International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 6 Issue 5, May - 2022, Pages: 214-221

Canonical Correlation Method for Studying the Relationship between Exports and Imports in Sudan (2002– 2020)

[1]Hisham Hamadna Allah Abdel Fattah Jobarat Allah, [2]Altaiyb Omer Ahmed Mohmmed

[1]Sudan University of Science and Technology - College of Science +249124444056 - hisho536@gmail.com
[2]Sudan University of Science and Technology - College of Science +249966693330- ibnomer8090@gmail.com

Abstract: *In this study, using the canonical correlation analysis to estimate the relationships between imports and exports in Sudan. For this purpose, it was designed to evaluate the relationship between two sets of imports variables: industrial and agriculture as the first set of variables (Y) and as the second set of variables (X) by using canonical correlation analysis. Estimated canonical correlations between the first and the second pair of canonical variates were significant (Chi-square calculated greater than tabulated). Canonical weights and loadings from canonical correlation analysis indicated that industrial had the largest contribution as compared with agriculture imports ant agricultures exports, number of variables represent exports and the others representing imports, which made us use the canonical correlation analysis method as a multivariate methods variables to figure out the co-linearity relationship between the variables in the two groups, so we calculate all the combinations of linear phase between the two groups and comparing this with some others, to reached greatest relationship,and has been found that there is a relationship between exports and imports for each variable in the two groups.*

Keywords: Trade Balance, Canonical Correlation, Exports, Correlation Matrix

Introduction:

The purpose of canonical correlation between two groups is to find the common relationship between them, researcher may face two large groups, for each group a number of variables to be exploration for the co-linearity correlations between those variables, and here the issue of Canonical correlation is considered one of the important topics in statistics that deal with such problems, in 1947 Bartlett had a paper entitled "General Distribution of Canonical Engagement." In 1952, Marriott published some tests of morality in Canonical analysis, and in 1959, the scholar Lawley also published some tests in this type of analysis. In 1971, Hodge and Klatzky presented an orthodox analysis of a mobile professional nature (Batlett 1947).

Canonical correlation analysis is a technique to extract common features from a pair of multivariate data. In complex situations, however, it does not extract useful features because of its linearity (Akaho 2007)

CCA is the main technique for two-set data dimensionality reduction such that the correlation between the pairwise variables in the common subspace is mutually maximized (Yang 2021)

Since only a few canonical variates are needed to represent the association between the two sets of variables, canonical correlation analysis is a data reduction technique (Sharma, 1996).

In the event that we imagine that we have two sets of variables that have a joint distribution, then the analysis is carried out for the correlations between the variables of one of the groups and the variables of the other groups with the aim of obtaining a harmonized system in the space of each group of variables, and in the same way, the new coordinate presents the system of correlations clearly There is no ambiguity in it. You will find that the linear combination of the variables in each group has a correlation that will be at its greatest degree (Rupnik & taylor 2010). These are the linear combinations or the harmonies in the new system. Thus, the second linear combinations in each group will require such a connection between the great correlations of the linear combinations as if they were not related to the first linear combinations. This method will continue until the two new coordinate systems are fully set (Frie & Christian 2009). Each group has a number of variables. Variables are sometimes used and applied in a misplaced manner. The importance of this research comes from the fact that it is one of the few studies that dealt with linking canonical correlation analysis as one of the methods of multivariate analysis, using some parametric statistical tests as a case study by applying it (Kettenring 1971).

Methods and material

The method used in this research is the analytical method or method, using the multivariate analysis method in some statistical tests as a new analysis method represented in the analysis of canonical correlation by dividing the data into two groups as independent variables and dependent variables, analyzing each group separately and finding the covariance between the two groups. The descriptive method describes how the multivariate method works and gives a general idea of statistical tests and how to apply, explain and explain the canonical link and the steps of the solution. Therefore the canonical variates representing the optimal linear combinations of dependent and independent variables and the canonical correlation showing the relationship between them are results of interest (Hair et al., 1998).

Deriving canonical correlation

To determine the relationship between $U_1 \; V_1$ The vector values must be established $\left(\alpha' \text{و} \delta\right)$ where: $E\left(U' X\right) = \alpha' \left(V_1 U\right)\alpha$ $= \alpha' E X_1 X_1' \alpha = \alpha' V_{11} \alpha$

So $E\left(V' V\right)$ he is

$$E\left(V'V\right) = \delta' V_{22} \delta^2$$

and

$$E\left(U'V\right) = \alpha' V_{22} \delta$$

if

$$\alpha' V_{11} \alpha = 1/$$
$$\therefore VAR\left(U\right) = \alpha' V_{11} \alpha = 1$$

$$E\left(V'V\right) = \delta' V_{22} \delta' = 1 = VAR\left(V\right)$$
$$E\left(U'V\right) = \alpha' V_{22} \delta = r_{u,v}$$

The variable has a conventional normal distribution because the anticipated values are zero $X \sim N\left(0,1\right)$ that is (Luijtens et al.1994):

$$r_{x,y} = \frac{\operatorname{cov}\left(X,Y\right)}{\sqrt{V\left(X\right)V\left(Y\right)}} = \operatorname{cov}\left(X,Y\right)$$

Therefore, it is:

$$r_{U,V} = \frac{\operatorname{cov}\left(U,V\right)}{\sqrt{V\left(U\right)}\sqrt{V\left(V\right)}} = \operatorname{cov}\left(U,V\right)$$

The following are some of the distinctions between legal association analysis and other multivariate statistical methods:

The main difference between canonical correlation analysis and other statistical methods like cluster analysis and factor analysis is that canonical correlation analysis investigates the relationship between two groups of variables after ensuring that they are linked by a linear relationship and determining the greatest correlation between them. Cluster analysis is one of the classification methods involved with classifying groups into two or more groups, whereas factor analysis is used to find the most relevant factors influencing a phenomenon based on a matrix of factor loadings and prevalence values (Lawley 1959). Assume there are two sets of variables: explanatory variables and outcome variables $\left(X_i\right)$ and by quantity $i = 1, 2, \ldots, q$ The second category includes response variables $\left(Y_i\right)$ and by quantity $i = 1, 2, \ldots, p$ ، $p, q > 1$ and that $\mu 2$, $\mu 1$ Both groups' arithmetic mean, respectively.

By studying the canonical correlation, we are attempting to identify the most appropriate linear correlation that provides the greatest correlation between the linear components, implying that our focus will be on determining the values of each linear component $W_{(indep)}$ and $(W_{(dep)}$ that strengthens the bond between V, U

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 6 Issue 5, May - 2022, Pages: 214-221

$$V = W\left(\overset{T}{indep.}\right).^{X_1}\ U = W\left(\overset{T}{dep.}\right).^{Y_i}$$

(transactions.), $(W_{(dep)}$ and Transactions of (X_i) and (Y_i) Which correspond to their appropriate training weights (Golob 1985).

And Knowing the explanatory factors' variance matrix $(\sum_{11})$ and the response variables matrix $(\sum_{22})$ and the matrix of their covariance $(\sum_{12})$ and the matrix. They have a link (K) where

$$\sum_{11} = E\{(x-u_1)(x-u_1)^T\} \text{ ، } \sum_{11} = E\{(x-u_1)(x-u_1)^T\}$$

$$\sum_{12} = E\{(x-u_1)(y-u_2)\}$$

$$K = \sum_{11}{}^{-\frac{1}{2}} \sum_{12} \sum_{22}{}^{-\frac{1}{2}}$$

So, let's say we have the typical vectors (Lawley 1959) let (α_i) for the matrix (KK^T) Vectors of characteristics -(β_i) for the matrix $(K^T K)$. As a result, we can calculate the value of the first Canonical correlation (the values that have the largest correlation) as follows:

$$W_{(dep.)} = \sum_{22}{}^{-\frac{1}{2}} \beta_i, W_{(indep.)} = \sum_{11}{}^{-\frac{1}{2}} \alpha_i,$$

Canonical correlation analysis is a less commonly utilized multivariate method, partly due to the difficulties in organizing and analyzing the data. Multiple regression, which is concerned with determining the relationship between variables, is presented as an extension of Canonical correlation (x_i), Variables that explain and predict (Y_i), The Canonical correlation, on the other hand, is concerned with determining the linear structure of a set of explanatory variables x_i , $i = 1, 2, \dots p$, A set of response variables is combined in a logical way , y_i , $i = 1, 2, \dots p$ That is, the former is concerned with a lot of expected variables, while the latter is not (Klatzky, & Hodge 1971) .

Statistically Canonical Correlation Analysis:

The following algorithm quantitatively summarizes the steps of Canonical correlation analysis. These stages are preceded, however, by computing the correlation matrix between the variables (explanatory and response), with the following algorithm (Luo te al 2015):

1- Using the relationship: Taking the deviation of the values of the variables from their arithmetic mean (Batlett 1947).

$$y_i = y_i - \bar{y},$$
$$x_i = x_i - \bar{x}$$

As a result, the average of each x_i, y_i the same as zero

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 6 Issue 5, May - 2022, Pages: 214-221

2- Determine a counter's beginning value (I) in which $(I = 0)$ Establish constants (q) Explanatory variables count (The first group). (p) The second category (the number of response variables) (Madrigal 2016).

3- Determine the linear structures of all explanatory variables, including their symbols, as well as the response variables (V_1) represents him (U_1) Which is the first Canonical correlationvariable at this point.

4- Calculate the correlation coefficient between $(U_1 V_1)$ which is the first Canonical correlationvariable.

5- Increase the counter value by (1) $I = I + 1$.

6- Finding the linear structures for each of the explanatory variables and its symbol (V_2) and It is denoted by for response variables (U_2) Which is referred to as the second Canonical correlationvariable at this point So that it meets the following criteria:

a. (V_2) Not connected with $(V_1\ ,\ U_1)$.

b. (U_∂) Not connected with $(V_1\ ,\ U_1)$.

c. Whoever $(U_2\ and\ V_2)$ they have the strongest potential correlation.

7- Calculate the correlation coefficient between $(U_2\ and\ V_2)$ As a result; condition (c) is satisfied.

8- Compare the resultant counter value to the smallest value between (p) We repeat the smallest value from step to step (5).

9- Using the hypothesis test to determine the significance of the canonical correlation coefficients $(: H\,0)$ The canonical correlation coefficients' values (first, second, ... =zero).

$: H_1$ Using the Bartlett test or Chi-square, the canonical correlation coefficients are $\neq$ zero.

Application and results:

The data was divided into two groups, the first representing exports and the second representing imports. The simple correlation was calculated for each group separately, and then the correlation between the two groups was calculated and then applied to the commodity groups on exports and imports, which include the following variables:

Table (1) descriptive statistics

variable		mean	standard deviation
exports	agricultural	383	184.063
	industrial	97.0526	92.013
imports	agricultural	87.283	75.278
	industrial	1517.915	1185.855

The average and standard deviation of the variables was calculated for both agricultural and industrial for the exports and imports table (1) shows the results.

Matrix calculation

Now find the matrix of imports and exports for the variables

Simple Export Matrix from 2002 to 2020:

$$H_0 : R_x = \begin{pmatrix} 1 & 0.866 \\ 0.866 & 1 \end{pmatrix}$$

Simple Import Matrix from 2002 to 2020:

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 6 Issue 5, May - 2022, Pages: 214-221

$$H_0 : R_y = \begin{pmatrix} 1 & 0.652 \\ 0.652 & 1 \end{pmatrix}$$

Creating a simple correlation matrix between variables:

$$H_0 : R_{xy} = \left(\begin{array}{cc|cc} 1 & 0.866 & 0.473 & 0.038 \\ 0.866 & 1 & 0.160 & -0.228 \\ \hline 0.473 & 0.160 & 1 & 0.652 \\ 0.038 & -0.228 & 0.652 & 1 \end{array}\right)$$

One set will be the first two variables, and another set will be the latter two variables. To get the set's principal correlation, solve the equation:

$$\left[\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12}\boldsymbol{R}_{22}^{-1}\boldsymbol{R}_{21} - \lambda^2 \boldsymbol{I} \right] = 0$$

$$\left[\boldsymbol{R}_{22}^{-1}\boldsymbol{R}_{21}\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{21} - \lambda^2 \boldsymbol{I} \right] = 0$$

And they both give $\left(\lambda^2\right)$ and A the same value:

$$\left(\begin{vmatrix} 3.999 & -3.463 \\ -3.462 & 3.999 \end{vmatrix}\right)^{-1} \left(\begin{vmatrix} 0.473 & 0.038 \\ 0.160 & -0.228 \end{vmatrix}\right) \left(\begin{vmatrix} -0.00236 & 0.00154 \\ 1.53736 & -0.00236 \end{vmatrix}\right)^{-1} \left(\begin{vmatrix} 0.473 & 0.038 \\ 0.160 & -0.228 \end{vmatrix}\right) - \left(\begin{vmatrix} \lambda^2 & 0 \\ 0 & \lambda^2 \end{vmatrix}\right)$$

where that $\left(\lambda^2 = r\right)$ if

$$\therefore \left(\begin{vmatrix} 0.683208 & 0.0549272 \\ -0.757505 & -0.0610795 \end{vmatrix}\right) - \left(\begin{vmatrix} r & 0 \\ 0 & r \end{vmatrix}\right) = 0$$

$$\therefore \left(\begin{vmatrix} 0.683208 - r & 0.0549272 \\ -0.757505 & -0.0610795 - r \end{vmatrix}\right) = 0$$

$$\left(0.683208 - r\right)\left(-0.0610795 - r\right) - \left(-0.757505\right)\left(0.0549272\right) = 0$$

$$r_1 = 0.6921 \quad \lambda_1 = \sqrt{0.6921} = 0.8319$$

$$r_2 = 0.4336 \quad \lambda_2 = \sqrt{0.4336} = 0.6585$$

We'll now perform the hypothesis test.

$H_0 :$ The main correlation is not significant

$H_1 :$ The main correlation is significant

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 6 Issue 5, May - 2022, Pages: 214-221

$$\therefore T_1 = \pi\left(1-\lambda_j^2\right)$$
$$=\left(1-0.6921\right)\left(1-0.4336\right)=0.1744$$

Whereas

$$-\left[\left(N-1\right)-\left(\frac{P_1+P_2+1}{2}\right)\right]InT_1 \square \chi^2_{(P_1-P_2)}$$
$$=-\left[\left(19-1\right)-\left(\frac{4+4+1}{2}\right)\right]In\left[0.1744\right]=-\left[18-4.5\right]In\left[0.1744s\right]$$
$$-13.5In\left[0.1744\right]=23.5764$$

The value above is calculated chi-square. We get the tabulated chi-square with 1 degree of freedom $(\chi^2{}_{(1,0.025)})$ equal (5.024)

So 5.024 < 23.5764

As a result, the null hypothesis is rejected, implying that either $\left(\lambda_1,\lambda_2\right)$ is significant, or at least one of them is relevant.

We must complete the solution to clarify the universe since we reject the null hypothesis $\left(\lambda_1\right)$ (that First C.C).

As a result, we propose a new hypothesis:

$$H_0:\lambda_2=0$$
$$H_0:\lambda_2\neq 0$$

This null hypothesis is based on the assumption that (First C.C) is eliminated or removed, and the test is undertaken for the rest, as previously stated in the mathematical aspect, and it is as follows:

$$T_2=-\left[N-1-\frac{\left(P_1+P_2+1\right)}{2}\right]InT_1$$

$$\therefore T_2=-13.5\ln\left(1-0.4336\right)=-13.5\left(-0.5685\right)=7.6748$$

the value of $\left(\chi^2\right)$ tabulated is the same as (5.024), that is means the null hypothesis is rejected and (λ_2) It is also sig, we use $\left(\lambda_1 \ and \ \lambda_2\right)$ To determine each value $\left(\alpha \ and \ \gamma\right)$.

$$\therefore \begin{pmatrix} 0.683208-\lambda_1^2 & 0.0549272 \\ -0.757505 & -0.0610795-\lambda_2^2 \end{pmatrix}\begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}$$

We find that:

$$\alpha_1=0.75668$$
$$\alpha_2=1.15871$$

Therefore:

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 6 Issue 5, May - 2022, Pages: 214-221

$$\left(R_{s11}^{-1}R_{12}R_{22}^{-1}R_{21}-\lambda^{2}I\right)\begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix}=\begin{pmatrix}0\\ 0\end{pmatrix}$$

We find that:

$$\alpha_1=-0.580609$$
$$\alpha_2=0.643962$$

And we can deduce from everything that has been said and the values that:

$$U=-0.580609X_1+0.643962X_2$$
$$V=0.8319X_3+0.6585X_4$$

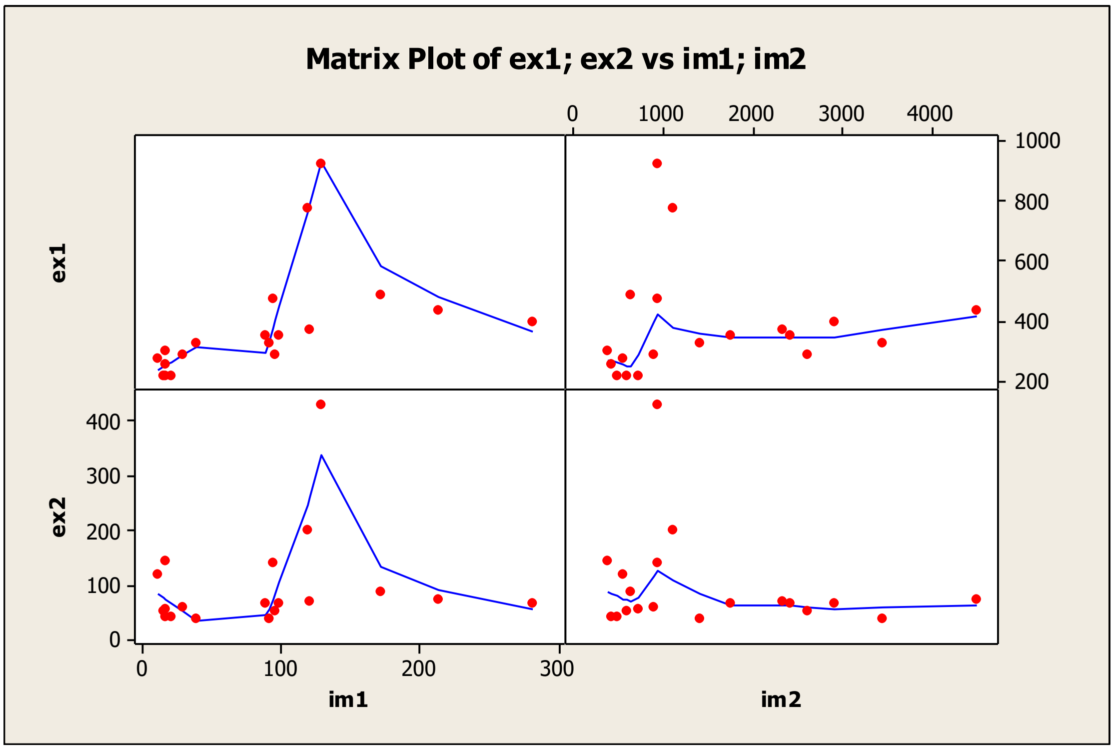

Figure (1) the matrix of exports and imports

where: ex1 represents agricultural exports, ex2 represents industrial exports, im1 represents agricultural imports and im2 represents industrial imports.

Results:

According to the report, average agricultural exports for the period were (383) million dollars, while average agricultural imports were (87,283) million dollars, indicating that exports are better than imports. The average industrial exports were $97.0526, whereas the average industrial imports were (1517.915) million dollars, indicating that industrial exports are in decline. The canonical link between the variables in the first group was (0.6921), while the relationship between the variables in the second group was (0.4336), indicating that the first group's link is stronger than the second group's.

Recommendations:

Preserving, developing and paying attention to agricultural exports, local industry and benefiting from the agricultural products that are exported, developing and promoting Sudan's bilateral and multilateral foreign trade relations, activating Sudan's role and presence in regional and international organizations and gatherings, benefiting from trade agreements and protocols in providing the country's needs for goods with the best Conditions and by increasing our exports to foreign markets, spreading a culture of knowledge of the importance of joining economic blocs and benefiting from Sudan's accession to regional groupings such as GAFZT and COMESA in increasing exports and providing goods,Obtaining technical assistance from international and regional organizations, as well as working and coordinating with them to benefit from the programs available to LDCs, particularly in the areas of development, technology transfer, and capacity building.

references:

1- Batlett, M.S.(1947); "The general canonical correlation distribution", Ann.Math. statist. 18.P.1.
2- Frie .G.K , Janssen . Christian , "Social inequality lifestyles and health – a non-linear canonical correlation analysis based on the approach of Pierre Bourdi ", Birkhäuser Verlag, Basel , 54 , 2009.
3- Golob , Thomas F.," A Non-linear Canonical Correlation Analysis of Weekly Trip Chaining Behavior", Institute of Transportation Studies University of California, Irvine , 1985.
4- Hun Lee Hun Lee ,J. et al.,"An Application of Canonical Correlation Analysis Technique to Land Cover Classification of LANDSAT Images", ETRI Journal, Volume 21, Number 4, December 1999 .
5- Klatzky, S. R, and Hodge, R.W.(1971)"A canonical analysis of occupational mobility". JASA, 66P.16.
6- Lawley, D .N.(1959)."test of significance in canonical analysis", Biometrika,46,P.50.
7- Luijtens .K , Symons .F & Vuylsteke-Wauters . M , " Linear and non-Linear canonical correlation analysis" : an exploratory tool for the analysis of group – structured data " , Volume 21, Issue 3, January ,1994, p 43-61.
8- Madrigal .P , " fCCAC: functional canonical correlation analysis to covariance between nucleic acid sequencing datasets " , bioRxiv preprint first posted online Jun. 28, 2016.
9- Ouali . D , Chebana . F, Ouarda . T. B. M. J ," Non-linear canonical correlation analysis in regional frequency analysis", Stoch Environ Res Risk Assess , 30:pp 449–462 , 2016 .
10- Weenink .D ,"Canonical correlation analysis". Institute of Phonetic Sciences , University of Amsterdam , Proceedings 25 ,pp 81–99 , 2003 .
11- Yazici . A .C , et , "An application of nonlinear canonical correlation analysis on medical data " , TÜBİTAK , 40 (3):pp 503-510 , 2010 .
12- Yin . Xiangrong ,"Canonical correlation analysis based on information theory ", Journal of Multivariate Analysis 91, pp161–176 , 2004 .
13- X. Yang, W. Liu, W. Liu and D. Tao, "A Survey on Canonical Correlation Analysis," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 6, pp. 2349-2368, 1 June 2021, doi: 10.1109/TKDE.2019.2958342.
14- Shotaro Akaho, A kernel method for canonical correlation analysis, arxiv cs cornell university 2007.
15- Hair J.F., Anderson R.E., Tatham R.H., William C.B. (1998): Multivariate Data Analysis. 5th Edition. Prentice Hall, Inc., 442–462.
16- J. R. Kettenring, "Canonical analysis of several sets of variables", Biometrika, vol. 58, no. 3, pp. 433-451, 1971.
17- J. Rupnik and J. Shawe-Taylor, "Multi-view canonical correlation analysis", Proc. Conf. Data Mining Data Warehouses, pp. 1-4, 2010.
18- Y. Luo, D. Tao, K. Ramamohanarao, C. Xu and Y. Wen, "Tensor canonical correlation analysis for multi-view dimension reduction", IEEE Trans. Knowl. Data Eng., vol. 27, no. 11, pp. 3111-3124, Nov. 2015.