T(x,y) \gamma_2(x,y,w) dx dy \\ &\simeq \int \delta \hat{f}_T(w,y) \gamma_0(w,y) dy + \int \delta \hat{f}_T(x,w) \gamma_1(x,w) dx \\ &\quad + \int \int \delta \hat{f}_T(x,y) \gamma_2(x,y,w) dx dy. \end{aligned}$$

From Appendix 5, point ii), it follows that:

$$\sqrt{Th_T} \frac{d\nu}{d\lambda}(w) \psi_T(w) \xrightarrow{d} N(0, \sigma^2(w)),$$

where:

$$\begin{aligned} \sigma^2(w) &= \left(\int K(z)^2 dz \right) \left(E \left[\gamma_{0t} \gamma'_{0t} \mid X_t = w \right] f_X(w) \right. \\ &\quad \left. + E \left[\gamma_{1t} \gamma'_{1t} \mid Y_t = w \right] f_Y(w) \right) \\ &= \left(\int K(z)^2 dz \right) \alpha_0(w). \end{aligned}$$

Appendix 10

Asymptotic expansion in the time series framework

In this Appendix we essentially derive the first order expansions to understand the form of the asymptotic distribution. The first order condition is:

$$0 = \int \int \frac{\left[\hat{f}_T(x|y) - f(x|y; \hat{A}_T) \right]}{\hat{f}_T(x|y)} \left\langle Df(x|y; \hat{A}_T), g \right\rangle \omega_T(x, y) \hat{f}_{Y,T}(y) dx dy,$$

$\forall g \in L^2(\nu)$. Let us expand this condition. We get:

$$\begin{aligned} 0 &\simeq \int \int \frac{\left[\hat{f}_T(x|y) - f(x|y) \right]}{f(x|y)} \left\langle Df(x|y; A_0), g \right\rangle f_Y(y) dx dy \\ &\quad - \int \int \frac{\left\langle Df(x|y; A_0), \delta \hat{A}_T \right\rangle}{f(x|y)} \left\langle Df(x|y; A_0), g \right\rangle f_Y(y) dx dy \\ &= \int \int \frac{\left[\hat{f}_T(x|y) - f(x|y) \right]}{f(x|y)} \left\langle D \log f(x|y; A_0), g \right\rangle f(x, y) dx dy \\ &\quad - \int \int \left\langle D \log f(x|y; A_0), \delta \hat{A}_T \right\rangle \left\langle D \log f(x|y; A_0), g \right\rangle f(x, y) dx dy \\ &= \left(\tilde{\psi}_T, g \right)_{L^2(\nu)} - \left(I_{X|Y} \delta \hat{A}_T, g \right)_{L^2(\nu)}. \end{aligned}$$

Thus the first order condition is equivalent to:

$$\left(g, I_{X|Y} \delta \hat{A}_T - \tilde{\psi}_T \right)_{L^2(\nu)} \simeq 0, \quad \forall g \in L^2(\nu).$$

Appendix 11

Nonparametric information bound

i) Cross-sectional framework.

Let us introduce a one dimensional parametric model $A(., \theta)$ and derive its Cramer-Rao bound. The score is given by:

$$\frac{\partial \log f}{\partial \theta}(x, y; A(\theta_0)) = \left\langle D \log f(x, y; A_0), \frac{dA}{d\theta}(\theta_0) \right\rangle.$$

The Fisher information is:

$$\begin{aligned} E_0 \left[\left(\frac{\partial \log f}{\partial \theta}(X_t, Y_t; A(\theta_0)) \right)^2 \right] &= E_0 \left[\left\langle D \log f(X, Y; A_0), \frac{dA}{d\theta}(\theta_0) \right\rangle^2 \right] \\ &= \left(\frac{dA}{d\theta}(\theta_0), I \frac{dA}{d\theta}(\theta_0) \right)_{L^2(\nu)}. \end{aligned}$$

Thus the Cramer-Rao bound is given by:

$$B_A(g, \theta) = \left(\frac{dA}{d\theta}(\theta_0), I \frac{dA}{d\theta}(\theta_0) \right)_{L^2(\nu)}^{-1}.$$

The parametric specification can be chosen such that:

$$\int g(v)' A(v, \theta) \nu(dv) = \theta,$$

which is equivalent (in a neighborhood of θ_0) to the constraint:

$$\int g(v)' \frac{dA}{d\theta}(v, \theta_0) \nu(dv) = 1,$$

that is:

$$\left(g, \frac{dA}{d\theta}(\theta_0) \right)_{L^2(\nu)} = 1. \quad (\text{a.15})$$

Thus both the Cramer Rao bound and the constraint (a.15) depend on the parameterization only by means of the function $\delta(.) = dA/d\theta(., \theta_0)$. Therefore problem (3.28) in the text is equivalent to:

$$\begin{aligned} \min_{\delta \in L^2(\nu)} & (\delta, I\delta)_{L^2(\nu)}, \\ \text{s.t.} \quad & (g, \delta)_{L^2(\nu)} = 1. \end{aligned}$$

By Cauchy-Schwarz inequality we have:

$$1 = (g, \delta)_{L^2(\nu)}^2 = (I^{-1/2}g, I^{1/2}\delta)_{L^2(\nu)}^2 \leq (I^{-1}g, g)_{L^2(\nu)} (\delta, I\delta)_{L^2(\nu)}.$$

Therefore $(\delta, I\delta)_{L^2(\nu)} \geq (I^{-1}g, g)_{L^2(\nu)}^{-1}$ and the bound is reached for $\delta^* = I^{-1}g \in L^2(\nu)$. Thus we deduce:

$$B_A(g) = (g, I^{-1}g)_{L^2(\nu)}.$$

ii) Time-series framework.

In this case the score is given by:

$$\frac{\partial \log f}{\partial \theta}(x | y; A(\theta_0)) = \left\langle D \log f(x | y; A_0), \frac{dA}{d\theta}(\theta_0) \right\rangle.$$

and the Fisher information is:

$$\begin{aligned} E_0 \left[\left(\frac{\partial \log f}{\partial \theta}(X_t | X_{t-1}; A(\theta_0)) \right)^2 \right] &= E_0 \left[\left\langle D \log f(X_t | X_{t-1}; A_0), \frac{dA}{d\theta}(\theta_0) \right\rangle^2 \right] \\ &= \left(\frac{dA}{d\theta}(\theta_0), I_{X|Y} \frac{dA}{d\theta}(\theta_0) \right)_{L^2(\nu)}. \end{aligned}$$

Thus the Cramer Rao bound is given by:

$$B_A(g, \theta) = \left(\frac{dA}{d\theta}(\theta_0), I_{X|Y} \frac{dA}{d\theta}(\theta_0) \right)_{L^2(\nu)}^{-1}.$$

The solution of the maximization problem is similar to that of the cross-sectional framework, and the nonparametric efficiency bound is immediately derived.

Appendix 12 Constrained estimation

i) Asymptotic expansions.

By arguments similar to those in Appendix 8, the first order condition is given by:

$$\left(h, I_H \delta \hat{A}_T - \psi_T \right)_{L^2(\nu)} = o_p \left(\|h\|_{L^2(\nu)} / \sqrt{T} \right), \quad h \in H.$$

This is equivalent to:

$$\left(g, I_H \delta \hat{A}_T \right)_{L^2(\nu)} = (g, P_H \psi_T)_{L^2(\nu)} + o_p \left(\|g\|_{L^2(\nu)} / \sqrt{T} \right), \quad g \in L^2(\nu). \quad (\text{a.16})$$

Let us first consider the asymptotic expansion of linear functionals. Since I_H is continuously invertible we get:

$$\left(g, \delta \hat{A}_T\right)_{L^2(\nu)} = \left(g, I_H^{-1} P_H \psi_T\right)_{L^2(\nu)} + o_p\left(\|g\|_{L^2(\nu)} / \sqrt{T}\right), \quad g \in L^2(\nu).$$

Thus for any $g \in L^2(\nu)$:

$$\begin{aligned} \sqrt{T} \left(g, \delta \hat{A}_T\right)_{L^2(\nu)} &\simeq \sqrt{T} \left(g, I_H^{-1} P_H \psi_T\right)_{L^2(\nu)} \\ &= \sqrt{T} \left(I_H^{-1} P_H g, \psi_T\right)_{L^2(\nu)}, \text{ since } I_H^{-1} \text{ and } P_H \text{ commute,} \\ &\xrightarrow{d} N\left[0, \left(P_H g, I_H^{-1} P_H g\right)_{L^2(\nu)}\right]. \end{aligned}$$

Let us now consider pointwise expansions. Equation (a.16) can be generalized to the case where $g = g_T \in L^2(\nu)$, such that $\|g_T\|_{L^2(\nu)} \leq C$ independent of T , for T large enough (see Appendix 8):

$$\sqrt{T} \left(g_T, I_H \delta \hat{A}_T\right)_{L^2(\nu)} = \sqrt{T} \left(g_T, P_H \psi_T\right)_{L^2(\nu)} + O_p\left(T^{-\beta}\right), \text{ for any } g_T. \quad (\text{a.17})$$

Let us apply (a.17) with $g_T = g_{T,x_0}$, $x_0 \in [0, 1]$, as defined in Appendix 8. Let us consider g_i , $i = 1, \dots, n$, an orthonormal basis of H^\perp . We have:

$$\begin{aligned} \sqrt{T} \left(g_{T,x_0}, P_H \psi_T\right)_{L^2(\nu)} &\simeq \text{const} \sqrt{T \tilde{h}_T} \frac{d\nu}{d\lambda}(x_0) P_H \psi_T(x_0) \\ &= \text{const} \sqrt{T \tilde{h}_T} \frac{d\nu}{d\lambda}(x_0) \left[\psi_T(x_0) - \sum_{i=1}^n (g_i, \psi_T)_{L^2(\nu)} g_i(x_0) \right] \\ &\simeq \text{const} \sqrt{T \tilde{h}_T} \frac{d\nu}{d\lambda}(x_0) \psi_T(x_0), \end{aligned}$$

where the last equivalence is due to $(g_i, \psi_T)_{L^2(\nu)} = O_p(1/\sqrt{T})$. Thus we can neglect in condition (a.17) the effect of the projector P_H on ψ_T and deduce with similar arguments as in Appendix 8 iv):

$$\sqrt{T \tilde{h}_T} \alpha_{0,H}(x_0) \delta \hat{A}_T(x_0) \simeq \sqrt{T \tilde{h}_T} \frac{d\nu}{d\lambda}(x_0) \psi_T(x_0), \quad \lambda\text{-a.s. in } x_0.$$

Therefore:

$$\sqrt{T \tilde{h}_T} \delta \hat{A}_T(x_0) \simeq \alpha_{0,H}(x_0)^{-1} \sqrt{T \tilde{h}_T} \frac{d\nu}{d\lambda}(x_0) \psi_T(x_0) \xrightarrow{d} N\left[0, \left(\int K(x)^2 dx\right) \alpha_{0,H}(x_0)^{-1}\right].$$

ii) The constrained nonparametric efficiency bound.

Let $A(\cdot; \theta)$ be a one-dimensional parametric model satisfying the constraints. Then we have $dA/d\theta(\theta_0) \in H$. It follows that the Fisher information is given by:

$$\left(\frac{dA}{d\theta}(\theta_0), I_H \frac{dA}{d\theta}(\theta_0) \right)_{L^2(\nu)},$$

and the constraint:

$$(g, A(\theta))_{L^2(\nu)} = \theta, \quad \theta \simeq \theta_0,$$

is equivalent to:

$$\left(P_H g, \frac{dA}{d\theta}(\theta_0) \right)_{L^2(\nu)} = 1.$$

Problem (3.28) becomes:

$$\begin{aligned} & \min_{\delta \in H} (\delta, I_H \delta)_{L^2(\nu)}, \\ \text{s.t.} \quad & (P_H g, \delta)_{L^2(\nu)} = 1. \end{aligned}$$

As in Appendix 11 it follows:

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

iii) Proof of Proposition 22.

The proof of the boundedness is the same as the proof of proposition 1. Let us now discuss the invertibility of the information operator I_H . Operator I_H can be written as:

$$\begin{aligned} I_H h(w) &= \frac{\alpha_{0,H}(w)}{d\nu/d\lambda(w)} h(w) + \int \frac{\alpha_{1,H}(w, v)}{d\nu/d\lambda(w)} h(v) dv \\ &= I_H^0 h(w) + I_H^1 h(w). \end{aligned}$$

As in the proof of Proposition 2, operators I_H^0 and I_H^1 extend to continuous operators on $L^2(\nu)$, such that I_H^0 is continuously invertible, and I_H^1 is compact. Let \tilde{I} be the operator with domain $L^2(\nu)$ defined by:

$$\tilde{I} = I_H P_H + P_{H^\perp}.$$

Then H and H^\perp are invariant subspaces of \tilde{I} , such that $\tilde{I}|_H = I_H$, and $\tilde{I}|_{H^\perp} = Id|_{H^\perp}$. Thus, if we show that \tilde{I} is invertible, invertibility of I_H will follow. We have:

$$\begin{aligned} \tilde{I} &= (I_H^0 + I_H^1) P_H + P_{H^\perp} \\ &= I_H^0 - I_H^0 P_{H^\perp} + I_H^1 P_H + P_{H^\perp}. \end{aligned}$$

Now, using that: i) the product of a compact and a bounded operator is compact; ii) the sum of two compact operators is compact; iii) an operator with finite dimensional range is compact, we get that $-I_H^0 P_{H^\perp} + I_H^1 P_H + P_{H^\perp}$ is compact. Thus \tilde{I} is the sum of a continuously

invertible operator and a compact operator. In addition, operator \tilde{I} has a zero null space. Indeed:

$$\begin{aligned}\tilde{I}h &= 0 \implies I_H P_H h + P_{H^\perp} h = 0 \implies I_H P_H h = P_{H^\perp} h = 0 \\ &\implies P_H h = P_{H^\perp} h = 0, \text{ since } I_H \text{ has zero null space,} \\ &\implies h = 0.\end{aligned}$$

By applying Lemma A.1 in Appendix 1, \tilde{I} is invertible, and the proof is concluded.

General Bibliography

Abdous, B., Ghoudi, K., and A., Khoudraji, (2000): "Nonparametric Estimation of the Limit Dependence Function of Multivariate Extremes", *Extremes*, 2, 245-268.

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

Aït-Sahalia, Y., (1993): "The Delta and Bootstrap Methods for Nonparametric Kernel Functionals", MIT Discussion Paper.

Aït-Sahalia, Y., and A., Lo, (1998): "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices", *Journal of Finance*, 53, 499-547.

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

Ang, A., and G., Bakaert, (2002a): "Short Rate Nonlinearities and Regime Switches", *Journal of Economic Dynamics and Control*, 26, 2343-1274.

Ang, A., and G., Bekaert, (2002b): "Regime Switches in Interest Rates", *Journal of Business and Economics Statistics*, forthcoming.

Back, K., and D., Brown, (1992): "GMM, Maximum Likelihood, and Nonparametric Efficiency", *Economic Letters*, 39, 23-28.

Back, K., and D., Brown, (1993): "Implied Probabilities in GMM Estimators", *Econometrica*, 61, 971-975.

Bakshi, G., and D., Madan, (2000): "Spanning and Derivative-Security Valuation", *Journal of Financial Economics*, 55, 205-238.

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

Barrett, J., and D., Lampard, (1955): "An Expansion for Some Second Order Probability Distributions and its Application to Noise Problems", *IRE Trans. PGIT*, 10-15.

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.

Bec, F., Salem, M. B., and M., Carrasco, (2001): "Tests for Unit-Root versus Threshold Specifications with an Application to the PPP", CREST, Université de Paris 1 and University of Rochester, Working Paper.

Bec, F., Salem, M. B., and M., Carrasco, (2002): "The Purchasing Power Parity Puzzle: Evidence from a LSTAR Specification", CREST, Université de Paris 1 and University of Rochester, Working Paper.

Begun, J. M., Hall, W. J., Huang, W.-M., and J. A., Wellner, (1983): "Information and Asymptotic Efficiency in Parametric-Nonparametric Models", *Annals of Statistics*, 11, 432-452.

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.

Bickel, P. J., and M., Rosenblatt, (1973): "On Some Global Measures of Deviations of Density Function Estimate", *Annals of Statistics*, 1, 1071-1095.

Billingsley, P., (1968): *Convergence of Probability Measures*, Wiley, New-York.

Bingham, N., Goldie, C., and J., Teugels, (1987): *Regular Variation*, Cambridge University Press.

Bollerslev, T., Chou, R., and K., Kroner, (1992) "ARCH Modelling in Finance: a Review of the Theory and Empirical Evidence", *Journal of Econometrics*, 52, 5-59.

Bollerslev, T., Engle, R., and D., Nelson, (1994) "ARCH Models", in *Handbook of Econometrics, Vol. IV*, R. F. Engle and D. L. McFadden eds., North Holland, Amsterdam.

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.

Brandt, M. W., (1999): "Estimating Portfolio and Consumption Choice: A Conditional Euler Equation Approach", *Journal of Finance*, Vol. LIV, 1609-1645.

Brémaud, P., (1981): *Point Processes and Queues, Martingale Dynamics*. Springer Verlag.

Buhlman, H., Delbaen, F., Embrechts, P., and A., Shyraev, (1996): "No Arbitrage, Change of Measure and Conditional Esscher Transforms in a Semi-Martingale Model of Stock Processes", *CWI Quarterly*, 9, 291-317.

Campbell, J., Lo, A., and C., MacKinlay, (1997): *The Econometrics of Financial Markets*, Princeton University Press, New Jersey.

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.

Chamberlain, G., (1987): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions", *Journal of Econometrics*, 34, 305-334.

Chen, X., Hansen, L., and J., Scheinkman, (1999): "Shape Preserving Estimation of Diffusion", University of Chicago Working Paper.

Cherubini, U., and E., Luciano, (2001): "Pricing Vulnerable Options With Copulas", University of Turin, Working Paper.

Clark, P., (1973): "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices", *Econometrica*, 41, 135-155.

Clayton, D. G., (1978): "A Model for Association in Bivariate Life Tables and its Application in Epidemiological Studies of Family Tendency in Chronic Disease Incidence", *Biometrika*, 65, 141-151.

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.

Cox, J., Ingersoll, J., and S., Ross (1985): "A Theory of the Term Structure of Interest Rates", *Econometrica*, 53, 385-408.

Darolles, S., Florens, J.-P., and C., Gouriéroux, (2001): "Kernel Based Nonlinear Canonical Analysis", forthcoming *Journal of Econometrics*.

Darolles, S., Gouriéroux, C., and J., Jasiak, (2002): "Structural Laplace Transform and Compound Autoregressive Models", CREST Working Paper.

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

Debnath, L., and P., Mikusinski, (1998): *Introduction to Hilbert Spaces with Applications*, 2nd ed., Academic Press.

Deheuvels, P., (1981): "A Nonparametric Test for Independence", *Pub. Instit. Stat. Univ. Paris*, 26, 29-50.

Demoura, S., (1993): "Theory and Application of Transition Operator to the Analysis of Economic Time Series. Necessary and Sufficient Conditions for Nonlinearities in Economic Dynamics; Aliasing Problem, Manuscript.

Dhaene, J., and M. J., Goovaerts, (1996): "Dependency of Risks and Stop-Loss Order", *Astin Bull.*, 26, 201-212.

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

Duffie, D., (1998): "First-To-Default Valuation", University of Paris, Dauphine, and Graduate School of Business, Stanford University, Working Paper.

Duffie, D., Pan, J., and K., Singleton, (2000): "Transform Analysis and Asset Pricing for Affine Jump Diffusions", *Econometrica*, 68, 1343-1376.

Duffie, D., and K., Singleton, (1999): "Simulating Correlated Defaults", Stanford University, Working Paper.

Duffie, D., and K., Singleton, (1999): "Modelling Term Structures of Defaultable Bonds", *The Review of Financial Studies*, 12, 687-720.

Dunford, N. and J., Schwartz, (1968): *Linear Operators: Part I*, Wiley, New York.

Durrleman, V., Nikeghbali, A., and T., Roncalli, (2000): "How to Get Bounds for Distribution Convolution? A Simulation Study and an Application to Risk Management", GRO Crédit Lyonnais, Working Paper.

Embrechts, P., Höing, A., and A., Juri, (2001): "Using Copulae to Bound the Value-at-Risk for Functions of Dependent Risks", Working Paper.

Embrechts, P., Mc Neil, A. J., and D., Straumann, (1999): "Correlation: Pitfalls and Alternatives", *Risk*, 12, 93-113.

Engle, R. (1982): "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of the U. K. Inflation", *Econometrica*, 50, 987-1008.

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.

Esscher, F., (1932): "On the Probability Function in the Collective Theory of Risk", *Skandinavisk Aktuarietidskrift*, 15, 165-195.

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 and Its Applications*, Second Edition, Vol. 2., Wiley.

Fermanian, J., and O., Scaillet, (2002): "Nonparametric Estimation of Copulas for Time Series", Working Paper.

Frey, R., and A. J., McNeil (2001): "Modelling Dependent Defaults", ETH Zurich working paper.

Gagliardini, P., and C., Gouriéroux, (2002a): "Constrained Nonparametric Dependence with Applications to Finance and Insurance", CREST Working Paper.

Gagliardini, P., and C., Gouriéroux, (2002b), "Constrained Nonparametric Copulas", CREST and USI Working Paper.

Gagliardini, P., and C., Gouriéroux, (2002c), "Efficient Nonparametric Estimation of Models with Nonlinear Dependence", CREST and USI Working Paper.

Garman, M., and J., Ohlson, (1980): "Information and the Sequential Valuation of Assets in Arbitrage Free Economics", *Journal of Accounting Research*, 18, 420-440.

Genest, C., (1987): "Frank's Family of Bivariate Distributions", *Biometrika*, 74, 549-555.

Genest, C., Ghoudi, K., and L.-P., Rivest, (1995): "A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions", *Biometrika*, 82, 543-552.

Genest, C., Ghoudi, K., and L.-P., Rivest, (1998): Discussion of "Understanding Relationships Using Copulas", by E. Frees and E. Valdez, *North American Actuarial Journal*, 3, 143-149.

Genest, C., Ghoudi, K., and L.-P., Rivest, (1998): "A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions", *Biometrika*, 82, 543-552.

Genest, C., and R. J., Mc Kay, (1986): "Copules Archimédiennes et familles de lois bidimensionnelles dont les marges sont données", *Can. J. Statistics*, 14, 145-159.

Genest, C., and L.-P., Rivest, (1993): "Statistical Inference Procedures for Bivariate Archimedean Copulas", *Journal of the American Statistical Association*, 88, 1034-1043.

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).

Gill, R. D., and A. W., van der Vaart, (1993): "Non- and Semi-parametric Maximum Likelihood Estimators and the von Mises Method-II", *Scandinavian Journal of Statistics*, 20, 271-288.

Ghysels, E., Gourioux, 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., Gourioux, C., and J., Jasiak, (2002): "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.

Gorgens, T., and J., Horowitz, (1999): "Semi-parametric Estimation of a Censored Regression Model with an Unknown Transformation of the Dependent Variable", *Journal of Econometrics*, 90, 155-191.

Gourioux, C., and J., Jasiak, (1999): "Nonlinear Persistence and Copersistence", CREST and University of Toronto Working Paper.

Gourioux, C., and J., Jasiak, (2000): "Autoregressive Gamma Processes", CREST Working Paper.

Gourioux, C., and J., Jasiak, (2001a): "State-space Models with Finite Dimensional Dependence", *Journal of Time Series Analysis*, 22, 665-678.

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

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

Gourioux, C., and J., Jasiak, (2002a): "Nonlinear Autocorrelograms: An Application to Intertrade Durations", *Journal of Time Series Analysis*, 33, 127-154.

Gourieroux, C., and J., Jasiak, (2002b): *Econometric Analysis of Individual Risks*, www.dept.econ.yorku.ca/~jasiakj.

Gourieroux, C., and A., Monfort, (2001a): "Econometric Specifications of Stochastic Discount Factor Models", CREST Working Paper.

Gourieroux, C., and A., Monfort, (2001b): "Pricing with Splines", CREST Working Paper.

Gourieroux, C., and A., Monfort, (2002a): "Equidependence in Qualitative and Duration Models", CREST Working Paper.

Gourieroux, C., and A., Monfort, (2002b): "Econometrics of Efficient Portfolios", CREST Working Paper.

Gourieroux, C., and A., Monfort, (2002c): "Age and Term Structure in Duration Models", CREST Working Paper.

Gourieroux, C., Monfort, A., and V., Polimenis, (2002): "Affine Term Structure Models", CREST Working Paper.

Gourieroux, C., and C. Y., Robert, (2001): "Tails and Extremal Behaviour of Stochastic Unit Root Models", CREST Working Paper.

Gumbel, E. J., (1960): "Distributions des valeurs extrêmes en plusieurs dimensions", Publ. Inst. Statist. Univ. Paris, 9, 171-173.

Hamilton, J., (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle", *Econometrica*, 57, 357-384.

Hamilton, J., (1990): "Analysis of Time Series Subject to Changes in Regime", *Journal of Econometrics*, 45, 39-70.

Hamilton, J., (1993): "Estimation, Inference, and Forecasting of Time Series Subject to Changes in Regime", in *Handbook of Statistics, Vol 11*, G. Maddala, C. Rao, and H. Vinod, eds., North Holland, New York.

Han, A., (1987a): "Nonparametric Analysis of a Generalized Regression Model", *Journal of Econometrics*, 35, 305-316.

Han, A., (1987b): "A Nonparametric Analysis of Transformations", *Journal of Econometrics*, 35, 191-209.

Hansen, L., and S., Richard, (1987): "The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models", *Econometrica*, 55, 587-613.

Hansen, L., Scheinkman, J., and N., Touzi, (1998): "Spectral Methods for Identifying Scalar Diffusions", *Journal of Econometrics*, 86, 1-32.

Härdle, W., (1990): *Applied Nonparametric Regression*, Econometric Society Monographs, Cambridge University Press.

Härdle, W., and T., Stocker, (1989): "Investigating Smooth Multiple Regression by the Method of Average Derivatives", *JASA*, 84, 986-995.

Härdle, W., and A., Tsybakov, (1997): "Local Polynomial Estimation of the Volatility Function", *Journal of Econometrics*, 81, 223-242.

Harrison, M., and D., Kreps, (1979): "Martingales and Arbitrage in Multiperiod Securities Markets", *Journal of Economic Theory*, 20, 381-408.

Hastie, T., and R., Tibshirani, (1990): *Generalized Additive Models*, in Monographs on Statistics and Applied Probability, 43, Chapman and Hall, London.

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.

Holly, A., (1995): "A Random Linear Functional Approach to Efficiency Bounds", *Journal of Econometrics*, 65, 235-261.

Horowitz, J., (1996): "Semiparametric Estimation of a Regression Model with an Unknown Transformation of the Dependent Variable", *Econometrica*, 64, 103-137.

Horowitz, J., and W., Härdle (1996): "Direct Semiparametric Estimation of Single Index Models with Discrete Covariates", *JASA*, 91, 1632-1640.

Ichimura, M., (1993): "Semi-parametric Least-Squares and Weighted SLS Estimation of Single Index Models", *Journal of Econometrics*, 58, 71-120.

Ingersoll, J., (1987): *Theory of Financial Decision Making*, Rowman & Littlefield, Totoway NJ.

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

Jouanin, J.-F., Rapuch, G., Riboulet, G., and T., Roncalli, (2001): "Modelling Dependence for Credit Derivatives with Copulas", GRO Crédit Lyonnais, Working Paper.

Juri, A., and M. V., Wuethrich, (2002a): "Copula Convergence Theorems for Tail Events", Insurance: Mathematics and Economics, 30, 411-427.

Juri, A., and M. V., Wuethrich, (2002b): "Tail Dependence from a Distributional Point of View", ETHZ Working Paper.

Kendall, M. G., (1938): "A New Measure of Rank Correlation", Biometrika, 30, 81-93.

Kimeldorf, G., and A. R., Sampson (1975): "Uniform Representations of Bivariate Distributions", Comm. Statist., 4, 617-627.

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.

Kitamura, Y., (1997): "Empirical Likelihood Methods with Weakly Dependent Processes", Annals of Statistics, 25, 2084-2103.

Koshevnik, Y., and B., Levit, (1976): "On a Nonparametric Analogue of the Information Matrix", Theory of Probability and its Applications, 21, 738-753.

Lancaster, H., (1968): "The Structure of Bivariate Distributions", Annals of Mathematical Statistics, 29, 719-736.

Lehmann, E. L., (1966): "Some Concepts of Dependence", Ann. Math. Statist., 37, 1137-1153.

Li, D. X., (2000): "On Default Correlation: a Copula Function Approach", Journal of Fixed Income, 9, 43-54.

Madan, D., and F., Milne (1994): "Contingent Claims Valued and Hedged by Pricing and Investing in a Basis", Mathematical Finance, 4, 223-245.

Mandelbrot, R. (1963): "The Variation of Certain Speculative Prices", The Journal of Business, 36, 394-411.

Markowitz, H. (1976): *Portfolio Selection*, Yale University Press, New Heaven.

Marshall, A. W., and I., Olkin, (1967): "A Multivariate Exponential Distribution", J. Amer. Statist. Assoc., 62, 30-44.

Merton, R. C., (1974): "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates", Journal of Finance, 29, 449-470.

Meyn, S. P., and R. L., Tweedie, (1993): *Markov Chains and Stochastic Stability*, Communications and Control Engineering Series, Springer.

Morgenstern, D., (1956): "Einfache Beispiele Zweidimensionaler Verteilungen", Mitt. Math. Statist., 8, 234-235.

Müller, A., and M., Scarsini, (2000): "Some Remarks on the Supermodular Order", Journal of Multivariate Analysis, 73, 107-119.

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

Neveu, J., (1968): "Processus aléatoires gaussiens", Presses de l'Université de Montréal.

Oakes, D., (1986): "Semiparametric Inference in a Model for Association in Bivariate Survival Data", Biometrika, 73, 353-361.

Oakes, D., (1989): "Bivariate Survival Models Induced by Frailties", Journal of the American Statistical Association, 84, 487-493.

Oakes, D., (1994): "Multivariate Survival Distributions", J. Nonparametric Stat., 3, 343-354.

Owen, A., (1990): "Empirical Likelihood Ratio Confidence Regions", Annals of Statistics, 18, 90-120.

Owen, A., (2001): *Empirical Likelihood*, Monographs on Statistics and Applied Probability, Chapman & Hall.

Parzen, E., (1962): "On Estimation of a Probability Density Function and the Mode", Annals of Mathematical Statistics, 33, 1065-1076.

Patton, A., (2002): "On the Importance of Skewness and Asymmetric Dependence in Stock Returns for Asset Allocation", UCSD Working Paper.

Polimenis, V., (2002): "Essays in Discrete Time Asset Pricing", PhD, The Wharton School, University of Pennsylvania.

Powell, J., Stock, J., and T., Stocker, (1989): "Semi-parametric Estimation of Index Coefficients", Econometrica, 57, 474-523.

Quin, J., and J., Lawless, (1994): "Empirical Likelihood and General Estimating Equations", Annals of Statistics, 22, 300-325.

Rahbek, A., and N., Shephard, (2002), "Autoregressive Conditional Root Model", University of Copenhagen and Nuffield College, Working Paper.

Rapuch, G., (2001): "Dépendance stochastique et pricing des actifs contingents", Mémoire de stage de DEA de Finance et Probabilités de Paris 6.

Reed, M., and B., Simon, (1980): *Methods of Modern Mathematical Physics. I: Functional Analysis*, Academic Press.

Ridder, G., (1990): "The Nonparametric Identification of Generalized Accelerated Failure Time Models", *Review of Economic Studies*, 57, 167-182.

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

Rosenberg, J. V., (1999): "Semiparametric Pricing of Multivariate Contingent Claims", Stern School of Business Working Paper.

Rosenberg, J. V., (2001): "Nonparametric Pricing of Multivariate Contingent Claims", Stern School of Business Working Paper.

Rosenblatt, M., (1956): "Remarks on Some Nonparametric Estimates of a Density Function", *Annals of Mathematical Statistics*, 27, 832-835.

Rudin, W., (1973): *Functional Analysis*, McGraw-Hill, New York.

Schönbucher, P. J., and D., Schubert, (2001): "Copula-Dependent Default Risk in Intensity Models", University of Bonn, Working Paper.

Scott, D., (1992): *Multivariate Density Estimation. Theory, Practice and Visualization*, Wiley.

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.

Shyraev, A., (1999): *Essentials of Stochastic Finance: Facts, Models, Theory*. World Scientific Publishing, London.

Silverman, B., (1978): "Weak and Strong Consistency of the Kernel Estimate of a Density and its Derivatives", *Annals of Statistics*, 6, 177-184.

Sklar, A., (1959): "Fonctions de répartition à n dimensions et leurs marges", *Publ. Inst. Statist. Univ. Paris*, 8, 229-231.

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

Stone, C., (1983): "Optimal Uniform Rates of Convergence of Nonparametric Estimators of a Density Function or its Derivatives", in *Recent Advances in Statistics, Essays in Honor of Herman Chernoff's Sixtieth Birthday*, Academic Press, New-York.

Stutzer, M., (1995): "A Bayesian Approach to Diagnosis of Asset Pricing Models", *Journal of Econometrics*, 68, 367-397.

Stutzer, M., (1996): "A Simple Nonparametric Approach to Derivative Security Valuation", *Journal of Finance*, 51, 1633-1652.

Tchen, A. H., (1980): "Inequalities for Distributions with Given Marginals", *Annals of Statistics*, 8, 814-827.

Teräsvirta, T., Tjøstheim, D., and C., Granger, (1994): "Aspects of Modelling Nonlinear Time Series", in R. Engle and D. McFadden (eds.), *Handbook of Econometrics*, Vol. IV, Elsevier, Amsterdam.

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

Vaart, A. W. Van der, (1994): "Infinite Dimensional M-Estimation", *Prob. Theory and Math. Stat.*, 1, 1-13.

Vaart, A. W. Van der, and J. A., Wellner, (1996): *Weak Convergence and Empirical Processes*, Springer.

Van den Berg, G., (2001): "Duration Models: Specification, Identification, and Multiple Durations", in *Handbook of Econometrics, Volume V*, J. J. Heckman and E. Leamer eds., North-Holland, Amsterdam.

Von Mises, R., (1947): "On the Asymptotic Distribution of Differentiable Statistical Functions", *Annals of Mathematical Statistics*, 18, 309-348.

Wiener, N., (1958): "Nonlinear Problems in Random Theory", MIT Press, Cambridge.

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

Yosida, K., (1995): *Functional Analysis*, Sixth Edition, Springer.

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.