AN APPLICATION OF THE CANONICAL CORRELATION

Reader, Dan Petru Vasiliu, Ph. D

Faculty of Economic Studies, University "Titu Maiorescu" – Bucharest

Senior lect. Ganea Tudor, Ph. D

Faculty of Economic Studies, University "Titu Maiorescu" – Bucharest

Abstract : *the paper deals with an experimental application of the canonical correlation in the study of the social environment in the last year Romania .*
To do this , two groups of variables are selected : one group repersenting the social effort and the other the social effect of the labour.
There studied the characteristic of the dependence between them.
The conclusion is that – in the considered period – the greater was the effort, the smaller was the effect .
Two different variants of the canonical correlation method were used to prove this
Both of them confirm this conclusion.

Keywords: canonical correlation ; canonical variates; the COREMAX method
multivariate analysis

The aim of the classical canonical correlation analysis is studying relationships between sets of multiple dependent and multiple independent variables.
So, let X be a p – dimensional and Y be a q – dimensional random variable .
We look for two vectors *a* , *b* such that $\mathbf{a} \in \mathbf{R}^p, \mathbf{b} \in \mathbf{R}^q$ and the quadratic **KENDALL** correlation coefficient $\mathbf{correl}^2(\mathbf{a} \cdot \mathbf{X}, \mathbf{b} \cdot \mathbf{Y})$ be maximal.

Note that the univariate random variable **U = a·X** and **V = b·Y** are known as ***the canonical variates*** of the variables X, resp. Y ,and that **a , b** are the corresponding ***canonical coefficients*** .
This problem was solved long before (from **Harold Hotteling** -1939 to **Joseph F. Hair** ,1998 for example).
In this paper the canonical correlation method will be applied to study dependence between the group

X = { civil employment ; employment rate }

as the group of independent variables , and

Y = { average net nominal monthly earnings ; total expenditure of household }
as the group of independent variables.
The data in our study are the next (http://www.insse.ro/cms/files/pdf/en/cp3.pdf)

Year	X1	X2	Y1	Y2
1997	11050	60.9	63.1	138.58*
1998	10845	59.6	104	196.31*
1999	10776	63.5	152	274.21*
2000	10508	63.6	214	375.92*
2001	10440	62.9	302	516.52
2002	9234	58	379	651.66
2003	9223	57.8	484	781.45

2004	9158	57.9	599	1049.44
2005	9147	57.7	746	1149.33

Here , as stated above :

X_1 = *civil employment*
X_2 = *employment rate*
Y_1 = *average net nominal monthly earnings*
Y_2 = *total expenditure of household* .

(The stared data are author's estimations).
First there proceeds to normalize this data ,in order to obtain a less error level (as is well known, the normalized variable $\hat{V}$ that corresponds to the variable V , is given by

$$\hat{V} = \frac{V - E(V)}{\sigma(V)} ,$$

where **E(V)** is the expected value ,and $\sigma(V)$ - the standard deviation of the variable V).
The normalized variables are then represented by the next table :

Year	$\hat{X}_1$	$\hat{X}_2$	$\hat{Y}_1$	$\hat{Y}_2$
1997	1.217	0.268	-1.174	-1.180
1998	0.970	-0.238	-0.999	-1.023
1999	0.886	1.281	-0.794	-0.810
2000	0.563	1.320	-0.530	-0.532
2001	0.480	1.047	-0.154	-0.147
2002	-0.977	-0.861	0.174	0.222
2003	-0.990	-0.939	0.623	0.577
2004	-1.068	-0.900	1.113	1.310
2005	-1.082	-0.978	1.741	1.583

The estimated total correlation matrix of this last dataset is then :

$$S = \begin{pmatrix} S_{XX} & S_{XY} \\ S_{YX} & S_{YY} \end{pmatrix}, \text{with:}$$

$$S_{XX} = \begin{pmatrix} 1 & 0,7834 \\ 0,7834 & 1 \end{pmatrix}; \quad S_{XY} = \begin{pmatrix} -0,9167 & -0,9268 \\ -0,6744 & -0,6783 \end{pmatrix}$$

$$S_{YX} = \begin{pmatrix} -0,9167 & -0,6744 \\ -0,9268 & -0,6783 \end{pmatrix}; \; S_{YY} = \begin{pmatrix} 1 & 0,9957 \\ 0,9957 & 1 \end{pmatrix}$$

There determines now the largest eigenvalues of the matrices

$$\mathbf{A} = (\mathbf{S}_{XX})^{-1} \cdot \mathbf{S}_{XY} \cdot (\mathbf{S}_{YY})^{-1} \cdot \mathbf{S}_{YX} = \begin{pmatrix} \mathbf{0{,}9586} & \mathbf{0{,}6959} \\ \mathbf{-0{,}1216} & \mathbf{-0{,}0850} \end{pmatrix}$$

and

$$\mathbf{B} = (\mathbf{S}_{YY})^{-1} \cdot \mathbf{S}_{YX} \cdot (\mathbf{S}_{XX})^{-1} \cdot \mathbf{S}_{XY} = \begin{pmatrix} \mathbf{-0{,}7034} & \mathbf{-0{,}7152} \\ \mathbf{1{,}5554} & \mathbf{1{,}5770} \end{pmatrix}$$

The (common) largest eigenvalue of the matrices **A, B** is **λ = 0,87 .**
This is equally the quadratic measure of dependence between the two groups ,
{ X_1,X_2} and **{ Y_1,Y_2 }** (details will be seen later).
The canonical coefficients are the eigenvectors , corresponding to **λ = 0,87 :** there obtain

a = (0,992 ; - 0,1263) , **b = (- 0,4138 ; 0,9104)**

therefore

U = 0,992·X_1 - 0,1263·X_2 ; V = - 0,4138·Y_1 + 0,9104·Y_2 .

U	V
1.17382	-0.5893
0.99194	-0.5178
0.71784	-0.4087
0.39175	-0.2649
0.34461	-0.0703
-0.8602	0.13015
-0.8636	0.26788
-0.9464	0.73189
-0.9498	0.72111

If compute now the correlation coefficient **correl (U,V) ,** there obtains

correl (U,V) = - 0,93285 ;

obviously , [**correl (U, V)**]^2 = **λ .**

For **correl (U,V) < 0 ,** there results that the dependence of **{ Y_1,Y_2 }** with respect to **{ X_1,X_2}** is of negative type : the greater **{ X_1,X_2}** become , the poor **{ Y_1,Y_2 }** become**.**

Otherwise, **correl (U,V)** being so near of -1 , the dependence is strong enough to let us consider the group **{ X_1,X_2}** determinative for **{ Y_1,Y_2 }.**

The values of the canonical variates are processed below :

$\tilde{X}_1$	$\tilde{X}_2$	$\tilde{Y}_1$	$\tilde{Y}_2$
1	0.542	0	0
0.892	0.322	0.060	0.057

0.856	0.983	0.130	0.134
0.715	1.000	0.221	0.235
0.679	0.881	0.350	0.374
0.046	0.051	0.463	0.508
0.040	0.017	0.616	0.636
0.006	0.034	0.785	0.901
0	0.000	1	1

A reason the canonical correlation method is not used so widely is that the practical meaning of the canonical coefficients is rarely clear .

To remediate that , different versions of the method are improved.

One of them is the **COREMAX** method (see (2)) , that gives the coefficients the meaning of ***degrees of contributions*** of each component of the group to the corresponding canonical variate.

To do this, we'll denote:

$$\eta=\alpha\cdot X_1+(1-\alpha)\cdot X_2\,;\quad 0\le\alpha\le1$$
$$\mu=\beta\cdot Y_1+(1-\beta)\cdot Y_2\,;\quad 0\le\beta\le1$$

Now, we look for values $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ that maximize **correl** $(\eta\,,\mu)$ **.**

To make economic interpretation possible , we'll enforce to the variables another kind of being comparable , instead of normalization : thus , the ***normed variables*** are more adequate .This is justified by the fact that negative values of the variables are equally difficult to interpret ; or, the normalization unavoidable gives negative values to normalized variables.

Remind that , for a variable series **V ,** the normed series $\tilde{V}$ is given by

$\tilde{X}_1$	$\tilde{X}_2$	$\tilde{Y}_1$	$\tilde{Y}_2$
1	0.542	0	0
0.892	0.322	0.060	0.057
0.856	0.983	0.130	0.134
0.715	1.000	0.221	0.235
0.679	0.881	0.350	0.374
0.046	0.051	0.463	0.508
0.040	0.017	0.616	0.636
0.006	0.034	0.785	0.901
0	0.000	1	1

$$\tilde{V}=\frac{V-\min\{V\}}{\max\{V\}-\min\{V\}}\ .$$

There obtaines :

Applying technique described in (2) , there obtains

α = 0,47 ; β = 0,70 ; max Correl(μ ; η) = - 0,8376

this value being close enough to **max correl (U,V) = - 0,93285 .**

Finally : in the next ,

- the canonical variate $\mu = 0{,}47 \cdot X_1 + 0{,}53 \cdot X_2$ gets the name of " synthetical employment"
- the canonical variate $\eta = 0{,}7 \cdot Y_1 + 0{,}3 \cdot Y_2$ gets the name of " synthetical earning"

Then: - the synthetical earning depends strongly from the synthetical employment ;

- the intensity of this dependence is about 83,76%
- the dependence is of the negative type : the greater the synthetical employment , the smaller the synthetical earning becomes .

In other words : in the mentioned period , the greater was the effort,
the smaller was the effect .

BIBLIOGRAPHY:

(1) :Multivariate Data Analysis, 5th edition by Joseph F. Hair, Jr., Rolph E. u.a. Prentice Hall, Inc. 1998.

(2) : Metode matematice pentru fundamentarea deciziilor in productie , by D.P.Vasiliu , Editura Tehnica , Bucuresti, 1986