

Modeling and Forecasting Agrofood Supply Chain Emissions Based on VECM-ARIMA Model

Mikidady I. Selungwi¹, Erick P. Massami^{2,*}

¹Department of Logistics and Transport Studies, National Institute of Transport, Dar es Salaam, Tanzania

²Department of Maritime Transport Management, National Institute of Transport, Dar es Salaam, Tanzania

Abstract Global efforts are being undertaken to reduce carbon emissions in agrofood supply chain due to escalating concerns over global warming and climate change. Despite such efforts, carbon emissions continue to rise, particularly in developing countries that prioritize economic growth over environmental sustainability. This study aims to analyse the causal relationship between agrofood processing, agrofood packaging, agrofood transportation, agrofood retailing, and agrofood household consumption as exogenous variables and agrofood supply chain emissions as an endogenous variable. In addition, the study forecasts Tanzania's agrofood supply chain emissions based on the data collected from FAO and IPCC over the period 1990-2020. The study uses the Vector Error Correction (VECM) model to examine the long-term and short-term effects between the variables whereas the Autoregressive Integrated Moving Average (ARIMA) model is used to forecast agrofood supply chain emissions from 2020 to 2030. The Johansen test of cointegration revealed the existence of both long-run and short-run relations between endogenous variable and exogeneous variables. More specifically, the findings reveal that agrofood transportation and agrofood retailing activities are dominant long-term drivers of agrofood supply chain emissions meanwhile agrofood processing activities are the major short-run influencer of agrofood supply chain emissions. In tandem, a 2030 forecast indicate a potential decline in agrofood supply chain emissions. Furthermore, the reduction of carbon emissions in agrofood supply chain can only be achieved provided all stakeholders take corrective actions including formulating and implementing green policies on agrofood supply chain, promoting agricultural resource efficiency, providing awareness training on green consumption, and using more sustainable production methods.

Keywords Agrofood, Supply Chain Emissions, VECM Model, ARIMA Model, Forecasting Model

1. Introduction

Global efforts are being undertaken to reduce carbon emissions in agrofood supply chain due to escalating concerns over global warming and climate change. In 2015, 196 states adopted the Paris Agreement on Climate Change with the aim of limiting global warming to well below 2 °C compared to pre-industrial levels [1]. This is achieved through nationally determined contributions (NDCs), where each country sets its own emission reduction goals and updates them every five years. The effectiveness of the Paris Agreement, in tandem with other environmental pollution policies, is closely linked to how each state party is responsible for reducing carbon emissions in every sector including agrofood supply chain. Globally, carbon emissions are escalating despite international efforts geared for their reduction [2]. This is particularly prevalent in developing countries that prioritize economic growth over environmental sustainability. In tandem with

this view, many governments in the world have been actively implementing various strategies to curb carbon emissions. Nonetheless, the overall level of carbon emissions has not declined [3,4]. As such, reducing global CO₂ emissions has become a focal objective of international policies aimed at mitigating the adverse effects of climate change. Globally, the agrofood supply chain accounts for about one-third of total human-induced emissions, but much of the research focuses on emissions from agricultural production, neglecting the downstream stages of the chain such as post-harvest handling, packaging, and consumer-level emissions.

In Tanzania the agrofood supply chain which encompasses agrofood production, agrofood processing, agrofood packaging, agrofood transportation, agrofood retailing, and agrofood household consumption activities are collectively responsible for a significant share of national GHG emissions [5]. These emissions are largely attributed by inefficient logistics systems, outdated food processing technologies, limited cold chain facilities, and fossil fuel-based energy consumption. It is worth mentioning that the Tanzania's agrofood sector is integral to both food security and economic development,

* Corresponding author:

erick.massami@nit.ac.tz (Erick P. Massami)

Received: Jul. 16, 2025; Accepted: Aug. 9, 2025; Published: Aug. 15, 2025

Published online at <http://journal.sapub.org/ajee>

contributing approximately 29% to the national GDP and employing about 65% of the population [6]. Despite its importance, the sector struggles with rising emissions linked to inefficient supply chain practices and poor energy infrastructure [7]. Regional transport challenges, such as a reliance on outdated road systems, and limited integration of green technology, further exacerbate these emissions [8]. Furthermore, less emphasis of the post-production activities of the agrofood supply chain has led to the knowledge gaps in how emissions from the non-production segments of the agrofood supply chain evolve over time [9,10].

While previous studies have analysed agrofood emissions in industrialized nations, the findings cannot be easily applied to Tanzania due to context-specific issues including policy enforcement, level of investment in sustainable infrastructure and facilities, and the prevalence of informal markets. Moreover, prior studies have rarely analysed emissions in an integrated framework that accounts for interdependencies between agrofood processing, agrofood packaging, agrofood transportation, agrofood retailing, and agrofood household consumption [11]. Consequently, the general objective of this study is to develop an evaluation and forecasting model of agrofood supply chain emissions in Tanzania. The specific objectives include examining the relationships between activities related to agrofood processing, agrofood packaging, agrofood transportation, agrofood retailing, and agrofood household consumption and agrofood supply chain emissions. Additionally, the study aims to forecast agrofood supply chain emissions up to the year 2030 based on ARIMA model. The rest of this paper is organized as follows: Section 2 presents a scrutiny of the relevant studies; Section 3 presents the VECM-ARIMA modeling framework and Section 4 applies VECM-ARIMA model to evaluate and forecast agrofood supply chain emissions in Tanzania. Lastly, Section 5 gives the conclusions.

2. Scrutiny of the Relevant Studies

A number of studies investigate agrofood supply chain emissions. Basri et al. [12] validate the significant role of global food provision while shaping the quality of the environment to achieve carbon peak targets, carbon neutrality, and promote agricultural modernization. Behvandi & Ghorbani [13] use a machine learning approach to predict GHG emissions in agri-food systems and find that on-farm energy use, pesticide manufacturing, and land use factors significantly influence GHG outputs. However, there is a lack of an integrated approach that holistically analyses emissions across multiple stages of the agrofood supply chain. Existing studies tend to isolate food system emissions into separate categories, such as transport emissions [11], pre-and post-production emissions [10], and trade-induced emissions [14], without considering how these factors interconnect within Tanzania's agrofood system. Without a comprehensive understanding of the relationship of each of these factors and agrofood supply chain emissions, it is difficult to develop effective policies for emission reduction in agrofood supply chains. However,

there are limited studies that investigate the relationship between agrofood supply chain emissions and one of these factors: agrofood processing, agrofood packaging, agrofood transportation, agrofood retailing, and agrofood household consumption.

Food processing is one of the most energy-intensive stages in the agrofood supply chain, contributing significantly to GHG emissions. Activities such as pasteurization, drying, refrigeration, and packaging require substantial amounts of energy, particularly in large-scale industrial operations [15]. In Tanzania, where the food processing sector relies primarily on fossil fuel-powered electricity grids, emissions are even higher compared to nations that have adopted renewable energy solutions [11]. Studies suggest that modernizing food processing plants with energy-efficient machinery and solar or hydroelectric power sources can significantly reduce carbon emissions. However, most small and medium-sized enterprises (SMEs) in Tanzania's agrofood industry lack the capital and infrastructure to implement such changes [14]. Despite the clear environmental impact of food processing emissions, there is limited research quantifying its contribution to Tanzania's agrofood supply chain emissions. This study addresses this gap by providing a model that evaluates the contribution of agrofood processing activities on agrofood supply chain emissions.

Food packaging plays a crucial role in protecting food products and extending shelf life, but it is also a major contributor to supply chain emissions. Most food packaging materials, particularly plastic and aluminium-based materials, have high carbon footprints due to energy-intensive production processes and end-of-life disposal challenges [16]. In Tanzania, limited recycling facilities mean that most packaging waste either ends up in landfills or is incinerated, releasing additional CO₂ and other pollutants into the atmosphere [10]. Globally, the push for sustainable packaging alternatives, such as biodegradable plastics, reusable containers, and minimal packaging designs, is gaining momentum. However, these innovations are less accessible in developing countries due to high production costs and limited consumer awareness [14]. The absence of strict regulatory policies on packaging waste further exacerbates environmental concerns in Tanzania's food industry. While packaging emissions have been well-documented in industrialized economies, there is limited empirical research on their impact in Tanzania. This study aims to quantify agrofood packaging-related emissions and assess potential mitigation strategies, such as incentivizing sustainable packaging alternatives and expanding recycling programs.

Food transportation is one of the largest contributors to food supply chain emissions, primarily due to the heavy reliance on fossil fuel-powered vehicles. Studies indicate that long-haul transportation significantly increases CO₂ emissions, especially in regions with poor infrastructure, where inefficient routes and extended travel times further raise fuel consumption [11]. Keshavarz-Ghorbani and Pasandideh [17] propose an optimization algorithm which integrates purchasing, transporting and storing decisions with the view

to minimize the total costs and CO₂ emissions in agro-supply chain. Accorsi et al. [18] present an optimization model that supports strategic decision-making on agriculture and food transportation issues while addressing climate stability. The findings reveal the interdependency between production, infrastructure, transportation, and environmental resources. Dwivedi et al. [19] propose an optimization model that minimizes the total transportation cost and carbon emission tax in the Indian agrofood grain supply chain. In Tanzania, the dominance of diesel-powered trucks and outdated logistics systems worsens the problem, making food transport one of the most carbon-intensive segments of the supply chain [10]. Efforts to mitigate transportation-related emissions in developed nations focus on electric freight vehicles, improved fuel efficiency, and optimized agrofood supply chains. However, electric freight vehicles remain largely inaccessible in Tanzania due to high costs, lack of charging infrastructure, and minimal policy support for green transport alternatives [15]. Additionally, the country's heavy dependence on road transport rather than rail or waterways contributes to higher emissions per ton-km. While existing studies explore global transportation emissions, few provide insights into Tanzania's transport logistics and inefficiencies. This study aims to fill that gap by evaluating the relationship between transportation activities and agrofood supply chain emissions, offering practical recommendations for decarbonizing food transportation in the country.

The retail sector plays a critical role in food supply chain emissions, with significant contributions from refrigeration, storage, packaging, and waste generation. Large supermarkets, convenience stores, and informal market vendors all require energy-intensive cooling systems to preserve perishable foods, leading to substantial indirect CO₂ emissions [10]. In many developing countries, including Tanzania, informal food markets lack adequate refrigeration, resulting in high food spoilage rates, which in turn lead to increased methane emissions from organic waste decomposition. In addition to refrigeration, food packaging in retail contributes significantly to the supply chain's carbon footprint. Many retailers rely on single-use plastics and non-biodegradable materials, leading to environmental pollution and increased emissions from incineration or landfill accumulation [14]. While developed countries are increasingly transitioning toward biodegradable packaging and sustainable retailing models, these solutions have yet to be fully implemented in Tanzania. Limited awareness, cost constraints, and weak regulatory enforcement continue to hinder progress in sustainable retail operations [11]. Despite its significant impact, research on Tanzania's retail emissions is scarce. Current global studies mainly focus on formal supermarket chains, which do not represent the predominantly informal retail structures in Tanzania. This study aims to bridge that gap by evaluating the relationship between food retailing activities and agrofood supply chain emissions.

Household food consumption is a significant contributor to greenhouse gas (GHG) emissions within the agrofood supply chain, particularly through energy intensive food

preparation, dietary choices, and food waste generation. Studies have consistently shown that diets high in animal-based products, such as beef and dairy, have a far greater carbon footprint than plant-based diets, mainly due to methane emissions from livestock [15]. In Tanzania, the situation is further worsened by the widespread use of firewood and charcoal for cooking, which not only increases CO₂ emissions but also leads to deforestation and loss of carbon sinks [10]. Beyond energy use, food waste at the household level is another major concern. When food is discarded, it leads to unnecessary production-related emissions, while its decomposition in landfills releases methane (CH₄), a greenhouse gas 25 times more potent than CO₂. Studies indicate that developed nations have implemented food waste reduction strategies, such as portion control and redistribution programs, yet these interventions remain largely absent in Tanzania's urban and rural settings [14]. Additionally, the lack of structured waste management systems means that much of this waste is either openly burned or left to decompose unmanaged, further contributing to emissions. Given the limited research on household emissions within Tanzania's food supply chain, this study addresses this gap by presenting a model that evaluates the contribution of household consumption activities on agrofood supply chain emissions. The findings will help inform policies aimed at reducing household-level emissions through sustainable food consumption and waste management strategies.

Despite increasing global research on greenhouse gas (GHG) emissions in the agro-food supply chain, studies focusing on Tanzania's specific food system emissions remain scarce. While existing literature by Cui et al. [20], Crippa et al. [10], Piñero et al. [14], Jehn et al. [11], and Stöckli et al. [16] provides valuable insights into emissions across various stages of the food supply chain, these studies primarily focus on global or industrialized economy contexts, leaving significant knowledge gaps for developing nations like Tanzania. In addition, most of prior studies are based on optimization models which focus on finding the best solution to a given agrofood supply chain problem not on forecasting time series data. More specifically, optimization models have difficulty in handling real-world complexities like uncertainty and dynamic environments. This study aims to bridge these gaps by conducting a comprehensive modeling and forecasting analysis of agrofood supply chain emissions from agrofood processing, agrofood packaging, agrofood transportation, agrofood retailing operations, and agrofood household consumption in Tanzania based on VECM-ARIMA model. The VECM-ARIMA model is a tool for time series analysis with several strengths including analyzing the short-run and long-run relationships of variables [12,21,22,23,24] and forecasting time series data. By integrating ARIMA and VECM models, this research will provide data-driven insights into key emission drivers, long-term trends, and potential policy interventions. The findings will support the development of targeted strategies for reducing emissions while enhancing supply chain efficiency in Tanzania's agrofood sector.

3. VECM-ARIMA Modeling Framework

A step-by-step framework for developing a Vector Error Correction (VECM) model with Autoregressive Integrated Moving Average (ARIMA) components involves identifying cointegration relationships between time series, developing ARIMA model, and integrating ARIMA model into a VECM framework.

3.1. Basic Concepts of Time Series

A time series is a set of observations $\{x_t\}_{t \in T}$ where each observation x_t is recorded over a given time period T . There are two types of time series: discrete and continuous. Discrete time series result when the observations are discretely recorded (i.e. take whole number values). Continuous time series result when observations are continuously recorded over a time interval [25]. Time series are often used for forecasting future outcomes. Prior recorded observations for a time series are normally presented at daily, weekly, monthly, quarterly or annually time intervals. The time series model uses statistical tools, where historical observations gives a foundation for estimating the unknown parameters of interest [26].

3.1.1. Autocorrelation

Autocorrelation (serial correlation) is the correlation of a time series and its lagged version over time. The autocorrelation coefficient ($\rho(k)$) measures the correlation between a time series and a lagged version of itself. It indicates how strongly related values are at different points in time within the same series, where one basic example is that future agrofood supply chain emissions are predicted by current agrofood supply chain emissions. The autocorrelation coefficient at lag k of time series $\{x_t\}$ is given by equation (1).

$$\rho(k) = \frac{\text{Cov}(x_t, x_{t-k})}{\sigma(x_t)\sigma(x_{t-k})} \quad (1)$$

Where $\text{Cov}(x_t, x_{t-k})$ represents the variance between the time series values at time t and time $t-k$ and measures how much these values tend to vary together; $\sigma(x_t)$ is the standard deviation of the time series values at time t ; and $\sigma(x_{t-k})$ is the standard deviation of the time series values at time $t-k$. Similarly, the autocorrelation coefficient ($\rho(k)$) at lag k for a time series $\{x_t\}$ is given by equation (2).

$$\rho(k) = \frac{\sum_{t=1}^{T-k} (x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum_{t=1}^T (x_t - \bar{x})^2} \quad (2)$$

Where \bar{x} is the mean of the time series; $\sum_{t=1}^{T-k} (x_t - \bar{x})(x_{t-k} - \bar{x})$ calculates the covariance between the time series values at time t and time $t-k$; and $\sum_{t=1}^T (x_t - \bar{x})^2$ calculates the variance of the time series which measures the overall variability of the data. $\rho(k) \in [-1, 1]$, represents the strength and direction of the linear relationship. If the autocorrelation is high (close to +1 or -1), it suggests a strong correlation between past and present values, indicating patterns or dependencies in the data. One of the fundamental features of a time series that can be identified through autocorrelation is stationarity.

3.1.2. Stationarity

A stationary time series has statistical properties that are constant over time. This means that statistics such as the mean, variance and autocorrelation, don't vary with time. Stationarity means that the time series does not have a trend, has a constant variance, a constant autocorrelation pattern, and no seasonal pattern. The autocorrelation function (ACF) declines to near zero rapidly for a stationary time series whereas the ACF drops slowly for a non-stationary time series. There are two fundamental types of stationary time series. Strong stationary (i.e. strict stationary) and weak stationary. A time series $\{x_t\}_{t \in T}$ is said to be strictly stationary if the joint distribution of $(x_{t_r}), r \in \{1, 2, \dots, n\}$ is identical to that of $(x_{t_r - k}), \forall r \in \mathcal{N}, \forall k \geq 1$. As a result, changing the origin time by lag k has no effect on the joint distribution. On the other hand, a time series is said to be weakly stationary provided: $E(x_t) = \mu, \forall t; E(x_t^2) < \infty, \forall t;$
 $\text{Cov}(x_t, x_{t-k}) = \gamma(k), \forall t, \forall k.$

Generally, data from real-life scenarios are mostly non-stationary. However, many time series models work under the assumption that the data is stationary. Consequently, the analysis of such model becomes more complicated when working with non-stationary data. One method to transform non-stationary data to stationary is by differencing. This method eliminates trends in the data. By using a difference operator, a non-stationary time series is transformed to a stationary time series. The difference of a time series $\{x_t\}_{t \in T}$ is defined by equation (3).

$$\Delta x_t = x_t - x_{t-1} \quad (3)$$

Another way of expressing the difference operator is by using a backshift operator (B) as defined in equation (4).

$$Bx_t = x_{t-1}, \forall t > 1 \quad (4)$$

Thus, the differencing operator can also be expressed in another form as given by equation (5).

$$\Delta x_t = x_t - Bx_t = (1 - B)x_t \quad (5)$$

If longer lags k are considered, the relation involving a backshift operator is given by equation (6).

$$Bx_t = x_{t-k}, \forall k \geq 2 \quad (6)$$

Since differencing can eliminate the effect of seasonality, the difference of a time series $\{x_t\}$ in terms of a seasonal difference operator is given by equation (7).

$$\Delta x_t = x_t - x_{t-k} = (1 - B^k)x_t \quad (7)$$

Where k denotes the number of seasons.

3.2. Statistical Tests

The following are some commonly used statistical tests for time series.

3.2.1. Augmented Dickey-Fuller Test

The Augmented Dickey-Fuller (ADF) is one of the methods to test whether the data of a time series is stationary or not. The method is used to detect the unit roots in the time series data. A unit root refers to a stochastic trend in a time

series, which shows an unpredictable systematic pattern. It is crucial to ensure that the variables are stationary since statistical inference assumes of constant means, variances and covariances. An ADF test has the following null hypothesis (Ho) and alternative hypothesis (Ha):

Ho: The time series is non-stationary,

Ha: The time series is stationary.

If the resulting p-value of the test is less than a specified α level, say $\alpha = 0.05$, then the null hypothesis is rejected at 5% significance level. This means that the alternative hypothesis is accepted and the time series is stationary. On the other hand, if the p-value is higher than a specified α level, the null hypothesis cannot be rejected and the time series is considered non-stationary [27,28].

3.2.2. The Kwiatkowski-Phillips-Schmidt-Shin Test

The Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) test is another method used to assess the stationarity of time series data. It has the following null hypothesis (Ho) and alternative hypothesis (Ha):

Ho: The time series is stationary,

Ha: The time series is non-stationary.

If the p-value of the test is less than a specified α level, then the null hypothesis is rejected at $\alpha\%$ significance level. Thus, the time series is non-stationary which suggests that differencing is required [29].

3.2.3. Shapiro-Wilk Test for Normality

The normality of a data set (i.e. residuals of VECM equation) is generally assessed by the use of a Q-Q plot, which gives a graphical visualisation of normality. If the data set scatter around the line of best fit inclined 45° with the horizontal-axis, the normality holds. However, a quantitative test is required to complement the visual inspection. The Shapiro-Wilk test is a hypothesis test that evaluates whether a data set is normally distributed. The Shapiro-Wilk test statistic (W) is computed by equation (8).

$$W = \frac{(\sum_{i=1}^n a_i x_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (8)$$

Where n represents the number of observations; x_i represents values of the ordered sample; and a_i represents tabulated coefficients. The Shapiro-Wilk test has the following null hypothesis (Ho) and alternative hypothesis (Ha):

Ho: The residuals are normally distributed,

Ha: The residuals are not normally distributed.

For a 95% confidence interval, the Shapiro-Wilk test resulting in a p-value < 0.05, indicates normality does not exist among residuals. On the other hand, a p-value > 0.05, indicates that the residuals are normally distributed [29].

3.3. Models for Stationary Time Series

3.3.1. White Noise Model

White noise model is the simplest time series model. Suppose $\{x_t\}_{t \in T}$ is a white noise process, which is defined as a model whose random variable is independently and

identically distributed. It is worth noting that $E(\epsilon_t) = 0$, $Var(\epsilon_t) = \sigma^2$, $Cov(\epsilon_t, \epsilon_{t-k}) = \sigma^2$ for $k = 0$, $Cov(\epsilon_t, \epsilon_{t-k}) = 0$ for $k \geq 1$, and $Corr(\epsilon_t, \epsilon_{t-k}) = 1$ for $k = 0$, $Corr(\epsilon_t, \epsilon_{t-k}) = 0$ for $k \geq 1$.

3.3.2. General Linear Model

For the general linear model each value of x_t is a linear combination of both the current and preceding white noise processes and it is expressed by equation (9).

$$x_t = \sum_{i=-\infty}^{\infty} \phi_i \epsilon_{t-i} \quad (9)$$

Where ϵ_t represents the white noise process and $\{\phi_i\}$ denotes a series of constants such that $\sum_{i=-\infty}^{\infty} |\phi_i| < \infty$.

3.3.3. Autoregressive Model

The autoregression processes are regressions on themselves. In an autoregressive model (AR(p) of order p, the model uses the p most recent values in the sequence to predict the next value. For example, an AR(1) model uses the two most recent values. An autoregressive model of order p at time t is defined by an expression given in equation (10).

$$x_t = \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \epsilon_t \quad (10)$$

Where ϵ_t is a zero-mean white noise process with a standard deviation of σ .

3.3.4. Moving Average Model

A moving average model of order q (referred to as MA(q) uses past forecast errors in a regression-like model, rather than using past values of the forecast variable like an autoregressive model does. In a moving average model, each value of x_t can be seen as a weighted moving average of the past few forecast errors. The model is defined by an expression given in equation (11).

$$x_t = \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t \quad (11)$$

Where ϵ_t is a zero-mean white noise process with a standard deviation of σ , θ_i 's are the moving average parameters, which results in different time series patterns when changed, and q is the order of the moving average [30].

3.3.5. Autoregressive Moving Average Model

Suppose the time series $\{x_t\}_{t \in T}$ is an autoregressive moving average model of order p and q, denoted by ARMA (p, q) and defined by an expression in equation (12).

$$x_t - \phi_1 x_{t-1} - \dots - \phi_p x_{t-p} = \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t \quad (12)$$

Where ϵ_t is a white noise process with a mean of zero and a standard deviation of σ . Notably, the ARMA process incorporates both AR and MA processes. It is interesting to note that if $\theta_1 = \theta_2 \dots = \theta_q = 0$, then equation (12) becomes an AR(p) process; however, if $\phi_1 = \phi_2 \dots = \phi_p = 0$, then equation (12) becomes an MA(q) model. In order to maintain a minimum order for p and q, AR(p) and MA(q) polynomials must have no common roots, where the AR(p) and MA(q) polynomials are defined by $\Phi(w) = 1 - \phi_1 w - \dots - \phi_p w^p$

and $\theta(w) = 1 + \theta_1 w + \dots + \theta_q w^q$, respectively.

3.3.6. Vector Autoregression Model

Vector autoregression (VAR) is used for multivariate time series (multiple related time series). Each variable in the system is modelled as a function of its own past values and the past values of all other variables in the system. The VAR model can be used to analyse relationships and predict future values across multiple related time series. A VAR model for a system of k variables and p lags, is represented by an expression given in equation (13).

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \epsilon_t \quad (13)$$

Where x_t : A $k \times 1$ vector of endogenous variables at time t .

ϕ_i : A $k \times k$ matrix of coefficients for each lag i .

x_{t-i} : A $k \times 1$ vector of endogenous variables at lag $t-i$.

ϵ_t : A $k \times 1$ vector of error terms (white noise).

3.4. VECM-ARIMA Model

A step-by-step framework for developing a Vector Error Correction (VECM) model with Autoregressive Integrated Moving Average (ARIMA) components involves identifying cointegration relationships between time series, developing ARIMA model, and integrating ARIMA model into a VECM framework.

3.4.1. VECM Model

A VECM model falls under the category of Vector Autoregressive (VAR) models. It is a specialized form of VAR model used when dealing with time series data that exhibit cointegration, meaning they have a long-run equilibrium relationship. A VECM models how variables deviate from their long-run equilibrium and how they adjust back to that equilibrium in the short-run. The general form of the VECM model is given by equation (14).

$$\Delta x_t = \pi x_{t-1} + \sum_{\forall i} \Gamma \Delta x_{t-i} + Cdt + \epsilon_t, i \in \{1, \dots, p\} \quad (14)$$

Where Δx_t : Represents the first difference of the vector of variables ($x_t - x_{t-i}$) at time t . This captures the short-run dynamics and could be the first difference of the dependent variable at time t .

πx_{t-1} : The error correction term which represents the long-run equilibrium relationship between the variables. π is a matrix that can be decomposed into $\alpha\beta$, where α is the adjustment speed matrix and β is the cointegrating vector, capturing the long-run relationships.

$\sum_{\forall i} \Gamma \Delta x_{t-i}$: Represents the summation of lagged differences of the variables, capturing short-run dynamics and feedback effects.

Cdt : Represents deterministic terms like constant or trend.

ϵ_t : Represents the error term, assumed to be a vector of white noise with zero mean and constant covariance matrix.

p : The lag order of the model. This needs to be determined through model selection criteria like Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) etc. The sections 3.4.1.1, 3.4.1.2, and 3.4.1.3 present essential steps in VECM modeling.

3.4.1.1. Data Pre-Processing and Stationarity

Data collection and visualization. Gather time series data and plot them to inspect for trends, seasonality, and outliers.

Stationarity testing. Apply unit root tests (e.g. ADF test) to each time series to determine if they are stationary. Non-stationary series need to be differenced to achieve stationarity.

3.4.1.2. Cointegration Analysis

VAR model estimation. Fit a Vector Autoregression (VAR) model to the time series. Determine the optimal lag order for the VAR model using information criteria like AIC, BIC etc.

Johansen test for cointegration. Perform the Johansen cointegration test to determine if the variables have a long-run equilibrium relationship. This test will identify the number of cointegrating equations (relationships).

Identify cointegrating vectors. If cointegration is found, identify the cointegrating vectors, which represent the long-run relationships.

3.4.1.3. VECM Model Building

VECM specification. Based on the cointegration results, specify the VECM model. This involves determining the number of lagged differences of the variables and the number of cointegrating equations.

Error correction term. Include the error correction term(s) derived from the cointegration analysis in the VECM. This term represents the adjustment of variables towards their long-run equilibrium.

3.4.2. ARIMA Model

The Autoregressive Integrated Moving Average (ARIMA) model integrates an AR and an MA model and thus combines differencing autoregression and moving average. It aims to describe the autocorrelations in the data. This model is denoted by ARIMA (p,d,q) where p is the order of the autoregressive (AR) part, indicating how many past values of x are used as predictors; d is the order of differencing, indicating how many times the time series needs to be differenced to be stationary; and q is the order of the moving average (MA) part, indicating how many past forecast errors are used as predictors. The general form of an ARIMA (p,d,q) model is given by equation (15).

$$x'_t = \phi_1 x'_{t-1} + \dots + \phi_p x'_{t-p} + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t \quad (15)$$

Where x'_t : The differenced time series at time t ; $\theta_1 \dots \theta_q$: The MA coefficients; $\phi_1 \dots \phi_p$: The AR coefficients; and ϵ_t : The white noise error term. Sections 3.4.2.1, 3.4.2.2, 3.4.2.3, 3.4.2.4, and 3.4.2.5 describe essential steps involved when ARIMA modeling is carried out.

3.4.2.1. Data Collection and Preparation

Gather historical data on agrofood supply chain emissions from various sources (e.g. agrofood transportation); ensure data quality and address missing values or outliers; and

transform the data to make it stationary (constant mean and variance over time), often by differencing.

3.4.2.2. Model Identification

Examine autocorrelation and partial autocorrelation plots (ACF and PACF) to identify the order of the ARMA model (p, d, q).

3.4.2.3. Model Estimation and Selection

Estimate the parameters of the chosen ARIMA model using the prepared data, and evaluate model fit using criteria like AIC or BIC to select the best model.

3.4.2.4. Forecasting

Use the fitted ARIMA model to generate forecasts for future emission levels, the model can be used for both short-term and long-term forecasting.

3.4.2.5. Validation and Refinement

Compare predicted emissions with actual emissions data as it becomes available. Refine the model (adjust parameters or consider adding other variables) if the forecast accuracy is not satisfactory.

3.5. Model Evaluation Measurements

There are four different model evaluation measurements which are commonly used for evaluating and comparing different forecasting models.

3.5.1. Akaike Information Criterion

Akaike Information Criterion (AIC) is based on a numeric value and likelihood, and aims to minimize the expression given by equation (16).

$$AIC = -2\log L(\theta) + 2k = -2l(\theta) + 2k \quad (16)$$

Where $L(\theta)$ is the likelihood function of the model, $l(\theta)$ is the log-likelihood function of the model and k is the number of parameters i.e. $k = p + q$ for ARMA (p, q) model. The log-likelihood function is the logarithm of the likelihood function. Since the expression is to be minimized, a numerical penalization occurs for complex models with poor likelihood. A lower AIC values indicate a better-fit model [31].

Generally, the AIC works better for a large sample size as such it can potentially lead to a high degree of negative bias and overfitting when working with a small sample. To rectify this bias, a Corrected Akaike Information Criterion (AICc) is used. The AIC is a first-order estimate meanwhile AICc is the second-order AIC. The AICc is defined by an expression given by equation (17).

$$AICc = AIC + \frac{2k^2 + 2k}{n - k - 1} \quad (17)$$

Where n represents the sample size and k denotes the number of parameters.

3.5.2. Mean Absolute Error

The Mean Absolute Error (MAE) refers to a measure of

variability in the absolute values of distances between observations and their corresponding predicted values i.e. the average magnitude of the errors. The MAE is commonly used as a measure of prediction of errors and is therefore useful when evaluating forecasting models. The MAE is defined by an expression given by equation (18).

$$MAE = \frac{1}{n} \sum_{t=1}^n |x_t - \tilde{x}_t| \quad (18)$$

Where n is the number of observations, x_t is the actual observation and \tilde{x}_t is the predicted value at time period t [26].

3.5.3. Mean Absolute Percentage Error (MAPE)

The Mean Absolute Percentage Error (MAPE) refers to a measure of variability in the absolute values of the ratio between the predicting errors and the observations. The MAPE is commonly used as a measure of prediction accuracy and is therefore useful when evaluating forecasting models. The MAPE is defined by an expression in equation (19).

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{x_t - \tilde{x}_t}{x_t} \right| \times 100 \quad (19)$$

Where n represents the number of observations, x_t and \tilde{x}_t represent the actual and the forecasted value at time period t respectively.

4. Application of VECM-ARIMA Model for Evaluation and Forecasting Agrofood Supply Chain Emissions in Tanzania

A VECM-ARIMA model combines the features of a VECM, which handles cointegrated time series, with the capabilities of ARIMA models for capturing autoregressive and moving average patterns within each variable. Specifically, the VECM model was applied in the analysis of the Tanzania's agrofood supply chain emissions to understand the short-run and long-run relationships between various variables influencing carbon emissions measured in kg tonnes within a supply chain. The variables that were analysed include agrofood processing activities, agrofood packaging activities, agrofood transportation, agrofood retailing activities, agrofood household consumption, and agrofood supply chain emissions. On the other hand, ARIMA model was applied to forecast the Tanzania's agrofood supply chain emissions up to 2030 (i.e. In line with the United Nations SDGs). This allowed for identification of patterns and predicting future emission levels based on past trends. This study used secondary data retrieved from publicly accessible databases of Food and Agriculture Organization (FAO) and the Intergovernmental Panel on Climate Change (IPCC). The two sources offered historical datasets from 1990 to 2020.

4.1. Trends of Emissions and Statistical Tests of Variables in the Tanzania Agrofood Supply Chain

4.1.1. The Trend of Agrofood Supply Chain Emissions from 1990 to 2020

Trend in emissions from agrofood supply chain from 1990 to 2020

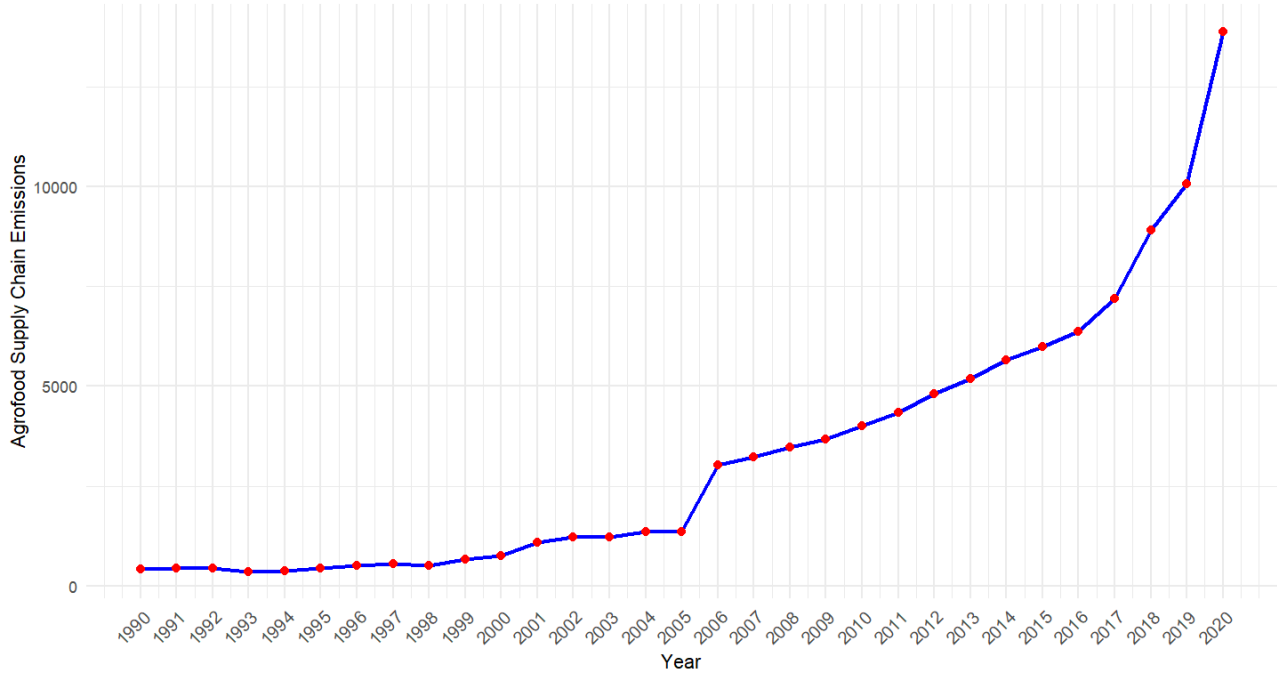


Figure 1. The trend of Tanzania’s agrofood supply chain emissions from 1990 to 2020. (Source: Authors’ compilation from RStudio)

Figure 1 depicts the trend of Tanzania’s agrofood supply chain emissions (kg tonnes of CO₂) from 1990 to 2020.

Figure 1 reveals an upward trend of agrofood supply chain emissions (kg tonnes of CO₂) from 1990 to 2020. The escalation of emissions signifies a growing environmental degradation in agrofood supply chain in Tanzania, most likely driven by the country’s population growth (i.e. from 26,206,012 in year 1990 till 61,704,518 in year 2020), expansion of agricultural production, and limited intervention on green logistics and public awareness towards environmental protection. Unless corrective actions are taken by all parties and other stakeholders in the agrofood supply chain, the agrofood sector in Tanzania will continue to contribute to global warming and climate change. Thus, reforms and effective measures are essential to facilitate the green economy transition, promoting strategies for agricultural resource efficiency and using more sustainable production methods, ultimately leading to the mitigation of climate change. Moreover, understanding trends, and the environmental impacts of agrofood systems is crucial for informing countries about the challenges and opportunities to mitigate GHG emissions in the food sector.

4.1.2. Stationarity Test

The ADF test was applied to test whether the data of a time series is stationary or not. All six (6) variables were tested: log(supply chain), log(food processing), log(food packaging), log(food transportation), log(food retailing), and log(household consumption). The results of the test before differencing (I(0)) and after differencing (I(1)) are shown in Table 1.

Table 1. ADF test at level I(0) and level I(1)

Variable	P - Value	Comments
Log(supply chain)	0.5183	Not Stationary
Log(food transport)	0.4985	Not Stationary
Log(household consumption)	0.5826	Not Stationary
Log(food retailing)	0.7288	Not Stationary
Log(food packaging)	0.3557	Not Stationary
Log(food processing)	0.8749	Not Stationary
At level I(1)		
D1_Log(supply chain)	0.01	Stationary
D1_Log(food transport)	0.01	Stationary
D1_Log(household consumption)	0.01	Stationary
D1_Log(food retailing)	0.01	Stationary
D1_Log(food packaging)	0.01357	Stationary
D1_Log(food processing)	0.01	Stationary

Source: Authors’ compilation from RStudio

In table 1, at level I(0) the $p - value > \alpha = 0.05$ for all variables, which implies that the time series is non-stationary since the null hypothesis is not rejected. However, at level I(1) the $p - value < \alpha = 0.05$ for all variables, which implies that the time series is stationary since the null hypothesis is rejected in favour of the alternative hypothesis.

4.1.3. Graphical Test of the Time Series Data

If the variable in a time series data has either an upward or downward trend, it is not stationary. However, by differencing the time series may turn into stationary. Stationary data has a constant mean and variance which are independent of time. Figures 2 and 3 depict the trend of all variables at level I(0)

(i.e. before differencing the time series) and level I(1) (i.e. after differencing the time series) respectively.

This means the mean and variance of each variable is changing over time hence the time series of each variable is not stationary at level I(0).

From Figure 2 all of the variables show an upward trend.

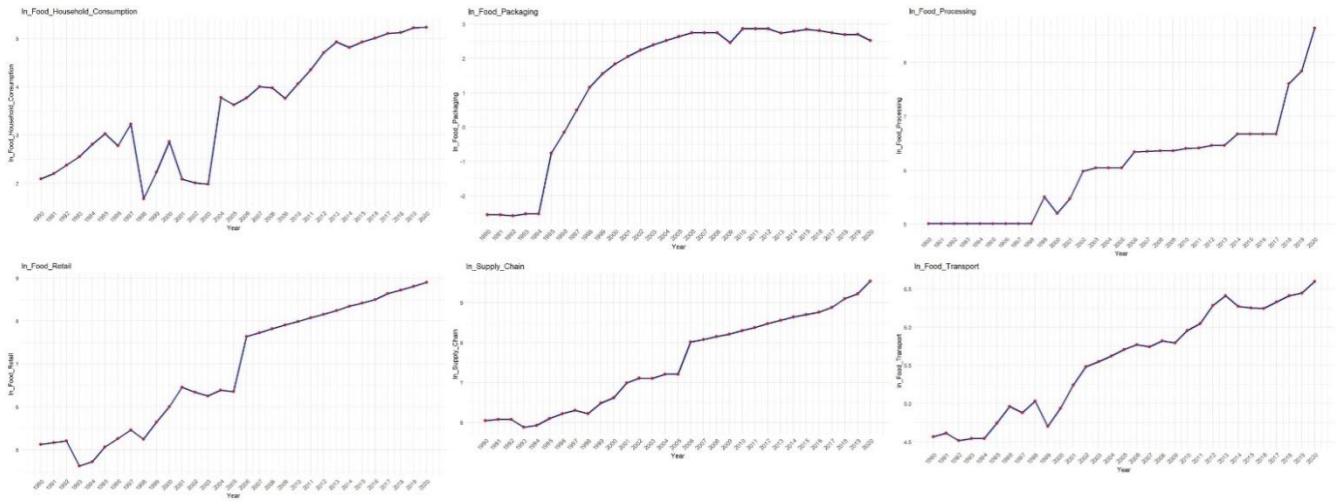


Figure 2. Trends of variables at level I(0) (Source: Authors' compilation from Rstudio)

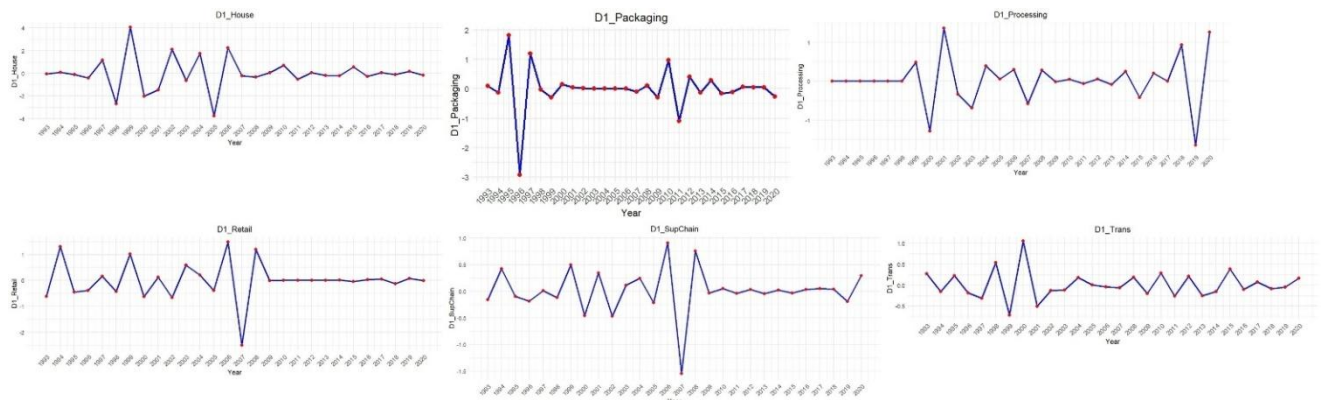


Figure 3. Trends of variables at level I(1) (Source: Authors' compilation from RStudio)

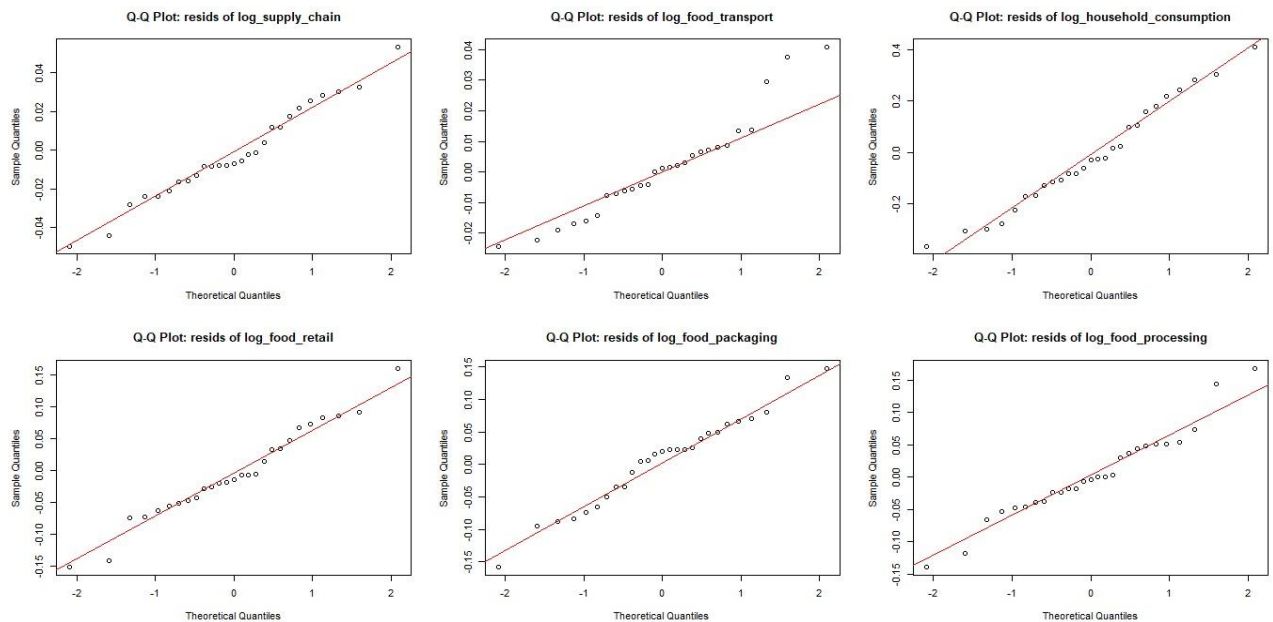


Figure 4. Q-Q plots for normality test of variables (Source: Authors' compilation from RStudio)

Figure 3 reveal that all of the variables have constant mean and variance over time, hence stationary at level I(1) i.e. after differencing the time series.

4.1.4. Normality Test

The Shapiro-Wilk test was applied to test whether the residuals are normally distributed. All residuals for the six (6) variables were tested. These variables include log (supply chain), log(food processing), log(food packaging), log(food transportation), log(food retailing), and log(household consumption). The W-statistic values and p-values are indicated in Table 2.

Table 2. Results of Shapiro-Wilk test

Residual	W-statistic	P-Value
Log(supply chain)	0.978	0.822
Log(food transport)	0.937	0.104
Log(household consumption)	0.975	0.749
Log(food retailing)	0.976	0.773
Log(food packaging)	0.978	0.805
Log(food processing)	0.961	0.386

Source: Authors' compilation from RStudio

Since in Table 2 the $p - value > \alpha = 0.05$ for all variables, the null hypothesis (H_0) is not rejected. This indicates that the residuals are normally distributed. The normality of residuals is further confirmed by the residual plots.

4.1.5. Residual Plots

Figure 4 depicts the residual plots to test the visualisation of normality for all variables.

The residual plots in Figure 4, are aligned at 45° with the horizontal-axis, indicating that the normality holds reasonably well for all variables. While there are slight deviations observed in the upper tails for log(food transport) and log(household consumption), they are not severe enough to invalidate the normality assumption.

4.1.6. Lag Order Selection

For the determination of the maximum lag to be used in the model, the Akaike Information Criteria (AIC) was used considering its strengths over other criteria. The chosen lag should be the one with the lowest AIC and it was automatically suggested by the RStudio program. The lag selection output and optimal value are indicated in Table 3 and Table 4 respectively. From Tables 3 and 4 it is revealed that the optimal lag to be used by the model is three.

Table 3. Lag selection output

```
> lag_select <- VARselect(vecm_data, lag.max = 9, type = "const")
>
> print(lag_select)
$selection
AIC(n)  HQ(n)  SC(n)  FPE(n)
      3      3      3      4

$criteria
      1      2      3      4      5      6      7      8      9
AIC(n) -1.687170e+01 -2.330188e+01 -Inf -Inf -Inf -Inf -Inf -Inf -Inf
HQ(n)   -1.651838e+01 -2.264571e+01 -Inf -Inf -Inf -Inf -Inf -Inf -Inf
SC(n)   -1.478399e+01 -1.942471e+01 -Inf -Inf -Inf -Inf -Inf -Inf -Inf
FPE(n)  5.853656e-08  4.746009e-10  NaN  0    0    0    0    0    0
```

Source: Authors' compilation from RStudio

Table 4. Lag selection

Lag	1	2	3	4	5	6	7	8	9
AIC	-16.8717	-23.3019	-∞	-∞	-∞	-∞	-∞	-∞	-∞
HQ	-16.5184	-22.6457	-∞	-∞	-∞	-∞	-∞	-∞	-∞
SC	-14.7839	-19.4247	-∞	-∞	-∞	-∞	-∞	-∞	-∞
FPE	-5.8537×10^{-8}	-4.746×10^{-10}	NaN	0	0	0	0	0	0

Source: Authors' compilation from RStudio

Table 5. Johansen test output

Null Hypothesis (H_0)	Trace Statistic	10% Critical Value	5% Critical Value	1% Critical Value	Decision (at 5%)
$r \leq 5$	11.73	7.52	9.24	12.97	Reject H_0
$r \leq 4$	32.20	17.85	19.96	24.60	Reject H_0
$r \leq 3$	59.28	32.00	34.91	41.07	Reject H_0
$r \leq 2$	106.93	49.65	53.12	60.16	Reject H_0
$r \leq 1$	157.80	71.86	76.07	84.45	Reject H_0
$r = 0$	226.76	97.18	102.14	111.01	Reject H_0

Source: Authors' compilation from RStudio

4.1.7. Cointegration Test

The Johansen test was used to check if there exists a long-run and short-run relationship between the endogenous variable (agrofood supply chain emissions) and exogeneous variables (agrofood processing, agrofood packaging, agrofood transportation, agrofood retailing operations, and agrofood household consumption). The tested hypotheses were: Ho: There is no cointegration between variables, Ha: There is cointegration between variables. The tests results are presented in Table 5.

From Table 5, the trace statistic is greater than 5% critical value for all scenarios and thus the null hypothesis is rejected for each case. The last rejection occurs at $r \leq 5$ which implies that there are 6 cointegrating relationships in the system. Consequently, there is long-run and short-run relationships between agrofood supply chain emissions and agrofood processing, agrofood packaging, agrofood transportation, agrofood retailing, and agrofood household consumption.

4.2. Analysis of Agrofood Supply Chain Emissions Based on VECM Model

Table 5 shows that all the variables: agrofood supply chain emissions, agrofood processing, agrofood packaging, agrofood transportation, agrofood retailing operations, and agrofood household consumption are co-integrated. Thus, VECM model suits for this study.

4.2.1. Speed of Adjustment

The speed of adjustment is obtained from the Error Correction Term (ECT). The Error Correction Term (ECT) measures the deviation from long run equilibrium corrected gradually through a series of partial short run adjustments. The ECT values for the variables are presented in Table 6.

From Table 6 the ECT2 is statistically significant at the 5% level and carries a negative coefficient (-1.2075), indicating a meaningful and stable correction mechanism. This suggests that deviations from the second long-run equilibrium relationship are corrected in the short run, with the agrofood supply chain emissions adjusting downward to restore balance. ECT5 is also statistically significant at the 1% level but has a positive coefficient (0.2102), implying that disequilibrium in the fifth long-run relation leads to short-run increase in the agrofood supply chain emissions. While this effect is significant, the positive sign may reflect a destabilizing influence that pushes the system away from equilibrium. ECT1 and ECT4 are marginally significant at the 10% level, suggesting weak corrective adjustments from the first and fourth long-run relationships, respectively. ECT3, however, is statistically insignificant, indicating that it does not play a notable role in short-run equilibrium correction. Overall, the results highlight that the agrofood supply chain partially corrects long-run disequilibrium, with some components stabilizing the system while others may introduce short-term volatility.

Table 6. Error Correction Term (ECT)

Variable	Estimate	Std. Error	t-Value	Pr(> t)	Significant
ECT1	4.3724	2.4228	1.805	0.099	Partially
ECT2	-1.2075	0.4567	-2.644	0.0228	Yes
ECT3	0.0275	0.1600	0.172	0.8668	No
ECT4	-2.6377	1.4334	-1.84	0.0929	Partially
ECT5	0.2102	0.0550	3.825	0.0028	Yes

Source: Authors' compilation from RStudio

Table 7. Short run dynamics of lagged differences

Variable	Estimate	Std. Error	t-Value	Pr(> t)	Significant
Log(supplychain).dl1	2.0158	1.3103	1.5380	0.1522	No
Log(foodtransport).dl1	-0.7986	0.4274	-1.8690	0.0885	No
Log(householdconsumption).dl1	-0.0272	0.0903	-0.3010	0.7690	No
Log(foodretail).dl1	-1.2716	0.6512	-1.9530	0.0768	No
Log(foodpackaging).dl1	-0.0008	0.1556	-0.0050	0.9960	No
Log(foodprocessing).dl1	-0.2975	0.2847	-1.0450	0.3185	No
Log(supplychain).dl2	2.5961	1.3196	1.9670	0.0749	No
Log (foodtransport).dl2	-0.8316	0.2984	-2.7870	0.0177	Yes
Log(householdconsumption).dl2	0.2765	0.1556	1.7770	0.1031	No
Log(foodretail).dl2	-1.7212	0.6832	-2.5190	0.0285	Yes
Log(foodpackaging).dl2	0.0960	0.1036	0.9270	0.3739	No
Log(foodprocessing).dl2	-0.3284	0.3264	-1.0060	0.3361	No

Source: Authors' compilation from RStudio

4.2.2. Short Run Dynamics

The short run dynamics are captured by the lagged differences of emissions, which show how emissions in each sector respond to past changes in emissions from other sectors benefiting from agrofood. The results of the short-run dynamics are displayed in Table 7.

From Table 7 all lagged values of study variables were insignificant except agrofood transportation in the second lag and agrofood retailing activities in the second lag. Agrofood transportation activities show a substantial short-run impact on agrofood supply chain emissions. This confirms the sector's responsiveness to supply chain fluctuations and external disruptions such as fuel prices and infrastructure constraints. Thus, a 1 unit increase of transport activities, leads to 83.16% decrease of agrofood supply chain emissions. This is based on the fact that agrofood transportation is an impetus of other downstream activities such as retailing practices. Nevertheless, this contradicts with Jehn et al. [11] and Crippa et al. [10] who argue that agrofood transportation contributes significantly to agrofood supply chain emissions. Thus, more studies in different setting and methodologies are called to justify the findings. On the other hand, Agrofood retailing operations have inconsistent but noticeable short-term effects. These were driven by energy demands for refrigeration and lighting in food outlets. From Table 7, 1 unit increase in retailing activities leads to 172.12% increase in agrofood supply chain emissions. The findings are in line with Crippa et al. [10] and Piñero et al. [14] who support that the retail sector plays a critical role in agrofood supply chain emissions. Other exogeneous variables including agrofood processing activities, agrofood packaging activities, and agrofood household consumption exhibit insignificant influence on agrofood supply chain emissions, suggesting that these variables may fluctuate seasonally depending on agrofood demand.

4.2.3. Long Run Effects

Table 8 reveals that all sectors except agrofood processing have a unit coefficient for their respective error correction terms (ECTs). In the long run, agrofood processing emerges the most influential sector on agrofood supply chain emissions. Its adjustment coefficients are strongly negative (-17.455, -11.497, -10.983, -27.582, -14.915), implying that this sector plays a stabilizing role. Changes in agrofood processing activities significantly pull the system back to equilibrium, highlighting its prime importance in the Tanzania's agrofood chain emissions. To this end, deviations in agrofood supply chain emissions are mostly corrected by changes in agrofood processing activities. These findings align with Smith et al. [5] who emphasise that agrofood processing activities are energy-intensive users. This is concomitant with the arguments of Love et al. [15] and Jehn et al. [11] that food processing significantly contributes to GHG emissions since is one of the most energy-intensive stages in the agrofood supply chain. Thus, modernizing food processing plants with energy-efficient machinery and solar or hydroelectric power sources can significantly reduce carbon emissions. However,

most SMEs in Tanzania's agrofood industry lack capital and infrastructure to implement such changes [14]. Therefore, the government should play a key role by devising a variety of financing and fiscal schemes for SMEs.

Table 8. Long run effects

Variable	ECT1	ECT2	ECT3	ECT4	ECT5
Supply Chain	1	0	0	0	0
Transport	0	1	0	0	0
Household	0	0	1	0	0
Retailing	0	0	0	1	0
Packaging	0	0	0	0	1
Processing	-17.455	-11.497	-10.983	-27.582	-14.915
Constant	0.137	0.096	0.083	0.223	0.133

Source: Authors' compilation from RStudio

Similarly, small constant values in Table 8 further imply that unexpected external shocks have limited long-term effect on agrofood supply chain emissions, suggesting that emissions in the agrofood supply chain are primarily driven by structural, rather than random factors. Furthermore, it is worth noting that other exogeneous variables including agrofood packaging activities, agrofood transportation, agrofood retailing operations, and agrofood household consumption exert insignificant long-run influence on agrofood supply chain emissions, suggesting that they don't serve as long-run anchors for stability of agrofood supply chain emissions. However, these variables are still relevant for crafting sustainable policies.

4.2.4. Autocorrelation Test

The Lagrange-multiplier was used to test the existence of autocorrelation. The two formulated hypotheses were: Ho: There is no autocorrelation present at lag order, Ha: There is autocorrelation present at lag order. The results of the Lagrange-multiplier are shown in Table 9.

Table 9. Lagrange multiplier test for auto correlation

Chi2	df	Prob>chi2
156	180	0.9015

Source: Authors' compilation from RStudio

Since the $p - value = 0.9015 > \alpha = 0.05$, the null hypothesis (Ho) is not rejected. This means that there is no significant autocorrelation in the residuals in the model, which is a good sign for model validity.

4.3. Forecasting of Agrofood Supply Chain Emissions Using ARIMA Model

An autoregressive integrated moving average (ARIMA) is a statistical analysis model that uses time-series data to either better understand the data set or to predict future trends which are used for analysing and forecasting time series data [32]. For the purpose of this study ARIMA model was applied to forecast and see the trend of agrofood supply chain emissions to 2030. It is worth noting that in order to achieve

stationarity, the endogenous variable (i.e. agrofood supply chain emissions) was transformed using natural logarithms prior to model estimation and forecasting.

Table 10. Auto selection ARIMA model

```
> print(fit)
Series: ts_forecast
ARIMA(1,0,2) with non-zero mean

Coefficients:
      ar1      ma1      ma2      mean
    0.9836  0.2576  0.3560  7.7624
s.e.  0.0227  0.1874  0.1776  1.5392

sigma^2 = 0.04259:  log likelihood = 4.28
AIC=1.44  AICc=4.17  BIC=8.1
```

Source: Authors' compilation from RStudio

The choice of using the best ARIMA model for forecasting based on the criteria of Akaike Information Criteria (AIC) was made automatically by RStudio. Table 10 reveals that ARIMA (1,0,2) was the most appropriate model for forecasting agrofood supply chain emissions.

The ARIMA (1,0,2) provides a forecast of agrofood supply chain emissions up to 2030 as depicted in Figure 5.

Figure 5 shows that ARIMA forecast projects a gradual decline in emissions from 2020 to 2030. Albeit modest, the forecast indicates that if current environmental practices are improved, the trend of agrofood supply chain emissions in Tanzania is promising. The stabilization of emissions is possible, provided that policy frameworks, technological upgrades, and stakeholder awareness continue to evolve positively. The decrease in agrofood supply chain emissions are in line with the achievement of several United Nations Sustainable Development Goals including SDG 15 (Life

on land), SDG 13 (Climate action), SDG 12 (Responsible consumption and production), SDG 11 (Sustainable cities and communities), and SDG 7 (Affordable and clean energy). The results are based on the premise of an ongoing implementation of measures to reduce agrofood supply chain emissions by all stakeholders including farmers, food processors, transporters, distributors, retailers, government and environmentalists. Belamkar et al. [33] contend that any agrofood supply chain should maximize profit while minimizing CO₂ emissions. This argument is supported by Basri et al. [12] who articulate the significant role of global food provision while shaping the quality of the environment to achieve carbon peak targets, carbon neutrality, and promote agricultural modernization. Moreover, understanding the environmental impact of agri-food sub-systems on agrofood supply chain emissions is crucial for informing all key stakeholders about the challenges and opportunities to mitigate GHG emissions in the food sector.

5. Conclusions

The study aimed to analyse and forecast carbon emissions across the Tanzania's agrofood supply chain emissions based on VECM-ARIMA model. The VECM model was applied to examine the short-run and long-run relationships between the research variables: agrofood processing activities, agrofood packaging activities, agrofood transportation, agrofood retailing activities, agrofood household consumption, and agrofood supply chain emissions. The ARIMA model was applied to forecast the Tanzania's agrofood supply chain emissions based on the data collected from FAO and IPCC over the time period from 1990 to 2020. The time series data were firstly subjected to the ADF test and found to be stationary after differencing. The same results were supported by the graphical test of the time series data.

Forecast of Supply Chain Emissions (1991-2030)

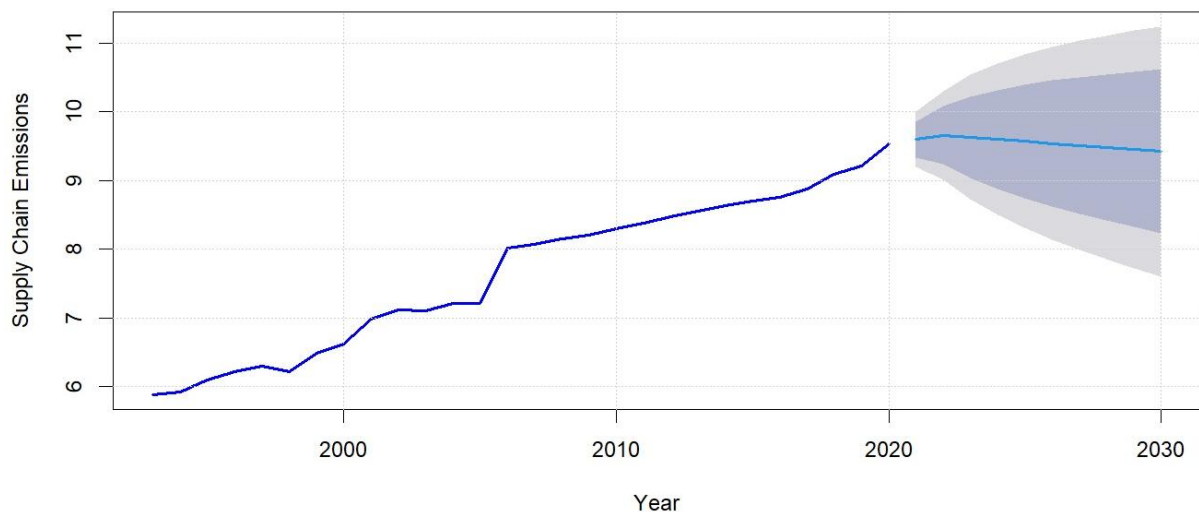


Figure 5. ARIMA (1,0,2) forecast of agrofood supply chain emissions (**Source:** Authors' compilation from RStudio)

The statistical tests were carried out to examine cointegration among variables and the normality of data set. The Shapiro-Wilk test confirmed the normality for all variables. Meanwhile the Johansen test of cointegration revealed the existence of both long-run and short-run relationships between agrofood supply chain emissions (i.e. endogenous variable) and agrofood processing activities, agrofood packaging activities, agrofood transportation, agrofood retailing operations, and agrofood household consumption (i.e. exogenous variables). More specifically, the short-run dynamics revealed that only agrofood transportation and agrofood retailing activities pose a significant influence on agrofood supply chain emissions. In the long-run, only agrofood processing activities exert a significant influence on agrofood supply chain emissions since its error correction term is strongly negative. Furthermore, the Lagrange-multiplier test revealed the non-existence of autocorrelation which confirms the model validity. On the other hand, the AIC was used to choose ARIMA (1,0,2) which provided a forecast of agrofood supply chain emissions from 2020 till 2030. The ARIMA forecasting model reveals a slightly decline of agrofood supply chain emissions up to the year 2030. This suggests that emission mitigation efforts by all stakeholders in the agrofood supply chain may be beginning to take place. Furthermore, understanding the environmental impact of agrofood sub-systems of the agrofood supply chain is crucial for informing all stakeholders about the challenges and opportunities to mitigate GHG emissions in the food sector. Specifically, the reduction of emissions is possible provided green policies and regulations, green technology and stakeholders' awareness towards sustainable supply chains continue to evolve. Furthermore, this study can serve as guidance for other countries where agrofood supply chain emissions are still escalating. Our future research direction is to model and forecast emissions from agri-food waste disposal systems based on VARIMA model.

Declarations

All authors declare that they have no conflicts of interest.

REFERENCES

- [1] Gao, Y., Gao, X. & Zhang, X. 2017. The 2°C global temperature target and the evolution of the long-term goal of addressing climate change – From the United Nations Framework Convention on Climate Change to the Paris Agreement. *Engineering*, Vol. 3, No. 2, pp. 171-193.
- [2] Acheampong, A.O. & Boateng, E.B. 2019. Modeling carbon emission intensity: Application of artificial neural network. *Journal of Cleaner Production*, Vol. 225, pp. 833-856.
- [3] Pollitt, M.G. 2019. A global carbon markets? *Frontiers of Engineering Management*, Vol. 68, No. 1, pp. 171-193.
- [4] Massami, E.P. 2025. Economic and environmental sustainability for BRT systems: A comparative analysis of electric and diesel buses based on comprehensive LCC model. *International Journal of Energy Engineering*, Vol. 15, No. 1, pp. 1-13.
- [5] Smith, P., Bustamante, M., Ahammad, H., Clark, H., Dong, H., Elsiddig, E. A., Haberl, H., Harper, R., House, J., Jafari, M. & Masera, O. 2014. Agriculture, forestry and other land use (AFOLU). In *Climate change 2014: mitigation of climate change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 811-922). Cambridge University Press.
- [6] Food and Agriculture Organization (FAO). 2020. Tanzania Agriculture Sector Overview. [Online]. Available: <https://www.fao.org>. [Accessed July 04, 2025].
- [7] World Bank. 2019. *Environmental Challenges in Sub-Saharan Africa: Sectoral Reports*. World Bank Publications. [Online]. Available: <https://www.worldbank.org>. [Accessed August 13, 2025].
- [8] Kusakabe, K., & Mwaniki, D. 2017. Carbon footprint of transportation in low-income countries: A policy analysis. *Environmental Sustainability Journal*, Vol. 15, No. 3, pp. 215-230.
- [9] United Nations Framework Convention on Climate Change (UNFCCC). 2015. Paris Agreement. [Online]. Available: <https://unfccc.int/process-and-meetings/the-paris-agreement>. [Accessed July 04, 2025].
- [10] Crippa, M., Tubiello, F. N., Karl, K., Flammini, A. & Gütschow, J. 2022. Pre- and post-production processes increasingly dominate greenhouse gas emissions from agri-food systems. *Earth System Science Data*, Vol.14, No. 4, pp. 1795–1809.
- [11] Jehn, F. U., Gajewski, L. G., Hedlund, J., Arnscheidt, C. W., Xia, L., Wunderling, N. & Denkenberger, D. 2024. Food trade disruption after global catastrophes. [Online]. Available: <https://eartharxiv.org>. [Accessed July 04, 2025].
- [12] Basri, Abdullah, Gunawan, S. & Hamid, A.M.A. 2024. Environmental sustainability development in Indonesia integrating with the green economy: A VECM analysis. *Proceedings of the 8th International Conference on Islamic Studies (ICONIS), July 30-31, 2024, IAIN Madura, Pamekasan, East Java, Indonesia*.
- [13] Behvandi, O. & Ghorbani, H. 2024. Predicting future climate scenarios: A machine learning perspective on greenhouse gas emissions in agrifood systems. *Frontiers in Environmental Science*, Vol. 12, pp. 1471599.
- [14] Piñero, P., Aguilera, E., Travieso, E., Infante-Amate, J., Bruckner, M., Gerber, J., Lassaletta, L., Mueller, N., & Sanz-Cobena, A. 2022. Agrofood greenhouse gas emissions are increasingly driven by foreign demand. [Online]. Available: <https://researchsquare.org>. [Accessed July 16, 2025].
- [15] Love, D. C., Conrad, Z., Altema-Johnson, D., Ramsing, R., Bassarab, K., Thorne-Lyman, A. L., Williams, D., Moore, E. H., Nussbaumer, E. M., & Anderson, P. 2024. *Nutrients, Environmental Sustainability, and Cost of the National School Lunch Program and School Breakfast Program in the United States: National Health and Nutrition Examination Survey, 2011-2018*.
- [16] Stoeckli, S., Merian, S., Wanner, S., Stucki, M. & Chaudhary, A. (2024). *Advancing Biodiversity Footprinting for Food-Related Behavior Change* (No. zpvq4_v1). Center for Open Science.

- [17] Keshavarz-Ghorbani, F. & Pasandideh, S.H.R. 2022. A lagrangian relaxation algorithm for optimizing a bi-objective agro-supply chain model considering CO₂ emissions. *Annals of Operations Research*, Vol. 314, No. 2, pp. 497-527.
- [18] Accorsi, R., Cholette, S., Manzini, R., Pini, C. & Penazzi, S. 2016. The land-network problem: Ecosystem carbon balance in planning sustainable agro-food supply chains. *Journal of Cleaner Production*, Vol. 112, pp. 158-171.
- [19] Dwivedi, A., Jha, A., Prajapati, D., Sreenu, N. & Pratap, S. 2020. Meta-heuristic algorithms for solving the sustainable agrofood grain supply chain network design problem. *Modern Supply Chain Research and Applications*, Vol. 2, No. 3, pp. 161-177.
- [20] Cui, Y., Diarrassouba, I., Joncour, C., & Michel Loyal, S. 2024. Optimization and Analysis of the Impact of Food Hub Location on GHG Emissions in a Short Food Supply Chain. *Sustainability*, Vol. 16, No. 17, pp. 7781.
- [21] Bughio, A., Teng, Y., Tunio, R.A., Das, G., Jamali, R. & Shar, R.U. 2023. The short- and long-run causal correlation between green finance, renewable energy consumption, and economic growth. *Energy & Environment*, Vol. 36, No. 2, pp. 888-907.
- [22] Tran, T.M., Phan, T.H., Tran, T.V., & Le, A.T.T. 2022. Examining the correlation among economic development, and CO₂ emissions by utilizing the VECM model – Empirical study in Vietnam. *Sustainability*, Vol. 14, No. 19, pp. 12621.
- [23] Hakimi, A. & Hamda, H. 2016. Trade liberalization, FDI flows, environmental quality and economic growth: A comparative analysis between Tunisia and Morocco. *Renewable and Sustainable Energy Reviews*, Vol. 58, pp. 1445-1456.
- [24] Soytaş, U. & Sari, R. 2007. The relationship between energy and production: Evidence from Turkish manufacturing industry. *Energy Economics*, Vol. 29, No. 6, pp. 1151-1165.
- [25] Brockwell, P.J. & Davis, R.A. 2016. Introduction to Time Series and Forecasting (3rd ed.). Springer.
- [26] Montgomery, D.C., Jennings, C.L. & Kulahci, M. 2015. Introduction to Time Series Analysis and Forecasting (2nd ed.). Wiley.
- [27] Jalil, A. & Rao, N.H. 2019. Chapter 8 - Time Series Analysis (Stationarity, Cointegration, and Causality). Environmental Kuznets Curve (EKC). Academic Press. pp. 85-99.
- [28] McDonald, J.H. 2014. Handbook of Biological Statistics (3rd ed.). Sparky House Publishing.
- [29] Thiese, M.S., Ronna, B. & Ott, U. (2016). P value interpretations and considerations. *Journal of Thoracic Disease*. Vol. 8, No. 9, pp. E928–E931.
- [30] Hyndman, R.J. & Athanasopoulos, G. 2018. Forecasting: principles and practice (2nd ed.). OTexts.
- [31] Lord, D., Qin, X. & Geedipally R, S. 2021. Chapter 2 - Fundamentals and datacollection. *Highway Safety Analytics and Modeling*. pp. 17-57.
- [32] Adhikari, R. & Agrawal, R. K. 2013. An introductory study on time series modeling and forecasting. *arXiv preprint arXiv: 1302.6613*.
- [33] Belamkar, P., Biswas, S., Baidya, A., Majumder, P. & Bera, U.K. 2023. Multi-objective optimization of agro-food supply chain networking problem integrating economic viability and environmental sustainability through type-2 fuzzy-based decision making. *Journal of Cleaner Production*, Vol. 421, pp. 138294.