Journal of Statistical Theory and Applications
Volume 11, Number 2, 2012, pp. 121-142 ISSN 1538-7887

Functional Analysis of Current and Noncurrent Balance Facilities of Iranian Export Development Bank

S. Mohammad E. Hosseininasab*, Mohammadreza Faridrohani and Anahita M.Golshan

Department of statistics, Faculty of Mathematical Sciences,
Shahid Beheshti University, Tehran, Islamic Republic of Iran.

Abstract

Canonical Correlation Analysis (CCA) deals with associations between two sets of random variables. CCA for data that are essentially in the form of curves is different from those that are multivariate, due to infinite dimensionality of the spaces that the data belong to. The main purpose of this research is to explore and understand the correlation structure and relationship between current and noncurrent balance of facilities, given by the Export Development Bank of Iran. Although they can be treated as multivariate observations, it may cause some problems because of not considering the functional nature of the data. We explore the correlation patterns between the two variables by using functional canonical correlation (FCCA), and then model noncurrent balance of facilities based on current balance of facilities. The model can be used to predict the values of the former variable by using the latter one.

Key Words: Functional data analysis, Functional canonical correlation, Covariance operator, Hilbert-Schmidt, Smoothing, Current and noncurrent balance facilities.

MSC: 6207, 62H20, 62J05

*m_hosseininasab@sbu.ac.ir

1 Introduction

Understanding dependence structure and relationship between two sets of variables is of main interest in statistics. When encountering two large sets of variables, a researcher can express the relationship between the two sets by extracting only finite linear combinations of the original variables that produce the largest correlations with the second set of variables.

The Export Development Bank of Iran (EDBI) is the unique state owned export -import bank in Iran that was founded in 1991. The main purpose of its establishment is to expand Iran's export and to grow trade relations with other countries. It supports Iranian exporters and buyers of the goods and services of Iranian origin, with financing facilities and banking services. Furthermore, it provides working capitals for Iranian exporter to purchase all equipments and raw materials that they need to produce, pack, and transport goods for export. It also supports exporters to sell Iranian goods and services. Both current balance of facilities, whose due dates for installment payments do not pass 2 months, and noncurrent balance of facilities, whose installment due dates pass 2 months are important to the EDBI, have positive and negative meaning for the bank performance, respectively. While high noncurrent balance of facilities indicates low chance of claims receipt, high current balance of facilities shows high number of the customers attracted to the bank.

We have considered both current and noncurrent balance of facilities as continuous functions of time, and taken functional nature of the two variables over time into account. It is worth noting that the corresponding data are generally available daily or monthly (discretely). However, if one treats those data as multivariate observations by using multivariate methods, it may cause some problems (Ramsay, 1982).

When data are continuous functions of another variable (generally time), the multivariate methods should not be used to analyze them. Therefore, some theoretical justification is needed to provide the required definitions and concepts regarding the essential nature of the data. This leads to defining canonical correlation for pairs of random functions called functional canonical correlation analysis (FCCA) (Ramsay and Silverman, 2005). If the data related to functional phenomena, are observed discretely at different times, the first task is to convert these observations to appropriate curves. This is due to functional nature of the phenomena that the data related to. On the other hand, the functional quantity of interest may be measured with error. In such cases, we should first remove the observational error by taking a smoothing procedure into account.

Canonical correlation analysis was first proposed by Hotelling (1936) to study the relationship between two vector random variables. In early work on this problem, Hannan (1961) and Brillinger (1975) described canonical analysis for multivariate stationary time series. Leurgans et al. (1993) also demonstrated the need for regularization in functional canonical correlation analysis and defined smoothed functional canonical correlation analysis. They achieved regularization by using modified smoothing splines and explained their technique with an application to the study of human gait movement data; see also Olshen et al. (1989) and Rice and Silverman (1991). Dauxios and Nkiet (1997) developed a theoretical approach, based on angels between subspace of functions. A different and promising approach aiming at covariance rather than correlation for pairs of random curves was proposed by Service et al. (1998). The role of regularization for canonical correlation is to restrict the dimension of the problem and can be done via a wise choice of the roughness penalty for smoothing splines (Leurgans et al. 1993), or by alternative approaches that allow to avoid the inversion problem (Service et al. 1998). He et.al (2000) extended basic concepts of multivariate analysis to those of functional data analysis. In particular, they drew the basic properties of functional canonical correlation and linear regression from their multivariate counterparts. He et.al (2004) proposed four computational methods for the estimation of functional canonical correlation and canonical weight functions.

In this paper, the data collected from the EDBI in 1388 persian solar calender called Shamsi(March 2009-March 2010), are treated by using FCCA. The data set contains discrete measurements of two phenomena: current and noncurrent balance of facilities, and was collected from 23 branches. We have fitted continuous curves to the original data and used finite expansions to approximate the corresponding processes and then applied canonical correlation analysis to the resulting curves. Then the correlation patterns between pairs of the two phenomena have been extracted. We have also modeled noncurrent balance of facilities based on current balance of facilities. The model can be used to predict the values of the former variable by using the latter one.

2 Functional Canonical Correlation Analysis

Suppose we observe a sample of bivariate processes (X, Y), where $X \in L_2[0, T_1]$ and $Y \in L_2[0, T_2]$,are jointly distributed and

$$E||Z||^2 = E[< Z, Z >] = E\int_0^T T(Z(s))^2 ds < \infty;\ for\ Z = X\ or\ Y.$$

Here $L_2[0, T_1], L_2[0, T_2]$ are two Hilbert spaces of square integrable functions on $[0, T_1]$ and $[0, T_2]$ with the inner product $< u, v >= \int u(s)v(s)ds$.

It should be mentioned that the definition of canonical correlation for functional data and for multivariate data is almost the same. While the inner product in multivariate for two random vectors is $\mathbf{X}$ and $\mathbf{Y}$, $< X, Y >= \mathbf{X}^T\mathbf{Y}$, the inner produce in Hilbert space for two random processes X and Y is defined as $< X, Y >= \int X(t)Y(t)dt$. Then, the covariance functions are defined as follows (Ramsay and Silverman, 2005):

$$\begin{aligned}
c_{XX}(s,t) &= Cov\{X(s), X(t)\} = E\{[X(s) - \mu_X(s)][X(t) - \mu_X(t)]\}, \quad s,t \in [0, T_1],\\
c_{YY}(s,t) &= Cov\{Y(s), Y(t)\} = E\{[Y(s) - \mu_Y(s)][Y(t) - \mu_Y(t)]\}, \quad s,t \in [0, T_2],\\
c_{XY}(s,t) &= Cov\{X(s), Y(t)\} = E\{[X(s) - \mu_X(s)][Y(t) - \mu_Y(t)]\}, s \in [0, T_1], t \in [0, T_2].
\end{aligned}$$

Also, the covariance operators $C_{XX} : L_2[0, T_1] \to L_2[0, T_1]$, $C_{XY} : L_2[0, T_2] \to L_2[0, T_1]$ and $C_{YY} : L_2[0, T_2] \to L_2[0, T_2]$, are defined as:

$$\begin{aligned}
C_{XX}u(s) &= \int_0^{T_1} c_{XX}(s,t)u(t)dt, \quad u \in L_2[0, T_1],\\
C_{XY}u(s) &= \int_0^{T_2} c_{XY}(s,t)u(t)dt, \quad u \in L_2[0, T_2], s \in [0, T_1],\\
C_{YY}u(s) &= \int_0^{T_2} c_{YY}(s,t)u(t)dt, \quad u \in L_2[0, T_2].
\end{aligned}$$

Operators C_{XX} and C_{YY} are compact, self-adjoint, and nonnegative definite, and C_{XY} is compact (He et al. 2003).

Since $Cov(< u, X >, < v, Y >) =< u, C_{XY}v >$, $Var(< u, X >) =< u, C_{XX}u >$, and $Var(< v, Y >) =< v, C_{YY}v >$, the kth canonical correlation and its corresponding weight function are given by

$$\rho_k = \sup_{\substack{u\in L_2[0,T_1],<u,C_{XX}u>=1,\\ v\in L_2[0,T_2],<v,C_{YY}v>=1}} < u, C_{YX}v >=< u_k, C_{YX}v_k >, \quad (1)$$

where, in addition, for $k > 1$, (u_k, v_k) is uncorrelated with (u_i, v_i) for $i = 1, 2, \dots k-1$.

Dauxois et al. (2004) introduced a canonical correlation analysis between two random variables relative to a third one, working only when the range of $C_{XX}^{1/2}$ and $C_{YY}^{1/2}$ are closed subspaces in $L_2[0, T_1]$ and $L_2[0, T_2]$, respectively, that is only held when the two operators are finite dimensional. This is similar to other "French school" works such as Dauxois and Pousse (1976) that are based on Hotelling's original results. However, Cupidon et al. (2008) showed that the solution may lie on the boundary of the range of $C_{XX}^{1/2}$ and $C_{YY}^{1/2}$ so that a pair of canonical variables may not exist in the underlying spaces, since the supremum in (1) may not be a maximum. He et.al (2003) tried to dispel the problem by imposing some unnecessary restrictions on the singular values of the operator $R = C_{XX}^{-1/2}C_{XY}C_{YY}^{-1/2}$, because their conditions do not always hold and so the canonical correlations may not generally exist (see Eubank and Hsing, 2008). The work of Kupersanin et al. (2010) gives a definition of FCCA based on the reproducing kernel Hilbert space representation of a stochastic process that extends the classical CCA successfully and fixes all these problems.

2.1 Empirical Estimates

Suppose the observed data are $\{(X_i(t), Y_i(t)),\ i = 1, 2, \dots, n\}$, then the sample covariance functions and cross-covariance function are defined as follows:

$$\begin{aligned}
\widehat{c}_{XX}(t_1,t_2) = \widehat{Cov}\{X(t_1), X(t_2)\} &= \frac{1}{n-1}\sum_{i=1}^{n}\{X_i(t_1) - \bar{X}(t_1)\}\{X_i(t_2) - \bar{X}(t_2)\},\\
\widehat{c}_{YY}(t_1,t_2) = \widehat{Cov}\{Y(t_1), Y(t_2)\} &= \frac{1}{n-1}\sum_{i=1}^{n}\{Y_i(t_1) - \bar{Y}(t_1)\}\{Y_i(t_2) - \bar{Y}(t_2)\},\\
\widehat{c}_{XY}(t_1,t_2) = \widehat{Cov}\{X(t_1), Y(t_2)\} &= \frac{1}{n-1}\sum_{i=1}^{n}\{X_i(t_1) - \bar{X}(t_1)\}\{Y_i(t_2) - \bar{Y}(t_2)\}.
\end{aligned}$$

Also the cross-correlation function is defined as:

$$\hat{r}_{XY}(t_1,t_2) = \widehat{Corr}\{X(t_1),Y(t_2)\} = \frac{\widehat{Cov}\{X(t_1),Y(t_2)\}}{\sqrt{\widehat{Var}\{X(t_1)\}\widehat{Var}\{Y(t_2)\}}}.$$

Sample canonical variables are obtained by replacing correlation operators in (1) with their corresponding sample versions, defined as follows:

$$\begin{aligned}
\widehat{C}_{XX}u(s) &= \int_0^{T_1} \widehat{c}_{XX}(s,t)u(t)dt, \quad u \in L_2[0,T_1],\\
\widehat{C}_{XY}u(s) &= \int_0^{T_2} \widehat{c}_{XY}(s,t)u(t)dt, \quad u \in L_2[0,T_2], s \in [0,T_1],\\
\widehat{C}_{YY}u(s) &= \int_0^{T_2} \widehat{c}_{YY}(s,t)u(t)dt, \quad u \in L_2[0,T_2].
\end{aligned}$$

3 Functional Smoothed Canonical Correlation Analysis

When data are observed with error, smoothed canonical correlation(SCCA) is usually more useful, compared to the ordinary canonical correlation(Leurgans et al., 1993). It involves defining a roughness penalty for each function, so that we can filter the errors from the data.

Let S be a space of functions on T with square integrable second derivative. The roughness penalty is then generated by a positive definite quadratic form defined for functions in S. Suppose $[.,.]$ is an inner product in S. For each f and g in S, we take $[f,g]$ to be $\int_T f''(t)g''(t)dt$. Then, the roughness penalty for $f \in S$ is defined as $[f,f] = \int_T f''(t)^2dt$.

A simple way of introducing smoothing is to modify the constraints in (1) with the following by adding the roughness penalty constraints:

$$< u, C_{XX}u > +\alpha_1[u,u] =< v, C_{YY}v > +\alpha_2[v,v] = 1, \tag{2}$$

where α_1 and α_2 are positive smoothing parameters. In this way, to obtain particular "candidate" canonical variables, we consider not only their variances but also their roughness.

Then, the problem of maximizing $< u, C_{XY}v >$ subject to constraint (1) is equivalent to maximizing the penalized squared sample correlation,

$$\rho_k = \sup_{\substack{u\in L_2[0,T_1],<u,C_{XX}u>+\alpha_1[u,u]=1,\\ v\in L_2[0,T_2],<v,C_{YY}v>+\alpha_2[v,v]=1}} < u, C_{YX}v >=< u_k, C_{YX}v_k >,$$

where, for $k > 1$, (u_k, v_k) is uncorrelated with (u_i, v_i) for $i = 1, 2, \dots k-1$. We shall refer to this procedure as functional smoothed canonical correlation analysis (FSCCA).

When considering a sample of size n of random functions, the kth canonical variable is given by

$$\widehat{\rho}_k = \sup_{\substack{u\in L_2[0,T_1],<u,\widehat{C}_{XX}u>+\alpha[u,u]=1,\\ v\in L_2([0,T_2],<v,\widehat{C}_{YY}v>+\alpha[v,v]=1}} < u, \widehat{C}_{YX}v >=< u_k, \widehat{C}_{YX}v_k >,$$

where, for $k > 1$, $(\hat{u}_k, \hat{v}_k)$ is uncorrelated with $(\hat{u}_i, \hat{v}_i)$ for $i = 1, 2, \dots k-1$. Alternatively, FSCCA is carried out by maximizing the penalized squared sample correlation

$$\widehat{Corr}^2_\alpha(u, v) = \frac{< u, \widehat{C}_{XY}v >^2}{(< u, \widehat{C}_{XX}u > +\alpha[u, u])(< v, \widehat{C}_{YY}v > +\alpha[v, v])}. \tag{3}$$

A good choice of the smoothing parameters leads to have a pair of canonical variables possessing fairly smooth weight functions and a correlation that is appropriate. Clearly the larger the values of α_1 and α_2, the more emphasis will be placed on the roughness penalty. On the other hand, smaller values of α_1 and α_2 take the true correlation more into consideration. Here, we only consider the special case where $\alpha_1 = \alpha_2 = \alpha$. The extension of the techniques and of the theoretical discussion to the more general case is straightforward. The smoothing parameter α can be chosen by cross-validation criterion. Let $\hat{u}_\alpha^{(-i)}$ and $\hat{v}_\alpha^{(-i)}$ be the maximizers of $\widehat{Cov}_\alpha^{-i}(u, v)$ with the ith curve omitted. The cross-validation score of α is then defined to be the squared correlation of the n pairs of numbers $(\hat{u}_\alpha^{(-1)T}X_1, \hat{v}_\alpha^{(-1)T}Y_1), \dots, (\hat{u}_\alpha^{(-n)T}X_n, \hat{v}_\alpha^{(-n)T}Y_n)$. We then choose the value of α that maximizes this correlation.

In practice, to maximize system (3), we should discretize the functions $u(t)$ and $v(t)$ and the covariance operators $\hat{C}_{XX}$, $\hat{C}_{YY}$ and $\hat{C}_{XY}$ onto a fine grid. As a result, we obtain a large linear system from system (3), whose leading eigenvalues and eigenvectors are found by the standard numerical methods.

4 Real Data Analysis

Data analyzed here contain current and noncurrent balance of facilities in the EDBI that have been collected monthly since 1388 persian solar calender, called Shamsi (March 2009-March 2010). The EDBI was established in 1991 with the objective of increasing Iran's export and developing trade with other countries. As the unique state-owned export-import bank of Iran, it was established to promote the country's exports and to develop economic and business exchanges with other countries. Now, it has 33 branches all across the country.

According to the EDBI, balance of facilities whose installment due dates do not pass more than 2 months are called current and those whose due dates for installment payments pass 2 months are called noncurrent. Both current and noncurrent balance of facilities are important and noticeable and conceptually have positive and negative meaning for the bank, respectively. High current balance of facilities (CBF) shows that the number of the customers that attracted to the bank is high, and consequently bank profit is enhanced. Raising noncurrent balance of facilities (NBF), however, decreases the chance of claims receipt.

The main purpose of this research is to explore and understand the correlation structure between current and noncurrent balance of facilities. Because noncurrent balance of facilities are affected by current balance of facilities, relationship between them is not deniable. These explanation shows that the research on relationship and correlation structure between these two variables is useful.

4.1 FCCA for the Bank Data

Since data are time-dependent, it could be considered as functions of time. In regard to the fact that the original discrete data must be firstly converted to continuous functions, we have fitted appropriate curves to the original data collected randomly from 23 branches of the EDBI. Figure 1 shows the appropriate curves fitted to the monthly average of CBF and NBF.

The mean and standard deviation functions for CBF and NBF are shown in figures 2 and 3, respectively. The measurement scale of CBF and NBF is in billion rial. As figure 2 shows, CBF and NBF are minimized in the second and ninth month of the year, respectively. Figure 3 displays higher variation in CBF than in NBF.

To start analyzing the data, we have first computed the correlation functions and their

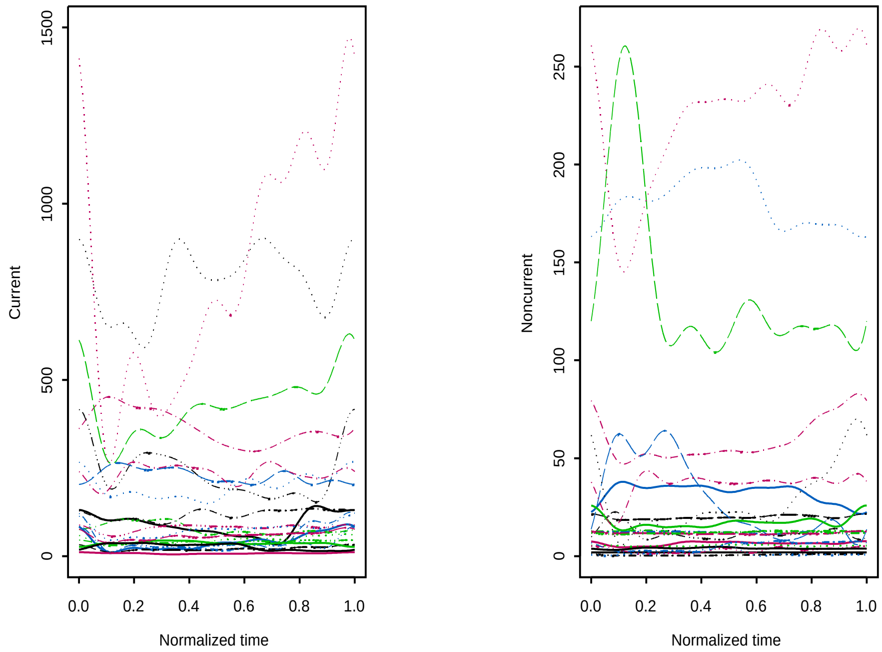

Figure 1: Monthly mean of Current and noncurrent balance of facilities collected from 23 branches of the EDBI (in billion rial).

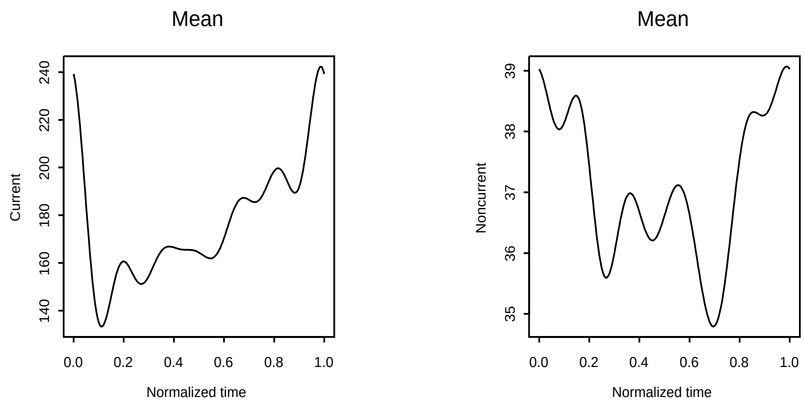

Figure 2: Mean function of current and noncurrent balance of facilities (in billion rial).

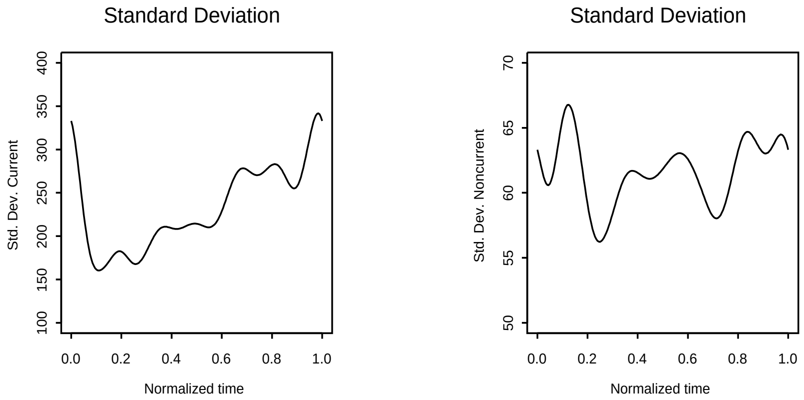

Figure 3: Standard deviation of current and noncurrent balance of facilities.

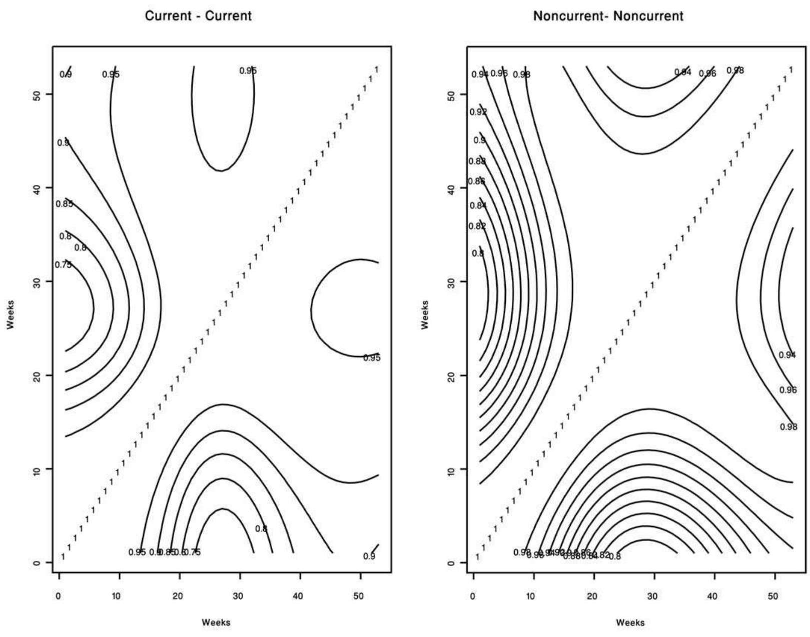

Figure 4: Correlation function of current and noncurrent balance of facilities.

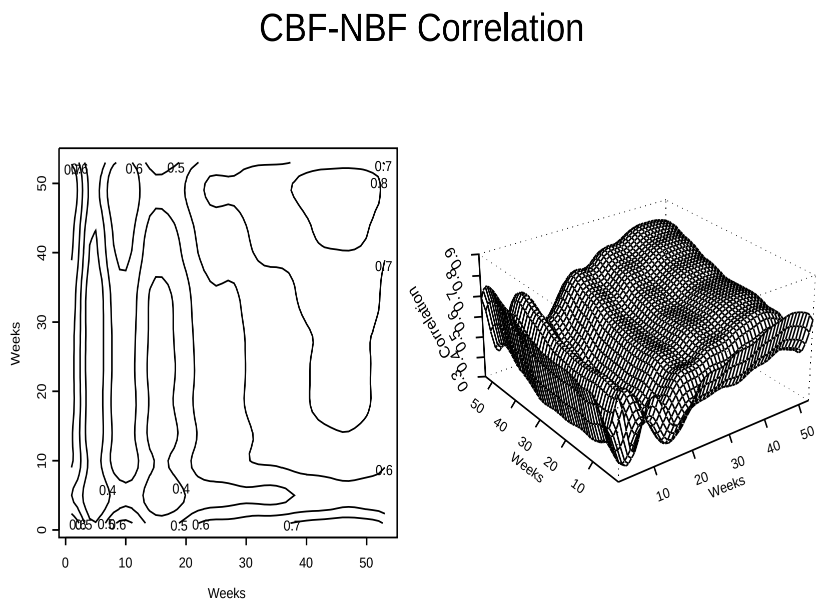

Figure 5: Cross-correlation function of current and noncurrent balance of facilities.

corresponding contour plots. The correlation functions of CBF and NBF are presented in figure 4. Generally, long-term dependency in NBF is higher than that in CBF. Furthermore, correlations in CBF(NBF) vary between 0.75 (0.78) and 1.00 (1.00). The correlation between CBF (NBF) at the beginning of the year and CBF (NBF) at the middle of the year is the lowest, 0.75 (0.78), compared to the other periods of time. However, it is the highest between the beginning of the year and the end of the year. This is the case for both CBF and NBF between the middle of the year and the end of the year. It is also seen that the highest correlations occur for periods whose length are at most 7 weeks. These periods are seen in a band with 7 weeks width, occurring symmetrically around the main diagonal of the counterplot appeared in figure 4. The correlation between CBF for weeks 1-5 and CBF for weeks 22-32 is the lowest (0.75). However, the correlation between CBF for weeks 1-5 and CBF for weeks 14 and 15 is 0.95. The lowest correlation for NBF (0.78) occurs between weeks 1-3 and weeks 24-34. Correlation for NBF between weeks 1-3 and weeks 8 and 9 is 0.98. It is clear that the correlation between CBF (NBF) at the beginning of the year and CBF (NBF) at the middle of the year is approximately 0.75 (0.78) and it increases when moving away from the middle of the year. The correlation between CBF (NBF) at the middle of the year and CBF (NBF) at

Canonical Correlation	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6
value	0.965	0.785	0.598	0.377	0.236	0.012

Table 1: Current and noncurrent balance of facilities canonical correlations.

the end of the year is approximately 0.90 (0.95), and it increases for NBF when moving to the beginning of the year. However, it is not changed for CBF in that period of time.

The cross-correlation function between CBN and NBF is given in figure 5. In this plot, the horizontal and vertical axes show CBF and NBF, respectively. They are scaled weekly so that we have a better interpretation of the plot. Figure 5 shows that there is a strong correlation between CBF and NBF for the last weeks of the year. In particular, this correlation is slightly more than 0.8 for the two variables since week 40. Furthermore, the correlation between NBF from week 30 and CBF from week 10 to the end of the year is a decreasing function of the difference between the corresponding times. The minimum correlation (about 0.5) between NBF since week 30 and CBF before week 10, occurs at week 5 for CBF, and when distancing from week 5 to week 1 or to week 10, this correlation increases until it reaches 0.7. For all weeks before week 30 of NBF, the correlation between CBF and NBF is periodic, so that in week 5 it is decreasing from 0.7 to less than 0.4, and then it is increasing until week 10 at which time it reaches about 0.6. From this week until week 15, the correlation again decreases to 0.4, and from week 15 to week 30 it increases to 0.6.

FSCCA is used to analyze the data and to extract correlation patterns between CBF and NBF. To do that, we have used cross-validation criterion for choosing the amount of the smoothing parameter. The first six canonical correlations values related to the two variables are shown in table 1. As we can see, the last three canonical correlations are small compared to the first three. Consequently, in this study the first three canonical variables are taken into consideration.

Weighted functions of the first three canonical variables are shown in figure 6. In this graph, time is scaled so that these functions take their own values between 0 and 1.

The first graph in figure 6 shows the first weighted function of canonical variables. It can clearly be seen that the shape of the weighted function curves of CBF and NBF are approximately similar. Also this graph shows CBF first minimum value of the weighted function occurs before that of NBF weighted function. It comes to the same conclusion for maximum value. We can conclude that CBF change as a result of variation in NBF. To make it clear, the first

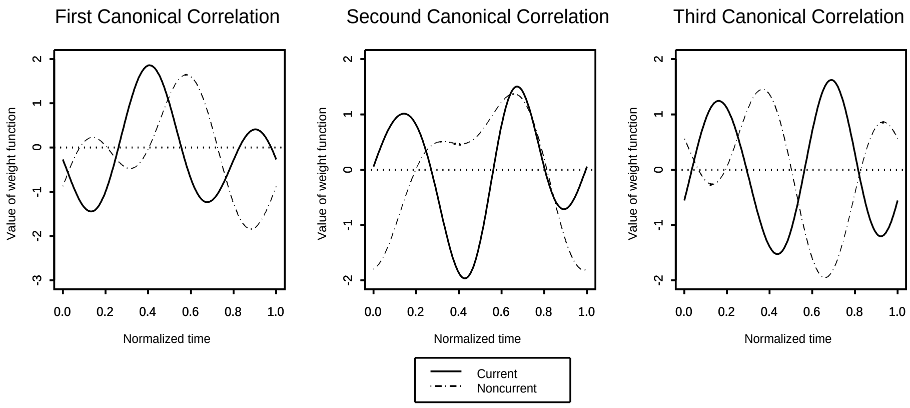

Figure 6: Weighted function canonical correlation of current and noncurrent balance of facilities.

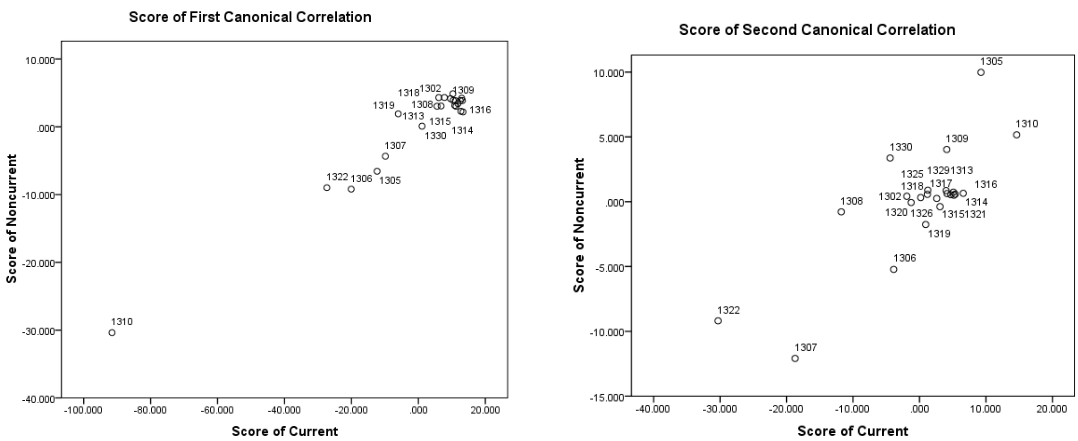

Figure 7: Scores of the first and second canonical correlation for current and noncurrent balance of facilities.

scores of current and noncurrent canonical correlation for all branches are shown in figure 7. The scores of each branch have two components. The first component shows the effects of first canonical correlation on CBF of each branch and the second one indicates the effects of the first canonical correlation on NBF of that branch. Figure 7 shows that there are five branches, whose CBF and NBF scores are approximately the same. Apart from these five branches, Keshavarz (1310) branch has minimum score for CBF and NBF. This means that the effect of the first canonical correlation on CBF and NBF of this branch is lower than other branches. This is due to a high rate of NBF for this branch at the end of the year and small values of the weighted function of the first canonical correlation for NBF at that time. Generally, the branches are put on the graph according to the amount of their work. Moreover, this figure shows that such branches as 1310, 1322, 1306, 1305, 1307 have a higher amount of work than other branches. These branches have given more facilities to the customers and as a result they have higher rate of NBF than other branches.

The second graph in figure 6 shows the second weighted function of the canonical variables. It can clearly be seen that the second CBF and NBF weighted functions have the same behavior from the 7th to the 9th month of the year.

To make it clear, the scores of the second canonical correlation for CBF and NBF for all branches are presented in figure 7. The graph shows that there are four branches,whose CBF and NBF scores are approximately the same. Among others, Mirdamad branch (1322) has minimum and Keshavarz branch has maximum score for CBF and NBF. This means that the effect of the second canonical correlation on CBF and NBF of Mirdamad branch is low and of Keshavarz branch is high in comparison to other branches. That is due to a high rate of NBF for Mirdamad branch at the end of the year and small values of the weighted function of the second canonical correlation for NBF at that time. It is worth noting that branches are put on the graph according to the number of their approved facilities. Therefore, such branches as 1307, 1322 and 1306 have maximum number of the approved facilities and 1310, 1309 and 1305 have minimum number of the approved facilities.

The third graph in figure 6 shows the third weighted function of the canonical variables. It can clearly be seen that the third CBF and NBF weighted functions have the same behavior.

The scores of the third canonical correlation of CBF and NBF show that Tabriz branch (1307) has minimum score for CBF and NBF. This means that the effect of the third canonical correlation of CBF and NBF for this branch is low. This is due to a low rate of NBF of this branch in the 4th month of the year and also large values of the weighted function of the third canonical correlation for NBF in this period of time. According to the scores, the main factor of putting the branches on the graph (not shown here) in this way is the place at which the facilities are approved. The scores show that branches such as 1307, have maximum number of the facilities approved by the branch, and the branches coded 1319 and 1322 have also maximum number of the facilities approved by the central office. The facilities approved by the central office are either the facilities that whose amount is higher than the maximum amount that the branches are allowed or those that are ordered to pay in any way.

5 Functional Modeling of NBF based on CBF

The simplest functional regression model is when dependent variable Y is scalar and the explanatory variable $X(t)$ is a function of another variable. Precisely, we assume

$$Y_i = \alpha + \int_{\mathcal{I}} \beta(t) X_i(t) dt + \epsilon_i, \quad i = 1, \cdots, n, \tag{4}$$

where α is the intercept, $\beta(.)$ is the slope of the regression, and the ϵ_i's are the error terms. It is assumed that $\beta(.)$ and $X_i(t)$ are square integrable functions from $\mathcal{I}$ to the real line, the pairs $(X_1, \epsilon_1), \cdots, (X_n, \epsilon_n)$ are independent and identically distributed, the random functions $X_i(.)$ are independent of the errors ϵ_i, $\sigma^2 = E(\epsilon^2) < \infty$, $E(\epsilon) = 0$, and $\int_{\mathcal{I}} E[X(t)^2] < \infty$, where ϵ and $X(.)$ are distributed as ϵ_i and $X_i(.)$, respectively.

In this way, we are interested in explaining variation in the total annual NBF by using the CBF variation pattern through the year. Because NBF variation depends on CBF variation, we are going to model NBF based on CBF(t), the amount of CBF at time t. Because the total annual NBF is distributed across the country, it is highly variable from one branch to another. Thus, we decided to use its logarithm as the dependent variable. Specially, we regarded the dependent variable Y as

$$Y_i = \log_{10}(\sum_{j=1}^{12} NBF_{ij}) - \frac{1}{23} \sum_{i=1}^{23} \log_{10}(\sum_{j=1}^{12} NBF_{ij}),$$

where $\mathrm{NBF_{ij}}$ denotes the amount of NBF reported by branch i at the end of month j. Then,

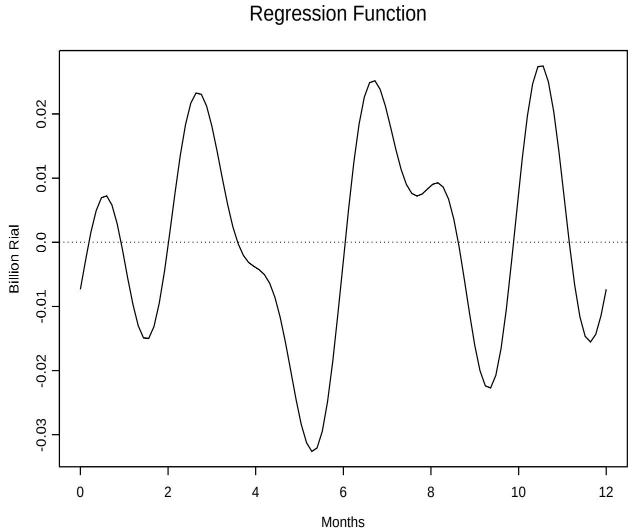

Figure 8: Estimated regression weight function $\beta(.)$ for prediction of log annual mean of NBF by the CBF profiles.

we estimated the slope $\beta(.)$ and intercept α from the data using the estimators proposed by Ramsay and Silverman (2005).

As figure 8 shows, the estimated slope of the functional linear regression gives larger weights to the weeks 2-3, 10-12, 27-29, and 44-46. Furthermore, the largest amount of the function occurs at week 45, followed by at week 28 and 11. The lowest weight, however, appears at week 22, followed by at week 40. The estimated intercept is $\hat{\alpha} = -0.43$, and the coefficient of determination (R^2) is also 0.74.

We now consider the response variable NBF as a function of t, and fit appropriate curves to the collected data. Then, we consider a regression model for the response function $NBF(t)$ involving the functional covariate CBF(t). In this way, we allow for CBF(.) influence on NBF(.) to extend over the whole year. Let

$$\log NBF_i(t) = \alpha(t) + \int_0^{12} CBF_i(s)\,\beta(s,t)ds + \epsilon_i(t)\,, \quad i = 1, \cdots, n, \tag{5}$$

where $\log NBF_i(t)$ is the logarithm of $NBF_i(t)$ for branch i , the bivariate function $\beta(s,t)$ is the regression slope, and can be interpreted as a function of s for each fixed value of t. Here, the $\epsilon_i(t)$'s are the error terms, and $\alpha(.)$ is the intercept function.

We can express $\beta(s,t)$ as a double expansion in terms of k_1 basis functions η_k and k_2 basis

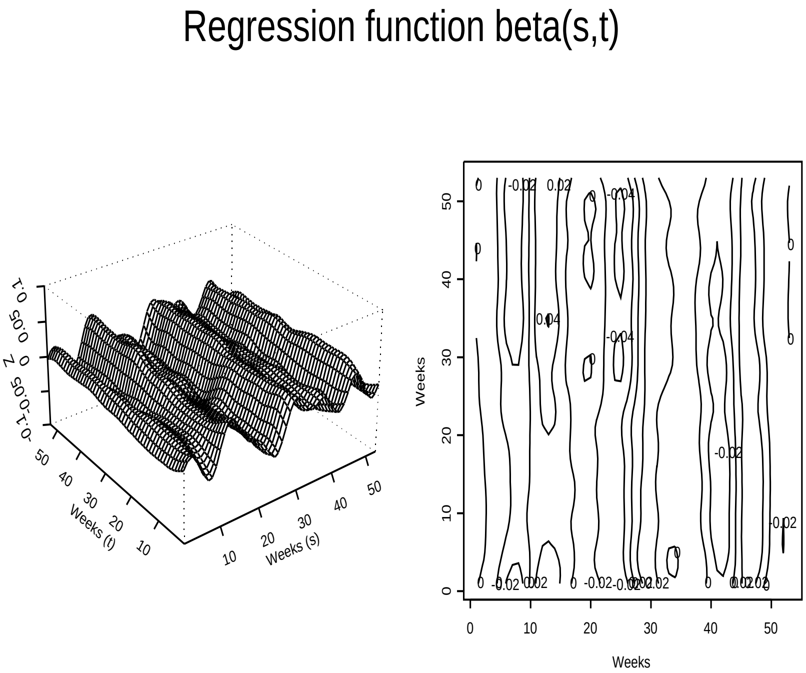

Figure 9: Estimated slope function $\beta(.,.)$ for prediction of log NBF(.) from CBF(.). The value $\beta(s,t)$ presents the influence of CBF at time s on log NBF at time t.

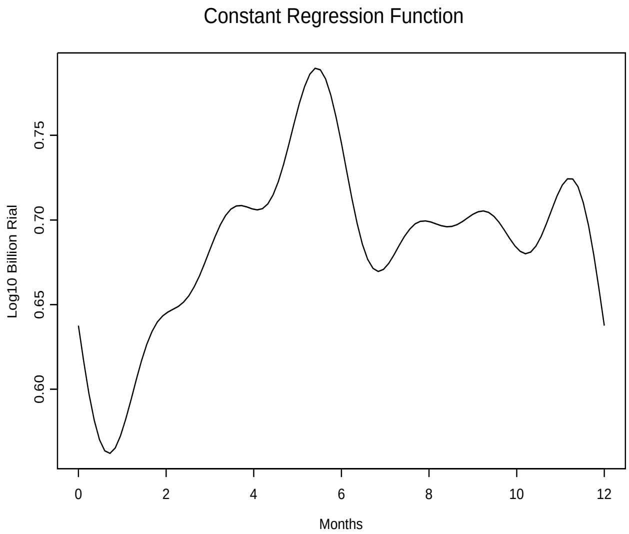

Figure 10: Estimated intercept function $\alpha(.)$ of the functional regression model for prediction of log NBF based on CBF.

functions θ_k as follows:

$$\beta(s,t) = \sum_{k=1}^{k_1}\sum_{l=1}^{k_2} b_{kl}\,\eta_k(s)\,\theta_l(t) = \eta(s)\,\mathbf{B}\,\theta(t)\,,$$

where $\mathbf{B} = [b_{lk}]$ is a $k_1 \times k_2$ matrix of coefficients b_{lk}. Also,

$$\alpha(t) = \sum_{j=1}^{k_2} c_j\,\theta_j(t) \qquad CBF_i(s) = \sum_{j=1}^{k_1} \xi_{ij}\,\eta_j(s)\,.$$

Therefore,

$$\log NBF_i(t) = \sum_{j=1}^{k_2} c_j\,\theta_j(t) + \sum_{l=1}^{k_2} b_{jl}\,\xi_{ij}\,\theta_l(t) = \sum_{j=1}^{k_2} d_{ij}\,\theta_j(t)\,,$$

where $d_{ij} = c_j + \sum_{l=1}^{k_1} b_{lj}\,\xi_{il}$. Here we have first used the Fourier basis for expanding both NBF and CBF curves and estimated $\beta(.,.)$ function with both bases truncated to 8 terms.

Figure 9 presents the estimated slope function $\beta(s,t)$. It is seen that the amount of the bivariate function $\beta(s,t)$ is less than -0.02 when $6 < s < 8$ and $t < 3$ or $t > 30$. This is also the case when $21 < s < 26$, $40 < s < 43$ and $51 < s < 54$, for all t. When $5 \leq 5$, $10 < s < 20$, $30 < s < 40$, and $43 \leq s < 50$, the amount of the slope function is non-negative. The estimated slope gives the largest weight (0.04) for $11 < s < 12$ and $32 < t < 36$, followed by 0.02 for $s = 47$ for all t. The lowest weight is less than -0.06 and occurs when $24 < s < 25$ and $28 < t < 32$ or $37 < t < 51$.

Figure 10 shows that the estimated intercept $\hat{\alpha}(t)$ decreases from 0.64 at week 1 to 0.51 at week 3. Then, it almost increases until week 22, and gets its maximum value (0.79) at week 22. After that, it decreases by week 28 and gets 0.68. Between weeks 28 and 42, it gradually changes, and from week 42 to week 47, it increases so that it is 0.72 at week 47. Then, by the end of the year, it decreases, and reaches 0.64. Therefore, $\hat{\alpha}(t)$ varies between 0.51 and 0.79 through the year.

We can assess the fit of a functional linear model as the standard linear models by computing the coefficient of determination (R^2). Here R^2 is a function of t, denoted by $R^2(t)$, and is defined as follows:

$$R^2(t) = 1 - \frac{\sum_{i=1}^{n}\{\hat{y}_i(t) - y_i(t)\}^2}{\sum_{i=1}^{n}\{y_i(t) - \bar{y}(t)\}^2}\,.$$

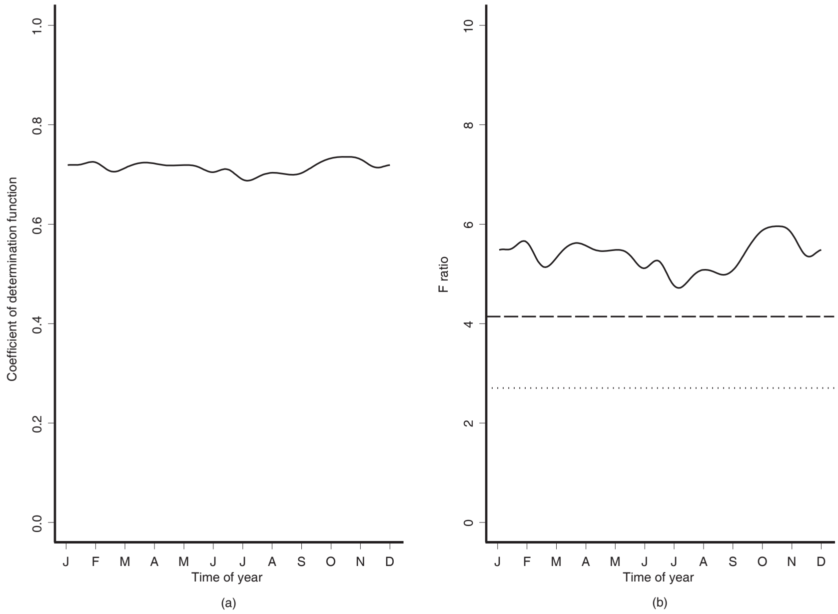

Figure 11: (a) Proportion of variance of log NBF(.) explained by a linear model on CBF(.), (b) F-ratio function for prediction of log NBF(.) from CBF(.). The horizontal lines show the upper 5% and 1% points of the F(7,15) distribution.

We can also compute an F-ratio function for the fit. Considering $SSR(t) = \sum_{i=1}^{n}\{\hat{y}_i(t) - \bar{y}(t)\}^2$ and $SSE(t) = \sum_{i=1}^{n}\{y_i(t) - \hat{y}(t)\}^2$ as the point-wise sum of squares of regression and sum of squares of errors, respectively, we can assign $k_0 - 1$ and $n - k_0$ degrees of freedom to SSR(t) and SSE(t) by analogy with the conventional linear model context, where k_0 is the number of the parameters used in the model. Therefore,

$$\mathrm{FRATIO(t)} = \frac{\mathrm{SSR(t)/k_0 - 1}}{\mathrm{SSE(t)/n - k0}},$$

which has approximately a F distribution with $k_0 - 1$ and $n - k_0$ degrees of freedom for each t.

Figure 11(a) plots $R^2(t)$ for the fit to the logNBF(.). It indicates that the fit is overall good, because of high values of $R^2(t)$ changing form 0.69 to 0.74 all over the year. Figure 11(b) shows the F-ratio for the fit to the logNBF(.). The upper 5% and 1% quantiles of the $F(7, 15)$ distribution are plotted by dotted and dashed lines, respectively. It is clear that the effect of the CBF on the NBF is highly significant overall.

6 Conclusion

We explore the correlation structures between CBF and NBF of the Iranian Export Development Bank by using Functional Canonical Correlation. It is obtained that the first correlation factor is related to the amount of work done by the branches and the second factor is a result of the number of approved facilities. The place at which facilities are approved, is the third factor. It is either the central office or the branches. We have then estimated the coefficients of a functional linear regression in which the independent variable is CBF(.) and the response variable is either scalar (logarithm of monthly mean of NBF) or function (log NBF(.)), and reported their related interpretation and models assessment.

Acknowledgements

The authors are very grateful to the reviewers for helpful comments.

References

[1] Brillinger, D.R. (1975). Time Seris, Data Analysis and Theory. *New York: Hlot, Rinehart and Winston.*

[2] Cupidon, J., Eubank, R., Gilliam, D. and Ruymgaart, F., (2008). Some properties of canonical correlations and variates in infinite dimensions. *J. Multivariate Anal.* **99**, 1083-1104.

[3] Dauxois, J. and Nkiet, G.M. (1997). Canonical analysis of two Euclidean subspaces and its applications. *Linear Algebra Appl.*, **264**, pp. 355-388.

[4] Dauxois, J. and Nkiet, G. and Romain, Y., (2004). Canonical analysis relative to a closed subspace, *Linear Algebra Appl.* **388**, 119-145.

[5] Dauxois, J. and Pousse, A., (1976). Une extension de l'analyse canonique. quelques applications, *Ann. Inst. H. Poincare Probab. Statist.* **11**, 355–379.

[6] Eubank, R. L. and Hsing T. (2008). Canonical correlation for stochastic processes. *Stochastic Processes and their Applications.*, **118**, pp. 1634-1661.

[7] Hannan, E.J. (1961). The general theory of canonical correlation and its relation to functional analysis. *J.Austral. Math.Soc.* , **2**, 229-242.

[8] Hannan, E.J. (1961). The general theory of canonical correlation and its relation to functional analysis. *J.Austral. Math.Soc.* , **2**, 229-242.

[9] He, G. and Müller, H.-G. and Wang, J.-L. (2000). Extending correlation and regression from multivariate to functional data. *Asymptotics in Statistics and Probability, M.L. Puri (Ed.)*, pp. 1-14.

[10] He, G., Müller, H.-G. and Wang, J.-L., (2003). Functional canonical correlation analysis for square integrable stochastic processes, *J. Multivariate Anal.* **85**, 54–77.

[11] He, G. and Müller, H.-G. and Wang, J.-L. (2004). Methods of canonical analysis for functional data. *Journal of Statistical Planing and Inference.*, **122**, pp. 141-159.

[12] Hannan, E.J. (1961). The general theory of canonical correlation and its relation to functional analysis. *J.Austral. Math.Soc.* , **2**, 229-242.

[13] Hannan, E.J. (1961). The general theory of canonical correlation and its relation to functional analysis. *J.Austral. Math.Soc.* , **2**, 229-242.

[14] He, G. and Müller, H.-G. and Wang, J.-L. (2003). Functional canonical analysis for square integrable stochastic processes. *Multivariate Analysis.*, **85**, 54-77.

[15] Hannan, E.J. (1961). The general theory of canonical correlation and its relation to functional analysis. *J.Austral. Math.Soc.* , **2**, 229-242.

[16] Hotelling, H. (1936). Relation between two sets of variates. *Biometrika*, **28**, 321-377.

[17] Leurgans, S.E. and Moyeed, R.A. and Silverman, B.W. (1993). Canonical corelation analysis when data are curves. *J.R. statist. Soc. B*, **55**, 725-740.

[18] Kupresanin, A., Shin, H., King, D., and Eubank, R. L. (2010). An RKHS framework for functional data analysis, *Journal of Statistical Planing and Inference.*, **140**, pp. 3627-3637.

[19] Olshen, R.A. and Biden, E.N. and Wayatt, M.P. and Sutherland, D.H. (1989). Gait analysis and bootstrap. *Ann. Statist*, **17**, 1419-1440.

[20] Ramsay, J.O. and Silverman, B.W. (2005). Functional data analysis. *Springer,2nd edition, New York*.

[21] Ramsay, J.O. (1982). When data are functions. *Psychometrika*,**47**, 4, 379-396.

[22] Rice, J.A. and Silverman, B.W. (1991). Estimating the mean and covariance strucure nonparametrically when the data are curves.*J. Roy. statist. Soc. B*, **53**, 233-243.

[23] Service, S.K. and Rice, J.A. and Ghavez, F.P. (1998). Relation between physical and biological variables during the upwelling period in Monterey Bay. *Ca, Deep-Sea Research II Vol.*, **45**, 1669-1685.