UNIVERSITY OF CAPE COAST

DISCRIMINANT CANONICAL CORRELATION ANALYSIS OF TIME-DEPENDENT MULTIVARIATE DATA STRUCTURE

KWAME ASARE GYASI-AGYEI

2024

UNIVERSITY OF CAPE COAST

DISCRIMINANT CANONICAL CORRELATION ANALYSIS OF TIME-DEPENDENT MULTIVARIATE DATA STRUCTURE

BY

KWAME ASARE GYASI-AGYEI

Thesis submitted to the Department of Statistics of the School of Physical Sciences, College of Agriculture and Natural Sciences, University of Cape Coast, in partial fulfilment of the requirements for the award of Doctor of Philosophy degree in Statistics

MARCH 2024

DECLARATION

Candidate's Declaration

I hereby declare that this thesis is the result of my own original research and that no part of it has been presented for another degree in this university or elsewhere.

Candidate's Signature .... Date ....

Name: Kwame Asare Gyasi-Agyei

Supervisors' Declaration

We hereby declare that the preparation and presentation of the thesis were supervised in accordance with the guidelines on supervision of thesis laid down by the University of Cape Coast.

Principal Supervisor's Signature: .... Date ....

Name: Prof. Bismark Kwao Nkansah

Co-Supervisor's Signature: .... Date ....

Name: Dr. Francis Eyiah-Bediako

ABSTRACT

Canonical Correlation Analysis (CCA), which is a widely used covariance analysis method is a technique that is not fundamentally designed for multivariate multiple time-dependent data (MMTDD) structure that could be suitably partitioned on two subsets of response and predictor variables. This means that for such data problems, the conventional CCA would not yield practical results. The literature also shows scanty work in this area. This study therefore designs and implements grouping scheme discriminant canonical correlation analysis (GSDCCA) for handling this problem so that the time effect is adequately captured in the computation of the correlation coefficient between the two sets of variables. It first identifies key matrices underlying the concert and presents the design in both theory and illustration. Using data on six weather conditions in Ghana spanning the period 2000 to 2021, the demonstrations show that correlation coefficient between heating and cooling sets of weather conditions varies at different time points, and that the overall correlations are quite higher than that obtained from data assumed to be time-independent. The procedure is therefore recommended as an innovative approach for handling MMTDD.

KEY WORDS

Canonical correlation analysis

Canonical Discriminant Functions

Eigenvalues and Eigenvectors

Grouping schemes

Time-dependent multivariate data

Weather conditions

ACKNOWLEDGEMENTS

My sincere thanks to God the Father and the Lord Jesus Christ for their protection and direction, which allowed me to finish this thesis. Professor Bismark Kwao Nkansah and Dr. Francis Eyiah Bediako, both of the Department of Statistics, who served as my supervisors for this thesis, have my sincere gratitude for their leadership, help, and overall professionalism. I genuinely appreciate it. I want to thank all of the Senior members at the University of Cape Coast's Statistics Department for their help, motivation, careful consideration during my presentations, and the subsequent comments that improved the quality of the thesis.

I also appreciate the guidance and kind efforts of my Vice Chancellor, Professor David Kofi Essumang, Koforidua Technical University, and all Senior members, particularly those in the Faculty of Applied Science and Technology. I would especially want to thank my elder brothers, Pastor Gabriel Boakye Dankwa, Doctor Nyamekye Gyasi-Agyei, Professor Yeboah Gyasi-Agyei, and Professor Amoakoh Gyasi-Agyei, as well as other family members, for their support, encouragement, and advice.

I acknowledged the special contribution from RUFORUM, University of Cape Coast, and Koforidua Technical University for their financial assistance. Finally, I would like to thank my wife Philomena Gyasi-Agyei and the children for their support during my three years academic career.

DEDICATION

To my late parents: Mr. and Mrs. Kofi Konadu Gyasi-Agyei

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATION

CCA	Canonical correlation analysis
CCC	Canonical correlation coefficient
CCF	Canonical correlation function
CL	Canonical loadings
C Class	Correct classification
CPct	Cumulative Percentage
DA	Discriminant analysis
DGCCA	Deep generalize canonical correlation analysis
DCCAFns	Discriminant canonical correlation Analysis functions
DCCA	Discriminant canonical correlation analysis
DV	Dummy variables
Eig V	Eigenvalue
EVD	Eigen Value decomposition
FDCC	First discriminant canonical correlation
fMRI	functional magnetic resonance imaging
GS	Grouping scheme
GSDCCA	Grouping scheme discriminant canonical correlation analysis
IEDisCA	Intra-class and extra-class discriminative correlation analysis
KCCA	Kernel canonical correlation analysis
LDA	Linear discriminant analysis
MANCOVA	Multiple analysis of covariance
MANOVA	Multiple analysis of variance
MDA	Multiple discriminant analysis
MGDA	Multiple-group discriminant analysis
MMTDD	Multivariate multiple time-dependent data
PMD	Penalized matrix decomposition
PMA	Penalized multivariate analysis

PVE	Proportion of variance explained
RCCA	Regularized canonical correlation analysis
SVD	Singular value decomposition
SCCA	Sparse canonical correlation analysis
Time-Dep	Time-dependent
Time-Ind	Time-independent
T-DCCA	Time-dependent canonical correlation analysis
T-DMD	Time-dependent multivariate data
T-DMGDCCA	Time-dependent Multiple-group discriminant CCA
UGC	Understanding global change

CHAPTER ONE

INTRODUCTION

The background includes the definitions of key concepts related to Canonical Correlation Analysis (CCA) and the characteristics of the study's illustrative data. It provides the primary motivation for the investigation, which will later be represented in the problem description. The study methodology and the data problem analysis are given adequate space in this chapter because of the study's relevance. CCA connects two sets of variables by finding linear combinations of variables that maximize correlation between them (Helwig, 2017; Hotelling, 1936). Helwig describes two typical uses for CCA - Data reduction and Data interpretation. A restricted set of linear combinations is utilized in data reduction to account for the co-variation between two sets of variables. Data interpretation is aided by identifying features (canonical variates) that are significant for explaining co-variation between two sets of variables. An attempt is made to clarify the ideas, provide evidence for their use, and talk about possible uses.

Background to the Study

One of the easiest methods for identifying commonalities among datasets is covariance analysis. CCA is one covariance analysis approach frequently used in statistics. It is a method used in many different fields to investigate the link between a set of variables that make up the same dataset—a set of predictor variables, **X**, and a set of response variables, **Y**. A variety of fields have used CCA in several applications (Samarov, 2009). Since its introduction by Hotelling (1935), Rencher (2002) led the development of a multivariate approach with the express goal of detecting the kinds of correlations that occur within one set of variables as well as between two groups of variables. The goal of CCA is to ascertain whether the criterion set of variables has an impact on the predictor sets of variables when it is known, based on some theory (Joshua, 2016).

Sharma, (1996) proposed the following four relevant situations for application of CCA.

1. Insurance providers are looking to see if there is a correlation between the types of insurance policies purchased and individual characteristics.

2. The availability of running water, heating and cooling conditions, kitchen and restroom facilities, and type of housing as quality-related factors that a health department is examining to determine if there is a correlation between these and incidences of minor and serious illness, as well as the number of disability days.

3. When it comes to several health-related issues including weight, stress levels, hypertension, and anxiety, a medical researcher is interested in whether people's dietary and lifestyle choices have an impact on their health.

4. A consumer goods company is curious to know if there is a correlation between the types of products purchased and consumer personalities and lifestyles.

Each of the aforementioned scenarios aims to establish whether there is a connection between two specific groups of variables. The best method for finding connections between two sets of data is the CCA. Meteorologists use CCA to look at the link between response variables like minimum temperature, maximum temperature, and solar radiation and a few other weather-related factors including wind, relative humidity, and precipitation. The intricacy of the data makes it necessary to model their interaction in order to predict weather conditions. Consequently, a natural framework for this kind of study is provided by CCA (Richardson, 1981).

A growing number of research make use of CCA as a key method. In the publications (Hair, Black, Babin, & Anderson, 2006; Rencher, 2002; Tandanai, 2015; Yang, Liu, Tao, & Cheng, 2017), it is established that regression analysis and CCA are two different things. Regression analysis is used to describe relationships and to predict outcomes when both independent and dependent variables are present. One dependent variable can be compared to a large number of independent variables using regression, and multiple regression analysis is a necessary step in the process. Regression analysis, then, focuses on one-to-many relationships, where as, CCA deals with multiple response variables versus multiple predictor variables (Dattalo, 2014; Zhihua & Zhen, 2010).

Same individuals are used to assess the p variables in the first group of variables and the q variables in the second group of variables. Given two datasets (Joshua, 2016), CCA generates as many canonical pairs of linear combinations as min(p, q). The objective of CCA in this part is to combine the two variables provided by

$$U_i = \alpha_i'\mathbf{Y} \qquad \text{and} \qquad V_j = \beta_j'\mathbf{X} \tag{1.1}$$

such that the correlation between the two linear functions U_i and V_j is maximized, where, $\alpha_i = (\alpha_{i1}, \alpha_{i2}, ..., \alpha_{1p})'$ and $\beta_j = (\beta_{j1}, \beta_{j2}, ..., \beta_{jq})'$ are the coefficient vectors of $\mathbf{Y}$ and $\mathbf{X}$, respectively. The very first step in the fundamental principle of CCA is to find one linear combination of the response variables, U_1, and one linear combination of the predictor variables, V_1 that constitute the same dataset (Lai & Fyfe, 1998).

Given the amount of attention that CCA has received over the past few years, it appears likely that the exploration of novel statistical modeling domains is just getting started. Due to its significance in mathematical and statistical analysis as well as other applications, the applicable data are those specifically designed for multivariate multiple regression that have more or less established names and notations (Mardia, Kent, & Bibby, 1979). Model generation and model optimization were the two viewpoints from which (Yang, Liu, Wei, & Tao; 2021) analyzed its theory. They have noticed that CCA has drawn a lot of interest as a potent approach for fusing multi-modal features.

Measuring the degree of the relationship between two sets of data is the goal of CCA, like that of regression analysis. It is comparable to factor analysis for constructing variable composites. It is also similar to discriminant analysis in that it can generate independent dimensions with the aim of obtaining the highest correlation between the dimensions for each set of variables. In order to maximize the relationship between response and predictor variable sets, CCA determines the best structure or dimensionality for each variable set (Roungu, Matair, Sanwar & Azizur, 2013).

The CCA technique, according to the literature (Benton, Huda, Biman, Reisinger, Zhang, & Arora, 2019), finds the maximally correlated linear projections of two random vectors. It is a basic multi-view learning strategy. When given two input views, $\mathbf{Y} \in \Re^{d_y}$ and $\mathbf{X} \in \Re^{d_x}$, using a cross-covariance matrix Σ_{12} and covariance matrices Σ_{11} and Σ_{22}, respectively, CCA finds the directions that maximize the correlation between the two input perspectives provided by

$$(U^*, V^*) = arg \max_{u,v} Corr\left(U'\mathbf{Y}, V'\mathbf{X}\right) \tag{1.2}$$

This formulation can be expressed as a constrained optimization as it is invariant to affine transformations of U and V given by

$$(U^*, V^*) = arg \max_{u,v} U'\Sigma_{12}V \tag{1.3}$$

under the conditions that $U'\Sigma_{11}U = V'\Sigma_{22}V = 1.$

Multivariate Multiple Time-dependent data (MMTDD), which is prevalent in many application domains, shows several attributes over time. Numerous other fields make use of the data that is produced when observations are made on a collection of units over time. Examples include, but are not limited to, the financial sector's asset price and volatility over time, the application of various weather conditions, the monitoring of heart or brain activity by medical equipment, the evolution of disease as indicated by appropriate biomarkers in epidemiology, and the nation's growth as indicated by economic indices (Casa, Bouveyron, Erosheva, & Menardi, 2021).

It is possible to treat each sample component as a function in this scenario because there are often multiple regularly sampled observations available. In contrast, longitudinal studies frequently only include a small number of observations throughout time with sparse and inconsistent measurements. A recent study of related techniques employed in CCA has revealed that, due to the fact that this is not always the case, there has been a growing interest in clustering methodologies that seek to describe heterogeneity among multivariate time-dependent observed trajectories. The methods that have been covered thus far are widely used when tracking a single attribute over an extended period of time for a large number of individuals.

The basic structure of the MMTDD is fairly represented in the literature (Casa et al., 2021) as $(n \times d \times T)$, where T represents the total number of time events, d represents the total number of time-dependent variables, and n represents the total number of observations (subject count). CCA and such data may actualy be regarded as time-independent if the effect of T is not appropriately captured in the designed of the procedure. In order to incorporate the time effect the stated basic structure of the MMTDD could further be partitioned as $(n_1 \times d \times t_1)$, $(n_2 \times d \times t_2)$, ..., $(n_r \times d \times t_r)$ for suitably determined t_r, $r < T$, such that the link between $\mathbf{Y}$ and $\mathbf{X}$ might be seasonally reflected in the average of the CCA. In this study we demonstrate how the optimal value of r is determined via multiple discriminant analysis. This procedure is what is referred to in this study as grouping scheme discriminant CCA (GSDCCA). Thus, the grouping scheme (GS) is an approach that is intended to incorporate the time effect into CCA via discriminant analysis (DA) in order to effectively handle the time-dependent multivariate data (TDMD).

Canonical correlation analysis shares fundamental implementation challenges with all other multivariate approaches. The impacts of measurement error, the categories of variables that can be used, and the varieties and transformations of those variables are all relevant to CCA (Luo, Dacheng, Yonggang, Kotagiri, & Chao, 2015). Without considering the effects on sample size, researchers are inclined to add a large number of variables in both the response and predictor variable sets. Very tiny sample sizes will not accurately reflect the connections, hiding any significant links. Even when practical significance is not indicated, very large samples will frequently imply statistical significance. A high dimensional link between two sets of variables is attempted to be reduced into a few pairs of canonical variables by the technique's maximizing component (Roungu et al., 2013).

The practical attractiveness of CCA is enhanced by the efficient and numerically stable computation of canonical correlations between linear sub-spaces. Canonical correlations are also generally unaffected by the uniform scaling of individual patterns, since they reflect changes within a set by a subspace. Because of this, CCA-based methods are particularly well-suited for a range of computer vision applications where variations may arise from variations in light or duration of exposure for the photosensitive material (Arandjelovic, 2013).

CCA is a data reduction strategy since it only requires a small number of canonical variates to indicate the relationship between two sets of variables (Sharma, 1996). According to Jaiswal, Poonia, and Kumar (1995), this successful multivariate technique has been widely accepted in a number of sectors, including education, marketing, psychology, social science, political science, ecology, and sociology-communication.

Statement of the Problem

CCA appears to underline several multivariate statistical techniques. In most of the techniques, the application of canonical correlation is only seen in the software output. As a result of the rigorous mathematical nature of the concept, its application is not clearly demonstrated. The structure of the data for which results are reported in standard texts are mostly non-time dependent in nature.

Specifically, the results of CCA on raw data, mean-corrected data, and standardized data are not clearly distinguished. Since the approach has been primarily theoretical, the execution of methods for data analysis has not been explicitly stated. Although numerous models of CCA have been suggested by various research, some conclusions are only stated in the text due to the mathematical complexity of the concept and the literature on the subject lacks a systematic presentation (Yang et al., 2021). A comparative analysis of the CCA in the literature at different times and eras reveals striking characteristics. This problem of finding an overall purpose correlation measure is approached by a number of research work. The literature does not, however, provide any techniques for performing CCA based on grouping scheme for time-dependent data structures.

In order to give a comprehensive and logical study of the interactions between various multivariate sets of variables that are also time series in nature, the aim of this thesis is to describe the links between CCA and Discriminant analysis. For effective presentation of the procedure, the fundamental approach would be the examination of key matrices that are involved in generating canonical variates, to make more explicit mathematical results that are stated in the standard text in the context of this study. The main problem from the literature, is identified to be an apparent lack of clarity and complexity in the presentation of applications of CCA. In spite of the numerous application and evolving extensions, statistical data still provides the need for further applications and extensions. There appears to be scanty in the literature a procedure designed for data that may be designed to have groupings with time-dependent (or constitute time series) nature.

Although CCA is described in widely used statistical computing packages and is available in many textbooks, there are certain technical and interpretive issues that prohibit practitioners from regularly using it. Problems with computing, interpretation, statistical significance, and the handling of discrete variables are some of these (Shafto, Degani, & Kirlik, 1997). Again, the benefits of expanded CCA methods have been shown in individual research, but the use of CCA with grouping scheme approach via discriminant analysis method is yet to be explored.

Purpose of the Study

The main aim of this thesis is to develop grouping scheme discriminant canonical correlation analysis method that is suitable for Time-dependent structure. Specific objectives are to:

1. provide a review of fundamental approaches to canonical correlation analysis in multivariate multiple data with time-dependent structure.
2. develop a procedure that identifies a grouping scheme for Multiple-group discriminant analysis that yields the optimal discriminant canonical correlation statistic measures.
3. use the identified optimal grouping scheme to obtain more precise canonical correlations among two sets of variables with time- dependent structure.

Significance of the Study

1. This thesis will provide a method for assessing the effectiveness of extension of canonical correlation analysis to Time-dependent multivariate data structure through a modification of multiple discriminant analysis. The study therefore provides tools for handling correlation analysis of multivariate time series data.

2. Through the presentation of a class of mathematical tools necessary to give a sufficient grasp of the concept, the study will present a focused and coherent presentation of the idea of canonical analysis.

3. This work will contribute to the research information on canonical correlation so that it can help in further work. The multivariate multiple Time-dependent dataset could then be modified using the generated codes to produce canonical variables for further use.

Justification of the Study

Research works, in general, are supposed to add to the store of knowledge, both in theory by helping to appreciate and understand certain occurrences and phenomena of life, and in practice, to assist in devising efficient means of addressing problems, if they exist. Consequent to the above, the justification of this work lies in the following:

1. Most of the well-known CCA researchers have not yet explored Discriminant Canonical Correlation Analysis of Time-Dependent Multivariate data structure. This area if effectively studied could enhance the application of CCA.

2. Multivariate multiple Time-dependent data structure research will add to the repository of knowledge in canonical correlation analysis research.

Definition of Terms

The following definitions of canonical correlation analysis will be used in this thesis throughout.

Canonical variable and canonical variate

The two supplied sets **Y** and **X** that are being examined both contain the original values or elements in their main sets, and are known as the canonical variables. The variables that are created from the original variables as weighted averages are known as canonical variates (Marden, 2015). The original variables of Y are always used to create a set of Y variates, and the same is true for the creation of a set of X variates. The collection of response variates and the set of predictor variates can be combined linearly to create a canonical variable.

Canonical weights

These are the weights that are utilized to build the linear combination, similar to regression coefficient. Larger weights have a greater impact on the function, whereas negative weights denote an inverse link to original variables. The variables, assuming they have weights with distinct signs, exhibit an inverse relationship. If and only if the weights have the same signs, the variables have a direct link (Marden, 2015; Nail, 2002).

However, there are issues with how the results should be interpreted, particularly when the weights are unstable because of multi-collinearity. It may be assigned a small weight or even no weight because the variance in that particular variable has already been explained by the other components. Weights are unable to accurately convey the significance of a variable in this context (Marden, 2015, Tandanai, 2015). Another problem with CCA is the variability of canonical weights from one sample to the next.

Canonical structures or loadings

The relationship between each variable and its variate in two groups with different variables is referred to as canonical loading (canonical structure). Its result is evaluated similar to principal component analysis. Using the correlations of each variable (Tandanai, 2015), it is possible to compute each variable's contribution to its own canonical variate. It measures the correlation between the variables and the corresponding canonical variates, as well as the variance that one variable contributes to the variance of another.

Canonical cross-loadings

This is the connection between every response variable that has been observed and a canonical variate. It gauges the degree to which the initial response variable and the predictor canonical variate are correlated. It offers a more accurate method for calculating the correlation between the response variables and the predictor variables (Marden, 2015).

Dependent variables set

The dependent (response) variables set covers p variables given by $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_p)$. The equation $\mathbf{U_i} = \alpha_{\mathbf{i}}'\mathbf{Y}$ is the linear combination for the i^{th} canonical variate in the set of response variables $\mathbf{Y}$. In the illustrative dataset, the response variable set is the monthly data on maximum temperature, minimum temperature, and solar radiation. Thus, $\mathbf{Y} = (Y_1, Y_2, Y_3)$ with 264 observations. Hence, the set of response canonical variates is given by $\mathbf{U} = \mathbf{AY}$, where

$$\mathbf{A} = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{pmatrix}$$

Independent variables set

The independent (predictor) variables covers q variables given by $\mathbf{X} = (X_1, X_2, \ldots, X_q)$. The j^{th} linear combination of canonical variates of the predictor variables set given by $\mathbf{V_j} = \beta_j'\mathbf{X}$. In the illustrative dataset, the predictor variables set is the monthly data on weather conditions which are precipitation, wind, and relative humidity. Thus, $\mathbf{X} = (X_1, X_2, X_3)$ with 264 observations. Hence, the set of predictor canonical variates is given by $\mathbf{V} = \mathbf{BX}$, where

$$\mathbf{B} = \begin{pmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \\ \beta_{31} & \beta_{32} & \beta_{33} \end{pmatrix}$$

Components of canonical correlation functions

CCA helps to estimate possible association among different variables, **Y** in the dependent set with variables, **X** in the independent set as shown in Figure 1. The figure gives the canonical loadings (CL), proportions of variance explained (PVE), canonical correlation coefficients (CCC), and the percentage of eigenvalues (EV) for both sets (Joshua, 2016).

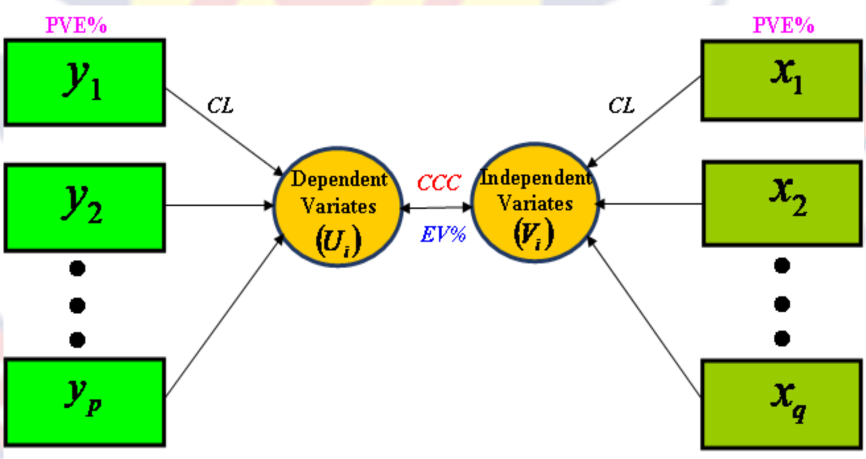

Figure 1: Structural Equation showing Components of CCA Functions

The number of canonical correlation functions (CCF) that can be extracted is equal to the number of components in the smaller set of canonical variate (Dattalo, 2014; Joshua, 2016).

Canonical Correlation Analysis Assumptions

The key presumptions for canonical correlation analysis (Priya, 2018) are as follows:

1. Just like multivariate regression analysis, CCA requires a sizable sample size to produce a trustworthy model.

2. Multicollinearity among one or more variable groups prevents the use of CCA. That is, none of the variables should have correlations close to 1.

3. The Gaussian distribution or multivariate normality of the variables in the population from which the sample was taken should be one of the fundamentals of CCA.

Checklists for Canonical Correlation Analysis Assumptions

The assumption that the variable distributions in the population from which the sample was taken are multivariate normal forms the basis of the tests for the significance of the canonical correlations. The following checklists (Tabachnick, 1989) guide the execution of CCA:

Check for missing data

When screening the data for outliers, attention should be given to patterns of missing numbers. When a row has missing values, the computer ignores it. The analysis should be skipped in order to collect data on the remaining variables if it appears that one or two variables account for the majority of the missing values (Priya, 2018).

Check for multivariate normality and outliers

Strong normality assumptions are not made by CCA. Outliers, like in all least squares methods, can pose major issues. The numerous univariate normality tests and graphs should be utilized to thoroughly examine the data for outliers (Priya, 2018).

Check for linearity

The CCA makes a claim that the factors are related to one another linearly. The scatter plots should be examined for each set of variables, giving particular attention to any curved or peculiar patterns. Analysis will be less helpful when there are curved relationships (Priya, 2018).

Check for multicollinearity and singularity

Since the analysis requires inverse matrices, multicollinearity must be examined. Multicollinearity occurs when one variable approaches a weighted average of the others. The exactness of this relationship is known as singularity. Separate Principal Components Analysis could be conducted on each set of variables. Eigenvalues at or close to zero indicate multicollinearity issues, hence the problematic variables could be removed (Priya, 2018).

Eigenvalues and eigenvectors of canonical factorization

The Eigenvalue eigenvector equation is given as follows if λ is a scalar, $\mathbf{v}$ is a vector that is not equal to zero, and $\mathbf{A}$ is a $(p \times p)$ square matrix.

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v} \tag{1.4}$$

The eigenvalue and eigenvector of the matrix $\mathbf{A}$ are λ and $\mathbf{v}$, respectively, according to Equation (1.4). (Magnus, 2019; Rencher, 2002; Gentle, 2017). Other names for eigenvalue include a latent root (λ), a characteristic value (C), or a suitable value. Equation (1.4) can now be written $(\mathbf{A} - \lambda\mathbf{I})\,\mathbf{v} = 0$ to determine λ and $\mathbf{v}$.

If $|\mathbf{A} - \lambda\mathbf{I}| \neq 0$, then there is only one solution, which is $\mathbf{v} = 0$. Consequently, $(\mathbf{A} - \lambda\mathbf{I})$ has an inverse. To determine non-trivial solution, we consider the characteristic equation $|\mathbf{A} - \lambda\mathbf{I}| = 0$.

The characteristic equation has p roots, as $\lambda_1, \lambda_2, ..., \lambda_p$. It is possible that not all of the λ's will be unique or non-zero. It is important to point out the following lemmas as the idea will be useful in Chapter 3.

1. If and only if **C** and **D** are two given matrices, then the matrices **CD** and **DC** have the same eigenvalues.

2. Let **C** and **D** be two matrices, where **D** is positive definite and suppose that the problem $\max_{\mathbf{v}}(\mathbf{v}'\mathbf{C}\mathbf{v})$ is given subject to the constraint: $\mathbf{v}'\mathbf{D}\mathbf{v} = 1$. When **v** is the eigenvector of $\mathbf{D}^{-1}\mathbf{C}$ corresponding to the largest eigenvalue, then the maximum is reached (Coleman and Hardin, 2013; Magnus, 2019).

Singular value and singular value decomposition

Singular value decomposition (SVD) aims to transform a dataset with many values into a dataset with significantly fewer values but with a significant amount of the variability that exists in the original data. Because of its uniqueness, the SVD is among the most important and useful decomposition in all of matrix theory and applications. The equation shows a factorization of the

$$\mathbf{S} = \mathbf{U}\mathbf{X}\mathbf{V}' \tag{1.5}$$

$(p \times p)$ matrix **S**, where **X** is a $(p \times m)$ diagonal matrix with non-negative components, **U** is a $(p \times p)$ orthogonal matrix, and **V** is a $(m\times m)$ orthogonal matrix (Gentle, 2017; Magnus, 2019). All other entries in a $(p \times m)$ diagonal matrix are zero, and its diagonal has $min(p, m)$ components. The SVD or canonical singular value factorization of **S** is the resolution of Equation (1.5). The singular values of **S** are the components along the diagonal of **X**, that is x_i (Gentle, 2017). If we rearrange the items in **X** such that $x_1 \geq x_2 \geq ... \geq x_{min(p,m)}$, nothing changes when we do the same with the columns of **U**.

Spectral Decomposition of Symmetric Matrices

The concept of the spectral decomposition of symmetric matrices (SDSM) is used widely, as will be seen throughout the course of the investigation. We state the SDSM as follows before introducing additional ideas that will be helpful in the methodology (Magnus, 2019; Nizamettin, Abasiyanik, Ersan, & Barik, 2006):

Theorem 1.1

Assume that the matrix $\mathbf{A} \in \Re^{p\times p}$ is symmetric. Then, $\mathbf{A} = \mathbf{PDP}'$ is the product of a diagonal matrix $\mathbf{D} \in \Re^{p\times p}$ and an orthogonal matrix $\mathbf{P} \in \Re^{p\times p}$. The eigenvalues of $\mathbf{A}$ are represented by the diagonal entries of $\mathbf{D}$, while the corresponding eigenvectors are represented by the columns of $\mathbf{P}$:

$$\mathbf{D} = diag(\lambda_1, \lambda_2, ..., \lambda_p); \mathbf{P} = [\mathbf{P}^{(1)}, \mathbf{P}^{(2)}, ..., \mathbf{P}^{(p)},]; \mathbf{AP}^{(i)} = \lambda\mathbf{P}^{(i)}, i = 1, 2, ..., p$$

By definition, $\mathbf{P}' = \mathbf{P}^{-1}$ is satisfied by an orthogonal matrix $\mathbf{P}$, indicating that the columns $\mathbf{P}$ are orthonormal. The Spectral Decomposition of $\mathbf{A}$ is given by the formula $\mathbf{A} = \mathbf{PDP}'$ of a symmetric matrix in terms of its eigenvalues and eigenvectors (Magnus, 2019).

Some important related matrices

In this study I shall rely very much on certain properties of the variance-covariance matrix ($\mathbf{C}$) to establish functional relationship among groups of variables. One important such property is the diagonalisability of the matrix. Another property is a matrix that has similarity with another matrix ($\mathbf{G}$). I will point out the relevance of having two matrices to be similar. There are important consequences for these two stated properties. In this section, I will discuss these properties and their consequences. Since a diagonalisable matrix must necessarily be similar to another matrix, I will first introduce similar matrices.

Similarity of matrices

If there is a non-singular matrix $\mathbf{S} \in \mathbf{M}_p$ such that $\mathbf{G} = \mathbf{S}^{-1}\mathbf{CS}$, then a matrix $\mathbf{C} \in \mathbf{M}_p$ is said to be comparable to $\mathbf{G} \in \mathbf{M}_p$. The transformation $\mathbf{C} - \mathbf{S}^{-1}\mathbf{CS}$ is called a similarity transformation (Magnus, 2019), and we write

$\mathbf{C} - \mathbf{G}$. Quite obviously, similarity is an equivalence relation. By this definition, we subsequently state the property of similar matrices in Theorem 1.2.

Theorem 1.2

Similar matrices have the same characteristic polynomial. This is because from the definition above, since $\mathbf{G} = \mathbf{S}^{-1}\mathbf{C}\mathbf{S}$, then considering the difference $\mathbf{G} - \lambda\mathbf{I}$, we should have

$$\mathbf{G} - \lambda\mathbf{I} = \quad \mathbf{S}^{-1}\mathbf{C}\mathbf{S} - \lambda\mathbf{I} = \quad \mathbf{S}^{-1}\mathbf{C}\mathbf{S} - \lambda\mathbf{S}^{-1}\mathbf{I}\mathbf{S} = \quad \mathbf{S}^{-1}\left(\mathbf{C} - \lambda\mathbf{I}\right)\mathbf{S}$$

It follows that

$$\begin{aligned} det\left(\mathbf{G} - \lambda\mathbf{I}\right) &= det\left(\mathbf{S}^{-1}\left(\mathbf{C} - \lambda\mathbf{I}\right)\mathbf{S}\right) \\ &= det\left(\mathbf{S}^{-1}\right) det\left(\mathbf{C} - \lambda\mathbf{I}\right) det\left(\mathbf{S}\right) \\ &= det\left(\mathbf{C} - \lambda\mathbf{I}\right) \end{aligned}$$

Thus, $\mathbf{C}$ and $\mathbf{G}$ have the same characteristic polynomial. Consequently, they should also have the same eigenvalues. Again, since

$$det\left(\mathbf{G}\right) = det\left(\mathbf{S}^{-1}\mathbf{C}\mathbf{S}\right) = det\left(\mathbf{S}^{-1}\right) det\left(\mathbf{C}\right) det\left(\mathbf{S}\right) = det\left(\mathbf{C}\right)$$

similar matrices have the same determinant. It also follows that $tr(\mathbf{G}) = tr(\mathbf{C})$.

Diagonalisability of a Matrix

Diagonal matrices provide the easiest forms of matrix analysis. In this thesis, diagonalising the variance-covariance matrix will provide additional information about the analysis of variance (ANOVA) (Magnus, 2019; Nail, 2002).

Definition:

A matrix $\mathbf{C} \in \mathbf{M}_p$ is diagonalisable if $\mathbf{C}$ is similar to a diagonal matrix. It should therefore be able to demonstrate that the characteristic equation of the matrix **C** and the matching diagonal matrix are the same, in accordance with the specification given above. Before proving this relation, some results should first be noted. Now let $\mathbf{C} \in \mathbf{M}_p$ be non-singular and $\mathbf{D} \in \mathbf{M}_p$, a diagonal matrix. The above definition means that $\mathbf{C} = \mathbf{QDQ}^{-1}$. Let the columns of **Q** being $v_1, v_2, ..., v_p$. Then

$$\mathbf{Cv}_i = \mathbf{QDQ}^{-1}\mathbf{v}_i = \mathbf{QDe}_i = \lambda_i\mathbf{De}_i = \lambda_i\mathbf{v}_i$$

where the i^{th} standard vector is $\mathbf{e}_i$, and the i^{th} diagonal element of **D** is λ_i (Magnus, 2019; Nail, 2002).

Cauchy-Schwarz Inequality

The Cauchy-Schwarz Inequality is useful for restricting difficult-to-calculate expected values. It enables us to divide E$[X_1, X_2]$ into two pieces, one for each random variable, to create an upper bound (Mukhopadhyay, 2000). This essentially means that the product of the expected squares of the two random variables, $E(Y^2)E(X^2)$, will always be less than or equal to the expected value of the product for Y and X, $E(YX)^2$. Supposing **s** and **t** are any two $(p \times 1)$ vectors, then the Cauchy-Schwarz inequality (Mukhopadhyay, 2000; Nelsen, 1994; Schwarz, 1888; Win & Wu, 2000) is given by

$$(\mathbf{s}'\mathbf{t})^2 \leq (\mathbf{s}'\mathbf{s})(\mathbf{t}'\mathbf{t}) \tag{1.6}$$

and equality is attained if and only if **s** = c**t** for some arbitrary constant c.

The difficulty with CCA has been introduced, and the purpose of the study has been explained, in this chapter. It is determined that the presentation of CCA obviously lacks both clarity and simplicity. The difficulty is attributed to the technique's mathematical complexity and the absence of clear instructions for using it in the standard texts. Within the constraints provided, the application will concentrate on applying CCA in the examination of multivariate linear relationships. The important matrices that may be helpful in CCA are anticipated and introduced in this work. In this chapter, the characteristics of these matrices have been covered in some detail.

The various extensions made for CCA though quite many do not cover data with all structures that are common. Some of these structures that equally requires attention are those that are multivariate multiple in nature with time-component. Datasets which are time- dependent in nature offers another type of multivariate multiple data for which CCA could be extended. Review is made of relevant application techniques involved with CCA such as singular value and singular value decomposition.

Another useful technique identified is the Cauchy-Schwarz Inequality- which is helpful for limiting difficult-to-calculate predicted product of two variables. The eigenvalues and eigenvectors of canonical factorization have been mentioned together with covariance matrices and vectors in CCA.

Organization of the Study

This thesis will consist of the following five chapters: The First Chapter is introduction. This chapter highlights the framework, problem statement, investigation's purpose, research questions, significance, rationale, and study organization. The intellectual underpinnings and other writers' perspectives will be considered in Chapter Two, Review of Related Literature, where the study's theoretical framework and conceptual framework will be emphasized. Chapter Three will be concerned with the method and procedure that will be adopted in carrying out the study. It discusses the procedures used to get the data, the theoretical underpinnings of the statistical methods employed in the study, and some techniques utilized to manage the data. The implementation of the procedure to data and the discussion are presented in Chapter Four. The summary of the entire results, conclusion and suggestions for effective data modeling decisions as well as for related studies will be presented in Chapter Five.

CHAPTER TWO

LITERATURE REVIEW

Introduction

To determine the usage of canonical correlation analysis (CCA), a thorough evaluation of pertinent literature is carried out. I begin by outlining the notations that will be used throughout this thesis. The literature on recommendation systems and other kinds of CCA are then briefly reviewed. A survey is made of the prior extensions and uses of CCA in the recommendation systems fields of research and present an overview of some essential CCA approaches as background for the suggested algorithms.

Theoretical Framework of Similar Studies of CCA

Nayir & Saridas (2022) found a link between culturally sensitive teaching responsibilities and creative job behavior. Their goal was to ascertain, from teachers' perspectives, the relationship between creative work practices and culturally sensitive teaching obligations. The first canonical function is computed to maximize the correlation between datasets on innovative work behavior and roles of culturally responsive teachers. Their preliminary research indicates that there is approximately 77% difference between data sets on innovative work behavior and roles of culturally sensitive teachers. The results of the CCA also showed a positive relationship between the variables of generating and implementing ideas and finding supporters for ideas in the dataset of innovative work behavior, as well as between the variables of the culturally mediating teacher and the culturally regulating teacher in the dataset of culturally responsive teacher roles. (Nayir & Saridas, 2022).

According to Samarov (2009), the process of finding chemicals with potentially significant biological action is known as drug discovery. The issue was that there were too many substances to look through, often in the millions, making experimental testing unfeasible. Therefore, substances that did not show a lot of biological activity were filtered out using computational techniques. This virtual screening process, also known as filtering, narrows the search field so that the remaining compounds can undergo experimental testing. Samarov (2009) offered a solution to the issue of virtual screening using CCA and a number of enhancements that made use of spectral and kernel learning concepts. These techniques were specifically used to address the protein-ligand matching issue. The behavior of CCA in the High Dimension Low Sample Size scenario was further investigated using the results of the theoretical study.

Roungu, Matair, Sanwar and Azizur (2013) presented CCA - an application to bank performance and client happiness. To determine the association, they used five consumer satisfaction indicators as their predictor variables and four bank performance indicators as their response variables. The findings of three trials demonstrated that the earning per share (0.7146) and branch count (0.6704) had a substantial impact on the coefficient of customer satisfaction variables. Additionally, they observed that bank performance variables liability had a stronger correlation with canonical performance variables (r = 0.7849). Total asset was one of the variables that had a less influence overall (r = 0.5916). The degree of Branch (r = 0.7181) and earning per share (r = 0.6291) were the most significant variables in terms of their customer satisfaction variables, proving the significance of all four examined variables. Additionally, the research of Roungu et al. (2013) focuses on assessing bank performance based on client satisfaction with CCA. The main causes of customers' dissatisfaction with banks are their high loan interest rates, inadequate deposit interest rates, and expensive preservation costs, despite the fact that banks firmly believe they are committed to meeting their customers' needs and becoming their first choice for banking. Since the indicators of bank performance and customer satisfaction are highly correlated, CCA is qualified to analyze this using its techniques and provide insightful commentary on the choice of variables and overall correlation. Roungu et al. (2013) only took into account private banks in their analysis that contributed to the stock market.

Shafto, Asaf and Kirlik (1997) investigated the problems associated with the interpretation, management, computation, and statistical significance of discrete variables and proposed viable solutions. They argued that a range of datasets used in human factors research, including field study data, data from part- and full-mission simulations, and data from flight recorders, should be used with CCA. In order to identify the issues and suggest answers, Shafto et al. (1997) used canonical correlation expertise to analyze a field study of crew-automation interaction in commercial aircraft. The study encompasses views from the cockpit of crew members interacting with the Boeing 757 / 767 aircraft's autonomous flight control system while on commercial flights operated by a prominent American airline. Each data record included a number of variables that described the circumstances of the change in mode selection as well as the change itself. Over 1500 records, each of which was distinguished by 75 different factors, made up the initial dataset used in the CCA analysis. The context or scenario accounted for around half of the variables, and the crew's response—that is, their selection of an autoflight mode configuration—accounted for the other half.

To gain a deeper understanding of the temporal fluctuations and coupling dynamics between the two time-varying sources, (Xuefei, Jun, Sandstedea, & Luoc, 2019) created the Time-Dependent CCA. From multilevel time series data, time-dependent canonical vectors can be extracted using the Time-dependent CCA technique. It was strongly suggested to use CCA, a method for generating linear projections in multivariate data analysis that improve the correlation between two sets of variables. Whilst using time series data, serial correlations between observations are usually ignored while calculating correlations. They suggested a convex formulation of the issue that makes use of the singular value decomposition (SVD) characteristics that are present in all answers to the CCA problem. They tested the proposed approach using simulated datasets. To examine how aging affects brain connections, they used it on a real resting state functional magnetic resonance imaging dataset. As a further method for assessing the dynamic pattern of brain connections, they also created a novel metric called canonical correlation variation. Additionally, Xuefei et al. (2019) looked into the temporal dynamics induced by various motor activities using their suggested strategy in a task-related functional magnetic resonance imaging. They showed that the Time-dependent CCA-based strategy outperforms previous techniques in both feature extraction and temporal variation detection.

Discriminative multiple CCA was introduced by (Lei-Gao, Lin-Qi, & Ling-Guan, 2021) for the analysis and synthesis of multimodal data. They discovered that it could extract more distinct features from multimodal information representations. They deliberately chose projected trajectories that boosted within-class correlation and minimized between-class correlation in order to maximize the use of the multimodal data. Lei-Gao et al. (2021) analytically proved that the optimally projected dimension via Discriminative multiple CCA could be rather precisely anticipated during this process, resulting in improved performance and a significant decrease in computational cost. By drawing the conclusion that CCA, multiple CCA, and discriminative CCA are particular examples of Discriminative multiple CCA, they created a unified structure for CCA. To demonstrate how well the Discriminative multiple CCA prototype could identify human emotions and handwritten numbers, they used it. Numerous tests revealed that multiple discriminative CCA performed better than conventional techniques such as discriminative CCA, discriminative canonical analysis with multiple canonical correlations, and serial fusion.

Akour, Rahamneh, Kurdi, Alhamad, Al-Makhariz, Alshurideh, & Al-Hawary (2023) examined student performance in Jordan's in-person and virtual learning environments using the CCA Method. By identifying the linear combinations of the two sets of data with the highest correlation, their research aims to determine the current correlations between the two sets of variables. The Al-Balqa Applied University researchers used the canonical correlation test to assess how strongly the degrees of in-person and online business education among their participants related to one another. For the objectives of analysis and result extraction, Akour et al. (2023) applied their study to a sample of students from Al-Balqa Applied University, specifically the faculty of Business. At the alpha level of significance, the findings of the Akour et al. (2023) study demonstrate that the canonical correlations of all three roots are statistically significant. For roots 1, 2, and 3, the response variables and the predictor variables have the first, second, and third canonical correlations of 0.98, 0.965, and 0.907, respectively. The canonical variables for the response and predictor account for 93.24, 4.7, and 2.047 percent of the variance, respectively. The study recommends assessing Jordan's experience with online learning at colleges and other institutions using the canonical correlation.

Benton, Khayrallah, Biman, Reisinger, Sheng, & Raman (2019)'s method for learning nonlinear transformations from arbitrary numbers of data views made sure that the resulting transformations were as mutually informative as feasible. While there are techniques for both linear many-view representation learning generalized the CCA and nonlinear two view representation learning. They demonstrated how Deep Generalize CCA blends nonlinear deep representation learning's adaptability with statistics' capacity to take into account data from many sources, or viewpoints. Both the Deep Generalize CCA formulation and a powerful stochastic optimization technique were described. They studied and evaluated Deep Generalize CCA representations for three downstream tasks: recommending hashtags, indicating acquaintances from a dataset of Twitter users, and phonetic transcription using auditory and articulatory measurements. Benton et al. (2019) presented Deep Generalize CCA, a non-linear multiview representation learning method from any number of views. By employing several views, they demonstrated that Deep Generalize CCA could learn Twitter user representations important for subsequent tasks, such as hashtag suggestion, and they outperformed earlier work in phoneme recognition when using labels as a third view. Until now, CCA-style multiview learning approaches have only been able to learn representations from a maximum of two views or to transform the input views strictly linearly.

Chacko (1986) discovered that a linear combination of variables from each of the two sets of variables is required to maximize the correlation between the two sets of data. The six interest groups—realistic, investigative, artistic, social, enterprising, and conventional—were combined to form his response variables. Additionally, the correlation between the Introversion or Extroversion Scores and the Academic Comfort Score was additionally looked at. His study indicates that 52.36% of the variation in the response variables can be replicated on average by the predictor variables for the first function. Typically, the predictor variables might account for 20.32% of the variation in the response variables for the second function. For the response variables, the first and second functions, respectively, can only reconstruct averages of 26.13% and 8.95% of their variance. When the variables in the set were predicted by every variable in the other set, the average multiple correlation for the variables in the set equals the pooled redundancy coefficients for the variables that were supplied. His response and predictor variables' pooled redundancy coefficients were, respectively, 72.67% and 35.08% (Chacko, 1986).

In addition to presenting a tutorial on CCA, (Hardoon, Szedmak, & Shawe-Talyor, 2004) developed a revolutionary universal method for obtaining pictures based only on their content. Then, this is used for retrieval that is mate-based and content-based. The generalized vector space model was not as accurate for picture retrieval, according to experiments. They illustrated how the ideal a priori regularization parameter k for various regimes might be selected. They therefore concluded that Kernel CCA is an effective method for image retrieval via content. Future tests by Hardoon et al. (2004) will include additional data sets. They could see that as the CCA process is generalized, the initial issue might be modified and reframed as a total distance issue or a variance minimization issue. To explain the structure of some specific spaces generated by more appropriate various kernels, more research into this special duality between the correlation and the distance is needed. When dealing with kernel space problems where the coordinates are unknown but the inner products and separations between the points are known, these methods can provide tools. For some issues, it is adequate to know the coordinates of a small number of specific locations, which can be stated using the inner product that is already known.

Likewise remarkable is the work of Akbas & Takma (2005). They found links between egg weight, sexual maturity age, and body weight and egg production. They planned their research to look at the connections between two groups of nesting hen-related factors. Their response variables were the number of eggs at three different dates; their predictor variables for CCA were the age at sexual maturity (ASM), body weight (BW), and egg weight (EW). They demonstrated a strong canonical link between the first and second sets of canonical variates. The age of sexual maturity, as opposed to body weight and egg weight, had the most influence on variation in the number of eggs produced at three distinct times (EN_1, EN_2, and EN_3) when using canonical weights and loadings with CCC, according to Akbas & Takma (2005). Only two pairs of canonical variables were considered since the canonical correlations between the first and second pair of canonical variables were determined to be significant. For the first pair of canonical variates, the estimated canonical correlation was 0.81, for the second pair, it was 0.152, and for the third pair, it was 0.024. Akbas and Takma (2005) noted that there were strong relationships between (ASM, BW, and EW) and the first set of canonical variates (EN_1, EN_2, and EN_3). The outcomes of their study demonstrated that the standardized coefficients' signs appropriately represented the impacts of ASM, BW, and EW on EN_1, EN_2, and EN_3. Additionally, the effects of ASM and BW on the amount of eggs produced at three distinct periods in comparison to EW were also accurately portrayed.

The work of Kabir1, Merrill, Shamim, Klemn, Labrique, Parul, Keith, and Nasser1 (2014) merits attention as well. Five birth size measures and ten maternal features serving as predictor variables and response variables, respectively, were compared using CCA. These researchers employed 14506 women who gave birth to singletons in a double-masked, cluster-randomized, placebo-controlled study of prenatal vitamin A or B-carotene supplementation in rural Bangladesh. With a value of 0.42 and a p-value less than 0.001, their investigation's initial Canonical Correlation Function showed a significant interaction between infant sex and premature delivery on birth size. Following the CCA, the Kabirl et al. (2014) study's multivariate analysis of variance showed a significant interaction impact on birth size between infant sex and early delivery. Male and female term newborns were larger than male and female preterm and term kids. Their research led to the discovery of this type of interaction impact on birth size in a large body of literature. The strong relationship between an infant's birth weight and maternal variables was investigated using CCA. The maternal characteristics that affected or did not affect infant size at delivery, as identified by CCA, were consistent with the literature's documentation of these kinds of correlations.

To test the viability of their proposed strategy, Lei-Gao et al. (2021) conducted information fusion experiments using the Ryerson Multimedia Lab (RML) and eNTERFACE (eNT) audiovisual databases, respectively. Users can access these through the RML database, which illustrates the six main emotions—disgust, fear, surprise, anger, sadness, and happiness—with video clips featuring eight people speaking in six different languages. The video was recorded at 30 frames per second, while the audio was sampled at 22,050 Hz. The facial region had an average dimension of about 112 x 96, and the image frames were 720 x 480 in size. Each of the six main emotions was captured on camera by 43 people in the eNT database at sampling rates of 25 frames per second for the video and 48,000 Hz for the audio channel. The facial region had an average size of around 260 x 300 over the 720 × 576 photo frames in the experiment, which used 456 audio samples from ten participants and eight patients from the eNT database and the RML database, respectively. For training and testing, they divided the two audio samples into subsets with 360 and 96 samples, respectively. Prosodic and Formant Frequency characteristics were initially evaluated as a standard for emotion recognition (Lei-Gao et al., 2021). The percentage of successfully identified samples to all testing samples was used to determine the recognition accuracy.

Arandjelovic (2013) addressed the issue of matching vector collections that were present in the same input space. He offered a strategy that was based on CCA, a statistical method that has been effective in tackling a number of pattern recognition issues. His CCA extension searches for the variability patterns that are most comparable between the two sets when matching sets. His first major contribution was the development of a sound framework for accurately inferring such modes from data in the environment of uncertainty caused by noise and sampling. Arandjelovic (2013) claimed that the extended CCA lacked free parameters that could not be found solely by data, despite maintaining the efficiency and closed form keywords of the CCA. His second major contribution is that he showed that extended CCA can match sets in a discriminative learning framework more easily than CCA. We then conducted an empirical evaluation of the theoretical work he had done on the problem of face detection from sets of rasterized appearance photographs. The results demonstrated that his method (Extended CCA) already outperformed CCA and its quasi-discriminative equivalent, limited CCA, for all values of their free parameters.

Convolutional neural networks have lately drawn more interest in the machine learning and computer vision fields as a result of the work of (Bernardo & Eulanda, 2017), enhancing the performance of a number of related applications. They disclosed that only a limited number of the most recent deep learning network architectures—principal component analysis network, CCA network, and linear discriminant analysis network, among others—have been suggested for the classification of objects and faces. They continued by saying that these architecture solutions had proven to be highly effective, had a straightforward implementation, and allowed for quick prototyping of effective picture categorization applications. The filters used in these methods, however, might not be able to extract highly discriminative features in increasingly difficult computer vision problems. Bernardo and Eulanda (2017) created a network of discriminative canonical correlations called the discriminative canonical correlation network, which employs filters created from the analysis of discriminative canonical correlations to yield more discriminative information. By learning filters from discriminative canonical correlation, they claim, the network will develop discriminative properties and generate more accurate and discriminative data. They established the value of discriminative canonical correlation Network with tests on four datasets.

Sisi, Hanyu, Wang, Zhou, & Shaochun (2020) looked into the connection between Yangtze River water quality and marine development. The Yangtze River's aquatic ecosystem and shipping data were assessed using CCA from 2006 to 2016. Using data from the shipping prosperity index and mainline freight traffic, they looked at the development of Yangtze River shipping. When evaluating the aquatic environment of the Yangtze River, researchers also took into account petroleum contamination, biochemical oxygen demand, waste water discharge, ammonia nitrogen concentration, and the potassium permanganate index. According to their investigations, the volume of mainline freight had a big impact on how much wastewater was produced and how much gasoline was polluting the water. They were also conscious of the interdependence between the aquatic environment and shipping on the river.

Hou, Heng, & Quansen (2018) provided a description of the sparse regularized discriminative CCA for multi-view semi-supervised learning. By incorporating label information from their primary samples, the traditional unsupervised CCA strategy for learning multi-view data representation has recently been turned into supervised methods. Their research showed that the need for a lot of labeled samples makes it difficult to deploy supervised CCA versions in practice. To extract the most discriminative information from the fewest practicable labeled samples, they proposed a unique sparse regularized discriminative CCA technique that maximizes the use of the label information. By creating sparse weighted matrices from different viewpoints and extracting fused multi-view features that were not only the most correlated but also contained the essential discriminative structure information, they were able to incorporate the structure information into the original CCA framework. To assess their method, both the handwriting dataset and the face dataset were employed. They eventually understood that its utility and superiority were demonstrated by experimental results and assessments of it against other pertinent algorithms.

By analyzing Discriminative Multiple CCA, (Lei-Gao et al., 2021) proposed a novel technique for efficient information fusion. Their usage of the provided algorithm to show that discriminating representation works best when the number of projected dimensions is fewer than or equal to the number of classes (c) in the fused space is the most significant addition to their work. The significance of the contribution lies in the fact that the discriminative representations can be derived without computing the complete transformation process by computing only the first c projected dimensions of the discriminative model. When handling complex issues, the capability is especially appealing. Additionally, they used mathematics to show that the suggested Discriminative Multiple CCA, which unifies correlation-based information fusion methods, was a special case of the CCA, Multi-modal CCA, and Discriminant CCA. The presented technique solves the multi-feature and multi-modal information fusion challenges by recognizing handwritten numerals and human emotions. The results of the experiments showed that by using effective practical pattern recognition, the suggested technique enhances recognition performance with a notably reduced dimensionality of the feature space.

With respect to a number of model spaces, Chao, Zongning, Zhao, and Zhou (2015) established rate-optimal non-asymptotic minimax estimation. Their literature did not offer many theoretical counterarguments for them. In high-dimensional situations, they thought about the issue of estimating the top canonical correlation directions. For many applications requiring massive amounts of data, many high-dimensional techniques have been proposed, based on the supposition that the primary canonical correlation routes are sparse. They used an enlarged sin-theta theorem and an empirical process limit for Gaussian quadratic forms with rank restrictions to create the minimax upper bounds and find two unexpected events.

Multimodal recognition happens, in accordance with study by (Tingkai, Songcan, Jingyu, Xuelei, & Pengfei, 2009), when the non-robustness of uni-modal recognition is noticed in real-world circumstances. For them, a potent method for multimodal recognition is feature fusion via CCA. But since paired samples are necessary for CCA, there are a number of unexpected reasons why this requirement might not be easily satisfied. They added that the class information in the sample data is not fully utilized by CCA. Due to these restrictions, CCA is unable to extract more distinguishing features for recognition. To address these issues, a brand-new multimodal identification technique called Discriminative CCA with Missing Samples is created. It incorporates class information into the CCA framework for recognition. In order to reduce calculation time and space, discriminative CCA with missing samples can accept missing samples without having to artificially make up for them. Their test results showed that discriminative CCA with missing samples outperformed comparable multimodal recognition methods and that the amount of missing data had little impact on how accurately discriminative CCA with missing samples recognized objects.

Langworthy (2020) showed that the robust correlation estimator yields consistent and asymptotically normal CCA estimates. Next, an accurate correlation estimator based on modifications to Kendall's tau rank correlation coefficient was used to extend CCA. Additionally, he described a bootstrap-based testing method for locating instructive canonical routes. His simulations demonstrated that for data from skewed and heavy-tailed distributions, this robust estimator outperforms traditional CCA. Using Diffusion Tensor Imaging and information on brain white matter structure, he used this method to show a connection between lateralization of white matter brain structure and higher EF test scores in six-year-old children. In the multivariate survival context, when failure time data could be appropriately suppressed, he then defined Principal Component Analysis. he covariance and correlation matrices were predicted to exist for the counting processes denoted by the corresponding martingales and failure times.

To get principal direction estimates, Langworthy (2020) used eigen decomposition of these covariance and correlation matrix estimations. These estimations were asymptotically normal and consistent. He used this approach to analyze data from a pancreatic cancer clinical trial and was able to identify groups of adverse events that were medically significant. He finally extended robust CCA to this environment when there were more than two sets of variables and at least one of those sets was high-dimensional. He applied changes to Kendall's tau to produce the same reliable correlation estimate. He also applied cross-validation to the testing and dimension reduction processes. It was discovered that this cross-validation testing technique was dependable when the data came from a heavy-tailed elliptical distribution, in contrast to traditional techniques. Langworthy (2020) expanded his analysis of Diffusion Tensor Imaging and EF data to include brain gray matter volume data from 88 distinct brain regions in order to more thoroughly examine the association between brain structure and EF test outcomes.

Extensions of Canonical Correlation Analysis

Various extensions have been so far made to the basic formulations of CCA according to the work of (Yang, Liu, Wei, & Tao, 2021). Those extensions are outlined in Table 1 together with the author and year of first mentioned.

Table 1: **Some Categorical Extensions of Canonical Correlation**

Category of Extension	Name of Author(s)	Year
Multi-View CCA	Vinograde	1950
Deep CCA	Asoh & Takechi	1994
Kernel CCA	Lai & Fyfe	2000
Regularization of CCA	Hardoon, Szedmak & Shawe-Talyor	2004
Probabilistic CCA	Bach & Jordan	2005
Discriminative CCA	Kim, Kittler & Cipolla	2006
Sparse CCA	Parkhmenko, Tritchler & Beyene	2007
Locality Preserving CCA	Sun & Chen	2007
Repeated CCA of KPM	Srivastava & Dayanand	2008
Tensor CCA	Luo, Dacheng, Yonggang, Kotagiri & Chao	2015

Source:Yang et al. (2021)

According to Yang et al. (2021), recent developments in data gathering and statistical analysis support the application of CCA in advanced research. The CCA decomposition, which maximizes the correlation between pairwise variables in the shared subspace, is the main technique for reducing the dimensionality of two-set data. Several CCA models have been put forth over the course of eighty years of development using various machine learning techniques. The field, however, lacks a comprehensive analysis of the most recent innovations. Their survey had as its goal to offer a comprehensive overview of canonical correlation analysis and its extensions. First, they looked explicitly at model construction and model optimization from the perspective of the CCA theory. Yang et al. (2021) introduced Eigen Value Decomposition (EVD) and Singular Value Decomposition (SVD), two commonly utilized solution techniques. There was also a taxonomy of current advances. It was classified into seven categories: Deep CCA, Probabilistic CCA, Kernel CCA, Locality preserving CCA, Multi-view CCA, Sparse CCA, and Discriminative CCA. They also presented two or three typical mathematical models for each group and talked about their advantages and disadvantages. They compiled the data sets and open-sources for use after providing explanations of the representative uses and quantitative findings of those seven groups in actual scenarios. Further, they provided a number of interesting new study directions that might help progress the current state of the art.

Yoshida, Junichiro, & Kenji (2017) suggested using sparse kernel CCA for high-dimensional data to find nonlinear relationships. They claim that as high-throughput technologies in genomics, transcriptomics, and metabolomics have advanced, there has been a rise in the need for bioinformatics tools that can integrate highly dimensional data from various sources. They found that feature selection and the recording of multiple canonical components were not possible with previous CCA extensions, including kernel CCA, which were used to record nonlinear connections. To choose the proper kernels within the context of multiple kernel learning, they created a novel technique known as Two-stage kernel CCA. They discovered that the Two-stage kernel CCA of the multiple kernel learning framework had initially selected the pertinent kernels. Weights were then established using non-negative matrix decomposition and L_1 regularization. They showed that Two-stage kernel CCA could identify many, nonlinear correlations among high-dimensional data and multiplicative interactions between components using fictitious datasets and real nutrigenomic data. They also discovered that Two-stage kernel CCA, as opposed to earlier nonlinear CCA approaches, could more precisely find nonlinear connections among high-dimensional data.

Vinograde (1950) was the first to investigate multi-view CCA, but his approach fell short of the second desired attribute. Steel (1951) created a system of complex equations. He used compound matrices to locate multi-group canonical solutions, but in actuality, his problems were quite challenging. The generalized CCA was first presented by (Horst, 1961). He created an iterative method for detecting canonical variates and applied it to several groups before discovering the method's convergence property in 1963, but the outcomes were unsatisfactory. The pairwise correlations of several points of view were evaluated using two formulations in the same year: the maximum variance (MAXVAR) and the total sum correlation (SUMCOR). His thesis focuses on two generalizations of CCA that have been published in the literature: sum of square correlations (SSCOR). From an optimization perspective, the SUMCOR issue formulation is noteworthy in and of itself because it also occurs in other circumstances. He originally introduced a novel, provably convergent strategy based on an iterative method for addressing multivariate eigenvalue problems in order to find non-linear higher order patterns.

Rupnik (2016) investigated a variety of the generalizations. SUMCOR is generally shown to be NP-hard, and then showed how to reformulate it into a computationally feasible Semi Denite Programming (SDP) issue. He developed various computationally workable bounds on global optimality based on the reformulation, which go in addition to the locally optimal solutions. Rupnik (2016) developed a novel preprocessing step to address large-scale SDP difficulties that arise from an application to cross-lingual text analysis. He looked into how to use his techniques on real datasets that had missing data. Due to the unique nature of the missing data in the topic under consideration, the SSCOR optimization problem, which was subsequently compressed to a manageable eigenvalue problem, is implicated. Then he showed how to develop cross-linguistic similarity models using the algorithms, and he utilized the models to do cross-linguistic cluster linking. Using news streams in many languages, a real-time global analysis was conducted using the cross-lingual cluster linking method.

Kettenring (1971) outlined five extensions of the traditional Two-group CCA theory. In these extensions, many models were created employing the cross-correlation with a wide range of methodologies. To exploit the significant nonlinear link between two sets of variables and produce a range of combination techniques, CCA was coupled with deep networks (Asoh & Takechi, 1994; Lai & Fyfe, 1998; Lai & Fyfe, 1999; Hsieh, 2000; Andrew, Arora, Bilmes & Livescu, 2013; Yang et al., 2017). In order to describe the nonlinear mappings for CCA, Asoh and Takechi developed the first combination in 1994 using two multilayer perceptrons. Three feed-forward neural networks were combined by Hsieh to create the non-linear CCA, which non-linearly generalizes CCA (Hsieh, 2000).

Ignacio, Sebastien, Pascal, & Alain (2008) developed the exploratory statistical technique known as CCA to identify correlations between two data sets that were gathered on the same experimental units. The calculations are done using R's cancor () function, but more work was needed to provide the researcher access to other tools that would make it easier to analyze the data. They created numerical and graphical data and provided the researcher with the capacity to manage missing values using the free CCA R package from the Comprehensive R Archive Network. To handle data sets with more variables, their package additionally includes a regularized form of CCA. Examples are provided through the examination of a data collection from a mouse nutrigenomic study (Ignacio et al., 2008).

Locality Preserving CCA was developed by Sun & Chen (2007) to reduce the overall non-linear complexity of data while preserving its local linear structure. They found that when the sample size is large, calculating the distances between neighbors takes time. After suggesting Least Squares based CCA, they noticed that their results were essentially identical to MAXVAR results. The output layers were to have the maximum linear correlation possible. Andrew et al. (2013) constructed a back propagation deep model, added the nonlinear component, and adjusted the network weights. In order to train two-view filters, (Yang et al., 2017) developed a network that applies CCA to each layer of a stacked convolutional network. The label information is crucial to them in terms of classification issues.

In order to effectively utilize the discriminant information, (Sun, Chen, Yang, & Shi, 2008) introduced discriminant CCA by taking into consideration the intra-class and inter-class similarities of different perspectives. As an alternative to discriminant CCA, they introduced Multi-view linear discriminant analysis, which merged CCA with linear discriminant analysis. Deep canonical correlation auto-encoders was first introduced by (Wang, Guan, & Venetsanopoulos, 2015). Deep neural network models of deep canonical correlation auto-encoders, however, are difficult to comprehend and need a lot of data to fit. Benton et al. (2019) developed the Deep neural network-based Deep Generalized CCA, which was a non-linear expansion of MAXVAR generalized CCA. They looked at the labels' one-hot encoding matrix as an additional viewpoint to make use of the classifier information. As a result, this dissertation offers an overview of numerous representative CCA techniques.

Yang et al. (2021) presented two or more exemplary mathematical models for each set of analyses, highlighting both their advantages and disadvantages. They talked about the relationship between two widely used CCA solution techniques, such as, Singular value decomposition techniques and Eigen-value decomposition methods. Bach & Jordan (2005) presented the first probabilistic explanation of CCA, which was a latent variable model for two Gaussian random vectors. Similar to the probabilistic justification for principal component analysis, they offered a probabilistic reading of CCA. Additionally, they considered Fisher linear discriminant analysis to be a CCA between vectors that had the required definitions. Their interpretation advances our understanding of CCA as a model-based approach and facilitates the incorporation of CCA models into more comprehensive probabilistic models. Following (Yang et al., 2021), the use of CCA and its extension increased. This demonstrates how essential CCA has become as a tool for today's rising tide of scholars.

Tensor CCA was first suggested by (Luo, Dacheng, Yonggang, Kotagiri, & Chao Xu, 2015) for multi-view dimension reduction. They demonstrated how it is possible to directly maximize the correlation of many points of view by looking at the high-order covariance tensor and developed Tensor CCA, a simple modification of CCA that can handle data from any number of points of view. Directly maximizing the canonical correlation of various perspectives is the goal of Tensor CCA. Most importantly, they showed how to solve the multi-view canonical correlation maximization problem by rapidly obtaining the optimal rank-one approximation of the data covariance tensor using the widely-known alternating least squares method. The high order correlation data found in the various viewpoints was analyzed by Luo et al. (2015) in an effort to identify a more precise common subspace shared by all features. A non-linear extension of Tensor CCA was proposed as a result of their research. Experiments on a range of challenging tasks, including the annotation of web photos, the prediction of large-scale biometric structures, and the classification of online advertisements, proved the validity of their proposed methodology.

Witten, Tibshrani, and Trevor-Hastie (2009) proposed the penalized matrix decomposition, a state-of-the-art technique for calculating a matrix's rank-K. The formula they arrived at was $\bar{\mathbf{X}} = \sum_{k=1}^{K} d_k \mathbf{u}_k \mathbf{v}_k'$. In this equation, the squared Frobenius norm of $\mathbf{X}$ - $\bar{\mathbf{X}}$ is minimized by $\mathbf{u}_k$, $\mathbf{v}_k$, and d_k. The singular value decomposition is regularized as a result. It would be interesting to look into how the L_1 - penalties on $\mathbf{u}_k$ and $\mathbf{v}_k$ lead to a sparse vector decomposition of $\mathbf{X}$. They showed how to construct a method for sparse main components where L_1 - penalty is given to $\mathbf{v}_k$ but not to $\mathbf{u}_k$ when applying the penalized matrix decomposition. In fact, their findings offer a helpful method for the ”SCoTLASS” technique that has been recommended for obtaining sparse main components. OThis approach is shown in a set of publicly accessible gene expression data. Witten et al. (2009) discovered parallels between the penalized CCA method generated by the penalized matrix decomposition applied to a cross-products matrix and the SCoTLASS sparse principal component analysis approach. To implement their penalized CCA approach, they used both simulated data and a genomic data collection consisting of assessments of gene expression and DNA copy number on the same set of samples.

Coleman and Hardin (2013) investigated the multivariate CCA approach. They carefully described how CCA identifies the most highly correlated linear combinations of variables from two datasets and examined the mathematics that supports it. They followed by comparing the performance of CCA with clean multivariate normal data and with contaminated data to show how contamination inhibits CCA from accurately capturing the structure of the population covariance matrices. After that, M-estimation was added to the sample covariance matrices, which was successful for large observation values but unsuccessful for observation values close to the total number of variables. To solve this problem and enhance interpret-ability, they looked at Sparse CCA, as it was defined by (Parkhomenko et al., 2007). They found that the results from extending Sparse CCA to produce multiple sparse canonical vectors and including robust estimate in the sample covariance matrices performed better than the results from robust CCA. They next examined PITCHf/x variables and traditional statistics' coefficients of linear combinations while using robust Sparse CCA to analyze baseball data.

Golugula, Lee, Master, Feldman, Tomaszewski, Speicher, and Anant (2011) introduced the Supervised Regularized CCA , a unique data fusion approach that, in contrast to CCA and Regularized CCA, may fuse with a feature selection scheme and is computationally affordable. They gave examples of how multi-scale, multi-modal imaging and non-imaging data can be statistically integrated and depicted using Supervised Regularized CCA. They created a combination quantitative histologic-proteomic classifier utilizing Supervised Regularized CCA in order to identify individuals with prostate cancer who are most likely to experience a 5-year biochemical recurrence after surgery.

According to Golugula et al. (2011), Supervised Regularized CCA is statistically significantly faster than Regularized CCA and can perform regularization as well as identify patients at risk of biochemical recurrence more accurately than principal component analysis, CCA, or Regularized CCA. It can also create a meta-space with samples that is more stratified than the meta-space created by CCA or Regularized CCA. Although the fused prognostic classifier they developed for their research to predict biochemical recurrence appears promising, they also identified certain disadvantages such as: They only examined 19 datasets since, as was already mentioned, mass spectrometry is rather expensive. They were able to ascertain that their fused Supervised Regularized CCA classifier would yield an accuracy of 93%, or more than 95% of the time, if their dataset were expanded to 56 studies. This was done by applying a minimal sample size derivation model.

In the future, Golugula et al. (2011) planned to assess their classifier on a cohort like this. The classifier should have been trained and evaluated using a randomized cross validation technique, ideally. Unfortunately, the size of the cohort also placed a limit on this. Although their research employed both parametric and non-parametric feature selection techniques, the use of parametric selection techniques would be feasible with Supervised Regularized CCA and a bigger dataset for classification, given that the underlying distribution could be calculated. For datasets with less sample sizes, a non-parametric feature selection method may be more appropriate. In the context of various application domains and problem domains, they will use supervised regularized CCA to new imaging and non-imaging datasets in their upcoming work.

Hwang, Jung, & Takane (2011) suggest that functional CCA should be extended to examine more than two sets of functional data. Their technique, which is a regularized variant of multiple-set CCA, uses a roughness penalty as a regularization term for each set of functional data. The recommended strategy was successful in identifying the brain regions that were engaged during a successful fMRI investigation on verbal working memory. Using their method, (Hwang et al., 2011) were able to create canonical variates of signal changes over scans that were distinct to each experimental condition. The neural networks that were commonly active during the experiment across different people were identified by the object evaluations of the voxels produced by integrating these canonical variates.

Conversely, the working memory trial settings could not be accurately matched to the canonical variates generated by traditional multiple-set CCA. According to this perspective, multiple-set CCA might have shown a correlation between several sets of BOLD signal changes that come from sources of brain activity apart from the study conditions. The results of multiple-set CCA are therefore unlikely to be worth interpreting from substantive perspectives. To enhance its data-analytic capabilities and application, (Hwang et al., 2011) may further develop and expand their suggested method. For more in-depth investigations, they can include linear limitations in their suggested method. In this case, the suggested methodology was used to examine the initial data before employing a design matrix to assess the results while accounting for the experimental conditions. But at the start of the research, they may categorize a design matrix like that as linear restrictions and divide the data according to it. Their suggested approach is only applicable to the data made available by the design matrix at the following stage. This could result in a solution that is more suited to the specifics of the experiment.

A method for Non-parametric CCA, was proposed by (Tomer, Weiran & Karen, 2016). For them, CCA is a traditional method of representation learning for identifying linked variables in multi-view data. There have been several non-linear extensions of the original linear CCA put forth, such as deep neural network and kernel techniques. These algorithms search for optimally correlated projections among user-defined families of functions. They are computationally intensive. It is interesting to note that Lancaster had already investigated the theory of nonlinear CCA in a population environment in the 1950s, but his findings did not lead to the development of useful algorithms. To create a workable method for Non-parametric CCA, (Tomer et al., 2016) went back to Lancaster's theory.

Specifically, Tomer et al. (2016) showed that the reaction can be expressed through the use of the singular value decomposition of a particular operator associated with the joint density of the views. As a result, Non-parametric CCA was reduced to solving an eigenvalue problem by inferring the population density from the data. While this seems on the surface to be similar to kernel CCA, it differs significantly in that no kernel matrices need to be inverted. According to Tomer et al. (2016), one of the views experienced a linear projection, while the other was non-parametric in a partially linear CCA variation. Their non-parametric, partially linear CCA algorithms outperformed kernel CCA and comparable to deep CCA, which employed a kernel density estimate derived from a sparse sample of nearest neighbors.

Tomer et al. (2016) were memory-efficient, frequently ran significantly quicker, and performed better than kernel CCA. They provided closed-form solutions to the Non-parametric CCA and partially linear CCA problems and demonstrated how the Singular value decomposition of a kernel defined by the point-wise mutual information between the views might yield the best non-parametric projections. This results in a straightforward technique that, for moderately sized data sets, beats Kernel CCA and matches deep CCA in terms of performance on many datasets while being more computationally economical than both.

Canonical correlation analysis was employed by Cankaya, Balkaya, & Karaagac (2011) to evaluate the associations between Plant Characteristics and Yield Components of fifty-six red pepper populations that were gathered from the Black Sea region of Turkey's Samsun area. The researchers determined that each of the canonical correlation coefficients (0.708, 0.635, and 0.413) between the canonical variable pairings was significant at ($P < 0.01$). As compared to other Yield Components, the results of their canonical correlation research showed that the number of fruits per plant had the greatest explanatory power of the canonical variables estimated from their dependent variables of 56 red pepper populations. When compared to other characters, Fruit Length and Plant Height made the largest contributions to the explanatory power of canonical factors generated from their independent variables. Their study's findings suggest that in order to maximize output per plant in red pepper genotypes, one should take advantage of plant height, fruit width, and fruit wall thickness.

Using CCA, Combes (2008) demonstrated the relationships between sensory and physicochemical indices in the meat of rabbits raised using three different breeding techniques. He claims that a range of physicochemical measurements were performed on meat from rabbits raised in accordance with a standard (STAND), high quality norm (LABEL), or low growth breeding (RUSSE) system. These measurements included weight of retail cuts, color parameters, ultimate pH, femur flexure test, Warner-Bratzler shear test, water holding capacities, and cooking losses. The meat on the rabbit's back and leg looked the best to him. Leg pain was significantly reduced by the rank order of STAND $>$ LABEL $>$ RUSSE. Combes (2008) found that there were substantial correlations ($R(2) = 0.73$ and 0.68 between the two first pairs of canonical variates) between sensory and physicochemical variables in CCA. Specifically, a link was observed between sensory discomfort and the WB shear test variables that were recorded on raw longissimus muscle (LL). There was a correlation between the fibrous characteristic in the rear and cooking loss in LL. When evaluated independently, only the RUSSE rabbits displayed the same connections between the variables as those found for the total dataset.

One of the main issues with high-dimensional data analysis in the real world is how to extract relevant and meaningful characteristics from multi-view data (Ankita & Pradipta, 2023). The multi-set CCA is a well-known statistical method for combining data from several perspectives. It finds a linear subspace that strengthens the relationships between different points of view. They demonstrated how the computationally expensive nature of the existing methods for identifying the multi-set canonical variables limits the usefulness of the multi-set CCA in real-world huge data research. Owing to the limited sample size, there is a chance that the covariance matrices of every high-dimensional viewpoint will encounter the singularity problem. Furthermore, many of the currently in use multi-set CCA-based feature extraction techniques are unsupervised (Ankita & Pradipta, 2023).

To them, how to extract essential and pertinent features from high-dimensional data in real-world applications is one of the key challenges. In light of this, a brand-new supervised feature extraction approach is put forth that combines multimodal, multidimensional data sets by resolving the multi-set CCA's maximal correlation problem. To make determining the canonical variables of the multi-set CCA less computationally complex, a new block matrix form is presented by (Ankita & Pradipta, 2023). The supervised ridge regression optimization technique can effectively compute the multi-set canonical variables than to the analytical formulation. It solves the "curse of dimensionality" issue brought on by high-dimensional data and makes it possible to generate pertinent features sequentially at a substantially lower cost. The effectiveness of the suggested multi-block data integration technique is illustrated on several real and benchmark cancer data sets, and a comparison with other currently employed methods is provided (Ankita & Pradipta, 2023).

Recently, with the introduction of pertinent computer software, this technique's utilization began to increase. However, certain studies in the fields of animal science and even poultry science (Akbas & Takma, 2005; Jaiswal et al., 1995) utilised CCA. Finding a linear combination of each pair of variables that maximizes the correlation between the two functions is the goal of CCA, according to Glahn (1967). He pointed out that this analysis is equal to multiple regression in some circumstances while DA is equivalent in other techniques. He suggested a generalized correlation coefficient, explained how his formulae for canonical variate prediction work, and gave an example of CCA. He also examined the relationships between these techniques. Srivastava & Dayanand (2008) used generalized CCA to look at the correlations between two groups of frequently or longitudinally seen data and a block Kronecker product matrix to illustrate the dependence of their chosen variables across time. They used their linked matrix to derive canonical correlations and canonical variates from their data.

Review of Discriminant Analysis in Multivariate Statistics

From the work of Shelley (2007), we use a method known as discriminant analysis (DA) to evaluate the available data when the response variable is categorical and the predictor variable is of an interval type. A response variable is divided into different categories when it is referred to as a categorical variable. One of the three dummy variables, Dummy Variables 1, 2, or 3, as an example, can serve as the category answer variable. In essence, a discriminant function is a linear collection of predictor variables that accurately differentiates between the response variable categories. Creating discriminant functions is the aim of the DA method. We can use this opportunity to look for significant differences in the predictor variables between the groups. It also evaluates the classification's accuracy (Shelley, 2007). The amount of categories the response variables have is how DA is defined. Since statistics holds that all propositions are true until infinity, the type used in this case is Two-group DA when the dependent variable has two categories. When the dependent variable has three or more categories, Multiple-group DA is utilized (Shelley, 2007). The fact that just one discriminant function can be generated for a Two-group distinguishes the different types of DA significantly from one another. Nevertheless, a multitude of discriminant functions can be computed using Multiple-group DA (Rencher, 2002).

Multivariate procedures called discrimination and classification are used to separate different groups of things and assign new objects to existing categories. Classification is the process of identifying and classifying objects or ideas into designated groups. Classification can also be defined as the process of grouping information based on similarities. It is a list of how frequently a variable has specific scores or ranges of scores. In data management, data can be divided and arranged in accordance with predetermined criteria for a variety of professional or individual objectives. Classification is used in predictive modeling with machine learning to give input data with a class label. An email security software, for instance, might use natural language processing to categorize emails as ’spam’ or ’not spam’ based on their content (Rencher, 2002; Shelley, 2007).

The DA process is largely exploratory. Insofar as they produce well defined criteria that may be used to fresh item assignment, classification techniques are less exploratory than other processes. Discrimination typically calls for less problem structure than classification does (Johnson & Wichern, 2007). First developed by Fisher (1936), linear discriminant analysis is one of the most widely used discriminant analysis methods.

Relationships with CCA and Discriminant Analysis Techniques

The relationship between discriminant analysis and CCA is covered in Chapter 3. Additionally discussed in that chapter are some benefits of CCA over discriminant analysis when dealing with numerous groups. Discriminant analysis is essential for statistical pattern identification. According to (Samarov, 2009)'s research, there are many difficulties that could be related to the problem of classifying a set of observations into various groups. Finding any logical groupings in the data is the goal when categorizing or clustering because the categories are unknown. Data classification is frequently done using linear discriminant analysis (Duda, Hart, & Stork, 2000; Fisher, 1936). This section provides an overview of linear discriminant analysis and illustrates how it applies to CCA.

Chapter Summary

The body of work demonstrates how thorough the idea of canonical correlation analysis has been researched. A few research delved into great detail, and publications also included data analytic applications. In addition to the subject's concentration on mathematics, it is clear that many authors have used a variety of strategies in dealing with the canonical correlation analysis and discriminant analysis techniques. In numerous presentations, key matrices have been stated without any justification. There are some research that appear to have derived the key matrices but with obvious disconnections. The diverse methods frequently result in misconceptions about the concept of canonical correlation analysis. This chapter has discussed a few extensions and applications of discriminant analysis and canonical correlation analysis. A number of canonical correlation analysis applications, such as the relationship between discriminant canonical correlation analysis and canonical correlation analysis, have been reviewed.

The literature makes it abundantly evident that little research has been done on the use of CCA in multivariate time-dependent data. Thus, this area still remains grey for further exploration.

CHAPTER THREE

RESEARCH METHODS

Introduction

This chapter looks at the primary technique this thesis employed and delves deeply into the main strategies it employed to arrive at its conclusions. Review of data notation for numerous multivariate multiple time-dependent variables are considered. Reviewing the broad key matrices as well as the fundamental prerequisites and presumptions for using the matrices are also pertinent. This chapter provides some multivariate multiple discriminant analysis methods for selecting a reasonable model, along with CCA methods for these models. All the formulae used in this thesis are thoroughly discussed in this chapter.

Research Design and Source of Data

This thesis is designed as both data-based and technique-based type of study. Secondary data is used for the modeling and analysis. The monthly types of weather conditions data which span from 2000 to 2021 with sample size of n, equals Two hundred and sixty-four (264) months was collected from Ghana Meteorological Agency (G.Met) through the Department of Geography, University of Cape Coast. This analysis' primary goal is to generate ideas of grouping scheme multiple discriminant canonical correlation analysis (GSMDCCA) and the CCA general results from the given software outputs to know to what extent is the response variables influenced by the predictor variables and what are those measures that should be taken based on the results obtained. The necessary data is empirically analyzed using Matlab, Minitab and IBM SPSS Statistical Softwares in addition to manual calculations. The Latex software is used to generate the entire work.

Descriptions of the link between the given weather conditions

The response variables set covers 3 variables given by $\mathbf{Y} = (Y_1, Y_2, Y_3)$, where Y_1 denotes maximum temperature, Y_2 is the minimum temperature, and Y_3 is the solar radiation, with 264 observations. Similarly, the predictor variables covers 3 variables given by $\mathbf{X} = (X_1, X_2, X_3)$, where X_1 is the precipitation, X_2 is the wind, and X_3 denotes relative humidity, also with 264 observations.

I must completely explain the relationship between the response variables and the predictor variables using the specified weather conditions from the literature in order to obtain the answers to the structural equations for each of the three roots that are displayed in Figure 1 of Chapter One. As shown in the literature, temperature is directly proportional to the global solar radiation and dew point also relates directly to the temperatures. This means that decrease in solar radiation decreases the temperatures and increase in solar radiation leads to an increase in temperatures (Knappenberger, 1993; Mehdi, 2020; Richardson, 1981). The solar radiation, minimum temperature and maximum temperature are known as the heating variables whereas, the precipitation, wind and relative humidity in the atmosphere are known as the cooling variables (UGC, 2022).

The average precipitation will rise as a result of low evaporation as heating variables at the surface of the planet fall or become lower. It follows that, the climate becomes cool which leads to an increase precipitation in several places. Similarly, if the heating variables become high or increase, the climate becomes very warm which is expected to decrease precipitation in several places (Knappenberger, 1993; Mehdi, 2020; Richardson, 1981).

Relative humidity in the atmosphere has an inverse relationship with heating variables. If the heating variables become very low or decrease it will lead to an increase in relative humidity, thus the air becomes wet. Also if the heating variables increase, the air becomes dry, which follows that the relative humidity will decrease (Knappenberger, 1993; Mehdi, 2020; Richardson, 1981).

Principles and Analytical Approach of Canonical Correlation Analysis

Given two interrelated random vectors $\mathbf{Y} = (Y_1, Y_2, ..., Y_p)'$ and $\mathbf{X} = (X_1, X_2, ..., X_q)'$, assume, for convenience, that $p \leq q$. The number of variables in each set of variables, $p \leq q$, is used to determine the random vectors. The combined covariance matrix is produced by using the enhanced random vectors. Let $E(\mathbf{Y}) = \mu_y$ and $E(\mathbf{X}) = \mu_x$ be their respective expectations. The resulting combined (Johnson & Wichern, 2007; Mazuruse, 2014) random vector, **Z**, and its mean vector, μ are, respectively, given as

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Y} \\ \cdots \\ \mathbf{X} \end{bmatrix} \text{ and } \mu = \begin{bmatrix} E(\mathbf{Y}) \\ \cdots\cdots \\ E(\mathbf{X}) \end{bmatrix}$$

The combined covariance matrix (Σ) for the enhanced random vector is

$$\begin{aligned} \Sigma &= E(\mathbf{Z} - \mu)(\mathbf{Z} - \mu)' \\ &= \left[\begin{array}{c:c} E(Y - \mu_y)(Y - \mu_y)' & E(Y - \mu_y)(X - \mu_x)' \\ \hdashline E(X - \mu_x)(Y - \mu_y)' & E(X - \mu_x)(X - \mu_x)' \end{array}\right] \end{aligned}$$

$$\Sigma = \left[\begin{array}{c:c} \Sigma_{11} & \Sigma_{12} \\ \hdashline \Sigma_{21} & \Sigma_{22} \end{array}\right] \quad (3.1)$$

where (Ronald, 2011) $D(Y) = Cov(Y, Y) = E(Y - \mu_y)(Y - \mu_y)' = \Sigma_{11}$ is a ($p \times p$) and $D(X) = Cov(X, X) = E(X - \mu_x)(X - \mu_x)' = \Sigma_{22}$ is a ($q \times q$) sample covariance matrices for variable sets Y and X, respectively, and $Cov(Y, X) = E(Y - \mu_y)(X - \mu_x)' = \Sigma_{12} = \Sigma'_{21}$ is a sample matrix of the cross-covariance between Y and X. Thus, Σ is a block matrix, where Σ_{11} and Σ_{22} stand for the respective within-sets covariance matrices and $\Sigma_{12} = \Sigma'_{21}$ is the between-sets covariance matrix. The notations used follow the convention in the literature (e.g; Borga, 2001; Ignacio et al., 2008; Ronald, 2011; Zhou, Lu,

& Cheung, 2017).

Constructions of canonical variables

The CCA aims to express any potential correlation structure between the first set of variables (**Y**) and the second set of variables (**X**) in terms of a few linear combinations, $\alpha_i'\mathbf{Y}$ and $\beta_j'\mathbf{X}$, respectively. Suppose two datasets are given, then CCA produces canonical pairs of linear combinations as many as min(p, q). The two linear combinations are presented as follows: Both the linear combination U and the linear combination V are obtained from the first set of variables, Y, and the second set of variables, X, respectively (Carroll, 2006). Let α_i and β_j be the canonical coefficient vectors for a fixed number of i = 1, 2,..., p and j = 1, 2,..., q. Then, $U_i = \alpha_i'\mathbf{Y}$ and $V_j = \beta_j'\mathbf{X}$ are the two linear combinations of Y and X, respectively. Consequently, the linear combinations for variance and covariance matrices of CCA variates are given by the following equations. $Var(U_i) = \alpha_i'\Sigma_{11}\alpha_i$, $Var(V_j) = \beta_j'\Sigma_{22}\beta_j$, and $Var(U_i, V_j) = \alpha_i'\Sigma_{12}\beta_j$. The coefficient vectors α_i and β_j are obtained (Mazuruse, 2014; Reiter, 2010) such that

$$Cov(\alpha_i'\mathbf{Y}, V_j) = Cov(U_i, \beta_j'\mathbf{X}) = 0$$

$$Var(\alpha_i'\mathbf{Y}, U_i) = Var(\beta_j'\mathbf{X}, V_j) = 1$$

Using Σ_{11}, Σ_{22}, and Σ_{12} as inputs (Amit & Sharmishtha, 1998), the fundamental goal of CCA is to identify α and β to maximize the correlations between U_i and V_j. The correlation between $U = \alpha'\mathbf{Y}$ and $V = \beta'\mathbf{X}$ is given as

$$\rho_i = Corr(U, V)$$

$$\Rightarrow \rho_i = \frac{\alpha'\Sigma_{12}\beta}{\sqrt{\alpha'\Sigma_{11}\alpha}\sqrt{\beta'\Sigma_{22}\beta}} \tag{3.2}$$

Let f = min (p, q), and i = 1, 2, ..., f, then the next goal of CCA is to identify canonical variates $U_i = \alpha'\mathbf{Y}$ and $V_j = \beta'\mathbf{X}$ so that the group of variables U_i and the group of variables V_j are both uncorrelated with unit variance. It follows that

$$Cov(U_i, U_j) = Cov(V_i, V_j) = \begin{cases} 0; & for\ i \neq j \\ 1; & for\ i = j \end{cases}$$

Since ρ_i is the canonical correlation coefficient between U_i and V_j, for i = 1, 2, ..., f, then

$$Cov(U_i, V_j) = Cov(V_j, U_i) = 0; for\ i \neq j$$

$$Cov(U_i, V_j) = Cov(V_j, U_i) = \rho_i;\ for\ i = j;\ i = 1, 2, ..., f.$$

Hence, for p = 3, we have U_1, U_2, U_3 as canonical variates for the first set (**Y**) and for q = 5, we have V_1, V_2, V_3, V_4, V_5 as canonical variates for the second set (**X**). Thus for p = 3 and q = 5, the correlation matrix for $\mathbf{U}' = [U_1, U_2, U_3]$ and $\mathbf{V}' = [V_1, V_2, V_3, V_4, V_5]$ has the form as given in the general canonical correlation matrix below (Nail, 2002; Samarov, 2009).

$$\begin{array}{c} U_1..U_2..U_3...V_1..V_2..V_3..V_4..V_5 \\ \begin{array}{c} U_1 \\ U_2 \\ U_3 \\ V_1 \\ V_2 \\ V_3 \\ V_4 \\ V_5 \end{array} \left(\begin{array}{ccc:ccc:cc} 1 & 0 & 0 & \rho_1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & \rho_2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & \rho_3 & 0 & 0 \\ \hdashline \rho_1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & \rho_2 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \rho_3 & 0 & 0 & 1 & 0 & 0 \\ \hdashline 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right) \end{array}$$

The variables V_4 and V_5 thus constitute a set $\mathbf{V}^R$ that is uncorrelated with $\mathbf{U}$ and $\mathbf{V}\backslash\mathbf{V}^R$ and are therefore the redundant set in $\mathbf{V}$. In order to optimize the correlation between U_i and V_j, the canonical coefficient vectors α and β must be chosen. The optimization issue can be found in Equation (3.3) for any normalization of α and β (Marden, 2015, Samarov, 2009).

$$\rho_i^* = \max_{\alpha,\beta} \frac{\alpha'\Sigma_{12}\beta}{\sqrt{\alpha'\Sigma_{11}\alpha}\sqrt{\beta'\Sigma_{22}\beta}} \tag{3.3}$$

The optimization problem in Equation (3.3) can be reduced to the form

$$\rho_i^* = \max_{\alpha,\beta} Corr(\alpha'\mathbf{Y}, \beta'\mathbf{X}) = \alpha'\Sigma_{12}\beta \tag{3.4}$$

subject to the constraints: $\alpha'\Sigma_{11}\alpha = 1$ and $\beta'\Sigma_{22}\beta = 1$.

Canonical variables by Cauchy Swartz inequality

Since Σ_{11} and Σ_{22} are positive definite, $Cov(U, V)$ can be written as

$$Corr(\alpha'\mathbf{Y}, \beta'\mathbf{X}) = \frac{\alpha'\Sigma_{12}\beta}{\sqrt{\alpha'\Sigma_{11}\alpha}\sqrt{\beta'\Sigma_{22}\beta}} \tag{3.5}$$

Let $\Sigma_{11}^{\frac{1}{2}}\alpha = \mathbf{w}_y$, $\Rightarrow$ $\alpha = \Sigma_{11}^{-\frac{1}{2}}\mathbf{w}_y$ and $\Sigma_{22}^{\frac{1}{2}}\beta = \mathbf{w}_x$, $\Rightarrow$ $\beta = \Sigma_{22}^{-\frac{1}{2}}\mathbf{w}_x$; where $\mathbf{w}_y$ and $\mathbf{w}_x$ are new coefficient vectors. Substituting α and β into Equation (3.5) gives

$$Corr(\alpha'\mathbf{Y}, \beta'\mathbf{X}) = \frac{\mathbf{w}_y'\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}\mathbf{w}_x}{\sqrt{\mathbf{w}_y'\mathbf{w}_y}\sqrt{\mathbf{w}_x'\mathbf{w}_x}} \tag{3.6}$$

By Cauchy Swartz inequality, we have

$$\mathbf{w}_y'\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}\mathbf{w}_x \leq \sqrt{\mathbf{w}_y'\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}}\sqrt{\mathbf{w}_x'\mathbf{w}_x} \tag{3.7}$$

If the maximum is achieved at **y** = e_1, then a real symmetric matrix **M** contains the eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p$ and eigenvectors $e_1, e_2, ..., e_p$ such that

$\max_{\mathbf{y}}\left(\frac{\mathbf{y}'\mathbf{My}}{\mathbf{y}'\mathbf{y}}\right) = \lambda_1$. Thus, $\mathbf{y}'\mathbf{My} \leq \lambda_1 \mathbf{y}'\mathbf{y}$ and Equation (3.8) is attained.

$$\mathbf{w}_y' \Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}} \mathbf{w}_y \leq \rho_1^{*2} \mathbf{w}_y' \mathbf{w}_y \tag{3.8}$$

In Equation (3.8) equality is attained at $\mathbf{w}_y = \mathbf{e}_1$ and in Equation (3.7) equality holds if $\mathbf{w}_x = \Sigma_{22}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{11}^{-\frac{1}{2}} \mathbf{e}_1$.

That is; $\Sigma_{11}^{\frac{1}{2}} \alpha = \mathbf{w}_y = \mathbf{e}_1$, $\Rightarrow \alpha = \Sigma_{11}^{-\frac{1}{2}} \mathbf{e}_1$ and $\beta = \Sigma_{22}^{-1} \Sigma_{12} \Sigma_{11}^{-\frac{1}{2}} \mathbf{e}_1$ Equation (3.8), therefore, gives

$$\max_{\alpha,\beta} Corr(\alpha'\mathbf{Y}, \beta'\mathbf{X}) \leq \sqrt{\frac{\rho_1^{*2} \mathbf{w}_y' \mathbf{w}_y}{\mathbf{w}_y' \mathbf{w}_y}} = \sqrt{\rho_1^{*2}} = \rho_1^*$$

Hence $\rho_i^* = \alpha' \Sigma_{12} \beta$ as required.

Computations of Canonical Variates and Canonical Correlations

Given linear combinations of the original variables, Y and X are utilized to determine all canonical pairs of variables.

The first canonical pair of variables

If the first canonical variate pair is (U_1, V_1), where the two linear combinations, U_1 and V_1 are given by $U_1 = \alpha_1' \mathbf{Y}$ and $V_1 = \beta_1' \mathbf{X}$, respectively, and let the correlation between U_1 and V_1 be equal to ρ_1 . We select the coefficients $\alpha_1 = (\alpha_{11}, \alpha_{12}, ..., \alpha_{1p})$ and $\beta_1 = (\beta_{11}, \beta_{12}, ..., \beta_{1q})$ to maximize the canonical correlation, given by ρ_1^*, between the first pair of canonical variate, subject to the constraints: $Var(U_1) = Var(V_1) = 1$. The required first pair of canonical variate equations are given as

$$U_1 = \alpha_{11} Y_1 + \alpha_{12} Y_2 + ... + \alpha_{1p} Y_p$$

$$V_1 = \beta_{11} X_1 + \beta_{12} X_2 + ... + \beta_{1q} X_q$$

The first pair of canonical variates (Marubayashi, André, Luciano, Rodrigo, & Elias, 2014) can be written in terms of Σ_{11} and Σ_{22} as: $U_1 = \mathbf{u}_1' \Sigma_{11}^{-\frac{1}{1}} \mathbf{Y}$ and $V_1 = \mathbf{v}_1' \Sigma_{22}^{-\frac{1}{1}} \mathbf{X}$, where $\mathbf{u}_1$ is the first eigenvector of the matrix $\mathbf{P}_1 = \Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}}$ and $\mathbf{v}_1$ is the first eigenvector of the matrix $\mathbf{P}_2 = \Sigma_{22}^{-\frac{1}{2}} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-\frac{1}{2}}$ (Helwig, 2017). The correlation between U_1 and V_1 is given by

$$\rho_1 = Corr(U_1, V_1) = Corr(\alpha'\mathbf{Y}, \beta'\mathbf{X}) = \frac{Cov(U_1, V_1)}{\sqrt{Var(U_1)}\sqrt{Var(V_1)}} \tag{3.9}$$

Thus, Equation (3.10) is the maximization of the first canonical correlation coefficient, where Σ_{11} and Σ_{22} are positive definite matrices.

$$\rho_1^* = \max_{\alpha,\beta} \frac{\alpha_{1i}' \Sigma_{12} \beta_{1j}}{\sqrt{\alpha_{1i}' \Sigma_{11} \alpha_{1i}} \sqrt{\beta_{1j}' \Sigma_{22} \beta_{1j}}} \tag{3.10}$$

where ρ_1^2 is the first eigenvalue of the matrix $\mathbf{P}_1 = \Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}}$ [ρ_1^2 is also the first eigenvalue of the matrix $\mathbf{P}_2 = \Sigma_{22}^{-\frac{1}{2}} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-\frac{1}{2}}$] (Borga, 2001; Helwig, 2017). The first canonical correlation's goal is to establish $\alpha_{11}, \alpha_{12}, ..., \alpha_{1p}$ and $\beta_{11}, \beta_{12}, ..., \beta_{1q}$ (Mazuruse, 2014; Sharma, 1996) in such a way that the canonical correlation, ρ_1^*, between U_1 and V_1, is at its highest.

The k^{th} canonical pair of variables

The k^{th} pair of canonical variates is identified such that ρ_k^* is maximum. Now,

$$U_k = \alpha_{k1} Y_1 + \alpha_{k2} Y_2 + ... + \alpha_{kp} Y_p$$

$$V_k = \beta_{k1} X_1 + \beta_{k2} X_2 + ... + \beta_{kq} X_q$$

where U_k and V_k are the k^{th} pair of canonical variates for $\mathbf{Y}$ and $\mathbf{X}$, respectively. The k^{th} pair of canonical variates can be written in terms of Σ_{11} and Σ_{22} as: $U_k = \mathbf{u}_k' \Sigma_{11}^{-\frac{1}{1}} \mathbf{Y}$ and $V_k = \mathbf{v}_k' \Sigma_{22}^{-\frac{1}{1}} \mathbf{X}$, where $\mathbf{u}_k$ is the k^{th} eigenvector of the matrix $\mathbf{P}_1 = \Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}}$ and $\mathbf{v}_k$ is the k^{th} eigenvector of the matrix

$\mathbf{P}_2 = \Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}$ (Helwig, 2017). The two linear combinations U_k and V_k, both of which have unit variance and maximize the correlation among all other possible linear combinations, make up the k^{th} canonical correlation. Hence, maximization of the correlation between U_k and V_k is given by Equation (3.11), where Σ_{11} and Σ_{22} are positive definite matrices.

$$\rho_k^* = \max_{\alpha,\beta} \frac{\alpha'_{ki}\Sigma_{12}\beta_{kj}}{\sqrt{\alpha'_{ki}\Sigma_{11}\alpha_{ki}}\sqrt{\beta'_{kj}\Sigma_{22}\beta_{kj}}} \tag{3.11}$$

where ρ_k^2 is the k^{th} eigenvalue of the matrix $\mathbf{P}_1 = \Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}$ and ρ_k^2 is also the k^{th} eigenvalue of the matrix $\mathbf{P}_2 = \Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}$ (Helwig, 2017).

To identify the k sets of canonical variate pairs; $(U_1, V_1), (U_2, V_2), ..., (U_k, V_k)$; is the general objective of CCA (Mazuruse, 2014) such that the corresponding canonical correlations; $\rho_1, \rho_2, ..., \rho_k$; are mutually maximized. The objective of CCA is to increase the correlation between two linear combinations of the two variables supplied by the initial canonical correlation, according to (Coleman & Hardin, 2013; Mazuruse, 2014). Clearly, canonical correlation is a maximization problem with restrictions.

Fundamental Approaches to Canonical Correlation Analysis

The two fundamental approaches of the concept of CCA need to be highlighted to identify the key matrices for its extraction. The main matrices involved in the construction of the canonical variables are generated through two fundamental approaches of the Lagrange Multiplier and the Cauchy-Schwarz Inequality.

Lagrange multipliers approach of the canonical variables

Assume that the Lagrange multipliers ρ_y and ρ_x are connected to the (Obben, 1992) variance-covariance matrices of $\mathbf{Y}$ and $\mathbf{X}$, respectively. Then, by, the corresponding Lagrangian, we have

$$L(\alpha, \beta, \rho_y, \rho_x) = \alpha' \Sigma_{12} \beta - \frac{\rho_y}{2} (\alpha' \Sigma_{11} \alpha - 1) - \frac{\rho_x}{2} (\beta' \Sigma_{22} \beta - 1) \quad (3.12)$$

Equation (3.13) is obtained by taking a partial derivative with regard to α on both sides of Equation (3.12), and then setting the resultant equation to zero.

$$\frac{\partial}{\partial \alpha} L(\alpha, \beta, \rho_y, \rho_x) = \frac{\partial}{\partial \alpha} \left[\alpha' \Sigma_{12} \beta - \frac{\rho_y}{2} (\alpha' \Sigma_{11} \alpha - 1) - \frac{\rho_x}{2} (\beta' \Sigma_{22} \beta - 1) \right]$$

$$\Sigma_{12} \beta - \rho_y \Sigma_{11} \alpha = 0 \quad (3.13)$$

Multiply both sides of Equation (3.13) by α' yields

$$\alpha' \Sigma_{12} \beta - \rho_y \alpha' \Sigma_{11} \alpha = 0 \quad (3.14)$$

Similarly, with respective to β and further simplification gives

$$\Sigma_{21} \alpha - \rho_x \Sigma_{22} \beta = 0 \quad (3.15)$$

and hence,

$$\beta' \Sigma_{21} \alpha - \rho_x \beta' \Sigma_{22} \beta = 0 \quad (3.16)$$

From Equations (3.16) and (3.14), we have

$$\alpha' \Sigma_{12} \beta - \rho_y \alpha' \Sigma_{11} \alpha - [\beta' \Sigma_{21} \alpha - \rho_x \beta' \Sigma_{22} \beta] = 0$$
$$\Rightarrow -\rho_y \alpha' \Sigma_{11} \alpha + \rho_x \beta' \Sigma_{22} \beta = 0$$
$$\Rightarrow \rho_x = \rho_y = \rho_F$$

Since $\rho_x = \rho_y = \rho_F$, the generalized eigenvalue decomposition problem can then be written from Equations (3.13) and (3.15) as

$$\begin{bmatrix} 0 & \Sigma_{12} \\ \Sigma_{21} & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \rho_F \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \tag{3.17}$$

If Σ_{11} and Σ_{22} are positive definite and invertible, then α and β can be expressed in the following ways:

From Equation (3.13),

$$\alpha = \frac{\Sigma_{11}^{-1}\Sigma_{12}\beta}{\rho_F}$$

Also from Equation (3.15),

$$\beta = \frac{\Sigma_{22}^{-1}\Sigma_{21}\alpha}{\rho_F}$$

Putting $\beta = \frac{\Sigma_{22}^{-1}\Sigma_{21}\alpha}{\rho_F}$ into Equation (3.13) and simplifying yields

$$\frac{\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\alpha}{\rho_F} - \rho_F\Sigma_{11}\alpha = 0$$

$$\Rightarrow \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\alpha - \rho_F^2\alpha = 0$$

$$\Rightarrow \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\alpha = \rho_F^2\alpha \tag{3.18}$$

Hence the eigenvalue of the matrix, $\mathbf{Q}_1$ given in Equation (3.18) is $\lambda_1 = \rho_F^2$, where

$$\mathbf{Q}_1 = \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \tag{3.19}$$

Let $\alpha = \Sigma_{11}^{-\frac{1}{2}}\mathbf{e}_1$, then from Equation (3.18), we have

$$\Sigma_{11}^{-\frac{1}{2}}\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\mathbf{e}_1 = \lambda_1\Sigma_{11}^{-\frac{1}{2}}\mathbf{e}_1$$

$$\Rightarrow \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\left(\Sigma_{11}^{-\frac{1}{2}}\mathbf{e}_1\right) = \lambda_1\left(\Sigma_{11}^{-\frac{1}{2}}\mathbf{e}_1\right)$$

Hence the first eigenvector of the matrix $\mathbf{Q}_1$ is $\Sigma_{11}^{-\frac{1}{2}}\mathbf{e}_1$. By expressing Equation (3.18) as a determinant equation as

$$\left|\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \lambda_1\mathbf{I}\right| = 0,$$

then

$$\left|\Sigma_{11}^{-\frac{1}{2}}\right|\left|\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}} - \lambda_1\mathbf{I}\right|\left|\Sigma_{11}^{\frac{1}{2}}\right| = 0$$

$$\left|\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}} - \lambda_1\mathbf{I}\right| = 0$$

$$\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\mathbf{e}_1 - \lambda_1\mathbf{e}_1 = 0$$

$$\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\mathbf{e}_1 = \lambda_1\mathbf{e}_1 \quad (3.20)$$

This means that the eigenvalue of the matrix,

$$\mathbf{P}_1 = \Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}} \quad (3.21)$$

is λ_1, which is the same as that of matrix $\mathbf{Q}_1$ and $\mathbf{e}_1$ is the normalized eigenvector associated with λ_1. Suppose the matrix $\mathbf{A} = \Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}$ is a parameter that maximizes the correlation between the i^{th} canonical pair of variables. Multiplying both sides of Equation (3.20) by $\mathbf{A}'$ and simplifying the resulting result yields

$$\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\left[\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\right]\mathbf{e}_1 = \lambda_1\left[\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\right]\mathbf{e}_1$$

$$\Rightarrow \Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}\left[\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\right]\mathbf{e}_1 = \lambda_1\left[\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\right]\mathbf{e}_1$$

Hence λ_1 is the eigenvalue of the matrix

$$\mathbf{P}_2 = \Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}} \quad (3.22)$$

which is the same as that of $\mathbf{Q}_1$ and $\mathbf{P}_1$. The corresponding eigenvector, f_1 of the matrix $\mathbf{P}_2$ is the normalized form of $\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\mathbf{e}_1$.

Similarly, putting $\alpha = \frac{\Sigma_{11}^{-1}\Sigma_{12}\beta}{\rho_F}$ into Equation (3.15) and simplify yields

$$\frac{\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\beta}{\rho_F} - \rho_F\Sigma_{22}\beta = 0$$

$$\Rightarrow \Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\beta - \rho_F^2\beta = 0$$

$$\Rightarrow \Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\beta = \rho_F^2\beta \tag{3.23}$$

This means that the eigenvalue of the matrix,

$$\mathbf{Q}_2 = \Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \tag{3.24}$$

is λ_1, which is the same as that of $\mathbf{Q}_1$, $\mathbf{P}_1$, and $\mathbf{P}_2$. It follows from the matrix results for $\mathbf{P}_2$ that the eigen equation is given as

$$\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}\left(\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\mathbf{e}_1\right) = \lambda_1\left(\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\mathbf{e}_1\right) \tag{3.25}$$

Multiply through Equation (3.25) by $\Sigma_{22}^{-\frac{1}{2}}$ gives

$$\Sigma_{22}^{-\frac{1}{2}}\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}\left(\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\mathbf{e}_1\right) = \lambda_1\left(\Sigma_{22}^{-\frac{1}{2}}\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\mathbf{e}_1\right)$$
$$\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\left(\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\mathbf{e}_1\right) = \lambda_1\left(\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\mathbf{e}_1\right)$$

Therefore, (λ_1, f_2) is the eigenvalue-eigenvector pair of the matrix $\mathbf{Q}_2$ and f_2 is the first column of the normalized form of the matrix $\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\mathbf{e}_1$. Noting that $\mathbf{AA}' = \mathbf{P}_1$ (see summary of key matrices), it implies that the first five matrices $\mathbf{P}_1, \mathbf{P}_2, \mathbf{Q}_1, \mathbf{Q}_2$, and $\mathbf{A}$ have the same non–zero eigenvalues. Thus, for the linear combinations $U = \alpha'\mathbf{Y}$ and $V = \beta'\mathbf{X}$

$$\rho_1^* = \max_{\alpha,\beta} Corr\left[\mathbf{e}_1'\Sigma_{11}^{-\frac{1}{2}}\mathbf{Y},\ \mathbf{f}_1'\Sigma_{22}^{-\frac{1}{2}}\mathbf{X}\right]$$

is attained for $U = \mathbf{e}_1'\Sigma_{11}^{-\frac{1}{2}}\mathbf{Y}$ and $V = \mathbf{f}_1'\Sigma_{22}^{-\frac{1}{2}}\mathbf{X}$.

Cauchy-Schwarz inequality approach of extracting canonical variates

Suppose that both Σ_{11} and Σ_{22} are positive definite in order to maximize the optimization problem. The optimization problem is then given by

$$\rho_i^* = \max_{\alpha,\beta} \frac{\alpha'\Sigma_{12}\beta}{\sqrt{\alpha'\Sigma_{11}\alpha}\sqrt{\beta'\Sigma_{22}\beta}}$$

$$\rho_i^* = \max_{\alpha,\beta} \frac{\alpha'\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}\beta}{\sqrt{\alpha'\alpha}\sqrt{\beta'\beta}} \tag{3.26}$$

Let $p \leq q$, and without loss of generality, let $\alpha = \Sigma_{11}^{-\frac{1}{2}}\theta_y$ and $\beta = \Sigma_{22}^{-\frac{1}{2}}\theta_x$, where θ_y and θ_x are the new correlation coefficient matrices of $\mathbf{Y}$ and $\mathbf{X}$, respectively. Substituting $\alpha = \Sigma_{11}^{-\frac{1}{2}}\theta_y$ and $\beta = \Sigma_{22}^{-\frac{1}{2}}\theta_x$ into Equation (3.26) and simplifying yields

$$\rho_i^* = \max_{\theta_y,\theta_x} \frac{\theta_y'\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}\theta_x}{\sqrt{\theta_y'\theta_y}\sqrt{\theta_x'\theta_x}} \tag{3.27}$$

The numerator of Equation (3.27) is maximized, subject to the constraints: $\theta_y'\theta_y = 1$ and $\theta_x'\theta_x = 1$. Hence, if the vectors $\mathbf{s} = \Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\theta_y$;
$\Rightarrow \mathbf{s}' = \theta_y'\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}$; $\mathbf{t} = \theta_x$.

To apply the Cauchy-Schwarz inequality (Mukhopadhyay, 2000; Nelsen, 1994; Schwarz, 1888; Win & Wu, 2000), the square root on both sides of the inequality yields

$$(\mathbf{s}'\mathbf{t}) \leq (\mathbf{s}'\mathbf{s})^{\frac{1}{2}}(\mathbf{t}'\mathbf{t})^{\frac{1}{2}} \tag{3.28}$$

Hence, the substitution gives

$$(\mathbf{s}'\mathbf{t}) = \left[\theta_y'\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\theta_y\right]^{\frac{1}{2}} [\theta_x{}'\theta_x]^{\frac{1}{2}}$$

$$= \left[\theta_y'\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\theta_y\right]^{\frac{1}{2}} [\theta_x{}'\theta_x]^{\frac{1}{2}} \tag{3.29}$$

From the maximization results, suppose that λ_1 is the largest eigenvalue of the matrix $\mathbf{P}_1$ given as

$$\mathbf{P}_1 = \Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}$$

Then $(\mathbf{s}'\mathbf{t})^2 \leq \lambda_1 {\theta_y}'\theta_y$ and the equality holds if $\theta_y = \mathbf{e}_1$, the normalized eigenvector associated with λ_1. Thus, the vector $\mathbf{s}$ can now be written as;

$\mathbf{s} = \Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\mathbf{e}_1$ and $\theta_x \propto \Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\mathbf{e}_1$.

It follows that the first linear transformation U_1, of the vector $\mathbf{Y}$ given the p linear transformation $U_1 = \mathbf{M}'\mathbf{Y}$ with coefficient as the first column $\alpha = \mathbf{M}_1$ of the transformation matrix $\mathbf{M}$ given by

$$\mathbf{M}_1 = \Sigma_{11}^{-\frac{1}{2}}\mathbf{e}_1$$

where; $\mathbf{e}_1 = \text{Eig}\left(\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\right)_1$. Thus, $\mathbf{M}_1 = \Sigma_{11}^{-\frac{1}{2}}\text{Eig}\left(\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\right)_1$ which is the matrix of transformation (Apanyin, 2021). It follows from definitions above that the eigen equation is given by

$$\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\mathbf{e}_1 = \lambda_1\mathbf{e}_1 \tag{3.30}$$

Given the matrix $\mathbf{A} = \Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}$, multiplying both sides of Equation (3.30) by $\mathbf{A}'$ and simplifying the result yields

$$\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\left[\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\right]\mathbf{e}_1 = \lambda_1\left[\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\right]\mathbf{e}_1$$

$$\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}\left[\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\right]\mathbf{e}_1 = \lambda_1\left[\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\right]\mathbf{e}_1$$

Hence λ_1 is the eigenvalue of the matrix $\mathbf{P}_2$ given as

$$\mathbf{P}_2 = \Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}$$

The corresponding eigenvector, f_1, is the normalized form of the matrix, $\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\mathbf{e}_1$. The first eigenvector of $\mathbf{P}_2$ is now given by Eig$\left(\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}\right)_1$

It follows that the first linear transformation V_1, of the vector $\mathbf{X}$ given the q linear transformation $V_1 = (\mathbf{N}'\mathbf{X})_1$ with coefficient as the first column $\beta = \mathbf{N}_1$ of the transformation matrix $\mathbf{N}$ given by

$$\mathbf{N}_1 = \Sigma_{22}^{-\frac{1}{2}} Eig\left(\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}\right)_1$$

and the i^{th} linear transformation is given as

$$\mathbf{N}_i = \Sigma_{22}^{-\frac{1}{2}} Eig\left(\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}\right)_i$$

Expressing Equation (3.30) as a determinant equation gives

$$\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\mathbf{e}_1 - \lambda_1\mathbf{e}_1 = 0$$

$$\left|\Sigma_{11}^{-\frac{1}{2}}\right|\left|\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}} - \lambda_1\mathbf{I}\right|\left|\Sigma_{11}^{\frac{1}{2}}\right| = 0$$

$$\left|\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \lambda_1\mathbf{I}\right| = 0 \tag{3.31}$$

This means that the eigenvalue of the matrix

$$\mathbf{Q}_1 = \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

is λ_1, which is the same as that of the matrices $\mathbf{P}_1$ and $\mathbf{P}_2$ in Equations (3.21) and (3.22). Multiply both sides of Equation (3.30) by $\Sigma_{11}^{-\frac{1}{2}}$ to determine the corresponding eigenvector.

$$\Sigma_{11}^{-\frac{1}{2}}\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\mathbf{e}_1 = \lambda_1\Sigma_{11}^{-\frac{1}{2}}\mathbf{e}_1$$

$$\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\left(\Sigma_{11}^{-\frac{1}{2}}\mathbf{e}_1\right) = \lambda_1\left(\Sigma_{11}^{-\frac{1}{2}}\mathbf{e}_1\right) \tag{3.32}$$

Hence the first eigenvector of $\mathbf{Q}_1$ is $\Sigma_{11}^{-\frac{1}{2}}\mathbf{e}_1$ which is the same as $\mathbf{M}_1$. Expressing the matrix $\mathbf{P}_2$ as a determinant equation gives

$$\left|\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} - \lambda_1\mathbf{I}\right| = 0$$

This means that the eigenvalue of the matrix

$$\mathbf{Q}_2 = \Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$$

is λ_1, which is the same as that of the matrices $\mathbf{P}_1$, $\mathbf{P}_2$, and $\mathbf{Q}_1$ in Equations (3.19), (3.21) and (3.22). It may be shown similarly that the first eigenvector of $\mathbf{Q}_2$ is the normalized form of the vector $\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\mathbf{e}_1$.

Generalization of the canonical correlation results

The salient results of the review thus far must be written in some concise form (Apanyin, 2021). By considering the matrix $\mathbf{P}_i$ given as

$$\mathbf{P}_i = \Sigma_{ii}^{-\frac{1}{2}}\Sigma_{ik}\Sigma_{kk}^{-1}\Sigma_{ki}\Sigma_{ii}^{-\frac{1}{2}}; k = \begin{cases} i+1; & i = 1 \\ i-1; & i = 2 \end{cases}$$

the matrix $\mathbf{Q}_i$ is obtained in terms of $\mathbf{P}_i$ as follows:

$$\begin{aligned}\mathbf{Q}_i &= \Sigma_{ii}^{-\frac{1}{2}}\mathbf{P}_i\Sigma_{ii}^{\frac{1}{2}} = \Sigma_{ii}^{-\frac{1}{2}}\left(\Sigma_{ii}^{-\frac{1}{2}}\Sigma_{ik}\Sigma_{kk}^{-1}\Sigma_{ki}\Sigma_{ii}^{-\frac{1}{2}}\right)\Sigma_{ii}^{\frac{1}{2}} \\ &= \Sigma_{ii}^{-1}\Sigma_{ik}\Sigma_{kk}^{-1}\Sigma_{ki}\end{aligned}$$

Thus, $\mathbf{Q}_i$ has the same eigenvalues as $\mathbf{P}_i$ with the corresponding i^{th} eigenvectors of $\Sigma_{ii}^{-\frac{1}{2}} E_{N_i}$, where $E_{N_i} = Eig(\mathbf{P}_i)$, i = 1, 2; and

$$E_{N_i} = \begin{cases} E_{N_1}; & i = 1 \\ \Sigma_{22}^{-\frac{1}{2}} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}} E_{N_1}; & i = 2 \end{cases}$$

Canonical correlation variables from the generalized results

I need to further summarize the generalized canonical variables from the generalized results given. Let p and q linear transformations be defined by $\mathbf{u} = \vartheta_y' \mathbf{Y}$ and $\mathbf{v} = \vartheta_x' \mathbf{X}$, respectively, for the sub-vectors of $\mathbf{Z}' = (\mathbf{Y}, \mathbf{X})$ using the matrices $\mathbf{P}_1$ and $\mathbf{P}_2$. Then the new variables **u** and **v** constitute canonical variables if $\vartheta_y = \Sigma_{11}^{-\frac{1}{2}} E_{N_1}$ and $\vartheta_x = \Sigma_{22}^{-\frac{1}{2}} E_{N_2}$; and $\mathbf{P}_1$ and $\mathbf{P}_2$ have the same non-zero eigenvalues (Ankita & Paradipta, 2023; Apanyin, 2021).

Theorem 3.1

The matrices of transformation $\mathbf{P}_i$ and $\mathbf{Q}_i$, where i = 1, 2 are similar matrices.

Proof

$$\begin{aligned} \mathbf{Q}_i - \gamma \mathbf{I} &= \Sigma_{ii}^{-\frac{1}{2}} \mathbf{P}_i \Sigma_{ii}^{\frac{1}{2}} - \mathbf{Q}_i \gamma \mathbf{Q}_i^{-1} \\ &= \Sigma_{ii}^{-\frac{1}{2}} \mathbf{P}_i \Sigma_{ii}^{\frac{1}{2}} - \gamma \left(\Sigma_{ii}^{-\frac{1}{2}} \mathbf{P}_i \Sigma_{ii}^{\frac{1}{2}} \right) \left(\Sigma_{ii}^{-\frac{1}{2}} \mathbf{P}_i \Sigma_{ii}^{\frac{1}{2}} \right)^{-1} \\ &= \Sigma_{ii}^{-\frac{1}{2}} \mathbf{P}_i \Sigma_{ii}^{\frac{1}{2}} - \Sigma_{ii}^{-\frac{1}{2}} \gamma \Sigma_{ii}^{\frac{1}{2}} \\ &= \Sigma_{ii}^{-\frac{1}{2}} \left(\mathbf{P}_i - \gamma \mathbf{I} \right) \Sigma_{ii}^{\frac{1}{2}} \end{aligned}$$

It follows that

$$\begin{aligned} det\left(\mathbf{Q}_i - \gamma \mathbf{I}\right) &= det\left(\Sigma_{ii}^{-\frac{1}{2}}\right) det\left(\mathbf{P}_i - \gamma \mathbf{I}\right) det\left(\Sigma_{ii}^{\frac{1}{2}}\right) \\ &= det\left(\mathbf{P}_i - \gamma \mathbf{I}\right) \end{aligned}$$

This indicates that $\mathbf{P}_i$ and $\mathbf{Q}_i$ have equal roots in their characteristic equations.

Theorem 3.2

The matrices of transformation $\mathbf{P}_1$ and $\mathbf{P}_2$ by themselves are similar matrices.

Proof

$$\begin{aligned}\mathbf{P}_1 - \gamma\mathbf{I} = \mathbf{AA}' - \gamma\mathbf{P}_1\mathbf{P}_1^{-1} &= \mathbf{AA}' - \gamma\left(\mathbf{AA}'\right)\left(\mathbf{AA}'\right)^{-1} \\ &= \mathbf{AA}' - \gamma\mathbf{AA}^{-1} = \mathbf{A}\left(\mathbf{A}' - \gamma\mathbf{A}^{-1}\right) \\ &= \mathbf{A}\left(\mathbf{A}'\mathbf{A} - \gamma\mathbf{I}\right)\mathbf{A}^{-1} \\ &= \mathbf{A}\left(\mathbf{P}_2 - \gamma\mathbf{I}\right)\mathbf{A}^{-1}\end{aligned}$$

It follows that

$$det\left(\mathbf{P}_1 - \gamma\mathbf{I}\right) = det\left(\mathbf{A}\right)det\left(\mathbf{P}_2 - \gamma\mathbf{I}\right)det\left(\mathbf{A}^{-1}\right) = det\left(\mathbf{P}_2 - \gamma\mathbf{I}\right)$$

This means that the matrices of $\mathbf{P}_1$ and $\mathbf{P}_2$ have the same characteristic roots. In general, I can conclude that the first five matrices $\mathbf{P}_1, \mathbf{P}_2, \mathbf{Q}_1, \mathbf{Q}_2$, and $\mathbf{A}$ have equal non-zero eigenvalues, since $\mathbf{P}_i$ and $\mathbf{Q}_i$, where i = 1, 2 have equal characteristic roots, then, $\mathbf{P}_1$ and $\mathbf{P}_2$ also have equal characteristic roots.

Summary of key matrices and their relationships

It is noted in the definitions that the matrix $\Sigma_{ii}^{-\frac{1}{2}}$, where i = 1, 2 is a crucial matrix that gives the new variables and their basic property of independence. The matrix

$$\mathbf{A} = \Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}$$

is the basic matrix for determining the new variables which constitutes the matrix factor by which (Ankita & Paradipta, 2023; Apanyin, 2021) the i^{th} correlation between the pair of variables $(\mathbf{Y}, \mathbf{X})$ is maximum among the remaining

$(r{-}i + 1)$ pairs, where r = min (p, q). Some important matrices are $\mathbf{P}_i$, where i = 1, 2 and the matrix $\mathbf{P}_1$ may be written in terms of $\mathbf{A}$ as

$$\mathbf{AA}' = \Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}} = \Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}} = \mathbf{P}_1$$

Similarly,

$$\begin{aligned}\mathbf{A}'\mathbf{A} &= \Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}} \\ &= \Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}} \\ &= \mathbf{P}_2\end{aligned}$$

It follows from the definitions that the canonical variables are the ordered eigenvectors of the matrices $\mathbf{Q}_i$, i = 1, 2. The required eigenvectors are however extracted as the product of $\Sigma_{ii}^{-\frac{1}{2}}$, and the eigenvectors of $\mathbf{P}_i$. Then, we have to express $\mathbf{Q}_i$ in terms of $\mathbf{P}_i$ as

$$\begin{aligned}\Sigma_{ii}^{-\frac{1}{2}}\mathbf{P}_i\Sigma_{ii}^{\frac{1}{2}} &= \Sigma_{ii}^{-\frac{1}{2}}\left(\Sigma_{ii}^{-\frac{1}{2}}\Sigma_{ik}\Sigma_{kk}^{-1}\Sigma_{ki}\Sigma_{ii}^{-\frac{1}{2}}\right)\Sigma_{ii}^{\frac{1}{2}} \\ &= \Sigma_{ii}^{-1}\Sigma_{ik}\Sigma_{kk}^{-1}\Sigma_{ki} \\ &= \mathbf{Q}_i\end{aligned}$$

This implies that the first five matrices $\mathbf{P}_1, \mathbf{P}_2, \mathbf{Q}_1, \mathbf{Q}_2$, and $\mathbf{A}$ have the same non–zero eigenvalues. These are seen to have the same eigenvalue or characteristic roots as already observed.

Let $\{Q_{it}\}$ be a sequence of input matrices for t = $\{1, 2, 3, ..., d\}$, $d < T$ where $t_g(g = 1, 2, ..., d)$ is some g partition of the original time period T and $t_1 = T$ under various schemes, $s = 1, 2, ..., T$. The optimal scheme is equivalent to the g-group discriminant analysis for that scheme that yields more optimal values than a two-group discriminant analysis statistics. Now let for Q_{it}, let the corresponding sequence of eigenvectors be $Eig(Q_{it})$ and $\lambda_t(Q_{it})$ be the corresponding values. For any time event (t), the eigen equation is given by

$$[Q_{it} - \Lambda_t(Q_{it})]\, Eig(Q_{it}) = 0$$

where Λ_t is $f \times f$ diagonal matrix , $f = min(p, q)$. The time-dependent eigen equations are therefore given as

$$\{\Sigma_{ii}^{-1}\Sigma_{ik}\Sigma_{kk}^{-1}\Sigma_{ki}^{-1}\}_t - \Lambda_t\{\Sigma_{ii}^{-\frac{1}{2}} Eig(P_i)\}_i = 0$$

The solution to the above equation gives the matrix Λ_t. The response variable, **Y**, and the predictor variable, **X**, have an average matrix of overall canonical correlation that is provided by

$$\Lambda = \frac{1}{g}\sum_{t=1}^{g}\Lambda_t$$

Computation of Canonical Coefficient Vectors

Before the canonical coefficient vectors α and β can be computed directly, the matrices $\Sigma_{11}, \Sigma_{22}, \Sigma_{12}$ and Σ_{21} must first be estimated from the given data. The generalized eigenvalue decomposition problem is solved (Chenfeng & Dongrui, 2021; Hardoon et al., 2004) by the equation

$$\begin{bmatrix} 0 & \Sigma_{12} \\ \Sigma_{21} & 0 \end{bmatrix}\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \rho_F \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{bmatrix}\begin{bmatrix} \alpha \\ \beta \end{bmatrix} \tag{3.33}$$

where $\Sigma_{11} = \frac{1}{N}\mathbf{YY'} + r_y\mathbf{I}$, $\Sigma_{22} = \frac{1}{N}\mathbf{XX'} + r_x\mathbf{I}$, and $\Sigma_{12} = \frac{1}{N}\mathbf{YX'}$. It follows that r_y and r_x are positive definite regularization coefficients of **Y** and **X**, respectively, (Bickel & Levina, 2008). The covariance matrix $\mathbf{YY'}$ or $\mathbf{XX'}$ is singular when the feature dimensionality is high and hence the optimization problem is under-determined (Chenfeng & Dongrui, 2021).

The second approach for computing the canonical coefficient vectors α

and β directly performs singular value decomposition (SVD) on matrix $\mathbf{A}$. Suppose ω_y and ω_x are the k^{th} leading left and right singular vectors of $\mathbf{A}$, then the canonical matrices are given by $\alpha = \Sigma_{11}^{-\frac{1}{2}}\omega_y$ and $\beta = \Sigma_{22}^{-\frac{1}{2}}\omega_x$ and hence the correlation $\rho\left(\alpha_k'\mathbf{Y}, \beta_k'\mathbf{X}\right)$ is equal to the k^{th} leading singular value of $\mathbf{A}$ (Chenfeng & Dongrui, 2021). Once α and β are found, the new projected features, the canonical variables, can also be computed by using $U = \alpha'\mathbf{Y}$ and $V = \beta'\mathbf{X}$. Once this is done, the singularity issue can be resolved by applying regularization to the covariance matrices.

Linear Discriminant Analysis

Linear discriminant analysis (LDA) is Bayes optimum when the class distributions are assumed to be identically distributed Gaussian (Ding & Li, 2007; Du & Swamy, 2014). Similar to principal component analysis (PCA), LDA is frequently used for pattern recognition, information retrieval, face recognition, and image retrieval. The assumption is that there are $(\mathbf{y}_i, \mathbf{x}_i) \in \Re^g \times (0,1)^g$, $i = 1, 2, ..., n$ observation-label pairings. Assume that $C_j, j = 1, 2, ..., g$ is the set of points $\mathbf{y}_i$ that are members of the class j (Duda et al., 2000; Fisher, 1936; Samarov, 2009). Let the observations $\mathbf{y}_i$ be a collection of climate heating variables (maximum temperature, minimum temperature, and solar radiation) and $\mathbf{x}_i$ be labels for cooling variables (precipitation, wind, and relative humidity).

Assume that $\mathbf{Y} \in \Re^{n\times p}$ is a matrix whose rows correspond to the observations $\mathbf{y}_i$. Declare $\mathbf{X} \in \Re^{n\times q}$ to be the label matrix, where the indicator function is $\mathbf{I}$ and the definition of the ij^{th} element is $x_{ij} = \mathbf{I}(y_j) \in C_j$. One approach to think about LDA is to find a vector of weights, $\mathbf{v}_y$, for each column of $\mathbf{Y}$ such that the linear combination of $\mathbf{Y}\mathbf{v}_y$ maximizes the ratio of within-class variance to between-class variance. This is what Equations (3.36) and (3.37) state (Duda et al., 2000). I presume that $\mathbf{Y}$ has been mean-centered to make notation easier. The number of observations in class j and the cardinality of C_j are indicated by the expressions $|C_j|$ and n_j, respectively. Considering that $\mathbf{m}_j$ is the average of the observations y_i for class j, then

$$\mathbf{m}_j = \frac{1}{n_j} \sum_{i:\mathbf{x}_i \in C_j}^{n_i} (\mathbf{y}_i) \tag{3.34}$$

The sum of squares total is defined by Duda et al. (2000) as

$$\mathbf{S}_T = \sum_{i=1}^{g} \sum_{j:\mathbf{x}_j \in C_i}^{n_i} \left(\mathbf{y}_j \mathbf{y}_j'\right) = (n-1)\,\mathbf{S}_{yy} \tag{3.35}$$

where the sample covariance matrix of $\mathbf{Y}$ is denoted by $\mathbf{S}_{yy}$. The sum of squares within class, $(\mathbf{S}_W)$ and between class, $(\mathbf{S}_B)$ can be added to obtain the sum of squares total, $(\mathbf{S}_T)$ (Duda et al., 2000), such that

$$\mathbf{S}_W = \sum_{i=1}^{g} \sum_{j:\mathbf{x}_j \in C_j}^{n_i} \left(\mathbf{y}_j - \mathbf{m}_i\right)\left(\mathbf{y}_j - \mathbf{m}_i\right)' \tag{3.36}$$

$$\mathbf{S}_B = \sum_{i=1}^{g} \left(n_i \mathbf{m}_i \mathbf{m}_i'\right) \tag{3.37}$$

and hence

$$\mathbf{S}_T = \mathbf{S}_B + \mathbf{S}_W \tag{3.38}$$

The LDA optimization problem (Duda et al., 2000; Samarov, 2009) is then given as

$$\mathbf{v}_y^* = arg \max_{y} \left(\mathbf{v}_y' \mathbf{S}_B \mathbf{v}_y\right) \tag{3.39}$$

Subject to the constraint: $\mathbf{v}_y' \mathbf{S}_W \mathbf{v}_y = 1$. Supposing δ is a Lagrange multiplier, then the modified Lagrangian is given by

$$L\left(\mathbf{v}_y, \delta\right) = \left(\mathbf{v}_y' \mathbf{S}_B \mathbf{v}_y\right) - \delta\left(\mathbf{v}_y' \mathbf{S}_W \mathbf{v}_y - 1\right) \tag{3.40}$$

The following equation can be obtained by taking the partial derivative with respect to $\mathbf{v}_y$ on both sides of Equation (3.40) and setting the resultant equation to zero.

$$\frac{\partial}{\partial \mathbf{v}_y} L(\mathbf{v}_y, \delta) = \frac{\partial}{\partial \mathbf{v}_y}\left[\left(\mathbf{v}_y' \mathbf{S}_B \mathbf{v}_y\right) - \delta\left(\mathbf{v}_y' \mathbf{S}_W \mathbf{v}_y - 1\right)\right] = 0$$

$$\Rightarrow \mathbf{S}_B \mathbf{v}_y - \delta \mathbf{S}_W \mathbf{v}_y = 0 \qquad \Rightarrow \mathbf{S}_B \mathbf{v}_y = \delta \mathbf{S}_W \mathbf{v}_y$$

It follows that

$$\left[\sum_{i=1}^{g}\left(n_i \mathbf{m}_i \mathbf{m}_i'\right)\right] \mathbf{v}_y = \delta\left[\sum_{i=1}^{g} \sum_{j:\mathbf{x}_j \in C_j}^{n_i}\left(\mathbf{y}_j - \mathbf{m}_i\right)\left(\mathbf{y}_j - \mathbf{m}_i\right)'\right] \mathbf{v}_y \tag{3.41}$$

which gives the generalized eigenvalue issue in LDA. The projected points are then applied to the resulting eigenvectors, $\mathbf{v}_y$, resulting in $\mathbf{y}_i^* = \mathbf{y}_i' \mathbf{v}_y$. A class is chosen for an observation $\mathbf{y}_i^*$ depending on which class center $\mathbf{m}_j = \mathbf{m}_j' \mathbf{v}_y$, where $j = 1, 2, ..., g$ is closest, given by

$$arg \min_j \left|\left|\mathbf{y}_i^* - \mathbf{m}_j^*\right|\right|^2$$

Two-group discriminant analysis

The total number of categories that the response variable has is what defines discriminant analysis (DA). In this scenario, Two-group DA is utilized when the response variable has two categories. Let a linear combination of X_j (Glahn, 1967) be given by

$$Z = \sum_{j=1}^{g} \lambda_j X_j \tag{3.42}$$

Let $\bar{z}_1$ be the mean of the Z values for n_1 observations in group one (1) and $\bar{z}_2$ be the mean of Z values for n_2 observations in group two (2). Now

$\sum_{j=1}^{n_1}\left(z_{1j} - \bar{z}_1\right)^2$ is a measure of variation in group one (1).

$\sum_{j}^{n_2}\left(z_{2j} - \bar{z}_2\right)^2$ is a measure of variation in group two (2).

$\sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(z_{ij} - \bar{z}_i\right)^2$ is a measure of variation in Z within the two groups of variables.

Let $(\bar{z}_1 - \bar{z}_2)^2$ be a measure of separation of values of Z between the two groups. Then Z is found such that λ maximizes (Glahn, 1967; Hardle & Simar, 2007) the expression by

$$\mathbf{G} = \frac{(\bar{z}_1 - \bar{z}_2)^2}{\sum_{i=1}^{2} \sum_{j=1}^{n_i} (z_{ij} - \bar{z}_i)^2} \tag{3.43}$$

where

$$\bar{z}_1 - \bar{z}_2 = \sum_{s=1}^{q} \lambda_s \bar{X}_{s1} - \sum_{s=1}^{q} \lambda_s \bar{X}_{s2} = \sum_{s=1}^{q} \lambda_s \left(\bar{X}_{s1} - \bar{X}_{s2}\right) = \lambda^T \left(\bar{X}_1 - \bar{X}_2\right)$$

$$(\bar{z}_1 - \bar{z}_2)^2 = \lambda' \left(\bar{X}_1 - \bar{X}_2\right) \left(\bar{X}_1 - \bar{X}_2\right)' \lambda = \lambda' \mathbf{B} \lambda = \beta \tag{3.44}$$

Let $\bar{X}_{p1} - \bar{X}_{p2} = d_p$, the difference in mean measurements on the p^{th} variable in the first and second groups (Hardle & Simar, 2007). It follows that

$$\left(\bar{Z}_1 - \bar{Z}_2\right)^2 = (\lambda_1 d_1 + \lambda_2 d_2 + ... + \lambda_r d_r + ... + \lambda_g d_g)^2 = \left(\sum_{p=1}^{g} \lambda_p d_p\right)^2$$

$$\begin{aligned} \Rightarrow \left(\bar{Z}_1 - \bar{Z}_2\right)^2 &= \sum_{p=1}^{g} \Sigma_{q=1}^{g} \lambda_p \lambda_q d_p d_q \\ &= \sum_{p=1}^{g} \sum_{q=1}^{g} \lambda_p \lambda_q \left(\bar{X}_{p1} - \bar{X}_{p2}\right) \left(\bar{X}_{q1} - \bar{X}_{q2}\right) \\ &= \lambda' \mathbf{B} \lambda \end{aligned}$$

The sequence in the denominator lead to the cross-product given by

$$S_{pq} = \sum_{i=1}^{2} \sum_{j=1}^{n_i} \left(X_{pij} - \bar{X}_{pi}\right) \left(X_{qij} - \bar{X}_{qi}\right) \tag{3.45}$$

Now

$$\begin{aligned}\sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(Z_{ij}-\bar{Z}_i\right)^2 &= \sum_{i=1}^{2}\sum_{j=1}^{n_i}\left[\lambda_1(X_{1ij}-\bar{X}_{1i})+\lambda_2(X_{2ij}-\bar{X}_{2i})+...+\lambda_g(X_{gij}-\bar{X}_{gi})\right]^2\\ &= \sum_{i=1}^{2}\sum_{j=1}^{n_i}\sum_{p=1}^{g}\sum_{q=1}^{g}\left[\lambda_p\lambda_q(X_{pij}-\bar{X}_{pi})(X_{qij}-\bar{X}_{qi})\right]\\ &= \sum_{p=1}^{g}\sum_{q=1}^{g}(\lambda_p\lambda_q)\sum_{i=1}^{2}\sum_{j=1}^{n_i}(X_{pij}-\bar{X}_{pi})(X_{qij}-\bar{X}_{qi})\\ &= \sum_{p=1}^{g}\sum_{q=1}^{g}(\lambda_p\lambda_q)\,S_{pq}\end{aligned}$$

$$\Rightarrow \sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(Z_{ij}-\bar{Z}_i\right)^2 = \lambda'\mathbf{W}\lambda$$

where S_{pq} is the Within-group Sum of Squares Cross-Product (Chang-Ha, 2011; Sun, Chen, Yang, Hu, & Shi, 2009).

Similarly,

$$\sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(z_{ij}-\bar{z}_i\right)^2 = \lambda'\mathbf{W}\lambda = \omega \tag{3.46}$$

where the total number of group covariance matrix is indicated by **G**, the within-class covariance matrix is represented by **W**, and the between-class covariance matrix is represented by **B**. The coefficients in Equation (3.42) are determined to maximize (Hardle & Simar, 2007) the ratio

$$\mathbf{G} = \frac{\lambda'\mathbf{B}\lambda}{\lambda'\mathbf{W}\lambda} = \frac{SSCP_{\mathbf{B}}}{SSCP_{\mathbf{W}}} = \frac{\beta}{\omega}$$

From $\frac{\partial \mathbf{G}}{\partial \lambda_r} = 0$; for all $r = 1, 2, ..., q$, and hence,

$$\frac{\partial \mathbf{G}}{\partial \lambda_r} = \frac{\omega\frac{\partial \beta}{\partial \lambda_r} - \beta\frac{\partial \omega}{\partial \lambda_r}}{\omega^2} = 0$$

$$\Rightarrow \frac{\partial \omega}{\partial \lambda_r} = \frac{\omega}{\beta}\frac{\partial \beta}{\partial \lambda_r} = \frac{1}{\mathbf{G}}\frac{\partial \beta}{\partial \lambda_r}$$

Now, if $d_r = \bar{X}_{r1} - \bar{X}_{r2} = X_{rij} - \bar{X}_{ri}$.

$$\frac{\partial \beta}{\partial \lambda_r} = 2\lambda^{/}\mathbf{B}d_r = 2\left(\lambda_1 d_1 + \lambda_2 d_2 + ... + \lambda_r d_r + ... + \lambda_q d_q\right) d_r.$$

Also if $C = \sum_{i=1}^{q} \lambda_i d_i$ is a constant, we get the following equation: $\frac{\partial \beta}{\partial \lambda_r} = 2Cd_r$. C eventually cancels out in **G**. Thus, any multiple of a set of λ that satisfies Equation (3.42) maximizes **G**. Therefore, there is no unique set of λ's maximizing **G**.

Similarly,

$$\frac{\partial \omega}{\partial \lambda_r} = 2\left(\lambda_1 S_{r1} + \lambda_2 S_{r2} + ... + \lambda_r S_{rr} + ... + \lambda_q S_{rq}\right);$$

The coefficients in $\lambda = (\lambda_1, \lambda_2, ..., \lambda_k)$ may be determined from Equation (3.42) in either of two ways. Now let us consider the following equation (Chang-Ha, 2011; Chu & Watterson, 1993; Sun et al., 2009).

$$S_{rt} = \sum_{j}^{ni}(X_{rj} - \bar{X}_r)(X_{tj} - \bar{X}_t) = \sum_{j=1}^{n_i} X_{rj}X_{tj} - n_i\bar{X}_r\bar{X}_t \tag{3.47}$$

where $S_{rt} = \sum_{j}^{n_i}\left(X_{rj} - \bar{X}_r\right)\left(X_{tj} - \bar{X}_t\right)$ is the sum of squares cross-product in X_r and X_t; where $r, t = 1, 2, ..., q$. Noting that **W** is of the form

$$\mathbf{W} = \lambda'\Sigma_{\omega}\lambda = \sum_{p=1}^{g}\sum_{q=1}^{g} \lambda_p\lambda_q S_{pq} \tag{3.48}$$

Hence, $\frac{\partial \omega}{\partial \lambda_r} = 2\left(\lambda_1 S_{r1} + \lambda_2 S_{r2} + ... + \lambda_r S_{rr} + ... + \lambda_p S_{rp} + ... + \lambda_g S_{rg}\right)$

Similarly,

$$\mathbf{B} = \sum_{p=1}^{g}\sum_{q=1}^{g} \lambda_p\lambda_q d_p d_q \tag{3.49}$$

Multiple-group discriminant analysis

There are many situations where discriminating between more than two categories may be of importance. That is, a goal of Multiple-group DA is to determine how few discriminatory functions are necessary to account for the majority of group discrimination. There are g groups with same set of variables, $\mathbf{X}_i = (X_1, X_2, ..., X_g)$, $i = 1, 2, ..., g$. Data matrix is given by $\mathbf{X} = (X_1, X_2, ..., X_g)'$. Let μ_i be the mean for $\mathbf{X}_i$. For groups $i = 1, 2, ..., g$, we have the following data matrix layout from the work of (Ali, 2019; Simo, Styan, & Jarkko, 2011):

1. If $\mu_1 = \mu_2 = ... = \mu_g$, sample means $\bar{X}_1, \bar{X}_2, ..., \bar{X}_q$ may differ.
2. Denote $\mathbf{X}_{sji}$, where $(s = 1, 2, ..., q; j = 1, 2, ..., n; i = 1, 2, ..., g)$ as the values on $\mathbf{X}_s$, the s^{th} variable for the j^{th} individual (observation) in the i^{th} group.

From a Two-group DA, the total sum of squares cross-product can be stated as

$$\mathbf{T} = \mathbf{X}'\left(\mathbf{I} - \frac{1}{n}11'\right)'\left(\mathbf{I} - \frac{1}{n}11'\right)\mathbf{X} = \mathbf{X}'\mathbf{H}\mathbf{X} \qquad (3.50)$$

where $\mathbf{H} = \left(\mathbf{I} - \frac{1}{n}11'\right)'\left(\mathbf{I} - \frac{1}{n}11'\right)$ is the centering matrix (Hardle & Simar, 2007; Simo et al., 2011). The discriminant scores can be obtained from total SSCP matrix given above as follows: Let $Z_i = (\lambda'\mathbf{X})\, i$, where

$$Z = \begin{bmatrix} \lambda' X_1 \\ \lambda' X_2 \\ \vdots \\ \lambda' X_g \end{bmatrix} = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_g \end{bmatrix} = \begin{bmatrix} X_1\lambda \\ X_2\lambda \\ \vdots \\ X_g\lambda \end{bmatrix} = \mathbf{X}\lambda \qquad (3.51)$$

The SSCP matrix of Z scores is $Z'\mathbf{H}Z$, given by

$$\begin{aligned} Z'\mathbf{H}Z &= \lambda'\mathbf{X}'\mathbf{H}\mathbf{X}\lambda = \lambda'\mathbf{T}\lambda \\ &= \lambda'\left(\mathbf{W} + \mathbf{B}\right)\lambda \\ &= \lambda'\mathbf{W}\lambda + \lambda'\mathbf{B}\lambda \\ &= W_Z + B_Z \end{aligned}$$

where $W_Z = \sum_{i=1}^{g} Z_i'\mathbf{H}Z_i$. By extending to the multiple case, where $\bar{Z}$ is the vector of grand means, we have

$$\begin{aligned} B_Z &= \sum_{i=1}^{g} n_i \left(\bar{Z}_i - \bar{Z}\right)^2 \\ &= \sum_{i=1}^{g} n_i \left(\lambda'\bar{X}_i - \lambda'\bar{X}\right)^2 \\ &= \sum_{i=1}^{g} n_i \left[\lambda'\left(\bar{X}_i - \bar{X}\right)\right]^2 \\ &= \lambda'\mathbf{B}\lambda \end{aligned}$$

From $Z = \mathbf{X}\lambda$, then $Z_i = \mathbf{X}_i\lambda = (\mathbf{X}_i\lambda)\,\mathbf{1}$ and $Z = \begin{bmatrix} X_1\lambda \\ X_2\lambda \\ \vdots \\ X_g\lambda \end{bmatrix}$.

By following the given procedure (Simo et al., 2011) we get

$$Z_i = \begin{bmatrix} X_1\lambda \\ X_2\lambda \\ \vdots \\ X_g\lambda \end{bmatrix} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0_g & 0_g & \cdots & 1_g \end{bmatrix} = (X\lambda)\,\mathbf{I}_g \tag{3.52}$$

Fisher linear discriminant function

Assume that Σ is the variance covariance matrix for the p variables in $\mathbf{X}$. The discriminant function will be represented by the expressions $\delta = \mathbf{v}'\mathbf{X}$, where $\mathbf{v}$ is a $p \times 1$ vector of weights. The resulting discriminant scores' sum of squares, δ, is then stated as

$$\delta'\delta = (\mathbf{X}'\mathbf{v})'\,(\mathbf{X}'\mathbf{v}) = \mathbf{v}'\mathbf{X}\mathbf{X}'\mathbf{v} = \mathbf{v}'\mathbf{S}_T\mathbf{v}$$

But $\mathbf{S}_T = \mathbf{S}_B + \mathbf{S}_W$. It follows that

$$\delta'\delta = \mathbf{v}'\left(\mathbf{S}_B + \mathbf{S}_W\right)\mathbf{v} = \mathbf{v}'\mathbf{S}_B\mathbf{v} + \mathbf{v}'\mathbf{S}_W\mathbf{v}$$

The vector of weights, **v** is estimated, such that, $\mathbf{G}^*$, described in Equation (3.53) is maximal (Fisher, 1936).

$$\mathbf{G}^* = \frac{\mathbf{v}'\mathbf{S}_B\mathbf{v}}{\mathbf{v}'\mathbf{S}_W\mathbf{v}} \tag{3.53}$$

By differentiating $\mathbf{G}^*$ with respect to **v** and then equating to zero, leads to

$$(\mathbf{S}_W^{-1}\mathbf{S}_B - \lambda\mathbf{I})\mathbf{v} = 0$$

Thus, the weight vectors, **v**, are the eigenvectors of $\mathbf{S}_W^{-1}\mathbf{S}_B$. The function, $\delta = \mathbf{x}'\mathbf{v}$, is thus called the Fisher Discriminant Function (Fisher, 1936; Sharma, 1996; Sun, Chen, Yang, & Shi, 2008). Only the ratio of the weights is unique; the weights are only unique relative to one another.

Statistical significance of multiple-group discriminant analysis

The discriminant functions might not all be numerically significant. The Chi-square (χ^2) value, is calculated to determine the overall statistical significance of all discriminant functions (Sharma, 1996; Sun et al. 2008) using the expression

$$\chi^2 = \left[n - 1 - \left(\frac{p+G}{2}\right)\right]\sum_{k=1}^{K} ln\left(1 + \lambda_k\right) \tag{3.54}$$

where there are n total observations in each set, p number of discriminator variables, G groups, K discriminant functions, and λ_k as the eigenvalue of the k^{th} discriminant function. The aforementioned process is continued until the χ^2 value is no longer significant in the case of K discriminant functions. In order to determine the statistical significance of the r discriminant function, the χ^2

value with $(p-r+1)(G-r)$ degrees of freedom is typically calculated as

$$\chi_r^2 = \left[n - 1 - \left(\frac{p+G}{2}\right)\right] \sum_{k=r}^{K} ln\,(1+\lambda_k) \tag{3.55}$$

Discriminative Canonical Correlation Analysis Extension

Regularization of CCA, Repeated CCA of Kronecker Product Matrix, Kernel CCA, Deep CCA, and Discriminative CCA are some of the CCA extensions that have been proposed to increase the flexibility of CCA for big dimensional data (Chu, Liao, Ng M, & Zhang, 2013; Hardoon & Shawe-Taylor, 2011). I consider the extension of Discriminative CCA in this thesis. Label information is used in Discriminative CCA to train the correlation matrix and integrate similarity across classes and within classes (Kim, Kittler, & Cipolla, 2007; Sun et al., 2009). Depending on whether local scattering is investigated or not, Discriminative CCA deals with the Global or Local Discriminative CCA multivariate data analysis models that are now in use.

Global discriminative canonical correlation analysis

Theoretically, I assume that the pairwise variables are divided into k classes, with n_i occurrences in the i^{th} class. As indicated by the accompanying vectors, let **Y** and **X**, respectively, represent the starting sets of variables in the first and second sets of variables:

$$\mathbf{Y} = (y_{1,1}, y_{1,2}, ..., y_{1,n_1}, .., y_{k,1}, y_{k,2}, ..., y_{k,n_k})$$

$$\mathbf{X} = (x_{1,1}, x_{1,2}, ..., x_{1,n_1}, .., x_{k,1}, x_{k,2}, ..., x_{k,n_k})$$

where $y_{i,p}$ represent the p^{th} data sample in **Y**'s i^{th} class and $x_{i,q}$ represent the q^{th} data sample in **X**'s i^{th} class. Equation (3.56) was developed using the conventional discriminant canonical correlation technique, which increases within-class correlation and decreased between-class similarity by using class label information (Sharma, 1996; Sun et al. 2008; Yang et al., 2021).

$$arg \max_{\alpha,\beta} \rho = \alpha'\mathbf{C}_w\beta - \tau\alpha'\mathbf{C}_b\beta \tag{3.56}$$

subject to the constraints: $\alpha'\mathbf{C}_{11}\alpha = \beta'\mathbf{C}_{22}\beta = 1$, where the coefficient vectors of **Y** and **X** are, respectively, α and β. The weights of the within-class and inter-class correlations are maintained in proportion to each other by the parameters $\tau > 0$. The within-class correlation is denoted by $\mathbf{C}_w$, and the between-class correlation by $\mathbf{C}_b$.

$$\mathbf{C}_w = \sum_{i=1}^{k}\sum_{p=1}^{n_i}\sum_{q=1}^{n_i} y_{i,p}x'_{i,q} = \mathbf{YM}_D\mathbf{X}' \tag{3.57}$$

$$\mathbf{C}_b = \sum_{i=1}^{k}\sum_{j=1}^{k}\sum_{p=1}^{n_i}\sum_{q=1}^{n_i} y_{i,p}x'_{j,q} = -\mathbf{YM}_D\mathbf{X}' \tag{3.58}$$

An all in one square matrix of dimension $n_i \times n_i$ makes up each block of the blocked diagonal matrix, $\mathbf{C}_{12}$. From the literature, Equation (3.56) can be rewritten as

$$arg \max_{\alpha,\beta} \rho = \alpha'\mathbf{YM}_D\mathbf{X}\beta \tag{3.59}$$

subject to the constraints: $\alpha'\mathbf{C}_{11}\alpha = \beta'\mathbf{C}_{22}\beta = 1$. In Equation (3.59), the global discriminative CCA difficulty can be resolved using the generalized eigenvalue decomposition technique. The Lagrangian equation can be modified (Hardoon et al., 2004; Samarov, 2009) from Equation (3.59) as

$$L(\delta_y, \delta_x, \alpha, \beta) = \alpha'\mathbf{C}_{12}\beta - \frac{\delta_y}{2}(\alpha'\mathbf{C}_{11}\alpha - 1) - \frac{\delta_x}{2}(\beta'\mathbf{C}_{22}\beta - 1) \tag{3.60}$$

where $\mathbf{C}_{12} = \mathbf{YM}_D\mathbf{X}$, δ_y and δ_x are the Lagrangian multipliers of **Y** and **X**, respectively. By taking partial derivatives with regard to α and setting the resulting equation to zero gives

$$\mathbf{C}_{12}\beta - \delta_y \mathbf{C}_{11}\alpha = 0 \tag{3.61}$$

Similarly, by taking partial derivatives with regard to β and setting the resulting equation to zero gives

$$\mathbf{C}_{21}\alpha - \delta_x \mathbf{C}_{22}\beta = 0 \tag{3.62}$$

Multiply both sides of Equation (3.61) by α' yields

$$\alpha'\mathbf{C}_{12}\beta - \delta_y \alpha'\mathbf{C}_{11}\alpha = 0 \tag{3.63}$$

Similarly, multiply both sides of Equation (3.62) by β' yields

$$\beta'\mathbf{C}_{21}\alpha - \delta_x \beta'\mathbf{C}_{22}\beta = 0 \tag{3.64}$$

It follows therefore that $\delta_x = \delta_y = \delta_F$. Supposing $\mathbf{C}_{11}$ is positive definite and invertible, then α can be made the subject from Equation (3.61) as

$$\alpha = \frac{\mathbf{C}_{11}^{-1}\mathbf{C}_{12}\beta}{\delta_F} \tag{3.65}$$

Similar assumption for $\mathbf{C}_{22}$ leads to

$$\beta = \frac{\mathbf{C}_{22}^{-1}\mathbf{C}_{21}\alpha}{\delta_F} \tag{3.66}$$

Now substituting β into Equation (3.61) follows that

$$\delta_F^2\alpha = \mathbf{C}_{11}^{-1}\mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{21}\alpha \tag{3.67}$$

Similarly, substituting α into Equation (3.62) yields

$$\delta_F^2\beta = \mathbf{C}_{22}^{-1}\mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{C}_{12}\beta \tag{3.68}$$

Equations (3.61) and (3.62) are the generalized eigenvalue problem for global discriminative CCA. Equations (3.63) and (3.64) may be written in terms of δ_F as

$$\alpha'\mathbf{C}_{12}\beta - \delta_F\alpha'\mathbf{C}_{11}\alpha = 0$$

$$\beta'\mathbf{C}_{21}\alpha - \delta_F\beta'\mathbf{C}_{22}\beta = 0$$

Hence, the generalized eigenvalue decomposition problem for global discriminative CCA is written formally as

$$\begin{pmatrix} 0 & \mathbf{YM}_D\mathbf{X} \\ \mathbf{XM}_D\mathbf{Y} & 0 \end{pmatrix}\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \delta_F \begin{pmatrix} \mathbf{YM}_D\mathbf{Y} & 0 \\ 0 & \mathbf{XM}_D\mathbf{X} \end{pmatrix}\begin{pmatrix} \alpha \\ \beta \end{pmatrix} \tag{3.69}$$

It is observed (Sakar & Kursun, 2017; Wilms & Christophe, 2015) that the model is highly sensitive to outliers and noisy samples because Equations (3.57) and (3.58) generate between-class and within-class correlation matrices that maximize correlation for any pair of samples.

Local discriminative canonical correlation analysis

The class separation characteristic persists even after the local discriminative CCA takes into account the local information of the data points (Peng, Zhang, & Zhang, 2010; Shin & Park, 2011; Yang et al., 2021; Zuobin, Kezhi, & Ng, 2017). Local designs are typically maintained by the k adjacent groupings. For example, (Peng et al., 2010) updated the local within-class covariance matrix, $\mathbf{C}_w^L$ and the local between-class covariance matrix, $\mathbf{C}_b^L$ based on the k nearest neighborhoods as

$$\mathbf{C}_b^L = \sum_{i=1,u_1}^{N}\sum_{j=1,u_2}^{N} \mathbf{y}_i\mathbf{x}_k' + \mathbf{y}_k\mathbf{x}_i' \tag{3.70}$$

$$\mathbf{C}_w^L = \sum_{i=1,t_1}^{N} \sum_{j=1,t_2}^{N} \mathbf{y}_i \mathbf{x}_k' + \mathbf{y}_k \mathbf{x}_i' \tag{3.71}$$

where $t_1 = \mathbf{y}_k \in N^w(\mathbf{y}_i)$, $t_2 = \mathbf{x}_k \in N^w(\mathbf{x}_i)$ and $u_1 = \mathbf{y}_k \in N^b(\mathbf{y}_i)$ $u_2 = \mathbf{x}_k \in N^b(\mathbf{x}_i)$; $N^w(\mathbf{y}_i)$ and $N^b(\mathbf{y}_i)$, respectively, are the k nearest neighbors of $\mathbf{y}_i$ between-class and within-class; The k nearest neighbors of $\mathbf{x}_i$ that are between-class and within-class are $N^b(\mathbf{y}_i)$ and $N^w(\mathbf{x}_i)$, respectively. Suppose that $\mathbf{C}_{12}^* = \mathbf{C}_b^L - \tau \mathbf{C}_w^L$, then the local discriminative CCA's optimization problem can be expressed as

$$arg \max_{\alpha,\beta} \rho = \alpha' C_{12}^* \beta \tag{3.72}$$

subject to the constraints: $\alpha' C_{11} \alpha = \beta' C_{22} \beta = 1$.

The normal CCA can be solved in this situation using the SVD problem (Peng et al., 2010; Yang et al., 2021). Local discriminant CCA minimizes within-class correlation and maximizes between-class similarity in terms of local neighbors as opposed to global samples to reduce the effects of outliers as seen in Equations (3.73) and (3.74). Shin & Park (2011) assessed the local distribution by integrating the nearest neighborhood scatter matrix between classes and among classes, without changing the correlation matrices given in the global discriminative CCA.

$$\mathbf{S}_b = \sum_{i=1}^{k} \sum_{p=1}^{n_i} \left[\mathbf{y}_{i,p} - \mu_w^k(\mathbf{y}_{i,p})\right] \left[\mathbf{y}_{i,p} - \mu_w^k(\mathbf{y}_{i,p})\right]' \tag{3.73}$$

$$\mathbf{S}_w = \sum_{i=1}^{k} \sum_{p=1}^{n_i} \left[\mathbf{x}_{i,p} - \mu_w^k(\mathbf{x}_{i,p})\right] \left[\mathbf{x}_{i,p} - \mu_w^k(\mathbf{x}_{i,p})\right]' \tag{3.74}$$

The within-class KNN mean of $\mathbf{y}_{i,p}$ is represented by $\mu_w^k(\mathbf{y}_i, p)$. The local scatterness for $\mathbf{Y}$ and $\mathbf{X}$ in projection subspace can be written as $\alpha' \mathbf{S}_b \alpha$ and $\beta' \mathbf{S}_w \beta$, respectively. The issue becomes optimizing these scatters given that they must be smaller to better classification performance and the Lagrangian equation

(Hardoon et al., 2004; Samarov, 2009) is given as

$$L(\alpha, \beta, \lambda_y, \lambda_x) = \alpha' \mathbf{C}_{12}\beta - \frac{\lambda_y}{2}\left(\alpha' \mathbf{S}_b \alpha - 1\right) - \frac{\lambda_x}{2}\left(\beta' \mathbf{S}_w \beta - 1\right) \tag{3.75}$$

where λ_y and λ_x are the Lagrangian multipliers of $\mathbf{Y}$ and $\mathbf{X}$, respectively.

By taking partial derivatives with regard to α and β and setting the resulting equations to zero leads to

$$\mathbf{C}_{12}\beta - \lambda_y \mathbf{S}_b \alpha = 0 \tag{3.76}$$

$$\text{and} \qquad \mathbf{C}_{21}\alpha - \lambda_x \mathbf{S}_w \beta = 0 \tag{3.77}$$

Subsequently, $\lambda_x = \lambda_y = \lambda$. Supposing $\mathbf{S}_b$ is positive definite and invertible, then

$$\alpha = \frac{\mathbf{S}_b^{-1}\mathbf{C}_{12}\beta}{\lambda}$$

Similarly,

$$\beta = \frac{\mathbf{S}_w^{-1}\mathbf{C}_{21}\alpha}{\lambda}$$

Making substitutions gives

$$\alpha = \frac{\mathbf{S}_b^{-1}\mathbf{C}_{12}}{\lambda}\left(\frac{\mathbf{S}_w^{-1}\mathbf{C}_{21}\alpha}{\lambda}\right)$$

$$\lambda\lambda\alpha = \left(\mathbf{S}_b^{-1}\mathbf{C}_{12}\right)\left(\mathbf{S}_w^{-1}\mathbf{C}_{21}\alpha\right)$$

$$\lambda^2\alpha = \mathbf{S}_b^{-1}\mathbf{C}_{12}\mathbf{S}_w^{-1}\mathbf{C}_{21}\alpha \tag{3.78}$$

Similarly,

$$\beta = \frac{\mathbf{S}_w^{-1}\mathbf{C}_{21}}{\lambda}\left(\frac{\mathbf{S}_b^{-1}\mathbf{C}_{12}\beta}{\lambda}\right)$$

$$\lambda\lambda\beta = \left(\mathbf{S}_w^{-1}\mathbf{C}_{21}\right)\left(\mathbf{S}_b^{-1}\mathbf{C}_{12}\beta\right)$$

$$\lambda^2\beta = \mathbf{S}_w^{-1}\mathbf{C}_{21}\mathbf{S}_b^{-1}\mathbf{C}_{12}\beta \tag{3.79}$$

Equations (3.78) and (3.79) are the generalized eigenvalue problem for local discriminative CCA, which may be written as

$$\begin{bmatrix} 0 & \mathbf{C}_{12} \\ \mathbf{C}_{21} & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \lambda \begin{bmatrix} \mathbf{S}_b & 0 \\ 0 & \mathbf{S}_w \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \tag{3.80}$$

By evaluating the scatter matrix, which measures the dispersion of data points that deviate from the local mean relative to the global mean, the model is better able to adapt to the intricate class distribution. To evaluate the local structure, (Zuobin et al., 2017) introduced the Intra-class and Extra-class Discriminative Correlation Analysis. Because the scatter matrix, $\mathbf{S}_w^1$, predicts the scatter of data points' deviations from the local mean instead of the global mean, the model is more robust to the complex class distribution. The between-class scatter matrix, $\mathbf{S}_b^1$, exhibits the following properties.

$$\mathbf{S}_b^1 = \sum_{i=1}^{k}\sum_{p=1}^{n_i}\left[\mathbf{y}_{i,p} - \mu_e^k(\mathbf{y}_{i,p})\right]\left[\mathbf{y}_{i,p} - \mu_e^k(\mathbf{y}_{i,p})\right]' \tag{3.81}$$

where $\mu_e^k(\mathbf{y}_{i,p})$, a weight matrix emphasizing boundary information, is the extra-class k closest neighbors mean of $\mathbf{y}_{i,p}$. By using the between-class scatter matrix, preprocessing rather than modeling retains more of the categorization structure of the input feature. The drawbacks of the global and local discriminative CCA (Zuobin et al., 2017) are as follows: 1. Because the correlation matrix $\mathbf{C}_w$ is of the typical order of (k-1), the maximum number of features that may be extracted is (k-1). Information loss could result from this, especially if the class size is tiny. 2. The computational and spatial complexity of the k nearest neighborhoods approach is considerable. 3. The application of k nearest neighborhoods would be restricted if the sample distribution across various groups is unbalanced. We shall show that, by modulo a scalar, the generalized eigen problem in all of the aforementioned eigen issues is nearly the same (Samarov, 2009; Zuobin et al., 2017).

Link Between Canonical Correlation Analysis and DA

It is possible to construct a multiple-group discriminant analysis as a canonical correlation problem with group membership as the dependent variable coded using dummy variables (Duda et al., 2000; Samarov, 2009; Zuobin et al., 2017). Let the observations $\mathbf{y}_i$ be a collection of heating variables, $\mathbf{x}_i$ be labels for cooling variables, C_i is the collection of points, n is the number of observations-label pairs, and n_i is the number of observations in class i, where $i = 1, 2, ..., g$ (Samarov, 2009) then, average of the observations for class i is

$$\mathbf{m}_i = \frac{1}{n_i} \sum_{i:\mathbf{x}_i \in C_i} (\mathbf{y}_i) \tag{3.82}$$

If I define $\mathbf{y}_i = \mathbf{I}_{(\mathbf{x_i} \in \mathbf{c_i})}$ as the label matrix with i^{th} entry being defined as the indicator function, the class labels matrix, $\mathbf{Y} \in \Re^{n \times g}$ and $\mathbf{I}$ as the indicator function, I have

$$\mathbf{Y} = \begin{bmatrix} \mathbf{1}n_1 & 0 & \cdots & 0 \\ 0 & \mathbf{1}n_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{1}n_g \end{bmatrix}$$

It is clear from this that if $\mathbf{m}_i$ is the average of the observations for class i

$$\mathbf{S}_{YX} = \mathbf{Y'X} = \begin{bmatrix} n_1 m_1' \\ n_2 m_2' \\ \vdots \\ n_g m_g' \end{bmatrix}$$

It follows that

$$\mathbf{S}_{YY}^{-1} = (\mathbf{Y}'\mathbf{Y})^{-1} = \begin{bmatrix} \frac{1}{n_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{n_g} \end{bmatrix}$$

Thus,

$$\mathbf{S}_{XY}\mathbf{S}_{YY}^{-1}\mathbf{S}_{YX} = \sum_{i=1}^{g} n_i m_i m_i' = \mathbf{S}_B \tag{3.83}$$

I begin with the sample canonical correlation matrix equation given in Equation (3.84), where ρ_H is the canonical correlation coefficient and α is the coefficient vector.

$$\mathbf{S}_{XY}\mathbf{S}_{YY}^{-1}\mathbf{S}_{YX}\alpha = \rho_H^2(n-1)\mathbf{S}_{YY}\alpha \tag{3.84}$$

From Equation (3.35), I have

$$\mathbf{S}_T = \sum_{i=1}^{g} \sum_{i:\mathbf{x}_i \in C_j}^{n_j} (\mathbf{y}_i\mathbf{y}_i') = (n-1)\,\mathbf{S}_{yy}$$

Substituting Equations (3.35) and (3.83) into Equation (3.84) yields

$$\mathbf{S}_B\alpha = \rho_H^2\mathbf{S}_T\alpha \tag{3.85}$$

Since $\mathbf{S}_T = \mathbf{S}_B + \mathbf{S}_W$ (Samarov, 2009; Zuobin et al., 2017), then

$$\mathbf{S}_B\alpha = \rho_H^2\left(\mathbf{S}_B + \mathbf{S}_W\right)\alpha$$

$$\mathbf{S}_B\alpha = \rho_H^2\mathbf{S}_B\alpha + \rho_H^2\mathbf{S}_W\alpha$$

$$\mathbf{S}_B\alpha = \left(\frac{\rho_H^2}{1-\rho_H^2}\right)\mathbf{S}_W\alpha \tag{3.86}$$

Hence, this link will be relevant by first determining the optimal groupings in the data for which discrimination is maximum.

Procedure to generalize the link between CCA and DA

I first go over the basic CCA problem in this section in order to create an algorithm that determines the optimum grouping schemes for every feasible Multiple Discriminant Analysis (MDA) that produces the highest CCA. Assuming that **Y** and **X** are the observations made at the same time and that both are column-centered, with $\mathbf{Y} = [y_1, y_2, ..., y_N] \in \Re^{N\times d_y}$, $\mathbf{X} = [x_1, x_2, ..., x_N] \in \Re^{N\times d_x}$, $\alpha \in \Re^{d_y\times l}$, and $\beta \in \Re^{d_x\times l}$. The number of observations is N, the number of canonical vector pairs I attempt to compute is l, and the feature dimensions for **Y** and **X**, are d_y and d_x, respectively. Let $r = rank(\mathbf{Y})$, $s = rank(\mathbf{X})$ and $t = min(r, s)$ (Abdeldjalil & Seghouane, 2016; Chu et al., 2013; Xuefei et al., 2019). The typical CCA problem can then be examined as

$$\min_{\alpha,\beta} ||\mathbf{Y}\alpha - \mathbf{X}\beta||\,|_F$$

subject to the constraints: $\alpha'\mathbf{Y}'\mathbf{Y}\alpha = 1$ and $\beta'\mathbf{X}'\mathbf{X}\beta = 1$.

Theorem 3.3 explains how to solve the above definition using the Singular Value Decomposition (SVD) technique. Let me examine the SVD for **Y** and **X** as follows:

$$\mathbf{Y} = Q_y\,[\Sigma_y, 0]\,[U_1, U_2]' = Q_y\Sigma_y U_1'$$

$$\mathbf{X} = Q_x\,[\Sigma_x, 0]\,[V_1, V_2]' = Q_x\Sigma_x V_1'$$

where $U_1 \in \Re^{d_y\times r}$, $U_2 \in \Re^{d_y\times(d_y-r)}$, $\Sigma_y \in \Re^{r\times r}$, $Q_y \in \Re^{N\times r}$, $V_1 \in \Re^{d_x\times s}$, $V_2 \in \Re^{d_x\times(d_x-s)}$, $\Sigma_x \in \Re^{s\times s}$, and $Q_x \in \Re^{N\times s}$. I also consider the SVD of $Q_y'Q_x$ as $Q_y'Q_x = P_y\Sigma P_x'$, such that $P_y \in \Re^{r\times r}$, $\Sigma \in \Re^{r\times s}$, and $P_x \in \Re^{s\times s}$. Now let me denote the distinct eigenvalue of $Q_y'Q_x$ as $\lambda_1 > \lambda_2 > ... > \lambda_k > 0$ with multiplicity for these k eigenvalues being $\sigma_1, \sigma_2, ..., \sigma_k$, it follows that $m_q = \sum_{i=1}^{q} \sigma_i$. P_y and P_x can be defined as follows, where $k_y\mathbf{I}d_y$ and $k_x\mathbf{I}d_x$ are the two regularization terms of $\mathbf{Y}$ and $\mathbf{X}$, respectively, with $k_y > 0$ and $k_x > 0$. Thus,

$$P_y = \mathbf{Y}' \left(\mathbf{Y}\mathbf{Y}' + k_y \mathbf{I} d_y\right)^{-1} \mathbf{Y}$$

$$P_x = \mathbf{X}' \left(\mathbf{X}\mathbf{X}' + k_x \mathbf{I} d_x\right)^{-1} \mathbf{X}$$

Theorem 3.3

Suppose $l = \sum_{i=1}^{q} \sigma_i$ for some $1 \leq q \leq k$ (Chu et al., 2013), then (α, β) is the solution of optimization problem if and only if

$$\alpha = U_1 \Sigma_y^{-1} P_y \left(:, 1 : l\right) \mathbf{B} + U_2 F_y$$

$$\beta = V_1 \Sigma_x^{-1} P_x \left(:, 1 : l\right) \mathbf{B} + V_2 F_x$$

where $\mathbf{B} \in \Re^{l \times l}$ is orthogonal vector, $F_y \in \Re^{(d_y - r) \times l}$, and $F_x \in \Re^{(d_x - s) \times l}$ are arbitrary vectors. According to (Xuefei et al., 2019), the sparse canonical correlation analysis can be viewed as fixing the issue if $l = \sum_{i=1}^{q} \sigma_i$ for some $1 \leq q \leq k$ as

$$\min_{\alpha, \beta} ||\alpha| \, |l_1 + ||\beta| \, |l_1$$

subject to the constraints of Theorem 3.3, are equivalently expressed as $U_1' \alpha = \Sigma_y^{-1} P_y \left(:, 1 : l\right)$ and $V_1' \beta = \Sigma_x^{-1} P_x \left(:, 1 : l\right)$, with corresponding optimal values $D_s(\rho_s)$ and $D(\rho)$, respectively. The element-wise l_1 penalty is defined as $||*| \, |l_1$ and $||*| \, |_F$ is the Frobenius norm.

Theorem 3.4

Suppose $l = \sum_{i=1}^{q} \sigma_i$ for some $1 \leq q \leq k$ and let $A = \frac{1}{\sqrt{l}}$, then $AD \leq D_s \leq D$.

Algorithms of CCA and multiple-group discriminant analysis

Let n be the total number of observations, d be the mean of the total number of time-dependent variables, and T be the total number of time events that were recorded. I present a number of computer simulations in this part to show how well the suggested algorithm works. I contrast the performance of the suggested algorithm with current cutting-edge discriminant canonical correlation analysis (DCCA) techniques.

Formulation of Grouping Schemes Discriminant CCA

Grouping scheme (GS) is an approach that is intended in this study to incorporate the time effect into the canonical correlation analysis via discriminant analysis in order to effectively handle the Time-dependent multivariate data (TDMD). The number of possible group-analysis that can be performed is denoted as $G = (2, 3, ..., t-1)$ for data on t years. Thus, $m \in G$. For a particular identified best grouping scheme for an m-group analysis $GS_m(:, j)$, I note that there would be multiple values of m in G. The following are the required formulations of grouping schemes discriminant canonical correlations for Two-group (GS_2) and Three-group (GS_3) discriminant analysis, respectively.

$$\mathbf{GS}_2 = \begin{bmatrix} 1_{y_1} & 1_{y_1} & 1_{y_1} & \cdots & 1_{y_1} \\ 2_{y_2} & 1_{y_2} & 1_{y_2} & \cdots & 1_{y_2} \\ 2_{y_3} & 2_{y_3} & 1_{y_3} & \cdots & 1_{y_3} \\ 2_{y_4} & 2_{y_4} & 2_{y_4} & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 1_{y_{t-1}} \\ 2_{y_t} & 2_{y_t} & 2_{y_t} & \cdots & 2_{y_t} \end{bmatrix} \quad \text{and} \quad \mathbf{GS}_3 = \begin{bmatrix} 1_{y_1} & 1_{y_1} & 1_{y_1} & \cdots & 1_{y_1} \\ 2_{y_2} & 1_{y_2} & 1_{y_2} & \cdots & 1_{y_2} \\ 3_{y_3} & 2_{y_3} & 1_{y_3} & \cdots & 1_{y_3} \\ 3_{y_4} & 3_{y_4} & 2_{y_4} & \ddots & \vdots \\ 3_{y_5} & 3_{y_5} & 3_{y_5} & \ddots & 1_{y_{t-2}} \\ \vdots & \vdots & \vdots & \ddots & 2_{y_{t-1}} \\ 3_{y_t} & 3_{y_t} & 3_{y_t} & \cdots & 3_{y_t} \end{bmatrix}$$

The general formulation of grouping scheme discriminant canonical correlation for Multiple-group discriminant analysis (GS_m) is given as

$$\mathbf{GS}_m = \begin{bmatrix} 1_{y_1} & 1_{y_1} & \dots & 1_{y_1} & \dots & 1_{y_1} & 1_{y_1} \\ 2_{y_2} & 1_{y_2} & \dots & 1_{y_2} & \dots & 1_{y_2} & 1_{y_2} \\ 3_{y_3} & 2_{y_3} & \ddots & \vdots & \dots & \vdots & \vdots \\ \vdots & \vdots & \ddots & 1_{y_j} & \dots & 1_{y_{(t-m)}} & 1_{y_{(t-m+1)}} \\ m_{y_m} & (m-1)_{y_m} & \dots & 2y_{j+1} & \ddots & \vdots & \vdots \\ m_{y_{m+1}} & m_{y_{m+1}} & \dots & \vdots & \ddots & (m-1)_{y_{t-2}} & (m-2)_{y_{t-2}} \\ \vdots & \vdots & \dots & (m-1)_{y_{t-1}} & \cdots & m_{y_{t-1}} & (m-1)_{y_{t-1}} \\ m_{y_t} & m_{y_t} & \dots & m_{y_t} & \dots & m_{y_t} & m_{y_t} \end{bmatrix}$$

where m = 3, 4, ..., t and j = 3, 4, ..., t.

Suppose $GS_m(:,j)$ is the best grouping scheme, where $\mathbf{Z} = (\mathbf{Y}|\mathbf{X})$ is the combined variables of Set 1 and Set 2, then the augmented data that incorporates the scheme is given by

$$\mathbf{Z}_F = [\mathbf{Y}|\mathbf{X} : GS_m(:,j)]$$

It follows that $\mathbf{Z}_F \in \Re^{p+q+m}$ and $GS_m(:,j) \in \Re^m$, where the number of variables in Sets 1 and 2 are denoted by the numbers p and q, respectively.

Hypotheses Testing of Discriminant Canonical Correlation Analysis

In order to arrive at the optimal grouping scheme, we need to analyze some related hypotheses in canonical correlation analysis and discriminant analysis. Equations below offer the null and alternative hypotheses for computing the statistical significance of canonical correlation analysis, where f is the $min(p,q)$. Let ρ_r be the r^{th} canonical correlation, where r = 1, 2,..., f.

$$H_0 : \rho_1 = \rho_2 = \rho_3 = ... = \rho_f = 0$$

$$H_1 : \rho_1 \neq \rho_2 \neq \rho_3 \neq ... \neq \rho_f \neq 0$$

If R_{12} is the correlation matrix describing the correlation between X and Y variables, then the null hypothesis asserts that $R_{12} = 0$. We can test the hypotheses using a variety of test statistics.

Wilks' Lambda text statistic

A statistic for testing the statistical significance of canonical correlation is Wilks' Lambda ($\wedge$) which may range from a value of zero to one. The test statistic based on Wilks' Lambda, $\wedge$ is given by

$$\Rightarrow \wedge = \Pi_{r=1}^{m}(1-\rho_r^2) \tag{3.87}$$

The closer Lambda is to zero the more likely canonical correlation will be statistically significant. The statistical significance of ($\wedge$) or the likelihood ratio (Checko, 1986; Sharma, 1996) is tested by using the test statistic

$$\mathbf{B} = -[n-1-\frac{1}{2}(p+q+1)]ln\wedge \tag{3.88}$$

Hotelling's T-square test

To perform a multivariate test of differences between the mean values of two groups, Hotelling's T^2 statistic is utilized. As per the null hypothesis, **S**, the centroid of the two groups, is identical. Multiple analysis of variance (MANOVA) and multiple analysis of covariance (MANCOVA) both make use of Hotelling's T^2 (Rathbun, Wiesner, Srabashi, Roths & Romer, 2023; Joungyoun, Youngrae, Johan & Sungim, 2023). The Hotelling's T^2 statistic is given by

$$T^2 = n\left(\bar{X}-\mu_0\right)'\mathbf{S}^{-1}\left(\bar{X}-\mu_0\right) \tag{3.89}$$

T^2 is roughly Chi-square, χ^2 distributed with p degrees of freedom when n very large. The population variance-covariance matrix, Σ, is substituted for the sample variance-covariance matrix, **S**, in equation (3.89).

$$\chi^2 = n\left(\bar{X} - \mu_0\right)' \Sigma^{-1} \left(\bar{X} - \mu_0\right) \tag{3.90}$$

When the data are normally distributed, the test is thus exactly Chi-square distributed with p degrees of freedom (Rathbun et al., 2023).

Pillais test statistic

The Pillais trace is the test statistic for a MANOVA. Its value ranges from 0 to 1. When Pillai's trace is close to 1, there is strong evidence that the explanatory variable influences the response variable values in a statistically significant way (Pillais, 1955). Thus,

$$Pillai's\ trace = \sum_{i=1}^{q} \frac{\lambda_i}{1+\lambda_i} \tag{3.91}$$

where λ_i is the i^{th} eigenvalue and q denotes the number of variables, for $i = 1, 2, ..., q$. This test is regarded as the most effective and reliable statistic, particularly for detecting assumptions that are not met. Pillais is the ideal choice when the sample size is small, the homogeneity of variance-covariance assumption of the MANOVA is broken, or your cells have varied sizes. However, Pillais tends to be less effective than the other three when the degree of freedom of the hypothesis exceeds one. Roy's Maximum Root is a far better choice when there is a significant divergence from the null hypothesis or when there are significant discrepancies in the eigenvalues (Pillais, 1955; Seber, 1984).

Roys test statistic

In contrast to some other widely used summary statistics (such as Hotelling's and Pillai's), Roy's Largest Root (Johnstone & Nadler, 2017; Warner, 2013) only reports explained variance for one discriminant function, as stated by

$$Roy's\ Largest\ Root = \lambda_i(1 + \lambda_i) \tag{3.92}$$

where λ_i is the i^{th} eigenvalue. Weighted linear composites of quantitative variables are produced via discriminative functions. Always, the trace of Hotelling is smaller than or equal to the biggest root of Roy. According to (Johnstone & Nadler, 2017; Warner, 2013), if they are equivalent, it could suggest one of the following:

(i) One dependent variable is largely linked to the effect,

(ii) Significant correlation exists between the dependent variables,

(iii) The effect's influence on the model is minimal.

Wilks test statistic

Wilks' test statistic is given in terms of λ by Equation (3.93), where λ_i is the i^{th} eigenvalue of the matrix $\mathbf{P}_1 = \Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}$; q denotes the number of variables (Cankaya et al., 2011).

$$Wilks\ statistic = \sum_{i=1}^{q} \frac{1}{1 + \lambda_i} \tag{3.93}$$

The F test statistic

The F test statistic for ρ_i^2's statistical significance can be found in Equation (3.94) (Cankaya et al., 2011).

$$F = \frac{1 - \lambda_i^{\frac{1}{k}}}{\lambda_i^{\frac{1}{k}}} \frac{f_{b_2}}{f_{b_1}} \approx F_{f_{b_2}, f_{b_1}} \tag{3.94}$$

In this case, $\lambda_i = \Pi_i^n(1 - \rho_i^2)$; $f = min(p, q)$; $f_{b_1} = pq$; $f_{b_2} = ck - \frac{1}{2}pq + 1$; $c = n - \frac{1}{2}(p + q + 3)$; $k = \sqrt{\frac{p^2q^2-4}{p^2+q^2-5}}$;

where the total number of cases is n, p denotes the number of response variables, q denotes the number of predictor variables, and ρ_i^2 denotes the i^{th} eigenvalue of the matrix, $\mathbf{Q}_1 = \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ or the squared canonical correlations (Cankaya et al., 2011).

Chapter Summary

The development of grouping scheme discriminant canonical correlation analysis (GSDCCA) and its application to data with time-dependent structure are the primary goals. The methodology has provided a thorough analysis of the canonical variable construction process. It has in the process discovered about six important matrices for creating canonical variables. The element matrices of the partitioned variance-covariance matrix were used to create, $\Sigma_{ii}^{-\frac{1}{2}}, i = 1, 2,$ and $\mathbf{A} = \Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}$, two of these matrices. In order to create the new variables, the review made it possible to determine the proper interpretation and uses of these matrices. It has been noted that the matrix $\beta = \Sigma_{12}\Sigma_{22}^{-1}$, whose columns represent the regression coefficients of the regression functions without intercepts, is an important matrix.

The connections between these factors have also been discovered. Also looked at are the methods for calculating the connection between CCA and DA functions. There have been two primary fundamental procedures looked at. These are the conditional distribution of Cauchy-Schwarz Inequality technique and the Lagrangian multiplier technique. These two fundamental approaches are thoroughly explained in this chapter. It has been noted that the outcomes of these two strategies are identical and generated similar results so far as the key matrices for CCA are concerned. The general methodology of various extensions of CCA have been so far made to the basic formulations. Those extensions are outlined and also explained into details using the two fundamental approaches given above together with the author and date of first mentioned.

This chapter looked at the canonical correlation variables' theoretical characteristics and described them in three ways. It can be shown that the new variables in each situation should have one of six essential characteristics, which, taken together, substantially cover their independence and unit variance. The partitioned sample variance-covariance matrix, which would produce new canonical correlation variables with amply supported attributes, has been studied for its structure. The methodology of the general formulation of grouping scheme discriminant canonical correlation analysis for Multiple-group discriminant analysis is discussed in this chapter. By incorporating the ideal grouping scheme into canonical correlation analysis, the time-effect of the data is expected to improve the true correlation between the two sets of variables.

CHAPTER FOUR

RESULTS AND DISCUSSION

Introduction

A number of theoretical concepts are presented in Chapter Three that are always important to keep in mind when performing canonical correlation analysis (CCA). This chapter's statistical modeling component of the work makes use of these ideas. Preliminary analyses in the chapter include discriminative statistics, CCA results, statistical significance and subsequent studies of higher inferential analyses including results of grouping scheme discriminant canonical correlation analysis (GSDCCA) that establishes a connection between CCA and DA. The problems developed in Chapters Two and Three are put to use in this chapter, which help to identify the key matrices involved in creating canonical variables using the illustrative dataset described in Chapter One. The goal of the implementation of the codes is to create the canonical variables while simultaneously ensuring that they have the necessary attributes. In this case, the first set of codes is described. The SPSS syntax uses the command for MANOVA and the sub-command ”discrim” in a one-factorial design. The response variables are listed first in the MANOVA command, followed by the predictor variables. I combine all response and predictor variables into a single factor and use the WITH command to separate the two groups. For all covariates, the subcommand, discrim, generates a CCA outputs. Covariates are defined after the word WITH. ALPHA specifies the level of significance required to extract a canonical variable.

Series Plots of the Six Monthly Weather Conditions in Ghana

The time series plots of the monthly weather conditions data in Ghana from January 2000 to December 2021 are shown in Figure 2. In the plots, two hundred and sixty-four (264) observations are provided. The two temperature

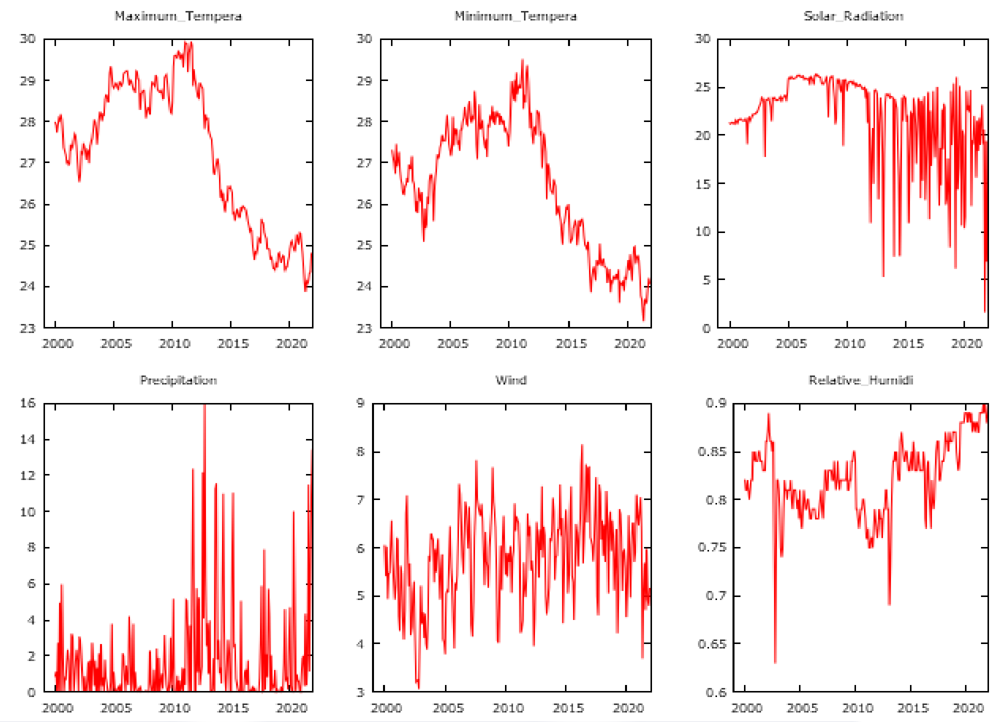

Figure 2: Series Plots of data on six monthly Weather Conditions

variables have similar characteristics with slightly higher variation in the second. In both cases, there is a gentle increasing linear trend with considerable variability up to about the year 2011, after which the pattern assumes a sharp negative linear trend till the end of the series; with variation in the last few portion being quite high. There appears to be a stationary trend for the other four conditions with very low variability particularly in Solar Radiation from the beginning to 2011. There is similarly quiet stable trend in Relative Humidity with large spikes at few time periods. The stable trend in Precipitation and Wind is rather characterized by large variability throughout the period. The general trend in all the six series however suggests evidence of non-stationarity in the mean. These generally appears to be a noticeable change in the behavior of all the series around the year 2011.

Table 2 shows the correlation matrix for all six variables. It can be seen from the table that there are some strong relationships, particularly among

Table 2: **Canonical Correlation Matrix for Set 1 and Set 2 Variables**

Variable	Max T	Min T	Solar Rad	Precip	Wind	Rel Hum
Max Temp	1.0000					
Min Temp	0.5785	1.0000				
Solar Rad	0.5502	0.5316	1.0000			
Precip	-0.0897	-0.0977	-0.1770	1.0000		
Wind	-0.1373	-0.0533	-0.1023	-0.0518	1.0000	
Rel Hum	-0.6890	-0.6425	-0.2805	0.1053	0.1240	1.0000

Source: Researcher's computations (2023)

Maximum Temperature, Minimum Temperature and Solar Radiation. There are also both positive and negative relationships among some variables, but the correlations between those variables are far less than unity. The highest correlations are observed between each of the two temperature variables and Relative Humidity. Generally, there is negative correlation between heat measure variables, **Y** = Maximum Temperature, Minimum Temperature, Solar Radiation and cooling measure variables, **X** = Precipitation, Wind, Relative Humidity.

Extraction of canonical correlation variables

The codes used in this chapter divide the data into the first three columns, which make up the sub-vector $\mathbf{Y} = (Y_1, Y_2, Y_3)$ of responses and the last three columns, which make up the sub-vector $\mathbf{X} = (X_1, X_2, X_3)$ of predictors to demonstrate the extraction of canonical variables from the dataset. The CCA is thus performed on p = 3 response variables and q = 3 predictor variables. In this case three canonical variate pairs or canonical roots are generated since the min(p, q) in both sets is three.

Derivation of Canonical Correlation Functions

Table 3 displays the canonical correlation coefficients (CCC) for every pair of the three fundamental canonical variates. The three degrees of freedom, their significance, and the percentage of variation that is explained are listed in the table. The first canonical roots as well as eigenvalues, Wilks statistics, the number of canonical variate pair has CCC of 0.731 and accounts for 87.27% of the correlation's explained variance. The results show that there is a high statistically significant association between the response variables and the other climatic parameters for all three of the canonical roots likelihood ratio tests.

Table 3: **Relevant CCA Statistic Measures from Original Data**

Root	CCC	ρ_i^2	Eig. V	Wilks L	Pct.	C. Pct	F	Sig.
1	0.731	0.534	1.145	0.398	87.270	87.270	32.112	0.000
2	0.345	0.119	0.135	0.854	10.316	97.586	10.654	0.000
3	0.175	0.031	0.032	0.969	2.414	100.000	8.236	0.004

Source: Researcher's computations (2023)

The high association between the test findings for the predictor variables and the response variables gives credibility to the validity of this study. Wilks Statistic or Lambda is interpreted as the opposite of R-squared statistic. If Wilks Lambda is close to zero, it means that there is high correlation between response variables and predictor weather conditions. There is little association between the response variables and the predictor weather conditions if Wilks Lambda is close to one. The findings indicate that each of the three correlation coefficients is significant.

Canonical correlation coefficient matrix

It is possible to reduce the structure of the canonical correlation coefficient matrix—which is used to verify the properties of the new variables—to that of the generic matrix shown below. The diagonal elements of the canonical correlation coefficient matrix's sub-matrix, $\Lambda = diag(0.731 + 0.345 + 0.175)$, demonstrate the correlation between the pair of variables (U_i, V_i), where i = 1, 2, and 3. All correlation coefficients are positive in this situation. The appropriateness of the components of the partitioned sub-vectors of the weather condition variables is demonstrated by the canonical correlation coefficient matrix.

$$CCC = \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 0.731 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0.345 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0.175 \\ \hline 0.731 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0.345 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0.175 & 0 & 0 & 1 \end{array}\right)$$

Results of canonical loadings

The findings include correlations between the observed variables, the canonical variate, the raw and normalized canonical coefficients, and the proportion of variation for each set that can be accounted for by the canonical root. The outcomes for the two set of variables are shown below. Table 4 gives the canonical loadings for response variables set. To precisely assess how well the initial response variables and the canonical variate pairs are correlated, structural coefficients or canonical loadings are used. The maximum temperature and the minimum temperature are the largest contributors to the first variate in this situation, according to the correlations between the response and canonical variables. Their various contributions to each variate are represented by these weights.

Table 4: **Canonical Loadings for Response Variables**

Variate Number	1	2	3
Maximum Temperature	0.93456	0.29508	-0.19882
Minimum Temperature	0.84927	0.48353	-0.21200
Solar Radiation	0.38873	0.01271	-0.92126

Source: Researcher's computations (2023)

The first canonical variate's key factor is its maximum temperature. The minimum temperature is the primary component of the second canonical variate. Solar radiation is the primary cause of the third canonical variate. The canonical loadings for the group of predictor variables are displayed in Table 5. It follows from the table that as far as the first canonical variate is concerned, relative humidity is the main contributor. The second and third canonical variates are primarily influenced by wind and precipitation, respectively.

Table 5: **Canonical Loadings for Predictor Variables**

Variate Number	1	2	3
Precipitation	-0.05945	-0.14456	0.98771
Wind	-0.33783	0.93899	0.06463
Relative Humidity	-0.97538	-0.21996	0.01566

Source: Researcher's computations (2023)

Dimension reduction analysis

Table 6 displays the Wilks' Statistics or Lambda, F-ratios, degrees of freedom, and significance of each of the three canonical roots. I put the canonical response and predictor variates' hierarchical structure to the test. The first significance test examines each of the three canonical roots of importance, from root 1 to root 3, here F = 32.112 with $p < 0.05$ and Wilks Lambda is 0.398.

Table 6: **Dimension Reduction Analysis - Likelihood Ratio**

Root Nos.	Wilks Lamb	F Ratio	Hypoth. DF	Error DF	Sig. of F
1 TO 3	0.398	32.112	9.000	628.055	0.000
2 TO 3	0.854	10.654	4.000	518.000	0.000
3 TO 3	0.969	8.236	1.000	260.000	0.004

Source: Researcher's computations (2023)

The second test does not include the first root and examines roots two to three which gives F = 10.654 with $p < 0.05$ and Wilks Lambda is 0.854. The final test examines root three alone. All the three roots are found significant.

Results of canonical correlation coefficient vectors

In order to optimize the expected correlations between U_i and V_i, the canonical coefficient vectors α and β are extracted as follows.

$$\alpha = \begin{pmatrix} 1.4934 & -2.3610 & 0.5976 \\ -0.9927 & 2.8886 & -0.3790 \\ -0.0438 & -0.0382 & -0.2625 \end{pmatrix}$$

and

$$\beta = \begin{pmatrix} 0.0108 & -0.0215 & 0.3689 \\ -0.2282 & 1.0214 & 0.1358 \\ -24.0480 & -8.4671 & -2.6866 \end{pmatrix}$$

where the columns represents the canonical variates in each of the sub-vectors of response and predictor variables, respectively.

Results of Canonical Variates and the Canonical Correlations

Up to three pairs of canonical variates can produce the three canonical correlation coefficients in this study since each set contains only three variables.

Table 7: **Raw CCCs for Response and Predictor Variables**

Unst Var	α_{1i}	α_{2i}	α_{3i}	Unst Var	β_{1i}	β_{2i}	β_{3i}
Max Temp	1.493	-2.361	0.598	Precip	0.011	-0.022	0.369
Min Temp	-0.993	2.889	-0.379	Wind	-0.228	1.021	0.136
Solar R	-0.044	-0.038	-0.263	R Hum	-24.048	-8.467	-2.687
$\sqrt{\Sigma\alpha_{ij}^2}$	1.794	3.731	0.755	$\sqrt{\Sigma\beta_{ij}^2}$	24.049	8.529	2.715

Stand Var	α_{1i}^*	α_{2i}^*	α_{3i}^*	Stand Var	β_{1i}^*	β_{2i}^*	β_{3i}^*
Max Temp	0.833	-0.633	0.792	Precip	0.0004	-0.003	0.136
Min Temp	-0.553	0.774	-1.024	Wind	-0.0095	0.120	0.050
Solar R	-0.024	-0.010	-0.093	R Hum	-1.0000	-0.993	-0.990

Source: Researcher's computations (2023)

The response and predictor variables' unstandardized raw canonical coefficients are described in Table 7. The table provides the canonical equations for the first, second, and third sets of canonical variates. The standardized variates are also given in the lower portion of the table. In each case, the standardized variate,

U_i^* has coefficients subject to the condition that

$$Var\left(U_i^*\right) = \alpha_i^{*\prime}\alpha_j = \Sigma\alpha_{ij}^{*2} = \frac{1}{\Sigma\alpha_{ij}^2}\Sigma\alpha_{ij}^{*2} = 1.$$

Given that for i = 1, 2, 3 and j = 1, 2, 3, then the unstandardized canonical variates are given as follows:

$$U_i = \sum_{j=1}^{3}\alpha_{ij}y_j \qquad and \qquad V_i = \sum_{j=1}^{3}\beta_{ij}x_j$$

The variates may be standardized by the normalization conditions as follows:

$$U_i^* = \frac{1}{\sqrt{\sum_{j=1}^{3}\alpha_{ij}^2}}\sum_{j=1}^{3}\alpha_{ij}y_j \qquad and \qquad V_i^* = \frac{1}{\sqrt{\sum_{j=1}^{3}\beta_{ij}^2}}\sum_{j=1}^{3}\beta_{ij}x_j$$

The equations for the first pair of unstandardized canonical variates denoted by U_1 and V_1 are given as

$$U_1 = 1.4934Y_1 - 0.9927Y_2 - 0.0438Y_3$$

$$V_1 = 0.0108X_1 - 0.2282X_2 - 24.0480X_3$$

For example, for U_1^*, we have

$$\sqrt{\Sigma\alpha_j^2} = \sqrt{(1.4934)^2 + (-0.9927)^2 + (-0.0438)^2} = 1.7938$$
$$U_1^* = \frac{1.4934}{1.7938}Y_1 + \frac{-0.9927}{1.7938}Y_2 + \frac{-0.0438}{1.7938}Y_3$$
$$U_1^* = 0.8325Y_1 - 0.5534Y_2 - 0.0244Y_3$$

Subject to: Variance $(U_1^*) = (0.8325)^2 + (-0.5534)^2 + (-0.0244)^2$ = 1, where Y_1 denotes Maximum Temperature, Y_2 represents Minimum Temperature and

Y_3 is Solar Radiation. Similarly, for V_1^* we have

$$\sqrt{\Sigma\beta_j^2} = \sqrt{(0.0108)^2 + (-0.2282)^2 + (-24.048)^2} = 24.0491$$
$$V_1^* = \frac{0.0108}{24.0491}X_1 - \frac{0.2282}{24.0491}X_2 - \frac{24.048}{24.0491}X_3$$
$$V_1^* = 0.0004X_1 - 0.0095X_2 - X_3$$

Subject to: Variance $(V_1^*) = (0.0004)^2 + (-0.0095)^2 + (-1)^2$ = 1, where X_1, X_2 and X_3 denote Precipitation, Wind and Relative Humidity, respectively. The normalized canonical equations for the first set are given by the following equations:

$$U_1^* = 0.8325Y_1 - 0.5534Y_2 - 0.0244Y_3$$

$$V_1^* = 0.0004X_1 - 0.0095X_2 - X_3$$

The standardized pairs (U_2^*, V_2^*) and (U_3^*, V_3^*) are the remaining two pairs of canonical variates and are given in the lower portion of Table 7.

Test of Statistical Significance of CCA Functions

It is vital to establish and validate the statistical significance of the canonical correlation coefficients prior to understanding the canonical variates and their canonical correlations. In this section, the null and alternative hypotheses are presented in order to evaluate the statistical significance of the traditional findings presented in Chapter 3. The test statistic is provided by

$$\mathbf{B}_i = -[n - 1 - \frac{1}{2}(p + q + 1)]ln\wedge_i$$

The values of $\wedge_i$ are given in Table 3. Using these data, Table 8—which contains a section of Table 6—gives the statistic values and a summary of the test findings.

Table 8: **Statistical Significance of all the CCA Functions**

Function	CCC	$\wedge_i$	df	B_i	Significance
1	0.735	0.3978	9	239.2071	0.0000
2	0.345	0.8537	4	41.0529	0.0000
3	0.175	0.9690	1	7.9223	0.0040

Source: Researcher's computations (2023)

The second and third canonical correlation variates can be used to determine their statistical significance once the first canonical variates' influence has been reduced. Within rounding errors, the value of $\wedge_2$ given in Table 8 is the same as the value of Wilk's Lambda given in Table 6. A 4-degree-of- freedom Chi square distribution for $\mathbf{B}_2$ is statistically significant.

The statistical significance of the third canonical correlation can be performed once the effects of the first and second canonical variates have been considered. Now within rounding errors, the value of $\wedge_3$ given in Table 8 is also the same as the value of Wilk's Lambda of the third canonical root given in Table 6. A 1-degree-of-freedom Chi-square distribution for $\mathbf{B}_3$ is also statistically significant.

For example, the statistical significance of the first canonical correlation is calculated using the following formula: The first root's Wilk's Lambda, $\wedge_1$, is given by

$$\begin{aligned}\wedge_1 &= (1-\rho_1^2)(1-\rho_2^2)(1-\rho_3^2)\\ &= (1-0.534)(1-0.119)(1-0.031)\\ &= (0.466)(0.881)(0.969)\\ &= 0.3978\end{aligned}$$

This value is identical, within rounding errors, to the likelihood ratios given in Tables 3 and 6. The corresponding statistic B_1 is given as

$$\mathbf{B}_1 = -[n-1-\frac{1}{2}(p+q+1)]ln\wedge_1$$

$$\Rightarrow \mathbf{B}_1 = -[264 - 1 - \frac{1}{2}(3 + 3 + 1)]ln(0.3978)$$
$$= -(259.5)(-0.9218)$$
$$= 239.2071$$

The 9-degree-of-freedom Chi square distribution for $\mathbf{B}_1$ in this case is statistically significant. The null hypothesis is rejected since there are non-zero canonical correlation values. By rejecting the null hypothesis, it is demonstrated that at least the first canonical correlation is statistically significant.

Total proportion of variance explained

Table 9 compares the overall percentage of variance explained by the response variables and the corresponding predictor variables.

Table 9: **Total Proportion of Variance Explained by the Two Sets**

Can Var	Set 1 by Set 1	Set 1 by Set 2	Set 2 by Set 2	Set 2 by Set 1
1.	0.582	0.311	0.356	0.190
2.	0.107	0.013	0.317	0.038
3.	0.311	0.010	0.327	0.010

Source: Researcher's computations (2023)

Graphical Representation of the CCA Functions

Figures 3 through 5 demonstrate the data from Tables 3, 4, and 5, as well as the correlation between the response and the predictor variates. Figure 3 shows the structural equation of canonical correlation results for the first root. Precipitation is quite near to zero, but all of the dependent variables have positive canonical loadings while the independent variables have negative canonical loadings.

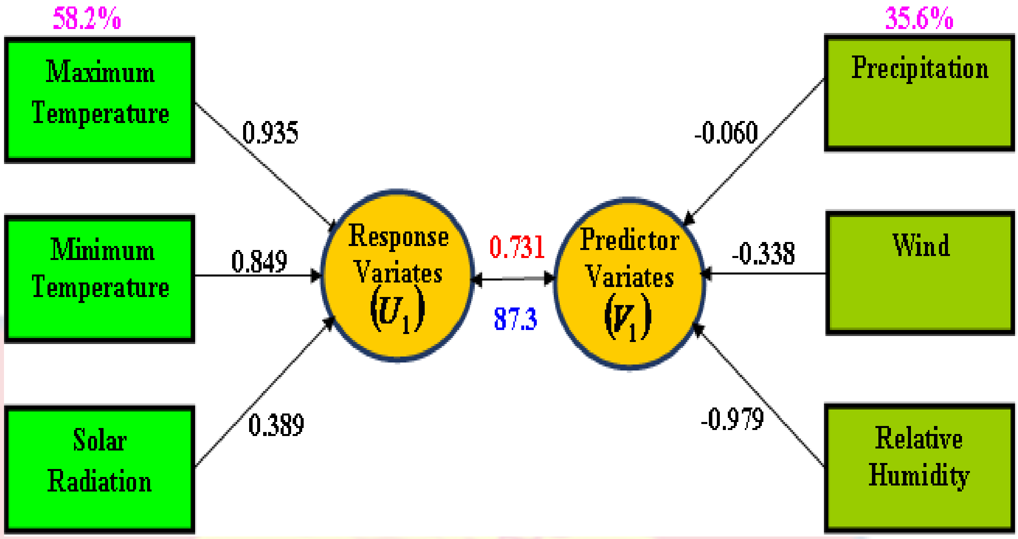

Figure 3: Structural Equation showing the First Canonical Root

Figure 4 depicts the structural equation of canonical correlation results for the second root. The canonical loading of criterion variables are positive and those of independent variables are negative except for Wind, which is also high. Total proportion of variance explained by both sets are 10.7% and 31.7%, which are very small and the variance in percentage explained from Eigenvalues is given as 10.3% which is also small.

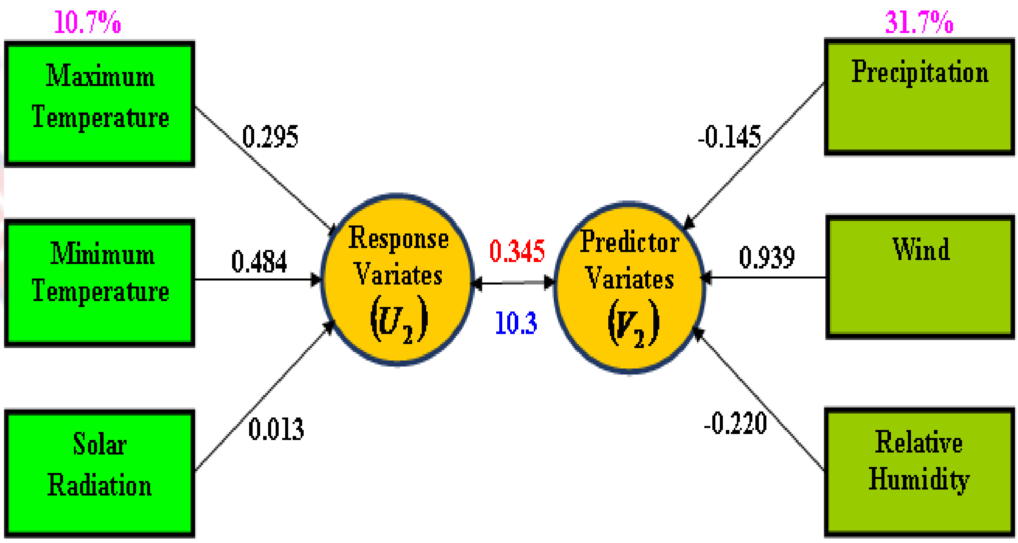

Figure 4: Structural Equation showing the Second Canonical Root

The explanations in Figure 4 demonstrate that this second root cannot sufficiently account for the correlation pattern between the criterion variables and various types of meteorological conditions. This second root indicates a very weak connection between the dependent and independent variables, although being statistically significant. Therefore, this canonical root cannot be used to depict the correlation structure between the response variables and the predictor factors.

The structural equation for the third root's canonical correlation results is shown in Figure 5. In contrast to the first two, the loading of the third root on the responses are all negatives. Similarly, the sign in the predictors have also reversed as into positive. This third root shows an approximate low correlation between the response variables and the predictor factors, even if it is also statistically significant.

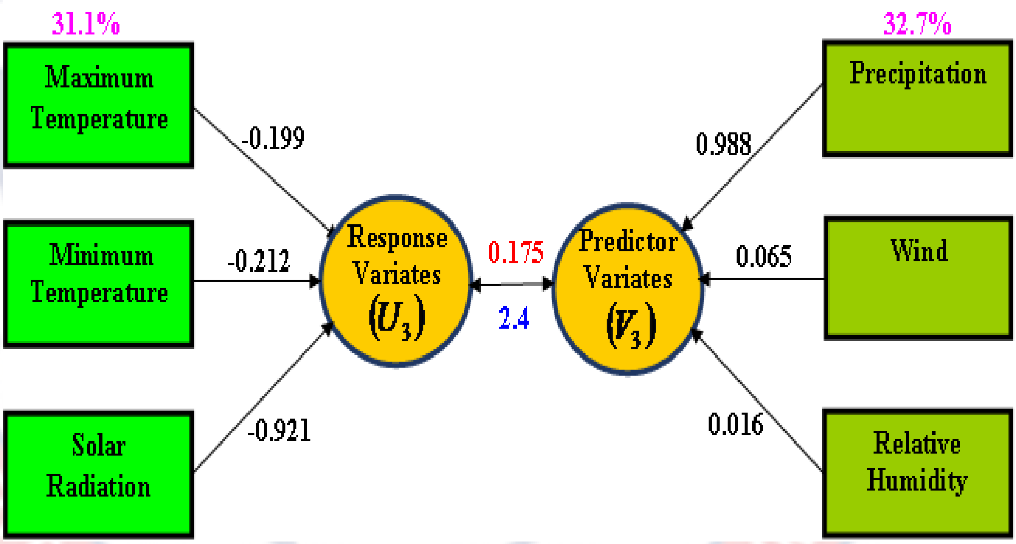

Figure 5: Structural Equation showing the Third Canonical Root

Grouping Scheme Discriminant CCA Results

Table 10 shows the overall canonical correlation coefficient and the other relevant statistics. Thus, the correlation coefficient between the subgroups of heating variables and cooling weather variable subsets is 0.9014 without taking into account discrimination resulting from the effect of the year. Now, I examine the effect of the year by first assuming that the year introduces only Two-group discrimination in the data. Higher statistics values from Two-group discriminant canonical correlation would provide the basis for Multiple-group discriminant canonical correlation to be studied subsequently.

Table 10: **Overall Tests of Significance for Original Data**

Test Statistic	Pillais	Hotellings	Wilks	Roys	Eig V	CCC
Value	0.8127	4.3388	0.1873	0.8127	4.3388	0.9014
Significance	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Source: Researcher's computations (2023)

Two-group discriminant canonical correlation analysis results

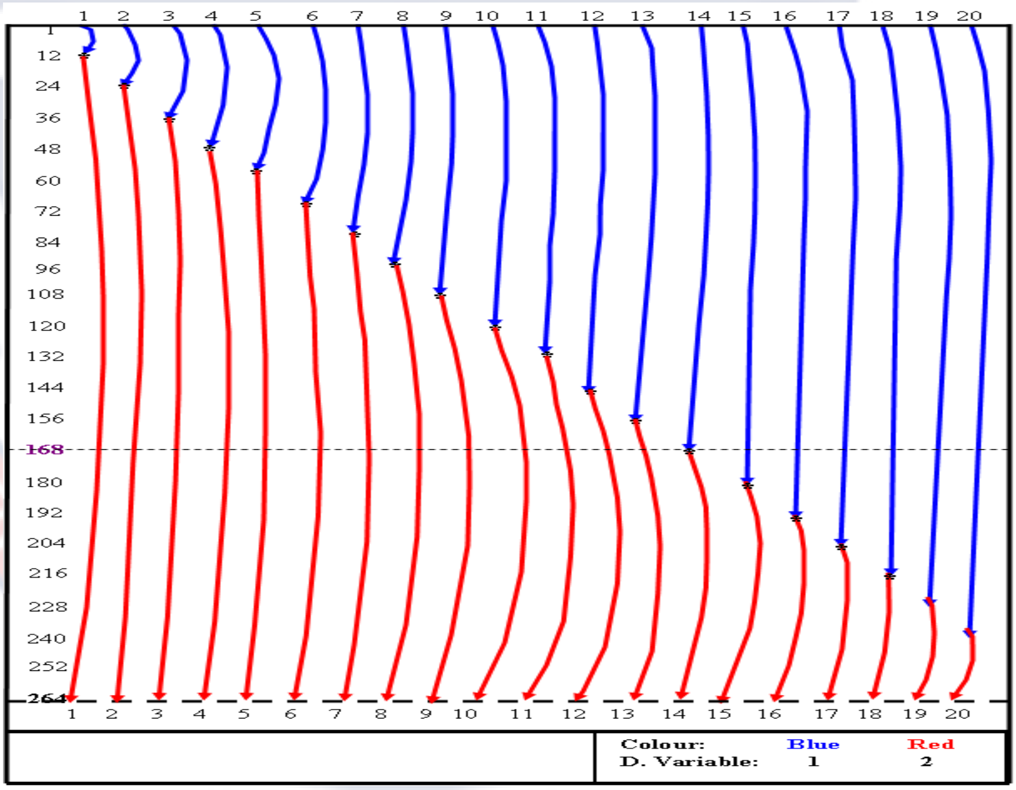

Figure 6: Grouping Scheme Pictorial map for 2-group DCCA Functions

The Two-group DCCA is designed in the GS pictorial map given in Figure 6, where blue colour denotes dummy variable one and red colour denotes dummy variable two. I generate the GSs pictorial map and use it to run several Two-group DCCA from GS one through GS twenty (r - 1), where r is the total number of years to check which one has the highest DCC, the highest Eigen-value, the lowest Wilks Lambda and the highest Chi-square value as shown in Table 10. GS is an approach that is intended to incorporate the time effect into the CCA via DA in order to effectively handle the time-dependent multivariate data (TDMD).

Table 11: **Statistic Measures of 2-Group GS DCCA Functions**

GS	DCCC	C Class	Eig V	W Lambda	Chi-Square
1	0.141	61.0	0.020	0.980	5.221
2	0.219	65.9	0.050	0.952	12.697
3	0.341	70.8	0.131	0.884	31.965
4	0.425	76.1	0.220	0.819	51.598
5	0.466	75.4	0.278	0.783	63.502
6	0.580	78.8	0.507	0.663	106.263
7	0.556	77.0	0.448	0.690	95.964
8	0.578	74.6	0.501	0.666	105.117
9	0.628	80.3	0.652	0.605	129.981
10	0.700	86.0	0.962	0.510	174.606
11	0.781	89.8	1.564	0.390	243.834
12	0.851	93.6	2.619	0.276	333.130
13	0.898	97.3	4.143	0.194	424.140
14	**0.907**	**98.9**	**4.627**	**0.178**	**447.452**
15	0.878	95.8	3.377	0.228	382.372
16	0.835	92.0	2.311	0.302	310.115
17	0.720	87.9	1.474	0.404	234.660
18	0.720	86.0	1.074	0.482	188.894
19	0.632	86.0	0.666	0.600	132.276
20	0.541	84.8	0.251	0.707	89.849

Source: Researcher's computations (2023)

Table 11 reports the first DCCCs and other relevant statistic measures of the 2-group discriminant canonical correlation analysis functions (DCCAFs). GS 14 has the best statistic values among all the schemes, which falls on the 168^{th} month, the end of 2013. That is, assuming the data may be suitably segmented into two, then the partition that is provided by month 168 is the partition that provides the best correlation between the two subsets weather conditions. This partition is given by the 14^{th} GS. It is worth noting that the highest discriminant canonical correlation coefficient (DCCC) obtained in this case is 0.907 which is slightly higher than that in Table 10.

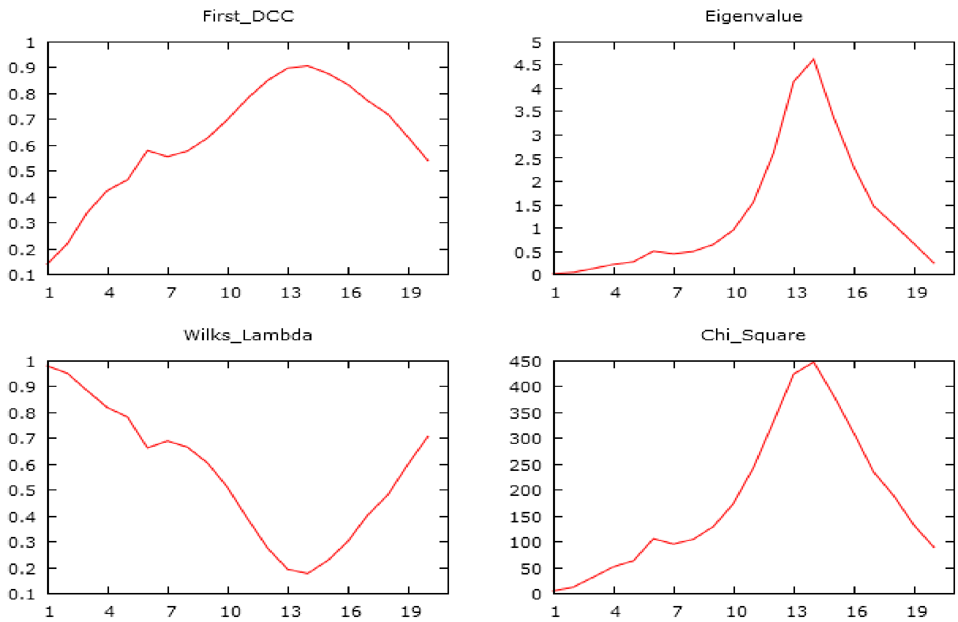

Figure 7: Series Plots of Statistic Measures of 2-group DCCA Functions

Figure 7 depicts the plots of the Two-group DCCAFs confirming that the optimal values are all in the GS 14. Hence, GS 14 is the best GS so far as Two-group discrimination is concerned. The overall CCC corresponds to one of the FDCCs of Two-grouping schemes. This CCC may not necessarily be the highest among all FDCCs of the GS but in my case they are the highest.

Multiple-group discriminant canonical correlation analysis results

I perform several grouping scheme discriminant canonical correlation analysis (GSDCCA) from Three-group up to Sixteen-group DCCA. This primary goal of this section is to determine whether all FDCCs, Eigenvalues and Chi-Square values in a given Multiple-group are greater than or equal to the statistics of Two-group's 14^{th} GS values in Table 11 with corresponding lowest Wilks

Lambdas. In all the Multiple-group DCCA that are examined, the various colours as defined in the graph are used to denote the dummy variables for the respective schemes.

Three-group grouping scheme DCCA functions results

Figure 8 depicts the grouping schemes pictorial map for 3-group DCCA functions. The details in this figure are used to generate the needed classification of Three-group discriminant functions shown in Table 12 to check whether all the relevant statistics are more optimal than those given in the 14^{th} GS shown in Table 11.

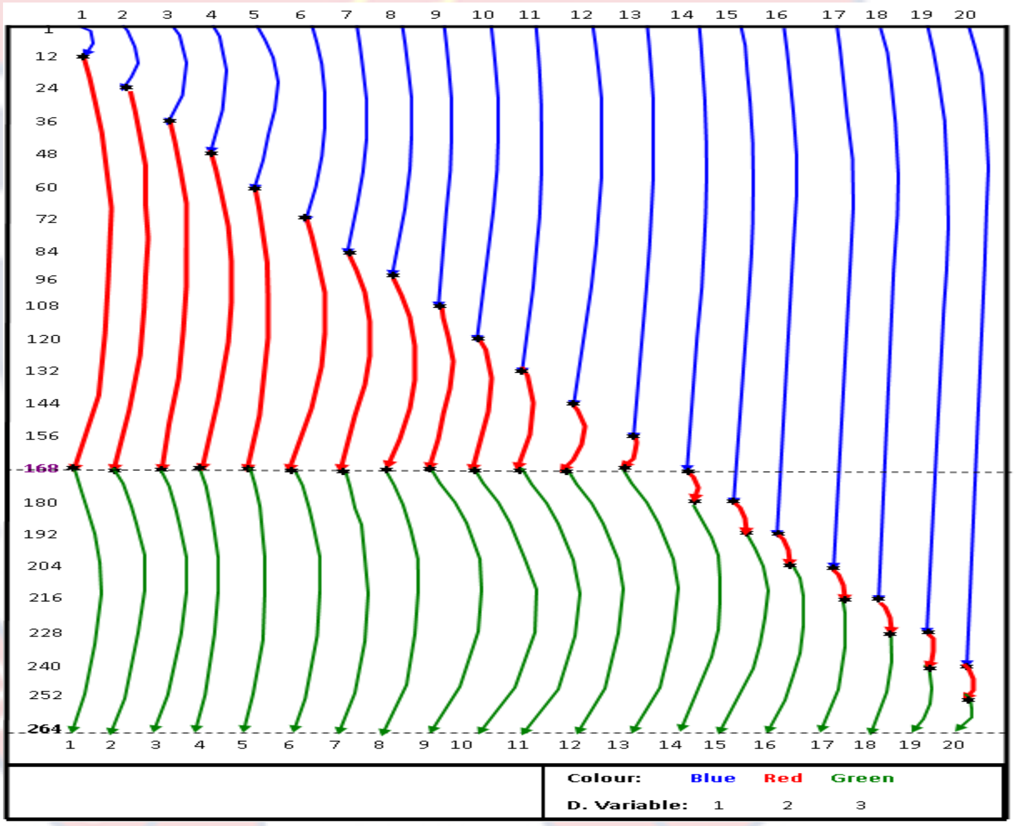

Figure 8: Grouping Scheme Pictorial map for 3-group DCCA Functions

Table 12 reports the statistics values of the Three-group canonical discriminant functions. Here, Grouping Scheme 4 has the most optimal statistics and fall on the 48^{th} month, at the end of 2003. However, the correct classification (92.4%) is not the highest among all the correct classifications. The 17^{th} GS reports the highest correct classification of 95.5%. All the FDCC values are greater than the overall CCC of 0.901, all the Wilks Lambda values are less than the overall value of 0.187 and the respective Eigenvalues are also greater than the overall Eigenvalue of 4.339 as expected.

Table 12: **Statistic Measures of 3-Group GS DCCA Functions**

GS	FDCC	C. Class.	Eig V	W. Lambda	Chi-Square
1	0.910	82.6	4.794	0.170	457.736
2	0.919	86.4	5.452	0.148	493.471
3	0.929	89.8	6.347	0.121	546.418
4	**0.934**	92.4	**6.854**	**0.105**	**582.402**
5	0.926	87.9	6.055	0.114	560.914
6	0.917	89.4	5.296	0.113	564.120
7	0.912	83.7	4.970	0.133	522.307
8	0.911	83.0	4.854	0.140	508.272
9	0.910	83.7	4.801	0.137	512.907
10	0.908	88.3	4.714	0.129	528.497
11	0.907	88.6	4.655	0.128	531.192
12	0.917	89.4	5.294	0.133	520.539
13	0.923	88.3	5.726	0.142	504.193
14	0.910	93.2	4.794	0.155	482.027
15	0.908	90.9	4.720	0.151	488.936
16	0.911	93.6	4.893	0.146	497.272
17	0.917	**95.5**	5.314	0.140	508.981
18	0.919	93.9	5.453	0.145	498.589
19	0.922	90.9	5.664	0.147	496.067
20	0.917	88.6	5.270	0.157	478.591

Source: Researcher's computations (2023)

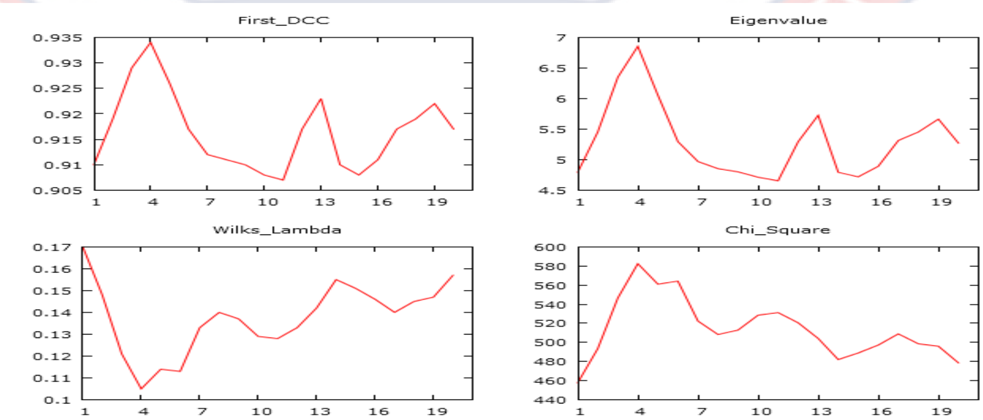

Figure 9: Series Slots of Statistic Measures of 3-group DCCA Functions

Figure 9 depicts the plots of the statistics of the Three-group discriminant CCA functions confirming the four optimal values in the GS 4.

Four-group grouping scheme DCCA Functions results

Figure 10 depicts the grouping schemes pictorial map for Four-group DCCA functions. The details in this figure are used to generate the needed table for statistic measures of Four-group DCCA functions.

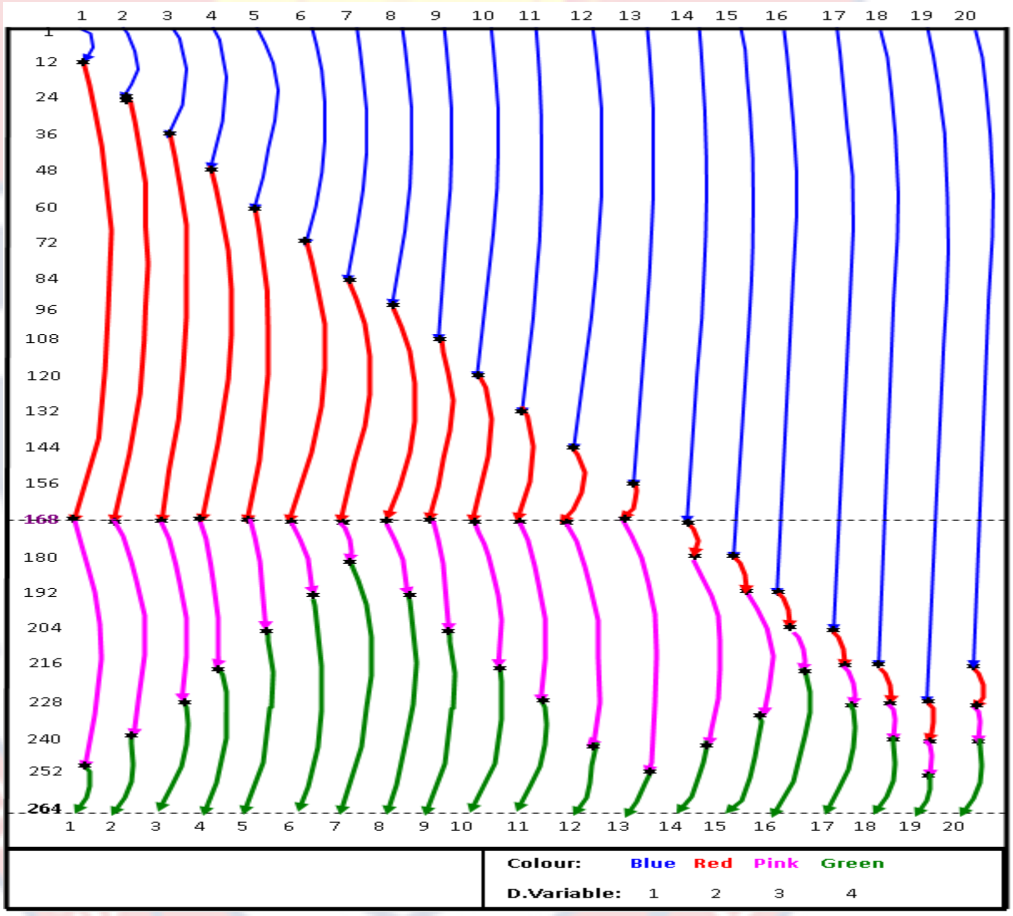

Figure 10: Grouping Scheme Pictorial map for 4-group DCCA Functions

Table 13 reports the relevant statistics of the Four-group GS DCCA functions. In this case, all the FDCC values are greater than the overall CCC of 0.902, the Wilks Lambda values are less than the overall value of 0.187 and the respective Eigenvalues are also greater than the overall Eigenvalue of 4.339 as required. It is observed that GS 4 has the most optimal values for all the statistic measures.

Table 13: **Statistic Measures of 4-Group GS DCCA Functions**

GS	FDCC	C. Class.	Eig V	W. Lambda	Chi-Square
1	0.919	77.7	5.468	0.150	489.375
2	0.934	83.0	6.813	0.120	547.881
3	0.945	86.6	8.342	0.089	623.733
4	**0.948**	**91.7**	**8.884**	**0.074**	**672.814**
5	0.933	84.8	6.760	0.089	624.132
6	0.920	80.7	5.476	0.094	609.164
7	0.916	78.0	5.186	0.115	558.404
8	0.921	75.0	5.596	0.122	542.141
9	0.927	77.7	6.082	0.110	569.401
10	0.922	84.8	5.679	0.103	586.450
11	0.917	86.7	5.317	0.100	595.289
12	0.920	85.2	5.534	0.110	570.155
13	0.924	82.2	5.825	0.121	545.563
14	0.920	84.8	5.508	0.116	555.945
15	0.909	81.8	4.739	0.108	573.591
16	0.914	84.5	5.056	0.104	584.822
17	0.921	80.3	5.626	0.106	579.601
18	0.925	81.4	5.964	0.109	572.374
19	0.929	78.4	6.289	0.106	578.697
20	0.928	81.1	6.185	0.097	600.922

Source: Researcher's computations (2023)

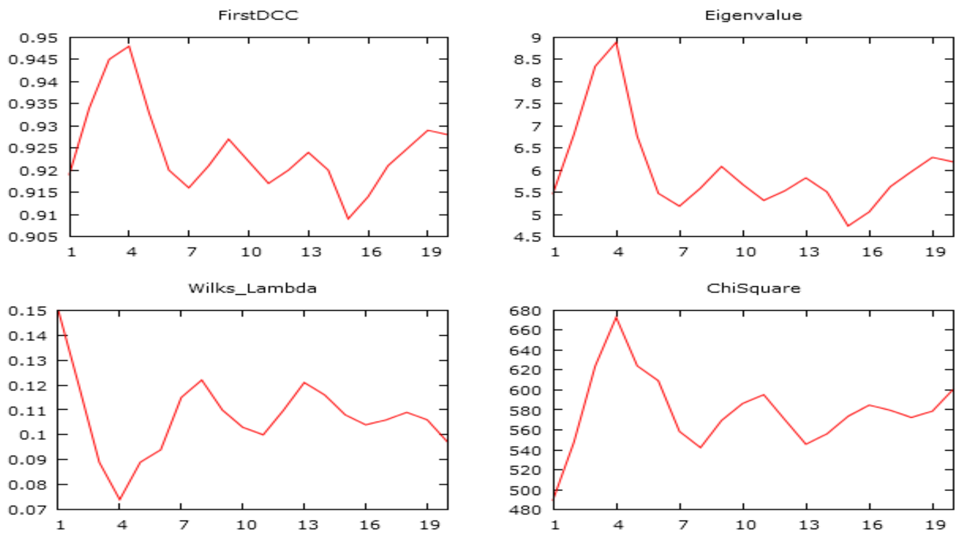

Figure 11: Series Plots of Statistic Measures of 4-group DCCA Functions

Figure 11 depicts the plots of the relevant statistic measures for Four-group discriminant canonical correlation analysis functions confirming that the most optimal values are all in grouping scheme 4, the 48^{th} month, at the end of 2003.

Five-group grouping scheme DCCA functions results

Figure 12 depicts the grouping scheme pictorial map for Five-group grouping scheme discriminant canonical correlation analysis functions. The details in this figure are used to generate the needed information.

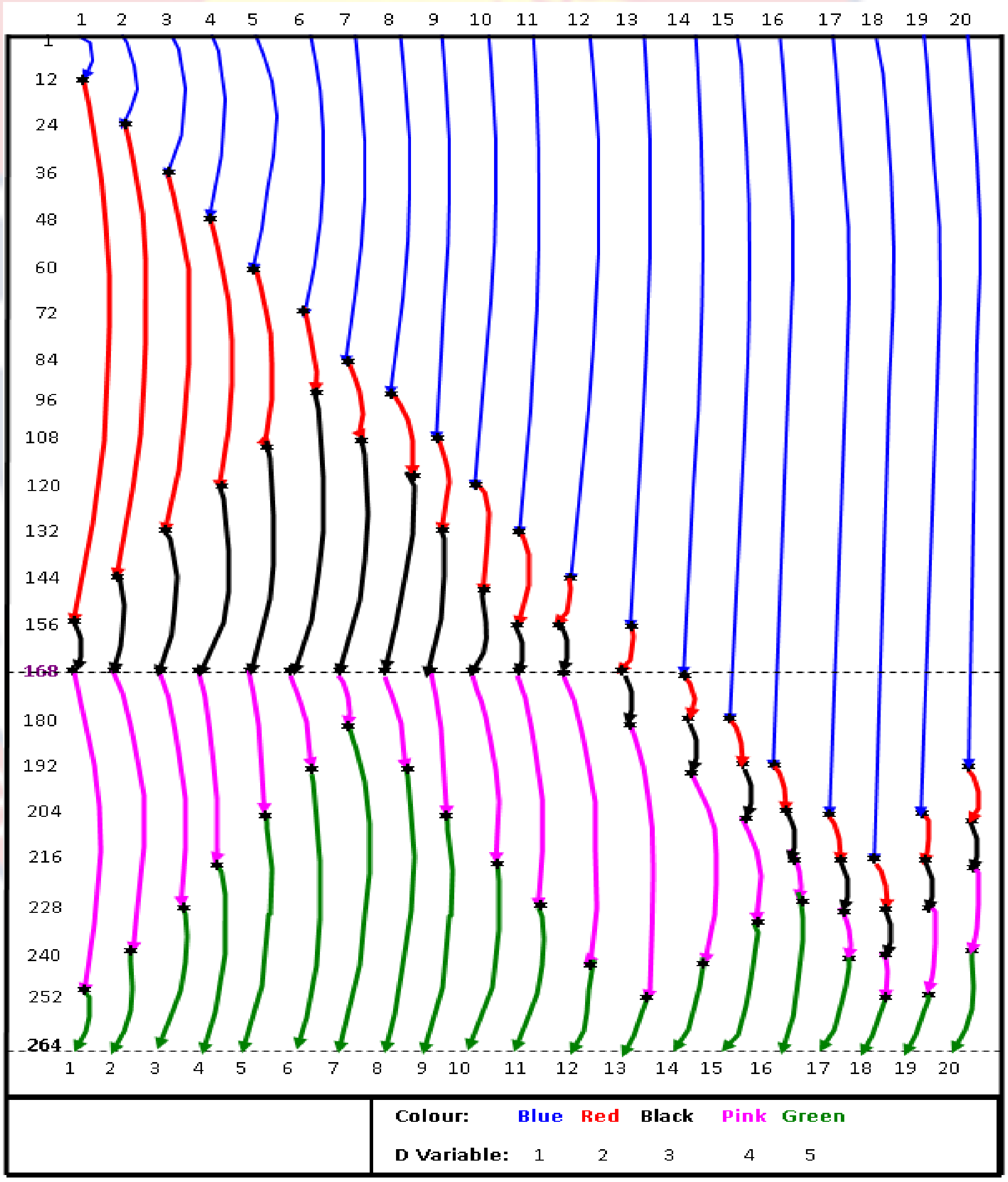

Figure 12: Grouping Scheme Pictorial map for 5-group DCCA Functions

Table 14 reports the statistic measures of the Five-group discriminant canonical correlation analysis functions. From the table, all the statistic measures for each GS are better than the overall canonical correlation coefficient statistic measures. Grouping scheme 14 has the most optimal values over all the statistic measures and are achieved on the 168^{th} month, at the end of 2013, which is the same as reported in the Two-group discriminant canonical correlation analysis functions scores given in Table 11.

Table 14: **Statistic Measures of 5-Group GS DCCA Functions**

GS	FDCC	C. Class.	Eig V	W. Lambda	Chi-Square
1	0.936	79.5	9.347	0.059	727.005
2	0.946	83.3	8.565	0.082	644.846
3	0.947	86.4	8.642	0.064	709.626
4	0.948	84.5	8.887	0.057	739.447
5	0.934	75.8	6.784	0.076	664.485
6	0.920	71.2	5.477	0.088	626.025
7	0.931	73.1	6.538	0.088	625.616
8	0.935	73.1	6.921	0.080	651.043
9	0.937	75.8	7.172	0.077	661.864
10	0.932	78.4	6.613	0.080	648.902
11	0.923	76.9	5.716	0.099	596.702
12	0.959	86.4	11.323	0.060	725.510
13	0.958	86.4	11.246	0.069	689.782
14	**0.967**	**89.0**	**14.375**	**0.045**	**799.903**
15	0.936	83.7	7.117	0.062	718.082
16	0.930	81.4	6.453	0.067	696.751
17	0.941	77.7	7.804	0.068	692.513
18	0.937	73.1	7.202	0.077	660.971
19	0.934	73.9	6.795	0.085	635.070
20	0.922	72.3	5.681	0.096	604.435

Source: Researcher's computations (2023)

Figure 13 depicts the plots of Five-group grouping scheme canonical discriminant statistics confirming the highest score of the first discriminant canonical correlation, the highest score of Eigenvalue, the highest score of Chi-Square value and the lowest score of Wilks Lambda, all in the grouping scheme 14, the 168^{th} month, at the end of 2013.

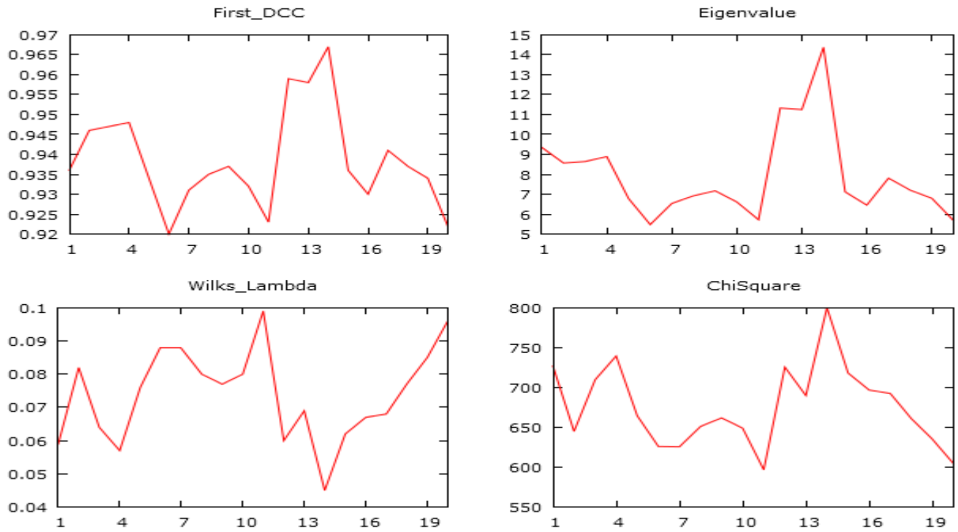

Figure 13: Series Plots of Statistic Measures of 5-group DCCA Functions

Six-group grouping scheme DCCA functions results

Table 15 depicts the five relevant statistic measures of the Six-group GS

Table 15: **Statistic Measures of 6-Group GS DCCA Functions**

GS	FDCC	C. Class.	Eig V	W. Lambda	Chi-Square
1	0.945	71.6	8.365	0.079	653.744
2	0.952	77.3	9.604	0.057	738.338
3	0.948	80.7	8.829	0.052	759.978
4	0.948	**81.1**	8.949	0.051	764.437
5	0.934	71.6	6.854	0.070	681.691
6	0.920	67.4	5.511	0.085	632.212
7	0.931	69.3	6.501	0.077	658.579
8	0.935	68.2	6.926	0.071	680.794
9	0.938	70.1	7.270	0.061	718.990
10	0.935	77.3	6.917	0.054	752.116
11	0.937	76.5	7.138	0.059	729.125
12	0.952	76.9	9.571	0.058	731.516
13	0.961	77.3	11.937	0.056	742.029
14	**0.961**	79.9	**12.123**	**0.048**	**777.800**
15	0.946	79.9	8.600	0.055	745.050
16	0.931	78.8	6.457	0.063	708.749
17	0.934	71.6	6.854	0.070	681.691
18	0.937	73.1	7.202	0.077	660.971
19	0.935	67.0	6.920	0.074	670.785
20	0.924	66.7	5.845	0.071	680.211

Source: Researcher’s computations (2023) discriminant canonical correlation analysis functions. From the table, all the first discriminant canonical correlation (FDCC) values are greater than the overall canonical correlation coefficient of 0.902, the Eigenvalues are greater than the overall Eigenvalue of 4.339 and the Wilks Lambda values are less than the overall value of 0.187 reported in Table 10. However, correct classification value reduces for all grouping schemes due to increased group-discrimination. The 4^{th} GS reports the highest correct classification of 81.1%. The 14^{th} grouping scheme reports the most optimal values of the remaining statistic measures which are achieved on the 168^{th} month, at the end of 2013. Figure 14 depicts the plots of the statistics for the Six-group grouping scheme discriminant canonical correlation analysis functions confirming the most optimal values all in grouping scheme 14.

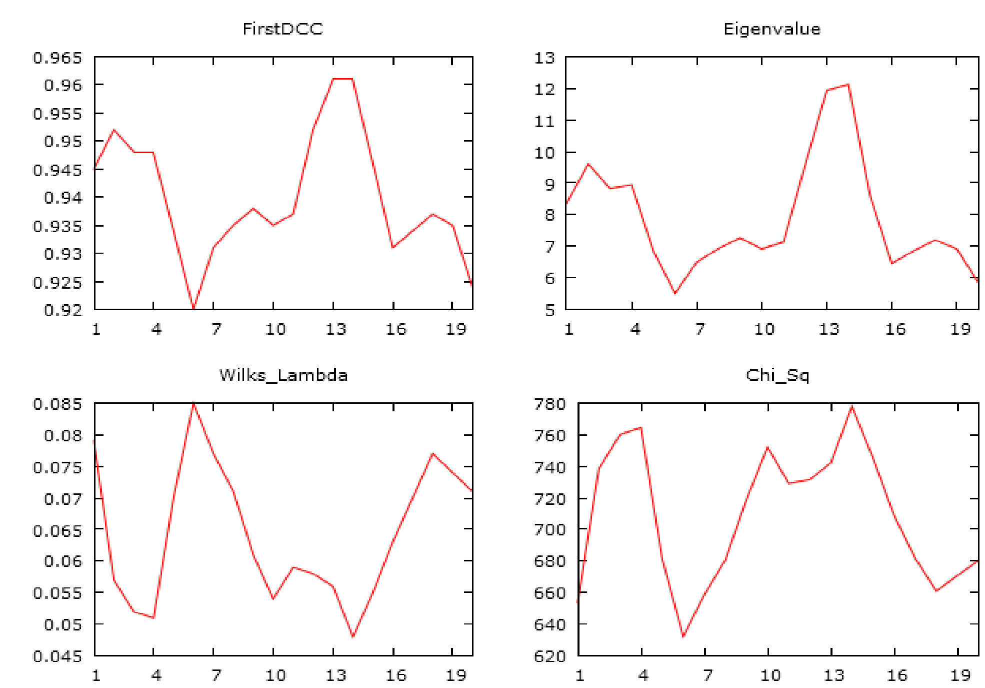

Figure 14: Series Plots of Statistic Measures of 6-group DCCA Functions

Seven-group grouping scheme DCCA functions results

Figure 15 depicts the grouping scheme pictorial map for Seven-group grouping scheme discriminant canonical correlation analysis functions which is used to generate the requisite information.

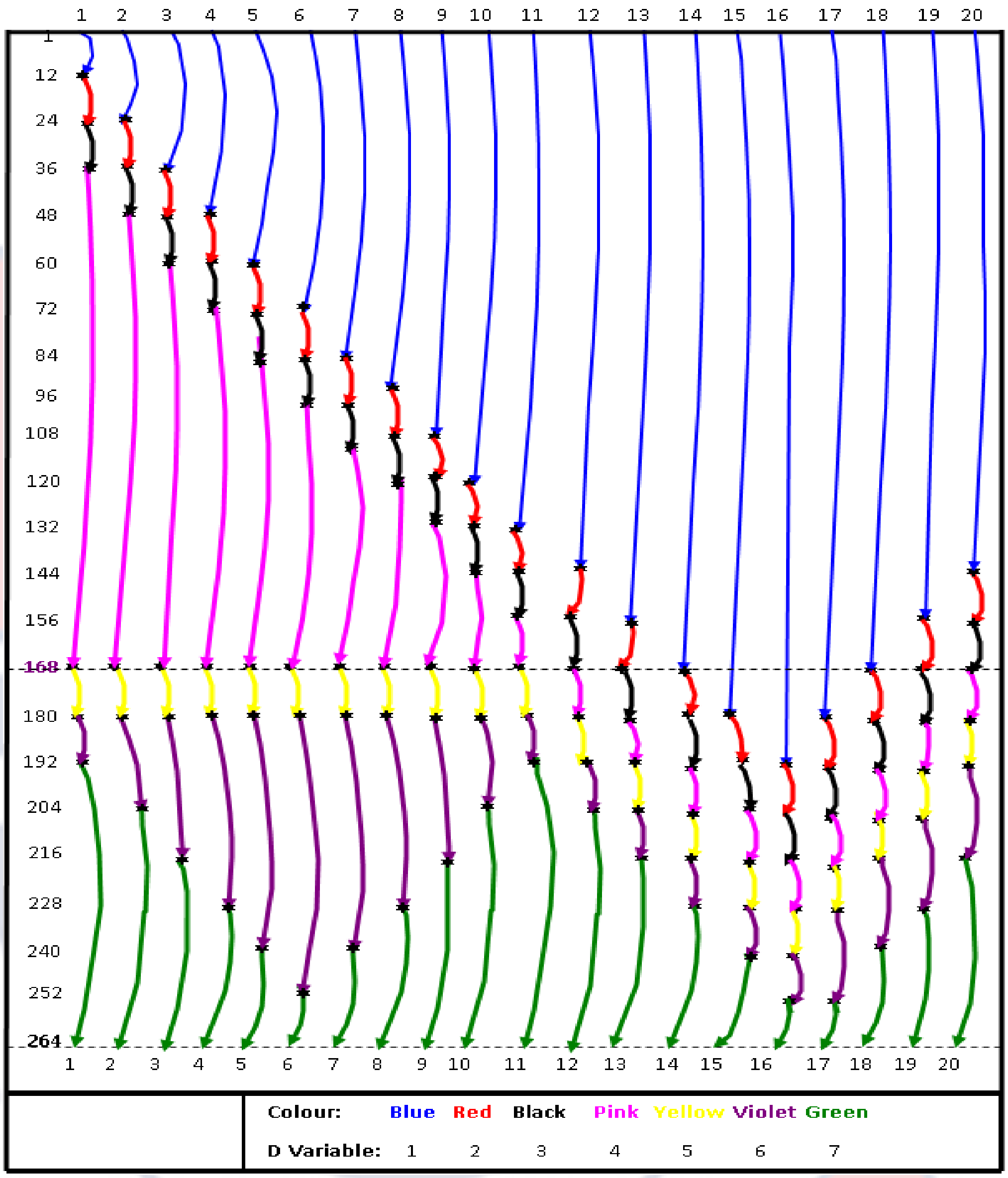

Figure 15: Grouping Schemes Pictorial map for 7-group DCCA Functions

Table 16 depicts the statistics of Seven-group discriminant canonical correlation analysis functions. From the table, for each grouping scheme, measures are more optimal than corresponding values of the overall canonical correlation coefficient except for the correct classifications that decrease as a result of increased group discrimination. The grouping scheme 14 reports the most optimal values for all five measures which also fell on the 168^{th} month, at the end of 2013.

Table 16: **Statistic Measures of 7-Group GS DCCA Functions**

GS	FDCC	C. Class.	Eig V	W. Lambda	Chi-Square
1	0.950	69.3	9.347	0.059	727.005
2	0.951	73.5	9.509	0.050	766.905
3	0.946	75.8	8.534	0.047	785.478
4	0.943	73.9	7.986	0.052	757.882
5	0.943	68.2	8.090	0.058	731.603
6	0.925	64.8	5.952	0.075	663.451
7	0.931	64.8	6.524	0.071	677.124
8	0.939	66.3	7.434	0.064	704.327
9	0.939	65.9	7.399	0.057	735.468
10	0.935	70.1	6.929	0.050	767.748
11	0.937	71.6	7.148	0.053	754.499
12	0.952	71.2	9.574	0.048	778.181
13	0.956	76.1	10.717	0.055	743.672
14	**0.968**	**81.8**	**14.928**	**0.032**	**881.004**
15	0.947	74.6	8.614	0.050	770.788
16	0.931	71.6	6.471	0.059	724.858
17	0.935	67.4	6.963	0.066	697.764
18	0.942	68.2	7.815	0.068	689.833
19	0.945	67.4	8.386	0.058	728.640
20	0.938	68.2	7.264	0.058	728.935

Source: Researcher's computations (2023)

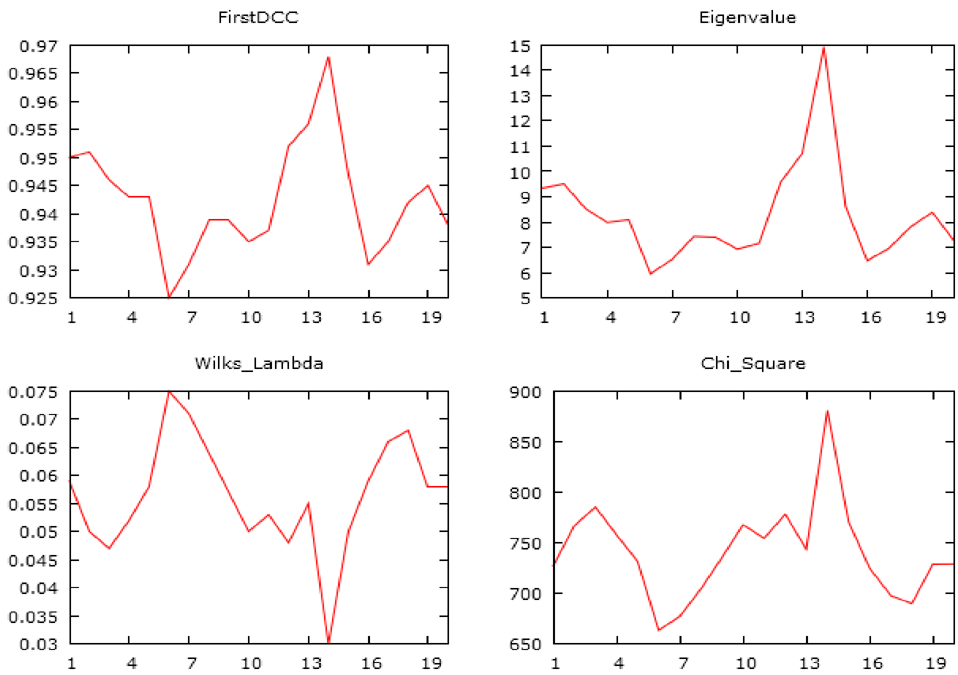

Figure 16: Series Plots of Statistic Measures of 7-group DCCA Functions

Figure 16 depicts the plots of Seven-group DCC functions confirming the most optimal values for all four measures of discrimination. My results show that from Group-seven up to Group-twenty, only first six canonical discriminant functions for each Multiple-group are used in the analysis.

It is noteworthy that, at this point, aside from the Two-group grouping scheme discriminant CCA, it is the Four-group, Five-group and Seven-group grouping scheme discriminant CCA that produce the most optimal statistics values over all five measures for the same best grouping scheme.

Summary statistic measures for all best multiple-groups DCCAFs

Table 17 depicts the summary of all the highest scores of the GS classifications of all the Multiple-groups from Two-group to Sixteen-group, reporting the highest FDCC values, the highest Eigenvalues, the lowest Wilks Lambda values and the highest Chi-square values and their respective grouping schemes.

Table 17: **Summary Statistic Measures for all Best Multiple-Groups**

Group	FDCC	Eig V	Wilks L.	Chi-Sq	GS	Month	Year
2	**0.907**	**4.627**	**0.178**	**447.452**	**14**	**168^{th}**	**2013**
3	0.934	6.854	0.105	582.402	4	48^{th}	2003
4	0.948	8.884	0.074	672.814	4	48^{th}	2003
5	0.967	14.375	0.045	799.903	14	168^{th}	2013
6	0.961	12.123	0.048	777.800	14	168^{th}	2013
7	**0.968**	**14.928**	**0.032**	**881.004**	**14**	**168^{th}**	**2013**
8	0.956	10.601	0.046	789.584	14	168^{th}	2013
9	0.964	13.190	0.047	784.399	14	168^{th}	2013
10	0.965	13.628	0.047	791.777	5	60^{th}	2004
11	0.965	13.715	0.040	866.505	4	48^{th}	2003
12	0.962	12.294	0.038	808.021	14	168^{th}	2013
13	0.963	12.924	0.036	828.047	14	168^{th}	2013
14	0.965	13.669	0.034	876.168	14	168^{th}	2013
15	0.964	12.996	0.034	841.741	4	48^{th}	2003
16	0.962	12.385	0.043	752.701	4	48^{th}	2003

Source: Researcher's computations (2023)

Since the best GS may not report the highest correct classification, the correct classification is not provided in the table. The table confirms that Seven-group grouping scheme multiple discriminant canonical correlation analysis outperforms all the other groupings, from Three-group to Sixteen-group. Again, out of all the Multiple-groups, it is GS 14 of Seven-group canonical discriminant function that gives the most optimal statistics for all four measures. Hence, GS 14 of the Seven-group GS discriminant CCA is used to perform the remaining analyses.

Again 60% ($\frac{9}{15}$) of the possible Multiple-group DA yields GS 14 as the best grouping scheme. The findings also show that when the time impact is considered at the canonical correlation analysis, the correlation between the subset of cooling variables and heating variables improves, rising from 0.901 to 0.968.

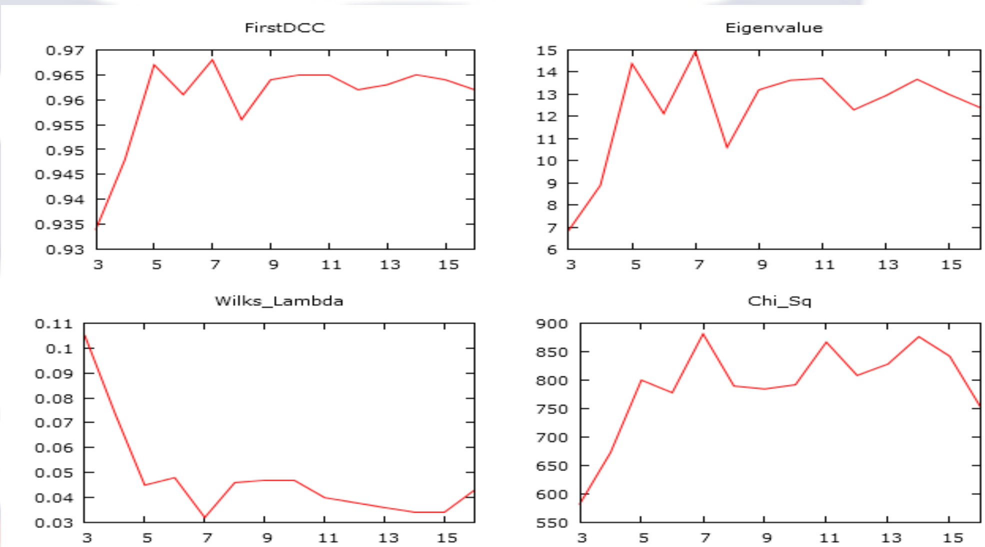

Figure 17: Plots of Overall Best-group DCCA Functions Statistic Measures

From Figure 17, the highest FDCC sharply increased from 3-group up to 5-group and slowly decrease from 5-group up to 7-group and slowly decreases again from there to Sixteen-group. The Eigenvalues shows similar results as the FDCC. The movement of the Wilks Lambda is the reverse of the corresponding FDCC. The figure confirms that 7-group reports the highest FDCC value for all Multiple-groups and hence outperforms all of them from 3-group to 17-group.

Figure 18 depicts the first nine plots from 3-group to 11-group of overall best highest FDCC values from GS 1 to GS 20. The figure confirms that GS 14 reports the highest values for a number of Multiple-group FDCC, for example, 5-group, 6-group, 7-group and 8-group.

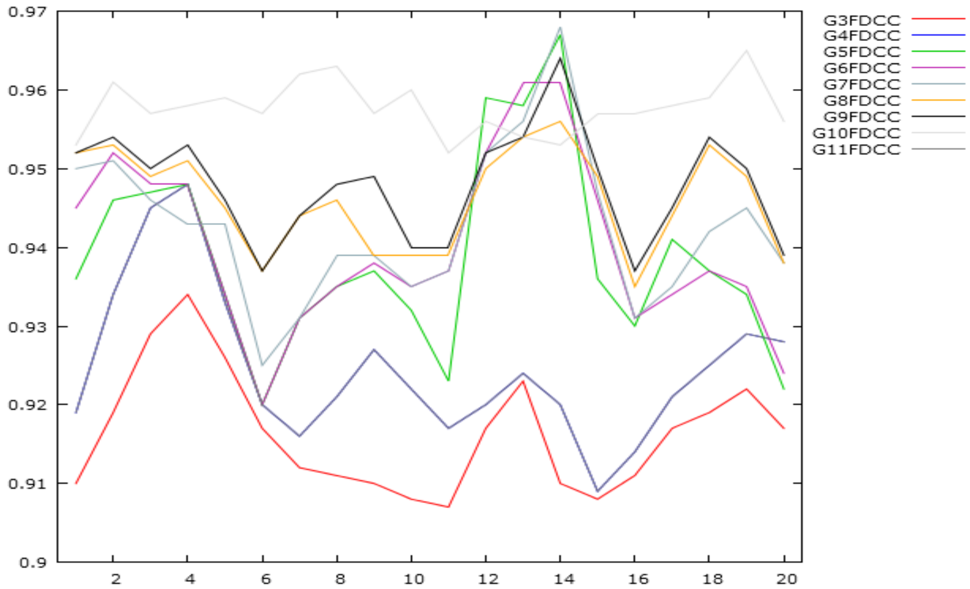

Figure 18: Plots of Multiple-group FDCCs from 3-group to 11-group

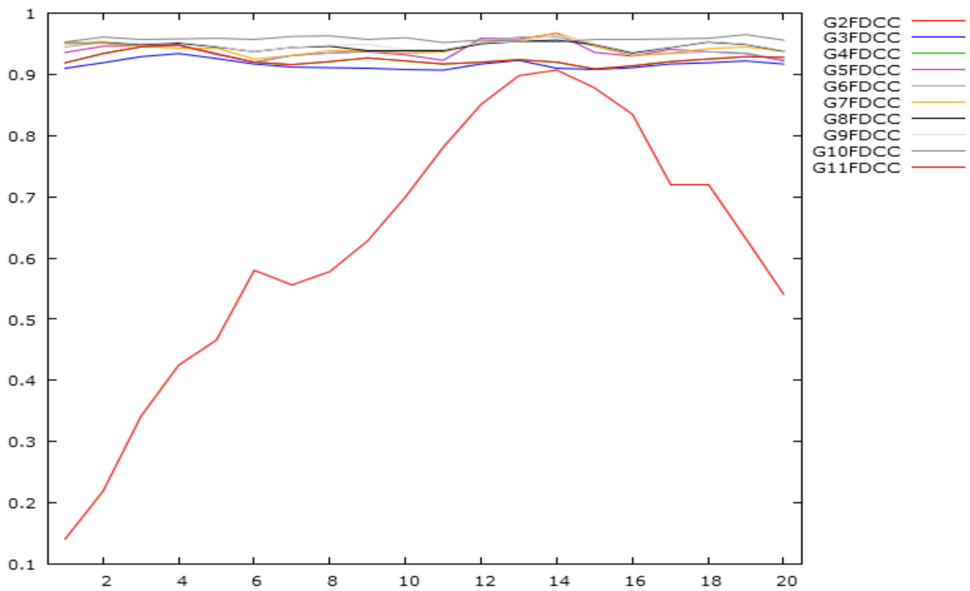

Figure 19: Plots of Multiple-group FDCCs from 2-group to 11-group

Figure 19 depicts the first ten series plots of the original data from Two-group to Eleven-group FDCC values from GS 1 to GS 20. The figure confirms that all the Multiple-group's FDCC values are higher than the highest Two-group DCC value of 0.907 in GS 14 given in Table 11 as well as the overall value of 0.901 also in Table 10.

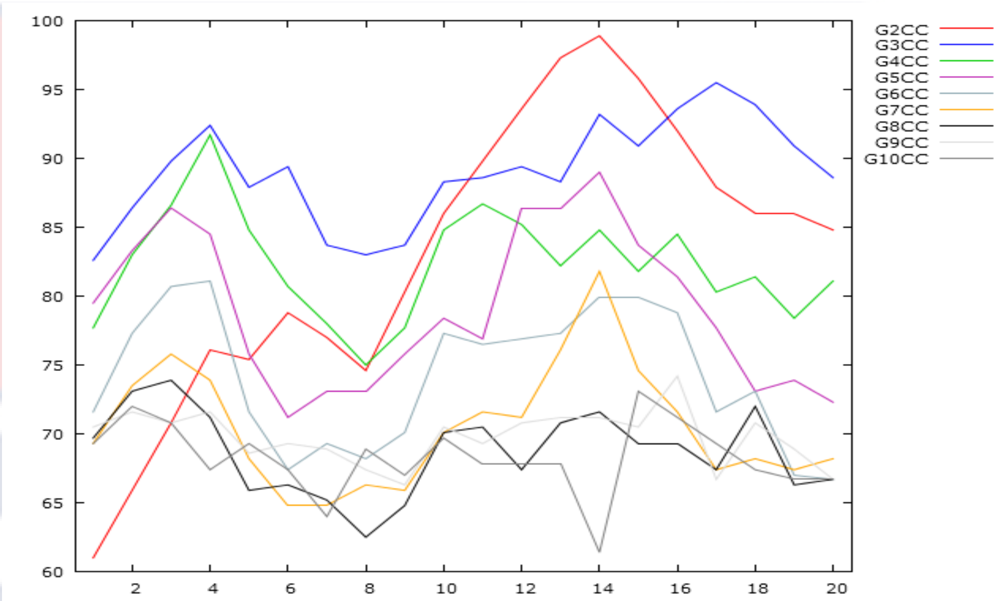

Figure 20: Plots of Correct Classifications from 2-group to 10-group

Figure 20 reports the first nine series plots from Two-group to Ten-group of correct classification values from GS 1 to GS 20. The figure confirms that the Two-group GSDCCA outperforms all the Multiple-group GSDCCA, follows by the Three-group, Five-group, Four-group, Seven-group GSDCCA, etc., and the Ten-group GSDCCA gives the lowest so far as GS 14 is concerned.

Statistical Significance of DCCA Functions

Table 18 displays a summary statistics of the DCCFs for GS 14 of the Seven-group discriminant canonical correlation analysis. The six (6) discriminant canonical correlation Functions are selected for the analysis in this study. According to the Chi-square test, the chi-square number provided in the table's first row indicates the statistical significance of every potential function, not just the first function. To evaluate the overall statistical significance of the CCA coefficients, I calculate the Chi-square values for each of the six functions.

Table 18: **Best Seven-Group Discriminant CCA Functions**

Fun	DCC	Chi-Sq	Eig V	Var	C Var	W Lamb	Sig.
1	0.968	881.004	14.928	95.0	95.0	0.032	0.000
2	0.578	170.995	0.503	3.2	98.2	0.513	0.000
3	0.432	66.515	0.230	1.5	99.7	0.772	0.000
4	0.207	13.488	0.045	0.3	99.9	0.949	0.142
5	0.091	2.205	0.008	0.1	100.0	0.991	0.698
6	0.017	0.075	0.000	0.0	100.0	1.000	0.785

Source: Researcher's computations (2023)

The first Chi-square (χ_1^2) value is computed as follows:

$$\begin{aligned}
\chi_1^2 &= \left[n - 1 - \left(\frac{p+G}{2}\right)\right] \sum_{k=1}^{6} ln\left(1+\lambda_k\right) \\
&= \left[264 - 1 - \left(\frac{6+7}{2}\right)\right] \left[ln\left(15.928\right) + ln\left(1.503\right) + ... + ln\left(1.008\right) + ln\left(1.0\right)\right] \\
&= (256.5)(3.434540902) \\
&= 880.9597
\end{aligned}$$

where λ_i are the eigenvalues. Within rounding error, the Chi-square value of 880.9597 is the same as the Chi-square value given in Table 17. The result means that at least the first DCC is statistically significant because the Chi-square value in this case is statistically significant. If the additional discriminant functions are statistically significant, it means that they explain substantially the difference between the Seven-group than the initial discriminant function.

The second Chi-square (χ_2^2) value, is obtained as follows:

$$\chi_2^2 = \left[n - 1 - \left(\frac{p+G}{2}\right)\right] \sum_{k=2}^{6} ln\left(1 + \lambda_k\right)$$
$$= \left[264 - 1 - \left(\frac{6+7}{2}\right)\right] \left[ln\left(1.503\right) + ln\left(1.23\right) + ... + ln\left(1.008\right) + ln\left(1.0\right)\right]$$
$$= (256.5)(0.6665)$$
$$= 170.9573$$

Noting that the initial discriminant function's eigenvalue is ignored. The third Chi-square (χ_3^2) value, is obtained as follows:

$$\chi_3^2 = \left[n - 1 - \left(\frac{p+G}{2}\right)\right] \sum_{k=3}^{6} ln\left(1 + \lambda_k\right)$$
$$= \left[264 - 1 - \left(\frac{6+7}{2}\right)\right] \left[ln\left(1.23\right) + ln\left(1.045\right) + ln\left(1.008\right) + ln\left(1.0\right)\right]$$
$$= (256.5)(0.2590)$$
$$= 66.4335$$

The fourth Chi-square (χ_4^2) value, is obtained as follows:

$$\chi_4^2 = \left[n - 1 - \left(\frac{p+G}{2}\right)\right] \sum_{k=4}^{6} ln\left(1 + \lambda_k\right)$$
$$= \left[264 - 1 - \left(\frac{6+7}{2}\right)\right] \left[ln\left(1.045\right) + ln\left(1.008\right) + ln\left(1.0\right)\right]$$
$$= (256.5)(0.05199)$$
$$= 13.3354$$

A significant Chi-square finding indicates that the second, third, and fourth functions significantly explain the group difference not captured by the first discriminant function. It may be concluded that the second discriminant function explains at least some of the difference between the Seven-group that is not described by the first discriminant function, as seen by its statistically significant Chi-square value (170.9573). The same is true for the third discriminant function Chi-square value (66.4335). The fourth, fifth, and sixth discriminant functions are not statistically significant as in Table 18.

Estimation of Discriminant CCA Function Scores

Table 19 depicts the unstandardized discriminant canonical correlation function coefficients that are used to generate the needed functions for dependent and independent variables. This section provides the various control parameters for calculating the discriminant functions. Six discriminant canonical correlation functions can be computed for Seven-group discriminant canonical correlation analysis.

Table 19: **Unstandardized Discriminant CCA Functions Coefficients**

Variable	Fn 1	Fn 2	Fn 3	Fn 4	Fn 5	Fn 6
Max. Temp.	2.411	1.986	-1.138	-0.778	2.470	-0.738
Min. Temp.	-0.185	-1.689	1.606	1.114	-2.503	0.263
Solar Rad.	0.035	-0.001	-0.108	0.078	-0.025	0.243
Precipitation	0.005	-0.180	0.231	0.041	0.212	0.164
Wind	-0.154	-0.397	-0.673	0.570	0.627	-0.331
Rel. Humid.	1.278	34.859	10.562	17.087	7.252	-8.249
Constant Term	-61.680	-35.431	-14.341	-27.394	-10.459	16.221

Source: Researcher's computations (2023)

Simulation Studies of Time-Independent Data

Consistent with the first part of this chapter, the first part of the simulation assumes a Time-independent data that does not incorporate the effect of the time (years). Using the simulated data thus obtained, the GSDCCA will be carried out to establish the relevance of the concept of grouping scheme proposed in this thesis. The second part of the simulation would then assume a time-dependent data that incorporates the effect of the year based on the best Seven-group GSDCCA identified at the end of the first part of this chapter.

Simulation of Multiple-group DCCA functions

This section is to build a simulation result tables that identifies the best grouping scheme for all possible Multiple-group Discriminant Analysis. Through the use of datasets for simulated weather conditions, we assess the effectiveness of the suggested approach. My approach and findings are compared with those of the original weather circumstances using discriminant canonical correlation analysis of Time-independent data structure. The summaries of the simulation data are presented in Table 20 where n represents the total number of observations, d represents the mean of the total number of time-dependent variables, and T is the total number of time events that were recorded.

Table 20: **Parameters for Time-Independent Simulation Data**

Sum Stats	Max T	Min T	Solar R	Precip	Wind	Rel Hum
n	264	264	264	264	264	264
d	27.281	26.500	22.071	1.779	5.793	0.824
T	1.761	1.612	4.555	2.786	0.978	0.040
Min Value	22.250	21.897	9.061	-6.168	3.000	0.708
Max Value	32.285	31.079	35.014	9.704	8.571	0.938

Source: Researcher's computations (2023)

Series plots of simulation for Time-independent weather condition

The plots of the monthly simulated weather conditions data in Ghana from January 2000 to December 2021 are shown in Figure 21. We can see from the figures that the data exhibit random changes with no noticeable patterns. The magnitudes of the seasonal fluctuations remain the same, although the data values tend to rise over time. Although it is not always the case, the tendency in simulated weather conditions data generally appears to be seasonal. Most of the data points depart marginally from the mean. This implies that mean non-stationarity is present in this situation.

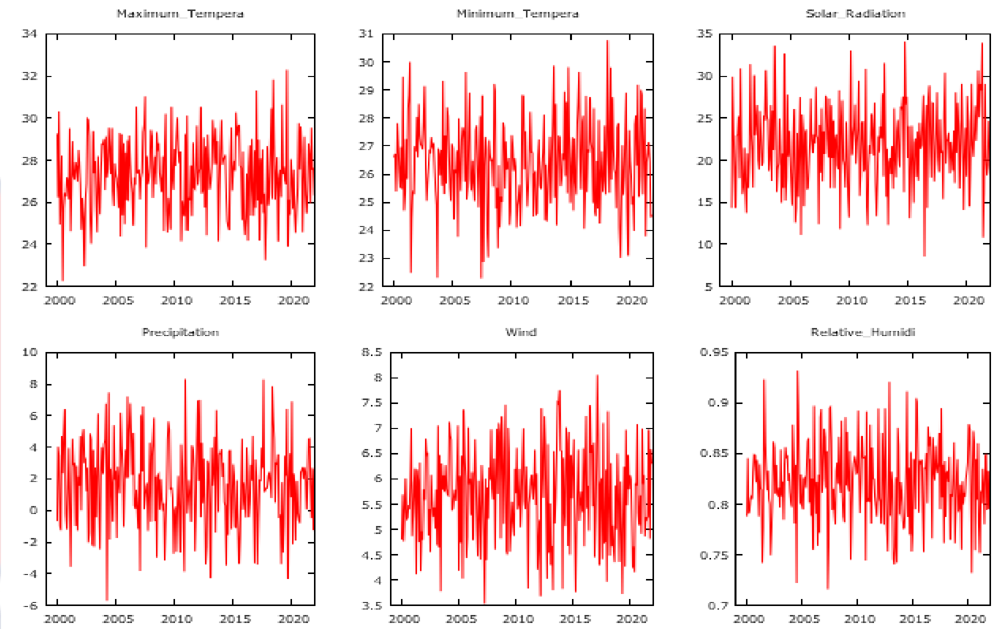

Figure 21: Series Plots of data on Six monthly Time-Ind. Simulation

Summary results of simulated data for DCCA functions for all 3 Roots

Table 21 shows the overall simulated CCC value of 0.7138, the overall Wilks Lambda value of 0.4905, and the overall Eigenvalue of 1.0386. The simulated data is used to generate 2-group up to 16-group GS DCCA to check if the results of the statistic measures are less than or greater than the overall results given in Table 21. It is also used to confirm the general results in the original data.

Table 21: **Overall Tests of Significance for Time-Ind. Simulation**

Test Statistic	CCC	Eig V	Wilks	Roys	Hotellings	Pillais
Value	0.7138	1.0386	0.4905	0.5095	1.0386	0.5095
Significance	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Source: Researcher's computations (2023)

Time-independent simulation results of 2-group DCCA functions

Table B1 in the appendix depicts Two-group GSDCCA. The 3^{rd} GS reports the highest FDCC of 0.163 which is less than the overall value, the highest correct classification of 61.4%, the highest eigenvalue of 0.027 which is less than the overall value, the lowest Wilks Lambda value of 0.973 which is greater than the overall value, and the highest Chi-square value of 7.014 as expected. Thus, optimal values for the Two-group GSDCCA are no better than the overall statistic measures. Similarly, Tables B2 and B3 show results for Three-group and Four-group GSDCCA, with GS 4 yielding the optimal statistic measures. Table B4 shows that GS 13 yields the optimal results in Six-group GSDCCA. In all of these, the highest correct classification is attained by the best identified grouping scheme.

Time-independent simulation results of Higher-group DCCA functions

Table 22 depicts Five-group grouping scheme discriminant canonical correlation analysis.

Table 22: **Time-Independent Statistics Measures of 5-Group DCCAFs**

GS	FDCC	C Class	Eig V	W Lambda	Chi-Square
1	0.201	21.6	0.042	0.923	20.764
2	0.183	23.5	0.035	0.926	19.681
3	0.216	25.8	0.049	0.914	23.186
4	0.214	27.3	0.048	0.924	20.482
5	0.181	25.0	0.034	0.940	15.940
6	0.170	28.0	0.030	0.943	15.090
7	0.185	25.4	0.036	0.926	19.710
8	0.211	27.3	0.047	0.929	18.987
9	0.214	26.5	0.048	0.922	21.051
10	0.164	27.3	0.028	0.937	16.621
11	0.162	21.6	0.027	0.942	15.407
12	0.169	21.2	0.030	0.930	18.581
13	0.216	23.9	0.049	0.911	24.070
14	**0.240**	28.4	**0.061**	**0.897**	**28.097**
15	0.176	25.8	0.032	0.927	19.496
16	0.164	22.0	0.028	0.936	16.996
17	0.213	27.3	0.047	0.924	20.445
18	0.213	26.9	0.048	0.929	19.092
19	0.219	**29.9**	0.050	0.919	21.836
20	0.171	26.5	0.030	0.942	15.435

Source: Researcher's computations (2023)

The grouping scheme 14 reports the most optimal statistics for all measures except for correct classification. All of these optimal values are no better than the overall values in Table 21.

Table 23: **Time-Independent Statistics Measures of 7-Group DCCFs**

GS	FDCC	C Class	Eig V	W Lamb	Chi-Square
1	0.213	19.3	0.048	0.880	32.728
2	0.181	22.0	0.034	0.891	29.513
3	0.236	**25.4**	0.059	0.886	31.128
4	0.211	22.0	0.047	0.896	28.145
5	0.187	17.8	0.036	0.905	25.556
6	0.191	20.8	0.038	0.913	23.474
7	0.194	17.4	0.039	0.911	23.962
8	0.215	19.7	0.049	0.904	25.897
9	0.215	20.8	0.049	0.899	27.421
10	0.170	20.8	0.030	0.917	22.248
11	0.177	20.1	0.032	0.910	24.114
12	0.186	22.3	0.036	0.900	27.070
13	0.213	22.3	0.047	0.876	34.102
14	**0.252**	20.8	**0.065**	**0.875**	**34.106**
15	0.212	22.0	0.047	0.893	28.942
16	0.182	20.5	0.034	0.916	22.609
17	0.193	16.7	0.039	0.900	27.102
18	0.215	24.2	0.048	0.909	24.536
19	0.221	17.8	0.051	0.877	33.611
20	0.179	20.8	0.033	0.904	25.981

Source: Researcher's computations (2023)

Time-independent simulation results for Seven-group grouping scheme discriminant canonical correlation analysis is shown in Table 23. The grouping scheme 14 reports the most optimal statistic for all measures except for correct classification but are no better than the overall values in Table 21.

Table 24 depicts the summary of all the highest scores of the grouping scheme Time-independent simulation results of all the Multiple-groups from Two-group to Sixteen-group, reporting the the optimal values for the relevant measures including their respective grouping schemes. The table confirms that Seven-group grouping scheme discriminant canonical correlation analysis has the optimal statistic measures and outperforms all the other Multiple-group grouping scheme discriminant canonical correlation analysis.

Table 24: **Summary Statistic Measures for all Best Time-Indep. DCCAFs**

Group	FDCC	Eig V	W Lamb	Chi-Sq	GS	Month	Year
2	**0.163**	**0.027**	**0.973**	**7.014**	**3**	**36**th	**2002**
3	0.214	0.048	0.945	14.532	4	48^{th}	2003
4	0.215	0.048	0.924	20.895	4	48^{th}	2003
5	0.240	0.061	0.897	28.097	14	168^{th}	2013
6	0.241	0.061	0.885	31.320	13	156^{th}	2012
7	**0.252**	**0.065**	**0.875**	**34.106**	**14**	**168**th	**2013**
8	0.224	0.053	0.877	33.592	14	168^{th}	2013
9	0.203	0.043	0.895	28.283	11	132^{nd}	2010
10	0.166	0.028	0.928	19.019	5	60^{th}	2004
11	0.216	0.049	0.904	25.617	4	48^{th}	2003
12	0.202	0.042	0.901	26.575	14	168^{th}	2013
13	0.211	0.046	0.886	30.823	14	168^{th}	2013
14	0.197	0.040	0.879	32.521	14	168^{th}	2013
15	0.192	0.038	0.914	22.631	4	48^{th}	2003
16	0.208	0.045	0.901	26.234	4	48^{th}	2003

Source: Researcher's computations (2023)

These results demonstrate that as predicted by my approaches, the Seven-group algorithm outperforms the other Multiple-groups of my generated Grouping Scheme Discriminant Canonical Correlation Analysis methods. Here, 40% ($\frac{6}{15}$) of the possible Multiple-group grouping scheme discriminant canonical correlation analysis yields grouping scheme 14 as the best grouping scheme as shown in Table 24 compared 60% ($\frac{9}{15}$) in the original data.

Figure 22 depicts the first nine series plots of the Time-independent FDCC values from Two-group to Ten-group ranging in size from GS 1 to GS 20. The figure confirms that all the FDCC values are no better than the overall value of 0.714 given in Table 21.

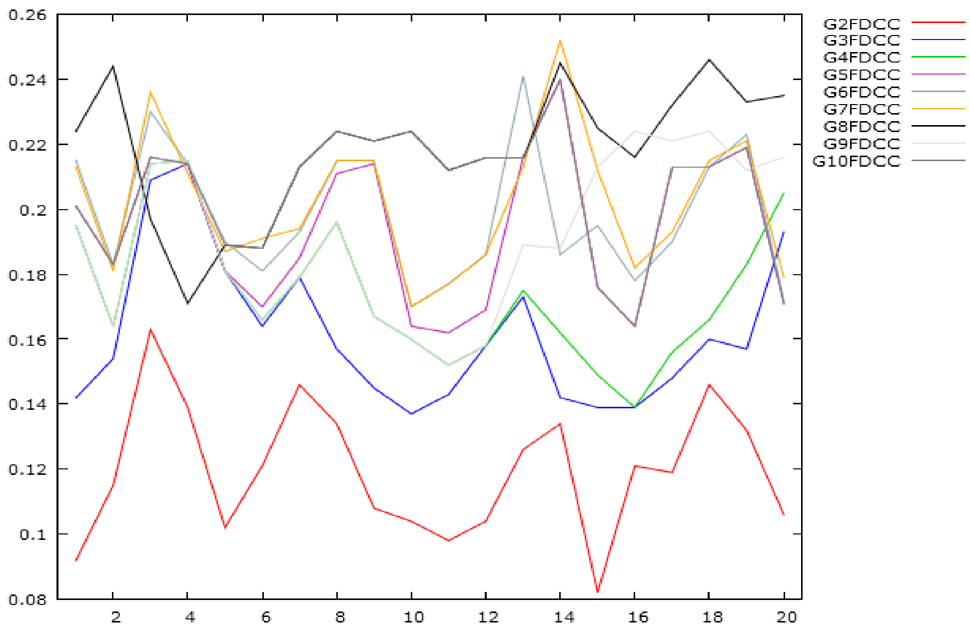

Figure 22: Plots of Time-Ind. Multiple-group FDCCs from 2 to 10-group

Simulation Studies of Time-Dependent Data

Using the simulated data thus obtained, the grouping scheme discriminant canonical correlation analysis will be carried out to establish the relevance of the concept of grouping scheme proposed in this thesis. This part of the simulation would then assume a Time-Dependent Data that incorporates the effect of the year based on the best Seven-group grouping scheme discriminant canonical correlation analysis identified at the end of the first part of this chapter.

Time-dep. simulated data summary results for DCCA functions

Table 25 shows the overall Time-dependent simulated canonical correlation coefficient (CCC) value of 0.8861, the overall Wilks Lambda value of 0.2148, and the overall Eigenvalue of 3.6546. The time-dependent simulated data is used to generate Two-group up to Sixteen-group GS discriminant canonical correlation analysis to check if the results of the first discriminant canonical correlation (FDCC) values and Eigenvalues are higher than the overall results of 0.8861 and 3.6546, respectively, and the Wilks Lambda values are also less than the overall result of 0.2148. It is also used to confirm the general results in the original data.

Table 25: **Overall Tests of Significance for Time-Dependent Data**

Test Statistic	CCC	Eig V	Wilks	Roys	Hotellings	Pillais
Value	0.8861	3.6546	0.2148	0.7852	3.6546	0.7852
Significance	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Source: Researcher's computations (2023)

The series plots of the Time-dependent monthly simulated weather conditions data are shown in Figure 23. All the six variables especially, the two temperature variables are seen to follow the pattern in the original data as expected. The data exhibit random changes with noticeable patterns. The magnitude of seasonal fluctuations do not remain the same, although the data values turn to rise over time. The Time-dependent monthly simulated weather conditions data, therefore, mimic the original data as expected.

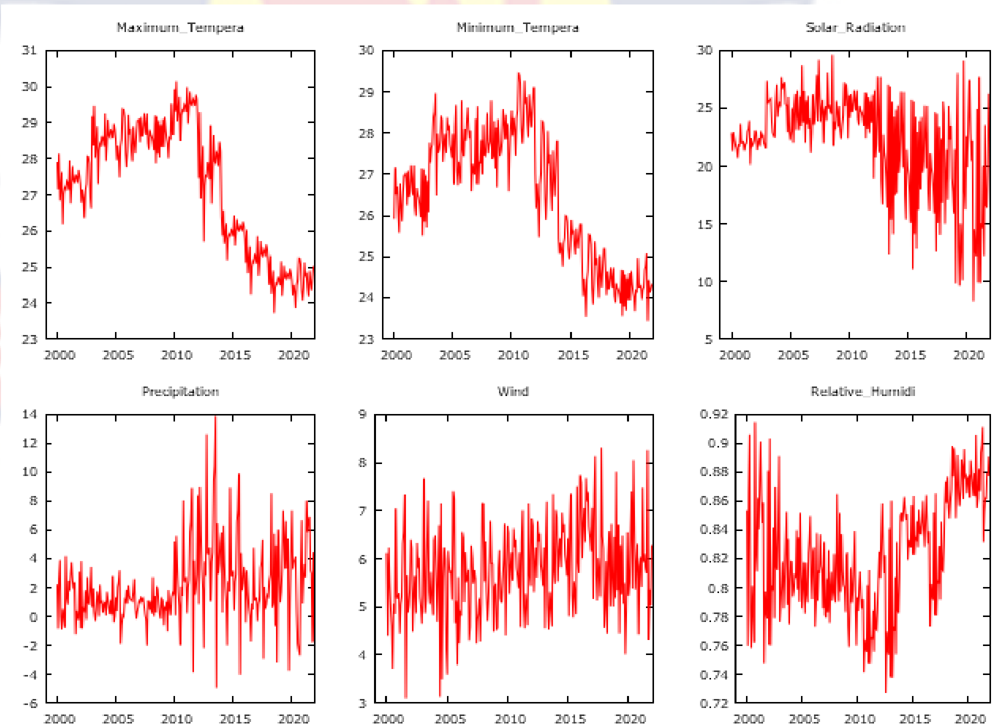

Figure 23: Series Plots of data on six monthly Time-dependent simulation

Table 26 reports the summary statistics of the response and predictor variables for the Time-dependent simulation data, based on the Seven-group discriminant canonical correlation analysis with best grouping scheme of 14.

Table 26: **Parameters for Time-Dependent Simulation Data**

DV	Sum Stats	Max T	Min T	Solar R	Precip	Wind	Rel H
	n_1	168	168	168	168	168	168
	d	28.39	27.47	23.65	1.56	5.59	0.81
1	T_1	0.84	0.89	2.84	2.51	0.96	0.03
	Min V	26.53	25.09	5.29	0.00	3.06	0.63
	Max V	29.94	29.52	26.35	15.96	7.82	0.89
	n_2	12	12	12	12	12	12
	d	26.23	25.64	19.10	2.56	6.07	0.85
2	T_2	0.24	0.28	6.21	3.19	0.77	0.01
	Min V	25.80	25.02	7.37	0.00	4.43	0.82
	Max V	26.66	25.96	24.05	11.00	7.31	0.87
	n_3	12	12	12	12	12	12
	d	25.86	25.39	20.07	0.51	5.93	0.84
3	T_3	0.24	0.29	4.01	3.62	0.84	0.01
	Min V	25.59	25.00	10.86	0.00	4.50	0.83
	Max V	26.36	26.00	23.85	11.05	7.28	0.86
	n_4	12	12	12	12	12	12
	d	25.53	25.01	20.21	0.42	6.88	0.81
4	T_4	0.37	0.56	4.35	0.61	0.88	0.03
	Min V	24.82	23.86	13.28	0.00	5.61	0.77
	Max V	25.95	25.63	24.47	1.76	8.15	0.85
	n_5	12	12	12	12	12	12
	d	25.16	24.50	18.95	1.49	6.23	0.83
5	T_5	0.32	0.23	4.18	2.69	0.87	0.02
	Min V	24.64	24.15	11.26	0.00	4.60	0.79
	Max V	25.64	25.04	25.02	7.90	7.46	0.86
	n_6	12	12	12	12	12	12
	d	24.69	24.28	17.63	1.42	6.28	0.86
6	T_6	0.25	0.22	5.14	1.84	0.57	0.01
	Min V	24.40	23.86	8.34	0.00	5.05	0.84
	Max V	25.17	24.66	24.59	5.72	7.18	0.87
	n_7	36	36	36	36	36	36
	d	24.71	24.19	18.58	2.46	5.75	0.88
7	T_7	0.37	0.44	5.83	3.66	0.84	0.02
	Min V	23.87	23.16	1.58	0.00	3.70	0.83
	Max V	25.31	24.97	26.05	13.41	7.10	0.90

Source: Researcher's computations (2023)

Simulation of Time-dependent multiple-group GS DCCA functions

In this section, I use the best GS 14 of the optimal Seven-group DCCA to build a simulation that identifies the best GS for all possible Time-dependent Multiple-group DCCA. Through the use of datasets for simulated weather conditions, I assess the effectiveness of the suggested approach. My approach

Table 27: **Statistic Measures of 2-Group GS DCCAFs for Time-Dep.**

GS	FirstDCC	C Class	Eig V	W Lambda	Chi-Square
1	0.150	68.9	0.023	0.977	5.919
2	0.288	65.2	0.037	0.965	9.339
3	0.209	67.0	0.046	0.956	11.537
4	0.230	63.6	0.056	0.947	14.064
5	0.321	64.4	0.115	0.897	28.147
6	0.379	67.0	0.168	0.856	40.127
7	0.453	69.3	0.258	0.795	59.454
8	0.528	74.6	0.387	0.721	84.707
9	0.598	79.9	0.557	0.642	114.657
10	0.676	85.2	0.843	0.543	158.317
11	0.748	90.9	1.272	0.440	212.530
12	0.848	94.3	2.552	0.282	328.259
13	0.857	95.5	2.776	0.265	344.119
14	**0.903**	**99.6**	**4.413**	**0.185**	**437.418**
15	0.860	95.1	2.839	0.261	348.388
16	0.833	90.9	2.275	0.305	307.230
17	0.764	86.7	1.401	0.416	226.888
18	0.740	89.0	1.209	0.453	205.288
19	0.626	85.2	0.645	0.608	128.973
20	0.496	82.2	0.327	0.754	73.221

Source: Researcher's computations (2023)

and findings are compared with those of the original weather circumstances discriminative CCA. I run several Multiple-group DCCA based on the Time-dependent simulation eventhough the underlying structure is Seven-group. First, Two-group DCCA is carried out for GS 1 through GS 20 to check which one has the best statistic values among all the schemes for the five statistic measures. As shown in Table 27, GS 14 has the best statistic values among all the schemes, which falls on the 168^{th} month, at the end of 2013. That is, assuming the data may be suitably generated into two, the partition that is provided with month 168 as the partition line provides the best correlation between the two subsets weather conditions. This partition is given by the 14^{th} GS. It is worth noting that the highest DCC obtained in this case is 0.903 which is slightly higher than the overall value of 0.886 given in Table 25.

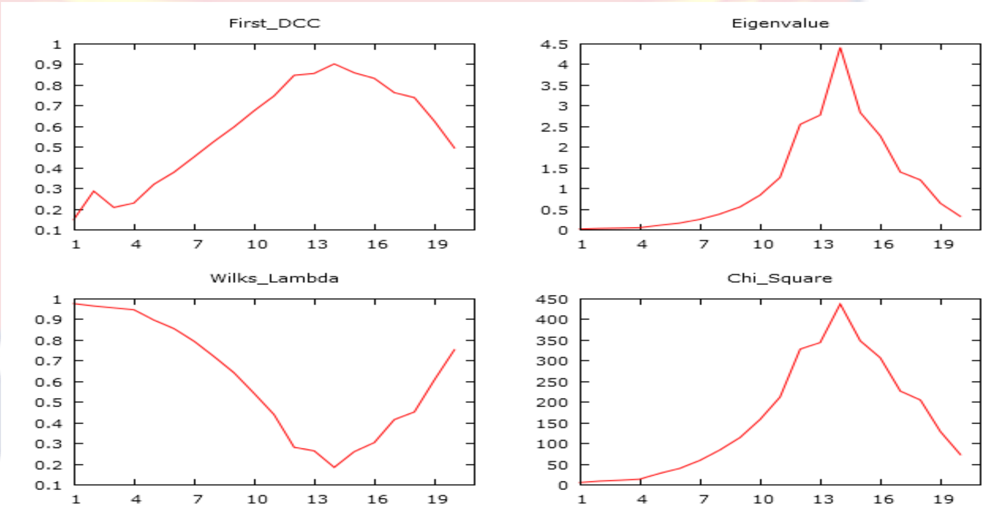

Figure 24: Series Plots of Time-Dep. Statistic Measures of 2-group DCCA

Figure 24 shows the plots of the statistics for the Two-group DCCA of the Time-dependent simulation, confirming that the optimal values are all in the GS 14. Hence, GS 14 is the best GS so far as Two-group discrimination is concerned. The overall CCC corresponds to (DCC = 0.903) of the Two-grouping schemes. Several functions are then performed from Three-group up to Sixteen-group GS-DCCA. This section's primary goal is to determine whether all the five statistics measures in a given Multiple-group are greater than or equal to the statistics of Two-group's 14^{th} GS values in Table 27.

Table 28 reports the statistics values of the Three-group discriminant canonical correlation analysis of Time-dependent data. Here, GS 3 has the most optimal statistics and fall on the 36^{th} month, at the end of 2002. However, the 17^{th} GS reports the highest correct classification value. All the FDCC values are greater than the overall CCC of 0.8861, all the Wilks Lambda values are less than the overall value of 0.2148 and the respective Eigenvalues are also greater than the overall Eigenvalue of 3.6548 as expected.

Table 28: **Statistic measures of 3-Group GS DCCAFs for Time-Dep.**

GS	FDCC	C Class	Eig V	W Lambda	Chi-Square
1	0.913	86.4	4.990	0.163	468.578
2	0.921	89.0	5.600	0.146	496.817
3	**0.937**	92.4	**7.181**	**0.117**	**554.647**
4	0.927	88.6	6.112	0.135	517.398
5	0.919	84.1	5.427	0.146	497.772
6	0.917	83.0	5.280	0.147	495.932
7	0.913	81.4	5.029	0.149	492.795
8	0.909	81.4	4.765	0.150	491.194
9	0.909	86.4	4.732	0.143	502.245
10	0.906	91.3	4.588	0.137	514.516
11	0.903	93.2	4.415	0.147	496.020
12	0.913	91.7	5.005	0.138	511.076
13	0.905	93.9	4.534	0.159	476.108
14	0.905	89.8	4.522	0.176	448.823
15	0.908	88.6	4.720	0.169	459.938
16	0.913	90.5	4.979	0.155	481.700
17	0.918	**95.8**	5.395	0.137	514.268
18	0.915	91.7	5.148	0.155	482.085
19	0.919	92.8	5.419	0.153	484.460
20	0.908	89.0	4.722	0.172	454.867

Source: Researcher's computations (2023)

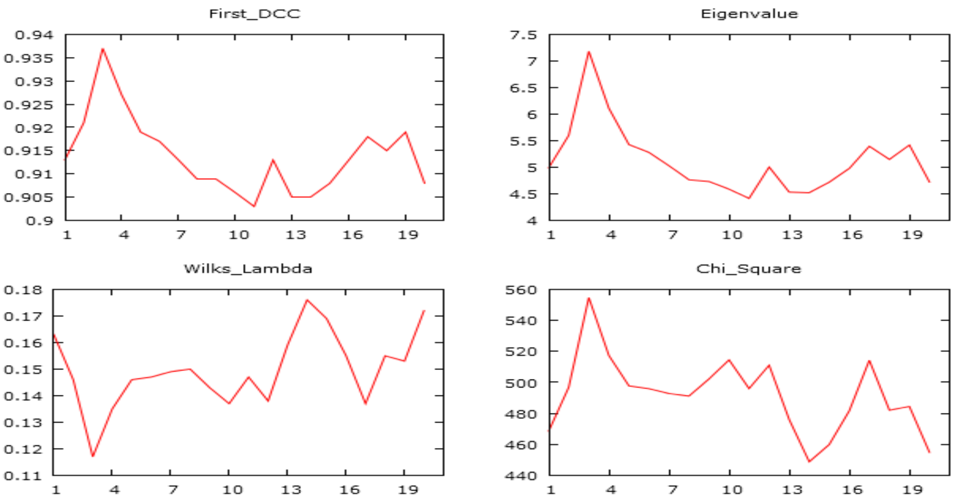

Figure 25: Series Plots of Time-Dep. Statistic Measures of 3-group DCCA

Figure 25 depicts the plots of the statistics measures in Table 28 confirming that the four optimal statistic measures are in the grouping scheme 3.

Table 29: **Statistic measures of 4-Group GS DCCAFs for Time-Dep.**

GS	FDCC	C Class	Eig V	W Lambda	Chi-Square
1	0.918	79.5	5.391	0.151	488.306
2	0.937	86.7	7.215	0.116	555.980
3	**0.950**	88.3	**9.317**	**0.089**	**625.480**
4	0.944	88.3	8.172	0.092	616.025
5	0.929	78.8	6.330	0.119	549.896
6	0.923	72.0	5.741	0.132	522.244
7	0.915	70.8	5.168	0.141	504.804
8	0.915	73.9	5.134	0.138	510.165
9	0.925	79.2	5.951	0.116	555.349
10	0.919	83.7	5.449	0.113	563.436
11	0.919	**89.4**	5.396	0.108	573.628
12	0.922	83.3	5.664	0.116	556.793
13	0.910	83.7	4.847	0.146	498.592
14	0.915	81.4	5.148	0.132	523.197
15	0.908	82.6	4.721	0.134	518.568
16	0.916	84.5	5.234	0.114	561.203
17	0.926	83.3	6.020	0.102	588.594
18	0.923	75.8	5.765	0.122	543.445
19	0.930	75.8	6.389	0.120	547.911
20	0.923	75.4	5.723	0.135	516.458

Source: Researcher's computations (2023)

Table 29 reports the relevant Four-group grouping scheme discriminant canonical correlation analysis of The Time-dependent simulated data. In this case, all the first discriminant canonical correlation values are greater than the overall canonical correlation coefficient of 0.886, the Wilks Lambda values are less than the overall value of 0.215 and the respective Eigenvalues are also greater than the overall Eigenvalue of 3.655 as shown in Table 25. It is observed that grouping scheme 3 has the most optimal discrimination variables but the grouping scheme 11 reports the highest correct classification.

Figure 26 depicts the plots of the relevant statistics for Four-group First DCCA in Table 29 confirming that the most optimal values are all in grouping scheme 3, the 36^{th} month, at the end of 2002.

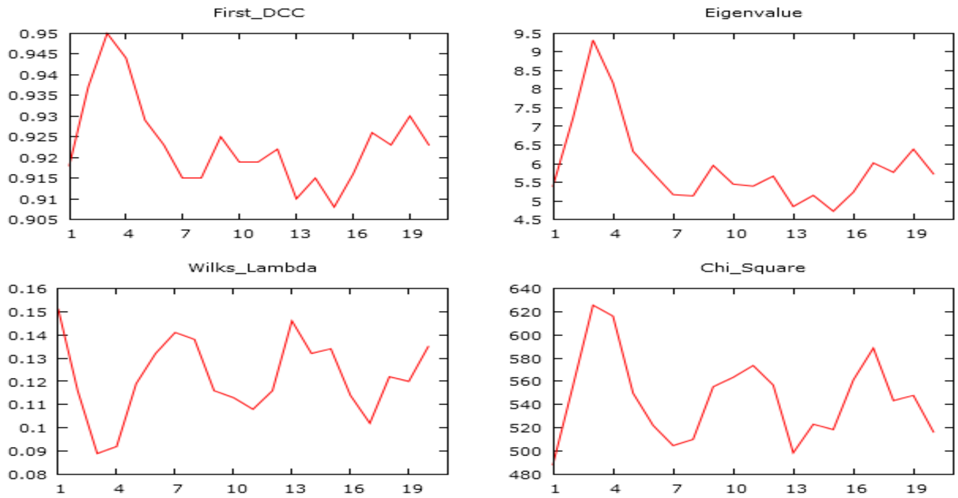

Figure 26: Series Plots of Time-Dep. Statistci Measures of 4-group DCCA

Table 30: **Statistic measures of 5-Group DCCAFs for Time-Dep.**

GS	FDCC	C Class	Eig V	W Lambda	Chi-Square
1	0.921	76.5	5.617	0.128	529.663
2	0.950	82.2	9.341	0.077	661.443
3	0.952	83.3	9.737	0.069	689.721
4	0.944	83.7	8.195	0.072	677.964
5	0.930	72.7	6.355	0.103	585.040
6	0.923	63.6	5.741	0.123	539.037
7	0.921	63.3	5.604	0.132	521.547
8	0.932	68.9	6.590	0.109	570.102
9	0.947	79.2	8.773	0.076	662.331
10	0.940	84.5	7.567	0.076	661.982
11	0.925	77.3	5.961	0.107	576.207
12	0.949	86.7	9.087	0.079	654.408
13	0.948	86.0	8.784	0.085	634.728
14	**0.968**	**90.9**	**14.959**	**0.050**	**773.702**
15	0.933	74.6	6.674	0.097	600.796
16	0.934	78.4	6.884	0.084	637.256
17	0.937	76.1	7.232	0.092	615.178
18	0.933	68.2	6.697	0.108	573.590
19	0.939	67.8	7.461	0.101	589.943
20	0.929	67.4	6.286	0.113	561.583

Source: Researcher's computations (2023)

Table 30 reports the statistics of the Time-dependent simulation Five-group grouping scheme discriminant canonical correlation analysis. From the table, all the statistic measures for each grouping scheme are better than those of the overall statistic measures given in Table 25. Grouping scheme 14 has the most optimal statistic measures over all the five measures and are achieved on the 168^{th} month, at the end of 2013, which is the same as obtained in Table 27 for Two-group grouping scheme discriminant canonical correlation analysis. Figure 27 depicts the series plots of the results given in Table 30.

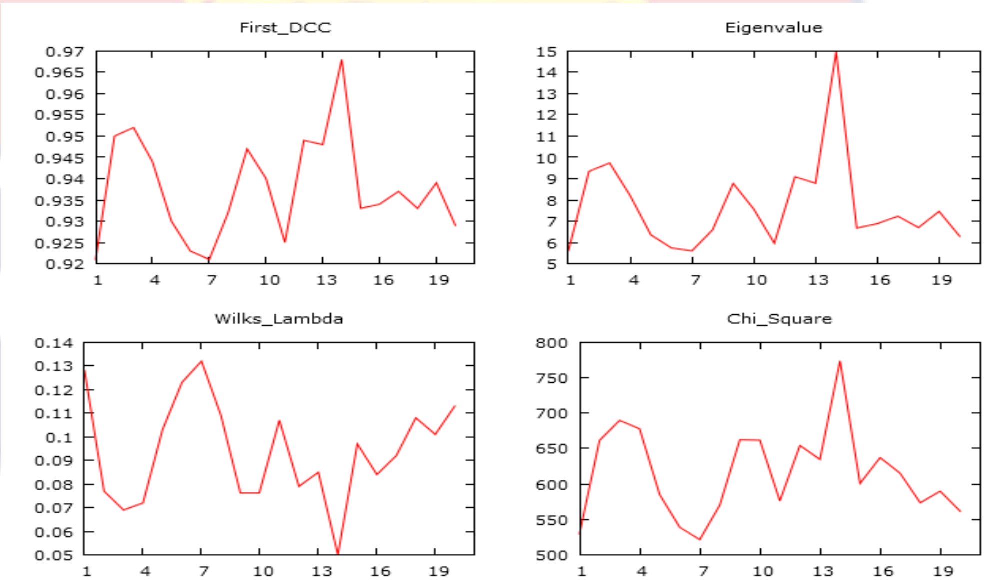

Figure 27: Series Plots of Time-Dep. Statistic Measures of 5-group DCCA

Table 31 shows the classification of Time-dependent simulation Six-group grouping scheme for five relevant statistics. All of the values in the table are more optimal than the overall statistic measures. However, correct classification value reduces for all grouping schemes due to increased group-discrimination. The grouping scheme 12 reports the most optimal statistics measures which are achieved on the 144^{th} month, at the end of 2011.

Figure 28 depicts the series plots of the statistic measures for the Time-dependent simulation Six-group discriminant canonical correlation analysis confirming that the most optimal statistic measures are all in the GS 12.

Table 31: **Statistic measures of 6-Group GS DCCAFs for Time-Dep.**

GS	FDCC	C Class	Eig V	W Lambda	Chi-Square
1	0.929	65.5	6.351	0.107	573.393
2	0.958	75.0	11.043	0.064	707.006
3	0.953	70.1	10.006	0.066	698.996
4	0.944	74.2	8.221	0.071	679.122
5	0.930	66.7	6.367	0.102	585.540
6	0.923	61.4	5.744	0.122	540.873
7	0.921	61.4	5.620	0.126	531.743
8	0.932	62.9	6.613	0.108	570.882
9	0.948	68.9	8.784	0.076	662.382
10	0.943	73.5	8.054	0.072	675.231
11	0.937	64.4	7.188	0.090	620.004
12	**0.958**	**75.0**	**11.092**	**0.064**	**707.630**
13	0.948	72.7	8.861	0.078	655.120
14	0.953	73.9	9.899	0.070	682.540
15	0.935	67.4	6.970	0.092	613.935
16	0.934	74.2	6.885	0.084	637.557
17	0.930	66.7	6.367	0.102	585.540
18	0.933	68.2	6.697	0.108	573.590
19	0.939	60.2	7.464	0.101	589.665
20	0.931	60.2	6.463	0.110	568.297

Source: Researcher's computations (2023)

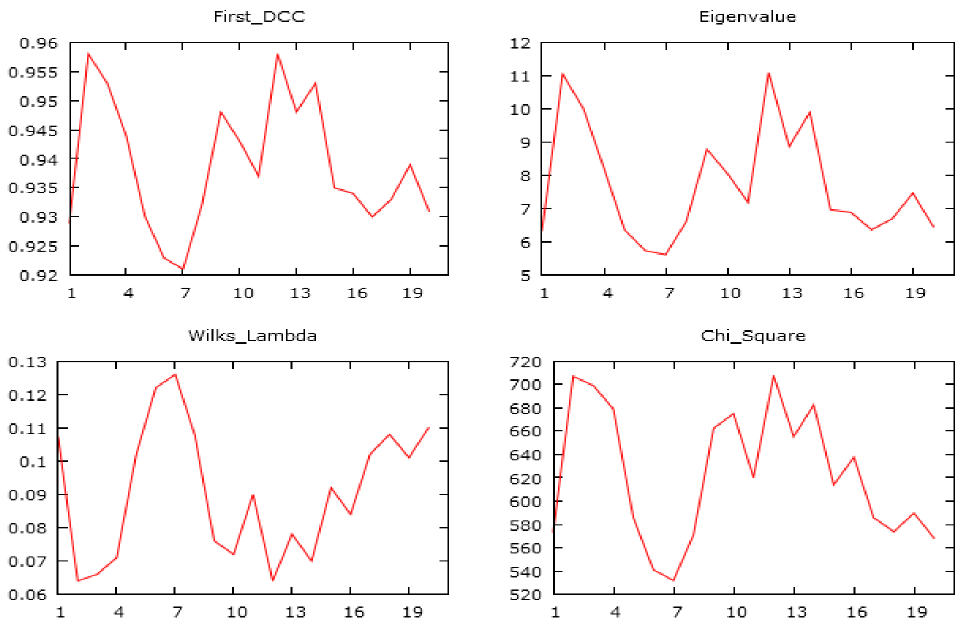

Figure 28: Series Plots of Time-Dep. Statistic Measures of 6-group DCCA

Time-dependent simulation of Seven-group DCCAFs results

Table 32 depicts the statistics of Seven-group DCCA functions of the Time-dependent simulation. From the table, for each grouping scheme, measures are more optimal than corresponding values of the overall statistic measures except for the correct classification values that decrease as a result of increased group discrimination. The GS 14 reports the most optimal values for all five measures which also fall on the 168^{th} month, at the end of 2013.

Table 32: **Statistic Measures of 7-Group GS DCCAFs for Time-Dep.**

GS	FDCC	C Class	Eig V	W Lambda	Chi-Square
1	0.935	61.0	6.971	0.092	611.413
2	0.956	62.9	10.557	0.063	708.341
3	0.957	66.7	10.944	0.055	743.455
4	0.937	64.0	7.236	0.084	634.443
5	0.937	59.1	7.180	0.094	607.818
6	0.927	61.4	6.140	0.110	566.127
7	0.925	59.1	5.956	0.118	548.747
8	0.932	59.8	6.617	0.107	573.304
9	0.948	62.9	8.786	0.076	662.627
10	0.943	64.8	8.055	0.072	676.383
11	0.939	57.6	7.410	0.083	638.225
12	0.959	65.9	11.417	0.053	725.137
13	0.942	64.8	7.843	0.083	638.118
14	**0.980**	**87.5**	**24.350**	**0.022**	**975.570**
15	0.937	60.2	7.192	0.085	632.267
16	0.934	65.9	6.886	0.083	638.787
17	0.930	61.0	6.368	0.102	585.923
18	0.937	67.0	7.222	0.096	602.695
19	0.948	62.9	8.850	0.076	661.376
20	0.939	56.4	7.450	0.094	607.101

Source: Researcher's computations (2023)

Figure 29 depicts the plots of Time-dependent simulation Seven-group DCCA confirming the most optimal values for all five statistic measures of discriminant canonical correlation analysis. It is amazing that up till now, besides the Two-group grouping scheme discriminant canonical correlation analysis, it is the Five-group, Six-group and Seven-group grouping scheme discriminant canonical correlation analysis that produce the most optimal statistic values over all five measures for the same best grouping scheme.

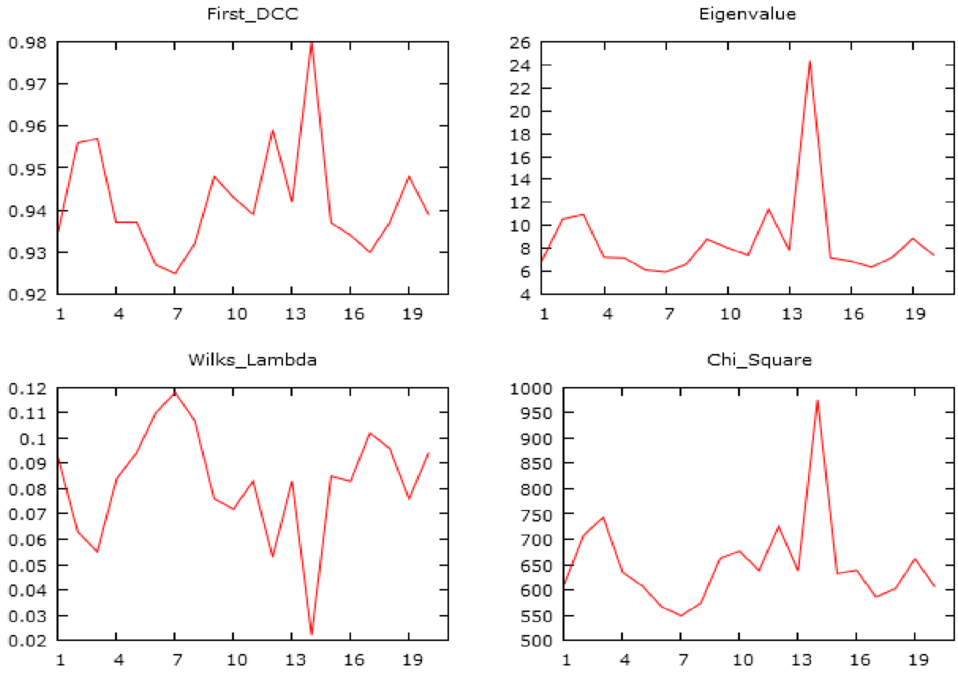

Figure 29: Series Plots of Time-Dep. Statistic Measures of 7-group DCCA

Summary Statistic Measures for all Best Time-Dep. *M*-Groups DCCA

Table 33 gives the summary of all the highest scores of the grouping scheme classifications of all the Multiple-groups from Two-group to Sixteen-group of the Time-dependent simulated data, reporting the five statistic measures and their respective grouping schemes. Since the best grouping scheme may not report the highest correct classification, it is not provided in the table.

The table confirms that 7-group GS discriminant canonical correlation analysis outperforms all the other multiple groupings, from 3-group to 16-group. Again, out of all the Multiple-groups, it is GS 14 of 7-group discriminant canonical correlation analysis that gives the most optimal statistics for all measures. In this section, 67% ($\frac{10}{15}$) of the possible Multiple-group discriminant analysis yields grouping scheme 14 as the best grouping scheme. The results of the study subsequently confirm the earlier claim that when the time impact is taken into account in the CCA, the correlation between the subsets of heating and cooling variables improves from 0.886 to 0.980.

Table 33: **Summary Statistic Measures for all Best Time-Dep. *m*-Groups**

Group	FDCC	Eig V	W Lamb	Chi-Sq	GS	Month	Year
2	**0.903**	**4.413**	**0.185**	**437.42**	**14**	**168^{th}**	**2013**
3	0.937	7.181	0.117	554.65	3	36^{th}	2002
4	0.950	9.317	0.089	625.48	3	36^{th}	2002
5	0.968	14.959	0.050	773.70	14	168^{th}	2013
6	0.958	11.092	0.064	707.63	12	144^{th}	2011
7	**0.980**	**24.350**	**0.022**	**975.57**	**14**	**168^{th}**	**2013**
8	0.966	14.101	0.046	787.56	14	168^{th}	2013
9	0.971	16.291	0.039	826.67	14	168^{th}	2013
10	0.960	11.638	0.050	765.34	5	60^{th}	2004
11	0.961	12.154	0.049	767.57	14	168^{th}	2013
12	0.954	10.027	0.058	723.09	14	168^{th}	2013
13	0.953	9.903	0.057	725.12	14	168^{th}	2013
14	0.967	14.363	0.036	837.06	3	36^{th}	2002
15	0.956	10.577	0.048	769.37	14	168^{th}	2013
16	0.962	12.530	0.039	812.25	14	168^{th}	2013

Source: Researcher's computations (2023)

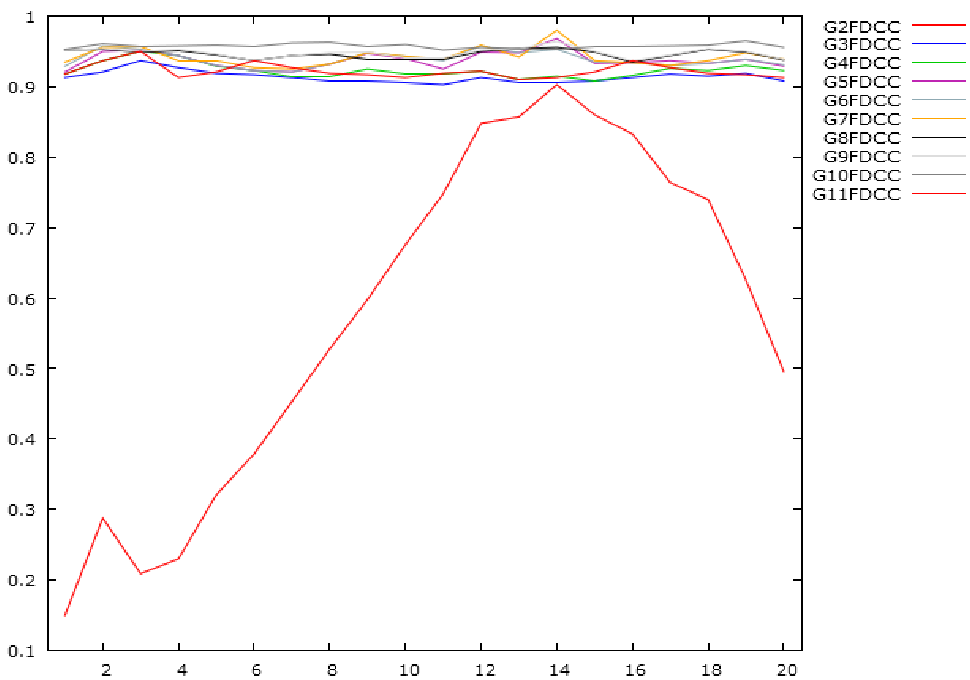

Figure 30: Plots of Time-Dep. Multiple-group FDCCs from 2 to 11-groups

Figure 30 depicts the first ten series plots of the Time-dependent data from Two-group to Eleven-group first discriminant canonical correlation values from grouping scheme 1 to grouping scheme 20. The figure confirms that all the first discriminant canonical correlation values of Time-dependent Multiple-group are higher than the highest Two-group discriminant canonical correlation value of 0.903 in grouping scheme 14 in Table 27 as well as the overall value of 0.886 also in Table 25.

Summary Statistic measures of all optimal values of 7-group DCCA

Table 34 reports the summary statistics results of optimal values of Seven-group discriminant canonical correlation analysis for the four statistical Measures all in grouping scheme 14. The table confirms that Time-dependent simulation of Seven-group grouping scheme discriminant canonical correlation analysis outperforms the original data and the Time-independent simulation so far as the statistical measures are concerned.

Table 34: **Summary Statistic Measures of Optimal 7-Group DCCA**

Stats Measures	Original Data	Time-Ind. Sim	Time-Dep. Sim
First DCC	0.968	0.252	0.980
Eigenvalue	14.928	0.065	24.350
Wilks Lambda	0.032	0.875	0.022
Chi-Square	881.004	34.106	975.570
Optimal Percent.	60%	40%	67%

Source: Researcher's computations (2023)

Again, out of all the Multiple-groups, it is grouping scheme 14 of Seven-group discriminant canonical correlation analysis that gives the most optimal statistics for all four measures. The results therefore shows that incorporating the time-effect into canonical correlation analysis achieves the time relationship between subsets variables within the data. It follows from Table 35 that Time-Dependent canonical correlation coefficient for roots one, two and three values of 0.774, 0.353 and 0.181, respectively, using the fourteenth grouping scheme of Seven-group discriminant canonical correlation analysis's dummy variables, outperforms the original data canonical correlation coefficient values of 0.731, 0.345 and 0.175, respectively.

Table 35: **Time-Dependent CCCs Using GS 14 of 7-Group DCCA**

CCC	Λ_1	Λ_2	Λ_3	Λ_4	Λ_5	Λ_6	Λ_7	Average
ρ_1	0.66	0.70	0.83	0.84	0.92	0.64	0.84	0.774
ρ_2	0.50	0.27	0.30	0.33	0.35	0.25	0.48	0.353
ρ_3	0.14	0.14	0.16	0.22	0.14	0.17	0.31	0.181

Source: Researcher's computations (2023)

The original canonical correlation coefficient matrix is given by

$$CCC = \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 0.731 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0.345 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0.175 \\ \hline 0.731 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0.345 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0.175 & 0 & 0 & 1 \end{array}\right)$$

The Time-dependent canonical correlation coefficient matrix is also given by

$$CCC = \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 0.774 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0.353 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0.181 \\ \hline 0.774 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0.353 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0.181 & 0 & 0 & 1 \end{array}\right)$$

Discussion

Canonical correlation was used in this thesis to show that it is an appropriate technique for evaluating correlations between sets of data. Using a grouping scheme discriminant canonical correlation analysis, the relationship between the heating and cooling variables is examined. Canonical variates in canonical correlation analysis are generated in a method that maximizes the correlation between each set of variables. There exists no correlation between any two sets of canonical variates, for example, $(\alpha' Y_2, \beta' X_2)$ is uncorrelated with $(\alpha' Y_1, \beta' X_1)$. A full discussion of two-group discriminant analysis is given. The method of discriminant analysis involves first determining which discriminant variables, also referred to as the best collection of variables, offer the greatest discrimination between the heating and the cooling variables. The discriminator variables are then combined linearly to create an estimated discriminant function. The values derived from the discriminant function are known as discriminant scores. In order to determine the discriminant scores with the biggest feasible ratio of between-groups sum of squares to within-groups sum of squares, the discriminant function is assessed. Sorting incoming data into one of the two groups according to the values of their discriminant scores is the ultimate goal of discriminant analysis.

This thesis also investigates multiple-group discriminant analysis, which is a generalization of two-group discrimination analysis, and the connection between canonical correlation analysis and discriminant analysis. It turned out that, in terms of geometry, multiple-group discriminant analysis reduced to selecting a new set of axes in order to best describe the major differences between the groups by projecting the points onto the original axis. The projection of the points onto the second axis described the maximum of what the first axis could not explain, and so on, until the axes, or $min(G-1, p)$, were identified. Our results show that from Group-seven up to Group-twenty, only first six canonical discriminant functions for each Multiple-group are used in the analysis. It is noteworthy that, at this point, aside from the Two-group grouping scheme discriminant canonical correlation analysis, it is the Four-group, Five-group and Seven-group grouping scheme discriminant canonical correlation analysis that produce the most optimal statistics values over all five measures for the same best grouping scheme.

In this thesis, I suggest a generalized form of the grouping scheme discriminant canonical correlation analysis. According to the nature of the objective function, a set of discriminant analysis parameters is provided. When choosing these parameters, cross validation is taken into account by comparing the estimated additive components. The association between heating variables and cooling variables is examined using a groping scheme-based test that is presented. The suggested method can successfully discover nonlinear relationships between heating factors and cooling variables, according to a simulation analysis, this reveals the relative weight of each variable in the groups as well. These benefits will be beneficial in a variety of research fields involving multivariate data. The suggested approach might not be able to handle cases where there are interactions between many variables within each group because of the additivity assumption.

The grouping scheme discriminant canonical correlation analysis test requires more calculation time than the traditional canonical correlation analysis, which utilizes a straightforward test statistic like canonical correlation coefficient, Wilks' lambda, and Chi-square. Distributed computing, on the other hand, effectively reduces the compute load. On the other hand, intensive computation is unavoidable for choosing the discriminant parameters in the grouping scheme discriminant canonical correlation analysis. Therefore, it is worthwhile to look into creating an algorithm to speed up processing or discovering a computationally more efficient technique of selection. The classical canonical correlation analysis can consider the second canonical variates that maximize the correlation $Corr\left(\alpha' Y_1, \beta' X_1\right)$ among all alternatives that are uncorrelated with the first canonical variates. Although canonical correlation analysis makes this difficult, it is nevertheless worth looking into for future research because it might disclose extra structural information about groups that the canonical correlation analysis model is insufficient to account for.

Chapter Summary

In the first place, the findings merely demonstrated a basic model fit, and several sections in-depthly evaluated the significance of each canonical root. The first canonical root performs better than the second and third, but all three possible canonical roots are shown to be statistically significant. Canonical weights and loadings from canonical correlation analysis demonstrate that, for all three roots, the contribution of precipitation to variance in maximum temperature, lowest temperature, and solar radiation is greater than that of the wind and the relative humidity. The three canonical correlation coefficients are discovered to not equal zero, rejecting the null hypothesis. The first canonical root is used to look at the correlation pattern between the response variables and the other types of weather conditions. Furthermore, the findings demonstrate that throughout the investigation, there was a very strong positive correlation between the response variables and the predictor factors.

Second, the theoretical underpinnings are examined and the link between canonical correlation analysis and discriminant analysis are presented. The canonical correlation analysis problem is created using the Pearson correlation coefficient, and it is then solved using the popular techniques of Eigen-Value Decomposition and Singular-Value Decomposition. Then, sixteen groups of the Discriminant canonical correlation analysis family algorithms are created, ranging in size from Three-group to Sixteen-group. The representative models are presented for each grouping using simulation methods, and analytical comparisons are used to summarize their strengths and flaws. The Seventh group discriminant canonical correlation analysis outperforms all other top Multiple-groups, achieving the greatest first discriminant canonical correlation value, the highest eigenvalue and the highest Chi-square value with corresponding lowest Wilks Lambda value. The normal canonical correlation of simulated data gives very low results. All the Multiple Discriminant Analysis yield a non significant results. It follows that canonical correlation analysis is meaningful when variance covariance matrix is well defined. The data in each of the variables should have quite distinct variations. If this fails, canonical correlation analysis may fail but the corresponding DA can work.

The non-significant results of the simulation is found to be as a results of assuming a Time-independent structure. This case arises by simulating the data with statistics of the original data without reference to time (year-effect). The statistics of partitioned data based on the best grouping scheme of 14 for optimal Multiple-group discriminant analysis of Seven (7) is used to obtain a more representative simulated data that closely mimics the original data. A number of grouping scheme discriminant canonical correlation analysis demonstrate that the optimal relationship between the two subset variables of the Time-dependent data is established by the 14^{th} grouping scheme of the Seven-group discriminant canonical correlation analysis. In light of the findings, grouping scheme techniques can be used for canonical correlation of Time-dependent data.

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

Overview

The summary, conclusions, and recommendations based on the study's findings are presented in this chapter. The purpose of this thesis is to examine the application of grouping scheme discriminant canonical correlation analysis (GSDCCA) of multivariate multiple Time-dependent data (MMTDD) structure. Finally, based on the results and findings, an appropriate conclusion and recommendation of the link between CCA and discriminant analysis are provided.

Summary

It is challenging to understand the concept of CCA because of how it is presented, which appears complicated. This might be because the method is mathematically intensive. As a result, it is necessary to conduct a study that provides a CCA in a rational and approachable manner. Therefore, the study's objectives are to provide simple procedures for producing canonical variables using generated codes and to explicitly state the logical justification for the results. It has facilitated the application of CCA to multivariate multiple time-dependent data structure.

The sort of data structure required for a multivariate linear connection is examined in the study. It is necessary to build such a dataset for the same individuals using a multivariate random vector that may be appropriately divided into two sub-vectors and whose components may have a linear connection with one another. A typical set of 264 observations on weather conditions in Ghana has been described in detail and considered pertinent for use in the study. The research anticipates and introduces a number of significant matrices that may be useful in CCA. The square root of the variance covariance matrix is a crucial matrix. It is discovered that in order to generate the required CCA, a critical matrix must generate predictor variables with zero means and unit variances. Important matrices used in the CCA are required to possess similarity and diagonalizability properties. CCA has reportedly been used for a while, and from the standpoint of data analysis, it is most frequently used in conjunction with other multivariate methods like factor analysis and principal component analysis. There are numerous methodologies used in the theoretical presentations of the ideas that can be found in renowned texts. In several of these methods, the salient matrices of the canonical extraction have been claimed without proof.

The theoretical properties of the canonical factors have been examined. It is observed that, the new variables should exhibit a few characteristics that broadly cover their independence to unit variance. By describing the multivariate multiple Time-Dependent in terms of the characteristics of the canonical variables, it is possible to express the canonical correlation matrices in terms of sums of matrices that incorporate diagonal matrices. On this foundation, it has been demonstrated that the matrix concatenation decomposition method may successfully recover the variation in the original subset of response variables that is explained by predictor canonical variables.

The body of work demonstrates how thorough the idea of CCA has been researched. A few research delved into great detail, and publications also included data analytic applications. In addition to the subject's concentration on mathematics, it is clear that many authors have used a variety of strategies in dealing with the CCA and DA techniques. The diverse methods frequently result in misconceptions about the concept of CCA. Some applications and extensions of DA and CCA have been reviewed. Several applications have been reviewed in dealing with canonical correlation analysis including the link between CCA and discriminant CCA. It is clear from the literature that scanty work is done regarding application of CCA in multivariate time-dependent data. Thus, this area still remains grey for further explanation.

The development of grouping scheme discriminant canonical correlation analysis (GSDCCA) and its application to data with time-dependent structure are the primary goals. The methodology has provided a thorough analysis of the canonical variable construction process. It has in the process discovered about six important matrices for creating canonical variables. In order to create the new variables, the review made it possible to determine the proper interpretation and uses of these matrices.

There have been two primary fundamental procedures looked at. These are the conditional distribution of Cauchy-Schwarz Inequality technique and the Lagrangian multiplier technique. These two fundamental approaches are thoroughly explained in this chapter. It has been noted that the outcomes of these two strategies are identical and generated similar results so far as the key matrices for CCA are concerned. The general methodology of various extensions of CCA have been so far made to the basic formulations. Those extensions are outlined and also explained into details using the two fundamental approaches. I looked at the canonical correlation variables' theoretical characteristics and described them in three ways. It can be shown that the new variables in each situation should have one of six essential characteristics, which, taken together, substantially cover their independence and unit variance. The methodology of the general formulation of grouping scheme discriminant CCA for Multiple-group discriminant analysis is discussed. The time-effect of the data is anticipated to enhance the genuine correlation between the two sets of variables by introducing the optimum grouping scheme into CCA.

In the first place, the findings merely demonstrated a basic model fit, and several sections in-depthly evaluated the significance of each canonical root. It is observed that all the three potential canonical roots are statistically significant but the first canonical root outperforms the second and third roots. Canonical weights and loadings from CCA demonstrate that, for all three roots, the contribution of precipitation to variance in maximum temperature, minimum temperature, and solar radiation is greater than that of the wind and the relative humidity. The first canonical root is used to look at the correlation pattern between the response variables and the other types of weather conditions. The results show that there is a very high positive association between the response variables and the predictor variables.

The theoretical underpinnings are examined and the link between CCA and DA are presented. The CCA problem is created using the Pearson correlation coefficient, and it is then solved using the popular techniques of Eigen-Value Decomposition and Singular-Value Decomposition. Then, sixteen groups of the Discriminant CCA family algorithms are created, ranging in size from Three-group to Sixteen-group. The representative models are presented for each grouping using simulation methods, and analytical comparisons are used to summarize their strengths and flaws. The Seventh group discriminant CCA outperforms all other top Multiple-groups, achieving the most optimal statistic values. The normal CCA of simulated data gives very low results. All the Multiple DA yield a non significant results. It follows that CCA is meaningful when variance covariance matrix is well defined. The data in each of the variables should have quite distinct variations. If this fails, CCA may fail but the corresponding DA can work.

The non-significant results of the simulation is found to be as a results of assuming a Time-independent structure. This case arises by simulating the data with statistics of the original data without reference to time (year-effect). The statistics of partitioned data based on the best GS of 14 for optimal Multiple-group DA of Seven is used to obtain a more representative simulated data that closely mimics the original data. A number of grouping scheme Multiple Discriminant CCA demonstrate that the optimal relationship between the two subset variables of the Time-dependent data is established by the 14^{th} GS of the Seven-group discriminant CCA. In light of the findings, GS techniques can be used for CCA of Time-dependent data.

Conclusions

In the thesis, the relationship between two groups of random vectors has been carefully examined. By removing the canonical variables from both sets, it has investigated the correlation between the two sets of variables. It gave the multivariate multiple Time-dependent model canonical factors in order to link one subset vector of response variables to another subset vector of predictor variables. The study of such a multivariate relationship makes use of a number of mathematical and statistical concepts. The Cauchy-Schwarz inequality, similar matrices, matrix diagonalization, spectral decomposition of symmetric matrices, vector orthogonalization, conditional distributions, multivariate least squares estimates, and matrix concatenation are some of the concepts covered in this procedure.

To fully understand extracting canonical variables, one must identify the crucial matrices that result in the required change of the original variables. The research also discovered generalized correlations between these significant matrices. The theoretical characteristics of the new canonical variables have been examined and explained using six main techniques that generally encompass the independence between the new variables and unit variance. It has been demonstrated that the inverse matrix must be a combination of the matrices that generate the required canonical variables.

The study determines appropriate dataset structure and partitioning, as well as the relevant matrices, that enable us to reach the intended theoretical result in order to achieve this. It has been proven that canonical variables can be extracted from normalized, centered, or unprocessed data. Knowledge of the pertinent matrices involved in extracting canonical variables and the identification of the required data structure have allowed for the creation of pertinent MATLAB codes that create canonical variables with the desired attributes. The research focused on multivariate multiple time-dependent data (MMTDD) and canonical correlation analysis. The general form of Multivariate multiple Time-dependent has been described in three distinct ways in order to describe the three different data structures. The data format for the multivariate multiple time-dependent applications does not appear to have been given consideration in the literature.

Even though the expanded CCA methods' benefits have been shown in individual research, a thorough comparison of CCA and DA methods is still lacking. Using a dataset of weather conditions, this study compares the existing CCA and MDA detection algorithms. Performance evaluation utilized real and results are validated by simulation. From the findings, Seven-group DA greatly enhances the determination of the true correlation between the two sets of variables with time-dependent structure. Additionally, for the combination method based on the conventional CCA and the multiple DA, Seven-group DA produced the best results.

It is observed that the adoption of grouping scheme in discriminant canonical correlation analysis incorporates quite effectively the time effect into the computation of the canonical correlation. This way, a more practical result is obtained. From the illustrative dataset, higher over all correlation coefficients are obtained for the two sets of variables when the time-dependent structure is considered than when the data is assumed to time-independent. In particular, correlations could be much higher between the two sets of variables for some years than others. The results therefore reflects the reality and thus provides justification for the technique adopted.

Recommendations

The research clearly demonstrates the implementation of the right procedures for multivariate multiple time-dependent data. For rapid extraction of time-dependent canonical variables from multivariate multiple time-dependent data, the provided procedures may be useful. The results of this work have demonstrated the usefulness of the proposed grouping scheme discriminant canonical correlation analysis. This will offer more explanation for future grouping scheme mechanisms and serve as a foundation for experimental validation and verification. It is demonstrated that the classical CCA may not be for multivariate multiple time-dependent data appropriate.

Future Work

For now, there is considerable interaction time with the data by the use of the proposed methodology. The grouping scheme discriminant CCA procedure therefore requires some enhancement in order to reduce implementation time. The approach could also serve as a fundamental step for obtaining what may be known as a Fundamental Multivariate Canonical Time Series Modelling (FMCTSM). By this, it should be possible to determine canonical correlations between the two sets of variables over time.

REFERENCES

Abdeldjalil, A. E. B. & Seghouane, A.-K. (2016). Sparse Canonical Correlation Analysis Based on Rank-1 Matrix Approximation and its Application for fMRI Signals. *41st IEEE International Conference on Acoustics*, Speech and Signal Processing, Shanghai, China, 4678 - 4682.

Akbas, Y. & Takman, C. (2005). Canonical correlation analysis for studying the relationship between egg production traits and body weight, egg weight and age at sexual maturity in layers. *Czech Journal of Animal Science.* 50, (4): 163–168.

Akour, I., AL-Rahamneh, A., Al-Kurdi, B., Alhamad, A., Al-Makhariz, I., Al-Shurideh, M., & Al-Hawary, S. (2023). Using the Canonical Correlation Analysis Method to Study Students' Levels in Face-to-Face and Online Education in Jordan. *Information Sciences Letters, An International Journal*. http://dx.doi.org/10.18576/isl/120229

Ali, G. (2019). *Probability, Random Variables, Statistics, and Random Processes*. Wiley and Sons.

Amit, M. & Sharmishtha, M. (1998). *Apllied Multivariate Analysis – Lecture* 40. National Programme on Technology Enhanced Learning. NP-TEL

Andrew, G., Arora, R., Bilmes, J., & Livescu, K. (2013). Deep canonical correlation analysis. *In Proc. Int. Comf. Mach. Learn*, 1247 – 1255.

Ankita, M., & Pradipta, M. (2023). Multiview regularized discriminant canonical correlation analysis: sequential extraction of relevant features from multiblock Data. *IEEE Transactions on Cybernetics*.

Apanyin, F. (2021). *Application of canonical correlation in multivariate multiple linear regression analysis*. Unpublished master's thesis. Dept. of Statistics, University of Cape Coast. Cape Coast. Ghana.

Arandjelovic, O. (2013). Discriminative extended canonical correlation analysis for pattern set matching. Centre for pattern recognition and data analytics (PRaDA).

Asoh, H. & Takechi, O. (1994). An approximation of nonlinear canonical correlation analysis by multilayer perceptrons. *In Proc. Int. Comf. Artif.* Neural Netw, 713 – 716.

Bach, F. R. & Jordan, M. I. (2005). *A probabilistic interpretation of canonical correlation analysis*. Dept. Statist., Univ. of California, Berkeley, CA, USA, Tech. Rep. 688.

Benton, A., Khayrallah, H., Biman, G., Dee Ann Reisinger, D. A., Sheng, Z., & Raman, A. (2019). *Deep generalized canonical correlation analysis*. NLP (RepL4NLP-2019), pages 1– 6, Florence, Italy, Association for Computational Linguistics.

Bernardo, B. G. & Eulanda, M. D. S. (2017). *Discriminative canonical correlation analysis network for image classification*. Institute of computing (ICOMP), Federal University of Amazonas, Brazil, Email: bernardo, emsantos@icomp.ufam.edu.br

Bickel, P. & Levina, E. N. (2008). Regularized Estimation of Large Covariance Matrices. *Annals of Statistics*, 36, 199-227. https://doi.org/10.1214/009053607000000758

Borga, M. (2001). *Canonical correlation analysis-a tutorial*. http://people.imt.line.se/'magnus/cca/.

Cankaya, S., Balkaya, A., & Karaagac, O. (2011). Canonical correlation analysis for the determination of relationships between plant characters and yield components in red pepper. [Capsicum annuum L. var. conoides (Mill.) Irish] genotypes. *Bulgarian Journal of Agricultural Science*.

Carroll, C. (2006). *Canonical correlation analysis: Assessing links between multiplex networks*. Social Networks, 10.

Casa, A., Bouveyron, C., Erosheva, E., & Menardi, G. (2021). Co-clustering of Time-dependent data via the shape invariant model.

Chang-Ha, H. (2011). On estimating the dimensionality in discriminant analysis. *Communications in Statistics* - Theory and Methods.

Chacko, H. E. (1986). An Example of the use of canonical correlation analysis. *Annual Meeting of the Mid-South Educational Research Association*. Memphis, 19-21.

Chao, G., Zongning, M., Zhao R., & Harrison, H. Z. (2015). Minimax estimation in sparse canonical correlation analysis.

Chenfeng, G. & Dongrui, W. (2021). Canonical correlation analysis based multi-view classification: an overview. ArXiv: 1907.01693v2 [cs. LG].

Chu, D., Liao, L. Z., Ng, M. K., & Zhang, X. (2013). Sparse canonical correlation analysis: new formulation and algorithm. *IEEE transactions on pattern analysis and machine intelligence*, 35 (12), 3050–3065.

Chu, M. T. & Watterson, J. L. (1993). *On a multivariate eigenvalue problem*. Part 1: Algebraic theory and a power method, SIAM J. Sci. Comput., 14 (5), 1089 – 1106.

Coleman J. & Hardin J. (2013). Robust sparse canonical correlation analysis and PITCHf/x. /C:/Users/HP/Desktop/JakeColeman-2013.pdf

Combes, S. (2008). Relationships between sensory and physicochemical measurements in meat of rabbit from three different breeding systems using canonical correlation analysis, Meat Science, 11.

Dattalo, P. (2014). *A demonstration of canonical correlation analysis with orthogonal rotation to facilitate interpretation*. Unpublished manuscript, School of Social Work, Virginia Commonwealth University, Richmond, Virginia.

Ding, C. & Li, T. (2007). *Adaptive dimension reduction using discriminant analysis and K-means clustering*. In: Proc.24rd Int. Conf. Mach. Learn., Corvallis, Oregon, USA, 521–528.

Du, K. L. & Swamy, M. N. S. (2014). *Neural networks and statistical learning in discriminant analysis*. Enjoyor Labs, Enjoyor Inc., China, Concordia University, Canada.

Duda, R. O., Hart, P. E., & Stork, D. G. (2000). *Pattern Classification*. Inter science. Wiley & Sons, 2nd edition.

Fisher, R. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*. 7, 179–188.

Gentle, E. J. (2017). *Matrix Algebra: Theory, computation, and applications in statistics*. 2nd Edition, Fairfax, VA, USA.

Glahn, R. H. (1967). *Canonical correlation analysis and its relationship to discriminant analysis and multiple regression analysis*. Weather Bureau, ESSA, Silver Spring, Md.

Golugula, A., Lee, G., Master, S. R., Feldman, M. D., Tomaszewski, J. E., Speicher, D. W., & Madabhushi, A. (2011). *Supervised Regularized Canonical Correlation Analysis: integrating histologic and proteomic measurements for predicting biochemical recurrence following prostate surgery*. BMC Bioinformatics, 12:483. http://www.biomedcentral.com/1471-2105/12/483

Hair (Jr), J. F., Black, W. C., Babin, B. J., & Anderson, R. E., (2006). *Multivariate data analysis*. New Jersey: Pearson Prentice Hall.

Hardle, W. & Simar, L. (2007). *Applied multivariate statistical analysis*. Berlin, Germany: Springer.

Hardoon, D. R., Szedmak, S., & Shawe-Talyor, J. (2004). *Canonical correlation analysis: An overview with application to learning methods*. Neural Computation 16, 2639 – 2664.

Hardoon, D. R. & Shawe-Taylor, J. (2011). *Sparse canonical correlation analysis*. Machine Learning 83 (3), 331–353.

Helwig, N. E. (2017). *Canonical Correlation Analysis*. Assistant Professor of Psychology and Statistics, University of Minnesota (Twin Cities).

Horst, P. (1961). *Generalized canonical correlations and their applications to experimental data*. J. Clin. Psychol., 17 (4), 331 – 347.

Hotelling, H. (1935). The most Predictable Criterion. *Journal of Educational Psychology*, 26 (2) 139-143.

Hotelling, H. (1936). *Relations between two sets of variates*. Biometri 28 (5), 321 – 377.

Hou, S., Heng, L., & Quansen, S. (2018). *Sparse regularized discriminative canonical correlation analysis for multi-view semi-supervised learning*. Neural Computing and Applications. https://doi.org/10.1007/s00521 -018-3582-2.

Hsieh, W. W. (2000). *Nonlinear canonical correlation analysis by neural networks*. Neural Network, 13 (10), 1095 – 1105.

Hwang, H., Kwanghee, J., & Yoshio, T. (2011). *Functional Multi-set canonical correlation analysis*. Psychometrika, DOI: 10.1007/S11336-011-9234-4

Ignacio, G., S´ebastien, D., Pascal, G., Martin, P., & Alain B. (2008). An R package to Extend Canonical Correlation Analysis. *Journal of Statistical Software*, 23 (12), http://www.jstatsoft.org/

Jaiswal, U. C., Poonia, J. S., Kumar, J. (1995). *Canonical correlation analysis for studying relationship among several traits: An example of calculation and interpretation*. Indian J. Anim. Sci., 65, 765–769

Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis*. 6th ed., Pearso prentice Hall, Upper Saddle River, New Jersey 07458.

Johnstone, I. M. & Nadler, B. (2017). *Roy's largest root test under rank-one alternatives*. Biometrika. Mar; 104(1): 181–193. Published online Jan 13. doi: 10.1093/biomet/asw060.

Joshua, C. (2016). *Introduction to canonical correlation analysis*. Video 1 https://www.youtube. com/watch?v=yz5JOfnGMA

Joungyoun K., Youngrae K., Johan L. & Sungim L (2023). Post Hotelling's T-Squared procedure to identify fault variables. *Journal of Statistical Computation and Simulation*, 94:1, 1-28, DOI: 10.1080/00949655.2023. 2228958.

Kabir1, A., Rebecca, D. M., Abu, A. S., Rolf, D. W. K., Alain. B. L., Parul, C., Keith, P. W. Jr, & Mohammed, N. (2014). *Canonical correlation analysis of infant's size at birth and maternal factors: A case study in rural north west bangladesh*. PLoS ONE 9(4): e94243, doi:10; 1371/journal. Pane.-0094243.

Kettenring, J. E. (1971). Canonical analysis of several sets of variables. Biometrika, 58 (3) 433–451.

Kim, T. K., Kittler, J. & Cipolla, R. (2006). Learning discriminative canonical correlations for object recognition with image sets. In Proc. Eur. Conf. Comput. Vis., 251–262.

Kim, T. K., Kittler, J., & Cipolla, R. (2007). *Discriminative learning and recognition of image set classes using canonical correlations*. IEEE Trans. Pattern Anal. Mach. Intell., 29 (6), 1005–1018.

Knappenberger, P. C. (1993). Cyclone tracks and wintertime climate in the mid-atlantic region of the USA. *International Journal of Climatology*, 07.

Lai, P. L. & Fyfe, C. (1998). *Canonical correlation analysis using artificial neural networks*. in Proc. 6th Eur. Symp. Artif. Netw., 363 – 368.

Lai, P. L. & Fyfe, C. (1999). *A neural implementation of canonical correlation analysis*. Neural Netw., 12 (10), 391 – 397.

Lai, P. L. & Fyfe, C. (2000). Kernel and nonlinear canonical correlation analysis. *Int. J. Neural Syst.*, 10 (5) 365–377.

Langworthy, W. B. (2020). *Extensions to Canonical Correlation Analysis and Principal Components Analysis with Applications to Survival and Brain Imaging Data*. Department of Biostatistics in the Gillings School of Global Public Health, Chapel Hill.

Lei-Gao, A., Lin, Q., & Ling, G. (2021). *Discriminative multiple canonical correlation analysis for information fusion*. arXiv:2103.00361v1 [cs.LG].

Luo, Y., Dacheng, T., Yonggang, W., Kotagiri, R., & Chao, X. (2015). *Tensor Canonical Correlation Analysis for Multi-view Dimension Reduction*. arXiv:1502.02330v1 [stat.ML]

Magnus, J. R. (2019). *Matrix Differential Calculus with Applications in Statistics and Econometrics* , Wiley & Sons.

Marden, I. J. (2015). *Multivariate Statistics*. Old School.

Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis*. Academic Press, London. MR0560319

Marubayashi, H., André, F., Luciano, P. S., Rodrigo, R. M., & Elias, N. M. (2014). Canonical-correlation analysis applied to selection-index methodology in quails. Livestock Science.

Mazuruse, P. (2014). Canonical correlation analysis. *Journal of Financial Economic Policy*

Mehdi, K. (2020). *Satellite Remote Sensing in Hydrological Data Assimilation*. Springer Science and Business Media LLC.

Mukhopadhyay, N. (2000). *Probability and Statistical Inference*. CRC Press.

Nail, H. T. (2002). *Applied Multivariate Analysis*. Springer-Verlag, New York, Berlin, Heideberg.

Nayir, F., & Saridas, G. (2022). The relationship between culturally responsive teacher roles and innovative work behavior: Canonical correlation analysis. *Journal of Educational Research and Practice*. 12, 36–50. https://doi.org/10.5590/JERAP.2022.12.1.03

Nelsen, R. (1994). *Proof without words: Cauchy-Schwarz inequality*. Math. Mag., 67 (1), 20.

Nizamettin, B., Abasiyanik, M. F., Ersan, S., & Barik, A. S. (2006). *Canonical Correlation Analysis of Factors Involved in the Occurrence of Peptic Ulcers*. Digestive Diseases and Sciences.

Obben, J. (1992). Performance of the Ghanaian rural banks: A canonical correlation analysis. Oxford Agrarian Studies.

Parkhomenko, E., Tritchler, D. & Beyene, J. (2007). Genome-wide sparse canonical correlation of gene expression with genotypes. BMC Proc., 1 (1). Art. no. S119.

Peng, Y., Zhang, D., & Zhang, J. (2010). *A new canonical correlation analysis algorithm with local discrimination*. Neural Process. Lett., 31 (1), 1–15.

Pillai, K. C. S. (1955). *Some New test criteria in multivariate analysis*. Ann Math Stat: 26 (1), 117–21.

Priya, C. (2018). *Performing canonical correlation anaylsis*. https://www.projectguru. in/publications/performing-canonical-correlation-analysis-cca/

Rathbun, S., Andrew, W., Srabashi, B., Roths, S., & Romer, M. (2023). *Applied multivariate statistical analysis*. STAT 505, Elberly Collage of Science, Pennsylvania State University.

Reiter, M. (2010). *Enhanced multiple output regression based on canonical correlation analysis with applications in computer vision*.

Rencher, A. C. (2002). *Methods of multivariate analysis*. Second Edition, John Wiley and Sons, INC. Publication, Brigham Young University.

Richardson, C. W. (1981). *Stochastic Simulation of Daily Precipitation, Temperature, and Solar Radiation*. Water Resources Research.

Ronald, C. (2011). *Regression Analysis*. Springer Texts in Statistics.

Roungu, A. M., Matair, R., Sanwar, H., & Azizur, R. (2013). *Canonical correlation analysis - an application to bank performance and consumers satisfaction*. International Journal of Advanced Scientific and Technical Research, 3 (1), http://www.rspublication.com/ijst/index.html.

Rupnik, J. (2016). *Multi-View canonical correlation analysis*. Faculty of Physics and Mathematics, University of Ljubljana, Slovenia.

Sakar, C. O. & Kursun, O. (2017). Discriminative feature extraction by a neural implementation of canonical correlation analysis. *IEEE Trans. Neural Netw. Learn. Syst.*, 28 (1), 164–176.

Samarov, V. D. (2009). *The analysis and advanced extensions of canonical correlation analysis*. Doctoral thesis, University of North Carolina, Chapel Hill. http://cdr.lib.unc.edu.gh.

Schwarz, K. (1888). Uber ein die fl¨achen kleinsten fl¨acheninhalts betreffendes problem der variationsrechnung. Retrieved January 7, 2018 from: https://link.springer.com/chapter/10.1007, 2F978-3-642-50665-9-11.

Seber, G. A. F. (1984). *Multivariate Observations*. New York: John Wiley and Sons.

Shafto, M., Asaf, D., & Kirlik, A. (1997). Canonical correlation analysis of data on human-automation interaction.

Sharma, S. (1996). *Applied Multivariate Techniques*. John Wiley and Sons Inc., New York.

Shelley, M. (2007). *Multivariate Techniques for Dichotomous Dependent Variables : An Application to Public Policy*. Public Administration and Public Policy.

Shin, Y. & Park, C. (2011). *Analysis of correlation based dimension reduction methods*. Int. J. Appl. Math. Comput. Sci., 21 (3), 549–558.

Simo, P., Styan, G. P. H., & Jarkko, I. (2011). *Matrix Tricks for Linear Statistical Models*. Springer Science and Business Media LLC.

Sisi, Q., Hanyu, L., Liang, W., Wenqiang, Z. & Shaochun, Y. (2020). *Canonical correlation study on the relationship between shipping development and water environment of the Yangtze river*.

Srivastava, J., & Dayanand, N. N. (2008), Canonical Correlation Analysis of Longitudinal Data, Denver JSM, Proceedings, Biometrics Section, 563-568.

Steel, R. G. (1951), Minimum generalized variance for a set of linear functions, *Annals Maths. Statist.*, 22 (7) 456 – 460.

Sun, T. & Chen, S. (2007). Locality preserving CCA with applications to data visualization and pose estimation, Image Vis. Comput., 25 (5) 531–543.

Sun, T., Chen, S., Yang, J., & Shi, P. (2008), A novel method of combined feature extraction for recognition. *In Proc. Conf. Data Mining*, 1043–1048.

Sun, T., Chen, S., Yang, J., Hu, X., & Shi, P. (2009), *Discriminative canonical correlation analysis with missing samples*, in Proc. World Congr. Comput. Sci. Inf. Eng., 95–99.

Tabachnick, B. (1989), *Using Multivariate Statistics*, 2nd ed., New York, Harper Collins Publishers, Inc.

Tandanai, R. (2015), *Canonical Correlation Analysis of Aggravated Robbery and Poverty in Limpopo Province*. Master's thesis. South Africa Univer. http://hdl.handle.net/10500/19629

Tingkai, S., Songcan, C., Jingyu, Y., Xuelei, H., & Pengfei, S. (2009), *Discriminative Canonical Correlation Analysis with Missing Samples*,

Tomer, M., Weiran, W., & Karen, L. (2016). Non-parametric Canonical Correlation Analysis. *Proceedings of the* 33^{rd} *International Conference on Machine Learning*, New York, NY, USA, 2016. JMLR: W & CP, 48 (2).

UGC (2022). *Solar radiation - Understanding Global Change*. A project of the University of California Museum of Paleontology. University of California Regents. https://ugc.berkeley.edu

Vinograde, B. (1950), *Canonical positive definite matrices under internal linear transformations*, Proc. Amcr. Math. Soc., 1 (2), 159 – 161.

Wang, Y., Guan, L., & Venetsanopoulos, A. N. (2015). "Kernel based fusion with application to audiovisual emotion recognition," IEEE Trans. on Multimedia, 14 (3), 597-607.

Warner, R (2013). *Applied Statistics: From Bivariate Through Multivariate Techniques*. SAGE.

Wilms, I., & Christophe, C. (2015). Sparse canonical correlation analysis from a predictive point of view. *Biometrical Journal*.

Win, H. & Wu, S. (2000). *Various proofs of the Cauchy-Schwarz inequality*. Retrieved January 1, 2018 from: http://www.ajmaa.org/RGMIA/papers/v12e/Cauchy-Schwarzinequality.pdf

Witten, M. D., Tibshrani, R., & Trevor, H. (2009), *A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis*, Advance Access publication, 515 – 534.

Xuefei, C., Jun, K., Sandstedea, B., & Luoc, X. (2019), *Time-dependent Canonical Correlation Analysis for Multilevel Time Series*.

Yang, X., Liu, W., Tao, D., & Cheng, J. (2017). *Canonical correlation analysis networks for two-view image recognition*. *Inf. Sci.*, 385 (9), 338 – 352.

Yang, X., Liu, W., Wei, L., & Tao, D. (2021). *A survey on canonical correlation analysis*. *Inf. Sci.*, 33 (6), 2349 – 2368.

Yoshida, K., Junichiro, Y., & Kenji, D. (2017). *Sparse kernel canonical correlation analysis for discovery of nonlinear interactions in high-dimensional data*. BMC Bioinformatics, 18:108 DOI 10.1186/s12859-017-1543-x

Zhihua, J. & Zhen, Y. (2010). On using non-linear canonical correlation analysis for voice conversion based on Gaussian mixture model. *Journal of Electronics* (China).

Zhou, Y, Lu, H. & Cheung, Y. M. (2017). Bilinear probabilistic canonical correlation analysis via hybrid concatenations. *In Proc.* 31st *AAAI Conf. Artif. Intell.*, 2949–2955.

Zuobin, W., Kezhi, M., & Ng, G. W. (2017). Effective feature fusion for pattern classification based on intra-class and extra-class discriminative canonical correlation analysis. *In Proc.*, 1–8.

APPENDICES

APPENDIX A: PROOF OF SOME LEMMAS

The following are the proofs of the eigenvalue and eigenvectors canonical correlation factorization for the two given lemmas.

Lemma A-1

If and only if $\mathbf{C}$ and $\mathbf{D}$ are two given matrices, then the matrices $\mathbf{CD}$ and $\mathbf{DC}$ have the same eigenvalue (Coleman and Hardin, 2013; Magnus, 2019).

Proof

If $[\mathbf{I} - \mathbf{CD}]$ is invertible, then $[\mathbf{I} - \mathbf{DC}]$ is also invertible, and let $\mathbf{M} = (\mathbf{I} - \mathbf{CD})^{-1}$ be the required matrix.

$$\begin{aligned}
\Rightarrow (\mathbf{I} - \mathbf{DC})(\mathbf{I} + \mathbf{DMC}) &= \mathbf{I} + \mathbf{DMC} - \mathbf{DC} - \mathbf{DCDMC} \\
&= \mathbf{I} + \mathbf{D}(\mathbf{MC} - \mathbf{CDMC}) - \mathbf{DC} \\
&= \mathbf{I} - \mathbf{DC} + \mathbf{D}(\mathbf{I} - \mathbf{CD})(\mathbf{MC}) \\
&= \mathbf{I} - \mathbf{DC} + \mathbf{D}(\mathbf{I} - \mathbf{CD})(\mathbf{I} - \mathbf{CD})^{-1}\mathbf{C} \\
&= \mathbf{I} - \mathbf{DC} + \mathbf{DC} = \mathbf{I}
\end{aligned}$$

Lemma A-2

Let $\mathbf{C}$ and $\mathbf{D}$ be two matrices, where $\mathbf{D}$ is positive definite and suppose that the equation $Max_{\mathbf{v}}(\mathbf{v}'\mathbf{C}\mathbf{v})$ is given subject to the constraint: $\mathbf{v}'\mathbf{D}\mathbf{v} = 1$. When $\mathbf{v}$ is the eigenvector of $\mathbf{D}^{-1}\mathbf{C}$ corresponding to the biggest eigenvalue, then the maximum is reached (Coleman and Hardin, 2013; Magnus, 2019).

Proof

Since $\mathbf{D}$ is positive definite and invertible, a positive square root exists. Suppose this square root is given by $\mathbf{D}^{\frac{1}{2}}$; which is positive definite, symmetric, and diagonalizable by the spectral theorem definition. If $\mathbf{w} = \mathbf{D}^{\frac{1}{2}}\mathbf{v} \Rightarrow \mathbf{v} = \mathbf{D}^{-\frac{1}{2}}\mathbf{w}$. Then

Equation (1.7) can be rewritten as

$$\max_{\mathbf{v}}(\mathbf{v}'\mathbf{C}\mathbf{v}) = \max_{\mathbf{w}}(\mathbf{D}^{-\frac{1}{2}}\mathbf{w})'\mathbf{C}(\mathbf{D}^{-\frac{1}{2}}\mathbf{w})$$

$$\max_{\mathbf{v}}(\mathbf{v}'\mathbf{C}\mathbf{v}) = \max_{\mathbf{w}} \mathbf{w}'\mathbf{D}^{-\frac{1}{2}}\mathbf{C}\mathbf{D}^{-\frac{1}{2}}\mathbf{w} \tag{A1}$$

Equation (A1) is subject to the following constraint:

$$\mathbf{v}'\mathbf{D}\mathbf{v} = \left(\mathbf{D}^{-\frac{1}{2}}\mathbf{w}\right)'\mathbf{D}\left(\mathbf{D}^{-\frac{1}{2}}\mathbf{w}\right) = \mathbf{w}'\mathbf{D}^{-\frac{1}{2}}\mathbf{D}\mathbf{D}^{-\frac{1}{2}}\mathbf{w} = \mathbf{w}'\mathbf{w} = 1$$

Let $\mathbf{D}^{-\frac{1}{2}}\mathbf{C}\mathbf{D}^{-\frac{1}{2}} = \Gamma\Lambda\Gamma'$ be the spectral decomposition (Coleman and Hardin, 2013; Magnus, 2019). It follows from the literature that: 1. For diagonal with eigenvalues of $\mathbf{D}^{-\frac{1}{2}}\mathbf{C}\mathbf{D}^{-\frac{1}{2}}$, Λ is diagonal.

2. $\Gamma = [\mathbf{v}_1|\mathbf{v}_2|\mathbf{v}_3|....\mathbf{v}_n]$ is a column matrix of eigenvectors corresponding to the entries of Λ. By the spectral theorem, they form an orthonormal basis in R^n. Now let $\mathbf{z} = \Gamma'\mathbf{w}$, then

$$\mathbf{z}'\mathbf{z} = (\Gamma'\mathbf{w})'(\Gamma'\mathbf{w}) = \mathbf{w}'\Gamma\Gamma'\mathbf{w} = \mathbf{w}'\mathbf{w}$$

The last equality follows since Γ is a matrix composed of orthonormal columns (Coleman and Hardin, 2013). Equation (A1) reduces to the form in Equation (A2) subject to the constraint: $\mathbf{z}'\mathbf{z} = 1$.

$$\max_{\mathbf{w}} \mathbf{w}'\mathbf{D}^{-\frac{1}{2}}\mathbf{C}\mathbf{D}^{-\frac{1}{2}}\mathbf{w} = \max_{\mathbf{w}} \mathbf{w}'\Gamma\Lambda\Gamma'\mathbf{w} = \max_{\mathbf{w}} (\Gamma'\mathbf{w})'\Lambda\Gamma'\mathbf{w} = \max_{\mathbf{z}} \mathbf{z}'\Lambda\mathbf{z}$$

$$\Rightarrow \max_{\mathbf{w}} \mathbf{w}'\mathbf{D}^{-\frac{1}{2}}\mathbf{C}\mathbf{D}^{-\frac{1}{2}}\mathbf{w} = \max_{\mathbf{z}} \sum_{i=1}^{n} \lambda_i \mathbf{z}_i^2 \tag{A2}$$

Because Λ is just a diagonal matrix with λ_i on the i^{th} diagonal, the final line naturally follows. Note that for $max_{\mathbf{z}}\mathbf{z}'\Lambda\mathbf{z}$, we can switch directly from maximizing over **w** to maximizing over **z** because Γ is constant given **C** and **D**. If we let λ_1 be the largest eigenvalue, then from Equation (A2) we have

$$\max_{\mathbf{z}} \sum_{i=1}^{n} \lambda_i \mathbf{z}_i^2 \leq \max_{\mathbf{z}} \sum_{i=1}^{n} \lambda_1 \mathbf{z}_i^2 \leq \max_{\mathbf{z}} \lambda_1 \sum_{i=1}^{n} \mathbf{z}_i^2 = \lambda_1 \tag{A3}$$

The final equality is determined by the constraint: $\mathbf{z}'\mathbf{z} = \sum_{i=1}^{n} \mathbf{z}_i^2 = 1$. Note that equality in Equation (A3) is attained for $\mathbf{z} = [1, 0, 0, ..., 0]'$. By the Spectral Theorem, $\mathbf{v}_1, \mathbf{z}_2, ..., \mathbf{z}_n$ form an orthonormal basis in R^n, it must be the case that $\mathbf{v}_i'\mathbf{v}_j = \delta_{ij}$. Thus, $\mathbf{v}_1 = \mathbf{w}$ is the unique answer to the set of equations. Remember that the largest eigenvalue of $\mathbf{D}^{-\frac{1}{2}}\mathbf{C}\mathbf{D}^{-\frac{1}{2}}$ is represented by the eigenvector $\mathbf{v}_1$. Therefore, $\mathbf{v} = \mathbf{D}^{-\frac{1}{2}}\mathbf{v}_1$.

By Lemma 1.1, $\mathbf{D}^{-\frac{1}{2}}\mathbf{C}\mathbf{D}^{-\frac{1}{2}}$ and $\mathbf{D}^{-1}\mathbf{C}$ have the same eigenvalues, so λ_1 is also the largest eigenvalue of $\mathbf{D}^{-1}\mathbf{C}$. Note that $\mathbf{D}^{-\frac{1}{2}}\mathbf{C}\mathbf{D}^{-\frac{1}{2}}\mathbf{v}_1 = \lambda_1\mathbf{v}_1$.

$$\begin{aligned}
\mathbf{D}^{-1}\mathbf{C}\mathbf{v} &= \mathbf{D}^{-1}\mathbf{C}\mathbf{D}^{-\frac{1}{2}}\mathbf{v}_1 \\
&= \mathbf{D}^{-\frac{1}{2}}\mathbf{D}^{-\frac{1}{2}}\mathbf{C}\mathbf{D}^{-\frac{1}{2}}\mathbf{v}_1 \\
&= \lambda_1\mathbf{v}_1 = \lambda_1\mathbf{D}^{-\frac{1}{2}}\mathbf{v}_1
\end{aligned}$$

$$\mathbf{D}^{-1}\mathbf{C}\mathbf{v} = \lambda_1\mathbf{v} \tag{A4}$$

λ_1 is the largest eigenvalue of $\mathbf{D}^{-1}\mathbf{C}$ and $\mathbf{v} = \mathbf{D}^{-\frac{1}{2}}\mathbf{w} = \mathbf{D}^{-\frac{1}{2}}\mathbf{v}_1$. Remember that $\mathbf{z}$ was maximized when $\mathbf{w} = \mathbf{v}_1$ leads to (Coleman and Hardin, 2013; Magnus, 2019) $\mathbf{v} = \mathbf{D}^{-\frac{1}{2}}\mathbf{v}_1$, which is the eigenvector of $\mathbf{D}^{-1}\mathbf{C}$ corresponding to the λ_1, the greatest eigenvalue of $\mathbf{D}^{-1}\mathbf{C}$.

APPENDIX B: PROOF OF CAUCHY-SCHWAZ INEQUALITY

The inequality holds if either $\mathbf{s} = 0$ or $\mathbf{t} = 0$. Suppose that $\mathbf{s}$ and $\mathbf{t}$ are not positive definite, that is, $\mathbf{s} \leq 0$ and $\mathbf{t} \leq 0$, and consider the vector$(\mathbf{s} - \theta\mathbf{t})$, where $\theta \neq 0$ is a scalar. It follows that $|\mathbf{s} - \theta\mathbf{t}| > 0$ is a positive definite and by expanding yields Equation (B1).

$$(\mathbf{s} - \theta\mathbf{t})' (\mathbf{s} - \theta\mathbf{t}) = \mathbf{s}'\mathbf{s} - \mathbf{s}'\theta\mathbf{t} - \theta\mathbf{t}'\mathbf{s} + \theta^2\mathbf{t}'\mathbf{t} = (\mathbf{t}'\mathbf{t})\theta^2 - 2(\mathbf{s}'\mathbf{t})\theta + \mathbf{s}'\mathbf{s}$$

$$(\mathbf{s} - \theta\mathbf{t})' (\mathbf{s} - \theta\mathbf{t}) = (\mathbf{t}'\mathbf{t})\theta^2 - 2(\mathbf{s}'\mathbf{t})\theta + \mathbf{s}'\mathbf{s} \tag{B1}$$

Equation (B1) is a quadratic equation in terms of θ. By the method of completing the squares and further simplification gives Equation (B2).

$$(\mathbf{s} - \theta\mathbf{t})' (\mathbf{s} - \theta\mathbf{t}) = \mathbf{s}'\mathbf{s} + \theta^2\mathbf{t}'\mathbf{t} - 2\theta\mathbf{s}'\mathbf{t} > 0 \Rightarrow \mathbf{s}'\mathbf{s} + (\mathbf{t}'\mathbf{t})\left[\theta^2 - \frac{2\theta\mathbf{s}'\mathbf{t}}{\mathbf{t}'\mathbf{t}}\right] > 0$$

$$\mathbf{s}'\mathbf{s} + (\mathbf{t}'\mathbf{t})\left[\theta^2 - \frac{2\theta\mathbf{s}'\mathbf{t}}{\mathbf{t}'\mathbf{t}} + \left(\frac{\mathbf{s}'\mathbf{t}}{\mathbf{t}'\mathbf{t}}\right)^2 - \left(\frac{\mathbf{s}'\mathbf{t}}{\mathbf{t}'\mathbf{t}}\right)^2\right] > 0$$

$$\mathbf{s}'\mathbf{s} - (\mathbf{t}'\mathbf{t})\left(\frac{\mathbf{s}'\mathbf{t}}{\mathbf{t}'\mathbf{t}}\right)^2 + (\mathbf{t}'\mathbf{t})\left[\theta^2 - \frac{2\theta\mathbf{s}'\mathbf{t}}{\mathbf{t}'\mathbf{t}} + \left(\frac{\mathbf{s}'\mathbf{t}}{\mathbf{t}'\mathbf{t}}\right)^2\right] > 0$$

$$\therefore \mathbf{s}'\mathbf{s} - \frac{(\mathbf{s}'\mathbf{t})^2}{\mathbf{t}'\mathbf{t}} + (\mathbf{t}'\mathbf{t})\left[\theta - \frac{\mathbf{s}'\mathbf{t}}{\mathbf{t}'\mathbf{t}}\right]^2 > 0 \tag{B2}$$

Since $(\mathbf{t}'\mathbf{t})\left[\theta - \frac{\mathbf{s}'\mathbf{t}}{\mathbf{t}'\mathbf{t}}\right]^2 \geq 0$ for all θ in Equation (B2), it follows that the inequality holds if $\mathbf{s}'\mathbf{s} - \frac{(\mathbf{s}'\mathbf{t})^2}{\mathbf{t}'\mathbf{t}} \geq 0$

$$\Rightarrow \mathbf{s}'\mathbf{s} - \frac{(\mathbf{s}'\mathbf{t})^2}{\mathbf{t}'\mathbf{t}} \geq 0 \Rightarrow (\mathbf{s}'\mathbf{s})(\mathbf{t}'\mathbf{t}) - (\mathbf{s}'\mathbf{t})^2 \geq 0 \Rightarrow (\mathbf{s}'\mathbf{s})(\mathbf{t}'\mathbf{t}) \geq (\mathbf{s}'\mathbf{t})^2$$

By changing the sides of the equation above yields the required Cauchy-Schwarz Inequality (Mukhopadhyay, 2000; Win and Wu, 2000). $\Rightarrow (\mathbf{s}'\mathbf{t})^2 \leq (\mathbf{s}'\mathbf{s})(\mathbf{t}'\mathbf{t})$

APPENDIX C: PROOF OF THEOREM 3.4

From the constraints of Equation (3.118) we have the following equations:

$\alpha = U_1\Sigma_y^{-1}P_y(:,1:l)\mathbf{B} + U_2F_y$ and $\beta = V_1\Sigma_x^{-1}P_x(:,1:l)\mathbf{B} + V_2F_x$

The above equations are equivalent to the following equations:

$U_1'\alpha = \Sigma_y^{-1}P_y(:,1:l)\mathbf{B}$ and $V_1'\beta = \Sigma_x^{-1}P_x(:,1:l)\mathbf{B}$

By taking $\mathbf{B} = I$, since $\mathbf{B}_f$ is also orthogonal, we have $D_s \leq D$. On the other hand, for $\mathbf{B}_f$ and α^f which satisfy the following equations,

$$U_1'\alpha^f = \Sigma_y^{-1}P_y(:,1:l)\mathbf{B}_f$$

$$V_1'\beta^f = \Sigma_x^{-1}P_x(:,1:l)\mathbf{B}_f$$

we obtain the following equations:

$$U_1'\alpha^f\mathbf{B}_f^{-1} = \Sigma_y^{-1}P_y(:,1:l)$$

$$V_1'\beta^f\mathbf{B}_f^{-1} = \Sigma_x^{-1}P_x(:,1:l)$$

It follows from the literature that $\left(\alpha^f\mathbf{B}_f^{-1}, \beta^f\mathbf{B}_f^{-1}\right) \in \rho$, for all $\alpha, \beta \in \rho$ and $\mathbf{B}$ orthogonal vector. Hence the above equations yield

$$U_1'\alpha = \Sigma_y^{-1}P_y(:,1:l)$$

$$V_1'\beta = \Sigma_x^{-1}P_x(:,1:l)$$

Multiplying both sides of the above equations by $\mathbf{B}$ yields the following eqns:

$$U_1'\alpha\mathbf{B} = \Sigma_y^{-1}P_y(:,1:l)\mathbf{B}$$

$$V_1'\beta\mathbf{B} = \Sigma_x^{-1}P_x(:,1:l)\mathbf{B}$$

This means that $(\alpha\mathbf{B}, \beta\mathbf{B}, \mathbf{B}) \in \rho_s$. Specifically, we can take $\alpha = \alpha^f \mathbf{B}_f^{-1}$, $\beta = \beta^f \mathbf{B}_f^{-1}$ and $\mathbf{B} = \mathbf{B}_f$ for all $\left(\alpha^f, \beta^f, \mathbf{B}_f\right) \in \rho_s$. This implies $(\alpha\mathbf{B}, \beta\mathbf{B}, \mathbf{B}) \mid (\alpha, \beta) \in \rho_s$, $\mathbf{B}$ is orthogonal = ρ_s, where $\mathbf{B} \in \Re^{l \times l}$, and as a result, the issue in Equation (3.118) is comparable to the issue in Equation (C1).

$$\min_{\alpha,\beta} ||\alpha\mathbf{B}|\,|l_1 + ||\beta\mathbf{B}|\,|l_1 \tag{C1}$$

Subject to the constraints: $U_1'\alpha = \Sigma_y^{-1} P_y\,(:, 1:l)$ and $V_1'\beta = \Sigma_x^{-1} P_x\,(:, 1:l)$. Based on the orthogonality of $\mathbf{B}$ and the norm equivalences of finite dimensional spaces, the norm results is finally given in Equation (C2).

$$\begin{aligned}
\frac{1}{\sqrt{l}} \sum_i ||\alpha|\,|l_1 &= \frac{1}{\sqrt{l}} \sum_i ||\alpha\,(i,:)|\,|l_1 \\
&\le \sum_i ||\alpha\,(i,:)|\,|l_2 \\
&= \sum_i ||\alpha\,(i,:)\,\mathbf{B}|\,|l_2 \\
&\le \sum_i ||\alpha\,(i,:)\,\mathbf{B}|\,|l_1 = \sum_i ||\alpha\mathbf{B}|\,|l_1
\end{aligned}$$

$$\Rightarrow \frac{1}{\sqrt{l}} \sum_i ||\alpha|\,|l_1 = \sum_i ||\alpha\mathbf{B}|\,|l_1 \tag{C2}$$

Equation (C2) depicts that $AD \le D_s$ and $D_s \le D$ is already known. It follows that $AD \le D_s \le D$ where $A = \frac{1}{\sqrt{l}}$ yields the proof as required.

APPENDIX D: TIME-IND. SIMULATION OF SOME GSDCCA

Table D1: Simulation Results of Two-Group GS Discriminant CCA

GS	FDCC	C Class	Eig V	Chi-Square	W Lambda
1	0.092	58.0	0.009	2.218	0.991
2	0.115	60.6	0.013	3.447	0.987
3	**0.163**	**61.4**	**0.027**	**7.014**	**0.973**
4	0.139	57.6	0.020	5.063	0.981
5	0.102	55.7	0.010	2.700	0.990
6	0.121	56.4	0.015	3.841	0.985
7	0.146	58.7	0.022	5.595	0.979
8	0.134	56.4	0.018	4.680	0.982
9	0.108	58.3	0.012	3.053	0.988
10	0.104	53.4	0.011	2.825	0.989
11	0.098	53.4	0.010	2.488	0.990
12	0.104	55.7	0.011	2.820	0.989
13	0.126	56.1	0.016	4.170	0.984
14	0.134	56.1	0.018	4.665	0.982
15	0.082	53.8	0.007	1.746	0.993
16	0.121	54.2	0.015	3.818	0.985
17	0.119	55.7	0.014	3.718	0.986
18	0.146	57.2	0.022	5.561	0.979
19	0.132	58.0	0.018	4.577	0.982
20	0.106	57.2	0.011	2.922	0.989

Source: Researcher's computations (2023)

Table D2: Simulation Results of Three-Group GS Discriminant CCA

GS	FDCC	C Class	Eig V	W Lambda	Chi-Square
1	0.142	37.9	0.021	0.972	7.235
2	0.154	42.0	0.024	0.966	8.875
3	0.209	39.4	0.046	0.950	13.285
4	**0.214**	40.5	**0.048**	**0.945**	**14.532**
5	0.181	39.8	0.034	0.962	10.134
6	0.164	39.8	0.028	0.961	10.242
7	0.179	41.7	0.033	0.953	12.439
8	0.157	**42.8**	0.025	0.961	10.317
9	0.145	39.0	0.022	0.967	8.569
10	0.137	36.0	0.019	0.973	7.155
11	0.143	33.0	0.021	0.972	7.295
12	0.158	34.5	0.026	0.965	9.172
13	0.173	39.8	0.031	0.955	12.024
14	0.142	42.0	0.020	0.968	8.400
15	0.139	41.3	0.020	0.977	6.087
16	0.139	41.3	0.020	0.969	8.273
17	0.148	39.4	0.022	0.961	10.290
18	0.160	38.3	0.026	0.961	10.320
19	0.157	40.9	0.025	0.961	10.283
20	0.193	42.4	0.039	0.958	11.069

Source: Researcher's computations (2023)

Table D3: Simulation Results of Four-Group GS Discriminant CCA

GS	FDCC	C Class	Eig V	W Lambda	Chi-Square
1	0.195	29.2	0.039	0.949	13.622
2	0.164	31.1	0.028	0.945	14.475
3	0.214	34.1	0.048	0.925	20.157
4	**0.215**	**34.5**	**0.048**	**0.924**	**20.895**
5	0.181	33.0	0.034	0.948	13.729
6	0.166	32.2	0.028	0.956	11.651
7	0.179	34.1	0.033	0.939	16.152
8	0.196	33.7	0.040	0.937	16.699
9	0.167	30.7	0.029	0.946	14.266
10	0.160	29.2	0.026	0.952	12.796
11	0.152	25.0	0.024	0.951	12.909
12	0.158	27.7	0.026	0.952	12.766
13	0.175	27.7	0.032	0.949	13.431
14	0.162	28.8	0.027	0.951	12.909
15	0.149	24.2	0.023	0.967	8.712
16	0.139	29.9	0.020	0.959	10.755
17	0.156	31.1	0.025	0.947	14.178
18	0.166	31.4	0.028	0.938	16.597
19	0.183	31.4	0.035	0.932	18.041
20	0.205	29.5	0.044	0.938	16.635

Source: Researcher's computations (2023)

Table D4: Simulation Results of Six-Group GS Discriminant CCA

GS	FDCC	C Class	Eig V	W Lambda	Chi-Square
1	0.215	20.1	0.048	0.897	27.821
2	0.183	24.2	0.035	0.904	25.806
3	0.230	25.4	0.056	0.894	28.876
4	0.214	25.4	0.048	0.905	25.697
5	0.190	21.2	0.038	0.916	22.667
6	0.181	21.6	0.034	0.937	16.852
7	0.193	20.8	0.039	0.918	22.029
8	0.215	23.9	0.048	0.919	21.669
9	0.215	25.0	0.048	0.914	22.993
10	0.170	22.7	0.030	0.925	19.989
11	0.177	21.2	0.032	0.922	20.800
12	0.186	25.0	0.036	0.903	26.181
13	**0.241**	24.2	**0.061**	**0.885**	**31.320**
14	0.186	22.3	0.036	0.907	24.953
15	0.195	23.9	0.040	0.905	25.598
16	0.178	21.2	0.033	0.924	20.351
17	0.190	21.2	0.038	0.916	22.667
18	0.213	**26.9**	0.048	0.929	19.092
19	0.223	23.5	0.052	0.912	23.744
20	0.172	26.1	0.031	0.933	17.785

Source: Researcher's computations (2023)