

# Kernel Based Nonlinear Canonical Analysis and Time Reversibility

Serge Darolles\*    Jean-Pierre Florens†    Christian Gouriéroux‡

## Abstract

We consider a kernel based approach to nonlinear canonical correlation analysis and its implementation for time series. We deduce a test procedure of the reversibility hypothesis. The method is applied to the analysis of stochastic differential equation from high frequency data on stock returns.

**Keywords:** Nonlinear Canonical Analysis, Kernel Estimators, Reversibility Hypothesis, Diffusion Equations, High Frequency Data.

**JEL Classification:** C14, C22, C40.

---

\*CREST and Société Générale Asset Management.

†GREMAQ and IDEI, Toulouse.

‡Corresponding Author, CEPREMAP and CREST, 15 Boulevard Gabriel Péri, 92245 Malakoff Cedex, France, tel: (33) 1 41 17 35 93, fax: (33) 1 41 17 76 66, e-mail: gouriéro@ensae.fr

# 1 Introduction

Let us consider a multivariate stationary process  $(X_t)$  with a continuous distribution, denote by  $f_h$  the joint density of  $(X_t, X_{t-h})$  and by  $f$  the marginal density of  $X_t$ . Under weak conditions, the joint density can be decomposed as (see e.g. Barrett, Lampard (1955), Dunford, Schwartz (1963), Lancaster (1968)):

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

where the canonical correlations  $\lambda_{i,h}$ ,  $i$  varying, are decreasing:  $\lambda_{1,h} \geq \lambda_{2,h} \geq \dots \geq 0$ ,  $\forall h$ , and the canonical directions satisfy the orthogonality conditions:

$$\begin{aligned} E[\varphi_{i,h}(X_t) \varphi_{k,h}(X_t)] &= 0, \quad \forall k \neq i, \quad \forall h, \\ E[\psi_{i,h}(X_t) \psi_{k,h}(X_t)] &= 0, \quad \forall k \neq i, \quad \forall h, \\ E[\varphi_{i,h}(X_t)] &= E[\psi_{i,h}(X_t)] = 0, \quad \forall i, h, \end{aligned} \quad (2)$$

and the normalization conditions:

$$V[\varphi_{i,h}(X_t)] = V[\psi_{i,h}(X_t)] = 1, \quad \forall i, h.$$

In this note we introduce kernel based nonparametric estimators of the canonical correlations and canonical directions.

The nonlinear canonical decompositions (1) are the basis for analyzing nonlinear dynamics. For instance they are used to estimate indirectly the drift and the volatility functions of an univariate stochastic differential equation from discrete time data (see Hansen, Scheinkman (1995), Kessler, Sorensen (1996), Darolles, Gourioux (2001), Hansen, Scheinkman, Touzi (1998), Chen, Hansen, Scheinkman (1998)). They are also used to define the nonlinear autocorrelograms, i.e. the sequence  $(\lambda_{i,h}, h \text{ varying})$  (see Ding, Granger (1996), Gourioux, Jasiak (2001)), to characterize the dynamic models with finite dimensional

predictor space (see Gouriéroux, Jasiak (2000)), to exhibit the dynamics of extreme risks in finance, and to analyze the nonlinear comovements between series, i.e. the nonlinear copersistence (see e.g. Gouriéroux, Jasiak (1999)).

In section 2 we introduce the unconstrained estimator of the canonical decomposition and describe its asymptotic properties. In section 3 we consider the corresponding estimation constrained by the reversibility property and discuss the test of the reversibility hypothesis. The method is applied in Section 4 to high frequency data on stock returns to illustrate the practical usefulness of the method. Section 5 concludes. The proofs are gathered in appendices.

## 2 Unconstrained estimator

We consider a pair of continuous random vectors  $(X, Y)$ , whose joint p.d.f. admits a nonlinear canonical decomposition:

$$f(x, y) = f(x, \cdot)f(\cdot, y)[1 + \sum_{i=1}^{\infty} \lambda_i \varphi_i(x) \psi_i(y)], \quad (1)$$

where the canonical correlations and canonical directions satisfy the standard orthogonality and normalization conditions. This decomposition exists whenever:

**Assumption A.1:**  $\int \frac{f^2(x, y)}{f(x, \cdot)f(\cdot, y)} dx dy < +\infty$ .

The elements of the canonical decomposition have important interpretations (see e.g. Hotelling (1936), Hamman (1961), Anderson (1963), Dauxois, Pousse (1975), Darolles, Florens, Renault (1998)). Indeed the first pair  $(\varphi_1, \psi_1)$  is a solution of the optimization problem  $\max_{\varphi, \psi} Corr[\varphi(X), \psi(Y)]$ , whereas  $\lambda_1$  is the maximal value of the correlation.  $(\varphi_2, \psi_2)$  is the second pair the most correlated

and  $\lambda_2$  the correlation between  $\varphi_2$  and  $\psi_2$ , and so on. These successive optimization problems are used to derive numerically the elements of the canonical decomposition.

In practice the distribution of the pair  $(X, Y)$  is not known and the theoretical canonical analysis cannot be performed. However we can estimate this canonical decomposition if some observations  $(X_n, Y_n)$ ,  $n = 1, \dots, N$ , of  $(X, Y)$  are available. We assume:

**Assumption A.2:** *The sequence  $(X_n, Y_n)$ ,  $n \geq 1$ , is a stationary process, whose marginal distribution coincides with the distribution of  $(X, Y)$ .*

The results will especially be applied to a stationary time series  $(X_t, t = 0, 1, \dots)$  observed until date  $T$ , with  $X_n = X_t$ ,  $Y_n = X_{t-h}$ . For this reason and without loss of generality we consider vectors  $X$  and  $Y$  with a same dimension:

**Assumption A.3:**  *$X$  and  $Y$  are  $d$ -dimensional.*

## 2.1 Definition of the estimator

When the joint p.d.f. is unknown, we perform the nonlinear canonical analysis on a nonparametric estimator of  $f$ . We consider a kernel estimator<sup>1</sup> of this joint p.d.f. (see Rosenblatt (1956), Silverman (1986)). Let us introduce two kernels  $K_1, K_2$  defined on  $\mathbf{R}^d$ . The unknown density function is estimated by:

$$\hat{f}_N(x, y) = \frac{1}{N} \sum_{n=1}^N \frac{1}{h_{1N}^d h_{2N}^d} K_1 \left( \frac{X_n - x}{h_{1N}} \right) K_2 \left( \frac{Y_n - y}{h_{2N}} \right), \quad (2)$$

where  $h_{1N}, h_{2N}$  are the bandwidths associated with the two components. Then we consider the associated canonical decomposition  $\hat{\lambda}_{i,N}, \hat{\varphi}_{i,N}, \hat{\psi}_{i,N}$ ,  $i \geq 1$ . It

<sup>1</sup>Other nonparametric estimators can be considered as sieve estimators (see Chen-Hansen-Scheinkman (1998), Darolles, Florens, Renault (1998)).

will be computed by solving the sequence of optimization problems corresponding to  $\hat{f}_N$ .

In this approach  $\hat{f}_N$  has to satisfy the properties of a density function, for any  $N$ , to ensure the validity of the empirical canonical analysis. This justifies the next assumption.

**Assumption A.4:** *The kernels  $K_1, K_2$  are non negative, with unit mass.*

## 2.2 Assumptions for consistency and normality of $\hat{f}_N$

The asymptotic properties of the estimated canonical decomposition have to be deduced from the asymptotic properties of the kernel density estimator. We consider the following assumptions to get the uniform consistency of the density kernel estimator and a central limit theorem.

**Assumption A.5<sup>2</sup>:** *The variables  $X$  and  $Y$  take values in the same compact set  $\mathcal{X} \subset \mathbf{R}^d$ ,  $\mathcal{X} = [0, 1]^d$ .*

**Assumption A.6:** *The probability density function  $f$  is continuous on  $\mathcal{X}^2$ .*

**Assumption A.7:** *The strictly stationary sequence  $(X_n, Y_n)$  is geometrically strong mixing, i.e. with  $\alpha$ -mixing<sup>3</sup> coefficients such that  $\alpha_k \leq c_0 \rho^k$ , for some fixed  $c_0 > 0$  and  $0 \leq \rho < 1$ .*

---

<sup>2</sup>The compactness assumption is not restrictive. Indeed, it is always possible to transform the initial data by a one to one transform onto a compact set, since the canonical analysis prior to transformation is easily deduced from the canonical analysis of the transformed data.

<sup>3</sup>The  $\alpha$ -mixing coefficients  $\alpha_k$  are defined as:

$$\alpha_k = \sup_{\substack{B \in \sigma(X_s, s \leq t) \\ C \in \sigma(X_s, s \geq t+k)}} |P(B \cap C) - P(B)P(C)|, k \geq 1.$$

**Assumption A.8:** *The kernels  $K_i$ ,  $i = 1, 2$ , are bounded, symmetric, of order 2, Lipschitzian, and satisfy  $\lim_{\|u\| \rightarrow \infty} \|u\|^d K_i(u) = 0$ ,  $i = 1, 2$ .*

**Assumption A.9:** *As  $N \rightarrow \infty$ ,  $h_{iN} \rightarrow 0$ ,  $\frac{Nh_{iN}^d}{(\log N)^2} \rightarrow +\infty$ ,  $i = 1, 2$ .*

Assumptions A.2-A.9 give the uniform consistency of the density kernel estimator (see Appendix A Lemma A.1). We introduce an additional assumption to obtain the uniform consistency of the associated inner product estimator (see Appendix A Lemma A.2):

$$\langle \varphi(X), \psi(Y) \rangle_N = \int \int \varphi(x) \psi(y) \hat{f}_N(x, y) dx dy.$$

**Assumption A.10:** *The probability density function  $f$  is bounded from below by  $\varepsilon > 0$ .*

Such assumption is now standard in the nonparametric literature (see e.g. Bosq (1998) and the references therein). If the density function were known, it would be possible to transform the data to get a uniform distribution on the compact set  $\mathcal{X} = [0, 1]^d$ . When the probability density function is unknown, a preliminary transformation can still be applied to satisfy assumption A.10, if we have a priori information on the tail behavior of the joint distribution. Without this a priori information and without assumption A.10, the results below can be strongly modified since they implicitly include a tail analysis which is out of the scope of this paper. In the special case of a Markov process  $(X_t)$ , and the choice  $X = X_t$  and  $Y = X_{t-1}$ , there exists another approach to circumvent the tail problem. It consists in censoring the time series from its extreme values while performing an appropriate change of time (Darolles, Gouriéroux (2001)). Indeed there is a simple relation between the canonical decompositions of the initial and transformed processes which can be exploited.

We now introduce the assumptions needed to derive a central limit theorem for the kernel density estimator.

**Assumption A.11:** *The p.d.f.  $f_{t_1, t_2, t_3, t_4}$  of  $\{(X_{t_1}, Y_{t_1}), (X_{t_2}, Y_{t_2}), (X_{t_3}, Y_{t_3}), (X_{t_4}, Y_{t_4})\}$  exists for any  $t_1 < t_2 < t_3 < t_4$ , and  $\sup_{t_1 < t_2 < t_3 < t_4} \|f_{t_1, t_2, t_3, t_4}\|_\infty < \infty$ .*

**Assumption A.12:** *The p.d.f.  $f_{t_1, t_2}$  of  $\{(X_{t_1}, Y_{t_1}), (X_{t_2}, Y_{t_2})\}$  satisfies the condition  $\sup_{t_1 < t_2} \|f_{t_1, t_2} - f \otimes f\|_\infty < \infty$ , where  $f \otimes f$  denotes the product of marginal p.d.f. of  $(X_{t_i}, Y_{t_i})$ ,  $i = 1, 2$ .*

**Assumption A.13:** *The p.d.f.  $f$  is twice continuously differentiable on  $]0, 1[^{2d}$ , and there exists  $b$  such that  $\|f\|_\infty < b$  and  $\|f^{(2)}\|_\infty < b$ .*

**Assumption A.14:** *As  $N \rightarrow \infty$ ,  $h_{iN} \rightarrow 0$ ,  $Nh_{iN}^d \rightarrow \infty$ ,  $Nh_{iN}^{d+4} \rightarrow 0$ ,  $i = 1, 2$ .*

Assumptions A.2-A.8 and A.11-A.14 allow to derive a central limit theorem and then obtain the asymptotic distribution of the kernel density estimator (see Appendix B Lemma B.1).

### 2.3 Consistency of the estimated canonical analysis

We are concerned by the consistency of the  $p$  first estimated canonical correlations and canonical directions, i.e. their convergence to their theoretical counterparts. We first need to introduce some identifiability conditions.

**Assumption A.15:** *The  $p$  first canonical correlations are distinct.*

Assumption A.15 is an identifiability condition for the canonical correlations. Since  $\sum_{i=1}^{\infty} \lambda_i^2 = \int \frac{f^2(x, y)}{f(x, \cdot)f(\cdot, y)} dx dy < +\infty$ , distinct nonzero canonical correlations are necessarily isolated. Another identifiability assumption has to

be introduced for the canonical directions, which are defined up to a change of sign.

**Assumption A.16:** *There exists a value  $x_0$  such that  $\varphi_j(x_0) \neq 0$ ,  $j = 1, \dots, p$ .*

Hence we select the pair of canonical directions with  $\hat{\varphi}_{jN}(x_0) > 0$  and  $\varphi_j(x_0) > 0$ ,  $j = 1, \dots, p$ .

Moreover the functional parameters of interest have to belong to the admissible values of the associated estimators. Note that the initial and approximated optimization problems associated with the nonlinear canonical analysis are not defined on the same spaces of functions. Indeed the approximated optimizations involve the spaces  $L_N^2(X)$ ,  $L_N^2(Y)$  of square integrable functions with respect to  $\hat{f}_N$ . A function  $\varphi$  is square integrable in the approximated optimization problem of order  $N$  if and only if:

$$\int \varphi^2(x) \frac{1}{h_{1N}^d} K_1 \left( \frac{X_n - x}{h_{1N}} \right) dx < +\infty, \quad n = 1, \dots, N.$$

Since the observations can take any value from the support of the marginal distribution of  $f$  and the bandwidth may vary, it is useful to introduce the following space:

$$L_{K_1}^2(X) = \left\{ \varphi : \int \varphi^2(x) \frac{1}{h_1^d} K_1 \left( \frac{\tilde{x} - x}{h_1} \right) dx < +\infty, \forall h_1 > 0, \forall \tilde{x} \in \text{supp } f \right\},$$

the space  $L_{K_2}^2(Y)$  being defined accordingly. The links between the spaces  $L_{K_1}^2(X)$  and  $L^2(X)$  ( $L_{K_2}^2(Y)$  and  $L^2(Y)$  respectively) involve the respective tails of the kernels  $K_1$ ,  $K_2$  and the p.d.f.  $f$ . We intuitively have to select kernels with rather thin tails to be sure that  $L_{K_1}^2(X)$  includes the theoretical canonical directions of interest. It explains the assumption below.

**Assumption A.17:** *The canonical directions  $\varphi_i$  and  $\psi_j$  are such that  $\varphi_i \in L_{K_1}^2(X)$ ,  $i = 1, \dots, p$ , and  $\psi_j \in L_{K_2}^2(Y)$ ,  $j = 1, \dots, p$ .*



Finally the assumption below is useful to study the estimators in appropriate topological space.

**Assumption A.18:** *i) The canonical variates  $\varphi_i, \psi_i, i = 1, \dots, p$ , are continuously differentiable.*

*ii) The kernels  $K_1, K_2$  are continuously differentiable.*

**Theorem 2.1 :** *Under the identification conditions A.15 and A.16, the estimability conditions A.17-A.18 and the technical assumptions A.1-A.10,  $\hat{\lambda}_{i,N}, \hat{\varphi}_{i,N}, \hat{\psi}_{i,N}, i = 1, \dots, p$  converge to their theoretical counterparts in the sense:*

$$\begin{aligned} \hat{\lambda}_{i,N} &\xrightarrow[N \rightarrow \infty]{} \lambda_i \text{ a.s.} \\ \int |\hat{\varphi}_{i,N}(x) - \varphi_i(x)|^2 f(x, \cdot) dx &\xrightarrow[N \rightarrow \infty]{} 0 \text{ a.s.}, \\ \int |\hat{\psi}_{i,N}(y) - \psi_i(y)|^2 f(\cdot, y) dy &\xrightarrow[N \rightarrow \infty]{} 0 \text{ a.s.} \end{aligned}$$

**Proof.** See Appendix A. ■

The a.s. convergence of the integrated mean square error does not imply in general the a.s. pointwise convergence of the estimated canonical variates. It implies (under some additional assumptions) the convergence of inner products  $\langle \hat{\varphi}_{i,N}, g \rangle$ , for any test function  $g \in L^2(X)$  (see Darolles, Florens, Renault (2001)). However, to obtain simple pointwise asymptotic distributions, we only consider  $x$  for which the pointwise convergence is achieved.

## 2.4 Asymptotic distributions

The convergence properties of the estimated canonical correlations and canonical directions allow us to expand the first order conditions and to derive the asymptotic distributions.

**Theorem 2.2** : Under assumptions A.1-A.18,

i) the asymptotic distribution of  $\hat{\lambda}_N = (\hat{\lambda}_{1,N}, \dots, \hat{\lambda}_{p,N})$  is a gaussian distribution:

$$\sqrt{N}(\hat{\lambda}_N - \lambda) \xrightarrow{d} \mathcal{N}(0, V),$$

where the elements of the matrix  $V$  are:  $V_{i,j} = \frac{1}{4} \sum_{k=-\infty}^{\infty} \text{Cov}[2\varphi_i(X_n) \psi_i(Y_n) - \lambda_i \varphi_i^2(X_n) - \lambda_i \psi_i^2(Y_n), 2\varphi_j(X_{n+k}) \psi_j(Y_{n+k}) - \lambda_j \varphi_j^2(X_{n+k}) - \lambda_j \psi_j^2(Y_{n+k})]$ .

ii) The asymptotic distribution of  $\hat{\varphi}_N(x) = (\hat{\varphi}_{1,N}(x), \dots, \hat{\varphi}_{p,N}(x))$  is a gaussian distribution:

$$\sqrt{Nh_{1N}^d}(\hat{\varphi}_N(x) - \varphi(x)) \xrightarrow{d} \mathcal{N}(0, W(x)),$$

where  $W_{i,j}(x) = \frac{1}{\lambda_i \lambda_j} \frac{1}{f(x, \cdot)} \int K_1^2(u) du \text{Cov}[\psi_i(Y), \psi_j(Y) | X = x]$ .

iii) The asymptotic distribution of  $\hat{\psi}_N(y) = (\hat{\psi}_{1,N}(y), \dots, \hat{\psi}_{p,N}(y))$  is a gaussian distribution:

$$\sqrt{Nh_{2N}^d}(\hat{\psi}_N(y) - \psi(y)) \xrightarrow{d} \mathcal{N}(0, U(y)),$$

where  $U_{i,j}(y) = \frac{1}{\lambda_i \lambda_j} \frac{1}{f(\cdot, y)} \int K_2^2(u) du \text{Cov}[\varphi_i(X), \varphi_j(X) | Y = y]$ .

**Proof.** See Appendix B ■

The rate of convergence is a parametric rate for the canonical correlations, whereas it is nonparametric for the canonical directions. The asymptotic variance  $W_{i,i}(x)$  of  $\hat{\varphi}_{i,N}$  coincides with the asymptotic variance of the Nadaraya-Watson estimator of  $\frac{1}{\lambda_i} E[\psi_i(Y) | X]$ . This corresponds to the interpretation of the canonical direction  $\varphi_i$  as the conditional expectation of the canonical direction  $\psi_i$ , up to a scale factor.

### 3 The reversibility hypothesis

In this section we are interested in reversible processes, i.e. processes with identical distributional properties in initial and reversed times. Discretized unidimensional diffusion processes are examples of reversible processes. Therefore, when rejecting the reversibility hypothesis, we also reject the existence of an underlying unidimensional diffusion process (see section 4). Some other procedures to test for unidimensional diffusion processes have been based on the embeddability hypothesis (see Florens, Renault, Touzi (1998)).

The reversibility condition implies that,  $\forall h$ ,  $f_h(x_t, x_{t-h}) = f_h(x_{t-h}, x_t)$ , i.e. the symmetry of the bivariate distribution  $f_h$  at any lag. We explain in the subsection below how to estimate the canonical decomposition of  $f(x, y)$  under these reversibility (i.e. symmetry) conditions. Then, by comparing the unconstrained and constrained estimators, we derive a test of the symmetry hypothesis.

#### 3.1 Constrained estimators

There exists different kernel estimators of the canonical decomposition taking into account the symmetry constraint.

*i)* We select identical kernels  $K_1 = K_2 = K$  and bandwidths  $h_{1N} = h_{2N} = h_N$ , whereas we artificially double the size of the sample by considering  $(X_i, Y_i)$ ,  $(Y_i, X_i)$ ,  $i = 1, \dots, N$ . The constrained kernel estimator of the density function is:

$$\hat{f}_N^R(x, y) = \frac{1}{2Nh_N^{2d}} \sum_{n=1}^N \left\{ K\left(\frac{X_n - x}{h_N}\right) K\left(\frac{Y_n - y}{h_N}\right) + K\left(\frac{X_n - y}{h_N}\right) K\left(\frac{Y_n - x}{h_N}\right) \right\}.$$

The constrained estimators of the canonical correlations and canonical di-

rections are deduced from the canonical decomposition of  $\hat{f}_N^R$  and denoted by  $\hat{\lambda}_{i,N}^R$  and  $\hat{\varphi}_{i,N}^R = \hat{\psi}_{i,N}^R$ ,  $i \geq 0$ .

ii) We look for the solutions of the spectral problem:

$$\frac{1}{2}(\hat{T}^* + \hat{T})\hat{\varphi}_{i,N}^R = \hat{\lambda}_{i,N}^R \hat{\varphi}_{i,N}^R, \langle \hat{\varphi}_{i,N}^R, \hat{\varphi}_{i,N}^R \rangle = 1,$$

where:

$$\begin{aligned} \hat{T}\varphi(y) &= \int \varphi(x) \frac{\hat{f}_N(x, y)}{\hat{f}_N(\cdot, y)} dx, \\ \hat{T}^*\varphi(x) &= \int \varphi(y) \frac{\hat{f}_N(x, y)}{\hat{f}_N(x, \cdot)} dy. \end{aligned}$$

iii) We look for the solutions of the spectral problem:

$$\frac{1}{2}(\hat{T}^*\hat{T} + \hat{T}\hat{T}^*)\hat{\varphi}_{i,N}^R = (\hat{\lambda}_{i,N}^R)^2 \hat{\varphi}_{i,N}^R, \langle \hat{\varphi}_{i,N}^R, \hat{\varphi}_{i,N}^R \rangle = 1.$$

The latter approaches are based on the property of self-adjoint conditional expectation operator under the reversibility hypothesis. These three constrained estimators do not provide the same results in finite sample, but share the same asymptotic properties under the reversibility hypothesis.

**Theorem 3.1** : *Under the assumptions of theorem 2.2, and if the reversibility hypothesis is satisfied,*

i)  $\hat{\lambda}_{i,N}^R, \hat{\varphi}_{i,N}^R$ ,  $i = 1, \dots, p$ , converge a.s. to their theoretical counterparts.

ii) *The asymptotic distribution of  $\hat{\lambda}_N^R = (\hat{\lambda}_{1,N}^R, \dots, \hat{\lambda}_{p,N}^R)$  is a gaussian distribution:*

$$\sqrt{N}(\hat{\lambda}_N^R - \lambda) \xrightarrow{d} \mathcal{N}(0, V^R),$$

where  $V_{i,j}^R = \frac{1}{4} \sum_{k=-\infty}^{\infty} \text{Cov}[2\varphi_i(X_n)\varphi_i(Y_n) - \lambda_i\varphi_i^2(X_n) - \lambda_i\varphi_i^2(Y_n), 2\varphi_j(X_{n+k})\varphi_j(Y_{n+k}) - \lambda_j\varphi_j^2(X_{n+k}) - \lambda_j\varphi_j^2(Y_{n+k})]$ .

iii) The asymptotic distribution of  $\hat{\varphi}_N^R(x) = (\hat{\varphi}_{1,N}^R(x), \dots, \hat{\varphi}_{p,N}^R(x))$  is a gaussian distribution:

$$\sqrt{Nh_N^d}(\hat{\varphi}_N^R(x) - \varphi(x)) \xrightarrow{d} \mathcal{N}(0, W^R(x)),$$

$$\text{where } W_{i,j}^R(x) = \frac{1}{2} \frac{1}{\lambda_i \lambda_j} \frac{1}{f(x, \cdot)} \int K_1^2(u) du \text{Cov}[\varphi_i(Y), \varphi_j(Y) \mid X = x].$$

**Proof.** See Appendix C. ■

We obtain asymptotic results, which may be directly compared to Theorem 2.2. The constrained and unconstrained estimators of the canonical correlations have the same asymptotic distribution under the reversibility hypothesis. In contrast the asymptotic variance of the constrained estimated canonical variate is half the variance of the unconstrained one. This is a correction for the double size of the sample (see interpretation *i*) of the constrained estimator).

## 3.2 Comparison of constrained and unconstrained estimators

Under the null hypothesis of reversibility we can compare the asymptotic properties of the constrained and unconstrained estimators of the canonical correlations and canonical directions. The difference between these two types of estimators can be used to construct testing procedures of the reversibility hypothesis.

### 3.2.1 Asymptotic equivalence

Under the reversibility hypothesis,  $\hat{\lambda}_{1N}$  and  $\hat{\lambda}_{1N}^R$  admit the same first order expansion. Therefore we need the second order expansion of the difference  $\hat{\lambda}_{1N} - \hat{\lambda}_{1N}^R$  to discuss testing procedures. The property below is proved in

Appendix D. The notation  $\sim$  means that the residual term can be neglected with respect to the terms of the left and right hand sides.

**Theorem 3.2** : *Under the reversibility hypothesis, we have the asymptotic equivalences:*

$$2(\hat{\varphi}_{1N} - \hat{\varphi}_{1N}^R) \sim \hat{\varphi}_{1N} - \hat{\psi}_{1N},$$

and

$$\begin{aligned} \hat{\lambda}_{1N} - \hat{\lambda}_{1N}^R &\sim 2\lambda_1 \int [\hat{\varphi}_{1N}(y) - \hat{\varphi}_{1N}^R(y)]^2 f(., y) dy \\ &\sim \frac{\lambda_1}{2} \int [\hat{\varphi}_{1N}(y) - \hat{\psi}_{1N}(y)]^2 f(., y) dy. \end{aligned}$$

**Proof.** See Appendix D. ■

The first relation means that it is equivalent to construct a test procedure based on the difference between the constrained and unconstrained estimators of the canonical variates  $\varphi_1$ , or the difference between the unconstrained estimators of the canonical variates  $\varphi_1$  and  $\psi_1$ . The second equation shows that a testing procedure based on the difference between the constrained and unconstrained canonical correlations consists in introducing an appropriate measure of discrepancy between the estimators of  $\varphi_1$ .

### 3.2.2 Test procedure

A test procedure can be deduced from the asymptotic properties of:

$$I_N = \mu_1 \int [\hat{\varphi}_{1N}(y) - \hat{\psi}_{1N}(y)]^2 f(., y) dy. \quad (3)$$

The difference between  $\hat{\varphi}_{1N}$  and  $\hat{\psi}_{1N}$  has been weighted by the marginal distribution  $f(., y)$  to keep the interpretation of Theorem 3.2. Similar results can be derived if  $f(., y)$  is replaced by another weighting function  $w(y)$  (see e.g.

Hall (1984a), Tenreiro (1997)). For instance we may use  $w(y) = f^2(\cdot, y)$  (see Pagan, Ullah (1999), p. 168).

The analysis is similar to the one usually followed for the integrated square error of nonparametric density or regression estimators (see e.g. Bickel, Rosenblatt (1975), Nadaraya (1983), Hall (1984a), Hall (1984b)). However these results are generally derived for i.i.d. observations, and we will use an extension of limit theorems established by Tenreiro (1997) (see also Meloche (1990)).

We assume that the process  $Z_i = (X_i, Y_i)'$  is strongly stationary and geometrically absolutely regular (see Bradley (1986)). Then, under regularity conditions described in Appendix E, we get the theorem below:

**Theorem 3.3** : *Under the reversibility hypothesis,*

*i) the asymptotic distribution of  $I_N - EI_N$  is a gaussian distribution:*

$$Nh_N^{d/2}(I_N - EI_N) \xrightarrow{d} \mathcal{N}(0, \eta^2),$$

*where:*

$$\eta^2 = 2 \left\{ \int \left[ \int K(u)K(u+v)du \right]^2 dv \right\} \int (\text{Var} [\varphi_1(X_{i+1}) | X_i = x_i])^2 dx_i.$$

*ii) The bias is:*

$$EI_N = \frac{2}{Nh_N^d} \int K^2(u)du \int \text{Var} [\varphi_1(X_{i+1}) | X_i = x_i] dx_i + o\left(\frac{1}{Nh_N^{d-1}}\right).$$

After replacement of  $I_N$  by a consistent estimator, we deduce from Theorem 3.3 a procedure for testing the reversibility hypothesis.

The test procedures described above can be used to check if discrete time data are compatible with an underlying continuous time diffusion model. To solve the question we can proceed in two steps. First apply a test for embeddability (see e.g. Florens, Renault, Touzi (1998)) to check if the eigenvalue are

strictly positive, that is if the data are compatible with a continuous time model (not necessarily a diffusion). If the the embeddability hypothesis is not rejected, the reversibility test can then be performed.

## 4 Applications

In this section we provide two illustrations of the approach. The first one is based on an artificial dataset consisting of simulated realizations of a reflected Brownian motion. This is a Markov reversible process providing a basis for a comparison of different estimation techniques. The second example involves high frequency data on returns on the Alcatel stock traded on the Paris-Bourse.

### 4.1 Reflected Brownian motion

In this example we consider a reflected Brownian motion, with zero drift, a variance  $\sigma^2$ , and two reflecting barriers at 0 and  $l$ . This process is stationary, markovian and reversible. Its infinitesimal generator:

$$\mathcal{A}f(x) = \lim_{h \rightarrow 0} \left[ \frac{E[f(X_{t+h}) | X_t = x] - f(x)}{h} \right] = \frac{1}{2} \sigma^2 \frac{d^2}{dx^2} f(x), \quad (4)$$

is defined for any function  $f$  belonging to  $D(\Lambda) = \{f \in L^2 : \mathcal{A}f \text{ exists and } f'(0) = f'(l) = 0\}$ , where  $L^2$  is the space of square integrable functions with respect to Lebesgue measure on  $[0, l]$ . The eigenelements  $(\rho_i, e_i)$  of  $\mathcal{A}$  are (see Darolles, Laurent (2000) for the computation):

$$\rho_i = -\frac{1}{2} \left( \frac{i\sigma\pi}{l} \right)^2, \quad e_i(x) = \cos \left( \frac{i\pi}{l} x \right), \quad i \text{ varying.}$$

Therefore, the canonical variates associated with discrete observations  $(X_t, X_{t-1})$  are  $\varphi_i = \psi_i = e_i$ , up to a change of sign, and the canonical correlations



are  $\lambda_i = \exp(\rho_i)$ ,  $i = 1$  varying. Moreover the reflected Brownian motion is reversible.

We simulate a path  $(X_t, t = 1, 2, \dots, T)$  of length  $T = 2500$ , for the volatility  $\sigma = 1$  and the barrier  $l = \pi$ . Next, we use these artificial observations to find the nonlinear canonical decomposition of the joint distribution of  $(X_t, X_{t-1})$ . Two estimation methods are successively applied to the data  $(X_t, X_{t-1}), t = 2, \dots, T$ :

- i*) an unconstrained kernel method, with gaussian kernels  $K_1(x) = K_2(x)$ , and bandwidths  $h_1 = h_2 = 0.1025$ ;
- ii*) the same kernel method constrained by the reversibility hypothesis.

We provide the estimated canonical correlations in Table 4.1.

[Insert here Table 4.1: Estimated canonical correlations]

The constrained and unconstrained estimators of the canonical correlations are close to each other, and close to the true values. Similar results are obtained when comparing the estimators of the canonical variates. Hence we only consider below the constrained estimation method.

[Insert here Figure 4.1: Constrained kernel estimator of the first canonical variate]

To study the asymptotic variance of the estimators, we perform the following Monte-Carlo study. We replicate 250 simulated paths using the same parameters values and we compute at each point of the support the mean and the standard deviation of the estimator of the first canonical variate. Figures (4.1) and (4.2)

provide the averaged estimators and the pointwise confidence bands, computed for the first and the second canonical variates. The figure presents the true function (dotted line) and its estimator (continuous line).

[Insert here Figure 4.2: Constrained kernel estimator of the second canonical variate]

The kernel estimators of the canonical variates satisfy approximatively the boundary constraints  $\varphi_1'(0) = \varphi_1'(l) = 0$ . This nice property is not satisfied in practice by the standard sieve method, which can create important finite sample bias (see Darolles, Gouriéroux (1997) for a modification of the sieve approach to integrate the boundary effects).

## 4.2 High Frequency Data

We consider a series of returns corresponding to the Alcatel stock traded on the Paris-Bourse. The prices are resampled every  $20mn$  from real time records and the returns are computed by differencing the log-prices. The observation period is May 2, 1997, to August 30, 1997, and includes 1705 observations. For this application, we can assume that returns take values in a compact set. Indeed, the tradings would automatically stop if the price modification with respect to the opening price was too large.

[Insert here Table 4.2: Estimated canonical correlations]

We implement the unconstrained and constrained kernel based methods, with a gaussian kernel and bandwidths  $h_{1N} = h_{2N} = 0.062$ . The reversibil-

ity hypothesis is clearly rejected when we compare the constrained and unconstrained estimated canonical correlations (see Table 4.2, their asymptotic variances are provided in Appendix *F*). The introduction of the reversibility constraint induces an underestimation of the first canonical correlation by about 30%.

[Insert here Figure 4.3: Estimated current canonical variates]

The estimated canonical variates are provided in Figures (4.3) and (4.4) for the unconstrained case. The first variate is represented by a continuous line, the second one by a dashed line, and the third one by a dotted line.

[Insert here Figure 4.4: Estimated lagged canonical variates]

It is commonly assumed in financial theory that the stock returns  $(X_t, t \geq 0)$  satisfy a stochastic differential equation:

$$dX_t = \mu(X_t) dt + \sigma(X_t) dW_t, \quad (\text{say}).$$

In such a case the process is necessarily reversible and the first canonical variate corresponds to a monotonous function, the second one to a function with one breakpoint, and so on. The comparison of the three figures shows clearly that the reversibility property has to be rejected, as the expected patterns of the canonical variates are. In particular, the observed returns are not compatible with an underlying unidimensional stochastic differential equation.

How to interpret the pattern of the first canonical variate ? It is well known that the (linear) autocorrelations of stock returns are generally insignificant,

which is consistent with the theory of market efficiency. In our case the first order linear correlation is 0.065 and is not significant. Therefore the linear transformation will not belong to the subspaces generated by the first canonical variates. Moreover the literature on ARCH models insists on the so-called volatility persistence, implying a large autocorrelation of squared returns. Therefore it is not surprising to find a first canonical variate with a parabolic form, even if the pattern also includes some leverage effect (see Black (1976)).

## 5 Concluding Remarks

In this paper we develop a nonlinear canonical correlation analysis based on kernel estimators of the density function. It allows to define nonparametric estimators of the canonical correlations and canonical directions either unconstrained, or constrained by the reversibility property. This nonparametric canonical analysis is especially useful for financial applications based on high frequency data and for the estimation of the drift and volatility functions of a diffusion equations. It may also be used to study the liquidity risk in an analysis of intertrade durations (see Gouriéroux, Jasiak (1998)), or to implement technical analysis based models for directions of price changes (see Darolles, Gouriéroux, Le Fol (2000)).

## A Proof of Theorem 2.1

From Roussas (1988), theorem 3.1, Bosq (1998), theorem 2.2, we get the following lemma.

**Lemma A.1** : *Under assumptions A.2-A.9, the kernel estimator of the p.d.f.*

is uniformly strongly consistent.

From this result, we deduce a uniform consistency property of integrals with respect to  $\hat{f}_N$ .

**Lemma A.2 :** *Under assumptions A.2-A.10,  $\int g(x, y)\hat{f}_N(x, y)dxdy$  converges a.s. uniformly to  $\int g(x, y)f(x, y)dxdy$ , for any function  $g$  in  $\mathcal{G} = \{g : \int |g(x, y)| f(x, y)dxdy \leq 1\}$ .*

**Proof.** For any function  $g$  in  $\mathcal{G} = \{g : \int |g(x, y)| f(x, y)dxdy \leq 1\}$ , we get:  $\sup_{g \in \mathcal{G}} | \int g(x, y)[\hat{f}_N(x, y) - f(x, y)]dxdy | \leq \sup_{(x, y) \in \mathcal{X}^2} | \hat{f}_N(x, y) - f(x, y) | / f(x, y)$ . Using assumption A.10, this uniform convergence is equivalent to the uniform convergence of  $\hat{f}_N(x, y)$  to  $f(x, y)$  given by Lemma A.1. ■

Let us now prove the consistency for  $i = 1$ . The case  $i \neq 1$  can be easily deduced by similar arguments. Let us introduce the three following maximization problems:

**Problem A.1:**  $(\hat{\varphi}_{1,N}, \hat{\psi}_{1,N})$  solution of  $\max_{\varphi, \psi} \int \varphi(x) \psi(y) \hat{f}_N(x, y) dxdy$ , subject to  $\int \varphi^2(x) \hat{f}_N(x, \cdot) dx = \int \psi^2(y) \hat{f}_N(\cdot, y) dy = 1$ .

**Problem A.2:**  $(\tilde{\varphi}_{1,N}, \tilde{\psi}_{1,N})$  solution of  $\max_{\varphi, \psi} \int \varphi(x) \psi(y) \hat{f}_N(x, y) dxdy$ , subject to  $\int \varphi^2(x) f(x, \cdot) dx = \int \psi^2(y) f(\cdot, y) dy = 1$ .

**Problem A.3:**  $(\varphi_1, \psi_1)$  solution of  $\max_{\varphi, \psi} \int \varphi(x) \psi(y) f(x, y) dxdy$ , subject to  $\int \varphi^2(x) f(x, \cdot) dx = \int \psi^2(y) f(\cdot, y) dy = 1$ .

Let us denote  $\mathcal{H} = \{(\varphi, \psi) : \int \varphi^2(x) f(x, \cdot) dx = \int \psi^2(y) f(\cdot, y) dy = 1\}$ . We first prove the consistency of  $\tilde{\varphi}_{1,N}$  and  $\tilde{\psi}_{1,N}$ .

*i) Consistency of  $\tilde{\varphi}_{1,N}$  and  $\tilde{\psi}_{1,N}$*

Using Cauchy-Schwarz inequality, we get  $g(x, y) = \varphi(x) \psi(y) \in \mathcal{G}$ ,  $\forall (\varphi, \psi) \in \mathcal{H}$ , and by Lemma A.2 we deduce the a.s. uniform convergence  $\int \varphi(x) \psi(y) \hat{f}_N(x, y) dx dy \rightarrow \int \varphi(x) \psi(y) f(x, y) dx dy$  as  $N \rightarrow \infty$ . Moreover the application  $(\varphi, \psi) \rightarrow \int \varphi(x) \psi(y) f(x, y) dx dy$ , is continuous from  $L^2(X) \times L^2(Y)$  to  $\mathbf{R}$ . We use a Jennrich's lemma valid for any metric space (see Jennrich (1969), proof of Theorem 6) to deduce the a.s. convergence of the solutions of the finite sample problem to the solution of the limit problem A.3. This implies:  $\|\tilde{\varphi}_{1,N} - \varphi_1\|_2 \rightarrow 0$  a.s. and  $\|\tilde{\psi}_{1,N} - \psi_1\|_2 \rightarrow 0$  a.s., where  $\|\cdot\|_2$  is the  $L^2$  distance.

*ii) Consistency of  $\hat{\varphi}_{1,N}$  and  $\hat{\psi}_{1,N}$*

Since  $g(x, y) = \varphi^2(x) \in \mathcal{G}$  and  $g(x, y) = \psi^2(y) \in \mathcal{G}$ , when  $(\varphi, \psi) \in \mathcal{H}$ , we deduce from Lemma A.2 the equicontinuity property, which implies the a.s. convergences:  $\alpha_N = \int \tilde{\varphi}_{1,N}^2(x) \hat{f}_N(x, \cdot) dx \rightarrow \int \varphi_1^2(x) f(x, \cdot) dx = 1$ , a.s., and  $\beta_N = \int \tilde{\psi}_{1,N}^2(y) \hat{f}_N(\cdot, y) dy \rightarrow \int \psi_1^2(y) f(\cdot, y) dy = 1$ , a.s.. The solutions  $(\tilde{\varphi}_{1,N}, \tilde{\psi}_{1,N})$  and  $(\hat{\varphi}_{1,N}, \hat{\psi}_{1,N})$  of problems A.1 and A.2 are proportional up to the terms  $(\alpha_N^{1/2}, \beta_N^{1/2})$ . Therefore we have  $\|\hat{\varphi}_{1,N} - \varphi_1\|_2 \leq \|\tilde{\varphi}_{1,N} - \varphi_1\|_2 + \|(1 - \alpha_N^{-\frac{1}{2}}) \tilde{\varphi}_{1,N}\|_2 = \|\tilde{\varphi}_{1,N} - \varphi_1\|_2 + |1 - \alpha_N^{-\frac{1}{2}}|$ , and we deduce  $\|\hat{\varphi}_{1,N} - \varphi_1\|_2 \rightarrow 0$  a.s. A similar computation gives the consistency of  $\hat{\psi}_{1,N}$ .

*iii) Consistency of  $\hat{\lambda}_{1,N}$*

It is a consequence of the equicontinuity property.

## B Proof of Theorem 2.2

From Bosq (1998), theorem 2.3, we get the following central limit theorem for the kernel density estimator.

**Lemma B.1** : *Under assumptions A.2-A.8, A.11-A.14, for any  $(x, y) \in ]0, 1[^{2d}$ ,*

$$\sqrt{Nh_{2N}^d h_{1N}^d} [\hat{f}_N(x, y) - f(x, y)] \xrightarrow{d} \mathcal{N}(0, W(x, y)), \text{ where } W(x, y) = f(x, y) \int K_1^2(u) du \int K_2^2(u) du.$$

To derive the asymptotic distributions, we need central limit theorems for specific transformations of the kernel density estimator.

**Lemma B.2 :** *Under assumptions A.2-A.8, A.11-A.14, we get for a given function  $g(x, y)$ :*

i)  $\sqrt{N} \int g(x, y) [\hat{f}_N(x, y) - f(x, y)] dx dy \xrightarrow{d} \mathcal{N}(0, V)$ , with asymptotic variance

$$V = \sum_{k=-\infty}^{\infty} \text{Cov}[g(X_n, Y_n), g(X_{n+k}, Y_{n+k})].$$

ii)  $\sqrt{Nh_{1N}^d} \int g(x, y) [\hat{f}_N(x, y) - f(x, y)] dy \xrightarrow{d} \mathcal{N}(0, W(x))$ , with asymptotic variance

$$W(x) = \int K_1^2(u) du \int g^2(x, y) f(x, y) dy.$$

iii) *If the reversibility hypothesis is satisfied,*

$$\sqrt{Nh_N^d} \int g(x, y) [\hat{f}_N^R(x, y) - f(x, y)] dy \xrightarrow{d} \mathcal{N}(0, W(x)), \text{ with asymptotic variance}$$

$$W(x) = \frac{1}{2} \int K^2(u) du \int g^2(x, y) f(x, y) dy.$$

We can now begin the proof of theorem 3.2. We derive asymptotic distributions for  $p = 1$ . The case  $p \neq 1$  is easily deduced by similar arguments.

**i) Expansion of the first order condition**

Let us denote by  $L^2(X)$  (resp.  $L^2(Y)$ ) the space of square integrable functions of  $X$  (resp.  $Y$ ), by  $\|\cdot\|_2$  the associated  $L^2$  norm, by  $T$  the conditional expectation operator from  $L^2(X)$  to  $L^2(Y)$ :

$$T : \varphi(X) \rightarrow T\varphi(Y) = E[\varphi(X) | Y] = \int \varphi(x) \frac{f(x, y)}{f(\cdot, y)} dx,$$

and by  $T^*$  the adjoint operator of  $T$ :

$$T^* : \psi(Y) \rightarrow T\psi(X) = E[\psi(Y) | X] = \int \psi(y) \frac{f(x, y)}{f(x, \cdot)} dy.$$

The first order conditions associated with Problem A.3 are:  $T^*T\varphi_1 = \mu_1\varphi_1$  and  $TT^*\psi_1 = \mu_1\psi_1$ , where  $\mu_1 = \lambda_1^2$ . Moreover  $\langle \varphi_1, \varphi_1 \rangle = 1$ ,  $\langle \psi_1, \psi_1 \rangle = 1$ , and:

$$T\varphi_1 = \mu_1^{\frac{1}{2}}\psi_1, T^*\psi_1 = \mu_1^{\frac{1}{2}}\varphi_1.$$

$T$ ,  $T^*$ ,  $\varphi_1$ ,  $\psi_1$  and  $\mu_1$  are function of the distribution  $F$ . We denote by  $dT_F(H)$  (resp.  $dT_F^*(H)$ ,  $d\varphi_{1F}(H)$ ,  $d\mu_{1F}(H)$ ) the Gâteaux derivative of  $T$  (resp.  $T^*$ ,  $\varphi_1$ ,  $\mu_1$ ) in the direction  $H$  at the point  $F$  ( $h(x, y)$  denotes the p.d.f. of  $H$ ). We will derive the explicit expressions of  $d\mu_{1F}(H)$  and  $d\varphi_{1F}(H)$ . The Gâteaux derivative of the first order condition corresponding to  $\mu_1$  and  $\varphi_1$  is:

$$d\mu_{1F}(H)\varphi_1 = dT_F^*(H)T\varphi_1 + T^*dT_F(H)\varphi_1 + (T^*T - \mu_1I)d\varphi_{1F}(H), \quad (5)$$

where  $I$  is the identity operator.

*i.a) Expression of  $d\mu_{1F}(H)$*

Let us focus on  $d\mu_{1F}(H)$ . To eliminate the term  $d\varphi_{1F}(H)$  in the expression above, we compute the inner product of this equation with  $\varphi_1$ . We find:

$$d\mu_{1F}(H)\langle \varphi_1, \varphi_1 \rangle = \langle \varphi_1, (dT_F^*(H)T + T^*dT_F(H))\varphi_1 \rangle + \langle \varphi_1, (T^*T - \mu_1I)d\varphi_{1F}(H) \rangle.$$

By the normalization condition  $\langle \varphi_1, \varphi_1 \rangle = 1$ , its derivative satisfies:

$$\langle \varphi_1, d\varphi_{1F}(H) \rangle = 0.$$

Moreover by using the definitions of the adjoint operator and of  $\psi_1$ , we get:

$$d\mu_{1F}(H) = \mu_1^{\frac{1}{2}}[\langle \varphi_1, dT_F^*(H)\psi_1 \rangle + \langle \psi_1, dT_F(H)\varphi_1 \rangle]. \quad (6)$$

Let us now compute the quantities  $dT_F(H)\varphi_1(y)$  and  $dT_F^*(H)\psi_1(x)$ . We get:

$$\begin{aligned} dT_F(H)\varphi_1(y) &= \int \varphi_1(x) \left[ \frac{h(x, y)}{f(., y)} - \frac{f(x, y)}{f(., y)^2} h(., y) \right] dx \\ &= \frac{1}{f(., y)} \left[ \int \varphi_1(x) h(x, y) dx - T\varphi_1(y) h(., y) \right] \\ &= \frac{1}{f(., y)} \int (\varphi_1(x) - \mu_1^{\frac{1}{2}}\psi_1(y)) h(x, y) dx, \end{aligned}$$



and

$$dT_F^*(H)\psi_1(x) = \frac{1}{f(x, \cdot)} \int (\psi_1(y) - \mu_1^{\frac{1}{2}} \varphi_1(x)) h(x, y) dy. \quad (7)$$

By introducing this expression in condition (6), we get the expression of  $d\mu_{1F}(H)$  as an integral with respect to  $h$ :

$$d\mu_{1F}(H) = \mu_1^{\frac{1}{2}} \int [2\varphi_1(x)\psi_1(y) - \mu_1^{\frac{1}{2}} \varphi_1^2(x) - \mu_1^{\frac{1}{2}} \psi_1^2(y)] h(x, y) dx dy. \quad (8)$$

*i.b) Expression of  $d\varphi_{1F}(H)$*

From (5), we get:

$$\begin{aligned} & (\mu_1 I - T^*T)d\varphi_{1F}(H) \\ &= dT_F^*(H)T\varphi_1 + T^*dT_F(H)\varphi_1 - d\mu_{1F}(H)\varphi_1 \\ &= dT_F^*(H)T\varphi_1 + T^*dT_F(H)\varphi_1 - \langle \varphi_1, dT_F^*(H)T\varphi_1 + T^*dT_F(H)\varphi_1 \rangle \varphi_1. \end{aligned}$$

The r.h.s. belongs to the null space of the operator  $\mu_1 I - T^*T$ , which implies the existence of a solution  $d\varphi_{1F}(H)$ . Moreover the solution is unique due to the constraint  $\langle \varphi_1, d\varphi_{1F}(H) \rangle = 0$ . It satisfies the Fredholm equation:

$$(\mu_1 I - T^*T)d\varphi_{1F}(H) = dv_F(H),$$

where  $dv_F(H)$  denotes the r.h.s.. It can be written as:

$$d\varphi_{1F}(H)(x) = \frac{1}{\mu_1} dv_F(H) - \int b(x, s) dv_F(H) ds,$$

where:  $b(x, s) = \frac{1}{\mu_1} \sum_{j \neq 1} \frac{\mu_1}{\mu_1 - \mu_k} \varphi_j(x) \varphi_j(s)$ . Assumption A.15 ensures the convergence of the series defining  $b$ .

*i.c) Frechet differentiability*

The directional Gâteaux derivatives can be used to derive the first order expansions of  $\mu_1(F + H) - \mu_1(F)$ ,  $\varphi_1(F + H) - \varphi_1(F)$  and  $\psi_1(F + H) - \psi_1(F)$ , whenever these functions are Frechet differentiable.

Let us consider the canonical variate  $\varphi_1(F)$  which is solution of the implicit equation:

$$T_F^* T_F \varphi_1 - \langle T_F \varphi_1, T_F \varphi_1 \rangle_F \varphi_1 = 0.$$

It is easily checked that the application:

$$g : (F, \varphi) \rightarrow T_F^* T_F \varphi - \langle T_F \varphi, T_F \varphi \rangle_F \varphi$$

is Frechet differentiable for the  $L^2(X)$  norm for  $\varphi$  and  $g$ , and the norm:

$$\|F\|_{(\infty,1)} = \sup_{x,y} |F(x,y)| + \sup_{x,y} \sum_{i=1}^d \left| \frac{\partial F}{\partial x_i}(x,y) \right| + \sup_{x,y} \sum_{i=1}^d \left| \frac{\partial F}{\partial y_i}(x,y) \right|,$$

for  $F$ . We deduce the differentiability of  $\varphi_1$  with respect to  $F$  by the implicit function theorem. Indeed we have:

$$\frac{\partial g}{\partial \varphi}(F, \varphi_1) d\varphi_{1F} = (T^* T - \mu_1 I) d\varphi_{1F} \neq 0.$$

Similarly we can check that:

$$\mu_1(F) = \max_{\varphi, \psi} \langle T_F \varphi, \psi \rangle_F,$$

where the maximum is computed on the pair  $(\varphi, \psi)$  of continuously differentiable functions, is Frechet differentiable for the norm:

$$\|F\|_{(\infty,0)} = \sup_{x,y} |F(x,y)|,$$

for  $F$  and the standard norm of  $\mathbf{R}$  for  $\mu$ .

*ii) Asymptotic distribution of  $\hat{\lambda}_{1,N}$*

By the Frechet differentiability of  $\mu_1(F)$  and the assumption A.18, we deduce the first order expansion of  $\hat{\mu}_{1,N}$ . We have:

$$\hat{\mu}_{1,N} - \mu_1 = \mu_1(\hat{F}_N) - \mu_1(F) = d\mu_{1F}(\hat{F}_N - F) + o\left(\|\hat{F}_N - F\|_{(\infty,0)}\right),$$

and:

$$\sqrt{N}(\hat{\mu}_{1,N} - \mu_1) = \sqrt{N}d\mu_{1F}(\hat{F}_N - F) + o\left(\sqrt{N}\|\hat{F}_N - F\|_{(\infty,0)}\right).$$

Since  $\sqrt{N}\|\hat{F}_N - F\|_{(\infty,0)} = O_p(1)$  (see Aït-Sahalia (1992), proof of theorem 2), the two terms  $\sqrt{N}(\hat{\mu}_{1,N} - \mu_1)$  and  $\sqrt{N}d\mu_{1F}(\hat{F}_N - F)$  have the same asymptotic distribution. We get:

$$\begin{aligned} \sqrt{N}d\mu_{1F}(\hat{F}_N - F) &= \sqrt{N} \int \mu_1^{\frac{1}{2}} [2\varphi_1(x)\psi_1(y) - \mu_1^{\frac{1}{2}}\varphi_1^2(x) - \mu_1^{\frac{1}{2}}\psi_1^2(y)] \\ &\quad \times [\hat{f}_N(x, y) - f(x, y)] dx dy, \end{aligned}$$

and the asymptotic distribution of  $\hat{\mu}_{1,N}$  follows from Lemma B.2 *i*) with  $g(x, y) = \mu_1^{\frac{1}{2}} [2\varphi_1(x)\psi_1(y) - \mu_1^{\frac{1}{2}}\varphi_1^2(x) - \mu_1^{\frac{1}{2}}\psi_1^2(y)]$ . We immediately deduce the asymptotic distribution of  $\hat{\lambda}_{1,N} = (\hat{\mu}_{1,N})^{\frac{1}{2}}$ .

*iii*) **Asymptotic distribution of  $\hat{\varphi}_{1,N}$**

By the Frechet differentiability of  $\varphi_1(F)$  and the assumption A.18, we deduce the first order expansion of  $\hat{\varphi}_{1,N}$ . We get:

$$\hat{\varphi}_{1,N} - \varphi_1 = \varphi_1(\hat{F}_N) - \varphi_1(F) = d\varphi_{1F}(\hat{F}_N - F) + o\left(\|\hat{F}_N - F\|_{(\infty,1)}\right),$$

and

$$\sqrt{Nh_{1N}^d}(\hat{\varphi}_{1,N} - \varphi_1) = \sqrt{Nh_{1N}^d}d\varphi_{1F}(\hat{F}_N - F) + o\left(\sqrt{Nh_{1N}^d}\|\hat{F}_N - F\|_{(\infty,1)}\right).$$

Since  $\sqrt{Nh_{1N}^d}\|\hat{F}_N - F\|_{(\infty,1)} = O_p(1)$  (see Aït-Sahalia (1992), proof of theorem 3),  $\sqrt{Nh_{1N}^d}(\hat{\varphi}_{1,N} - \varphi_1)$  and  $\sqrt{Nh_{1N}^d}d\varphi_{1F}(\hat{F}_N - F)$  have the same asymptotic distribution (see Serfling (1980), Lemma B p. 218). Moreover by using the expression of  $d\varphi_{1F}(H)$  and comparing the rates of convergence, we get:

$$\begin{aligned} \sqrt{Nh_{1N}^d}d\varphi_{1F}(\hat{F}_N - F)(x) &= \frac{1}{\mu_1} \sqrt{Nh_{1N}^d}dT_F^*(H)T\varphi_1(x) + o_p(1) \\ &= \frac{1}{\lambda_1} \frac{1}{f(x, \cdot)} \sqrt{Nh_{1N}^d} \int [\psi_1(y) - \lambda_1\varphi_1(x)] \\ &\quad \times [\hat{f}_N(x, y) - f(x, y)] dy + o_p(1). \end{aligned} \quad (9)$$

We apply Lemma B.2 *ii*) with  $g(x, y) = \frac{1}{\lambda_1} \frac{1}{f(x, \cdot)} [\psi_1(y) - \lambda_1 \varphi_1(x)]$  to get the asymptotic distribution of  $\hat{\varphi}_{1,N}(x)$ :

$$\sqrt{Nh_{1N}^d}(\hat{\varphi}_{1,N}(x) - \varphi_1(x)) \xrightarrow{d} \mathcal{N}(0, W_{1,1}(x)),$$

where

$$\begin{aligned} W_{1,1}(x) &= \int K_1^2(u) du \frac{1}{\lambda_1^2} \frac{1}{f^2(x, \cdot)} \int [\psi_1(y) - \lambda_1 \varphi_1(x)]^2 f(x, y) dy \\ &= \frac{1}{\lambda_1^2} \frac{1}{f(x, \cdot)} \int K_1^2(u) du V[\psi_1(Y) | X = x]. \end{aligned}$$

The proof is similar for deriving the asymptotic distributions of  $(\hat{\varphi}_{1,N}(x), \dots, \hat{\varphi}_{p,N}(x))$  and of  $(\hat{\psi}_{1,N}(x), \dots, \hat{\psi}_{p,N}(x))$ .

## C Proof of Theorem 3.1

The proof of consistency is similar to the proof given for theorem 2.1 in Appendix A. We derive asymptotic distributions of the estimators under the reversibility hypothesis. We only present the expansions for the constrained estimators of type *iii*) defined by:

$$\frac{1}{2}(\hat{T}^* \hat{T} + \hat{T} \hat{T}^*) \hat{\varphi}_{i,N}^R = \hat{\mu}_{i,N}^R \hat{\varphi}_{i,N}^R, \quad \langle \hat{\varphi}_{i,N}^R, \hat{\varphi}_{i,N}^R \rangle = 1,$$

where  $\hat{\mu}_{i,N}^R = (\hat{\lambda}_{i,N}^R)^2$ . For expository purpose, we consider the case  $p = 1$ . The Gâteaux derivative of the first order condition is:

$$d\mu_{1F}^R(H)\varphi_1 = \frac{1}{2}d(T_F^* T_F + T_F T_F^*)(H)\varphi_1 + \left(\frac{1}{2}(T^* T + T T^*) - \mu_1 I\right) d\varphi_{1F}^R(H).$$

The Gâteaux derivatives are computed at a point  $F$  which satisfies the re-

reversibility hypothesis and is such that  $T^* = T$ ,  $\varphi_1 = \psi_1$ . We get:

$$\begin{aligned} d\mu_{1F}^R(H)\varphi_1 &= \frac{1}{2} [dT_F^*(H)T + TdT_F(H) + dT_F(H)T + TdT_F^*(H)]\varphi_1 \\ &\quad + (T^2 - \mu_1 I)d\varphi_{1F}^R(H), \end{aligned} \quad (10)$$

where  $I$  is the identity operator.

*i) Expression of  $d\mu_{1F}^R(H)$*

To eliminate the terms  $d\varphi_{1F}(H)$  in the previous expression, we compute the inner product of this equation with  $\varphi_1$ . We find:

$$d\mu_{1F}^R(H) = \mu_1^{\frac{1}{2}} [\langle \varphi_1, dT_F^*(H)\varphi_1 \rangle + \langle \varphi_1, dT_F(H)\varphi_1 \rangle]. \quad (11)$$

By comparing with the Gâteaux derivative of the unconstrained estimator (see appendix B), we note that:

$$d\mu_{1F}^R(H) = d\mu_{1F}(H),$$

under the reversibility hypothesis. In particular the constrained and unconstrained estimated canonical correlations have the same asymptotic distributions.

*ii) Expression of  $d\varphi_{1F}^R(H)$*

From (10), we get:

$$(\mu_1 I - T^2)d\varphi_{1F}^R(H) = dv_F^R(H),$$

where

$$dv_F^R(H) = \frac{1}{2} [(dT_F^*(H) + dT_F(H))T + T(dT_F^*(H) + dT_F(H))] \varphi_1 - d\mu_{1F}^R(H)\varphi_1.$$

When  $H$  is replaced by  $\hat{F}_N - F$  in the expansion and by comparing the rates

of convergence, we get:

$$\begin{aligned}
& \sqrt{Nh_N^d} d\hat{\varphi}_{1,N}^R(\hat{F}_N - F)(y) \\
&= \frac{\sqrt{Nh_N^d}}{\mu_1} \frac{1}{2} \left[ dT_F^*(\hat{F}_N - F) + dT_F(\hat{F}_N - F) \right] T\varphi_1(y) + o_p(1) \\
&= \frac{\sqrt{Nh_N^d}}{\lambda_1} \frac{1}{f(\cdot, y)} \int (\varphi_1(x) - \mu_1^{\frac{1}{2}} \varphi_1(y)) \left[ \frac{\hat{f}_N(x, y) + \hat{f}_N(y, x)}{2} - f(x, y) \right] dx + o_p(1) \\
&= \frac{\sqrt{Nh_N^d}}{\lambda_1} \frac{1}{f(\cdot, y)} \int (\varphi_1(x) - \mu_1^{\frac{1}{2}} \varphi_1(y)) \left[ \hat{f}_N^R(x, y) - f(x, y) \right] dx + o_p(1).
\end{aligned}$$

We apply Lemma B.2 *iii*) with  $g(x, y) = \frac{1}{\lambda_1} \frac{1}{f(\cdot, y)} [\varphi_1(x) - \lambda_1 \varphi_1(y)]$  to obtain the asymptotic distribution of  $\hat{\varphi}_{1,N}^R$ :

$$\sqrt{Nh_N^d} (\hat{\varphi}_{i,N}^R(x) - \varphi_i(x)) \xrightarrow{d} \mathcal{N}(0, W_{1,1}^R(x)),$$

where

$$\begin{aligned}
W_{1,1}^R(x) &= \frac{1}{2} \frac{1}{\lambda_1^2} \frac{1}{f^2(x, \cdot)} \int K_1^2(u) du \int [\varphi_1(y) - \lambda_1 \varphi_1(x)]^2 f(x, y) dy \\
&= \frac{1}{2} \frac{1}{\lambda_1^2} \frac{1}{f(x, \cdot)} \int K_1^2(u) du V[\varphi_1(Y) | X = x].
\end{aligned}$$

## D Proof of Theorem 3.2

### *i*) Second order Gâteaux derivatives

Let us derive the second order Gâteaux derivative of  $d\mu_{1,F} - d\mu_{1,F}^R$ , at a point  $F$  satisfying the reversibility hypothesis. From (8) and (11) we have:

$$d\mu_{1F}(H) = \int [2\mu_1^{\frac{1}{2}} \varphi_1(x) \psi_1(y) - \mu_1 \varphi_1^2(x) - \mu_1 \psi_1^2(y)] h(x, y) dx dy,$$

and

$$d\mu_{1F}^R(H) = \int [2\mu_1^{\frac{1}{2}} \varphi_1(x) \varphi_1(y) - \mu_1 \varphi_1^2(x) - \mu_1 \varphi_1^2(y)] h(x, y) dx dy.$$

Then we compute  $d^2\mu_{1F}(H, L) - d^2\mu_{1F}^R(H, L)$  as

$$\begin{aligned}
& \int d \left[ 2\mu_1^{\frac{1}{2}}\varphi_1(x)\psi_1(y) - 2\mu_1^{\frac{1}{2}}\varphi_1(x)\varphi_1(y) - \mu_1\psi_1^2(y) + \mu_1\varphi_1^2(y) \right] (L)h(x, y)dxdy \\
= & \int d \left[ \left( 2\mu_1^{\frac{1}{2}}\varphi_1(x) - \mu_1(\psi_1(y) + \varphi_1(y)) \right) (\psi_1(y) - \varphi_1(y)) \right] (L)h(x, y)dxdy \\
= & 2 \int \mu_1^{\frac{1}{2}} \left[ \varphi_1(x) - \mu_1^{\frac{1}{2}}\varphi_1(y) \right] d(\psi_{1F} - \varphi_{1F})(L)(y)h(x, y)dxdy, \tag{12}
\end{aligned}$$

since  $\psi_1 = \varphi_1$  under the reversibility hypothesis. The notation  $d(\psi_{1F} - \varphi_{1F})(L)$  is used for  $d\psi_{1F}(L) - d\varphi_{1F}(L)$ .

*ii) Second order expansion of  $\hat{\mu}_{1,N} - \hat{\mu}_{1,N}^R$*

The first nonzero term of the expansion is  $d^2(\mu_{1F} - \mu_{1F}^R)(L, L)$ , where  $L = \hat{F}_N - F$ . From (12), we get:

$$\begin{aligned}
& \hat{\mu}_{1,N} - \hat{\mu}_{1,N}^R \\
\sim & 2\mu_1 \int \frac{1}{\lambda_1} \frac{1}{f(\cdot, y)} (\varphi_1(x) - \lambda_1\varphi_1(y)) d(\psi_{1F} - \varphi_{1F})(L)(y)l(x, y)f(\cdot, y)dxdy \\
= & 2\mu_1 \int \left[ \frac{1}{\lambda_1} \frac{1}{f(\cdot, y)} \int (\varphi_1(x) - \lambda_1\varphi_1(y))l(x, y)dx \right] \\
& \quad \times d(\psi_{1F} - \varphi_{1F})(L)(y)f(\cdot, y)dy. \tag{13}
\end{aligned}$$

We know from (9) that:

$$d\varphi_{1F}(L)(y) \sim \frac{1}{\lambda_1} \frac{1}{f(\cdot, y)} \int (\varphi_1(s) - \lambda_1\varphi_1(y))l(y, s)ds,$$

and

$$d\psi_{1F}(L)(y) \sim \frac{1}{\lambda_1} \frac{1}{f(\cdot, y)} \int (\varphi_1(s) - \lambda_1\varphi_1(y))l(s, y)ds;$$

then

$$\begin{aligned}
& d(\psi_{1F} - \varphi_{1F})(L)(y) \\
\sim & \frac{1}{\lambda_1} \frac{1}{f(\cdot, y)} \int (\varphi_1(s) - \lambda_1\varphi_1(y)) [l(s, y) - l(y, s)] ds,
\end{aligned}$$

when  $F$  satisfies the reversibility hypothesis. Moreover we note that:

$$\begin{aligned}
& d(\varphi_{1F} - \varphi_{1F}^R)(L)(y) \\
& \sim \frac{1}{\lambda_1} \frac{1}{f(\cdot, y)} \int (\varphi_1(s) - \lambda_1 \varphi_1(y)) \left[ l(y, s) - \frac{l(y, s) + l(s, y)}{2} \right] ds \\
& \sim \frac{1}{2} \frac{1}{\lambda_1} \frac{1}{f(\cdot, y)} \int (\varphi_1(s) - \lambda_1 \varphi_1(y)) [l(y, s) - l(s, y)] ds \\
& \sim -\frac{1}{2} d(\psi_{1F} - \varphi_{1F})(L)(y). \tag{14}
\end{aligned}$$

By replacing in (13), we get:

$$\begin{aligned}
\hat{\mu}_{1,N} - \hat{\mu}_{1,N}^R & \sim -4\mu_1 \int d\psi_{1F}(L)(y) d(\varphi_{1F} - \varphi_{1F}^R)(L)(y) f(\cdot, y) dy \\
& = -4\mu_1 \langle d\psi_{1F}(L), d(\varphi_{1F} - \varphi_{1F}^R)(L) \rangle. \quad (\text{say}).
\end{aligned}$$

### iii) Additional equivalences

Since the computations are symmetrical in  $\varphi$ ,  $\psi$ , we deduce:

$$\langle d\psi_{1F}(L), d(\varphi_{1F} - \varphi_{1F}^R)(L) \rangle = \langle d\varphi_{1F}(L), d(\psi_{1F} - \psi_{1F}^R)(L) \rangle.$$

By applying (14), we get:

$$\langle d\psi_{1F}(L), d(\psi_{1F} - \varphi_{1F})(L) \rangle \sim -\langle d\varphi_{1F}(L), d(\psi_{1F} - \varphi_{1F})(L) \rangle,$$

or equivalently:

$$\langle d(\psi_{1F} + \varphi_{1F})(L), d(\psi_{1F} - \varphi_{1F})(L) \rangle \sim 0. \tag{15}$$

We deduce that:

$$\begin{aligned}
& \langle d\psi_{1F}(L), d(\varphi_{1F} - \varphi_{1F}^R)(L) \rangle \\
& \sim -\frac{1}{2} \langle d\psi_{1F}(L), d(\psi_{1F} - \varphi_{1F})(L) \rangle \quad (\text{from (14)}) \\
& \sim -\frac{1}{4} \langle d(\psi_{1F} - \varphi_{1F})(L), d(\psi_{1F} - \varphi_{1F})(L) \rangle \quad (\text{from (15)}),
\end{aligned}$$



and therefore:

$$\hat{\mu}_{1,N} - \hat{\mu}_{1,N}^R \sim \mu_1 \int [d(\psi_{1F} - \varphi_{1F})(L)(y)]^2 f(\cdot, y) dy.$$

## E Proof of Theorem 3.3

### i) Decomposition of $I_N - EI_N$

By applying (14), we get:

$$\begin{aligned} I_N &= \mu_1 \int [\hat{\varphi}_{1N}(y) - \hat{\psi}_{1N}(y)]^2 f(\cdot, y) dy \\ &= \iiint \frac{1}{f(\cdot, y)} (\varphi_1(x) - \lambda_1 \varphi_1(y)) (\varphi_1(s) - \lambda_1 \varphi_1(y)) \\ &\quad \times [\hat{l}_N(x, y) - \hat{l}_N(y, x)] [\hat{l}_N(s, y) - \hat{l}_N(y, s)] dx ds dy, \end{aligned}$$

where:  $\hat{l}_N(x, y) = \hat{f}_N(x, y) - f(x, y)$ . Under the reversibility hypothesis, we

have:  $E\hat{l}_N(X, Y) = E\hat{l}_N(Y, X)$ . We deduce that:

$$\begin{aligned} I_N - EI_N &= \frac{1}{Nh_N^{d/2}} \frac{2}{N} \sum_{i < j} [H_N(Z_i, Z_j) - EH_N(Z_i, Z_j)] \\ &\quad + \frac{1}{N^{\frac{3}{2}} h_N^{d/2}} \frac{1}{\sqrt{N}} \sum_i [H_N(Z_i, Z_i) - EH_N(Z_i, Z_i)], \end{aligned} \quad (16)$$

where:

$$\begin{aligned} &H_N(Z_i, Z_j) \\ &= \iiint \frac{1}{f(\cdot, y)} (\varphi_1(x) - \lambda_1 \varphi_1(y)) (\varphi_1(s) - \lambda_1 \varphi_1(y)) \\ &\quad \times \frac{1}{h_N^{7d/2}} \left[ K\left(\frac{x-x_i}{h_N}\right) K\left(\frac{y-y_i}{h_N}\right) - E \left[ K\left(\frac{x-x_i}{h_N}\right) K\left(\frac{y-y_i}{h_N}\right) \right] \right. \\ &\quad \left. - K\left(\frac{y-x_i}{h_N}\right) K\left(\frac{x-y_i}{h_N}\right) + E \left[ K\left(\frac{y-x_i}{h_N}\right) K\left(\frac{x-y_i}{h_N}\right) \right] \right] \\ &\quad \times \left[ K\left(\frac{s-x_j}{h_N}\right) K\left(\frac{y-y_j}{h_N}\right) - E \left[ K\left(\frac{s-x_j}{h_N}\right) K\left(\frac{y-y_j}{h_N}\right) \right] \right. \\ &\quad \left. - K\left(\frac{y-x_j}{h_N}\right) K\left(\frac{s-y_j}{h_N}\right) + E \left[ K\left(\frac{s-x_j}{h_N}\right) K\left(\frac{y-y_j}{h_N}\right) \right] \right] ds dx dy. \end{aligned} \quad (17)$$

*ii) Asymptotic distribution of  $I_N - EI_N$*

We can now apply Tenreiro (1997), Theorem 1 concerning the asymptotic behavior of  $U$ -statistics under dependence conditions. Let us denote by  $\bar{Z}_0$  an independant copy of  $Z_0$ . Under the assumptions below:

$$A.1^* \quad EH_N(Z_0, z) = 0;$$

$$A.2^* \quad H_N(Z_i, Z_j) = H_N(Z_j, Z_i);$$

$$A.3^* \quad E [H_N(Z_0, \bar{Z}_0)^2] = 2\eta^2 + o(1);$$

$$A.4^* \quad \text{There exist } \delta_0 > 0, \gamma_0 < \frac{1}{2} \text{ such that: } u_N(4+\delta_0) = O(N^{\gamma_0}), \text{ where } u_N(r) = \max \{ \max_{1 \leq i \leq N} (\|H_N(Z_i, Z_0)\|_r), \|H_N(Z_0, \bar{Z}_0)\|_r \} \text{ and } \|\zeta\|_r = (E|\zeta|^r)^{\frac{1}{r}};$$

the term:

$$\frac{1}{N} \sum_{1 \leq i < j \leq N} [H_N(Z_i, Z_j) - EH_N(Z_i, Z_j)],$$

is asymptotically normal with zero mean and variance  $\eta^2$ .

We easily check that assumptions  $A.1^*$  and  $A.2^*$  are satisfied, and the expression of  $\eta^2$  is computed below. Moreover since the second term of the decomposition (16) is of order  $O(N^{-\frac{3}{2}}h_N^d)$ , we deduce that:

$$Nh_N^{d/2}(I_N - EI_N) \xrightarrow{d} \mathcal{N}(0, \eta^2).$$

*iii) Computation of  $\eta^2$*

We get:

$$\begin{aligned}
& H_N(Z_i, \bar{Z}_i) \\
&= \frac{1}{h_N^{7d/2}} \iiint \frac{1}{f(\cdot, y)} (\varphi_1(x) - \lambda_1 \varphi_1(y)) (\varphi_1(s) - \lambda_1 \varphi_1(y)) \\
&\quad \times \left[ K\left(\frac{x-x_i}{h_N}\right) K\left(\frac{y-x_{i-1}}{h_N}\right) - K\left(\frac{x-x_{i-1}}{h_N}\right) K\left(\frac{y-x_i}{h_N}\right) \right] \\
&\quad \times \left[ K\left(\frac{s-\bar{x}_i}{h_N}\right) K\left(\frac{y-\bar{x}_{i-1}}{h_N}\right) - K\left(\frac{s-\bar{x}_{i-1}}{h_N}\right) K\left(\frac{y-\bar{x}_i}{h_N}\right) \right] ds dx dy \\
&\sim A_{N,1} + A_{N,2} + A_{N,3} + A_{N,4},
\end{aligned}$$

where:

$$\begin{aligned}
A_{N,1} &= \int \frac{1}{f(\cdot, y)} (\varphi_1(x_i) - \lambda_1 \varphi_1(y)) (\varphi_1(\bar{x}_i) - \lambda_1 \varphi_1(y)) \\
&\quad \times \frac{1}{h_N^{3d/2}} K\left(\frac{y-x_{i-1}}{h_N}\right) K\left(\frac{y-\bar{x}_{i-1}}{h_N}\right) dy, \\
A_{N,2} &= - \int \frac{1}{f(\cdot, y)} (\varphi_1(x_i) - \lambda_1 \varphi_1(y)) (\varphi_1(\bar{x}_{i-1}) - \lambda_1 \varphi_1(y)) \\
&\quad \times \frac{1}{h_N^{3d/2}} K\left(\frac{y-x_{i-1}}{h_N}\right) K\left(\frac{y-\bar{x}_i}{h_N}\right) dy, \\
A_{N,3} &= - \int \frac{1}{f(\cdot, y)} (\varphi_1(x_{i-1}) - \lambda_1 \varphi_1(y)) (\varphi_1(\bar{x}_i) - \lambda_1 \varphi_1(y)) \\
&\quad \times \frac{1}{h_N^{3d/2}} K\left(\frac{y-x_i}{h_N}\right) K\left(\frac{y-\bar{x}_{i-1}}{h_N}\right) dy, \\
A_{N,4} &= \int \frac{1}{f(\cdot, y)} (\varphi_1(x_{i-1}) - \lambda_1 \varphi_1(y)) (\varphi_1(\bar{x}_{i-1}) - \lambda_1 \varphi_1(y)) \\
&\quad \times \frac{1}{h_N^{3d/2}} K\left(\frac{y-x_i}{h_N}\right) K\left(\frac{y-\bar{x}_i}{h_N}\right) dy,
\end{aligned}$$

are functions of the independent copies  $(x_i, x_{i-1}), (\bar{x}_i, \bar{x}_{i-1})$ . When computing the quantity:

$$E [H_N(Z_i, \bar{Z}_i)^2] = E (A_{N,1} + A_{N,2} + A_{N,3} + A_{N,4})^2,$$

we get square terms of the type  $E[A_{N,j}^2]$  and cross terms like  $E[A_{N,j}A_{N,k}]$ ,  $k \neq j$ . We first check that the cross terms can be neglected. For instance we have:

$$\begin{aligned}
& E[A_{N,1}A_{N,2}] \\
= & - \int \int \int \int \left\{ \int \frac{1}{f(\cdot, y)} (\varphi_1(x_i) - \lambda_1 \varphi_1(y)) (\varphi_1(\bar{x}_i) - \lambda_1 \varphi_1(y)) \right. \\
& \quad \left. \times \frac{1}{h_N^{3d/2}} K\left(\frac{y - x_{i-1}}{h_N}\right) K\left(\frac{y - \bar{x}_{i-1}}{h_N}\right) dy \right\} \\
& \left\{ \int \frac{1}{f(\cdot, y^*)} (\varphi_1(x_i) - \lambda_1 \varphi_1(y^*)) (\varphi_1(\bar{x}_{i-1}) - \lambda_1 \varphi_1(y^*)) \right. \\
& \quad \left. \times \frac{1}{h_N^{3d/2}} K\left(\frac{y^* - x_{i-1}}{h_N}\right) K\left(\frac{y^* - \bar{x}_i}{h_N}\right) dy^* \right\} \\
& \quad f(x_i, x_{i-1}) f(\bar{x}_i, \bar{x}_{i-1}) dx_i dx_{i-1} d\bar{x}_i d\bar{x}_{i-1} \\
= & - \int \int \int \int \left\{ \int \frac{1}{f(\cdot, x_{i-1} + uh_N)} (\varphi_1(x_i) - \lambda_1 \varphi_1(x_{i-1} + uh_N)) \right. \\
& \quad \left. \times (\varphi_1(x_{i-1} - v^* h_N) - \lambda_1 \varphi_1(x_{i-1} + uh_N)) K(u) K(u + v) du \right\} \\
& \left\{ \int \frac{1}{f(\cdot, x_{i-1} + u^* h_N)} (\varphi_1(x_i) - \lambda_1 \varphi_1(x_{i-1} + u^* h_N)) \right. \\
& \quad \left. \times (\varphi_1(x_{i-1} - v h_N) - \lambda_1 \varphi_1(x_{i-1} + u^* h_N)) K(u^*) K(u^* + v^*) du^* \right\} \\
& \quad f(x_i, x_{i-1}) f(x_{i-1} - v^* h_N, x_{i-1} - v h_N) h_N^d dx_i dx_{i-1} dv dv^* \\
= & O(h_N^d).
\end{aligned}$$

Let us now consider a square term. We get:

$$\begin{aligned}
& E [A_{N,1}^2] \\
&= \int \int \int \int \left\{ \int \frac{1}{f(\cdot, y)} (\varphi_1(x_i) - \lambda_1 \varphi_1(y)) (\varphi_1(\bar{x}_i) - \lambda_1 \varphi_1(y)) \right. \\
&\quad \left. \times \frac{1}{h_N^{3d/2}} K\left(\frac{y - x_{i-1}}{h_N}\right) K\left(\frac{y - \bar{x}_{i-1}}{h_N}\right) dy \right\}^2 f(x_i, x_{i-1}) f(\bar{x}_i, \bar{x}_{i-1}) dx_i dx_{i-1} d\bar{x}_i d\bar{x}_{i-1} \\
&= \int \int \int \int \left\{ \int \frac{1}{f(\cdot, x_{i-1} + uh_N)} (\varphi_1(x_i) - \lambda_1 \varphi_1(x_{i-1} + uh_N)) (\varphi_1(\bar{x}_i) - \lambda_1 \varphi_1(x_{i-1} + uh_N)) \right. \\
&\quad \left. \times K(u) K(u+v) du \right\}^2 f(x_i, x_{i-1}) f(\bar{x}_i, x_{i-1} - vh_N) dx_i d\bar{x}_i dx_{i-1} dv \\
&\sim \int \int \int \int \left\{ \int \frac{1}{f(\cdot, x_{i-1})} (\varphi_1(x_i) - \lambda_1 \varphi_1(x_{i-1})) (\varphi_1(\bar{x}_i) - \lambda_1 \varphi_1(x_{i-1})) \right. \\
&\quad \left. \times K(u) K(u+v) du \right\}^2 f(x_i, x_{i-1}) f(\bar{x}_i, x_{i-1}) dx_i d\bar{x}_i dx_{i-1} dv \\
&= \int \left[ \int K(u) K(u+v) du \right]^2 dv \int (\text{Var} [\varphi_1(X_i) \mid X_{i-1} = x_{i-1}])^2 dx_{i-1}.
\end{aligned}$$

It is easily checked that  $EA_{N,j}^2 = EA_{N,1}^2, \forall j$ .

#### *iv) Computation of $EI_N$*

By the strong stationarity of the process  $(Z_i)$  and the symmetry of the function  $H_N$ , we get:

$$\begin{aligned}
EI_N &= \frac{1}{N^2 h_N^{d/2}} \sum_{i=1}^N \sum_{j=1}^N [EH_N(Z_i, Z_j)] \\
&= \frac{1}{N^2 h_N^{d/2}} \sum_{k=-\infty}^{+\infty} [\max(N - |k|, 0) EH_N(Z_i, Z_{i+k})] \\
&\sim \frac{1}{N h_N^{d/2}} \sum_{k=-\infty}^{+\infty} EH_N(Z_i, Z_{i+k}). \tag{18}
\end{aligned}$$

The analysis of the limiting behavior of  $EH_N(Z_i, Z_{i+k})$  is based on the equivalences below:

$$\int K_N(x, x_i) a(x) dx = a(x_i) + o(h_N), \tag{19}$$

$$\begin{aligned}
& \int K_N(y, y_i) K_N(y, y_j) a(y) dy \\
&= \int K(u) \frac{1}{h_N^d} K\left(u + \frac{y_i - y_j}{h_N}\right) a(y_i + u h_N) du, \\
&= \begin{cases} o(h_N) & \text{if } y_i \neq y_j \\ \frac{1}{h_N^d} [a(y_i) \int K^2(u) du + o(h_N)] & \text{if } y_i = y_j, \end{cases} \tag{20}
\end{aligned}$$

whenever  $K$  tends to zero at infinity at a sufficient rate (see Assumption A.8).

In the application on time series, we have:  $y_i = x_{i-1}$ . From the decomposition (17) and the equivalence (20), we deduce that the expectations:

$$EH_N(Z_i, Z_{i+k}) = o(h_N),$$

if  $|k| \geq 2$ . Moreover we have:

$$\begin{aligned}
EH_N(Z_i, Z_{i+1}) &= - \int K^2(u) du \int f(x_{i+1} | x_i) f(x_i | x_{i-1}) (\varphi_1(x_{i-1}) - \lambda_1 \varphi_1(x_i)) \\
&\quad \times (\varphi_1(x_{i+1}) - \lambda_1 \varphi_1(x_i)) dx_{i-1} dx_i dx_{i+1} + o(h_N), \\
&= o(h_N),
\end{aligned}$$

since  $\int (\varphi_1(x_{i+1}) - \lambda_1 \varphi_1(x_i)) f(x_{i+1} | x_i) dx_{i+1} = 0$ . Then we compute:

$$\begin{aligned}
EH_N(Z_i, Z_i) &= \frac{1}{h_N^{d/2}} \int K^2(u) du \left\{ \int \int (\varphi_1(x_i) - \lambda_1 \varphi_1(x_{i-1}))^2 f(x_i | x_{i-1}) dx_i dx_{i-1} \right. \\
&\quad \left. + \int \int (\varphi_1(x_{i-1}) - \lambda_1 \varphi_1(x_i))^2 f(x_{i-1} | x_i) dx_{i-1} dx_i \right\} \\
&= \frac{2}{h_N^{d/2}} \int K^2(u) du \int \text{Var} [\varphi_1(X_{i+1}) | X_i = x_i] dx_i,
\end{aligned}$$

due to the reversibility hypothesis. We get:

$$EI_N = \frac{2}{N h_N^d} \int K^2(u) du \int \text{Var} [\varphi_1(X_{i+1}) | X_i = x_i] dx_i.$$

## F Variance Matrices

We get in Tables *F.1* and *F.2* the constrained and unconstrained estimated canonical correlations asymptotic variances

[Insert here Table F.1: Estimated canonical correlations and standard error]

[Insert here Table F.2: Estimated canonical correlations and standard error]

To construct a procedure for testing the reversibility hypothesis from Theorem 3.3, we get in Table *F.3* the covariance matrix of  $\hat{\lambda}_i - \hat{\lambda}_i^R$ ,  $i = 1, \dots, 6$ .

[Insert here Table F.3: Estimated covariance of difference of canonical correlations]

**Acknowledge:**



- Darolles, S., Florens, J.P. and E. Renault, 2001, *Nonparametric Intrusmental Regression*, Discussion Paper CREST.
- Darolles, S. and C. Gouriéroux, 2001, *Truncated Dynamics and Estimation of Diffusion Equations*, Journal of Econometrics, **102**, 1-22.
- Darolles, S., Gouriéroux, C. and G. Le Fol, 2000, *Intraday Transaction Price Dynamics*, Annales d'Economie et Statistique, 60, 208-238.
- Darolles, S. and J.P. Laurent, 2000, *Approximating Payoffs and Pricing Formulas*, Journal of Economic Dynamics and Control, **24**, 1721-1746.
- Dauxois, J. and G. Nkiet, 1998, *Nonlinear Canonical Analysis and Independence Tests*, Annals of Statistics, **26**, 4, 1254-1278.
- Dauxois, J. and A. Pousse, 1975, *Une extension de l'analyse canonique. Quelques applications*, Ann. Inst. H. Poincaré, **11**, 355-379.
- Ding, Z. and C. Granger (1996), *Modelling Volatility Persistence of Speculative Returns: A New Approach*, Journal of Econometrics, **73**, 185-216.
- Dunford, N. and J. Schwartz, 1963, *Linear Operators 2*, Wiley, New York.
- Florens, J.P., Renault, E. and N. Touzi, 1998, *Testing for Embeddability by Stationary Reversible Continuous-Time Markov Processes*, Econometric Theory, **14**, 744-769.
- Gouriéroux, C. and J. Jasiak, 1998, *Nonlinear Autocorrelograms with Application to Intertrade Durations*, forthcoming Journal of Time Series Analysis.
- Gouriéroux, C. and J. Jasiak, 1999, *Nonlinear Persistence and Copersistence*, Working Paper CEPREMAP 9920.
- Gouriéroux, C. and J. Jasiak, 2001, *State-Space Models with Finite Dimensional Dependence*, Journal of Time Series Analysis, **22**, 665-678.

- Hall, P. 1984a, *Central Limit Theorem for Integrated Square Error Properties of Multivariate Nonparametric Density Estimators*, Journal of Multivariate Analysis, **14**, 1-16.
- Hall, P. 1984b, *Integrated Square Error Properties of Kernel Estimators of Regression Functions*, Annals of Statistics, **12**, 241-260.
- Hannan, E. 1961, *The General Theory of Canonical Correlation and its Relation to Functional Analysis*, J. Austral. Math. Soc., **2**, 229-242.
- Hansen, L. and J. Scheinkman, 1995, *Back to the Future: Generating Moment Implications for Continuous Time Markov Processes*, Econometrica **73**, 767-807.
- Hansen, L., Scheinkman, J., and N., Touzi, 1998, *Spectral Methods for Identifying Scalar Diffusions*, Journal of Econometrics, **86**, 1-32.
- Hinich, M. 1982, *Testing For Gaussianity and Linearity of a Stationary Time Series*, Journal of Time Series Analysis, **3**, 168-176.
- Hotelling, H. 1936, *Relations Between Two Sets of Variables*, Biometrika, **28**, 321-377.
- Jenrich, R. 1969, *Asymptotic Properties of Nonlinear Least Squares Estimators*, The Annals of Mathematical Statistics, **40**, 633-643.
- Kessler, M. and M. Sorensen, 1996, *Estimating Equations Based on Eigenfunctions for Discretely Observed Diffusion Process*, manuscript.
- Lancaster, H. 1968, *The Structure of Bivariate Distributions*, Ann. Math. Statist., **29**, 719-736.
- Meloche, J. 1990, *Asymptotic Behavior of the Mean Integrated Squared Error of Kernel Density Estimators for Dependent Observations*, Canadian Journal of Statistics, **18**, 205-224.

Nadaraya, E. 1983, *A Limit Distribution of the Square Error Deviation of Nonparametric Estimators of the Regression Functions*, Z. Wahrsch. Verw. Gebiete, **64**, 37-48.

Pagan A. and A. Ullah, 1999, *Nonparametric Econometrics*, Cambridge University Press.

Rosenblatt, M. 1956, *Remarks on Some Nonparametric Estimates of a Density Function*, Annals of Mathematical Statistics, **27**, 832-837.

Rosenblatt, M. 1975, *A Quadratic Measure of Deviation of Two Dimensional Density Estimates and a Test of Independence*, Annals of Statistics, **3**, 1-14.

Roussas, G. 1988, *Nonparametric Estimation in Mixing Sequences of Random Variables*, Journal of Statistical Planning and Inference, **18**, 135-149.

Tenreiro, C. 1995, *Théorèmes limites pour les erreurs quadratiques intégrées des estimateurs à noyau de la densité et de la régression sous des conditions de dépendance*, C.R. Acad. Sci. Paris, Ser. I, **320**, 1535-1538.

Tenreiro, C. 1997, *Loi asymptotique des erreurs quadratiques intégrées des estimateurs à noyau de la densité et de la régression sous des conditions de dépendance*, Portugalia Mathematica, **53**, 187-213.

Serfling, R.J. 1980, *Approximation Theorems of Mathematical Statistics*, Wiley, New-York.

Silverman, B. 1986, *Density Estimation for Statistics and Data Analysis*, Monographs on Statistics and Applied Probability, Chapman and Hall.

Canonical correlations $\lambda_i$			
Order	True	Unconstrained	Constrained
1	0.6065	0.6146	0.6139
2	0.1353	0.1795	0.1717
3	0.0111	0.0897	0.0734
4	0.0003	0.0781	0.0682

Table 4.1: Estimated canonical correlations

Canonical correlations $\lambda_i$		
Order	Unconstrained	Constrained
1	0.3595	0.2569
2	0.1829	0.1459
3	0.1467	0.1260
4	0.030	0.1255

Table 4.2: Estimated canonical correlations

Constrained Estimation		
	$\hat{\lambda}_i^R$	$\hat{\sigma}_{\hat{\lambda}_i^R}$
1	0.25659518	0.048155669
2	0.14579950	0.030734183
3	0.12594681	0.015937796
4	0.12531014	0.032680527
5	0.00543565	0.025264084
6	0.00166780	0.007468832

Table F.1: Estimated canonical correlations and standard error

Unconstrained Estimation		
	$\hat{\lambda}_i$	$\hat{\sigma}_{\hat{\lambda}_i}$
1	0.35914099	0.031038080
2	0.18268443	0.033228646
3	0.14661676	0.092862605
4	0.03072694	0.067672767
5	0.00338121	0.029166852
6	0.00017670	0.0078029362

Table F.2: Estimated canonical correlations and standard error

Unconstrained Estimation		
	$\hat{\lambda}_i - \hat{\lambda}_i^R$	$\hat{\sigma}_{\hat{\lambda}_i - \hat{\lambda}_i^R}$
1	0.10254581	0.02038768
2	-0,03688493	0.01122321
3	-0,02066995	0.03007654
4	0,09458319	0.01987775
5	0,00205443	0.00616685
6	0,00149110	0.00250076

Table F.3: Estimated covariance of difference of canonical correlations



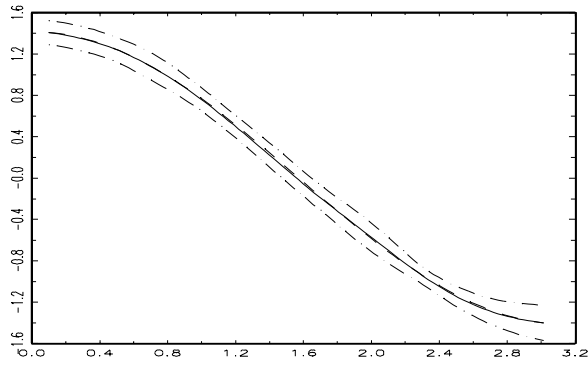


Figure 4.1: Constrained kernel estimator of the first canonical variate

Darolles Florens Gouriéroux

Kernel Based Nonlinear Canonical Correlation Analysis ...

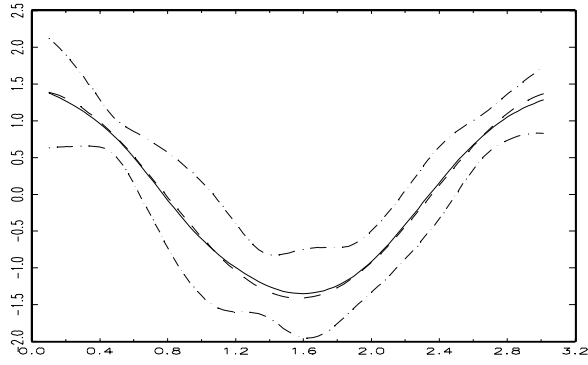


Figure 4.2: Constrained kernel estimator of the second canonical variate

Darolles Florens Gouriéroux

Kernel Based Nonlinear Canonical Correlation Analysis ...

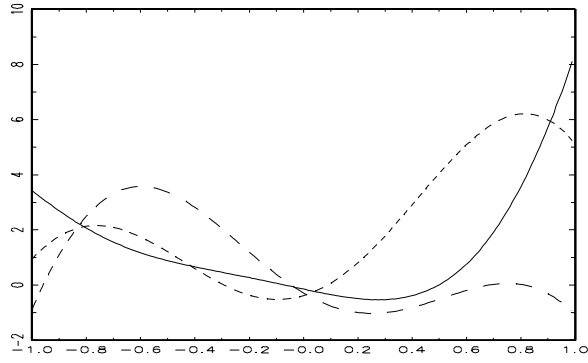


Figure 4.3: Estimated current canonical variates

Darolles Florens Gouriéroux

Kernel Based Nonlinear Canonical Correlation Analysis ...

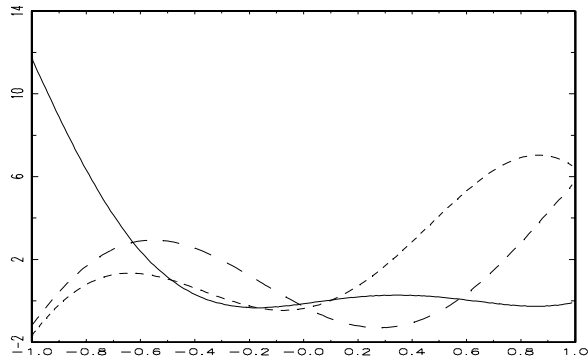


Figure 4.4: Estimated lagged canonical variates

Darolles Florens Gouriéroux

Kernel Based Nonlinear Canonical Correlation Analysis ...