QATAR UNIVERSITY

COLLEGE OF ENGINEERING

A LASSO-BASED DEA FOR ECO-EFFICIENCY PERFORMANCE ASSESSMENT FOR THE GLOBAL FOOD AND BEVERAGES INDUSTRY

BY

DOA'A MOHAMMAD FUAD SALAMEH

A Thesis Submitted to

the College of Engineering

in Partial Fulfillment of the Requirements for the Degree of

Masters of Science in Engineering Managemen

June 2022

COMMITTEE PAGE

The members of the Committee approve the Thesis of
Doa'a M. F. Salameh defended on 05/12/2022.

Name
Dr. Galal M. Abdella

Name
Dr. Murat Kucukvar

Name
Dr. Murat Gunduz
Committee Member

Name
Dr. Nuri Onat
Committee Member

Name
Dr. Charitha Dias
Committee Member

Name
Dr.Oktay Karabakh
External Examiner

Approved:

Khalid Kamal Naji, Dean, College of Engineering

ABSTRACT

SALAMEH, DOA'A, M. F., Master Degree: June 2022, Masters of Science in Engineering Management

Title: A Lasso-Based DEA for Eco-Efficiency Performance Assessment for Global Food and Beverages Industry.

Supervisor of Thesis: Dr. Galal M Abdella.

Sustainable food systems are essential to secure food and nutrition for society and preserve the economic, social, and environmental aspects. A sustainable food system has become a significant demand for survival due to the dramatic growth of urbanization and global economic and health disruptions. Recently, food supply chains of global industries are encountering economic and environmental challenges, resulting in a significant decrease in their eco-efficiency performance. Therefore, there is a great need to identify possible reasons and their potential relationship across the eco-environmental pillars of sustainability. To this end, this thesis proposes an approach for eco-efficiency assessment integrating both the Least Absolute Shrinkage Squared Operator (LASSO) with the Data Envelope Analysis. The new approach constitutes two stages. First, the LASSO regression is applied to reduce the dimension-space of the eco- and find the relative weights estimates for each indicator in the new dimension. Second, the DEA is used to estimate the eco-efficiency ratio for all the food industries. The mathematical and operational procedures of the new approach are demonstrated using the economic and environmental footprints of 30 food and beverages industries in the USA. The new strategy is expected to provide food and beverage industries with a powerful tool for assessing their contribution toward achieving sustainable development goals.

DEDICATION

This Thesis is dedicated to my beloved family and friends for their endless support and encouragement throughout this journey.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my supervisors Dr. Galal Abdella and Dr.Murat Kucukvar for their encouragement and continued support for the past four years of my journey at the college of engineering and for the best future doctor Eng. Adeeb who guided, assisted, and motivated me through all the stages of this Thesis.

I would also like to thank Qatar University and the engineering management department who provided me with this opportunity and informative journey and every individual who contributed to completing this Thesis."

LIST OF ACRONYMS

Nomenclature

Symbol	Description
β_j	Values of the coefficients of the eco-environmental indicators in the model
$x_{ij} \in \mathbb{R}^n$	The ith observation of the jth indicator
ε	Error term
X	The design matrix
p	The dimension space (number of indicators)
$\lambda \geq 0$	Pre-chosen penalization (or shrinkage) parameter
$\|\boldsymbol{\beta}\|_1$	The squared Euclidean norm
$\lambda\|\boldsymbol{\beta}\|_1 = \lambda \sum_{j=1}^{p} \lvert\beta_j\rvert$	LASSO penalty function
$\beta^{\wedge Lasso}$	Regression coefficients of the eco-environmental indicators
λ	Penalty parameter
n	Decision Making Units (DMUs).
$\lambda_j = (\lambda_1, \lambda_2, \ldots..\lambda_{nj})^T$	The intensity vector
s^- / s^+	Slacks
x_{ij}	DMU ith input
y_{rj}	DMU rth output
Pd	Value of The Objective Function
m	Inputs of DMUs Used in DEA Analysis
s	Outputs of DMUs Used in DEA Analysis
j	Number of DMUs

Abbreviations

Symbol	Description
DEA	Data Envelopment Analysis
DMU	Decision Making Units
LASSO	Least Absolute Shrinkage And Selection Operator
FSC	Food Supply Chain
GHG	Greenhouse Gases
SDGs	Sustainable Development Goals
LCA	Life Cycle Assessment
GIS	Geographical Information System
SAFA	Sustainability Assessment Of Food And Agriculture Systems
RISE	Response-Inducing Sustainability Evaluation
EIO-LCA	Hybrid Economic Input-Output And Life Cycle Assessment
ECO-LCA	Ecologically-Based Life Cycle Assessment
SD	System Dynamic Modeling
DAG	Directed Acyclic Graphs
VECM	Vector Error Correction Model
AHP	Analytical Hierarchy Process
PCA	Principle Component Analysis
OLS	Ordinary Least Squares
SBM	Slacks-Based Measure
UM-LCA	Hybrid Urban Metabolism - Life Cycle Assessment

TABLE OF CONTENTS

DEDICATION...iv

ACKNOWLEDGMENTS...v

LIST OF ACRONYMS...vi

LIST OF TABLES...ix

LIST OF FIGURES...x

Chapter 1: Introduction...1

1.1. Motivation...3

1.2. Research Problem and Research Questions...4

1.2.1. Research Problem...4

1.2.2. Research Questions...4

1.3. Research Aims Objective...5

Chapter 2: Literature Review...6

2.1. Overview...6

2.2. Environmental Assessment...6

2.3. Socio-Economic Assessment...7

2.4. Eco-environmental Assessment...9

2.5. DEA for sustainability studies:...11

2.6. Modeling and Optimization Techniques...12

2.7. Indicators and Variable Selection Approaches...13

Chapter 3: Methodology...17

3.1. Proposed Methodology 17

3.2. LASSO-based Eco-Environmental Indicator Selection 17

3.3. DEA-based Eco-efficiency Assessment 19

Chapter 4: data analysis and Results 21

4.1. Data Generation 21

4.2. Data Description 22

4.3. Selecting sustainability indicators using LASSO 25

4.3.1. Sustainability Impact Normalization 26

4.3.2. Measure Correlation 26

4.3.3. Model Selection using LASSO 28

4.4. DEA Based Efficiency Assessment 32

4.4.1. Eco-Efficiency Assessment of LASSO-based un-weighted DEA 32

4.4.2. Eco-Efficiency Assessment LASSO-based Weighted DEA 34

4.4.3. Eco-Efficiency of Full Dimension DEA 35

4.5. Results Discussion 36

Chapter 5: Conclusion and future work 43

5.1. Conclusion and Recommendations 43

References 45

Appendix A 56

Appendix B 58

LIST OF TABLES

Table 1: Selected environmental Indicators...22

Table 2: Food Industries. 24

Table 3: Optimal model settings..30

Table 4: Optimal model statistics. 31

Table 5: Ranking of 30 indutries among the three scenarios...38

Table 6: Projection Analysis Data. 40

LIST OF FIGURES

Figure 1: Most used indicators in the previous studies. 14

Figure 2: Two-stage LASSO based DEA for eco-efficiency assessment 17

Figure 3: Selected Indicators Correlation. 27

Figure 4: K-Fold Cross-validation settings under the IBM®SPSS®. 29

Figure 5: LASSO regression settings under the IBM®SPSS®. 30

Figure 6: Eco-efficiency of LASSO-based Un-weighted DEA 33

Figure 7: Eco-efficiency of LASSO Based Weighted DEA 34

Figure 8: DMUs Eco-efficiency Using 14 Indicators. 35

CHAPTER 1: INTRODUCTION

Food is an essential need for creatures in this universe. Securing and sustaining food production is a significant demand for survival. It has been noticed that the food supply chain has been jeopardized recently and lost its efficiency, which is mainly caused by greenhouse gas emissions (GHG), deforestation, water pollution, biodiversity loss, uneven water extraction, and carbon footprint. The carbon footprint continues to grow as a result of the continued use of conventional vehicles and the failure to switch to electric vehicles, which have proven to be environmentally friendly (Al-Buenain et al., 2021; Kucukvar et al., 2021a; Al-Abadi et al., 2021; Onat et al., 2021; Kutty et al., 2021a). Food consumption sustainability necessitates understanding multidimensional environmental, economic, and social impacts through a comprehensive and inclusive sustainability assessment and model-based framework (Abdella et al., 2020a). To stand and survive, food production needs clean air and water and healthy soils and climate; however, its sustainability requires a conscious and intentional decision considering the continuous societal growth(Kutty et al., 2020d). Sustainable development is critical to balancing the needs of current and future generations. A sustainable agricultural system assures food security and nutrition while simultaneously maintaining the economic, social, and environmental underpinnings essential for future generations' food security and nutrition. Between 1950 and 1960, sustainable food development was initiated; in the meantime, the green revolution exported high-technology agriculture. Using a two-stage Data Envelopment Analysis (DEA), this thesis proposes a novel approach to analyzing the sustainability performance of 30 food industries in the United States. This novel method concentrates on identifying the most crucial variables that substantially impact sustainability performance first, then assessing the efficiency performance using DEA. The importance of these sustainability indicators is that they are quantitative attributes of environmental and economic systems used to detect effects on system characteristics required to maintain human and environmental well-being. This method, therefore, increases the veracity of results. The first stage involves selecting the most significant eco-environmental indicators using the Least Absolute Shrinkage Squared Operator (LASSO) that eliminate the insignificant indicators; based on the LASSO results, an input-based Data Envelopment Analysis (DEA) model will be run to determine the efficient food manufacturing industry, accompanied by a projection level analysis to determine how inefficient food industries can improve their performance to achieve the efficacy. Considering 102 indicators arranged around the 17 Sustainable Development Goals (SDGs) adopted by the united nations, one of the most challenging difficulties for sustainability indicators is to portray historical gaps, trade-offs between the short and long term, and the differentiation between weak and strong sustainability dimensions (Eurostat, 2021). Each goal is typically linked to six indicators, 37 of the 102 metrics are multifunctional (Eurostat, 2021), which means they are used to measure several SDGs and have a good association with one another, leading most previous DEA-based studies to run DEAs assessments with a relatively large number of indicators that may be insignificant to the context or have high relation to other indicators, resulting in inaccurate results (López et al., 2016; Chen et al., 2021). The goal of this thesis is to bridge the gap in selecting appropriate eco-efficiency indicators from the vast space dimension of indicators, to assist decision-makers in situations where there is frequently a bias in the decision-making process, and to help create a standard practice to be followed by the food businesses to improve their efficiency performance.

1.1. Motivation

An unprecedented increase in the demand for food associated with the world's growing population pose serious concerns over the supply of food, particularly for developing countries. Geopolitical conflicts as well as natural disasters, exacerbated by climate crisis, lead to food market volatility and reduction in the amount of food available for import. As an integral part of ensuring long-term sustainability of food systems, scientific community has long been concerned on tracking the efficiency of the food systems with global peers to enhance the local production, consumption, and resource utilization. Indicator-based assessments are often used to track the efficiency of food systems across multiple dimensions of sustainability. Data Envelopment Analysis (DEA) has proved to be one of the most reliable tools for conducting such research (Martín-Gamboa et al.,2021). However, DEA pose certain drawbacks when dealing with high dimensional data, affecting the accuracy of the efficiency scores (López et al., 2016; Chen et al., 2021). To address these caveats, the research proposes a novel two-stage DEA approach, where stage 1 deals with dimensionality reduction followed by stage 2 running an envelopment model utilizing the selected set of indicators from stage 1 to identify the efficient decision-making units (DMU). The proposed approach is then applied to the case of 30 United States (U.S) food industries to assess the eco-efficiency performance. Least Absolute Shrinkage and Selection Operator (LASSO) is used to dimensionally reduce the set of indicators for the eco-efficiency assessment. LASSO is proved to be a powerful variable selection technique under high dimensionality and its integration with DEA will result in improving the uncertainties and vagueness despite high correlation between indicators.

1.2. Research Problem and Research Questions

This section discusses the problem that this thesis was created to address and the significant need identified when investigating the food system sustainability subject, followed by the questions that this thesis aims to answer.

1.2.1. Research Problem

This thesis aims to tackle the difficulties of assessing eco-efficiency using DEA under a high-dimensional indicator setting. High dimensional indicators settings are usually characterized by the high correlation between indicators, which leads to less accurate conclusions when utilized jointly, as revealed by prior research based only on DEA analysis(Lee et al., 2020).

1.2.2. Research Questions

The study aims to evaluate the sustainability performance of 30 food industries in the United States by measuring eco-efficiency using three different scenarios illustrated in chapter 4 to examine the applicability of the proposed method. Eco-efficiency is primarily an aspect used for assessing both environmental and economic factors by measuring the food and beverage industry's efficiency in consuming water-related to water stress and fossil energy resources in order to promote financial compensation for workers and, indirectly, to provide greater food security for a population that is more sensitive to economic insecurity. The study suggests a two-stage DEA approach that utilizes LASSO to filter the space dimension and assess the efficiencies. The research attempt to address the following questions:

1. How could LASSO be used to generate valuable inputs/outputs for analysis that do have a significant impact on eco-efficiency?
2. Could DEA analysis be conducted with a considerably smaller number of indicators

yet get valid results?

1.3. Research Aims Objective

This thesis aims to develop a new weighting method for efficiency assessment that can identify effective and efficient sustainable food systems to discover the successful practices and generalize them to food supply chains to maximize their efficiency using a novel multistage DEA approach. The aim of this study will be achieved through the following objectives:

1. LASSO analysis will analyze the critical indicators that significantly affect the industry's eco-efficiency performance.
2. To measure the eco-efficiency performance of the 30 food industries using LASSO-based DEA.
3. To examine the applicability and operational performance of the proposed weighting method.

CHAPTER 2: LITERATURE REVIEW

2.1. Overview

Several methodologies and techniques have been utilized to analyze the food system's sustainability and assess the sustainability dimensions. Although boosting green consumption and production practices at various stages of the food supply chain (FSC), is critical to achieving global food security and sustainability, several challenges must be addressed, ranging from food waste accumulation in the FSC to tackling gender inequalities and climate-related concerns (Kutty et al., 2020b). As a result, this literature review provides a background on the tools and techniques used to assess the three categories of efficiency (environmental assessment, socio-economic assessment, and eco-environmental assessment), DEA usage for sustainability studies, modeling & optimization techniques, and the indicators and variable selection approaches. Each section summarizes the tools, techniques, and studies found in the Scopus database for the last decade of research.

2.2. Environmental Assessment

Food and agriculture organizations indicate that a sustainable food system should positively or neutral impact the environment. In this section, all the tools and techniques used to quantify and assess the food system's environmental impact will be mentioned to improve these systems further, reduce emissions, and improve plant, soil, water, and animal health, biodiversity, and food loss. For example, an increase in greenhouse gases (GHG) has been indicated in China between 1989 and 2017 caused by food production and consumption shown by using a hybrid economic input-output and life cycle assessment (EIO-LCA) to determine all kinds of environmental emissions as EIOA can compute environmental footprints holistically and reliably of China's food production and consumption (Zhang et al., 2022). According to the findings, population growth and urbanization negatively influence CO2 emissions (Alsarayreh et al., 2020). Stone et al. (2021) studied the three vegetables production systems(small, medium, and large scale) and 18 vegetable crops normally grown in Des Moines, Iowa, to examine their environmental impact and make improvement decisions. The environmental impact analysis was done using the LCA approach for grapes production, including three models (early harvesting, ordinary harvesting, and delayed harvesting) that relied on the Italian system (Roselli et al., 2020). Another study evaluated the potential environmental savings that could be achieved in southern Sweden on broccoli crops if specific actions were taken using the LCA approach (Eriksson et al., 2021). Another novel bottom-up approach that utilizes a "Hybrid Urban Metabolism - Life Cycle Analysis (UM-LCA)" assessment to assess the food system's environmental impact on land use, freshwater quality, and global warming by Stellwagen et al. (2021). In Malaysia, the ecological effects of the production processes of rice crops have been studied using LCA to assess their performance (Harun et al., 2021).

2.3. Socio-Economic Assessment

Social-economic sustainability establishes a safe and prosperous workplace that supports humanity's wellness and needs (Kucukvar et al., 2021; Kutty et al., 2020c). Considering food sectors, it has been recognized that social sustainability is assessed and measured along the food supply chain (FSC), which consists of five stages: production, processing, wholesale, retailer/food services, and consumer (Desiderio et al., 2021). Social impact on sustainability could be understood once being analyzed at each stage of the food supply chain, as mentioned by Desiderio et al. (2021) and several authors in their systematic review that investigated social sustainability in the food sector and realized that most papers are done in that area studies tools used to assess the social impact in specific stages of the FSC only, without considering its effect on others.

Considering the growth of the population living in the urban areas, which is expected to increase by 2050 to reach 70% of people as indicated by the Food and Agriculture Organization. That increase will lead to enormous challenges facing conventional food production and supply chain that drive food and nutrition insecurity to urban and rural residents. Not using the resources efficiently due to the impractical practices of farmers and staffing along the FSC systems is the reason behind that. However, these unsustainable practices and degraded natural resources can once be fixed and standardized to link rural and urban communities since they are vital in designing stable and comprehensive linkages. A few countries and organizations recently acknowledged the importance of sustainable food systems and initiated guidelines and efficient practices to improve the food sector. For example, the Milan Urban Food Policy Pact lets over 120 cities get involved in developing food systems based on sustainability and social justice as well. At the UN Conference on "Housing and Sustainable Urban Development" in 2016, a new urban plan was established to regulate the international efforts revolving around urbanization for the next two decades. Moving to local initiatives adopted by a few countries, Sri Lanka, for instance, had initiated a fertilizer plant for generating compost from the solid wastes collected from cities. The Sri Lanka urban council placed this plant in the rural area to provide farmers with an easily accessible organic fertilizer.

In contrast, in Argentina, they focus on teaching farmers effective agricultural practices besides providing them with technical and financial support, which was launched by the municipality of Rosario to switch to ecological agriculture (FAO,2017). In the

United States, people lack access to fresh food stores or take a long distance to reach them, as discovered by Benez-Secanho et al. (2021), who studied this phenomenon in Georgia, which suffers from this phenomenon the freshest food deficiency. Sophisticated space-related tools in geographical information systems (GIS) are used to spot fresh food stores and consider the population density as the dependent variable in the spatial lag regression model to figure out the factors affecting the accessibility of fresh food in Georgia (Benez-Secanho et al., 2021). Another universal assessment tool developed by the "Food and Agricultural Organization of the United Nations (FAO)" to assess the sustainability systems across each dimension is the "Sustainability Assessment of Food and Agriculture systems (SAFA)." It has been used among government, academia and research, private sectors, and research and projects as it helps to implement SAFA guidelines designed by FAO (FAO,2016). "Response-Inducing Sustainability Evaluation (RISE)" assesses the sustainability dimensions across farming operations discovered by the Swiss college of Agriculture in 2011. It is an interview-based way that collects information about 54 different parameters rated by farmers using a scale from 1 (worst) to 100 (perfect) are then being summarized into ten indicators their scores are displayed in a radar chart for further studies (Grenz et al., 2012). Furthermore, an economic analysis has been done using life cycle costing assessments of agriculture food systems specialized in grapes crops production (Roselli et al., 2020).

2.4. Eco-environmental Assessment

A critical dimension of sustainability is the eco-environmental dimension, a combination of the economic and environmental impact of the system (Saling, 2016; Kutty et al., 2020). Eco-efficiency is defined as a proportion of economic output to environmental effect (Abdella et al., 2021f). Being able to assess and quantify its benefits in terms of materials and products significantly influences future improvement and takes further resolutions. It requires a holistic look at the system to build new processes and products that adopt sustainable principles(Elhmoud et al., 2020; Elhmoud et al., 2020a). Many techniques were established to analyze the entire product life cycle from early implementation. Eco-efficiency assessment has set standards for industries and products sustainability by researchers, as identified by Abdella et al. (2020). A novel approach using "Economic Input-Output Life Cycle Assessment (EIO-LCA)" has been used to assess the effects of consumption and production tasks considering all possible impacts from the supply chain (Abdella et al., 2020). This approach is regarded as a top-down strategy that utilizes environmental indicators and monetary flows to decide (Kucukvar et al., 2019). EIO-LCA, associated with Data Envelopment Analysis (DEA), has been used by Egilmez et al. (2014) to assess the sustainability dimensions in the United States for food manufacturing industries. Another methodology adopted by Park et al. (2016) utilized ecologically-based life cycle assessment (Eco-LCA) to establish a sustainability benchmarking model for 54 agriculture and food organizations in the United States that provide a benchmark for land and water footprints beside ecological resource consumption and atmospheric emissions. Eco-LCA is considered a complement to the EIO-LCA since it has additional ecological footprint categories for renewable and non-renewable resources (Tatari & Kucukvar, 2011). To address the difficulties of regression-based weights methods that are considered to be one of the recent sustainability models and give valuable relative weights for eco-efficiency composite indicators, a unique weighting methodology integrating linear mixed-effect models with Johnson's relative weights was developed by Abdella and several researchers (Abdella et al., 2021g). Life cycle assessment coupled with economic equilibrium modeling to prepare a feasible, realistic plan for the aquaculture sector has been studied by Bohnes et al. (2022), whereas Spykman et al. (2021) have initiated a newly modular method for assessing the eco-efficiency of the production of dried Hermetia illucens larvae followed by environmental life cycle and cost assessment to analyze these two sustainability dimensions.

2.5. DEA for sustainability studies:

Charnes et al. (1978) presented data envelopment analysis (DEA) as a non-parametric technique for assessing the efficacy of a collection of similar decision-making units (DMUs) wherein one or more inputs are utilized to produce one or more outputs (An et al., 2015; Cook et al. 2009; Wu et al. 2016). The main idea underlying DEA is to maximize the ratio of the sum of weighted outputs to overall weighted inputs for the DMU under consideration while ensuring that the ratios of all other DMUs are smaller than one. The technique is repeated for all DMUs, and the highest ratios obtained are indicated as the DMUs' efficiency ratings.Wong et al. (2008) and Mahdiloo et al. (2011) revealed several benefits of DEA that contribute to its popularity and effectiveness compared to other methods that are: DEA is an effective method for determining the relative efficiency of DMUs when several assessment criteria are present, second, the "efficient frontier" defines best practice as an evidence-based level of excellence that serves as a benchmark and identifies the best strategies that the least efficient DMUs could have used to improve. In the presence of several evaluation metrics, DEA is an effective method for determining the relative efficiency of DMUs.According to studies, it has been proved that DEA models serve in several contexts of sustainability, especially the traditional model that translates the data and deals with undesired outcomes as inputs (Choi & Zhang, 2011). DEA models have been used extensively to clarify and cope with the challenges that face food systems locally, regionally, and globally. Several researchers used the DEA model to investigate the food and beverage industry performance efficiency in Thailand and Vietnam to understand its capability and suggest productivity improvements coupled with the resampling method to quantify the efficiency of 40 enterprises, 20 Vietnamese and 20 Thai companies, as indicated by Wang et al., (2020). Following an investigation, it was discovered that DEA has some drawbacks related to high dimensional input indicator space and their association with each other because as the correlation of the inputs increases, either positively or negatively, along with the number of indicators, the accuracy of the DEA results will be affected (López et al., 2016; Chen et al., 2021).

2.6. Modeling and Optimization Techniques

Several modeling techniques have been applied in the area of food system assessments. Modeling and optimization techniques help identify the relationships between indicators collected and utilized with the system's responses, minimizing production costs and enhancing quality. One of the most used techniques is Data envelopment analysis (DEA). DEA models have been used extensively to clarify and cope with the challenges that face food systems locally, regionally, and globally. For example, several researchers used the DEA model to investigate the food and beverage industry performance efficiency in Thailand and Vietnam to understand its capability and suggest productivity improvements coupled with the resampling method to quantify the efficiency of 40 enterprises, 20 Vietnamese and 20 Thai companies, as indicated by Wang et al., (2020). Another modeling technique that has been recently developed used by Olagunju et al. (2021) "Directed Acyclic Graphs (DAG) "beside a Vector Error Correction Model (VECM), which is used to assess the relationship between phosphate rocks, plant fertilizers, and wheat selling prices to give perception for the concerned people about how to respond to phosphate rocks supply shocks. Müller et al. (2020) used the following multiple modeling techniques metamodeling techniques, non-equilibrium approaches, and behavioral-based modeling endeavors to establish a detailed reflection on three multiple subjects food security volatility, technology, and transformation. A system dynamic (SD) modeling approach was used by Sampedro et al. (2020) to determine the primary reasons that drive Galapagos island food systems to develop plans and test the impact on the supply system structures. Lastly, few researchers in the United States have used simulation models to visualize food system capacity and evaluate its performance (Conrad et al., 2018).

2.7. Indicators and Variable Selection Approaches

The models built for quick monitoring of results can benefit from variable selection approaches since they simplify the modeling process and improve the accuracy of the models (Abdella et al., 2016h; Abdur-Rouf K et al., 2018; Abdella et al., 2019i). This literature highlights some widely used variable selection approaches and their uses. Using variable selection techniques within the analysis removes the irrelevant attributes, which reduces the cost of processing unrequired data on the model (Sagar et al., 2021). However, choosing an indicator set that gives a complete system picture. Considering many indicators would raise the expenses of collecting and analyzing them to study the system (Reisi et al., 2014).

Furthermore, the variables can be indicators reflecting a particular set of data to explore or measure a model, where indicators must be limited, valuable, and well figured out. Utilizing many indicators is undesirable and complicates decision-making (Reisi et al., 2014). Grenz et al. (2009) have identified the most common indicators that researchers have used to investigate the sustainability dimensions of food industries, as illustrated in Fig 1.

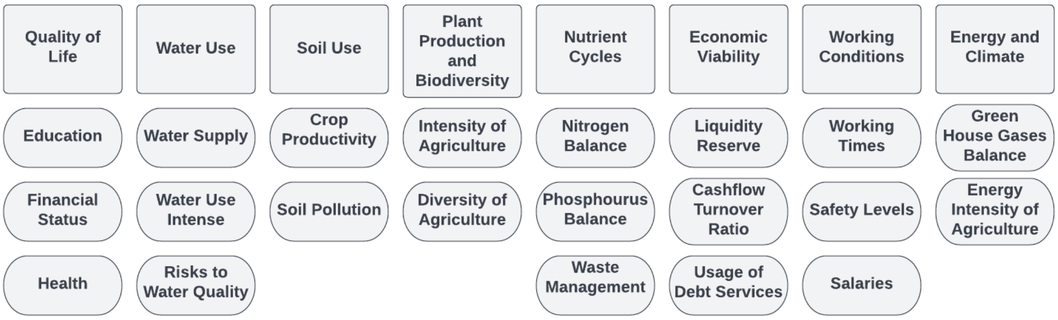

Figure 1: Most used indicators in the previous studies.

One of the most widespread approaches used is the analytical hierarchy process (AHP) which is a method that associates weights to the indicators in the model. Still, this method is subjective and inconsistent because it is based on various individuals' viewpoints (Reisi et al., 2014). Nevertheless, it has been applied to the process of selecting transport sustainability indicators to evaluate transport sustainability strategies in local governments in Taiwan, where the weights of the indicators and the selection of them for the governments have been developed through a panel of committee members (Shiau et al., 2013). Zheng et al. (2013)developed performance metrics for evaluating transportation sustainability has used AHP to assign relative weights to the variables, often determined by an expert panel based on theoretical backgrounds and considerations. However, this can pose the same problems as choosing a weighting pattern by a panel of experts. The principle component technique (PCA) compares distinct indicators on many aspects and ranks them (Reisi et al., 2014). "The least absolute shrinkage and selecting operator (LASSO)" is another method used mainly to set certain variables to further using it as a regression technique, where LASSO studies the correlation between the independent variable (x) and the dependent variable (y) to select the most valuable variables (Marami Milani et al., 2016; Sagar et al., 2021; Abdella et al., 2020). To find the potentially relevant indicators based on exploratory data analysis, the Adaptive LASSO approach is used before the prediction of air quality and pollutants in multiple cities in India, where the use of adaptive LASSO resulted in pointing out essential components that affect the air that is measured (Sethi et al., 2021). Variable selection approaches are recently applied in many sustainability studies to select the best-fit indicators to measure the progress on sustainability (Reisi et al., 2014). For example, the LASSO approach is utilized in a paper to determine the best regression models between milk's essential components (protein, fat, and milk yield) as predictions and environmental factors as predictors (Marami Milani et al., 2016). Another study used the LASSO technique and a cross-validation approach to measure weather conditions' impact on pedestrians' injury (Abdella et al., 2020d).

Feature selection in data science and machine learning seeks to exclude the less important indications before completing the study with strong data analysis tools(Tang et al., 2014). This study suggested employing a multistage DEA technique because of the powerful analysis that DEA can perform. However, one of the key downsides of DEA is its potential sensitivity to the number of inputs and outputs chosen since it does not analyze their appropriateness, and its accuracy decreases as input indicators have a strong association with one another. Variable selection in DEA seeks to pick the smallest collection of variables possible to (1) influence the efficacy of the production function approximation; (2) estimate the actual inefficiency distribution of each observation; and (3) provide more significant inputs for a clearer understanding of the production conversion from inputs to outputs. Excluding insignificant variables in DEA, whether inputs or outputs, allows frontiers to be adequately calculated and eliminates the computational burden associated with such datasets. DEA loses explanatory value as the spatial space expands (Nataraja et al., 2011). While bigger datasets are better, the DEA literature has identified several basic criteria to be followed. Golany et al. (1989) state that a wider collection of DMUs allows for more precise identification of the normal connections between the set's inputs and outputs. They suggest that DMUs should be at least twice the number of inputs and outcomes evaluated. According to Boussofiane et al. (1991), the number of inputs and outputs considered should be less than the total number of DMUs for successful discrimination and flexibility in weight selection. As a result, to address the DEA limitations related to accuracy and dimensional space complexity in the food sustainable systems domain and the need to conduct a food sustainability study to achieve food security, this thesis selected to utilize the LASSO approach to minimize dimensional space complexity. The benefits of employing LASSO over other regression-based approaches include that only LASSO achieves genuine dimensionality reduction by forcing numerous beta coefficients to be zero. In contrast, Ridge and Elastic Net force tiny coefficients near zero. The penalty factor in LASSO affects how many features are maintained; utilizing cross-validation to calculate the penalty factor ensures that the model will generalize well enough to subsequent sample data, which will be explained in detail in chapter 3.

CHAPTER 3: METHODOLOGY

This section introduces the penalization-based DEA methodology. The methods and tools used are presented, such as LASSO-based regression and DEA.

3.1. Proposed Methodology

This study proposes a two-stage methodology for estimating operational efficiency; see Figure 2. The penalization-based regression is firstly applied to reduce the dimension space of the economic and environmental indicators. Second, the non-weighted DEA, weighted DEA, and whole data set models estimates operational efficiency using the reduced dimension of the indicators. Three DEA-based efficiency models have been generated in this study to test the applicability of the proposed method.

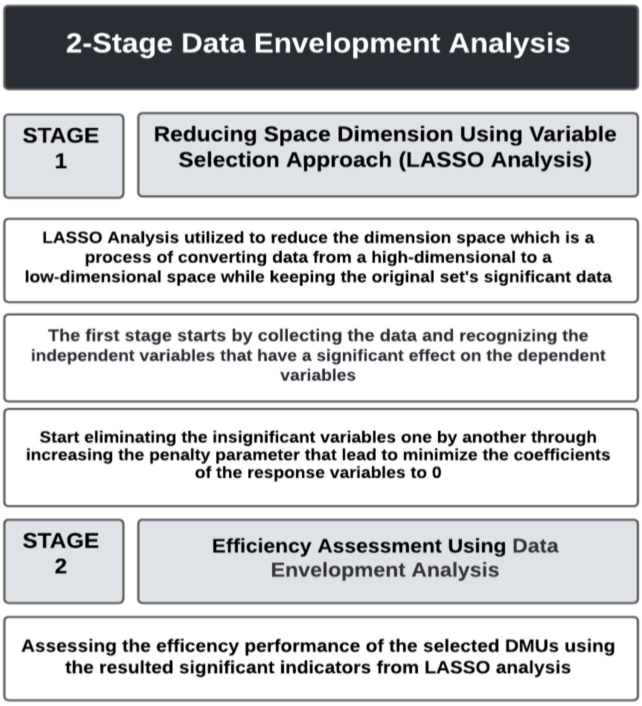

Figure 2: Two-stage LASSO based DEA for eco-efficiency assessment

3.2. LASSO-based Eco-Environmental Indicator Selection

This study utilizes penalized regression to reduce the dimension of the eco-environmental indicators only to include the most significant indicators. This step will provide the management with insight into the most significant eco-environmental indicators. However, in practice, the LASSO (least absolute shrinkage and selection operator) and the ridge penalization are the most frequently used penalty functions. The shrinkage is intended to avoid overfitting the eco-environmental indicators. Both of these penalty functions are capable of shrinking regression parameters to zero. However, only the LASSO penalty can eliminate the regression coefficients. Due to this property, LASSO-based regression is the most well-known technique for high-dimensional applications. For further reading, see Hoerl & Kennard (1970), Verweij & Van Houwelingen (1994), and Tibshirani (1996 and 1997).

The selection of eco-environmental indicators is formulated using a classical linear model expressing the value-added (response variable y_i) as a linear function of one or more eco-environmental indicators (model predictors x's) as follows:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i \qquad i = 1,2, \ldots \tag{1}$$

where $x_{ij} \in \mathbb{R}^n$ is the ith observation of the jth indicator. The β_j values are the coefficients of the eco-environmental indicators in the model. The error term ε is usually assumed to have a normal distribution.

The LASSO based solution is often formulated as an optimization problem as follows:

$$\beta^{\wedge Lasso} = (\boldsymbol{y} - \boldsymbol{X}\beta)^{T}(\boldsymbol{y} - \boldsymbol{X}\beta) + \lambda \sum_{j=1}^{p} |\beta_j| \tag{2}$$

Here $\boldsymbol{X}$ is the design matrix,p is the dimension space (number of indicators), $\lambda \geq 0$ is a pre-chosen penalization (or shrinkage) parameter that can be selected to achieve a particular penalty strength, $\|\boldsymbol{\beta}\|_1$ is the squared Euclidean norm, and the notation $\lambda\|\boldsymbol{\beta}\|_1 = \lambda \sum_{j=1}^{p} |\beta_j|$ represents the LASSO penalty function. The vector $\beta^{\wedge Lasso}$

represnts the regression coefficients of the eco-environmental indicators involved in the study. However, the penalization parameter selection, λ, is critical for the regression model's accuracy. When λ=0, the optimization problem in Equation (1) is considered a general linear regression (no dimension reduction). However, numerous techniques for determining the optimal value of this parameter have been developed and validated over the years; see, for example, Tibshirani (1996) and Park & Hastie (2007). This study uses the LASSO regression in Equation (2). The results of the LASSO regression will be reported in the case study.

3.3. DEA-based Eco-efficiency Assessment

A non-radial Slacks-Based Measure (SBM) Data Envelopment Analysis (DEA) model is used to assess the efficiency of ***n*** Decision-Making Units (DMUs). Each DMU generates ***s*** outputs from ***m*** inputs. The i^{th} input and r^{th} output of DMU_j are indicated as x_{ij} where, i = 1, 2,...,m and y_{rj} where, r = 1, 2,...,s, respectively. The production possibility set (PPS) is defined using the non-negative combination of the DMUs in the set *J* as;

$$P = \left\{ (x_i, y_r) : x_i \geq \sum_{j=1}^{n} \lambda_j x_{ij},\ 0 \leq y_r \leq \sum_{j=1}^{n} \lambda_j y_{rj},\ \sum_{j=1}^{n} \lambda_j = 1,\ \lambda_j \geq 0;\ \forall\ j, i, r \right\} \quad (3)$$

$\lambda_j = (\lambda_1, \lambda_2, \ldots..\lambda_{nj})^T$ is called the intensity vector. The SBM model presented by Tone (2001) under the variable returns to scale form is presented in Eq. 4, as follows;

$$\min p_d^* = \frac{1 - \left(\frac{1}{m}\right) \sum_{i=1}^{m} \frac{s_i^-}{x_{id}}}{1 + \frac{1}{s} \sum_{r=1}^{s} \frac{s_r^+}{y_{rd}}} \quad (4)$$

The inequalities in (3) can be transformed into equalities by introducing slacks, which forms the set of constrains of the non-oriented SBM model in Eq. (4) as follows;

$$x_{id} = \sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- \ (i = 1,2,\ldots,m) \quad (5)$$

$$y_{rd} = \sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+ \ (r = 1,2,\ldots,s) \tag{6}$$

$$\sum_{j=1}^{n} \lambda_j = 1 \tag{7}$$

$$\lambda_j \geq 0 \ (j = 1,2,\ldots, n)$$

$$s_i^-, s_r^+ \geq 0 \ , \forall \ i, r$$

The proposed SBM model minimizes the mean rate of input and maximizes the inverted mean rate of output through $1 - (\frac{1}{m}) \sum_{i=1}^{m} s_i^- / x_{id}$ and $1 + \frac{1}{s} \sum_{r=1}^{s} \frac{s_r^+}{y_{rd}}$ respectively. $s_i^- = (s_1^-, s_2^- \ldots, s_m^-)^T \in R^m$ and $s_r^+ = (s_1^+, s_2^+ \ldots, s_s^+)^T \in R^s$ are the slacks of the inputs and outputs of the DMUs. Then these slacks are used in the objective function to assess the efficiency of the DMUs. When this model for DMU_d is solved, the optimal value of the objective function $(\rho^*{}_d)$ is obtained, and DMU_d is considered efficient when $\rho^*{}_d = 1$ (Seiford and Zhu 2002; Yang and Pollitt 2009; Badau 2015).

CHAPTER 4: DATA ANALYSIS AND RESULTS

In this chapter, the proposed approach (2-Stage LASSO-based DEA) has been described using an actual case study of 30 food industries in the United States to provide concerned individuals with an indication of plans to enhance their food systems.

4.1. Data Generation

The statistics used to evaluate the sustainability performance of the 30 food industries in the United States were developed using an input-output model based on data from the EORA database, which has proved to be an accurate international database whose data has been used in numerous significant researches(Wiedmann et al., 2015). Eora comprises national input-output tables that span roughly the international economy (Lenzen et al., 2013). Eora is based on credible data sources, such as the UN System of National Accounts and the COMTRADE databases, and several national organizations like Eurostat and IDE/JETRO (Sen et al., 2020). In Eora, input-output tables are generated by converting Supply and Use Tables (SUTs) from 190 nations into Make and Use matrices. Information is collected using specialized satellites. For this research objective, the 429 US domestic SUTs were coupled with various environmental variables via the specialized environmental satellite. The multiplier data were acquired from the input-output model to estimate the environmental consequences of the food consumption sectors in the United States. The information gathered was from the year 2015. Energy use, renewable resource consumption, air pollution, water footprint, and non-renewable resource consumption are the six major categories of environmental indicators. The input-output model for the US economy has been used to determine the multipliers for US food sectors following obtaining environmental effect data for the US food consumption sector and matching food consumption in monetary units ($). Sen et al. (2019) give extensive instructions on creating input-output models and their formalization for further understanding.

4.2. Data Description

Fourteen indicators were retrieved from the Eora database to be utilized for the 30 food industry analyses. The selection of these 14 indicators and food industries shown in Tables 1 and 2 is based on a study of existing research in that sector and the influence of these indicators on food industry eco-efficiency based on the literature review.

Table 1: Selected environmental Indicators.

Environmental Indicator	Abbreviation	Symbol	Unit	Mean
	CO2	X1		5.5E+09
	CH4	X2		1.5E+08
	CO2	X3		8.0E+07
	HFC	X4		1.3E+07
Air Pollutants	PM10	X5	Kt	9.6E+06
	PM2.5	X6		3.7E+02
	N20	X7		1.0E+07
	VOC	X8		1.8E+07
	SO2	X9		1.7E+07
Minerals	M	X10	t	9.3E+10
Mining and Quarrying	MQ	X11	t	6.1E+08
Non-Renewable Energy Consumption	NREC	X12	TJ	9.6E+10
Renewable Energy Consumption	REC	X13	TJ	6.9E+09
Water Footprint	WF	X14	Mm3/ yr	2.5E+10
Economic Value added	GDP	Y	$M	9.6E+09

The indicators evaluated to execute the suggested approach are shown in Table 1. The independent variables (x1,x2...x14) are the environmental indicators, and the dependent variable (Y) is the economic value-added, where they will be used in the first stage of the proposed method (LASSO analysis) to identify the significant indicators before moving on to the second stage.

Table 2: Food Industries.

	Food industries	Abbreviation
1.	All other food manufacturing	AOFF
2.	Beet sugar manufacturing	BSM
3.	Bread and bakery product manufacturing	BBPM
4.	Breakfast cereal manufacturing	BCM
5.	Breweries	BW
6.	Cheese manufacturing	CM
7.	Chocolate and confectionery manufacturing from cacao beans	CCCB
8.	Coffee and tea manufacturing	CTM
9.	Confectionery manufacturing from purchased chocolate	CMPC
10.	Cookie, Cracker, and pasta manufacturing	CCPM
11.	Dog and cat food manufacturing	DCFM
12.	Dry, condensed, and evaporated dairy product manufacturing	DCEPM
13.	Fats and oils refining and blending	FORB
14.	Flour milling and malt manufacturing	FMMM
15.	Fluid milk and butter manufacturing	FMBM
16.	Frozen food manufacturing	FFM
17.	Fruit and vegetable canning, pickling, and drying	FVPD
18.	Ice cream and frozen dessert manufacturing	ICFM
19.	Non-chocolate confectionery manufacturing	NCM
20.	Other animal food manufacturing	OAFM
21.	Poultry processing	PP
22.	Seafood product preparation and packaging	SPP
23.	Seasoning and dressing manufacturing	SDM
24.	Snack food manufacturing	SFM
25.	Soft drink and ice manufacturing	SDIM
26.	Soybean and other oilseed processing	SOP
27.	Sugar cane mills and refining	SCMR
28.	Tortilla manufacturing	TM
29.	Wet corn milling	WCM
30.	Wineries	WINE

4.3. Selecting sustainability indicators using LASSO

This section introduces the steps to select the best subset of indicators to model the operation efficiency of the corresponding food and beverage industries. However, with the current development in data collection methods and techniques, it becomes crucial to practice several dimension reduction techniques to overcome the over-fitting issues in various scientific and business fields. This is widely known in statistics as "Model Selection." Model selection refers to choosing the best subset of model predictors (sustainability indicators) to enhance the model performance. Traditional regression methods, such as simple linear regression and multiple regression using least square estimation methods, have some drawbacks under high dimension settings, such as assigning nonzero values to all model predictors, making them difficult to interpret, and producing overfitted model having low model prediction performance.

This study uses LASSO-based regression as a very well-known tool for model selection. Lasso regression is a type of linear regression that uses shrinkage. In addition, lasso regression performs L1 regularization, which adds a penalty equal to the absolute value of the magnitude of coefficients. This type of regularization can result in sparse models with few coefficients; Some coefficients can become zero and be eliminated from the model.

On the other hand, larger penalties result in coefficient values closer to zero, ideal for producing simpler models. However, LASSO improves the model's stability and accuracy by continuous shrinkage and only includes the most significant model predictors (indicators). The model selection provides the decision-makers with a simple, accurate, and interpretable model.

Since the LASSO outcome depends on the value of the shrinkage parameter, several ways can be used to select the best model. For instance, K-fold cross-validation, Mean

Square Error (MSE) and Mean Absolute Deviation (MAD). This study applies the K-fold cross-validation to optimize the shrinkage parameter. The procedures for the LASSO application will be described in the following sections.

4.3.1. Sustainability Impact Normalization

The sustainability impacts of the selected indicators contain different scales and units. Therefore, the "Feature Scaling" method is applied to have comparable indicators. The normalization step helps make the data comparable across all the indicators so that the information can be combined meaningfully. The normalized measures can be calculated using:

$$x'_{ij} = b_0 + \frac{\left(x_{ij} - \mathrm{X}_{min,j}\right)(b_1 - b_0)}{\left(\mathrm{X}_{max,j} - \mathrm{X}_{min,j}\right)}; \quad i = 1,2, \ldots\ldots, n\,, \quad j = 1,2, \ldots, p \qquad (8)$$

where x_{ij} ix the i^{th} observation under the j^{th} sustainability indicator, $\mathrm{X}_{max,j}$ and $\mathrm{X}_{min,j}$ are the maximum and minimum values of the j^{th} indicator, respectively, n is the number of food and beverages industries (n=30), p is the number of indicators (p=14), and $b_0 < b_1$ are predetermine min-max values for the range of n'_{ij}. This study uses b_0= 0 and b_1=1. e set at zero and 1. The normalized sustainability matrix is reported in Appendix A.

4.3.2. Measure Correlation

In this section, we measure collinearity among the sustainability indicators; see Figure 3. This step is added to justify the importance of penalization regression to avoid overfitting. This study uses the correlation of determination (r^2) as the most popular method for correlation measure. The r^2 statistic determines the percentage of variation in one model predictor (indicator) that is predictable from the other predictors (indicators). The r^2 can take any value from 0 to 1. The r^2 means more percentages can be predicted and vice versa. This study uses r_{ij}^2 to refer to the coefficient of determination between the i^{th} and j^{th} indicators. The r_{ij} values, refer to the sample-based Pearson correlation coefficients, for all the pairs of the sustainability indicators were estimated and reported in the correlation matrix below:

		X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X4
		CO2	**CH4**	**CO**	**HFC**	**PM10**	**PM2.5**	**N20**	**VOC**	**SO2**	**M**	**MQ**	**NREC**	**REC**	**W**
X1	**CO2**	1	0.727	**0.976**	0.823	0.937	0.661	0.89	0.949	0.936	0.898	0.932	0.958	0.964	0.653
X2	**CH4**	0.727	1	0.757	**0.458**	0.711	0.539	0.758	0.603	0.541	0.551	0.737	0.621	0.602	0.56
X3	**CO**	0.976	0.757	1	0.821	0.933	0.715	0.914	0.933	0.905	0.872	0.944	0.932	0.941	0.681
X4	**HFC**	0.823	0.458	0.821	1	0.906	0.630	0.808	0.907	0.934	0.806	0.769	0.847	0.863	0.527
X5	**PM10**	0.937	0.711	0.933	0.906	1	0.738	0.928	0.957	0.939	0.805	0.89	0.906	0.913	0.686
X6	**PM2.5**	0.661	0.539	0.715	0.63	0.738	1	0.87	0.68	0.646	0.476	0.779	0.709	0.68	0.931
X7	**N20**	0.89	0.758	0.914	0.808	0.928	0.87	1	0.872	0.85	0.7	0.928	0.874	0.865	0.834
X8	**VOC**	0.949	0.603	0.933	0.907	0.957	0.68	0.872	1	0.971	0.89	0.891	0.945	0.965	0.642
X9	**SO2**	0.936	0.541	0.905	0.934	0.939	0.646	0.85	0.971	1	0.881	0.863	0.933	0.953	0.59
X10	**M**	0.898	0.551	0.872	0.806	0.805	0.476	0.7	0.89	0.881	1	0.828	0.882	0.888	0.464
X11	**MQ**	0.932	0.737	0.944	0.769	0.89	0.779	0.928	0.891	0.863	0.828	1	0.915	0.911	0.776
X12	**NREC**	0.958	0.621	0.932	0.847	0.906	0.709	0.874	0.945	0.933	0.882	0.915	1	0.984	0.703
X13	**REC**	0.964	0.602	0.941	0.863	0.913	0.68	0.865	0.965	0.953	0.888	0.911	0.984	1	0.661
X14	**W**	0.653	0.56	0.681	0.527	0.686	0.931	0.834	0.642	0.59	0.464	0.776	0.703	0.661	1

Figure 3: Selected Indicators Correlation.

The correlation measures show that the sustainability indicators have medium to high correlations. The{CO, CO2}shows the highest correlation (0.976), while the {CH4, HFC} shows the lowest correlation.

4.3.3. Model Selection using LASSO

This section introduces the procedures for solving the LASSO regression using IBM®SPSS® software. The IBM®SPSS® is a well-known statistical platform that offers advanced techniques and methods to help make high-quality decisions.

The value of the tuning parameter, λ, is determined using the K-Fold Cross Validation (CV). This method is based on the trade-off between the value of λ and the model accuracy measured. Let y be the normalized value of the ith observation (the added-value of the ith industry); then model accuracy can be calculated as follows:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{Y}_i\right)^2 \tag{9}$$

Where n is the number of observations and $\hat{Y}_i$ is the estimated response value of the i^{th} observation. In this study, n=30 is the number of food and beverages industry.

The CV is also conducted using the IBM®SPSS®. Cross-validation tests the model's accuracy in predicting the new dataset to overcome statistical problems such as overfitting. The original dataset randomly divides the k-fold CV into k equal-sized subsets. Then, a single subsample is retained as the validation dataset for testing the model accuracy, and the remaining subsets are used as training subsets. The process is repeated *k* times, with each of the *k* subsamples used exactly once as the validation data. The *k* results can then be averaged to produce a single estimation. The CV procedure will be repeated at all suggested levels λ. Finally, the value of λ with the minimum MSE is selected as the optimal λ for conducting LASSO regression. The range of λ is selected to be from {0.01 to 1}, with an increment of 0.01. That means that 100 values λ are tested under the cross-validation framework. The run-time using the IBM®SPSS® is around 30 seconds.

The CV is available as an option under the regularization regression; see Figure 4. The number of folds is selected to 10. The CV outcome is reported in Appendix B. The optimal value of λ and the selected best model were highlighted in gray and reported in the following section.

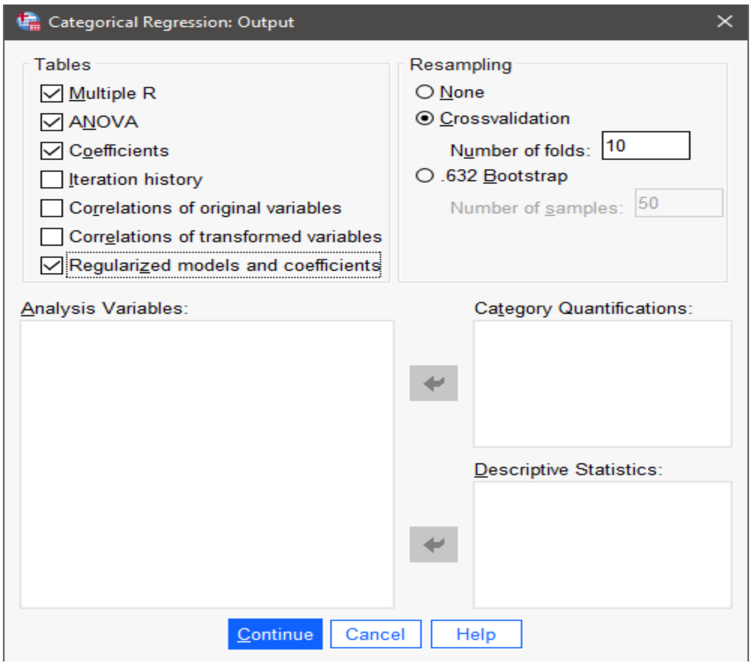

Figure 4: K-Fold Cross-validation settings under the IBM®SPSS®.

From the CV procedures, the optimal value of λ is 0.080. The LASSO regression is then conducted using this value λ. Figure 5 shows the settings of the LASSO regression using the IBM®SPSS®.

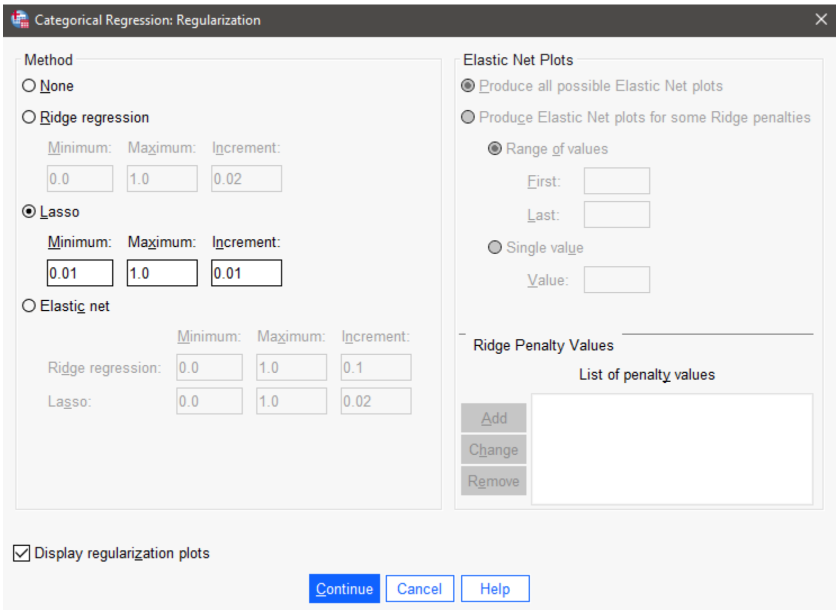

Figure 5: LASSO regression settings under the IBM®SPSS®.

The optimal model based on LASSO regression is found to be as shown in Table 3:

Table 3: Optimal model settings

	X3	X4	X8	X9	X10	X13
Indicators	CO	HFC	VOC	SO2	M	REC
Coefficients	0.01229	0.77137	0.00036	0.07317	0.07833	0.09215
Weight	0.01196	0.7506	0.00035	0.0712	0.07622	0.08967

The optimal model shows that six indicators are selected to model better the relationship between the sustainability indicators and the added-value (Gross Domestic Product (GDP). The statistics of the selected model are shown in Table 4.

Table 4: Optimal model statistics.

Penalty	Regularization "R Square"	Number of Selected Predictors	Prediction Error
0.080	0.992	6	0.008

Table 4 shows that the optimal model has achieved a very small "prediction Error" with a high R^2 value equal to 0.992. Other models, for instance, when λ=0.10 and 0.110 have shown a higher R^2 (0.994), but considering that these models have only two indicators selected, the model with λ=0.08, six indicators selected, becomes more practical.

The most critical challenge for completing the aggregation process is determining the extent to which each indicator contributes to the eco-efficiency value. It is customary to refer to the contribution level as a weight-value – or relative importance.

The operational efficiency assessment challenges large business organizations with many sustainability impacts. One main challenge is determining the relative weight – or importance of each sustainability indicator. There are two different weighting methods in practice. These are weighting methods based on expert opinion and weighting methods based on statistical approaches; see (Saisana et al., 2002). The statistics-based weighting methods are widely common as they use a data-driven approach, such as principal component analysis, factor analysis, and multiple regression. (Saisana, 2002; Reisi, et al., 2014). In this study, we are using the outcome of the LASSO methods, mainly the coefficient estimates, to refer to the weight of the selected indicator. The weights reported in Table 4 were calculated as follows:

$$\mathrm{W}_j = \frac{\beta_j}{\sum_{i=1}^{s} \beta_i} \qquad i, j = 1,2, \dots \dots, s \tag{10}$$

where W_j and β_j are the weight and the regression coefficients of the jth indicator, and s is the number of indicators of the selected model.

4.4. DEA Based Efficiency Assessment

The DEA approach has shown to be a powerful, effective analytical tool that has been employed in various sectors over the years and has provided valuable results by notable researchers internationally that assist decision-makers in enhancing their firms and practices by providing a benchmark for them. Based on the literature and research on DEA, it was discovered that DEA is sensitive to the high correlation dimensional space of indicators(López et al., 2016). As the number of indicators with relatively substantial associations increases, the accuracy of DEA decreases (López et al., 2016). Thus, the accuracy of DEA improved by carrying out the proposed first stage, which limited the indicators to those with a significant impact only. To better understand the proposed method, three scenarios will be examined using the DEA technique using the SBM DEA model run using the MaxDEA 8 Ultra software and their outcomes will be compared in the sections that follow. The three scenarios are 1) LASSO-based un-weighted DEA, 2) LASSO-based Weighted DEA, and 3) Full-dimension DEA (including the 14 indications collected).

4.4.1. Eco-Efficiency Assessment of LASSO-based un-weighted DEA

This section presents the results of the first scenario that deploys the indicators set generated from the LASSO analysis.

Figure 6 shows that DMUs BBPM, BCM, CCCB, CTM, CMPC, FMBM, FFM, FVPD, ICFM, SPP, SFM, SDIM, SOP, and WINE got 100% eco-efficiency, followed by DMUs BW and AOFM with an approximately 88% eco-efficiency, whereas all the remaining DMUs lagged at 80% efficiency, having the least efficiency percent of 43.68% and 44.18% for the OAFM and FORB industries, respectively.

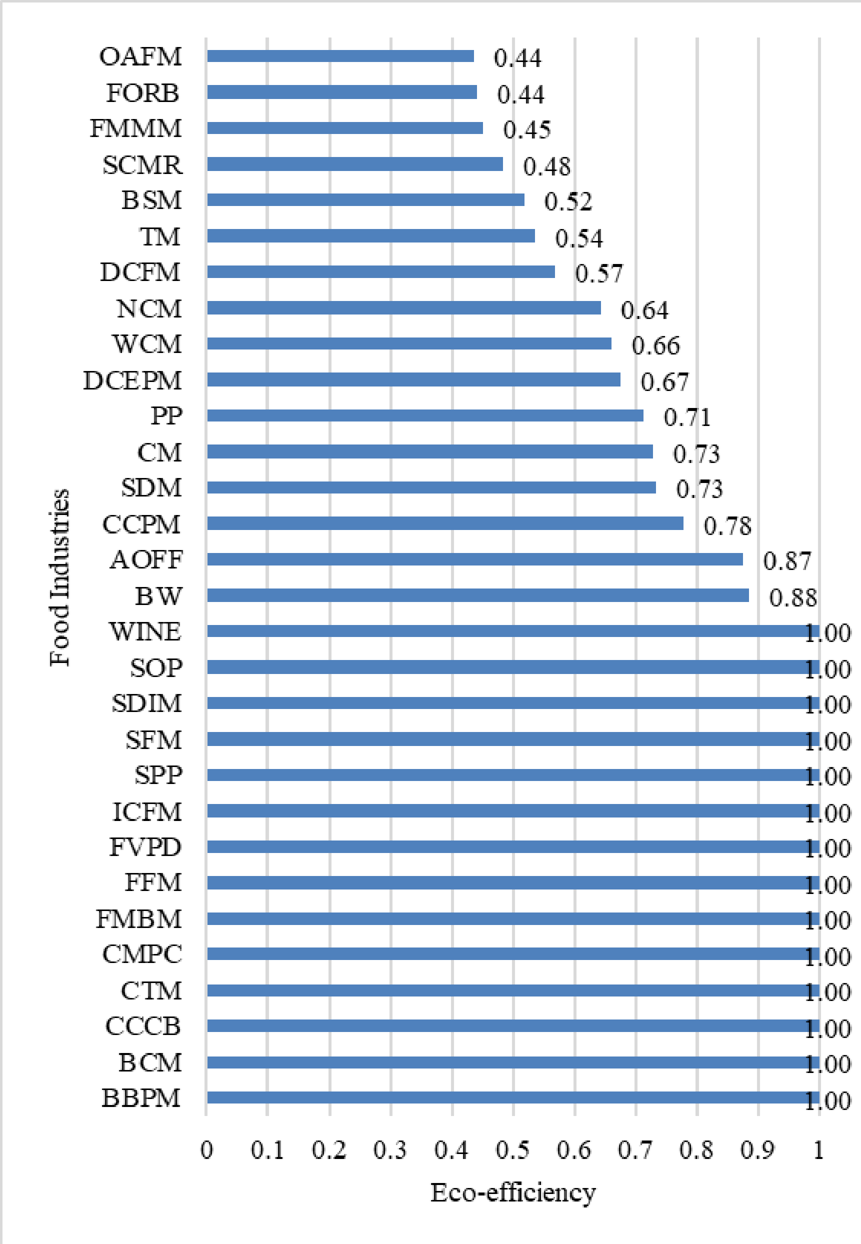

Figure 6: Eco-efficiency of LASSO-based Un-weighted DEA

4.4.2. Eco-Efficiency Assessment LASSO-based Weighted DEA

This section presents the second scenario results that deploy the indicators set generated from LASSO analysis with their associated weights generated from computed as illustrated in section 4.3.3.

Figure 7 shows that the DMUs BBPM, BCM, BW, CCCB, CTM, CMPC, FMBM, FFM, FVPD, ICFM, SPP, SFM, SDIM, SOP, and WINE adopt the best eco-efficiency among the 30 DMUs, with 100% efficiency, followed by DMU AOFF with 88.4%, whereas all the remaining DMUs fall behind 83% eco-efficiency. Whereas we have five food industries with less than 50% efficiency, the least efficient percent is equal to 30.6%, acquired by the TM industry.

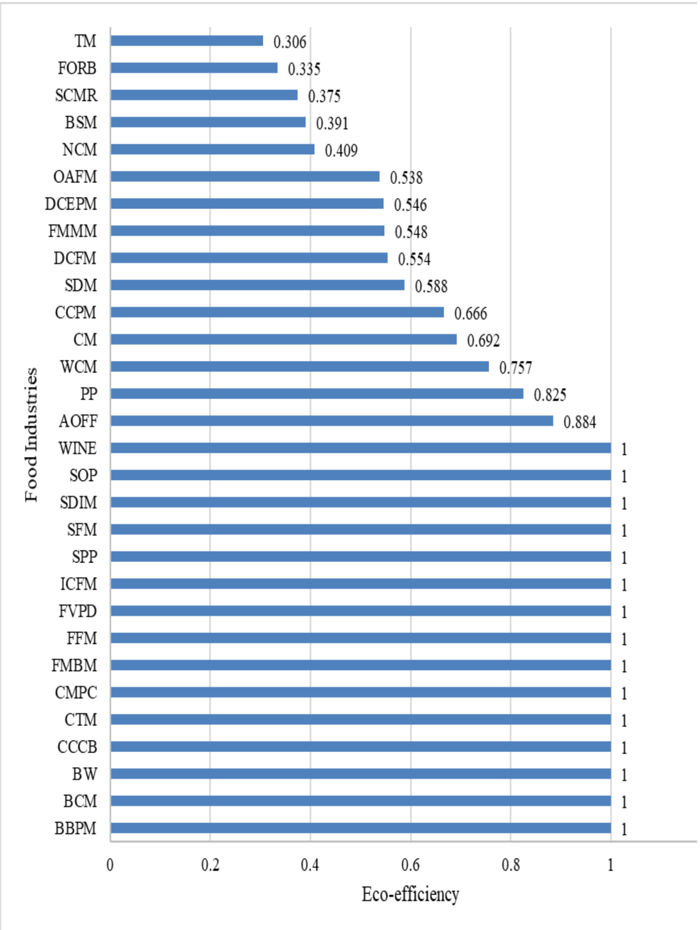

Figure 7: Eco-efficiency of LASSO Based Weighted DEA

4.4.3. Eco-Efficiency of Full Dimension DEA

This section presents the results of the third scenario that deploys the whole indicators set generated, collected, and normalized without using LASSO.

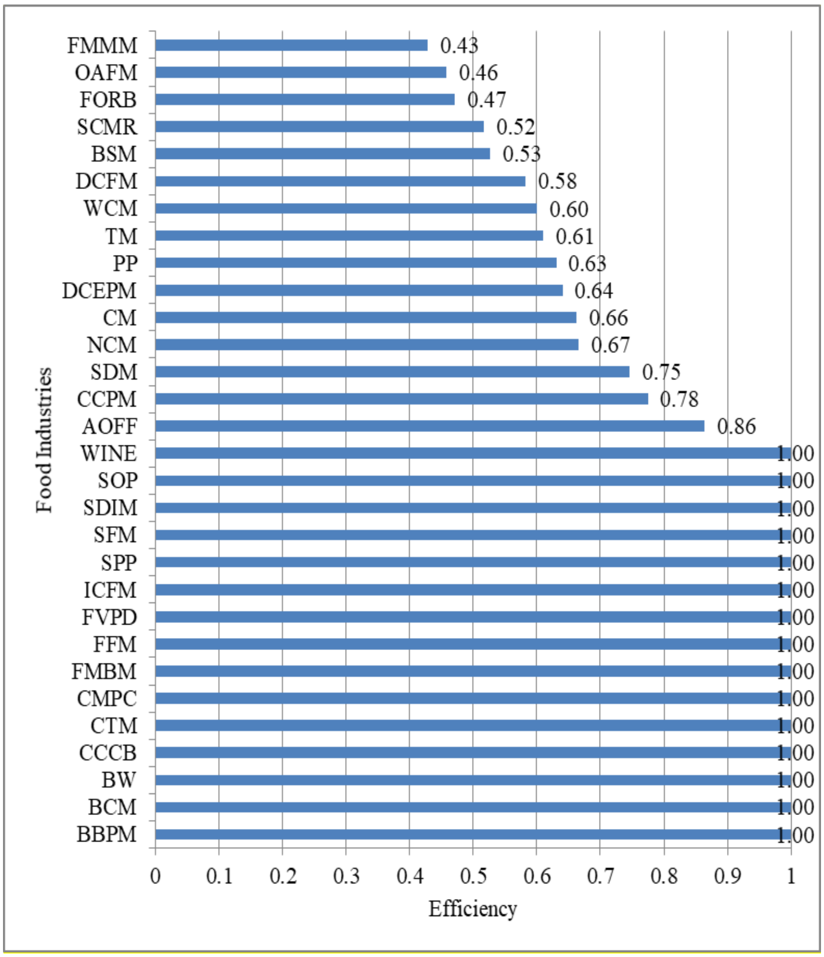

Figure 8: DMUs Eco-efficiency Using 14 Indicators.

Figure 8 shows that several DMUs have 100% eco-efficiency. These are the BBPM, BCM, BW, CCCB, CTM, CMPC, FMBM, FFM, FVPD, ICFM, SPP, SFM, SDIMM SOP, and WINE. The results showed that AOFF has an eco-efficiency of 86.4%, whereas all the remaining food industries are behind 78% eco-efficiency. Having the least eco-efficiency equals 42.87%, acquired by the FMMM industry.

4.5. Results Discussion

The outcomes of the three scenarios are illustrated in Table 5. In the first model (LASSO-based DEA-non-weighted), the top-ranked industries with 100 percent efficiency were: BBPM, BCM, CCCB, CTM, CMPC,FMBM,FFM, FVPD,ICFM, SPP, SFM, SDIM, SOP, and WINE. The highly performing industries in the second scenario (LASSO based weighted DEA) were BBPM, BCM, BW, CCCB, CTM, CMPC, FMBM, FFM, FVPD, ICFM, SPP, SFM, SDIM, SOP, and WINE, whereas the top-ranked industries in the third scenario (DEA alone) were BBPM, BCM, BW, CCCB, CTM, CMPC, FMBM, FFM, FVPD, ICFM, SPP, SFM, SDIMM SOP, and WINE. This clearly illustrates that the top-ranked industries created by both non-weighted and weighted LASSO-based DEA have 14 out of 15 same DMU efficiency, while the results from the second scenario are identical to the results from the third.

Examining the eco-efficiency of the following five ranking industries: the second, third, fourth, fifth, and sixth performing industries, the first scenario said that BW, AOFF, CCPM, SDM, and CM, respectively. In the second scenario, the results were AOFF, PP, WCMM CM, and CCPM. Finally, the third scenario, the 14 dimensional DEA, displayed the AOFF, CCPM, SDM, NCM, and CM. Observing that the similarity between the first and third scenarios was high, as four of the five DMUs were in common, while when comparing the results of the second and third scenarios, there were only three DMUs in common, concluding that the first and third scenarios got nearly the same results.

Finally, after comparing the five lowest-ranked eco-efficiency businesses, the first scenario revealed that BSM, SCMR, FMMM, FORB, and OAFM are the least efficient industries. The NCM, BSM, SCMR, FORB, and TM were in the second. And according to the findings of 14 dimensional DEA, the third scenario, BSM, SCMR, FORB, OAFM, and FMMM are the least eco-efficient performing industries. Observing that the first and third scenarios provide identical results but in a slightly different order of the industries. At the same time, the second and third scenarios have three out of four industries in common with some different sequencing.

After comparing three categories of results from the three scenarios, it was discovered that there is a high similarity between the three models, most notably between the non-weighted LASSO-based DEA and the entire dimension DEA model. This indicates that by using the most significant set of indicators only generated by LASSO, we achieved the best results rather than using a large set of indicators, implying that these selected variables and other variables primarily influence the DMUs' efficiency have little influence on the eco-efficiency. This illustrates the method's applicability while also addressing the DEA's limitation with high input/output dimensions and the difficulty in evaluating efficiency with many inputs.

Table 5: Ranking of 30 indutries among the three scenarios

Model 1		Model 2		Model 3	
Industry	Rank	Industry	Rank	Industry	Rank
BBPM	1	BBPM	1	WINE	1
BCM	1	BCM	1	SPP	1
CCCB	1	BW	1	SOP	1
CTM	1	CCCB	1	SFM	1
CMPC	1	CMPC	1	SDIM	1
FMBM	1	CTM	1	ICFM	1
FFM	1	FFM	1	FVPD	1
FVPD	1	FMBM	1	FMBM	1
ICFM	1	FVPD	1	FFM	1
SPP	1	ICFM	1	CTM	1
SFM	1	SDIM	1	CMPC	1
SDIM	1	SFM	1	CCCB	1
SOP	1	SOP	1	BW	1
WINE	1	SPP	1	BCM	1
BW	15	WINE	1	BBPM	1
AOFF	16	AOFF	16	AOFF	16
CCPM	17	PP	17	CCPM	17
SDM	18	WCM	18	SDM	18
CM	19	CM	19	NCM	19
PP	20	CCPM	20	CM	20
DCEPM	21	SDM	21	DCEPM	21
WCM	22	DCFM	22	PP	22
NCM	23	FMMM	23	TM	23
DCFM	24	DCEPM	24	WCM	24
TM	25	OAFM	25	DCFM	25
BSM	26	NCM	26	BSM	26
SCMR	27	BSM	27	SCMR	27
FMMM	28	SCMR	28	FORB	28
FORB	29	FORB	29	OAFM	29
OAFM	30	TM	30	FMMM	30

4.4 Benchmark Learning Pathways

This section performs projection analysis for the eco-efficiency first model, which used non-weighted LASSO-based DEA and had a practically identical ranking to full dimension DEA analysis (third scenario). The projection study focuses on improving the eco-efficiency of the inefficient food industry and getting a 100% score. The percentage reduction values linked with each inefficient DMU to the frontier were calculated based on the distance to frontier approach, which helps determine how much each indicator must be decreased to reach the efficient frontier. In other words, it forecasts the future activities that will be taken to improve each food business's sustainability performance. As a result, Table 6 shows the appropriate activities, represented with percentages, to improve the eco-efficiency performances of each DMU across all eco-efficiency metrics, which will be reflected in an increase in the economic value added to reach the maximum efficiency frontier.

Table 6: Projection Analysis Data.

DMU	CO	HFC	VOC	SO2	WFM	MLS
	Diff (%)					
AOFF	-32.807	-9.949	-18.05	-10.791	0	-3.736
BSM	-54.353	-66.343	-58.922	-55.075	-17.387	-36.677
BBPM	0	0	0	0	0	0
BCM	0	0	0	0	0	0
BW	-0.346	-16.48	0	0	0	-52.528
CM	-56.234	-35.663	-31.723	-6.317	-1.638	-31.383
CCCB	0	0	0	0	0	0
CTM	0	0	0	0	0	0
CMPC	0	0	0	0	0	0
CCPM	-28.444	-39.547	-21.967	-14.939	-18.087	-9.929
DCFM	-47.822	-44.053	-32.429	-15.419	-65.094	-54.861
DCEPM	-43.967	-35.291	-34.82	-22.562	-20.131	-38.282
FORB	-60.755	-71.102	-45.51	-32.376	-84.828	-40.334
FMMM	-70.296	-42.478	-56.555	-52.192	-71.244	-36.602
FMBM	0	0	0	0	0	0
FFM	0	0	0	0	0	0
FVPD	0	0	0	0	0	0
ICFM	0	0	0	0	0	0
NCM	-19.472	-60.316	-41.613	-26.64	-30.803	-35.766
OAFM	-62.371	-40.146	-46.579	-27.858	-82.424	-78.541
PP	-52.714	-11.381	-30.394	0	-46.341	-30.806
SPP	0	0	0	0	0	0
SDM	-36.253	-35.983	-30.333	-14.09	-11.277	-32.452
SFM	0	0	0	0	0	0
SDIM	0	0	0	0	0	0
SOP	0	0	0	0	0	0
SCMR	-50.842	-65.255	-56.935	-54.822	-62.182	-19.301
TM	-30.001	-69.306	-46.118	-42.394	-85.355	-4.805
WCM	-45.136	-21.685	-40.671	-39.66	-28.757	-27.79
WINE	0	0	0	0	0	0

The percentage indicates that the FMMM industry has the most significant CO reduction, with a 70% decrease, which, when reduced, will have a significant positive influence on its performance. In terms of HFC, FORB has the highest drop percentage, equaling 70%. At the same time, VOC levels in the BSM business should be decreased by 59%. Whereas SO2 should be reduced by 55% in the BSM sector, WFM should be reduced by 85% in the TM sector. Finally, OAFM necessitates a 78.5 percent decrease in MLS.

4.5 Improvement Actions

Developing novel ways to reduce the environmental effect of food production and support the universe potential to produce food in the future is a demand. According to the findings, several indicators contribute significantly on the industry's sustainability performance such as emissions and energy consumption.

Several actions could be taken to reduce carbon emissions according to the Alkaabneh et al., (2021) It has been discovered that research and development that leads to storage solutions with lower carbon emission rates has the greatest promise for reducing emissions. Carbon taxes have the ability to cut emissions as well, but at the expense of reducing production output and raising consumer costs. The fastest-growing greenhouse gases are hydrofluorocarbons (HFCs), which are chemicals commonly used in refrigeration and cooling systems. HFCs have the potential to have a large influence on climate change. Short-term initiatives to minimize HFC emissions will dramatically lower expected temperature rises over the next few decades. The American Innovation and Manufacturing Act of 2020 urges the Environmental Protection Agency (EPA) to execute a 15-year phasedown of regulated HFC production and consumption of 85 percent as well handle these HFCs and their alternatives and make the transition to next-generation technology easier (Controlling industrial greenhouse gas emissions, 2021). Choosing eco-friendly packaging is another way to reduce wastes and emissions as food packaging generates a significant amount of waste and pollution. Every year, about 78 million metric tons of plastic packaging are manufactured, with just 14 percent of that amount recycled. Because the great majority of plastic is manufactured from nonrenewable resources – either oil or natural gas – and will end up in a landfill (Royte, 2021). In terms of water footprint reductions, research has shown that decreasing the environmental effect of food waste and loss will assist to reduce water footprint

(Marston et al., 2021). Improving energy and water efficiency would help to improve the food industry's sustainable performance. Food processing and manufacturing need a lot of energy and water. Processing and manufacturing account for approximately 23% of total energy expenditure in the food business in the United States. Water is widely utilized in food preparation, both as an ingredient and in a variety of industrial operations (e.g. washing, disinfecting, conditioning, and cooking). Reducing energy and water usage can be especially difficult in the food processing and manufacturing industries, where production demands and safety regulations must take precedence. Smart metering and the adoption of energy efficiency and sustainable water management systems can assist in improving the industry sustainability (University of Michigan, 2019). Another way to improve sustaibility performance is to employ eco-friendly ingredients because the production of a variety of agricultural commodities, in particular, generates environmental and social sustainability problems. Deforestation and habitat degradation have been connected to palm oil, cocoa, and coffee, for example.

CHAPTER 5: CONCLUSION AND FUTURE WORK

5.1. Conclusion and Recommendations

Food system Food eco-efficiency assessment is important for assessing and monitoring economic and environmental management systems. The growing concern of sustainability management is attributed to the need for qualitative and quantitative management tools to provide business owners, stakeholders, and suppliers with a data-driven framework for selecting the most appropriate decisions related to both environmental and socio-economic sustainability issues. However, the eco-efficiency assessment for large organizations with multiple environmental and economic impacts is computationally expensive and complex. Therefore, there is a demand for developing a reliable and accurate method of assessing food eco-efficiency. To this point, this thesis presents a two-stage methodology integrating both dimension reduction with DEA for assessing food eco-efficiency performance under high dimensional settings of sustainability indicators. The LASSO method is well-known for its efficacy as a variable selection technique. The incorporation of LASSO into the DEA aids in resolving the DEA's issues with high-dimensional space and highly correlated indicators. The proposed method was evaluated and validated using a real-world dataset that represented the environmental impact of fourteen indicators across 30 food and beverage industries. The food and beverage industries dataset was obtained from the Eora database.

The first stage of the proposed methodology is known as dimension reduction. This stage has been complet3ed using IBM-SPSS software. The outcome of this stage resulted in a new dimension of six sustainability indicators with MSE=0.008. In addition, the new subset has shown a correlation coefficient equal to 0.992.

The second stage of the proposed methodology is the eco-efficiency-based DEA. This step has been completed using the SBM DEA model run using the MaxDEA 8 Ultra software. Three DEA eco-efficiency models were constructed in this study. These are 1) LASSO-based un-weighted DEA, 2) LASSO-based Weighted DEA, and 3) Full-dimension DEA. An eco-efficiency rank comparison of the three models has been conducted, and important conclusions have been extracted.

Generally speaking, the results revealed that the three eco-efficiency models performed quite similarly. However, the non-weighted LASSO-based DEA and the full-DEA models demonstrated the highest similarity. This finding allows the management decision-makers to interpret the DEA-based ecoefficiency outcomes using few indicators. Also, it demonstrates the proposed method's applicability in addressing the DEA's weaknesses with high input/output dimensions and the difficulty in interpreting efficiency with a large number of inputs. The LASSO-base weighted DEA has shown a reasonable performance compared with the LASSO-based unweighted DEA.

The regression coefficients were used as weights in this study. However, other weighting techniques, such as PCA, and Elastic Net, can be used in conjunction with the proposed method to improve the LASSO-based weighted DEA. Additionally, machine-learning-based clustering is a possible direction for future work, in which comparisons are made based on cluster group rather than rank.

REFERENCES

Abdella G. M., Al-Khalifa, K. N., Kim, S., Jeong, M. K., Elsayed, E. A., and Hamouda, A.M.S., Variable Selection-based Multivariate Cumulative Sum Control Chart, Quality and Reliability Engineering International, Vol. 33, pp. 565–78, 2017c.

Abdella, G. M., and Shaaban, K., Modeling the Impact of Weather Conditions on Pedestrian Injury Counts Using LASSO-Based Poisson Model, *Arabian Journal for Science and Engineering*, pp. 1-12, 2020d.

Abdella, G. M., Kim, J., Al-Khalifa, K. N., and Hamouda, A. M. S., Penalized Conway-Maxwell-Poisson regression for modeling dispersed discrete data: The case study of motor vehicle crash frequency, *Safety Science*, Vol. 120, pp. 157-63, 2019i.

Abdella, G. M., Kim, J., Al-Khalifa, K. N., and Hamouda, A.M.S., Double EWMA-based polynomial profile monitoring, *International Journal of Quality and Reliability*, Vol. 32, pp. 2639-52, 2016h.

Abdella, G. M., Kucukvar, M., Ismail, R., Abdelsalam, A. G., Onat, N. C., & Dawoud, O., A mixed model-based Johnson's relative weights for eco-efficiency assessment: The case for global food consumption. *Environmental Impact Assessment Review*, Vol. 89, 106588, 2021g.

Abdella, G. M., Kucukvar, M., Kutty, A. A., Abdelsalam, A. G., Sen, B., Bulak, M. E., & Onat, N. C., A novel approach for developing composite eco-efficiency indicators: The case for US food consumption. *Journal of Cleaner Production*, Vol. 299, p.126931, 2021f.

Abdella, G. M., Kucukvar, M., Onat, N. C., Al-Yafay, H. M., and Bulak, M. E., Sustainability assessment and modeling based on supervised machine learning techniques: The case for food consumption, *Journal of Cleaner Production*, Vol.

251, 2020a.

Abdella, G. M., Maleki, M. R., Kim, S., Al-Khalifa, K. N., and Hamouda, A. M. S., Phase-I monitoring of high-dimensional covariance matrix using an adaptive thresholding LASSO rule. *Computers & Industrial Engineering*, 106465, 2020.

Abdur-Rouf K. B., Abdella, G. M., Al-Khalifa, K. N., and Alhajyaseen, W., Ridge penalization-based generalized linear model (GzLM) for predicting risky-driving index, *In the Proceedings of the International Conference on Industrial Engineering and Operations Management*, Washington DC, USA, pp. 1462-73, 2018.

Al-Buenain, A., Al-Muhannadi, S., Falamarzi, M., Kutty, A. A., Kucukvar, M., & Onat, N. C. The Adoption of Electric Vehicles in Qatar Can Contribute to Net Carbon Emission Reduction but Requires Strong Government Incentives. Vehicles 2021, Vol. 3,pp. 618–63,2021.

Alkaabneh, F. M., Lee, J., Gómez, M. I., & Gao, H. O. (2021). A systems approach to carbon policy for fruit supply chains: Carbon tax, technology innovation, or land sparing?. *Science of The Total Environment*, *767*, 144211

Al-Obadi, M., Kutty, A. A., Abdella, G. M., Kucukvar, M., & Bulak, M. E.. Carbon footprints across the food supply chain: A microscopic review on food waste-related sustainability assessment. In *Proceedings of the 11th Annual International Conference on Industrial Engineering and Operations Management* ,pp. 7-11, 2021.

Alsarayreh, M. M., AlSuwaidi, M. F., Sharif, R. A., and Kutty, A. A., The Factors Affecting CO_2 Emission in the European Union Countries: A Statistical Approach to Sustainability across the Food Industry, *2020 IEEE 7th International Conference on Industrial Engineering and Applications (ICIEA)*, IEEE, April 16-

21, 2020.

Al-Sheeb, B., Abdella, G. M., Hamouda, A. M. S., and Abdulwahed, M. S., Predictive modeling of first-year student performance in engineering education using sequential penalization-based regression, *Journal of Statistical Management Systems*, Vol. 22, pp. 31-50, 2019.

An, Q., Chen, H., Wu, J., & Liang, L. . Measuring slacks-based efficiency for commercial banks in China by using a two-stage DEA model with undesirable output. Annals of Operations Research, 235(1), pp.13–35, 2015.

Benez-Secanho, F. J., Miner, J., & Dwivedi, P. Using advanced spatial statistical analyses to determine socio-economic constructs of fresh food availability in Georgia, United States. *Journal of Agriculture and Food Research*, Vol. *6* , 100204, 2021.

Bohnes, F. A., Hauschild, M. Z., Schlundt, J., Nielsen, M., & Laurent, A.. Environmental sustainability of future aquaculture production: Analysis of Singaporean and Norwegian policies. *Aquaculture*, Vol. *549*, 737717, 2022.

Boussofiane, A., Dyson, R. G., & Thanassoulis, E. Applied data envelopment analysis. *European journal of operational research*, *52*(1), pp.1-15, 1991.

Chen, Z., & Han, S. Comparison of dimension reduction methods for DEA under big data via Monte Carlo simulation. *Journal of Management Science and Engineering, 2021*.

Conrad, Z., Johnson, L. K., Peters, C. J., & Jahns, L. . Capacity of the US food system to accommodate improved diet quality: A biophysical model projecting to 2030. *Current developments in nutrition*, *2*(4), nzy007, 2018.

Controlling industrial greenhouse gas emissions. Center for Climate and Energy Solutions. (2021, September 15). Retrieved May 20, 2022, from https://www.c2es.org/content/regulating-industrial-sector-carbon-emissions/

Cook, W. D., Liang, L., Zha, Y., & Zhu, J. . A modified super-efficiency DEA model for infeasibility. Journal of the Operational Research Society, 60(2), pp. 276–281, 2009.

Council, F. A. O. (1989). Sustainability assessment of food and agriculture systems (SAFA). *Dostupno na: http://www. fao. org/nr/sustainability/sustainabilityassessments-safa/en/, 2016*.

De Olde, E. M., Oudshoorn, F. W., Sørensen, C. A., Bokkers, E. A., & De Boer, I. J. Assessing sustainability at farm-level: Lessons learned from a comparison of tools in practice. *Ecological Indicators*, Vol. *66* ,pp. 391-404, 2016.

Desiderio, E., García-Herrero, L., Hall, D., Segrè, A., & Vittuari, M.. Social sustainability tools and indicators for the food supply chain: a systematic literature review. *Sustainable Production and Consumption, 2021*.

Egilmez, G., Gumus, S., Kucukvar, M., & Tatari, O. (2016). A fuzzy data envelopment analysis framework for dealing with uncertainty impacts of input–output life cycle assessment models on eco-efficiency assessment. *Journal of cleaner production*, *129*, 622-636.

Egilmez, G., Kucukvar, M., Tatari, O., & Bhutta, M. K. S. Supply chain sustainability assessment of the US food manufacturing sectors: A life cycle-based frontier approach. *Resources, Conservation and Recycling*, Vol. *82*, pp. 8-20, 2014.

Elhmoud, E. R., and Kutty, A. A., Sustainability Assessment in Aviation Industry: A Mini-Review on the Tools, Models and Methods of Assessment. *In the Proceedings of the International Conference on Industrial Engineering and Operations Management. Harare, Zimbabwe*, October 20-22, 2021.

Elhmoud, E. R., Kutty, A. A., Abdalla, G. M., Kucukvar, M., Bulak, M. E., & Elkharaz, J. M.. Eco-Efficiency Performance of Airlines in Eastern Asia: A Principal Component Analysis Based Sustainability Assessment. In *Proceedings of the 11th Annual International Conference on Industrial Engineering and Operations Management, Singapore* , pp. 7-11, 2021a.

Eurostat. (2021). Sustainable development in the European Union — Monitoring report on progress towards the SDGs in an EU context — 2021 edition . Retrieved from https://ec.europa.eu/eurostat/documents/3217494/12878705/KS-03-21-096-EN-N.pdf/8f9812e6-1aaa7823928f03d8dd74df4f?t=1623741433852.

Food and Agriculture Organization of the United Nations. *Improved rural urban linkages: Building sustainable food systems*. Retrieved January 26, 2022, from https://youtu.be/DJgMzxUTx2U, 2017.

Golany, B., & Roll, Y. An application procedure for DEA. *Omega*, *17*(3), pp.237-250, 1989.

Grenz, J., Thalmann, C., Schoch, M., Stalder, S., & Studer, C.. RISE (Response-Inducing Sustainability Evaluation), version 2.0. *School of Agriculture, Forest and Food Science, HAFL. Bern University of Applied Science, Bern*, 2012.

Grenz, J., Thalmann, C., Stämpfli, A., Studer, C., & Häni, F.. RISE–a method for assessing the sustainability of agricultural production at farm level. *Rural Development News*, *1*(2009), pp. 5-9, 2009.

Harun, S. N., Hanafiah, M. M., & Aziz, N. I. H. A. . An LCA-based environmental performance of rice production for developing a sustainable agri-food system in Malaysia. *Environmental Management*, *67*(1), pp. 146-161, 2021.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55-67.

Kucukvar, M., Ismaen, R., Onat, N. C., Al-Hajri, A., Al-Yafay, H., & Al-Darwish, A. Exploring the social, economic and environmental footprint of food consumption: a supply chain-linked sustainability assessment. In *2019 IEEE 6th International Conference on Industrial Engineering and Applications (ICIEA)*, IEEE. pp. 733-742, 2019.

Kucukvar, M., Kutty, A. A., Al-Hamrani, A., Kim, D., Nofal, N., Onat, N. C., ... & Al-Nahhal, W..How circular design can contribute to social sustainability and legacy of the FIFA World Cup Qatar 2022™? The case of innovative shipping container stadium. *Environmental Impact Assessment Review*, Vol. *91*, 106665, 2021.

Kucukvar, M., Onat, N. C., Kutty, A. A., Abdella, G. M., Bulak, M. E., Ansari, F., & Kumbaroglu, G.. Environmental efficiency of electric vehicles in Europe under various electricity production mix scenarios. *Journal of Cleaner Production*, 130291, 2021a.

Kutty, A. A., Abdella, G. M., and Kucukvar, M., Ridge Penalization-based weighting approach for Eco-Efficiency assessment: The case in the food industry in the United States. *In IOP Conference Series: Materials Science and Engineering*, Vol. 947, no. 1, pp. 012003, IOP Publishing, 2020.

Kutty, A. A., Abdella, G. M., Kucukvar, M., Onat, N. C., and Bulu, M., A system thinking approach for harmonizing smart and sustainable city initiatives with United Nations sustainable development goals. *Sustainable Development*, Vol. 28, pp. 1347-1365, 2020d.

Kutty, A. A., Al-Jondob, R., Abdella, G. M., and El-Mekkawy, T. A., Frontier Based Eco-Efficiency Assessment of Electric Vehicles: The Case of European Union Countries Using Mixed and Renewable Sources of Energy, *Proceedings of the*

First Central American and Caribbean International Conference on Industrial Engineering and Operations Management, Port-au-Prince, Haiti, June 15-16, 2021a.

Kutty, A. A., and Abdella, G. M., Tools and Techniques for Food Security and Sustainability Related Assessments: A focus on the Data and Food Waste Management System, *Proceedings of the 5th NA Conference on Industrial Engineering and Operations Management*, Detroit, Michigan, USA, August 10-14, 2020b.

Kutty, A. A., Yetiskin, Z., Abraham, M. M., Nooh, M. A., Kucukvar, M., and Abdalla, G. M., An Empirical Assessment on the Transportation Sustainability Indicators and their Impact on Economic Productivity, *Proceedings of the 5th NA Conference on Industrial Engineering and Operations Management,* Detroit, Michigan, USA, August 10-14, 2020c.

Lee, C. Y., & Cai, J. Y. , LASSO variable selection in data envelopment analysis with small datasets. *Omega*, *91*, 102019, 2020.

Lenzen, M., Moran, D., Kanemoto, K., & Geschke, A. Building Eora: a global multi-region input–output database at high country and sector resolution. *Economic Systems Research*, *25*(1), pp.20-49, 2013.

López, F. J., Ho, J. C., & Ruiz-Torres, A. J. A computational analysis of the impact of correlation and data translation on DEA efficiency scores. *Journal of Industrial and Production Engineering*, *33*(3), pp.192-204, 2016.

Marami Milani, M. R., Hense, A., Rahmani, E., & Ploeger, A. . Applying least absolute shrinkage selection operator and akaike information criterion analysis to find the best multiple linear regression models between climate indices and components of cow's milk. Foods, 5(3), Vol. 52, 2016.

Marston, L. T., Read, Q. D., Brown, S. P., & Muth, M. K. (2021). Reducing water scarcity by reducing food loss and waste. *Frontiers in Sustainable Food Systems*, *5*, 651476.

Martín-Gamboa, M., & Iribarren, D. Coupled life cycle thinking and data envelopment analysis for quantitative sustainability improvement. In *Methods in Sustainability Science* (pp. 295-320). Elsevier, 2021.

Müller, B., Hoffmann, F., Heckelei, T., Müller, C., Hertel, T. W., Polhill, J. G., ... & Webber, H. . Modelling food security: Bridging the gap between the micro and the macro scale. *Global Environmental Change*, Vol. *63*, 102085, 2020.

Nataraja, N. R., & Johnson, A. L. Guidelines for using variable selection techniques in data envelopment analysis. *European Journal of Operational Research*, *215*(3), pp.662-669, 2011.

Olagunju, K. O., Feng, S., & Patton, M. . Dynamic relationships among phosphate rock, fertilisers and agricultural commodity markets: Evidence from a vector error correction model and Directed Acyclic Graphs. *Resources Policy*, Vol. *74*, 102301, 2021.

Onat, N. C., Abdella, G. M., Kucukvar, M., Kutty, A. A., Al-Nuaimi, M., Kumbaroğlu, G., & Bulu, M.. How eco-efficient are electric vehicles across Europe? A regionalized life cycle assessment-based eco-efficiency analysis. *Sustainable Development*, *29*(5), pp.941-956, 2021.

Park, M. Y., & Hastie, T. (2007). L1-regularization path algorithm for generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *69*(4), 659-677.

Park, Y. S., Egilmez, G., & Kucukvar, M.. Emergy and end-point impact assessment of agricultural and food production in the United States: A supply chain-linked Ecologically-based Life Cycle Assessment. *Ecological indicators*, Vol. *62*, pp.117-

137, 2016.

Reisi, M., Aye, L., Rajabifard, A., & Ngo, T. (2014). Transport sustainability index: Melbourne case study. *Ecological Indicators*, Vol. *43*, 288-296.

Reisi, M., Aye, L., Rajabifard, A., & Ngo, T. (2014). Transport sustainability index: Melbourne case study. *Ecological Indicators*, *43*, 288-296, 2014.

Roselli, L., Casieri, A., De Gennaro, B. C., Sardaro, R., & Russo, G. . Environmental and economic sustainability of table grape production in Italy. *Sustainability*, *12*(9), 3670, 2020.

Royte, E. (2021, May 3). *Innovative Food Packaging, explained*. Environment. from https://www.nationalgeographic.com/environment/article/food-packaging-plastics-recycle-solutions

Sagar, P., Gupta, P., & Kashyap, I. . A forecasting method with efficient selection of variables in multivariate data sets. *International Journal of Information Technology*, *13*(3), pp. 1039-1046, 2021.

Saisana, M., & Tarantola, S. State-of-the-art report on current methodologies and practices for composite indicator development, 2002.

Saling, P. Eco-efficiency assessment. In *Special types of life cycle assessment* . Springer, Dordrecht, pp. 115-178, 2016.

Sampedro, C., Pizzitutti, F., Quiroga, D., Walsh, S. J., & Mena, C. F. (2020). Food supply system dynamics in the Galapagos Islands: Agriculture, livestock and imports. *Renewable Agriculture and Food Systems*, *35*(3), pp.234-248, 2020.

Sen, B., Kucukvar, M., Onat, N. C., & Tatari, O. Life cycle sustainability assessment of autonomous heavy-duty trucks. *Journal of Industrial Ecology*, *24*(1), pp.149-164, 2020.

Sethi, J. K., & Mittal, M. (2021). An efficient correlation based adaptive LASSO regression method for air quality index prediction. *Earth Science*

Informatics, *14*(4), pp.1777-1786, 2021.

Shiau, T. A., & Liu, J. S. . Developing an indicator system for local governments to evaluate transport sustainability strategies. *Ecological indicators,* Vol. *34,* pp. 361-371, 2013.

Spykman, R., Hossaini, S. M., Peguero, D. A., Green, A., Heinz, V., & Smetana, S.. A modular environmental and economic assessment applied to the production of Hermetia illucens larvae as a protein source for food and feed. *The International Journal of Life Cycle Assessment*, *26*(10), pp. 1959-1976, 2021.

Stelwagen, R. E., Slegers, P. M., de Schutter, L., & van Leeuwen, E. S. . A bottom-up approach to model the environmental impact of the last mile in an urban food system. *Sustainable Production and Consumption*, Vol. *26*, pp. 958-970, 2021.

Stone, T. F., Thompson, J. R., Rosentrater, K. A., & Nair, A.. A Life Cycle Assessment Approach for Vegetables in Large-, Mid-, and Small-Scale Food Systems in the Midwest US. *Sustainability*, *13*(20), 11368, 2021.

Tang, J., Alelyani, S., & Liu, H. Feature selection for classification: A review. *Data classification: Algorithms and applications*, Vol.37, 2014.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267-288.

Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. *Statistics in medicine*, *16*(4), 385-395.

Tone, K. A slacks-based measure of efficiency in data envelopment analysis. European Journal of Operational Research, 130(3), pp.498–509, 2001.

University of Michigan. Home | Center for Sustainable Systems. (n.d.). Retrieved May 20, 2022, from https://css.umich.edu/

Verweij, P. J., & Van Houwelingen, H. C. (1994). Penalized likelihood in Cox regression. *Statistics in medicine*, *13*(23-24), 2427-2436.

Wang, C. N., Nguyen, M. N., Le, A. L., & Tibo, H. . A DEA Resampling Past-Present-Future Comparative Analysis of the Food and Beverage Industry: The Case Study on Thailand vs. Vietnam. *Mathematics*, *8*(7), 1140, 2020.

Wiedmann, T. O., Schandl, H., Lenzen, M., Moran, D., Suh, S., West, J., & Kanemoto, K. (2015). The material footprint of nations. *Proceedings of the national academy of sciences*, *112*(20), pp.6271-6276,2015.

Wu, J., Chu, J., Sun, J., & Zhu, Q. DEA cross-efficiency evaluation based on Pareto improvement. European Journal of Operational Research, 248(2), pp. 571–579, 2016.

Zhang, H., Xu, Y., & Lahr, M. L.. The greenhouse gas footprints of China's food production and consumption (1987–2017). *Journal of Environmental Management*, Vol. *301*, 113934, 2022.

Zheng, J., Garrick, N. W., Atkinson-Palombo, C., McCahill, C., & Marshall, W. . Guidelines on developing performance metrics for evaluating transportation sustainability. *Research in Transportation Business & Management*, Vol. *7*, pp. 4-13, 2013.

APPENDIX A

Data collected and analyzed:

Y	X1	X2	X3	X4	X5	X6	X7	
	CO_2	CH_4	CO	HFC	PM10	PM2.5	N20	Industry
0.736823	8.05E-01	9.75E-01	8.49E-01	5.75E-01	8.91E-01	8.59E-01	9.09E-01	AOFF
0.98637	9.79E-01	9.99E-01	9.82E-01	8.66E-01	9.75E-01	9.55E-01	9.78E-01	BSM
0.206456	4.88E-01	9.47E-01	6.76E-01	7.18E-02	7.37E-01	5.10E-01	7.78E-01	BBPM
0.814606	8.89E-01	9.90E-01	8.98E-01	5.45E-01	8.96E-01	7.62E-01	9.05E-01	BCM
0.483263	6.26E-01	9.75E-01	7.28E-01	2.07E-01	8.15E-01	7.40E-01	8.51E-01	BW
0.785475	7.77E-01	8.90E-01	8.08E-01	6.29E-01	8.50E-01	8.43E-01	8.55E-01	CM
0.98836	9.90E-01	9.99E-01	9.91E-01	9.32E-01	9.88E-01	9.85E-01	9.90E-01	CCCB
0.921569	9.52E-01	9.98E-01	9.56E-01	7.74E-01	9.61E-01	9.83E-01	9.71E-01	CTM
0.739592	8.46E-01	9.87E-01	8.89E-01	1.91E-01	8.54E-01	8.06E-01	8.86E-01	CMPC
0.647861	7.52E-01	9.71E-01	8.00E-01	3.21E-01	8.30E-01	7.13E-01	8.55E-01	CCPM
0.787576	8.32E-01	9.71E-01	8.34E-01	5.57E-01	8.64E-01	6.26E-01	8.64E-01	DCFM
0.882247	8.91E-01	9.57E-01	9.11E-01	6.95E-01	9.13E-01	9.19E-01	9.24E-01	DCEPM
0.92024	9.29E-01	9.94E-01	9.21E-01	6.96E-01	9.18E-01	6.72E-01	9.13E-01	FORB
0.961943	9.50E-01	9.93E-01	9.44E-01	9.21E-01	9.66E-01	8.90E-01	9.65E-01	FMMM
0.609013	6.44E-01	8.25E-01	7.03E-01	4.34E-01	7.66E-01	7.52E-01	7.72E-01	FMBM
0.621072	6.49E-01	9.36E-01	7.04E-01	5.24E-01	8.29E-01	6.48E-01	8.22E-01	FFM
0.52784	6.31E-01	9.72E-01	7.38E-01	4.45E-01	8.52E-01	8.82E-01	8.86E-01	FVPD
0.939681	9.56E-01	9.90E-01	9.69E-01	8.36E-01	9.67E-01	9.79E-01	9.72E-01	ICFM
0.849192	8.93E-01	9.90E-01	9.22E-01	4.14E-01	8.91E-01	8.24E-01	9.11E-01	NCM
0.957835	9.58E-01	9.93E-01	9.52E-01	9.17E-01	9.66E-01	8.68E-01	9.62E-01	OAFM
0.448697	4.55E-01	6.39E-01	5.07E-01	2.41E-01	5.82E-01	4.93E-01	5.79E-01	PP
0.948628	9.67E-01	9.96E-01	9.82E-01	8.88E-01	9.80E-01	9.95E-01	9.85E-01	SPP
0.842628	8.84E-01	9.89E-01	8.98E-01	6.30E-01	9.12E-01	8.39E-01	9.22E-01	SDM
0.580604	6.97E-01	9.81E-01	7.43E-01	2.32E-01	8.10E-01	5.31E-01	8.09E-01	SFM
0.262738	5.52E-01	9.80E-01	7.22E-01	1.21E-01	8.04E-01	8.68E-01	8.72E-01	SDIM
1	9.99E-01	1.00E+00	9.98E-01	9.99E-01	9.99E-01	9.91E-01	9.98E-01	SOP
0.96479	9.59E-01	9.97E-01	9.66E-01	7.12E-01	9.49E-01	9.21E-01	9.58E-01	SCMR
0.964642	9.72E-01	9.96E-01	9.73E-01	6.26E-01	9.40E-01	9.60E-01	9.50E-01	TM
0.99589	9.93E-01	9.99E-01	9.94E-01	9.92E-01	9.96E-01	9.82E-01	9.95E-01	WCM
0.828873	8.94E-01	9.95E-01	9.14E-01	6.10E-01	9.26E-01	9.87E-01	9.45E-01	WINE

X8 VOC	X9 SO2	X10 M	X11 MQ	X12 NREC	X13 REC	X14 W	Industry
7.99E-01	7.56E-01	8.51E-01	8.73E-01	7.81E-01	6.97E-01	8.03E-01	AOFF
9.71E-01	9.65E-01	9.86E-01	9.85E-01	9.81E-01	9.71E-01	9.84E-01	BSM
4.87E-01	3.12E-01	5.28E-01	6.17E-01	2.95E-01	7.24E-02	2.30E-01	BBPM
8.58E-01	8.28E-01	9.03E-01	8.91E-01	8.25E-01	7.48E-01	6.50E-01	BCM
6.20E-01	5.37E-01	1.01E-01	4.66E-01	4.71E-01	2.79E-01	6.21E-01	BW
8.07E-01	8.12E-01	8.23E-01	7.84E-01	7.87E-01	7.59E-01	7.88E-01	CM
9.85E-01	9.82E-01	9.91E-01	9.92E-01	9.87E-01	9.83E-01	9.93E-01	CCCB
9.37E-01	9.28E-01	9.49E-01	9.68E-01	9.59E-01	9.39E-01	9.92E-01	CTM
7.85E-01	7.40E-01	8.40E-01	8.99E-01	8.03E-01	6.69E-01	9.23E-01	CMPC
7.10E-01	6.43E-01	7.69E-01	7.98E-01	6.11E-01	4.53E-01	5.75E-01	CCPM
7.98E-01	7.84E-01	7.21E-01	7.92E-01	7.44E-01	6.99E-01	3.82E-01	DCFM
8.78E-01	8.63E-01	8.91E-01	9.02E-01	8.86E-01	8.51E-01	8.90E-01	DCEPM
9.12E-01	9.06E-01	9.26E-01	9.27E-01	8.82E-01	8.79E-01	4.41E-01	FORB
9.42E-01	9.30E-01	9.63E-01	9.34E-01	9.15E-01	9.17E-01	8.22E-01	FMMM
6.88E-01	6.99E-01	6.98E-01	6.45E-01	6.50E-01	5.74E-01	6.71E-01	FMBM
7.06E-01	6.99E-01	7.31E-01	7.08E-01	6.13E-01	4.66E-01	4.13E-01	FFM
6.34E-01	5.69E-01	4.96E-01	7.68E-01	6.28E-01	4.45E-01	8.29E-01	FVPD
9.53E-01	9.48E-01	9.57E-01	9.69E-01	9.50E-01	9.15E-01	9.72E-01	ICFM
8.29E-01	8.18E-01	8.66E-01	9.25E-01	8.43E-01	7.72E-01	8.27E-01	NCM
9.49E-01	9.50E-01	8.81E-01	9.29E-01	9.31E-01	9.32E-01	6.90E-01	OAFM
4.94E-01	5.48E-01	5.24E-01	5.83E-01	3.88E-01	2.69E-01	3.41E-03	PP
9.71E-01	9.61E-01	9.80E-01	9.86E-01	9.71E-01	9.58E-01	9.91E-01	SPP
8.52E-01	8.40E-01	8.68E-01	8.56E-01	8.57E-01	8.10E-01	8.59E-01	SDM
6.87E-01	6.91E-01	6.81E-01	6.81E-01	6.49E-01	4.96E-01	5.13E-01	SFM
5.18E-01	3.74E-01	0.00E+00	7.56E-01	4.58E-01	2.44E-01	8.45E-01	SDIM
9.99E-01	9.99E-01	9.99E-01	9.98E-01	9.99E-01	9.99E-01	9.84E-01	SOP
9.39E-01	9.23E-01	9.78E-01	9.76E-01	9.37E-01	9.25E-01	9.73E-01	SCMR
9.44E-01	9.30E-01	9.78E-01	9.72E-01	9.59E-01	9.51E-01	9.31E-01	TM
9.93E-01	9.92E-01	9.95E-01	9.90E-01	9.91E-01	9.89E-01	9.72E-01	WCM
8.74E-01	8.47E-01	7.98E-01	8.94E-01	8.99E-01	8.74E-01	9.83E-01	WINE

APPENDIX B

Cross-Validation Outcome

	Penalty	Regularization "R Square" (1-Error)	Number of Selected Predictors	Apparent Prediction Error
1	0.010	0.996	14	0.004
2	0.020	0.992	14	0.008
3	0.030	0.988	14	0.012
4	0.040	0.988	13	0.012
5	0.050	0.975	14	0.025
6	0.060	0.977	13	0.023
7	0.070	0.971	13	0.029
8	0.080	0.992	6	0.008
9	0.090	0.990	7	0.010
10	0.100	0.994	2	0.006
11	0.110	0.994	2	0.006
12	0.120	0.993	2	0.007
13	0.130	0.993	2	0.007
14	0.140	0.919	12	0.081
15	0.150	0.992	2	0.008
16	0.160	0.991	2	0.009
17	0.170	0.990	2	0.010
18	0.180	0.989	2	0.011
19	0.190	0.989	2	0.011
20	0.200	0.988	2	0.012
21	0.210	0.987	2	0.013
22	0.220	0.986	2	0.014
23	0.230	0.985	2	0.015
24	0.240	0.983	2	0.017
25	0.250	0.982	2	0.018
26	0.260	0.981	2	0.019
27	0.270	0.980	2	0.020
28	0.280	0.978	2	0.022
29	0.290	0.977	2	0.023
30	0.300	0.975	2	0.025
31	0.310	0.974	2	0.026
32	0.320	0.972	2	0.028

33	0.330	0.971	2	0.029
34	0.340	0.969	2	0.031
35	0.350	0.967	2	0.033
36	0.360	0.966	2	0.034
37	0.370	0.964	2	0.036
38	0.380	0.962	2	0.038
39	0.390	0.960	2	0.040
40	0.400	0.958	2	0.042
41	0.410	0.956	2	0.044
42	0.420	0.954	2	0.046
43	0.430	0.952	2	0.048
44	0.440	0.950	2	0.050
45	0.450	0.947	2	0.053
46	0.460	0.945	2	0.055
47	0.470	0.943	2	0.057
48	0.480	0.940	2	0.060
49	0.490	0.938	2	0.062
50	0.500	0.936	2	0.064
51	0.510	0.933	2	0.067
52	0.520	0.931	2	0.069
53	0.530	0.928	2	0.072
54	0.540	0.925	2	0.075
55	0.550	0.922	2	0.078
56	0.560	0.920	2	0.080
57	0.570	0.917	2	0.083
58	0.580	0.914	2	0.086
59	0.590	0.911	2	0.089
60	0.600	0.908	2	0.092
61	0.610	0.905	2	0.095
62	0.620	0.902	2	0.098
63	0.630	0.899	2	0.101
64	0.640	0.896	2	0.104
65	0.650	0.893	2	0.107
66	0.660	0.889	2	0.111
67	0.670	0.886	2	0.114
68	0.680	0.883	2	0.117
69	0.690	0.879	2	0.121

70	0.700	0.876	2	0.124
71	0.710	0.872	2	0.128
72	0.720	0.869	2	0.131
73	0.730	0.865	2	0.135
74	0.740	0.861	2	0.139
75	0.750	0.858	2	0.142
76	0.760	0.854	2	0.146
77	0.770	0.850	2	0.150
78	0.780	0.846	2	0.154
79	0.790	0.842	2	0.158
80	0.800	0.838	2	0.162
81	0.810	0.834	2	0.166
82	0.820	0.830	2	0.170
83	0.830	0.826	2	0.174
84	0.840	0.822	2	0.178
85	0.850	0.818	2	0.182
86	0.860	0.813	2	0.187
87	0.870	0.809	2	0.191
88	0.880	0.805	2	0.195
89	0.890	0.800	2	0.200
90	0.900	0.796	2	0.204
91	0.910	0.791	2	0.209
92	0.920	0.787	2	0.213
93	0.930	0.782	2	0.218
94	0.940	0.777	2	0.223
95	0.950	0.773	2	0.227
96	0.960	0.768	2	0.232
97	0.970	0.763	2	0.237
98	0.980	0.758	2	0.242
99	0.990	0.753	2	0.247
100	1.000	0.748	2	0.252