In this document we provide some proofs and additional results, which have been omitted in the paper for the sake of saving space. To avoid confusion, the numbering of sections and propositions (equations, respectively) in this document is preceded by letter B (resp. letter b). Any other section, proposition or equation number refers to the paper.

B.1 The functional score and the information operator

In the paper we assume that the Hadamard derivative $D \log f(X, Y; A_0)$ is a bounded operator, and that the information operator I is invertible (Assumptions A.2 and Assumption A.4, respectively). In this section we provide sufficient conditions for these assumptions to be valid, when the information operator satisfies the decomposition in Assumption A.3.

B.1.1 Boundedness of the Hadamard derivative

Under Assumption A.3, the boundedness of the Hadamard derivative $D \log f(X, Y; A_0)$ can be ensured by appropriate regularity conditions on functions α_0, α_1. We have the following proposition.

Proposition B.1: *Let us assume that the information operator satisfies the decomposition in Assumption A.3. For any $A \in \mathcal{A}$, let $\alpha(.; A)$ be a positive definite $q \times q$ matrix function such that:*

$$\int \int \left\| \alpha\left(v; A\right)^{-1/2} \alpha_1(v, w; A) \alpha\left(w; A\right)^{-1/2} \right\|^2 dv dw < \infty, \forall A \in \mathcal{A},$$

where $\|.\|$ is a matrix norm on $R^{q \times q}$. Let $\lambda_{\max}(v; A)$ denote the largest element in the set of eigenvalues of matrix $\alpha_0\left(v; A\right)$ and eigenvalues of matrix $\alpha\left(v; A\right)$. Suppose:

$$\sup_v \lambda_{\max}(v; A) < +\infty, \quad \forall A \in \mathcal{A}. \tag{b.1}$$

Then $D \log f(X, Y; A)$ is a bounded operator from $L^2(\lambda)$ to $L^2(P_A)$, for any $A \in \mathcal{A}$.

Proof: Let $h \in L^2(\lambda)$. We have:

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

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

$$\int h(v)' \alpha_0\left(v; A\right) h(v) dv \leq \int \lambda_{\max}(v; A) h(v)' h(v) dv \leq C_A \|h\|^2_{L^2(\lambda)},$$

where $C_A = \sup_v \lambda_{\max}(v; A)$. Let us now consider the second term, and denote:

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

We get:

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

Thus:

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

and Proposition B.1 is proved. Q.E.D.

B.1.2 Invertibility of the information operator

Assume that the differential has a zero null space:

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

[which is the sufficient condition for invertibility given in Section 3.1 iii) of the paper]. Since $I = D \log f_0^* D \log f_0$, this condition is equivalent to the fact that the information operator I has a zero null space:

$$Ih = 0, h \in L^2(\lambda) \Rightarrow h = 0.$$

The following proposition shows that, under some regularity conditions on functions α_0, α_1 in the decomposition of Assumption A.3, a zero null space of I implies (and thus is equivalent to) the invertibility of the information operator.

Proposition B.2: *Let us assume the conditions in Proposition B.1 and the invertibility of matrix $\alpha_0(v; A)$ for any $v, A \in \mathcal{A}$. Let $\lambda_{\min}(v; A)$ be the smallest eigenvalue of matrix $\alpha_0(v; A)$, and assume that:*

$$\inf_v \lambda_{\min}(v; A) > 0, \quad \forall A \in \mathcal{A}.$$

If the information operator I has a zero null space, then the information operator is continuously invertible.

Proof: The information operator can be decomposed in two components:

$$Ih(w) = \alpha_0(w; A_0)h(w) + \int \alpha_1(w, v; A_0)h(v)dv \equiv I^0 h(w) + I^1 h(w).$$

The invertibility of I is proved by using classical results of operator theory. In particular, let us consider the following Theorem B.3, which is a consequence of the so-called Fredholm alternative (see e.g. Rudin, 1973, Theorem 4.25 and Exercise 15 in Chapter 4; see also Van der Vaart, 1994, Lemma 4, for another application in statistics) and is proved below for completeness.

Theorem B.3: *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 theorem are satisfied by operators I^0 and I^1 defined above. In the proof of Proposition B.1 it has been shown that they are both bounded operators of $L^2(\lambda)$ into itself. Moreover:

$$\left\| I^{0^{-1}} h \right\|^2_{L^2(\lambda)} = \int h(v)' \alpha_0 (v; A_0)^{-2} h(v) dv \leq \widetilde{C}_A^{-2} \|h\|^2_{L^2(\lambda)},$$

where $\widetilde{C}_A = \inf_v \lambda_{\min}(v; A)$; thus I^0 is continuously invertible. Let us now consider the operator I^1:

$$I^1 h(w) = \int \alpha_1(w, v; A_0)h(v)dv.$$

We have:

$$\begin{aligned}
& \int\int \|\alpha_1(v, w; A_0)\|^2 \, dvdw \\
\leq \; & \int\int \|\alpha(v; A)\| \left\| \alpha(v; A)^{-1/2} \alpha_1(v, w; A)\alpha(w; A)^{-1/2} \right\|^2 \|\alpha(w; A)\| \, dvdw \\
\leq \; & c^2 C_A^2 \int\int \left\| \alpha(v; A)^{-1/2} \alpha_1(v, w; A)\alpha(w; A)^{-1/2} \right\|^2 dvdw < \infty,
\end{aligned}$$

where c is a constant from the equivalence of norms. Thus we deduce from Hilbert-Schmidt theorem (see e.g. Example 2 in Section 10.2 of Yosida, 1995) that I^1 is a compact operator. All conditions of Theorem B.3 are satisfied, and Proposition B.2 is proved. Q.E.D.

We conclude this section by proving Theorem B.3.

Proof of Theorem B.3: Write $I = I^0 \left(Id + \left(I^0\right)^{-1} I^1 \right) \equiv I^0 (Id + K)$. Since $\left(I^0\right)^{-1}$ is continuous and I^1 is compact, operator $K = \left(I^0\right)^{-1} I^1$ is compact. From Theorem 4.25 a) in Rudin (1973), operator $Id + K$ is continuously invertible, and the conclusion follows. Q.E.D

B.2 Examples

In this section we prove some results for the differential and the information operators in the examples of constrained nonparametric families discussed in Section 3.2 of the paper.

B.2.1 Archimedean copula

i) Proof of Lemma 1

The Jacobian of the transformation is:

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

Thus:

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

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

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

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 the generator ϕ and the p.d.f. 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 $\phi^{-1}(1) = 0$, we get:

$$\phi^{-1}(y) = \int_y^1 \frac{dv}{\int_0^v f^*(w)dw}, \quad y \in (0,1)\,.$$

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 obvious. 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. Lemma 1 is proved.

ii) The density of the variable W

The c.d.f. of $W = C(U,V)$ is given by (see Genest and 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:

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

B.2.2 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 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):

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

where the generator χ is a function defined on $[0,1]$, is convex, and satisfies:

$$\max(v, 1-v) \leq \chi(v) \leq 1.$$

The extreme value copula is parameterized by one-dimensional functional parameter χ.

i) 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:

$$C(u,v) = \exp\left[D(x,y)\right],$$
$$\frac{\partial C(u,v)}{\partial u} = \frac{C(u,v)}{u}\frac{\partial D(x,y)}{\partial x},$$
$$\frac{\partial^2 C(u,v)}{\partial u \partial v} = \frac{C(u,v)}{uv}\left\{\frac{\partial D(x,y)}{\partial x}\frac{\partial D(x,y)}{\partial y} + \frac{\partial^2 D(x,y)}{\partial x \partial y}\right\}.$$

The partial 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:

$$\begin{aligned}
c_{\chi}(u,v) &= \frac{C_{\chi}(u,v)}{uv}\left\{-\frac{\widetilde{u}\widetilde{v}}{\log u + \log v}\chi^{''}\left(\widetilde{u}\right)\right.\\
&\left. + \left[\chi\left(\widetilde{u}\right) + \widetilde{v}\chi^{'}\left(\widetilde{u}\right)\right]\left[\chi\left(\widetilde{u}\right) - \widetilde{u}\chi^{'}\left(\widetilde{u}\right)\right]\right\}, \qquad \text{(b.2)}
\end{aligned}$$

where $\widetilde{u} = \log u / (\log u + \log v)$, $\widetilde{v} = \log v / (\log u + \log v)$. This copula p.d.f. does not satisfy Assumption A.3 if generator χ is chosen as the functional parameter. To introduce another parameterization satisfying Assumption A.3 we need a specific characterization of generator χ.

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

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

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:

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

The boundary conditions on χ are equivalent to:

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

that is F^* is a c.d.f. such that $\int_0^1 z dF^*(z) = 1/2$.

iii) Expression of the generator and of its derivatives

When F^* admits a density f^*, we get:

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

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

$$\chi^{'}(v) = 2 \int_0^v (1-z) f^*(z)dz - 2 \int_v^1 z f^*(z)dz = 2 \int_0^v f^*(z)dz - 1,$$

and:

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

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

$$\begin{aligned}
\chi(v) &= 2v \int_0^v a(w)dw - 2 \int_0^v w a(w)dw + 1 - v, \\
\chi^{'}(v) &= 2 \int_0^v a(w)dw - 1, \quad \chi^{''}(v) = 2a(v).
\end{aligned}$$

By substitution of these expressions in the copula p.d.f. given in equation (b.2), it is easily verified that the copula family is differentiable with respect to functional parameter a and satisfies Assumptions A.2 and A.3. Functions α_0 and α_1 are available upon request by the authors.

B.3 Kernel estimators

In this section we provide some results on kernel estimators, which are used in the proofs of the asymptotic properties of the minimum chi-square estimator.

B.3.1 Functionals of kernel estimators

Let us first recall the following theorem for functionals of kernel estimators (see Theorem 3 of Aït-Sahalia, 1993).

Theorem B.4: *Let us consider a functional Φ from an open subset of $C^s\left(\mathbb{R}^d\right)$*

to $\mathbb{R}$. *Suppose that* Φ *is Hadamard differentiable at the true c.d.f.* F *with Hadamard derivative* $\langle D\Phi(F), H\rangle = \int \varphi[F](x,y)dH(x,y)$*:*

$$\Phi(F+H) = \Phi(F) + \int \varphi[F](x,y)dH(x,y) + R[F,H],$$

with $R[F,H] = O\left(\|H\|^2_{L^\infty}\right)$*, uniformly on* H *in the class of compact sets. Let* $\widehat{F}_T$ *be the c.d.f. of a d-dimensional kernel estimator, and assume that the bandwidth* h_T *is such that:*

$$h_T \to 0, \qquad Th_T^d \to \infty.$$

Then under Assumptions A.5, A.6, A.8 and A.9:

i) *if* $\varphi[F]$ *is a cadlag function:*

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

where:

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

ii) *If* $d=2$*, and* $\varphi[F]$ *admits the decomposition* $\varphi[F](x,y) = \gamma_0(x,y)\delta_{x_0}(x) + \gamma_1(x,y)\delta_{y_0}(y)$*, where* $\gamma_0, \gamma_1 \in C^0$*, then:*

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

where:

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

(*The same result applies for any dimension d when the Dirac measures of the singular components of* $\varphi[F]$ *have dimension* 1).

iii) *If* $\varphi[F]$ *is of the form* $\varphi[F](x,y) = \alpha(x,y)\delta_{(x_0,y_0)}(x,y)$*, with* α *cadlag*:

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

where:

$$V_\Phi(F) = \left(\int K(u)^2 du\right)^d \alpha(x_0,y_0) f(x_0,y_0).$$

Let us apply this theorem to deduce the asymptotic distribution of some relevant statistics.

i) Density estimators

Let us consider the kernel density estimator 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 A.5, A.6, A.8 and A.9, and if $Th_T^d \to \infty$ we have:

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

ii) Partial moment estimators

Let us assume $d = 2$ and consider a partial moment of the type:

$$\begin{aligned} g(x_0, y_0) &= \int \gamma_0(x_0, y) f(x_0, y)\, dy + \int \gamma_1(x, y_0) f(x, y_0)\, dx \\ &= f_X(x_0) E[\gamma_0(X, Y) \mid X = x_0] + f_Y(y_0) E[\gamma_1(X, Y) \mid Y = y_0], \end{aligned}$$

where $\gamma_0, \gamma_1 \in C^0$, and $x_0, y_0 \in \mathbb{R}$. The functional $\Phi(F) = g(x_0, y_0)$ is Hadamard differentiable, with $\varphi[F](x, y) = \gamma_0(x, y)\, \delta_{x_0}(x) + \gamma_1(x, y)\, \delta_{y_0}(y)$, and $R(F, H) = 0$. Then under Assumptions A.5, A.6, A.8 and A.9, and if $Th_T^d \to \infty$, the partial moment estimator:

$$g_T(x_0, y_0) = \int \gamma_0(x_0, y)\, \widehat{f}_T(x_0, y)\, dy + \int \gamma_1(x, y_0)\, \widehat{f}_T(x, y_0)\, dx,$$

is asymptotically normal, with:

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

where:

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

Formula (10) in the paper is a special case. For a general d, the results extend to partial moments involving integrals of dimension $d - 1$.

iii) Moment estimators

Finally let us consider a moment estimator $\int\int \gamma(x, y)\widehat{f}_T(x, y)dxdy$, where γ is cadlag. The functional $\Phi(F) = \int\int \gamma(x, y)f(x, y)dxdy = E[\gamma(X, Y)]$ is

Hadamard differentiable, with $\varphi[F](x,y)=\gamma(x,y)$ and $R[F,H]=0$. Thus, under Assumptions A.5, A.6, A.8 and A.9, and if $Th_T^d\to\infty$, we get:

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

where:

$$V_\Phi\left(F\right)=\sum_{k=-\infty}^{\infty}cov\left[\gamma\left(X_t,Y_t\right),\gamma\left(X_{t-k},Y_{t-k}\right)\right].$$

B.3.2 Convergence of kernel density estimators

In this section we prove Lemma A.1 of the paper and we provide some additional results on the convergence of kernel density estimator $\widehat{f}_T$. The following theorem proves the uniform convergence and is established in Bosq (1998), Theorem 2.2.

Theorem B.5 : *Under Assumptions A.5, A.6, A.9 and if the bandwidth h_T is such that $\sqrt{\frac{Th_T^d}{(\log T)^3}}\to\infty$, we have:*

$$\sup_{(x,y)\in[0,1]^d}\left|\widehat{f}_T(x,y)-f\left(x,y\right)\right|=O\left(h_T^m\right)+o_p\left(\frac{\log T}{\sqrt{Th_T^d}}\right).$$

We deduce Lemma A.1 of the paper as a corollary.

Proof of Lemma A.1: Let us prove point ii) [the proof of point i) is similar]. We have:

$$\begin{aligned}\left|\frac{\widehat{f}_T(x,y)-f(x,y)}{\widehat{f}_T(x,y)}\right| &= \left|\frac{\widehat{f}_T(x,y)-f(x,y)}{f(x,y)}\frac{1}{1+\frac{\widehat{f}_T(x,y)-f(x,y)}{f(x,y)}}\right|\\ &\leq \left|\frac{\widehat{f}_T(x,y)-f(x,y)}{f(x,y)}\right|\frac{1}{1-\left|\frac{\widehat{f}_T(x,y)-f(x,y)}{f(x,y)}\right|},\end{aligned}$$

whenever the last term is positive. Moreover:

$$\begin{aligned}\sup_{(x,y)\in\Omega_T}\left|\frac{\widehat{f}_T(x,y)-f(x,y)}{f(x,y)}\right| &\leq \frac{\sup_{(x,y)\in\Omega_T}\left|\widehat{f}_T(x,y)-f(x,y)\right|}{\inf_{(x,y)\in\Omega_T}f(x,y)}\\ &= O\left(\left(\log T\right)^\gamma h_T^m\right)+o_p\left(\left(\log T\right)^{1+\gamma}\cdot\frac{1}{\sqrt{Th_T^d}}\right)\\ &\quad \text{(by Theorem B.5 and Assumption A.7)}\\ &= O\left(\left(\log T\right)^\gamma\cdot T^{-\alpha m}\right)+o_p\left(\left(\log T\right)^{1+\gamma}\cdot T^{-(1/2-d\alpha/2)}\right)\\ &= o_p(T^{-\beta_1}),\end{aligned}$$

for any $\beta_1 < \min\{\alpha m, 1/2 - d\alpha/2\}$. By Assumption A.10 we get:
$\min\{\alpha m, 1/2 - d\alpha/2\} > \frac{1}{4}\left(1 + \frac{1}{2}\frac{2m-1}{4m^2+2m+1}\right)$, and the conclusion follows. Q.E.D.

The following theorem proves the convergence of the L^2-norm and is derived in Gouriéroux and Tenreiro (2001).

Theorem B.6: *Under Assumptions A.5, A.6, A.8, and A.9:*

i) *If the bandwidth h_T is such that $Th_T^d \to \infty$ and $\limsup_T T^\delta h_T^d < \infty$, for some $\delta \in (0,1)$, then:*

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

ii) *If moreover $\limsup_T Th_T^{d+2m} < \infty$ then:*

$$\int\int \left[\widehat{f}_T(x,y) - f(x,y)\right]^2 dxdy = O_p\left(h_T^{2m} + \frac{1}{Th_T^d}\right).$$

We deduce the following Corollary.

Corollary B.7: *Under Assumptions A.5-A.10:*

$$\int\int \frac{\left[\widehat{f}_T(x,y) - f(x,y)\right]^2}{f(x,y)} \mathbb{I}_{\Omega_T}(x,y)dxdy = o_p(1).$$

Proof: We have:

$$\begin{aligned}
&\int\int \frac{\left[\widehat{f}_T(x,y) - f(x,y)\right]^2}{f(x,y)} \mathbb{I}_{\Omega_T}(x,y)dxdy \\
\leq\ & \left(\inf_{(x,y)\in\Omega_T} f(x,y)\right)^{-1} \int\int \left[\widehat{f}_T(x,y) - f(x,y)\right]^2 dxdy \\
\leq\ & O_p\left((\log T)^\gamma \cdot h_T^{2m} + \frac{1}{Th_T^d}(\log T)^\gamma\right) = o_p(1),
\end{aligned}$$

by Theorem B.6 ii) and Assumption A.10 on the bandwidth. Q.E.D.

B.4 Asymptotic properties of the minimum chi-square estimator

In this section we prove some results which are used in Appendix A.2 of the paper to derive the asymptotic properties of the minimum chi-square estimator.

B.4.1 Consistency of the minimum chi-square estimator

Proof of Lemma A.2 i): To show the continuity of the limiting criterion $Q_\infty = Q$, we have to prove:

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

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

$$\begin{aligned}
Q(A+h) &= \int\int \frac{[f(x,y) - f(x,y;A+h)]^2}{f(x,y)} dxdy \\
&= \int\int \frac{[f(x,y) - f(x,y;A) - \langle Df(x,y;A), h\rangle - R(x,y;A,h)]^2}{f(x,y)} dxdy \\
&= Q(A) + \int\int \frac{\langle Df(x,y;A), h\rangle^2}{f(x,y)} dxdy + \int\int \frac{R(x,y;A,h)^2}{f(x,y)} dxdy \\
&\quad -2 \int\int [f(x,y) - f(x,y;A)] \frac{\langle Df(x,y;A), h\rangle}{f(x,y)} dxdy \\
&\quad +2 \int\int \frac{\langle Df(x,y;A), h\rangle}{f(x,y)} R(x,y;A,h) dxdy \\
&\quad -2 \int\int \frac{f(x,y) - f(x,y;A)}{f(x,y)} R(x,y;A,h) dxdy, \qquad \text{(b.3)}
\end{aligned}$$

where $R(x,y;A,h)$ is defined in Assumption A.13. Let us now upper bound the terms in the last three lines. For the first one we have:

$$\begin{aligned}
&\left| \int\int \frac{[f(x,y) - f(x,y;A)]}{\sqrt{f(x,y)}} \frac{\langle Df(x,y;A), h\rangle}{\sqrt{f(x,y)}} dxdy \right| \\
\leq &\left(\int\int \frac{[f(x,y) - f(x,y;A)]^2}{f(x,y)} dxdy \right)^{1/2} \left(\int\int \frac{\langle Df(x,y;A), h\rangle^2}{f(x,y)} dxdy \right)^{1/2} \\
= &\; Q(A)^{1/2} \left(\int\int \frac{\langle Df(x,y;A), h\rangle^2}{f(x,y)} dxdy \right)^{1/2}.
\end{aligned}$$

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

$$\begin{aligned} Q(A+h) &= Q(A) + \int\int \frac{\langle Df(x,y;A),h\rangle^2}{f(x,y)}dxdy + \int\int \frac{R(x,y;A,h)^2}{f(x,y)}dxdy \\ &+O\left[Q(A)^{1/2}\left(\int\int \frac{\langle Df(x,y;A),h\rangle^2}{f(x,y)}dxdy\right)^{1/2}\right] \\ &+O\left[\left(\int\int \frac{\langle Df(x,y;A),h\rangle^2}{f(x,y)}dxdy\right)^{1/2}\left(\int\int \frac{R(x,y;A,h)^2}{f(x,y)}dxdy\right)^{1/2}\right] \\ &+O\left[Q(A)^{1/2}\left(\int\int \frac{R(x,y;A,h)^2}{f(x,y)}dxdy\right)^{1/2}\right]. \end{aligned} \tag{b.4}$$

Under Assumption A.13, continuity follows. Q.E.D.

Proof of Lemma A.2 ii): For $A = A_0$, the first, the fourth and the sixth term in equation (b.3) are zero, whereas the second term is equal to $(h, Ih)_{L^2(\lambda)}$. As in equation (b.4) we get:

$$Q(A_0 + h) = (h, Ih)_{L^2(\lambda)} + O\left(\|h\|^3_{L^2(\lambda)} + \|h\|^4_{L^2(\lambda)}\right), \; A_0 + h \in \Theta. \tag{b.5}$$

Since Θ is bounded (Assumption A.15), the result follows. Q.E.D.

Proof of Lemma A.2 iii): From Lemma A.2 ii) we deduce that:

$$\sup_{A\in\Theta} Q(A) = \sup_{h:A_0+h\in\Theta} Q(A_0+h) \leq \sup_{h:A_0+h\in\Theta}\left((h,Ih)_{L^2(\lambda)} + O\left(\|h\|^3_{L^2(\lambda)}\right)\right) < \infty,$$

since the operator I is bounded [Assumption A.13 i)] and Θ is a bounded set (Assumption A.15). Q.E.D.

Finally let us consider the rate of convergence of the minimum chi-square discrepancy as $A \to A_0$. Since I is a bounded operator (Assumption A.2) and Θ is a bounded set (Assumption A.15), from Lemma A.2 ii), we have:

$$Q(A_0 + h) = O\left(\|h\|^2_{L^2(\lambda)}\right),$$

for $A_0 + h \in \Theta$, and thus:

$$\inf_{h\in(\Theta - A_0)\setminus B_\varepsilon(0)} Q(A_0 + h) \leq C\varepsilon^2, \tag{b.6}$$

for some constant C. From inequality (a.6) in Appendix A.2.3 of the paper and inequality (b.6) we deduce the rate of convergence of the chi-square proximity measure as $\varepsilon \to 0$:

$$\inf_{A\in\Theta\setminus B_\varepsilon(A_0)} Q(A) \simeq \varepsilon^2.$$

B.4.2 Asymptotic expansions of the minimum chi-square estimator

i) A bound for the residual term

Proof of Lemma A.3: Let us bound separately the six terms in the expression of $R\left(\delta\widehat{A}_T, g\right)$.
i) The first term is such that:

$$
\begin{aligned}
\left|R_1\left(\delta\widehat{A}_T, g\right)\right| &\leq \sup_{(x,y)\in\Omega_T}\left|\frac{\delta\widehat{f}_T(x,y)}{\widehat{f}_T(x,y)}\right|^2 \int \left|\left\langle D\log f\left(x,y;A_0\right), g\right\rangle\right| \left|\widehat{f}_T(x,y)\right| \mathbb{I}_{\Omega_T}(x,y)dxdy \\
&\leq \tau_{T,1}^2 \left\{ \int \left|\left\langle D\log f\left(x,y;A_0\right), g\right\rangle\right| \sqrt{f(x,y)} \frac{\left|\widehat{f}_T(x,y) - f(x,y)\right|}{\sqrt{f(x,y)}} \mathbb{I}_{\Omega_T}(x,y)dxdy \right. \\
&\quad \left. + \int \left|\left\langle D\log f\left(x,y;A_0\right), g\right\rangle\right| \sqrt{f(x,y)}\sqrt{f(x,y)}dxdy \right\} \\
&\leq \tau_{T,1}^2 (g, Ig)_{L^2(\lambda)}^{1/2} \left\{ \left(\int\int \frac{\left[\widehat{f}_T(x,y) - f(x,y)\right]^2}{f(x,y)} \mathbb{I}_{\Omega_T}(x,y)dxdy \right)^{1/2} + 1 \right\} \\
&\quad \text{(by Cauchy-Schwarz inequality applied to both components)} \\
&= O_p\left[\|g\|_{L^2(\lambda)}\, \tau_{1,T}^2\right],
\end{aligned}
$$

by continuity of the information operator I (Assumption A.2) and Corollary B.7.

ii) The second term is such that:

$$
\begin{aligned}
\left|R_2\left(\delta\widehat{A}_T, g\right)\right| &\leq \int\int \left|\left\langle D\log f\left(x,y;A_0\right), \delta\widehat{A}_T\right\rangle \left\langle D\log f\left(x,y;A_0\right), g\right\rangle\right| \\
&\quad f(x,y)\left|\omega_T(x,y) - 1\right| dxdy \\
&\quad + \int\int \left|\left\langle D\log f\left(x,y;A_0\right), \delta\widehat{A}_T\right\rangle \left\langle D\log f\left(x,y;A_0\right), g\right\rangle\right| \\
&\quad f(x,y)\left|\frac{\delta\widehat{f}_T(x,y)}{\widehat{f}_T(x,y)}\right| \omega_T(x,y)dxdy
\end{aligned}
$$

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

by Assumptions A.2, A.11 and A.20.

iii) The third term satisfies:

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

by Assumptions A.2, A.13 and Lemma A.1 i).

iv) The term R_4 is such that:

$$
\begin{aligned}
\left| R_4 \left(\delta \widehat{A}_T, g \right) \right| &\leq (1 + \tau_{1,T}) \tau_{1,T} \int\int \frac{\left| \widetilde{R}\left(x,y;\delta \widehat{A}_T, g\right) \right|}{f(x,y)} \left| \widehat{f}_T(x,y) \right| \mathbb{I}_{\Omega_T}(x,y) dxdy \\
&\leq (1 + \tau_{1,T}) \tau_{1,T} \left\{ \int\int \frac{\left| \widetilde{R}\left(x,y;\delta \widehat{A}_T, g\right) \right|}{\sqrt{f(x,y)}} \frac{\left| \widehat{f}_T(x,y) - f(x,y) \right|}{\sqrt{f(x,y)}} \mathbb{I}_{\Omega_T}(x,y) dxdy \right. \\
&\quad \left. + \int\int \frac{\left| \widetilde{R}\left(x,y;\delta \widehat{A}_T, g\right) \right|}{\sqrt{f(x,y)}} \sqrt{f(x,y)} dxdy \right\}
\end{aligned}
$$

$$
\begin{aligned}
&\leq (1+\tau_{1,T})\,\tau_{1,T}\left(\int\int \frac{\widetilde{R}\left(x,y;\delta\widehat{A}_T,g\right)^2}{f(x,y)}dxdy\right)^{1/2} \\
&\quad\cdot\left[\left(\int\int \frac{\left[\widehat{f}_T(x,y)-f(x,y)\right]^2}{f(x,y)}\mathbb{I}_{\Omega_T}(x,y)dxdy\right)^{1/2}+1\right] \\
&= O_p\left(\tau_{1,T}\left\|g\right\|_{L^2(\lambda)}\left\|\delta\widehat{A}_T\right\|_{L^2(\lambda)}\right),
\end{aligned}
$$

by Assumption A.19, Lemma A.1 i) and Corollary B.7.

v) The fifth term is bounded by:

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

due to Assumptions A.2, A.19 and Lemma A.1 i).

vi) Finally, the last term:

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

by Assumptions A.13, A.19 and Lemma A.1 i). By gathering the dominant terms, the bound for $R\left(\delta\widehat{A}_T,g\right)$ is proved.

Q.E.D.

ii) The residual term is negligible pointwise

We provide a detailed proof of Lemma A.5.

Proof of Lemma A.5: Since the first order condition holds for any given T:

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

From Lemma A.1 ii), A.3, A.4 i) and Assumptions A.10, A.21 we get:

$$\begin{aligned}
& R\left(\delta\widehat{A}_T, g_T\right) \\
= \ & \|g_T\|_{L^2(\lambda)} O_p\left[T^{-2\beta_1} + \left(T^{-\beta_1} + T^{-\beta_2/q}\right)\left(T^{-1/2} + h_T^m\right) + \left(T^{-1/2} + h_T^m\right)^2\right] \\
= \ & O_p\left(T^{-1/2}\left(T^{-2(\beta_1 - 1/4)} + T^{-\beta_1} + T^{-\beta_2/q}\right.\right. \\
& \left.\left. + \ T^{1/4}h_T^m + T^{1/4}h_T^m + T^{-1/2} + h_T^m + h_T^{2m}T^{1/2}\right)\right) \\
& [\text{since } \ \beta_1, \beta_2/q > 1/4\,], \\
= \ & O_p\left(T^{-1/2}\left(T^{-2(\beta_1-1/4)} + T^{-\beta_1} + T^{-\beta_2/q} \ + T^{-(\alpha m - 1/4)} + T^{-1/2} + T^{-2(\alpha m - 1/4)}\right)\right) \\
= \ & O_p(T^{-\beta^* - 1/2}),
\end{aligned}$$

for $\beta^* = \min\left\{2\left(\beta_1 - \frac{1}{4}\right), \beta_1, \frac{\beta_2}{q}, \alpha m - \frac{1}{4}, \frac{1}{2}, 2\left(\alpha m - \frac{1}{4}\right)\right\} > \frac{1}{4}\frac{2m-1}{4m^2+2m+1}$. Q.E.D.

iii) Expansion of the constrained density estimator

Proof of Proposition 6: Since $(\psi_T - E\psi_T)(w) = O_p\left(1/\sqrt{Th_T}\right)$ [Lemma 7] and $I^{-1}E\psi_T(w) = O\left(h_T^m\right)$ [Section B.4.3 ii) below], we deduce from Corollary 5 ii):

$$\delta\widehat{A}_T(w) = O_p\left(\frac{1}{\sqrt{Th_T}} + h_T^m\right), \quad \lambda\text{-a.s. in } w \in [0,1]\,. \qquad \text{(b.7)}$$

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

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

Under Assumption A.13 iii), we deduce from Lemma A.4 i), equation (b.7) and the bandwidth condition A.10: $R(x,y;\delta\widehat{A}_T) = o_p\left(1/\sqrt{Th_T}\right)$. Proposition 6 follows. Q.E.D.

B.4.3 The efficient score ψ_T

i) Definition

For any $T \in \mathbb{N}$, function $\left(\delta\widehat{f}_T/f\right)\omega_T \in L^2(P_{A_0})$ with probability 1. By Riesz representation theorem there exists $\psi_T \in L^2(\lambda)$ such that:

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

Function ψ_T is given by $\psi_T = \left\langle D \log f_0^*, \left(\delta \widehat{f}_T / f\right) \omega_T \right\rangle$. In particular, when the differential admits decomposition (5) in the paper, function ψ_T is given by:

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

ii) Asymptotic bias

Let us now consider the expected score:

$$
\begin{aligned}
E\psi_T(w) &\simeq \int \left(E\widehat{f}_T - f\right)(w,y)\gamma_0(w,y)\,dy + \int \left(E\widehat{f}_T - f\right)(x,w)\gamma_1(x,w)dx \\
&\quad + \int \int \left(E\widehat{f}_T - f\right)(x,y)\gamma_2(x,y,w)dxdy.
\end{aligned}
$$

By the standard argument we have:

$$
\begin{aligned}
& \int \left(E\widehat{f}_T(w,y) - f(w,y)\right)\gamma_0(w,y)\,dy \\
= & \int \left(\int \int K(u)K(v)\left[f(w - h_T u, y - h_T v) - f(w,y)\right] dudv \right) \gamma_0(w,y)\,dy \\
= & \frac{h_T^m}{m!} \left(\int K(u)u^m du \right) \int \left(\frac{\partial^m f}{\partial x^m}(w,y) + \frac{\partial^m f}{\partial y^m}(w,y) \right) \gamma_0(w,y)\,dy + o\left(h_T^m\right),
\end{aligned}
$$

and similarly for the other terms. We get:

$$
E\psi_T(w) = \frac{h_T^m}{m!} \left(\int K(u)u^m du \right) b(w) + o\left(h_T^m\right),
$$

where function b is given by:

$$
b(w) = \int \Delta^m f(w,y)\gamma_0(w,y)\,dy + \int \Delta^m f(x,w)\gamma_1(x,w)dx + \int \int \Delta^m f(x,y)\gamma_2(x,y,w)\,dxdy,
$$

with:

$$
\Delta^m f(x,y) = \frac{\partial^m f}{\partial x^m}(x,y) + \frac{\partial^m f}{\partial y^m}(x,y).
$$

Using the boundedness of I^{-1} (Assumption A.4), equation (18) in the paper follows.

B.5 Nonparametric efficiency bound

In this section we derive the nonparametric efficiency bound in the time series framework.

Proof of Proposition 14 ii): 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:

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

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

Thus the Cramer-Rao bound is given by:

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

The solution of maximization problem (21) in the paper is similar to that of the i.i.d framework in Appendix A.3, and the nonparametric efficiency bound:

$$B_A\left(g\right) = \left(g, I_{X|Y}g\right)_{L^2(\lambda)},$$

immediately follows. Q.E.D.

References

Aït-Sahalia, Y., 1993, The delta and bootstrap methods for nonparametric kernel functionals, MIT Discussion Paper.

Bosq, D., 1998, Nonparametric statistics for stochastic processes (Springer, New-York).

Genest, C., and L.-P. Rivest, 1993, Statistical inference procedures for bivariate Archimedean copulas, Journal of the American Statistical Association 88, 1034-1043.

Gouriéroux, C., and C. Tenreiro, 2001, Local power properties of kernel based goodness of fit tests, Journal of Multivariate Analysis 78, 161-190.

Joe, H., 1997, Multivariate models and dependence concepts (Chapman and Hall, London).

Rudin, W., 1973, Functional analysis (McGraw-Hill, New York).

Vaart, A. W. van der, 1994, Infinite dimensional M-estimation, Probability Theory and Mathematical Statistics 1, 1-13.

Yosida, K., 1995, Functional analysis (Springer, New York).