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

n° 2002-21

Duration Time Series Models with Proportional Hazard

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

May 2002

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

CONSTRAINED NONPARAMETRIC COPULAS

P. Gagliardini[1], C. Gourieroux[2]

July 8, 2002

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

Constrained Nonparametric Copulas

Abstract

In this paper we introduce models with constrained nonparametric dependence, for which the copula is characterized by a one-dimensional functional parameter. They provide an appropriate specification for the analysis of nonlinear dependence in financial applications, as an intermediate case between standard parametric specifications (which are in general too restrictive) and a totally unrestricted approach (which incurs in 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

Copules non paramétriques contraints

Résumé

Dans cet article nous introduisons des modèles à dépendance non paramétrique contrainte, où le copule est caractérisé par un paramètre fonctionnel de dimension un. Ils fournissent des spécifications adaptées à l'analyse des dépendances non linéaires rencontrées dans les applications financières et constituent un intermediaire entre les formulations paramétriques standard, en genéral trop contraintes, et les approches non paramétriques purs, qui butent sur les questions de manque de données. Un estimateur non paramétrique naturel est defini en minimisant la distance du khi-deux entre les densités contraintes de la famille et celle estimée sans contrainte par une méthode de noyau. Nous dérivons les propriétés asymptotiques de cet estimater et de ses fonctionnels linéaires. Nous montrons que, sous un choix approprié du paramètre fonctionnel, il est possible d'obtenir le taux de convergence espéré, c'est à dire celui non paramétrique pour la dimension 1. Finalement nous calculons les bornes d'efficacité non paramétrique et montrons que l'estimateur du khi-deux est non paramétriquement efficace.

Mots clefs: dépendance non linéaire, copule, estimation non paramétrique, efficacité non paramétrique

Classification JEL: C14, C51

1 Introduction

The copulas have been introduced as a tool for specifying the joint distribution of a pair of continuous variables X and Y. Let $F(x, y)$ denote the bivariate cumulative distribution function (c.d.f.) of (X, Y), $F_X(x)$ [resp. $F_Y(y)$] the marginal c.d.f. of X [resp. Y]. The joint c.d.f. can always be written as [Sklar (1959)]:

$$F(x, y) = C\left[F_X(x), F_Y(y)\right],$$

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

$$c(u, v) = \frac{\partial^2 C}{\partial u \partial v}(u, v),$$

is the copula p.d.f. (simply called copula in the rest of the paper). Thus it is possible 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.

There is a large literature on copulas, which focuses on the analysis of positive dependence and on the research of parametric families of copulas [see Joe (1997), and Nelsen (1999) for general presentations and the references therein]. More recently risk management in finance requires a careful analysis of dependence between risks, for instance:

· between the risk on interest rate and the default risk to analyze the term structure of the spread between T-bonds and corporate bonds,

· between the default risk of different corporates to capture the so-called default correlation, that is some clustering in corporate failure[1],

· between the risks in different budget lines of a bank's balance sheet, in order to aggregate the Value at Risk (and the required capital) computed per line[2].

This dependence is not well-captured by a gaussian model, or even by standard parametric copulas. Indeed the standard parametric copulas are not appropriate for describing the dependence between quantitative and qualitative risks (as risk on interest rate and default risk), or for performing a separate analysis of the dependence between low, medium and high risk for two quantitative risks. An alternative followed by Deheuvels (1981) and Scaillet (2001) consists in applying a completely nonparametric approach where F_X, F_Y and c are let unconstrained. However the copula is bivariate and its nonparametric estimation is not very accurate due to the curse of dimensionality.

The aim of this paper is to study constrained nonparametric copulas, which depend on a one-dimensional functional parameter a. The parameterized copula is denoted by $c(u, v; a)$, where function a is defined on $[0, 1]$. Such a constrained copula can be used for different purposes. In a cross-sectional framework, it will

[1] See e.g. Li (2000), Schönbucher, Schubert (2001), and Gourieroux, Monfort (2002).

[2] See e.g. Durrleman, Nikeghbali, Roncalli (2000), and Embrechts, Höing, Juri (2001).

be used to specify the joint distribution $F(x, y)$ of two risk variables X and Y. The bivariate distribution is generally parameterized 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, that is a time series X_t, t varying. 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 marginals are identical because of the stationarity. The bivariate distribution 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.

Since the functional parameters are one-dimensional, we can expect consistent estimators converging at the one-dimensional nonparametric rate $\sqrt{Th_T}$, where h_T is a bandwidth. However it is well-known that the rate of convergence is not invariant by one to one change of 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, with a nondegenerate differential.

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, where the joint p.d.f. is specified either by means of the conditional density and a marginal distribution, or by the copula and the marginal distributions. For each example we discuss 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. The idea is to minimize a chi-square distance between the constrained bivariate density $f(x, y; A)$ [resp. the constrained conditional density $f(x_t | x_{t-1}; A)$] and an unconstrained kernel estimator of the bivariate density [resp. the conditional density] in the cross-sectional framework [resp. in the time series framework]. 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 integrate to 1, or to identification constraints. The extension of the results to these cases is considered in section 6. Proofs are gathered in Appendices.

2 The information operator

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 univariate functions, equipped with a norm $\|.\|_{L^2(\nu)}$, where the measure ν will be precised later on [see section 2.3 iii)]. 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 explain why it is necessary to normalize the functional parameter A. This normalization 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. $D \log f(., .; A) : L^2(\nu) \to L^2(P_A)$ *is a bounded linear operator, for any $A \in \mathcal{A}$;*

ii. *the residual term $R(x, y; A, h)$ is such that $\|R(X, Y; A, h)\|_{L^2(P_A)} = o\left(\|h\|_{L^2(\nu)}\right)$, uniformly on h in the class of compact sets, for any $A \in \mathcal{A}$* [3]*.*

[3] Precisely: $\forall A \in \mathcal{A}$, $K \subset \mathcal{A}$ compact: $\|R(X, Y; A, h)\|_{L^2(P_A)} / \|h\|_{L^2(\nu)} \to 0$, uniformly in $h \in K$ [see Aït-Sahalia (1993), Van der Vaart, Wellner (1996)].

2.2 Identification and Information.

Let $A_0 \in \mathcal{A}$ denote the true value of the functional parameter, and $f(.,.) = f(.,.;A_0)$ 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:

$$Q(A) = \int\int \frac{[f(x,y) - f(x,y;A)]^2}{f(x,y)} dxdy,\ A \in \mathcal{A}.$$

Under Assumption A.2 and an additional technical condition[4], Q is well-defined in a neighborhood of A_0 (w.r.t $\|.\|_{L^2(\nu)}$) 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

Under the global identification condition:

$$f(x,y;A) = f(x,y;A_0) \ \ \lambda_2\text{-a.s.},\ A \in \mathcal{A} \Longrightarrow A = A_0,$$

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 by[5]:

$$(g, Ih)_{L^2(\nu)} = E_0 \left[\langle D \log f(X,Y;A_0), g \rangle \langle D \log f(X,Y;A_0), h \rangle \right], \qquad (1)$$

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:*

$$\langle D \log f(X,Y;A_0), h \rangle = 0 \ P_0\text{-a.s.},\ h \in L^2(\nu) \Longrightarrow h = 0.$$

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:

$$Ih = 0, h \in L^2(\nu) \Rightarrow h = 0;$$

[4] See Assumption A.2.bis in Appendix 2.

[5] see 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 (1) to those adopted in the literature.

ii. I is a positive operator:

$$(h, Ih)_{L^2(\nu)} = 0,\ h \in L^2(\nu) \ \Rightarrow h = 0.$$

Under Assumption A.3. i. and an additional technical condition[6], 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 $\Theta \subset \mathcal{A}$ containing A_0, and:

$$\forall \varepsilon > 0 : \inf_{A \in \Theta \backslash B_\varepsilon(A_0)} Q(A) > Q(A_0) = 0,$$

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 will show that, if the differential 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:

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:

$$\forall \varepsilon > 0 : \inf_{A \in \Theta \backslash B_\varepsilon(A_0)} Q(A) > Q(A_0) = 0.$$

Assumption A.3 ii. implies in particular that operator I is invertible[7].

2.3 Representation by measures

The differential operator and the information operator can often be represented in terms of measure. We discuss below the link between both representations, select the basic measure ν and characterize the invertibility of the information operator.

[6] See Assumption A.3.* in Appendix 2.

[7] Since 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 $\inf_{\lambda \in \sigma(I)} \lambda > 0$. The invertibility of I just requires $0 \notin \sigma(I)$.

i) Representation of the differential operator.

The differential operator can generally be written in terms of a measure:

$$\langle D \log f(x,y;A), h \rangle = \int h(w)^{'} \mu(x,y,A;dw), \tag{2}$$

where $\mu(x,y,A;.)$ is a q-vector of measures, $\forall x, y$. When this measure $\mu(x,y,A;.)$ has both a discrete and a continuous part, we get for instance:

$$\begin{aligned} \langle D \log f(x,y;A), h \rangle &= \gamma_0(x,y;A)^{'} h(x) + \gamma_1(x,y;A)^{'} h(y) \\ &\quad + \int \gamma_2(x,y,w;A)^{'} h(w) dw, \end{aligned} \tag{3}$$

where γ_0, γ_1 and γ_2 are $\mathbb{R}^q$-valued functions, that is:

$$\mu(x,y,A;dw) = \gamma_0(x,y;A)\delta_x(dw) + \gamma_1(x,y;A)\delta_y(dw) + \gamma_2(x,y,w;A)\lambda(dw).$$

ii) Representation of the information operator.

We can deduce the form of the information operator I when the differential operator $D \log f(X,Y;A_0)$ admits representation (2). We get:

$$(g, Ih)_{L^2(\nu)} = \int g(w)^{'} Ih(w) \nu(dw),$$

where[8]:

$$Ih(w)\nu(dw) = \int E_0 \left[\mu(X,Y;A_0;dw)\, \mu(X,Y;A_0;dv)^{'} \right] h(v),$$

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:

$$\begin{aligned} E_0 \left[\mu(X,Y;A_0;dw)\, \mu(X,Y;A_0;dv)^{'} \right] &= \alpha_0(w;A_0)\delta_w(dv)\lambda(dw) \\ &\quad + \alpha_1(w,v;A_0)\lambda_2(dv,dw), \end{aligned} \tag{4}$$

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:

$$(g, Ih)_{L^2(\nu)} = \int g(w)^{'} \alpha_0(w;A_0) h(w) dw + \int \int g(w)^{'} \alpha_1(w,v;A_0) h(v) dv dw, \tag{5}$$

[8] We assume that the integrals with respect to μ and P_0 can be commuted.

that is:

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

Thus the information operator I admits a decomposition in two components, corresponding to functions α_0 and α_1. This decomposition is common in applied examples (see section 3.2). In order to get intuition on it, let us consider the case where the differential operator admits the form (3). We get:

$$\begin{aligned}
\alpha_0(w;A) &= \int \gamma_0(w,y;A)\gamma_0(w,y;A)^{'} f(w,y)dy \\
&\quad + \int \gamma_1(x,w;A)\gamma_1(x,w;A)^{'} f(x,w)dx \\
&\equiv E\left[\gamma_{0,t}\gamma_{0,t}^{'} \mid X_t = w\right] f_X(w) + E\left[\gamma_{1,t}\gamma_{1,t}^{'} \mid Y_t = w\right] f_Y(w),
\end{aligned} \tag{6}$$

$$\begin{aligned}
\alpha_1(w,v;A) &= \gamma_0(w,v;A)\gamma_1(w,v;A)^{'} f\,(w,v) \\
&\quad + \int \gamma_0(w,z;A)\gamma_2(w,z,v;A)^{'} f(w,z)dz \\
&\quad + \int \gamma_1(z,w;A)\gamma_2(z,w,v;A)^{'} f\,(z,w)\,dz \\
&\quad + \frac{1}{2}\int\int \gamma_2(z,y,w;A)\gamma_2(z,y,v;A)^{'} f(z,y)dzdy + sym\,(w \leftrightarrow v)^{'} \\
&= \gamma_0(w,v;A)\gamma_1(w,v;A)^{'} f\,(w,v) \\
&\quad + E\left[\gamma_{0,t}\gamma_{2,t}^{'}\,(v) \mid X_t = w\right] f_X(w) \\
&\quad + E\left[\gamma_{1,t}\gamma_{2,t}^{'}\,(v) \mid Y_t = w\right] f_Y(w) \\
&\quad + \frac{1}{2}E\left[\gamma_{2,t}(w)\gamma_{2,t}^{'}\,(v)\right] + sym\,(w \leftrightarrow v)^{'}.
\end{aligned} \tag{7}$$

where $\gamma_{0,t} = \gamma_0\,(X_t,Y_t)$, $\gamma_{1,t} = \gamma_1\,(X_t,Y_t)$, $\gamma_{2,t}\,(v) = \gamma_2\,(X_t,Y_t,v)$. 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.

iii) Choice of the measure ν

Let us assume that the measure μ satisfies equation (4), 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 1 *: Assume that the measure μ satisfies equation (4). For any $A \in \mathcal{A}$, let $\alpha(.;A)$ be a positive definite matrix function such that:*

$$\int\int \left\| \alpha\left(x;A\right)^{-1/2} \alpha_1(x,y;A) \alpha\left(y;A\right)^{-1/2} \right\|^2 dxdy < \infty, \forall A, \tag{8}$$

where $\|.\|$ is a matrix norm on $\mathbb{R}^{q\times q}$. Let the measure ν be such that:

$$\forall A : \exists C_A > 0 : C_A \frac{d\nu}{d\lambda}(v) Id_q \geq \max\left\{\alpha_0(v;A), \alpha(v;A)\right\}, \ \ \forall v. \tag{9}$$

Then $D\log f(.,.;A)$ is a bounded operator from $L^2\left(\nu\right)$ to $L^2(P_A)$, for any $A \in \mathcal{A}$.
Proof. *See 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 a bivariate independent family: $f\left(x,y;A\right) = f_X(x;A)f_Y(y;A)$.
i) If parameter A consists of the marginals themselves, $A = (f_X, f_Y)$, we get:

$$\left\langle D\log f(x,y;A), h\right\rangle = \frac{h_X(x)}{f_X(x;A)} + \frac{h_Y(y)}{f_Y(y;A)}, \quad h = \left(h_X, h_Y\right)',$$

and:

$$E_0\left[\left\langle D\log f(X,Y;A_0), g\right\rangle \left\langle D\log f(X,Y;A_0), h\right\rangle\right] = \int \frac{g_X(x)h_X(x)}{f_X(x;A_0)}dx$$
$$+ \int \frac{g_Y(y)h_Y(y)}{f_Y(y;A_0)}dy + \int g_X(x)h_Y(y)dxdy + \int h_X(x)g_Y(y)dxdy.$$

Thus:

$$\alpha_0(w;A_0) = \begin{pmatrix} 1/f_X(w;A_0) & 0 \\ 0 & 1/f_Y(w;A_0) \end{pmatrix}, \ \ \alpha_1(w,v;A_0) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The choice $\alpha = \alpha_0$ satisfies condition (8) in Proposition 1. Condition (9) becomes then:

$$\forall A : \exists C_A : C_A \frac{d\nu}{d\lambda}(v) \geq \max\left\{\frac{1}{f_X(v;A)}, \frac{1}{f_Y(v;A)}\right\}, \forall v.$$

ii) If instead we choose $A = \left(f_X^{1/2}, f_Y^{1/2}\right)$, we get:

$$E_0\left[\left\langle D\log f(X,Y;A_0), g\right\rangle \left\langle D\log f(X,Y;A_0), h\right\rangle\right] = 4\left\{\int g_X(x)h_X(x)dx\right.$$
$$\left. + \int g_Y(y)h_Y(y)dy + \int \left[g_X(x)h_Y(y) + h_X(x)g_Y(y)\right] f_X(x;A)^{1/2} f_Y(y;A)^{1/2} dxdy\right\},$$

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 (8) and (9) are satisfied by $\alpha = Id_2$, $\nu = \lambda$.

iii) Finally, if $A = (\log f_X, \log f_Y)$, we get:

$$E_0 \left[\langle D \log f(X, Y; A_0), g \rangle \langle D \log f(X, Y; A_0), h \rangle \right] = \int g_X(x) h_X(x) f_X(x; A) dx$$

$$+ \int g_Y(y) h_Y(y) f_Y(y; A) dy + \int \left[g_X(x) h_Y(y) + h_X(x) g_Y(y) \right] f_X(x; A) f_Y(y; A) dx dy,$$

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 (8) in Proposition 1. Condition (9) is equivalent to:

$$\forall A : \exists C_A : C_A \frac{d\nu}{d\lambda}(v) \geq \max \left\{ f_X(v; A), f_Y(v; A) \right\}, \ \forall v,$$

that is the measure ν dominates both marginal distributions in the family.

iv) Invertibility of the information operator

When the measure μ satisfies decomposition (4) with additional restrictions on α_0, a zero null space of the information operator I is sufficient for its invertibility[9].

Proposition 2 *: Assume the conditions of Proposition 1, and in addition let $\alpha_0(v; A)$ be invertible, $\forall v, \forall A$, such that:*

$$\forall A : \exists \widetilde{C}_A > 0 : \widetilde{C}_A \frac{d\nu}{d\lambda}(v) Id_q \leq \alpha_0(v; A), \ \forall v.$$

Assume further that the information operator I has a zero null space. Then the information operator is continuously invertible.

Proof. *See Appendix 1.*

[9] The following proposition uses the theory of Fredholm operators [see Van der Vaart (1994) for similar results].

3 Examples

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 $f_{X|Y}(x \mid y; A)$ [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 $D \log f_{X|Y}$, and $D \log f_Y$ denote their differentials with respect to A. A family of bivariate densities is defined by:

$$f(x, y; A) = f_{X|Y}(x \mid y; A) f_Y(y; A).$$

We have (see Appendix 3) [10]:

Proposition 3 *: The differential of* $\log f(x, y; A)$ *is given by:*

$$D \log f(x, y; A) = D \log f_{X|Y}(x \mid y; A) + D \log f_Y(y; A).$$

Moreover:

$$\begin{aligned} D \log f_Y(y; A) &= \underset{A}{E}\left[D \log f(X, Y; A) \mid Y = y\right] \\ &= \int D \log f(x, y; A) f_{X|Y}(x \mid y; A)\, dx. \end{aligned}$$

Thus $D \log f_{X|Y}(x \mid y; A)$ is the residual in the projection of $D \log f(x, y; A)$ on Y; in particular it is orthogonal to $D \log f_Y(y; A)$:

$$\underset{A}{E}\left[\left\langle D \log f_{X|Y}(X \mid Y; A), h\right\rangle \left\langle D \log f_Y(Y; A), g\right\rangle\right] = 0, \ \ \forall h, g \in L^2(\nu). \quad (10)$$

As a consequence the information operator I is the sum of a conditional and a marginal information operators:

$$I = I_{X|Y} + I_Y,$$

[10] The 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:

$$\left\langle \int \varphi(y) D \log f_Y(y; A)\, dy, h \right\rangle := \int \varphi(y) \left\langle D \log f_Y(y; A), h\right\rangle dy.$$

Similarly for the other differentials.

where $I_{X|Y}$ and I_Y are defined by:

$$\begin{aligned}(g, I_{X|Y}h)_{L^2(\nu)} &= E_0\left[\left\langle D\log f_{X|Y}(X \mid Y; A_0), g\right\rangle \left\langle D\log f_{X|Y}(X \mid Y; A_0), h\right\rangle\right], \\ (g, I_Y h)_{L^2(\nu)} &= E_0\left[\langle D\log f_Y(Y; A_0), g\rangle \langle D\log f_Y(Y; A_0), h\rangle\right],\end{aligned}$$

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:

$$f(x; A) = \int f_{X|Y}(x \mid y; A)\, f(y; A)\, dy. \tag{11}$$

By differentiating this equation, we get the relationship satisfied by the associated differentials.

Proposition 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 $f_X(x; A)$, $f_Y(y; A)$:

$$f(x, y; A) = c\left[F_X(x; A), F_Y(y; A); A\right] f_X(x; A) f_Y(y; A).$$

Proposition 5 *: The differential of the density $f(x, y; A)$ is given by:*

$$\begin{aligned}D\log f(x, y; A) = {} & D\log c\left[F_X(x; A), F_Y(y; A); A\right] \\ & + D\log f_X(x; A) + D\log f_Y(y; A) \\ & + \frac{\partial \log c}{\partial u}\left[F_X(x; A), F_Y(y; A); A\right] \int_{-\infty}^{x} f_X(z; A) D\log f_X(z; A) dz \\ & + \frac{\partial \log c}{\partial v}\left[F_X(x; A), F_Y(y; A); A\right] \int_{-\infty}^{y} f_Y(z; A) D\log f_Y(z; A) dz.\end{aligned} \tag{12}$$

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 effect of the different functional parameters are distinguished.

Corollary 6 *: The differential of the density $f(x, y; A)$ is given by:*

$$\begin{aligned} D_a \log f(x, y; A) &= D \log c \left[F_X(x), F_Y(y); a\right], \\ \langle D_{f_X} \log f(x, y; A), h \rangle &= \frac{\partial \log c}{\partial u} \left[F_X(x), F_Y(y); a\right] \int_{-\infty}^{x} h(z) dz + \frac{h(x)}{f_X(x)}, \\ \langle D_{f_Y} \log f(x, y; A), h \rangle &= \frac{\partial \log c}{\partial v} \left[F_X(x), F_Y(y); a\right] \int_{-\infty}^{y} h(z) dz + \frac{h(y)}{f_Y(y)}. \end{aligned}$$

Let us define the information operator I_{cop} associated with the copula density:

$$(g, I_{cop} h)_{L^2(\nu)} = E_0 \left[\langle D \log c(U, V; A_0), g \rangle \langle D \log c(U, V; A_0), h \rangle\right],$$

for $h, g \in L^2(\nu)$. Since:

$$E_A \left[\langle D \log c \left(U, V; A_0\right), h \rangle \mid U\right] = E_A \left[\langle D \log c \left(U, V; A\right), h \rangle \mid V\right] = 0,$$

$\forall h \in L^2(\nu)$, the first term in the decomposition of the differential [see equation (12)] 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 $(g, I_{XY} h)_{L^2(\nu)} = E_0 \left[\langle D \log f_X(X; A_0), g \rangle \langle D \log f_Y(Y; A_0), h \rangle\right]$, 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 (12).

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]).

3.2 Examples.

We consider below different constrained nonparametric families of bivariate densities, and give the expressions of the differential of either 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 (5).

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 $X^* < y$. The conditional p.d.f. of X given Y is:

$$f(x \mid y) = \frac{f^*(x)}{\int_{-\infty}^{y} f^*(z)dz} \mathbb{I}_{x \leq y}.$$

By choosing the parametrization $A = \log f^*$, the differential of $\log f(x \mid y; A)$, for $x \leq y$, is given by:

$$\begin{aligned} \langle D \log f(x \mid y; A), h \rangle &= h(x) - \int f(z \mid y; A) h(z)dz \\ &= h(x) - E_A [h(X) \mid Y = y]. \end{aligned}$$

Let us now consider the conditional information operator $I_{X|Y}$. By definition we have:

$$\begin{aligned} \left(g, I_{X|Y} h\right)_{L^2(\nu)} &= E_0 \{(g(X) - E_0 [g(X) \mid Y]) (h(X) - E_0 [h(X) \mid Y])\} \\ &= E_0 Cov_0 (g(X), h(X) \mid Y). \end{aligned}$$

It admits the measure representation with:

$$\begin{aligned} \alpha_0 (x; A) &= f_X(x; A), \\ \alpha_1(x, y; A) &= -\int f(x \mid z; A) f(y \mid z; A) f_Y(z; A)dz. \end{aligned}$$

Let us now discuss the boundedness of the differential operator (Proposition 1). If we choose $\alpha(x; A) = f_X(x; A)$ we get:

$$\int \int \frac{\alpha_1(x, y; A)^2}{\alpha(x; A)\alpha(y; A)} dxdy = \int \int \frac{\left[\int f(x \mid z; A) f(y \mid z; A) f_Y(z; A)dz\right]^2}{f_X(x; A) f_X(y; A)} dxdy.$$

Thus condition (8) of Proposition 1 requires[11]:

$$\int \int \frac{\left[\int f(x \mid z; A) f(y \mid z; A) f_Y(z; A)dz\right]^2}{f_X(x; A) f_X(y; A)} dxdy < \infty. \tag{13}$$

Moreover the measure ν has to satisfy:

$$\forall A : \exists C_A > 0 : C_A \frac{d\nu}{d\lambda}(x) \geq f_X(x; A), \ \ \forall x. \tag{14}$$

The measure ν must dominate the marginal density of X, for any distribution in the family.

[11] Note that $\int f(x \mid z; A) f(y \mid z; A) f_Y(z; A)dz$ is the joint density of two observations of X having the same (unknown) conditioning value Y. This distribution has marginals equal to $f_X(.; A)$, and the expression in the LHS of (13) is the sum of its squared canonical correlations, see e.g. Dunford, Schwartz (1968), and Lancaster (1968).

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:

$$P\left(X_t \geq x_t \mid X_{t-1} = x_{t-1}\right) = \frac{S\left[x_t + a(x_{t-1})\right]}{S\left[a(x_{t-1})\right]}.$$

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 censored at $a(x_{t-1})$, $X^* \geq a(x_{t-1})$, and $X^* \sim S$.

Let us denote by g (resp. λ) the density (resp. the hazard function) corresponding to S. The conditional distribution is given by:

$$f\left(x_t \mid x_{t-1}; A\right) = \frac{g\left[x_t + a(x_{t-1})\right]}{\int_{a(x_{t-1})}^{+\infty} g(z)dz}, \quad A = (a, \log g)^{'}.$$

The differential is:

$$\begin{aligned}
\left\langle D_a \log f\left(x \mid y; A\right), h\right\rangle &= \left(\frac{d\log g}{dz}\left[x + a(y)\right] + \lambda\left[a(y)\right]\right) h(y), \\
\left\langle D_{\log g} \log f\left(x \mid y; A\right), h\right\rangle &= h\left(x + a(y)\right) - E_A\left[h\left(X_t + a(X_{t-1})\right) \mid X_{t-1} = y\right].
\end{aligned}$$

The information operator admits the representation (5), with local component:

$$\alpha_0\left(w; A\right) = \begin{pmatrix} E_A\left[\left(\frac{d\log g}{dz}\left[X_t + a_{t-1}\right] + \lambda\left[a_{t-1}\right]\right)^2 \mid X_{t-1} = w\right] f(w; A) & 0 \\ 0 & f_{X_t + a_{t-1}}(w; A) \end{pmatrix},$$

where $a_{t-1} = a\left(X_{t-1}\right)$ 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 Gourieroux 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:

$$X_t = \begin{cases} X_{t-1} + \varepsilon_t & \text{, with prob. } \pi\left(X_{t-1}\right), \\ \varepsilon_t & \text{, with prob. } 1 - \pi\left(X_{t-1}\right), \end{cases}$$

where the ε_t are i.i.d. errors independent from $\underline{X_{t-1}}$, 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 noise. 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.

The conditional density is given by:

$$f(x \mid y) = \pi(y) g(x - y) + [1 - \pi(y)] g(x).$$

For parameterization $A = (\log \pi, \log g)'$, the differential is given by:

$$\begin{aligned}
\langle D_{\log \pi} \log f(x \mid y; A), h \rangle &= r(x, y; A) h(y), \\
\langle D_{\log g} \log f(x \mid y; A), h \rangle &= p_1(x, y; A) h(x - y) + p_0(x, y; A) h(x),
\end{aligned}$$

where $r(x, y; A) = [f(x \mid y; A) - g(x; A)] / f(x \mid y; A)$, and p_0, p_1 are the filtering probabilities:

$$\begin{aligned}
p_1(x_t, x_{t-1}; A) &= P_A \left[Z_t = 1 \mid \underline{x_t} \right] = P_A [Z_t = 1 \mid x_t, x_{t-1}] \\
&= \pi(x_{t-1}) g(x_t - x_{t-1}) / [\pi(x_{t-1}) g(x_t - x_{t-1}) + (1 - \pi(x_{t-1})) g(x_t)],
\end{aligned}$$

and:

$$p_0(x_t, x_{t-1}; A) = P_A \left[Z_t = 0 \mid \underline{x_t} \right] = 1 - p_1(x_t, x_{t-1}; A).$$

In this example the differential operator is associated with a measure which involves a discrete component only. The information operator admits the representation (5) with:

$$\alpha_0(z; A_0) = \begin{pmatrix} E_0 \left[r_t^2 \mid X_{t-1} = z \right] f(z) & 0 \\ 0 & \begin{array}{c} E_0 \left[p_{1,t}^2 \mid X_t - X_{t-1} = z \right] f_{X_t - X_{t-1}}(z) \\ + E_0 \left[p_{0,t}^2 \mid X_t = z \right] f(z) \end{array} \end{pmatrix}$$

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 $E_0 \left[r_t^2 \mid X_{t-1} = z \right]$, that 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) Copula with proportional hazard.

Let (U_t, V_t) be variables with uniform margins, and let us assume that their distribution features proportional hazard:

$$P[U_t \geq u \mid V_t = v] = \exp[-a(v) H_0(u)],$$

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:

$$1-u = E\left[P\left[U_t \geq u \mid V_t\right]\right], \forall u \in [0,1],$$

that is

$$H_0^{-1}(z) = 1 - \int_0^1 \exp\left[-za(v)\right] dv, \; z \geq 0. \tag{15}$$

Thus the proportional hazard copula of (U_t, V_t) is characterized by the functional parameter a and it is given by:

$$c(u,v;a) = a(v)h_0(u;a)\exp\left[-a(v)H_0(u;a)\right],$$

where $H(u;a)$ is defined by (15), and $h_0 = dH_0/du$. Note that two functional parameters differing by a multiplicative constant, a and ka (say), define the same proportional hazard copula[12].

The differential of the copula density is given by [see Gagliardini, Gourieroux (2002)]:

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

where $a_{t-1} = a(U_{t-1})$, $H_{0t} = H_0(U_t, a)$,

$$\gamma_0\left(u,v;a\right) = \frac{1-a(v)H_0(u;a)}{a(v)},$$

and γ_1 is given in Gagliardini, Gourieroux (2002), formula (a.13), Appendix 7. The information operator admits the form (5) with local component:

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

and α_1 given in Appendix 8 of Gagliardini, Gourieroux (2002) [13] [14].

[12] Gagliardini, Gourieroux (2002) use the proportional hazard copula for specifying duration time series models with proportional hazard, and for discussing their serial dependence properties.

[13] It is possible to consider the example of general distributions (X,Y) with proportional hazard:

$$P\left[X \geq x \mid Y = y\right] = \exp\left[-a\left(y\right)\Lambda(x)\right],$$

where a is a positive function, and Λ is the baseline cumulated hazard.

[14] The results on proportional hazard copula can be extended to more general transformation copulas, that is the c.d.f. of variables (U_t, V_t) with uniform margins and 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)]:

$$C(u,v) = \phi\left[\phi^{-1}(u) + \phi^{-1}(v)\right], \tag{16}$$

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 (for failure, called frailty models, see e.g. Joe [1997]). In this case ϕ is the Laplace transform of a positive random variable Z representing a latent factor with common effect on X and Y:

$$\phi(s) = E\left[\exp\left(-sZ\right)\right], \; s \geq 0. \tag{17}$$

Assume ϕ is twice continuously differentiable. The copula p.d.f. is:

$$c(u,v) = \frac{\phi^{''}\left[\phi^{-1}(u) + \phi^{-1}(v)\right]}{\phi^{'}\left[\phi^{-1}(u)\right]\phi^{'}\left[\phi^{-1}(v)\right]}.$$

However 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 7 : *The joint p.d.f. of W and Z is given by:*

$$f(w,z) = \frac{f^*(w)}{\int_0^z f^*(v)dv} 1_{w \leq z}, \;\; w,z \in (0,1),$$

where the latent measure density f^* is given by:

$$f^*(w) = -\frac{\phi^{''}\left[\phi^{-1}(w)\right]}{\phi^{'}\left[\phi^{-1}(w)\right]}, \;\; w \in (0,1). \tag{18}$$

Moreover we have a one-to-one relationship between the measure F^* and the generator ϕ^{-1} since:

$$F^*(w) = -\phi^{'}\left[\phi^{-1}(w)\right] \iff \phi^{-1}(y) = \int_y^1 \frac{dw}{F^*(w)},$$

under the condition $\int_0^1 1/F^*(w)dw = \infty$ [15].

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 (18) and $Z \sim U(0,1)$.

We choose to parameterize the copula density by means of function $a = f^*$. Thus the copula density is:

$$c(u, v; a) = a\left[C(u, v; a)\right] \frac{F^*\left[C(u, v; a); a\right]}{F^*\left(u; a\right) F^*\left(v; a\right)},$$

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:

$$\langle D \log c(u, v; a), h \rangle = \frac{h\left[C(u, v; a)\right]}{a\left[C(u, v; a)\right]} + \int_0^1 \gamma\left(u, v, w; a\right) h(w)dw,$$

where function γ is given in Appendix 4. The information operator is of the form (5), where the local component is given by:

$$\alpha_0(w, a) = \frac{f_W(w; a)}{a(w)^2} = \frac{\phi^{-1}(w; a)}{a(w)},$$

where $f_W(.; a)$ 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), $i = 1, ..., n$ be independent pairs of random variables. Extreme value bivariate copulas are associated with the limiting joint distribution of

[15] By 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:

$$f^*\left(0\right) = \lim_{w \to 0} f^*\left(w\right) = \lim_{w \to \infty} -\frac{\phi^{''}(w)}{\phi^{'}(w)},$$

this condition is equivalent to:

$$\lim_{w \to \infty} -\frac{\phi^{''}(w)}{\phi^{'}(w)} < \infty.$$

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:

$$\begin{aligned} c_\chi(u,v) &= \frac{C(u,v)}{uv}\left\{-\frac{\widetilde{u}\widetilde{v}}{\log u + \log v}\chi^{''}(\widetilde{u})\right. \\ &\quad \left. + \left[\chi(\widetilde{u}) + \widetilde{v}\chi^{'}(\widetilde{u})\right]\left[\chi(\widetilde{u}) - \widetilde{u}\chi^{'}(\widetilde{u})\right]\right\}, \end{aligned}$$

where $\widetilde{u} = \log u / (\log u + \log v)$, $\widetilde{v} = \log v / (\log u + \log v)$. 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 $\chi^{''}$. 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 $a = f^* = \chi^{''}/2$:

$$\begin{aligned} c(u,v;a) &= \frac{C(u,v;a)}{uv}\left\{-\frac{2\widetilde{u}\widetilde{v}}{\log u + \log v}a(\widetilde{u})\right. \\ &\quad \left. + \left[\chi(\widetilde{u};a) + \widetilde{v}\chi^{'}(\widetilde{u};a)\right]\left[\chi(\widetilde{u};a) - \widetilde{u}\chi^{'}(\widetilde{u};a)\right]\right\}, \end{aligned}$$

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:

$$\langle D\log c(u,v;a), h\rangle = \gamma_0(u,v;a)\,h(\widetilde{u}) + \int_0^1 \gamma_1(u,v,w;a)\,h(w)dw,$$

where:

$$\gamma_0(u,v;a) = \left\{ a(\widetilde{u}) - \frac{\log u + \log v}{2\widetilde{u}\widetilde{v}} \left[1 - \int_0^{\widetilde{u}} wa(w)dw \right] \left[\int_0^{\widetilde{u}} a(w)dw - \int_0^{\widetilde{u}} wa(w)dw \right] \right\}^{-1}.$$

The copula information operator admits representation (5) with local component:

$$\alpha_0(w;a) = E_a\left[\gamma_0(U,V;a)^2 \mid \widetilde{U} = w\right] f_{\widetilde{U}}(w;a),$$

where $\widetilde{U} = \log U / (\log U + \log V)$, and $f_{\widetilde{U}}$ is the density of $\widetilde{U}$.

vii) Copula with one 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 dependence[16]. Functions φ_j, ψ_j, j varying, are called canonical directions, and $\lambda_j = corr\left[\varphi_j(X_t), \psi_j(X_{t-1})\right]$, j varying, are the associated canonical correlations. The copula of a stationary Markov process with one dimensional canonical decomposition is obtained when $\lambda_j = 0$, $j \geq 2$, and $\lambda_1 = \lambda > 0$ [see Gourieroux, Jasiak (2001)]. It 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:

$$\int_0^1 a(v)dv = 0.$$

[16] See Lancaster (1968), and Dunford, Schwartz (1968).

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 admits the form (3) and it is given by:

$$D \log c(u, v; a) = \frac{a(v)}{1 + a(u)a(v)} h(u) + \frac{a(u)}{1 + a(u)a(v)} h(v).$$

From (6) and (7) the information operator admits representation (5), 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.

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.

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)]:

$$\widehat{f}_T(x, y) = \frac{1}{Th_T^2} \sum_{t=1}^{T} K\left(\frac{x - X_t}{h_T}\right) K\left(\frac{y - Y_t}{h_T}\right), \tag{19}$$

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:

$$\sqrt{Th_T^2}\left[\widehat{f}_T(x,y) - f(x,y;A)\right] \xrightarrow{d} N\left(0, \sigma^2\left(x,y;A\right)\right), \tag{20}$$

where $\sigma^2\left(x,y;A\right) = f(x,y;A)\left(\int K^2(w)dw\right)^2$. 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:

$$\sqrt{Th_T}\left[\int g(x)\widehat{f}_T(x,y)dx - \int g(x)f(x,y;A)dx\right] \xrightarrow{d} N\left(0, \sigma^2\left(y,g;A\right)\right), \tag{21}$$

where $\sigma^2\left(y,g;A\right) = \underset{A}{E}\left[g(X_t)^2 \mid Y_t = y\right] f_Y(y) \int K^2(w)dw$, and

$$\sqrt{T}\left[\int\int g(x,y)\widehat{f}_T(x,y)dxdy - \int\int g(x,y)f(x,y;A)dxdy\right] \xrightarrow{d} N\left(0, \sigma^2\left(g;A\right)\right), \tag{22}$$

where $\sigma^2\left(g\right) = V_A\left[g\left(X_t, Y_t\right)\right]$.

The unconstrained estimator of the bivariate density can be used to derive a minimum chi-square estimator of A:

$$\widehat{A}_T = \arg\min_{A\in\Theta} Q_T(A) = \int_0^1\int_0^1 \frac{\left[\widehat{f}_T(x,y) - f(x,y;A)\right]^2}{\widehat{f}_T(x,y)}\omega_T(x,y)dxdy, \tag{23}$$

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 $\widehat{A}_T$ 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 $\|.\|_{L^2(\nu)}$; or*

ii. *the criterion Q_T is weakly lower semicontinuous and the set Θ is bounded and closed with respect to the norm $\|.\|_{L^2(\nu)}$* [17].

[17] Let $(X, \|.\|)$ 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^*: $l\left(x_n\right) \to l(x)$. A function Φ on X is weakly lower semicontinuous (w.l.s.c.) if: $x_n \xrightarrow{w} x$ implies $\Phi(x) \leq \liminf \Phi\left(x_n\right)$. 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:

$$\widehat{f}_T^0(x,y) = f(x,y;\widehat{A}_T). \tag{24}$$

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 $\widehat{f}$ or a distance between f^* and the transformation of $\widehat{f}$ 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^*.*

4.2 Consistency of the estimators

Let us consider the consistency of the minimum chi-square estimator $\widehat{A}_T$. In Appendix 6 it is shown that under the following two assumptions and additional regularity conditions (see Assumptions A.7 - 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 $\widetilde{\Omega}_T, \Omega_T$ such that $\widetilde{\Omega}_T \subset \Omega_T \subset [0,1]^2$, ω_T has support in Ω_T, is smaller than 1 with restriction $\omega_T|_{\widetilde{\Omega}_T} = 1$, $T \in N$, and $\lambda_2\left(\widetilde{\Omega}_T\right) \to 1$, as $T \to \infty$, 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:

$$E_0\left[\langle D\log f(X,Y;A), g\rangle \langle D\log f(X,Y;A), h\rangle\right] = (g, I_A h)_{L^2(\nu)},$$

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 8 *: Under Assumptions A.1-A.11* [18] *the chi-square estimator $\widehat{A}_T$ is consistent in norm:*

$$\left\|\widehat{A}_T - A_0\right\|_{L^2(\nu)} \xrightarrow{p} 0.$$

[18] We assume that either A.3 i. and A.5 i., or A.3 ii. and A.5 ii. hold.

Let us now consider the constrained density estimator $\widehat{f}_T^0$, and show its consistency in L^1-norm[19]. Uniform convergence of Q_T to Q implies convergence of $Q(\widehat{A}_T)$ to $Q(A_0) = 0$. By using the Cauchy-Schwarz inequality:

$$\left\| f(.,.;\widehat{A}_T) - f(.,.) \right\|_{L^1} \leq \left\| \frac{f(.,.;\widehat{A}_T) - f(.,.)}{\sqrt{f(.,.)}} \right\|_{L^2} \left\| \sqrt{f(.,.)} \right\|_{L^2} = Q(\widehat{A}_T)^{1/2},$$

we deduce the following proposition.

Proposition 9 *: Under Assumptions of Proposition 8, the constrained density estimator $\widehat{f}_T^0$ is consistent in L^1 norm:*

$$\left\| \widehat{f}_T^0 - f \right\|_{L^1} \xrightarrow{p} 0.$$

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 sense[20].

Assumption A.12 *For any $g \in L^2(\nu)$: $\widehat{A}_T + tg \in \Theta$ 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.

Proposition 10 *: Under Assumptions A.1-A.15 the minimum chi-square estimator $\widehat{A}_T$ is such that:*

$$I\left(\widehat{A}_T - A_0\right) \simeq \psi_T, \tag{25}$$

where the efficient score $\psi_T \in L^2(\nu)$ is defined by[21]*:*

$$(\psi_T, h)_{L^2(\nu)} = \int\int \delta\widehat{f}_T(x,y)\,\omega_T(x,y)\,\langle D\log f(x,y;A_0), h\rangle\,dxdy,\ h \in L^2(\nu),$$

[19] Let $\Omega \subset [0,1]^2$ 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 $L^p \equiv L^p([0,1]^2)$.

[20] This 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.

[21] The differential operator $D\log f(x,y;A_0)$ smoothed by the kernel density estimator, that is $\int\int \delta\widehat{f}_T(x,y)\,\omega_T(x,y)\,D\log f(x,y;A_0)dxdy$ 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.

where $\delta \widehat{f}_T = \widehat{f}_T - f$.

As an example, when the differential operator is of the form (3), function ψ_T is given by:

$$\frac{d\nu}{d\lambda}(w)\psi_T(w) = \int \delta \widehat{f}_T(w,y)\omega_T(w,y)\gamma_0(w,y)\,dy + \int \delta \widehat{f}_T(x,w)\omega_T(x,w)\gamma_1(x,w)dx + \int\int \delta \widehat{f}_T(x,y)\omega_T(x,y)\gamma_2(x,y,w)dxdy. \quad (26)$$

Moreover when the information operator admits the representation (5), the first order condition is equivalent to:

$$\alpha_0(w)\delta \widehat{A}_T(w) + \int \alpha_1(w,v)\delta \widehat{A}_T(v)dv \simeq \frac{d\nu}{d\lambda}(w)\psi_T(w), \quad (27)$$

where $\delta \widehat{A}_T = \widehat{A}_T - A_0$. 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 iv) for sufficient conditions].

Corollary 11 : *When I is invertible and continuous at zero:*

$$\widehat{A}_T - A_0 \simeq I^{-1}\psi_T. \quad (28)$$

Since $I = D\log f_0^* D\log f_0$ and $\psi_T = D\log f_0^*\left(\omega_T \delta \widehat{f}_T/f\right)$, where $D\log f_0^*$ denotes the adjoint of the differential operator $D\log f_0 \equiv D\log f(.,.;A_0)$, the asymptotic expansion in (28) can be written as:

$$\widehat{A}_T - A_0 \simeq \left[D\log f_0^* D\log f_0\right]^{-1} D\log f_0^*\left(\omega_T \delta \widehat{f}_T/f\right),$$

that is a regression of the "errors" $\delta \widehat{f}_T/f$ 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 12 : *The constrained estimator is such that:*

$$\widehat{f}_T^0(x,y) - f(x,y) \simeq \left\langle Df(x,y;A_0), \delta \widehat{A}_T \right\rangle.$$

4.4 The asymptotic distribution of the minimum chi-square estimator

The asymptotic distribution of the minimum chi-square estimator $\widehat{A}_T$ is derived from the asymptotic expansion given in Corollary 11. To simplify the presentation we assume decomposition into measures of both differential and information operators. We distinguish the pointwise estimation of A and the estimation of linear functionals of A, such as $\int_0^1 g(w)' \widehat{A}_T(w)\nu(dw)$, 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 (27). For pointwise estimation, the second term of order $1/\sqrt{T}$ can be neglected leading to [see Appendix 8 iv)]:

$$\sqrt{Th_T}\delta\widehat{A}_T(w) \simeq \alpha_0(w)^{-1}\sqrt{Th_T}\frac{d\nu}{d\lambda}(w)\psi_T(w).$$

When the differential operator admits the measure representation (3) we directly deduce from (26), (21), and (22) that $\sqrt{Th_T}\psi_T(w)$ is pointwise asymptotically normal (see Appendix 9).

Lemma 13 *: When the differential admits the measure decomposition (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 $\widehat{A}_T$ follows.

Proposition 14 *: Under Assumptions A.1-A.15 the estimator $\widehat{A}_T$ is λ-a.s. pointwise asymptotically normal:*

$$\sqrt{Th_T}\left(\widehat{A}_T(w) - A_0(w)\right) \xrightarrow{d} N\left(0, \left(\int K^2(x)dx\right)\alpha_0(w)^{-1}\right),$$

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), 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.

Corollary 15 *: The constrained estimator* $\sqrt{Th_T}\left(\widehat{f}_T^0(x,y)-f(x,y)\right)$ *is asymptotically normal, with asymptotic variance:*

$$\left(\int K^2(x)dx\right) f\left(x,y\right)^2\left[\gamma_0\left(x,y\right)\alpha_0\left(x\right)^{-1}\gamma_0\left(x,y\right)^{'}+\gamma_1\left(x,y\right)\alpha_0\left(y\right)^{-1}\gamma_1\left(x,y\right)^{'}\right].$$

In particular the constrained estimator has a one-dimensional nonparametric convergence rate, and:

$$\begin{aligned}\sqrt{Th_T^2}\left[\widehat{f}_T(x,y)-\widehat{f}_T^0(x,y)\right] &\simeq \sqrt{Th_T^2}\left[\widehat{f}_T(x,y)-f(x,y)\right]\\ &\xrightarrow{d} N\left[0,f(x,y)\left(\int K^2(w)dw\right)^2\right].\end{aligned}$$

The discrepancy $\sqrt{Th_T^2}\left[\widehat{f}_T(x,y)-\widehat{f}_T^0(x,y)\right]$, 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 $G=\int g(v)^{'}A_0(v)\nu\left(dv\right)$, with $g\in L^2\left(\nu\right)$. We expect the estimator $\widehat{G}_T=\int g(v)^{'}\widehat{A}_T(v)\nu\left(dv\right)$ to have a parametric rate, so that the second term of equation (27), which is of order $1/\sqrt{T}$, can no longer be neglected. We deduce from Corollary 11:

$$\begin{aligned}\sqrt{T}\left(\widehat{G}_T-G\right) &= \sqrt{T}\int g(v)^{'}\delta\widehat{A}_T(v)\nu\left(dv\right)=\sqrt{T}\left(g,\delta\widehat{A}_T\right)_{L^2(\nu)}\\ &\simeq \sqrt{T}\left(g,I^{-1}\psi_T\right)_{L^2(\nu)},\text{ from (28),}\\ &= \sqrt{T}\left(I^{-1}g,\psi_T\right)_{L^2(\nu)},\ \text{ since } I^{-1} \text{ is self-adjoint on } L^2\left(\nu\right).\end{aligned}$$

The following Lemma provides the asymptotic distribution of $\sqrt{T}\left(g,\psi_T\right)_{L^2(\nu)}$, $g\in L^2\left(\nu\right)$.

Lemma 16 *For* $g\in L^2\left(\nu\right)$*:*

$$\sqrt{T}\left(g,\psi_T\right)_{L^2(\nu)}\xrightarrow{d} N\left[0,\left(g,Ig\right)_{L^2(\nu)}\right].$$

Proof. *We have:*

$$\begin{aligned}\sqrt{T}\left(g,\psi_T\right)_{L^2(\nu)} &= \sqrt{T}\int\int\delta\widehat{f}_T(x,y)\omega_T(x,y)\left\langle D\log f(x,y;A_0),g\right\rangle dxdy\\ &\simeq \sqrt{T}\int\int\delta\widehat{f}_T(x,y)\left\langle D\log f(x,y;A_0),g\right\rangle dxdy.\end{aligned}$$

By using (22), the latter expression is asymptotically normal. Its variance is given by:

$$\begin{aligned} \sigma^2(g) &= V_0\left[\langle D\log f(X_t, Y_t; A), g\rangle\right] \\ &= E_0\left[\langle D\log f(X_t, Y_t; A), g\rangle^2\right] \\ &= (g, Ig)_{L^2(\nu)}. \end{aligned}$$

Q.E.D.

The asymptotic distribution of a linear functional follows.

Proposition 17 *Under Assumptions A.1-A.15 the estimator* $\widehat{G}_T = \int g(v)' \widehat{A}_T(v)\nu(dv)$ *of a linear functional of* A *is asymptotically normal, with parametric rate of convergence:*

$$\sqrt{T}\left(\widehat{G}_T - G\right) \xrightarrow{d} N\left(0, \left(g, I^{-1}g\right)_{L^2(\nu)}\right).$$

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* $f(x \mid y; A)$*, and* β*-mixing coefficients such that:* $\beta_k = O\left(k^{-\delta}\right)$*,* $\delta > 1$*. 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:

$$\widehat{A}_T = \arg\min_{A\in\Theta} Q_T(A) = \int_0^1\int_0^1 \frac{\left[\widehat{f}_T(x|y) - f(x|y;A)\right]^2}{\widehat{f}_T(x|y)}\omega_T(x,y)\widehat{f}_{Y,T}(y)dxdy. \tag{29}$$

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 information operator $D\log f(x|y;A)$. They are denoted by adding TS.

Proposition 18 *: Under Assumptions A.1.TS-A.11.TS the minimum chi-square estimator* $\widehat{A}_T$ *is consistent.*

The asymptotic expansion of the chi-square estimator in the time series framework is given by:

$$I_{X|Y}\left(\widehat{A}_T - A_0\right) \simeq \widetilde{\psi}_T,$$

where the function $\widetilde{\psi}_T \in L^2(\nu)$ is defined by:

$$\left(\widetilde{\psi}_T, h\right)_{L^2(\nu)} = E_0\left[\frac{\delta \widehat{f}_T(X \mid Y)}{f(X \mid Y)}\omega_T(X,Y)\left\langle D\log f(X \mid Y; A_0), h\right\rangle\right],\ h \in L^2(\nu).$$

In particular, when the conditional information operator $I_{X|Y}$ admits a measure representation with $\widetilde{\alpha}_0$, $\widetilde{\alpha}_1$, say, the asymptotic expansion becomes:

$$\widetilde{\alpha}_0(w)\delta\widehat{A}_T(w) + \int \widetilde{\alpha}_1(w,v)\,\delta\widehat{A}_T(v)dv \simeq \frac{d\nu}{d\lambda}(w)\,\widetilde{\psi}_T(w).$$

The asymptotic distribution of $\widehat{A}_T$ is immediately deduced from that of $\widetilde{\psi}_T$:

$$\sqrt{Th_T}\frac{d\nu}{d\lambda}(w)\,\widetilde{\psi}_T(w) \xrightarrow{d} N\left[0, \left(\int K^2(x)dx\right)\widetilde{\alpha}_0(w)\right],\ \ \lambda\text{-a.s. in } w,$$

$$\sqrt{T}\left(g, \widetilde{\psi}_T\right)_{L^2(\nu)} \xrightarrow{d} N\left[0, \left(g, I_{X|Y}g\right)_{L^2(\nu)}\right],\ \text{ for } g \text{ in } L^2(\nu).$$

Note that the asymptotic variance $\left(g, I_{X|Y}g\right)_{L^2(\nu)} = V_0\left[\left\langle D\log f(X_t \mid X_{t-1}; A_0), g\right\rangle\right]$ includes no cross-term, since $\left\langle D\log f(X_t \mid X_{t-1}; A_0), g\right\rangle$ is a martingale difference sequence.

We deduce:

Proposition 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:

$$\sqrt{T}\left(g, \widehat{A}_T - A_0\right)_{L^2(\nu)} \xrightarrow{d} N\left[0, \left(g, I^{-1}_{X|Y}g\right)_{L^2(\nu)}\right],\ \text{ for } g \text{ in } L^2(\nu).$$

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.

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:

$$B_A(g) = \max B_A(g,\theta),\ g \text{ varying},$$

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:

$$\begin{aligned} B_A(g) &= \max B_A(g,\theta), \\ \text{s.t} &: \int g(v)^{'} \frac{\partial A}{\partial \theta}(v;\theta_0)\,\nu(dv) = 1, \end{aligned} \tag{30}$$

g varying, where maximization is performed over all parameterizations $A(.,\theta)$.

Proposition 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:

$$(g, Ih)_{L^2(\nu)} = E_0\left[\langle D\log f(X,Y;A_0), g\rangle \langle D\log f(X,Y;A_0), h\rangle\right].$$

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:

$$\left(g, I_{X|Y}h\right)_{L^2(\nu)} = E_0\left[\langle D\log f(X_t \mid X_{t-1};A_0), g\rangle \langle D\log f(X_t \mid X_{t-1};A_0), h\rangle\right].$$

From Propositions 17 and 19, we immediately deduce that the estimator $\widehat{G}_T = \int g(v)^{'}\widehat{A}_T(v)\nu(dv)$ reaches this bound.

Corollary 21 *: The minimum chi-square estimator $\widehat{A}_T$ 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).

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: $\int a(v)dv = 1$.

6.1 Restricted information operator.

Let us assume that functional parameter A satisfies n linear constraints:

$$\int A(v)' g_i(v)\nu(dv) = (A, g_i)_{L^2(\nu)} = k_i, \;\; i = 1, ..., n,$$

where $g_i \in L^2(\nu)$, $k_i \in \mathbb{R}$, $i = 1, ..., n$. Let us denote by $\widetilde{\mathcal{A}} \subset \mathcal{A}$ the subset of functional parameters satisfying these restrictions. The tangent space H of $\widetilde{\mathcal{A}}$ at $A_0 \in \widetilde{\mathcal{A}}$ does not depend on A_0, has a finite codimension, and it is given by:

$$H = \{h \in L^2(\nu) : (h, g_i)_{L^2(\nu)} = 0, \quad i = 1, ..., n\}.$$

The differential operator $D\log f(.,.;A_0)$ can be restricted to the linear space $H \subset L^2(\nu)$, and we assume that $D\log f(.,.;A_0) : H \to L^2(P_0)$ is a bounded operator. The information operator I_H is the bounded linear operator from H in itself defined by:

$$(g, I_H h)_{L^2(\nu)} = E_0 \left[\langle D\log f(X, Y; A_0), g\rangle \langle D\log f(X, Y; A_0), h\rangle\right], \;\; h, g \in H.$$

Then we can extend to the constrained framework the definitions of identification and measure decomposition.

i) Local identification

Let us introduce the following local identification condition:

Assumption A.3. *i. Local identification:*

$$\langle D\log f\left(X,Y;A_0\right),h\rangle = 0 \ P_0\text{-a.s.},\ h\in H \Longrightarrow h=0.$$

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 $\widetilde{\Theta}\subset\widetilde{\mathcal{A}}$ 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) Measure decomposition

When the information operator I_H admits a measure decomposition:

$$I_H h\left(w\right)=\frac{\alpha_{0,H}(w)}{d\nu/d\lambda(w)}h(w)+\int\frac{\alpha_{1,H}(w,v)}{d\nu/d\lambda(w)}h(v)dv,\ \ h\in H,$$

it is possible to characterize boundedness and invertibility of I_H in terms of $\alpha_{0,H}$ and $\alpha_{1,H}$ [22].

Proposition 22 *:*

i. Assume that for any A there exists a positive definite matrix $\alpha_H(.,A)$ such that:

$$\int\int\left\|\alpha_H\left(x;A\right)^{-1/2}\alpha_{1,H}(x,y;A)\alpha_H\left(y;A\right)^{-1/2}\right\|^2 dxdy<\infty,\forall A,\quad(31)$$

where $\|.\|$ is a matrix norm on $\mathbb{R}^{q\times q}$. Let the measure ν be such that:

$$\forall A:\exists C_A>0: C_A\frac{d\nu}{d\lambda}(v)Id_q\geq\max\left\{\alpha_{0,H}(v;A),\alpha_H(v;A)\right\},\ \ \forall v.\quad(32)$$

Then I_H is a bounded operator from H in itself.

[22] Let $I: L^2(\nu)\to L^2(\nu)$ denote the unrestricted information operator defined by the differential $D\log f(.,.;A_0)$ with domain $L^2(\nu)$. Since $I_H=P_H I P_H = I - P_{H^\perp}I - IP_{H^\perp} - P_{H^\perp}IP_{H^\perp}$, 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 measure decomposition if I has such a decomposition. Moreover, both decompositions have identical local component: $\alpha_{0,H}=\alpha_0$.

ii. Assume further that $\alpha_{0,H}(v;A)$ is invertible, $\forall v, \forall A$, and such that:

$$\forall A : \exists \widetilde{C}_A > 0 : \widetilde{C}_A \frac{d\nu}{d\lambda}(v) Id_q \leq \alpha_{0,H}(v;A), \;\; \forall v. \tag{33}$$

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:

$$H = \left\{ h \in L^2(\nu) : \int_0^1 h(v)dv = 0 \right\}.$$

Boundedness and invertibility of the copula information operator I_H^c on H is discussed in Gagliardini, Gourieroux (2002) 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:

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

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.

6.2 The minimum chi-square estimator.

Let $\widetilde{\Theta}$ be a subset of $\widetilde{\mathcal{A}}$. The minimum chi-square estimator is obtained by minimizing the chi-square divergence under the constraints:

$$\widehat{A}_T = \arg\min_{A \in \widetilde{\Theta}} Q_T(A) = \int_0^1 \int_0^1 \frac{\left[\widehat{f}_T(x,y) - f(x,y;A) \right]^2}{\widehat{f}_T(x,y)} \omega_T(x,y) dx dy. \tag{34}$$

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 $\widehat{A}_T$ satisfies the first order condition in the sense that $\widehat{A}_T + th \in \widetilde{\Theta}$ with probability approaching to 1, for t small enough, for any $h \in H$. The first order condition is equivalent to (see Appendix 12):

$$\left(h, I_H \delta \widehat{A}_T - \psi_T\right) \simeq 0, \forall h \in H,$$

that is:

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

If the operator I_H is invertible, the asymptotic expansion of $\widehat{A}_T$ is:

$$\widehat{A}_T - A_0 \simeq I_H^{-1} P_H \psi_T.$$

By using:

$$\sqrt{T} \left(h, \psi_T\right)_{L^2(\nu)} \xrightarrow{d} N\left[0, \left(h, I_H h\right)_{L^2(\nu)}\right], \quad h \in H,$$

we get (see Appendix 12):

Proposition 23 *: Under Assumptions A.1-A.15:*

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

When the differential operator admits a measure decomposition (3):

$$\sqrt{T h_T}\left(\widehat{A}_T(v) - A_0(v)\right) \xrightarrow{d} N\left(0, \left(\int K^2(x) dx\right) \alpha_{0,H}(v)^{-1}\right),$$

λ-a.s in v.

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 $A \in \widetilde{\mathcal{A}}$.

Proposition 24 *: The nonparametric efficiency bound is given by:*

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

The constrained minimum chi-square estimator is therefore nonparametrically efficient.

7 Concluding remarks.

The analysis of nonlinear dependence is crucial for financial applications and requires an appropriate copula specification. To avoid the curse of dimensionality the copula cannot be let unconstrained. At the opposite a standard parametric specification of the copula is generally too restrictive to get the expected fit. In this paper we have considered the intermediate case in which the copula depends on a one-dimensional functional parameter. 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 copula.

Appendix 1
The information operator

i) Definition.

Let us relate the definition of the information operator given in (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 $f^{1/2} : L^2(\nu) \to L^2(P_0)$ is differentiable at $h = 0$, with continuous derivative:

$$df_0^{1/2} : L^2(\nu) \to L^2(P_0).$$

Then, following Begun, Hall, Huang, Wellner [1983], and Gill, Van der Vaart [1993], the information operator can be defined as:

$$I = df_0^{1/2*} df_0^{1/2} : L^2(\nu) \to L^2(\nu).$$

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 $f^{1/2} : L^2(\nu) \to L^2(P_0)$ 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:

$$E_0 \left[\langle D \log f(X,Y;A_0), g \rangle \langle D \log f(X,Y;A_0), h \rangle \right] = (g, Ih)_{L^2(\nu)}, \quad h, g \in L^2(\nu).$$

This is the definition adopted in our paper, and in Holly (1995).

ii) Choice of the measure ν

Let us prove Proposition 1. We have:

$$\begin{aligned}
\|\langle D\log f(X,Y;A), h\rangle\|^2_{L^2(P_A)} &= E_A\left[\langle D\log f(X,Y;A), h\rangle^2\right] \\
&= \int h(v)'\alpha_0(v;A)\,h(v)dv + \int\int h(v)'\alpha_1(v,w;A)\,h(w)dvdw.
\end{aligned}$$

Both terms are easily bounded. For the first one we get:

$$\int h(v)'\alpha_0(v;A)\,h(v)dv \leq C_A\int h(v)'h(v)\nu(dv) = C_A\|h\|^2_{L^2(\nu)}.$$

Let us now consider the second one, and denote:

$$k_A = \left(\int\int\left\|\alpha(v;A)^{-1/2}\alpha_1(v,w;A)\,\alpha(w;A)^{-1/2}\right\|^2 dvdw\right)^{1/2} < \infty.$$

We get:

$$\begin{aligned}
& \int\int h(v)'\alpha_1(v,w;A)\,h(w)dvdw \\
= & \int\int\left(\alpha(v;A)^{1/2}h(v)\right)'\left[\alpha(v;A)^{-1/2}\alpha_1(v,w;A)\,\alpha(w;A)^{-1/2}\right]\alpha(v;A)^{1/2}h(w)dvdw \\
\leq & \int\int\left\|\alpha(v;A)^{1/2}h(v)\right\|\left\|\alpha(v;A)^{-1/2}\alpha_1(v,w;A)\,\alpha(w;A)^{-1/2}\right\|\left\|\alpha(w;A)^{1/2}h(w)\right\|dvdw \\
\leq & \left(\int\int\left\|\alpha(v;A)^{-1/2}\alpha_1(v,w;A)\,\alpha(w;A)^{-1/2}\right\|^2 dvdw\right)^{1/2} \\
& \left(\int\left\|\alpha(v;A)^{1/2}h(v)\right\|^2 dv\right), \text{ by applying twice Cauchy-Schwarz inequality,} \\
= & k_A\int h(v)'\alpha(v;A)\,h(v)dv \\
\leq & k_A C_A\int h(v)'h(v)\nu(dv) \\
= & k_A C_A\|h\|^2_{L^2(\nu)}.
\end{aligned}$$

Thus:

$$\|\langle D\log f(X,Y;A), h\rangle\|^2_{L^2(P_0)} \leq C_A(1+k_A)\|h\|^2_{L^2(\nu)},$$

and Proposition 1 is proved.

iii) Invertibility

Let us prove Proposition 2. The information operator can be decomposed in two components:

$$
\begin{aligned}
Ih(w) &= \frac{1}{d\nu/d\lambda(w)}\alpha_0(w;A_0)h(w) + \int \frac{1}{d\nu/d\lambda(w)}\alpha_1(w,v;A_0)h(v)dv \\
&\equiv I^0h(w) + I^1h(w).
\end{aligned}
$$

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 $I^0 : H \to H$ be a continuously invertible operator, and let $I^1 : H \to H$ 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:

$$
\begin{aligned}
\left\| I^{0^{-1}} h \right\|^2_{L^2(\nu)} &= \int \left(\frac{d\nu}{d\lambda}(v) \right)^2 h(v)' \alpha_0(v)^{-2} h(v)\nu(dv) \\
&\le \widetilde{C}_A^{-2} \int h(v)' h(v)\nu(dv) = \widetilde{C}_A^{-2} \|h\|^2_{L^2(\nu)};
\end{aligned}
$$

thus I^0 is continuously invertible. Let us now consider the operator I^1. We have:

$$
I^1h(w) = \int K(w,v;A_0)\, h(v)\nu(dv),
$$

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:

$$
\begin{aligned}
& \int\int \|K(w,v;A_0)\|^2 \nu(dw)\nu(dv) \\
\le & \int\int \frac{\|\alpha(x;A)\| \left\| \alpha(x;A)^{-1/2} \alpha_1(x,y;A)\alpha(y;A)^{-1/2} \right\|^2 \|\alpha(x;A)\|}{(d\nu/d\lambda(x)\ d\nu/d\lambda(y))^2} \nu(dx)\nu(dy) \\
\le & \ C_A^2 \int\int \frac{\left\| \alpha(x;A)^{-1/2} \alpha_1(x,y;A)\alpha(y;A)^{-1/2} \right\|^2}{d\nu/d\lambda(x)\ d\nu/d\lambda(y)} \nu(dx)\nu(dy) \\
= & \ C_A^2 \int\int \left\| \alpha(x;A)^{-1/2} \alpha_1(x,y;A)\alpha(y;A)^{-1/2} \right\|^2 dxdy < \infty.
\end{aligned}
$$

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:

$$\begin{aligned}
\mathcal{K}(A) &= -E_0 \log\left[1 + \frac{f(X,Y;A) - f(X,Y)}{f(X,Y)}\right] \\
&\simeq -E_0\left[\frac{f(X,Y;A) - f(X,Y)}{f(X,Y)}\right] + \frac{1}{2}E_0\left[\left(\frac{f(X,Y;A) - f(X,Y)}{f(X,Y)}\right)^2\right] \\
&= \frac{1}{2}Q\left(A\right).
\end{aligned}$$

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:

$$\begin{aligned}
Q(A_0 + h) &= (h, Ih)_{L^2(\nu)} + \int\int \frac{R(x,y;A_0,h)^2}{f(x,y)} dxdy \\
&\quad + O\left[\left(\int\int \frac{R(x,y;A_0,h)^2}{f(x,y)} dxdy\right)^{\frac{1}{2}} (h, Ih)^{\frac{1}{2}}_{L^2(\nu)}\right],
\end{aligned}$$

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:*

$$\int\int \frac{R(x,y;A_0,h)^2}{f(x,y)} dxdy = O\left(\|h\|^4_{L^2(\nu)}\right),\ A_0 + h \in \mathcal{N}_0.$$

We get:

$$Q(A_0 + h) = (h, Ih)_{L^2(\nu)} + r(h), \ \ A_0 + h \in \mathcal{N}_0,$$

where $r(h) = O\left(\|h\|^2_{L^2(\nu)} (h, Ih)^{1/2}_{L^2(\nu)}\right) = O\left(\|h\|^3_{L^2(\nu)}\right)$. 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:

$$\frac{|r(h)|}{(h, Ih)_{L^2(\nu)}} \leq \frac{C\,\|h\|^2_{L^2(\nu)}}{(h, Ih)^{1/2}_{L^2(\nu)}},$$

for some constant C.

Assumption A.3*:

$$\inf_{\substack{h\in(\Theta-A_0)\\ h\neq 0}} \frac{1}{\|h\|^2_{L^2(\nu)}} \frac{(h, Ih)_{L^2(\nu)}}{\|h\|^2_{L^2(\nu)}} > 4C^2.$$

Thus: $|r(h)| \leq \frac{1}{2}(h, Ih)_{L^2(\nu)}, \;\; h \in \Theta - A_0$, and we get:

$$Q(A_0 + h) = (h, Ih)_{L^2(\nu)} + r(h) \geq \frac{1}{2}(h, Ih)_{L^2(\nu)}, \;\; h \in \Theta - A_0.$$

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:

$$\begin{aligned}
\inf_{A \in \Theta \setminus B_\varepsilon(A_0)} Q(A) &= \inf_{h \in (\Theta - A_0) \setminus B_\varepsilon(0)} Q(A_0 + h) \\
&\geq \frac{1}{2} \inf_{h \in (\Theta - A_0) \setminus B_\varepsilon(0)} \|h\|^2_{L^2(\nu)} \frac{(h, Ih)_{L^2(\nu)}}{\|h\|^2_{L^2(\nu)}} \\
&\geq \frac{1}{2} \varepsilon^2 \inf_{h \neq 0} \frac{(h, Ih)_{L^2(\nu)}}{\|h\|^2_{L^2(\nu)}} > 0.
\end{aligned}$$

v) Equivalence of Assumption A.3 i. and the conditions on the information operator.

ii) $\Longrightarrow$ *i)*: Let $h \in L^2(\nu)$ such that $Ih = 0$. It follows $(h, Ih)_{L^2(\nu)} = 0$ and thus $h = 0$.
i) $\Longrightarrow$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 $h = 0$.
A.3 i.$\Longrightarrow$ *ii)*: Let $h \in L^2(\nu)$ such that $(h, Ih)_{L^2(\nu)} = 0$. Then $E_0 \left[\langle D \log f(X, Y; A_0), h \rangle^2 \right] = 0$. 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:

$$\begin{aligned} F_X(x; A+h) &= \int_{-\infty}^{x} f_X(z; A+h) dz \\ &\simeq \int_{-\infty}^{x} \left[f_X(z; A) + \langle Df_X(z; A), h \rangle \right] dz \\ &= F_X(x; A) + \int_{-\infty}^{x} f_X(z; A) \langle D \log f_X(z; A), h \rangle \, dz, \end{aligned}$$

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:

$$\begin{aligned}
\log f(x|y;a+h,\log g) &= \log g\left[x+a(y)+h(y)\right]-\log\left(\int_{a(y)+h(y)}^{+\infty}g(z)dz\right)\\
&\simeq \log f(x\mid y;a,\log g)+\frac{d\log g}{dz}\left[x+a(y)\right]h(y)\\
&\quad+\frac{g\left[a(y)\right]}{\int_{a(y)}^{+\infty}g(z)dz}h(y).
\end{aligned}$$

Thus we get:

$$\langle D_a \log f(x\mid y;A)\rangle=\left(\frac{d\log g}{dz}\left[x+a(y)\right]+\lambda\left[a(y)\right]\right)h(y).$$

Let us now derive the differential with respect to $\log g$. The first order expansion is given by:

$$\begin{aligned}
\log f(x|y;a,\log g+h) &= \log g\left[x+a(y)\right]+h\left[x+a(y)\right]-\log\left(\int_{a(y)}^{+\infty}g(z)e^{h(z)}dz\right)\\
&\simeq \log f(x\mid y;a,\log g)+h\left[x+a(y)\right]-\frac{\int_{a(y)}^{+\infty}g(z)h(z)dz}{\int_{a(y)}^{+\infty}g(z)dz},
\end{aligned}$$

and thus:

$$\langle D_{\log g}\log f(x\mid y;A)\rangle=h\left[x+a(y)\right]-E_A\left[h\left(X_t+a(X_{t-1})\right)\mid X_{t-1}=y\right].$$

iii) Stochastic unit roots

Let us first compute the differential of $f(x\mid y;\pi,g)$ with respect to π and g. By linearity, we have:

$$\begin{aligned}
\langle D_\pi f(x\mid y;\pi,g),h\rangle &= \left[g\left(x-y\right)-g(x)\right]h(y),\\
\langle D_g f(x\mid y;\pi,g),h\rangle &= \pi\left(y\right)h(x-y)+\left[1-\pi\left(y\right)\right]h(x).
\end{aligned}$$

Thus the differential of $\log f(x \mid y; A)$ with respect to $A = (\log \pi, \log g)$ is given by:

$$
\begin{aligned}
\langle D_{\log \pi} \log f(x \mid y; A), h \rangle &= \frac{[g(x-y) - g(x)]\,\pi(y)}{f(x \mid y; A)} h(y) \\
&= \frac{f(x \mid y; A) - g(x)}{f(x \mid y; A)} h(y) \\
&= r(x, y; A)\, h(y), \\
\langle D_{\log g} \log f(x \mid y; A), h \rangle &= \frac{\pi(y)\, g(x-y)}{f(x \mid y; A)} h(x-y) + \frac{[1 - \pi(y)]\, g(x)}{f(x \mid y; A)} h(x) \\
&= p_1(x, y; A) h(x-y) + p_0(x, y; A) h(x).
\end{aligned}
$$

Let us now derive the information operator. We compute separately each block. We get:

$$
\begin{aligned}
\left(\widetilde{h}, I_{\log \pi, \log \pi} h\right)_{L^2(\nu)} &= E_A \left[\left\langle D_{\log \pi} \log f(X \mid Y; A), \widetilde{h} \right\rangle \langle D_{\log \pi} \log f(X \mid Y; A), h \rangle \right] \\
&= E_A \left[r(X, Y; A)^2\, \widetilde{h}(Y) h(Y) \right] \\
&= E_A \left[E_A \left[r(X, Y; A)^2 \mid Y \right] \widetilde{h}(Y) h(Y) \right] \\
&= \int E_A \left[r(X, Y; A)^2 \mid Y = z \right] f_Y(z; A) \widetilde{h}(z) h(z) dz,
\end{aligned}
$$

$$
\begin{aligned}
\left(\widetilde{h}, I_{\log g, \log g} h\right)_{L^2(\nu)} &= E_A \left[\left\langle D_{\log g} \log f(X \mid Y; A), \widetilde{h} \right\rangle \langle D_{\log g} \log f(X \mid Y; A), h \rangle \right] \\
&= E_A \left[E_A \left[p_1(X, Y; A)^2 \mid X - Y \right] \widetilde{h}(X - Y) h_2(X - Y) \right] \\
&\quad + E_A \left[p_1(X, Y; A) p_0(X, Y; A) \widetilde{h}(X - Y) h(X) \right] \\
&\quad + E_A \left[p_1(X, Y; A) p_0(X, Y; A) \widetilde{h}(X) h(X - Y) \right] \\
&\quad + E_A \left[E_A \left[p_0(X, Y; A)^2 \mid X \right] \widetilde{h}(X) h(X) \right]
\end{aligned}
$$

$$
\begin{aligned}
= &\int E_A \left[p_1(X, Y; A)^2 \mid X - Y = z \right] f_{X-Y}(z) \widetilde{h}(z) h(z) dz \\
&+ \int \widetilde{h}(x) \left(\int p_0(z, z - x; A) p_1(z, z - x; A) f(z, z - x; A) h(z) dz \right) dx \\
&+ \int \widetilde{h}(x) \left(\int p_0(x, x - z; A) p_1(x, x - z; A) f(x, x - z; A) h(z) dz \right) dx \\
&+ \int E_A \left[p_0(X, Y; A)^2 \mid X = z \right] f_X(z) \widetilde{h}(z) h(z) dz,
\end{aligned}
$$

and finally:

$$\begin{aligned}
\left(\widetilde{h}, I_{\log\pi,\log g} h\right)_{L^2(\nu)} &= E_A\left[\left\langle D_{\log\pi}\log f(X \mid Y;A), \widetilde{h}\right\rangle \left\langle D_{\log g}\log f(X \mid Y;A), h\right\rangle\right] \\
&= E_A\left[r\left(X,Y;A\right)\widetilde{h}(Y)p_1(X,Y;A)h(X-Y)\right] \\
&\quad + E_A\left[r\left(X,Y;A\right)\widetilde{h}(Y)p_0(X,Y;A)h(X)\right] \\
&= \int \widetilde{h}(x)\left(\int \left[f(z+x \mid x;A) - g(z+x)\right]p_1(z+x,x;A)h(z)dz\right)f_Y(x)dx \\
&\quad + \int \widetilde{h}(x)\left(\int \left[f(z \mid x;A) - g(z)\right]p_0(z,x;A)h(z)dz\right)f_Y(x)dx.
\end{aligned}$$

Thus the information operator admits a measure decomposition with:

$$\alpha_0(z;A_0) = \begin{pmatrix} E_0\left[r_t^2 \mid X_{t-1}=z\right]f(z) & 0 \\ 0 & \begin{array}{c} E_0\left[p_{1,t}^2 \mid X_t - X_{t-1} = z\right]f_{X_t-X_{t-1}}(z) \\ +E_0\left[p_{0,t}^2 \mid X_t = z\right]f(z) \end{array} \end{pmatrix}$$

and:

$$\begin{aligned}
\alpha_1(x,z;A_0) &= \begin{pmatrix} 0 & \begin{array}{c} f(x)\left\{\left[f\left(x+z|x\right) - g(x+z)\right]p_1(x+z,x) \right. \\ \left. + \left[f\left(z|x\right) - g(z)\right]p_0(z,x)\right\} \end{array} \\ 0 & p_0(x,x-z)p_1(x,x-z)f\left(x,x-z\right) \end{pmatrix} \\
&\quad + \left(x \quad \longleftrightarrow \quad z\right)'.
\end{aligned}$$

v) Archimedean Copulas

a) Proof of Proposition 7.

The Jacobian of the transformation is:

$$\det \frac{\partial\left(w,z\right)}{\partial\left(u,v\right)} = \frac{\phi'\left[\phi^{-1}\left(u\right)+\phi^{-1}\left(v\right)\right]}{\phi'\left[\phi^{-1}\left(u\right)\right]} \equiv J(u,v).$$

Thus:

$$\frac{c\left(u,v\right)}{J(u,v)} = \frac{\phi''\left\{\phi^{-1}\left[C\left(u,v\right)\right]\right\}}{\phi'\left\{\phi^{-1}\left[C\left(u,v\right)\right]\right\}\phi'\left[\phi^{-1}\left(v\right)\right]},$$

and the joint p.d.f. of W and Z is given by:

$$f(w,z) = \frac{\phi''\left[\phi^{-1}\left(w\right)\right]}{\phi'\left[\phi^{-1}\left(w\right)\right]\phi'\left[\phi^{-1}\left(z\right)\right]}\mathbb{I}_{w\le z}.$$

Let us define the function:

$$f^*(w) = -\frac{\phi^{''}\left[\phi^{-1}(w)\right]}{\phi^{'}\left[\phi^{-1}(w)\right]} = -\frac{d}{dw}\phi^{'}\left[\phi^{-1}(w)\right], \quad w \in [0,1].$$

Since $\phi^{'}\left[\phi^{-1}(0)\right] = \phi^{'}[+\infty] = 0$, we have:

$$\phi^{'}\left[\phi^{-1}(z)\right] = -\int_0^z f^*(v)dv = -F^*(z), \text{ say.}$$

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:

$$F^*(w) = -\phi^{'}\left[\phi^{-1}(w)\right],$$

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:

$$\frac{d}{dy}\phi^{-1}(y) = -\frac{1}{\int_0^y f^*(w)dw} \leq 0,$$

$$\frac{d^2}{dy^2}\phi^{-1}(y) = \frac{f^*(y)}{\left(\int_0^y f^*(w)dw\right)^2} \geq 0,$$

and thus ϕ^{-1} is decreasing and convex.

b) Differential of the copula.

The log copula density is given by:

$$\begin{aligned}\log c(u,v;a) &= \log a\left[C(u,v;a)\right] + \log F^*\left[C(u,v;a);a\right] \\ &\quad - \log F^*(u;a) - \log F^*(v;a),\end{aligned}$$

where:

$$a = f^*.$$

The general expression

Let us derive the differential with respect to a. We get:

$$\begin{aligned}
\langle D \log c(u,v;a), h \rangle &= \frac{h\left[C(u,v;a)\right]}{a\left[C(u,v;a)\right]} + \frac{d \log a}{dw}\left[C(u,v;a)\right] \langle DC(u,v;a), h \rangle \\
&\quad + \langle D \log F^*\left[C(u,v;a);a\right], h \rangle + \frac{a\left[C(u,v;a)\right]}{F^*\left[C(u,v;a);a\right]} \langle DC(u,v;a), h \rangle \\
&\quad - \langle D \log F^*(u;a), h \rangle - \langle D \log F^*(v;a), h \rangle \\
&= \frac{h\left[C(u,v;a)\right]}{a\left[C(u,v;a)\right]} + \langle D \log F^*\left[C(u,v;a);a\right], h \rangle \\
&\quad - \langle D \log F^*(u;a), h \rangle - \langle D \log F^*(v;a), h \rangle \\
&\quad + \left(\frac{d \log a}{dw}\left[C(u,v;a)\right] + \frac{a\left[C(u,v;a)\right]}{F^*\left[C(u,v;a);a\right]} \right) \langle DC(u,v;a), h \rangle .
\end{aligned} \tag{a.1}$$

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:

$$\begin{aligned}
\langle DC(u,v;a), h \rangle &= \left\langle D\phi\left[\phi^{-1}(u;a) + \phi^{-1}(v;a);a\right], h \right\rangle \\
&\quad + \phi^{'}\left[\phi^{-1}(u;a) + \phi^{-1}(v;a);a\right]\left\{\left\langle D\phi^{-1}(u;a), h \right\rangle + \left\langle D\phi^{-1}(v;a), h \right\rangle\right\} \\
&= \left\langle D\phi\left(\phi^{-1}\left[C(u,v;a);a\right];a\right), h \right\rangle \\
&\quad + \phi^{'}\left(\phi^{-1}\left[C(u,v;a);a\right];a\right)\left\{\left\langle D\phi^{-1}(u;a), h \right\rangle + \left\langle D\phi^{-1}(v;a), h \right\rangle\right\}.
\end{aligned}$$

By the implicit function theorem we have:

$$\left\langle D\phi\left[\phi^{-1}(y;a);a\right], h \right\rangle = -\phi^{'}\left[\phi^{-1}(y;a);a\right]\left\langle D\phi^{-1}(y;a), h \right\rangle,$$

and thus we get:

$$\begin{aligned}
\langle DC(u,v;a), h \rangle &= F^*\left[C(u,v;a);a\right]\left\{\left\langle D\phi^{-1}\left[C(u,v;a);a\right], h \right\rangle \right. \\
&\quad \left. - \left\langle D\phi^{-1}(u;a), h \right\rangle - \left\langle D\phi^{-1}(v;a), h \right\rangle\right\}.
\end{aligned} \tag{a.2}$$

Differential of $F^*(y;a)$

Let us now derive the differential of $F^*(y;a)$. We get:

$$\begin{aligned}
\langle D \log F^*(y;a), h \rangle &= \frac{1}{F^*(y;a)} \int_0^y h(v)dv \\
&= E_a\left[h(W)/a(W) \mid Z = y\right].
\end{aligned} \tag{a.3}$$

By inserting (a.2) and (a.3) in (a.1) we get:

$$\begin{aligned}
\langle D\log c(u,v;a),h\rangle &= \frac{h\left[C(u,v;a)\right]}{a\left[C(u,v;a)\right]}+E_a\left[h(W)/a(W)\mid Z=C(u,v;a)\right] \\
&\quad -E_a\left[h(W)/a(W)\mid Z=u\right]-E_a\left[h(W)/a(W)\mid Z=v\right] \\
&\quad +\left\{a\left[C(u,v;a)\right]+\frac{d\log a}{dw}\left[C(u,v;a)\right]F^*\left[C(u,v;a);a\right]\right\} \\
&\quad \cdot\left\{\left\langle D\phi^{-1}\left[C(u,v;a);a\right],h\right\rangle-\left\langle D\phi^{-1}(u;a),h\right\rangle-\left\langle D\phi^{-1}(v;a),h\right\rangle\right\}.
\end{aligned} \tag{a.4}$$

Let us finally compute the differential of $\phi^{-1}(y;a)$ with respect to a.

<u>Differential of $\phi^{-1}(y;a)$</u>

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:

$$\begin{aligned}
\left\langle D\phi^{-1}(y;a),h\right\rangle &= -\int_y^1\frac{1}{F^*(w;a)^2}\left(\int_0^w h(v)dv\right)dw \\
&= \left(\int_w^1\frac{dv}{F^*(v;a)^2}\right)\left(\int_0^w h(v)dv\right)|_y^1-\int_y^1\left(\int_w^1\frac{dv}{F^*(v;a)^2}\right)h(w)dw \\
&= -\left(\int_y^1\frac{dv}{F^*(v)^2}\right)\int_0^y h(v)dv-\int_y^1\left(\int_w^1\frac{dv}{F^*(v;a)^2}\right)h(w)dw. \\
&= k(y;a)\int_0^y h(v)dv+\int_y^1 k(w;a)h(w)dw,
\end{aligned} \tag{a.5}$$

where $k(y;a)=-\int_y^1\left(1/F^*(v)^2\right)dv$.

By inserting (a.5) in (a.4), we get the differential of the copula density, which is of the form:

$$\langle D\log c(u,v;a),h\rangle=\frac{h\left[C(u,v;a)\right]}{a\left[C(u,v;a)\right]}+\int_0^1\gamma\left(u,v,w;a\right)h(w)dw,\ \text{ say.}$$

c) The information operator.

Let us now compute the information operator I_c of the copula. We get:

$$\begin{aligned} & E_0\left[\langle D\log c(U,V;a_0),g\rangle\langle D\log c(U,V;a_0),h\rangle\right] \\ = \; & E_0\left\{g\left[C_0(U,V)\right]h\left[C_0(U,V)\right]/a_0\left[C_0(U,V)\right]^2\right\} \\ & +\int E_0\left\{g\left[C_0(U,V)\right]\gamma\left(U,V,y\right)/a_0\left[C_0(U,V)\right]\right\}h(y)dy \\ & +\int E_0\left\{\gamma\left(U,V,y\right)h\left[C_0(U,V)\right]/a_0\left[C_0(U,V)\right]\right\}g(y)dy \\ & +\int\int E_0\left\{\gamma\left(U,V,x\right)\gamma\left(U,V,y\right)\right\}g(x)h(y)dxdy. \end{aligned}$$

Let us consider the four terms separately. The first one is:

$$E_0\left\{g\left[C_0(U,V)\right]h\left[C_0(U,V)\right]/a_0\left[C_0(U,V)\right]^2\right\}=\int g(w)h(w)\frac{f_W(w;a_0)}{a_0(w)^2}dw,$$

where $f_W(.;a_0)$ is the density of W. The second term is:

$$\begin{aligned} & \int E_0\left\{g\left[C_0(U,V)\right]\gamma\left(U,V,y\right)/a_0\left[C_0(U,V)\right]\right\}h(y)dy \\ = \; & \int E_0\left\{g(W)\widetilde{\gamma}\left(W,Z,y\right)/a_0(W)\right\}h(y)dy, \text{ say,} \\ = \; & \int\int g(w)\frac{E_0\left\{\widetilde{\gamma}\left(W,Z,y\right)\mid W=w\right\}f_W\left(w,a_0\right)}{a_0(w)}h(y)dy. \end{aligned}$$

Similarly we get for the third and fourth terms:

$$\begin{aligned} & \int E_0\left\{\gamma\left(U,V,y\right)h\left[C_0(U,V)\right]/a_0\left[C_0(U,V)\right]\right\}g(y)dy \\ = \; & \int\int g(y)\frac{E_0\left\{\widetilde{\gamma}\left(W,Z,y\right)\mid W=w\right\}f_W\left(w,a_0\right)}{a_0(w)}h(w)dw, \end{aligned}$$

and:

$$\begin{aligned} & \int\int E_0\left\{\gamma\left(U,V,x\right)\gamma\left(U,V,y\right)\right\}g(x)h(y)dxdy \\ = \; & \int\int g(x)E_0\left\{\widetilde{\gamma}\left(W,Z,x\right)\widetilde{\gamma}\left(W,Z,y\right)\right\}h(y)dxdy. \end{aligned}$$

Thus the information operator I_c admits representation (5), with local component:

$$\alpha_0(w;a)=\frac{f_W\left(w;a\right)}{a_0(w)^2},$$

and:

$$\begin{aligned}\alpha_1(x,y;a) &= E_a\left\{\widetilde{\gamma}\left(W,Z,y\right) \mid W=x\right\} f_W\left(x,a\right)/a(x) \\ &+E_a\left\{\widetilde{\gamma}\left(W,Z,x\right) \mid W=y\right\} f_W\left(y,a\right)/a(y) \\ &+E_0\left\{\widetilde{\gamma}\left(W,Z,x\right)\widetilde{\gamma}\left(W,Z,y\right)\right\}.\end{aligned}$$

d) The density of the variable W

The c.d.f. of W is given by [see Genest, Rivest (1993)]:

$$\begin{aligned}F_W(w) &= P\left[C(U,V)\leq w\right] = w - \frac{\phi^{-1}(w)}{d\phi^{-1}(w)/dw} \\ &= w-\phi^{-1}(w)\phi^{'}\left[\phi^{-1}(w)\right] \\ &= w+\phi^{-1}(w)F^*(w).\end{aligned}$$

Thus the density of W is given by:

$$\begin{aligned}f_W\left(w\right) &= 1+\frac{1}{\phi^{'}\left[\phi^{-1}(w)\right]}F^*(w)+\phi^{-1}(w)f^*(w) \\ &= \phi^{-1}(w)f^*(w).\end{aligned}$$

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:

$$C(u,v)=\exp\left[D(x,y)\right],$$

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 $A(x,y) = -\log C\left(e^{-x}, e^{-y}\right)$ admits the representation:

$$A(x,y) = \int_{S^1} \max\left\{q_1 x, q_2 y\right\} \sigma\left(dq\right),$$

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:

$$\begin{aligned}
\chi\left(v\right) &= 2\int_0^1 \max\left\{\left(1-z\right)v, z\left(1-v\right)\right\} dF^*(z), \\
\chi(0) &= \chi(1) = 1.
\end{aligned}$$

The boundary conditions on χ are equivalent to:

$$\int_0^1 \left(1-z\right) dF^*(z) = \int_0^1 z dF^*(z) = \frac{1}{2},$$

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:

$$\chi\left(v\right) = 2v\int_0^v \left(1-z\right) f^*(z)dz + 2\left(1-v\right)\int_v^1 z f^*(z)dz.$$

Let us now compute the derivatives of χ. We get:

$$\begin{aligned}
\chi^{'}\left(v\right) &= 2\int_0^v \left(1-z\right) f^*(z)dz - 2\int_v^1 z f^*(z)dz \\
&= 2\int_0^v f^*(z)dz - 1,
\end{aligned}$$

and:

$$\chi^{''}(v) = 2f^{*}(v).$$

Let us introduce functional parameter $a = f^{*}$. Using the restrictions on f^{*}, we deduce the expressions of χ and $\chi^{'}$ in terms of functional parameter a:

$$\begin{aligned}
\chi(v) &= v\int_0^v a(w)dw - \int_0^v wa(w)dw + 1 - v, \\
\chi^{'}(v) &= \int_0^v a(w)dw - 1.
\end{aligned}$$

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: $\beta(k) = O\left(k^{-\delta}\right)$, $\delta > 1$.*

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 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 $|R[F,H]| = O\left(\|H\|_\infty^2\right)$, uniformly on H in the class of compact set. Assume the bandwidth h_T is such that $h_T \to \infty$, $Th_T^2 \to \infty$. Then under Assumptions B.1-B.3:

i. *if $\varphi[F]$ is a cadlag function, and $Th_T^{2m} \to 0$:*

$$\sqrt{T}\left[\Phi\left(\widehat{F}_T\right) - \Phi(F)\right] \xrightarrow{d} N\left[0, V_\Phi(F)\right],$$

where:

$$V_\Phi(F) = \sum_{k=-\infty}^{\infty} cov\left(\varphi[F](Z_t), \varphi[F](Z_{t-k})\right).$$

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 $Th_T^{2m+1} \to 0$:*

$$\sqrt{Th_T}\left[\Phi\left(\widehat{F}_T\right) - \Phi(F)\right] \xrightarrow{d} N\left[0, V_\Phi(F)\right],$$

where:

$$\begin{aligned} V_\Phi(F) &= \left(\int K(z)^2 dz\right)\left(E\left[\gamma_0(Z_t)^2 \mid X_t = x_0\right] f_X(x_0)\right. \\ &\quad \left. + E\left[\gamma_1(Z_t)^2 \mid Y_t = y_0\right] f_Y(y_0)\right). \end{aligned}$$

iii. *If* $\varphi[F]$ *is of the form* $\varphi[F](x,y) = \alpha(x,y)\,\delta_{(x_0,y_0)}(x,y)$, *and* $Th_T^{2m+2} \to 0$*:*

$$\sqrt{Th_T^2}\left[\Phi\left(\widehat{F}_T\right) - \Phi(F)\right] \xrightarrow{d} N\left[0, V_\Phi(F)\right],$$

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* $h_T \to \infty$, $Th_T^2 \to \infty$, $Th_T^{2m} \to 0$.

i) Density estimators.

Let us consider the kernel estimator for the density at (x_0,y_0), $\widehat{f}_T(x_0,y_0)$. 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:

$$\sqrt{Th_T^2}\left(\widehat{f}_T(x_0,y_0) - f(x_0,y_0)\right) \xrightarrow{d} N\left[0, f(x_0,y_0)\left(\int K(z)^2 dz\right)^2\right].$$

ii) Conditional moment estimators.

Let us consider a conditional moment of the type:

$$\begin{aligned} g(x_0,y_0) &= \int \gamma_0(x_0,y)\, f(x_0,y)\, dy + \int \gamma_1(x,y_0)\, f(x,y_0)\, dx \\ &\quad + \int\int \gamma_2(x,y)\, f(x,y) dx, \end{aligned}$$

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:

$$\begin{aligned} g_T(x_0,y_0) &= \int \gamma_0(x_0,y)\, \widehat{f}_T(x_0,y)\, dy + \int \gamma_1(x,y_0)\, \widehat{f}_T(x,y_0)\, dx \\ &\quad + \int\int \gamma_2(x,y)\, \widehat{f}_T(x,y) dx, \end{aligned}$$

is asymptotically normal, with:

$$\sqrt{Th_T}\left[g_T(x_0,y_0) - g(x_0,y_0)\right] \xrightarrow{d} N\left(0, V_\Phi(F)\right)$$

where:

$$\begin{aligned} V_\Phi(F) &= \left(\int K(z)^2 dz\right)\left(E\left[\gamma_0(Z_t)^2 \mid X_t = x_0\right] f_X(x_0)\right. \\ &\quad \left. + E\left[\gamma_1(Z_t)^2 \mid Y_t = y_0\right] f_Y(y_0)\right). \end{aligned}$$

Formula (21) is a special case.

iii) Moment estimators.

Finally let us consider a moment estimator $\int\int g(x,y)\widehat{f}_T(x,y)dxdy$, where g is cadlag. The functional $\Phi(F) = \int\int g(x,y)f(x,y)dxdy$ is Hadamard differentiable, with $\varphi[F](x,y) = g(x,y)$ and $R[F,H] = 0$. Thus, under Assumptions B.1-B.4:

$$\sqrt{T}\left(\int\int g(x,y)\widehat{f}_T(x,y)dxdy - \int\int g(x,y)f(x,y)dxdy\right) \xrightarrow{d} N(0, V_\Phi(F)),$$

where:

$$V_\Phi(F) = \sum_{k=-\infty}^{\infty} cov\left[g(Z_t), g(Z_{t-k})\right].$$

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 $\|.\|_{L^2(\nu)}$.

In the proof of these three points we use the following technical assumptions.

Assumption A.8: *There exist $p > 1$ such that:*

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

Assumption A.9: *For $q > 1$ such that $1/p + 1/q = 1$:*

$$\left\| \frac{\widehat{f}_T(.,.) - f(.,.)}{\widehat{f}_T(.,.)} \right\|_{L^q(\Omega_T)} \xrightarrow{p} 0.$$

Assumption A.10:

$$\int_0^1 \int_0^1 \frac{\left[\widehat{f}_T(x,y) - f(x,y)\right]^2}{f(x,y)} dxdy = O_p(1).$$

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 $A \in \Theta$:

$$\int\int \frac{R(x,y;A,h)^2}{f(x,y)} dxdy = O\left(\|h\|^4_{L^2(\nu)}\right), \; h \in L^2(\nu).$$

i) Uniform Convergence.

We have:

$$\begin{aligned} Q_T(A) &= \int\int \widehat{f}_T(x,y)\omega_T(x,y)dxdy \\ &\quad -2\int\int f(x,y;A)\omega_T(x,y)dxdy \\ &\quad +\int\int \frac{f(x,y;A)^2}{\widehat{f}_T(x,y)}\omega_T(x,y)dxdy, \end{aligned}$$

and:

$$Q_\infty(A) = Q(A) = \int\int \frac{f(x,y;A)^2}{f(x,y)}dxdy - 1.$$

Thus:

$$\begin{aligned}
Q_T(A) - Q_\infty(A) &= \int\int \widehat{f}_T(x,y)\omega_T(x,y)dxdy - 1 \\
&\quad -2\left(\int\int f(x,y;A)\omega_T(x,y)dxdy - 1\right) \\
&\quad +\int_0^1\int_0^1 f(x,y;A)^2\left(\frac{1}{\widehat{f}_T(x,y)} - \frac{1}{f(x,y)}\right)\omega_T(x,y)dxdy \\
&\quad +\int_0^1\int_0^1 \frac{f(x,y;A)^2}{f(x,y)}\left(\omega_T(x,y) - 1\right)dxdy \\
&\equiv S_{1,T} + S_{2,T} + S_{3,T} + S_{4,T}, \text{ say.}
\end{aligned}$$

Let us now check that each term converges in probability to 0, uniformly in $A \in \Theta$. We have:

$$\begin{aligned}
&|S_{1,T}| = \left|\int\int \widehat{f}_T(x,y)\left(\omega_T(x,y) - 1\right)dxdy\right| \\
\leq\ & \int\int \widehat{f}_T(x,y)\left|\omega_T(x,y) - 1\right|dxdy \\
\leq\ & \int\int \widehat{f}_T(x,y)\mathbb{I}_{\widetilde{\Omega}_T^c}(x,y)dxdy \\
\leq\ & \left\|\frac{\widehat{f}_T}{\sqrt{f}}\right\|_{L^2}\left\|\sqrt{f}\mathbb{I}_{\widetilde{\Omega}_T^c}\right\|_{L^2} \\
\leq\ & \left(\int_0^1\int_0^1 \frac{\widehat{f}_T(x,y)^2}{f\left(x,y\right)}dxdy\right)^{\frac{1}{2}}\left(\int_0^1\int_0^1 \mathbb{I}_{\widetilde{\Omega}_T^c}(x,y)f\left(x,y\right)dxdy\right)^{\frac{1}{2}} \\
\leq\ & \left(\int\int \frac{\left[\widehat{f}_T(x,y) - f(x,y)\right]^2}{f\left(x,y\right)}dxdy + 1\right)^{\frac{1}{2}} P_0\left[(X_t,Y_t) \in \widetilde{\Omega}_T^c\right]^{1/2} \xrightarrow{p} 0\ ,
\end{aligned}$$

due to Assumptions A.6 and A.10.
The proof is similar for $S_{2,T}$:

$$
\begin{aligned}
|S_{2,T}| &\leq \int\int f(x,y;A)\mathbb{I}_{\widetilde{\Omega}_T^c}(x,y)dxdy \\
&\leq \left(\int\int \frac{f(x,y;A)^2}{f(x,y)}dxdy\right)^{\frac{1}{2}} \left(\int\int \mathbb{I}_{\widetilde{\Omega}_T^c}(x,y)f(x,y)\,dxdy\right)^{\frac{1}{2}} \\
&\leq \left(\sup_{A\in\Theta} Q(A)+1\right)^{\frac{1}{2}} P_0\left[(X_t,Y_t)\in\widetilde{\Omega}_T^c\right]^{1/2} \\
&\to 0, \text{ in probability uniformly in } A\in\Theta,
\end{aligned}
$$

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:

$$
Q(A_0+h) = C_1\|h\|^2_{L^2(\nu)} + C_2\|h\|^3_{L^2(\nu)} + C_3\|h\|^4_{L^2(\nu)},
$$

for some constants C_1, C_2, C_3 [see ii) below].
Let us now consider $S_{3,T}$:

$$
\begin{aligned}
|S_{3,T}| &\leq \int_0^1\int_0^1 \frac{f(x,y;A)^2}{f(x,y)}\left|\frac{\widehat{f}_T(x,y)-f(x,y)}{\widehat{f}_T(x,y)}\right|\omega_T(x,y)dxdy \\
&\leq \sup_{A\in\Theta}\left\|\frac{f(.,.;A)^2}{f(.,.)}\right\|_{L^p}\left\|\frac{\widehat{f}_T-f}{\widehat{f}_T}\right\|_{L^q(\Omega_T)} \\
&\to 0, \text{ in probability uniformly in } A\in\Theta,
\end{aligned}
$$

due to Assumptions A.8 and A.9.
Finally, the last term $S_{4,T}$ is such that:

$$
\begin{aligned}
|S_{4,T}| &\leq \int\int \frac{f(x,y;A)^2}{f(x,y)}\left|\omega_T(x,y)-1\right|dxdy \\
&\leq \int\int \frac{f(x;y;A)^2}{f(x,y)}\mathbb{I}_{\widetilde{\Omega}_T^c}(x,y)dxdy \\
&\leq \left\|\frac{f(.,.;A)^2}{f(.,.)}\right\|_{L^p}\left\|\mathbb{I}_{\widetilde{\Omega}_T^c}\right\|_{L^q}, \\
&\leq \sup_{A\in\Theta}\left\|\frac{f(.,.;A)^2}{f(.,.)}\right\|_{L^p}\cdot\lambda_2\left(\widetilde{\Omega}_T^c\right)^{1/q}, \\
&\to 0, \text{ in probability uniformly in } A\in\Theta,
\end{aligned}
$$

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:

$$\lim_{h\to 0} Q(A+h) = Q(A), \quad \forall A \in \Theta,$$

where $h \to 0$ denotes convergence in norm $\|.\|_{L^2(\nu)}$. For this purpose let us consider the expansion of the chi-square criterion:

$$\begin{aligned}
Q(A+h) &= \int\int \frac{[f(x,y)-f(x,y;A+h)]^2}{f(x,y)} dxdy \\
&= \int\int \frac{[f(x,y)-f(x,y;A)-\langle Df(x,y;A),h\rangle - R(x,y;A,h)]^2}{f(x,y)} dxdy \\
&= Q(A) + \int\int \langle D\log f(x,y;A),h\rangle^2 f(x,y)dxdy \\
&\quad + \int\int \frac{R(x,y;A,h)^2}{f(x,y)} dxdy \\
&\quad -2\int\int [f(x,y)-f(x,y;A)] \langle D\log f(x,y;A),h\rangle \, dxdy \\
&\quad +2\int\int \langle D\log f(x,y;A),h\rangle \, R(x,y;A,h)dxdy \\
&\quad -2\int\int \frac{f(x,y)-f(x,y;A)}{f(x,y)} R(x,y;A,h)dxdy.
\end{aligned}$$

Let us now bound the terms in the last three lines. For the first one we have:

$$\begin{aligned}
&\left| \int\int [f(x,y)-f(x,y;A)] \langle D\log f(x,y;A),h\rangle \, dxdy \right| \\
= &\left| E_0\left[\frac{f(X,Y)-f(X,Y;A)}{f(X,Y)} \langle D\log f(X,Y;A),h\rangle \right] \right| \\
\leq &\ E_0\left[\left(\frac{f(X,Y)-f(X,Y;A)}{f(X,Y)} \right)^2 \right]^{1/2} E_0\left[\langle D\log f(X,Y;A),h\rangle^2 \right]^{1/2} \\
= &\ Q(A)^{1/2} \, (h, I_A h)^{1/2}_{L^2(\nu)} .
\end{aligned}$$

Similar upper bounds can be obtained for the last two terms. Thus the expansion of Q is:

$$\begin{aligned}
Q(A+h) &= Q(A) + (h, I_A h)_{L^2(\nu)} + \int\int \frac{R(x,y;A,h)^2}{f(x,y)} dxdy \\
&\quad + O\left[(h, I_A h)^{1/2}_{L^2(\nu)} \, Q(A)^{1/2} \right] \\
&\quad + O\left[\left(\int\int \frac{R(x,y;A,h)^2}{f(x,y)} dxdy \right)^{1/2} (h, I_A h)^{1/2}_{L^2(\nu)} \right] \\
&\quad + O\left[\left(\int\int \frac{R(x,y;A,h)^2}{f(x,y)} dxdy \right)^{1/2} Q(A)^{1/2} \right].
\end{aligned}$$

Under Assumptions A.7 and A.11 we get:

$$Q(A+h) = Q(A) + O\left(\|h\|^2_{L^2(\nu)}\right),$$

and the continuity follows.

iii) Identification.

Under Assumption A.3 i. or ii. we have (see Appendix 2):

$$\sup_{A \in \Theta \backslash B_\varepsilon(A_0)} Q(A) > 0.$$

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:* $\sup\limits_{A \in \Theta} \|A\|_{L^p} < \infty$;

ii. $\sup_{A \in \Theta} \|A(.+u) - A(.)\|_{L^p} \longrightarrow 0$, *as* $u \to 0$;

iii. $lim_{\alpha \to \infty} \sup_{A \in \Theta} \int_{|x|>\alpha} A(x)^p dx = 0$.

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 $g(.,.)/f(.,.;A_0) \in L^2(P_0)$. By Riesz representation theorem, there exists $\psi(g) \in L^2(\nu)$ such that:

$$(\psi(g), h)_{L^2(\nu)} = E_0\left[\frac{g(X,Y)}{f(X,Y)}\langle D\log f(X,Y;A_0), h\rangle\right], \ \forall h \in L^2(\nu).$$

It is given by $\psi(g) = \langle D\log f(.,.;A_0)^*, g/f\rangle$. When the differential admits a measure decomposition (3), function $\psi(g)$ is given by:

$$\begin{aligned}\psi(g)(z) &= \frac{1}{d\nu/d\lambda(z)}\left[\int g(z,y)\gamma_0(z,y)\,dy + \int g(x,z)\gamma_1(x,z)dx\right.\\ &\quad \left. + \int\int g(x,y)\gamma_2(x,y,z)dxdy\right].\end{aligned}$$

Let us now apply these results to function $g_T = \left[\left(\widehat{f}_T - f\right)/f\right]\omega_T = \left(\delta\widehat{f}_T/f\right)\omega_T$. For any $T \in \mathbb{N}$, $\left(\delta\widehat{f}_T/f\right)\omega_T \in L^2(P_0)$ with probability 1. Thus there exists $\psi_T \in L^2(\nu)$ such that:

$$(\psi_T, h)_{L^2(\nu)} = E_0\left[\frac{\delta\widehat{f}_T(X,Y)}{f(X,Y)}\omega_T(X,Y)\langle D\log f(X,Y;A_0), h\rangle\right], \ \forall h \in L^2(\nu).$$

When the differential admits a measure decomposition (3), function ψ_T is given by:

$$\begin{aligned}\psi_T(z) &= \frac{1}{d\nu/d\lambda(z)}\left[\int \delta\widehat{f}_T(z,y)\omega_T(z,y)\gamma_0(z,y)\,dy + \int \delta\widehat{f}_T(x,z)\omega_T(x,z)\gamma_1(x,z)dx\right.\\ &\quad \left. + \int\int \delta\widehat{f}_T(x,y)\omega_T(x,y)\gamma_2(x,y,z)dxdy\right].\end{aligned}$$

Appendix 8
Asymptotic expansion of first order conditions

i) Expansion of the first order condition

From Assumption A.12, $\widehat{A}_T$ satisfies the set of first order conditions:

$$\int\int \frac{\widehat{f}_T(x,y) - f(x,y;\widehat{A}_T)}{\widehat{f}_T(x,y)} \left\langle Df\left(x,y;\widehat{A}_T\right), g\right\rangle \omega_T(x,y)dxdy = 0, \ \forall g \in L^2(\nu).$$

Let us denote $\delta\widehat{A}_T = \widehat{A}_T - A_0$. We can expand the functions involved in the first order condition. We get:

$$\begin{aligned} f\left(x,y;\widehat{A}_T\right) &= f(x,y) + \left\langle Df\left(x,y;A_0\right), \delta\widehat{A}_T\right\rangle + R\left(x,y;\delta\widehat{A}_T\right), \\ \left\langle Df\left(x,y;\widehat{A}_T\right), g\right\rangle &= \left\langle Df\left(x,y;A_0\right), g\right\rangle + \widetilde{R}\left(x,y;\delta\widehat{A}_T, g\right). \end{aligned}$$

The behaviour of the residual terms R and $\widetilde{R}$ has to be constrained to ensure that they are negligible for small h. This is achieved for R by Assumption A.2.bis. For $\widetilde{R}$ we assume:

Assumption A.13: *The residual term $\widetilde{R}$ is such that:*

$$\int\int \frac{\widetilde{R}\left(x,y;h,g\right)^2}{f(x,y)} dxdy = O\left(\|h\|^2_{L^2(\nu)} \|g\|^2_{L^2(\nu)}\right).$$

By writing:

$$\frac{1}{\widehat{f}_T(x,y)} = \frac{1}{f(x,y)}\left(1 - \frac{\delta\widehat{f}_T(x,y)}{\widehat{f}_T(x,y)}\right),$$

where $\delta\widehat{f}_T = \widehat{f}_T - f$, the first order condition becomes:

$$\begin{aligned} 0 &= \int\int \delta\widehat{f}_T(x,y)\omega_T(x,y)\left\langle D\log f\left(x,y;A_0\right), g\right\rangle \left(1 - \frac{\delta\widehat{f}_T(x,y)}{\widehat{f}_T(x,y)}\right) dxdy \\ &- \int\int \frac{\left\langle Df\left(x,y;A_0\right), \delta\widehat{A}_T\right\rangle \left\langle Df\left(x,y;A_0\right), g\right\rangle}{f(x,y)} \left(1 - \frac{\delta\widehat{f}_T(x,y)}{\widehat{f}_T(x,y)}\right) \omega_T(x,y)dxdy \\ &- \int\int R\left(x,y;\delta\widehat{A}_T\right) \frac{\left\langle Df\left(x,y;A_0\right), g\right\rangle}{f(x,y)} \left(1 - \frac{\delta\widehat{f}_T(x,y)}{\widehat{f}_T(x,y)}\right) \omega_T(x,y)dxdy \end{aligned}$$

$$+\int\int \frac{\delta\widehat{f}_T(x,y)}{f(x,y)}\widetilde{R}\left(x,y;\delta\widehat{A}_T,g\right)\left(1-\frac{\delta\widehat{f}_T(x,y)}{\widehat{f}_T(x,y)}\right)\omega_T(x,y)dxdy$$

$$-\int\int \frac{\left\langle Df\left(x,y;A_0\right),\delta\widehat{A}_T\right\rangle}{f(x,y)}\widetilde{R}\left(x,y;\delta\widehat{A}_T,g\right)\left(1-\frac{\delta\widehat{f}_T(x,y)}{\widehat{f}_T(x,y)}\right)\omega_T(x,y)dxdy$$

$$-\int\int \frac{R\left(x,y;\delta\widehat{A}_T\right)}{f(x,y)}\widetilde{R}\left(x,y;\delta\widehat{A}_T,g\right)\left(1-\frac{\delta\widehat{f}_T(x,y)}{\widehat{f}_T(x,y)}\right)\omega_T(x,y)dxdy.$$

The leading terms are the first one [where we recognize $(g,\psi_T)_{L^2(\nu)}$, see Appendix 7] and the second one [with $\left(g,I\delta\widehat{A}_T\right)_{L^2(\nu)}$]. Thus the first order condition can be rewritten as:

$$\left(g,\psi_T-I\delta\widehat{A}_T\right)_{L^2(\nu)}+R\left(\delta\widehat{A}_T,g\right)=0,\ \forall g\in L^2\left(\nu\right),$$

where the residual term $R\left(\delta\widehat{A}_T,g\right)$ is:

$$R\left(\delta\widehat{A}_T,g\right)=-\int\int \delta\widehat{f}_T(x,y)\left\langle D\log f\left(x,y;A_0\right),g\right\rangle \frac{\delta\widehat{f}_T(x,y)}{\widehat{f}_T(x,y)}\omega_T(x,y)dxdy$$

$$-\int\int \left\langle D\log f\left(x,y;A_0\right),\delta\widehat{A}_T\right\rangle\left\langle D\log f\left(x,y;A_0\right),g\right\rangle f(x,y)$$

$$\cdot\left[\left(1-\frac{\delta\widehat{f}_T(x,y)}{\widehat{f}_T(x,y)}\right)\omega_T(x,y)-1\right]dxdy$$

$$-\int\int R\left(x,y;\delta\widehat{A}_T\right)\left\langle D\log f\left(x,y;A_0\right),g\right\rangle\left(1-\frac{\delta\widehat{f}_T(x,y)}{\widehat{f}_T(x,y)}\right)\omega_T(x,y)dxdy$$

$$+\int\int \frac{\delta\widehat{f}_T\left(x,y\right)}{f(x,y)}\widetilde{R}\left(x,y;\delta\widehat{A}_T,g\right)\left(1-\frac{\delta\widehat{f}_T(x,y)}{\widehat{f}_T(x,y)}\right)\omega_T(x,y)dxdy$$

$$-\int\int \frac{\left\langle Df\left(x,y;A_0\right),\delta\widehat{A}_T\right\rangle}{f(x,y)}\widetilde{R}\left(x,y;\delta\widehat{A}_T,g\right)\left(1-\frac{\delta\widehat{f}_T(x,y)}{\widehat{f}_T(x,y)}\right)\omega_T(x,y)dxdy$$

$$-\int\int \frac{R\left(x,y;\delta\widehat{A}_T\right)}{f(x,y)}\widetilde{R}\left(x,y;\delta\widehat{A}_T,g\right)\left(1-\frac{\delta\widehat{f}_T(x,y)}{\widehat{f}_T(x,y)}\right)\omega_T(x,y)dxdy$$

$$\equiv\ R_1\left(\delta\widehat{A}_T,g\right)+R_2\left(\delta\widehat{A}_T,g\right)+R_3\left(\delta\widehat{A}_T,g\right)+R_4\left(\delta\widehat{A}_T,g\right)+R_5\left(\delta\widehat{A}_T,g\right)$$

$$+R_6\left(\delta\widehat{A}_T,g\right).$$

ii) A bound for the residual term

The following Lemma provides a bound for the residual term $R\left(\delta\widehat{A}_T, g\right)$ under the additional assumption:

Assumption A.14: *There exists* $p>1$ *such that:*

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

Lemma A.2: *Under Assumptions A.13 and A.14 the residual term* $R\left(\delta\widehat{A}_T, g\right)$ *is such that:*

$$R\left(\delta\widehat{A}_T, g\right) = \|g\|_{L^2(\nu)}\ O_p\left[\tau_{T,1}^2 + (\tau_{T,1}+\tau_{T,2})\left\|\delta\widehat{A}_T\right\|_{L^2(\nu)} + \left\|\delta\widehat{A}_T\right\|_{L^2(\nu)}^2\right],$$

where

$$\tau_{T,1} = \left\|\frac{\delta\widehat{f}_T}{\widehat{f}_T}\right\|_{L^\infty(\Omega_T)},\quad \tau_{T,2} = \lambda_2\left(\widetilde{\Omega}_T^c\right)^{1/q},\quad 1/p+1/q=1,$$

and p *is defined as in Assumption A.14.*
Proof. *We bound each of the six terms in the expression of* $R\left(\delta\widehat{A}_T, g\right)$.
i) *The first term is such that:*

$$\begin{aligned}
\left|R_1\left(\delta\widehat{A}_T, g\right)\right| &\leq \left\|\frac{\delta\widehat{f}_T}{\widehat{f}_T}\right\|_{L^\infty(\Omega_T)}^2 \int \left|\langle D\log f(x,y;A_0), g\rangle\right|\sqrt{f(x,y)}\frac{\widehat{f}_T(x,y)}{\sqrt{f(x,y)}}dxdy \\
&\leq \left\|\frac{\delta\widehat{f}_T}{\widehat{f}_T}\right\|_{L^\infty(\Omega_T)}^2 E_0\left[\langle D\log f(X,Y;A_0), g\rangle^2\right]^{1/2}\left(\int \frac{\widehat{f}_T(x,y)^2}{f(x,y)}dxdy\right)^{1/2} \\
&= \left\|\frac{\delta\widehat{f}_T}{\widehat{f}_T}\right\|_{L^\infty(\Omega_T)}^2 (g, Ig)_{L^2(\nu)}^{1/2}\left(\int\int \frac{\left[\widehat{f}_T(x,y)-f(x,y)\right]^2}{f(x,y)}dxdy + 1\right)^{1/2} \\
&= O_p\left[\|g\|_{L^2(\nu)}\tau_{1,T}^2\right],
\end{aligned}$$

by continuity of the information operator I *and Assumption A.10.*

ii) *The second term is such that:*

$$\begin{aligned}\left|R_2\left(\delta\widehat{A}_T,g\right)\right| &\leq \int\int\left|\left\langle D\log f\left(x,y;A_0\right),\delta\widehat{A}_T\right\rangle\left\langle D\log f\left(x,y;A_0\right),g\right\rangle\right| \\ &\quad f(x,y)\left|\omega_T(x,y)-1\right|dxdy \\ &\quad+\int\int\left|\left\langle D\log f\left(x,y;A_0\right),\delta\widehat{A}_T\right\rangle\left\langle D\log f\left(x,y;A_0\right),g\right\rangle\right| \\ &\quad f(x,y)\frac{\left|\delta\widehat{f}_T(x,y)\right|}{\widehat{f}_T(x,y)}\omega_T(x,y)dxdy \\ &\leq \left\|\left\langle D\log f(.,.;A_0),g\right\rangle\left\langle D\log f(.,.;A_0),\delta\widehat{A}_T\right\rangle f(.;.)\right\|_{L^p}\left\|\omega_T-1\right\|_{L^q} \\ &\quad+\left\|\frac{\delta\widehat{f}_T}{\widehat{f}_T}\right\|_{L^\infty(\Omega_T)}\left\|\left\langle D\log f(.,.;A_0),g\right\rangle\left\langle D\log f(.,.;A_0),\delta\widehat{A}_T\right\rangle f(.,.)\right\|_{L^1} \\ &= O_p\left[\|g\|_{L^2(\nu)}\left\|\delta\widehat{A}_T\right\|_{L^2(\nu)}\right]\left(\lambda\left(\widetilde{\Omega}_T^c\right)^{1/q}+\left\|\frac{\delta\widehat{f}_T}{\widehat{f}_T}\right\|_{L^\infty(\Omega_T)}\right) \\ &= O_p\left[\|g\|_{L^2(\nu)}\left\|\delta\widehat{A}_T\right\|_{L^2(\nu)}\left(\tau_{1,T}+\tau_{2,T}\right)\right].\end{aligned}$$

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:

$$\begin{aligned}\left|R_3\left(\delta\widehat{A}_T,g\right)\right| &\leq (1+\tau_{1,T})\int\int\left|\left\langle D\log f\left(x,y;A_0\right),g\right\rangle\right|\sqrt{f(x,y)}\frac{\left|R\left(x,y;\delta\widehat{A}_T\right)\right|}{\sqrt{f(x,y)}}dxdy \\ &\leq (1+\tau_{1,T})\left(g,Ig\right)_{L^2}^{1/2}\left(\int\int\frac{R\left(x,y;\delta\widehat{A}_T\right)^2}{f(x,y)}dxdy\right)^{1/2} \\ &= O_p\left(\|g\|_{L^2(\nu)}\left\|\delta\widehat{A}_T\right\|_{L^2(\nu)}^2\right),\end{aligned}$$

by Assumption A.2.bis.

iv) *The term R_4 is such that:*

$$\begin{aligned}
\left|R_4\left(\delta\widehat{A}_T, g\right)\right| &\leq (1+\tau_{1,T})\,\tau_{1,T}\int\int \frac{\left|\widetilde{R}\left(x,y;\delta\widehat{A}_T,g\right)\right|}{\sqrt{f(x,y)}}\frac{\widehat{f}_T(x,y)}{\sqrt{f(x,y)}}dxdy \\
&\leq (1+\tau_{1,T})\,\tau_{1,T}\left(\int\int \frac{\widetilde{R}\left(x,y;\delta\widehat{A}_T,g\right)^2}{f(x,y)}dxdy\right)^{1/2}\left(\int \frac{\widehat{f}_T(x,y)^2}{f(x,y)}dxdy\right) \\
&= O_p\left(\tau_{1,T}\,\|g\|_{L^2(\nu)}\left\|\delta\widehat{A}_T\right\|_{L^2(\nu)}\right),
\end{aligned}$$

by Assumptions A.10, A.13.

v) *The fifth term is bounded by:*

$$\begin{aligned}
\left|R_5\left(\delta\widehat{A}_T, g\right)\right| &\leq (1+\tau_{1,T})\left(\delta\widehat{A}_T, I\delta\widehat{A}_T\right)^{1/2}_{L^2(\nu)}\left(\int\int \frac{\widetilde{R}\left(x,y;\delta\widehat{A}_T,g\right)^2}{f(x,y)}dxdy\right)^{1/2} \\
&= O_p\left(\|g\|_{L^2(\nu)}\left\|\delta\widehat{A}_T\right\|^2_{L^2(\nu)}\right).
\end{aligned}$$

vi) *Finally, the last term:*

$$\begin{aligned}
\left|R_6\left(\delta\widehat{A}_T, g\right)\right| &\leq (1+\tau_T)\left(\int\int \frac{R\left(x,y;\delta\widehat{A}_T\right)^2}{f(x,y)}dxdy\right)^{1/2} \\
&\quad \cdot\left(\int\int \frac{\widetilde{R}\left(x,y;\delta\widehat{A}_T,g\right)^2}{f(x,y)}dxdy\right)^{1/2} \\
&= O_p\left(\|g\|_{L^2(\nu)}\left\|\delta\widehat{A}_T\right\|^3_{L^2(\nu)}\right).
\end{aligned}$$

By gathering the dominant terms, the bound for $R\left(\delta\widehat{A}_T, g\right)$ 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:

$$\tau_{T,1} = \left\| \frac{\delta \widehat{f}_T}{\widehat{f}_T} \right\|_{L^\infty(\Omega_T)} = o_p(T^{-1/4}).$$

Lemma A.3: *Under Assumptions A.1-A.15:*

i.

$$\left\| \delta \widehat{A}_T \right\|_{L^2} = O_p \left(1/\sqrt{T} \right).$$

ii.

$$\sqrt{T} \left(g, \delta \widehat{A}_T \right)_{L^2(\nu)} = \sqrt{T} \left(g, I^{-1} \psi_T \right)_{L^2(\nu)} + o_p(1), \quad g \in L^2(\nu).$$

Proof. *From Lemma A.2, Assumptions A.6 and A.15 we get:*

$$R \left(\delta \widehat{A}_T, g \right) = o_p \left(\|g\|_{L^2} / \sqrt{T} \right) + o_p \left(\|g\|_{L^2} \left\| \delta \widehat{A}_T \right\|_{L^2(\nu)} \right).$$

Then the first order condition is such that:

$$\left(g, I \delta \widehat{A}_T \right)_{L^2} = (g, \psi_T)_{L^2} + o_p \left(\|g\|_{L^2} / \sqrt{T} \right) + o_p \left(\|g\|_{L^2} \left\| \delta \widehat{A}_T \right\|_{L^2(\nu)} \right),$$

for any $g \in L^2(\nu)$, and since I^{-1} is bounded we get:

$$\begin{aligned} \left(g, \delta \widehat{A}_T \right)_{L^2(\nu)} &= \left(g, I^{-1} \psi_T \right)_{L^2(\nu)} + o_p \left(\|g\|_{L^2} / \sqrt{T} \right) \\ &\quad + o_p \left(\|g\|_{L^2(\nu)} \left\| \delta \widehat{A}_T \right\|_{L^2(\nu)} \right), \end{aligned} \tag{a.6}$$

for any $g \in L^2(\nu)$.

Let us now deduce a bound for $\left\| \delta \widehat{A}_T \right\|_{L^2(\nu)}$. Since $\sqrt{T} \left(I^{-1} g, \psi_T \right)_{L^2(\nu)} \xrightarrow{d} N[0, \left(g, I^{-1} g \right)_{L^2(\nu)}]$ (see Lemma 16 in the text) and I^{-1} is bounded:

$$\left(g, I^{-1} \psi_T \right)_{L^2(\nu)} = O_p \left(\|g\|_{L^2} / \sqrt{T} \right).$$

Thus:

$$\left(g, \delta \widehat{A}_T \right)_{L^2(\nu)} = O_p \left(\|g\|_{L^2} / \sqrt{T} \right) + o_p \left(\|g\|_{L^2(\nu)} \left\| \delta \widehat{A}_T \right\|_{L^2(\nu)} \right), \ \ g \in L^2(\nu).$$

We get:

$$\begin{aligned}\left\|\delta\widehat{A}_T\right\|_{L^2(\nu)} &= \sup_{g\in L^2(\nu):\|g\|_{L^2(\nu)}=1}\left(g,\delta\widehat{A}_T\right)_{L^2(\nu)} \\ &= O_p\left(1/\sqrt{T}\right)+o_p\left(\left\|\delta\widehat{A}_T\right\|_{L^2(\nu)}\right),\end{aligned}$$

that is $\left\|\delta\widehat{A}_T\right\|_{L^2(\nu)}=O_p\left(1/\sqrt{T}\right)$. *From* (a.*6) 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 $T\to\infty$, 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 $\widetilde{h}_T$ be a bandwidth converging to 0. For any $x_0\in[0,1]$, define the function:

$$g_{T,x_0}(x)=\frac{1}{\sqrt{\widetilde{h}_T}}\varphi\left(\frac{x-x_0}{\widetilde{h}_T}\right),\quad x\in[0,1].$$

Then:

$$\begin{aligned}\|g_{T,x_0}\|^2_{L^2(\nu)} &= \int\frac{1}{\widetilde{h}_T}\varphi\left(\frac{x-x_0}{\widetilde{h}_T}\right)^2\frac{d\nu}{d\lambda}(x)dx \\ &= \int\varphi(u)^2\frac{d\nu}{d\lambda}(x_0+\widetilde{h}_Tu)du\simeq\left(\int\varphi(u)^2du\right)\frac{d\nu}{d\lambda}(x_0).\end{aligned}$$

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 $T\to\infty$. In addition, for any $h\in L^2(\nu)$:

$$\begin{aligned}(g_{T,x_0},h)_{L^2(\nu)} &= \int\frac{1}{\sqrt{\widetilde{h}_T}}\varphi\left(\frac{x-x_0}{\widetilde{h}_T}\right)h(x)\frac{d\nu}{d\lambda}(x)dx \\ &= \sqrt{\widetilde{h}_T}\int\varphi(u)\,h(x_0+\widetilde{h}_Tu)\frac{d\nu}{d\lambda}(x_0+\widetilde{h}_Tu)du \\ &= \sqrt{\widetilde{h}_T}h(x_0)\frac{d\nu}{d\lambda}(x_0) \\ &\quad+\sqrt{\widetilde{h}_T}\int\varphi(u)\left[\left(h\frac{d\nu}{d\lambda}\right)(x_0+\widetilde{h}_Tu)-\left(h\frac{d\nu}{d\lambda}\right)(x_0)\right]du.\end{aligned}$$

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':

$$\tau_{T,1} = O_p\left(T^{-1/4-\beta_1}\right), \quad \tau_{T,2} = O_p\left(T^{-\beta_2}\right), \quad \beta_1, \beta_2 > 0.$$

Lemma A.4: *Let $g_T \in L^2(\nu)$ for any T, such that $\|g_T\|_{L^2(\nu)} \leq const$, independent of T, for T sufficiently large. Then under Assumption A.15':*

$$\sqrt{T}\left(g_T, I\delta\widehat{A}_T\right)_{L^2(\nu)} = \sqrt{T}\left(g_T, \psi_T\right)_{L^2(\nu)} + O_p\left(T^{-\beta}\right),$$

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 $g_T \in L^2(\nu)$:*

$$\left(g_T, I\delta\widehat{A}_T\right)_{L^2(\nu)} = \left(g_T, \psi_T\right)_{L^2(\nu)} + R\left(\delta\widehat{A}_T, g_T\right).$$

From Lemma A.2, Lemma A.3 i., and using A.15', we get:

$$\begin{aligned} R\left(\delta\widehat{A}_T, g_T\right) &= \|g_T\|_{L^2(\nu)} O_p\left[T^{-1/2-2\beta_1} + \left(T^{-1/4-\beta_1} + T^{-\beta_2}\right)T^{-1/2} + T^{-1}\right] \\ &= O_p\left(T^{-1/2-\beta}\right). \end{aligned}$$

Q.E.D.

Let us apply Lemma A.4 to $g_T = g_{T,x_0}$, where the bandwidth for localization $\widetilde{h}_T$ is selected such that:

$$\widetilde{h}_T = o\left(h_T\right), \;\; h_T = o(\widetilde{h}_T T^{2\beta}).$$

We get:

$$\sqrt{Th_T/\widetilde{h}_T}\left(g_{T,x_0}, I\delta\widehat{A}_T\right)_{L^2(\nu)} = \sqrt{Th_T/\widetilde{h}_T}\left(g_{T,x_0}, \psi_T\right)_{L^2(\nu)} + O_p\left(T^{-\beta}\sqrt{h_T/\widetilde{h}_T}\right). \tag{a.7}$$

Let us consider the RHS of (a.7). We get:

$$\begin{aligned} & \sqrt{Th_T/\widetilde{h}_T}\left(g_{T,x_0}, \psi_T\right)_{L^2(\nu)} + O_p\left(T^{-\beta}\sqrt{h_T/\widetilde{h}_T}\right) \\ \simeq\; & \sqrt{Th_T}\frac{d\nu}{d\lambda}(x_0)\psi_T\left(x_0\right) \\ & +\sqrt{Th_T}\int \varphi\left(u\right)\left[\left(\frac{d\nu}{d\lambda}\psi_T\right)\left(x_0 + \widetilde{h}_T u\right) - \left(\frac{d\nu}{d\lambda}\psi_T\right)\left(x_0\right)\right]du. \end{aligned}$$

Let us now consider the LHS of (a.7). We get:

$$
\begin{aligned}
&\sqrt{Th_T/\widetilde{h}_T}\left(g_{T,x_0}, I\delta\widehat{A}_T\right)_{L^2(\nu)} \\
= \ & \sqrt{Th_T}\frac{d\nu}{d\lambda}(x_0)\, I\delta\widehat{A}_T(x_0) \\
&+\sqrt{Th_T}\int\varphi(u)\left[\left(\frac{d\nu}{d\lambda}I\delta\widehat{A}_T\right)\left(x_0+\widetilde{h}_T u\right)-\left(\frac{d\nu}{d\lambda}I\delta\widehat{A}_T\right)(x_0)\right]du \\
\simeq \ & \sqrt{Th_T}\alpha_0(x_0)\delta\widehat{A}_T(x_0) \\
&+\sqrt{Th_T}\int\varphi(u)\left[\left(\alpha_0\delta\widehat{A}_T\right)\left(x_0+\widetilde{h}_T u\right)-\left(\alpha_0\delta\widehat{A}_T\right)(x_0)\right]du.
\end{aligned}
$$

Thus, from (a.7) we get:

$$
\begin{aligned}
&\sqrt{Th_T}\alpha_0(x_0)\delta\widehat{A}_T(x_0) \\
&+\sqrt{Th_T}\int\varphi(u)\left[\left(\alpha_0\delta\widehat{A}_T\right)\left(x_0+\widetilde{h}_T u\right)-\left(\alpha_0\delta\widehat{A}_T\right)(x_0)\right]du \\
\simeq \ & \sqrt{Th_T}\frac{d\nu}{d\lambda}(x_0)\psi_T(x_0) \\
&+\sqrt{Th_T}\int\varphi(u)\left[\left(\frac{d\nu}{d\lambda}\psi_T\right)\left(x_0+\widetilde{h}_T u\right)-\left(\frac{d\nu}{d\lambda}\psi_T\right)(x_0)\right]du.
\end{aligned}
$$

Let us now show that the second term on the RHS is negligible, since $\widetilde{h}_T = o(h_T)$. We have:

$$
\begin{aligned}
&\sqrt{Th_T}\int\varphi(u)\left[\left(\frac{d\nu}{d\lambda}\psi_T\right)\left(x_0+\widetilde{h}_T u\right)-\left(\frac{d\nu}{d\lambda}\psi_T\right)(x_0)\right]du \\
\simeq \ & \sqrt{Th_T}\frac{\widetilde{h}_T^2}{2}\frac{d^2}{dx^2}\left(\frac{d\nu}{d\lambda}\psi_T\right)(x_0)\int u^2\varphi(u)\,du, \text{ (since the kernel } \varphi \text{ is symmetric)}, \\
= \ & o_p(1). \qquad\qquad \text{(a.8)}
\end{aligned}
$$

Indeed, since $\frac{d\nu}{d\lambda}\psi_T(x_0)$ involves conditional moments of kernel estimators of a density [see (26)], 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:

$$
\begin{aligned}
&\sqrt{Th_T}\alpha_0(x_0)\delta\widehat{A}_T(x_0) \\
\simeq \ & -\sqrt{Th_T}\int\varphi(u)\left[\left(\alpha_0\delta\widehat{A}_T\right)\left(x_0+\widetilde{h}_T u\right)-\left(\alpha_0\delta\widehat{A}_T\right)(x_0)\right]du \\
&+\sqrt{Th_T}\frac{d\nu}{d\lambda}(x_0)\psi_T(x_0), \ \ \lambda\text{-a.s. in } x_0\in[0,1].
\end{aligned}
$$

This is an integral equation for $\sqrt{Th_T}\alpha_0\delta\widehat{A}_T$ which has a unique solution [see e.g. Theorem 5.2.3 in Debnath, Mikusinski (1998)]. By subsitution and using (a.8), we see that the solution is of the form $\sqrt{Th_T}\alpha_0\delta\widehat{A}_T = \sqrt{Th_T}\frac{d\nu}{d\lambda}\psi_T + o_p(1)$. We conclude:

$$\sqrt{Th_T}\alpha_0(x_0)\delta\widehat{A}_T(x_0) \simeq \sqrt{Th_T}\frac{d\nu}{d\lambda}(x_0)\psi_T(x_0), \quad \lambda\text{-a.s. in } x_0 \in [0,1].$$

v) Expansion of the constrained estimator.

Let us now consider the asymptotic expansion of the constrained estimator $\widehat{f}_T^0(x,y)$. We get:

$$\begin{aligned}\widehat{f}_T^0(x,y) - f(x,y) &= f(x,y;\widehat{A}_T) - f(x,y;A_0) \\ &= \left\langle Df(x,y;A_0), \delta\widehat{A}_T\right\rangle + R(x,y;\delta\widehat{A}_T).\end{aligned}$$

Let us now derive a bound for $R(x,y;\delta\widehat{A}_T)$. By Assumption A.2.bis and Lemma A.3.i. we get:

$$\left(\int\int \frac{R(x,y;\delta\widehat{A}_T)^2}{f(x,y)}dxdy\right)^{1/2} = O_p\left(\left\|\delta\widehat{A}_T\right\|^2_{L^2(\nu)}\right) = O_p(1/T). \tag{a.9}$$

For any $x_0, y_0 \in [0,1]$ let us introduce the function:

$$g_{T,x_0,y_0}(x,y) = \frac{1}{\widetilde{h}_T}\varphi\left(\frac{x-x_0}{\widetilde{h}_T}\right)\varphi\left(\frac{y-y_0}{\widetilde{h}_T}\right),$$

where $\varphi \in C_0^\infty$ is a symmetric kernel with compact support and the localization bandwidth $\widetilde{h}_T$ is selected such that $\widetilde{h}_T = o(h_T)$ and $\sqrt{h_T/T} = o\left(\widetilde{h}_T\right)$ [23]. Then:

$$\begin{aligned}\|g_{T,x_0,y_0}\|^2_{L^2(P_0)} &= \int \frac{1}{\widetilde{h}_T^2}\varphi\left(\frac{x-x_0}{\widetilde{h}_T}\right)^2\varphi\left(\frac{y-y_0}{\widetilde{h}_T}\right)^2 f(x,y)dxdy \\ &\simeq \left(\int \varphi(u)^2\,du\right)^2 f(x_0,y_0).\end{aligned}$$

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:

$$\begin{aligned}\int\int g_{T,x_0,y_0}(x,y)R(x,y;\delta\widehat{A}_T)dxdy &\leq \|g_{T,x_0,y_0}\|_{L^2(P_0)}\left(\int\int\frac{R(x,y;\delta\widehat{A}_T)^2}{f(x,y)}dxdy\right)^{1/2} \\ &= O_p(1/T), \text{ [from (a.9)]}.\end{aligned} \tag{a.10}$$

[23] This is possible since $Th_T \to \infty$ by Assumption B.4 in Appendix 5.

On the other hand:

$$\int\int g_{T,x_0,y_0}(x,y)R(x,y;\delta\widehat{A}_T)dxdy = \int\int \frac{1}{\widetilde{h}_T}\varphi\left(\frac{x-x_0}{\widetilde{h}_T}\right)\varphi\left(\frac{y-y_0}{\widetilde{h}_T}\right)R(x,y;\delta\widehat{A}_T)dxdy$$
$$= \widetilde{h}_T\int\int \varphi(u)\,\varphi(v)\,R(x_0+\widetilde{h}_Tu, y_0+\widetilde{h}_Tv;\delta\widehat{A}_T)$$
$$\simeq \widetilde{h}_T R(x_0,y_0;\delta\widehat{A}_T), \tag{a.11}$$

since $\widetilde{h}_T = o(h_T)$. In particular, from (a.10) and (a.11) we get:

$$R(x_0,y_0;\delta\widehat{A}_T) = O_p(1/T\widetilde{h}_T).$$

Since $1/\widetilde{h}_T = o(\sqrt{T/h_T})$ it follows:

$$R(x,y;\delta\widehat{A}_T) = O_p(1/T\widetilde{h}_T) = o_p\left(1/\sqrt{Th_T}\right), \;\; \lambda\text{-a.s. in } x,y.$$

Thus:

$$\widehat{f}_T^0(x,y) - f(x,y) = \left\langle Df(x,y;A_0),\delta\widehat{A}_T\right\rangle + o_p\left(1/\sqrt{Th_T}\right), \;\; \lambda\text{-a.s. in } x,y.$$

Appendix 9
Asymptotic distribution of ψ_T

Let us consider the case where the differential operator admits the measure decomposition (3) [24]. From Appendix 7, function ψ_T is such that:

$$\begin{aligned}
\frac{d\nu}{d\lambda}(w)\,\psi_T(w) &= \int \delta\widehat{f}_T(w,y)\omega_T(w,y)\,\gamma_0(w,y)\,dy + \int \delta\widehat{f}_T(x,w)\omega_T(x,w)\,\gamma_1(x,w)dx \\
&\quad + \int\int \delta\widehat{f}_T(x,y)\omega_T(x,y)\,\gamma_2(x,y,w)dxdy \\
&\simeq \int \delta\widehat{f}_T(w,y)\gamma_0(w,y)\,dy + \int \delta\widehat{f}_T(x,w)\gamma_1(x,w)dx \\
&\quad + \int\int \delta\widehat{f}_T(x,y)\gamma_2(x,y,w)dxdy.
\end{aligned}$$

From Appendix 5, point ii), it follows that:

$$\sqrt{Th_T}\frac{d\nu}{d\lambda}(w)\,\psi_T(w) \xrightarrow{d} N\left(0,\sigma^2(w)\right),$$

where:

$$\begin{aligned}
\sigma^2(w) &= \left(\int K(z)^2dz\right)\left(E\left[\gamma_{0t}^2 \mid X_t = w\right] f_X(w)\right. \\
&\quad \left. + E\left[\gamma_{1t}^2 \mid Y_t = w\right] f_Y(w)\right) \\
&= \left(\int K(z)^2dz\right)\alpha_0(w).
\end{aligned}$$

[24] This is the case 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.

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:

$$0=\int\int\frac{\left[\widehat{f}_T(x|y)-f\left(x|y;\widehat{A}_T\right)\right]}{\widehat{f}_T(x|y)}\left\langle Df\left(x|y;\widehat{A}_T\right),g\right\rangle\omega_T(x,y)\widehat{f}_{Y,T}(y)dxdy,$$

$\forall g\in L^2(\nu)$. Let us expand this condition. We get:

$$\begin{aligned}
0 &\simeq \int\int\frac{\left[\widehat{f}_T(x|y)-f(x|y)\right]}{f(x|y)}\left\langle Df(x|y;A_0),g\right\rangle f_Y(y)dxdy\\
&\quad-\int\int\frac{\left\langle Df(x|y;A_0),\delta\widehat{A}_T\right\rangle}{f(x|y)}\left\langle Df(x|y;A_0),g\right\rangle f_Y(y)dxdy\\
&= \int\int\frac{\left[\widehat{f}_T(x|y)-f(x|y)\right]}{f(x|y)}\left\langle D\log f(x|y;A_0),g\right\rangle f(x,y)dxdy\\
&\quad-\int\int\left\langle D\log f(x|y;A_0),\delta\widehat{A}_T\right\rangle\left\langle D\log f(x|y;A_0),g\right\rangle f(x,y)dxdy\\
&= \left(\widetilde{\psi}_T,g\right)_{L^2(\nu)}-\left(I_{X|Y}\delta\widehat{A}_T,g\right)_{L^2(\nu)}.
\end{aligned}$$

Thus the first order condition is equivalent to:

$$\left(g,I_{X|Y}\delta\widehat{A}_T-\widetilde{\psi}_T\right)_{L^2(\nu)}\simeq 0,\quad\forall g\in L^2(\nu).$$

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:

$$E_0\left[\left(\frac{\partial \log f}{\partial \theta}(X_t,Y_t;A(\theta_0))\right)^2\right] = E_0\left[\left\langle D\log f(X,Y;A_0), \frac{dA}{d\theta}(\theta_0)\right\rangle^2\right]$$
$$= \left(\frac{dA}{d\theta}(\theta_0), I\frac{dA}{d\theta}(\theta_0)\right)_{L^2(\nu)}.$$

Thus the Cramer-Rao bound is given by:

$$B_A(g,\theta) = \left(\frac{dA}{d\theta}(\theta_0), I\frac{dA}{d\theta}(\theta_0)\right)^{-1}_{L^2(\nu)}.$$

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:

$$\left(g, \frac{dA}{d\theta}(\theta_0)\right)_{L^2(\nu)} = 1. \qquad \text{(a.12)}$$

Thus both the Cramer Rao bound and the constraint (a.12) depend on the parameterization only by means of the function $\delta(.) = dA/d\theta(.,\theta_0)$. Therefore problem (30) in the text is equivalent to:

$$\min_{\delta\in L^2(\nu)} (\delta, I\delta)_{L^2(\nu)},$$
$$\text{s.t} \quad : \quad (g,\delta)_{L^2(\nu)} = 1.$$

By Cauchy-Schwarz inequality we have:

$$1 = (g,\delta)^2_{L^2(\nu)} = \left(I^{-1/2}g, I^{1/2}\delta\right)^2_{L^2(\nu)} \leq \left(I^{-1}g, g\right)_{L^2(\nu)} (\delta, I\delta)_{L^2(\nu)}.$$

Therefore $(\delta, I\delta)_{L^2(\nu)} \geq \left(I^{-1}g, g\right)^{-1}_{L^2(\nu)}$ and the bound is reached for $\delta^* = I^{-1}g \in L^2(\nu)$. Thus we deduce:

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

ii) Time-series framework.

In this case the score is given by:

$$\frac{\partial \log f}{\partial \theta}\left(x \mid y; A\left(\theta_0\right)\right) = \left\langle D \log f(x \mid y; A_0), \frac{dA}{d\theta}(\theta_0)\right\rangle.$$

and the Fisher information is:

$$E_0\left[\left(\frac{\partial \log f}{\partial \theta}\left(X_t \mid X_{t-1}; A\left(\theta_0\right)\right)\right)^2\right] = E_0\left[\left\langle D \log f(X_t \mid X_{t-1}; A_0), \frac{dA}{d\theta}(\theta_0)\right\rangle^2\right]$$
$$= \left(\frac{dA}{d\theta}(\theta_0), I_{X|Y}\frac{dA}{d\theta}(\theta_0)\right)_{L^2(\nu)}.$$

Thus the Cramer Rao bound is given by:

$$B_A(g, \theta) = \left(\frac{dA}{d\theta}(\theta_0), I_{X|Y}\frac{dA}{d\theta}(\theta_0)\right)^{-1}_{L^2(\nu)}.$$

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:

$$\left(h, I_H \delta \widehat{A}_T - \psi_T\right)_{L^2(\nu)} = o_p\left(\|h\|_{L^2(\nu)}/\sqrt{T}\right), \quad h \in H.$$

This is equivalent to:

$$\left(g, I_H \delta \widehat{A}_T\right)_{L^2(\nu)} = (g, P_H \psi_T)_{L^2(\nu)} + o_p\left(\|g\|_{L^2(\nu)}/\sqrt{T}\right), \quad g \in L^2(\nu). \quad \text{(a.13)}$$

Let us first consider the asymptotic expansion of linear functionals. Since I_H is continuously invertible we get:

$$\left(g, \delta \widehat{A}_T\right)_{L^2(\nu)} = \left(g, I_H^{-1} P_H \psi_T\right)_{L^2(\nu)} + o_p\left(\|g\|_{L^2(\nu)}/\sqrt{T}\right), \quad g \in L^2(\nu).$$

Thus for any $g \in L^2(\nu)$:

$$\begin{aligned}
\sqrt{T}\left(g, \delta \widehat{A}_T\right)_{L^2(\nu)} &\simeq \sqrt{T}\left(g, I_H^{-1} P_H \psi_T\right)_{L^2(\nu)} \\
&= \sqrt{T}\left(I_H^{-1} P_H g, \psi_T\right)_{L^2(\nu)}, \text{ since } I_H^{-1} \text{ and } P_H \text{ commute,} \\
&\xrightarrow{d} N\left[0, \left(P_H g, I_H^{-1} P_H g\right)_{L^2(\nu)}\right].
\end{aligned}$$

Let us now consider pointwise expansions. Equation (a.13) 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):

$$\left(g_T, I_H \delta \widehat{A}_T\right)_{L^2(\nu)} = (g_T, P_H \psi_T)_{L^2(\nu)} + O_p\left(T^{-\beta}\right), \text{ for any } g_T. \quad \text{(a.14)}$$

Let us apply (a.14) with $g_T = g_{T,x_0}$, $x_0 \in [0,1]$, as defined in Appendix 8. Let us consider g_i, $i = 1, ..., n$, an orthonormal basis of $H^{\perp}$. We have:

$$\begin{aligned}
\sqrt{T}\left(g_{T,x_0}, P_H \psi_T\right)_{L^2(\nu)} &\simeq const\sqrt{T\widetilde{h}_T \frac{d\nu}{d\lambda}}(x_0) P_H \psi_T(x_0) \\
&= const\sqrt{T\widetilde{h}_T \frac{d\nu}{d\lambda}}(x_0)\left[\psi_T(x_0) - \sum_{i=1}^{n} (g_i, \psi_T)_{L^2(\nu)} g_i(x_0)\right] \\
&\simeq const\ \sqrt{T\widetilde{h}_T \frac{d\nu}{d\lambda}}(x_0) \psi_T(x_0),
\end{aligned}$$

where the last equivalence is due to $(g_i, \psi_T)_{L^2(\nu)} = O_p\left(1/\sqrt{T}\right)$. Thus we can neglect in condition (a.14) the effect of the projector P_H on ψ_T and deduce with similar arguments as in Appendix 8 iv):

$$\sqrt{Th_T}\alpha_{0,H}\left(x_0\right)\delta\widehat{A}_T\left(x_0\right) \simeq \sqrt{Th_T}\frac{d\nu}{d\lambda}(x_0)\psi_T\left(x_0\right), \ \lambda\text{-a.s. in } x_0.$$

Therefore:

$$\sqrt{Th_T}\delta\widehat{A}_T\left(x_0\right) \simeq \alpha_{0,H}\left(x_0\right)^{-1}\sqrt{Th_T}\frac{d\nu}{d\lambda}(x_0)\psi_T\left(x_0\right) \xrightarrow{d} N\left[0, \left(\int K(x)^2dx\right)\alpha_{0,H}(x_0)^{-1}\right].$$

ii) The constrained nonparametric efficiency bound.

Let $A\left(.;\theta\right)$ be a one-dimensional parametric model satisfying the constraints. Then we have $dA/d\theta\left(\theta_0\right) \in H$. It follows that the Fisher information is given by:

$$\left(\frac{dA}{d\theta}\left(\theta_0\right), I_H\frac{dA}{d\theta}\left(\theta_0\right)\right)_{L^2(\nu)},$$

and the constraint:

$$\left(g, A\left(\theta\right)\right)_{L^2(\nu)} = \theta, \ \ \theta \simeq \theta_0,$$

is equivalent to:

$$\left(P_H g, \frac{dA}{d\theta}\left(\theta_0\right)\right)_{L^2(\nu)} = 1.$$

Problem (30) becomes:

$$\begin{aligned} &\min_{\delta\in H}\left(\delta, I_H\delta\right)_{L^2(\nu)}, \\ \text{s.t} \quad : \quad &\left(P_H g, \delta\right)_{L^2(\nu)} = 1. \end{aligned}$$

As in Appendix 11 it follows:

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

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 $\widetilde{I}$ be the operator with domain $L^2(\nu)$ defined by:

$$\widetilde{I} = I_H P_H + P_{H^\perp}.$$

Then H and $H^\perp$ are invariant subspaces of $\widetilde{I}$, such that $\widetilde{I}_{|H} = I_H$, and $\widetilde{I}_{|H^\perp} = Id_{|H^\perp}$. Thus, if we show that $\widetilde{I}$ is invertible, invertibility of I_H will follow. We have:

$$\begin{aligned} \widetilde{I} &= \left(I_H^0 + I_H^1\right) P_H + P_{H^\perp} \\ &= I_H^0 - I_H^0 P_{H^\perp} + I_H^1 P_H + P_{H^\perp}. \end{aligned}$$

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 $\widetilde{I}$ is the sum of a continuously invertible operator and a compact operator. In addition, operator $\widetilde{I}$ has a zero null space. Indeed:

$$\begin{aligned} \widetilde{I}h &= 0 \Longrightarrow I_H P_H h + P_{H^\perp} h = 0 \Longrightarrow I_H P_H h = P_{H^\perp} h = 0 \\ &\Longrightarrow P_H h = P_{H^\perp} h = 0, \text{ since } I_H \text{ has zero null space,} \\ &\Longrightarrow h = 0. \end{aligned}$$

By applying Lemma A.1 in Appendix 1, $\widetilde{I}$ is invertible, and the proof is concluded.

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.

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.

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 Hopkings 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.

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.

Gagliardini, P., and C., Gourieroux, (2002): "Duration Time Series Models with Proportional Hazard", 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 Archimediennes et familles de lois bidimensionelles 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.

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

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

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

Gourieroux, C., and C., Tenreiro, (2001): "Local Power Properties of Kernel Based Goodness of Fit Tests", Journal of Multivariate Analysis, 78, 161-190.

Holly, A., (1995): "A Random Linear Functional Approach to Efficiency Bounds", Journal of Econometrics, 65, 235-261.

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.

Quin, J., and J., Lawless, (1994): "Empirical Likelihood and General Estimating Equations", Annals of Statistics, 22, 300-325.

Reed, M., and B., Simon, (1980): *Methods of Modern Mathematical Physics. I: Functional Analysis,* Academic Press.

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.

Scaillet, O., (2001): "Nonparametric Estimation of Copulas for Time Series", Working Paper.

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

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.

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.

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.