

Canonical correlation analysis for functional data

Mirosław Krzyśko, Łukasz Waszak

Faculty of Mathematics and Computer Science, Adam Mickiewicz University,
Umultowska 87, 61-614 Poznań, Poland
e-mail: mkrzysko@amu.edu.pl, lwaszak@amu.edu.pl

SUMMARY

Classical canonical correlation analysis seeks the associations between two data sets, i.e. it searches for linear combinations of the original variables having maximal correlation. Our task is to maximize this correlation, and is equivalent to solving a generalized eigenvalue problem. The maximal correlation coefficient (being a solution of this problem) is the first canonical correlation coefficient. In this paper we propose a new method of constructing canonical correlations and canonical variables for a pair of stochastic processes represented by a finite number of orthonormal basis functions.

Key words: functional data, orthonormal basis, stochastic processes, canonical correlation analysis

1. Introduction

In classical canonical correlation analysis we are interested in the relationship between two random vectors \mathbf{Y} and \mathbf{X} . We seek weight vectors \mathbf{u} and \mathbf{v} such that the linear combinations of vectors $\mathbf{u}'\mathbf{Y}$ and $\mathbf{v}'\mathbf{X}$, called canonical variables, are maximally correlated with each other. In the functional case we are interested in the relationship between two stochastic processes $Y(t)$ and $X(t)$. We seek weight functions $u(t)$ and $v(t)$ such that $\int uY$ and $\int vX$ are maximally correlated (all integrals are definite integrals on the interval $[0, T]$). The values $\int uY$ and $\int vX$ are called functional canonical variables. The problems of functional canonical correlation analysis were first raised in the work of Leurgans, Moyeed and Silverman (1993), and are presented in a more mature form in the monograph of Ramsay and Silverman (2005). Canonical correlations, weight functions and functional canonical variables are constructed on the basis of N independent realizations of the stochastic processes $Y(t)$ and $X(t)$. These are called functional data, after Ramsay and Danzell (1991). In practice we have available discrete data in the form of time series, and it is necessary to transform these

to functional data. In section 2 of this work a method of transforming discrete data to functional data is presented. Section 3 shows the relationship between the techniques of canonical correlation analysis for random vectors and for random processes. Section 4 is devoted to canonical correlations, weight functions and functional canonical variables constructed on the basis of functional data. In the final section of the work a concrete real-life example is given.

2. Transformation of discrete data to functional data

Many financial, meteorological and other data are recorded at discrete moments in time. Let $\{y_{ij}, x_{ij}\}$ denote an observed value of a pair of statistical features Y and X on the i th individual at the j th time point, where $i = 1, 2, \dots, N$, $j = 1, 2, \dots, J_i$. Our data then consist of N triples $\{t_{ij}, y_{ij}, x_{ij}\}$, where $t_{ij} \in [0, T]$, $i = 1, 2, \dots, N$, $j = 1, 2, \dots, J_i$. Actually, we have N two-dimensional time series, each observed at a specific sequence of time points not necessarily of the same length. However in many cases it is more convenient to use continuous functions of time, $y(t)$ and $x(t)$, $t \in [0, T]$, i.e. functional data. In this case the discrete data $\{t_{ij}, y_{ij}, x_{ij}\}$ are transformed to functional data $\{y_i(t), x_i(t), t \in [0, T]\}$. The transformation process is identical for all functions $y_i(t)$ and $x_i(t)$, hence our further considerations will concern a single function $y(t)$ and $x(t)$, $t \in [0, T]$. The process of transformation of discrete to continuous data requires smoothing (Ramsey and Silverman, 2005). One of the methods of smoothing is to present the function $y(t)$ as a linear combination of a finite number $K_1 + 1$ of orthonormal basis functions φ_k and to present the function $x(t)$ as a linear combination of a finite number $K_2 + 1$ of orthonormal basis functions ψ_k :

$$y(t) = \sum_{k=0}^{K_1} \alpha_k \varphi_k(t), \quad t \in [0, T_1], \quad x(t) = \sum_{l=0}^{K_2} \beta_l \psi_l(t), \quad t \in [0, T_2]. \quad (1)$$

The orthonormal basis functions usually chosen are the Fourier system:

$$\varphi_0(t) = 1, \quad \varphi_{2k-1}(t) = \sqrt{\frac{2}{T}} \sin \frac{2\pi kt}{T}, \quad \varphi_{2k}(t) = \sqrt{\frac{2}{T}} \cos \frac{2\pi kt}{T},$$

for $k = 1, 2, \dots, t \in [0, T]$, or a system of Legendre polynomials over the space $L_2([-1, 1])$:

$$\tilde{p}_k(x) = \sqrt{\frac{2k+1}{2}} p_k(x),$$

where $p_{k+1}(x) = \frac{1}{k+1}[(2k+1)xp_k(x) - kp_{k-1}(x)]$ for $k \geq 1$, $p_0(x) = 1$ and $p_1(x) = x$.

Every finite interval $[a, b]$ can be transformed to the interval $[-1, 1]$ using the substitution

$$x = \frac{2}{b-a}t - \frac{b+a}{b-a}, \quad t \in [a, b], \quad x \in [-1, 1].$$

Let $\mathbf{y} = (y_1, y_2, \dots, y_{J^*})'$, $\mathbf{x} = (x_1, x_2, \dots, x_{J^{**}})'$, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \dots, \alpha_{K_1})'$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_{K_2})'$ and let Φ be a matrix of dimension $J^* \times (K_1 + 1)$ containing the values $\varphi_k(t_j)$, $k = 0, 1, \dots, K_1$, $j = 1, 2, \dots, J^*$, and let Ψ be a matrix of dimension $J^{**} \times (K_2 + 1)$ containing the values $\psi_l(t_j)$, $l = 0, 1, \dots, K_2$, $j = 1, 2, \dots, J^{**}$.

The coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in the expansion of (1) are selected by the least squares method, that is, so as to minimize the functions:

$$S(\boldsymbol{\alpha}) = (\mathbf{y} - \Phi\boldsymbol{\alpha})'(\mathbf{y} - \Phi\boldsymbol{\alpha}),$$

or

$$S(\boldsymbol{\beta}) = (\mathbf{x} - \Psi\boldsymbol{\beta})'(\mathbf{x} - \Psi\boldsymbol{\beta}).$$

Differentiating $S(\boldsymbol{\alpha})$ with respect to the vector $\boldsymbol{\alpha}$ and differentiating $S(\boldsymbol{\beta})$ with respect to the vector $\boldsymbol{\beta}$, we obtain

$$\hat{\boldsymbol{\alpha}} = (\Phi'\Phi)^{-1} \Phi'\mathbf{y}, \quad \hat{\boldsymbol{\beta}} = (\Psi'\Psi)^{-1} \Psi'\mathbf{x}. \quad (2)$$

The degree of smoothness of the function $y(t)$ depends on the value of K_1 , because small (large) values of K_1 cause more (less) smoothing of the curves. The optimum value of K_1 is selected for each function $y_i(t)$ using the Bayesian Information Criterion (BIC), and then from the values of K_1 corresponding to all functions a modal value is selected, as the common value for all $y_i(t)$, $i = 1, 2, \dots, N$. The same applies to $x(t)$. The BIC measures the exactness of fit (see for example Shmueli (2010)). The BIC value for $y(t)$ is expressed by the following formula:

$$BIC(y(t)) = \ln \left(\frac{\mathbf{e}'\mathbf{e}}{2} \right) + (K_1 + 1) \left(\frac{\ln J^*}{J^*} \right),$$

where $\mathbf{e} = (e_1, \dots, e_{J^*})'$, $e_j = y_j - \sum_{k=0}^{K_1} \hat{\alpha}_k \varphi_k(t_j)$, $j = 1, 2, \dots, J^*$.

We compute the BIC value for $x(t)$ in an analogous manner.

3. Canonical correlation analysis for random vectors and random processes

Let us assume that we observe the realization of a two-dimensional random process (Y, X) , where $Y(t) \in L_2(I_1)$ and $X(t) \in L_2(I_2)$. Here $L_2(I_1)$ and $L_2(I_2)$ are the spaces of square integrable functions on the intervals I_1 and I_2 , equipped with the inner product

$$\langle u, v \rangle = \int u(t)v(t)dt.$$

We further assume that $E(\langle Z, Z \rangle) < \infty$ for $Z = Y$ or $Z = X$.

Canonical correlation for finite-dimensional random vectors $\mathbf{Y} \in \mathbb{R}^{K_1+1}$, $\mathbf{X} \in \mathbb{R}^{K_2+1}$ and for stochastic processes $Y(t) \in L_2(I_1)$, $X(t) \in L_2(I_2)$ is defined as follows. Let $H_i = \mathbb{R}^{K_i+1}$ in the vector case and $H_i = L_2(I_i)$ in the functional case, $i = 1, 2$. Then the first canonical correlation ρ_1 and the associated vectors u_1 and v_1 or weight functions $u_1(t)$ and $v_1(t)$ are defined as follows:

$$\rho_1 = \sup_{u \in H_1, v \in H_2} Cov(\langle u, Y \rangle, \langle v, X \rangle) = Cov(\langle u_1, Y \rangle, \langle v_1, X \rangle), \quad (3)$$

where u and v are subject to the restrictions

$$Var(\langle u, Y \rangle) = 1, \quad Var(\langle v, X \rangle) = 1. \quad (4)$$

In general, the k th canonical correlation ρ_k and the associated vectors u_k and v_k or weight functions $u_k(t)$ and $v_k(t)$ are defined as follows:

$$\rho_k = \sup_{u \in H_1, v \in H_2} Cov(\langle u, Y \rangle, \langle v, X \rangle) = Cov(\langle u_k, Y \rangle, \langle v_k, X \rangle), \quad (5)$$

where u and v are subject to the restrictions of (4), and the k th pair of canonical variables (U_k, V_k) is not correlated with the first $k - 1$ pairs $\{(U_i, V_i), i = 1, 2, \dots, k - 1\}$, where

$$U_k = \langle u_k, Y \rangle, \quad V_k = \langle v_k, X \rangle. \quad (6)$$

The expression (ρ_k, u_k, v_k) will be called the k th canonical system.

We will say that Y and X are uncorrelated if all of their canonical correlations are equal to zero. This is equivalent to the equality $\rho_1 = 0$, because $\rho_1 \geq \rho_2 \geq \dots \geq 0$.

Considering the case where Y and X are stochastic processes, we will assume that Y and X can be represented by a finite number of orthonormal

basis functions. For such finite-dimensional processes, functional canonical analysis is equivalent to ordinary canonical analysis for random vector coefficients. To demonstrate this fact, we adopt the following notation. Let

$$Y(t) = \sum_{k=0}^{K_1} \alpha_k \varphi_k(t), \quad t \in I_1, \quad X(t) = \sum_{l=0}^{K_2} \beta_l \psi_l(t), \quad t \in I_2, \quad (7)$$

where $\{\varphi_k\}$ and $\{\psi_k\}$ are the first $(K_i + 1)$, $(i = 1, 2)$ elements of the bases of the spaces $L_2(I_i)$, and $\{\alpha_k\}$ and $\{\beta_k\}$ are random variables with zero expected values and finite variances. This means that $E[Y(t)] = E[X(t)] = 0$. This fact does not lead to loss of generality, because the canonical systems are determined on the basis of the covariances of the processes $Y(t)$ and $X(t)$ which do not depend on their means. Let

$$\boldsymbol{\varphi}(t) = (\varphi_0(t), \varphi_1(t), \dots, \varphi_{K_1}(t))', \quad (8)$$

$$\boldsymbol{\psi}(t) = (\psi_0(t), \psi_1(t), \dots, \psi_{K_2}(t))', \quad (9)$$

$$\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \dots, \alpha_{K_1})', \quad (10)$$

$$\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_{K_2})', \quad (11)$$

where $E(\boldsymbol{\alpha}) = \mathbf{0}$, $E(\boldsymbol{\beta}) = \mathbf{0}$,

$$\text{Var}(\boldsymbol{\alpha}) = \boldsymbol{\Sigma}_{11}, \quad \text{Var}(\boldsymbol{\beta}) = \boldsymbol{\Sigma}_{22}, \quad \text{Cov}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = E(\boldsymbol{\alpha}\boldsymbol{\beta}') = \boldsymbol{\Sigma}_{12}. \quad (12)$$

Then the processes $Y(t)$ and $X(t)$ can be written in the form

$$Y(t) = \boldsymbol{\alpha}'\boldsymbol{\varphi}(t), \quad X(t) = \boldsymbol{\beta}'\boldsymbol{\psi}(t). \quad (13)$$

Theorem 1. *The k th canonical system $(\gamma_k, \mathbf{u}_k, \mathbf{v}_k)$ of the pair of random vectors $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is related to the k th canonical system $(\rho_k, u_k(t), v_k(t))$ of the pair of random processes $(Y(t), X(t))$ by the equations*

$$\rho_k = \gamma_k, \quad u_k(t) = \mathbf{u}'_k \boldsymbol{\varphi}(t), \quad v_k(t) = \mathbf{v}'_k \boldsymbol{\psi}(t). \quad (14)$$

Proof. The covariance function of the process $Y(t)$ is equal to

$$r_{YY}(s, t) = E[Y(s)Y(t)] = \boldsymbol{\varphi}'(s)E(\boldsymbol{\alpha}\boldsymbol{\alpha}')\boldsymbol{\varphi}(t) = \boldsymbol{\varphi}'(s)\boldsymbol{\Sigma}_{11}\boldsymbol{\varphi}(t). \quad (15)$$

Similarly we have

$$r_{XX}(s, t) = \boldsymbol{\psi}'(s)\boldsymbol{\Sigma}_{22}\boldsymbol{\psi}(t), \quad r_{YX}(s, t) = \boldsymbol{\varphi}'(s)\boldsymbol{\Sigma}_{12}\boldsymbol{\psi}(t) \quad (16)$$

Without loss of generality we may assume that the covariance matrices $\boldsymbol{\Sigma}_{11}$ and $\boldsymbol{\Sigma}_{22}$ are of full rank. We notice that every function $u(t) \in L_2(I_1)$

can be written as the sum of two components $u_0 \in \text{span}\{\varphi_0, \dots, \varphi_{K_1}\}$ and $u_1 \in \text{span}\{\varphi_0, \dots, \varphi_{K_1}\}^\perp$, namely

$$u(t) = u_0(t) + u_1(t) = \mathbf{u}'\boldsymbol{\varphi}(t) + u_1(t)$$

where $\mathbf{u} \in \mathbb{R}^{K_1+1}$, (Schott (2005), p. 51). Then

$$\begin{aligned} \langle u, Y \rangle &= \langle u_0 + u_1, \boldsymbol{\alpha}'\boldsymbol{\varphi} \rangle = \langle \mathbf{u}'\boldsymbol{\varphi}, \boldsymbol{\alpha}'\boldsymbol{\varphi} \rangle = \mathbf{u}' \langle \boldsymbol{\varphi}, \boldsymbol{\varphi}' \rangle \boldsymbol{\alpha} = \\ &= \mathbf{u}' \mathbf{I}_{K_1+1} \boldsymbol{\alpha} = \mathbf{u}'\boldsymbol{\alpha}, \end{aligned}$$

where \mathbf{I}_{K_1+1} is an identity matrix of order $K_1 + 1$.

Similarly for $v(t) \in L_2(I_2)$, $\langle v, X \rangle = \mathbf{v}'\boldsymbol{\beta}$, where $\mathbf{v} \in \mathbb{R}^{K_2+1}$. Hence

$$\begin{aligned} E(\langle u, Y \rangle) &= \mathbf{u}'E(\boldsymbol{\alpha}) = 0, \\ E(\langle v, X \rangle) &= \mathbf{v}'E(\boldsymbol{\beta}) = 0, \\ \text{Var}(\langle u, Y \rangle) &= E(\mathbf{u}'\boldsymbol{\alpha}(\mathbf{u}'\boldsymbol{\alpha})') = \mathbf{u}'E(\boldsymbol{\alpha}\boldsymbol{\alpha}')\mathbf{u} = \mathbf{u}'\boldsymbol{\Sigma}_{11}\mathbf{u}, \\ \text{Var}(\langle v, X \rangle) &= E(\mathbf{v}'\boldsymbol{\beta}(\mathbf{v}'\boldsymbol{\beta})') = \mathbf{v}'E(\boldsymbol{\beta}\boldsymbol{\beta}')\mathbf{v} = \mathbf{v}'\boldsymbol{\Sigma}_{22}\mathbf{v}, \\ \text{Cov}(\langle u, Y \rangle, \langle v, X \rangle) &= E(\mathbf{u}'\boldsymbol{\alpha}\mathbf{v}'\boldsymbol{\beta}) = \mathbf{u}'E(\boldsymbol{\alpha}\boldsymbol{\beta}')\mathbf{v} = \mathbf{u}'\boldsymbol{\Sigma}_{12}\mathbf{v}. \end{aligned}$$

Let us consider the first canonical correlation between the processes $Y(t)$ and $X(t)$:

$$\rho_1 = \sup_{u \in L_1(I_1), v \in L_2(I_2)} \text{Cov}(\langle u, Y \rangle, \langle v, X \rangle) = \text{Cov}(\langle u_1, Y \rangle, \langle v_1, X \rangle),$$

subject to the restriction

$$\text{Var}(\langle u, Y \rangle) = 1, \quad \text{Var}(\langle v, X \rangle) = 1.$$

This is equivalent to saying that

$$\rho_1 = \sup_{\mathbf{u} \in \mathbb{R}^{K_1+1}, \mathbf{v} \in \mathbb{R}^{K_2+1}} \mathbf{u}'\boldsymbol{\Sigma}_{12}\mathbf{v} = \mathbf{u}'_1\boldsymbol{\Sigma}_{12}\mathbf{v}_1,$$

subject to the restriction

$$\mathbf{u}'\boldsymbol{\Sigma}_{11}\mathbf{u} = 1, \quad \mathbf{v}'\boldsymbol{\Sigma}_{22}\mathbf{v} = 1.$$

This is the definition of the first canonical correlation between the random vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. On the other hand, if we begin with the first canonical system $(\rho_1, \mathbf{u}_1, \mathbf{v}_1)$ between random vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, we will obtain the first canonical system for the processes $Y(t)$ and $X(t)$ from the equation

$$(\rho_1, u(t), v(t)) = (\rho_1, \mathbf{u}'_1\boldsymbol{\varphi}(t), \mathbf{v}'_1\boldsymbol{\psi}(t)).$$

We may extend these considerations to the second canonical system and so on. \square

We have thus shown that for $1 \leq k \leq \min\{K_1+1, K_2+1\}$ the canonical correlations ρ_k for a pair of random processes $Y(t)$ and $X(t)$ are the same as the canonical correlations for a pair of random vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, and that the weight functions are expressed by the formula

$$u_k(t) = \mathbf{u}'_k \boldsymbol{\varphi}(t), \quad v_k(t) = \mathbf{v}'_k \boldsymbol{\psi}(t),$$

and the canonical variables are of the form

$$U_k = \langle u_k(t), Y(t) \rangle, \quad V_k = \langle v_k(t), X(t) \rangle.$$

Let $\mathbf{C} = \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12}$ and $\mathbf{D} = \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21}$. Then the matrices \mathbf{CD} and \mathbf{DC} have the same nonzero eigenvalues ρ_k^2 , and the corresponding eigenvectors \mathbf{u}_k and \mathbf{v}_k are given by the equations

$$(\mathbf{CD} - \rho_k^2 \mathbf{I}_{p_1}) \mathbf{u}_k = \mathbf{0}, \quad (\mathbf{DC} - \rho_k^2 \mathbf{I}_{p_2}) \mathbf{v}_k = \mathbf{0},$$

$1 \leq k \leq \min\{K_1+1, K_2+1\}$ (see for example Krzyśko (2009)).

4. Canonical correlations from a sample

Canonical correlation analysis for random vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ is based on the matrices $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\Sigma}_{22}$ and $\boldsymbol{\Sigma}_{12}$. In practice these matrices are unknown. We estimate them on the basis of N independent realizations of these vectors, forming the matrices

$$\hat{\mathbf{A}} = \begin{bmatrix} \hat{\alpha}_{10} & \hat{\alpha}_{11} & \dots & \hat{\alpha}_{1K_1} \\ \hat{\alpha}_{20} & \hat{\alpha}_{21} & \dots & \hat{\alpha}_{2K_1} \\ \dots & \dots & \dots & \dots \\ \hat{\alpha}_{N0} & \hat{\alpha}_{N1} & \dots & \hat{\alpha}_{NK_1} \end{bmatrix}, \quad \hat{\mathbf{B}} = \begin{bmatrix} \hat{\beta}_{10} & \hat{\beta}_{11} & \dots & \hat{\beta}_{1K_2} \\ \hat{\beta}_{20} & \hat{\beta}_{21} & \dots & \hat{\beta}_{2K_2} \\ \dots & \dots & \dots & \dots \\ \hat{\beta}_{N0} & \hat{\beta}_{N1} & \dots & \hat{\beta}_{NK_2} \end{bmatrix},$$

where $\hat{\alpha}_{ik}$ and $\hat{\beta}_{il}$ are estimates obtained by the least squares method for the parameters α_{ik} and β_{il} in the representation

$$y_i(t) = \sum_{k=0}^{K_1} \alpha_{ik} \varphi_k(t), \quad x_i(t) = \sum_{l=0}^{K_2} \beta_{il} \psi_l(t).$$

of processes $Y(t)$ and $X(t)$. It should be mentioned that each time series leading to estimates of matrices $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ uses the least squares approximation with the same set of basis functions. The optimal selection of the

number of basis functions is presented at the end of Section 2. Then the estimates of the unknown matrices Σ_{11} , Σ_{22} and Σ_{12} are of the form:

$$\hat{\Sigma}_{11} = \frac{1}{n} \hat{\mathbf{A}}' \hat{\mathbf{A}}, \quad \hat{\Sigma}_{22} = \frac{1}{n} \hat{\mathbf{B}}' \hat{\mathbf{B}}, \quad \hat{\Sigma}_{12} = \frac{1}{n} \hat{\mathbf{A}}' \hat{\mathbf{B}}.$$

If $N > K_1 + 1$, the matrix $\hat{\Sigma}_{11}$ is positive definite with probability 1. If $N > K_2 + 1$, the matrix $\hat{\Sigma}_{22}$ is positive definite with probability 1. The positive definiteness of these matrices is desired because only then will the canonical correlations from the sample, and the weight functions from the sample corresponding to those correlations, be maximum likelihood estimates of those expressions built from the matrices Σ_{11} , Σ_{22} and Σ_{12} (Seber, 1984, p. 197).

Let $\hat{\mathbf{C}} = \hat{\Sigma}_{11}^{-1} \hat{\Sigma}_{12}$ and $\hat{\mathbf{D}} = \hat{\Sigma}_{22}^{-1} \hat{\Sigma}_{21}$, where $\hat{\Sigma}_{21} = \hat{\Sigma}_{12}'$. Then the matrices $\hat{\mathbf{C}} \hat{\mathbf{D}}$ and $\hat{\mathbf{D}} \hat{\mathbf{C}}$ have the same nonzero eigenvalues $\hat{\rho}_k^2$, and their corresponding eigenvectors $\hat{\mathbf{u}}_k$ and $\hat{\mathbf{v}}_k$ are given by the equations:

$$(\hat{\mathbf{C}} \hat{\mathbf{D}} - \hat{\rho}_k^2 \mathbf{I}_{K_1+1}) \hat{\mathbf{u}}_k = \mathbf{0}, \quad (\hat{\mathbf{D}} \hat{\mathbf{C}} - \hat{\rho}_k^2 \mathbf{I}_{K_2+1}) \hat{\mathbf{v}}_k = \mathbf{0},$$

$1 \leq k \leq \min\{K_1 + 1, K_2 + 1\}$. Having determined the eigenvectors $\hat{\mathbf{u}}_k$ and $\hat{\mathbf{v}}_k$ we determine the weight functions:

$$\hat{u}_k(t) = \hat{\mathbf{u}}_k' \boldsymbol{\varphi}(t), \quad \hat{v}_k(t) = \hat{\mathbf{v}}_k' \boldsymbol{\psi}(t), \quad t \in [0, T].$$

Hence the coefficients of the projection of the i th realization $y_i(t)$ of process $Y(t)$ on the j th functional canonical variable are equal to

$$\hat{U}_{ij} = \langle \hat{u}_j(t), y_i(t) \rangle = \int \hat{u}_j(t) y_i(t) dt = \sum_{k=0}^{K_1} \hat{\alpha}_{ik} \hat{u}_{jk} = \hat{\boldsymbol{\alpha}}_i' \hat{\mathbf{u}}_j,$$

$i = 1, 2, \dots, N, j = 1, \dots, \min(K_1 + 1, K_2 + 1)$, and the coefficients of the projection of the i th realization $x_i(t)$ of process $X(t)$ on the j th functional canonical variable are equal to

$$\hat{V}_{ij} = \hat{\boldsymbol{\beta}}_i' \hat{\mathbf{v}}_j, \quad i = 1, 2, \dots, N, \quad j = 1, \dots, \min(K_1 + 1, K_2 + 1).$$

5. Example

The data come from the online database (<http://data.worldbank.org/>) of the World Bank. For the analysis, eight regions of the world were chosen ($N = 8$): the European Union (EUU), low and medium income countries (LMY), the whole world (WLD), Bulgaria (BGR), Poland (POL), Russia (RUS), Uganda (UGA), and the United States (USA). These were characterized by two variables: GDP growth (Y) and rate of growth in direct foreign investment (X), recorded in the years 1993–2011 ($J^* = J^{**} = 19$). These data are presented in Figure 1 and Figure 2 respectively.

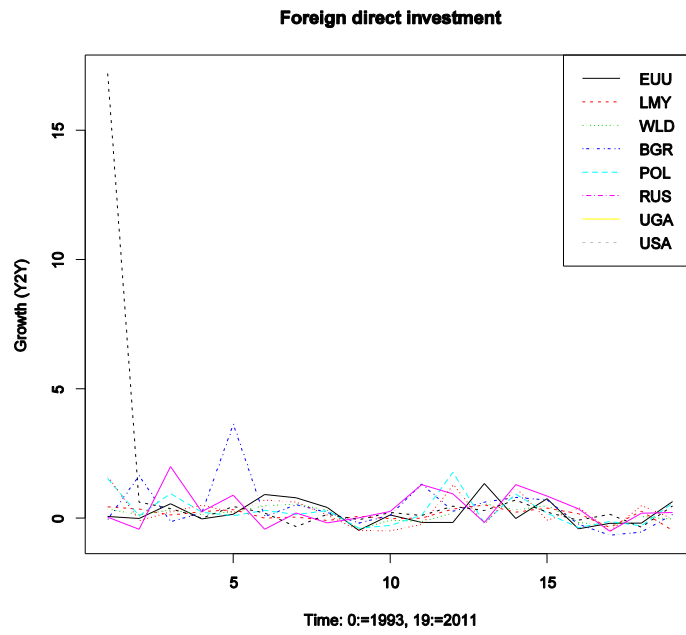


Figure 1. Time series for GDP growth (Y)

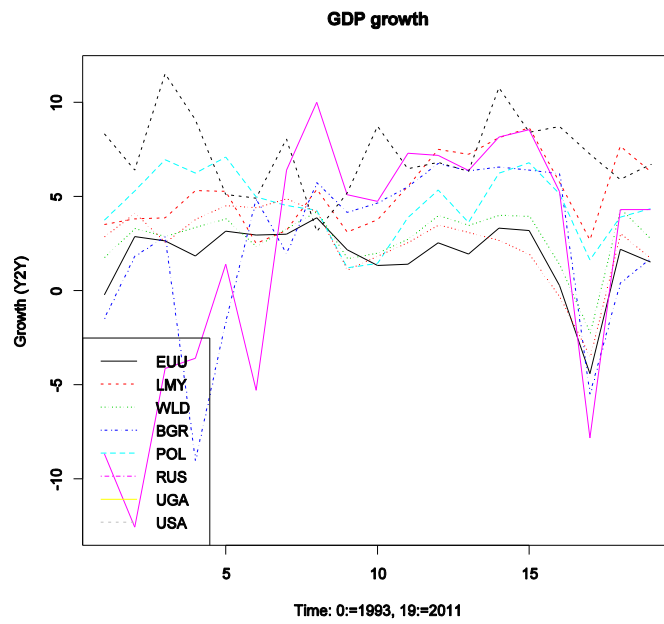


Figure 2. Times series for rate of growth in direct foreign investment (X)

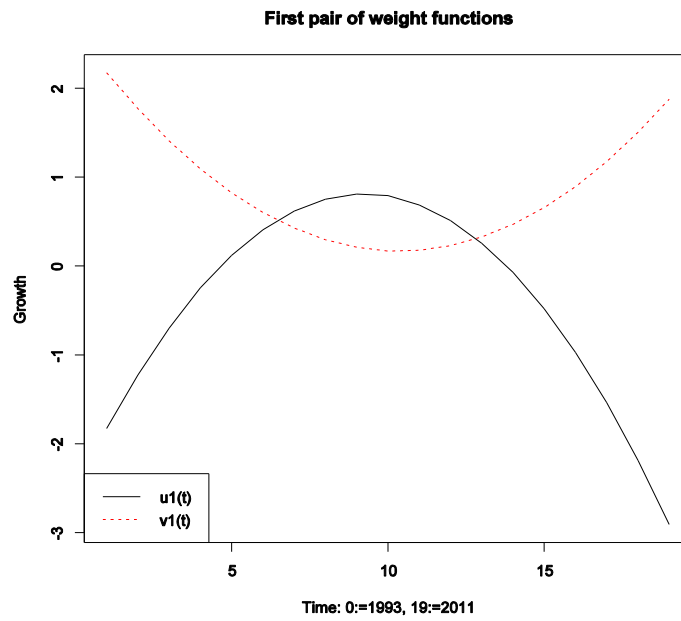


Figure 3. Weight functions $(u_1(t), v_1(t))$

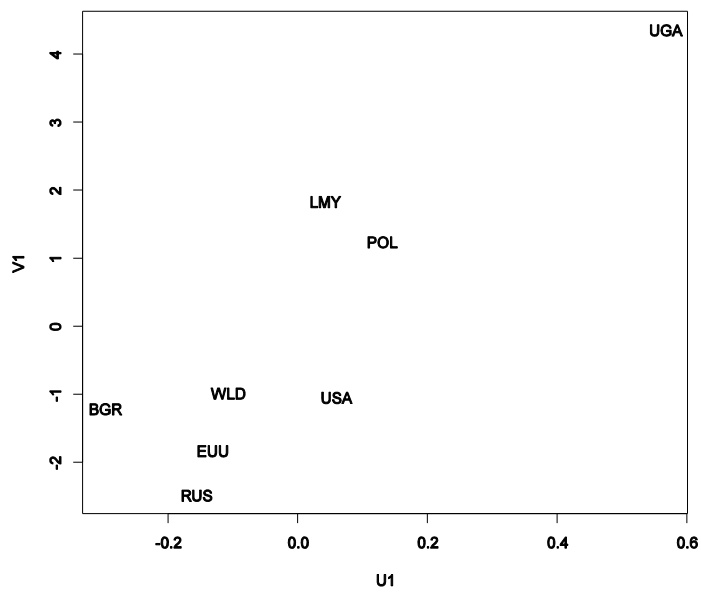


Figure 4. Projection of the eight selected regions of the world on the plane (U_1, V_1)

The series were centred, and calculations were performed using the Legendre basis on $L_2([-1, 1])$. Optimum values of K_1 and K_2 , selected using the BIC, take the value 2 for both series ($K_1 = 2 = K_2$). The first pair of weight functions $(u_1(t), v_1(t))$ is presented in Figure 3.

The first functional canonical correlation between the processes $Y(t)$ and $X(t)$ is equal to $\rho_1 = 0,88$, and the projection of the eight selected regions of the world on the plane of the first functional canonical variables (U_1, V_1) is shown in Figure 4.

Figure 4 shows the relative position of the eight selected regions of the world due to the degree of relationship between the two variables characterizing them: GDP growth (Y) and rate of growth in direct foreign investment (X), recorded in the years 1993–2011. The extreme positions are occupied by Uganda and Bulgaria. Poland is the most similar to the low and medium income countries (LMY). The utility of the proposed method can be seen on observing the time series shown in Figures 1 and 2.

Acknowledgments

The authors wish to thank the anonymous referee for his comments, which helped to improve the presentation of this paper.

REFERENCES

- Krzyśko M. (2009): Podstawy wielowymiarowego wnioskowania statystycznego [Foundations of multidimensional statistical inference]. Wydawnictwo Naukowe UAM, Poznań.
- Leurgans S.E., Moyeed R.A., Silverman B.W. (1993): Canonical correlation analysis when the data are curves. *Journal of the Royal Statistical Society B* 55(3): 725–740.
- Ramsay J.O., Danzell C.J. (1991): Some tools for functional data analysis. *Journal of the Royal Statistical Society B* 53: 539–572.
- Ramsay J.O., Silverman B.W. (2005): *Functional Data Analysis*. Second Edition, Springer.
- Schott J.R. (2005): *Matrix Analysis for Statistics*. Second Edition, Wiley, New York.
- Seber G.A.F. (1984): *Multivariate Observations*. Wiley, New York.
- Shmueli G. (2010): To explain or to predict? *Statistical Science* 25(3): 289–310.