Kernel Based Nonlinear Canonical Analysis

Serge Darolles[1], Jean-Pierre Florens[2], Christian Gouriéroux[3]

December 16, 1998

[1]CERMSEM-Université Paris I and CREST, Paris.
[2]GREMAQ and IDEI, Toulouse.
[3]CEPREMAP and CREST, Paris.

Abstract

We consider a kernel based approach to nonlinear canonical correlation analysis and its implementation for time series. We deduce various diagnostics for reversible processes and gaussian processes. The method is first applied to a simulated series satisfying a diffusion equation, allowing us to estimate nonparametrically the drift and volatility functions. The second application involves high frequency data on stock returns.

Keywords : Nonlinear Canonical Analysis, Gaussian Processes, Reversibility, Diffusion Equations, Kernel.

Classification JEL : C14, C15, C22.

Résumé

Nous considérons une analyse canonique non linéaire fondée sur un estimateur à noyau de la densité et sa mise en oeuvre sur séries temporelles. Nous en déduisons divers diagnostics pour les hypothèses de processus réversibles et gaussiens. Cette approche est ensuite appliquée à des données simulées selon une équation de diffusion, ce qui permet d'estimer de façon non paramétrique les fonctions de translation et de volatilité, et sur des séries haute fréquence de rendements.

Mots clés : Analyse Canonique non Linéaire, Processus Gaussiens, Réversibilité, Equation de Diffusion, Noyau.

Classification JEL : C14, C15, C22.

1 Introduction

Canonical correlation analysis has been introduced by [29, Hotelling (1936)], and is in general applied to linear transformations of either vectors (see [33, Lawley-Maxwell (1971)], [50, Tuckey (1977)], [34, Muirhead (1982)]) or individual histories (see [**?**, Dauxois-Pousse (1975)], [38, Rice-Silverman (1991)], [46, Silverman (1996)]). In this paper, we consider nonlinear canonical analysis which determines the most correlated nonlinear transformations of two vectors of interest (see [49, Tsai-Sen (1990)], [13, Dauxois-Nkiet (1998)]). In section 2, we first recall the principle of canonical analysis in Hilbertian framework and discuss its implementation to time series. In particular, we find a nonlinear factor representation for Markov processes and characterize the gaussian processes in terms of their nonlinear canonical correlations and canonical covariates. Statistical inference is studied in section 3. We propose a new approach called kernel canonical analysis, where the unknown joint density function is replaced by a kernel estimator. We derive the asymptotic properties of the corresponding estimators of the canonical correlations and covariates. We also examine the estimation under the reversibility constraint and propose a test of the reversibility hypothesis. The method is applied in the last section to a simulated series satisfying a diffusion equation allowing us to estimate nonparametrically the drift and volatility functions, and to high frequency data on stock returns.

2 Nonlinear Canonical Analysis

In this section, we first recall the principle of canonical analysis in Hilbertian framework. This allows us to extend the classical idea of principal component analysis to the nonlinear setting. Next, we detail the results for stationary time series and gaussian processes.

2.1 The principle

Let us consider two square integrable random vectors X and Y defined on a probability space $(\Omega, \mathcal{A}, P)$. The space of square integrable random variables (resp. functions of X, Y) is denoted by L^2 (resp. $L^2(X), L^2(Y)$), and the associated inner product by $\langle ., . \rangle$. The problem of canonical analysis is to reveal $p-$dimensional subspaces $I_p^*(X)$, $I_p^*(Y)$ of $L^2(X)$ and $L^2(Y)$ respectively which maximize the minimal inner product between the vectors of $I_p^*(X)$ and $I_p^*(Y)$:

$$
\begin{aligned}
& \left[I_p^*(X), I_p^*(Y)\right] \\
= & \underset{\substack{I_p(X) \subset L^2(X) \\ I_p(Y) \subset L^2(Y)}}{\operatorname{argmax}} \; \min_{\substack{\varphi(X) \in I_p(X) \\ \psi(Y) \in I_p(Y)}} \frac{\langle \varphi(X), \psi(Y) \rangle}{\|\varphi(X)\| \, \|\psi(Y)\|} \qquad (2.1) \\
= & \underset{\substack{I_p(X) \subset L^2(X) \\ I_p(Y) \subset L^2(Y)}}{\operatorname{argmax}} \; \min_{\substack{\varphi(X) \in I_p(X) \\ \psi(Y) \in I_p(Y)}} \frac{E\left[\varphi(X)\psi(Y)\right]}{\left(E\varphi^2(X)\right)^{\frac{1}{2}} \left(E\psi^2(Y)\right)^{\frac{1}{2}}}.
\end{aligned}
$$

In particular, when $p = 1$, we simply maximize $E\left[\varphi(X)\psi(Y)\right] / \left(E\varphi^2(X)\right)^{\frac{1}{2}} \left(E\psi^2(Y)\right)^{\frac{1}{2}}$ with respect to the functions φ and ψ, or alternatively, search for the solution of the optimization problem:

$$
\max_{\varphi,\psi} E\left[\varphi(X)\psi(Y)\right], \qquad (2.2)
$$

$$
s.t.\ E\varphi^2(X) = E\psi^2(Y) = 1.
$$

In the general case, the solution to problem (2.1) is based on the spectral decompositions of appropriate conditional expectation operators which are defined below (see e.g. [18, Dunford-Schwartz (1963), chapter XI]). We introduce:

1. the conditional expectation operator T for the mapping from $L^2(X)$ to $L^2(Y)$:
$$
\varphi(X) \to T\varphi(Y) = E\left[\varphi(X) \mid Y\right]; \qquad (2.3)
$$

2. the conditional expectation operator T^* for the mapping from $L^2(Y)$ to $L^2(X)$:
$$
\psi(Y) \to T^*\psi(X) = E\left[\psi(Y) \mid X\right]. \qquad (2.4)
$$

From the projection interpretation of conditional expectation operators, these operators are bounded with a norm equal to one and we get:

$$
E\left[\varphi(X)\psi(Y)\right] = E\left[\varphi(X) E\left[\psi(Y) \mid X\right]\right] = E\left[E\left[\varphi(X) \mid Y\right]\psi(Y)\right],
$$

or equivalently:

$$
\langle \varphi(X), \psi(Y) \rangle = \langle \varphi(X), T^*\psi(X) \rangle = \langle T\varphi(Y), \psi(Y) \rangle. \qquad (2.5)
$$

Therefore, T^* is the adjoint operator of T, which justifies ex-post the notation *. The main result is easy to present under the assumption that TT^* and T^*T have the same spectrum with isolated eigenvalues (see [10, Darolles-Florens-Renault (1998)] for a discussion of this assumption).

Assumption A.1 : *The operators TT^* and T^*T have a discrete spectrum:*

$$1 = \lambda_0^2 > \lambda_1^2 > ... > \lambda_k^2 > \lambda_{k+1}^2 > ... > 0.$$

In the sequel, λ_k denotes the (positive) square root of λ_k^2. The condition of strictly positive eigenvalues eliminates the standard case arising in finite dimensional linear canonical analysis, where X and Y have different dimensions. The condition of isolated eigenvalues is related to the general decomposition of the spectrum into a continuous spectrum, a residual spectrum and a point spectrum (see [54, Yoshino (1993), chapter 2], and to the compactness of the operators. Finally note that $\lambda_0 = 1$ is an eigenvalue associated with the constant eigenfunction and is the largest one since $\|T\| = 1$. Hence, we have the two theorems below (see e.g. [35, Naylor-Snell (1982)], [8, Buja (1990)]):

Theorem 2.1 : *Under assumption A.1, there exist two joint hilbertian basis of eigenfunctions $\varphi_i(X)$, $i \geq 0$, and $\psi_j(Y)$, $j \geq 0$ of T^*T and TT^*, respectively, satisfying:*

i) $T^*T\varphi_i(X) = \lambda_i^2 \varphi_i(X)\,,\ i \geq 0;$

ii) $TT^*\psi_i(Y) = \lambda_i^2 \psi_i(Y)\,,\ i \geq 0;$

iii) $\varphi_0(X) = 1$, $\psi_0(Y) = 1;$

iv) $\left\langle \varphi_i(X), \varphi_j(X) \right\rangle = \delta_{ij}$, $i, j \geq 0$, where δ_{ij} is the Kronecker symbol;

v) $\left\langle \psi_i(Y), \psi_j(Y) \right\rangle = \delta_{ij}$, $i, j \geq 0;$

vi) $\left\langle \varphi_i(X), \psi_j(Y) \right\rangle = \lambda_i \delta_{ij}$, $i, j \geq 0;$

vii) $T\varphi_i(Y) = E[\varphi_i(X) \mid Y] = \lambda_i \psi_i(Y)\,,\ i \geq 0;$

viii) $T^*\psi_i(X) = E[\psi_i(Y) \mid X] = \lambda_i \varphi_i(X)\,,\ i \geq 0.$

In particular, the orthogonality conditions *iv*) and *v*) applied to φ_0 and ψ_0 imply that the other eigenfunctions are zero-mean:

$$E\varphi_i(X) = E\psi_i(Y) = 0,\ i \geq 1. \tag{2.6}$$

The two previous hilbertian basis can be used to decompose the vectors of $L^2(X)$ and $L^2(Y)$ (see e.g. [54, Yoshino (1993)]). We get the Fourier expansions:

$$\begin{aligned}\varphi(X) &= \sum_{i=0}^{\infty} \langle \varphi(X), \varphi_i(X) \rangle \varphi_i(X) \\ &= E\varphi(X) + \sum_{i=1}^{\infty} \langle \varphi(X), \varphi_i(X) \rangle \varphi_i(X), \qquad (2.7)\end{aligned}$$

and

$$\psi(Y) = E\psi(Y) + \sum_{i=1}^{\infty} \langle \psi(Y), \psi_i(Y) \rangle \psi_i(Y). \tag{2.8}$$

We deduce the variance and the covariance formulas from the Parseval's identities:

$$\begin{aligned} V\varphi(X) &= \sum_{i=1}^{\infty} \left[cov\left(\varphi(X), \varphi_i(X)\right)\right]^2, \\ V\psi(Y) &= \sum_{i=1}^{\infty} \left[cov\left(\psi(Y), \psi_i(Y)\right)\right]^2, \\ cov\left[\varphi(X), \psi(Y)\right] &= \sum_{i=1}^{\infty} \lambda_i cov\left(\varphi(X), \varphi_i(X)\right) cov\left(\psi(Y), \psi_i(Y)\right). \end{aligned}$$

We have now to understand how the canonical analysis is related to the decomposition given in proposition 2.1.

Theorem 2.2 : *Under assumption A.1, there exists a unique pair of subspaces $I_p^*(X)$ and $I_p^*(Y)$ solving the optimization problem* (2.1).

$I_p^*(X)$ is the subspace generated by $\varphi_i(X)$, $i = 0, ..., p-1$,
$I_p^*(Y)$ is the subspace generated by $\psi_i(Y)$, $i = 0, ..., p-1$.

Proof. See Appendix A. ■

The successive pairs of canonical variates (φ_i, ψ_i), i varying, are defined up to a change of sign. The eigenfunctions φ_i, ψ_i are called canonical variates, whereas the square roots of the eigenvalues λ_i are the canonical correlations. In practice, the canonical analysis is usually performed on zero-mean variables. The optimization problem (2.1) becomes:

$$\left[I_p^*(X), I_p^*(Y)\right] = \underset{\substack{I_p(X) \subset L_0^2(X) \\ I_p(Y) \subset L_0^2(Y)}}{\operatorname{argmax}} \min_{\substack{\varphi(X) \in I_p(X) \\ \psi(Y) \in I_p(Y)}} \frac{Cov\left[\varphi(X), \psi(Y)\right]}{\sqrt{V\varphi(X)}\sqrt{V\psi(Y)}}, \tag{2.9}$$

where $L_0^2(X)$ (resp. $L_0^2(Y)$) is the subspace of $L^2(X)$ (resp. $L^2(Y)$) of zero-mean variables.

Corollary 2.1 : *Under assumption A.1, there exists a unique pair of subspaces $I_p^*(X)$ and $I_p^*(Y)$ solving the optimization problem* (2.9).

$I_p^*(X)$ is the subspace generated by $\varphi_i(X)$, $i = 1, ..., p$,
$I_p^*(Y)$ is the subspace generated by $\psi_i(Y)$, $i = 1, ..., p$.

For continuous variables, the previous results can also be written in terms of the joint density function $f(x,y)$ of (X,Y). For instance, the operators are defined by:

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

with

$$c(x,y) = \int \frac{f(y,z)\, f(x,z)}{f(x,.)\, f(.,z)} dz, \tag{2.10}$$

whereas the Parseval identity becomes:

$$\frac{f(x,y)}{f(x,.)\, f(.,y)} = 1 + \sum_{i=1}^{\infty} \lambda_i \varphi_i(x)\, \psi_i(y), \tag{2.11}$$

which is a decomposition of the functional measure of dependence between X and Y (see [4, Barrett-Lampard (1955)], [32, Lancaster (1958)]). This decomposition may be the basis of independence tests (see [13, Dauxois-Nkiet (1998)]).

2.2 Application to time series

The interest in nonlinear canonical analysis of time series has recently increased with the availability of large datasets in finance, especially the high frequency datasets provided by the electronic trading systems. Methods related to the canonical analysis have already been introduced for either continuous time diffusion models (see [27, Hansen-Scheinkman-Touzi (1998)], [15, Demoura (1995)], [31, Kessler-Sorensen (1996)], [11, Darolles-Gouriéroux (1997)], [10, Darolles-Florens-Renault (1998)], [19, Florens-Renault-Touzi (1998)], [9, Chen-Hansen-Scheinkman (1998)]) or for discrete time models (see [37, Ray-Tsay (1998)], [21, Gouriéroux-Jasiak (1998)], [12, Darolles-Gouriéroux-Le Fol (1998)]). This technique has several advantages for investigating the nonlinear dynamics, especially for examining in detail the price-volume relationship (see [20, Ghysels-Gouriéroux-Jasiak (1998)]) or the dynamics of extreme returns. We first consider Markov processes and next discuss the general case. The time series of interest, denoted by $(X_t,\ t = 0,1,...)$, may be multidimensional of dimension d, and is assumed to be stationary.

2.2.1 Markov process of order one

We can apply the canonical analysis to a current and a lagged values of the time series: $X = X_t$, $Y = X_{t-1}$. In this framework, the canonical decomposition characterizes the whole distribution of the process. Therefore, a number of constraints imposed on this distribution can be easily analysed in the framework of the canonical decomposition. For simplicity, we call $\varphi_i(X_t)$, $\forall i \geq 1$, the current canonical variates, and $\psi_i(X_{t-1})$, $\forall i \geq 1$, the lagged canonical variates. We discuss below the reversibility property and introduce a factor decomposition for reversible and irrreversible processes. Let us recall that a process is reversible if and only if its distributional properties are identical in the initial and in reversed time. We deduce the following characterization of the reversibility property.

Theorem 2.3 *: Under assumption A.1, the stationary Markov process is reversible if and only if:*

$$\varphi_i = \pm\, \psi_i, \ \forall i \geq 1.$$

Proof. The necessary part is obvious since the computation is the same in the initial and reversed time, and the canonical variates are defined up to a change of sign. Conversely, if $\varphi_i = \pm\, \psi_i$, $\forall i \geq 1$, we deduce from (2.11) that:

$$\begin{aligned} f(x,y) &= f(x,.)\,f(.,y)\left(1+\sum_{i=1}^{\infty}\lambda_i\varphi_i(x)\,\psi_i(y)\right), \\ &= f(x,.)\,f(.,y)\left(1\pm\sum_{i=1}^{\infty}\lambda_i\varphi_i(x)\,\varphi_i(y)\right). \end{aligned}$$

This expression is symmetric in x and y, since the marginal distribution of $X = X_t$, $Y = X_{t-1}$ are identical. ■

***i)* Reversible process**

A reversible Markov process admits a factor autoregressive representation.

Proposition 2.1 *: Under assumption A.1, a reversible Markov process can be written as:*

$$X_t = a_0 + \sum_{j=1}^{\infty} a_j Z_{j,t},$$

where the Z_j's processes satisfy:

$$Z_{j,t} = \lambda_j Z_{j,t-1} + u_{j,t},$$

with $E\left[u_{j,t} \mid \underline{X_{t-1}}\right] = 0$, *and* $cov\left[u_{j,t}, u_{l,t}\right] = \left(1-\lambda_j^2\right)\delta_{jl}$.

Proof. We have to select $Z_{j,t} = \varphi_j(X_t)$. The various conditions are the consequences of the Fourier decomposition of X_t and also follow from theorem 2.1. ■

The existence of a factor decomposition with AR(1) components has already been noted in the case of transformed gaussian processes by [23, Granger-Newbold (1976)].

Remark 2.1 *: Let us consider the limiting case corresponding to:* $\lambda_1 > 0$, $\lambda_j = 0$, $\forall j > 1$. *The previous factor decomposition becomes:*

$$X_t = a_0 + a_1 Z_{1,t} + v_{1,t},$$

where the error term $v_{1,t}$ *is a martingale difference sequence and the* Z_1 *process satisfies the autoregressive relation:*

$$Z_{1,t} = \lambda_1 Z_{1,t-1} + u_{1,t}.$$

Remark 2.2 *: In general, the error terms* $u_{j,t}$ *are conditionally heteroscedastic. More precisely, let us introduce the Fourier decomposition of the squared eigenfunctions:*

$$\varphi_j^2(X_t) = Z_{j,t}^2 = 1 + \sum_{i=1}^{\infty} c_{j,i} Z_{i,t}.$$

We get:

$$\begin{aligned} E\left[u_{j,t}^2 \mid \underline{X_{t-1}}\right] &= V\left[Z_{j,t} \mid \underline{X_{t-1}}\right] \\ &= E\left[Z_{j,t}^2 \mid \underline{X_{t-1}}\right] - E\left[Z_{j,t} \mid \underline{X_{t-1}}\right]^2 \\ &= 1 + \sum_{i=1}^{\infty} \lambda_i c_{j,i} Z_{i,t-1} - \lambda_j^2 Z_{j,t-1}^2 \\ &= 1 - \lambda_j^2 + \sum_{i=1}^{\infty} c_{j,i}\left(\lambda_i - \lambda_j^2\right) Z_{i,t-1}. \end{aligned}$$

Hence, the error terms $u_{j,t}$ *are conditionally homoscedastic if and only if* $c_{j,i}\left(\lambda_i - \lambda_j^2\right) = 0$, $\forall i \geq 1$, *which is satisfied if there exists* i_0, $i_0 \geq 1$, *with* $c_{j,i} = 0$, $\forall i \neq i_0$ *and* $\lambda_{i_0} = \lambda_{i_0}^2$.

Remark 2.3 *: Note that the factor decomposition is valid for any transformation of the process:*

$$\varphi(X_t) = b_0 + \sum_{i=1}^{\infty} b_i Z_{i,t} \qquad \textit{(say)}.$$

We directly deduce the predictions of the transformed variable at various horizons using the decomposition:

$$E\left[\varphi(X_{t+h}) \mid \underline{X_t}\right] = b_0 + \sum_{i=1}^{\infty} \lambda_i^h b_i Z_{i,t}.$$

Remark 2.4 : *In particular, the previous result is valid when the process* $(X_t,\ t=0,1,...)$ *consists of discrete time observations on a diffusion process:*

$$dX_t = \mu(X_t)\, dt + \sigma(X_t)\, dW_t, \tag{2.12}$$

where $(W_t,\ t \geq 0)$ *is a multidimensional standard brownian motion. In this framework, we can introduce the infinitesimal generator defined by:*

$$\mathcal{A}\varphi(X_t) = \lim_{h\to 0} \left[\frac{E[\varphi(X_{t+h}) \mid X_t] - \varphi(X_t)}{h} \right]. \tag{2.13}$$

It is related to the conditional expectation operator at horizon one by:

$$\mathcal{A} = \lim_{h\to 0} \frac{T^h - Id}{h}, \tag{2.14}$$

and takes the form of a differential operator:

$$\mathcal{A}\varphi(x) = \mu(x) \frac{\partial\varphi(x)}{\partial x} + \frac{1}{2}\sigma(x)' \frac{\partial^2\varphi(x)}{\partial x \partial x'}\sigma(x). \tag{2.15}$$

From the nonlinear canonical analysis of the discrete time series $(X_t,$ $t=0,1,...)$*, we can deduce the conditional expectation* T*, the infinitesimal generator* $\mathcal{A}$ *by* (2.14)*, and the drift and volatility functions by* (2.15)*. This technique is especially simple in the unidimensional case, since the continuous time process is reversible (see [26, Hansen-Scheinkman (1995)]). Let us take a closer look at this case. If* $\lambda_i,\ \varphi_i = \pm\ \psi_i,\ i \geq 0$ *denote the canonical decomposition corresponding to the discrete time process, the infinitesimal generator admits eigenvalues* $\ln\lambda_i = \rho_i$ *(say) with corresponding eigenfunctions* φ_i*. Then, the drift and volatility functions can be identified by considering the two first canonical variates since we get the bivariate system:*

$$\begin{aligned}
\mathcal{A}\varphi_1(x) &= \mu(x)\frac{d\varphi_1(x)}{dx} + \frac{1}{2}\sigma^2(x)\frac{d^2\varphi_1(x)}{dx^2} = \ln\lambda_1\ \varphi_1(x), \\
\mathcal{A}\varphi_2(x) &= \mu(x)\frac{d\varphi_2(x)}{dx} + \frac{1}{2}\sigma^2(x)\frac{d^2\varphi_2(x)}{dx^2} = \ln\lambda_2\ \varphi_2(x),
\end{aligned}$$

which may be solved with respect to μ *and* σ^2 *(see [15, Demoura (1993)]).*

ii) Irreversible process

In the irreversible case, a Markov process also admits a factor decomposition, but the factor dynamics is more complicated.

Proposition 2.2 : *Under assumption A.1, a Markov process can be written as:*

$$X_t = a_0 + \sum_{j=1}^{\infty} a_j Z_{j,t},$$

where the Z_j's processes satisfy:

$$Z_{j,t} = \lambda_j \sum_{k=1}^{\infty} b_{jk} Z_{k,t-1} + u_{j,t},$$

with $E\left[u_{j,t} \mid \underline{X_{t-1}}\right] = 0$, *and* $cov\left[u_{j,t}, u_{l,t}\right] = \left(1 - \lambda_j^2\right) \delta_{jl}$.

Proof. We have to select $Z_{j,t} = \varphi_j (X_t)$ and $\tilde{Z}_{j,t} = \psi_j (X_t)$. The various conditions are consequences of the Fourier decomposition of X_t and $\tilde{Z}_{j,t} = \psi_j (X_t)$. From theorem 2.1, we obtain the following dynamics:

$$Z_{j,t} = \lambda_j \tilde{Z}_{j,t-1} + u_{j,t},$$

for the Z_j's processes appearing in the decomposition of X_t. Using the decomposition formula for $\tilde{Z}_{j,t} = \psi_j (X_t)$, we get:

$$\tilde{Z}_{j,t} = \sum_{k=1}^{\infty} b_{jk} Z_{k,t},$$

since the $\tilde{Z}_{j,t}$ have zero mean. Finally, we obtain the factor dynamics equation. ∎

Note that, in this case, the factor dynamics is an infinite autoregressive process of order one.

Remark 2.5 *: Let us now consider the limiting case corresponding to:* $\lambda_1 > 0$, $\lambda_j = 0$, $\forall j > 1$. *Therefore, the factor decomposition becomes:*

$$X_t = a_0 + a_1 Z_{1,t} + v_{1,t},$$

$$Z_{1,t} = \lambda_1 \tilde{Z}_{1,t-1} + u_{1,t},$$

and

$$\tilde{Z}_{1,t} = b_{11} Z_{1,t} + w_{1,t},$$

where the error terms $v_{1,t}$ and $w_{1,t}$ are martingale difference sequences obtained by aggregating the effects of the $Z_{j,t}$ variables for $j \geq 2$. Therefore, the dynamics of the Z_1 process satisfies the ARMA(1,1)-type relation:

$$Z_{1,t} = \lambda_1 b_{11} Z_{1,t-1} + u_{1,t} + \lambda_1 w_{1,t-1}.$$

2.2.2 Nonlinear autocorrelogram

In the general case, we can apply nonlinear canonical analysis to construct nonlinear autocorrelograms as suggested in [21, Gouriéroux-Jasiak (1998)]. By choosing $X = X_t$, $Y = X_{t-h}$ with h varying, we construct a bivariate sequence of spectral decompositions: $\lambda_{i,h}$, $\varphi_{i,h}$, $\psi_{i,h}$, $i, h \geq 1$, and examine how the eigenvalues and eigenfunctions depend on the lag. The comparison of the canonical variates at different lags is an extension of the comparison of nonlinear expectations $E\left[Y_t \mid Y_{t-h}\right]$, h varying, variances $V\left[Y_t \mid Y_{t-h}\right]$, h varying, or bivariate histograms $f\left(Y_t, Y_{t-h}\right)$, h varying, proposed in [48, Tong (1993), p. 364-374].

The factor decompositions introduced above can be used to compare the linear and nonlinear predictions of a reversible markov process (see [17, Donelson-Matz (1972)], [23, Granger-Newbold (1976)] for nonlinear transformations of gaussian processes). Indeed, let us consider the factor decomposition of the process:

$$X_t = \sum_{j=0}^{\infty} a_j \varphi_j\left(X_t\right).$$

The nonlinear prediction is: $E\left[X_t \mid X_{t-1}\right] = \sum_{j=0}^{\infty} \lambda_j a_j \varphi_j\left(X_{t-1}\right)$, whereas the quadratic prediction error is: $\gamma_{NL} = \sum_{j=1}^{\infty}\left(1-\lambda_j^2\right) a_j^2$. The linear prediction is:

$${}_{t-1}\hat{X}_t = a_0 + \varrho\left(1\right)\left(X_{t-1} - a_0\right),$$

where $\varrho\left(1\right) = \sum_{j=1}^{\infty} \lambda_j a_j^2 / \sum_{j=1}^{\infty} a_j^2$, and the quadratic linear prediction error is:

$$\gamma_L = \sum_{j=1}^{\infty} a_j^2 \left(1 - \frac{\left(\sum_{j=1}^{\infty} \lambda_j a_j^2\right)^2}{\left(\sum_{j=1}^{\infty} a_j^2\right)^2}\right).$$

We directly note that: $\gamma_L - \gamma_{NL} = \sum_{j=1}^{\infty} a_j^2 \; V_a\left(\lambda\right) > 0$, where $V_a\left(\lambda\right)$ is the variance of the canonical correlations λ_j computed with the weights a_j.

Finally, it is interesting to note that the nonlinear canonical analysis is also suitable for density forecasting (see [24, Granger-Pesaran (1996)], [16, Diebold-Gunther-Tay (1997)]), since in the reversible case, the canonical distribution of X_{t+h} given X_t is:

$$f\left(X_{t+h} \mid X_t\right) = f\left(X_{t+h}, .\right)\left(1 + \sum_{j=1}^{\infty} \lambda_j^h \varphi_j\left(X_{t+h}\right) \varphi_j\left(X_t\right)\right).$$

2.2.3 Gaussian process

The canonical variates have been fully described for gaussian vectors (see [4, Barrett-Lampard (1955)], [51, Wiener (1958), lecture 5], [53, Wong-Thomas (1962)], [1, Abramowitz-Stegun (1965), formula 26.3.29], [36, Neveu (1968)]). We summarize below the main results.

Proposition 2.3 *: If $(X, Y)'$ is a bidimensional gaussian vector with zero mean, and variance-covariance matrix:*

$$\begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix},$$

with $\rho \geq 0$, its canonical correlations are $\lambda_i = \rho^i$, $i \geq 1$, and the corresponding canonical variates are:

$$\varphi_i(x) = \psi_i(x) = \frac{1}{\sqrt{i!}} H_i\left(\frac{x}{\sigma}\right),$$

up to a joint change of sign, where the Hermite polynomials H_i's are defined by:

$$H_i(x) = \sum_{0 \leq m \leq \left[\frac{i}{2}\right]} \frac{i!}{(i-2m)!\ m!\ 2^m} (-1)^m x^{i-2m}.$$

The first Hermite polynomials are: $H_1(x) = -x$, $H_2(x) = x^2 - 1$, $H_3(x) = -x^3 + 3x$. For a negative autocorrelation, we have only to replace Y by $-Y$ to deduce that the canonical correlations are $\lambda_i = |\rho|^i$, whereas the canonical variates are: $\varphi_i(x) = \frac{1}{\sqrt{i!}} H_i\left(\frac{x}{\sigma}\right)$, $\psi_i(x) = \frac{1}{\sqrt{i!}} H_i\left(-\frac{x}{\sigma}\right)$, up to a joint change of sign.

The property above can be used to check if a given unidimensional process is gaussian. Indeed, if the process is gaussian with zero mean and autocorrelation function $\rho(h)$:

i) the canonical correlations are geometrically decreasing for any lag:

$$\forall h, \quad \lambda_{i,h} = (\lambda_{1,h})^i \left[= |\rho(h)|^i\right];$$

ii) the current and lagged canonical variates $\varphi_{i,h}$, $\psi_{i,h}$, are equal and independent of the lag, up to a change of sign of the argument if $\rho(h)$ is negative. Up to the changes of sign, they coincide with the Hermite polynomials.

The first condition has been proposed under an equivalent form as a test for gaussianity by [23, Granger-Newbold (1976)]. Indeed, we have the corollary below.

Corollary 2.2 *: If $(X_t,\ t = 0, 1, ...)$ is an unidimensional gaussian process with autocorrelation function $\rho(h)$, then the autocorrelation function of the process $(H_p(X_t),\ t = 0, 1, ...)$, $p \in \mathbf{N}^*$, is:*

$$\rho^{(p)}(h) = [\rho(h)]^p.$$

3 Statistical Inference

In practice, the distribution of the pair (X, Y) is not known and the theoritical canonical analysis described in section 2 cannot be performed. However, we can approximate the canonical decomposition if some observations (X_n, Y_n), $n = 1, ..., N$ 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 be in particular applied to a stationary time series (X_t, $t = 0, 1, ...$) observed until date T, with $X_n = X_t$, $Y_n = X_{t-h}$. Each vector is assumed to be of dimension d.

3.1 The estimators

A natural idea is to replace the initial optimization problem (2.1) by its empirical counterpart, i.e. to replace the inner product $\langle \varphi(X), \psi(Y) \rangle = E[\varphi(X)\psi(Y)]$ by the empirical cross moment $\frac{1}{N}\sum_{n=1}^{N} \varphi(X_n)\psi(Y_n)$. However, the empirical operators $\widehat{T^*T}$ and $\widehat{TT^*}$ do not converge to their theoritical counterpart. This convergence problem may be solved by replacing the operators by their restrictions to some finite dimensional subspaces of $L^2(X)$ and $L^2(Y)$ whose dimension increases with the number of observations (see [27, Hansen-Scheinkman-Touzi (1998)], [10, Darolles-Florens-Renault (1998)], [9, Chen-Hansen-Scheinkman (1998)]). In our approach, we consider instead a kernel based approximation of the density in the optimization problem yielding to operators with a finite spectrum whose dimension is equal to the number of observations. If the distribution of (X, Y) is continuous with a probability distribution function $f(x, y)$, the inner product becomes:

$$\langle \varphi(X), \psi(Y) \rangle = \int\int \varphi(x)\psi(y) f(x, y)\, dxdy. \tag{3.1}$$

Let us introduce two kernels K_1, K_2 defined on $\mathbf{R}^d$; the unknown density function can be approximated by (see [39, Rosenblatt (1956)]):

$$\hat{f}_N(x, y) = \frac{1}{N}\sum_{n=1}^{N} \frac{1}{h_{1N}^d h_{2N}^d} K_1\left(\frac{X_n - x}{h_{1N}}\right) K_2\left(\frac{Y_n - y}{h_{2N}}\right), \tag{3.2}$$

where h_{1N}, h_{2N} are the bandwidths associated with the two components. Then, we consider the canonical decomposition $\hat{\lambda}_{i,N}$, $\hat{\varphi}_{i,N}$, $\hat{\psi}_{i,N}$, $i \geq 0$, after replacing in the optimization problem (2.1) the initial inner product by:

$$\langle \varphi(X), \psi(Y) \rangle_N = \int\int \varphi(x)\psi(y)\hat{f}_N(x, y)\, dxdy. \tag{3.3}$$

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

Assumption A.3 : *The kernels K_1, K_2 are non negative, with unit mass.*

Finally, note that the initial and approximated optimization problems 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 $\hat{f}_N$. We can easily see that a function φ is square integrable in the approximated optimization problem of order N if and only if:

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

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:

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

The space $L_{K_2}^2(Y)$ is 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) will involve the respective tails of the kernels K_1, K_2 and the p.d.f. f. Intuitively, we have to select kernels with rather thin tails to be sure that $L_{K_1}^2(X)$ includes the theoretical canonical variates of interest (see Assumption A.15).

Using the definition of $\hat{f}_N$, we may also introduce the associated estimated conditional expectation operators:

$$\hat{T}_N \varphi(y) = \frac{\int \varphi(x) \hat{f}_N(x, y)\, dx}{\int \hat{f}_N(x, y)\, dx}, \tag{3.4}$$

$$\hat{T}_N^* \psi(x) = \frac{\int \psi(y) \hat{f}_N(x, y)\, dy}{\int \hat{f}_N(x, y)\, dy}, \tag{3.5}$$

which are the Nadarahya-Watson estimators of the corresponding regression functions.

3.2 Numerical implementation

There exist various numerical methods for deriving accurate approximations of the p first elements of the estimated canonical decomposition. We can for instance consider the restriction limiting the domain of the operator $\hat{T}_N$ to finite dimensional subspaces of $L^2(X)$ and $L^2(Y)$, with respective dimensions M_1 and M_2. This restricted operator may be written in a matrix form of size $M_1 \times M_2$, and the spectral decomposition of this matrix is obtained by a standard algorithm, such as the power method (see [52, Wilkinson-Reinsch (1971)], [22, Gourlay-Watson (1973)]), or the bi-iteration method due to Bauer and adapted to symmetric positive definite matrices by [43, Rutishauser (1969)]. It is important to note that the dimensions M_1 and M_2 can be chosen arbitrarily and are not related to the number of observations. This is the advantage of projecting on finite subspaces after the kernel estimation of the operators, compared to defining the estimators directly from projections ([27, Hansen-Scheinkman-Touzi (1998)], [10, Darolles-Florens-Renault (1998)], [9, Chen-Hansen-Scheinkman (1998)]).

Some of the numerical approaches can be directly applied to the estimated infinite dimensional operator itself. For instance, let us consider the determination of $\hat{\lambda}_{i,N}$, $\hat{\varphi}_{i,N}$, once $\hat{\lambda}_{j,N}$, $\hat{\psi}_{j,N}$, $j \leq i-1$ have been derived. We have to solve the problem:

$$\max_{\varphi} \left\langle \hat{T}_N \varphi(X), \hat{T}_N \varphi(X) \right\rangle_N, \tag{3.6}$$

where $\langle \varphi(X), \varphi(X) \rangle_N = 1$, φ belongs to the orthogonal subspace of $\Big(\varphi_{0,N}, \ldots, \varphi_{i-1,N}\Big)$. We can choose an initial function φ^1 (say) in this subspace, and then compute recursively:

$$\varphi^k = \hat{T}_N^* \hat{T}_N \varphi^{k-1}.$$

The desired eigenvalue is the limit of the Rayleigh quotients:

$$\hat{\lambda}_{i,N} = \lim_{k \to \infty} \frac{\left\langle \hat{T}_N \varphi^k(X), \hat{T}_N \varphi^k(X) \right\rangle_N}{\langle \varphi^k(X), \varphi^k(X) \rangle_N},$$

whereas

$$\hat{\varphi}_{i,N} = \lim_{k \to \infty} \varphi^k.$$

3.3 Asymptotic properties

The asymptotic properties of the estimated canonical decomposition have been studied in the linear case for vectors ([3, Anderson (1963)]) and curves ([14, Dauxois-Pousse-Romain (1982)], [38, Rice-Silverman (1991)], [46, Silverman (1996)]). In our nonlinear framework, they have to be deduced from the asymptotic properties of the kernel density estimator.

3.3.1 Properties of the kernel density estimator

These properties are standard. We first describe the uniform strong consistency properties and then discuss a functional central limit theorem. The proof of the two theorems below are deduced from results by [41, Roussas (1988)], [6, Bosq (1996)] and adapted to the case of a compact set of values. We consider the following assumptions.

Assumption A.4 : *The variables* X *and* Y *take values in the same compact set* $\mathcal{X} \subset \mathbf{R}^d$, $\mathcal{X} = [0,1]^d$ *say.*

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.

Assumption A.5 : *The probability density function* f *is continuous on* $\mathcal{X}^2 = [0,1]^{2d}$.

Assumption A.6 : *The strictly stationary sequence* (X_n, Y_n) *is geometrically strong mixing, i.e. with* α*-mixing coefficients*[1] *such that:*

$$\alpha_k \leq c_0 \rho^k,$$

for some fixed $c_0 > 0$ *and* $0 \leq \rho < 1$.

Assumption A.7 : *The kernels* K_i ,$i = 1, 2$, *are:*

i) *bounded,*

ii) *symmetric,*

iii) *of order*[2] 2,

iv) *Lipschitzian,*

v) *and satisfy* $\lim_{\|u\| \to \infty} \|u\|^d K_i(u) = 0$, $i = 1, 2$.

The following assumption concerns the choice of the bandwidths h_{iN}, $i = 1, 2$.

Assumption A.8 : *As* $N \to \infty$, $h_{iN} \to 0$, $\frac{N h_{iN}^d}{(\log N)^2} \to +\infty$, $i = 1, 2$.

Hence, from [41, Roussas (1988), theorem 3.1], [6, Bosq (1996), theorem 2.2], we get the following property.

[1] The α-mixing coefficients α_k are defined as:

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

[2] The kernel K is of order r if:

$$\forall \alpha \in N^d, \ \alpha_1 + ... + \alpha_d \in \{1, ..., r-1\}, \ \int x_1^{\alpha_1} ... x_d^{\alpha_d} K(x)\, dx = 0;$$

$$\exists \alpha \in N^d, \ \alpha_1 + ... + \alpha_d = r, \ \int \int x_1^{\alpha_1} ... x_d^{\alpha_d} K(x)\, dx \neq 0.$$

Note that, if K_1 and K_2 are of order r, then $K_1 K_2$ is also of order r.

Theorem 3.4 *: Under assumptions A.2-A.8, the kernel estimator of the p.d.f. is uniformly strongly consistent:*

$$\sup_{(x,y)\in\mathcal{X}^2}\left|\hat{f}_N(x,y)-f(x,y)\right|\to 0 \text{ a.s.}.$$

We can deduce a uniform consistency property of integrals with respect to $\hat{f}_N$. Let us introduce the additional assumption.

Assumption A.9 : *The probability density function f is bounded from below by $\varepsilon > 0$.*

Such assumption in now standard in the nonparametric literature. If the density function were known, it would be sufficient to transform the data by the cumulative density function, since the transformed data would follow the uniform distribution on the compact set $\mathcal{X} = [0,1]^d$. Of course, in practice, we do not know the probability density function generating the data and an appropriate one to one transformation is more difficult to find.

Theorem 3.5 *: Under assumptions A.2-A.9,*

$$\int\int g(x,y)\,\hat{f}_N(x,y)\,dxdy,$$

converges a.s. uniformly to

$$\int\int g(x,y)\,f(x,y)\,dxdy,$$

for any function g in $\mathcal{G} = \{g : \int\int |g(x,y)|\,f(x,y)\,dxdy \le 1\}$.

Proof. See Appendix B. ■

Let us now consider the asymptotic distribution and the assumptions below (see [6, Bosq (1996)]).

Assumption A.10 *: 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.11 : *The p.d.f. f_{t_1,t_2} of $\{(X_{t_1},Y_{t_1}),(X_{t_2},Y_{t_2})\}$ satisfies $\sup_{t_1<t_2}\|f_{t_1,t_2}-f\otimes f\|_\infty<\infty$, where $f\otimes f$ denotes the product of marginal p.d.f. of (X_{t_i},Y_{t_i}), $i=1,2$.*

Assumption A.12 : *The p.d.f. f is twice continuously differentiable on $]0,1[^{2d}$, and there exists b such that $\|f\|_\infty < b$ and $\left\|f^{(2)}\right\|_\infty < b$.*

Assumption A.13 : *As $N\to\infty$, $h_{iN}\to 0$, $Nh_{iN}^d\to\infty$, $Nh_{iN}^{d+4}\to 0$, $i=1,2$.*

Hence, from [6, Bosq (1996), theorem 2.3], we get the following central limit theorem.

Theorem 3.6 : *Under assumptions A.2-A.7, A.10-A.13, for any* $(x,y) \in]0,1[^{2d}$,

$$\sqrt{Nh_{2N}^{d}h_{1N}^{d}}\left[\hat{f}_N(x,y)-f(x,y)\right]dy \xrightarrow{d} \mathcal{N}(0,W(x,y)),$$

where $W(x,y)=f(x,y)\int K_1^2(u)\,du\int K_2^2(u)\,du.$

In the sections below, we need central limit theorems for specific transformations of the kernel density estimator.

Theorem 3.7 : *Under assumptions A.2-A.7, A.10-A.13,*

i) If $g(x,y)$ *is a given function, we have:*

$$\sqrt{N}\int\int g(x,y)\left[\hat{f}_N(x,y)-f(x,y)\right]dxdy \xrightarrow{d} \mathcal{N}(0,V),$$

where $V=\sum_{k=-\infty}^{\infty} Cov\left(g(X_n,Y_n),g(X_{n+k},Y_{n+k})\right).$

ii) If $g(x,y)$ *is a given function, we have:*

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

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

3.3.2 Consistency of the estimated canonical analysis

We are concerned by the consistency of the p first estimated canonical correlations and canonical variates, i.e. their convergence to their theoretical counterparts. We need first to introduce some identifiability conditions. Assumption A.1 is an identifiability condition for the canonical correlations. Another identifiability assumption has to be introduced for the canonical variates, which are defined up to a change of sign.

Assumption A.14 : *There exists a value* x_0 *such that* $\varphi_j(x_0)\neq 0$, $j=1,...,p$.

Hence, we select the pair of canonical variates with $\hat{\varphi}_{jN}(x_0)>0$, $\varphi_j(x_0)>0$, $j=1,...,p$. Moreover, the functional parameters of interest have to belong to the admissible values of the associated estimators.

Assumption A.15 : *The canonical variates* φ_i and ψ_j are such that*:*

i) $\varphi_i \in L^2_{K_1}(X)$, $i=1,...,p$.

ii) $\psi_j \in L^2_{K_2}(Y)$, $j=1,...,p$.

Theorem 3.8 : *Under the identification conditions A.1, A.14, the estimability condition A.15 and the assumptions of Theorem 3.5,* $\hat{\lambda}_{i,N}$, $\hat{\varphi}_{i,N}$, $\hat{\psi}_{i,N}$, $i = 1, ..., p$ *converge to their theoretical counterparts in the sense:*

$$\hat{\lambda}_{i,N} \underset{N\to\infty}{\to} \lambda_i \text{ a.s.}$$

$$\int \left|\hat{\varphi}_{i,N}(x) - \varphi_i(x)\right|^2 f(x,.)\,dx \underset{N\to\infty}{\to} 0 \text{ a.s.},$$

$$\int \left|\hat{\psi}_{i,N}(y) - \psi_i(y)\right|^2 f(.,y)\,dy \underset{N\to\infty}{\to} 0 \text{ a.s.},$$

Proof. See Appendix C.1 ■

Remark 3.6 : *The convergence result of the canonical correlations is easily understood. Indeed, under assumptions A2-A9, the operators* $\hat{T}_N$ *converge a.s. uniformaly to the conditional expectation operator* T. *From [18, Dunford-Schwartz (1963), chapter XI], we deduce that the canonical correlations* $\hat{\lambda}_{i,N}$, $i = 1, 2, ...$ *converge a.s. uniformaly to* λ_i, $i = 1, 2, ...$ *However, this uniformity with respect to the order of the canonical analysis is not valid when we consider the canonical variates (see [13, Dauxois-Nkiet (1998), remark 5.1]).*

3.3.3 Asymptotic distributions

The convergence properties of the estimated canonical correlations and canonical variates under the identification conditions A.1, A.14 allow us to expand the first order conditions. These conditions are of the same type for the estimators $\left(\hat{\lambda}_{i,N}, \hat{\varphi}_{i,N}, \hat{\psi}_{i,N}\right)$, i varying. We present the case $i = 1$. We can directly consider the pair $\mu_1 = \lambda_1^2$, φ_1 and its estimated counterpart $\hat{\mu}_{1,N} = \hat{\lambda}_{1,N}^2$, $\hat{\varphi}_{1,N}$. The pair (μ_1, φ_1) satisfies:

$$\int \varphi_1(y)\, c(x,y)\,dy = \mu_1 \varphi_1(x), \tag{3.7}$$

$$\int \varphi_1^2(x)\, f(x,.)\,dx = 1,$$

where $c(x,y) = \int f(y,z)\, f(x,z) / (f(x,.)\, f(.,z))\,dz$ (see (2.10)). Similarly, the pair $\left(\hat{\mu}_{1,N}, \hat{\varphi}_{1,N}\right)$ satisfies:

$$\int \hat{\varphi}_{1,N}(y)\, \hat{c}_N(x,y)\,dy = \hat{\mu}_{1,N} \hat{\varphi}_{1,N}(x), \tag{3.8}$$

$$\int \hat{\varphi}_{1,N}^2(x)\, f(x,.)\,dx = 1,$$

where $\hat{c}_N(x,y) = \int \hat{f}_N(y,z)\, \hat{f}_N(x,z) / \left(\hat{f}_N(x,.)\, \hat{f}_N(.,z)\right) dz.$

Theorem 3.9 : *Under the assumptions of theorem 3.8, $\hat{\mu}_{1,N}$, $\hat{\varphi}_{1,N}$ have the asymptotic expansions:*

$$\begin{aligned}
\sqrt{N}\left(\hat{\mu}_{1,N}-\mu_1\right) &= \sqrt{N}\left\langle \hat{A}_{1,N},\varphi_1\right\rangle + o\left(1\right), \\
\sqrt{Nh_{1N}^d}\left(\hat{\varphi}_{1,N}\left(x\right)-\varphi_1\left(x\right)\right) &= \sqrt{Nh_{1N}^d}\left[\hat{B}_{1,N}\varphi_1\left(x\right)+\sum_{j=2}^{\infty}\frac{\left\langle \hat{A}_{1,N},\varphi_j\right\rangle}{\mu_1-\mu_j}\varphi_j\left(x\right)\right]+o\left(1\right), \\
&\simeq \frac{1}{\mu_1}\sqrt{Nh_{1N}^d}\hat{A}_{1,N}\left(x\right),
\end{aligned}$$

where:

$$\begin{aligned}
\hat{A}_{1,N}\left(x\right) &= \int \varphi_1\left(y\right)\left[\hat{c}_N\left(x,y\right)-c\left(x,y\right)\right]dy \\
&\simeq \frac{1}{f\left(x,.\right)}\left(\mu_1^{\frac{1}{2}}\int \psi_1\left(z\right)\delta\hat{f}_N\left(x,z\right)dz-\mu_1\varphi_1\left(x\right)\delta\hat{f}_N\left(x,.\right)\right), \\
\hat{B}_{1,N} &= -\frac{1}{2}\int \varphi_1^2\left(x\right)\left[\hat{f}_N\left(x,.\right)-f\left(x,.\right)\right]dx.
\end{aligned}$$

Proof. See Appendix C.2 ■

Theorem 3.10 : *Under assumptions A.1-A.15,*

i) the asymptotic distributions of $\hat{\mu}_{i,N}$, $i=1,...,p$ are:

$$\sqrt{N}\left(\hat{\mu}_{i,N}-\mu_i\right)\xrightarrow{d}\mathcal{N}\left(0,V_i\right),$$

where

$$\begin{aligned}
V_i &= \mu_i\sum_{k=-\infty}^{\infty}Cov\left[\varphi_i\left(X_n\right)\psi_i\left(Y_n\right)+\psi_i\left(X_n\right)\varphi_i\left(Y_n\right)\right. \\
&\quad -\mu_i^{\frac{1}{2}}\varphi_i^2\left(X_n\right)-\mu_i^{\frac{1}{2}}\psi_i^2\left(Y_n\right),\ \varphi_i\left(X_{n+k}\right)\psi_i\left(Y_{n+k}\right) \\
&\quad \left.+\psi_i\left(X_{n+k}\right)\varphi_i\left(Y_{n+k}\right)-\mu_i^{\frac{1}{2}}\varphi_i^2\left(X_{n+k}\right)-\mu_i^{\frac{1}{2}}\psi_i^2\left(Y_{n+k}\right)\right].
\end{aligned}$$

ii) The asymptotic distributions of $\hat{\varphi}_{i,N}$, $i=1,...,p$ are:

$$\sqrt{Nh_{1N}^d}\left(\hat{\varphi}_{i,N}\left(x\right)-\varphi_i\left(x\right)\right)\xrightarrow{d}\mathcal{N}\left(0,W_i\left(x\right)\right),$$

where

$$W_i\left(x\right)=\frac{1}{\mu_i}\frac{1}{f\left(x,.\right)}\int K_1^2\left(u\right)du\ V\left[\psi_i\left(Y\right)\mid X=x\right].$$

iii) The asymptotic distributions of $\hat{\psi}_{i,N}$*,* $i = 1, ..., p$ *are:*

$$\sqrt{Nh_{2N}^d}\left(\hat{\psi}_{i,N}(y) - \psi_i(y)\right) \xrightarrow{d} \mathcal{N}(0, U_i(y)),$$

where

$$U_i(y) = \frac{1}{\mu_i}\frac{1}{f(.,y)}\int K_2^2(u)\,du\,V[\varphi_i(X) \mid Y = y].$$

Proof. See Appendix C.3 ■

Remark 3.7 *The asymptotic variance* $W_i(x)$ *of* $\hat{\varphi}_{i,N}$ *coincide with the asymptotic variance of the Nadarayha-Watson estimator of* $\frac{1}{\lambda_i}E[\psi_i(Y) \mid W]$*, which corresponds to the interpretation of the canonical variate as conditional expectation.*

Remark 3.8 *Theorem 3.9 can also be used to get the joint distributions of either two canonical correlations, or two canonical variates of different orders. For instance, we get the asymptotic normality for:*

$$\sqrt{Nh_{1N}^d}\left(\hat{\varphi}_{i,N}(x) - \varphi_i(x), \hat{\varphi}_{j,N}(x) - \varphi_j(x)\right),\ i \neq j,$$

with asymptotic variance:

$$W_{i,j}(x) = \frac{1}{\mu_i^{\frac{1}{2}}}\frac{1}{\mu_j^{\frac{1}{2}}}\frac{1}{f(x,.)}\int K_1^2(u)\,du\,cov\left[\psi_i(Y), \psi_j(Y) \mid X = x\right].$$

It has to be noted that the estimated canonical variates are generally correlated.

3.4 Reversibility property

Under the reversibility hypothesis, we can perform the constrained canonical analysis. From the comparison of the constrained and unconstrained estimators, we get some insights for testing the reversibility hypothesis. Under the hypothesis of reversibility, the distribution of (X, Y) and (Y, X) coincide, i.e. the p.d.f. $f(x, y)$ is symmetric in x and y. Hence, we can introduce a kernel estimator of the density taking into account this symmetry constraint. For this purpose, 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 $(X_i, Y_i), (Y_i, X_i), i = 1, ..., N$. The constrained kernel estimator of the density is:

$$\begin{aligned} \hat{f}_N^R(x,y) &= \frac{1}{2N}\sum_{n=1}^{N}\frac{1}{h_N^{2d}}\left(K\left(\frac{X_n - x}{h_N}\right)K\left(\frac{Y_n - y}{h_N}\right)\right. \\ &\left. + K\left(\frac{X_n - y}{h_N}\right)K\left(\frac{Y_n - x}{h_N}\right)\right). \end{aligned} \quad (3.9)$$

Hence, we replace in the initial canonical analysis (2.1) the inner product by:

$$\langle \varphi(X), \psi(Y) \rangle_N^R = \int\int \varphi(x)\,\psi(y)\,\hat{f}_N^R(x,y)\,dxdy. \tag{3.10}$$

The constrained estimators of the canonical correlations and canonical covariates are denoted: $\hat{\lambda}_{i,N}^R$, $\hat{\varphi}_{i,N}^R = \hat{\psi}_{i,N}^R$, $i \geq 0$. We give below the joint asymptotic properties of the constrained canonical analysis under the null hypothesis of reversibility.

Theorem 3.11 *: Under the assumptions of theorem 3.8, and if the reversibility hypothesis is satisfied:*

i) $\hat{\lambda}_{i,N}^R$, $i = 1, ..., p$ *converge a.s. to their theoretical counterparts, whereas* $\int \left| \hat{\varphi}_{i,N}^R(x) - \varphi_i(x) \right|^2 f(x,.)\,dx \underset{N\to\infty}{\to} 0$ *a.s.,* $i = 1, ..., p$.

ii) They admit the asymptotic expansions:

$$\begin{aligned}\sqrt{N}\left(\hat{\lambda}_{i,N}^R - \lambda_i\right) &= \sqrt{N}\left\langle \hat{A}_{i,N}^R, \varphi_i \right\rangle + o(1), \\ \sqrt{Nh_N^d}\left(\hat{\varphi}_{i,N}^R(y) - \varphi_i(y)\right) &= \frac{1}{\lambda_i}\sqrt{Nh_N^d}\hat{A}_{i,N}^R(y) + o(1),\end{aligned}$$

where:

$$\hat{A}_{i,N}^R(y) = \int \varphi_i(x)\left[\hat{f}_N^R(x \mid y) - f(x \mid y)\right]dx.$$

Proof. The proof of consistency is similar to the proof given for theorem 3.8, whereas the asymptotic expansions are developped in appendix C.4. ■

Hence, we deduce the corresponding following asymptotic distributions.

Theorem 3.12 *: Under assumptions A.1-A.15, and if the reversibility hypothesis is satisfied,*

i) the asymptotic distributions of $\hat{\lambda}_{i,N}^R$, $i = 1, ..., p$ *are:*

$$\sqrt{N}\left(\hat{\lambda}_{i,N}^R - \lambda_i\right) \xrightarrow{d} \mathcal{N}\left(0, V_i^R\right),$$

where

$$\begin{aligned}V_i^R = \frac{1}{2}\sum_{k=-\infty}^{\infty} Cov&\left[\varphi_i(X_n)\,\varphi_i(Y_n) - \lambda_i\varphi_i^2(Y_n),\right. \\ &\left.\varphi_i(X_{n+k})\,\varphi_i(Y_{n+k}) - \lambda_i^2\varphi_i(Y_{n+k})\right].\end{aligned}$$

ii) The asymptotic distributions of $\hat{\varphi}^R_{i,N}$, $i = 1, ..., p$ *are:*

$$\sqrt{Nh_N^d}\left(\hat{\varphi}^R_{i,N}(x) - \varphi_i(x)\right) \xrightarrow{d} \mathcal{N}\left(0, W_i^R(x)\right),$$

where

$$W_i^R(x) = \frac{1}{2}\frac{1}{\lambda_i^2}\frac{1}{f(x,.)}\int K_1^2(u)\, du\, V\left[\varphi_i(Y) \mid X = x\right].$$

Proof. See Appendix C.5 ■

3.5 Comparison of the 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 variates. The difference between these two types of estimators can be used to construct testing procedures of the reversibility hypothesis.

3.5.1 Canonical correlations

Under the null hypothesis, the unconstrained estimator of μ_i has an asymptotic variance given by:

$$V_a^R\left(\sqrt{N}\left(\hat{\mu}_{i,N} - \mu_i\right)\right) = 4\mu_i V_i^R,$$

using the formula of theorem 3.10 with $\psi_i = \varphi_i$. By applying the δ-method, we deduce that:

$$V_a^R\left(\sqrt{N}\left(\hat{\lambda}_{i,N} - \lambda_i\right)\right) = V_i^R = V_a^R\left(\sqrt{N}\left(\hat{\lambda}^R_{i,N} - \lambda_i\right)\right). \tag{3.11}$$

The two estimators have the same asymptotic precision, and their difference $\hat{\lambda}_{i,N} - \hat{\lambda}^R_{i,N}$ will tend to zero at a rate of convergence strictly larger than $N^{-\frac{1}{2}}$.

3.5.2 Canonical variates

Under the reversibility hypothesis, we have:

$$\begin{aligned} &\sqrt{Nh_N^d}\left(\hat{\varphi}^R_{i,N}(y) - \varphi_i(y)\right) \\ = \; &\frac{1}{\lambda_i}\frac{1}{f(.,y)}\left(\int \varphi_i(x)\sqrt{Nh_N^d}\,\delta\hat{f}_N^R(x,y)\, dx - \lambda_i\varphi_i(y)\sqrt{Nh_N^d}\,\delta\hat{f}_N^R(.,y)\right), \end{aligned}$$

from theorem 3.12 and appendix $C.4$ (formula $(C.15)$). Moreover, under the reversibility hypothesis, we deduce from theorem 3.12 and appendix $C.2$ (formula $(C.7)$):

$$\begin{aligned}&\sqrt{Nh_N^d}\left(\hat{\varphi}_{i,N}(y)-\varphi_i(y)\right)\\ \simeq\ &\frac{1}{\lambda_i}\frac{1}{f(.,y)}\left(\int\varphi_i(x)\sqrt{Nh_N^d}\delta\hat{f}_N(x,y)\,dx\ -\lambda_i\varphi_i(y)\sqrt{Nh_N^d}\delta\hat{f}_N(.,y)\right).\end{aligned}$$

We note that:

$$\begin{aligned}&\sqrt{Nh_N^d}\left(\hat{\varphi}_{i,N}(y)-\hat{\varphi}_{i,N}^R(y)\right)\\ \simeq\ &\frac{1}{\lambda_i}\frac{1}{f(.,y)}\left(\int\varphi_i(x)\sqrt{Nh_N^d}\left[\delta\hat{f}_N(x,y)-\delta\hat{f}_N^R(x,y)\right]dx\right.\\ &\left.-\lambda_i\varphi_i(y)\sqrt{Nh_N^d}\delta\hat{f}_N(.,y)\right)\left[\delta\hat{f}_N(.,y)-\delta\hat{f}_N^R(.,y)\right],\end{aligned}$$

is a simple linear transformation of the difference between the constrained and unconstrained estimators of the p.d.f. We can note (see theorems 3.10 and 3.12) that under the null hypothesis, the asymptotic variance of $\hat{\varphi}_{i,N}^R(y)$ is half the asymptotic variance of $\hat{\varphi}_{i,N}(y)$. A simple computation shows that, under the null hypothesis, we have:

$$\begin{aligned}&\sqrt{Nh_N^d}\left(\hat{\varphi}_{i,N}(y)-\hat{\varphi}_{i,N}^R(y)\right)\\ \simeq\ &\frac{\sqrt{Nh_N^d}}{2}\left(\hat{\varphi}_{i,N}(y)-\hat{\psi}_{i,N}(y)\right)\xrightarrow{d}\mathcal{N}\left(0,W_i^R(y)\right).\end{aligned}$$

4 Applications

In this section, we provide two illustrations of the previous approach. The first one is based on an artificial dataset consisting of simulated realizations of an Ornstein-Uhlenbeck process. 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. The nonlinear canonical analysis of this series shows that its dynamics is not compatible with an underlying diffusion model.

4.1 Ornstein-Uhlenbeck process

We consider a continuous time process $(X_t,\ t\geq 0)$ satisfying the stochastic differential equation:

$$dX_t=\beta(\alpha-X_t)\,dt+\sigma dW_t. \tag{4.1}$$

It is well known that this equation has a stationary gaussian solution. Moreover, the associated discrete time process has a linear autoregressive representation of order one, with a positive autoregressive coefficient. Therefore, the canonical variates $\varphi_i = \psi_i$ are the Hermite polynomials, up to a change of sign, and the canonical correlations are $\lambda_i = \exp(-i\beta)$, $i \geq 0$.

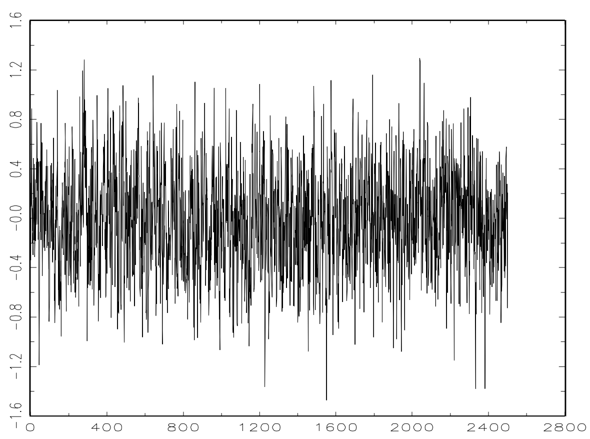

Figure 4.1: Simulated Trajectory

We simulate a path $(X_t, \ \ t = 1, 2, ..., T)$ of length $T = 2500$, of the model with parameter values: $\alpha = 0$, $\beta = 0.8$, $\sigma = 0.5$. This path is plotted in Figure (4.1). Next, we use these artificial observations to find a nonlinear canonical decomposition and deduce the associated nonparametric estimators of the drift function: $\mu(x) = -0.8\,x$, and the volatility function: $\sigma(x) = 0.5$. Three estimation methods are successively considered:

i) an unconstrained sieve method based on a finite basis of polynomials with degree smaller then six;

ii) an unconstrained kernel method, with gaussian kernels $K_1(x) = K_2(x) = \frac{1}{\sqrt{2\pi}} \exp -\frac{x^2}{2}$, and bandwidths $h_1 = h_2 = 0.1025$;

iii) the same kernel method constrained by the reversibility hypothesis.

They are applied to the data X_t, X_{t-1}, without any preliminary transformation of the data to get a compact set values.

4.1.1 Sieve method and kernel based method

Figures (4.2)-(4.3) present the estimator $\hat{\varphi}_{1,N}$ of the first canonical variate computed by a polynomial based sieve method and by the kernel method, respectively. Each Figure presents the true function (dotted line) and its estimator (continuous line).

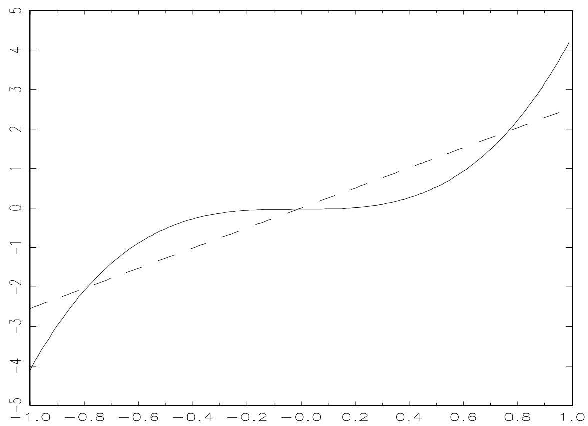

Figure 4.2: Sieve estimator of the first current canonical variate

Even though the sieve approach is consistent when the number of elements in the basis increases, it may be difficult to select an appropriate ordering of functions of the basis in practice. To give an idea of this problem, we artificially impose a null coefficient for the term in x. Since the true first canonical variate corresponds to the Hermite polynomial of order one, the expected result is outside the finite dimensional subspace in which we compute the estimator. It explains the poor fit obtained in Figure (4.2). Moreover, the estimated canonical correlation is equal to $\hat{\lambda}_{1,N} = 0.30$, which is much lower than the true value equal to 0.449.

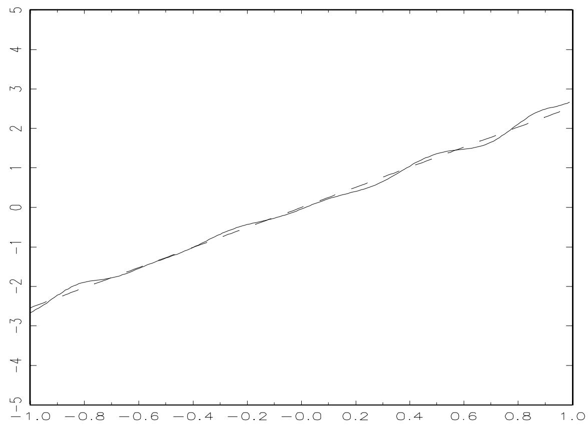

Figure 4.3: Kernel estimator of the first current canonical variate

However, this sieve estimator can be used as the initial canonical variate in the iterated estimation scheme introduced in subsection 3.2. By applying this procedure, we obtain a consistent estimator of the first canonical variate. After three iterations, we get the estimator plotted in Figure (4.3). The estimated first canonical correlation is now equal to 0.445, which is close to the true value.

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, for the kernel method and the sieve method. Each Figure presents the averaged estimators and the pointwise confidence bands. In sieve method, without a priori knowledge on the pattern of the canonical variate, we can make a bad selection of the functions of the finite basis. As noted in Figure (4.4), this may imply a confidence band which only unfrequently includes the true canonical variate.

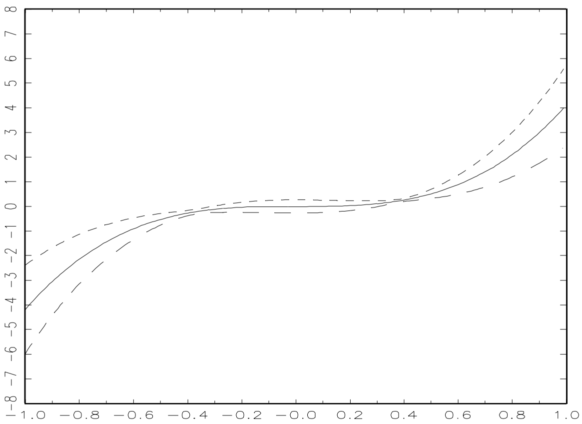

Figure 4.4: Confidence band for the first current canonical variate (sieve method)

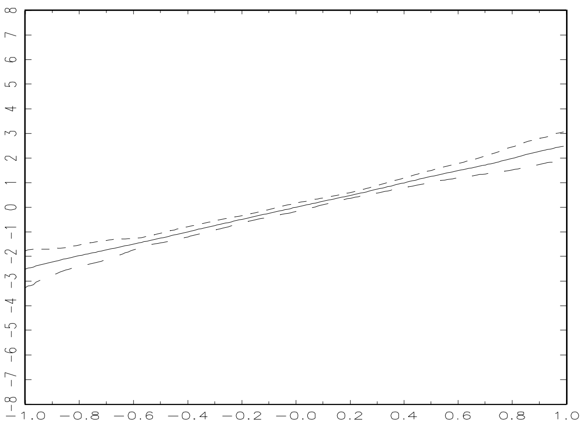

Figure 4.5: Confidence band for the first current canonical variate (kernel method)

The bias and variance of the estimated first canonical correlations are summarized in table 4.1. With the previous choice of the basis of polynomials, the sieve method underestimates the first canonical correlation but, when this biased estimator is used as initial input for the iterative kernel method, the kernel iteration procedure will reduce the bias (and also the variance).

	bias	variance
sieve estimator	0.085	0.028
kernel estimator	0.003	0.011

Table 4.1: Properties of the estimated first canonical correlation

4.1.2 Comparison of constrained and unconstrained kernel based methods

The Ornstein-Uhlenbeck process is reversible. In this section, we compare the results of the constrained and unconstrained kernel based methods. We provide in Table 4.2 the estimated canonical correlations.

Canonical correlations λ_i			
0rder	True	Unconstrained	Constrained
1	0.4493	0.4506	0.4503
2	0.2018	0.2155	0.2133
3	0.0907	0.1074	0.1019
4	0.0407	0.0531	0.0472
5	0.0183	0.0271	0.0264
6	0.0082	0.0082	0.0108

Table 4.2: Estimated canonical correlations

The estimates are close to each other, and close to the true values. The current estimated canonical variates are plotted in Figure (4.6) for the unconstrained case, and in figure (4.7) for the constrained one. The first variate is represented by a continuous line, the second one by a dashed line, and the third one by a dotted line.

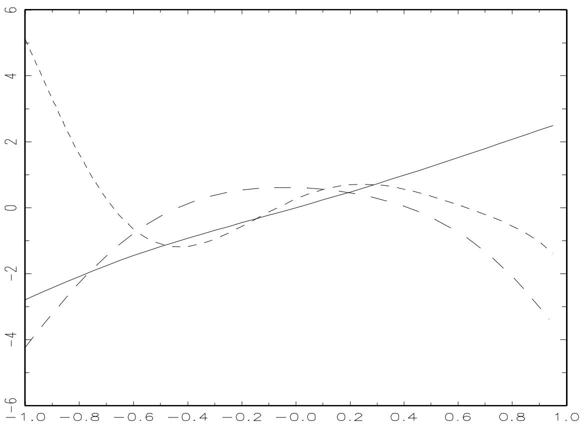

Figure 4.6: Unconstrained estimated canonical variates

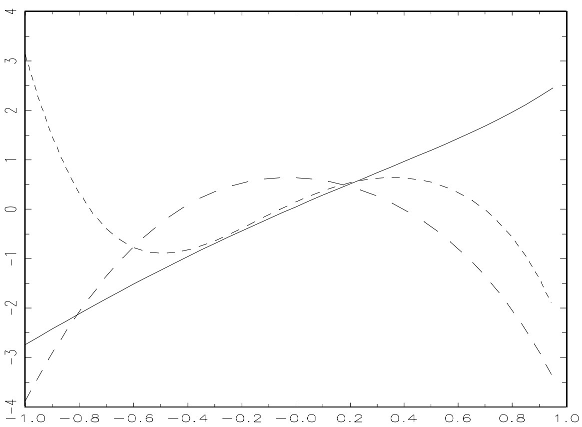

Figure 4.7: Constrained estimated canonical variates

The constrained estimated canonical variates closely imitate the functional forms of the Hermite polynomials, whereas we observe some asymmetry in the second canonical variate estimated without constraint. By comparing the constrained and unconstrained estimators, we obtain diagnostics for the reversibility hypothesis (see subsection 3.5).

Finally, we can propose some diagnostic procedures for the normality hypothesis referred to in section 2.2.3. Let us consider the canonical correlations and corollary 2.2, for instance. Under the normality hypothesis, we get: $\lambda_i = \lambda_1^i$, or $\ln \lambda_i = i \ln \lambda_1$. We give below the result of linear regressions of the log estimated canonical correlations on 1 and i, $i = 1, ..., 6$, both for the constrained and unconstrained cases.

	constant coefficient	i coefficient	estimated λ_1	R^2
unconstrained	0.0228	-0.7650	0.4652	0.9967
constrained	-0.038	-0.7416	0.4763	0.9993

Table 4.3: Pattern of the log canonical correlations

We observe large values of the multiple correlation coefficient R^2 indicating the high adequacy of the fits. The two following figures provide the patterns of the log estimated canonical correlations for the unconstrained and constrained cases. Each dot represents the log estimated canonical correlation. The dotted lines correspond to the linear adjustment and the confidence intervals, respectively.

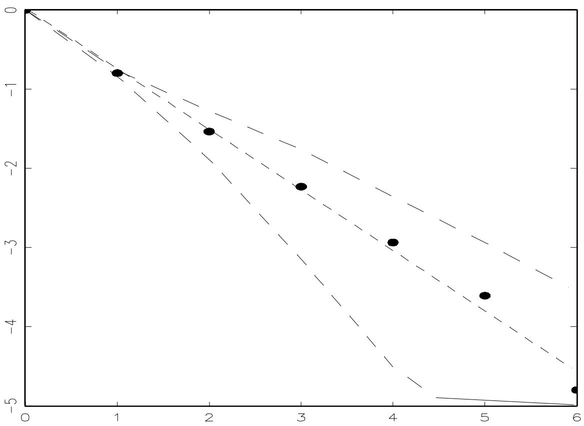

Figure 4.8: Pattern of the log canonical correlations (unconstrained method)

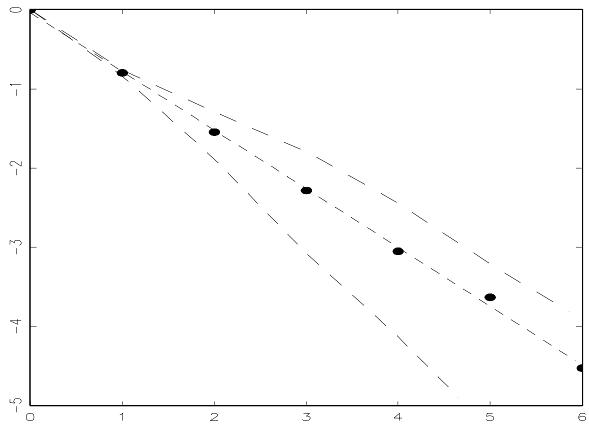

Figure 4.9: Pattern of the log canonical correlations (constrained method)

4.1.3 Estimation of the drift and volatility functions

In the case of univariate diffusion equation, the nonlinear canonical analysis allows to identify both the drift and the volatility functions, from the two first canonical correlations and variates (see the approach by [15, Demoura (1993)] described in subsection 2.2.1). The two following figures provide the true functions with a continuous line, their estimators and their confidence bands with dotted lines.

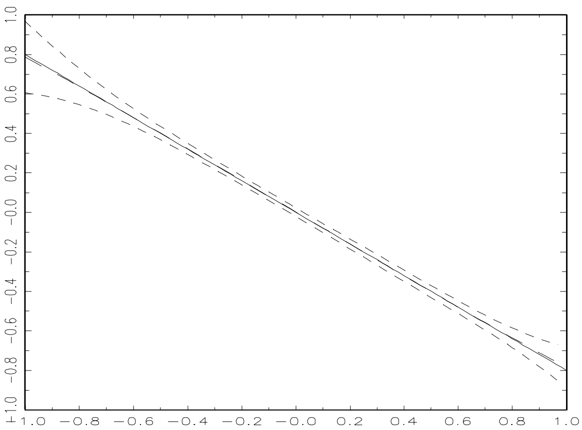

Figure 4.10: Estimated drift function

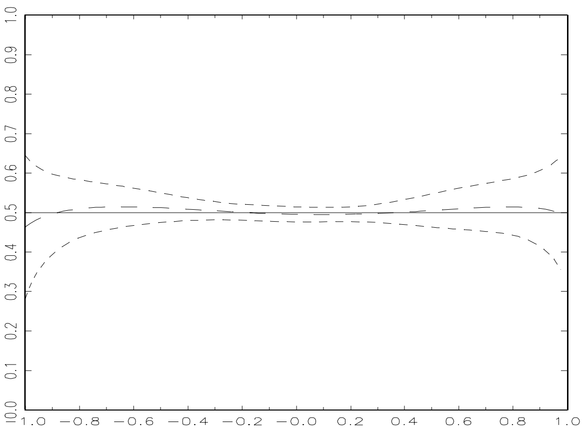

Figure 4.11: Estimated volatility function

4.2 High Frequency Data

We apply the previous approach to a series of returns corresponding to the Alcatel stock traded on the Paris-Bourse.

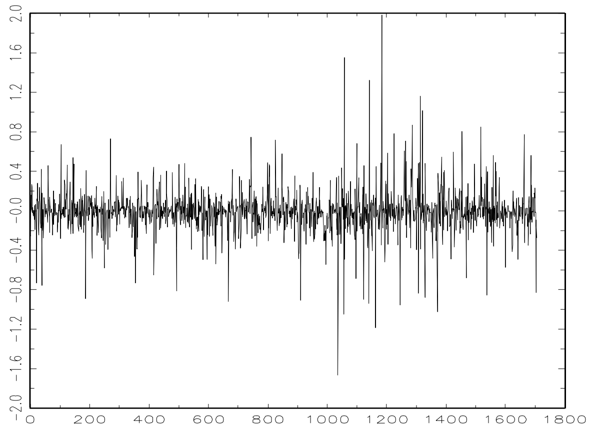

Figure 4.12: Returns for Alcatel stock

The prices are resampled from real time records at a constant interval of $20mn$ and the returns are computed by differencing the log-prices. The sampling period is May 2, 1997 to August 30, 1997 and contains 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.

We implemented the unconstrained and constrained kernel based methods, with a gaussian kernel and bandwiths $h_{1N} = h_{2N} = 0.062$, and horizon $h = 1$. The estimated canonical variates are provided in Figures (4.13) and (4.14) for the unconstrained case, and in Figure (4.15) for the constrained one. The first variate is represented by a continuous line, the second one by a dashed line, and the third one by a dotted line.

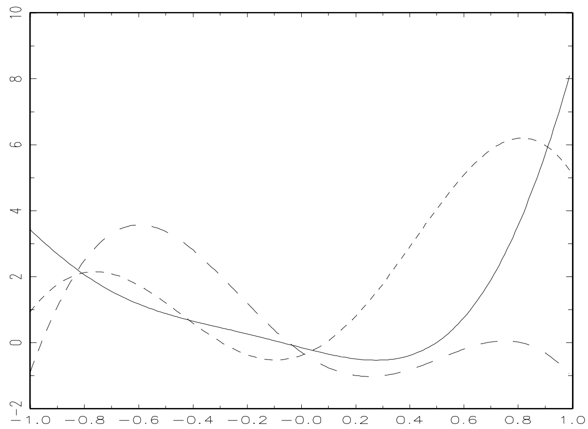

Figure 4.13: Estimated current canonical variates

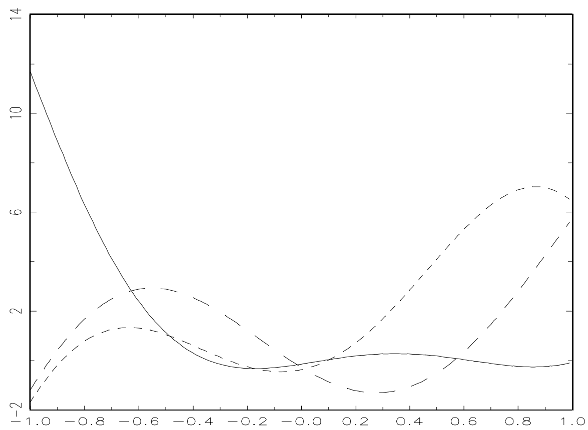

Figure 4.14: Estimated lagged canonical variates

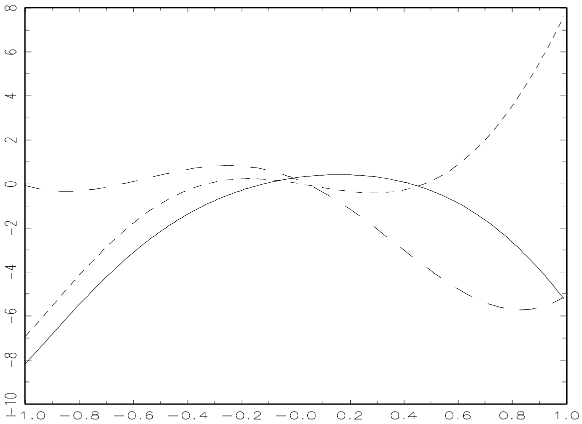

Figure 4.15: Estimated canonical variates under reversibility

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

$$dr_t = \mu\,(r_t)\,dt + \sigma\,(r_t)\,dW_t \qquad \text{(say)}.$$

In such case, the process is necessarily reversible and the first canonical variate corresponds to a monotone 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 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 unsignificant, which is consistent with the the theory of market efficiency. In our case, the first order linear correlation is 0.065. 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 the 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 to distinguish bull and bear markets. The reversibility hypothesis is also clearly rejected when we compare the constrained and unconstrained estimated canonical correlations. The introduction of the reversibility constraint induces an underestimation of the first canonical correlation by about 30%.

Canonical correlations λ_i		
0rder	Unconstrained	Constrained
1	0.3595	0.2569
2	0.1829	0.1459
3	0.1467	0.1260
4	0.030	0.1255
5	0.003	0.0054
6	0.001	0.0016

Table 4.3: Estimated canonical correlations

5 Concluding Remarks

In this paper, we developped a nonlinear canonical correlation analysis based on kernel estimators of the density function. This approach has been applied to a high frequency series of returns. There exist in the literature other nonparametric estimation methods involving the sieves (see [9, Chen-Hansen-Scheinkman (1998)]) or the generalized kernels (see [7, Bosq-Lecoutre (1987)]). The analysis of the asymptotic properties of the estimated canonical correlations and variates are similar. We have shown that our method outperforms the sieve method which artificially restraint the domain of estimation.

Due to the nonlinearities, these nonparametric techniques require a large set of observations, which explains the interest for financial data, including high frequency data. But, even if we restrict to this field, nonlinear canonical analysis is certainly a useful tool for understanding the risk of liquidity in an analysis of intratrade durations (see [21, Gouriéroux-Jasiak (1998)]), or to implement technical analysis based models for directions of price changes (see [12, Darolles-Gouriéroux-Le Fol (1998)]).

References

[1] Abramowitz, M., and I., Stegun (1965), *The Handbook of Mathematical Functions*, Dover, New-York.

[2] Aït-Sahalia, Y. (1995), *The Delta and Bootstrap Methods for Nonparametric Kernel Functionals,* Discussion Paper, MIT.

[3] Anderson, T.W. (1963), *Asymptotic Theory for Principle Component Analysis*, Annals of Statistics, **34**, 122-148.

[4] Barrett, J.F., and D.G., Lampard (1955), *An Expansion for some Second Order Probability Distributions and its Application to Noise Problems*, I.R.E. Trans. PGIT, **IT-1**, 10-15.

[5] Bosq, D. (1973), *Sur l'estimation de la densité d'un processus stationnaire et mélangeant,* C.R. Aca. Sci. Paris, Série A 281, 1095-1098.

[6] Bosq, D. (1996), *Nonparametric Statistics for Stochastic Processes*, Lecture Notes in Statistics, Springer-Verlag, New York.

[7] Bosq, D., and J.P., Lecoutre (1987), *Théorie de l'estimation fonctionnelle*, Economica.

[8] Buja, A. (1990), *Remarks on Functional Canonical Variates, Alternating Least Squares Methods and ACE,* Annals of Statistics, **18**, 1032-1069.

[9] Chen, X., Hansen, L., and J., Scheinkman (1998), *Shape Preserving Estimation of Diffusions,* Discussion Paper, University of Chicago.

[10] Darolles, S., Florens, J.P. and E., Renault (1998), *Nonlinear Principal Components and Inference on a Conditional Expectation Operator*, Discussion Paper, CREST.

[11] Darolles, S., and C., Gouriéroux (1997), *Truncated Dynamics and Estimation of Diffusion Equations*, Discussion Paper CREST 9736.

[12] Darolles, S., Gouriéroux, C., and G., LeFol (1998), *Intraday Transaction Price Dynamics*, CREST DP.

[13] Dauxois, J., and G., Nkiet (1998), *Nonlinear Canonical Analysis and Independance Tests,* Annals of Statistics, **26**, 4, 1254-1278.

[14] Dauxois, J., and A., Pousse (1975), *Une extension de l'analyse canonique. Quelques applications*, Ann. Inst. H. Poincaré, **11**, 355-379.

[15] Demoura, S. (1993), *Theory and Application of Transition Operator to the Analysis of Economic Time Series. Necessary and Sufficient Conditions for Nonlinearities in Economic Dynamics. Aliasing Problem*, manuscript.

[16] Diebold, F., Gunther, T., and A., Tay (1997), *Evaluating Density Forecasts*, Discussion Paper, University of Pennsylvania.

[17] Donelson, J., and F., Maltz (1972), *A Comparison of Linear versus Nonlinear Prediction for Polynomial Functions of the Ornstein-Uhlenbeck Process*, Journal of Applied Probability, **9**, 725-744.

[18] Dunford, N., and J., Schwartz (1963), *Linear Operators 2,* Wiley, New York.

[19] Florens, J.P., Renault, E. and N., Touzi (1998), *Testing for Embeddability by Stationary Reversible Continuous-Time Markov Processes*, forthcoming in Econometric Theory.

[20] Ghysels, E., Gouriéroux, C., and J., Jasiak (1998), *Causality Between Returns and Traded Volumes*, CREST DP 9840.

[21] Gouriéroux, C., and J., Jasiak (1998), *Nonlinear Autocorrelograms with Application to Intratrade Durations*, CREST DP 9841.

[22] Gourlay, A., and G., Watson (1973), *Computational Methods for Matrix Eigenproblems*, Wiley, New York.

[23] Granger, C., and P., Newbold (1976), *Forecasting Transformed Series,* Journal of the Royal Statistical Society, B, **38**, 189-203.

[24] Granger, C., and M., Pesaran (1996), *A Decision Theoretic Approach to Forecast Evaluation,* Discussion Paper, University of California San Diego.

[25] Hannan, E. (1961), *The General Theory of Canonical Correlation and its Relation to Functional Analysis,* J. Austral. Math. Soc., **2**, 229-242.

[26] Hansen, L., and J., Scheinkman (1995), *Back to the Future : Generating Moment Implications for Continuous Time Markov Processes*, Econometrica **73**, 767-807.

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

[28] Hinich, M. (1982), *Testing For Gaussianity and Linearity of a Stationary Time Series,* Journal of Time Series Analysis, **3**, 168-176.

[29] Hotelling, H. (1936), *Relations Between Two Sets of Variables*, Biometrika, **28**, 321-377.

[30] Jennrich, R. (1969), *Asymptotic Properties of Nonlinear Least Squares Estimators,* The Annals of Mathematical Statistics, **40**, 633-643.

[31] Kessler, M., and M., Sorensen (1996), *Estimating Equations Based on Eigenfunctions for Discretely Observed Diffusion Process*, manuscript.

[32] Lancaster, H. (1968), *The Structure of Bivariate Distributions,* Ann. Math. Statist., **29**, 719-736.

[33] Lawley, D., and A., Maxwell (1971), *Factor Analysis as a Statistical Method*, Wiley, New York.

[34] Muirhead, R. (1982), *Aspects of Multivariate Statistical Theory*, Wiley, New York.

[35] Naylor, A., and G., Snell (1982), *Linear Operator Theory in Ingineering and Science,* Springer, New York.

[36] Neveu, J. (1968), *Processus Aléatoires Gaussiens*, Presses de l'Université de Montréal.

[37] Ray, B., and R.S., Tsay (1998), *Common Long Run Dependencies in Daily Stock Volatilities*, Discussion Paper, Graduate School of Busines, University of Chicago.

[38] Rice, J., and B., Silverman (1991), *Estimating the Mean and Covariance Structure Nonparametrically when the Data are Curves*, Journal of the Royal Statistical Society, B, **53**, 233-243.

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

[40] Rosenblatt, M. (1970), *Density Estimates and Markov Sequences*, in Puri, Ed., Nonparametric Techniques in Statistical Inference, Cambridge University Press.

[41] Roussas, G. (1969), *Nonparametric Estimation in Markov Process*, Annals of the Institute of Statistical Mathematics, Vol. 21, 73-87.

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

[43] Rutishauser, H. (1969), *Computational Aspects of F.L. Bauer's Simultaneous Iteration Method*, Numerisch Mathematik, **13**, 4-13.

[44] Sarmanov, O. (1963), *Investigating of Stationary Markov Processes by the Method of Eigenfunction Expansion*, Translated in Selected Translations in Mathematical Statistics and Probability Theory, **4**, 245-269. Amer. Math. Soc., Providence.

[45] Silverman, B. (1986), *Density Estimation for Statistics and Data Analysis*, Monographs on Statistics and Applied Probability, Chapman and Hall.

[46] Silverman, B. (1996), *Smoothed Functional Principle Component Analysis by Choice of Norm*, Annals of Statistics, **24**, 1-24.

[47] Tiao, G.C., and R.S., Tsay (1989), *Model Specification in Multivariate Time Series*, Journal of the Royal Statistical Society, B, **51**, 157-213.

[48] Tong, H. (1993), *Nonlinear Time Series*, Oxford Science.

[49] Tsai, M., and P., Sen (1990), *Tests for Independance in Two-Way Contingency Tables Based on Canonical Analysis,* Calcutta Statist. Assoc. Bull., **40**, 109-123.

[50] Tuckey, J. (1977), *Exploratory Data Analysis*, Addison Wesley.

[51] Wiener, N. (1958), *Nonlinear Problem in Random Theory*, MIT Press, Cambridge.

[52] Wilkinson, J., and C., Reinsch (1971), *Handbook for Automatic Computation*, Vol. 2, Linear Algebra, Springer-Verlag, New-York.

[53] Wong, E., and J., Thomas (1962), *On Polynomial Expansions of the Second Order Distributions,* J. Soc. Ind. Applied Math., **10**, 507-516.

[54] Yoshino, T. (1993), *Introduction to Operator Theory*, Longman Scientific and Technical.

A Proof of Theorem 2.2

We consider the class of optimization problems:

$$\max_{\varphi,\psi} E\left[\varphi(X)\psi(Y)\right], \tag{A.1}$$

$$\text{s.t. } E\varphi^2(X) = E\psi^2(Y) = 1,$$

where $\varphi(X) \in I(X)$, $\psi(Y) \in I(Y)$, $I(X)$, $I(Y)$ are subspaces of $L^2(X)$, $L^2(Y)$ respectively, such that:

$$TI(X) \subset I(Y),\ T^*I(Y) \subset I(X).$$

If $\tilde{\varphi}$, $\tilde{\psi}$ is a pair solution of this problem, $\tilde{\varphi}$ is also a solution of:

$$\max_{\varphi,\psi} E\left[\varphi(X)\tilde{\psi}(Y)\right],$$

$$s.t.\ E\varphi^2(X) = 1, \varphi(X) \in I(X),$$

or equivalently of the problem:

$$\max_{\varphi,\psi} E\left[\varphi(X)T^*\tilde{\psi}(X)\right],$$

$$s.t.\ E\varphi^2(X) = 1, \varphi(X) \in I(X).$$

It is a direct consequence of the Cauchy-Schwarz inequality that $\tilde{\varphi}$ and $T^*\tilde{\psi}$ are proportional (and also are $T\tilde{\varphi}$ and $\tilde{\psi}$). Therefore:

$$\begin{aligned} \exists\tilde{\mu} &: T^*\tilde{\psi} = \tilde{\mu}\tilde{\varphi}, \\ \exists\tilde{v} &: T\tilde{\varphi} = \tilde{v}\tilde{\psi}. \end{aligned}$$

In particular: $\left\langle \tilde{\varphi}, \tilde{\psi} \right\rangle = \left\langle \tilde{\varphi}, T^*\tilde{\psi} \right\rangle = \tilde{\mu} = \left\langle T\tilde{\varphi}, \tilde{\psi} \right\rangle = \tilde{v}$, and:

$$\begin{aligned} TT^*\tilde{\psi} &= \tilde{\mu}^2\tilde{\psi}, \\ T^*T\tilde{\varphi} &= \tilde{\mu}^2\tilde{\varphi}. \end{aligned}$$

The solutions $\tilde{\varphi}$, $\tilde{\psi}$ of the problem (2.1) can be deduced from the simultaneous spectral decomposition of the non-negative auto-adjoint operators TT^* and T^*T. Of course, the pair $\left(\tilde{\varphi}, \tilde{\psi}\right)$ is defined up to a change of sign.

Hence, the solution of the initial problem (2.1) is derived by considering a sequence of problems of type $(A.1)$, with first $I(X) = L^2(X)$, $I(Y) = L^2(Y)$, then $I(X)$, $I(Y)$ the subspaces orthogonal to the solutions of the first problem, and so on.

B Proof of Theorem 3.5

For any function g in $\mathcal{G} = \{g : \int\int |g(x,y)| f(x,y)\, dxdy \leq 1\}$, we get:

$$\sup_{g \in \mathcal{G}} \left| \int \int g(x,y) \hat{f}_N(x,y)\, dxdy - \int \int g(x,y) f(x,y)\, dxdy \right|$$
$$\leq \sup_{(x,y) \in \mathcal{X}^2} \frac{\left| \hat{f}_N(x,y) - f(x,y) \right|}{f(x,y)}.$$

Using assumption A.9, this uniform convergence is equivalent to the uniform convergence of $\hat{f}_N(x,y)$ to $f(x,y)$ on $\mathcal{X}^2$, i.e.:

$$\sup_{(x,y) \in \mathcal{X}^2} \left| \hat{f}_N(x,y) - f(x,y) \right| \to 0 \ a.s,$$

given by Theorem 3.4.

C Asymptotic Properties

C.1 Proof of Theorem 3.8

We write the proof in the case $i = 1$. The case $i \neq 1$ can be easily deduced using the same type of arguments. Let us introduce the three following maximization problems:

Problem 1 $\left(\hat{\varphi}_{1,N}, \hat{\psi}_{1,N}\right)$ solution of:

$$\max_{\varphi,\psi} \int\int \varphi(x)\,\psi(y)\,\hat{f}_N(x,y)\,dxdy,$$

$$\text{s.t.} \quad \int \varphi^2(x)\,\hat{f}_N(x,.)\,dx = \int \psi^2(y)\,\hat{f}_N(.,y)\,dy = 1.$$

Problem 2 $\left(\tilde{\varphi}_{1,N}, \tilde{\psi}_{1,N}\right)$ solution of:

$$\max_{\varphi,\psi} \int\int \varphi(x)\,\psi(y)\,\hat{f}_N(x,y)\,dxdy,$$

$$\text{s.t.} \quad \int \varphi^2(x)\,f(x,.)\,dx = \int \psi^2(y)\,f(.,y)\,dy = 1.$$

Problem 3 (φ_1, ψ_1) solution of:

$$\max_{\varphi,\psi} \int\int \varphi(x)\,\psi(y)\,f(x,y)\,dxdy,$$

$$\text{s.t.} \quad \int \varphi^2(x)\,f(x,.)\,dx = \int \psi^2(y)\,f(.,y)\,dy = 1.$$

Let us denote $\mathcal{H}$ the compact set:

$$\mathcal{H} = \left((\varphi,\psi) : \int \varphi^2(x)\,f(x,.)\,dx = \int \psi^2(y)\,f(.,y)\,dy = 1\right).$$

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

Using Cauchy-Schwarz inequality, we have $g(x,y) = \varphi(x)\,\psi(y) \in \mathcal{G}, \forall(\varphi,\psi) \in \mathcal{H}$, and by Lemma 3.5, we deduce the a.s. uniform convergence:

$$\int\int \varphi(x)\,\psi(y)\,\hat{f}_N(x,y)\,dxdy \underset{N\to\infty}{\to} \int\int \varphi(x)\,\psi(y)\,f(x,y)\,dxdy,$$

where $(\varphi,\psi) \in \mathcal{H}$. Moreover, the application:

$$(\varphi,\psi) \to \int\int \varphi(x)\,\psi(y)\,f(x,y)\,dxdy,$$

is continuous from $L^2(X) \times L^2(Y)$ to $\mathbf{R}$. Now, we can use the Jennrich's theorem (see [30, Jennrich (1969)]) to deduce that the solutions of the finite sample problem converge a.s. to the solution of the limit problem:

$$\begin{aligned}
\left\|\tilde{\varphi}_{1,N} - \varphi_1\right\|_2 &\rightarrow 0 \; a.s., \\
\left\|\tilde{\psi}_{1,N} - \psi_1\right\|_2 &\rightarrow 0 \; a.s.,
\end{aligned}$$

where $\|.\|_2$ is the L^2 distance.

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

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 3.5 the equicontinuity property, which implies the two following a.s. convergences:

$$\begin{aligned}
\alpha_N &= \int \tilde{\varphi}_{1,N}^2(x) \hat{f}_N(x,.)\,dx \rightarrow \int \varphi_1^2(x) f(x,.)\,dx = 1, \; a.s., \\
\beta_N &= \int \tilde{\psi}_{1,N}^2(y) \hat{f}_N(.,y)\,dy \rightarrow \int \psi_1^2(y) f(.,y)\,dy = 1, \; a.s.
\end{aligned}$$

The solutions $\left(\tilde{\varphi}_{1,N}, \tilde{\psi}_{1,N}\right)$ and $\left(\hat{\varphi}_{1,N}, \hat{\psi}_{1,N}\right)$ of problems 1 and 2 are proportional up to the terms $\left(\sqrt{\alpha_N}, \sqrt{\beta_N}\right)$. Therefore, we have:

$$\begin{aligned}
\left\|\hat{\varphi}_{1,N} - \varphi_1\right\|_2 &= \left\|\frac{\tilde{\varphi}_{1,N}}{\sqrt{\alpha_N}} - \varphi_1\right\|_2 \\
&\leq \left\|\tilde{\varphi}_{1,N} - \varphi_1\right\|_2 + \left\|\left(1 - \frac{1}{\sqrt{\alpha_N}}\right)\tilde{\varphi}_{1,N}\right\|_2 \\
&= \left\|\tilde{\varphi}_{1,N} - \varphi_1\right\|_2 + \left|1 - \frac{1}{\sqrt{\alpha_N}}\right|,
\end{aligned}$$

and we deduce that $\left\|\hat{\varphi}_{1,N} - \varphi_1\right\|_2 \rightarrow 0 \; a.s.$ An analog computation gives $\left\|\hat{\psi}_{1,N} - \psi_1\right\|_2 \rightarrow 0 \; a.s.$

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

It is a consequence of the equicontinuity property.

C.2 Asymptotic expansion of $\hat{\mu}_{1,N}$, $\hat{\varphi}_{1,N}$

i) Expansion of the first order condition

The equations (3.7) can be written as:

$$\int \left[\varphi_1(y) + \delta\hat{\varphi}_{1,N}(y)\right] \left[c(x,y) + \delta\hat{c}_N(x,y)\right] dy$$
$$= \left(\mu_1 + \delta\hat{\mu}_{1,N}\right)\left[\varphi_1(x) + \delta\hat{\varphi}_{1,N}(x)\right],$$
$$\int \left[\varphi_1(x) + \delta\hat{\varphi}_{1,N}(x)\right]^2 \left[f(x,.) + \delta\hat{f}_N(x,.)\right] = 1,$$

where the differential terms are $\delta\hat{\mu}_{1,N} = \hat{\mu}_{1,N} - \mu_1$, $\delta\hat{\varphi}_{1,N} = \hat{\varphi}_{1,N} - \varphi_1$, $\delta\hat{c}_N = \hat{c}_N - c$ and $\delta\hat{f}_N = \hat{f}_N - f$. A first order expansion provides the system below:

$$\int \delta\hat{\varphi}_{1,N}(y)\, c(x,y)\, dy + \int \varphi_1(y)\, \delta\hat{c}_N(x,y)\, dy \tag{C.1}$$
$$= \delta\hat{\mu}_{1,N}\varphi_1(x) + \mu_1\delta\hat{\varphi}_{1,N}(x),$$
$$\int \varphi_1^2(x)\, \delta\hat{f}_N(x,.)\, dx + 2\int \varphi_1(x)\, f(x,.)\, \delta\hat{\varphi}_{1,N}(x)\, dx = 0. \tag{C.2}$$

Let us denote:

$$\hat{A}_{1,N}(x) = \int \varphi_1(y)\, \delta\hat{c}_N(x,y)\, dy,\ \hat{B}_{1,N} = -\frac{1}{2}\int \varphi_1^2(x)\, \delta\hat{f}_N(x,.)\, dx.$$

The conditions $(C.1)$ and $(C.2)$ become:

$$\delta\hat{\mu}_{1,N}\varphi_1(x) + \mu_1\delta\hat{\varphi}_{1,N}(x) - \int \delta\hat{\varphi}_{1,N}(y)\, c(x,y)\, dy = \hat{A}_{1,N}(x), \tag{C.3}$$

$$\int \varphi_1(x)\, f(x,.)\, \delta\hat{\varphi}_{1,N}(x)\, dx = \hat{B}_{1,N}. \tag{C.4}$$

$\delta\hat{\varphi}_{1,N}(x)$ admits a Fourier expansion on the eigenfunction basis $\varphi_j(x)$, $j \geq 1$,

$$\delta\hat{\varphi}_{1,N}(x) \simeq \sum_{j=1}^{\infty} \hat{b}_{j,N}\varphi_j(x),$$

where the constant term is zero since $\delta\hat{\varphi}_{1,N}$ is asymptotically zero mean. By replacing in system $(C.3) - (C.4)$, we get:

$$\begin{aligned} \hat{A}_{1,N}(x) &= \delta\hat{\mu}_{1,N}\varphi_1(x) + \sum_{j=2}^{\infty} \hat{b}_{j,N}\left(\mu_1 - \mu_j\right)\varphi_j(x), \\ \hat{B}_{1,N} &= \hat{b}_{1,N}. \end{aligned}$$

We deduce the explicit form of the solution:

$$\delta\hat{\mu}_{1,N} = \left\langle \hat{A}_{1,N}, \varphi_1 \right\rangle = \int \hat{A}_{1,N}(x)\,\varphi_1(x)\,f(x,.)\,dx, \tag{C.5}$$

$$\delta\hat{\varphi}_{1,N}(x) = \hat{B}_{1,N}\varphi_1(x) + \sum_{j=2}^{\infty} \frac{\left\langle \hat{A}_{1,N}, \varphi_j \right\rangle}{\mu_1 - \mu_j}\varphi_j(x). \tag{C.6}$$

ii) Asymptotic expansions of $\delta\hat{c}_N(x,y)$ and $\left\langle \hat{A}_{1,N}, \varphi_j \right\rangle$

$$\begin{aligned}
\delta\hat{c}_N(x,y) &= \delta\left[\int \frac{\hat{f}_N(y,z)\,\hat{f}_N(x,z)}{\hat{f}_N(x,.)\,\hat{f}_N(.,z)}dz\right] \\
&\simeq \int \frac{\delta\hat{f}_N(y,z)\,f(x,z)}{f(x,.)\,f(.,z)}dz + \int \frac{f(y,z)\,\delta\hat{f}_N(x,z)}{f(x,.)\,f(.,z)}dz \\
&\quad - \int \frac{f(y,z)\,f(x,z)}{f^2(x,.)\,f(.,z)}\delta\hat{f}_N(x,.)\,dz - \int \frac{f(y,z)\,f(x,z)}{f(x,.)\,f^2(.,z)}\delta\hat{f}_N(.,z)\,dz,
\end{aligned}$$

which allows to simplify $\left\langle \hat{A}_{1,N}, \varphi_j \right\rangle$ as follows:

$$\begin{aligned}
\left\langle \hat{A}_{1,N}, \varphi_j \right\rangle &= \int\int \varphi_1(y)\,\varphi_j(x)\,\delta\hat{c}_N(x,y)\,f(x,.)\,dxdy \\
&\simeq \int\int\int \varphi_1(y)\,\varphi_j(x)\,\frac{f(x,z)}{f(.,z)}\delta\hat{f}_N(y,z)\,dxdydz \\
&\quad + \int\int\int \varphi_1(y)\,\varphi_j(x)\,\frac{f(y,z)}{f(.,z)}\delta\hat{f}_N(x,z)\,dxdydz \\
&\quad - \int\int\int \varphi_1(y)\,\varphi_j(x)\,\frac{f(y,z)\,f(x,z)}{f(x,.)\,f(.,z)}\delta\hat{f}_N(x,.)\,dxdydz \\
&\quad - \int\int\int \varphi_1(y)\,\varphi_j(x)\,\frac{f(y,z)\,f(x,z)}{f^2(.,z)}\delta\hat{f}_N(.,z)\,dxdydz \\
&= \int\int \varphi_1(y)\left[\int \varphi_j(x)\,f(x \mid z)\,dx\right]\delta\hat{f}_N(y,z)\,dydz \\
&\quad + \int\int \varphi_j(x)\left[\int \varphi_1(y)\,f(y \mid z)\,dy\right]\delta\hat{f}_N(x,z)\,dxdz \\
&\quad - \int\int \varphi_1(y)\,\varphi_j(x)\left[\int \frac{f(y,z)\,f(x,z)}{f(x,.)\,f(.,z)}dz\right]\delta\hat{f}_N(x,.)\,dxdy \\
&\quad - \int\left[\int \varphi_j(x)\,\frac{f(x,z)}{f(.,z)}dx\right]\left[\int \varphi_1(y)\,\frac{f(y,z)}{f(.,z)}dy\right]\delta\hat{f}_N(.,z)\,dz \\
&= \mu_j^{\frac{1}{2}}\int\int \varphi_1(x)\,\psi_j(y)\,\delta\hat{f}_N(x,y)\,dxdy \\
&\quad + \mu_1^{\frac{1}{2}}\int\int \varphi_j(x)\,\psi_1(y)\,\delta\hat{f}_N(x,y)\,dxdy
\end{aligned}$$

$$
\begin{aligned}
& -\mu_1 \int \varphi_1(x)\,\varphi_j(x)\,\delta\hat{f}_N(x,.)\,dx \\
& -\mu_1^{\frac{1}{2}}\mu_j^{\frac{1}{2}} \int \psi_1(y)\,\psi_j(y)\,\delta\hat{f}_N(.,y)\,dy \\
= & \int\int \Big[\mu_j^{\frac{1}{2}}\varphi_1(x)\,\psi_j(y) + \mu_1^{\frac{1}{2}}\varphi_j(x)\,\psi_1(y) \\
& -\mu_1\varphi_1(x)\,\varphi_j(x) - \mu_1^{\frac{1}{2}}\mu_j^{\frac{1}{2}}\psi_1(y)\,\psi_j(y)\Big]\,\delta\hat{f}_N(x,y)\,dxdy.
\end{aligned}
$$

In particular:

$$
\begin{aligned}
\left\langle \hat{A}_{1,N},\varphi_1\right\rangle \simeq\ & \mu_1^{\frac{1}{2}}\int\int [\varphi_1(x)\,\psi_1(y) + \psi_1(x)\,\varphi_1(y) \\
& -\mu_1^{\frac{1}{2}}\varphi_1^2(x) - \mu_1^{\frac{1}{2}}\psi_1^2(y)\Big]\,\delta\hat{f}_N(x,y)\,dxdy.
\end{aligned}
$$

A similar computation provides:

$$
\begin{aligned}
\hat{A}_{1,N}(x) \simeq\ & \frac{1}{f(x,.)}\Big(\int\int \varphi_1(y)\,f(x\mid z)\,\delta\hat{f}_N(y,z)\,dydz + \mu_1^{\frac{1}{2}}\int \psi_1(z)\,\delta\hat{f}_N(x,z)\,dz \\
& -\mu_1\varphi_1(x)\,\delta\hat{f}_N(x,.) - \mu_1^{\frac{1}{2}}\int \psi_1(z)\,f(x\mid z)\,\delta\hat{f}_N(.,z)\,dz\Big) \\
\simeq\ & \frac{1}{f(x,.)}\left(\mu_1^{\frac{1}{2}}\int \psi_1(z)\,\delta\hat{f}_N(x,z)\,dz - \mu_1\varphi_1(x)\,\delta\hat{f}_N(x,.)\right), \qquad \text{(C.7)}
\end{aligned}
$$

after the elimination of the terms of higher order.

iii) Convergence rates of the estimators

Finally, we can note that the estimators of the canonical correlations and canonical variates do not converge at the same rate. Indeed, from theorem 3.7, we deduce that $\hat{B}_{1,N}$ and $\left\langle \hat{A}_{1,N},\varphi_1\right\rangle$ converge at rate $N^{-\frac{1}{2}}$, whereas $\hat{A}_{1,N}(x)$ converge at rate $N^{-\frac{1}{2}}h_{1N}^{-\frac{d}{2}}$. We deduce from $(C.3)$ and the comparison of rates that:

$$
\sqrt{Nh_{1N}^d}\,\delta\hat{\varphi}_{1,N}(x) \simeq \frac{1}{\mu_1}\sqrt{Nh_{1N}^d}\,\hat{A}_{1,N}(x).
$$

C.3 Asymptotic distribution of $\hat{\mu}_{i,N}$, $\hat{\varphi}_{i,N}$, $\hat{\psi}_{i,N}$

i) Asymptotic distribution of $\hat{\mu}_{i,N}$

From the expansion $\delta\hat{\mu}_{1,N} \simeq \left\langle \hat{A}_{1,N}, \varphi_1 \right\rangle$, we deduce by theorem 3.7 that:

$$\sqrt{N}\delta\hat{\mu}_{1,N} \xrightarrow{d} \mathcal{N}\left(0, V_1\right),$$

where

$$\begin{aligned} V_1 &= \mu_1 \sum_{k=-\infty}^{\infty} Cov\left[\varphi_1\left(X_n\right)\psi_1\left(Y_n\right) + \psi_1\left(X_n\right)\varphi_1\left(Y_n\right)\right. \\ &\quad -\mu_1^{\frac{1}{2}}\varphi_1^2\left(X_n\right) - \mu_1^{\frac{1}{2}}\psi_1^2\left(Y_n\right), \varphi_1\left(X_{n+k}\right)\psi_1\left(Y_{n+k}\right) + \\ &\quad \left.\psi_1\left(X_{n+k}\right)\varphi_1\left(Y_{n+k}\right) - \mu_1^{\frac{1}{2}}\varphi_1^2\left(X_{n+k}\right) - \mu_1^{\frac{1}{2}}\psi_1^2\left(Y_{n+k}\right)\right]. \end{aligned}$$

ii) Asymptotic distribution of $\hat{\varphi}_{i,N}$

From the expansion $\delta\hat{\varphi}_{1,N}\left(x\right) \simeq \frac{1}{\mu_1}\hat{A}_{1,N}\left(x\right)$, we deduce:

$$\delta\hat{\varphi}_{1,N}\left(x\right) \simeq \frac{1}{\mu_1^{\frac{1}{2}}}\frac{1}{f\left(x,.\right)}\int\left[\psi_1\left(y\right) - \mu_1^{\frac{1}{2}}\varphi_1\left(x\right)\right]\delta\hat{f}_N\left(x,y\right)dy,$$

and, by theorem 3.7:

$$\sqrt{Nh_{1N}^d}\left(\hat{\varphi}_{1,N}\left(x\right) - \varphi_1\left(x\right)\right) \xrightarrow{d} \mathcal{N}\left(0, W_1\left(x\right)\right),$$

where

$$\begin{aligned} W_1\left(x\right) &= \int K_1^2\left(u\right)du\,\frac{1}{\mu_1}\frac{1}{f^2\left(x,.\right)}\int\left[\psi_1\left(y\right) - \mu_1^{\frac{1}{2}}\varphi_1\left(x\right)\right]^2 f\left(x,y\right)dy, \\ &= \frac{1}{\mu_1}\frac{1}{f\left(x,.\right)}\int K_1^2\left(u\right)du\left(\int\psi_1^2\left(y\right)f\left(y\mid x\right)dy - \mu_1\varphi_1^2\left(x\right)\right), \\ &= \frac{1}{\mu_1}\frac{1}{f\left(x,.\right)}\int K_1^2\left(u\right)du\,V\left[\psi_1\left(Y\right)\mid X = x\right]. \end{aligned}$$

iii) Asymptotic distribution of $\hat{\psi}_{i,N}$

The result is immediate by symmetry.

C.4 Asymptotic expansion of $\hat{\lambda}_{1,N}^{R}$, $\hat{\varphi}_{1,N}^{R}$

$i)$ Expansion of the first order condition

We can directly consider the pair (λ_1, φ_1) and its estimated counterpart $\left(\hat{\lambda}_{1,N}, \hat{\varphi}_{1,N}\right)$. The pair (λ_1, φ_1) satisfies:

$$\int \varphi_1(x) f(x \mid y)\, dx = \lambda_1 \varphi_1(y), \tag{C.8}$$

$$\int \varphi_1^2(x) f(x,.)\, dx = 1,$$

where $f(x \mid y) = f(x,y)/f(.,y)$. Similarly, the pair $\left(\hat{\lambda}_{1,N}^{R}, \hat{\varphi}_{1,N}^{R}\right)$ satisfies:

$$\int \hat{\varphi}_{1,N}^{R}(x) \hat{f}_N^R(x \mid y)\, dx = \hat{\lambda}_{1,N}^{R} \hat{\varphi}_{1,N}^{R}(y), \tag{C.9}$$

$$\int \hat{\varphi}_{1,N}^{2}(x) \hat{f}_N^R(x,.)\, dx = 1,$$

where $\hat{f}_N^R(x \mid y) = \hat{f}_N^R(x,y)/\hat{f}_N^R(.,y)$. The two previous equations can be written as:

$$\int \left[\varphi_1(x) + \delta\hat{\varphi}_{1,N}^{R}(x)\right]\left[f(x \mid y) + \delta\hat{f}_N^R(x \mid y)\right] dx$$

$$= \left(\lambda_1 + \delta\hat{\lambda}_{1,N}^{R}\right)\left[\varphi_1(y) + \delta\hat{\varphi}_{1,N}^{R}(y)\right],$$

$$\int \left[\varphi_1(x) + \delta\hat{\varphi}_{1,N}(x)\right]^2 \left[f(x,.) + \delta\hat{f}_N^R(x,.)\right] = 1,$$

where the differential terms are $\delta\hat{\lambda}_{1,N}^{R} = \hat{\lambda}_{1,N}^{R} - \lambda_1$, $\delta\hat{\varphi}_{1,N} = \hat{\varphi}_{1,N} - \varphi_1$ and $\delta\hat{f}_N^R = \hat{f}_N^R - f$. A first order expansion provides the system below:

$$\int \delta\hat{\varphi}_{1,N}^{R}(x) f(x \mid y)\, dx + \int \varphi_1(x)\, \delta\hat{f}_N^R(x \mid y)\, dx \tag{C.10}$$

$$= \delta\hat{\lambda}_{1,N}^{R} \varphi_1(y) + \lambda_1 \delta\hat{\varphi}_{1,N}^{R}(y),$$

$$\int \varphi_1^2(x)\, \delta\hat{f}_N^R(x,.)\, dx + 2\int \varphi_1(x) f(x,.)\, \delta\hat{\varphi}_{1,N}^{R}(x)\, dx = 0. \tag{C.11}$$

Let us denote:

$$\hat{A}_{1,N}^{R}(y) = \int \varphi_1(x)\, \delta\hat{f}_N^R(x \mid y)\, dx, \ \hat{B}_{1,N}^{R} = -\frac{1}{2}\int \varphi_1^2(x)\, \delta\hat{f}_N^R(x,.)\, dx.$$

The conditions $(C.10)$ and $(C.11)$ become:

$$\delta\hat{\lambda}_{1,N}^{R} \varphi_1(y) + \lambda_1 \delta\hat{\varphi}_{1,N}^{R}(y) - \int \delta\hat{\varphi}_{1,N}^{R}(x) f(x \mid y)\, dx = \hat{A}_{1,N}^{R}(y), \tag{C.12}$$

$$\int \varphi_1(x) f(x,.)\, \delta\hat{\varphi}^R_{1,N}(x)\, dx = \hat{B}^R_{1,N}. \qquad \text{(C.13)}$$

Let us now multiply both sides of equation $(C.12)$ by $\varphi_1(y) f(.,y)$ and integrate with respect to y, we obtain:

$$\begin{aligned}\left\langle \hat{A}^R_{1,N}, \varphi_1 \right\rangle &= \delta\hat{\lambda}^R_{1,N} + \lambda_1 \int \varphi_1(y)\, \delta\hat{\varphi}^R_{1,N}(y) f(.,y)\, dy \\ &\quad - \int\int \delta\hat{\varphi}^R_{1,N}(x)\, \varphi_1(y) f(x \mid y) f(.,y)\, dxdy.\end{aligned}$$

By using now the reversibility property, we get:

$$\begin{aligned}&\int\int \delta\hat{\varphi}^R_{1,N}(x)\, \varphi_1(y) f(x \mid y) f(.,y)\, dxdy \\ =\ &\int\int \delta\hat{\varphi}^R_{1,N}(x)\, \varphi_1(y) f(y \mid x) f(x,.)\, dxdy \\ =\ &\lambda_1 \int \delta\hat{\varphi}^R_{1,N}(x)\, \varphi_1(x) f(x,.)\, dx,\end{aligned}$$

and we conclude that:

$$\delta\hat{\lambda}^R_{1,N} \simeq \left\langle \hat{A}^R_{1,N}, \varphi_1 \right\rangle = \int\int \varphi_1(x)\, \varphi_1(y)\, \delta\hat{f}^R_N(x \mid y) f(.,y)\, dxdy. \qquad \text{(C.14)}$$

ii) Asymptotic expansion of $\delta\hat{f}^R_N(x \mid y)$

We get: -

$$\delta\hat{f}^R_N(x \mid y) \simeq \frac{\delta\hat{f}^R_N(x,y)}{f(.,y)} - \frac{\delta\hat{f}^R_N(.,y)}{f^2(.,y)} f(x,y),$$

and we deduce:

$$\begin{aligned}\hat{A}^R_{1,N}(y) &\simeq \frac{1}{f(.,y)} \left[\int \varphi_1(x)\, \delta\hat{f}^R_N(x,y)\, dx - \delta\hat{f}^R_N(.,y) \int \varphi_1(x) f(x \mid y)\, dx\right] \\ &= \frac{1}{f(.,y)} \left[\int \varphi_1(x)\, \delta\hat{f}^R_N(x,y)\, dx - \lambda_1 \varphi_1(y)\, \delta\hat{f}^R_N(.,y)\right], \qquad \text{(C.15)}\end{aligned}$$

and, in particular:

$$\delta\hat{\lambda}^R_{1,N} \simeq \left\langle \hat{A}^R_{1,N}, \varphi_1 \right\rangle = \int\int \left[\varphi_1(x)\, \varphi_1(y) - \lambda_1 \varphi_1^2(y)\right] \delta\hat{f}^R_N(x,y)\, dxdy.$$

iii) Convergence rates of the estimators

As in the non reversible case, $\hat{A}^R_{1,N}(y)$ tends to zero at rate $N^{-\frac{1}{2}} h_{1N}^{-\frac{d}{2}}$, whereas $\left\langle \hat{A}^R_{1,N}, \varphi_1 \right\rangle$ and $\hat{B}^R_{1,N}$ tend to zero at rate $N^{-\frac{1}{2}}$. We conclude that:

$$\sqrt{N h^d_{1N}}\, \delta\hat{\varphi}^R_{1,N}(y) \simeq \frac{1}{\lambda_1} \sqrt{N h^d_{1N}}\, \hat{A}^R_{1,N}(y).$$

C.5 Asymptotic distribution of $\hat{\lambda}^R_{i,N}$, $\hat{\varphi}^R_{i,N}$

$i)$ **Asymptotic distribution of** $\hat{\lambda}^R_{i,N}$

From the expansion of $\delta\hat{\lambda}^R_{1,N} \simeq \left\langle \hat{A}^R_{1,N}, \varphi_1 \right\rangle$, we deduce by theorem 3.7 that:

$$\sqrt{N}\delta\hat{\lambda}^R_{1,N} \xrightarrow{d} \mathcal{N}\left(0, V^R_1\right),$$

where

$$\begin{aligned} V^R_1 &= \frac{1}{2}\sum_{k=-\infty}^{\infty} Cov\Big[\varphi_1(X_n)\varphi_1(Y_n) - \lambda_1\varphi_1^2(Y_n), \\ &\qquad \varphi_1(X_{n+k})\varphi_1(Y_{n+k}) - \lambda_1^2\varphi_1(Y_{n+k})\Big]. \end{aligned}$$

$ii)$ **Asymptotic distribution of** $\hat{\varphi}^R_{i,N}$

From the expansion $\delta\hat{\varphi}_{1,N}(x) \simeq \frac{1}{\lambda_1}\hat{A}_{1,N}(x)$, we deduce:

$$\delta\hat{\varphi}^R_{1,N}(x) \simeq \frac{1}{\lambda_1}\frac{1}{f(x,.)}\int[\varphi_1(y) - \lambda_1\varphi_1(x)]\,\delta\hat{f}^R_N(x,y)\,dy,$$

and, by theorem 3.7, that:

$$\sqrt{Nh^d_N}\left(\hat{\varphi}^R_{i,N}(x) - \varphi_i(x)\right) \xrightarrow{d} \mathcal{N}\left(0, W^R_i(x)\right),$$

where

$$\begin{aligned} W^R_1(x) &= \frac{1}{2}\frac{1}{\lambda_1^2}\frac{1}{f^2(x,.)}\int K_1^2(u)\,du\int[\varphi_1(y) - \lambda_1\varphi_1(x)]^2 f(x,y)\,dy \\ &= \frac{1}{2}\frac{1}{\lambda_1^2}\frac{1}{f(x,.)}\int K_1^2(u)\,du\left(\int\varphi_1^2(y) f(y \mid x)\,dy - \lambda_1^2\varphi_1^2(x)\right) \\ &= \frac{1}{2}\frac{1}{\lambda_1^2}\frac{1}{f(x,.)}\int K_1^2(u)\,du\,V[\varphi_1(Y) \mid X = x]. \end{aligned}$$