Proposed Statistical Techniques for Combining Parameter Estimates: A Case of Food Production in Sub-Saharan Africa

by

Busanga Jerome Kanyama

Submitted in accordance with the requirements for
the degree of

DOCTOR OF PHILOSOPHY

In the subject of

STATISTICS

at the

UNIVERSITY OF SOUTH AFRICA

SUPERVISOR: Professor Peter M. Njuho

February 2022

DECLARATION

Name: BUSANGA JEROME KANYAMA

Student number: 36460672

Degree: Doctor of Philosophy in Statistics

Exact wording of the title of the thesis as appearing on the copies submitted for examination:

Proposed Statistical Techniques for Combining Parameter Estimates: A Case of Food Production in Sub-Saharan Africa.

I declare that the above thesis is my own work and that all the sources that I have used or quoted have been indicated and acknowledged by means of complete references.

____

SIGNATURE

04 February 2022

Acknowledgement

I am grateful to many people who gave me the opportunity to complete this thesis. There are a few of them that I am privileged to identify.

(1) Professor Peter Njuho as my supervisor for his knowledge and encouragement throughout the study.

(2) I would like to thank the master's and doctoral support programme applications (MDSP) and teaching development grants (TGD) for their financial support.

(3) I am also grateful to the Head of the Department of Statistics, Professor Eeva Rapoo, for her support, the continued recess leaves and motivation during my studies.

(4) I am thankful for the contributions of my colleagues: Prof John Olaomi, Prof Principal Ndlovu, Dr Jean-Claude Malela-Majika and Mr Serge Malandala for their contributions and friendship.

Finally, I am grateful to my wife Apoline Kijana Kanyama and my children Akilimali Heritier Kanyama, Gracia Mwamini Kanyama, Coltide Sakina Kanyama, Beatrice Mwaibwa Kanyama and Apoline Immaculate Kanyama for their prayers, patience and sacrifices. I love you so much.

Busanga Jerome Kanyama.

Abstract

The underperforming agricultural sector in Sub-Saharan Africa (SSA) has left African countries with insufficient food production in the face of challenges related to climate change, diseases and increasing population growth. The agricultural sector is the main source of food, generates income, employs a large portion of the population, and produces raw materials for agribusinesses. The improvement of agricultural food production contributes to food security, poverty alleviation, the development of trade, and a country's economy. The challenges facing the SSA countries include ineffective farming system, loss of soil fertility, limited access to land, climate change, water scarcity, outdated production technology that needs to change, restricted market access due to poor infrastructure, and high transaction costs among others. To address these challenges, the combination of multiple nutrients was proposed to increase grain yield of crop simply because of the contribution of each nutrient rather than the use of a single fertiliser.

Research conducted in SSA with the aim of improving food production miss the opportunity to share the findings across the various sectors. This points out the lack of appropriate statistical techniques to address the challenges. We can understand better the real situation on food production by developing a comprehensive scientific and statistical approach that can gather all published single information to a unified finding. The process of collecting and combining research outputs require the use of meta-analysis (MA) to provide precise estimates on various parameters associated with food production. Various factors can be considered in making significant contribution to agricultural food production such as fertiliser, access to market, energy use, trade, etc. To establish the diverse set of relationships that can be developed among the factors, structural equation model (SEM) statistical technique is used. In some conditions, this procedure can be more restrictive and inflexible since the approach requires the specification of latent variables in the mix of a huge diversity of sets of variables. In the

body of this work, we propose a more suitable, flexible and accurate approach in determining the number of linear regressions based on the observed data in a clear and precise manner through factor analysis and principal component analysis (PCA). In addition, to test the large number of variables or factors of the parameters obtained in SEM, we propose to synthesise all this information by integrating MA into SEM. The incorporation of MA into SEM allows us to account simultaneously all effects of factors of the food production in a single model. In MA, the effect sizes are assumed independent from each study and univariate MA is used. A single study could involve multiple tests of the same hypothesis, resulting in reporting multiple outcomes (MOs). In such situation, the researcher developed MOs approach to determine the multiple linear regression model that tested and analysed the relations between the factors of interests in the food production.

The results of MA were expressed in terms of fixed- and random-effects. The fixed-effects models were more appropriate simply because of the presence of homogenous effects in the studies. The random effect models helped to control unobserved heterogeneity when the between-studies variance was large. It was more productive to apply the combined inorganic fertilizer by the raisin yield grain of maize. The findings of SEM provide efficient results in the evaluation of the relations among variables and for testing a statistical theoretical model. The findings from the integration approach of MA into SEM permitted to combine parameter estimates within a single model. Researchers in agricultural and related field can use these techniques positively.

We hope that many researchers can benefit from the methodological approach to estimate and draw inference in addressing the food production situation. The outcomes of this work contribute to science by providing scientifically comprehensive statistical approaches to evaluate and synthesise the more suitable results. The benefit can be extended to the development of suitable food production.

Keys terms

Combined multiple outcomes, Combined model, Factor analysis, Fixed effects model, Meta-analysis, Multivariate meta-analysis, Principal component analysis, Parameter estimate, Random-effects model, Structural equation model

Table of Contents

List of Figures

List of Tables

List of Abbreviations

CFL: Confirmation Factor Analysis

EFA: Exploratory Factor Analysis

ES: Effect sizes

FA: Factor Analysis

FAO: Food and Agriculture Organisation

GLS: Generalised Least Square

GDP: Gross Domestic Products

KZN: KwaZulu Natal

LISREL: Linear Structural Relation Modelling

MA: Meta-Analysis

MOs: Multiple Outcomes

NGO: Non-Governmental Organisation

OLS: Ordinary Least Squares

PA: Path Analysis

PCA: Principal Component Analysis

SEM: Structural Equation Model

SMD: Standardised Mean Diffrence

SSA: Sub-Saharan Africa

UN: United Nations

List of research outputs

The research outputs include peer-reviewed published articles. All these publications are associated with this thesis.

Publication from this thesis

Kanyama, B. J., Njuho, P. and Majika, J. M. (2018). Improved Structural Equation Models Using Factor Analysis. Pak.stat.oper.res., 14(4), pp. 995-1012.

Kanyama, B. J. and Njuho, P. (2020). Use of Linear Mixed Effects Model in Meta-Analysis for Studies with Multiple Outcomes, International Journal of Agricultural and Statistical Sciences. 16(1), pp. 11-21.

Kanyama, B. J. and Njuho, P. (2021). A Comprehensive Approach for Integrating Meta-Analysis into Structural Equation Modeling: Case of Food Production, International Journal Agricultural Statistical Sciences. 17(2), pp. 479 – 492.

CHAPTER 1

GENERAL ORIENTATIONS

1.1 INTRODUCTION

The agricultural food production

Agriculture plays a vital role in a country's economy through its share in the gross domestic product (GDP), which is the monetary value of all finished products of goods and services made within a country. Several studies have revealed that agriculture is the engine of growth in most developing countries, running from agricultural growth to economy-wide growth (Chauvin et al., 2012). It provides the primary source of food, generates income, employs many the population, and produces raw materials for the agribusiness. The improvement of agricultural food production contributes to food security, poverty alleviation, and developing a country's trade and economy. Developing these entities is a major challenge facing sub-Saharan African (SSA) countries. The challenges include an ineffective farming system, loss of soil fertility, climate change, water scarcity and limited market access. According to Demment et al. (2003), increased contribution by agricultural factors has collectively improved people's lives, productivity, social well-being, and the national economy. As one of the poorest regions in the world, the SSA has failed over a long period to register significant growth, resulting in high levels of poverty (Chauvin et al., 2012). This failure is because agricultural development is characterised by low productivity with insignificant application of science and technology.

According to the United Nations (UN) (UN, 2009), three crises that are currently engulfing Africa and impact development relate to the finance and economy, food security, climate change and energy. These crises occurred when African countries were achieving remarkable population growth. The underperforming agricultural sector leads to insufficient food to feed the high population in the SSA countries (Mwichabe, 2013). Agricultural products such as fertiliser, irrigation systems, crop management, and new technology are needed to grow food production. The African farmers still have issues with market participation due to the standard of food safety, poor infrastructure, and high transaction costs (Holloway et al., 2000; Shiferaw et al., 2011). A large portion of the increase in world population by 2050 is expected to come from SSA. In a way, the management of this growth with limited resources could be considered a risk for African countries' development (Gashu et al., 2019).

In consideration of the challenges mentioned above in SSA agriculture, researchers need to take their role seriously. Such roles entail investigating how to increase food production significantly for a wide-range food crops and animal feed. Improved food production could trigger the development of SSA countries.

Agricultural land availability

Access to land and utilisation correlates with food production and agricultural growth in general. The small-scale farmers form the largest group of producers of food for home consumption in SSA. The issue of land in SSA intertwines with cultural, political and environmental factors to the detriment of food production. The variation in land access among farmers, especially smallholder farmers, threatens their viability. The use of new agricultural technology, which leads to commercialisation of the agriculture, has affected land utilisation (Jayne et al., 2010).

The high cost of farm inputs, continuous population growth and ever-decreasing farm size due to sub-division affects agricultural production. The crop yield in SSA had been dropping compared to the amount needed to satisfy the demand of the growing population (Josephson et al., 2014; Gashu et al., 2019). The migration from the rural to the urban areas in search of a better life can be attributed to the land issue. The move is an adaptive reaction due to land availability and production. This involves income diversification from non-farming activities. The rural population in SSA has been unable to leave poverty behind simply because they cannot transform their basic economic activities, which is agriculture (Chauvin et al., 2012).

Climate change effects on agricultural production

Climate change has seriously affected agricultural productivity globally. These changes have affected weather conditions and have caused drought, increased variation of diseases and pests, all associated with agricultural production. The increase in temperature impacts the irrigated agriculture directly by losing the revenue and drying the land that reduces agriculture productivity (Kurukulasuriya, 2006). Irrigation requires a massive capital investment that the majority of SSA countries cannot provide. Increased temperature due to climate change affects the rate of precipitation and evaporation, leading to water scarcity, and subsequently, food production decreases. Interventions such as farming system adjustment to prevent food production losses become necessary. The changes in climatic condition create costs that overwhelm the resources of the poor farmers (Howden et al., 2007). The climate change in SSA has created the necessity of having a mechanism that can help farmers deal with these possible risks.

Water scarcity is an important factor in food production. The availability of quality water has directly influenced agricultural activities such as crop, livestock and fishing, etc. The scarcity of water exerts significant pressure on smallholder farming activities (Namara et al., 2010). Water shortage impacts food production due to agricultural activity that uses a proportion of freshwater (Wallace et al., 2002). Farming by irrigation systems needs to be promoted to overcome the effects of climate change. The approach is better than rainfall, but it necessitates high capital investment from small and challenged African farmers.

Infrastructure effect on agricultural food production

Infrastructure development remains a concern as transportation facilities have considerable limitations. People living in rural areas have limited access both to the regional cities and the international market destination. The improved infrastructure in SSA will boost productivity by an increase of about 40% of African firms in general (Ondiege et al., 2013). The transport infrastructure correlates with an increase in agricultural productivity. The farmers need their products to reach the market on time. The SSA countries strive to attain the status of developed countries in terms of high-speed railways, modern highways, improved maritime transport and the aviation industry. Agenda 2063 outlines the political commitments towards achieving these goals (Gashu et al., 2019).

Need for an innovative approach to agricultural food production

The agricultural food production in SSA requires technological change that includes innovative research on production, distribution and marketing of value-added products. The attention of many researchers in agricultural and related fields is needed to articulate the challenges facing the SSA countries. The success of revolutionising agricultural food production lies in adapting theoretical and practical (or empirical) approaches to innovative research. New technologies affect markets and prices, and for the farmers to adopt them, extra incentives need to be factored in (Chavas et al., 2020). The improvement of agricultural practices and the reduction of the impact of climate change require that farmers implement a combination of technologies which include making decisions to adopt new specific practices (Branca et al., 2020). The government needs considerable dedication to understand various aspects of technological change. It is essentially the function of the state and planning departments to make decisions (Lawson et al., 2020). All stakeholders, civil society, the private sector, and others must take responsibility jointly with the government when making country-level policy decisions (Ikram et al., 2020). In the framework of the current crisis, the role and contribution of agriculture cannot be fully utilised to apply resources efficiently for-profit maximisation because of the current shortcomings of agriculture due to the imbalances created by the agricultural policies (Loizou et al., 2019). The following are some areas in agricultural food production where technological innovation is required:

1) Using tools such as mechanical improvement: tractors, biological advances with new seed varieties, fertilisers, and new farm management practice.
2) Finding ways of improving irrigation to the extent that it has to save water.
3) Impacting the market by improving existing agricultural approaches' performance and products' value-add.
4) Achieving multiple effects such as the saving of labour, better usage of energy and capital.
5) A desire to increase the yield by increasing the amount of water used and capital invested.

Firms with access to capital, technology, and financing have increased food by using new products or herbicide mixtures (Caceres et al., 2020). Farmers are now adopting technologies, including organic matter levels, that have sufficiently improved their fertiliser, enabling them to sell enough livestock and crops (Pretty et al., 2003).

Many types of innovation are needed to articulate the challenges in agricultural food production. Each of the innovation categories associated with technological change affects the economy. Technological development brings economic growth by causing an increase in communication and providing easy and fast access to new markets; therefore, it makes a positive impact on the economy by paving ways to produce a new chapter of goods and capital accumulation (Caliskan, 2015).

Methodology approach to agricultural food production challenges

In this study, we investigated the perception that the adoption of new technology and related agricultural products increases food production results in SSA. The approach focuses on the scientific and published studies in the SSA. The factors that are considered as contributing to agricultural food production include such as fertilisers (nitrogen and phosphate), access to market, energy use and trade (import and export). These factors are assumed to bring change in the development of agricultural process activities and therefore articulate the challenges facing the SSA countries.

Over the years, many documented reviews have addressed the positive impact of the adoption of fertilizer for the growing in crop yield and improvement in soil health (Akinnifesi et al, 2009). These studies have discussed potential intervention to increase food production in SSA whereas technologies depend on the resources available to each country (Tovihoudji et al., 2019; Pasley et al., 2019). "This research conducted in SSA fall into the form of standalone categories, "silos", depending on the funder with the aim to improve the food production for subsistence farmers" (Kim et al., 2016; Muluh et al., 2019). The funds for these research groups come from national research institutions, donors from different countries, non-governmental organisations (NGOs), etc. We develop a mechanism that combines the research outputs generated by different research groups, by having precise estimates on various parameters in food production, thus, providing a clear understanding of the challenges facing the SSA countries. The process of collecting research outputs and combining them requires the use of the meta-analysis (MA) technique. In applying the technique, we review various studies and identify the appropriate common effects to estimate the parameters. In combining these effects, we will provide a common estimate that is stable and informative. Based on stable estimates generated through our approach, the policymakers can formulate interventions on agricultural food production in SSA countries. The livelihood of the small-scale farmers in SSA will be improved once the challenges in agricultural production are addressed.

Many studies have addressed the negative impact of factors such as poor soil fertility, water scarcity, limited resources and so on that have caused in the low crop of food production and several recommendations have been proposed. If these challenges continue to persist, this is because of lack of appropriate and effective statistical techniques to address the real problem facing the SSA countries. There is a need to develop a comprehensive statistical approach that scientifically can gather all independent research findings to a unified model to address effectively these challenges facing SSA. In this study, MA is presented to gather the results from scientific published studies and unify this information into a single model. To establish a set of relationships that can be developed among diverse factors involved in agricultural food production, a structural equation modelling (SEM) is proposed based on a more suitable, flexible, and accurate approach using factor analysis through principal component analysis (PCA). To test the causal models and establish the combined parameters estimates, we propose to combine MA and SEM within a single model simply because the two techniques use the correlations (or covariance) matrices as inputs.

Food production is a combination of factors such as genotype (varieties), environment (location) and their interactions simply because a genotype can react differently from one location to the next. Understanding the cause for the increase of food production, whether this can be attributed to the quality of a genotype, location, or some combination of factors, leads to determining the performance of a genotype across locations using stability statistics. We construct an adaptive model that identifies the genotypes with outstanding performance across different locations and indicate how to validate the model and test the stability parameter estimates in an experiment.

MA enables us to synthesise data and summarise the results of individual studies by showing the effect that exists from one study to the next. The MA results help increase the power of individual studies by synthesising data and identifying the sources of diversity across studies of several natures. Three models can be formulated in MA:

- The fixed model: To remove omitted variables bias that can occur by measuring changes within group across time. That is, the approach does not estimate the effects of variables whose values do not change across time.
- The random effects model: allow to make inferences on the population data based on the assumption of the normal distribution. That is, it assumes that the individual specific effects are uncorrelated with the independent variables.
- The mixed effects model: To estimate the model parameters between and within the subject variability simply because in the analysis, the observations within a subject may be correlated.

In this work, constructed models are expressed in terms of mixed models, fixed and random effects models. In the structural equation model, we test the relations between variables in a model by examining the statistical significance. Meta-analysis and structural equation modelling are the common multivariate techniques widely used in medical and social sciences (Yukiko, 2019).

Hence, we aim to borrow the approach of these various techniques to use them to combine various results produced independently in agricultural studies within the SSA. This is because the development process of food production has a multivariate structure that involves more than one correlation per study by applying an MA procedure. The integration of MA into SEM requires combining correlations from N independent studies by using fixed and random effects to test the homogeneity of correlation that can fit a structural equation model. The new approach uses all the data sets using factor analysis through the PCA to generate the important components; the variables with the highest sample variances are among the retained components taken simply because each variable receives its weight in the analysis by using standardised variables. The components created enable the researcher to estimate the parameters in the models because it is expected that the estimated parameters from the models hold good loadings for the models that fit the data well. The generalised least square model (GLS) is used to generate significant results. The process of combining results permits us to combine parameter estimates within a single model, and researchers in the agricultural field can use these techniques positively. A simulation study is used to test the effectiveness of the improved model over a dataset of SEM approach. The results indicate that the proposed approach is preferred.

Meta-analysis differs from both primary and secondary analysis. Primary analysis refers to the analysis of the data to provide solutions to the study's research questions. In contrast, secondary analysis refers to re-analysis of the data, and most often, secondary analysis answers the research questions in a different way (Card, 2012). In this work, SEM tests the relationships between variables in a model by examining the statistical significance. SEM tests all defined relationships between the variables simultaneously because it is a powerful tool capable of solving complex problems involving diverse factors. SEM is a tool that provides efficient results in the evaluation of the relations among variables and in testing theoretical models. In this study, we develop an adaptive approach that is unrestrictive since the current SEM specifies latent variables involved in the analysis and creates theoretical relations between variables. We present a more reliable approach that provides a guideline on evaluating the suitability of a given SEM. Research in the agriculture sector uses all possible variables that might be identified for a set of data. Using factor analysis through the PCA, researchers can use the most significant variables in the model simply because the high dimensions are transformed into lower dimensional data.

The formulation of the combined approach tests the hypotheses about the parameters involved in the model in terms of fixed effects or random effects. These models test for homogeneity or heterogeneity to produce significant results and after that, tests whether the model fits the data well.

The next question is whether the combined approach can be more flexible and suitable. This approach can be established by using the factor analysis to determine the hypothetical relations between variables. The benefits of the combined approach are to obtain all effects within a single model. It is anticipated that the integration approach could generate significant results when using the generalised least square procedure.

In this study, we use factor analysis through the PCA upon which the highly correlated variables are the best-selected variables for the underlying causal relationship. This technique is suitable simply because it uses the reduced set of data so that the variables are significant. We hope that this work can stimulate more research development in the agricultural field and many other areas. It will improve the quality of the data required to improve food production in the agricultural field in SSA. Many countries and researchers are expected to benefit from this information by knowing the methodological approach to estimate and draw an inference in addressing the food production situation. The outcomes of this work have to contribute to science by providing evidence of the techniques that are comprehensive to evaluate and synthesise the more suitable results. The benefit is extended to the development of suitable food production. Examples in the agricultural field are used to show the practical use of each approach. In the process, all these statistical methodologies from different approaches are introduced into the agriculture field.

1.2 PROBLEM STATEMENT

The challenges facing the SSA countries originate from multiple factors such as the farming system, the loss of soil fertility, climate change, the energy crisis, water scarcity, an unbalance in accessing land and a poor infrastructure to access markets. These factors created insufficient food security, a high level of poverty and hunger, decreased trade, and negatively impacted the country's economy. Much research conducted in SSA falls into silos. It emerged that a food crisis is overwhelming

African countries while African countries are attaining a significant population growth. There is a need for rapid intervention to increase agricultural food production, thereby improving the lives of people, the social wellbeing, and the growth of the country's economy. To better understand the reality and gravity of the problem in agricultural research food production, we need to develop a mechanism that combines the research outputs generated by different research groups. The question is how to combine the research outputs generated by different research groups across SSA to have precise estimates of various parameters in the food production system. The multivariate techniques such as meta-analysis (MA), structural equation modelling (SEM) and the combined approach of the two techniques with data availability enabled us to address the challenges facing the SSA countries.

1.3 OBJECTIVES OF THE STUDY

The core objectives of this research work are to:

(1) develop an improved statistical model that is flexible and suitable in estimating parameters and drawing inferences in food production.

(2) test structural relations by examining how the data fits a theoretical model.

(3) synthesise information to establish the actual situation of food production using meta-analysis and structural equation modelling.

(4) provide a methodological approach with mathematical integration of MA into SEM with the help of guidelines by indicating how to test the hypotheses of homogeneity against heterogeneity.

(5) estimate the stability of genotypes and indicate those that are stable across other components; and

(6) test the integration of the mixed-effects model into MA with multiple outcomes within a single model.

In the next chapter, we develop the MA statistical technique, and an example is used as an application of the proposed approach.

1.4 ORGANISATION OF THE THESIS

This thesis is organised as follows:

Seven chapters are presented, including the introduction as Chapter 1. The meta-analysis technique is presented in Chapter 2. Chapter 3 discusses the SEM techniques. In Chapter 4, an approach is developed that allows the integration of MA into SEM. Chapter 5 discusses the estimating parameters of stability measures on the performance of the yield model. The integration of the mixed-effects model into MA with multiple outcomes within a single model is given in Chapter 6, while Chapter 7 provides the concluding discussion, recommendations, and future work.

CHAPTER 2

META-ANALYSIS

2.1 INTRODUCTION

MA is an overarching collection of analysis results from individual studies with the aim of integrating the findings (Glass, 1976). The MA began to appear as a leading part of research in the late 70s. Since then, it has become a common way for synthesising evidence and summarising the results of individual studies (Chalmers et al., 1997). Smith and Glass (1977) used MA in social sciences to investigate whether psychotherapy was beneficial for patients. Many research studies have been done, for example, in medicine, by examining the effect of aspirin after a myocardial infraction (Brockwell et al., 2001). In social marketing campaigns, the MA technique has been used to validate data for behaviour intention and estimate mean effect sizes for each variable across all campaigns. Path analysis has been used to measure relationships among variables, including across different models (Green et al., 2019). It has also been used in medical sciences to systematically assess the association of circulating inflammation markers with the future risk of hypertension (Musi et al., 2011). This approach has made a significant contribution to plant ecology. It allowed a review of the evidence for various ecological hypotheses and theories and provided an estimation of the effects of major environmental drivers (climate change, habitat fragmentation, invasive species, and air pollution) as well as research gap identification (Koricheva et al., 2014).

The MA approach was applied in interventions for loneliness reduction by quantifying the effects of each strategy and examining the potential role of variables of interests (Jayedth et al., 2018). Rajabi et al. (2019) used the MA technique to quantify the association between tobacco, opioid use, and opioid use disorders where tobacco smoking was a major risk factor. Recently, MA has been applied in various human studies. As an example, it was found that there were inconsistent associations between the length ratio of the second finger to the fourth finger, which is a proxy for prenatal androgen load and substance or computer use in adolescents and adults (Siegmann et al., 2019). In medicine, Muhie (2019) has found a positive effect on human's health, leading to a reduction in mortality and the disease burden, therefore improving the quality of life. Hence, MA plays a key role in food production in a new area of applied research studies. We intend to borrow this technique to combine various results produced in an independent agricultural study within the SSA.

Agricultural research plays a very important role in addressing the challenges that affect the lives of people. Food production in its development process involves many factors such as type of crop, fertiliser, kinds of manure, trade, labour, water etc. There is a need for gathering all this information to estimate and get significant treatment effects (Mavridis et al., 2011). Since we intend to make decisions about the strength of the estimated parameters, we propose synthesising all the results from different studies into a single model approach. The effect sizes are the correlations. The steps in the MA approach are as follows:

1 indicate the research interest.
2 provide a number of critical inclusion criteria with indications of the method for searching in the literature by using the keywords of searching the past studies.
3 test whether or not the treatment effect is statistically significant.
4 test for heterogeneity of the effect on the outcome between the included studies; and
5 formulate and test the fixed-effects and random-effects models.

The fixed effects model helps us draw conclusions on the included studies in the MA. The random-effects model enables us to simplify the results further than the included studies. Section 2.2 presents the basic characteristics of MA. The contributions to the field of agricultural science are given through the methodology by using the standardised mean difference (SMD) as given in Section 2.3. Section 2.4 is the multivariate meta-analysis for several outcomes. The data analysis is given in Section 2.5. Conclusion and recommendations are given in Section 2.6.

The method of MA summarises the results of independent studies into a single estimate. Since the inputs of MA are the effect sizes (ESs), we need to determine the effect that exists from one study to the next. For instance, when the ESs are constant from one study to the next, then the fixed effects model is used to identify the common effect, but when the ESs vary from one study to the next, then the random-effects model is used to identify the reason for the variation.

Since we have to describe the relationships between two factors to address the challenges in the SSA, fixed and random effects estimates are reported in terms of linear equations. A standard linear regression has only fixed effects while mixed models are an extension of simple linear models to allow and to include both fixed and random effects. For instance, past studies could be sampled from within authors. In case of generalisation of ordinary linear regression, the generalisation linear model is used.

2.2 METHODOLOGY OF META-ANALYSIS

2.2.1 Introduction

The current approach of meta-analysis is more restrictive. For instance, when the number of past studies is too small, and therefore the method does not effectively reveal the error associated with the parameter estimation, it becomes difficult to identify the sources of heterogeneity between studies (Turner et al., 2013). We present a more consistent and understandable approach that uses systematic review under the set of inclusion/exclusion criteria to calculate the effect sizes (ESs). This approach enables us to measure each study as an independent estimate of the primary accurate ESs. Effect sizes enable the researcher to quantify the difference between the two groups. It represents a quantitative measure that facilitates the comparison of any two groups to understand how the two groups differ from each other. There are many ways to calculate the ESs; for instance, Cohen proposed that a calculated value “d” equal to 0.2 indicates the small effect size, a calculated value “d” = 0.5 represents a medium effect size, and “d” = 0.8 is considered as a large effect size. That is, if “d” = 0.2, we conclude that the two groups don’t differ by 0.2 standard deviations or more. This means the difference between the two groups is trivial. This principle remains valid even if the statistical test can be significant. In practice, the effect size is calculated by taking the mean of the treatment group minus the mean of the control group, and the difference is divided by the standard deviation of one of the groups. Most researchers in the literature use the standard deviation of the control group. The hypotheses for testing the effect sizes are: The null hypothesis H_0: The variables of concern (of interests) remain the same, and the alternative hypothesis H_1: The variables of concern differ from each other.

2.2.2 Meta-Analysis in Calculating Effect Sizes

The following guidelines provide a summary that will enable the researcher to understand the calculations of the ESs as presented in Section 2.3.1. The different steps involved are as follows:

Step 1: Screen the data for suitability testing.

Step2: Calculate the descriptive statistics from the data.

Step 3: Calculate the ESs using the SMD from each study.

Step 4: Calculate the weighted mean and the test statistic of the ESs.

Step 5: Calculate the standard error and the confidence interval; and

Step 6: Calculate and test the fixed and random-effects model.

The procedure enables us to use data from scientific and published articles that provide a good understanding of the effects of food production in the use of the combined of fertilisers compared to the use of unfertilized (control) farming system through MA. We investigate the adoption of the new technology approach and related tools under the conditions of the rainfall that are theoretically likely to increase food production in the SSA. In the next section, we consider more than one factor of independent variables that influence the variability of the dependent variables. The conclusion drawn in the form of multivariate approach is more accurate and realistic because these conclusions are nearer to the real-life situation. The multivariate analysis is based in analysing more than one statistical outcome variable at a time for large datasets.

2.3 MULTIVARIATE-ANALYSIS FOR MULTIPLE OUTCOMES

2.3.1 Multivariate Meta-Analysis for Fixed-Effects Model

The meta-analysis technique enables the researcher to combine quantitative evidence from the past and related studies. This approach provides the results for the research question. In some situations, the variables of interests may be correlated to more than one variable, or some studies may not provide direct evidence about some particular outcomes. As a result, such variables or studies are often rejected from the MA. Because research studies incur high costs and are time-consuming, rejecting variables in this way can be viewed as a research waste. Multivariate meta-analysis outcome can analyse multiple outcomes simultaneously by allowing more studies to contribute towards each outcome. That is, the approach is likely to provide the summary results that are not created for each piece of evidence; instead, these outcomes depend on correlated results from other evidence. Applying this approach leads to a gain of information that includes both indirect evidence and those outcomes of some relevant studies that may not be available. The outcomes are the overall results simply because of the borrowing of strength in the statistical analysis. Suppose that there are p studies and each study have p outcomes denoted by

$\mathbf{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_p)'$ a $p \times p$ matrix, where $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \dots, y_{ip})'$, ($i$ = 1…,p).

The variance-covariance matrix of **X** is denoted by **V** where

$$\boldsymbol{V} = \begin{pmatrix} var(x_{i1}) & cov(x_{2i}, x_{i2}) & \dots cov(x_{ip}, x_{p1}) \\ cov(x_{i1}, x_{1i}) & var(x_{i2}) & \cdots cov(x_{ip}, x_{p2}) \\ \vdots & \vdots & \vdots \quad \vdots \\ cov(x_{i1}, x_{1p}) & \cdots & \cdots \ var(x_{ip}) \end{pmatrix}.$$

The fixed-model effects for the i^{th} study is

$$\mathbf{Y} = \mathbf{X\beta} + \mathbf{e}, \tag{2.1}$$

where $\mathbf{Y} = \begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_p \end{pmatrix}$, $\mathbf{X} = \begin{pmatrix} \mathrm{x}_{11} & \mathrm{x}_{12} & \cdots & \mathrm{x}_{1p} \\ \mathrm{x}_{21} & \mathrm{x}_{22} & \cdots & \mathrm{x}_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ \mathrm{x}_{p1} & \mathrm{x}_{p2} & \cdots & x_{pp} \end{pmatrix}$, $\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$, $\mathbf{e} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_p \end{pmatrix}$,

$\boldsymbol{y}_i$ represents the $p \times 1$ vector of the observed ESs (correlation from the sample), $\boldsymbol{X}$ represents a $p \times p$ matrix of the explanatory variables, which is an identity matrix and $\boldsymbol{e}_i \sim N(\mathbf{0}, {\sigma_i}^2 I)$ $i = 1,2, \dots, p$. The distribution of $\boldsymbol{y}_i$ in the fixed effects has an MVN distribution with the mean $\boldsymbol{\mu}$ and the matrix of variance-covariance $\boldsymbol{V}$.

$E(\boldsymbol{y}_i) = E(y_{i1}, y_{i2}, \dots, y_{ip})' = (\mu_1, \mu_2, \dots, \mu_p)' = \boldsymbol{\mu}$ is a vector of means for each outcome i = 1, 2... p. $V(\boldsymbol{x}_i)$ is the variance-covariance matrix. The distribution is equal to

$$\begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_p \end{pmatrix} \sim MVN\left(\begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}, \begin{pmatrix} Var(x_{i1}) & \cdots & \cdots & Cov(x_{1p}, x_{p1}) \\ Cov(x_{i1}, x_{1i}) & Var(x_{i2}) & \cdots & Cov(x_{ip}, x_{p2}) \\ \vdots & \vdots & & \vdots \quad \vdots \\ Cov(x_{i1}, x_{1p}) & \cdots & \cdots & Var(x_{ip}) \end{pmatrix} \right) \tag{2.2}$$

The above distribution is multivariate with a mean vector and the variance-covariance matrix with $\boldsymbol{e} \sim$MVN $(\mathbf{0}, \boldsymbol{V})$, where $\boldsymbol{V}$ is assumed to be an identity matrix.

The univariate statistical model can be extended to a multivariate model to have multiple outcome variables. This is like the analysis of variance that can be extended to multivariate analysis. The problem of multivariate analysis is the choice between the fixed effects and the random-effects model. The approach requires:

(1) the null hypothesis H_0: The random-effects model is preferred.

(2) The alternative H_1: The fixed effects model is preferred.

(3) If the *p*-value of the test is greater than the level of significance (0.05), we fail to reject the null hypothesis H_0. That is, the model is consistent and efficient.

(4) If the *p*-value of the test is less than the level of significance (0.05), we reject the null hypothesis H_0. That is, the fixed-effects model is preferred.

2.3.2 Multivariate Meta-Analysis for Random-Effects Model

The multivariate random effect model is like the multivariate fixed effect model as described in Section 2.4.1, from which the model is extended with a factor $\boldsymbol{\nu}^{\mathbf{2}}$, representing some studies that can come from a larger population which causes a random effect. This factor $\boldsymbol{\nu}^{\mathbf{2}}$ is a multivariate normal distribution with the mean equal to $\mathbf{0}$ and the between-study covariance matrix as shown below.

$$\boldsymbol{\nu}^{\mathbf{2}} \sim \text{MVN}\left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} {\nu_1}^2 & \nu_1\nu_2 & \cdots & \nu_1\nu_p \\ \nu_2\upsilon_1 & {\nu_2}^2 & \cdots & \nu_2\nu_p \\ \vdots & \vdots & \vdots & \vdots \\ \nu_p\nu_1 & \nu_p\nu_2 & \dots & {\nu_p}^2 \end{pmatrix}\right) \tag{2.3}$$

The implication of the random effects model is to consider the studies not to be independent; therefore, the researcher can generalise the results to other population. The multivariate random effect model for a meta-analysis is given in Equation 2.4.

$$\boldsymbol{Y} = \mathbf{X\beta} + \mathbf{UZ} + \mathbf{e}, \tag{2.4}$$

where

$$\mathbf{Y} = \begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_p \end{pmatrix}, \ \mathbf{X} = \begin{pmatrix} \mathrm{x}_{11} & \mathrm{x}_{12} & \cdots & \mathrm{x}_{1\mathrm{p}} \\ \mathrm{x}_{21} & \mathrm{x}_{22} & \cdots & x_{2\mathrm{p}} \\ \vdots & \vdots & & \vdots \\ \mathrm{x}_{\mathrm{p}1} & \mathrm{x}_{\mathrm{p}2} & & x_{\mathrm{pp}} \end{pmatrix}, \ \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}, \ \mathbf{e} = \begin{pmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \\ \vdots \\ \boldsymbol{e}_p \end{pmatrix},$$

$$\mathbf{U} = \begin{pmatrix} u_{11} & 0 & \cdots & 0 \\ 0 & u_{22} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & u_{pp} \end{pmatrix}, \ \mathbf{Z} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_p \end{pmatrix} \tag{2.5}$$

The distribution of the multivariate is the combination of Equation (2.2) and (2.3) as follows

$$\begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \vdots \\ \boldsymbol{y}_p \end{pmatrix} \sim \boldsymbol{MVN}(\boldsymbol{\mu}, \boldsymbol{V}), \tag{2.6}$$

where $\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}$,

$$\boldsymbol{V} = \begin{pmatrix} Var(x_{i1}) + \nu_1{}^2 & Cov(x_{2i}, x_{i2}) + \nu_1 \upsilon_2 & \cdots & Cov(x_{pi}, x_{p1}) + \nu_1 \upsilon_p \\ Cov(x_{i1}, x_{2i}) + \nu_2 \upsilon_1 & Var(x_{i2}) + \nu_2{}^2 & \cdots & Cov(x_{ip}, x_{p2}) + \nu_2 \upsilon_p \\ \vdots & \vdots & & \vdots \ \vdots \\ Cov(x_{i1}, x_{1p}) + \nu_p \upsilon_1 & Cov(x_{i2}, x_{2p}) + \nu_p \upsilon_2 & \cdots & Var(x_{ip}) + \nu_p{}^2 \end{pmatrix},$$

where $\boldsymbol{e}$ is distributed with mean vector $\mathbf{0}$ and variance-covariance matrix $\mathbf{V}$, thus having $\mathbf{X}$, defined the same way as in fixed-effect sizes, and $\boldsymbol{\beta}$ is a random population mean vector of the effect sizes under the random-effects model. $\boldsymbol{ZU}$ is the study-specific effect in the i^{th} study with $\boldsymbol{Z}_i$ being a selection of a matrix of zeros and ones. $\boldsymbol{U}_i \sim N(\mathbf{0}, \ \boldsymbol{\nu}^{\mathbf{2}})$ is the study-specific random effects in the i^{th}study, with $\boldsymbol{\nu}^2$ being a $p \times p$ no negative definite matrix (Becker et al., 1994;

Raudenbush et al., 2002; Cheung, 2013). Under the fixed effects model, we assume that the effect size for all past studies is identical. That is, the effect size varies between studies because of the sampling error. Under the random effects model, we estimate the mean of the distribution of the effects and the covariance structures. In applying the random effects, no assumption is made, such as common effects size, simply because the random effects model aims to generalise the effect sizes to other populations. That is, we consider some of the dispersion in the observed effects to reflect the real differences in the effect size across the studies. In the following section, we present the approach that estimates the parameter for the fixed and random effects model.

2.3.3 Estimating Parameters in the Multivariate Meta-Analysis Model

The formulas to estimate parameters such as the vector $\boldsymbol{\mu}$, the matrix $\boldsymbol{\nu}^{\mathbf{2}}$ and the test

statistic are given as $$\boldsymbol{\nu} = \frac{Q_{cal}-(k-1)}{\sum w - \left(\frac{\sum w^2}{\sum w}\right)} \qquad (2.7)$$

where *k* is the number of effect sizes, Q_{cal} represents the test statistic of the effect size and w is the optimal weight.

In MA, we are testing whether a set of a single study is heterogeneous or not. To apply this test requires a measure of heterogeneity called Cochran's Qtest (West et al., 2010). This test tells us about the presence against the absence of heterogeneity. In this study, the performance of the Q test is used to determine the weighted sum of squared differences between individual study effects and the pooled effect across studies. The Q calculated test statistic is

$$Q = \sum w_i \times {ES_i}^2 + \frac{[(\sum w_i \times ES_i)]^2}{\sum w_i} \qquad (2.8)$$

where w_i is the weight of each study and ES_i is the effect size (correlation) for each study.

The estimate of the ESs is calculated based on the inverse variance weight. Hedges indicated that the optimal weight for the fixed effects model in MA is $w = \frac{1}{SE^2}$,where

$$SE = \sqrt{\frac{{S_1}^2}{n_1} + \frac{{S_2}^2}{n_2}} \quad (2.9)$$

is the standard error of a direct index of effect size precision, S_1 and S_2 are the standard deviations of the samples and n_1 and n_2 are the sample sizes.

2.4 DATA ANALYSIS

2.4.1 Introduction

To understand how to estimate the parameters in the fixed or random-effects model, we have to proceed with data analysis. In this section, the data analysis will enable us to gain more information from the data to derive knowledge that can help us make decisions, for instance, on the choice of the model to be used in estimating the parameters. In applying the proposed approach, systematic review and MA explored the response yield of maize production in soil that went through experiments for the use of the combine fertilizers compared to control (unfertilised use) in SSA. Many documented reviews have addressed the positive impact of the adoption of fertiliser for the growing in crop yield and improvement in soil health (Akinnifesi et al, 2009). These studies have discussed potential intervention to increase food production in SSA. In most countries of SSA, poverty contributes to the reduced of fertilizer adoption by farmers while fertilizer products enhance crop productivity of farmers. In this section, we review the published literature and performed MA to examine the benefits of the use of the combined of all diversities of fertilizers compared to control in growing yield maize response under rainfall status.

This work reviews scientifically all significant published literature reporting on fertilizers applications and unfertilized soil reserves within a sample of the response yield maize production. Maize was a high valued crop and farmers use several technologies for soil fertility preservation to increase grain yield. These technologies depend on the resources available to each country (Tovihoudji et al., 2019; Pasley et al., 2019). This crop accounts for 50% of the calories consumed in most African countries thus, it is cultivated under widely varying rainfall and edaphic conditions (Sileshi et al., 2008). Most region in SSA are among the leading maize producing areas (Okebelema, 2016). Maize is predominantly grown under rainfed conditions with limited resources resulting in the low yield due to little inherent soil fertility with high cost of mineral fertilizer (Van der Velde et al., 2013; Ouedraogo et al., 2019). Poor fertility in SSA is caused by the degrading nature of soils, poor farming techniques and very poor fertility inputs. Fertilizer use is recommended as a way of resolving the poor soil fertility problem in the SSA (Abunyewa et al., 2007; Okebalama, 2016). Results indicated a positive trend in maize crop response to fertilizer.

Considering the benefits of manures for soil health, combining fertilizer with all kind of inorganic fertilizer would be approved to increase the sustainability of the food production. The combined use of manure and inorganic fertilizer is an approach that seek to reduce cost of external inputs, increase food production and create safeguard environment for future generation (Abunyawa et al., 2007, 2002). Vicedo-Cabrera et al. 2019 reported that MA based on linear models was used to present a general framework for projections of climate change on health. In medicine, MA with generalized linear models have been used to show that the serum-neutralizing antibody titter can be associated with the protection against SHIV challenges (Pegu et al., 2019).

Material and Methods

The review was conducted with indication to MA based on Cochrane collaboration. MA approach requires to precise both the population of studies of interest and the criteria of eligibility of the studies to be included when the data is extracted, the next step is to determine what comparisons will be made (Sileshi et al., 2008). The criteria used for inclusion for the studies are

- published in the journal or peer-reviewed proceeding.
- originated from SSA resulting from an electronic library database.
- reported maize yield coming from the application with or without fertilizer.
- Excluded all published articles not dealing with maize or fertiliser and unfertilised (control) where not both reported.
- Reported the annual average rainfall.

Outcomes of Interest

The principal focus is to identify the maize yield response to the use of fertiliser compared to control applications in SSA. Preferred and defined keywords were only in English such as application of fertiliser, mineral fertiliser, nitrogen fertilisers, phosphorus fertilisers, calcium, magnesium, manures, compost, potassium are the most critical organic and inorganic fertilisers. These are followed by refining terms: Africa, central Africa, East Africa, Western Africa, Southern Africa, Northern Africa and SSA. All these were used as dominant factors to establish the impact on the maize yield response. Afterward, the rainfall outcome was regarded as a contributor factor for growing the crop. The literature research was performed by examining the Scopus network, google scholar search engine to collect scientific journal articles, thesis and conference proceedings concerning the use of fertilizer compared to unfertilized (control) application. The search of this work was limited to the period from 2007 to 2019 as given in Figure 2.1.

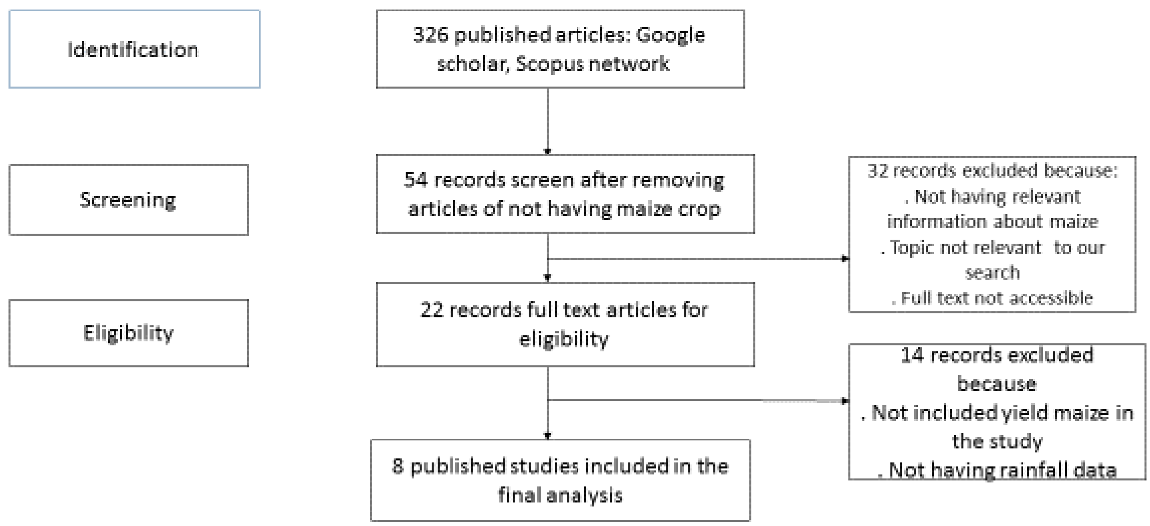


Figure 2.1: Study flowchart demonstrating the identification and inclusion process for the quantitative synthesis

A total of 326 past studies were screened for potential inclusion according to the titles, abstract but 8 articles were eligible for quantitative analysis in SSA based on the criteria as presented in Table 1. A selected study provided the author(s), the year, the country, the unfertilized data, and yield response data compared to fertilizer data with yield response data respectively for the maize-based crop. Identical data reported in a few studies were included only once. To better understand the real situation, systematic review is used prior to MA to combine all past studies into a common estimate parameter.

Table 2.1: Data collected from past published studies: Averages

	Author	Year	Country	Cont.	Fert.	Yield.C	Yield.F	Rain.
1	Tovihoudji et al,	2019	Benin	3000	300	200	5611	1114
2	Kisinyo et al.	2015	Kenya	770	1550	4000	37500	1600
3	Sime et at	2014	Ethiopia	260	260	416	551	815

Table 2.1 (Continued)

	Author	Year	Country	Cont.	Fert.	Yield.C	Yield.F	Rain.
4	Kamanga et al.	2014	Malawi	2200	3000	150	407	492
5	Mashingaidze et al.	2017	Zimbabwe	7700	3900	3245	4254	730
6	Pasley et al.	2019	Zimbabwe	230	130	720	474	100
7	Opala et al.	2007	Kenya	1010	8000	4500	8500	1800
8	Abunyewa et al.	2007	Ghana	451	108	13800	23024	1025

Cont.= Control (in tha^{-1}) Fert.= Fertiliser (in tha^{-1}) Yield.C = Yield.Control (tha^{-1}) Yield.F= Yield.Fertiliser (tha^{-1}) Rain= Rainfall

In the next section, the MA approach was used by summarising the data selected from the major factors. The visualization of the data was checked, the data cleaning, the checking of outliers and then the transformed data were used in the analysis.

Visualization of the data

To examine the data without making assumption, exploratory data analysis was used as shown in Chapter 9: Appendix, section A.1. This approach enables the researcher to understand the patterns within the data. This helps as well to detect outliers and find interesting relations among the variables. The results indicate that the variables did not follow the normal distribution based on the graphs and boxplot. The test for normality of the variables reveals that all the variables were normally distributed except the variable “Rainfall”.

2.4.2 Statistical Model in Meta-Analysis

Studies vary in size, and therefore larger studies must hold more weight in the analysis than smaller studies. The estimate of the effect sizes is calculated based on the inverse variance weight. Hedges and Vevea (1994) indicated that the optimal weight for the fixed effects model in MA is $w = \frac{1}{SE^2}$, where SE is the standard error of a direct index of effect size.

Table 2.2 indicates the weighted mean, standard error of the effect sizes, the Z-test for the mean of the effect sizes and the 95% confidence interval computed as shown below:

The weighted mean effect, $\overline{ES} = \frac{\sum w \times ES}{\sum w}$ (2.10)

The standard error of the mean effect sizes, $SE_{\overline{ES}} = \sqrt{\frac{1}{\sum w}}$ (2.11)

The Z-test statistic is, $Z = \frac{\overline{ES}}{SE}$ (2.12)

The confidence interval by the 95% confidence interval is $\overline{ES} \pm Z_{\frac{\alpha}{2}}\, SE_{\overline{ES}}$ (2.13)

Table 2.2: Effect sizes calculated based on the standardised mean difference, standard error, test statistic and 95% confidence interval (in quantities of production)

Variables	Effect Sizes	Standard error	Weighted mean $\overline{ES}$	Test statistics Z	95% confidence interval
Control	-0.245	892.08	-0.01081	-1.210E-01	(-357.32; 367.299)
Fertiliser	-0.07	975.14	-0.00299	-2.651E-06	(-357.313; 357.3078)
Yield-control	0.582	1618.2	0.00761	4.8237E-07	(- 357.30; 367.3182)
Yield-fertiliser	0.999	4729.2	0.00157	3.3171E-07	(-367.309; 367.312)
Rainfall	0.636	197.1	0.57491	0.002917	(-366.736; 367.8853)

The findings presented in Table 2.2 indicate that at 95% confidence interval, the results are all statistically significant by giving a plausible value for the parameter of interest. The yield-fertiliser is large being the ideal as we are more likely to capture the true value of the parameter.

Fixed and random-effects model

The fixed and random effects model has different purposes in making inferences. The fixed effects model is about the effect of parameters in the studies. We use the random-effects model if we want to make inferences about the effect of parameters in a population of the studies from a random sample of the studies. From each approach model, we have to evaluate the performance of the null hypothesis, testing that combine of fertiliser application does not perform better than the control application to increase yield maize food production. The results will enable the researcher to make inferences from each model.

2.4.2.1 The fixed-effects model

In the fixed-effects model, we assume that the variability between the ESs is due to sampling error under the null hypothesis that the distribution of the ESs is homogeneous. In other words, all the ESs are estimating the same population mean. If the null hypothesis is rejected, then we can fit a random effect model. Table 2.3. gives the fixed effects model and the test statistic Q for the variables of interest.

Table 2.3: The fixed effects size test

Variables	Q_{cal} Test statistic	*p-value*
Control	7.54271E-08	0.0001
Fertiliser	5.15302E-09	0.0001
Yield-Control	1.29355E-07	0.0001
Yield-Fertiliser	4.46231E-08	0.0001
Rainfall	1.04116E-05	0.0001

Since the $p-value$ is less than the 0.05, we reject the null hypothesis at 5% level of significance in favour of the alternative hypothesis that the combined of fertiliser does perform better than the control application in increasing of maize response food production.

2.4.2.2 The random-effects model

In a random effect model, when the total calculated Q is significant, that means the surplus variability across effect sizes comes from the random variations across studies that we cannot identify or measure the source of. In the random-effects model, we assume that the variability between effect sizes is due to sampling error and variability in the population (Wilson, 1999). We weighted each study by the inverse of the sampling variance plus a constant that represents the variability across the population effects.

The inverse of the sampling variance is $w = \frac{1}{SE^2 + \nu}$ (2.14)

where ν is the random effects variance component. The value of the random-effects variance is calculated as indicated below, and Table 2.4 gives the calculations for the random effects model.

$$\nu = \frac{Q_{cal} - (k-1)}{\sum w - \left(\frac{\sum w^2}{\sum w}\right)} \quad (2.15)$$

where k is the number of ESs used in the study.

Table 2.4: The weighted approach in the random-effects model

Component	Control	Fertiliser	Yield-control	Yield-Fertiliser	Rainfall
The random effects variance component ν	-783162.4	2.165E-07	2.855E-08	3.914E-10	0.0001297
Weight on random effects model w	7.911E-05	1.051E-06	3.818E-07	4.471E-08	2.573E-05

In Table 2.5, we present the weighted mean, standard error, test statistic and the 95% confidence interval for the random-effects model (in quantities of production).

Table 2.5: The random effects size test

Variables	Standard error	Weighted mean $\overline{ES}$	Test statistics Z	95% confidence interval
Control	96.97787	-0.18229	-0.00188	(-190.259; 189.894)
Fertiliser	96.97787	-0.00069	-7.1E-06	(-190.077; 190.0759)
Yield-Control	96.97787	-0.00069	-7.1E-06	(-190.077; 190.0759)
Yield-Fertiliser	96.97787	0.00042	4.33E-06	(-190.076; 190.077)
Rainfall	96.97787	0.153955	0.001588	(-189.923; 190.2306)

Effect sizes for types of both combine fertiliser and control applications are considered statistically significant because the 95% confidence interval limits of the average ESs did include 0; therefore, we fail to reject the null hypothesis. Thus, the variability across the ESs does not exceed what could be expected based on the sampling error.

The meta-analysis technique has enabled the researcher to synthesise information. In applying the approach, a suitable statistical summary of the results was provided upon which we gain clear information in drawing inferences in the agricultural research field on the maize food production in the SSA.

Forest plots

Forest plots graphically display the information from the individual studies like shown through the vertical axis upon which each country represents a scientific publish article as given in Table 2.1. The results as shown in Figure 1 represent the odds ratios, and the 95% confidence intervals represent an overall pooled estimate.

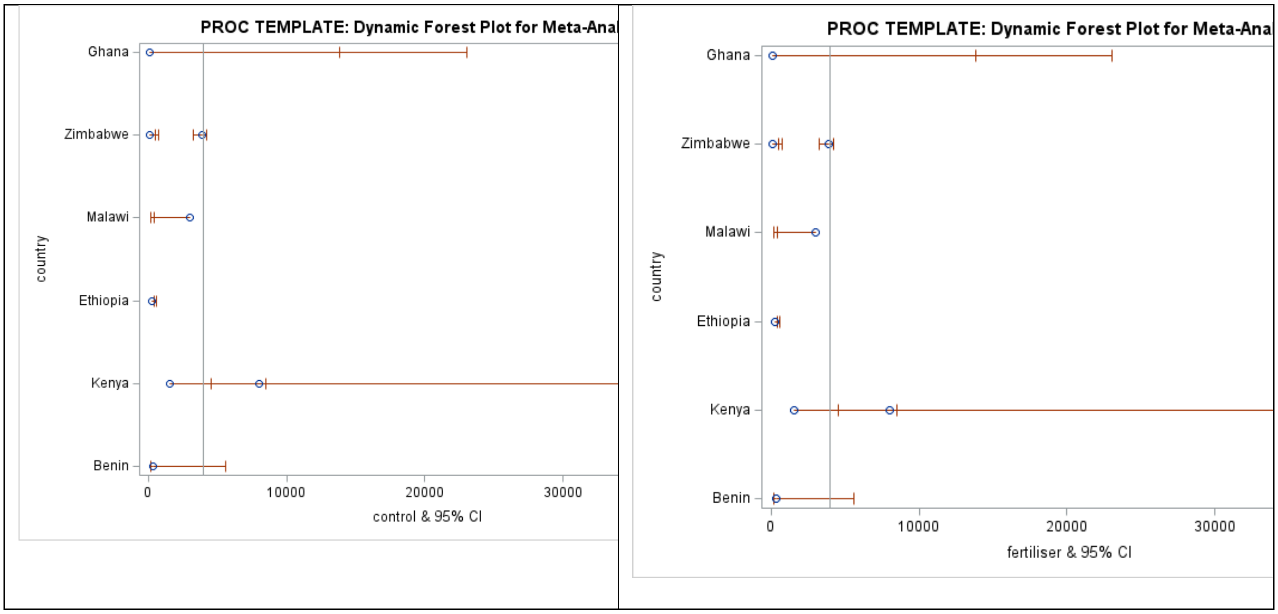


Figure 2.2: Forest plot

There was a difference in response in ratio between experiment under the combined fertilisers compared to the control experiment. That is, variability in response yield with the combined fertilisers were increased.

CHAPTER 3

STRUCTURAL EQUATION MODEL

3.1 INTRODUCTION

The SEMs are powerful tools that can be used to solve complex problems involving diverse factors. In particular, these tools can provide efficient results in evaluating the relations among variables and testing theoretical models. The SEM(s) and path analysis are introduced in agricultural science as powerful tools to solve complex problems. Worldwide, agricultural studies play a significant role in human beings' lives, particularly in SSA, where countries are characterised by many hungry people (Mwichabe, (2013)). SEM was used to examine the pathways from adverse childhood experiences to physical health, directly or indirectly, through many variables within a framework (Nurius et al., 2019). Lee et al. (2011) and Nitzl (2016)) used SEMs and partial least squares (PLS) techniques in accounting. More recently, Hair et al. (2017) used a series of ordinary least squares regressions to estimate partial model structures of composite-based SEM models. Henseler (2017) developed a variance-based SEM. Goodboy et al. (2017) presented a statistical and practical concern with published research featuring SEM.

In this study, we develop an adaptive approach that is unrestrictive as the current SEMs are involved in the analysis and create the theoretical relations between variables. There is a huge diversity in a set of relationships that could be developed among the variables. The variability of a set of relationships points to inconsistent conclusions about the level at which a model is truly equivalent to the observed data. We present a more reliable approach that provides a guideline on evaluating the suitability of a given SEM. We demonstrate how to include many factors by using factor analysis through the PCA, leading to a reduced number of variables. In Section 3.2.6, we will show how useful the application of the covariance is to estimate the parameters in the linear regression models.

The structural equation model comprises:

(1) a set of linear equations identifying or detailing the causal relationship between the variables in the model, and

(2) several supporting assumptions. Similarly, to linear equations, SEM establishes a direct relationship between any cause and any effect that is generally specified by the coefficients connecting or associating two variables in the equation. As a result, the coefficient is the variation in effect generated by a one-unit variation in the level of the cause holding the other causes constant. Generally, the value of the coefficient is unknown.

We have noticed a great need for the development and improvement of new analytical methods in agricultural science. The SEM and path analysis are presented by developing appropriate structural equations and path diagrams. The linear relationship in a system of equation models can be presented in different ways, but in this study, these equations are offered as given in Equations 3.1a to 3.1c.

Section 3.3 presents the basic characteristics of SEMs and path analysis. Section 3.4 highlights the methodology. In Section 3.5, through the estimation of parameters, we develop a model of observable facts of an interesting SEM. The developed model is tested by means of the variance-covariance technique based on factor analysis in the SEM. Their contributions to the field of agricultural science are illustrated through a practical example given in Section 3.6. In Section 3.7, a conclusion and useful recommendations are given, and Section 3.8 presents a simulation study for testing the proposed model to determine if this model will perform better than the existing model.

3.2 OBJECTIVES

The study aims to:

1) develop an improved approach of SEM that is more flexible and unrestrictive by using factor analysis through PCA in the application of variables.

2) demonstrate and test the new model by the covariance technique when the maximum likelihood is used to estimate the parameters in the model.

3) introduce the SEM in the agricultural field by solving complex problems as new applied research.

4) demonstrate through a practical example the systematic application of the new technique by showing its contributions to the studies in the agricultural field.

3.3 CHARACTERISTICS OF THE STRUCTURAL EQUATION MODEL

The concept of causality has always been an alarming issue in various scientific fields, like in social sciences. Similarly, SEMs come across the causality hypothesis that is normally tested in non-empirical study models. Wright (1921) was the first to suggest SEM in a complete approach with regression analysis as a foundation to test the relationship between observed and implicit variables (Raykov & Marcoulides, 2000). In addition, SEM can perform multiple regression tests with two or more indirect or hidden variables subject to a few display variables associated with error terms. In general, SEM remains wholly subjected to the theoretical suggestion that the SEM will demonstrate whether the previously defined connection pattern could be supported or not by the collected data. In other words, we use SEM in the prediction of unknown parameters on a linear structure of equations. The variables in a set of SEMs are directly and indirectly observed. In SEM, we assume the existence of a causality (or interconnection) structure between the directly observed variables and the indirectly measured variables.

Technically, SEMs hold one or more linear regressions that explain how endogenous structures are determined upon exogenous structure. That means, in SEM the focus is in terms of measurement of variables that define just how theoretical (indirect) structures depend on observed variables when assuming a causality relationship between indirect variables. Path analysis (PA) and confirmatory factor analysis (CFA) are special types of SEM. PA examines how independent variables are statistically related to a dependent variable. Moreover, PA can allow causal interpretation of statistical dependencies, and most importantly, PA allows for the examination of how the data fits a theoretical model. PA enables us to draw a path diagram based on the theory and conduct one or more regression analyses (see Figures 3.1 and 3.2).

The estimation process in SEM involves different techniques, which include maximum likelihood, commonly used by software programs. It assumes either multivariate normality or generalised least square of robust estimators. Linear structural relation modelling (LISREL) is used, which is a powerful and flexible approach to analyse complex data. In applying SEM, we can assess hypothesised relations between variables that are consistent with the observed relations between the variables. The researcher starts by conceiving how a set of variables are related to another. In some instances, a model that is assumed may not fit the observed data. In such a condition, we can use LISREL, since it provides researchers with special quantitative estimates of the hypothesised model that can fit the observed data. Many researchers have used the approach, for instance Oiu et al. (2020), in E-learning assessment for tourism education, in which LISREL enabled them to assess intercultural tourism perception. Gumus et al. (2020) applied LISREL to investigate sportsmanship behaviour of in university students. Jöreskog et al. (1982) described new developments in SEM in using LISREL. The research in this work for the application of SEM uses Jöreskog's LISREL notations as presented by Bentler et al. (1980) and defined as follows:

$$\boldsymbol{\eta} = \mathbf{B}\ \boldsymbol{\eta} + \boldsymbol{\Gamma}\ \boldsymbol{\xi} + \boldsymbol{\varsigma} \tag{3.1a}$$

$$\mathbf{Y} = \boldsymbol{\Lambda}^{(y)}\ \boldsymbol{\eta} + \boldsymbol{\varepsilon} \tag{3.1b}$$

$$\mathbf{X} = \boldsymbol{\Lambda}^{(x)}\, \boldsymbol{\xi} + \boldsymbol{\delta} \tag{3.1c}$$

where, $\boldsymbol{\eta}$ represents the random vector latent dependent variables, **B** indicates the weights (parameter matrices) for predicting dependent variables from each other, **Γ** represents weights (parameter matrices) for predicting dependent variable from independent variables, **B** and **Γ** are coefficient matrices for linear relations of all variables involved in SEM), $\boldsymbol{\xi}$ denotes the random vector latent independent variables, $\boldsymbol{Y}$ indicates the vector of the observed indicator for a latent dependent variable, and $\boldsymbol{X}$ denotes the vector of the observed indicator for a latent independent variable. $\boldsymbol{\Lambda}^{(x)}$ and $\boldsymbol{\Lambda}^{(y)}$ are parameter matrices, and $\boldsymbol{\varepsilon}$, $\boldsymbol{\varsigma}$ and $\boldsymbol{\delta}$ are random vectors.

Path diagrams present the models graphically, enabling researchers to visualize the research's conceptual models and show statistical results. Moreover, path diagrams present functional relationships among multiple regression models that are a special case of structural equation model. From the output given by the path diagram, when the *p*-value is greater than a 5% level of significance, we conclude that the theoretical model is not a good one for the data. To illustrate this process, we use an example. Consider four dependent and four independent latent variables upon which we want to establish the system of equations of the observed vectors ($\boldsymbol{Y}$ and $\boldsymbol{X}$) and theoretical model ($\boldsymbol{\eta}$) using (3.1a), (3.1b) and (3.1c) as shown in Figure 3.1.

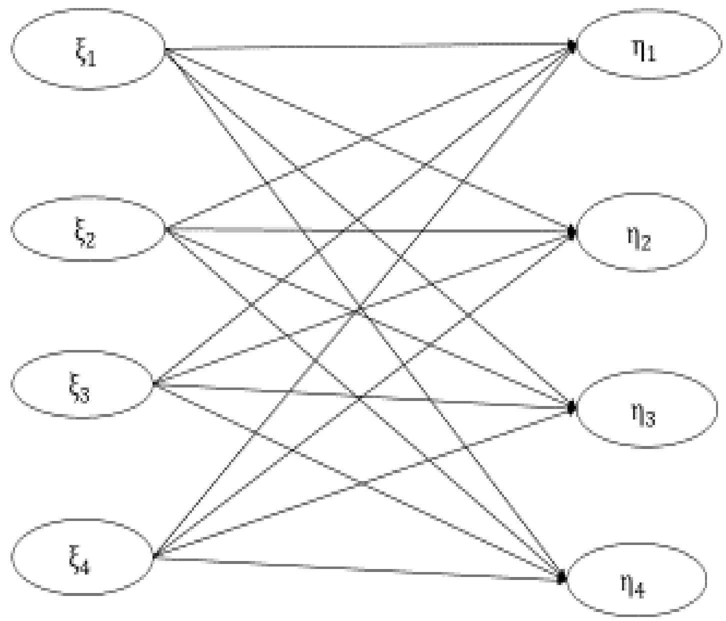


Figure 3.1: Path diagram

Figure 3.1 presents a path diagram provide the relationships between the exogenous and endogenous variables.

Usually, path analysis provides a diversity of a set of relationships that can be developed among the variables. However, some of these variables are similar. Therefore, there is a need for a more advanced technique (or method) that allows us to reduce a huge number of variables into a small number. Factor analysis (FA) serves this purpose. FA is a multivariate statistical method for reducing large numbers of variables to fewer underlying dimensions. This method involves the grouping of similar variables into dimensions. This process is used to identify latent variables or constructs. Most often, factors are rotated after extraction. FA has several different rotation methods, and some of them ensure that the factors are orthogonal (i.e. uncorrelated), which eliminates multi-collinearity problems in regression analysis. There are many techniques for FA, with principal component analysis being the most frequently used, followed by the exploratory factor analysis (EFA). PCA is used if the components can be derived or/and summarised. It has been used by many researchers in medical science, education, social science, and many other related fields (Bolt et al., 2018; Wang & Staver 2001). However, EFA is used if the variables have unmeasured variables. It is not as popular as PCA. In this study, we integrate FA into SEM to provide an optimal and cost-effective model that explains the key factors in the food production system better.

3.4 METHODOLOGY

3.4.1 The current approach of the structural equation model

The traditional approach of SEM is more restrictive since it specifies the latent variables that are involved in the analysis and creates the theoretical relations between the variables. There is a huge diversity in a set of relationships that could be developed among the variables. The variability of a set of relationships points to inconsistent conclusions about the level at which a model is truly equivalent to the observed data. Therefore, a variety of the options of the path diagram is utilised. We present a more reliable approach that provides a guideline on how to evaluate the suitability of a given SEM. Researchers in the agricultural sector uses all possible variables that might be identified for a set of data, but by using factor analysis through the PCA, researchers can use the most important variables in the model. SEM has been commonly applied in many fields after it has been introduced in the agriculture field.

3.4.2 The Proposed Structural Equation Models using Factor Analysis

We outline the necessary steps to take in producing SEM using factor analysis after obtaining provisional factors via PCA as follows:

1) Screen the data for suitability through testing.

2) Apply PCA on the correlation matrix to obtain provisional factors when the test in Step (1) is statistically significant. Using the Factor analysis (FA), calculate the communalities accounting for the pre-set proportion of total variation.

3) Determine the number of principal components to retain and rotate to obtain orthogonality.

4) Interpret the new variables (FAs) based on the factor loading for each variable.

5) Consider rotating the factors to attain orthogonality. Thus, the final factors are orthogonal.

6) Determine the component score coefficient matrix for the possible models.

7) The process assists with the selection of original variables for an easy interpretation of the components. The new variables ($\boldsymbol{F}_i$) are unknown while the original variables are observed. This can be referred to as the unobserved and observed variables in the current approach (3.1b and 3.1c).

8) After the selection of the variables has been done, the path diagrams that represent the model graphically can be completed. This enables the researcher to visualise the conceptual model behind the research question by showing the statistical results. Path diagrams represent functional relationships among the multiple regression models.

9) The correlations between the latent (unobserved) variables and latent (observed) variables in the current approach are equivalent to factor loading in PCA with the new approach.

10) Parameter estimates in SEM are calculated by the maximum likelihood method.

The linear combination of the original variables can be constructed in general as in Equation 3.2.

$F_i = a_{i1}X_1 + a_{i2}X_2 + ... + a_{im}X_m$ (3.2)

where F_i (i = 1, 2, ..., m) represent the new variables called factors which give a linear combination of the original variables (X_1,X_2, ..., X_n) with $m < n$. Let a_1,a_2, ..., a_m be the weights associated with the original variables. They represent the correlation between the components and the original variables. The set of equations obtained through the PCA can then be formulated as

$F_1 = a_{11}X_1 + a_{12}X_2 + ... + a_{1m}X_m$ (3.2a)

$F_2 = a_{21}X_1 + a_{22}X_2 + ... + a_{2m}X_m$ (3.2b)

$F_m = a_{m1}X_1 + a_{m2}X_2 + ... + a_{mm}X_m$ (3.2c)

$Var(F_i) = \sum_{k=1}^{m}\sum_{l=1}^{n} a_{ik}\ a_{il}\sigma_{kl} = a'_i\Sigma a_i$ (3.3)

3.5 ESTIMATION OF PARAMETERS

3.5.1 Introduction

Estimation of parameters in SEM is obtained by the maximum likelihood method. It provides estimates for the linear equations that reduce the deviation between the observed and the proposed model. We incorporate the selected factors into a few SEMs and then test for the different inter-associations among the latest variables. The correlations between the latent (unobserved) variables and latent (observed) variables were equivalent to factor loading in PCA. The general SEM as given in Equation 3.1a is equivalent to the Equation 3.2 summarised as

$\boldsymbol{\eta} = \boldsymbol{\beta\eta} + \boldsymbol{\Sigma\xi} + \boldsymbol{\varphi}$, (3.4)

where

$$\boldsymbol{\eta} = \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} 0 & \beta_{12} & \beta_{13} & \beta_{14} \\ 0 & 0 & \beta_{23} & \beta_{24} \\ 0 & 0 & 0 & \beta_{34} \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \boldsymbol{\Gamma} = \begin{pmatrix} \delta_{11} & \delta_{12} & \delta_{13} & \delta_{14} \\ \delta_{21} & \delta_{22} & \delta_{23} & \delta_{24} \\ \delta_{31} & \delta_{32} & \delta_{33} & \delta_{34} \\ \delta_{41} & \delta_{42} & \delta_{43} & \delta_{44} \end{pmatrix}, \boldsymbol{\xi} = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \end{pmatrix} \text{ and}$$

$$\boldsymbol{\varphi} = \begin{pmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_3 \\ \varphi_4 \end{pmatrix}. \tag{3.5}$$

These structures of random vectors and parameter matrices are used in the data analysis.

3.5.2 Constructive Relations Based on Endogenous and Exogenous Variables

3.5.2.1 Constructive relations based on the endogenous variables

In using the SEM technique, we want to demonstrate the use of the proposed LISREL procedure that can be able to analyse a complex of data. The LISREL approach requires defining the hypothesised model that includes a system of equations of the observed relations model **X** and **Y** representing the exogenous and endogenous variables given in the Equations (3.1a) to (3.1c). In this section, we present the estimation of parameter process through LISREL. For illustration purposes, suppose we are given four endogenous variables and four exogenous variables. Considering the endogenous variables, we would like to construct the equations from the four dependent latent variables model **Y,** using the concept of Equation 3.1b. The endogenous **Y** comprises $(y_1, y_2, y_3, \ldots, y_m)$, the parameter matrix Λ^y represented by $(\lambda_{11}, \lambda_{21}, \lambda_{31}, \ldots, \lambda_{m1})$, the random vector latent dependent variable $\boldsymbol{\eta}$ = ($\eta_1, \eta_2, \eta_3, \ldots, \eta_m$) and the random error $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots, \varepsilon_m)$ as shown below

$$y_1 = \lambda_{11}\eta_1 + \varepsilon_1$$
$$y_2 = \lambda_{21}\eta_1 + \varepsilon_2$$
$$y_3 = \lambda_{31}\eta_1 + \varepsilon_3$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$y_m = \lambda_{m1}\eta_1 + \varepsilon_m$$
$$y_i = \lambda_{i1}\eta_1 + \varepsilon_i.$$

In matrix notation,

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} \lambda_{11} \\ \lambda_{21} \\ \vdots \\ \lambda_{m1} \end{pmatrix} \eta_1 + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{pmatrix}.$$

The model is then

$\mathbf{Y} = \boldsymbol{\lambda}^{(y)} \eta_1 + \boldsymbol{\varepsilon}$.

and

$$\begin{aligned} y_{m+1} &= \lambda_{(m+1)_2}\eta_2 + \varepsilon_{m+1} \\ y_{m+1} &= \lambda_{(m+2)_2}\eta_2 + \varepsilon_{m+2} \\ &\ldots\ldots\ldots\ldots \\ y_s &= \lambda_{s_2}\eta_2 + \varepsilon_s \\ y_i &= \lambda_{i2}\eta_2 + \varepsilon_i. \end{aligned} \qquad (3.6)$$

Using the matrix notation,

$$\begin{pmatrix} y_{m+1} \\ y_{m+2} \\ \vdots \\ y_s \end{pmatrix} = \begin{pmatrix} \lambda_{(m+1)_2} \\ \lambda_{(m+2)_2} \\ \vdots \\ \lambda_{s_2} \end{pmatrix} \eta_2 + \begin{pmatrix} \varepsilon_{m+1} \\ \varepsilon_{m+2} \\ \vdots \\ \varepsilon_s \end{pmatrix}.$$

The model is $\mathbf{Y} = \boldsymbol{\lambda}^{(y)} \eta_2 + \boldsymbol{\varepsilon}$ (3.7)

and

$$\begin{aligned} y_{s+1} &= \lambda_{(s+1)_3}\eta_3 + \varepsilon_{s+1} \\ y_{s+2} &= \lambda_{(s+2)_3}\eta_3 + \varepsilon_{s+2} \\ &\ldots\ldots\ldots\ldots \\ y_t &= \lambda_3\eta_3 + \varepsilon_t \\ y_i &= \lambda_{i3}\eta_3 + \varepsilon_i \end{aligned}$$

Using the matrix notation,

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} \lambda_{11} \\ \lambda_{21} \\ \vdots \\ \lambda_{m1} \end{pmatrix} \eta_1 + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{pmatrix}.$$

The model becomes

$\boldsymbol{Y} = \boldsymbol{\lambda}^{(y)} \eta_3 + \boldsymbol{\varepsilon}$ (3.8a)

and

$$y_{t+1} = \lambda_{(t+1)_4}\eta_4 + \varepsilon_{t+1}$$
$$y_{t+2} = \lambda_{(t+2)_4}\eta_4 + \varepsilon_{t+2}$$
……………………........
$$y_n = \lambda_{n_4}\eta_4 + \varepsilon_n$$
$$y_i = \lambda_{i4}\eta_4 + \varepsilon_i.$$

Using the matrix notation,

$$\begin{pmatrix} y_{t+1} \\ y_{t+2} \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \lambda_{(t+1)_4} \\ \lambda_{(t+2)_4} \\ \vdots \\ \lambda_{n_4} \end{pmatrix} \eta_1 + \begin{pmatrix} \varepsilon_{t+1} \\ \varepsilon_{t+2} \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

The model is

$$\mathbf{Y} = \boldsymbol{\lambda}^{(y)}\, \eta_4 + \boldsymbol{\varepsilon}. \qquad (3.8b)$$

The combined equations of the four dependent latent variables altogether in matrix notation is

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \\ y_{m+1} \\ y_{m+2} \\ \vdots \\ y_s \\ y_{s+1} \\ y_{s+2} \\ \vdots \\ y_t \\ y_{t+1} \\ y_{t+2} \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \lambda_{11} & 0 & 0 & 0 \\ \lambda_{21} & 0 & 0 & 0 \\ \vdots & 0 & 0 & 0 \\ \lambda_{m1} & 0 & 0 & 0 \\ 0 & \lambda_{m+1} & 0 & 0 \\ 0 & \lambda_{m+2} & 0 & 0 \\ 0 & \vdots & 0 & 0 \\ 0 & \lambda_s & 0 & 0 \\ 0 & 0 & \lambda_{s+1} & 0 \\ 0 & 0 & \lambda_{s+2} & 0 \\ 0 & 0 & \vdots & 0 \\ 0 & 0 & \lambda_t & 0 \\ 0 & 0 & 0 & \lambda_{t+1} \\ 0 & 0 & 0 & \lambda_{t+2} \\ 0 & 0 & 0 & \vdots \\ 0 & 0 & 0 & \lambda_n \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_{m1} \\ \varepsilon_{m+1} \\ \varepsilon_{m+2} \\ \vdots \\ \varepsilon_s \\ \varepsilon_{s+1} \\ \varepsilon_{s+2} \\ \vdots \\ \varepsilon_t \\ \varepsilon_{t+1} \\ \varepsilon_{t+2} \\ \vdots \\ \varepsilon_n \end{pmatrix}. \qquad (3.9)$$

In a general display, the latent dependent model is

$$\mathbf{Y} = \boldsymbol{\lambda}^{(y)} \boldsymbol{\eta} + \boldsymbol{\varepsilon}$$

Using the LISREL approach, we can construct a model through the endogenous **Y** that is observed, the parameter matrix $\boldsymbol{\lambda}^{(y)}$ that will be estimated by the maximum likelihood and the random vector latent dependent variable $\boldsymbol{\eta}$ that is unobserved. The final model will be expressed in terms of $\boldsymbol{\eta}$.

3.5.2.2 Constructive relations based on the exogenous variables

Similarly, to the application of LISREL in Section 3.5.2.1, considering the exogenous variables we would like to construct the equations from the four dependent latent variables as per model **X** using the concept of Equation 3.1c. The exogenous **X** comprises $(x, x_2, x_3, \ldots, x_m)$, the parameter matrix $\boldsymbol{\Lambda}^{x}$ represented by $(\lambda_{11}, \lambda_{21}, \lambda_{31}, \ldots, \lambda_{m1})$, the random vector latent dependent variable $\boldsymbol{\eta}$ = $(\eta_1, \eta_2, \eta_3, \ldots, \eta_m)$ and the random error $\boldsymbol{\delta} = (\delta_1, \delta_2, \delta_3, \ldots, \delta_m)$ as shown below

$$x_1 = \lambda_{11}\xi_1 + \delta_1$$
$$x_2 = \lambda_{21}\xi_1 + \delta_2$$
$$x_3 = \lambda_{31}\xi_1 + \delta_3$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$x_m = \lambda_{m1}\xi_1 + \delta_m$$
$$x_i = \lambda_{i1}\xi_1 + \delta_i$$

In matrix notation

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} = \begin{pmatrix} \lambda_{11} \\ \lambda_{21} \\ \vdots \\ \lambda_{m1} \end{pmatrix} \xi_1 + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_m \end{pmatrix}$$

The model is $\mathbf{X} = \boldsymbol{\lambda}^{(x)}\, \xi_1 + \boldsymbol{\delta}$

$$x_{m+1} = \lambda_{(m+1)_2}\ \xi_2 + \delta_{m+1}$$
$$x_{m+1} = \lambda_{(m+2)_2}\ \xi_2 + \delta_{m+2} \qquad (3.10a)$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$x_s = \lambda_{s_2}\xi_2 + \delta_s$$
$$x_i = \lambda_{i2}\xi_2 + \delta_i$$

Using the matrix notation

$$\begin{pmatrix} x_{m+1} \\ x_{m+2} \\ \vdots \\ x_s \end{pmatrix} = \begin{pmatrix} \lambda_{(m+1)_2} \\ \lambda_{(m+2)_2} \\ \vdots \\ \lambda_{s_2} \end{pmatrix} \xi_2 + \begin{pmatrix} \delta_{m+1} \\ \delta_{m+2} \\ \vdots \\ \delta_s \end{pmatrix}$$

The model is $\mathbf{X} = \boldsymbol{\lambda}^{(x)} \xi_2 + \boldsymbol{\delta}$ (3.10b)

$$x_{s+1} = \lambda_{(s+1)_3}\xi_3 + \delta_{s+1}$$

$$x_{s+2} = \lambda_{(s+2)_3}\xi_3 + \delta_{s+2}$$

. . . .

$$x_t = \lambda_3\xi_3 + \delta_t$$

$$x_i = \lambda_{i3}\xi_3 + \delta_i$$

Using the matrix notation

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} = \begin{pmatrix} \lambda_{11} \\ \lambda_{21} \\ \vdots \\ \lambda_{m1} \end{pmatrix} \xi_3 + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_m \end{pmatrix}$$

The model is $\mathbf{X} = \boldsymbol{\lambda}^{(x)} \xi_3 + \boldsymbol{\delta}$ (3.10c)

$$x_{t+1} = \lambda_{(t+1)_4}\xi_4 + \delta_{t+1}$$

$$x_{t+2} = \lambda_{(t+2)_4}\xi_4 + \delta_{t+2}$$

. . . .

$$x_n = \lambda_{n_4}\xi_4 + \delta_n$$

$$x_i = \lambda_{i4}\xi_4 + \delta_n$$

Using the matrix notation

$$\begin{pmatrix} x_{t+1} \\ x_{t+2} \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \lambda_{(t+1)_4} \\ \lambda_{(t+2)_4} \\ \vdots \\ \lambda_{n_4} \end{pmatrix} \xi_4 + \begin{pmatrix} \delta_{t+1} \\ \delta_{t+2} \\ \vdots \\ \delta_n \end{pmatrix}.$$

The model is $\mathbf{X} = \boldsymbol{\lambda}^{(x)} \xi_4 + \boldsymbol{\delta}$ (3.10d)

The combination of the four independent latent variables altogether is

$$
\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \\ x_{m+1} \\ x_{m+2} \\ \vdots \\ x_s \\ x_{s+1} \\ x_{s+2} \\ \vdots \\ x_t \\ x_{t+1} \\ x_{t+2} \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \lambda_{11} & 0 & 0 & 0 \\ \lambda_{21} & 0 & 0 & 0 \\ \vdots & 0 & 0 & 0 \\ \lambda_{m1} & 0 & 0 & 0 \\ 0 & \lambda_{m+1} & 0 & 0 \\ 0 & \lambda_{m+2} & 0 & 0 \\ 0 & \vdots & 0 & 0 \\ 0 & \lambda_s & 0 & 0 \\ 0 & 0 & \lambda_{s+1} & 0 \\ 0 & 0 & \lambda_{s+2} & 0 \\ 0 & 0 & \vdots & 0 \\ 0 & 0 & \lambda_t & 0 \\ 0 & 0 & 0 & \lambda_{t+1} \\ 0 & 0 & 0 & \lambda_{t+2} \\ 0 & 0 & 0 & \vdots \\ 0 & 0 & 0 & \lambda_n \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_{m1} \\ \delta_{m+1} \\ \delta_{m+2} \\ \vdots \\ \delta_s \\ \delta_{s+1} \\ \delta_{s+2} \\ \vdots \\ \delta_t \\ \delta_{t+1} \\ \delta_{t+2} \\ \vdots \\ \delta_n \end{pmatrix}
$$

In a general display, the independent latent model is $\mathbf{X} = \boldsymbol{\lambda}^{(x)} \boldsymbol{\xi} + \boldsymbol{\delta}$. (3.11)

The LISREL approach enables the researcher to construct a model through the exogenous **X** that is observed, the parameter matrix $\boldsymbol{\lambda}^{(x)}$, that will be estimated by using the maximum likelihood and the random vector latent dependent variable $\boldsymbol{\xi}$ that is unobserved. The final model will be expressed in terms of $\boldsymbol{\xi}$.

Linear structural relation based on the LISREL approach is a powerful and flexible approach to analyse complex data. This technique assesses hypothesised relations between variables that are consistent with the observed relations between the variables. The researcher starts by conceiving how a set of variables are related to another. If that model assumes a specific latent variable that does not fit the observed data, then LISREL can provide the researcher with specific quantitative estimates of the hypothesised model that would fit the observed data.

Based on the theoretical LISREL approach Equation 3.4 is

$\boldsymbol{\eta} = \boldsymbol{B}\ \boldsymbol{\eta} + \boldsymbol{\Gamma}\ \boldsymbol{\xi} + \boldsymbol{\varphi}$.

The Equation (3.4) is equivalent to

$$\begin{pmatrix}\eta_1\\ \eta_2\\ \eta_3\\ \eta_4\end{pmatrix} = \begin{pmatrix}0 & \beta_{12} & \beta_{13} & \beta_{14}\\ 0 & 0 & \beta_{23} & \beta_{24}\\ 0 & 0 & 0 & \beta_{34}\\ 0 & 0 & 0 & 0\end{pmatrix}\begin{pmatrix}\eta_1\\ \eta_2\\ \eta_3\\ \eta_4\end{pmatrix} + \begin{pmatrix}\delta_{11} & \delta_{12} & \delta_{13} & \delta_{14}\\ \delta_{21} & \delta_{22} & \delta_{23} & \delta_{24}\\ \delta_{31} & \delta_{32} & \delta_{33} & \delta_{34}\\ \delta_{41} & \delta_{42} & \delta_{43} & \delta_{44}\end{pmatrix}\begin{pmatrix}\xi_1\\ \xi_2\\ \xi_3\\ \xi_4\end{pmatrix} + \begin{pmatrix}\varphi_1\\ \varphi_2\\ \varphi_3\\ \varphi_4\end{pmatrix}$$

which can also be written as

$$\begin{pmatrix}\eta_1\\ \eta_2\\ \eta_3\\ \eta_4\end{pmatrix} = \begin{pmatrix}\beta_{12}\,\eta_2 + \beta_{13}\eta_3 + \beta_{14}\,\eta_4\\ \beta_{23}\,\eta_3 + \beta_{24}\eta_4\\ \beta_{34}\,\eta_4\\ 0\end{pmatrix} + \begin{pmatrix}\delta_{11}\,\xi_1 & \delta_{12}\xi_2 & \delta_{13}\xi_3 & \delta_{14}\,\xi_4\\ \delta_{21}\,\xi_1 & \delta_{22}\xi_2 & \delta_{23}\xi_3 & \delta_{24}\,\xi_4\\ \delta_{31}\xi_1 & \delta_{32}\xi_2 & \delta_{33}\xi_3 & \delta_{34}\xi_4\\ \delta_{41}\xi_1 & \delta_{42}\xi_2 & \delta_{43}\xi_3 & \delta_{44}\xi_4\end{pmatrix} + \begin{pmatrix}\varphi_1\\ \varphi_2\\ \varphi_3\\ \varphi_4\end{pmatrix}.$$

The above equations can be simplified as follows:

$$\eta_1 = \beta_{12}\,\eta_2 + \beta_{13}\eta_3 + \beta_{14}\,\eta_4 + \delta_{11}\,\xi_1 + \delta_{12}\,\xi_2 + \delta_{13}\,\xi_3 + \delta_{14}\,\xi_4 + \varphi_1 \quad (3.12a)$$

$$\eta_2 = \beta_{23}\,\eta_3 + \beta_{24}\eta_4 + \delta_{21}\,\xi_1 + \delta_{22}\,\xi_2 + \delta_{23}\,\xi_3 + \delta_{24}\,\xi_4 + \varphi_2 \quad (3.12b)$$

$$\eta_3 = \beta_{34}\,\eta_4 + \delta_{31}\,\xi_1 + \delta_{32}\,\xi_2 + \delta_{33}\,\xi_3 + \delta_{44}\,\xi_4 + \varphi_3 \quad (3.12c)$$

$$\eta_4 = \delta_{41}\,\xi_1 + \delta_{42}\,\xi_2 + \delta_{43}\,\xi_3 + \delta_{44}\,\xi_4 + \varphi_4. \quad (3.12d)$$

Using SEM through the LISREL approach, we can construct a model through the endogenous **Y** and exogenous **X** that are observed variables, the parameter matrices $\boldsymbol{\lambda}^{(y)}$ and $\boldsymbol{\lambda}^{(x)}$ are estimated by the maximum likelihood, and the random vector latent dependent variables $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ are the unobserved variables. The general model is expressed in terms of $\boldsymbol{\eta}$ by using the Equation 3.1a. To illustrate the use of SEM through the LISREL approach, an example in agricultural research food production is used upon which we demonstrate the proposed approach.

3.6 DATA ANALYSIS

3.6.1 Introduction

To illustrate the modelling process in Section 3.5, real-life data from the Food and Agriculture Organisation (FAO, 2021) from 45 African countries has been used. The data components were crops (in tons), livestock (number of heads), and contributors such as fertiliser (in tons), trade (in dollars), labour (in number of people), land (in hectares) and water (in litters). Table 3.1 presents the structure of the data from 33 countries between 2012 and 2018 with the LISREL notations according to Jöreskog (2000).

Table 3.1: Crop components classified into three vital factors (crop, livestock and contributors) with various factor levels denoted by LISREL

Components		Description of variables	LISREL notations
Crop (in tons)		Banana	Y_1
		Beans	Y_2
		Cassava	Y_3
		Rice	Y_4
		Groundnut	Y_5
		Maize	Y_6
		Sugar cane	Y_7
		Vegetables	Y_8
		Cereals	Y_9
		Fruits	Y_{10}
Livestock (in numbers of heads)		Cattle and Buffaloes	Y_{11}
		Pigs	Y_{12}
		Poultry	Y_{13}
		Sheep and Goats	Y_{14}
Contributors	Fertiliser (Factor 1) (in tons)	Nitrogen	X_1
		Phosphate	X_2
	Trade (Factor 2) (in Dollars)	Export values	X_3
		Import values	X_4
	Labour (Factor 3) (in numbers of people)	Rural	X_5
		Urban	X_6
	Land (Factor4) (hectares)	Arable	X_7
		Permanent	X_8
	Water (in Littre) (Factor 5)	Rainfall	X_9
		Irrigated land	X_{10}
	Energy used (in Kilowatts) (Factor 6)	Electricity	X_{11}
		Diesel	X_{12}
		Transport	X_{13}

In this section, we aim to estimate parameters using SEM through LISREL approach. This technique requires formulating the hypothesised model that can fit the observed data, and therefore this approach will enable the researcher to determine the relationships between the variables. In the literature, the relationships are specified by the researcher, leading to inconsistent results in some situations. Using the proposed approach of SEM through LISREL, we define the observed variables represented by the dependent variable denoted by **Y** and the independent variable denoted by **X**. To determine **X** and **Y** in the complexity of variables and data, PCA is used to determine the directly observed variables to decide about the number of factors to be retained in the model. The PCA approach is strongly related to factor analysis, indicating the correlations or associations between the variables, and determining the small number of latent variables. For an illustration of this technique, countries in SSA were grouped into crop production, livestock and contributing factor dimensions from which inferences will be drawn to obtain stable estimated parameters for the solutions to the problem facing African countries. We used the PCA approach to determine direct and indirect variables based on the 26 variables. The exogenous variable **X** comprises 12 factors denoted by X_1 to X_{12}, and the endogenous variable **Y** is represented by 14 factors denoted by Y_1 to Y_{14} as provided in Table 3.1.

In applying the PCA, the correlation matrix was used to determine the variables that were the most strongly correlated with each component, but one variable was eliminated because no correlation was found. The screening of variables reduced the number of highly correlated variables from 25 to 10 new independent variables, as indicated in Table 3.2. The retained variables account for much of the total variation in the variable of interest, which is explained by each component, as this cannot be performed in multiple regressions. The results of PCA determined the levels at which the variables were measured. The variables with the highest sample variances were among the few components taken as each variable received its weight in the analysis. To receive equal weight in the analysis, we have then standardised variables before carrying out the PCA (performing PCA on a correlated matrix). Table 3.2 shows the number of components and the eigenvalues (initial and rotation eigenvalues).

Table 3.2: Screening of different variables through PCA based on the total variance explained

Component	Initial Eigenvalues			Rotation Sums of Squared loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.746	18.985	18.985	3.077	12.310	12.310
2	3.254	13.017	32.002	2.950	11.799	24.109
3	2.742	10.969	42.971	2.930	11.720	35.829
4	2.435	9.741	52.712	2.681	10.725	46.554
5	1.885	7.542	60.254	2.123	8.492	55.046
6	1.545	6.181	66.435	2.052	8.207	63.253
7	1511	6.044	72.479	1.706	6.825	70.078
8	1.402	5.606	78.085	1.532	6.127	76.205
9	1.154	4.615	82.700	1.415	5.661	81.866
10	1.028	4.111	86.811	1.236	4.945	86.811
11	0.804	3.217	90.028			

Extraction Method: Principal Component Analysis

From Table 3.2, the rule proposed by Kaiser (1960) about a common practice criterion for the number of factors to rotate having eigenvalues greater than one, as shown in Column 2 of the Table 3.2, was used. It is given that there are as many factors as there are eigenvalues greater than one, resulting in the formula for internal consistency reliability. An eigenvalue less than one implies that the scores on the component can have negative reliability. About 87% of the total variation is accounted for by 10 out of 25 original variables. Thus, we rotate the ten principal components using FA to attain orthogonality.

3.6.2 Illustrative Example on Agricultural Data Analysis Using SEMs

In SSA countries, agriculture is one of the most dominant activities providing jobs for the population. Productivity in this part of the world remains low because of many challenges that go beyond weather, pests and lack of fertiliser. For instance, in the Northern part of the African continent, less than thirty per cent of the land is irrigated, and Africa is far behind in the use of more advanced agricultural technology. We have used food production to display the values of this modelling method.

The factor components are represented by component 1 to 10, as given in Table 3.3. The bold values are the highest correlations between the original variables and the components in the array.

Table 3.3: The rotated component matrix

	Factor components									
Original variables	1	2	3	4	5	6	7	8	9	10
Bananas	.100	-.045	.049	-.099	-.009	-.191	-.105	**.802**	-.038	-.039
Beans	-.050	-.029	.057	-.076	.114	.139	.075	**.821**	-.001	.073
Cassava	-.017	.040	**.872**	-.072	.089	.012	-.019	.114	-.044	.053
Rice	-.072	-.024	.090	.000	-.038	-.117	-.005	.050	-.031	**.886**
Groundnut	.033	**.960**	.208	-.041	-.009	-.017	.002	-.095	-.050	-.031
Maize	.074	**.990**	-.064	.034	.005	-.004	-.034	.003	-.027	.004
Sugar cane	**.804**	.083	.013	.421	.093	-.050	.203	.061	-.017	.062
Vegetables	.023	**.993**	.068	-.037	.001	-.005	-.032	.004	-.030	.013
Cereals	.382	.116	**.843**	.095	.008	-.088	-.022	.068	.081	.146
Fruit	.041	.059	**.933**	-.114	.023	-.056	.007	-.069	-.024	-.060
Export	**.707**	.016	.124	.254	.029	-.054	-.088	.243	.216	-.124
Import	**.659**	.071	**.558**	.255	.059	-.077	-.010	.126	.005	-.111
Irrigated	**.775**	.030	.200	.109	.144	-.102	-.065	-.169	-.078	-.051

Table 3.3 (Continued)

Original Variables	Factor Components									
	1	2	3	4	5	6	**7**	8	9	10
Rainfall	-.131	-.073	-.009	-.020	.274	.070	**.868**	-.045	.026	.191
Nitrogen	-.111	-.019	-.044	-.049	.024	**.919**	-.008	.062	-.105	-.076
Phosphate	-.011	-.012	-.055	-.017	.174	**.922**	-.010	-.098	.069	-.027
Rural	.031	.019	.092	.010	**.961**	.070	.088	.112	-.054	-.040
Urban	.221	-.018	.029	.112	**.934**	.142	-.025	.002	.029	-.010
Electricity	.335	-.044	-.047	**.860**	-.007	.090	-.149	-.067	.058	-.025
Diesel	.409	-.060	-.012	**.855**	-.015	-.010	.032	-.053	-.049	.000
Transport	-.016	.029	-.023	**.895**	.130	-.144	.037	-.106	.044	-.006
Cattle- Buffaloes	.120	.000	-.023	-.026	-.158	-.082	**.907**	.009	.031	-.189
Pigs	.120	-.042	-.077	.026	-.173	.076	-.028	-.080	**.889**	.115
Poultry	.128	.102	-.111	-.033	-.304	.222	-.154	-.073	**-.726**	.361
Sheep - goats	**-.565**	-.064	.348	.162	-.034	-.367	.013	.125	.006	-.396

The dominant variables explaining each of the ten factors accounting for 87% of the total variation are outlined below:

Factor 1 --- Sugar cane, Import, Irrigated crops and Sheep/Goat

Factor 2 --- Groundnut, Maize, and Vegetables
Factor 3 --- Cassava, Cereals and Fruits
Factor 4 --- Electricity, Diesel and Transport
Factor 5 --- Rural and Urban
Factor 6 --- Nitrogen and Phosphate
Factor 7 --- Rainfall and Cattle - Buffalos
Factor 8 --- Bananas and Beans
Factor 9 --- Pigs and Poultry
Factor 10 --- Rice.

The test for normality of the variables in each of the observed indicators for endogenous and exogenous variables is validated as shown in Tables 3.4 and 3.5

Table 3.4: Test for normality for the endogenous variables

Observation	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Chi-squared	88.942	263.882	113.417	18.676	8.068	6.940
Degrees of freedom	10	3	6	1	1	1
p-value	0.0001	0.0001	0.0001	0.0001	0.005	0.008

Table 3.5: Test for normality for the exogenous variables

Observation	Factor 1	Factor 2	Factor 3
Chi-squared	105.636	69.642	48.157
Degrees of freedom	3	1	1
p-value	0.0001	0.0001	0.0001

The variables were normally distributed since the *p*-value is less than 0.05 (level of significance), therefore, the maximum likelihood estimation can be used. The general linear SEM is given in Equations 3.1a, 3.1b and 3.1c. The latent endogenous and exogenous models are the most highly correlated of the factor loads in which the measurement model is obtained by the maximum likelihood. The model fit resulted from the goodness-of-fit statistical tests that explain the discrepancy between latent and unobserved variables. In this practice, the model fits the data well, as this indicates that no important paths have been omitted from the model.

After estimating the endogenous and exogenous latent measurement model separately, a joint model that includes the total latent model can now be estimated (Figure 3.2). Since latent variables are observed, the measurement is obtained indirectly through the latent endogenous and exogenous variables. The latent unobserved variables are represented as ellipses, and the latent observed variables are represented as rectangles. Because we cannot measure or estimate the unknown factors or parameters perfectly; the measurement is subjected to an error. Therefore, the error terms are associated with each of the latent observed variables as they form part of the overall model. The error terms are also represented as ellipses (Figure 3.2).

Based on the type of regression and relationship indicated in the diagram, the SEMs result from a potentially complex interplay between many observed and unobserved variables, including error terms. Using the variables in the data and corresponding identifier notations, we illustrate the inter-relationship using the path diagram. The path diagram represents the model in line with the overall outcome of this study. The maximum likelihood estimates were obtained using Equation 3.2, and the results in Table 3.3.

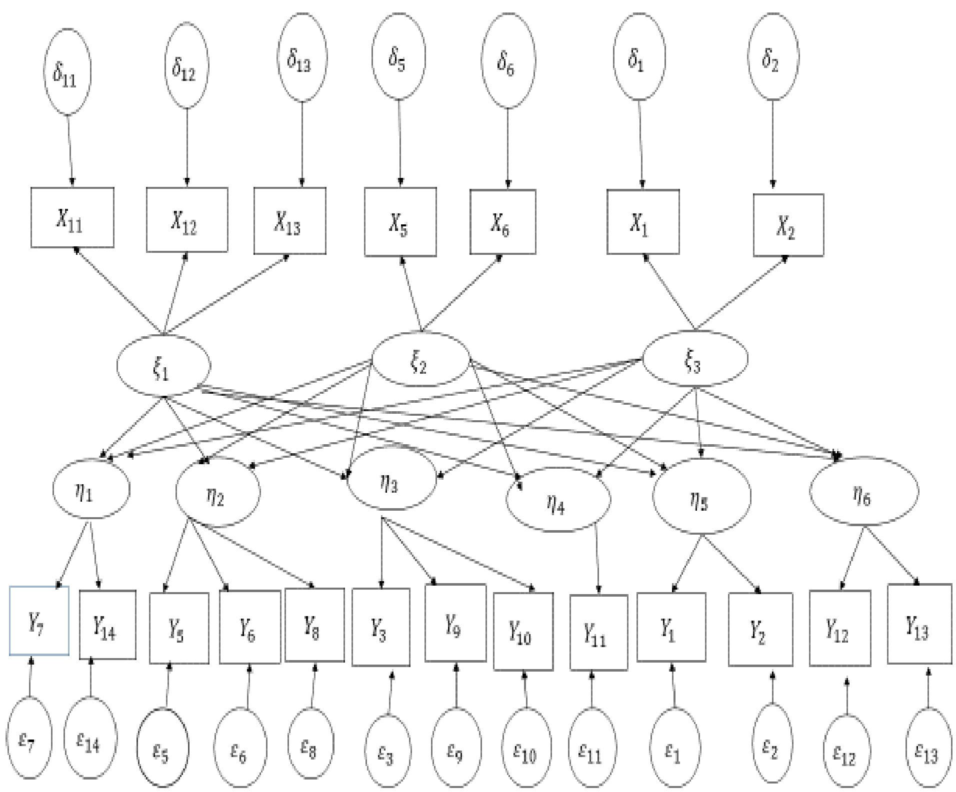


Figure 3.2: Conceptual path diagram for the structural model

Table 3.6 presents the endogenous variables under different models based on the factor loadings obtained from rotated provisional factors. The model equations, measurement model parameters and associated score components, in addition to the goodness-of-fit test statistics are also included.

Table 3.6: The endogenous descriptions model

Model	Factor load	Correlation	MEq	MModel	Component score coefficient	Goodn
1	Sugar Export Import Irrigation Sheep & goats	0.804 0.707 0.659 0.775 0.565	$Y_1 = \Lambda^y \eta_1 + \varepsilon$	$\begin{pmatrix} y_7 \\ y_{14} \\ x_3 \\ x_4 \\ x_{10} \end{pmatrix} = \begin{pmatrix} \lambda_{71}{}^y \\ \lambda_{14,1}{}^y \\ \lambda_{31}{}^y \\ \lambda_{41}{}^y \\ \lambda_{10,1}{}^y \end{pmatrix} \eta_1 + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{pmatrix}$	$\begin{pmatrix} y_7 \\ x_3 \\ x_4 \\ x_{10} \\ y_{14} \end{pmatrix} = \begin{pmatrix} 0 \\ 0.649 \\ 0.445 \\ 0 \\ 0 \end{pmatrix} \eta_1 + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{pmatrix}$	χ^2 = 8.018 df = 4 *p*-value= 0.005
2	Ground Maize Vegies	0.960 0.990 0.993	$Y_2 = \Lambda^y \eta_2 + \varepsilon$	$\begin{pmatrix} y_5 \\ y_6 \\ y_8 \end{pmatrix} = \begin{pmatrix} \lambda_{52}{}^y \\ \lambda_{62}{}^y \\ \lambda_{82}{}^y \end{pmatrix} \eta_2 + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$	$\begin{pmatrix} y_5 \\ y_6 \\ y_8 \end{pmatrix} = \begin{pmatrix} 0.055 \\ 0.936 \\ 0.013 \end{pmatrix} \eta_2 + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$	χ^2 = 0.637 df = 2 *p*-value= 0.0001

Table 3.6 (Continued)

Model	Factor Load	Correlation	MEq	MMode	Component score coefficient	Goodn
3	Cassava Cereals Fruits	0.872 0.843 0.933	$Y_3 = \Lambda^y \eta_3 + \varepsilon$	$\begin{pmatrix} y_3 \\ y_9 \\ y_{10} \end{pmatrix} = \begin{pmatrix} \lambda_{33}{}^y \\ \lambda_{93}{}^y \\ \lambda_{10,3}{}^y \end{pmatrix} \eta_3 + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$	$\begin{pmatrix} y_3 \\ y_9 \\ y_{10} \end{pmatrix} = \begin{pmatrix} 0.152 \\ 0.186 \\ 0.688 \end{pmatrix} \eta_3 + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$	χ^2 = 15.49 df = 2 *p*-value = 0.0001
4	Rainfall Cattle	0.868 0.907	$Y_4 = \Lambda^y \eta_4 + \varepsilon$	$\begin{pmatrix} X_9 \\ y_{11} \end{pmatrix} = \begin{pmatrix} \lambda_{94}{}^y \\ \lambda_{11,4}{}^y \end{pmatrix} \eta_4 + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$	$\begin{pmatrix} X_9 \\ y_{11} \end{pmatrix} = \begin{pmatrix} 0.560 \\ 0.560 \end{pmatrix} \eta_4 + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$	χ^2 = 16.68 df = 1 *p*-value = 0.0001
5	Banana Beans	0.802 0.821	$Y_5 = \Lambda^y \eta_5 + \varepsilon$	$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \lambda_{15}{}^y \\ \lambda_{25}{}^y \end{pmatrix} \eta_5 + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$	$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 0.594 \\ 0.594 \end{pmatrix} \eta_5 + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$	χ^2 = 8.068 df= 1 *p*-value = 0.005
6	Pigs Poultry	0.889 -0.776	$Y_6 = \Lambda^y \eta_6 + \varepsilon$	$\begin{pmatrix} y_{12} \\ y_{13} \end{pmatrix} = \begin{pmatrix} \lambda_{12,6}{}^y \\ \lambda_{13,6}{}^y \end{pmatrix} \eta_6 + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$	$\begin{pmatrix} y_{12} \\ y_{13} \end{pmatrix} = \begin{pmatrix} 0.600 \\ -0.600 \end{pmatrix} \eta_6 + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$	χ^2 = 6.940 df = 1 *p*-value = 0.008

MEq = Model equation MModel = Measurement model, Vegies = Vegetables, Ground = Groundnut, Goodn = Goodness-of-fit.

In applying SEM through LISREL, we present the relationships obtained in terms of endogenous variables given in Table 3.6 and in terms of exogenous variables, as shown in Table 3.7. Table 3.6

summarises all the information, such as the six endogenous models presented in Column 1, while the selected endogenous variables are in Column 2. Column 3 gives the correlations of the variables with the component factors denoted by η_i (i = 1, 2, …, 6). In Column 4, the six models are presented, the matrix notation of each model is given in Column 5. The estimated parameter for each model is shown in Column 6 and in Column 7, the goodness-of-fit test for each model. Similarly, Table 3.7 summarises all the information concerning the three exogenous models in the same order as in the endogenous model except that in Column 3, the correlations are expressed in terms of component factor ξ_j (j = 1, 2, 3).

Table 3.7: The exogenous descriptions model

	Factor load	Correlation	M.Eq.	MModel	Component score coefficient	Goodness-of-fit test
1	Electricity Diesel Transport	0.860 0.855 0.895	$X_1 = \Lambda^x \xi_1 + \delta$	$\begin{pmatrix} X_{11} \\ X_{12} \\ X_{13} \end{pmatrix} = \begin{pmatrix} \lambda_{33}{}^x \\ \lambda_{93}{}^y \\ \lambda_{10,3}{}^x \end{pmatrix} \xi_1 + \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix}$	$\begin{pmatrix} X_{11} \\ X_{12} \\ X_{13} \end{pmatrix} = \begin{pmatrix} 0.376 \\ 0.582 \\ 0.057 \end{pmatrix} \xi_1 + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{pmatrix}$	Chi-squared = 15.49 df = 2 *p*-value = 0.0001
2	Rural Urban	0.961 0.934	$X_2 = \Lambda^x \xi_2 + \delta$	$\begin{pmatrix} X_5 \\ X_6 \end{pmatrix} = \begin{pmatrix} \lambda_{52}{}^x \\ \lambda_{62}{}^x \end{pmatrix} \xi_2 + \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix}$	$\begin{pmatrix} X_5 \\ X_6 \end{pmatrix} = \begin{pmatrix} 0.513 \\ 0.513 \end{pmatrix} \xi_2 + \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix}$	Chi-squared = 69.64 df = 1 *p*-value = 0.0001

Table 3.7 (Continued)

	Factor Load	Correlation	M.Eq.	MModel	Component score coefficient	Goodness-of-fit test
3	Nitrogen Phosphate	0.919 0.922	$X_3 = \Lambda^x \xi_3 + \delta$	$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} \lambda_{1,3}{}^x \\ \lambda_{2,3}{}^x \end{pmatrix} \xi_3 + \begin{pmatrix} \delta_1 \\ \delta_3 \end{pmatrix}$	$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} 0.524 \\ 0.057 \end{pmatrix} \xi_3 + \begin{pmatrix} \delta_1 \\ \delta_3 \end{pmatrix}$	Chi-squared = 48.16 df = 1 *p*-value = 0.0001

MModel – Measurement model, M.Eq. = Model Equation

For instance, in this exogenous model, three principal components were the main factors responsible for improving food production in SSA. These components are obtained through the analysis of the data when using the PCA approach. The three components causing the changes derived from the data were energy (ξ_1), labour (ξ_2), and fertiliser (ξ_3), as indicated in the path diagram (Figure 3.2). These variables are called "exogenous variables" in this context. This is because they were governed by external factors in relation to the food products. In addition, these variables appear to be random. In another illustration, the exogenous variables may be fixed by the researcher (Sobel, 1986). On the other hand, we had six effects that were derived from the data; these variables were: η_1 , "Sugar cane and sheep - goat"; η_2, "Groundnut, maize and vegetable"; η_3, "Cassava, cereals and fruit"; $\eta_{4,}$, "cattle - buffaloes"; η_5, "bananas and beans"; and $\eta_{6,}$ "pigs and poultry". These variables are called "endogenous variables" given that their impact depends stochastically on the operational systems relating to food, required to solve the problem of hunger in SSA. The arrows between these variables indicate that one variable was a cause of the other variable and $\boldsymbol{\varepsilon}_i(i = 1, 2, \ldots, 6)$, and $\boldsymbol{\delta}_i$ (i = 1, 2, and 3) are random variables that are assumed to have a multivariate normal distribution.

This means the expectation of the vector $\boldsymbol{\varepsilon}$ or $\boldsymbol{\delta}$ is equal to zero. For instance, the matrix variance-covariance of $\boldsymbol{\varepsilon}$ or $\boldsymbol{\delta}$ was assumed to be zero and the $Cov\ (\ \varepsilon_1, \varepsilon_2\ ) = Cov\ (\ \varepsilon_2, \varepsilon_3\ ) = \ldots = Cov\ (\ \varepsilon_i, \varepsilon_j\ ) = 0$, where i = 1, 2, ..., n and j = 1, 2, ..., m.

The path diagram showed that the absence of curved arrows between the variables in $\boldsymbol{\varepsilon}$ or $\boldsymbol{\delta}$ indicated that the covariance matrix is equal to zero as assumed above.

This results from the power of the exploratory properties of factor analysis, showing a strong indication against orthogonality in the solutions in this complexity of data. Therefore, the six-measurement model in matrix notation for the exogenous model are equivalent to Path diagram 2 represented by Equation 3.1b, given by

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \\ y_{10} \\ y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \end{pmatrix} = \begin{pmatrix} 0.000 & 0.000 & 0.000 & 0.000 & 0.594 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.594 & 0.000 \\ 0.445 & 0.000 & 0.152 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.055 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.936 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.013 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.186 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.688 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.600 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & -0.60 \\ 0.649 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \\ \eta_6 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \\ \varepsilon_{10} \\ \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{14} \end{pmatrix} \quad (3.3)$$

In the same way, the exogenous measurement model represented by Equation 3.1c is given by

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \\ X_7 \\ X_8 \\ X_9 \\ X_{10} \\ X_{11} \\ X_{12} \\ X_{13} \end{pmatrix} = \begin{pmatrix} 0.376 & 0.582 & 0.057 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0.513 & 0 \\ 0 & 0.513 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0.524 \\ 0 & 0 & 0 \\ 0 & 0 & 0.057 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \\ \delta_8 \\ \delta_9 \\ \delta_{10} \\ \delta_{11} \\ \delta_{12} \\ \delta_{13} \end{pmatrix} \quad (3.4)$$

The SEM given by Jöreskog's (2000) formulas is given by Equation 3.2, as shown in Table 3.8.

Table 3.8: The parameters estimates and measurement model matrices:

$\boldsymbol{B} = \begin{pmatrix} 0 & 0.527 & 0.098 & 0.008 & 0.094 & -0.025 \\ 0 & 0 & -0.196 & 0.265 & -0.102 & 0.066 \\ 0 & 0 & 0 & 0.217 & 0.548 & -0.634 \\ 0 & 0 & 0 & 0 & -0.454 & 0.207 \\ 0 & 0 & 0 & 0 & 0 & 0.244 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$	$\boldsymbol{\eta} = \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \\ \eta_6 \end{pmatrix}$
$\boldsymbol{\Gamma} = \begin{pmatrix} 0.707 & 0.673 & 0.193 & 0 & 0 & 0 \\ -0.087 & -0.036 & 0.761 & 0 & 0 & 0 \\ 0.384 & -0.141 & -0.567 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$ $\boldsymbol{\varphi} = \begin{pmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_3 \\ \varphi_4 \\ \varphi_5 \\ \varphi_6 \end{pmatrix}$	$\boldsymbol{\xi} = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \\ \xi_5 \\ \xi_6 \end{pmatrix}$

The structural model estimated with the class of the linear model as given in Equation 3.2 is equivalent to

$$
\begin{pmatrix}\eta_1\\ \eta_2\\ \eta_3\\ \eta_4\\ \eta_5\\ \eta_6\end{pmatrix} = \begin{pmatrix} 0 & 0.527 & 0.098 & 0.008 & 0.094 & -0.025\\ 0 & 0 & -0.196 & 0.265 & -0.102 & 0.066\\ 0 & 0 & 0 & 0.217 & 0.548 & -0.634\\ 0 & 0 & 0 & 0 & -0.454 & 0.207\\ 0 & 0 & 0 & 0 & 0 & 0.244\\ 0 & 0 & 0 & 0 & 0 & 0\end{pmatrix}\begin{pmatrix}\eta_1\\ \eta_2\\ \eta_3\\ \eta_4\\ \eta_5\\ \eta_6\end{pmatrix} +
$$

$$
\begin{pmatrix} 0.707 & 0.673 & 0.193 & 0 & 0 & 0\\ -0.087 & -0.036 & 0.761 & 0 & 0 & 0\\ 0.384 & -0.141 & -0.567 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\end{pmatrix}\begin{pmatrix}\xi_1\\ \xi_2\\ \xi_3\\ \xi_4\\ \xi_5\\ \xi_6\end{pmatrix} + \begin{pmatrix}\varphi_1\\ \varphi_2\\ \varphi_3\\ \varphi_4\\ \varphi_5\\ \varphi_6\end{pmatrix}.
$$

Having the latent scores for η_1,η_2, $\eta_{3,}$ η_4, η_5 and η_6, and ξ_1, ξ_2 and ξ_3, we can use the information from the model to compare the productivity level for all the identified components. Based on this information, Figure 3.2 shows that the primary crop production level was simultaneously controlled by the support of livestock (using manure) and the contributing factors. The SEMs obtained extracted more information about the food production than using a single linear model for instance, maize. In so doing, with latent scores, we were able to estimate a single linear equation by using ordinary least squared (OLS) estimation through η_1 as an endogenous variable. This procedure generates the equation η_1 = - 0.0479ξ_1 – 0.0182 ξ_2 + 0.404 ξ_3. As an illustration of the model, this suggested that η_1 was a linear function of ξ_1 , ξ_2 and ξ_3 and as a result, the component units can be ranked either based on η_1 or - 0.0479 ξ_1 – 0.0182 ξ_2 + 0.404 ξ_3.

As indicated earlier, the approach adopted by the SEM was based on the variance-covariate matrix between the variables in the data and the initial path diagrams that hypothesise the causal relationships among the variables. These path diagrams were later translated into a diverse set of linear equations describing the relationships that define a certain pattern when using the variance-covariance matrix.

The results were remarkably natural, as the correlations between latent (unobserved) variables and observed variables were found highly correlated (all above 0.80) and in a positive direction except Y (poultry) that was negatively strong (- 0.73) and Y representing sheep and goats ($-$ 0.57) that was an acceptable relationship. By contrast, the relationship between the latent (unknown) variables was positively weak but statistically significant.

Given these patterns, it indicates both a direct and indirect effect between exogenous and endogenous variables. The six endogenous variables derived from the diverse type of crop and kind of livestock mutually affect the three direct exogenous cause-factors: energy, labour, and fertilizer, as this is likely to confirm the supporting structure. Conversely, the energy used, labour and fertiliser factor types were likely to be correct with a high degree of confidence. These factors were key strategies to create a more abundant crop and rear healthy livestock.

3.7 CONCLUSION AND RECOMMENDATIONS

Latent variables were identified based on the effects of observable variables. Therefore, in applying SEM to food production data, the results have shown that the livestock's products and the crop varieties had a set of relationships when using the latent variable as a set of predictors. The results also indicated that factors such as energy, labour and fertiliser have contributed positively to the development of food production in SSA; therefore, application of new technology is likely to increase food production.

3.8 SIMULATION STUDY FOR TESTING STRUCTURAL EQUATION MODEL

3.8.1 Introduction

In the previous sections of this chapter, the SEM approach provides efficient results in evaluating the relations among variables and testing theoretical models. These variables can be directly or indirectly observed, and SEM enables the researcher to specify the set of causal relationship between variables in the model. In some conditions, the variability of a set of relationships points to inconsistent conclusions about the level at which a model is indeed equivalent to the observed data. In this thesis, we present an approach to evaluate the suitability of a given SEM by employing factor analysis through the PCA to select the variables upon which the relationships can be established. In literature, the estimation of the parameters when using SEM assumes of multivariate normality or generalised least square of robust estimators. In this study, the researcher presents the use of LISREL from Jöreson. In applying this approach, we can assess hypothesised relations between variables that are consistent with the observed relations among the variables. In the literature, researchers can start by conceiving how a set of variables is related to one another. The researcher defines a hypothesised model that is assumed to fit the data. A good model fits the data well, and therefore the results can then be interpreted. In some conditions, a model that is assumed to fit a specific latent variable may not fit the observed data; therefore, LISREL can be used. This is because LISREL provides the researcher with special quantitative estimates of the hypothesised model that fits the observed data. To demonstrate the mechanism of using LISREL, we intend to compare the results obtained from SEM through the LISREL approach to be confident of the technique used. The use of SEM has some restrictions since it specifies the latent variables involved in the analysis. There is a large, diverse set of relationships that can be developed among variables. There is a sub-set of the variability of relationships that can point to inconsistent conclusions about the level upon which a model is indeed equivalent to the observed data. This inconsistency may arise in applying factor analysis to test the component factor patterns of the loadings that reflect the performance of the relations between the original variables and the corresponding simulated component factors. In the next section, a simulation is examined to investigate whether there is consistency between the loadings of the rotated components and the simulated rotated components. If the simulation approach presents similar components to the loadings obtained from the SEM technique, then the proposed SEM approach can be used with confidence in particular research. The approach that will be used in the simulation analysis will have a form of multivariate normal distribution with the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$.

3.8.2 Concept of Simulation

Simulation is a technique that conducts tests on the computer by associating random sampling from a probability distribution. Any measurement resulting from a sampling approach includes some level of sampling error. Using simulation, we can examine the performance of the test statistics and the interval estimation procedure when the observed parameters are known. The technique of simulation requires running enough tests for the uncertain variables in the model by collecting results from the given outcomes to produce the summary statistics, for instance, point estimates, standard errors and covariance with charts and graphs. In applying simulation, we can understand the solution obtained for a better picture of the decisions made. To test the performance of the improved approach in the SEM, we are now proceeding to use simulation to calculate the loading through the factor analysis. These quantities will determine the consistency of the pattern of the model, and whether they remain the same or not.

The approach described in Table 3.10 has a form of an MVN distribution with the mean vector $\boldsymbol{\mu} = (\bar{y}_1, \bar{y}_2, .. \bar{y}_{25})^t$ and covariance matrix $\boldsymbol{\Sigma}$ as given below

$$\boldsymbol{\Sigma} = \begin{pmatrix} S_{11} & S_{12} & \cdots & S_{1n} \\ S_{21} & S_{22} & \cdots & S_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ S_{n1} & S_{n2} & \cdots & S_{nn} \end{pmatrix}.$$

The multivariate normal distribution can be noted, such as Model $\mathbf{Y}$ ~MVN $(\boldsymbol{\mu}; \boldsymbol{\Sigma})$.

In this study, simulation is used to test the effectiveness of the improved model proposed in Section 3.4.2. The application of this approach is used over an extensive range of data for the food production process through factor analysis techniques, simply because factor analysis helps us reduce a large number of variables by improving the understanding of the relationships among the variables in a single study. We present factor analysis to test the component factor patterns of the loadings that reflect the performance of the correlations between the original variables and the corresponding component factors. The operation of the simulation is illustrated by an example of a multivariate normal distribution with the parameters of a vector mean $\bar{\boldsymbol{y}}$ $(25 \times 1\; column\; vector)$ and the covariance matrix $\sum$ $(25 \times 25\; matrix)$. The approach is processed through a simulation of ten independent samples from the sample size of thirty-three.

The following section will use the example given in Section 3.2.6 to examine the consistency between the loadings of the rotated components as per Table 3.10 and the simulated rotated components as shown in Table 3.9. This example has generated two parameters ($\bar{\boldsymbol{y}}$ and $\sum$) that have been used in the MVN distribution with a sample size of thirty-three using 20000 iterations to generate the factor that has been analysed. Therefore, sixteen independent samples are generated in the MVN distribution, but the size of this experiment has become very demanding to compute the rotated components, as this requires far more computer capacity than what is available on the computer. In such a situation, to reduce the memory requirement, an assessment of the run of the loadings are made based on “Factor 1” of the simulation specification factors. However, the first factor resulting from the 20000 iterations, Factor 1, by the improved method and Factor 1 by simulation of the sixteen independent samples, are provided in Table 3.9. The total proportion of variance accounted for is given in Table 3.9. These substantial proportions of variance are of particular interest simply because they indicate the degree to which the relative performance of the two approaches varied with Factor 1 by using the R software.

Table 3.9: The rotated component of Factor 1 using simulation compared to the true Factor 1

No	True value	Simulation of factor components								
		S1	S2	S3	…	S8	S9	…	S16	AS
1	0.100	0.019	-0.110	0.350	…	-0.222	-0.220	…	-0.218	0.058
2	-.050	-0.003	-0.220	0.085	…	-0.038	-0.036	…	-0.304	-0.019
3	-.017	-0.158	-0.028	0.118	…	0.207	-0.206	…	-0.204	0.258
4	-.072	0.397	-0.267	0.156	…	-0.102	-0.100	…	-0.100	-008
5	0.033	-0.216	-0.010	-0.017	…	-0.405	-0.402	…	-0.400	0.276
6	0.074	-0.207	0.047	-0.056	...	-0.344	-0.342	…	-0.341	0.250
7	**0.804**	**0.906**	**0.882**	**0.749**	**…**	**0.765**	**0.761**	**…**	**0.760**	**0.598**
8	0.023	-0.294	-0.062	-0.064	…	-0.426	-0.424	…	-0.422	0.240
9	0.382	0.555	0.289	0.728	…	0.021	0.020	…	0.019	0.532
10	0.041	0.070	-0.054	0.324	…	-0.214	-0.213	…	-0.210	0.284
11	**0.707**	**0.684**	**0.618**	**0.690**	**…**	**0.520**	**0.519**	**…**	**0.518**	**0.517**
12	**0.659**	**0.818**	**0.648**	**0.903**	**…**	**0.630**	**0.629**	**…**	**0.628**	**0.649**
13	**0.775**	**0.759**	**0.583**	**0.723**	**…**	**0.478**	**0.476**	**…**	**0.474**	**0.510**
14	-.131	-0.122	0.238	-0.184	…	0.141	0.140	…	0.138	-0.084
15	-.111	-0.153	-0.120	-0.381	…	-0.390	-0.389	…	-0.388	-0.142
16	-.011	-0.121	-0.100	-0.057	…	-0.274	0.272	…	-0.270	-0.076
17	0.031	0.204	0.197	0.277	…	0.113	0.112	…	0.110	0.118
18	0.221	0.340	0.318	0.475	…	0.473	0.472	…	0.470	0.224
19	0.335	0.761	0.884	0.637	…	0.994	0.991	…	0.990	0.445
20	0.409	0.764	0.949	0.682	…	0.953	0.952	…	0.950	0.493
21	-.016	0.730	0.723	0.479	…	0.904	0.902	…	0.900	0.314
22	0.120	0.078	0.366	-0.094	…	0.028	0.026	…	0.024	0.038
23	0.120	0.052	-0.083	0.459	…	0.379	0.377	…	0.376	0.024
24	0.128	-0.147	0.149	-0.581	…	-0.052	-0.050	…	-0.051	-0.020
25	**-.565**	-0.053	-0.145	-0.014	…	-0.269	-0.268	…	-0.267	-0.104

S1 = Sample 1, S2 = Sample 2, S3 = Sample 3, S16 = Sample 16, AS = Average of samples

Table 3.9 displays the high loadings per component factor, representing the form of the factor pattern for consistency. A correlation of the true value above 0.5 is regarded as important; therefore, we indicated them in bold for easy visibility.

Table 3.10: The proportion variance

True value	Simulation in factor analysis								
	S1	S2	S3	…	S8	S9	…	S15	S16
0.190	0.206	0.190	0.213	…	0.225	0.224	…	0.220	0.204

S1 = Sample 1, S2 = Sample 2, S8 = Sample 8, S9 = Sample 9, S16 = Sample 16

The form of the factor pattern for consistency is specified in terms of high loadings per component factor in Table 3.9. In this study, a correlation of the true value above 0.5 is regarded as important, therefore for easy visibility we have put all the values above 0.5 in bold. The findings reveal an accurate consistency of the high and low loadings. That is, the simulation procedure fits the pattern directly to the true factor components. The 25^{th} variable is a poorer fit that has no similar patterns to the true loadings of Factor 1. This might be caused by an unexpected error in the calculation or from the dataset.

3.8.3 Mean Squared Error

Statistical procedures are usually associated with some type of errors that might arise during data collection or by the time of the calculation of the results. To quantify these errors requires applying the method of mean squared error (MSE). This technique will enable us to measure the performance of an estimator simply because it provides precision, bias, and accuracy during the statistical assessment. MSE needs an estimator, which is a function of a given data. Suppose $\widehat{\omega}$ represents the estimator, the MSE will then measure the average of the squares of the errors, that is, the average squared difference between the observed and the predicted values.

$MSE(\hat{\varphi}) = E((\hat{\omega} - \omega)^2)$ alternatively,

$MSE = \frac{1}{n}\sum(x_i - \hat{x}_i)^2$.

where n represents the number of the data, x_i represents the observed values and $\hat{x}_i$ the predicted values. MSE is a value that provides information about the goodness of fit of the model that is used, it reveals how close a model such as a regression line is to a set of points. In addition, MSE enables us to compare two or more statistical models. The smaller the value of MSE, the better the fit, simply because it minimises the variance, and a larger value of MSE indicates that the data values are scattered largely around the mean value. For instance, when MSE is zero, it means that there is a perfect precision found between the estimator $\hat{\omega}$ and the parameter ω. The simulation will help us to calculate MSE and to construct the confidence intervals when the observed data are given. Using simulation, we can examine the performance of the test statistic and the interval estimation procedure when the observed parameters are known.

Table 3.11: Mean squared values

No	MSE values for simulation components: factor 1 for the eight generated samples								Average sample
	1	2	3	…	8	9	…	16	
1	0.003	0.002	0.003	…	0.005	0.002	…	0.001	0.000
2	0.000	0.001	0.001	…	0.003	0.000	…	0.000	0.000
3	0.000	0.000	0.001	…	0.001	0.000	…	0.001	0.003
4	0.010	0.002	0.002	…	0.000	0.000	…	0.000	0.000
5	0.003	0.000	0.000	…	0.008	0.004	…	0.002	0.003
6	0.003	0.000	0.001	…	0.008	0.006	…	0.004	0.001
7	0.000	0.000	0.000	…	0.000	0.001	…	0.000	0.000
8	0.004	0.000	0.000	…	0.009	0.002	…	0.000	0.002
9	0.001	0.000	0.005	…	0.006	0.000	…	0.000	0.001
10	0.000	0.000	0.003	…	0.016	0.004	…	0.001	0.002
11	0.000	0.000	0.000	…	0.002	0.000	…	0.000	0.002
12	0.001	0.002	0.003	…	0.000	0.000	…	0.000	0.000

Table 3.11 (Continued)

No	MSE values for simulation components: factor 1 for the eight generated samples								Average Sample
	1	2	3	…	8	9	…	16	
13	0.000	0.006	0.000	…	0.004	0.002	…	0.000	0.003
14	0.000	0.000	0.000	…	0.003	0.000	…	0.000	0.000
15	0.000	0.000	0.003	…	0.003	0.000	…	0.000	0.000
16	0.001	0.001	0.000	…	0.003	0.001	…	0.000	0.000
17	0.001	0.000	0.003	…	0.000	0.000	…	0.000	0.000
18	0.001	0.013	0.003	…	0.003	0.002	…	0.000	0.000
19	0.008	0.013	0.004	…	0.009	0.004	…	0.002	0.000
20	0.005	0.024	0.003	…	0.013	0.011	…	0.010	0.000
21	0.024	0.003	0.011	…	0.000	0.000	…	0.000	0.005
22	0.000	0.002	0.002	…	0.000	0.000	…	0.000	0.000
23	0.000	0.008	0.005	…	0.000	0.000	…	0.000	0.000
24	0.003	0.000	0.022	…	0.001	0.001	…	0.001	0.000
25	0.011	0.008	0.013	…	0.004	0.001	…	0.001	0.009

The findings of Table 3.11 range from 0.00 to 0.05, indicating that the model fits the data well. In other words, there is a perfect precision simply because the data values are all closely around the mean value.

3.8.4 Discussion

This study tested the improved SEM approach using factor analysis through the PCA with the simulation factor analysis under the same conditions of a dataset for a multivariate normal distribution. The two approaches were able to generate the same pattern of loadings accurately. Overall, this was true for sixteen independent samples even though one value (sheep and goats) differed. The sample size, the number of variables used, and the loadings range did not significantly impact the accuracy factor pattern. The simulation approach performed similarly to the loadings; therefore, the proposed improved approach presents a suitable and reliable methodology. Thus, the proposed approach can be used with confidence in agricultural and related fields of research. The simulation R program is given in the appendix at the end of the thesis.

Although the proposed model can improve better than the traditional approach, we note this can only be valid in the following conditions:

1. The model can perform better under the assumption that the parameter estimates calculated are normally distributed and the sample size is 25 or more.
2. When the parameter estimates are either not normally distributed or when the sample size is small, this model will not perform well unless other robust methods be used.
3. When the sample size has increased from 8 to 16, the findings revealed an accurate consistency of the high and low loadings. That is, the simulation procedure fits the pattern directly to the true factor components.

CHAPTER 4

INTEGRATING META-ANALYSIS INTO STRUCTURAL EQUATION MODELLING

4.1 INTRODUCTION

MA and SEM are the two statistical techniques most used in medical sciences (Sutton et al., 2000), psychology (Cooper, 2010) and education (Hedges & Olkin, 1985), more than in the agricultural field. Its benefit is that we can combine estimated parameters that strengthen each other in a single model. Many research studies have produced important contributions by combining the two techniques in an application, such as in the world values survey, by testing a theory on how job control predicts job satisfaction at the cultural level (Cheung, 2013). We construct a systematic approach that provides guidelines on how a MA model can be integrated into structural equation modelling. In Section 4.3, we present a methodological approach of MA. The methodology of SEM is described in Section 4.4. In Section 4.5, we show how to integrate MA into SEM since the new approach offers additional research opportunities in terms of methodological development in both MA and SEM techniques. MA synthesises the research findings, while SEM is a multivariate technique that tests hypothetical models. In Section 4.6, the advantages and disadvantages of the combined approach are described. An illustrative example is presented in Section 4.7, and the conclusion is given in Section 4.8.

Traditionally, MA and SEM are two different techniques by nature, based on their assumptions, models, respective methodological approaches, and how one summarises the results. In statistical theory, the two techniques are multivariate techniques that use correlations or covariance matrices as inputs. The productivity in the results obtained from MA in many domains had motivated researchers to use MA and SEM in testing causal models (Becker & Scram, 1994; Viswesvaran &

Ones, 1995; Cheung & Chan, 2005). Many research studies have produced important contributions by combining MA and SEM, for instance in analysing longitudinal data (Ployhart & Ward, 2011). Other examples include the application of polynomial regression (Shanock et al., 2010) and in the state of the World Values Survey that tested a theory on how job control can predict job satisfaction at the cultural level (Cheung, 2013). We intend to introduce this technique in the agricultural research field as it is applied in other disciples.

The food production process is a combination of activities developed separately, but the final activity adds together all the separate partial processes. The methodology to address such cases in the agricultural sector has not yet been sufficiently developed as it is done across various disciplines. This study proposes to synthesise correlation matrices in an adaptable methodology that allows for the incorporation of MA into SEM enables us to combine the data from past independent studies by drawing the overall conclusions while SEM examines and tests the relationships all together in the model. The data in MA are the effect sizes. Integrating MA into SEM aims to present a methodological approach that involves testing an extensive data set of the parameters in SEM. Guidelines will enable researchers to combine correlations from independent studies by using fixed and random effects to test homogeneity of correlation to fit a SEM. In this study, we show how to test the hypothesis against the model's parameters. An agricultural research example is used to illustrate the procedure. The variables are standardised to have a common scale across the studies. In the next section, we briefly present the methodology concepts of MA, SEM, and the integration of MA into SEM. The illustrated example in the agricultural field is given in Section 4.7.

4.2 OBJECTIVES

The principal objectives are as follows

1 To introduce MA and SEM into the agricultural field as a new area accessible to applied research.

2 To synthesise all studies that use MA and to model for the suitability of the SEM into a single combined model; and

3. To provide a methodological approach with mathematical integration of MA into SEM by providing guidelines on how to test the hypotheses of homogeneity against heterogeneity.

4.3 METHODOLOGY OF META-ANALYSIS

Using MA, we integrate the findings from past independent studies. The input in MA is called effect size statistics that can be the standardised in respect of the mean difference (denoted as SMD), the correlation coefficient or the odds ratio. The methodology in MA is identical as given in Section 2.3.

Through the results obtained in the statistical test, we draw meaningful inferences, but the challenge is when the sample of past studies is too small, it becomes difficult to make conclusions about the population, as Jak (2015) reported that we need at least two studies to be included in MA that summarise the research question quantitatively. In addition, Landis (2013) indicated that the more studies collected, the bigger the total sample size, usually, the more confidence we have in our estimate of the actual correlation between the variables of interest.

Individual studies can be combined in either the fixed-effects or random-effects model. The fixed-effects model is represented by the equation $\boldsymbol{\eta} = \boldsymbol{X\xi} + \boldsymbol{\delta}$, and the assumption is that the population of effect sizes remains the same for all studies. That is, the effects are homogeneous across studies. In the random-effects models, which are represented by the equation $\boldsymbol{\eta} = \boldsymbol{X\xi} + \boldsymbol{U} + \boldsymbol{\delta}$, the assumption is that the population of effect sizes vary from one study to the next. The difference between the two models is in calculating the standard errors associated with the combined effect sizes.

4.4 METHODOLOGY OF STRUCTURAL EQUATION MODELLING

Structural equation modelling estimates structural relations between latent variables generated from observed variables such as structural relations that are identified through a system of simultaneous linear equations (Olmos et al., 2018). This technique establishes the relationships between many variables, and therefore it requires constructing the model(s) of interest before the analysis of the data (Hoyle, 2012). Jöreskog (1973) coined the phrase LISREL to describe the technique. Hence, SEM is expressed based on a theory that must be tested to determine whether the model constructed can be rejected or not. A good model fits the data well, and the results obtained can then be interpreted. The inputs in SEM are the covariance or correlation matrices. The methodology used in SEM is as given in Section 3.4.

4.5 INTEGRATING META-ANALYSIS INTO STRUCTURAL EQUATION MODELLING

4.5.1 Materials and Methods from the Current Approach

The current approach uses the same steps as given in Section 2.1. Some of the challenges from the current approach are in terms of synthesising past studies that are not suitable due to the eligibility criteria in the selection of the inclusion of the studies. Jak and Cheung (2018) reported that if the number of studies became too small, the reliability of the results inevitably became a problem.

We need to improve the approach by developing new objectives because individual works may provide unbiased, perfect, and reliable results. Synthesising criteria of processes by combining past studies is not suitable and flexible because SEM is a multivariate technique that fits, and tests hypothesised models. The steps for the current approach are:

Step 1: Apply a clear sense of the hypothesis of interest by anticipating the relationship between MA and SEM (Viswesvaran & Ones, 1995).

Step 2: Include all studies with existing results related and any associated interest.

Step 3: Synthesise correlation matrices.

Step 4: Fit SEM.

Step 5: Apply MA on a series of correlation matrices to create a pooled correlation matrix.

Step 6: Analyse the pooled correlation matrix obtained in Step 3 into SEM.

Step 7: Make use of available software that performs the integration of MA into SEM, such as MASEM (Viswesvaran and Ones, 1995; Colquitt et al., 2000) or TTSEM (Cheung & Chan, 2005; Cheung, 2009).

4.5.2 Materials and Methods from the Improved Approach

The inputs of MA are the effect sizes that can be the standard error, sample size, covariance or correlations. In this study, we propose to get all possible effect sizes collected from the past studies. The methodology of the newer approach facilitates the integration of the MA into SEM by proposing a suitable way of testing complex theories involving multiple variables that cannot be measured. The steps are as follows:

Step 1: Set clear eligible criteria that enable a complete literature search in locating all relevant past studies.

Set 2: Calculate the effect sizes from the past studies.

Step 3: Use factor analysis on the data to identify the most important variables for the integration. PCA enable us to reduce the number of dimensions without much loss of information.

Step 4: Calculate the variance-covariance matrix from the standardised data because using PCA on the standardised data is equivalent to PCA using the correlation matrix.

Step 5: Rotate components with the reduced factors that have the high loading on the factor; in so doing, the interpretation of the factors becomes simple.

Step 6: Conduct a test of homogeneity effects of the correlation matrices.

Step 6.1: If the test is not statistically significant, then we proceed by combining the correlations matrices to form a pooled correlation matrix.

Step 6.2: If the test is statistically significant, then we use the random-effects model to explain the variations across the studies.

Step 7: Determine the average of the correlation if Step 6.1 has been performed by employing the pooled correlations to fit the SEMs. The diagonals of the linear model indicate that the correlation matrix is one. The benefit of integrating MA into SEM is to account simultaneously for all effects of factors in a single model.

4.5.3 Test for Structural Equation Modelling: Fitting SEM

Suppose two factors are used in the process of integrating MA into SEM. Factor A comprises five indicators for agricultural food production products such as “the number of heads of animals”, “the quantity of crops used in kg/ha”; “soil texture”, “fertiliser-use measured in kg/ha”, and “the impact of agroforestry adoption of the livelihood of farmers”. Factor B represents the five regions in the SSA that are classified based on their ecological zones. The five regions are “Eastern Africa”, “Middle Africa”, “Northern Africa”, “Southern Africa” and “Western Africa as given in Table 4.1. In this study, factors A and B enable us to integrate MA into SEM. The SEM approach is used based on the guideline of the methodology provided in the previous sections. In applying the approach, we employ all the data sets using the factor analysis through the PCA. This process enables us to use the most important variables in which the path diagram displays the suitable linear relations between exogenous and endogenous variables. The results provide a set of relations of four exogenous factor models ($\xi_1, \xi_2, \xi_3, \xi_4$) such that: Eight indicators ($X_2, X_4, X_5, X_7, X_8, X_{10}, X_{11}, X_{15}$) are related to Factor 1, five indicators ($X_1, X_3, X_9, X_{12}, X_{14}$) are linked to Factor 2, 0ne indicator X_6 are associated with Factor 3 and one indicator X_{13} is connected to Factor 4. From the factor B side of models there is one factor $\eta_1$upon which the indicators Y_1, Y_2, Y_3, Y_4, Y_5 are linked to Factor 1, as shown in Figure 4.1

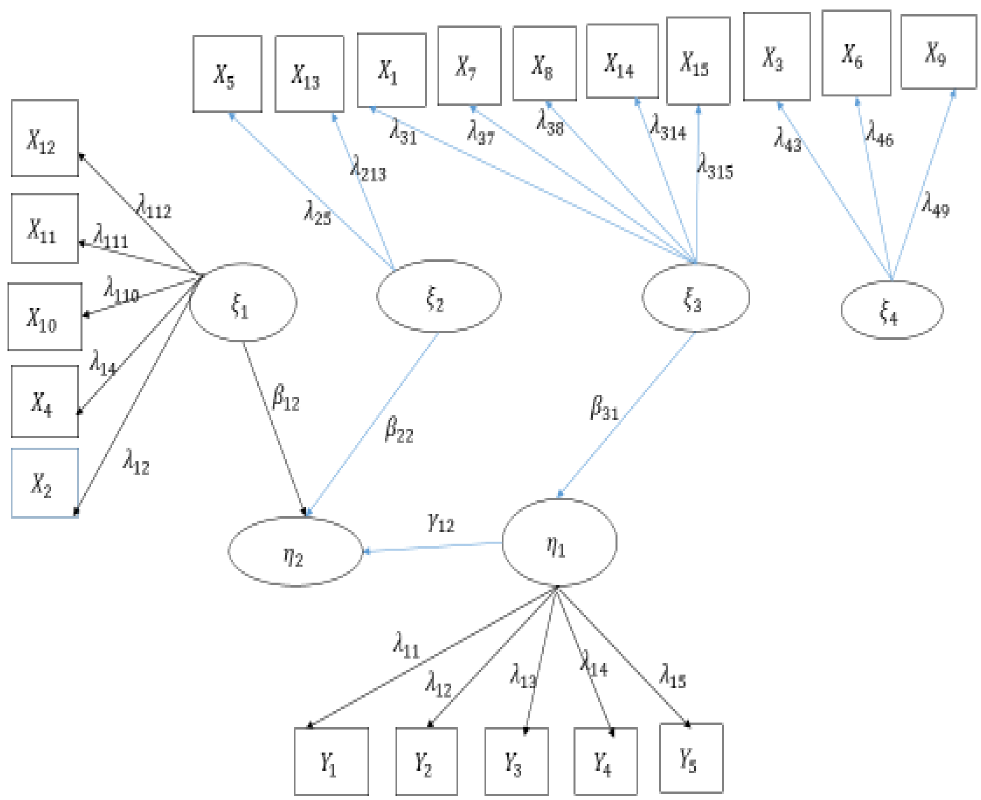


Figure 4.1: Path Diagram, Structural Equation Model

X_1: Goat, X_2: Sheep, X_3: Pig, X_4: Maize, X_5: Groundnut, X_6: Pea, X_7: Sand, X_8: Loan, X_9: Clay, X_{10}: Nitrogen, X_{11}: Phosphorus, X_{12}: Potassium, X_{13}: Increase.Income, X_{14}: Increase.Savings, X_{15}: Increase.Agro, Y_1: represents Eastern Africa, Y_2: Middle Africa, Y_3: Northern Africa, Y_4: Southern Africa and Y_5: Western Africa.

The fifteen observed variables ($X_1, X_2, ..., X_{15}$) are given in squares, and their values regressed on the exogenous variables. The values of the five categorical regions ($Y_1, Y_2, ..., Y_5$) are regressed on the endogenous variable η_1. In a latent factor model, (ξ_1, ξ_2, ξ_3 and ξ_4) are linked to the latent indicator factor (η_1, η_2, and η_3) and by the estimated parameters denoted by (β_{12}, β_{21}, $\beta_{31,}\beta_{32,}$ and β_{41}). In addition, the latent factors are connected to the original variables. The distribution of latent factors is explained in terms of the expected value and the covariance of the residual of δ.

Suppose we define a series of correlations denoted η_k for k = 1, 2, 3... N. For each independent study η_k, the covariance matrix $\boldsymbol{V}$ is known, with $n \times n$ being a matrix with non-zero off-diagonal elements, so that the observations remain correlated. The regression model to estimate the parameters is

$$\boldsymbol{\eta} = \boldsymbol{X\xi} + \boldsymbol{\delta} \tag{4.1}$$

assuming that $E(\boldsymbol{\delta})$ = 0 and the covariance $V(\boldsymbol{\delta}) = \sigma^2\boldsymbol{I}$.

$\boldsymbol{X}$ is an $n \times n$ matrix of explanatory variables, which is an identity matrix, $\boldsymbol{\eta}$ is a vector of the observed effect sizes (correlations from the sample), $\boldsymbol{\xi}$ is a vector of the specific effect sizes (correlations from the population), and $\boldsymbol{\delta}$ is a vector of sampling error with $\boldsymbol{\delta_i}$ ~$N(\boldsymbol{0}, \boldsymbol{V_i})$. $\boldsymbol{\eta}$ is distributed with a multivariate normal distribution with a mean equal to $\boldsymbol{\xi}$ and the covariance equal to $\boldsymbol{V}$. Alternatively, we can write $\boldsymbol{\eta}$ ~MVN ($\boldsymbol{\xi}, \boldsymbol{V}$). $\boldsymbol{\xi}$ and $\boldsymbol{V}$ need to be estimated. In this study, the estimation of the weighted correlation is done through the approach of the generalised least squares (GLS) test to determine the summary of the correlation matrix $\hat{\boldsymbol{\xi}}$, with variance denoted by Var $(\hat{\boldsymbol{\xi}}$). The least square's function is

$$(\boldsymbol{\eta} - \boldsymbol{X\xi})^T \, \boldsymbol{V}^{-1}(\boldsymbol{\eta} - \boldsymbol{X\xi}) \tag{4.2}$$

The partial derivative with respect to $\boldsymbol{\xi}$ is $(\boldsymbol{X^T V^{-1} X})\boldsymbol{\xi}$ = $\boldsymbol{X^T V^{-1} \eta}$ and the estimator $\hat{\boldsymbol{\xi}}$ is

$$\hat{\boldsymbol{\xi}} = (\boldsymbol{X^T V^{-1} X})^{-1} \boldsymbol{X^T V^{-1} \eta} \ , \tag{4.3}$$

$$\text{where } E(\hat{\boldsymbol{\xi}}) = E((\boldsymbol{X^T V^{-1} X})^{-1} \boldsymbol{X^T V^{-1} \eta}) = \boldsymbol{\xi} \tag{4.4}$$

$$Var(\hat{\boldsymbol{\xi}}) = E((\boldsymbol{X^T V^{-1} X})^{-1} \boldsymbol{X^T V^{-1} \eta}) = \sigma^2 (\boldsymbol{X^T V^{-1} X})^{-1}. \tag{4.5}$$

The covariance **V** is equal to $(\boldsymbol{X^T V^{-1} X})^{-1}$,

where $\boldsymbol{V} = \begin{pmatrix} S_{11} & 0 & 0 & \dots & 0 \\ 0 & S_{22} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & S_{nn} \end{pmatrix}$, $\quad \boldsymbol{\eta} = \begin{pmatrix} \eta_{11} \\ \eta_{21} \\ \vdots \\ \eta_{n1} \end{pmatrix}$, and

$$\boldsymbol{X} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix},$$

The off-diagonal terms $\boldsymbol{V_{ij}}$ are the conditional sampling covariance matrix of effect sizes, therefore,

$\boldsymbol{V} = Diag(S_{11}, S_{22}, \dots, S_{nn})$.

4.5.4 Test for Fixed Effects Model

Suppose that from the past studies we obtain the correlations $\eta_1, \eta_2, \dots, \eta_n$ and the sample variances $S_1, S_2, \dots, S_n$ that form the elements of $\boldsymbol{V}$. When the past studies are independent under the generalised least squares (GLS) approach, then $\boldsymbol{V}$ = diag $(S_{11}, S_{22} \ \dots, S_{nn})$. The test statistic Q for homogeneity of all effects sizes across all the studies is

$$Q = (\hat{\boldsymbol{\eta}} - \boldsymbol{X^T} \hat{\boldsymbol{\xi}}) \boldsymbol{V^{-1}} (\hat{\boldsymbol{\eta}} - \boldsymbol{X^T} \hat{\boldsymbol{\xi}}) \tag{4.6}$$

where Q follows a Chi-squared distribution with $(n-1)$ degrees of freedom.

The hypotheses to test for heterogeneity are

H_0 : Homogeneity effect is identified versus H_1: Heterogeneity effect is identified.

If the null hypothesis is rejected, then we proceed with the random effects. We will begin by recalculating the summary of the effects and the variance, including between studies. The model for the fixed effects is given in Equation 4.1: $\boldsymbol{\eta} = \boldsymbol{X\xi} + \boldsymbol{\delta}$

where $E(\boldsymbol{\delta}) = 0$ and $Var(\boldsymbol{\delta}) = {\sigma_n}^2\mathbf{I}$.

The estimated variance is given by

$$E(Q) = (n-1) + {\sigma_n}^2[tr\ (\boldsymbol{V}^{-1}) - tr(\boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1} \times \boldsymbol{X}^T\boldsymbol{V}^{-1})]$$

$$Q = (n-1) + {\sigma_n}^2[tr\ (\boldsymbol{V}^{-1}) - tr(\boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1} \times \boldsymbol{X}^T\boldsymbol{V}^{-1})]$$

$$Q - (n-1) = {\sigma_n}^2[tr\ (\boldsymbol{V}^{-1}) - tr(\boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1} \times \boldsymbol{X}^T\boldsymbol{V}^{-1})]$$

$${\sigma_n}^2 = \frac{(Q-(n-1))}{tr\ (\boldsymbol{V}^{-1}) - tr(\boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1} \times \boldsymbol{X}^T\boldsymbol{V}^{-1})} \qquad (4.7)$$

4.5.5 Test for Random Effects Model

The random-effects model to test the heterogeneity between the studies is similar to the fixed-effects model given in Equation 4.1, with the same assumptions but together with an explained portion $\boldsymbol{U}$ of the random effects. Therefore, the model for random effects is defined by

$$\boldsymbol{\eta} = \boldsymbol{X\xi} + \boldsymbol{U} + \boldsymbol{\delta}. \qquad (4.8)$$

Similarly, for the random-effects model, we need to recalculate the estimated parameter $\boldsymbol{\xi}$, the expected value $\boldsymbol{\xi}$ and the variance of $\boldsymbol{\xi}$ in the random effects under the model of Equation 4.8, where the covariance of the random effects of the unexplained $\boldsymbol{U}$ is Cov ($\boldsymbol{U_{ij}}$) = D (Laird & Ware, 1982; Jennrich & Schluchter, 1986). Suppose that $\boldsymbol{U} + \boldsymbol{\delta} = \boldsymbol{\tau}$, then the random model of Equation 4.8 can be written as $\boldsymbol{\eta} = \boldsymbol{X\xi} + \boldsymbol{\tau}$, where Cov ($\boldsymbol{\tau}$) = Cov (U+ $\boldsymbol{\delta}$) $= D + {\sigma_n}^2\boldsymbol{I}$. Similarly, the test statistic Q

for heterogeneity can be shown by replacing $\boldsymbol{V}$ from the fixed effects model with $D + {\sigma_n}^2\mathbf{I}$, as given in Equation 4.9. The test statistic Q becomes

$$Q = \left(\widehat{\boldsymbol{\eta}} - \boldsymbol{X}^{\boldsymbol{T}}\widehat{\boldsymbol{\xi}}\right)(\boldsymbol{S} + {\sigma_n}^2\mathbf{I}\,)^{-1}\left(\widehat{\boldsymbol{\eta}} - \boldsymbol{X}^{T}\widehat{\boldsymbol{\xi}}\right) \tag{4.9}$$

The partial derivative with respect to $\boldsymbol{\xi}$ is $(\boldsymbol{X}^{T}(\boldsymbol{S} + {\sigma_n}^2\mathrm{I}\,)^{-1}\boldsymbol{X})\boldsymbol{\xi} = \boldsymbol{X}(\boldsymbol{S} + {\sigma_n}^2\mathbf{I}\,)^{-1}\boldsymbol{\eta}$ and the estimator $\widehat{\boldsymbol{\xi}}$ is

$$\hat{\xi}=(\boldsymbol{X}^{T}(\boldsymbol{S} + {\sigma_n}^2\mathbf{I}\,)^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{T}(\boldsymbol{S} + {\sigma_n}^2\mathbf{I}\,)^{-1}\boldsymbol{\eta} \tag{4.10}$$

$$\text{where } E\left(\widehat{\boldsymbol{\xi}}\right)=E((\boldsymbol{X}^{T}(\boldsymbol{S} + {\sigma_n}^2\mathbf{I}\,)^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\boldsymbol{T}}(\boldsymbol{S} + {\sigma_n}^2\mathrm{I}\,)^{-1}\boldsymbol{\eta}) = \xi \tag{4.11}$$

$$Var\left(\widehat{\boldsymbol{\xi}}\right)=E((\boldsymbol{X}^{\boldsymbol{T}}(\boldsymbol{S} + {\sigma_n}^2\mathrm{I})^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{T}(\boldsymbol{S} + {\sigma_n}^2\mathrm{I}\,)^{-1}\boldsymbol{\eta})= \sigma^2(\boldsymbol{X}^{T}(\boldsymbol{S} + {\sigma_n}^2\mathbf{I}\,)^{-1}\boldsymbol{X})^{-1} \tag{4.12}$$

$$E(Q) = (n-1) + {\sigma_n}^2[tr\;((\boldsymbol{S} + {\sigma_n}^2\mathbf{I}\,)^{-1}) - tr((\boldsymbol{S} + {\sigma_n}^2\mathbf{I}\,)^{-1}\boldsymbol{X}(\boldsymbol{X}^{T}(\boldsymbol{S} + {\sigma_n}^2\mathbf{I}\,)^{-1}\boldsymbol{X})^{-1} \times \boldsymbol{X}^{T}(\boldsymbol{S} + {\sigma_n}^2\mathbf{I}\,)^{-1})]$$

$$Q = (n-1) + {\sigma_n}^2[tr\;((\boldsymbol{S} + {\sigma_n}^2\mathbf{I}\,)^{-1}) - tr((\boldsymbol{S} + {\sigma_n}^2\mathbf{I}\,)^{-1}\boldsymbol{X}(\boldsymbol{X}^{T}(\boldsymbol{S} + {\sigma_n}^2\mathbf{I}\,)^{-1}\boldsymbol{X})^{-1} \times \boldsymbol{X}^{T}(\boldsymbol{S} + {\sigma_n}^2\mathbf{I}\,)^{-1})]$$

$$Q - (n-1) = {\sigma_n}^2[tr\;((\boldsymbol{S} + {\sigma_n}^2\mathbf{I}\,)^{-1}) - tr((\boldsymbol{S} + {\sigma_n}^2\mathbf{I}\,)^{-1}\boldsymbol{X}(\boldsymbol{X}^{T}(\boldsymbol{S} + {\sigma_n}^2\mathbf{I}\,)^{-1}\boldsymbol{X})^{-1} \times \boldsymbol{X}^{T}(\boldsymbol{S} + {\sigma_n}^2\mathbf{I}\,)^{-1})]$$

$${\sigma_n}^2= \frac{\left(Q-(n-1)\right)}{tr\;((\boldsymbol{S}+{\sigma_n}^2\mathbf{I}\,)^{-1}) - tr((\boldsymbol{S}+{\sigma_n}^2\mathbf{I}\,)^{-1}\boldsymbol{X}(\boldsymbol{X}^{T}(\boldsymbol{S}+{\sigma_n}^2\mathbf{I}\,)^{-1}\boldsymbol{X})^{-1} \times \boldsymbol{X}^{T}(\boldsymbol{S}+{\sigma_n}^2\mathbf{I}\,)^{-1})} \tag{4.13}$$

These estimates are likely to improve the results. If the heterogeneity is detected, the next step is to quantify the dimension of heterogeneity.

The size of heterogeneity: If the heterogeneity of the correlation coefficients is identified and significant, then the size of the heterogeneity is calculated by I^2 as proposed by Higgins and Thompson (2002), as demonstrated below

$$I^2 = \frac{U^2}{U^2+\vartheta} = \frac{Q-(k-1)}{Q} \tag{4.14}$$

where Q is the overall heterogeneity calculated based on Equations 4.6 and 4.9 for the fixed and random effect-models respectively. If Q is less than (k – 1), that means $I^2 = 0$. When the calculated I^2 values are approximately equal to 0.25, 0.50 and 0.75; then this is equivalent to a low, medium, and high level of heterogeneity (Higgins et al., 2003). These numbers help the researcher to evaluate the level of heterogeneity. A higher value indicates that heterogeneity hold therefore, the random-effects models can be used, and the researcher can further perform the meta-regression model. The meta-regression approach explores potential heterogeneity, resulting in the changes that might occur in practice over time and evaluate the effect of the variables of interests. A lower number indicates less heterogeneity, therefore, more consistency across the included studies.

4.6 ADVANTAGES AND DISADVANTAGES OF THE COMBINED MODEL

In this section, we present some of the advantages and disadvantages of using the combined model.

4.6.1 The Advantages of the Combined Model

The major benefits are especially:

- Single models can help us predict, but in case these models are consistently unable to predict the risk accurately or when the models do not fit the data well. In such a condition, the combined model is used to improve the overall precision of the prediction.
- In selecting the model, suppose that there is uncertainty on the choice of the model to use. Under such a condition, a combined approach might be the desired approach to use, among others. For instance, in botanical epidemiology, there is usually significant uncertainty in the selected models for the analysis (Shah et al., 2014 and Xu et al., 2014).
- When no single model is strongly supported by the data, it becomes difficult to choose one model over another. The combined model provides a means to address the research question. Parameter estimates obtained by the combined model are more robust than the single model

simply because the combined approach reduces both the model selection bias and the model selection uncertainty (Yang, 2003; Yuan & Yang; 2005; Hansen, 2007).

- The combined model increases the predictive power, especially for cases that have a sizeable residual discrepancy; therefore, the number of predictor variables in the models will be small.
- In a real-life situation, the researcher might not have enough information about the variables of interest. In such a condition, the single model might be inappropriate to estimate the parameter or make predictions. Instead, the combined model will perform better than the single models. For instance, agricultural research is concerned with techniques that can describe the observed data to test hypotheses and predict future situations. In other words, if the observed data fails significantly to support any single model, the combined model can be used to estimate model parameters and make a prediction. For example, in agriculture, most of the research problems involve many variables, and some might be correlated. Therefore, the data collected under such conditions require multiple approaches simply because a way to build a best single model with correct number of variables is usually problematic.

4.6.2 The Disadvantages of the Combined Model

The major disadvantages among others are especially:

- In real life, we usually use the models of which we have clear knowledge, but when we combine them, sometimes the combined model will have less reusability since the researcher might have too much unclear knowledge about the combined model, more than the application of the single model.
- Statistical techniques might be improperly used by researchers. That is, the single models are inappropriately used in the analysis; therefore, using the combined approach would be a risk as taking decisions based on incorrect results might cause significant mistakes.

- Subconsciously, the combined model can allow one technique to influence or distort the other model simply because the information may not be well evaluated, or it may be the case of the wrong choice of a model. Hence, the combined model must be used with precaution.

4.7 DATA ANALYSIS

The hypotheses that test how to integrate MA into SEM are now presented. The approach uses factor analysis through the PCA to create essential variables for a valid correlation matrix. The loadings allow us to test the null hypothesis of equality of factors under the fixed- and random-effects model. The data were extracted from each eligible article based on the inclusion criteria in the SSA countries. Published reports on “the crops production measured in kg/ha”, “life heads of animal in stock”, “fertilizer use measured in kg/ha”, “soil texture in Kg N ha^{-1}" and “the number of farmers with positive impact on livelihood” from 2005 to 2017 as identified by factor A with five indicators. Factor B uses a set of data selected in African countries such as Malawi, Zambia, Zimbabwe, Tanzania, Mozambique, and Kenya as given in Table 4.1.

Table 4.1: Data set (average production)

Factor A	Factor B: Regions					
	Average Prod	Nimpaye et al. 2011	Cordon-Obra et al. 2009	Ayama et al 2015	Tadese and Megerssa 2010	Lelisa et al. 2016
Live animal Number of heads (in stock)	Goat	264	456	227	234	204
		Sow et al.	Ndetchi et al.	Kebele et al.	Sim Shaw et al.	Amina et al.
	Sheep	1009	192	222	122	250
		Nakayima	Karshina et	Von	Karshina et al.	Anene et al
	Pig	248	600	312	712	300
	Total	1521	1248	761	1068	754

Table 4.1 (Continued)

Crop (in kg /ha)	Kg/ha	Malawi	Tanzania	Zambia	Zimbabwe	Zambia
	Maize	3900	2300	2800	3000	3200
		Ndula et al.	Ndula et al.	Ndula et al.	Ndula et al.	Ndula et al.
	Groundnut	204	224	198	322	340
		Ndula at el.	Ndula et al.	Ndula et al.	Ndula et al.	Ndula et al.
	Pea	220	297	184	100	444
	Total	4324	2821	3182	3422	3984
Soil (in Kg N ha^{-1})	Kg N	Ndula et al.	Ndula et al.	Ndula et al.	Ndula et al.	Ndula et al.
	Sand	98	127	99	143	168
	Loan	45	77	152	151	184
	Clay	71	85	112	46	169
	Total	214	289	363	340	521
Fertilizer (in Kg/ha)	Kg/ha	Muheza	Nsekera	Kagoro	Kalunga	Mukoka
	Nitrogen	34	66	72	67	74
		Meheza	Nsekera	Makoka	Msekera	Kalunga
	Phosphorus	20	44	52	46	43
		Muheza	Muheza	Msekora	Makuka	Kalunga
	Potassium	22	46	31	43	25
	Total	76	156	155	156	142
Live of farmers						
	Number	Malawi	Zambia	Mozambique	Kenya	Zimbabwe
	Increase	58	68	53	84	65
	Increase	87	94	71	82	59
	Increase in	55	87	65	77	86
	Total	200	243	189	243	210

Live = Livehood

MA employs the findings from Table 4.1 by combining the results as given in Table 4.2. Table 4.2 provides the correlation matrices having the same dimensions across the studies.

Table 4.2: The correlation matrices

Study			1.Goat	2.Sheep	3.Pig	4.East	5.Middle	6.North	7.South	8.West
1	1		-0.089							
	2		0.424	-0.602	1					
	3		0.981	0.874	0.287	1				
	4		0.968	0.142	0.223	0.998	1			
	5		0.993	-0.077	0.337	0.987	0.975	1		
	6		0.989	0.006	0.295	0.997	0.991	0.961	1	
	7		0.990	-0.128	0.976	0.961	0.998	0.990	1	
	8		0.990	-0.128	0.976	0.961	0.998	0.990	0.997	1
2			1.Crop	2.Groundnut	3.Pea	4.East	5.Midd	6.North	7.South	8.West
	1		1							
	2		0.023	1						
	3		-0.025	0.291	1					
	4		-0.452	0.444	0.188	1				
	5		-0.395	-0.482	0.177	0.998	1			
	6		-0.587	-0.412	0.181	0.987	0.975	1		
	7		-0.518	-0.434	0.188	0.997	0.991	0.996	1	
	8		-0.632	-0.406	0.185	0.976	0.961	0.998	0.990	1
3	1		1.Nitro	2.Phos	3.Pota	4.East	5.Midd	6.North	7.South	8.West
	1		1							
	2		0.949	1						
	3		0.419	0.558	1					
	4		-0.205	-0.179	0.461	1				
	5		-0.269	-0.239	0.414	0.998	1			
	6		-0.056	-0.021	0.555	0.987	0.975	1		
	7		-0.138	-0.108	0.502	0.997	0.991	0.996	1	
	8		-0.001	0.038	0.577	0.976	0.961	0.998	0.990	1

Table 4.2 (Continued)

4	1		1.Sand	2.Loan	3.Clay	4.East	5.Midd	6.North	7.South	8.West
	1		1							
	2		0.633	1						
	3		0.440	0.547	1					
	4		-0.189	-0.688	-0.221	1				
	5		-0.235	-0.734	-0.237	0.998	1			
	6		-0.144	-0.583	-0.187	0.987	0.975	1		
	7		-0.171	-0.641	-0.701	0.997	0.991	0.996	1	
	8		-0.132	-0.540	-0.163	0.976	0.961	0.998	0.990	1
5	1		Inco	2.Savin	3.Agro	4.East	5.Midd	6.North	7.South	8.West
	1		1							
	2		0.219	1						
	3		0.515	-0.152	1					
	4		-0.030	0.743	0.280	1				
	5		-0.068	0.758	0.219	0.998	1			
	6		0.000	0.699	0.372	0.987	0.975	1		
	7		-0.022	0.722	0.322	0.997	0.991	0.998	1	
	8		-0.003	0.673	0.399	0.976	0.961	0.998	0.990	1

East = Eastern Midd. = Middle, Nort = Northern West = Western Inco: Income, Savin: Saving. Agro: Agroforestry Motor gas, Mac. = Macaroni, Elect. = Electricity, Power. = Power of irrigation.

In MA, the selection of the model can be fixed- or random effects. We use the Chi-squared test to select the model, and the results in Table 4.6 show that homogeneity holds; therefore, the fixed effects model is applicable. If the fixed effects model is used, the integration of MA into SEM requires the correlations matrices to be averaged to obtain the pooled correlation matrix. It is with the pooled correlation matrix that the process of integration of MA into SEM begins. The parameters in the model need to be estimated and tested for significance.

The fixed- and random-effects models account for homogeneity or heterogeneity through the analysis of correlations. We assume that all studies included contribute equally to the population of effect sizes simply because all the studies have the same effects. Since individual tests were not significant, we must average all corresponding correlations to form a pooled estimate to be used in the SEM, as given below in Table 4.3.

Table 4.3: The pooled correlation matrix

	1.Goat	2 Sheep	3. Pig	4.East	5. Middle	6.North	7.South	8.West
1	1							
2	0.347	1						
3	0.365	0.129	1					
4	0.004	-0.155	0.199	1				
5	0.017	-0.079	0.159	0.998	1			
6	0.041	-0.091	0.252	0.987	0.975	1		
7	0.028	-0.073	0.460	0.996	0.992	0.986	1	
8	0.044	-0.073	0.269	0.973	0.968	0.990	0.990	1

The hypothesis of homogeneity of the correlation matrices is rejected if at least one of the correlation coefficients is heterogeneous across studies. Similarly, the variance of the correlations matrices is given below.

The variance for correlations matrices is given in Table 4.4.

Table 4.4:Variance for correlation matrices

	1.Goat	2.Sheep	3.Pig	4.East.	5.Mid.	6.North.	7.South.	8. West
1	1							
2	0.189	1						
3	0.038	0.250	1					
4	0.317	0.308	0.065	1				
5	0.300	0.322	0.057	0.001	1			
6	0.337	0.243	0.078	0.001	0.001	1		
7	0.323	0.273	0.103	0.001	0.001	0.001	1	
8	0.347	0.225	0.078	0.001	0.001	0.001	0.001	1

East. = Eastern Africa, Midd. = Middle Africa, North=. Northern Africa, South= Southern Africa West = Western Africa

The values of the variance for the correlation matrix are small, indicating that the correlations in the dataset are closer to the correlation mean; therefore, the studies used in the analysis contribute equally to the population of the effect sizes simply because the studies have the same effects across studies.

The pooled correlation matrix provides the true values with no errors that the researcher uses to integrate MA into SEM. In applying the approach, the researcher can test the causal models (Viswesvaran & Ones 1995; Shadish, 1996).

Fitting SEM. The new approach uses all the data sets using the factor analysis to generate the important components through the PCA. The variables with the highest sample variances were among the four components taken as each variable received its weight in the analysis by using the standardised variables as given in Table 4.5.

Table 4.5: Screening of different variables through PCA based on the total variance explained

Comp.	Initial Eigenvalues			Rotation Sums of Squared loading		
	Total	% of variance	Cum. %	Total	% of variance	Cum. %
1	6.753	43.819	43.819	4.955	33.036	33.036
2	4.604	30.692	74.511	3.541	23.608	56.644
3	2.089	13.928	88.439	3.319	22.127	78.771
4	1.734	11.561	100.000	3.184	21.229	100.00
5	4.85E-16	3.28E-15				

Comp. = Component, Extraction method: Principal Component Analysis, Cum. - Cumulative

From Table 4.5, 100% of the total variation is accounted for four out of 15 original variables. We rotate the four principal components using factor analysis to attain orthogonality, as shown in Table 4.6.

Table 4.6: The rotated components matrix

Component				
	1	2	3	4
Goat	0.035	0.714	-0.492	0.498
Sheep	-0.952	-0.134	0.153	0.230
Pig	0.574	0.764	0.294	0.448
Maize	-0.701	-0.454	0.547	0.053
Groundnut	0.677	-0.322	0.622	0.225
Pea	0.220	-0.606	-0.334	0.688
Nitrogen	0.924	-0.166	-0.255	-0.231
Phosphorus	0.833	0.0.37	-0.310	-0.457
Potassium	0.579	0.813	-0.062	-0.016
Increase.Inco	0.559	0.455	0.677	0.148
Increase.Savi	-0.355	0.908	-0.040	0.217
Increase.Agr	0.879	0.062	-0.108	0.460
Sand	0.765	-0.309	0.361	0.433

Table 4.6 (Continued)

	Component			
	1	2	3	4
Loan	0.752	-0.534	0.136	-0.361
Clay	0.301	-0.843	-0.393	0.211

Increase.Inco= Increase.Income, Increase.Savi= Increase.Savings, Increase.Agr = Increase.Agroforestry

The results indicate that there are four-exogenous factor models in which eight indicators (X_2, X_4, X_5, X_7, X_8, X_{10}, X_{11}, X_{15}) are connected to Factor 1, Five (X_1, X_3, X_9, X_{12}, X_{14}) are linked to Factor 2, one indicators X_{13} is related to Factor 3 and one indicator X_6 is associated with Factor 4, as displayed by the path diagram in Figure 4.2.

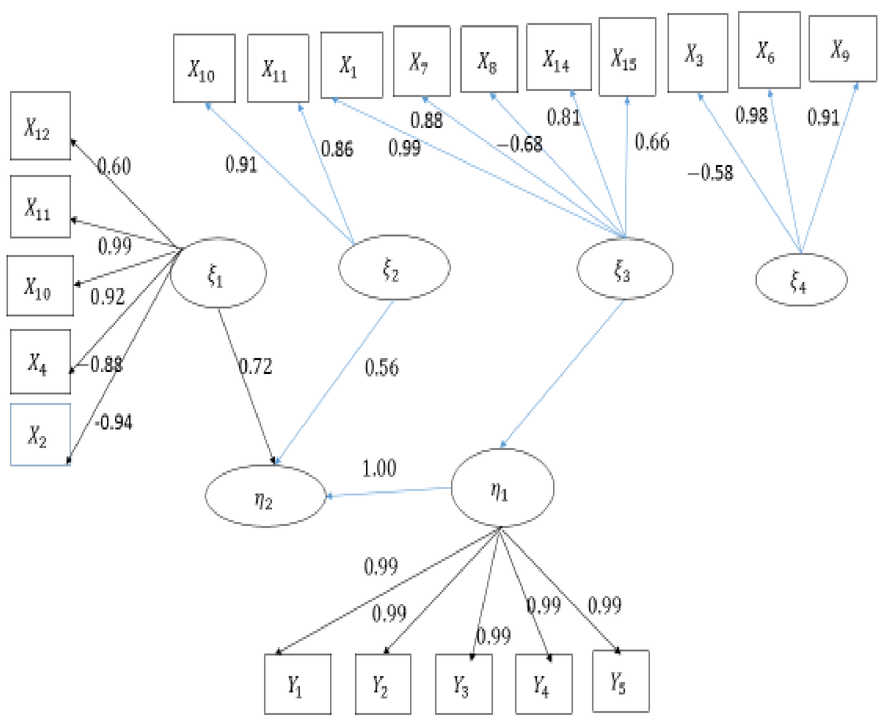


Figure 4.2: Path diagram for the structural equation model

The methodological approach of SEM explores the conditions of an agricultural theory in the complexity of a set of associations among variables that enable agricultural researchers to pursue a new understanding of hypothetical models. The path model in Figure 4.2 illustrates different models based on the correlations or the factor loadings obtained from rotated provisional factors.

The model specification: The confirmatory factor analysis enables us to determine the measurement model based on the factors that are measured by the indicators (observed variables). The path diagram described two effects:

(1) Effect of factor A in terms of four latent variables that have been evaluated with fifteen variables.

(2) Effect of factor B in term of one latent variable that has been evaluated with five variables.

The latent variables in factor A and in factor B have a direct effect on each other as shown in Figure 4.2. The parameters to be estimated are the regression coefficients, the variances, and covariance/correlations of the independent variables in the model (Bentler, 2007). We can specify the relationships of the measurement model in a set of equations as provided below

$$\eta_1 = \beta_{31}\xi_3 + \delta_3 \qquad \eta_2 = \beta_{12}\xi_1 + \delta_1 \qquad \eta_2 = \beta_{22}\xi_2 + \delta_2 \qquad \eta_2 = \gamma_{12}\eta_1 + \varepsilon_2 \qquad (4.15)$$

δ_i $(i = 1, 2, 3)$ is an error, that is, the unique factor affecting η_i and β_{ij} is the estimated parameter of the factors. The regression models based on the path diagram SEM without the intercept are given below

$$\eta_1 = 0.98\xi_3 \qquad \eta_1 = 0.72\xi_1 \qquad \eta_3 = 0.56\xi_3 \qquad \eta_3 = -0.91\eta_2 \qquad (4.16)$$

Since the heterogeneity is absent across studies, we did not apply the random effects model in this study. The test for the regression models was presented using the ANOVA as shown in Table 4.7.

Table 4.7: The ANOVA tables for testing the models

		Sum Squares	Degrees of freedom	Mean squares	F	*p*-value
Model 1	Regression	3.858	1	3.858	81.312	0.003
	Residual	0.142	3	0.047		
	Total	4.000	4			
Model 2	Regression	0.072	1	0.072	0.055	0.830
	Residual	3.928	3	1.309		
	Total	4.000	4			
Model 3	Regression	0.002	1	0.002	0.002	0.970
	Residual	3.998	3	1.333		
	Total	4.000	4			
Model 4	Regression	0,674	1	0.674	0.607	0.493
	Residual	3.328	3	1.109		
	Total	4.000	4			

The results show that the models do not fit the data, excluding Model 1, and therefore the estimated parameters cannot be interpreted. Table 4.8 summarises the un-standardised and standardised parameter estimates of the path models, t-value, and their *p*-values.

Table 4.8: Unstandardised and standardised parameter estimates of the model, t-value and their p-values.

Model	Parameter	Un-standardised Estimate	Coefficient of Standard Error	Standardized estimate	t-value	*p*-value
1	β_{31}	0.982	0.109	0.982	9.017	0.003
2	β_{12}	-0.134	0.572	-0.134	-0.235	0.830
3	β_{22}	0.024	0.577	0.024	0.041	0.970
4	γ_{12}	0.410	0.526	0.410	0.779	0.493

The parameter estimates from the models do not have good loadings on the specific factor except Model 1, that has a strong negative loading. Because the p-value > 0.05, all the parameters in the model fail to differ significantly from zero; thus, we cannot interpret these parameters although the models fit the data well. The standardised β_{12} means that one standard deviation increases in positive relationships is related to -0.134 standard deviations that increase in Factor 1 in the endogenous variable when we control the effect of the other four endogenous variables.

$\beta_{31,}$ for example, indicates the regression coefficient of the variable "3" on the variable "1".

The regression models calculated based on the ANOVA approach without the intercept are given below

$$\eta_1 = 0.982\xi_1 \qquad \eta_2 = -0.134\xi_1 \qquad \eta_3 = 0.024\xi_2 \quad \eta_3 = 0.410\eta_2 \tag{4.17}$$

In this study, we work on the direct effects of the factors only and not with their interactions. The model fit of the integration of MA into SEM is given in Equation 4.1.

The size of heterogeneity: We want to quantify the dimension of the heterogeneity of the coefficient of correlation using Equation 14, given that Q = 20 and k = 15 as indicated below

$$I^2 = \frac{U^2}{U^2+\vartheta} = \frac{Q-(k-1)}{Q} = \frac{20-(15-1)}{20} = 0.30$$

The above finding indicates that the heterogeneity is at a low level across the studies. Multiple tests used for heterogeneity assessment indicate the same results; namely, that heterogeneity is negligible.

4.8 CONCLUSION

The new approach tests the hypotheses about the parameters involved in the models. The agricultural food production is used to illustrate the approach. The fixed-effects model is more appropriate since the test of homogeneity produces significant results; therefore, we suggest collecting more studies in the MA for feasibility tests of the random-effects model. The model obtained through the fixed effects fits the data well. The newer approach is more flexible and suitable for using the factor analysis through the PCA to determine the hypothetical relations between the variables. The integration of MA into SEM using the GLS procedure generates significant results. The analysis did not consider the multivariate nature since the tests depend on the regression coefficients and not the coefficients of correlation, simply because the heterogeneity is absent across the studies. The estimated parameters are not significant; as a result, no interpretation is made of the estimated parameters.

In the previous chapters, adoption of the new technology is likely to increase the results in food production in SSA. The use of agricultural products makes a positive impact in the development of economic growth, thereby they are keys to articulate the challenges in the agricultural food production in SSA. In the next chapter, we will investigate the impact of genotype (varieties) against location (environment). This is because food production depends on the combination of factors such as genotype, environment, or their interactions simply because a type of genotype reacts differently from one location to the next. Understanding the cause for the increase of food production, whether this can be attributed to the quality of genotype, location, or some combination of factors, leads to determining the performance of a genotype across locations with the use of stability statistics.

CHAPTER 5

STATISTICAL MODEL TO ESTIMATE PARAMETERS OF GENOTYPE ACROSS LOCATION

5.1 INTRODUCTION

Food production consists primarily of crops and meats in SSA. There are many constraining factors for food production, including drought, weeds, pests and diseases, potassium and nitrogen deficiency, soil erosion, soil acidity, flooding, salinity, and many others. Plant breeding has made huge contributions to increase food production (Borlaug, 1983; Passioura, 2005). Plant breeding had increased crop yields and improved the nutritional values of numerous crops, including corn, soybeans, and different kinds of wheat (Hefferon, 2015). The increase of food production, for instance crops, depends on the achievement of today's plant breeding technology. Yield from food production as observed in data is mostly a combination of the factors of genotype, location, and the interactions between $Genotype \times locations$, since genotypes can react in a different way to different environments. There is a need to estimate stability and parameters in food production. The concept of stability has been applied in different instances, such as reducing the influence of the main environmental effects on the pattern analysis of plant breeding (Fox et al., 1982). This concept of stability has been used to estimate parameters in family selection with adaptability and genotype stability. Freiria et al. (2018) had applied it in the adaptability and stability in breeding lines of the food-type of soybeans. The analysis of variance (ANOVA) is used to demonstrate the combination of genotype and locations. The study originated by Makongwana (2016) was conducted in two consecutive summer seasons, namely, 2013 and 2014, in Kwazulu-Natal. Many factors are involved in the plant breeding process to increase crop yields based on experimental trials. We present an adaptive model to identify the genotypes with excellent performance across different locations. We perform the validation of the model and test the stability of the parameter estimates through an experiment.

This study considers several scenario options to construct many factors simultaneously in a single model that contributes to the yield in food production. The approach enables us to estimate parameters, validate the model and introduce stability measures. The focus is mainly on the main and interaction effects of genotype and the cropping systems model. Plant breeding produces the best crop under various conditions and can increase crop yields for crops such as corn, soybeans, and wheat (Hefferon, 2015). The results of food production as reflected in terms of data is a combination of the factors of genotype, location, and the interactions with genotypes simply because genotypes can react differently from one location to the next. We aimed to compare the results of different varieties in different environments. This approach would enable us to decide about a variety that is achieving the highest yield simply because we need to know the cause for the increase, whether this could be attributed to the quality of the genotype, location, or some combination of factors. The ANOVA is used to demonstrate the combination of genotype and locations. Table 5.1 presents the structure of n genotypes associated with m locations denoted by a $n \times m$ matrix. The discussion is addressed in the unique single model of yield, the ANOVA model.

5.2 OBJECTIVES

The following objectives are proposed:

1) Understanding the important features of genotype, locations, cropping systems and their interactions through a constructed model over an experiment for testing hypotheses.
2) Obtaining accurate yield estimates using the ANOVA in developing a model that fits the data.

3) Developing an approach of selecting a genotype that can increase the main and interaction effects.

4) Providing an easy interpretable measure of yield stability.

In this study, a methodological approach is formulated for assessing adaptability and stability measures of the main components and their interactions in a generic model.

5.3 MODEL BUILDING

5.3.1 Introduction

The goal is to obtain combined estimated parameters using factors such as location and genotype (variety of crops). The model is based on a scenario that assesses the genotype, location, and interaction effects under a particular type of randomisation. Suppose we have independent random samples of size n from genotype. The j^{th} value from the genotype i^{th} is denoted by y_{ijk}, that is:

Genotype 1:

$y_{111}, y_{121}, y_{131}, \cdots, y_{1m1}$ (5.1)

$y_{112}, y_{122}, y_{132}, \cdots, y_{1m2}$

....

$y_{11r}, y_{12r}, y_{13r}, \cdots, y_{1mr}$

Genotype 2:

$y_{211}, y_{221}, y_{231}, \cdots, y_{2m1}$ (5.2)

$y_{212}, y_{222}, y_{232}, \cdots, y_{2m2}$

....

$y_{21r}, y_{22r}, y_{23r}, \cdots, y_{2mr}$

Genotype 3:

$y_{311}, y_{321}, y_{331}, \cdots, y_{3m1}$ (5.3)

$y_{312}, y_{322}, y_{332}, \cdots, y_{3m2}$

$y_{313}, y_{323}, y_{333}, \cdots, y_{3m3}$

$y_{31r}, y_{32r}, y_{33r}, \cdots, y_{3mr}$

Genotype n

$y_{n11}, y_{n21}, y_{n31}, \cdots, y_{nm1}$ (5.4)

$y_{n12}, y_{n22}, y_{n32}, \cdots, y_{nm2}$

$y_{n13}, y_{n23}, y_{n33}, \cdots, y_{nn3}$

....

$y_{n1r}, y_{n2r}, y_{n3r}, \cdots, y_{nnr}$

We can summarise this information in an ANOVA table as shown in Table 5.1.

Table 5.1: Yields replicated from a genotype in a location

	Location levels					Genotype Totals	Genotype Means	True genotype Effects
Genotypes	1	2	3	....	m			
1	y_{111}	y_{121}	y_{131}	....	y_{1m1}	$y_{1..}$	$\bar{y}_{1.}$	G_1
	y_{112}	y_{122}	y_{132}	....	y_{1m2}			
	⋮	⋮	⋮	⋮	⋮			
	y_{11r}	y_{12r}	y_{13r}	⋯	y_{1mr}			
2	y_{211}	y_{221}	y_{231}	....	y_{2m1}	$y_{2..}$	$\bar{y}_{2.}$	G_2
	y_{212}	y_{222}	y_{232}	....	y_{2m2}			
	⋮	⋮	⋮	⋮	⋮			
	y_{21r}	y_{22r}	y_{23r}	....	y_{2mr}			
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	y_{n11}	y_{n21}	y_{n31}	....	y_{nm1}	$y_{n..}$	$\bar{y}_{n.}$	G_n
	y_{n12}	y_{n22}	y_{n32}	⋯	y_{nm2}			
	⋮	⋮	⋮	⋮	⋮			
	y_{n1r}	y_{n2r}	y_{n3r}	....	y_{nmr}			

Table 5.1 (Continued)

	Location levels					Genotype Totals	Genotype Means	True genotype Effects
Genotypes	1	2	3	....	m			
Location	$y_{.1.}$	$y_{.2.}$	$y_{.3.}$	....	$y_{.m.}$	$y_{...}$		
Location	$\bar{y}_{.1.}$	$\bar{y}_{.2.}$	$\bar{y}_{.3.}$	....	$\bar{y}_{.m.}$	$\bar{y}_{...}$	$\bar{y}_{...}$	
True	μ_1	μ_2	μ_2	....	μ_2			

The constructed model for the yield from food production for a randomised model involving the location design is

$$y_{ijk} = \mu + G_i + L_j + \varepsilon_{ijk}. \tag{5.5}$$

The genotype is denoted by G_i (i = 1, 2, ..., n), the location is denoted by L_j (j = 1, 2, ..., m) and replication is represented by (k= 1, 2, ..., r).

In the equation above, y_{ijk} is the yield (variety of crop) of the k^{th} replication of the i^{th} genotype in the j^{th} location.

μ is the overall average of all locations, and it is given by $\mu = \sum_{k=1}^{r} \mu_k$,

G_i represents the fixed effect of the i^{th} genotype,

L_j represents the fixed effect of the j^{th} location and

ε_{ijk} is the error term associated with y_{ijk}.

Similarly, the model that involves the main effects and the interaction factor is

$$y_{1jk} = \mu_1 + G_1 + L_1 + (G \times L)_{11} + \varepsilon_{1jk} \tag{5.6}$$

$y_{2jk} = \mu_2 + G_2 + L_2 + (G \times L)_{22} + \varepsilon_{2jk}$

$y_{3jk} = \mu_3 + G_3 + L_3 + (G \times L)_{33} + \varepsilon_{3jk}$

$y_{njk} = \mu_n + G_n + L_m + (G \times L)_{nm} + \varepsilon_{njk}$

The generic model can be written as

$$y_{ijk} = \mu + G_i + L_j + (G \times L)_{ij} + \varepsilon_{ijk} \quad (5.7)$$

where i= 1, 2, 3, ..., n; j = 1, 2, 3, ..., m and k = 1, 2, 3, ..., r and $(G \times L)_{ij}$ represents the interaction between the i^{th} genotype and the j^{th} location. The following assumptions are applied:

1. The random variables y_{ijk} are all-independent and are all normally distributed with a mean μ_i and the common variance σ^2.
2. The variance of each location is identical.
3. The core factors (genotype G_i and the location L_j) are normally independent. That is, the genotype and location effects are uncorrelated denoted, as such, as $G_i \sim^{iid} N(0, {\sigma_G}^2)$ and $L_j \sim^{iid} N(0, {\sigma_L}^2)$.
4. $\varepsilon_{ijk} \sim^{iid} N(0, {\sigma_\varepsilon}^2)$, where N denotes a normal distribution and iid means independent and identically distributed.

$$\sum_{i=1}^{n} G_i = \sum_{j=1}^{m} L_j = \sum_{i=1}^{n} (G \times L)_{ij} = \sum_{j=1}^{m} (G \times L)_{ij} = 0 \quad (5.8)$$

The core factors such as genotype and location can be considered as fixed-effect, random-effect, or subject to joint regression analysis or the Finlay-Wilkinson model (Yates & Cochran 1938; Finlay & Wilkinson, 1963).

5.3.2 Testing the Effectiveness of the Factors: Location and Genotypes

We consider a two-way analysis of variance, a method that simultaneously examines the effect of two factors on the dependent variable together with the effects of their interactions. There are three tests to perform as given below

1. Testing the genotype effects. The hypotheses are

$H_0 : G_i = 0$ for each level of genotype, with i =1, 2, ..., n. In other words, no level of genotype factor has an effect.

$H_1 : G_i \neq 0$ for at least one value of i, with i = 1, 2, ..., n . In other words, at least one level of genotype has an effect.

2. Testing the location effects. The hypotheses are

$H_0 : L_j = 0$ for each level of location, with j = 1, 2, ..., m. In other words, no level of location factor has an effect.

$H_1 : L_j \neq 0$ for at least one value of j, with j = 1, 2, ..., m. In other words, at least one level of location has an effect.

3. Testing for interaction effects between levels of the factors genotype and location. The hypotheses are

$H_0 : (G \times L)_{ij} = 0$ for each combination of i and j. In other word, there are no interaction effects.

$H_1 : (G \times L)_{ij} \neq 0$ for at least one combination of i and j. In other word, at least one level of genotype has an effect.

5.3.3 Validation, Consistency and Stability of Estimate

The success of plant breeding is about having high production stability, but the fundamental reason for the differences between genotypes is the implication of the interaction of the terms $Genotype \times$

$Location$. In other words, the high production stability depends on the performance of genotypes in the specific environmental conditions where the varieties must grow. We can define or explain some of the interaction effects through the environmental factors, for instance, the number of diseases or pests, rainfall, dry season, soil fertility, etc. (Ferreira et al., 2006). The interaction $G \times L$ is the main concept that is addressed in this study.

The second challenge to overcome to ensure the success of plant breeding is having high productivity stability through new genotypes that are grown in several locations under diverse climate and soil fertility scenarios, including different seasons of the year (Becker and Leon, 1988). In other words, accommodating changes in environmental conditions due to location and seasons is a major factor in the achievement of the desired performance and high productivity. Hence, the experiments designed in this study and the associated special randomised complete block design (RCBD) techniques to model the estimated genotype means in the jth environment have been created. The regression model is as given in Equation 5.5.

We aimed to identify the genotype performance that will remain stable irrespective of the environmental conditions that might have changed. This constant analysis is valid only when the interaction factor ($G \times L$) is present (Hussein et al. 2000). Becker and Leon (1988) indicated that the two most important cases in practice for the concept of stability were:

a) The genotype that maintains its performance independently of the variations in the ecological conditions. This concept is static, and this kind of stability is called biological stability.
b) The expected approach is that the genotype varies its performance when the environment changes. The concept describes a dynamic approach, and this kind of stability is agronomic stability.

The stability of a genotype is measured by the sample variance, which calculates the variability of yields across locations. Plant breeders have used these approaches mainly for assessing stability; they are equally applicable in the context of agronomic experiments (Piepho 1998). Plant breeders are interested in classifying the varieties accessible in terms of changes in the yields due to changes in the location.

5.4 CONSTRUCTED MODEL THROUGH CROPPING SYSTEM

To understand the contribution of each factor that can occur in food production, we must document the constituents of each factor. In this section, the increase of crops depends on the way the sowing and harvesting systems are managed. In other words, this is the cropping system that can be expressed by the seasonal period (summer and winter) in the way that crop cultivated through cropping system is described as the combination of management practices and plant genotype to produce the desired crops. The construction of the second model involves two core factors such as the system and location. From the breeder's point of view, location and year by location combinations are random effects since the breeder's objective is to generate seed for a bigger area. By contrast, the farmer's perspective considers location effects as fixed, whereas variations from year to year are unpredictable.

The constructed model involves two core factors, such as the cropping system and location, in a two-way table as shown in Table 5.1. The model in terms of system, location and year is given by

$$y_{rs} = \mu + Y_r + L_s + \delta_{rs} \tag{5.9}$$

The model with interaction is

$$y_{rs} = \mu + Y_r + L_s + (Y \times L)_{rs} + \delta_{rs} \tag{5.10}$$

The year is denoted by Y_r (r = 1, 2, ..., n) and the location is denoted by L_s (s = 1, 2, ..., m), where y_{rs} is the yield of a system of $the\ r^{th}$ year in the s^{th} location,

where μ is a general mean,

Y_r represents the random effect of the r^{th} year,

L_s represents the random effect of the s^{th} location, and

δ_{rs} is the error term associated with y_{rs}.

Assumptions:

i) We assume that the effects of $\mu,\ Y_r$, L_s and δ_{rs} differ among systems.
ii) The terms Y_r , L_s and δ_{rs} are random with independent variances ${\sigma_r}^2$, ${\sigma_s}^2$ and ${\sigma_\delta}^2$ respectively.

Therefore, the variance of Equation 5.9 is

$$Var(Y_{rs}) = Var(\mu + Y_r + L_s + \delta_{rs})$$

$$= Var(\mu) + Var(Y_r) + Var(L_s) + Var(\delta_{rs})$$

$$Var(Y_{rs}) = {\sigma_Y}^2 + {\sigma_L}^2 + {\sigma_\delta}^2 \qquad (5.11)$$

In the assessment of the planting systems, the terms years and locations are considered random since year effects are unpredictable.

The variance of an observation in a random year at a given location S is

$Var(Y_{rs}/S) = \sigma^2(r/s) = {\sigma_Y}^2 + {\sigma_\delta}^2$; representing the stability measures for the variability factors.

1) The effects Y_r, L_s and δ_{rs} of a pair of systems are correlated (Piepho et al., 1998). In other words, the effect Y_r of one system is correlated with the other system L_s. Similarly,

the effects L_s is correlated with another system, the effects δ_{rs} are also correlated and so on.

2) The variances of interaction effects associated with systems differ among systems. (Piepho, 1994).

The stability of yield is an important characteristic to be considered when judging the value of a cropping system relative to others (Piepho et al., 1998). Plant breeders and geneticists have mainly applied the methods for assessing yield stability in experiments; therefore, these techniques can be applied by agronomists in the context of experiments, which calculate different cropping systems in several locations (Piepho, 1998).

5.5 CONSTRUCTING A GENERIC MODEL

This section aims to combine all the approaches from the core factors to form a generic modelling framework. It is expected that the combined model will be suitable for stability analysis simply because of sharing strength among the variables. The variance can quantify the variation of yields from location to location. Similarly, the mean is considered a fixed effect and the variance as a random effect. The two parameters (mean and variance) describe the response pattern of the generic model.

We define the generic model by different conditions such as genotype, locations, crop-year cycles and management practices that are measured independently or in combination with the environmental conditions. The components of interactions are as follows:

1. $Genotype \times Location (G \times L)$
2. $Genotype \times Year (G \times Y)$
3. $Genotype \times Management\ Practices\ (G \times MP)$
4. $Genotype \times Location \times Year (G \times L \times Y)$
5. $Genotype \times Location \times Management\ Practices (G \times L \times MP)$
6. $Genotype \times Year \times Management\ Practices (G \times Y \times MP)$

7. $Genotype \times Location \times Year \times Management\ Practices (G \times L \times Y \times MP)$

The model is helpful by identifying the stable genotype and its adapted components. The model can be written as shown below

$$y_{ijst} = \mu + G_i + L_j + Y_s + MP_t + (G \times L)_{ij} + (G \times Y)_{is} + (L \times Y)_{ls} + (G \times MP)_{it} + (L \times MP)_{lt} + (G \times L \times Y)_{ijs} + (G \times L \times MP)_{ijt} + (G \times Y \times MP)_{ist} + (L \times Y \times MP)_{jst} + (G \times L \times Y \times MP)_{ijst} + \varepsilon_{ijst} \quad (5.12)$$

where y_{ijst} is the yield value of the effect of the i^{th} level of genotype (G), $at\ the\ j^{th}$ level of the location (L), $the\ s^{th}$level of the crop-year (Y) and $the\ t^{th}$ level of the management practice. On this representation μ is a grand mean, G_i is the effect of the i^{th} genotype, L_j is the j^{th} effect location, Y_s is the effect of $the\ s^{th}$ level of the crop-year and $(MP)_t$ is the effect of t^{th} management practice. In addition, $(G \times L)_{ij}$ is the interaction effect between the i^{th} genotype and j^{th} location, $(G \times L \times (MP))_{ijt}$ is the interaction effect between the i^{th} genotype, $the\ j^{th}$ of location and $the\ t^{th}$ management practice, $(G \times L \times Y)_{ijs}$ is the interaction effect between the i^{th} genotype, $j^{th}$of location and s^{th} of the crop-year, $(G \times L \times Y \times (MP))_{ijrst}$ is the interaction effect between the i^{th}genotype, $j^{th}$of location, s^{th} of crop-year and t^{th} management practice and ε_{ijrst} is the error term.

5.6 METHODS OF ESTIMATING STABILITY IN THE GENERIC MODEL: MAIN AND INTERACTIONS EFFECTS

The concept of high stability is associated with high mean performance simply because a stable genotype has a high mean performance (Yan & Tinker, 2006). Estimating the stability of the genotype is very important in the interpretation of the results to indicate a stable genotype across all the components involved. There is a need to describe the stability of the genotype against any main factor effect such as location (L), crop-year (Y) or management practices (MP) so that we can maintain relative performance across a range of environments. The main effect in the model of Equation 5.12 is assumed to be a fixed or random effect.

5.6.1 Estimating Stability when the Main Effects are Fixed

The objective is to estimate and test the hypotheses about the main and interaction effects, for example, for G_i, G_L and $(G \times L)_{ij}$. In this case, we assume that all the term-factors in Equation 5.12 have a fixed effect since the stability is measured by the mean and the variance. The variance estimates of yield Y_{ijst} is given by the variance of

$$Var(Y_{ijst}) = Var\big(\mu + G_i + L_j + Y_s + (MP)_t + (G \times L)_{ij} + (G \times Y)_{is} + (G \times (MP))_{it} + (G \times L \times Y)_{ijs} + (G \times L \times (MP))_{ijt} + (G \times Y \times (MP))_{ist} + (G \times L \times Y \times (MP))_{ijst} + \varepsilon_{ijst}\big)$$

$$Var(Y_{ijst}) = Var(\varepsilon_{ijst}) = \sigma^2 \qquad (5.13)$$

This could indicate that the genotype cannot have the results changed by changes in the location (L), year (Y), and management practices (MP) under the hypothesis

$H_0 : \sigma^2 = 0$ Vs $H_0 : \sigma^2 \neq 0$, in which H_0 is not rejected.

5.6.2 Estimating Stability when the Main Effects are Random

The objective is to estimate and test the hypotheses about the variances of the main effects and their interaction effects. For example, for ${\sigma^2}_G$, ${\sigma^2}_L$ and ${\sigma^2}_{G\times L}$.

This is the dynamic concept measured, including the interactions as it is most recommended (Ferreira et al., 2006). The variance estimates of yield Y_{ijst} is given by the variance of

$$Var(Y_{ijrst}) = Var\big(\mu + G_i + L_j + Y_s + MP_t + (G \times L)_{ij} + (G \times Y)_{is} + (G \times (MP))_{it} + (G \times L \times Y)_{ijs} + (G \times L \times (MP))_{ijt} + (G \times Y \times (MP))_{ist} + (G \times L \times Y \times (MP))_{ijst} + \varepsilon_{ijst}\big)$$

$$Var(Y_{ijst}) = Var(G_i) + Var(L_j) + Var(Y_s) + Var((MP)_t) + Var(G \times L)_{ij} + Var(G \times Y)_{is} + Var(G \times (MP))_{it} + Var(G \times L \times Y)_{ijs} + Var(G \times L \times (MP))_{ijt} + Var(G \times Y \times (MP))_{ist} + + Var(G \times L \times Y \times (MP))_{ijst} + Var(\varepsilon_{ijst})$$

$$Var(Y_{ijst}) = \sigma^2{}_G + \sigma^2{}_L + \sigma^2{}_R + \sigma^2{}_{MP} + \sigma^2{}_{GL} + \sigma^2{}_{G(MP)} + \sigma^2{}_{GLY} + \sigma^2{}_{GL(MP)} + \sigma^2{}_{GY(MP)} + \sigma^2{}_{GLY(MP)} + \sigma^2 \quad (5.14)$$

This could mean that the genotype can have the results changed with the change in Location (L), year (Y), and management practices (MP). It implies that $H_0 : \sigma^2 = 0$ is rejected.

$$Var(Y_{ijst}) = \sigma^2{}_{ijs} \quad (5.15)$$

The variance $\sigma^2{}_{ijs}$ is the stability measure under the assumption that the variations among main effects and their interactions are equally important.

5.6.3 Estimating Stability when the Main Effects are Fixed or Random

Estimating stability when the genotype is fixed, and other main effects are random

This covers the case when the genotype is fixed, and other main effects (location, year, and management practices) are random. This means that we are interested in a particular set of genotypes that is grown in a randomly set location selected purposefully by an investigator for a specific year and special management practices. If the experience is repeated, the location and other factors are most likely to differ each time.

The variance estimates of yield Y_{ijst} is given by the variance of

$$Var(Y_{ijst}) = Var(\mu + G_i + L_j + MP_t + (G \times L)_{ij} + (G \times Y)_{is} + (G \times MP)_{it} + (G \times L \times Y)_{ijs} + (G \times L \times MP)_{ijt} + (G \times Y \times MP)_{ist} + (G \times L \times Y \times MP)_{ijst} + \varepsilon_{ijst})$$

$$Var(Y_{ijst}) = Var(G_i) + Var(L_j) + Var(Y_s) + Var((MP)_t) + Var(G \times L)_{ij} + Var(G \times Y)_{is} + Var(G \times (MP))_{it} + Var(G \times L \times Y)_{ijs} + Var(G \times L \times (MP))_{ijt} + Var(G \times Y \times (MP))_{ist} + Var(G \times L \times Y \times (MP))_{ijrst} + Var(\varepsilon_{ijst})$$

$$Var(Y_{ijst}) = \sigma^2{}_L + \sigma^2{}_R + \sigma^2{}_{MP} + \sigma^2 \qquad (5.16)$$

This means testing the null hypothesis $H_0 : \sigma^2{}_L = 0$; $H_0 : \sigma^2{}_R = 0$; $H_0 : \sigma^2{}_{MP} = 0$. This is because the genotype indicates variation responding according to the changing of the locations (or environment). There are significant effects from the population variances for the fixed effects. The distribution is independent and identically normally distributed with the mean zero and the variance σ^2. A chi-squared test is used to test if the variance of a population is equal to a specified value. Snedecor and Cochran (1983) is used for the test. This test will enable us to detect the differences in variance of a quantitative trait between independent factors of genotype, location and management practice. For simplicity, we consider the specified value equal to zero. To increase the yield of food production and sustain it requires the combined analysis of variance to be significant (p-value < 0.05). In other words, the combination of genotype (G), location (L), Year (Y), and management practice (MP), and their interactions must be significant. The analysis of variance is used in this approach while G, L, Y, and MP are the main effects therefore, the fixed effects hold.

The variance estimates of yield Y_{ijst} is given by the variance of

$$\begin{aligned} Var(Y_{ijst}) = Var\big(\mu \ + \ G_i \ + \ L_j \ + Y_s + \ (MP)_t + \ (G \times L)_{ij} + \ (G \times Y)_{is} + (G \times (MP))_{it} \\ + (G \times L \times Y)_{ijs} + \ (G \times L \times (MP))_{ijt} \ + \ (G \times Y \times (MP))_{ist} \\ + (G \times L \times Y \times (MP))_{ijrst} + \ \varepsilon_{ijst}\big) \end{aligned}$$

$$Var(Y_{ijst}) = Var(G_i) + Var(L_j) + Var(Y_s) + Var((MP)_t) + Var(G \times L)_{ij} + Var(G \times Y)_{is} + Var(G \times (MP))_{it} + Var(G \times L \times Y)_{ijs} + Var(G \times L \times (MP))_{ijt} + Var(G \times Y \times (MP))_{ist} + Var(G \times L \times Y \times (MP))_{ijrst} + Var(\varepsilon_{ijst})$$

$$Var(Y_{ijrst}) = \sigma^2{}_G + \sigma^2 \qquad (5.17)$$

This means that we are testing the null hypothesis $H_0 : \sigma^2{}_G = 0$.

An example is used to determine the performance of the genotype across locations using the stability approach to understand the cause for the increase of food production under the genotype, location, or their interaction effects.

5.7 DATA ANALYSIS

5.7.1 Site Selection and Description

The study was conducted in two consecutive summer seasons, as shown in Table 5.2. This table presents the locations selected based on their heterogeneity in terms of attitude, longitude, year and average rainfall. The experiments were established in six locations: Bethal, Carolina, Cedara, Potchefstroom, Winterton and Mooi River during the same seasons.

Table 5.2:Geographical positions and information of where the experiments were conducted

Location	Code	Province	Latitude	Longitude	Year	AR (mm)
Bethal	L_1	Mpumalang	26.4579^0S	29.4667^0E	2	710
Carolina	L_2	Mpumalang	26.0731^0S	30.1070^0E	2	614
Cedara	L_3	Kwazulu	29.5478^0S	30.2667^0E	2	900
Potchefstroom	L_4	Northwest	26.7145^0S	27.0970^0E	2	615
Winterton	L_5	Kwazulu	28.8166^0S	29.5296^0E	2	789
Mooi River	L_6	Kwazulu	29.2106^0S	30.0030^0E	2	900

AR = Average annual rainfall

5.7.2 Treatment and Experimental Design

Thirty-eight genotypes were used for the study in which 29 were experimental hybrids (G1 to G29) and nine were commercial check hybrid (G30 to G38), as reflected in an ANOVA in Table 5.3. In the next section, the data for each location is analysed separately and then combined across locations for the maize main yield with two replications. To determine other measures such as the interactions: $genotype \times location\ (G \times L),\ genotype \times year\ (G \times Y),\ genotype \times location \times year\ (G \times L \times Y)$. A combined ANOVA is used in which location and year are considered random factors while genotype is considered a fixed factor. The equation used for the model is as follows:

$$y_{ijs} = \mu + G_i + L_j + Y_s + (G \times L)_{ij} + (G \times Y)_{is} + (G \times L \times Y)_{ijs} + \varepsilon_{ijs} \qquad (5.18)$$

5.8 PERFORMANCE OF GENOTYPES ACROSS INDIVIDUAL LOCATIONS

To determine the performance of a genotype across all locations, we construct an ANOVA table between genotypes for each specific yield to determine the genotypes that have the highest significant mean across all locations. The *p*-value, least significant difference (LSD) or coefficient variation (CV) are used for testing hypotheses but in this study, the CV is used to test the performance of genotypes.

Table 5.3: Genotype and location factors: grain yield of maize evaluated across six locations in KZN in 2013 and 2014.

Genotype		Location						Gen- T	GenM	CV_i %
Code	N	Betha	Carol	Cedar	Pot-ch	Wint	Moo			
03C47	G_1	8.35	8.69	8.351	8.351	8.351	8.01	50.105	8.351	2.60
11C34	G_2	7.31	9.14	7.158	7.840	7.822	7.67	46.932	7.822	8.96
11C32	G_3	8.79	8.49	7.349	10.22	10.19	7.70	52.739	8.790	13.8
11C29	G_4	7.70	8.18	6.619	8.177	10.21	8.18	49.063	8.177	14.3
02C31	G_5	8.12	8.57	7.688	8.091	8.116	8.12	48.696	8.116	3.44
11C63	G_6	8.26	8.26	6.756	8.258	10.03	7.99	49.548	8.258	12.7
11C25	G_7	7.47	8.42	6.455	7.436	7.436	7.44	44.615	7.436	8.34
13C70	G_8	7.46	8.45	8.449	8.449	9.77	8.12	50.694	8.449	8.92
13C70	G_9	7.23	8.06	6.497	7.609	10.42	7.96	47.78	7.963	16.7
13C70	G_{10}	8.58	7.91	7.671	8.583	10.17	8.58	51.498	8.583	10.2
13C70	G_{11}	8.46	8.91	7.466	8.050	9.87	7.98	50.736	8.456	10.0
13C70	G_{12}	7.42	8.39	8.385	8.385	9.98	7.76	50.311	8.385	10.5
13C71	G_{13}	7.98	7.52	6.502	7.98	9.92	7.98	47.88	7.98	13.9
13C71	G_{14}	7.69	7.21	8.249	8.249	9.84	8.09	49.331	8.222	10.8
13C71	G_{15}	7.456	8.344	8.344	7.847	10.11	7.96	50.064	8.344	11.1
11C17	G_{16}	7.429	7.860	6.780	8.208	10.05	8.07	48.392	8.065	13.7
11C15	G_{17}	10.5	10.5	10.46	10.46	10.46	10.5	62.76	10.46	0.00
11C15	G_{18}	6.40	6.40	6.705	8.638	10.26	6.40	44.803	7.467	21.7
11C22	G_{19}	8.12	7.12	8.118	8.908	8.118	8.32	48.709	8.118	7.09
11C13	G_{20}	7.78	7.78	7.775	7.775	7.775	7.78	46.65	7.775	0.00
11C15	G_{21}	7.77	7.31	8.128	8.128	9.68	7.76	48.768	8.128	10.1
11C22	G_{22}	7.69	7.69	7.686	7.686	7.686	7.69	46.116	7.686	0.00
11C14	G_{23}	7.78	8.06	7.921	7.921	7.921	7.92	47.526	7.921	1.12
11C22	G_{24}	7.422	7.125	7.433	7.433	7.433	7.75	44.597	7.433	2.66

Table 5.3 (Continued)

Genotype		Location						Gen- T	GenM	CV_i %
Code	N	Betha	Carol	Cedar	Pot-ch	Wint	Moo			
10HDT	G_{25}	7.874	8.386	8.386	8.898	8.386	8.39	50.316	8.386	3.86
14XH1	G_{26}	7.865	7.865	7.865	7.542	7.865	8.19	47.19	7.865	2.60
14XH1	G_{27}	7.885	7.885	6.944	8.406	7.885	8.30	47.309	7.885	6.55
14XH0	G_{28}	7.245	7.100	6.520	7.534	7.100	7.10	42.599	7.100	4.65
PAN6	G_{29}	8.496	8.446	8.518	11.72	11.46	8.01	56.654	9.442	17.8
PAN66	G_{30}	8.233	9.182	8.788	8.705	9.64	8.17	52.727	8.788	6.57
DKC78	G_{31}	8.207	8.116	8.463	7.819	9.71	8.46	50.778	8.463	7.76
DKC80	G_{32}	8.599	8.907	6.971	9.596	10.46	8.91	53.44	8.907	13.0
PAN53	G_{33}	7.558	7.601	7.676	7.568	7.601	7.60	45.605	7.601	0.54
PAN67	G_{34}	7.308	7.174	7.040	7.174	7.174	7.17	43.044	7.174	1.18
SC633	G_{35}	7.776	7.776	7.776	7.776	7.776	7.78	46.656	7.776	0.00
SC506	G_{36}	8.01	8.158	7.002	8.761	8.01	8.11	48.053	8.001	7.09
SC301	G_{37}	6.871	6.871	6.871	6.871	6.871	6.87	41.226	6.871	0.00
SC403	G_{38}	7.250	7.497	7.862	7.862	7.862	8.84	47.173	7.862	6.88
Location		298.2	305.296	289.6	314.9	339.418	303.	1851.0		
Location mean		7.849	8.034	7.622	8.287	8.932	7.99	8.071	8.119	

Betha = Bethal, Carol = Carolina. Cedar = Cedara, Pot-ch = Pot-chefstroom, Wint = Winterton, Moo = MooiRiver, Gen-T = Genotype Total, GenM = Genotype mean

The performance of genotypes across individual locations is reflected in an ANOVA table for genotypes for maize yield across all locations, revealing the genotype with the highest significant mean across all locations. The lower the value of the coefficient of variation, the more precise the estimate. For example, the genotype such as G_{17} , G_{20}, G_{22}, G_{35}, and G_{37} followed by G_{23}, G_{33} and G_{34}.

There is a need for providing new genotypes that will indicate the performance in the specific environmental conditions based on the stability statistics measures to understand and improve the yield production and livelihoods.

5.9 STABILITY ANALYSIS

The purpose of this section is to investigate the four-stability statistic approaches and to define the comparison measures of their relationships concerning the reaction of the genotypes against locations. A genotype is stable if its variation among locations is small or if the residual mean square from a regression on a location index is small (Lin et al., 1985). This type of stability is biological stability that breeders use in expectation of a high level of yield in selecting a location under experiment. We have selected four stability statistic measures, among others, that we will apply to compare the relationship measures based on the performance of each genotype across locations to provide the genotypes that increase yield production most.

Environmental conditions such as climate, soil conditions and diseases have increased the concerns of the plant breeders (Makongwana, 2016). These factors emphasise the benefit and importance of stability statistic measures. This section examines and uses the stability statistic measures to calculate the stability statistics of the genotypes. A two-way model is used as shown in Table 5.6. The four approaches used for stability statistics are:

1 The coefficient of variability CV_i presented by Francis et al. (1978). This can be denoted as

$CV_i = \frac{S_i}{\bar{Y}_{i.}} \times 100.$

$$S_i^{\,2} = \sum_{j=1}^{m} \frac{\left(Y_{ij} - \overline{Y_{i.}}\right)^2}{m-1} \tag{5.19}$$

where Y_{ij} is the yield of the i^{th} genotype in the j^{th} location and $\overline{Y_{i.}}$ is the genotype means. A small CV_i % (less than 20%) is good (Lin et al., 1985).

2 To calculate the stability measure for genotype i, we consider the mean of the variance component for the interaction $genotype \times location$ $(G \times L)$, denoted by $\hat{\theta}_{i,}$ as presented by Plaisted and Peterson (1959).

$$\hat{\theta}_i = \frac{m}{2(n-1)(m-1)} \sum_{j=1}^{m} \left(Y_{ij} - \overline{Y}_{i.} - \overline{Y_{.j}} + \overline{Y}_{..}\right)^2 + \frac{SS(G \times L)}{2(n-1)(m-1)}$$

$$SS(G \times L) = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(Y_{ij} - \overline{Y}_{i.} - \overline{Y_{.j}} + \overline{Y}_{..}\right)^2$$

3 A lower value for the $\hat{\theta}_i$ % estimate of the genotype shows more stability. The genotype with the smallest mean variance component contributor for the total interaction is the most stable.

Wricke (1962) proposed the "ecovalency" parameter ${W_i}^2$ that calculates the interaction of $genotype \times location$ $(G \times L)$. The parameter is used to measure for a genotype i. It is given by

$${W_i}^2 = \sum_{j=1}^{m} \left(Y_{ij} - \overline{Y}_{i.} - \overline{Y_{.j}} + \overline{Y}_{..}\right)^2$$

where $\overline{Y_{.j}}$ the location means, $\overline{Y}_{i.}$ is the genotype means and $\overline{Y}_{..}$ is the grand mean in the ANOVA as reflected in Table 5.4. If the value of W_i = 0, then the genotype is stable and if W_i is greater than zero, the genotype is unstable and therefore Wricke called this parameter "ecovalency". A genotype with the smallest ecovalence ${W_i}^2$ is thought to be the most stable.

4 A stability statistic to measure the residuals in a two-way classification was proposed by Shukla (1972). The stability statistic is a fixed estimate measure of the variance for the genotype across locations, as given below.

$${\sigma_i}^2 = \frac{n}{(n-2)(m-1)} \sum_{j=1}^{m} \left(Y_{ij} - \overline{Y}_{i.} - \overline{Y_{.j}} + \overline{Y}_{..}\right)^2 - \frac{SS(G \times L)}{(n-1)(n-2)(m-1)}$$

The null hypothesis is $H_0 : {\sigma_i}^2 = 0$. A genotype is stable if the sample estimate is zero. That is, the genotype will not have the results changed with the changes in the location, therefore H_0 is not rejected.

Table 5.4: Genotypes mean and ecovalency estimates for 38 common maize varieties

Genotype		GenM $\overline{y_{i..}}$	S.es $W^2{}_i$	CV_i %	$\hat{\theta}_i$	$\sigma_i{}^2$	SE
Code	Number						
03C475	G_1	8.351	0.073	2.60	0.029	0.324	0.089
11C3417	G_2	7.822	0.263	8.96	0.169	1.890	0.286
11C3201	G_3	8.790	0.000	13.82	1.028	11.468	0.496
11C2974	G_4	8.177	0.225	14.25	0.871	9.712	0.476
02C3156	G_5	8.116	0.000	3.44	0.003	0.032	0.114
11C6363	G_6	8.258	0.000	12.66	0.566	6.310	0.427
11C2557	G_7	7.436	0.000	8.34	0.070	0.780	0.253
13C7060	G_8	8.449	0.984	8.92	0.152	1.698	0.303
13C7065	G_9	7.963	0.538	16.71	1.485	16.553	0.543
13C7071	G_{10}	8.583	0.000	10.16	0.274	3.056	0.356
13C7082	G_{11}	8.456	0.000	10.01	0.242	2.702	0.345
13C7083	G_{12}	8.385	0.931	10.49	0.283	3.161	0.359
13C7109	G_{13}	7.980	0.000	13.91	0.718	8.009	0.453
13C7110	G_{14}	8.222	0.279	10.79	0.293	3.266	0.362
13C7122	G_{15}	8.344	0.789	11.11	0.350	3.900	0.379
11C1774	G_{16}	8.065	0.405	13.65	0.694	7.743	0.449
11C1579	G_{17}	10.46	0.000	0.00	0.000	0.000	0.000
11C1566	G_{18}	7.467	1.139	21.73	3.281	36.581	0.663
11C2245	G_{19}	8.118	0.000	7.09	0.052	0.579	0.235
11C1350	G_{20}	7.775	0.000	0.00	0.000	0.000	0.000
11C1511	G_{21}	8.128	0.131	10.07	0.212	2.362	0.334
11C2242	G_{22}	7.686	0.000	0.00	0.000	0.000	0.000
11C1483	G_{23}	7.921	0.020	1.12	0.000	0.000	0.036
11C2243	G_{24}	7.433	0.000	2.66	0.001	0.008	0.081
10HDTX11	G_{25}	8.386	0.262	3.86	0.005	0.058	0.132
14XH146	G_{26}	7.865	0.000	2.60	0.001	0.009	0.083
14XH149	G_{27}	7.885	0.000	6.55	0.034	0.374	0.211
14XH082	G_{28}	7.100	0.021	4.65	0.006	0.063	0.135
PAN6Q445	G_{29}	9.442	0.896	17.76	3.742	41.726	0.685
PAN6611	G_{30}	8.788	0.308	6.57	0.046	0.518	0.229
DKC78-4	G_{31}	8.463	0.066	7.76	0.088	0.980	0.268
DKC80-40	G_{32}	8.907	0.095	13.03	0.857	0.857	0.474
PAN53	G_{33}	7.601	0.002	0.54	0.000	0.000	0.017
PAN67	G_{34}	7.174	0.018	1.18	0.000	0.000	0.035
SC633	G_{35}	7.776	0.000	0.00	0.000	0.000	0.000
SC506	G_{36}	8.001	0.000	7.09	0.049	0.548	0.232
SC301	G_{37}	6.871	0.000	0.00	0.000	0.000	0.000
SC403	G_{38}	7.862	0.375	6.88	0.040	0.452	0.221

GenM = Genotype mean, S.es = Sample estimator, SE = Standard error

There is consistency between the four methods in the arrangement of the best genotypes G_{17}, G_{20}. G_{22}, G_{23}, G_{24}, G_{33}, G_{34}, G_{35} and G_{37} for both adaptability and stability statistic measures simply because they provide the best average gain of maize yield, and they maintain their performance independently of the variation of the location.

The next section examines thirty-eight hybrids (experimental and commercial check) genotypes grown in six locations over two years. The analysis of variance is used to evaluate the significance of the main and interaction effects as shown in Table 5.5.

5.10 PERFORMANCE OF THE COMBINED COMPONENTS THROUGH A GENERIC MODEL

Multiple components are now used simultaneously, such as the main components of genotypes and locations that we combine through a two-way ANOVA to evaluate the main effect genotype (G), location (L), and interactions. The calculated F test statistics and *p*-value are used to determine the significant results.

Table 5.5: Summarised analysis of variance among thirty-eight maize genotypes in six locations

Source	df	SS	MS	F	*p*-value
Genotype	37	93121.85	2516.8068	140.2472	0.0001
Location	5	2093.8152	418.7630	23.3353	0.0001
$G \times L$	185	20277.87	109.6101	6.1079	0.0001
Error	227	4073.636	17.9455		

A model fit describes the relationship between a response variable and one or more predictor(s). In this study, we investigate the relationship between the genotype (G), the location (L) and the interaction effects ($G \times L$) from the period of 2013 to 2014. Using the ANOVA test, the models were statistically significant. In other words, the levels of genotypes were associated with different performance based on the location. The same approach indicated that the levels of locations were related to differences in performance. These results indicate overall strong differences that are statistically highly significant between the main and the interaction effect because there is a diversity within hybrids, allowing an opportunity to accurately select a suitable hybrid for the different locations. We also explore the R-squared (R^2) statistic to further test the regression model separately for the main effects. The results show that the $R^2 = 0.21$ and 0.17 for genotype and location model, respectively. This finding indicates a large difference between the observed and the model predicted since the points appeared randomly scattered on the plot.

5.11 CONCLUSION

Many challenges still exist in determining the performance of genotypes across locations. Several methods and models exist that prove the relationship between multiple methods or models, but each one has its limitations that need to be addressed. The stability statistics approach can be used to evaluate the performance of genotypes, but we still need further clarification on how to describe the location response of some patterns of the genotypes that might be required. The solution is to obtain an approach that can offer a possible improvement of the techniques used to evaluate stability with the understanding of the causes around interaction; thus, that approach will have the potential to be successful. We propose identifying the best-adapted, responsive, and stable genotypes because this worthwhile task is done up to the final phases of the plant breeding.

CHAPTER 6

THE USE OF LINEAR MIXED-EFFECTS MODEL IN META-ANALYSIS WITH MULTIPLE OUTCOMES

6.1 INTRODUCTION

Meta-analysis (MA) has become a common way of synthesising evidence and summarising the results of individual studies (Chalmers et al., 1997). Multivariate statistical techniques help in solving complex problems to find the relationships between several variables simultaneously. MA is a technique that allows the joint synthesis of multiple correlated outcomes. One of MA's benefits is to transform the findings obtained from different studies into a joint measure of standardised effect sizes (ESs) as inputs of MA before combining the results. With the ESs, we can initiate comparing and synthesising the results across similar studies (Lipsey & Wilson, 2001).

The meta-analysis approach assumes the ESs are independent from one study to the other, but in reality, a single study could involve multiple tests of the same hypothesis, resulting in reporting multiple outcomes (MOs). So, in such a situation, what could be considered independent if a decision had to be made regarding a single study that reports MOs? The independency assumption of the ESs holds when independent studies are considered. However, this may not be the case when ESs are from studies involving many variables measured within the same subjects (Moeyaert et al., 2018). The research of this kind requires MOs with MA rather than the univariate MA. Research done in the agricultural field uses univariate MA in synthesising ESs simply because ESs are assumed to be independent. However, applied research that produces more than one outcome within a study, resulting in multiple ESs, is common in the agricultural field. Jackson et al. (2011) reported the potential of the multivariate MA and promised to describe the areas of application that multivariate

MA has found, the methods available, the difficulties with the approach. Riley (2009) examined the role within a study of the correlation in multivariate MA, which might have the effect of it being ignored. White (2009) developed a multivariate MA random effects approach that combines estimates of several related parameters over several studies since the parameters referred to MOs.

Houwelingen et al. (2002) discussed methods that analyse univariate and bivariate treatment effects and meta-regression methods. In medicine, Malcolm (2019) published an empirical comparison of the use of univariate and multivariate meta-analysis in Cochrane Pregnancy and Childbirth Reviews with multiple binary outcomes. The applicability of meta-analysis for studying multiple ecosystem outcomes in rangelands was explored by Gravuer (2018). Many studies that have been done in the agricultural field use separate univariate MA in synthesising ESs. However, research work that produces more than one outcome within a study, resulting in multiple ESs, are popular in the agricultural field. Gilbels et al. (2005) used meta-analysis with MOs to investigate the influence of assessment on the effects of problem-based learning. In addition, Virués-Ortega (2010) used the approach in clinical trials to measure the comprehensive applied behaviour analytic intervention for young children with autism. Davis et al. in (2003) also performed MA with MOs for randomised efficacy trials computing second-generation antipsychotics and first-generation antipsychotics to reduce efficacy due to the use of a high-dose comparator. More recently, Doyle et al. (2019) used multivariate meta-analysis in trials of pharmacological, psychotherapeutic, and exercise care interventions for depressive symptoms in patients with coronary artery disease.

Identify systematic patterns in a MA, we propose four basic categories of study characteristics for classification and apply them in the agricultural field. The four categories of study characteristics, discussed below, are: environmental research, model specification, estimation methods and data reuse.

(1) **Environmental research**: This comprises the type of factors in the food production that could indicate the type of food production, looking at labour productivity and crops, with a strong linkage to human health, affecting the livelihoods of many inhabitants in SSA. This involves the name of authors, the year of publication of the studies, the categories of food such as food for human health, animal health, food security, food aid and the country or region of the experiment.

(2) **Model specification**: This contains variables in the food production model that have been stretched to embrace fertiliser, the type of cropping system, and farmer selection**.**

(3) **Estimation**: This is the effect of estimation methods on parameter estimates on the type of food production that needs to be investigated. Estimation procedures include statistical techniques such as ordinary least squares estimation (OLS), maximum likelihood (ML) procedures and others (nonlinear programming, numerical solution techniques) that might be used in the studies.

(4) **Effects included in the reuse of data**: The reuse of data reduces the reliability of a partial repeated measure design, as suggested by Pedhazur and Kerlinger (1982).

We present multiple linear regression analysis that uses correlations among the study characteristic variables that determine the mathematical expression by estimating the variables of interests. This will allow us to establish the fundamental relationships between each category of study characteristic and the factors involved. We use the multiple linear regression method to test and analyse the relations between the variables of interests. Agriculture plays a vital role in economic development in many developing countries, especially in the Sub-Saharan African (SSA) countries (Ogundari & Awokuse, 2016). An increase in agricultural productivity is the concern of many researchers, non-government organisations (NGOs), donors etc., to initiate greater food production output, which could result in lower food prices and greater food availability in SSA (Fuglie 2013; Awokuse et al., 2015). Food, and by extension nutrient intakes, has been found to have a strong empirical linkage with human health and labour productivity (Aromolaran, 2004) because food insecurity is highly correlated with the risk of infectious diseases and malnutrition (WHO, 2015). The World Bank (2007) reveals that the failure to realise the potential of SSA in agriculture has significantly compromised agriculture's role in reducing poverty and achieving food security. This chapter proposes a meta-analysis approach using MOs in the study where we use the multiple linear regression (MLR) to analyse MOs jointly within a single model. Multiple outcomes in MA have been applied in different fields, for instance, in education. We use an illustrative example to demonstrate the model characteristics with the categories of study in the SSA as shown in Table 6.1.

6.2 THE COMBINED APPROACH WITH MULTIPLE OUTCOMES

Multiple outcomes are integrated into the food production process simply because multiple outcomes will influence a range of responses depending on the number of variables within the subject. Factors such as fertiliser and crops are specifically involved in multiple outcomes. Reporting more than one outcome in factors such as crop and fertiliser in their experimental process is more appropriate simply because a single measure may not be enough to describe the effect of the treatment (or process). Food production involves a complex set of factors (known and unknown) that we are jointly synthesising in a single model with MOs measured at each of the factors. A joint model provides an overall test that is more powerful, realistic and useful than the separate reports of an individual model. The constructed approach is useful due to collecting several outcomes in the study that allow different analytical relations for analysis. This situation fits perfectly in the multivariate methods framework. The advantages of the combined approach are:

(1) The ability to convert the information collected from a variety of conditions into a potential distillation of important variables of the data.

(2) It allows an easy weighting of the variables since the outcomes are of a different nature and measured on different scales such as crop, fertiliser and trade.

(3) Most importantly, it allows testing for the overall treatment effect since a separate analysis does not offer such an estimate without further work and for each outcome, this creates the question of correcting the *p*-values for multiple comparisons (Bland & Altman, 1995; Pinto et al., 2009).

(4) It allows analysis of each outcome, generating unbiased estimates for treatments.

Analysing the outcomes separately do not require that the outcomes be measured on the same scale because the outcomes were not observed together (Pinto, et al., 2009). The data collected were the effect sizes (correlations) with no missing data observed.

We present a combined model that has the characteristics of a multivariate method to

1 analyse all the outcomes together by using the correlations as effect sizes.
2 allow combinations of different types of outcomes.
3 use real data as an example to demonstrate the effectiveness of the approach by showing the results in separate models compared to the combined approach.

6.3 CONSTRUCTING MULTIPLE LINEAR REGRESSION MODELS USING META-ANALYSIS

Food production carries multiple factors due to more than one feature that occur to help to account for variability among the experimental units. We use a scenario to construct the combined multiple regression model that determines the simplest relationship between the factors involved in the model. Thereafter, we then compare the treatments or populations using the MLR models. We assume that the model with correlated outcomes of factors of interest follows a multivariate normal distribution. The scenarios are based on the four characteristics-based study discussed earlier. The four scenarios can be described as follows:

Scenario 1 presents the multivariate structure of the data that analyses internal factors with MOs. In this section, we assume that there is a joint association between the outcomes and the covariates. Let us assume that m studies from the past and n factors that produced MOs are selected. Each study provides the correlations measures denoted by Ω_{ij} (i =1, 2, …, m and j = 1, 2, …, n) associated with the total correlation of the internal factors to food production treatment effect denoted by X_{ij} . The association between past studies and the internal factors can be defined by the following mathematical model:

$$\boldsymbol{X} = \boldsymbol{\Omega\beta} + \boldsymbol{\varepsilon}, \tag{6.1}$$

with

$$\mathbf{X} = \begin{pmatrix} X_{i1} \\ X_{i2} \\ X_{i3} \\ \vdots \\ X_{in} \end{pmatrix}, \quad \boldsymbol{\Omega} = \begin{pmatrix} 1 & \Omega_{11} & \Omega_{12} & \dots & \Omega_{1n} \\ 1 & \Omega_{21} & \Omega_{22} & \dots & \Omega_{2n} \\ 1 & \Omega_{31} & \Omega_{32} & \dots & \Omega_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \Omega_{m1} & \Omega_{m2} & \dots & \Omega_{mn} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_{00} \\ \beta_{i1} \\ \beta_{i2} \\ \vdots \\ \beta_{in} \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \vdots \\ \varepsilon_{in} \end{pmatrix},$$

where Ω_{ij} denotes the value of i^{th} study on the j^{th} internal factor experimental unit, β_{ij} denotes the slope of i^{th}study in the direction of the j^{th} the internal factor for treatment, and we assume that ε_{ij} are iid N(0, $\sigma^2_{\varepsilon_{ij}}$).

Scenario 2 is based on the multivariate structure of the data that analyses external factors with MOs. Similarly, we assume m past studies and n factors that provide MOs. Each study offers the correlation measures denoted by γ_{ij} $(i = 1, 2, \dots, m \ and \ j = 1, 2, \dots, n)$ associated with the total correlation of the external factors for food production treatment effect denoted by Y_{ij} . The following mathematical model can define the association between the past studies and the external factors:

$\boldsymbol{Y} = \boldsymbol{\gamma\alpha} + \boldsymbol{\xi},$ (6.2)

with

$$\boldsymbol{Y} = \begin{pmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \\ \vdots \\ Y_{in} \end{pmatrix}, \quad \boldsymbol{\gamma} = \begin{pmatrix} 1 & \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1n} \\ 1 & \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2n} \\ 1 & \gamma_{31} & \gamma_{32} & \cdots & \gamma_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \gamma_{m1} & \gamma_{m2} & \ldots & \gamma_{mn} \end{pmatrix}, \quad \boldsymbol{\alpha} = \begin{pmatrix} \alpha_{00} \\ \alpha_{i1} \\ \alpha_{i2} \\ \vdots \\ \alpha_{in} \end{pmatrix}, \quad \boldsymbol{\xi} = \begin{pmatrix} \xi_{i1} \\ \xi_{i2} \\ \xi_{i3} \\ \vdots \\ \xi_{in} \end{pmatrix},$$

where γ_{ij} denotes the value of i^{th} study on the j^{th} external factor experimental unit, α_{ij} denotes the slope of i^{th}study in the direction of the j^{th} external factor for treatment, and we assume that ξ_{ij} are iid N$(0, \sigma^2_{\xi_{ij}})$.

Scenario 3: presents the multivariate structure of the data that considers the statistical techniques used to estimate the parameters in the factors with MOs. Similarly, we assume that there is a joint association between the outcomes and the covariates. Let us assume that m studies from the past and n factors that produced MOs are selected. Each study provides the measures denoted by U_{ij} (i =1, 2, …, m and $\boldsymbol{j} = 1, 2, \ldots, n$) associated with the total of the estimated parameters on the factors of the food production treatment effect denoted by Z_{ij} . The following mathematical model can define the association between the total of the estimated parameters in past studies and the statistical methods used in the factors**:**

$\boldsymbol{Z} = \boldsymbol{U\delta} + \boldsymbol{\varsigma},$ (6.3)

with

$$\boldsymbol{Z} = \begin{pmatrix} Z_{i1} \\ Z_{i2} \\ Z_{i3} \\ \vdots \\ Z_{in} \end{pmatrix}, \quad \boldsymbol{U} = \begin{pmatrix} 1 & U_{11} & U_{12} & \ldots & U_{1n} \\ 1 & U_{21} & U_{22} & \ldots & U_{2n} \\ 1 & U_{31} & U_{32} & \ldots & U_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & U_{m1} & U_{m2} & \ldots & U_{mn} \end{pmatrix}, \quad \boldsymbol{\delta} = \begin{pmatrix} \delta_{00} \\ \delta_{i1} \\ \delta_{i2} \\ \vdots \\ \delta_{in} \end{pmatrix}, \quad \boldsymbol{\varsigma} = \begin{pmatrix} \varsigma_{i1} \\ \varsigma_{i2} \\ \varsigma_{i3} \\ \vdots \\ \varsigma_{in} \end{pmatrix},$$

where U_{ij} denotes the value of i^{th} study on the j^{th} statistical techniques used on the factor experimental unit, δ_{ij} denotes the slope of i^{th}study in the direction of the j^{th} statistical technique on the factor for treatment, and we assume that ς_{ij} are iid N(0, $\sigma^2_{\varsigma_{ij}}$).

Scenario 4: Similarly, the total effects of the reuse of data (R_{ij}) associated with the measure of the reuse data on the factors can be expressed as

$$\boldsymbol{R} = \boldsymbol{C\theta} + \boldsymbol{\zeta}, \tag{6.4}$$

with

$$\boldsymbol{R} = \begin{pmatrix} R_{i1} \\ R_{i2} \\ R_{i3} \\ \vdots \\ R_{in} \end{pmatrix}, \quad \boldsymbol{C} = \begin{pmatrix} 1 & C_{11} & C_{12} & \dots & C_{1n} \\ 1 & C_{21} & C_{22} & \dots & C_{2n} \\ 1 & C_{31} & C_{32} & \dots & C_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & C_{m1} & C_{m2} & \dots & C_{mn} \end{pmatrix}, \quad \boldsymbol{\theta} = \begin{pmatrix} \theta_{00} \\ \theta_{i1} \\ \theta_{i2} \\ \vdots \\ \theta_{in} \end{pmatrix}, \quad \boldsymbol{\zeta} = \begin{pmatrix} \zeta_{i1} \\ \zeta_{i2} \\ \zeta_{i3} \\ \vdots \\ \zeta_{in} \end{pmatrix}$$

where C_{ij} denotes the value of i^{th} study on the j^{th} reuse of data on the factor experimental unit, θ_{ij} denotes the slope of i^{th}study in the direction of the j^{th} reuse of data on the factor for treatment and we assume that ζ_{ij} are iid N(0, $\sigma^2_{\zeta_{ij}}$).

The combined multivariate model $\boldsymbol{F} = \boldsymbol{X} + \boldsymbol{Y} + \boldsymbol{Z} + \boldsymbol{R}$ is a joint model of Equation 6.1 to 6.4 given by

$$\boldsymbol{F} = \boldsymbol{\Omega\beta} + \boldsymbol{\gamma\alpha} + \boldsymbol{U\delta} + \boldsymbol{C\theta} + \boldsymbol{\tau} \tag{6.5}$$

The matrix is

$$\begin{pmatrix} F_{i1} \\ F_{i2} \\ F_{i3} \\ \vdots \\ F_{in} \end{pmatrix} = \begin{pmatrix} 1 & \Omega_{11} & \Omega_{12} & \Omega_{13} & \cdots & \Omega_{1n} & \cdots & C_{11} & C_{12} & C_{13} & \cdots & C_{1n} \\ 1 & \Omega_{21} & \Omega_{22} & \Omega_{23} & \cdots & \Omega_{2n} & \cdots & C_{21} & C_{22} & C_{23} & \cdots & C_{2n} \\ 1 & \Omega_{31} & \Omega_{32} & \Omega_{33} & \cdots & \Omega_{3n} & \cdots & C_{31} & C_{32} & C_{33} & \cdots & C_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & \Omega_{m1} & \Omega_{m2} & \Omega_{m3} & \cdots & \Omega_{mn} & \cdots & C_{m1} & C_{m2} & C_{m3} & \cdots & C_{mn} \end{pmatrix} \begin{pmatrix} \beta_{00} \\ \beta_{i1} \\ \beta_{i2} \\ \vdots \\ \beta_{in} \\ \vdots \\ \theta_{00} \\ \theta_{i1} \\ \theta_{i2} \\ \vdots \\ \theta_{in} \end{pmatrix} + \begin{pmatrix} \tau_{i1} \\ \tau_{i2} \\ \tau_{i3} \\ \vdots \\ \tau_{in} \end{pmatrix} \quad (6.6)$$

This is like $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{Y}$ is the data vector, $\boldsymbol{X}$ is the design matrix, $\boldsymbol{\beta}$ is the vector of parameters, and $\boldsymbol{\varepsilon}$ is the vector errors.

The GLS regression model is used in the fixed and random effects model since the approach provides the benefit of analysing multiple outcomes by combining them into a single model. In addition, the approach can explain the heterogeneity of the results among the factors better. If the heterogeneity does not hold, then the fixed model provides the estimated coefficients, the standard errors, and the *p*-values. Therefore, the findings obtained from the fixed effects model will not be used to make inferences to avoid misleading conclusions (Berkey et al., 1998).

The hypotheses are H_0 : Homogeneity effect is detected versus H_1: Heterogeneity effect is detected.

Reject H_0 if the Chi-square test Q is greater than the critical value at the degrees of freedom df = (k-1) or *p*-value < 0.05. Equation 6.5 represents the fixed effects model, and similarly, the random-effects model is

$$\boldsymbol{F} = \boldsymbol{\Omega\beta} + \boldsymbol{\gamma\alpha} + \boldsymbol{U\delta} + \boldsymbol{C\theta} + \boldsymbol{\upsilon} + \boldsymbol{\tau} \quad (6.7)$$

Based on Equation 6.5 and 6.7, the covariance of the random effects error is denoted by

Cov (υ_{ij}) = value D and the covariance of the random sampling error denoted by Cov (τ_{ij}) =S_{ij} ,

$E(\tau_{ij}) = E(\upsilon_{ij}) = 0$

The distribution τ_{ij}~MVN ($\boldsymbol{0}$;$\boldsymbol{S}_{ij}$) (6.8)

Because the vector $\boldsymbol{F}$ is a collection of random vectors, $\boldsymbol{F}_{i1}, \boldsymbol{F}_{i2}, \dots, \boldsymbol{F}_{in}$ are assumed independent and normally distributed, then $\boldsymbol{F}_{ij}$ ($i = 1, 2, \dots, m$ and $j = 1, 2, \dots, n$) is MVN distributed; therefore the distributions for $\boldsymbol{F_{ij}}$ for the fixed-effects and random-effects are given, such as the distribution $\boldsymbol{F}_{ij}$~MVN ($\boldsymbol{\Omega}_{\text{ij}}\boldsymbol{\beta} + \boldsymbol{\gamma}_{\text{ij}}\boldsymbol{\alpha} + \mathbf{U}_{\text{ij}}\boldsymbol{\delta} + \mathbf{C}_{\text{ij}}\boldsymbol{\theta}$; $\boldsymbol{S}_{ij}$) for the fixed-effects and the distribution $\boldsymbol{F}_{ij}$~MVN ($\boldsymbol{\Omega}_{\text{ij}}\boldsymbol{\beta} + \boldsymbol{\gamma}_{ij}\boldsymbol{\alpha} + \mathbf{U}_{\text{ij}}\boldsymbol{\delta} + \boldsymbol{C}_{ij}\boldsymbol{\theta}$; $D + \boldsymbol{S}_{ij}$) for the random-effects model. In the fixed-effects model, we assume that the off-diagonal elements of each $\boldsymbol{S}_{ij}$ are zero simply because the observations i and j are independent therefore, the matrix variance-covariance is

$$\boldsymbol{S}_{ij} = \begin{pmatrix} S_{11} & 0 & 0 & \cdots & 0 \\ 0 & S_{22} & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots & \vdots \\ 0 & \vdots & 0 & \ddots & 0 \\ 0 & 0 & 0 & \dots & S_{mn} \end{pmatrix} \tag{6.9}$$

6.4 ESTIMATION OF PARAMETERS IN THE COMBINED MODEL

The estimation of parameters is performed through least square regression to minimise the sum of squared residuals. We are comparing each independent variable to the dependent variable, and more especially because the variables are measured with different scales. The technique requires standardising the coefficients in the regression model using the Z-scores. The normal scaling entails subtracting the sample mean and dividing by the standard deviation from Equation 6.3 as follows

$$\mathbf{F}_{\text{ij}} = \frac{\text{F}_{\text{ij}} - \overline{\text{F}}}{\text{S}_{\text{F}_{ij}}},\ \Omega_{\text{ij}} = \frac{\Omega_{\text{ij}} - \overline{\Omega}_{ij}}{\text{S}_{\Omega_{\text{ij}}}},\ \boldsymbol{\gamma}_{\text{ij}} = \frac{\gamma_{\text{ij}} - \overline{\gamma}_{ij}}{\text{S}_{\gamma_{\text{ij}}}},\ \mathbf{U}_{\text{ij}} = \frac{\text{U}_{\text{ij}} - \overline{\text{U}}_{ij}}{\text{S}_{\text{s}_{\text{ij}}}} \text{ and } \mathbf{C}_{\text{ij}} = \frac{\text{C}_{\text{ij}} - \overline{\text{C}}_{ij}}{\text{S}_{\text{c}_{\text{ij}}}}$$

where $S_{\boldsymbol{F_{ij}}}$, $S_{\Omega_{ij}}$,$S_{\gamma_{ij}}$, $S_{S_{ij}}$ and $S_{C_{ij}}$ represent the estimated sample standard deviation of the dependent variable and the independent variables in the model. Using the standardised formulas in Equation 6.5, the model becomes

$$\frac{F_{ij}-\bar{F}}{\sigma_F} = \frac{\tau_{ij}-\bar{\tau}}{\sigma_\tau} + \beta_{i1}\frac{\Omega_{i1}-\bar{\Omega}}{\sigma_{\Omega i1}} + \beta_{i2}\frac{\Omega_{i2}-\bar{\Omega}}{\sigma_{\Omega i2}} + \ldots + \Omega_{in}\frac{\Omega_{,ni}-\bar{\Omega}}{\sigma_{\Omega in}} + \alpha_{i2}\frac{\gamma_{ij}-\bar{\gamma}}{\sigma_{\gamma i2}}$$

$$+\ldots+ \gamma_{in}\frac{y_{in}-\bar{y}}{\sigma_{\gamma in}} + \delta_{1i}\frac{U_{i1}-\bar{U}}{\sigma_{Ui1}} + \ldots + \delta_{in}\frac{U_{in}-\bar{U}}{\sigma_{Uin}} + \theta_{i1}\frac{C_{i1}-\bar{C}}{\sigma_{Ci1}} + \ldots + \theta_{in}\frac{C_{in}-\bar{C}}{\sigma_{Cin}} \qquad (6.10)$$

Equation 6.10 can be written as

$$\boldsymbol{F}^{*}_{ij} = \tau^{*} + \beta_{i1}\,\boldsymbol{\Omega}^{*}_{i1} + \ldots + \beta_{in}\,\boldsymbol{\Omega}^{*}_{in} + \alpha_{i1}\,\boldsymbol{\gamma}^{*}_{i1} + \ldots + \alpha_{in}\,\boldsymbol{\gamma}^{*}_{in} + \delta_{i1}\,\boldsymbol{U}^{*}_{i1} + \ldots + \delta_{in}$$

$$\boldsymbol{U}^{*}_{in} + \theta_{i1}\,\boldsymbol{C}^{*}_{i1} + \cdots + \theta_{in}\,\boldsymbol{C}^{*}_{in} + \boldsymbol{\tau}^{*}_{in} \qquad (6.11)$$

where $\boldsymbol{F}^{*}_{ij}$,$\boldsymbol{\Omega}^{*}_{ij}$, $\boldsymbol{\gamma}^{*}_{ij}$, $\boldsymbol{U}^{*}_{ij}$ and $\boldsymbol{C}^{*}_{ij}$ are the standardised food production for the dependent variable and the independent variables respectively. In addition, β_{ij}, $\alpha_{ij,}$ δ_{ij} and θ_{ij} represent the parameters to estimate and $\boldsymbol{\tau}^{*}_{ij}$ represent the error term. Using the least-squares method to minimise the sum of squared residuals equal to the form $\sum_{i=1}^{n} e^2{}_i = \sum_{i=1}^{n}\left(y_i - \mu^* - \sum_{i=1}^{n}\beta_{ij}X_{ij}\right)^2$ and taking the derivative with respect to the model parameters β_{i1} ,β_{i2}, β_{i3},... and β_{in} set to zero, the expression becomes $n\,\mu^* + \hat{\beta}_{i1}\sum_{i=1}^{n}X^*{}_{i1} + \hat{\beta}_{i2}\sum_{i=1}^{n}X^*{}_{i2} + \ldots + \hat{\beta}_{in}\sum_{i=1}^{n}X^*{}_{in} = \sum_{i=1}^{n}y^*{}_i$

For each factor, the simpler form, using vector and matrix notations, is

$$\boldsymbol{Y}^* = \begin{pmatrix} y^*_1 \\ y^*_2 \\ y^*_3 \\ \vdots \\ y^*_m \end{pmatrix}, \quad \boldsymbol{X}^* = \begin{pmatrix} 1 & X^*_{11} & X^*_{12} & X^*_{13} & \dots & X^*_{1n} \\ 1 & X^*_{21} & X^*_{22} & X^*_{23} & \dots & X^*_{2n} \\ 1 & X^*_{31} & X^*_{32} & X^*_{33} & \dots & X^*_{3n} \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & X^*_{m1} & X^*_{m2} & X^*_{m3} & \dots & X^*_{mn} \end{pmatrix}, \quad \boldsymbol{\beta}^* = \begin{pmatrix} \mu^* \\ \beta_{i1} \\ \beta_{i2} \\ \vdots \\ \beta_{mn} \end{pmatrix}, \tag{6.12}$$

$$\boldsymbol{\varepsilon}^* = \begin{pmatrix} \varepsilon^*_1 \\ \varepsilon^*_2 \\ \varepsilon^*_3 \\ \vdots \\ \varepsilon^*_m \end{pmatrix},$$

The linear regression model is denoted as $\boldsymbol{Y}^* = \boldsymbol{X}^* \boldsymbol{\beta}^* + \boldsymbol{\varepsilon}^*$and the least-squares parameter estimates $\boldsymbol{\beta}^*$ are the vectors that minimise:

$$\sum_{i=1}^{n} e^2{}_i = \varepsilon^{*\prime}\varepsilon^* = (y - X^* \beta^*)'(y - X^* \beta^*) \tag{6.13}$$

After determining the derivative of Equation 6.13 with respect to mixed effects β^*, the vector of the parameters is

$$\widehat{\boldsymbol{\beta}^*} = (\boldsymbol{X}^{*\prime}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*\prime}\boldsymbol{Y}^* \tag{6.14}$$

The ordinary least squares (OLS) technique is used to estimate the parameters for the fixed and random effects models for multiple outcomes. We aim to determine in the analysis which factor in each model describes the mean effects of the dependent variable and test each model's fit. The null hypothesis is that all the estimated parameters are equal.

6.5 DATA ANALYSIS

6.5.1 Introduction

Each study of food production reported the results about the environmental research for crops and the type of specification for fertiliser with three outcomes that have been measured on each of eight African countries. Meta-analysis in agriculture performs a single synthesis for the results of each outcome. In this study, we present a simultaneous analysis that uses the correlations as the effect sizes, and the MLR model is employed to estimate parameters. The data are averages from the published articles and the effect sizes are the correlations from each selected country for each factor respectively. Three outcomes are assessed from the environmental research category (maize, vegetables and fruit) and two from the model specification category (potassium and urea) in the classification of the study as indicated in Table 6.1.

Table 6.1: The correlation data for the effect sizes and in bracket are the standard error for the correlations

Country	Classification of the study						
	Environmental Research				Model Specification		
	Maize	Vegies	Fruits	FoodProd	Potas	Urea	Prod
Burk. F	0.8013 (0.4088)	0.2773 (0.1415)	0.0970 (0.0495)	0.9471 (0.4832)	0.4276 (0.2182)	0.3915 (0.1997)	0.5638 (0.288)
Camer	0.8019 (0.4091)	0.2811 (0.1434)	0.9530 (0.4862)	0.9528 (0.4861)	0.2091 (0.1067)	0.2874 (0.1466)	0.5022 (0.256)
Cote Iv	0.8117 (0.4141)	0.4802 (0.2450)	0.1084 (0.0553)	0.9527 (0.4861)	0.2091 (0.1067)	0.2874 (0.1466)	0.5022 (0.256)
Gabon	0.8547 (0.4361)	0.4730 (0.2413)	0.1321 (0.0674)	0.9570 (0.4883)	0.4297 (0.2192)	0.5351 (0.2730)	0.6660 (0.339)
Ghana	0.8821 (0.4501)	0.3619 (0.1846)	0.8840 (0.4510)	0.9520 (0.4857)	0.3464 (0.1767)	0.5344 (0.2727)	0.4255 (0.425)
Kenya	0.8996 (0.4590)	0.2198 (0.1121)	0.9086 (0.1636)	0.9556 (0.4876)	0.4768 (0.2433)	0.4245 (0.2166)	0.5661 (0.289)
Sou. Afr	0.8984 (0.4584)	0.1829 (0.0933)	0.9144 (0.4665)	0.9858 (0.5030)	0.4994 (0.2548)	0.4871 (0.2485)	0.4015 (0.205)
Senegal	0.1190 (0.0807)	0.4372 (0.2231)	0.8032 (0.4098)	0.9634 (0.4915)	0.7071 (0.3608)	0.7071 (0.3608)	0.7071 (0.361)

Burk. F.; Burkina Faso; Camer: Cameroon; Cote Iv.: Ivory Coast; Sou. Afr: South Africa.
FoodProd: food production

The data are the effect sizes called correlations. In Table 6.2, we present the single and the combined approach for testing the estimated parameter. The multivariate test provides the overall treatment, while individual tests raise the issue of adjusting the *p*-values for multiple comparisons (Pinto et al., 2009; Bland et al., 1995). Equation 6.1 and 6.2 are used in the model for estimation. The ANOVA tests the hypothesis that the factor model is adequate for the data, as shown below.

Table 6.2: Model estimation using the ANOVA procedure

Model		Sources	Sum of Square	Df	Mean Squares	F	*p*-value
Separate Model	Environmental Research	Regression	5.598	3	1.666	5.322	0.070
		Residual	1.402	4	0.351		
		Total	7.000				
	Model Specification	Regression	3.095	2	1.548	1.981	0.232
		Residual	3.905	5	0.781		
		Total	7.000				
Combine Model	Multivariate Model	Regression	0.163	5	0.033	1.642	0.420
		Residual	0.040	2	0.020		
		Total	0.202	7			

Since the *p*-value is greater than the 0.05 level of significance, it implies that the model and the data are statistically equivalent. We will calculate the fixed and the random effects from each model as given in Table 6.3 for the single and combined approach.

Table 6.3: Effect sizes for calculations of fixed and random effects

Environmental research					Model specification			
Study	ES	SE	W	$W \times ES$	ES	SE	W	$W \times ES$
1	-0.119	-0.014	5110.12	-606.57	0.428	0.07566	174.688	74.766
2	0.9881	0.1165	73.7447	72.8671	-1.8914	-0.3344	8.9451	-16.91
3	-2.416	-0.285	12.3350	-29.801	1.3057	0.2308	18.7700	24.507
4	3.0154	0.3554	7.9185	23.8774	1.3955	0.2467	16.432	22.930
5	2.5839	0.3045	10.7840	27.8649	-1.0334	-0.1825	30.0230	-30.99
6	-2.722	-0.321	9.7154	-26.448	0.3454	0.0611	268.229	92.646
7	2.1254	0.2505	15.9386	33.8760	0.1137	0.0201	2475.31	281.44
8	-0.525	-0.062	261.424	-137.20	-0.1773	-0.0313	1017.96	-180.4
Total			5501.98	641.53			4010.36	267.89

ES: Effect sizes SE: Standard Error W: weights for Meta-analysis

The tests to estimate the model for the fixed and random effects regression models are as indicated below.

6.5.2 Fixed Effects for Single Model

The following test examines the homogeneity of the ESs. If the homogeneity holds according to the testing, then all the ESs are estimating the same population, meaning that the fixed effects model is applicable.

Table 6.4: Fixed effects calculations for single model

Model	SE.Mean	M.ES $\overline{ES}$	Z-test	95 % CI of ES	Q	P-value
Environmental Research	0.014	0.12	0.0668	(0.09; 0.14)	501.1977	0.0001
Model Specification	0.0158	0.067	4.2278	(0.051; 0.083)	238.1045	0.0024

SE: Standard error, SE.Mean = SE of the mean of the effect sizes, CI of ES = Confidence interval of effect sizes, Q: The calculated test statistic value for heterogeneity.

The results are significant (p-value < 0.05), which implies that not all ESs estimate the same population. This means that the distribution is heterogeneous; therefore, the single mean of the ESs is not suitable to describe the distribution. That means the studies are estimating different populations and mean of ESs. The Q test statistic is greater than the critical value (14.10); thus, these results confirmed the previous test. The Q-test statistics are distributed as per the Chi-squared with the degrees of freedom (df = 7), representing the number of ESs minus 1. The findings indicate that all the models are significant. That is, the variability across the ESs does exceed what could be expected based on the sampling error.

6.5.3 Random Effects Model for Single Model

The random-effects model is used under the assumption that the variability between effects sizes is due to the sampling error.

Table 6.5: Fixed effects calculations for single model

Environmental research					Model specification			
Study	ES	W	$W \times ES$	W^2	ES	W	$W \times ES$	W^2
1	-0.119	5110.	-606.57	2611332	0.428	174.688	74.7664	30515
2	0.988	73.74	72.8671	5438.277	-1.8914	8.9451	-16.919	80.014
3	-2.416	12.33	-29801	152.1519	1.3057	18.7700	24.5079	352.31
4	3.015	7.918	23.8774	62.7026	1.3955	16.432	22.9309	270.01
5	2.583	10.78	27.8649	116.2953	-1.0334	30.0230	-30.996	901.38
6	-2.722	9.715	-26.448	94.3889	0.3454	268.229	92.6462	71946.
7	2.125	15.93	33.8760	254.0401	0.1137	2475.31	281.442	612714
8	-0.525	261.4	-137.20	68342.31	-0.1773	1017.96	-180.49	103625
Total			641.53	2618778			267.894	726746

ES: Effect sizes; SE: Standard Error; W: weights for Meta-analysis

The following tests examine the heterogeneity of the ESs. If the heterogeneity is shown to hold by testing the null hypothesis, then all the ES estimates vary from one population mean to the next;

therefore, the random-effects model is applicable. The weight for each study is $W_i = \frac{1}{SE_i{}^2 + \hat{\Delta}_\theta}$

where $\hat{\Delta}_\theta = \frac{Q_i-(k-1)}{\sum W - \left(\frac{\sum W^2}{\sum W}\right)}$, the random effects variance component. The calculated value of $\hat{\Delta}_\theta$ and

W_i are as follows:

For environmental research: $\hat{\Delta}_\theta$= 0.6658 and W_i = 1.5015.

For specification model: the weight for each study is $W_i = \frac{1}{SE_i{}^2+\hat{\Delta}_\theta}$ where $\hat{\Delta}_\theta$, the random effects variance component is $\hat{\Delta}_\theta$ = 0.1051 and W_i = 9.4922

Table 6.6: Random effects calculations for single model:

Model	SE.Mean	M.ES $\overline{ES}$	Z-test	95 % CI of ES	P-value
ER	$SE_{\overline{ES}}$ = 0.8161	$\overline{ES}$ = 0.1166	0.1429	(-1.483; 1.716)	0.9999
MS	$SE_{\overline{ES}}$ = 0.1053	$\overline{ES}$ = 0.0668	0.6344	(-0.139; 0.273)	0.9888

SE: Standard error, SE.Mean = SE of the mean of effect sizes, M.ES = Mean effect sizes, CI of ES = Confidence interval of effect sizes, Q: The calculated test statistic value for heterogeneity, ER – Environmental Research, MS = Model Specification

We fail to reject the null hypothesis because there is a presence of homogeneity effects, therefore the variability across the effects sizes is not going beyond what could be expected based on the sampling error.

6.5.4 Fixed and Random Effects Model with Combined Multiple Outcomes

The combined approach in MA is used through the fixed and the random-effects models by combining information across studies to test the heterogeneity effects before calculating the parameters that represent the effects of particular interest. The aim is to describe the relations between the environmental research and model specification level. Analysing heterogeneity through a combined model when using fixed- and random-effects is shown in Table 6.7.

Table 6.7: Fixed and random effects for the combined model

Study	ES	SE	W	$W \times ES$
1	0.3093	0.0210	2257.844	698.3511
2	-0.9033	-0.0615	264.7218	-239.123
3	-1.1103	-0.0756	175.2157	-194.542
4	4.4109	0.3001	11.10195	48.9696
5	1.5515	0.1056	89.7326	139.2201
6	-2.3769	-0.1617	38.2324	-90.8747
7	2.2391	0.1524	43.0831	96.4673
8	-0.7021	-0.0478	438.1833	-307.648
Total			3318.115	150.8198

ES: Effect sizes; SE: Standard error; W: weights for meta-analysis

The test for homogeneity and heterogeneity of the ESs is given in Table 6.8. If the homogeneity holds as shown by testing, that means that all the ESs estimate the same population mean, therefore the fixed effects model is applicable.

Table 6.8: Fixed and random effects calculations for the combined model

Model	SE.Mean	M.ES $\overline{ES}$	Z-test	95 % CI of ES
Fixed effects	$SE_{\overline{ES}}$ = 0.0174	$\overline{ES}$ = 0.0003	0.0172	(-0.0334; 0.0344)
Random effects	$SE_{\overline{ES}}$= 1.0073	$\overline{ES}$ = 0.0455	0.0452	(-1.9288; 2.0198)

SE: Standard error, SE.M = SE of the mean, M.ES = Mean effect sizes, Q: The calculated test statistic value for heterogeneity

In the random-effects model, the weight for each study is $W_i = \frac{1}{SE_i^2 + \hat{\Delta}_\theta}$ and the random effects variance component $\hat{\Delta}_\theta$ = 1.0143. The fixed and random effects model in the combined approach fails to reject the null hypothesis of homogeneity of effects since the true effect size is the same in all the studies.

The comparison between the single and the combined models using the OLS procedure is estimated as given in Table 6.9.

Table 6.9: Comparison between the single approach and combined approach

		Single model				Combined model			
Study	Outcome	Coeff	SE	*p*-v	t	Coeff	SE	*p*-v	t
ER	Maize	-0.44	0.262	0.173	-1.66	-0.395	0.088	0.525	-0.763
	Vegies	-0.94	0.281	0.029	-3.35	-0.852	0.091	0.253	-1.588
	Fruits	0.050	0.262	0.859	0.19	-0.024	0.087	0.967	-0.046
MS	Potas	0.621	0.634	0.372	0.980	0.033	0.139	0.971	0.040
	Urea	0.051	0.634	0.939	0.080	0.351	0.144	0.718	0.415
Model fit		ER: Single model R^2= 0.800 *p*-value= 0.070 SE: 0.5921 F: 5.322 Residual variance: 0.5715 MS: Single model R^2= 0.442 *P*-value= 0.232 SE : 0.884 F : 1.981 Residual variance: 0.7140				Combined model: R^2= 0.804 *p*-value= 0.420 SE: 0.141 F: 1.642 Residual variance: 0.2862			

ER: Environmental research; MS: Model specification; SE: Standard error; Coeff: Coefficient, *pv* = *p-value*

The coefficient of determination estimates the goodness-of-fit R^2 values for all the models as the results indicated that the models fit the data well. That is, the regression equations exhibit the associations among the observed values well. We suggest using the combined multiple outcomes model instead of the single univariate models because the combined model is more economical due to its multivariate property of borrowing the strength of the correlations between variables of interests. The smaller residual variance (0.2862) obtained from the combined model is better than the results of the single outcome models. The single model is Y = -0.44 maize – 0.94 vegies + 0.05 fruits + 0.621 potassium + 0.051 urea and the combined model is Y = -0.395 maize – 0.852 vegies – 0.024 fruits + 0.033 potassium + 0.351 urea. The F-values tests, equal to 5.322, 1.981 and 1.642 respectively, and all the *p*-values are larger; therefore, no differences between the population means were found. These results are consistent with the above tests indicating significant linear correlation between the dependent variables and the independent variables. Because all our regression models parameters were insignificant, we noted to refrain from interpreting the parameters, but we suggest increasing the number of variables in the models for suitable regression models.

CHAPTER 7

DISCUSSION

7.1 INTRODUCTION

The soil cultivation practices, water scarcity and climate change have led the SSA countries to a loss of soil fertility simply because these factors negatively influence the results of agricultural food production; therefore, they have harmfully affected the lives of people due to the political instability and natural disasters in the SSA. Farmers who have adopted technologies had sufficiently improved their productivity by raising their economic growth and enhancing the conditions of the people. The development of these entities is likely to increase food security and to alleviate poverty, therefore advancing the trade and the economy of the country. The results indicate a positive significant relationship between the adoption of new technology practices and the increase of food production, thus, improving lives and the social wellbeing of the people. In applying MA and SEM approaches, the results indicated that the use of hybrid seed, fertiliser, irrigation, and new equipment were some of the agricultural products in the modern process of food production. By adopting this approach, farmers are likely to change their agricultural food production activities and therefore address the challenges facing the SSA countries.

This study develops a procedure of combining the research outputs produced by different research collections to establish accurate parameter estimates using various statistical techniques to address the challenges facing agricultural research food production in SSA countries. These techniques are:

1 Meta-analysis (MA).

2 Structural equation modelling (SEM) under the conditions of factor analysis and principal component analysis (PCA).

3 An approach that combines MA and SEM.

4 An approach that determines the performance of the genotype across locations by testing the stability statistics; and

5 The use of multiple linear regression to analyse various studies involving multiple outcomes within a single model.

This work demonstrates how the above techniques were used to address the research questions based on the evaluation of suitable and stable parameter estimates on the practicability of existing techniques. These techniques were able to estimate the parameters based on the fixed, random, and mixed-effects model accurately. The models were more reliable than the existing approaches simply because the usual models are more restrictive under some conditions. To apply the present approach, the researcher has provided guidelines for each technique to calculate the suitability of the parameter estimates as presented below

Meta-analysis is a statistical technique that enables the researcher to improve the definition and analysis of the data from the past studies, therefore, getting a better estimate of the effect sizes. The application of MA requires underlining the sample size of the past studies, the standard errors and applying the eligibility criteria for selecting studies for obtaining the effect sizes that are inputs in the analysis. The choice of these elements has been inconsistent in some instances and has therefore greatly affected the estimation parameters and could identify the sources of heterogeneity between studies difficult. The present approach is more consistent and understandable because of the use of original data to generate the effect sizes that are the correlation coefficients used in a multivariate meta-analysis. Since the correlations are obtained from independent studies and conducted by multiple researchers, the measures of the variables of interest are likely to be different. Therefore, the new approach has standardised the correlations before pooling them together to form a pooled correlation matrix using the multivariate meta-analysis procedure. We have used the pooled correlation matrix to test various mixed theoretical models. The illustrative example had shown how multivariate meta-analysis could be applied in agricultural studies and related fields simply because the model was reliable with the specific data.

This technique increases the statistical power since it has improved the estimates of the effect sizes and resolved any uncertainty of the summary statistics collected. In addition, the proposed approach provides an educational methodology for obtaining reliable effect sizes to the researcher, thus, giving both consistent effect sizes from one study to the next and providing a better understanding of the entire process, for example, in the process of food production for subsistence farming. It is expected that the use of MA by integrating the findings from multiple studies will assist subsistence agriculture, for example, to obtain the best assessment of the food production in dealing with multiple results from past studies. The summarised description of the improved adaptive techniques approach enabled us to draw important conclusions and extend the univariate model to a multivariate MA concept.

The second substantive result is in SEM, where the traditional approach is more restrictive in how inter-correlations are defined among the factors. The proposed structure is more suitable, flexible, and accurate in determining the number of linear regressions that explain the effects of the endogenous structure upon the exogenous structure. The theoretical structure is now clearly and precisely defined based on the observed data with the help of the guidelines because a factor analysis procedure was used through the PCA. This principle mainly involved examining the factor loadings with the highest loadings (values above 0.5 or equal). It can be stated with confidence that the researcher would not be constrained when the guidelines are followed. In addition, using all the variables under factor analysis for the analysis and rotating the correlation matrix specifies the highest component together with the nature of the exogenous and endogenous components. On the other hand, the traditional approach uses huge diversity in a set of relationships that points to inconsistent conclusions when the model should be truly equivalent to the observed data.

Using SEM in estimating the parameters, the current approach in SEM involves various techniques to estimate parameters in the model; for instance, the maximum likelihood is usually determined using computer programs. This approach requires assumptions of either multivariate normality or generalised least squares for robust estimates. In this study, we propose the use of SEM by applying Jöreson’s approach in linear structural relations (LISREL) notation as presented by Bentler and Weeks (1980), by establishing the observed inter-relationship among the variables of interest. We introduce the concept of meta-analysis using the covariance matrix procedure to provide the parameter estimates for linear equation models. The technique yields more advantages in reducing the deviation between the observed and the proposed model. The selected inter-relationship between the variables of interest was obtained by using factor analysis through the principal component analysis in a more reliable and concise approach. These models are then tested for goodness-of-fit statistical tests to explain the discrepancy between latent and unobserved variables. The illustrative example indicated that the model fits the data well. This signifies the power of the approach implying that no important paths (or inter-relationship) among the variables have been omitted from the model. We further extended the confirmation of the improved approach with a simulation study. Simulation enables us to examine the performance of the test’s statistics and the interval estimation of the observed parameters. The tests statistics include points estimates, standard errors, covariance with charts and graphs. In applying simulation, we can understand the solution obtained for a better picture of the decisions made. In this work, we test the performance of the improved approach in the SEM. These findings provided perfect evidence that the proposed approach can be recommended for analysis to estimate parameters because of its clear and good statistical properties. The proposed model can deal with a larger and more complex situation.

The third practical result is the integration approach of MA into SEM. The two techniques are different multivariate statistical techniques based on their assumptions, models, and respective methodology approaches. Since the two techniques use correlations or covariance matrices as inputs, we can integrate the MA into SEM. Meta-analysis enables the combination of data from past independent studies by drawing the overall conclusions. SEM examines and tests all the relationships together in the model. We have proposed a suitable way of testing complex theories involving multiple variables that cannot be measured. The guidelines show how to:

(1) calculate the effect sizes through the standardised data.

(2) conduct the test of homogeneity or the heterogeneity of the effects.

(3) form the pooled correlation matrix to fit the SEM.

Researchers can combine correlations from independent studies and test the homogeneity or the heterogeneity of the correlations to fit a structural equation model. The traditional approach worked with the small samples due to the eligibility criteria in selecting the inclusion of the studies. This produces biases in the weighting of the studies, and therefore, the samples are not well represented. In particular, the reliability of the results inevitably became a problem. Instead, the proposed approach covers all possible dimensions by using the effect sizes (correlations). This technique calculates the variance-covariance matrix from the standardised data using the PCA on the standardised data that is equivalent to the PCA on the use of the correlation matrix. The benefit of this technique is in integrating MA into SEM by using all the effects of the factors simultaneously in the single model. The confirmatory factor analysis had enabled us to determine the measurement of the model with precision based on the factors measured by the observed variables. The agricultural food production was used to illustrate the approach. The findings indicate that the set of correlations were closer to the population correlation. The provided guidelines present step-by-step instructions to the user who wants to apply the technique. Formulas are offered for consistent correlation matrices of the combined effect sizes.

The study's fifth substantive conclusion concerns the findings identifying the performance of stable genotypes unaffected by environmental conditions. This approach offers a potential improvement of the methods applied to evaluate the stability with a better understanding and classification of the responsive locations.

The sixth conclusion concerns integrating the mixed-effects model into MA with multiple outcomes within a single model. The proposed approach with a multivariate structure model of MA is more comprehensive than in the univariate model. The advantages of the proposed model go beyond just explaining the observed heterogeneity of the results. The data used in MA are the correlations, and the results show that:

1) When the test for homogeneity holds, both the conventional and the proposed approaches are unbiased because the underlying effects are equal. That is, the effects measured by individual studies are sufficiently comparable to certify their combinations.

2) Irrespective of the approach used, the number of studies employed in the analysis plays an important role in the precision and accuracy of the estimated parameters each time heterogeneity was present.

3) The findings indicate how the newer approach works on simulation data simply because the means of the parameter estimates were unbiased for homogeneity or had a small degree of heterogeneity. The test statistic for the parameter estimates was equitably estimated. The

simulation study confirms the consistent pattern of the proposed model by offering similar results that have indicated no serious issues.

7.2 CONCLUSION, LIMITATION AND CHALLENGES

Agriculture is a primary sector that contributes to the welfare of SSA by providing employment, livelihood, and raw materials. Any changes in the agricultural food production directly or indirectly affect the country's economy based on its significant share in many sectors of the country. Hence, rapid development and transformation in agricultural research on food production and related fields is urgently needed. This worth of these resolutions requires the government to understand the role of agriculture and its important relationships to a country's economy growth. The nature of the relationships between the variables involved in agricultural food production influence each other affecting the growth of the agriculture sector. This explains the causal relationship that exists between the modern agriculture and the increase in food production.

Growing social services and the infrastructure are recommended since these factors are expected to raise the living standard of Africans, given that more commitments are provided by the government and stakeholders in agriculture. The move of the agricultural food production in SSA to the industrial development of agricultural food production is likely to increase the productivity of the agriculture products in the SSA. African countries must promote competitiveness among the markets across different food production processes simply because agricultural products have to be standardised to meet the market obligation. African government and independent producer organisation must take collective actions to address the issue such as agricultural prices, infrastructure for investment, the security measures to access the land, the protection of the natural resource, and the promotion of substantial farming for small scale farmers. The contribution of these entities and the use of new technology are deemed to increase the productivity in agricultural food production and, as a result, address the challenges facing the SSA.

Agricultural research on food production in SSA should be prioritised in the development agenda as government adopts new strategies to attract more foreign investments into its industrial development. The achievement of such a plan requires a review of policies by the government and its policymakers. This study introduced various improved approaches over the conventional techniques. The illustrative example of each technique presents a methodological approach that will help researchers in applying the methods. Readers with the necessary knowledge of multivariate statistics can use these techniques in their respective studies because the examples providing straightforward applications can address more complex topics. Statistical formulas and the step-by-step guidelines had made it possible to test the theoretical models. The use of univariate to multivariate meta-analysis holds great potential for addressing the study's research questions in a single model. The traditional approaches had some limitations that raised concerns about the need for improvement. Meta-analysis helped us to test and establish parameter estimates by synthesising findings from past studies.

Traditionally, structural equation modelling had two stages, namely the measurement and the structural process. The structural stage defines the causal relations between the variables, and the measurement stage determines the latent variables to be used in the model. These stages must be well-defined by the researcher. That is, the researcher determines which variable is related to which latent variable. In some instances, such situations are unpredictable, and therefore, more uncertainty can be developed, thereby causing an unreliable conclusion.

The proposed approaches are multivariate with a methodology for conducting meta-analysis structural equation modelling and mixed linear model with multiple outcomes. These techniques were developed to overcome the limitations of the univariate or single model when estimating parameters and drawing inferences. The results of the techniques have provided both meaningful information and a better understanding of the agricultural and related fields for subsistence farming in SSA. The improved approaches have now been introduced in the agricultural field. This field is a new area accessible to applied research. The techniques used are multivariate statistical techniques for making decisions, but they also have their own strengths and limitations. The best method should be the one addressing the purpose of the research. Despite good confirmation of these approaches, requirements for improvement and important gaps remain. Closing the current gaps in food production yield represent the greatest challenges and uncertainties facing SSA. The outcome of this study contributes to science in terms of the methodological approach of each technique and support through guidelines to estimate parameters and draw the inferences. This addresses the gaps that were identified between the existing methods and the developed approaches. In addition, this work contributes to the development of suitable food production.

Single and combined models were used based on the fixed and random effects models in the MA. We present an approach that analyses multiple individuals' models simultaneously into a single model. The analysis begins by standardising the effect sizes (correlations) followed by performing the multiple linear regression meta-analysis procedure to estimate parameters of interest both for single and combined multivariate models. Both fixed- and random-effects were computed, as the two approaches were necessary. We have found that combined multivariate MA presented more potential ability above MA in the single approach. The combined approach made a valid contribution to MA technique due to its power to describe the relationships between the estimates of effects since these estimates provided better statistical properties. That is, the combined model is based on the borrowing of strength compared to the single model. The utilisation of the combined correlation model had allowed the appropriate calculations of a combined confidence interval around the pooled results of environmental research and model specification categories. It also allowed for joint prediction values for the actual environmental research and model specification rather than a single study setting. The more we increase the correlation by combining the results from the single model, the more we obtain additional precise estimates.

The most significant benefit of the combined approach is obtaining data for all effects within a single model approach and getting hold of the parameter estimates simultaneously in a single analysis rather than doing a separate analysis. This work explored the feasibility of linking all the classification studies discovered in a single combined model. We have disclosed the practicability and value of further research by using all the classifications of study and test the effect in the fixed and random effects models for multiple linear regression meta-analysis. This kind of methodological approach is rarely used in agricultural practice.

The limitations and challenges of the techniques

In MA, this technique combines and summarises past studies. The methodology of summarising large amount of information requires that a single number represent such information. This process might ignore the fact, for instance, that the treatment effects vary from one study to the next. We also suggest that more efforts should be made during the identification of the relevant studies because the results in MA depend only on the studies included. In SEM, this technique allows the researcher to establish a relationship between variables, including a complexity process case that can offer multiple parameters. These parameters must correspond to various hypotheses that are evaluated simultaneously. Under some conditions, the model identification does not allow the estimation of more model parameters simply because the testing of the hypothesis requires that there are fewer parameters to be estimated than there are individual observed correlations/covariance.

7.3 FUTURE WORK

1) Every univariate MA application can finally be amenable to a multivariate approach because the multivariate approach is likely to make a genuine contribution to MA except if the single effect is considered sufficient.

2) In the improved model in the SEM, using factor analysis through the PCA, the simulation study tested one factor among the ten components created with the PCA. Using a simulation study, we have succeeded to generate sixteen independent samples from Factor 1 due to the large size of the covariance matrix that was processed on the computer. Future studies should develop a way to incorporate all these factors and test for both the highest loadings and the proportion of variance.

3) From all the constructed models in the current study, we did not consider the missing values in the development of the models. Future studies should consider developing the model in case of missing data, both in the effect sizes and in the model, because the presence of missing values may bring some statistical challenges for the proposed models.

4) Research on agricultural food production with adoption to technology needs to be introduced speedily to the level of industrialisation in the agricultural sector to cater for the growth of the population in the SSA therefore, this requires new sources of funding to boost farmers to become more successful.

REFERENCES

Abunyewa, A. A., Osei, C., Asiedu, E. K. and Safo, E. Y. (2007). Integrated Manure and Fertilizer Use, Maize Production and Sustainable Soil Fertility in Sub Humid Zone of West Africa. Journal of Agronomy, 6(2), pp. 302 – 309.

Akinnifesi, F. K., Sileshi, G., Franzel, S., Ajayi, O. C., Harawa, R., Makumba, W., Wolf, J. J. and Chianu, J. N. (2009). On-Farm Assessment of Legume Fallows and other Soil Fertility Management Options Used by Smallholder Farmers in Southern Malawi, Agricultural Journal, 4(6), pp. 260 - 271.

Aromolaran, A., (2004). Household Income, Women's Income Share and Food Calorie Intake in South-Western Nigeria. *Food Policy, 29* (5), pp. 507-530. https://doi.org/ 10.1016/j.foodpol.2004.07.002.

Awokuse, T. O. and Xie, R., (2015). Does Agriculture Really Matter for Economic Growth in Developing Countries? *Canadian Journal of Agricultural Economics, 65*, pp. 77-99. Viewed on 29 July 2019. https://ageconsearch.umn.edu/record/49762/.

Becker, BJ and Schram, C.M., (1994). Examining Explanatory Models through Research Synthesis. In Cooper H, Hedges L. V. (eds). *The Handbook of Research Synthesis* (pp. 357-381). New York, NY: SAGE Publications.

Becker, H. C. and Leon, J., (1988). Stability Analysis in Plant Breeding. *Plant Breeding, 101*(1), pp.1-23.

Bentler, P.M. and Weeks, D.G., (1980). Linear Structural Equations with Latent Variables. *Psychometrika, 45*(3), pp. 289-308.

Bentler, P.M., (2007). On Tests and indices for evaluating structural models. Personality and Individual Differences; 42(5), pp. 825-829.

Berkey, C.S., Hoaglin, D.C., Bouckoms, A.A., Mosteller, F., and Colditz, G.A., (1998). A meta-analysis of multiple outcomes by regression with random effects; *Statistics in Medicine*, 17, pp. 2537-2550.

Bland, J. M., and Altman, D.G., (1995). Multiple Significance Tests: The Bonferroni method. *BMJ, 310*(6973) p.170. https:10.1136/bmj.310.6973.170.

Bolt, T., Prince, E.B., Nomi, J.S., Messinger D., llabre, M.M. and Uddin, L.Q., (2018). Combining region-and network-level brain-behaviour relationships in a Structural equation model. *Neuroimage, 165*, pp. 158-16. DOI: 10.1016/j.neuroimage.2017.10.007.

Borlaung, N.E. (1983). Contributions of Conventional Plant Breeding to Food Production, Science, 219(4585), pp. 689-693. https://doi.org/I:10.1126/science219.4585.689.

Branca, G. and Parelli, C., (2020). Clearing the Air: Common Drivers of Climates Smart Smallholder Food Production in Eastern and Southern Africa. *Journal of Cleaner Production*, 270, https://oloi.org/10.1016/j.jclepro.2020.121900.

Brockwell, S.E. and Gordon, I.R., (2001). A Comparison of Statistical Methods for Meta-Analysis. *Statistics in medicine, 20*(6), pp.825-840. DOI: 10.1002/sim.650.

Caceres, I., Khoury, A. Ei, Khoury, R. Ei., Lorber, S., Oswald, I., Khoury, A., Ei, Atoni, A., Puel, O., and Bailly, J.D. (2020). Aflatoxin, Biosynthesis and Genetic Regulation: A Review Toxins (Bassel), 12, 150; doi:10.3390/toxins12030150.

Caliskan, M., Boker, S.W., Gilad, Y., Ober, C. Genet, P., (2015). Host Genetic Variation Influences Gene Expression Response to Rhinovirus Infection, *Plos Genetics*, https://doi.org/10.1371/Journal.pgen.100511.

Chalmers, I., Sackett, D. and Silagy, C., (1997). *The Cochrane Collaboration: Non-Random Reflections on Health Services Research*. London: BMJ Publishing 231-249.

Chauvin, D., Mulangu, F. and Porto, G., (2020). Food Production and Consumption Trends in Sub-Saharan Africa: Prospects for the Transformation of the Agricultural Sector, Research gates.

Chavas, J.P. and Nauges, C. (2020). Uncertainty, Learning, and Technology Adoption Agriculture. *Applied Economic Perspectives and Policy*, 42(1), pp. 42-53.

Card, N. A., (2012). *Applied meta-analysis for social science research,* Guilford, New York.

Chauvier, Y. W., Thuiller, W., Brun, P., Laverghe, S., Descombes, P., Karger, D. N., Renauld, J., and Zimmermann, (2021). Influence of Climate, Soil, and Land Cover on Plant Species Distribution in the European Alps. *Ecological Monographs* 00(00): e01433.10.1002/em.

Cheung, M.W.L., (2013). Fixed-and Random-Effects Meta-Analytic Structural Equation Modeling: Examples and Analyses in R. *Behavior Research Methods*, *46*(1), pp. 29-40. https://doi.org/10.3758/s13428-013-0361-y.

Cheung, M.W.L and Chan, W., (2005). Meta-Analytic Structural Equation Modeling: A Two-Stage Approach. *Psychological Methods*, *10*(1), pp. 40-64. Viewed on 27 July 2019. https://psycnet.apa.org/record/2005-03264-003.

Cheung, M.W.L., (2009). Comparison of Methods for Constructing Confidence Intervals of Standardized Indirect Effects. *Behavior Research Methods, 41*(2), pp. 425-438.

Colquitt, J.A., LePine, J.A. and Noe, R.A., (2000). Toward an Integrative Theory of Training Motivation: A Meta-Analytic Path Analysis 20 Years of Research. *Journal of Applied Psychology, 85*(5), pp. 678-707. DOI: 10.1037/0021-9010.85.5.678.

Cooper, H. M. and Hedges, L. V., (1994). A Handbook of Research Synthesis *Biometrics, 51*(4), pp. 1-610.

Cooper, H.M., (2010). *Research Synthesis and Meta-Analysis: A step by step Approach*, Los Angeles, CA: Sage ISBN 13: 978-1483331157.

Davis, J. M., Chen, N., Glick, I. D., (2003). A Meta-Analysis of the Efficacy of Second-Generation Antipsychotics. *Archives of general psychiatry*, *60*(6), pp. 553-564. DOI: 10.1001/archpsyc.60.6.553.

Demment, M.W., Young, M.M. and Sensenig, R.L., (2003). Providing Micronutrients Through Food-Based Solutions: A Key to Human and National Development, *Journal of Nutrition, American Society for Nutritional Sciences* 0022-3166/03.

Doyle, F., Freedlank, K., Carney, R., De Jonge, P., Dickens, CH., Pedersen, S., Sorensen, J., and Dempster, M., (2019). Network meta-analysis of randomised trials of pharmacological, psychotherapeutic, exercise and collaborative care interventions for depressive symptoms in patients with coronary artery disease: hybrid systematic review of systematic reviews protocol *Systematic Reviews, 8*, pp. 71. 6. https://doi.org/10.1186/s13643-019-0985-9.

Ferreira, D. F., Demétrio, C.G.B., Manly, BFJ. Machado, A. de A. and Vencovsky, R., (2006). Statistical Models in Agriculture: Biometrical Methods for Evaluating Phenotypic Stability in Plant Breeding. *Cerne, 12*(4), pp.373-388. Viewed on 27 July 2019. https://www.redalyc.org/pdf/744/74412409.pdf.

Freiria, G.H., Goncalves, L.S.A., Furlan, F.F., Fonseca Junior, N.S., Lima, W.F., and Prete, C.E.C., (2018). Statistical Methods to Study Adaptability and Stability in Breeding lines of Foot-type Soybeans, Bragantia, 77(2), 253-264.

Finlay, K.W., Wilkinson, G.N., (1963). The Analysis of Adaptation in a Plant Breeding Programme. Australian Journal of Agricultural Research, 14(6), pp.742-754. https://doi.org/10.1071/AR963074.

Food and Health Organisation of the United Nations (FAO). (2021). Data. Viewed 27 July 2019. <http://www.fao.org/faostat/en/#data>.

Fox, P.N., Skovmand, B., Thompson, B.K., Braun, H.J. and Cormier, R., (1990). Yield and Adaptation of Hexaploid Spring. *Tricicate Euphytica, 47*(1), pp. 57-64. https://doi.org/10.1007/BF00040364.

Fuglie, K. O. and Rada, N.N.E., (2013). Resources, Policies, and Agricultural Productivity in Sub-Saharan Africa. *USDA-ERS Economic Research Report*, (145).

Gashu, D., Demment, M.W. and Stocker, B.J., (2019). Challenges and Opportunities to the African Agriculture and Food Systems. *African Journal of Food Agriculture, 19*(1), pp. 14190-14217. D0I: 1018697/ajfand.84.BLFB 2000.

Gravuer, K., Gennet, S., and Throop, H. L., (2018). Organic amendment Additions to Rangelands: A Meta-\Analysis of Multiple Ecosystem Outcomes. *Global Change Biology., 25*, pp. 1152-1170.

Glass, G.V., (1976). Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher, 5*(10), pp.3-8.

Gilbels, D. Dochy, F., Bossche, P.V. and Sergers, M. (2005). A Meta-Analysis from the Angle of Assessment. *Review of Educational Research, 75*(1), pp. 27-61.

Goodboy, A.K. and Kline, R.B., (2017). Statistical and Practical Concerns with Published Communication Research Featuring Structural Equation Modeling, *Communication Research Reports, 34*(1), 68-77. http://dx.doi.org/10.1080/08824096.2016.1214121.

Green, J.M.H., Goft, S.A., Duran, P.A., (2019). Linking Global Drivers of Agricultural Trade to on-the-Ground Impacts on Biodiversity. *PNAS, 116*(51), pp. 26085-26086.

Gumus, I., Karataş, Y. and Gülcan, M., (2020). Silver nanoparticles stabilized by a metal–organic framework (MIL-101 (Cr)) as an efficient catalyst for imine production from the dehydrogenative coupling of alcohols and amines. *Catalysis Science & Technology*, *10*(15), pp.4990-4999.

Hair, J. F., Hurt, G.T.M., Ringle, C. M., Sarstedt, M. and Thiele, K. O., (2017). Mirror, Mirron on the wall: A Comparative Evaluation of composite-based Structural Equation Modeling Methods. Journal of the Academy of Marketing Science. 45(5), 616-632. https://doi.org/10.1007/s11747-0517-x.

Hansen, Ch.M., (2007). *Hansen Solubility Parameters a User's Handbook. Second Edition*.

Hefferon, K.L. (2015). Nutritionally Enhanced Food Crops: Progress and Perspectives, *International Journal of Molecular Sciences*, *16* (2) pp. 3895-3914.

Henseler, J., (2017). Bridging Design and Behavioral research with variance-based Structural Equation Modeling: *Journal of Advertising*, *46*(1), pp. 178-192. DOI: 10.1080/00913367.201780.

Hauwelingen, H. C., Arens, L. R. and Stijnen, T., (2002)., Advanced methods in meta-analysis: multivariate approach and meta-regression, *Statistics in Medicine, 21*, pp. 589-624.

Hedges, L.V. and Olkin, I. (1985). Statistical Methods in Meta-Analysis. *Journal of Educational Statistics*. 20(1).

Hedges, L.V. and Vevea, J.L., (1994)., Fixed and Random-Effects. *Models in Meta-Analysis. Psychological Methods, 3*(4), pp. 486-504.

Higgins, J.P.T. and Thompson, S.G., (2002). Quantifying Heterogeneity in Meta-Analysis. *Statistics in medicine*, *21*(11), pp.1539-1558. https://doi.org/10.1002/sim.1186.

Higgins, J P.T., Thompson, S.G., Deeks, J. and Altman, D., (2003). Measuring inconsistency in meta-analysis. *British Medical Journal*, *327*(7414), pp.557-60.

Holloway, G., Nicholson, Ch. And Delgado, Ch. (2006). A Groindustralization through Institutional Innovation: Transmissions costs, Cooperatives and Milk-Market Development in The Ethiopian Highlands. *Agricultural Economics, 23*(3), 2000.

Holloway, G., Nicholson, C., Delgado, C., (2000). Agroindustrialization though Institutional Innovation Transaction Costs, Cooperatives and Milk-Market Development in the East-African Highlands. *Wiley, 23*(3), pp: 279-288.

Hoyle, R.H., (2012). *Handbook of Structural Equation Modeling*. Guilford; New York, NY.

Howden, S.M., Soussane, J.F., Tubiello, F.N., Chhetri, N., Dunlop, M. and Meinke, H., (2007). Adapting Agriculture to Climate Change. *Proceedings of the national academy of sciences*, *104*(50), pp.19691-19696

Hussein, G., Miyashiro, H., Nakamura, N. Hattori, M., Kakiuchi, N, Shimotohno, K., (2000). Inhibitory Effects of Sudanese Medical Plant Extracts on Hepatitis C. Virus (HCV) Protease. *Phytotherapy Research: An International Journal Devoted to Pharmacological and Toxicological Evaluation of Natural Product Derivatives*, *14*(7), pp.510-516. DOI: 10.1002/1099-1573(200011)14:7< 510: AID-PTR646>3.0.CO;2-B.

Ikram, M., Zhang, Q., Sroufe, R. Shah, S., (2020). Towards a Sustainable Environment the Nexus between Iso 14001, Renewable Energy Consumption, Access to Electricity, Agriculture and C02 Emissions in SAARC. *Sustainable Production and Consumption, 22*, pp. 218-230.

Jackson, D, Riley, R. D., and White, I. R., (2011). Multivariate meta-analysis: Potential and Promise. Statist. Med., 30 (20), pp. 2481-2498.

Jak, S., (2015). "Meta-Analytic Structural Equation Modeling". Springer via DOI: 10.1007/978-3-319-27174-3.

Jak, S. & Cheung, M.W.L., (2018). Testing Moderator Hypotheses in Meta-Analytic Structural Equation Modeling Using Subgroup Analysis. *Behavior Research Methods*, *50* (4) pp. 1359-1373. DOI: 10.3758/s13428-018-1046-3.

Jayne, T.S., (2012). Managing Food Price Instability in east and Southern Africa. *Global Food Security 1*, pp. 143-149.

Jenmirch, R.I. and Schluchter, (1986). Unbalanced repeated measures Models with Structural Covariance matrices. *Biometrics. 42*, pp. 802-820.

Jöreskog, K.G., (1975). Analysis of covariance structures. In *Multivariate analysis*–III (pp. 263-285). Academic Press.

Jöreskog, K.G., (2000). Latent variable scores and their uses. *Scientific Software International. Inc., Lincolnwood.* Available at http://www.ssicentral.com/Liserl /advancedtopics.html.

Jöreskog, K.G. and Sörbom., (1962). Recent Developments in Structural Equation Modeling, *Journal of Marketing Research, 19*(4), https://doi.org/10.1177/002224378201900402.

Josephson, A.J., Ricker-Gilbert, R. Florax, D., Headey, Dereje, M., (2014). How does Population density influence agricultural intensification and Productivity? Evidence from Ethiopia. *Food Policy* 48 (2014): 142-152. http://dx.doi.org/10.1016/j.foodpol.2014.03.004.

Jayedth, A., Rahimi, K., Bautisha, L., Nazarzadeth, M., Zargar, M.Z., Shab-Bidar, S., (2018). Inflammation Markers and Risk of Developing Hypertension: A Meta-Analysis of Cohort Studies. *Heart*, *105*(9), pp. 686-692. http://dx.doi.org/10.1136/hearjnl-2018-314216.

Kaiser, H.F., (1960). Application of Electronic Computers to Factor Analysis. Educational and Psychological Measurement. *Educational and psychological measurement*, *20*(1), pp.141-151.

Kamanga, B. C. G., Waddington, S. R., Whitbread, A. M., Almekinders, C. J. M., and Giller, K. E. (2014). Improving the Efficiency of Use of Small Amounts of Nitrogen and Phosphorus Fertilizer on Smallholder Maize in Central Malawi, Expl.Agric, 50(2), pp. 229 – 249.

Kim, D. G., Thomas, A. D., Pelster, D., Rosenstock, T. S., and Cobena, A. S., (2016). Greenhouse gas Emissions from Natural Ecosystems and Africa: Synthesis of Available data and Suggestions for futher Research. Biogeosciences, 13, pp. 4789 – 4809. Doi:10.10.51941bg.13-4789-2016.

Kisinyo, P. O., Opala, P. A., Kalapala, V. A., Gudu, S. O., Othieno, C. O., Ouma, E. (2015). Micro-Dosing of Lime, Phosphorus and Nitrogen Fertilizers Effect on Maize Performance on an Acid in Kenya, sustain. Agric, Res. 4, pp. 21 – 30.

Koricheva, J. and Gurevitch, Jessica., (2014). Uses and Misuses of Meta-analysis in Plant Ecology, *Journal of Ecology, 102*(4), pp. 828-844.

Kurukulasuriya, P. and Mendelsohn, R., (2006). Crop Selection: Adopting to Climate Change in Africa. Centre for Environmental Economics and Policy in Africa (CEEPA); Discussion Paper No.26. University of Pretoria, South Africa.

Landis, R.S., (2013). Successfully Combining Meta-Analysis and Structural Equation Modeling: Recommendations and Strategies. *Journal of Business and Psychology, 28*(3), pp.251-261. DOI: 10.1007/s10869-013-9285-x.

Laird and J.H. Ware, (1982). Random effects Models for Longitudinal data, *Biometrics, 38*, pp. 963-974.

Lee, L., Petter, S., Fayard, D., and Robinson, S., (2011). On the use of partial least squares path modelling in accounting, *Research International of Accounting Information System, 12*(4), pp. 305-328.

Lin, Fu-K., Suggs, S., Lin, Chi-H., Browne, J.K., Smalling, R. Engrie, J. C., Chen, K. K., Fox, G.M., Martin, F., Stabinsky, Z., Badrawi, S. M., Lai, Por-H. and Golwasser, E., (1985). Cloning and expression of human erythropoietin gene. *Proceedings of the National Academy of Sciences*, *82*(22), pp.7580-7584

Lipsey, M., and Wilson, D., (2001). *Practical Meta-analysis*. Thousand Oaks, CA: Sage.

Lazou, E., Karelakis, Ch., Galanopoulos, K. and Mattas, K., (2019). The Role of Agriculture as a Development Tool for a Regional Economy. *Agricultural Systems*, *173*, pp.482-490

Lawson, K., Kent, K., Rampofd, S., Telg, R.W. and McLeon-Morin, A., (2020). Impacting Agriculture and Natural Resource Policy: Country Commissioners Decision Marketing Behaviors and Communication Preferences. *Journal of Applied Communication, 1004*(1).

Malcolm, P. and Dan, J., (2019). Empirical Comparison of Univariate and Multivariate Meta-Analysis in Cochrane Pregnancy and Childbirth Reviews with Multiple Binary Outcomes. *Research synthesis methods*, *10*(3), pp.440-451.DOI: 10.1002/jrsm.1353.

Masi, C. M., Chen, H. Y., Hawkley, L. C. and Cacioppo., (2011). On the use of partial least squares path modeling in accounting. *Research International Journal of Accounting Information Systems, 12*(4), pp. 305-328.

Mavridis, D. and Salanti, G.., (2011). A Practical Introduction to Multivariate Meta-Analysis. *Statistical methods in medical research*, *22*(2), pp.133-158. DOI: 10.1177/0962280211432219.

Makongwana, M.P., (2016). *GGE-Biplot and genetic diversity analysis of maize hybrids and in bread lines from the breeding programme at KZN*. (Doctoral Dissertation, University of KZN).

Mashingaidze, N., Twomlow, S., Madakadze, J. C., Mupangwa, W., Mavungadidze, Z. Weed Growth and Crop Yield Responses to Tillage and Mulching under Different Crop Rotation Sequences in Semi-Arid Conditions., J CRISAT, 33, pp. 311 – 327.

Moeyaert, M, Ugille, M., Beretvas, S., Ferron, J & Bunuan, R.l & Van den Noortgate, W., (2018). Methods for dealing with multiple outcomes in meta-analysis: a comparison between averaging effect sizes, robust variance estimation and multilevel meta-analysis. *International Journal of Social Research Methodology*, pp. 1-14. DOI: 10.1080/13645579.2016.1252189.

Mwichabe, S., (2013). The African Agrarian Ideology and food security challenge in Sub-Saharan Africa. *Development*, *56*(3), pp. 412-420. https://doi.org/10.1057/dev.2014.14.

Muluh, G. N., Kimengsi, J. N., and Azibo, N. K. (2019). Challenges and Prospects of Sustaining Donor-Funded Projects in Rural Cameroom. Sustainability, 11, 6990; doi:10.3390/su11246990.

Namara, R.E., Hanjra, M.A., Castillo, G.E., Ravnborg, H.M., Smith, L., Van Koppen, B., (2010). Agricultural Water Management and Poverty Linkages. Agricultural Water Management, *Comprehensive Assessment of Water Management Agriculture, 97(*4), pp. 520-527.

Nitzi, C., (2016). The use of Partial least squares Structural equation modelling (PLS+SEM) in Management Accounting Research: Direction for future theory development. Journal Accounting Research: direction for future theory development. *Journal Accounting Literature, 37*, 19-35. DOI: 10.1016/ j. acclit.2016.09.003.

Nurius, P. S., Fleming, C. M. and Brindle, E., (2019). Life Course Pathways from Adverse Childhood Experiences to Adult Physical Health: A Structural Equation Model. *Journal of Aging and Health**, 31***(2), pp. 211-230. DOI: 10.1177/0898264317726448.

Nuhie, O.A., (2019). Antibiotic Use and Resistance Pattern in Ethiopia: Systematic Review and Meta-Analysis, *International Journal of Microbiology*, volume 2019, article ID 2489063.

Okebalama, C. B., Safa, E. Y., Yeboah, E., Abaidoo, R. C. and Logah, V. (2016). Fertilizer Use and Management Practices among Maize and Cowpea Smallholders Farmers in Ghana. Tropicultura, 34(2), pp. 113 – 126.

Olmos, J.G., Cebollero, M.P. and Ferr*é* E.G., (2018). Meta-Analysis of the Structural Equation. Models Parameters for the Estimation of Brain Connectivity with fMRI. *Frontiers in Behavioural Neuroscience, 12,* p.19. | DOI: 10.3389/fnbeh.2018.00019.

Ogundari, K. and Awokuse, T., (2016). Assessing the Contribution of Agricultural Productivity to Food Security Levels in Sub-Saharan African Countries. *Research in Agricultural and Applied Economic, 333.* Viewed on 29 July 2019.

Opala, P., Odendo, M. and Muyekho, F. (2019). Effects of Lime and Fertilizer on Soil Properties and Maize Yields in Acid Soils of Western Kenya. African Journal of Agricultural Research, 13(13), pp. 658 – 663.

Oiu, L. and Qi, L. (2020). E-Learning Assessment for Tourism Education LISREL Assisted Intercultural Tourism Perception and Data Integrated Satisfaction Perspectives, *Journal of Computing in Higher Education, 32*(1), pp. 89-108.

Ondiege, P., Moyo, J.M., Vendier-Chouchane, A., (2013). Developing Africa's infrastructure for enhanced Competitiveness. *Report 2013,* pp. 69-91.

Oue'draogo, M., Dao, K., Millogo, Y., Aubert, J-E., Messan, A., Seynou, M., Zerbo, L., Gomina, M. (2019). Physical, Thermal and Mechanical Properties of Adobes Stabilized with Fonio (Digitana exilis) Straw, J. Building Eng., 23, pp. 250 – 258.

Pasley, H. R., Cairus, J. E., Camberato, J. J. and Vyn, T. J. (2009). Nitrogen Fertilizer Rate Increases Plant Uptake and Soil Availability of Essential Nutrients in Continuous Maize Production in Kenya and Zimbabwe, Nutr Cycl Agroecosyst, 115, pp. 373 – 389.

Passioura, J., (2005). The drought Environment Physical Biological and Agricultural Perspectives, *Journal of Experimental Botany, 58*(2), pp. 113-117.

Pedhazur, E.J., (1997). Multiple Regression in Behavioral Research. (3rd ed). New York, NY: Holt, Rinehart, and Winston. Viewed on 27 July 2019. http://agris.fao.org/agris-search/search.do?recordID=US201300123426.

Pedhazur, E. J. & Kerlinger, F.N., (1982). *Multiple-Regression in Behavioral Research* (2nd Ed.). Fort Woth, TX: Holt, Rinehart and Winston. pp. 555-7.

Pegu, A., Borate, B., …Ying Huang. (2019). A Meta-Analysis of Passive Immunization Studies Shows that Serum-Neutralizing Titter Associates with Protection against SHIV Challenge, Cell Host and Microbe, 26, pp. 336 – 346. https://doi.org/10.1016/j.chom.2019.08.014.

Pinto, B. M., Rabin, C. and Dunsiger S., (2009). Home-Based Exercise among Cancer Survivors. Adherence and its Predictors. *Psycho‐Oncology: Journal of the Psychological, Social and Behavioral Dimensions of Cancer*, *18*(4), pp.369-376. DOI: 10.1002/pon.1465

Ployhart, R. E. and Ward, A.K., (2011). The "Quick Start Guide" for Conducting and Publishing Longitudinal Research. *Journal of Business and Psychology*, *26*(4), pp. 413-422. DOI: 10.1007/s10869-011-9209-6.

Piepho, H.P. (1994)., Best Linear Unbiased Prediction (BLUP) for Regional Yield Trials: A Comparison to Additive Main Effects and Multiplicative Interaction (AMMI) Analysis. *Theoretical and Applied Genetics, 89* (5), pp. 647-654.

Piepho, H.P., (1998). Methods for Comparing the Yield Stability of cropping Systems-A review. *Journal of Agronomy and Crop Science*, *180*(4), pp.193-213.

Pretty, J. Attwood, S., Bawden, R., Yang, P., (2020). Assessment of Growth in Social Groups for Sustainable Agriculture and Land Management. *Global Sustainability, 3*, e23.

Rajabi, A., Dehghani, M., Shojaei, A. Farjam, M. and Motevalian, S.A., (2019). Association between Tobacco Smoking and Opioid Use: A Meta-Analysis, *Addictive Behaviors, 92*, pp. 225-235.

Raudenbush, S. W. and Bryk, A.S., (2002). H*ierarchical Linear Models: Applications and Data Analysis Methods* (Second Ed.). Thousand Oaks, CA: SAGE Publications.

Raykov, T. and Marcoulides G.A., (2000). *First Course in structural equation modeling*. Viewed on 27 July 2019. https://psycnet.apa.org/record/2000-03763-000.

Riley, R. D., (2009). Multivariate meta-analysis: the effect of ignoring within-study correlation. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *172*(4), pp.789-811.

Shadish, W.R., (1996). Meta-analysis and the exploration of causal mediating processes: A Primer of examples, methods and issues. *Psychological Methods;* 1(1), pp. 47-65.

Shal, A.A., and Jafari, A., (2014). Study of Structural and Magnetic Properties of Superparamagnetic Fe3O4-ZnO Core-Shell Nanoparticles. Journal of Super conductivity and Novel Magnetism, 27, 1531-1538.

Shanock, L.R., Baran, B.E., Gentry, W.A., Pattison, S.C. and Heggestad, E.D., (2010). Polynomial Regression with Response Surface Analysis: A Powerful Approach for Examining Moderation and Overcoming Limitations of Difference Scores. *Journal Business and Psychology*, *25*(4), pp. 543-554. DOI: 10.1007/s10869-010-9183-4.

Shiferaw, B., Hellin, J., and Murisho, G., (2011). Improving market access and agricultural productivity growth in Africa: What role for producer organisations and collective action institutions? *Food security, 3(*4), pp. 475-489.

Shukla, G.K., (1972). Some Statistical Aspects of Partitioning Genotype-Environmental Components of Variability. *Heredity, 29*(2), pp.237-245.

Siegmann, E., Bouna-Pyrrou, P. Lenz, B. and Komhuber, J., (2019). J*ournal of Neural Transmission, 126*, pp. 623-636.

Sileshi, G., Akinnifesi, F. K., Ajayi, O. C. and Place, F. (2008). Meta-Analysis of Maize Yield Response to Woody and Herbaceous Legumes in Sub-Saharan Africa, Plant Soil, 307, pp. 1 – 9.

Sime, G. and Aune, J. B. (2014). Maize Response to Fertilizer Dosing at Three Sites in the Central Rift Valley of Ethiopia. Agronomy, 4, pp. 436 – 451.

Smith, M.L. and Glass, G.V., (1977). Meta-Analysis of Psychotherapy Outcome Studies. *American psychologist*, *32*(9), pp.752.

Snedecor, G. W. and Cochrane, W. G. (1986). *Statistical Methods*, Eight Edition, Iowa State University,

Sobel, M.E., (1986). Some New Results on Indirect Effects and their Standard Errors in Covariance Structure Models. *Sociological Methodology, 16,* pp. 159-186.

Tovihoudji, P. G., Akponikpe, I., Euloge, A., Bielders, C. (2019). Using the DSSAT Model to Support Decision Making Regarding Fertilizer Microdosing for Maize Production in the Sub-Humid Region of Benin, Frontiers in Environmental Science, DOI:10.3389/fenvs.2019.00013.

Turner, R.M., Bird, S.M. and Higgins, J.P.T., (2013). *The Impact of Study Size on Meta-Analysis: Examination of Underpowered Studies in Cochrane Reviews.* https://doi:org/10.1371/ Journal pone 0059202.

United Nations (UN), (2009). World Economic and Social Survey. Report.

Van der Velde, M. Folberth, C., Balkovic, J., Ciais, P., Fritz, S., Janssens, I. A., Obersteiner, M., See, L., Skalsky, R., Xiong, W., Penwelas, J. (2013). African Crop Yield Reductions due to Increasingly Unbalanced Nitrogen and Phosphorus Consumption, Global Change Biology, 20(4), pp. 1278 – 1288.

Viswesvaran, C. and Ones D.S., (1995). Theory Testing: Combining Psychometric Meta-Analysis and Structural Equations Modeling. *Pesonnel psychology, 48*(4), pp. 865-885. DOI : 10.1111/j.1744-6570. 1995.tb01784. x.

Vicedo-Cabrera A. M., Sera F. and Gaspamini, A. (2019). Hands-on Tutorial on a Modeling Framework for Projections of Climat Change on Heath, PMC Epidemiology, 30 (3). Pp. 321 – 329.

Virués-Ortega J., (2010). Applied Behavior Analytic Intervention for Autism in Early Childhood: Meta-Analysis, Meta-Regression and Dose-Response Meta-Analysis of Multiple Outcomes. *Clinical Psychology Review 30*(4), pp. 387-399.

Wallace, J.S. and Gregory, P.J., (2002). Water Resources and their Use Food Production Systems, *Aquatic Sciences*, 64(4), pp. 363-375.

Wang, J. and Staver J.R., (2001). Examining relationships between factors of science education and student career aspiration. *The Journal of Educational Research*, 94(5), pp.312-319.

West, M.A., Richter, A. and Dawson, J.F., (2010). The Effectiveness of Teams in Organizations: A Meta-Analysis. The International Journal of Human Resource Management, 22(13), pp. 2749-2769. D0I: 10.1080/09585192.2011. 573971.

White, I. R., (2009). Multivariate random-effects meta-analysis. *The Statistical Journal Vol. 9(*1), pp. 40-56.

Wilson, T.D., (1999). Exploring Models of Information Behaviour: The "Uncertainty.*"* Project. *Information Processing and Management*, 35(6), pp. 839-849.

World Health Organization (WHO)., (2015). World health statistics 2015. Viewed on 29 July 2019. <https://apps.who.int/iris/bitstream/handle/10665/170250/9789240694439_eng.pdf?sequence=1.

World Bank., (2007). World Development Report, 2007. Development and the Next Generation. Viewed on 29 July 2019.

Wolf, F M., (1986). *Meta-Analysis: Quantitative Methods for Research Synthesis*. Thousand Oakes, CA: Sage Publishers 1st Edition,

Wricke, G., (1962). On a method of understanding the biological diversity in field research. Z. Pfl. *Züc̈ht, 47*, pp. 92-146.

Wright, S., (1921). Systems of mating. I. the Biometric Relations between parent and offspring. *Genetics*, *6*(2), pp.111.

Xu, G., Strathearn, L. Liu, B. and Bao, W., (2014). Prevalence of Autism Spectrum Disorder Among Us Children and Adolescents, *JAMA, 319*(1), pp. 81-82.

Yang, J., Kolari, J.W., and Min, I., (2003). Stock Market Integration and Financial Crises: The Case of Asia. *Applied Financial Economics, 13*(7), pp. 477-486.

Yan, W. and Tinker, N. A., (2006). Biplot Analysis of Multi-Environment Trial Data: Principles and Applications. *Canadian Journal of Plant Science, 86*(3), pp.623-645.

Yates, F. Cochran, W.G., (1938). The Analysis of Groups of Experiments. *The Journal of Agricultural Science, 28*(4), pp.556-580. DOI: 10.1017/S0021859600050978.

Yuan, Z. and Yang, Y., (2005). Combining Linear Regression Models: When and How? *Journal of the American Statistical Society, 100*, pp. 1202-1214.

Yuniko, F. (2009). The Correlation of Aspect of Patient Relationship to staff Toward the Satisfaction Level of Outpatient Installation at RSUD. *Pariaman Proceeding International Conference Syedza Saintika, 1*(1).

APPENDIX

A.1. VISUALIZATION OF THE DATA

Chapter 2: Data

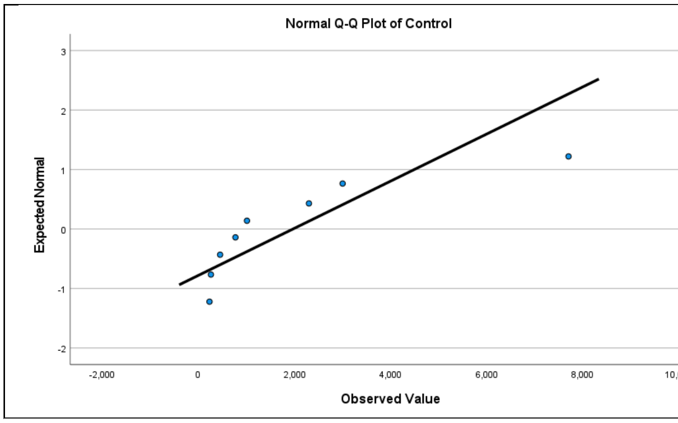
Summaries statistics						
	Mean	Sdt dev	Median	Skewness	Kurtosis	95% Confidence Interval
Control	1965.0	2524.8	890	2.054	4.453	(-145.8; 4075.8)
Fertilizer	2156.0	2767.5	925	1.574	2.322	(-157,7; 4469.7)
Yield.cont	3378.9	4579.0	1982	2032	4.580	(-447.6; 7205.3)
Yield.fert	10040.1	13376.1	4932	1.591	1.779	(-1142.8; 21222)
Rainfall	959.5	557.5	920	0.119	-0.412	(493.4; 1425.6)

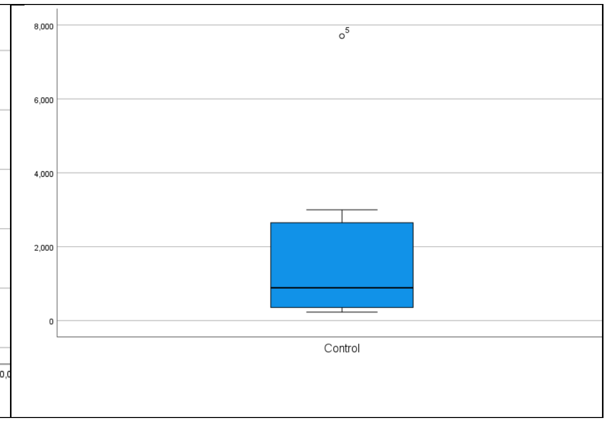
Tests of Normality

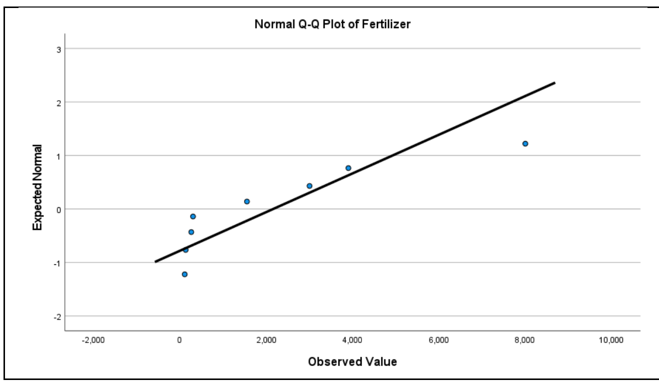
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Control	.272	8	.082	.734	8	.005
Fertilizer	.249	8	.156	.792	8	.024
Yield.control	.278	8	.068	.736	8	.006
Yield.fertilizer	.296	8	.038	.770	8	.013
Rainfall	.141	8	.200*	.977	8	.948

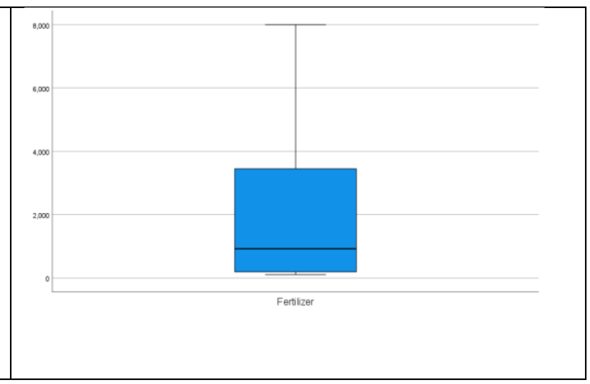
*. This is a lower bound of the true significance.

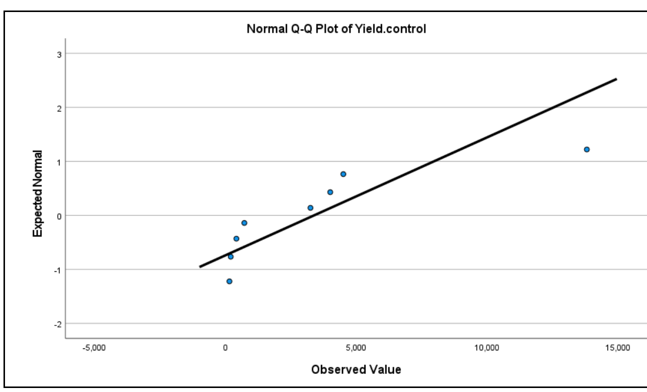
a. Lilliefors Significance Correction

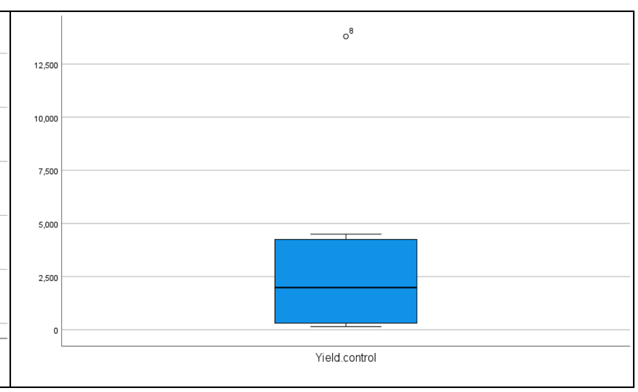

Normal Q-Q Plot of Control
Expected Normal
Observed Value
3
2
1
0
-1
-2
-2,000
0
2,000
4,000
6,000
8,000
10,0

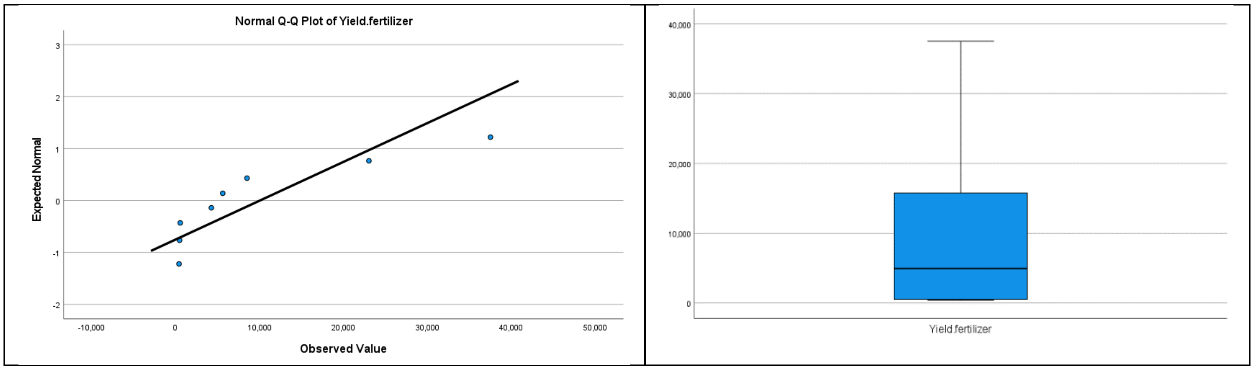

8,000
6,000
4,000
2,000
0
5
Control

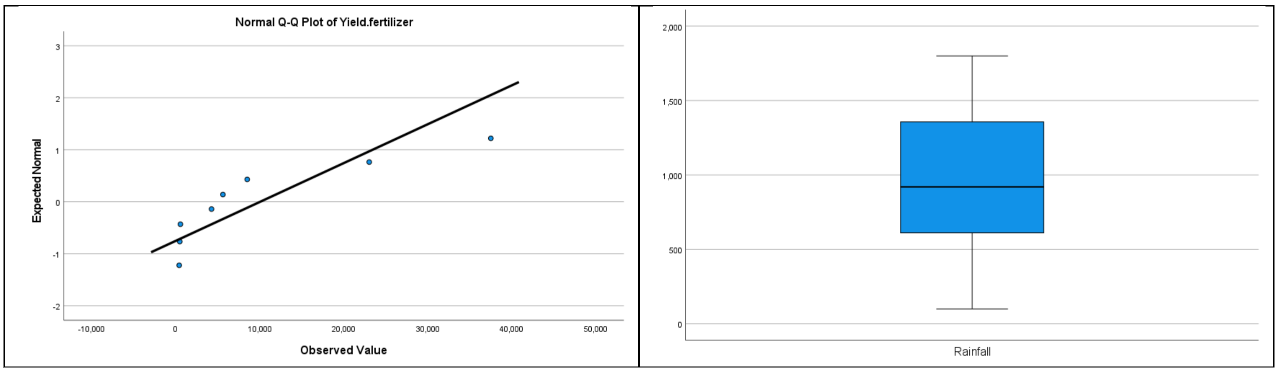

Normal Q-Q Plot of Fertilizer
Expected Normal
Observed Value
3
2
1
0
-1
-2
-2,000
0
2,000
4,000
6,000
8,000
10,000


8,000
6,000
4,000
2,000
0
Fertilizer


Normal Q-Q Plot of Yield.control
Expected Normal
Observed Value
3
2
1
0
-1
-2
-5,000
0
5,000
10,000
15,000

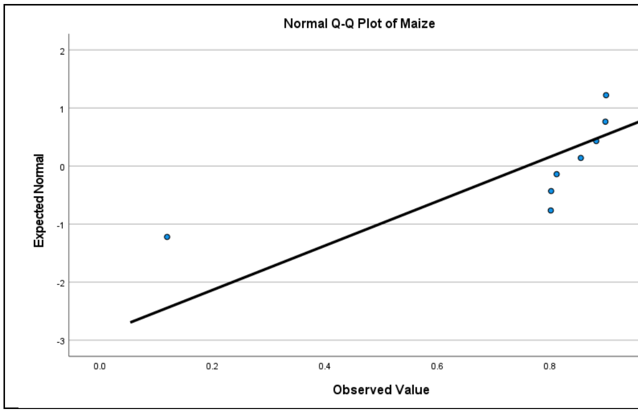

12,500
10,000
7,500
5,000
2,500
0
8
Yield.control

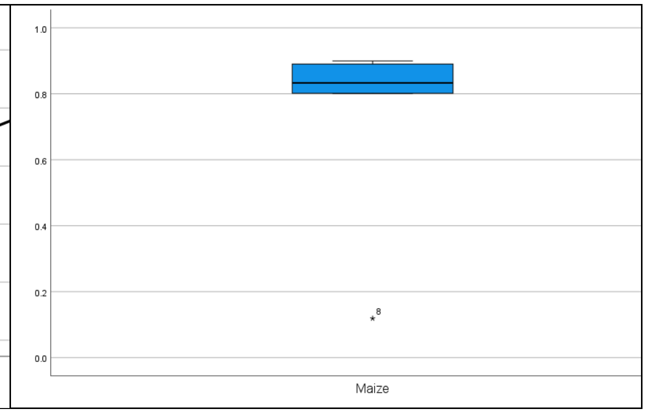


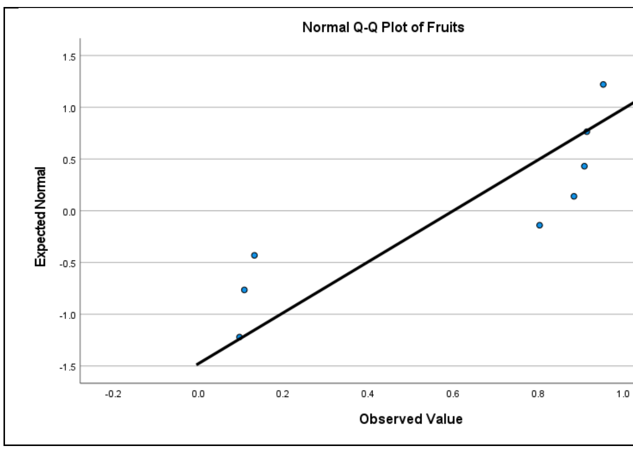


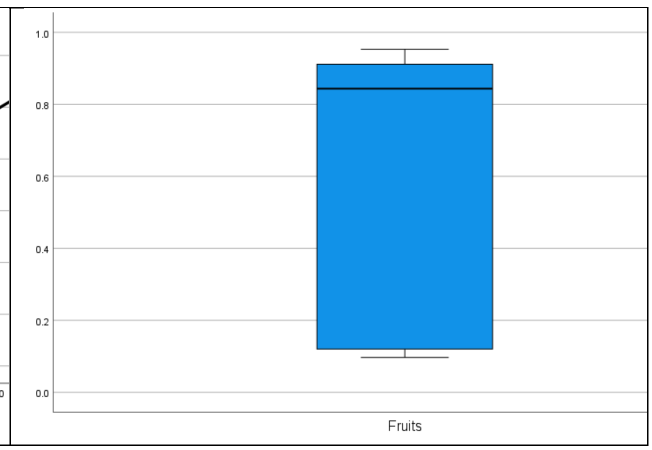
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Maize	.440	8	.000	.563	8	.000
Fruits	.317	8	.018	.723	8	.004
Vegies	.137	8	.200*	.972	8	.913

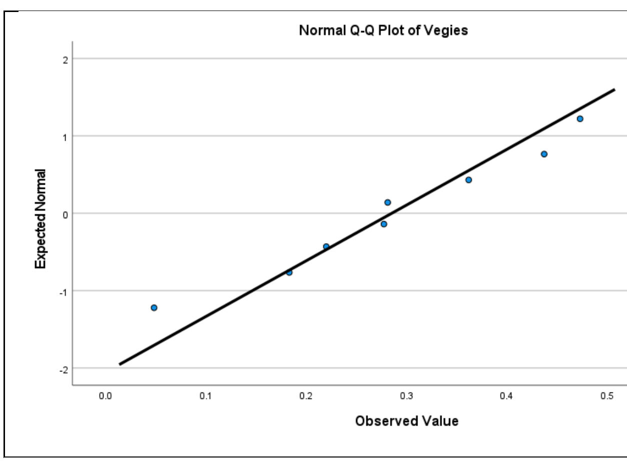
*. This is a lower bound of the true significance.

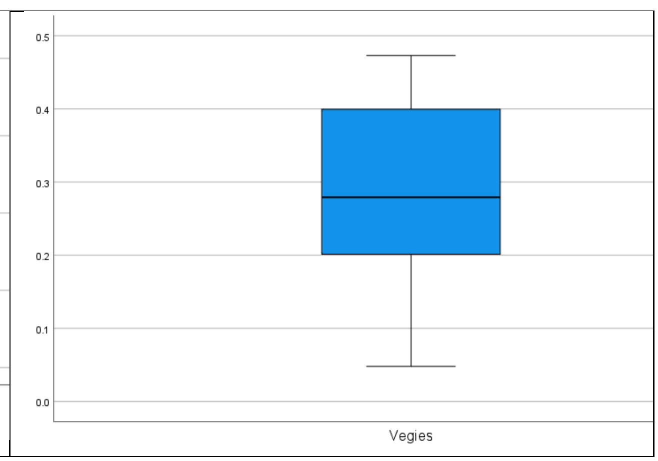

Normal Q-Q Plot of Maize
Expected Normal
Observed Value

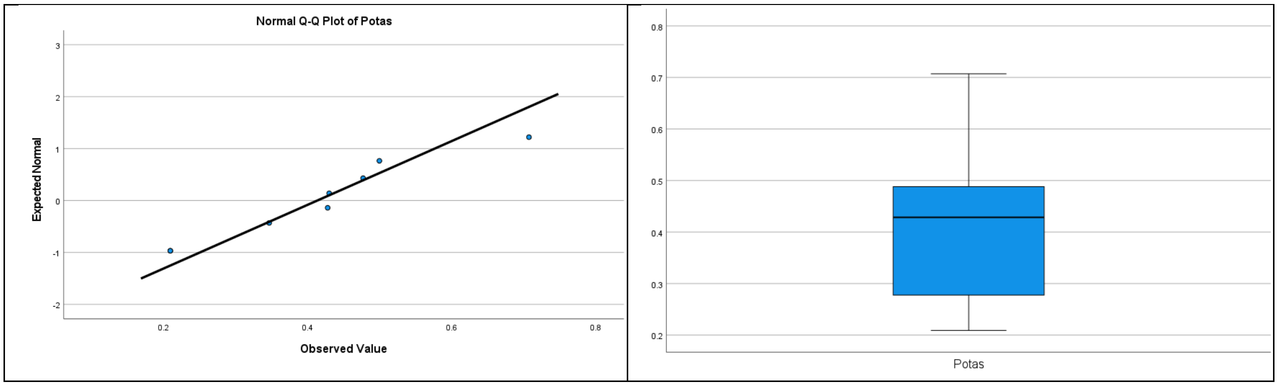

8
Maize

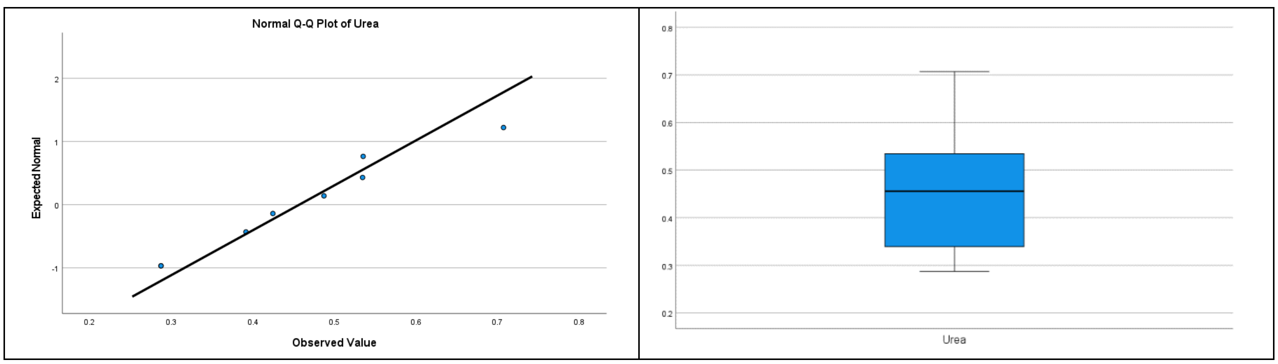

Normal Q-Q Plot of Fruits
Expected Normal
Observed Value


Fruits


Normal Q-Q Plot of Vegies
Expected Normal
Observed Value


Vegies





A.2. R SOFTWARE

Using the R software, the Input for simulation is the normal distribution that has a mean μ and sigma matrix ∑ as described below

mean<-

c(3.00E+06,3.00E+08,1.00E+07,4.00E+07,1.00E+07,4.00E+07,3.00E+06,5.00E+06,1.00E+07,3.00E+06,6.00E

+08,7.00E+08,1.00E+05,1.14E+03,7.74E+00,3.28E+00,1.15E+04,7.24E+03,4.00E+06,1.00E+08,

```
1.00E+08,2.00E+06,9.00E+05,6.00E+06,2.00E+06)

sigma<-matrix(c(1,0.416,0.171,0.044,-0.081,-0.04,0.059,-0.02,0.081,0.045,0.156,0.189,0.004,-0.129,-
0.12,-0.188,0.045,0,-0.093,-0.065,-0.128,-0.035,-0.087,-0.089,0.092,

0.416,1,0.11,-0.001,-0.099,-0.046,-0.02,-0.029,0.113,0.043,0.079,0.013,-
0.103,0.074,0.119,0.007,0.273,0.069,-0.119,-0.126,-0.153,0.014,-0.004,0.029,0.046,

0.171,0.11,1,0.043,0.204,-0.013,-0.014,0.108,0.668,0.762,0.076,0.48,0.031,0.027,-0.073,-
0.073,0.179,0.122,-0.057,-0.048,-0.139,-0.064,-0.095,-0.033,0.174,

0.044,-0.001,0.043,1,-0.027,-0.025,0.011,-0.008,0.181,-0.02,-0.048,-0.023,-0.047,0.095,-0.07,-0.085,-
0.091,-0.079,-0.117,-0.064,-0.006,-0.121,-0.054,0.185,-0.099,

-0.081,-0.099,0.204,-0.027,1,0.927,0.067,0.963,0.274,0.269,0.007,0.201,0.123,-0.095,-0.047,-
0.027,0.031,-0.021,-0.059,     -0.077,-0.002,0.029,-0.087,0.113,-0.016,

-0.04,-0.046,-0.013,-0.025,0.927,1,0.162,0.984,0.093,-0.009,0.082,0.091,0.079,-0.102,-0.02,-
0.013,0.01,0.006,0.006,-0.001,0.055,-0.034,-0.059,0.134,-0.109,

0.059,-0.02,-0.014,0.011,0.067,0.162,1,0.082,0.387,-0.025,0.674,0.654,0.65,0.106,-0.147,-
0.037,0.12,0.304,0.562,0.65,0.369,0.205,0.067,0.063,-0.302,

-0.02,-0.029,0.108,-0.008,0.963,0.984,0.082,1,0.177,0.132,0.028,0.105,0.06,-0.094,-0.022,-0.025,0.02,-
0.015,-0.065,-0.069,-0.021,-0.037,-0.075,0.12,-0.066,

0.081,0.113,0.668,0.181,0.274,0.093,0.387,0.177,1,0.794,0.443,0.717,0.464,-0.08,-0.165,-
0.154,0.092,0.12,0.128,0.218,0.078,-0.014,0.046,-0.059,0.087,
```

0.045,0.043,0.762,-0.02,0.269,-0.009,-0.025,0.132,0.794,1,0.05,0.459,0.29,-0.001,-0.122,-

0.091,0.118,0.021, -0.107,-0.07,-0.083,-0.003,-0.066,-0.079,0.267,

0.156,0.079,0.076,-0.048,0.007,0.082,0.674,0.028,0.443,0.05,1,0.658,0.462,-0.181,-0.068,-

0.081,0.035,0.217,0.377,0.431,0.22,-0.002,0.104,-0.221,-0.155,

0.189,0.013,0.48,-0.023,0.201,0.091,0.654,0.105,0.717,0.459,0.658,1,0.603,-0.137,-0.143,-

0.094,0.119,0.246,0.389,0.431,0.213,0.077,-0.038,-0.114,-0.039,

0.004,-0.103,0.031,-0.047,0.123,0.079,0.65,0.06,0.464,0.29,0.462,0.603,1,-0.14,-0.155,-

0.032,0.195,0.239,0.312,0.382,0.216,0.019,0.053,0.093,-0.221,

-0.129,0.074,0.027,0.095,-0.095,-0.102,0.106,-0.094,-0.08,-0.001,-0.181,-0.137,-

0.14,1,0.053,0.096,0.298,0.201,-0.186,-0.06,0.063,0.596,-0.014,-0.143,-0.045,

-0.12,0.119,-0.073,-0.07,-0.047,-0.02,-0.147,-0.022,-0.165,-0.122,-0.068,-0.143,-

0.155,0.053,1,0.823,0.082, 0.113,-0.062,-0.122,-0.156,-0.085,-0.103,0.172,-0.181,

-0.188,0.007,-0.073,-0.085,-0.027,-0.013,-0.037,-0.025,-0.154,-0.091,-0.081,-0.094,-

0.032,0.096,0.823,1,0.193,0.284,0.06,-0.053,-0.092,-0.092,0.082,0.072,-0.302,

0.045,0.273,0.179,-

0.091,0.031,0.01,0.12,0.02,0.092,0.118,0.035,0.119,0.195,0.298,0.082,0.193,1,0.898,0.005,0.013,0.118,-

0.044,-0.19,-0.247,-0.03,

0,0.069,0.122,-0.079,-0.021,0.006,0.304,-

0.015,0.12,0.021,0.217,0.246,0.239,0.201,0.113,0.284,0.898,1,0.194,0.193,0.158,-0.139,-0.099,-0.249,-

0.186,

-0.093,-0.119,-0.057,-0.117,-0.059,0.006,0.562,-0.065,0.128,-0.107,0.377,0.389,0.312,-0.186,-0.062,0.06,0.005,0.194,1,0.917,0.675,-0.098,0.16,0.026,-0.178,

-0.065,-0.126,-0.048,-0.064,-0.077,-0.001,0.65,-0.069,0.218,-0.07,0.431,0.431,0.382,-0.06,-0.122,-0.053,0.013,0.193,0.917,1,0.688,0.086,0.051,0.057,-0.159,

-0.128,-0.153,-0.139,-0.006,-0.002,0.005,0.369,-0.021,0.078,-0.083,0.22,0.213,0.216,0.063,-0.156,-0.092,0.118,0.158,0.675,0.688,1,-0.016,0.056,-0.104,0.19,

-0.035,0.014,-0.064,-0.121,0.029,-0.034,0.205,-0.037,-0.014,-0.003,-0.002,0.077,0.019,0.596,-0.085,-0.092,-0.044,-0.139,-0.098,0.086,-0.016,1,0.009,-0.203,0.029,

-0.087,-0.004,-0.095,-0.054,-0.087,-0.059,0.067,-0.075,0.046,-0.066,0.104,-0.038,0.053,-0.014,-0.103,0.082,-0.19,-0.099,0.16,0.051,0.056,0.009,1,-0.388,-0.187,

-0.089,0.029,-0.033,0.185,0.113,0.134,0.063,0.12,-0.059,-0.079,-0.221,-0.114,0.093,-0.143,0.172,0.072,-0.247,-0.249,0.026,0.057,-0.104,-0.203,-0.388,1,-0.378,

0.092,0.046,0.174,-0.099,-0.016,-0.109,-0.302,-0.066,0.087,0.267,-0.155,-0.039,-0.221,-0.045,-0.181,-0.302,-0.03,-0.186,-0.178,-0.159,0.19,0.029,-0.187,-0.378,1),25,25)

The program is

```
for(i in 1:1){

data<-mvrnorm(n=33,mean,sigma);

#out<-princomp(data,cor=TRUE, scores=FALSE);

#out2<-summary(out);

outf<-factanal(data,factors=1,rotation="varimax");
```

```
#outp<-prcomp(data,retx=TRUE,center=TRUE,scale=TRUE,rank=10);

#mp<-as.matrix(out)

#mptot<-mptot+mp;

#mptot2<-mptot2+mp*mp;

mf8<-as.matrix(loadings(outf));

#Sample10<-mf8;

#mf8tot<-mf8tot+mf8;

#mf8tot2<-mf8tot2+mf8*mf8;

}

#mf8aver<-mf8tot/10000;
```