UNIVERSITY of RWANDA

COLLEGE OF SCIENCE AND TECHNOLOGY
SCHOOL OF SCIENCE
DEPARTMENT OF MATHEMATICS

STATISTICAL MODELLING OF GREEN COFFEE PRICE IN SOME SELECTED EAST AFRICAN COUNTRIES USING PANEL DATA

By

UWIMBABAZI Jeannette
Student Number: 220004378

Thesis Submitted in Partial Fulfillment of the Academic Degree of
Master of Science
in
Applied Mathematics
(Statistical modeling and actuarial sciences)

Supervisors: Prof. Joseph NZABANITA and
Dr. Emelyne UMUNOZA GASANA

August 29, 2024

Abstract

Coffee is a critical agricultural commodity in East African countries, significantly contributing to their economies by generating foreign exchange, creating employment opportunities and supporting the livelihoods of millions of smallholder farmers.Despite its importance, the region's coffee sector faces numerous challenges, including fluctuating international prices, climate change impacts, limited access to quality inputs, policy and trade barriers, and economic instability. While various researchers have explored aspects of coffee price in East Africa, there has been no comprehensive study utilizing panel data to analyze these dynamics across multiple countries over an extended period. This study addresses this gap by developing a statistical model to analyze the evolution of coffee price in East African countries from 1996 to 2022. Data were collected from FAOSTAT and the World Bank, and a panel data analysis model was employed. The findings reveal that regional trade agreements play a vital role in enhancing coffee price and trading in the region. Additionally, factors such as exchange rates, fertilizers, inflation rates, agricultural land, and foreign direct investment are significant determinants of coffee price. The study also identifies that agricultural employment and inflation are currently under performing relative to their potential contribution to the global market, indicating substantial room for improvement. The objectives of this research are to identify key determinants of coffee price ,To evaluate the economic implications of green coffee price in East Africa using panel data analysis techniques and provide actionable policy recommendations. The insights gained from this study aim to inform policymakers and stakeholders in developing strategies to promote sustainable coffee price, enhance economic stability, and improve the livelihoods of coffee farmers in East Africa.

DECLARATION

This thesis is my original work and has not in part or in whole been presented for a degree award in any other university.
Signature

Jeannette UWIMBABAZI

Date 29/8/2024

Certificate

This thesis has been submitted with our approval as University supervisors.

Prof. Joseph NZABANITA

Signature

Date 29/8/2024

Dr. Emelyne UMUNOZA GASANA

Signature

Date 29/8/2024

ACKNOWLEDGEMENT

I am very delighted to express my sincere appreciation to the University of Rwanda which has provided a conducive environment for learning. I would like to express my sincere gratitude to my supervisors Prof. Joseph NZABANITA and Dr. Emelyne UMUNOZA GASANA for their guidance, motivation, pieces of advice, necessary explanations, and their bright ideas for this study.

Contents

LIST OF ABBREVIATIONS AND SYMBOLS

FE: Fixed Effect Model

RE: Random Effect Model

EAC: East African Community

FAO: Food and Agriculture Organization

UNCTAD: United Nation Conference on Trade and Development

WB: World Bank

EU: European Union

VAR: Vector autoregression

GDP:Gross domestic product

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Millions of people in East Africa depend on the coffee business for their livelihoods, making it a significant contributor to the region's GDP. The quality and sustainability of coffee production are heavily influenced by the green coffee processing methods. Coffee originated in the south-western side of the Great Rift Valley, including regions like the Kaffa region in Ethiopia, the Boma Plateau in Sudan, and Mount Marsabit in Kenya [6]. Coffee gained popularity before the late Middle Ages, making it a relatively recent beverage compared to tea, wine, and cocoa, which have been known for thousands of years [14].

Coffee was first introduced to Yemen from Ethiopia between the 15th and 16th centuries. Subsequently, coffee spread globally: Yemen to India in the 1660s by Baba Budan, to Sri Lanka [15] by the Dutch in the 1690s, to Indonesia in 1699, and to Amsterdam in 1706. French missionaries brought coffee to Réunion in 1715, while coffee reached Brazil in 1727 via French Guiana and was planted in Para. Jamaica adopted coffee by 1730, and it reached Costa Rica by 1779. In 1784, coffee arrived in Venezuela, and in 1822, coffee plantations were established in Angola. Australia began coffee cultivation in the 1880s. French missionaries introduced coffee to Kenya in 1893, and it reached Tanzania in 1898 and Rwanda in 1904 under German colonization. Robusta coffee expanded across Africa in the 1920s [4]

Urbanization significantly increased coffee consumption, leading to the proliferation of coffee shops. However, the sustainability of coffee supply is at risk due to persistently low prices over recent decades [9]. Consequently, many smallholder farmers have switched to more profitable crops. Countries within the tropics, between the Tropic of Cancer and the Tropic of Capricorn, produce nearly all global coffee. These regions form the "coffee belt" or "coffee growing zone" [8]. Unfortunately, coffee-consuming countries are primarily located in the Northern Hemisphere, where value is added through roasting, branding, packaging, and retailing, resulting in greater benefits compared to producing countries in the coffee belt [3].

Over 500 years, the coffee industry has grown exponentially [5], becoming the most consumed product worldwide. According to the International Coffee Organization [5], millions depend on coffee production, with over 25 million people in poverty. Coffee has become the most valuable commodity traded in developing countries, with global trade worth \$30.7 billion in 2006 and significantly more by 2017. In East Africa, which comprises 5.71% of the world's population, over 5 million people are either coffee growers or work in the coffee sector [13]. Coffee contributes significantly to foreign currency earnings and plays a vital role in export, offering a sustainable solution to rural poverty through employment.

East African coffees are renowned for their quality. Ethiopia is famous for unique varieties like Yirgacheffe, Sidamo, and Harar Arabica. Kenya produces high-altitude Arabica near Mount Kenya. Burundi's Kayanza coffees are celebrated with numerous international awards. Tanzanian Kilimanjaro coffees are prized in the Japanese market. Rwanda's Maraba coffee also commands high value. However, smallholder coffee farmers and workers are vulnerable to low coffee prices, lacking bargaining power, which benefits larger actors who control the global market [3], creating significant imbalances in value distribution [13]. African coffee exports totaled approximately \$2 billion annually over the past decade, with 17.123 million bags exported in 2016, representing 11% of global green coffee exports, but declining by 13.5% to 2.8 million bags in 2022. About 10 million people derive their livelihoods from the African coffee economy.

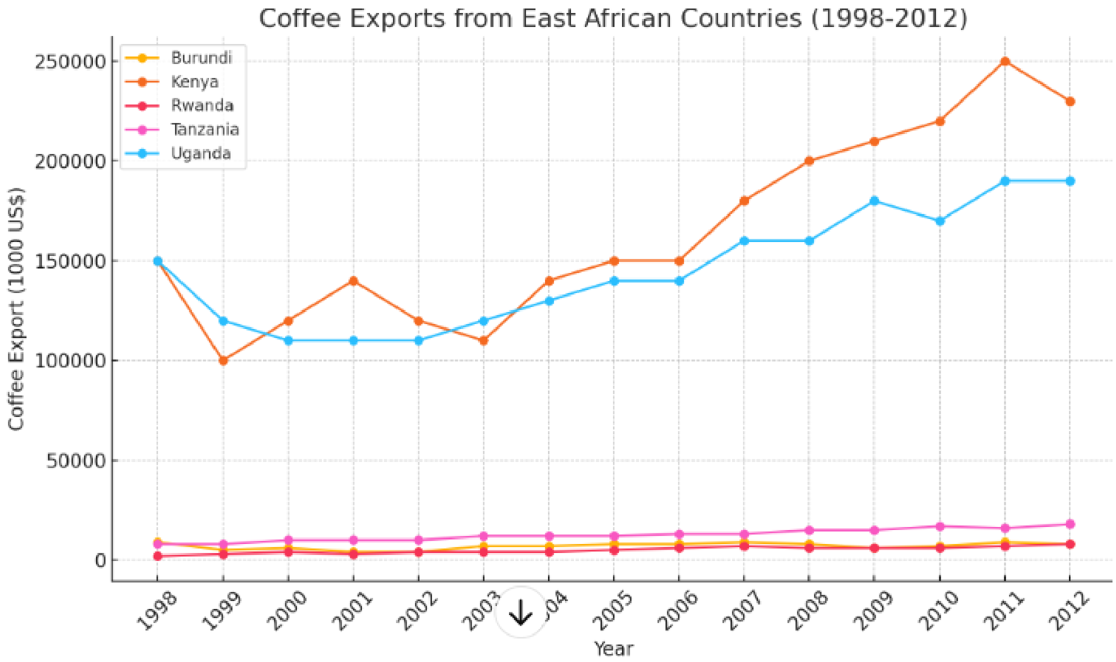

Figure 1.1: Coffee export as a percentage of country's total export to the EU

Figure 1.1 visually represents the contribution of coffee to various countries' total exports to the European Union (EU).

The East Africa sub-region dominates African coffee production, accounting for over 80% of the continent's total production. Coffee export earnings are significant for many East African countries, known for their diverse coffee specialties. Arabica and Robusta coffees dominate production, grown on small plots ranging from half a hectare to 10 hectares per farm [7].

Different harvesting methods are employed worldwide. Hand-picking is the most common method, where only ripe cherries are picked, ensuring high-quality green beans. Strip-picking involves removing all cherries from the branches, ripe or not, which is cost-effective but results in lower-quality coffee. Mechanical harvesting using vibrators or rotating brushes is efficient for large farms but not commonly used in East Africa. The comb method selectively removes ripe cherries, ensuring higher yields and quality but is more time-consuming.

Agriculture in Rwanda contributes 31.5% to GDP and employs 75% of the population [10]. Coffee alone accounts for 27% of export revenues, with 95% of plantations growing Arabica and 5% Robusta [12]. Rwanda faces challenges like low productivity due to poor soil fertility, inadequate fertilizer application, pests and diseases, lack of good agricultural practices, and aging trees [1]

Figure 1.2: Coffee seedling establishment and management practices on polyethylene bag in nursery site (L, 2017)

The figure 1.2 likely depicts the setup of a coffee seedling nursery where polyethylene bags are used to hold the growing medium (soil or compost) for the seedlings. This method is common in nurseries as it allows for easier management of individual plants.

Figure 1.3: Harvested green coffee

This figure 1.4 shows the initial stage of post-harvest processing, where coffee cherries have been collected but not yet processed.

Figure 1.4: Machine used to remove the core of Coffee

This figure 1.4 illustrates the machinery used in the coffee processing stage, specifically for pulping or removing the outer layers of the coffee cherry.

1.1.1 Statistics and Econometrics in Coffee Research

In statistics and econometrics, various models are employed to analyze data. In this study, panel data models were applied to investigate green coffee production in East African countries. Pooled Ordinary Least Squares (OLS), Fixed Effects models (including Least Squares Dummy Variables (LSDV), Fixed Effect model (FE)), and Random Effects models were utilized to estimate the panel data.

The Pooled OLS results indicated significant heterogeneity, reflecting unobserved firm-specific characteristics such as location and board diversity. These characteristics, assumed to be time-invariant, were aggregated in the error term, causing endogeneity issues where the error term correlates with regressors. This biases and inconsistencies in estimated coefficients, leading to erroneous inferences. Fixed Effects and Random Effects models were employed to account for unobserved heterogeneity.

1.2 Problem Statement

East African coffee price faces multiple challenges, including price volatility, mass urban migration, changing societal attitudes, and a significant lack of involvement from younger generations. These issues, combined with the time and financial investments required for coffee farming, often result in minimal profits for farmers, leading to the abandonment of coffee plantations in favor of more lucrative crops. The instability of the coffee market, insufficient organization within the industry, inadequate financial services, and aging coffee trees further exacerbate the problem. Given the crucial economic role of coffee as a cash crop in East Africa, there is an urgent need to enhance the sustainability and profitability of coffee price. To address this, we propose developing a comprehensive statistical model to assist decision-making for coffee producers, government agencies, and investors in Rwanda and the rest of the East African Community.

1.3 Objectives

1.3.1 Main Objectives

To analyze and model the factors influencing green coffee price in East Africa using panel data analysis from 1996 to 2022.

1.3.2 Specific Objectives

1. To identify Key Determinants of Green Coffee Price
2. To evaluate the economic implications of green coffee price in some selected East Africa countries by using panel data analysis techniques
3. To provide actionable policy recommendations.

1.4 Expected users and significance

This study provides the required theoretical framework to practitioners on statistical modelling of Green coffee price in East African countries. The results of this study should help the policy makers and other agencies that intend to promote the development of agribusiness. It will help the investors and the government to minimize the losses and risks that may occur due to lack of deep study as the usual techniques seem to be inadequate. In fact, there is a need of conducting this study as it will contribute to the literature on how to construct statistical model that may give insight on green coffee price in some selected East Africa countries. The findings of this study are expected to be useful to investors, policy makers, researchers and practitioners in their decision making and future planning.

Chapter 2

LITERATURE REVIEW

2.1 Empirical Literature

Lanfranchi (2016) agreed that coffee is the second most marketed product worldwide after petroleum. The countries located in the Global South have more than 20 million farmers, while more than 25 million people make a living from coffee.

Issa (2015) investigated the factors that push farmers to cultivate coffee in Huye District. The findings show that gender, education level, farm size, and access to credits, among others, are crucial factors that influence decision-making.

Sabari (2020) studied the influence of coffee pricing on reviving coffee production in cooperative societies in Meru County, Kenya. The results show that coffee pricing significantly influences the revival of coffee production in cooperatives.

Newman (2012) examined the nature of price fluctuation and transmission in the Tanzanian coffee chain. The study shows the varying roles of domestic marketing and local-level institutions in shaping price fluctuations .

(Abel, 2017) reported that over 25 million farmers produce coffee across four continents, and coffee provides ecosystem services. However, due to coffee architecture, production systems, planting arrangements, and cropping cycles, there are many challenges that hinder the application of traditional remote sensing .

(Tröster, 2015) examined global commodity chains, financial markets, and local market structures' price risks in the coffee sector in Ethiopia. The study highlights the crucial importance of price setting and transmission within the global commodity chain from a developmental and distributional perspective .

2.2 Theoretical Review

Sutton (2002) explained that we have various types of data, such as cross-sectional data, where observations of the subjects are obtained at the same point in time. Data (2011) highlighted that we also have time series data, where observations are generated over time, which can be problematic due to issues of stationarity. Additionally, there is panel data, also called longitudinal data, which combines both time series and cross-sectional observations.

Kerstens and Vanden Eeckaut (2014) identified various types of panel data. For instance, in a balanced panel, there is an equal number of observations for each subject (firm) throughout the entire period. They also noted the existence of a short panel (Micro panel), where the number of subjects (firms) exceeds the number of periods. Lillo-Bañuls and Casadesús-Masanell (2018) described a long panel (Macro panel), where the number of periods surpasses the number of subjects.

Bond (2002) and Wooldridge (2001) explained that dynamic panel data includes lagged values of the dependent variable as regressors and employs the General Method of Moments (GMM) estimation. Baltagi (2006) discussed the concept of an unbalanced panel, where the number of observations varies across subjects (firms).

DeSouza (2005) and Dismuke and Egede (2006) discuss that, by using time series regression, we can estimate a time series model for each subject using the Ordinary Least Square (OLS) method. According to n years, we will end up with n different regression models for each subject (or firm). However, this method leads to disparate pieces of information, which do not allow for a comprehensive assessment of how the independent variables jointly affect the dependent variable. This model is not ideal as it ignores information about other firms operating in the same environment. Additionally, Phillips and Perron (1987) highlight that we encounter the problem of serial correlation due to the time-dependent nature of the dependent variable. Therefore, time series regression is not a suitable approach in this case.

Andrews (2005) points out that, when using a Cross-Section Regression model, a cross-sectional regression is estimated for each year. This means that if we have n years, we will have n cross-section regressions, one for each year. This approach severely limits the degrees of freedom required to perform a meaningful and comprehensive analysis. It is limited because, for each year, we have only r observations of subjects (or firms) with which to estimate r parameters. Therefore, this approach is not effective.

Borenstein et al. (2010) suggest that in the Between-Group Cross-Section Regression model, all subjects (e.g., firms) are included in the model by taking the average of each subject's (firm's) values over the entire period. However, this method, as discussed in the literature, faces the challenge of erasing time-dependent information due to averaging and also limits the sample size. Therefore, this approach is also not ideal.

Panel data regression is highlighted as the best approach in research, according to Borenstein et al. (2010). This model allows for more observations, more cross-sectional information, and more degrees of freedom. It incorporates changes within a firm as well as changes across firms. Additionally, it accounts for the impact of subject-specific (firm-specific) attributes, such as location, management philosophy, customer orientation, and culture. Therefore, this is the approach adopted in this research.

Ahn and Schmidt (1996) explain that we may estimate the panel data model using pooled OLS (Ordinary Least Square). In this technique, we simply pool all given observations and estimate one large OLS regression. The assumption is that regression coefficients are the same for all firms, regressors are non-stochastic (i.e., errors are not correlated with explanatory variables), and the error terms are identically and independently distributed with a mean of zero and constant variance. Additionally, we can estimate the panel data model using Fixed Effects Model (FEM) and Random Effects Model (REM).

The Fixed Effects Model includes techniques such as Least Squares Dummy Variables (LSDV), Within-Group (WG), and First Differences (FD). The Pooled OLS result often shows a high likelihood of heterogeneity, which refers to unobserved firm-specific characteristics, such as location and board diversity. Since these firm characteristics vary across subjects but are time-invariant (i.e., fixed through time), pooling firms with different characteristics into one OLS estimation camouflages these fixed effects. This can lead to endogeneity problems, where the error term is correlated with one or more regressors, causing the estimated regression coefficients to be biased and inconsistent, leading to erroneous inferences. To address this issue, FEM and REM are employed, as they account for unobserved heterogeneity.

Borenstein et al. (2010) state that Fixed Effects Models explicitly account for firm heterogeneity by breaking down the error term into a firm-dependent error term and an idiosyncratic error term. The Least Squares Dummy Variable (LSDV) model captures heterogeneity by allowing different intercepts for each firm in the pooled data using dummy variables. The differences in intercepts reflect the unique characteristics of the firms. The term ”fixed effect” arises because, although the intercept varies across firms, it remains fixed over time, making it time-invariant.

The coefficients from the Fixed Effects Model produce estimators known as fixed effect estimators. This is a one-way fixed effects model because the intercepts vary only across firms (to account for heterogeneity) but not across time. A decision between models can be made using a Restricted or Partial F test and a Wald test of differential intercepts. The FEM is accepted, and the Pooled OLS model is rejected if the p-value is significant and if the F-calculated value is greater than the critical value of F.

When heterogeneity exists, the Fixed Effects Model within Group can be used. By expressing each variable as a deviation from its time-mean, we effectively eliminate heterogeneity in the data. This method removes long-run effects, leaving only short-run dynamics.The Random Effects Model (REM), also known as the error components model, incorporates firm heterogeneity within the error term rather than specifying it as dummy variables. It allows for a common intercept but treats it as a random variable with a mean and a random firm-specific error term. Unlike the FEM, where each firm has its own fixed intercept value, the REM uses the average of all firms' intercepts as the common intercept. The firm-specific error component measures the random deviation of each firm's intercept from the common intercept. Estimating with pooled OLS can lead to serially correlated errors; therefore, the REM addresses this issue using the Generalized Least Squares (GLS) estimation approach.

In hypothesis testing, Ahn and Schmidt (1996) explain that the null hypothesis considers the REM as the appropriate estimator, while the alternative hypothesis suggests that the FEM is more suitable. If the null hypothesis (Ho) is rejected, it indicates that the REM is not appropriate because the random effects are likely correlated with the independent variables.

Chapter 3

METHODOLOGY

This chapter highlights the methodology used in this study. This chapter dwells on type and sources of data collected, research design, the study countries, data processing and analysis and model estimation.

3.1 Type and Sources of Data

The study employed panel data by compiling and analyzing green coffee data of EAC selected countries. Data retrieved from FAO, United Nations Conference on Trade and Development (UNCTAD), World Bank (WB) and other online sources for the period 1996 to 2022.
The annually recorded data of green coffee recorded from selected East African Countries(Rwanda, Burundi, Tanzania, Uganda, Kenya) collected by FAO (Food and Agriculture Organization) from 2000 to 2021 will be applied in this study. Given that the method used allows for the insertion of data and the generation of estimates from data that were actually measured by the nation, data gaps will not be filled .

3.2 Study Countries

The study was conducted in selected EAC countries. Currently, there are seven EAC partner states beneficiaries under this arrangement; Tanzania, Kenya, Uganda, Rwanda, Burundi, Democratic Republic of Congo(DRC) and South Sudan. This study selected the top five coffee-producing countries in the East African Community (EAC). EAC is a regional bloc mandated by the governments of Tanzania, Kenya, Uganda, Rwanda, Burundi, Democratic Republic of Congo(DRC) and South Sudan to spearhead the East African economic, social, and political integration agenda. EAC countries can enter into international agreements either through bilateral or multilateral systems to seek trading opportunities. Focusing on coffee export, this study reveals how selected EAC countries have comparative advantages in the region in producing and exporting coffee.

3.3 Model construction

To construct an empirical model to estimate the effects of various factors on coffee price fluctuations, we will define the variables and specify the regression equation. This model will help us understand how different factors contribute to changes in coffee prices.

Variables Definition

- Coffeep_{it}: Coffee price for entity i at time t.

- Agriland$_{it}$: Land covered by coffee plantation for entity i at time t.
- Fertilizerp$_{it}$: Fertilizer price for entity i at time t.
- Interest$_{it}$: Interest rate for entity i at time t.
- Inf$_{it}$: General price inflation rate for entity i at time t.
- Agriem$_{it}$: Agricultural employment for entity i at time t.
- Ex$_{it}$: Official exchange rate of country i at time t.
- ϵ_{it}: Error term capturing other factors affecting Coffeep$_{it}$.

Empirical Model Specification

The empirical model is specified as follows:

$$\begin{aligned} Coffeep_{it} &= \beta_0 + \beta_1 Agriland_{it} + \beta_3 Interest_{it} + \beta_4 inf_{it} \\ &+ \beta_5 Agriem_{it} + \beta_6 Fertilizerp_{it} + \beta_7 Ex_i t + \epsilon_{(t)} \end{aligned} \tag{3.1}$$

where

Intercept (β_0): Represents the baseline level of coffee prices when all other explanatory variables are zero.

To estimate this equation (3.1), we will run a Panel-VAR model, for a sample of some of East African countries. Furthermore, some preliminary tests, such as a unit root test and a stability test are used to ensure the reliability of Panel-VAR estimation results.

3.4 Unit Root Stochastic Process

Time series variables can fail to be stationary in various ways, but two are especially relevant for regression analysis of economic time series data: First, the series can have persistent, long-run movements, that is the series can have trends. Second, the population regression may be unstable over time, that population regression can have breaks[11]. There are two types of random walk: Random walk without drift (i.e no constant or intercept term) and Random walk with drift (i.e a constant term is present)

3.5 Random walk without drift

A random walk without drift is a simpler version of the random walk model where there is no deterministic trend or consistent directional movement. Instead, the process is purely driven by random fluctuations around a starting point.

In a random walk without drift, the formula is:

$$y_t = y_{t-1} + \epsilon_t \tag{3.2}$$

where:

- y_t is the value at time t,
- y_{t-1} is the value at time $t-1$,
- ϵ_t is the random error term, typically assumed to be drawn from

a distribution with a mean of zero (e.g., normal distribution with mean 0).

3.6 Random walk with drift

The random walk model is an example of what is known as unit root process. a random walk with drift can be represented as:

$$y_t = y_{t-1} + \mu + \epsilon_t \tag{3.3}$$

where:

- y_t is the value at time t,
- y_{t-1} is the value at time $t-1$,
- μ is the drift term, and
- ϵ_t is the random error term.

3.7 Testing for Unit Roots

Autocorrelation coefficients for different time lags for a variable is used to answer different problems, namely, are the data nonstationary or stationary? Are the data random? etc. If a series is random, the autocorrelations (ACF) between Y_t and Y_{t-k} for any lag k are close to zero[2]. If a series has a trend, successive observations are highly correlated and (ACF) coefficients are significantly different from zero.

3.8 Augmented Dickey-Fuller Test for Unit Roots

There are three possible forms of the ADF test as follows:

$\Delta Y_t = \delta Y_{t-1} + \sum_{i=1}^{p} \beta_i \Delta Y_{t-i} + u_t$

$\Delta Y_t = \alpha + \delta Y_{t-1} + \sum_{i}^{p} \beta_i \Delta Y_{t-i} + u_t$

$\Delta Y_t = \alpha + \gamma^T + \delta Y_{t-1} + \sum_{i}^{p} \beta_i \Delta Y_{t-i} + u_t$

The above three equations are different due to the presence of the deterministic elements α and γ^T. The nature of the structural modeling is to discover the dynamic causal relationship between Y_t and X_t we must distinguish between short-run and long-run relationship. We consider the simple autoregressive distributed lag model [ARDL (1,1)] in the following

$$Y_t = A_0 + A_1 + Y_{t-1} + B_0 X_t + B_1 X_{t-1} + u_t \tag{3.4}$$

The short-run relationship between Y_t and x_t can be represented as $\frac{\delta Y_t}{X_t} = \beta_0$

This relationship is instantaneous and does not consider any lagged effects.

$$\bar{Y_t} = \frac{A_0 + (B_0 + B_1)\bar{X}_t}{1 - A_1} \tag{3.5}$$

This relationship considers the effects of lagged variables and represents the long-run relationship between Y_t and X_t

3.9 ARDL and Error Correction Model (ECM)

By subtracting Y_{t-1} both sides of equation (3.4) and rearranging, we have

$$Y_t - Y_{t-1} = A_0 + A_1 Y_{t-1} - Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + u_t$$

Rearrange the terms:

$$Y_t - Y_{t-1} = A_0 + (A_1 - 1)Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + u_t$$

Define the first difference ΔY_t and ΔX_t:

$$\Delta Y_t = Y_t - Y_{t-1}$$

$$\Delta X_t = X_t - X_{t-1}$$

Substitute $X_t = X_{t-1} + \Delta X_t$ into the equation:

$$\Delta Y_t = A_0 + (A_1 - 1)Y_{t-1} + \beta_0 (X_{t-1} + \Delta X_t) + \beta_1 X_{t-1} + u_t$$

Simplify by combining like terms:

$$\Delta Y_t = A_0 + (A_1 - 1)Y_{t-1} + \beta_0 X_{t-1} + \beta_0 \Delta X_t + \beta_1 X_{t-1} + u_t$$

$$\Delta Y_t = A_0 + (A_1 - 1)Y_{t-1} + (\beta_0 + \beta_1) X_{t-1} + \beta_0 \Delta X_t + u_t$$

$$\mathrm{ECT}_{t-1} = Y_{t-1} - \gamma X_{t-1}$$

where γ represents the long-run coefficient on X in the ARDL model. To find γ, compare the coefficient of X_{t-1}:

$$\gamma = -\frac{\beta_0 + \beta_1}{1 - A_1}$$

Thus:

$$\mathrm{ECT}_{t-1} = Y_{t-1} - \frac{\beta_0 + \beta_1}{1 - A_1} X_{t-1}$$

Substitute ECT_{t-1} into the equation:

where $\pi = 1 - A_1$.

Thus:

$$\Delta Y_t = \beta_0 \Delta X_t - \pi \mathrm{ECT}_{t-1} + u_t \tag{3.6}$$

The ECM and ARDL models are basically the same if the series Y_t and X_t are integrated of the same order (often $I(1)$) and cointegrated. Y_t and X_t are assumed to be in long-run equilibrium. If Y_{t-1} deviates from the optimal value, there is a correction. Speed of adjustment is given by $\pi = (1 - A_1)$, which is between 0 and 1.

3.9.1 Empirical model

Variable	**Abbreviation**	**Sources**
Agriculture Production Index	Coffeep	FAOSTAT (Dependent)
Exchange Rate	Exch	WORLD BANK (Independent)
Agriculture Employment	Agriemp	WORLD BANK (Independent)
Fertilizer	Fertilizerp	WORLD BANK (Independent)
Agriculture Land	Agriland	FAOSTAT (Independent)
Inflation	Infl	WORLD BANK (Independent)
Foreign Direct Investment	FDI	WORLD BANK (Independent)

Table 3.1: Types and sources of data in the research

Table 3.1 shows the variable specifications in the model presented in (3.1). Cointegration tests have been introduced since the 1980s, mainly dealing with the problem of regressing non-stationary time series, given that several economic time series are non-stationary. The cointegration tests in the Autoregressive Distributed Lag (ARDL) approach introduced by Pesaran and Shin (1999) and developed by Pesaran (2001) are suitable compared to the rest of cointegration tests, as they are suitable for finite samples with small sizes. The ARDL approach is used for detecting cointegration between time series in the long run unless any of the time series variables is level two (I_2). It starts by testing the stationarity of the variables using unit root tests. If the variables are not I_2 variables, the Autoregressive Distributed Lag is not recommended. Bounds testing is recommended to examine the existence of cointegration between the model variables in the long run, as proposed by Pesaran (2001) and Youssef (2023).

If the computed F-Statistic of the Bounds test is above the critical value for the upper bound, then the regressors are considered cointegrated. Conversely, if the F-Statistic is below the critical value for the lower bound, there is no cointegration. If the model variables are cointegrated based on the results of the bounds test, long-run coefficients of the model variables will be estimated using the conditional ARDL model with optimal lags $(p, q_1, q_2, q_3, q_4, q_5, q_6)$ of the following form:

$$
\begin{aligned}
&ln\text{Coffeep}_{it} = \beta_0 + \sum_{i=1}^{p} \beta_{1i} ln\text{Coffeep}_{t-1} + \sum_{i=0}^{q_1} \beta_{2i} \ln \text{Exch}_{t-1} + \\
&\sum_{i=0}^{q_2} \beta_{3i} \ln \text{Agriemp}_{t-1} + \sum_{i=0}^{q_3} \beta_{4i} \Delta \ln \text{Fertilizerp}_{t-1} + \sum_{i=0}^{q_4} \beta_{5i} \ln \text{infl}_{t-1} \\
&\sum_{i=0}^{q_5} \beta_{6i} \ln \text{Agriland}_{t-1} + \sum_{i=0}^{q_6} \beta_{7i} \ln \text{FDI}_{t-1} + \varepsilon_t .
\end{aligned}
\tag{3.7}
$$

In the final step, an Error Correction Model (ECM) is estimated using the Ordinary Least Squares (OLS) method to estimate the short-run coefficients of the model variables, in addition to an error correction term (ECT_{t-1}) that shows the speed of adjustment of the variables towards the long-run equilibrium. The conditional ECM $(p, q_1, q_2, q_3, q_4, q_5, q_6)$ will take this form:

$$
\begin{aligned}
&\Delta \ln \text{CoffeeP}_{it} = \beta_0 + \sum_{i=1}^{p} \beta_{1i} \Delta \ln \text{CoffeeP}_{t-1} + \sum_{i=0}^{q_1} \beta_{2i} \Delta \ln \text{Exch}_{t-1} \\
&+ \sum_{i=0}^{q_2} \beta_{3i} \Delta \ln \text{Agriemp}_{t-1} + \sum_{i=0}^{q_3} \beta_{4i} \Delta \ln \text{Fertilizerp}_{t-1} + \sum_{i=0}^{q_4} \beta_{5i} \Delta \ln \text{infl}_{t-1} \\
&+ \sum_{i=0}^{q_5} \beta_{6i} \Delta \ln \text{Agriland}_{t-1} + \sum_{i=0}^{q_6} \beta_{7i} \Delta \ln \text{FDI}_{t-1} + \delta \text{ECT}_{t-1} + \varepsilon_{it} .
\end{aligned}
\tag{3.8}
$$

where β_0 is the constant term, β_{1i} to β_{7i} are the coefficients, ε_{it} is the error term, i represents countries, and t denotes the time period (1996, 1997, ..., 2022).

3.9.2 Decision Criteria for Bounds Test:

If series are integrated of different orders, i.e., a combination of both level and first difference or second difference stationarity, it is necessary to perform a cointegration test to establish the long-run relationship. However, in this case, the Johansen cointegration test is no longer valid. Hence, the appropriate cointegration test is the Bounds test proposed by Pesaran and Smith (2001). The hypotheses are stated as:

H_0: no cointegration

H_1:Cointegration exists (i.e., H_0 is not true)

Rejection at the 10%, 5%, or 1% level.

If the calculated F-statistic is greater than the critical value for the upper bound I(1), then we can conclude that there is cointegration, indicating a long-run relationship. We reject the null hypothesis and estimate the long-run model, which is the Error Correction Model (ECM). If the calculated F-statistic is lower than the critical value for the lower bound I(0), then we conclude that there is no cointegration, hence no long-run relationship. We do not reject the null hypothesis and estimate the short-run model, which is the Autoregressive Distributed Lag (ARDL) model.

If the F-statistic falls between the lower bound, $I(0)$, and the upper bound, $I(1)$, the test is considered inconclusive.

3.10 General linear panel data model

The general linear panel data model used in this study is specified as follows:

$$\begin{aligned} \text{CoffeePr}_{it} &= \beta_0 + \beta_1\text{Exch}_{it} + \beta_2\text{Agriemp}_{it} + \beta_3\text{Fertilizer}_{it} \\ &+ \beta_4\text{infl}_{it} + \beta_5\text{Agriland}_{it} + \beta_6\text{FDI}_{it} + \varepsilon_t. \end{aligned} \tag{3.9}$$

Where $\varepsilon_t = u_i + v_{it}$ and u_i represents anything we cannot observe and remains constant over time, while v_{it} is the idiosyncratic error term. We will control for all the fixed effects u_i by including dummies for all i's, but this becomes problematic for large N. Hence, panel data models offer techniques to remove u_i. Here we mention the Random Effects model, which should be used when there is no correlation between u_i and explanatory variables. If you have reason to believe that differences across entities have some influence on your dependent variable, then you should use the Random Effects model.
Fixed effect models should be used when there is a rationale behind the assumption of correlation between the entity's error term and predictor variables. It is designed to study the causes of changes within a person or entity.

Pesaran (2021) claimed that without taking into consideration cross-sectional dependency in panel data analysis, this may result in errors and misleading information. To identify cross-sectional dependency, Pesaran (Breusch and Pagan, 1980) came up with Pesaran CD and

Standardized Lagrange Multiplier (LM) tests. Pesaran suggested that for large panel data sizes N and time T, the standardized test should be used and may be approximated as follows.

$$LM = \sqrt{\frac{1}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (T_{ij}\mu_{ij}^2 - 1) \rightarrow N(0,1). \tag{3.10}$$

This formula is used for big size and variable time T

$$CD = \sqrt{\frac{2}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} T_{ij}\mu_{ij}^2 \rightarrow N(0,1). \tag{3.11}$$

This formula is used for large N and constant T. While Breusch and Pagan (Pesaran, 2021) proposed the Breusch-Pagan LM test, which is efficient for small size N and time T. It is approximated as follows:

$$LM = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} T_{ij}\mu_{ij}^2 \rightarrow \chi^2 \left(\frac{N(N-1)}{2} \right) \tag{3.12}$$

Where μ_{ij}^2 is the correlation coefficient obtained from the residuals of equation (3.14) and computed as follows.

$$\mu_{ij} = \mu_{ji} = \frac{\sum_{t=1}^{T} \varepsilon_{ij,t}\varepsilon_{ji,t}}{\left(\sum_{t=1}^{T} \varepsilon_{ij,t}^2\right)^{\frac{1}{2}} \left(\sum_{t=1}^{T} \varepsilon_{ji,t}^2\right)^{\frac{1}{2}}} \tag{3.13}$$

Where $\varepsilon_{ij,t}$ and $\varepsilon_{ji,t}$ represents the ij^{th} error term at time **t**.

Chapter 4

RESULTS AND DISCUSSION

In this chapter, we present analysis and interpretation of various data especially, coffee consumption, Exchange rate, Agriculture, Employment, Fertilizer used in coffee plantation, Inflation, Agriculture land and Foreign direct investment in some selected East African countries. We explore the long-run and short-run effect of above variables of coffee production and consumption in East Africa countries using Fixed Effect and Random Effect model.

Figure 4.1 shows the use of Urea fertilizers in five countries of East

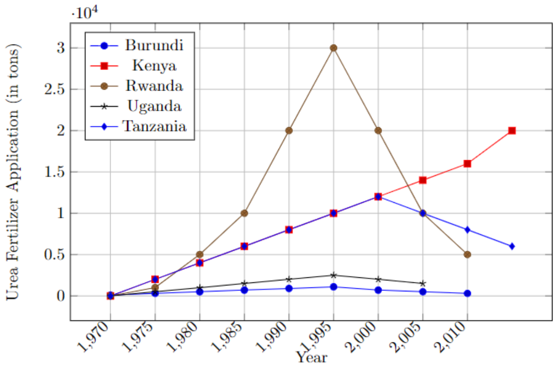

Figure 4.1: Urea Fertiliser application in selected EAC countries

Africa. This indicates that some of these countries are no longer using Urea fertilizers as it is indicated in figure 4.1, which should be one of the main reasons for the reduction of coffee production in Easter Africa. But Figure 4.2 shows that East African countries adopt the use of insecticide for agriculture which is increasingly significant since 2000 except Tanzania and Burundi.

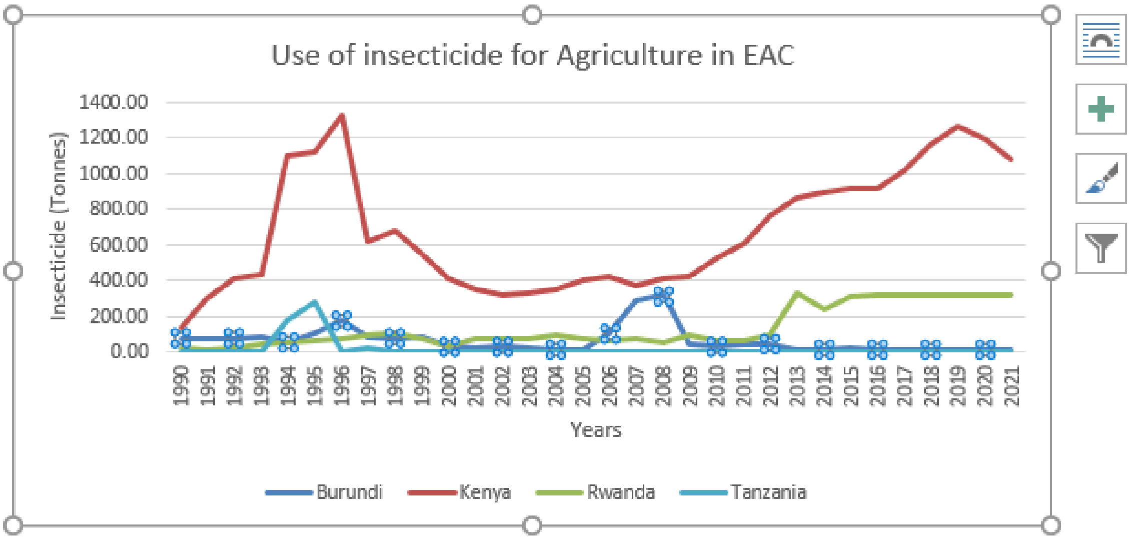

Figure 4.2: Use of Insecticide for agriculture in selected EAC countries

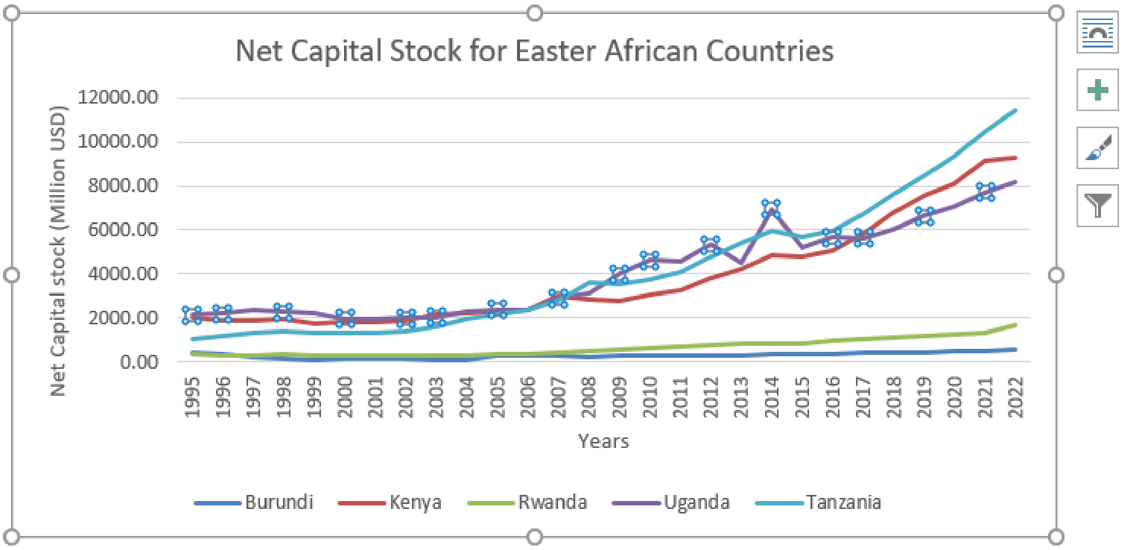

Figure 4.3: Net Capital Stock for selected EAC Countries

We see that Figure 4.3 reflects the increase in Net Capital Stock in the five East African countries, namely Burundi, Kenya, Rwanda, Uganda, and Tanzania. This indicates that the market value of the stock of fixed assets in the economy is accelerating, contributing to the overall wealth of the country.

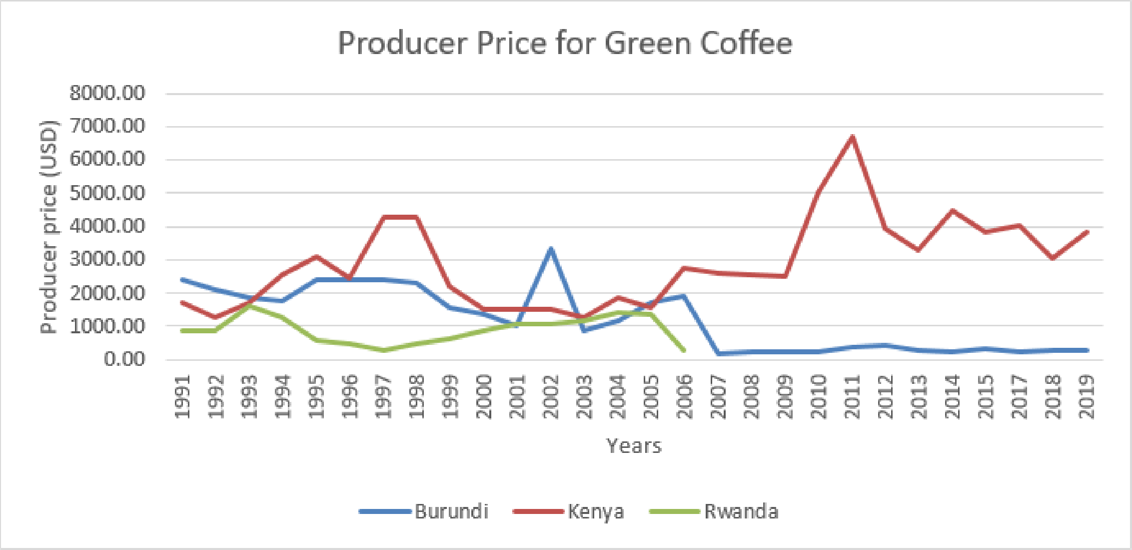

Figure 4.4: Producer Price for green coffee in selected EAC countries

Figure 4.4 above indicates that the producer price of green coffee , which shows too much change fluctuation in Kenya and Rwanda and seems to be constant in Burundi. Unfortunately, Data from Uganda and Tanzania are not available.

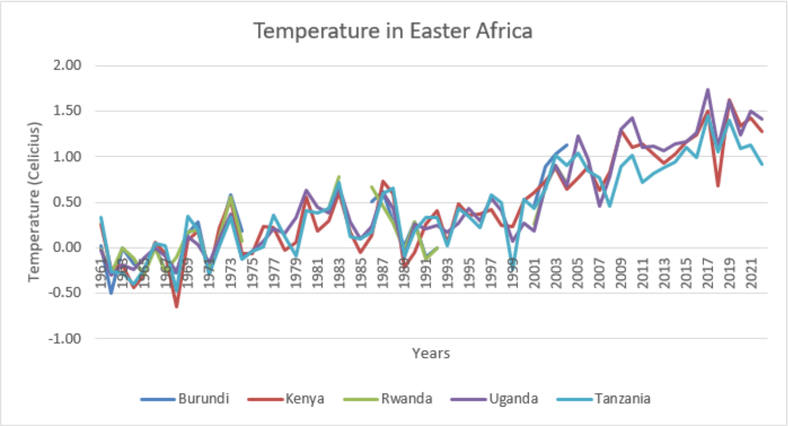

Figure 4.5: Temperature in selected EAC Countries

Temperature in East African Countries is increasing with time as it is indicated in Figure 4.5, and this may also affect the coffee price in

East African Countries.

Table 4.1: Area Covered by Herbaceous Crops in selected EAC Countries

Year	Burundi	Kenya	Rwanda	Uganda	Tanzania
2020	772.28	2089.37	810.17	1592.19	15890.23
2021	672.65	2638.55	798.94	2714.28	15028.28

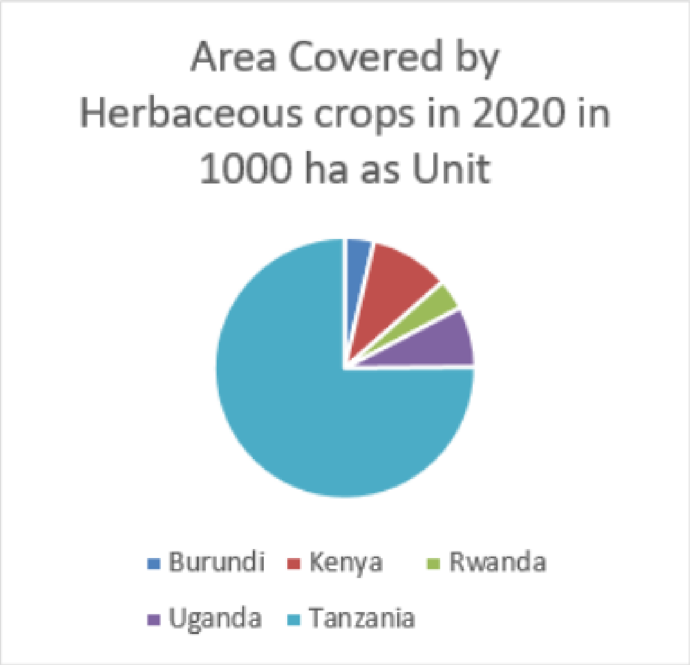

Figure 4.6: Area Covered by Herbaceous Crops in selected EAC Countries

The following Figure 4.7 shows the credit used to agriculture in Burundi, Uganda and Tanzania. We see that the curve is increasing except Burundi which seems to have little credit oriented to agriculture. Data for countries namely Rwanda and Kenya their data are not available. Without enough credit to agriculture may also reduce the green coffee price in EAC.

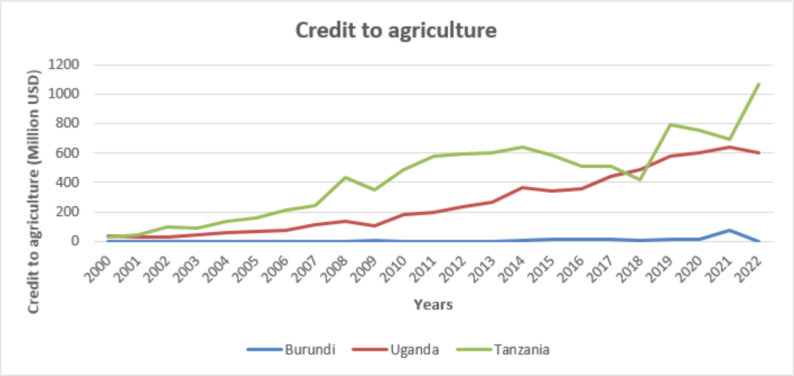

Figure 4.7: Credit to agriculture in EAC Countries

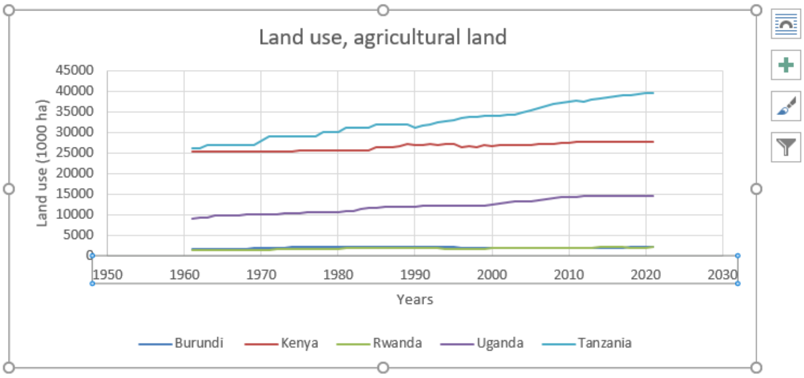

Figure 4.8: Land Use for agriculture Crops in selected EAC Countries

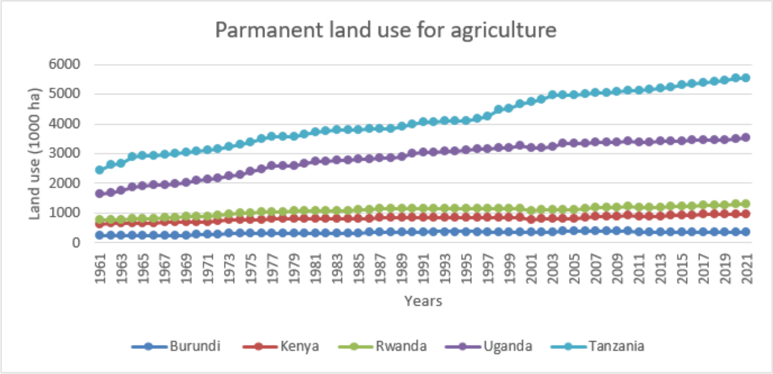

Figure 4.9: Permanent land use in selected EAC countries

We can see that Figure 4.8 and Figure 4.9 show that the permanent land and agricultural land use for agriculture has increased as people shift from rural areas and live in town, boosting the increase in land use.

Table 4.2: Population growth in East African Countries

Year	Country	Population	Population in 2024
2012	Burundi	12889576	13452423 (with 2.7% increase)
2014	Uganda	47249585	49924252 (with 3.2% increase)
2015	Kenya	54027487	56203030(with 1.9% increase)
2018	Tanzania	65497748	69419073(with 3.0% increase)
2019	Rwanda	13776698	14288145 (with 2.4% increase)

From Table 4.2, it is clear that the population of Uganda, Tanzania, Burundi, Rwanda, and Kenya is increasing dramatically, respectively.

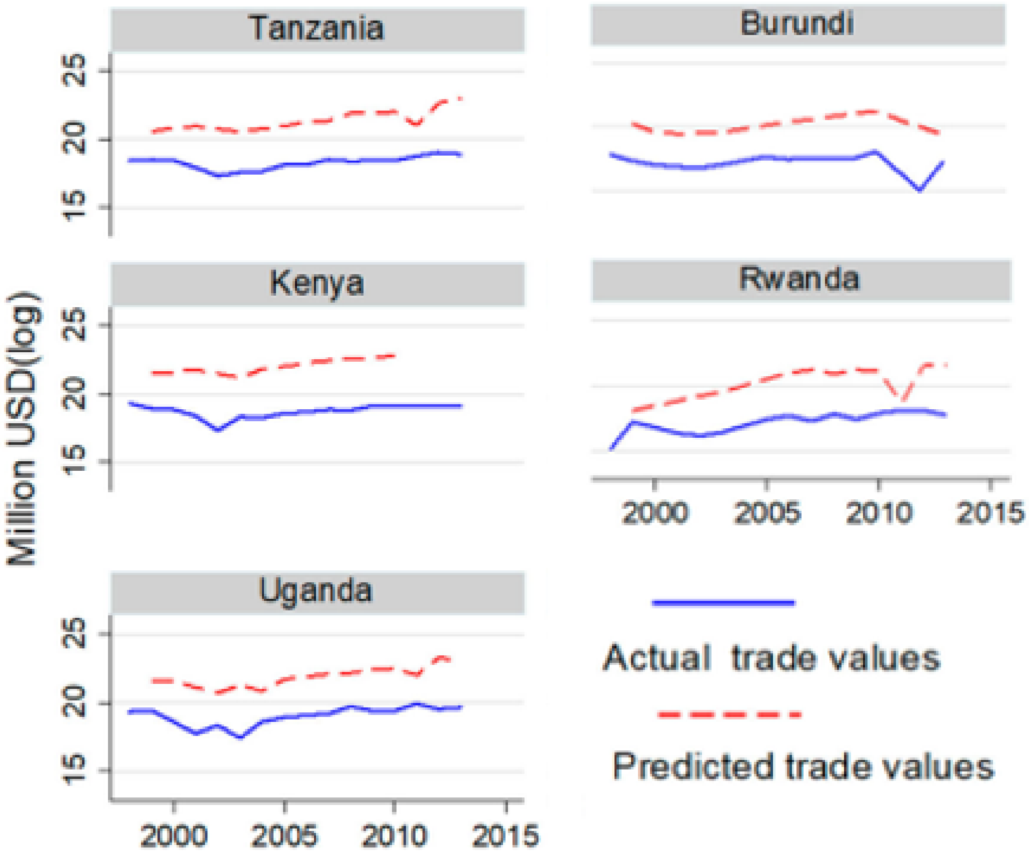

Figure 4.10: Coffee trade potentials of Eastern African countries

Figure 4.10 shows the trade of eastern African coffee; we see that eastern African coffee trade is promising as it has been increasing with time. Unfortunately, we did not succeed in finding current data for all Eastern African Countries.

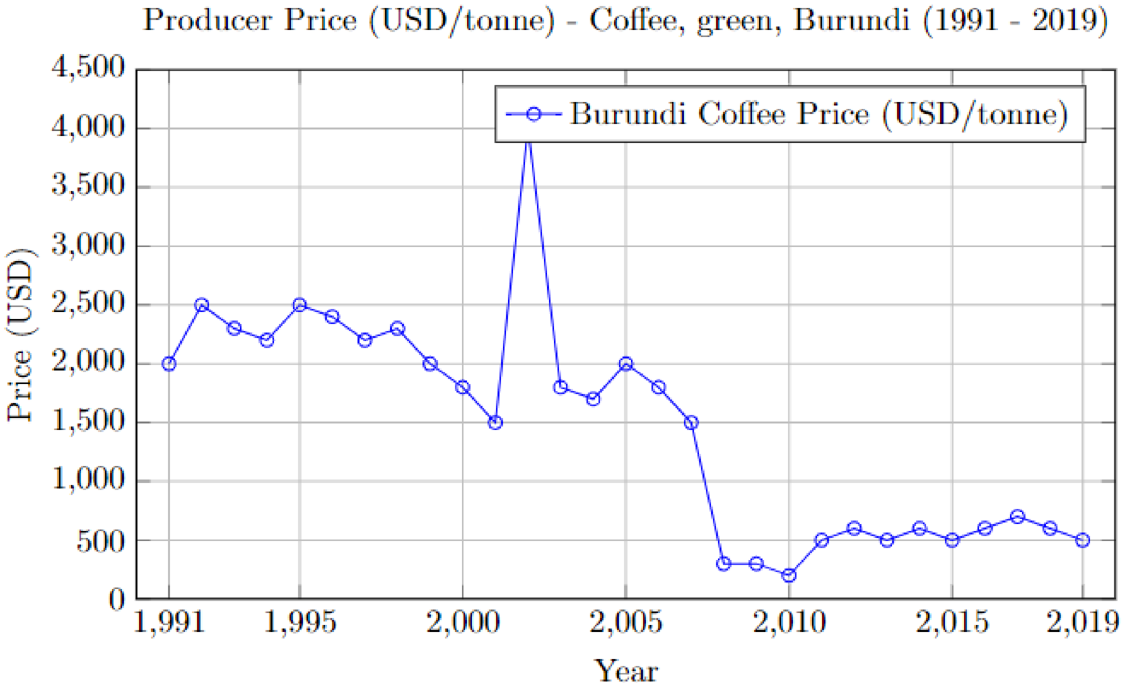

Figure 4.11: producer price of green coffee in Burundi from 1991 to 2019

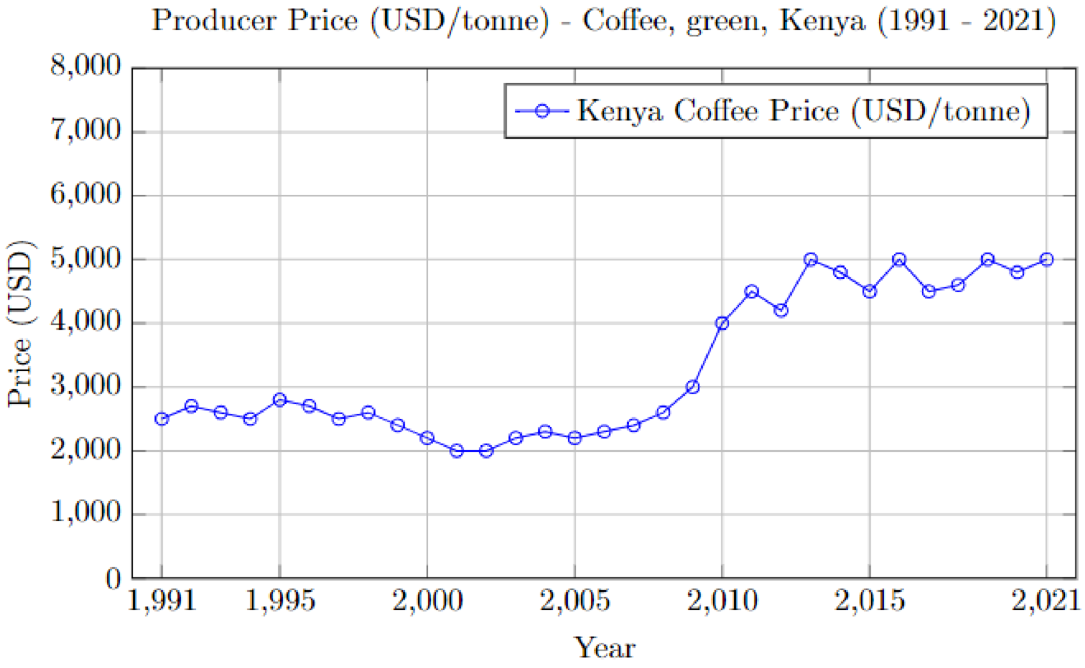

Figure 4.12: producer price of green coffee in Kenya from 1991 to 2021

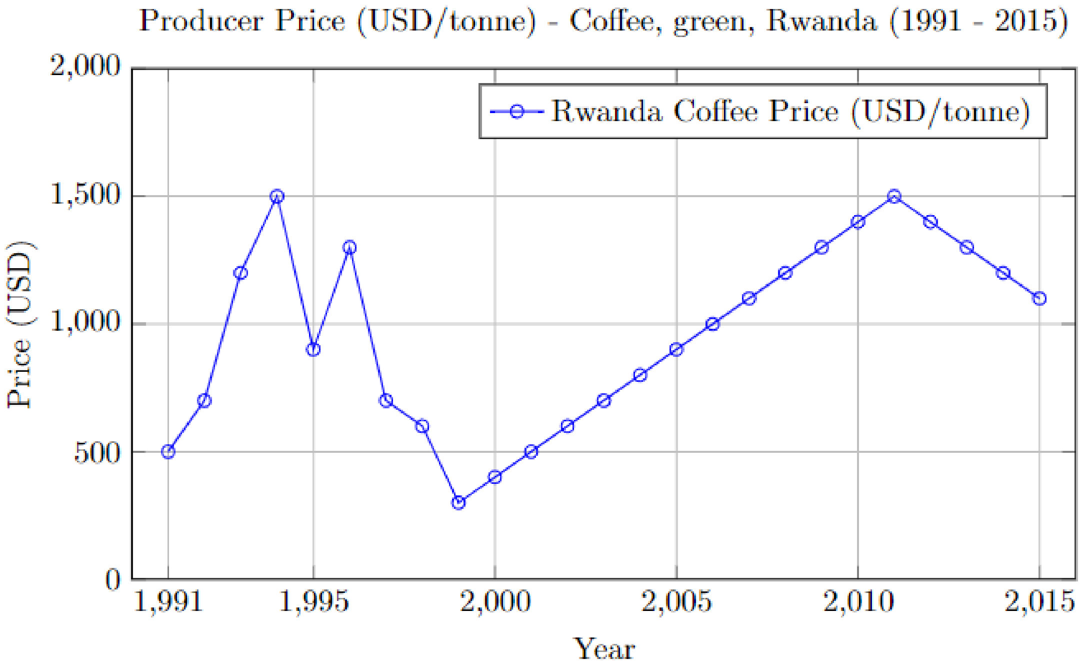

Figure 4.13: Rwanda coffee prices from 1991 to 2021

The graph 4.11,4.12 and 4.13 illustrates a volatile period of producer prices for green coffee (USD/tonne) of Rwanda, Kenya, and Burundi from 1991 to 2021, with significant price changes over the years. Data from Uganda and Tanzania are not Available .

Table 4.3: Descriptive Statistics

Variable	Observation	Mean	Std.Dev	Min	Max
Consumpr	135	7.912763	5.78077	-2.814698	31.11159
Exch	135	1151.772	942.4379	57.11487	3727.069
Agriemp	135	69.4336	17.62473	33.0281	92.59169
Fertilizer	135	13.27564	16.11367	0.0793919	65.815
Infl	135	7.912763	5.78077	-2.814698	31.11159
Agriland	135	61.74908	14.8463	37.73085	81.89252
FDI	135	1.834028	1.597258	-0.0588052	6.656597

4.1 Summary output by using pooled ordinary linear regression

In the analysis of coffee price in East Africa, a pooled ordinary linear regression model was employed, assuming :

- Regression coefficients are the same for all countries.
- Regressors are nonstochastic, i.e., errors are not correlated with explanatory variables $\text{cov}(V_{it}, X_{it}) = 0$.
- Error term $V_{it} \sim \text{iid}(0, \sigma_v^2)$ (i.e., error term should be white noise).

Table 4.4: Regression Statistics

Multiple R	0.86
R Square	0.73
Adjusted R Square	0.72
Standard Error	11.73
Observations	135

Table 4.5: ANOVA

	df	SS	MS	F
Regression	6	48109.4	8018.24	58.26
Residual	128	17617.1	137.63	
Total	134	65726.5		

We can see that the regression as a whole is statistically significant from the ANOVA table. Indeed, Exchange rate, Fertilizer, Inflation, Agriculture land, and Foreign

Direct Investment have a significant impact on Coffee Production in the five countries of East Africa. Only Inflation and agriculture employment do not influence coffee price in selected East Africa.

By observing the p-value of Exchange rate, Fertilizer, Inflation, and Agriculture land, we see that they are both statistically significant. For good measure, R-squared, which is the coefficient of determination and a measure of goodness of fit, tells us that more than 73% of the variation of coffee price is explained by the pooled ordinary linear regression.

Unfortunately, linear regression is not an appropriate method for the data we have as it gives us general information about the five counties considered in the study. It does not capture the relationship of all explanatory variables to coffee price in each individual country among the five countries. As we are analyzing time series data, it is very crucial also to test the stationarity in our data which is not the case for linear regression. Therefore, the panel model is the right model for the data we have.

Since panel data include different countries with different characteristics, the likelihood of heterogeneity exists which refers to unobserved country-specific characteristics (i.e., location, management, etc.). As a remedy, we use fixed and random effect models, both take into account unobserved heterogeneity which may cause the correlation among explanatory variables.

4.2 Choosing between Fixed Effect and Random Effect Models

We perform Hausman's (1978) specification test where:

H_0: Random effects model is consistent

H_a: Fixed effects model is consistent

If the p-value of the test is greater than 0.05, we accept the null hypothesis (H_0).

If the p-value of the test is less than 0.05, we reject the null hypothesis (H_0). Hence, the Fixed effect model (FE) is preferred and consistent

Table 4.6: Random Effects Model Coefficients

CoffeePr	**Coef.**	**Std. Err.**	**z**	**P> |z|**	**[95% Conf. Interval]**
Exch	0.0121738	0.0013553	8.98	0.000	[0.0095174, 0.0148301]
Agriemp	-0.0945582	0.1221551	-0.77	0.439	[-0.3339778, 0.1448615]
Fertilizer	0.8566795	0.1396765	6.13	0.000	[0.5828109, 1.1305482]
Inf	-0.4022946	0.1827242	-2.20	0.028	[-0.7604213, -0.0441678]
Agriland	0.0087022	0.0797024	12.66	0.000	[0.8524883, 1.1649156]
FDI	1.705903	0.7734373	2.21	0.027	[0.1991739, 3.2126319]
_cons	10.57371	11.01149	0.96	0.337	[-11.00841, 32.15583]

Random-effects GLS regression	
Number of obs	135
Group variable: id	
Number of groups	5
R-sq:	
within	0.6856
between	0.9625
overall	0.7320
corr(u_i, X)	0 (assumed)
Wald chi2(6)	349.55
Prob > chi2	0.0000

Table 4.7: Random Effects Model Summary

σ_u	0
σ_e	10.876566
ρ	0 (fraction of variance due to u_i)

Table 4.8: Variance Components

Table 4.9: Fixed Effects Model Coefficients

CoffeePr	**Coef.**	**Std. Err.**	**t**	**P> |t|**	**[95% Conf. Interval]**
Exch	0.0079592	0.0027831	2.86	0.005	[0.0024506, 0.0134678]
Agriemp	-0.8681348	0.2126447	-4.08	0.000	[-1.289058, -0.4472117]
Fertilizer	0.6857405	0.0741244	9.23	0.000	[0.5395035, 0.8319775]
Inf	-0.3775008	0.0924043	-4.08	0.000	[-0.5603475, -0.1946541]
Agriland	0.089871	0.4474421	0.20	0.839	[-0.7949229, 0.9746649]
FDI	0.546107	0.2523937	2.16	0.033	[0.049472, 1.042742]
_cons	68.32289	31.68554	2.16	0.033	[5.61753, 131.0282]

Table 4.10: Fixed Effects Model Summary

Fixed-effects (within) regression	
Number of obs	135
Group variable: id	
Number of groups	5
R-sq:	
within	0.7192
between	0.3097
overall	0.5204
corr(u_i, Xb)	-0.5373
F(6,124)	52.94
Prob > F	0.0000

Table 4.11: Variance Components

σ_u	14.810517
σ_e	10.876566
ρ	0.64963926 (fraction of variance due to u_i)
F test that all $u_i = 0$: F(4, 124) = 6.23	Prob > F = 0.0001

Table 4.12: Haussmann test

Coefficients	(b) **FE**	(B)**RE**	(b-B)**Difference**	$Sqrt(diag(v_b - v_B))$ **S.E.**
Exch	0.0079592	0.0121738	-0.0042146	0.0024308
Agriemp	-0.8681348	-0.0945582	-0.7735766	0.1740816
Fertilizer	0.6857405	0.8566795	-0.170939	0.1045184
Inf	-0.3775008	-0.4022946	0.0247938	0.1942338
Agriland	0.089871	0.0087022	0.0811688	0.4402649
FDI	0.546107	1.705903	-1.159796	0.4852757

Where

Coefficients: These are the variables included in the model.

(b) FE: Coefficients from the Fixed Effects model.

(B) RE: Coefficients from the Random Effects model.

(b-B) Difference: The difference between the Fixed Effects and Random Effects coefficients.

$Sqrt(diag(v_b - v_B))$ **S.E:** The standard error of the differences.

Test:

Ho: Difference in coefficients not systematic

$$\chi^2(6) = (b - B)' \left[V_b - V_B\right]^{-1} (b - B) \tag{4.1}$$

$\chi^2(6) = 31.30$

(Prob $> \chi^2$): 0.0000

Note: $(v_b - v_B)$is not positive definite

From the Hausman test, it is clear that the probability of Chi-squared (Prob $> \chi^2$) is significant. Therefore, we reject the null hypothesis and conclude that the Fixed Effect model is consistent.

Table 4.13: Regression Statistics

Multiple R	0.881
R Square	0.777
Adjusted R Square	0.751
Standard Error	10.877
Observations	135

Table 4.14: ANOVA for Housman test

Source	**Df**	**SS**	**MS**	**F**	**F**
Regression	11	51057.37	4641.5794	43.15934975	0.00
Residual	124	14669.16	118.29968		
Total	135	65726.53			

Table 4.15: Coefficients

Variable	**Coefficients**	**Standard Error**	**t- Stat**	**P-value**	**Lower 95%**	**Upper 95%**
Intercept	80.90	38.93	2.08	0.04	3.85	157.95
D1	(10.07)	4.88	(2.06)	0.04	(19.73)	(0.41)
D2	-	-	65535.00	#NUM!	-	-
D3	(37.60)	15.82	(2.38)	#NUM!	(68.92)	(6.28)
D4	(3.27)	16.70	(0.20)	0.84	(36.34)	29.79
D5	(11.91)	8.06	(1.48)	0.14	(27.86)	4.03
Exch	0.01	0.00	2.86	0.00	0.00	0.01
Agriemp	(0.87)	0.21	(4.08)	0.00	(1.29)	(0.45)
Fertilizer	0.69	0.17	3.93	0.00	0.34	1.03
Infl	(0.38)	0.18	(2.08)	0.04	(0.74)	(0.02)
Agriland	1.09	0.45	2.44	0.02	0.20	1.98
FDI	0.55	0.91	0.60	0.55	(1.26)	2.35

The Least Square Dummy Variable approach removes heterogeneity by allowing different intercepts, one for each firm in the pooled data. It does this with the use of dummy variables. Differences in intercepts capture the unique characteristics of the firms. The term ”fixed effect” is due to the intercept β_{0i}. Even if it varies across firms, it is fixed over time. It is time-invariant and, as a result, has no subscript of t. For the intercept to vary among the countries, we run the following model, the differential intercept dummy variable regression model:

$$\text{Coffeep}_{it} = \beta_{0i} + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 D_{3i} + \beta_4 D_{4i} + \beta_5 D_{5i} + \beta_6 \text{Exch}_{it} + \beta_7 \text{Agriem}_{it} + \beta_8 \text{Fertilizerp}_{it} + \beta_9 \text{infl}_{it} + \beta_{10} \text{Agriland}_{it} + \beta_{11} \text{FDI}_{it} + \epsilon_{it} \tag{4.2}$$

Where:

i. $D_1 = 1$ if country 1, 0 otherwise,

ii. $D_2 = 1$ if country 2, 0 otherwise,

iii. $D_3 = 1$ if country 3, 0 otherwise,

iv. $D_4 = 1$ if country 4, 0 otherwise,

v. $D_5 = 1$ if country 5, 0 otherwise.

Country 5 is the reference category, determined if

$$D_1 = D_2 = \ldots = D_4 = 0$$

. We see that the regression is statistically significant with a p-value of zero. The coefficient of determination shows that about 77% increase in coffee price is explained by this regression. We see that the fixed effect is more accurate compared to pooled OLS where we had only 73% as the coefficient of determination. We may say that the fixed effect model gives the best fit. We see that Exchange rate, Fertilizer, Agriculture land, and FDI (Foreign Direct Investment) have a positive impact on coffee price while Agriculture employment and Inflation have a negative impact on coffee production.

4.3 Comparison of Pooled OLS vs Fixed Effect

We can make our decision in one of two ways: Restricted or Partial F test, or Wald test of differential intercepts.

$$F = \frac{(SSE_R - SSE_C)/k_c^*}{SSE_C/(n - k_C)} = \frac{(17617.10 - 14669.16)/5}{17617.10/(130 - 12)} = 4.00$$

Alternatively,

$$F = \frac{(R_C^2 - R_R^2)/k_c^*}{(1 - R_C^2)/(n - k_C)} = \frac{(0.77 - 0.73)/5}{(1 - 0.77)/(130 - 12)} = 4.00$$

Where:

1. SSE_R is SSE for the restricted model (pooled OLS).
2. SSE_C is SSE for the complete model (FE-LSDV).
3. R_C^2 is R^2 from the complete model.
4. R_R^2 is R^2 from the restricted model.
5. n is the sample size.
6. k_c is the number of coefficients in the complete model ($\beta = 12$).
7. k_c^* is the number of additional coefficients in the complete model ($\beta = 5$).

4.4 Discussion

Both F-tests result and F-statistic is 4.00. This value must be compared to a critical value from the F-distribution with appropriate degrees of freedom to determine significance. For this test, the critical value at the 5% significance level (with 5 and 118 degrees of freedom) is approximately 2.29. Since our calculated F-value of 4.00 exceeds this critical value, we reject the null hypothesis that the additional coefficients in the Fixed Effect model do not improve the model fit. Therefore, the Fixed Effect model is preferred over the Pooled OLS model.

4.4.1 Analysis of Findings

The significant F-test indicates that individual-specific effects play a crucial role in our data, which are not captured by the Pooled OLS model. This suggests that ignoring these effects can lead to biased or inconsistent estimates, highlighting the importance of considering individual heterogeneity in panel data analysis.

4.4.2 Comparison with Existing Literature

Existing literature on panel data econometrics frequently supports the preference for Fixed Effect models over Pooled OLS when individual-specific effects are present. Studies such as those by [2] and [16] emphasize that Fixed Effect models account for time-invariant heterogeneity, leading to more reliable and robust parameter estimates. [2] In "Econometric Analysis," Greene argues that Fixed Effect models are particularly useful when dealing with unobserved heterogeneity, as they control for omitted variable bias that may arise in Pooled OLS models.

[16] In "Econometric Analysis of Cross Section and Panel Data," Wooldridge provides a comprehensive discussion on the advantages of Fixed Effect models, especially in the context of policy analysis and behavioral studies where individual-specific characteristics are significant.

Therefore Our findings align with these perspectives, reinforcing the argument that Fixed Effect models provide a better framework for analyzing panel data with significant individual-specific variations.

This means that accounting for heterogeneity is important in determining how Exchange rate, Agriculture land, Agriculture employment, Fertilizer, Inflation, and Foreign Direct Investment (FDI) affect Coffee production.

Chapter 5

CONCLUSION AND RECOMMENDATION

5.1 CONCLUSION

The study analyzes factors affecting coffee price in five East African countries using panel data. It evaluates exchange rates, inflation, agricultural land, agricultural employment, Foreign Direct Investment (FDI), and fertilizers. The analysis finds agricultural employment and inflation to be statistically insignificant. The Fixed Effects model was used for more reliable results after Ordinary Least Squares (OLS) and Random Effects models proved unreliable. The study notes a significant decline in coffee production and a sharp rise in consumption. Coffee prices are highly volatile, indicating a need for greater grower involvement in price setting. Despite difficulties in obtaining comprehensive price data, the study successfully gathered consumption price data, highlighting progress and challenges in understanding the region's coffee market dynamics.

5.2 RECOMMENDATION

Conduct regular market assessments to identify and monitor key determinants such as exchange rates, inflation, and agricultural practices. This can be achieved through the establishment of a centralized database that collects and analyzes data on these factors across East African countries. Additionally, engaging with local farmers and cooperatives can provide qualitative insights that complement quantitative data.

Implement a comprehensive training program for stakeholders, including farmers, government officials, and investors, on the findings of the panel data analysis. This program should focus on understanding the economic implications of coffee pricing and how to leverage this information for better decision-making. Furthermore, policymakers should consider creating supportive policies that stabilize coffee prices, such as price floors or subsidies during periods of market volatility. These recommendations aim to enhance the sustainability, productivity, and economic resilience of the coffee sector in East Africa, addressing the challenges identified in the study.

References

[1] J. Bigirimana, E. Ndayitabi, and L. Mukaruziga. Title of the article. *Journal Name*, Volume(Issue):Page numbers, 2019.

[2] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung. *Time Series Analysis: Forecasting and Control*. Wiley, 5th edition, 2015.

[3] J. Caswell et al. *Coffee, Society, and Power in Latin America*. Publisher, City, 2012.

[4] W. Clarence-Smith and S. Topik. *The Global Coffee Economy in Africa, Asia, and Latin America, 1500-1989*. Publisher, City, 2003.

[5] B. Daviron and S. Ponte. *Title of the Book*. Publisher, City, Year.

[6] M. Davis. *Coffee History*. Publisher, City, 2012.

[7] S. D. Elder, H. Zerriffi, and P. Le Billon. Effects of fair trade certification on social capital: The case of rwandan coffee producers. *World Development*, 40(11):2355–2367, 2012.

[8] N. Karanja and J. Nyoro. *Coffee and Development: A Global History*. Publisher, City, 2002.

[9] C. Moss et al. *The Coffee Paradox: Global Markets, Commodity Trade, and the Elusive Promise of Development.* Publisher, City, 2017.

[10] National Institute of Statistics of Rwanda. Statistical yearbook 2022, 2022. Accessed: 2024-07-09.

[11] C. R. Nelson and C. I. Plosser. Trends and random walks in macroeconomic time series: Some evidence and implications. *Journal of Monetary Economics*, 10(2):139–162, 1982.

[12] I. Nzeyimana, A. E. Hartemink, and C. J. Ritsema. Land use and cover changes in rwanda: The agrarian crisis and the forest lost. 2013.

[13] S. Ponte. Thelatte revolution'? regulation, markets and consumption in the global coffee chain. *World development*, 30(7):1099–1122, 2002.

[14] S. Topik. *The Global Coffee Economy in Africa, Asia, and Latin America, 1500-1989.* Publisher, City, 2015.

[15] M. Tuchscherer. *Coffee and Coffeehouses: The Origins of a Social Beverage in the Medieval Near East.* Publisher, City, 2003.

[16] J. M. Wooldridge. Applications of generalized method of moments estimation. *Journal of Economic Perspectives*, 15(4):87–100, 2001.

APPENDIX : List of data used in the study

Table 5.1: Data for Rwanda (1996-2022)

Year	Consumpr	Ex	Agriemp	Fertilizer	Inflation	Agriland	FDI
1996	7.41	306.82	89.82	0.4	7.41	64.9	0.16
1997	12.02	301.53	89.44	0.5	12.02	66.92	0.14
1998	6.21	312.31	89.38	0.37	6.21	67.73	0.36
1999	-2.41	333.94	89.39	0.35	-2.41	69.6	0.08
2000	3.90	389.70	89.10	0.3	3.90	74.02	0.39
2001	3.34	442.99	88.69	0.27	3.34	76.90	0.94
2002	1.99	475.37	87.71	2.19	1.99	76.53	0.08
2003	7.45	537.66	87.38	2.19	7.45	76.45	0.22
2004	12.25	577.45	86.55	1.82	12.25	76.33	0.32
2005	9.01	557.82	85.48	3.02	9.01	76.25	0.27
2006	8.88	551.71	84.22	3.45	8.88	76.37	0.92
2007	9.08	546.96	83.02	7.81	9.08	76.12	2.02
2008	15.44	546.85	81.54	9.50	15.44	76.21	1.97
2009	12.94	568.28	80.54	1.25	12.94	77.26	2.09
2010	-0.25	583.13	79.30	0.08	-0.25	77.91	3.53
2011	3.08	600.31	78.06	0.09	3.08	78.76	1.63
2012	10.27	614.30	76.61	4.72	10.27	79.41	3.52
2013	5.92	646.64	72.92	10.48	5.92	80.26	2.99
2014	2.35	682.44	68.45	11.27	2.35	81.31	3.81
2015	2.53	719.86	65.61	17.61	2.53	81.35	1.90
2016	7.17	787.25	62.97	8.77	7.17	81.27	3.22
2017	8.28	831.55	60.36	17.39	8.28	81.27	2.96
2018	-0.31	861.09	56.81	14.93	-0.31	79.00	3.80
2019	3.35	899.35	52.68	25.04	3.35	77.18	2.54
2020	9.85	943.28	55.49	29.10	9.85	79.39	1.50
2021	-0.39	988.62	54.67	23.90	-0.39	81.25	1.92
2022	17.69	1030.31	54.89	24.80	17.69	81.25	2.99

Table 5.2: Data for Burundi (1996-2022)

Year	Consumpr	Ex	Agriemp	Fertilizer	Inflation	Agriland	FDI
1996	26.44	302.75	92.52	2.98	26.44	77.61	0
1997	31.11	352.35	92.59	1.05	31.11	76.01	0
1998	12.50	447.77	92.21	3.81	12.50	74.49	0
1999	3.39	563.56	92.07	4.14	3.39	73.48	-0.06
2000	24.43	720.67	91.91	3.65	24.43	72.70	1.34
2001	9.30	830.35	91.65	3.59	9.30	72.08	-0.00
2002	-1.37	930.75	91.29	1.34	-1.37	71.42	0
2003	10.65	1082.62	91.17	0.30	10.65	71.38	0
2004	8.18	1100.9	91.14	1.11	8.18	71.96	0.00
2005	13.25	1081.58	90.76	3.55	13.25	70.83	0.05
2006	2.75	1028.68	90.26	3.20	2.75	70.95	0.00
2007	8.41	1081.87	89.73	1.82	8.41	71.07	0.04
2008	24.41	1185.69	89.32	2.07	24.41	71.18	0.24
2009	10.56	1230.18	88.91	1.86	10.56	71.30	0.02
2010	6.49	1230.75	88.36	3.63	6.49	71.38	0.04
2011	9.59	1261.07	87.79	5.64	9.59	71.69	0.15
2012	18.16	1442.51	87.20	6.21	18.16	72.43	0.03
2013	7.94	1555.09	86.73	10.59	7.94	73.48	4.76
2014	4.41	1546.69	86.19	12.01	4.41	74.53	3.02
2015	5.54	1571.90	86.35	9.03	5.54	75.58	1.60
2016	5.56	1654.63	86.32	15.24	5.56	76.64	0.00
2017	16.05	1729.06	86.31	18.33	16.05	77.69	0.01
2018	-2.81	1782.88	86.18	24.00	-2.81	78.74	0.04
2019	-0.69	1845.62	86.06	19.92	-0.69	79.79	0.04
2020	7.32	1915.05	86.09	19.49	7.32	80.84	0.32
2021	8.40	1975.95	85.85	19.07	8.40	81.89	0.28
2022	18.80	2034.31	85.95	19.01	18.80	81.89	0.27

Table 5.3: Data for Kenya (1996-2022)

Year	Consumpr	Ex	Agriemp	Fertilizer	Inflation	Agriland	FDI
1996	8.86	57.11	46.31	34.43	8.86	46.48	0.90
1997	11.36	58.73	46.16	27.92	11.36	46.73	0.47
1998	6.72	60.37	45.89	27.17	6.72	46.49	0.19
1999	5.74	70.33	45.66	29.06	5.74	47.22	0.40
2000	9.98	76.18	45.57	29.79	9.98	46.86	0.87
2001	5.74	78.56	45.18	29.25	5.74	47.16	0.04
2002	1.96	78.75	44.95	27.31	1.96	47.12	0.21
2003	9.82	75.94	44.59	33.10	9.82	47.22	0.55
2004	11.62	79.17	44.08	27.68	11.62	47.43	0.29
2005	10.31	75.55	43.51	34.33	10.31	47.44	0.11
2006	14.45	72.10	42.74	33.15	14.45	47.53	0.20
2007	9.76	67.32	41.96	36.40	9.76	47.56	2.28
2008	26.24	69.18	41.56	33.29	26.24	47.59	0.27
2009	9.23	77.35	40.97	34.04	9.23	47.94	0.27
2010	3.96	79.23	40.13	32.79	3.96	47.98	0.39
2011	14.02	88.81	39.44	39.59	14.02	48.53	3.09
2012	9.38	84.53	38.70	48.08	9.38	48.55	2.45
2013	5.72	86.12	38.01	56.58	5.72	48.57	1.81
2014	6.88	87.92	37.26	47.73	6.88	48.59	1.20
2015	6.58	98.18	36.59	38.88	6.58	48.61	0.88
2016	6.30	101.50	35.87	48.38	6.30	48.63	0.63
2017	8.01	103.41	35.18	63.08	8.01	48.65	1.64
2018	4.69	101.30	34.30	48.12	4.69	48.67	0.83
2019	5.24	101.99	33.59	57.29	5.24	48.69	0.47
2020	5.40	106.45	33.64	65.82	5.40	48.69	0.42
2021	6.11	109.64	33.03	60.66	6.11	48.69	0.42
2022	7.66	117.87	33.03	60.65	7.66	48.69	0.42

Table 5.4: Data for Tanzania (1996-2022)

Year	Consumpr	Ex	Agriemp	Fertilizer	Inflation	Agriland	FDI
1996	20.98	579.98	84.46	3.47	20.98	37.73	2.31
1997	16.09	612.12	84.29	4.49	16.09	38.05	2.05
1998	12.80	664.67	84.05	3.54	12.80	38.16	1.40
1999	7.89	744.76	83.60	2.41	7.89	38.27	4.06
2000	5.92	800.41	83.08	2.61	5.92	38.38	3.46
2001	5.15	876.41	82.29	0.92	5.15	38.50	4.04
2002	5.32	966.58	80.63	3.70	5.32	38.61	2.80
2003	5.30	1038.42	78.90	4.46	5.30	38.69	2.09
2004	4.74	1089.34	76.88	5.52	4.74	39.24	2.65
2005	5.03	1128.93	74.51	5.78	5.03	39.87	5.08
2006	7.25	1251.90	72.18	5.13	7.25	40.50	2.16
2007	7.03	1245.04	71.38	4.71	7.03	41.11	2.66
2008	10.28	1196.31	71.18	4.68	10.28	41.74	4.95
2009	12.14	1320.31	70.73	7.52	12.14	41.98	3.28
2010	6.20	1395.63	70.25	8.72	6.20	42.22	5.66
2011	12.69	1557.43	69.64	8.92	12.69	42.46	3.55
2012	16.00	1571.70	69.37	8.69	16.00	42.33	4.54
2013	7.87	1597.56	68.70	12.38	7.87	42.94	4.57
2014	6.13	1653.23	68.05	9.68	6.13	43.18	2.83
2015	5.59	1991.39	67.50	9.56	5.59	43.42	3.18
2016	5.17	2177.09	66.88	13.32	5.17	43.66	1.74
2017	5.32	2228.86	66.34	12.63	5.32	43.90	1.76
2018	3.49	2263.78	65.77	16.27	3.49	44.14	1.70
2019	3.46	2288.21	65.08	8.98	3.46	44.38	1.99
2020	3.29	2294.15	64.83	16.44	3.29	44.62	1.04
2021	3.69	2297.76	64.27	9.25	3.69	44.62	1.30
2022	4.35	2298.76	64.33	9.25	4.35	44.62	1.30

Table 5.5: Data for Uganda (1996-2022)

Year	Consumpr	Ex	Agriemp	Fertilizer	Inflation	Agriland	FDI
1996	7.19	1046.08	70.71	0.12	7.19	60.92	2.00
1997	8.17	1083.01	70.48	0.12	8.17	60.92	2.79
1998	0.07	1240.31	70.15	0.70	0.07	61.27	3.19
1999	5.78	1454.83	69.51	0.88	5.78	61.37	2.34
2000	3.39	1644.48	69.50	1.25	3.39	62.62	2.59
2001	1.87	1755.66	69.32	1.07	1.87	63.12	2.59
2002	-0.29	1797.55	68.62	1.33	-0.29	64.12	2.99
2003	8.68	1963.72	68.70	1.60	8.68	65.62	3.06
2004	3.72	1810.31	72.16	1.47	3.72	66.37	3.72
2005	8.45	1780.54	75.30	0.97	8.45	66.37	4.11
2006	7.31	1831.45	74.57	1.25	7.31	67.38	6.46
2007	6.14	1723.49	74.21	1.20	6.14	68.38	6.66
2008	12.05	1720.44	74.00	2.94	12.05	69.64	5.05
2009	13.02	2030.49	73.79	2.08	13.02	70.64	3.35
2010	3.98	2177.56	71.32	1.72	3.98	71.14	2.04
2011	16.56	2522.80	68.80	1.76	16.56	71.64	3.21
2012	12.68	2504.56	66.14	2.13	12.68	71.89	4.41
2013	4.91	2586.89	71.92	2.44	4.91	71.89	3.79
2014	3.08	2599.79	68.47	1.82	3.08	71.89	3.25
2015	5.59	3240.65	67.82	1.59	5.59	71.89	2.28
2016	5.71	3420.10	66.78	1.91	5.71	71.89	2.14
2017	5.21	3611.22	65.10	2.06	5.21	71.89	2.61
2018	2.62	3727.07	64.28	3.30	2.62	71.89	3.21
2019	2.87	3704.05	63.51	3.37	2.87	71.89	3.60
2020	3.31	3718.25	63.48	2.44	3.31	71.89	2.32
2021	2.20	3587.05	62.92	2.44	2.20	71.89	2.72
2022	7.20	3689.82	62.92	2.44	7.20	71.89	2.72