Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2024, 2024:86
https://doi.org/10.28919/cmbn/8673
ISSN: 2052-2541

A PROPOSED ROBUST REGRESSION MODEL TO STUDY CARBON DIOXIDE EMISSIONS IN EGYPT

ABEER R. AZAZY[1], MOHAMED R. ABONAZEL[2], ABANOUB M. SHAFIK[3], TAREK M. OMARA[4,*], NESMA M. DARWISH[5]

[1]Al-Alsun Higher Institute for Tourism, Hotels and Computer, Cairo, Egypt

[2]Department of Applied Statistics and Econometrics, Faculty of Graduate Studies for Statistical Research, Cairo University, Egypt

[3]Department of Statistics, Faculty of Politics and Economics, Beni-Suef University, Beni-Suef, Egypt

[4]Department of Mathematics, Faculty of Sciences, Islamic University of Madinah, Madinah, Saudi Arabia

[5]Business Administration Department, Faculty of Politics, Economics and Business Administration, May University in Cairo (MUC), Egypt

Abstract: Developing countries face environmental challenges as they rely on non-renewable energy for economic growth. This study examined the dynamic relationship between CO_2 emissions and four economic factors (manufacturing, trade, urban population, and cereal production land) in Egypt from 1990 to 2021. It introduced a new approach to estimate the autoregressive distributed lag (ARDL) model using robust methods (M, MM, and S) and compared them with ordinary least squares (OLS) to determine the best estimate. The ARDL methodology was employed to test short-term and long-term relationships. Results showed ARDL (1,0,2,1,3) as the most suitable model. Manufacturing and Trade variables negatively impacted CO_2 emissions, while Urban population and Land

[*]Corresponding author
E-mail address: Tarek_em@iu.edu.sa
Received May 31, 2024

variables had positive long-term effects. The period lags of Trade and Land variables have significantly affected CO_2 levels. The error correction model indicated economic adjustments occur within about 25 months. The study found that robust methods (M, MM, and S) outperformed the non-robust OLS method. Among these, the S estimation method proved most effective, showing the lowest Akaike information criterion (AIC) and Bayesian information criterion (BIC) values.

Keywords: carbon emissions; manufacturing; trade; urban population; land under cereal production; ARDL model; robust regression.

2020 AMS Subject Classification: 62F35, 37M10.

1. INTRODUCTION

The energy policy in Egypt has a vital role in raising both the environmental issues induced by CO_2 and the economic circumstances. Energy consumption and urbanization dropped by 2.47% and 1%, respectively in 2014, while gross capital formation and trade increased by 15.71% and 6.6% in 2018 (World Bank 2020), and CO_2 per capita in Egypt corresponded to 2.32 tons per capita in 2016, a rise of 0.06 over the figure of 2.27 tons of CO_2 per capita recorded in 2015; this symbolizes a 2.5% change in CO_2 per capita (World Bank 2020). Recently, Egypt has been categorized by rapid economic growth with an increasingly urban population, a rigid demand for energy, and increased CO_2. Egypt needs to reorganize national environmental policies, particularly those attributed to reducing CO_2 [1].

Unquestionably, economic growth is one of the main elements influencing a nation's success. As a result, several nations have looked for strategies to promote economic expansion even at the expense of the environment. Nevertheless, it is important to remember that there is a reciprocal link between economic progress and the environment. Economic expansion may result in environmental deterioration, while reducing ecological quality can also harm economic expansion. According to theory, resource depletion, health issues, and natural catastrophes that are caused by environmental degradation, can impede economic progress [2]. Global warming, which is mostly brought on by CO_2, is the most important environmental issue [3].

Global economies have changed as a result of the Industrial Revolution from being based on organic production methods driven by humans and animals to inorganic processes, which are typically based on non-renewable resources for energy use, like fossil fuels, which significantly accelerate global warming by trapping heat in the atmosphere and producing greenhouse gases (GHGs). According to both environmentalists and legislators, one of the most important environmental concerns of the past few decades has been global warming. Human economic activity harms the environment, the rate at which greenhouse gases are released into the earth's atmosphere worldwide has increased due to economic growth in both developed and emerging nations [1].

Methane (CH4), CO_2, hydrofluorocarbons (HCFC), nitrous oxide (N2O), sulfur hexafluoride (SF6), and chlorofluorocarbons (CFC) are the main components of GHGs that are trapped in the atmosphere. Due to its tremendous heat-trapping capacity and the fact that it is the most concentrated GHG in the earth's atmosphere, CO_2, which is mostly produced by burning fossil fuels, making cement, and changing land use, has been identified as the primary cause of global warming (World Bank (2018)).

The primary goal of emerging nations is to raise their quality of life, which may be done via accelerating economic growth. Increasing gross capital formation from raising investment, increasing employment, and opening work opportunities can lead to faster economic growth. Higher savings are the result, which inspires confidence to make larger investments and boosts output production. These procedures create a domino effect that leads to an ongoing rise in output production and economic expansion. Faster economic expansion also means more urbanization and energy consumption, which increases CO_2 mostly in poor (developing countries) nations because of the misuse or abuse of non-renewable energy sources. As a result, this problem has come to be a concern for policymakers as well as scholars. The commitment of the world's largest emitters is crucial for the efficacy of CO_2 reduction. Since CO_2 is a byproduct of energy production and controlling, it has become more difficult. As a result, efforts to avoid CO_2 might negatively impact economic growth, particularly in emerging nations [1].

A key idea in modern times is sustainable development, which is described as growth that satisfies current demands without jeopardizing the capacity of the next generations to fulfill their own needs. This idea, which initially came to light in the 1970s, addresses social, environmental, and economic growth. One of the effects of unsustainable growth is climate change. This is because people have overused fossil fuels, a natural resource, to create power by clearing forests [4], neglecting the effects of this development on the environment and future generations. For example, the International Energy Agency (EIA) reported that global CO_2 was between 17.78 billion and 32.1 billion tonnes in 1980 and 2015, respectively, and the ensuing warming has impacted biodiversity, economic activity, agriculture, human health, and ecosystem functioning. The concentration of GHG in the atmosphere might double its preindustrial levels by as early as 2035 if no steps are made to curb such emissions [5].

The dynamic moderation of social and economic capacities from rural to urban regions is known as urbanization. Approximately half of the world's population resides in urban areas, and by 2050, 64% of people living in developing nations will do so as well. Practically all these activities occur in metropolitan areas; this phenomenon has contributed to increased energy consumption and CO_2 emissions from industry and other economic activities. Additionally, as a result of urbanization, the majority of rural dwellers have raised their standards of living and modified certain aspects of their lifestyles. Urban residents' consumption levels are rising along with their incomes and living conditions, and their consumption patterns are gradually changing from survival to enjoyment mode [6]. This could lead to an increase in urban energy use due to consumption patterns shifting from survival mode to enjoyment mode. Urbanization and CO_2 are positively correlated, as several prior studies have shown. The main sources of CO_2 emissions in urban areas are residential homes, transportation, and the construction material industry [5].

Zhu and Peng [7] stated that there are three main ways that urbanization influences CO_2: first, energy consumption in residential and industrial settings; second, energy utilized by the construction industry to build better transportation, residential buildings, and infrastructure; and third, conversion of grasslands and woodlands to make way for urban development. Moreover, increased usage of domestic equipment (such as air conditioners and water heaters) has resulted in high electrical power consumption and has an indirect impact on CO_2 levels.

These similar series of trends make Egypt an interesting country for re-examining the main determinants of environmental degradation and the interaction between the main indicators, where the independent variables include Manufacturing, Trade, Urban population, and Land under cereal production. Using recently developed techniques for short-term and long-term dynamics, as well as modern statistical theories such as robust and non-robust, we will propose a new method for estimating ARDL model by M, MM, and S robust estimation methods, then compare OLS and robust methods to describe the best estimate by using the AIC and BIC metrics. All previous procedures serve as a resource for policymakers to design sustainable and effective policies.

The remaining sections of this paper are as follows: Section 2 includes a summary of some important previous studies. Section 3 involves our methodology, which is divided into two subsections, non-robust and robust estimation methods. The results of our empirical study are presented in Section 4, while Section 5 includes the main conclusions of our study.

2. Previous Studies

This section focuses on applied research related to CO_2 and the economic factors affecting it and is interested in applications using the ARDL model.

Hossain and Hasanuzzaman [8] aimed to analyze the link between CO_2, energy consumption, economic growth, urbanization, financial development, and trade openness in Bangladesh. by using the ARDL bounds testing approach to investigate the short- and long-term dynamics and cointegrating. The empirical findings for Bangladesh spanning from 1975 to 2010 indicate the presence of a significant long-term connection among the variables in Bangladesh at a 1% level of significance. The positive and highly significant estimated coefficients for energy consumption and urbanization suggest that the increasing levels of energy consumption and urbanization contribute to CO_2 in Bangladesh. The research shows, nevertheless, that a rise in real GDP per person is associated with a decrease in per-person CO_2 emissions. CO_2 emissions, on the other hand, are not causally related to trade openness or financial progress.

Bekhet and Othman [5] aimed the connection between CO_2, the rate of urbanization, domestic investment, energy consumption, financial development, and GDP was the goal of this study between the years 1971 - 2015. Granger causality is tested using the VECM and F-bounds test. Long-term inverted U-shaped links between CO_2 and urbanization as well as the dynamic relationship among factors are analyzed. Early in the urbanization process, the elasticity of CO_2 is shown to be positively elastic; but, as urbanization progresses, it becomes negatively inelastic. Moreover, there is a five percent significance level for the bidirectional causative relationship between urbanization and CO_2 over the long run and a one percent significance level for the unidirectional causation relationship between urbanization and CO_2 in the short term. Additionally, at least at a 5 percent significant level, we were able to identify unidirectional causation from financial development to CO_2 accompanied by bidirectional causality between domestic investment, energy consumption, GDP, and CO_2.

Rayhan et al [9] aimed to analyze the influence of urbanization and energy consumption on CO_2 within the theoretical framework of the Environmental Kuznets Curve (EKC) hypothesis. The EKC hypothesis posits that during the initial stages of economic growth, CO_2 rises alongside development levels but begins to decrease after reaching a peak at a higher level of economic advancement. The study employed the ARDL model. The results confirm the integration of our variables, as deviations from the long-term equilibrium are annually adjusted by 77.19% towards the long-term equilibrium path. The empirical results offer support for the credibility of the EKC hypothesis over both short and long durations within the Bangladeshi context. Furthermore, the correlation between urbanization, energy consumption, and CO_2 is statistically significant and positive, while the effects of FDI and economic openness are statistically deemed insignificant.

Danish et al. [10] determined how that transportation economic expansion, energy consumption, and CO_2 from the transportation sector which includes urbanization and foreign direct investment relate to each other. ARDL and VECM are used in this study conducted in Pakistan from 1990 to 2015. The empirical evidence indicates that transportation energy consumption has a significant influence on CO_2 from the transportation industry. Moreover, CO_2

emissions are also a result of foreign direct investment. It's interesting to note that economic expansion and urbanization have no statistically significant impact on CO_2 from transportation.

Ali et al. [11] explored to investigate the relationship between Pakistan's GDP, land under cereal crops (LCC), agricultural value-added (AVA), and CO_2 from 1961 to 2014 based on the ARDL models, descriptive statistical analysis, and pairwise Granger causality tests. The research utilizes the Phillips-Perron (PP) and augmented Dickey-Fuller (ADF) tests to verify the stationarity of the variables. The analysis's findings show that the nation's agricultural output, economic expansion, and CO_2 are all correlated over the short and long run. According to the long-term data, CO_2, LCC, and AVA are all positively and marginally correlated. The short-run analysis's findings indicate that the relationship between CO_2 and GDP is negative and statistically insignificant.

Karedla et al. [12] aimed to analyze the influence of trade accessibility, economic expansion, and industrial activities on India's CO_2. The study utilized an ARDL bounds test method and incorporated GDP, manufacturing, trade, and CO_2 to scrutinize the correlation using data from 1971 - 2016. Findings illustrate a sustained correlation between CO_2 and other factors. Trade openness notably diminishes CO_2, whereas GDP and manufacturing exert a substantial and positive influence on CO_2 in the long term.

Jiang et al. [13] conducted a useful study that examined the connection between China's economic growth, urbanization, CO_2 emissions, and foreign commerce between 1971 and 2020. Applying the ARDL model, the results show that, when urbanization, economic growth, and CO_2 are taken into account as explanatory variables, there is a persistent cointegration connection between the variables. Urbanization has a significantly beneficial effect on CO_2 and economic growth within this long-lasting connection, with coefficients of 0.2921 and 2.2172, respectively. With a long-term elasticity value of 0.4864, economic growth and urbanization are more significant than the 1% criterion. In the short run, some relationships reinforce each other between urbanization and CO_2, economic growth and CO_2, and urbanization and economic development, whereas foreign trade is seen to curb CO_2 in the short term.

Priyanto and Chanthavy [14] attempted to determine the short- and long-term relationships between several factors associated with economic growth in Cambodia, including internet literacy, CO_2, and educational attainment. As an additional source of statistical data, we used data between 2000 - 2020. Our estimating results teach us that the variables we predicted economic development, internet literacy, and growth have both long- and short-term relationships, as education. In the short run, internet literacy significantly boosts economic growth. Accordingly, A robust positive association exists between economic growth and educational attainment. There is a correlation between economic growth and CO_2 in addition to the statistically significant correlation between the economic growth variable for this year and the economic growth for the previous year.

Liu and Zhang [15] delved into the effects of trade-offs between ecological construction and urban expansion on CO_2 by utilizing diverse data sources and conducting statistical and spatial analysis. The study focused on ten urban agglomerations in China representing varying levels of development. Findings revealed that (1) urban growth in the selected areas increased by 90% to 288.9%, while ecological land experienced changes ranging from 28.1% to 2.3%. (2) CO_2 emissions in urban agglomerations, particularly in the southern and eastern regions of China, exhibited a growth of 154.2% to 291.2%. Concentrations of CO_2 with high values within all urban agglomerations expanded outward from the cities. (3) The expansion of urban areas led to increased emissions, although ecological construction efforts partially mitigated this growth in emissions.

Tabassum et al. [16] aimed to investigate the influence of Institutional quality, Employment, and trade openness on environmental conditions (specifically CO_2 in metric tonnes) within the timeframe spanning from 2002 to 2021. The analysis utilizes data from the top ten CO_2-producing nations: Iran, China, India, Russia, the United States, South Korea, Germany, Japan, Saudi Arabia, and Indonesia. The study employs the panel-based ARDL model, which incorporates a unit root test to assess the stationarity characteristics of the dataset. The regression coefficients of the long-term equation reveal that CO_2 levels are impacted by the predictors EMP, IMP, and IQ. Following the identification of significant cointegration, both the short-term and long-term coefficients for the models are computed. The model exhibits negative and statistically significant lagged Error Correction Term coefficients, suggesting a highly stable long-term relationship among the variables. The panel ARDL outcomes in both the long-run and short-run demonstrate that Institutional Quality exerts a notable influence on CO_2; specifically, an enhancement in IQ leads to a reduction in CO_2. Employment and Trade openness also play significant roles in affecting the CO_2 environment.

Tanveer et al. [17] studied and explored the effect of GHGs (CO_2, N2O, and CH4) and ecological impressions on various aspects such as urbanization, deforestation, economic growth, AVA, and globalization in Pakistan by using the ARDL model from 1990 to 2017 in Pakistan. Our research findings indicate a significant positive relationship between CO_2 and deforestation, AVA, globalization, and urbanization, leading to prolonged environmental deterioration. Furthermore, our analysis reveals that in the long term, ecological footprint plays a crucial and positive role in linking deforestation, agricultural land, and economic growth. Additionally, the study shows that methane emissions contribute to increased deforestation, AVA, globalization, and economic progress. On the other hand, a minor proportion of nitrous oxide demonstrates a negative association with deforestation and AVA but exhibits a positive correlation with economic growth, urbanization, and globalization. Further analysis using variance decomposition and impulse response function is conducted to investigate the causative relationships with the variables.

Through previous studies, there is a gap in the previous studies that does not cover in terms of variables explaining carbon dioxide emissions. The research paper is interested in studying the effect of Manufacturing, Trade, Urban population, and Land under cereal production on the dependent variable carbon dioxide emissions for data from 1990 to 2021. We will also propose a new ARDL model based on the Robust estimation method instead of the regular model that relies on OLS, and we will compare the results to prove that the proposed model is better.

Table 1 presents an overview of previous studies on CO_2 in different countries.

Table 1: Overview of previous studies on CO_2

Study	Country	Study period	Main variables that affect CO_2
Hossain and Hasanuzzaman [8]	Bangladesh	1975 - 2010	- energy consumption - economic growth - urbanization - financial development - trade openness
Bekhet and Othman [5]	Malaysia	1971 - 2015	- urbanization - energy consumption - GDP - domestic investment - financial development
Rayhan et al. [9]	Bangladesh	1971 - 2010	- urbanization - GDP
Danish et al. [10]	Pakistan	1990 - 2015	- transportation energy - urbanization - foreign direct investment - economic expansion
Ali et al. [11]	Pakistan	1961 - 2014	- land under cereal crops - GDP - agricultural value-added
Karedla et al. [12]	India	1971 - 2016	- influence of economic expansion - trade accessibility - industrial activities
Jiang et al. [13]	China	1971 - 2020	- economic growth - urbanization - foreign commerce
Priyanto and Chanthavy [14]	Cambodia	2000 - 2020	- internet literacy - educational attainment
Liu and Zhang [15]	China	2000 - 2018	- urban expansion - ecological construction
Tabassum et al. [16]	Top ten CO_2-producing nations	2002 - 2021	- institutional quality - employment - trade openness
Tanveer et al. [17]	Pakistan	1990 - 2017	- urbanization - deforestation - economic growth - agricultural land - globalization

Through the previous presentation of studies, we find that the research gap lies in the studies that comprehensively analyze the combined effect of Manufacturing, Trade, Urban population, and Land under cereal production on CO_2 in Egypt. While individual studies may exist these factors in a comprehensive framework allow for a more comprehensive understanding of dynamics and interactions between these variables and their combined impact on CO_2.

3. METHODOLOGY

This section is divided into three parts. In the first part, we will explain OLS ARDL, or by another name, non-robust ARDL. In the second part, we will explain the proposed robust ARDL models. Finally, we will explain the tools for comparing the two types using AIC and BIC.

3.1 OLS ARDL Model

The ARDL model is the major workhorse in dynamic unconstrained single-equation regression models in econometric literature. It is concerned with analyzing the long-term relationships between integrated variables and re-measuring the relationship between those studied in the error correction model, ARDL models typically start with a large and general dynamic model and gradually lower their mass and change variables by imposing linear and non-linear constraints. The ARDL model is distinguished from the rest of the co-integration models that preceded it in that it applies co-integration analysis in cases of variables that are combined from different orders such as I (0) and I (1). The ARDL model uses the current and delayed values of the independent variable and the delayed values of the dependent variable, therefore it should be able to tackle the correlation problem [18, 19, 20].

The general ARDL $(p, q_1, q_2, \dots, q_k)$ is given by the following equation:

$$y_t = \beta_0 + \sum_{i=1}^{p} \beta_i y_{t-i} + \sum_{j=1}^{k} \sum_{h=0}^{\mathrm{q}_j} \alpha_{j,h} x_{\mathrm{j,t-h}} + e_t \qquad t = 1,2, \dots, T, \tag{1}$$

where e_t is the error term, p is the number of lags in the dependent variable, q_k is the number of lags of the k^{th} and $\beta_0, \beta_i, \alpha_{j,h}$ are the parameters of the model.

The assumption of the ARDL $(p, q_1, q_2, \dots, q_k)$ model in Eq. (1), could be written as follows:

1. Linear in parameter.
2. $E(e_t) = 0; \quad t = 1, 2, \dots, T.$
3. $Var(e_t) = \sigma^2; \ t = 1, 2, \dots, T.$

4. Cov $(e_t, e_s) = 0$; $t \neq s$, $t = 1, 2, \dots, T$, $s = 1,2, \dots, T$.
5. Cov $(e_t, x_{jt}) = 0$; $\forall\, t = 1, 2, \dots, T$; $j = 1,2, \dots, k$.
6. $e_t \sim N\,(0, \sigma^2)$; $t = 1, 2, \dots, T$.

Considering the premise that estimation corresponds to the express method of least squares, it follows that the assumption warrants using estimation as a suitable technique [21].

To analyze the impact of the Manufacturing, Trade, Urban population, and Land under cereal production on the CO_2, the linear relationship. According to [22], the error correction representation of the ARDL model is:

$$\Delta CO_{2_t} = \beta_0 + \sum_{i=1}^{p} \beta_{1i} \Delta CO2_{t-i} + \sum_{h=0}^{q_1} \beta_{2h} \Delta \mathrm{MANUF}_{t-h} + \sum_{h=0}^{q_2} \beta_{2h} \Delta Trade_{t-h} + \sum_{h=0}^{q_3} \beta_{2h} \Delta \mathrm{Urban}_{t-h} + \sum_{h=0}^{q_4} \beta_{2h} \Delta \mathrm{Land}_{t-h} + \alpha_1 CO_{2_{t-1}} + \alpha_2 \mathrm{MANUF}_{t-1} + \alpha_3 Trade_{t-1} + \alpha_4 \mathrm{Urban}_{t-1} + \alpha_5 \mathrm{Land}_{t-1} + \mathcal{E}_t. \quad (2)$$

The long-run equation is:

$$CO_{2_t} = \beta_0 + \alpha_1 \mathrm{MANUF}_t + \alpha_2 Trade_t + \alpha_3 \mathrm{Urban}_t + \alpha_4 \mathrm{Lan}_t + u_t. \quad (3)$$

CO_2 represents the dependent variable. The independent variables expected to influence CO_2 are Manufacturing, Trade, Urban population, and Land under cereal production. $\alpha_1, \alpha_2, \alpha_3$ and α_4 are the regression coefficients, representing the magnitude and direction of the influence of each independent variable on the CO_2. While β_0 is the intercept of the regression Eq. (3). $\mathcal{E}$ is the error term, capturing the unexplained variation in the CO_2 not accounted for by the independent variables.

The estimated coefficients of $\alpha_1, \alpha_2, \alpha_3$ and α_4 provide insights into the relationship between each independent variable and the CO_2. A positive coefficient suggests a positive relationship, indicating that an increase in the independent variable leads to an increase in the inflation rate, while a negative coefficient suggests a negative relationship.

3.2 Proposed Robust ARDL Models

This part will be divided into three methods: M-estimation method, S-estimation method, MM-estimation method.

3.2.1 M-Estimation Method

The M-estimation method, which has been noted as the most popular strategy for robust regression, was examined by [23]. Huber [24] introduced this approach and applied maximum likelihood estimation to location models. The M-estimation approach is virtually as effective as OLS; however, it concentrates on minimizing the residual function rather than the sum of squared errors.

Hampel et al. [25] established a system of normal equations to address this minimization problem that will be detected by taking partial derivatives concerning β and defining them equal to zero, $X^T W X \beta^T = X^T W y$. Where W is an $(n \times n)$ diagonal matrix of weight, top-rated functions for M-estimators:

$$\hat{\beta}_M = (X^T W X)^{-1} X^T W y, \tag{4}$$

where X here is the regressors matrix, y is the vector of dependent variable, and β is the vector of unknown parameters of the regression model.

In other words, we can say that the M-estimation is a generalization of maximum likelihood estimation. It aims to minimize a function of the residuals:

$$min \sum_{t=1}^{T} \rho(r_t/\sigma_m), \tag{5}$$

where $\rho(\cdot)$ is a loss function, $r_t = y_t - x_t^T \hat{\beta}_{ols}$ are the residuals of the OLS model, x_t^T is the t-th row of X matrix (where X is an T × (k+1) matrix of predictors including a column of 1s for the intercept), and σ_m is a scale parameter. The estimator is defined implicitly by the following normal equations:

$$\sum_{t=1}^{T} \psi(r_t/\sigma_m)\, x_{tj} = 0;\ for\ j = 1, \dots, k+1, \tag{6}$$

where $\psi = \rho'$ is the derivative of ρ function. The iteratively reweighted least squares (IRLS) algorithm is often used to solve this:

$$\beta_{(a+1)} = (X^T W_{(a)} X)^{-1} X^T W_{(a)} y, \tag{7}$$

where $W_{(a)}$ is a diagonal matrix with entries $w_{t(a)} = \psi(r_{t(a)}/\sigma_m) \,/\, (r_{t(a)}/\sigma_m)$.

The common choice for $\rho(z)$ is the Huber function:

$$\rho(z) = \begin{cases} z^2/2 & if\ |z| \leq c; \\ c|z| - c^2/2 & if\ |z| > c, \end{cases} \tag{8}$$

where $c = 1.345$. M-estimation is particularly effective against outliers in the response variable but can be sensitive to leverage points.

3.2.2 S-Estimation Method

The S-estimation method that proposed by Rousseeuw and Yohai [26] is essentially based on the residual scale of the M-estimation method and is an expansion of the least median of squares (LMS) and least trimmed squares (LTS) robust methods. S-estimators possess equivalent asymptotic properties as M-estimators but can handle up to 50% of the outliers present in the data. However, one major drawback of the M-estimation method is its failure to consider the data distribution and its inability to function based on overall data, since it only relies on the median as the weighted value. Thus, the S-estimator can be defined as follows:

$$\hat{\beta}_s = min_{\beta}\ \hat{\sigma}_s(r_1, r_2, \dots, r_T). \tag{9}$$

S-estimators are more resistant than M-estimators because of their smaller asymptotic bias and variance when handling contaminated data. This unique characteristic makes them an ideal choice when dealing with data that may be outliers [27].

3.2.3 MM-Estimation Method

The MM estimation technique is an exclusive variant of the M-estimation approach, conceptualized by Yohai [28]. The MM estimation method combines a high breakdown value estimation methodology with an effective estimation mechanism, which culminates into the MM estimator. The MM estimator stands out as the pioneering approach that assembles a high breakdown point and high efficiency under normal error. MM estimation method is designed to get the estimators who have high breakdown values, and they are more efficient.

MM-estimation follows these steps: Step 1: Compute an initial S-estimate $\hat{\beta}_s$ with a high breakdown point but possibly low efficiency. Step 2: Compute an M-scale estimate $\hat{\sigma}_m$ based on the residuals from $\hat{\beta}_s$. Step 3: Compute the final MM-estimate $\hat{\beta}_{MM}$ by M-estimation, using $\hat{\sigma}_m$ as a fixed scale.

The MM-estimator gets the breakdown point of the initial S-estimator and can achieve high efficiency under normal errors.

3.3 Model Selection Criteria

It is a measure of model performance that accounts for model complexity. This part will be divided into two methods.

3.3.1 Akaike information criterion

Akaike [29] suggested a specific measure of goodness of fit of the model. The model with a smaller AIC is considered the best. It is administered:

$$AIC = -2\ ln(L) + 2k, \tag{10}$$

where k is the number of estimated parameters in the model and L is the maximum value of the likelihood function for the model.

3.3.2 Bayesian information criterion

Schwarz [30] developed the Bayesian information criterion (BIC) as follows:

$$BIC = -2\ ln(L) + ln(T)\, k. \tag{11}$$

4. EMPIRICAL STUDY

This section is divided into five parts. In the first part, we will explain the data source. In the second part, we will display descriptive statistics for the variables. In part three, we will display the application for non-robust (OLS ARDL). Part four includes the results of new ARDL models by using the three robust estimation methods. Finally, we will compare the results of robust and non-robust models.

4.1 Data Source

The CO_2 data for Egypt and its economic determinants (Manufacturing, Trade, Urban population, and Land under cereal production) from 1990 to 2021 were obtained from the World Bank (https://data.worldbank.org/) and CO_2 data from the International Energy Agency (IEA) (https://www.iea.org/data-and-statistics/data-sets).

4.2 Descriptive Statistics

We divide this section into three parts: first, some descriptive measures of data. The correlation matrix is second. Third, test the stationarity of each variable.

4.2.1 Descriptive measures

The disregard for environmental conditions and ongoing exploitation of the environment can lead to environmental damage. One of the consequences of such harm is climate change, triggered by the release of greenhouse gases [31]. Greenhouse gas emissions, including CO_2, N2O, and CH4, contribute to the greenhouse gas effect. Among these gases, CO_2 emissions play the most significant role in intensifying the greenhouse effect, ultimately contributing to climate change [32].

Energy is a vital component of daily living, and production, and one of the main drivers of economic expansion. The relationship between the two has a direct bearing on the energy strategy of an area [33].

Table 2: Some descriptive statistics of the used variables crossing the period from 1990 to 2021.

Variable	**CO_2 emissions**	**Manufacturing**	**Trade**	**Urban population**	**Land under cereal production**
Abbreviation	**CO_2**	**MANUF**	**TRADE**	**URBAN**	**LAND**
Minimum	76.63	15.36992	29.85697	42.658	2283426
Mean	143.3928	16.51075	47.26851	42.91828	2889629
Median	148.581	16.36567	45.58333	42.891	2836634
Standard Deviation	49.76894	0.777708	10.77199	0.193734	318617.3
Maximum	218.557	18.49675	71.68063	43.478	3623430

Table 2 presents some descriptive statistics (means, minimum values, medians, maximum values, and standard deviations) for CO_2 Emissions, Manufacturing, Trade, Urban population, and Land under cereal production. Over the sample period, we note that the minimum CO_2 emissions are 76.63. This was at the beginning of the nineties century when the economy had not developed in Egypt, the stage of economic openness had not begun, and environmental pollution was not in its form today. It reached its maximum level, which is 218.557, The average value was 148.581. As for Manufacturing, its minimum was 15.36992, Manufacturing reached its peak in 2002 as a result of economic openness, and encouragement of industry and investment to reach a value of 18.49675, The average value was 16.36567. As for trade, its minimum was 29.85697 in 2021 due to the outbreak of the corona epidemic due to foreign wars. This did not happen before the events of 2011, as trade flourished and peaked in 2008 to reach a value of 71.68063, The average value was 45.58333. This is the case with the Urban population. It reached its minimum in the ninety's century during the beginning of the industrial revolution in Egypt at a value of 42.658 then it returned and achieved its highest value at 43.478, and the average value was 42.91828. This is the case with the Land under cereal production. It reached its minimum at a value of 2283426, then returned and achieved its highest value at 3623430. The average value was 2836634. The data's time series appears to be consistent, as shown by the reduced standard deviation findings.

Table 3 demonstrates how to determine whether the data is normal by using the Jarque-Bera (JB) [34] test. The null (H_0) and alternative (H_1) hypotheses of the JB test are H_0: the data is normal distribution versus H_1: the data is not normal distribution. The main conclusion drawn from these statistical studies is that every variable has a normal distribution, as shown by the p-value in the Jarque-Bera test being greater than 0.05. Where the p-value of CO_2 variable is 0.18973 more than 0.05, the p-value of Manufacturing variable is 0.220698 more than 0.05, the p-value of Trade variable is 0.18973 more than 0.05, the p-value of Urban variable is 0.18973 more than 0.05, and the p-value of Land under cereal production variable is 0.18973 more than 0.05. Therefore, in this case, accepting the null hypothesis is considered unattainable, which asserts that the data has a normal distribution [34].

Table 3: JB test of each variable

Variable	Jarque-Bera statistic	p-value
CO_2	3.324908	0.189673
MANUF	3.021924	0.220698
TRADE	1.451764	0.483898
URBAN	5.624542	0.060068
LAND	0.968567	0.616138

4.2.2 Correlation Matrix

Table 4 presents the correlation matrix of the independent variables. It seems that every pair of independent variables correlates with one another, with a very low correlation between the Land under cereal production and Manufacturing equal to (-0.1283), a very low correlation between the Land under cereal production and the Urban population equal to (-0.18057), a low correlation between the Land under cereal production and Trade equal to (-0.39406), a low correlation between Manufacturing and Trade equal to (-0.43988), a low correlation between Manufacturing and Urban equal to (-0.41375). There is a moderate relationship between Trade and Urban equal to (0.665445). Since all the correlation coefficients are less than 0.8. The prior correlation data indicates that the independent variables do not have a multicollinearity issue [35, 36].

Table 4: Correlation matrix of the independent variables

Variable	MANUF	TRADE	URBAN	LAND
MANUF	1			
TRADE	-0.43988	1		
URBAN	-0.41375	0.665445	1	
LAND	-0.1283	-0.39406	-0.18057	1

4.2.3 Stationarity test

The cointegration approach is not appropriate when working with integrated variables of different orders, such as the I(1) and I(0) series, as Pesaran and Shin [22] was showed. In certain situations, however, the ARDL cointegration process can be applied. As said, Johansen and Juselius [37] showed that even while pre-testing for unit roots is not necessary, it is essential first to check the stationary condition of every chain to keep the ARDL model from collapsing when there is an integrated stochastic trend of I(2). A time series is said to be stationary if its meaning, variance, and structure remain constant during the series. In contrast, unit roots or structural breakdowns identify nonstationary time series as stochastic processes. First, non-stationarity is mostly caused by unit roots. If a unit root is present, the time series under study is non-stationary; if it is absent, the time series is stationary. The unit root in the time series stationarity test was first presented by Dickey and Fuller [38] in 1979. The reasoning behind the unit root test is that a non-stationary series (X) has d unit roots at its level and must be integrated of order d, represented as I (d) if it takes d times to become stationary. White noise in the disturbance term is assumed for the DF test. Autocorrelation in the dependent variable will thus lead to autocorrelation in the error term. Therefore, the DF test is no longer valid. By taking p lag values into account, Dickey and Fuller [39] developed the DF test to improve the ADF. Both the DF and ADF tests use the null hypothesis and crucial values table.

The null (H_0) and alternative (H_1) hypotheses of the ADF test may be stated as follows: H_0 assumes that the series being tested has a unit root, whereas H_1 assumes that the series does not have a unit root.

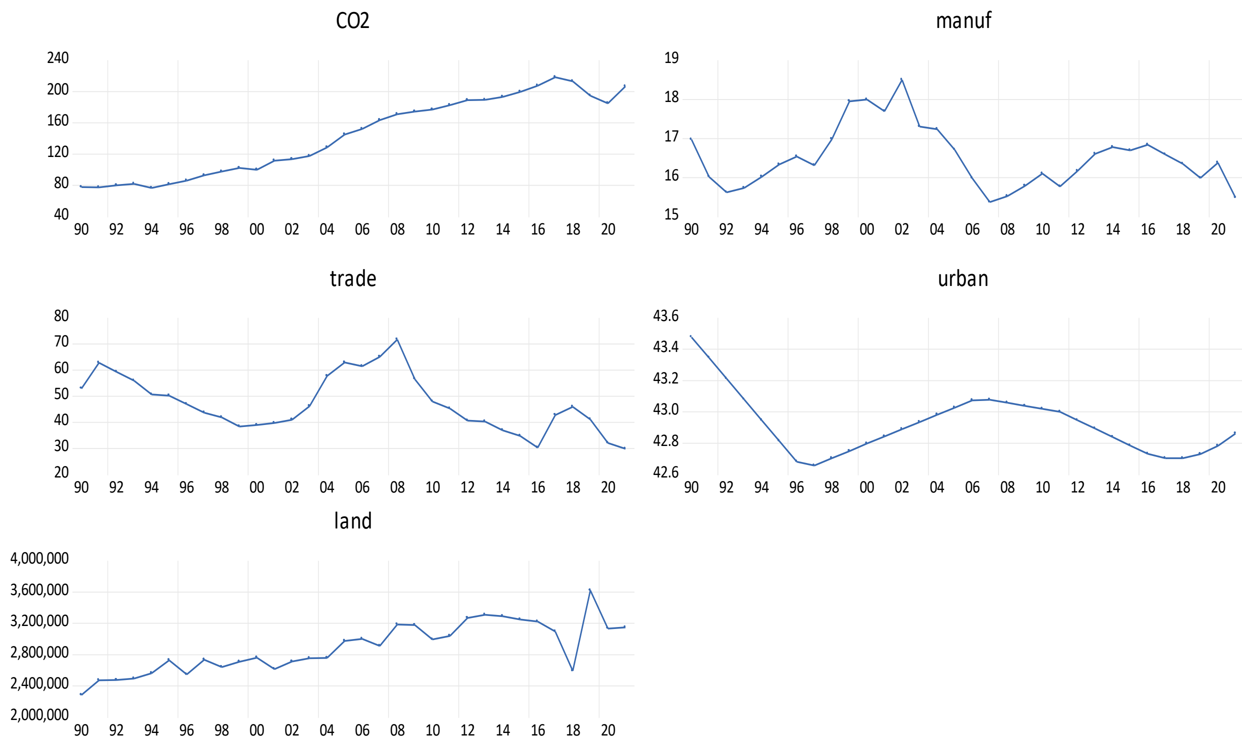

Figure 1: Time series graphs of the variables over the period from 1990 to 2021.

Table 5 presents the results of the ADF test for each variable. Table 5 shows that the stationarity variables after taking the first difference are CO_2 emissions, Manufacturing, and Trade (i.e., these variables are integrated of order 1). While the stationarity variables at the level are Urban population and Land under cereal production (i.e., these variables are integrated of level). Figure 1 displays the time series graphs depicting the variables over the period from 1990 to 2021. Figure 2 displays the time series graphs depicting the variables over the period from 1990 to 2021 after taking differences.

Table 5: Results of the ADF test

Series	CO_2	MANUF	TRADE	URBAN	LAND
Integrated order	I (1)	I (1)	I (1)	I (0)	I (0)

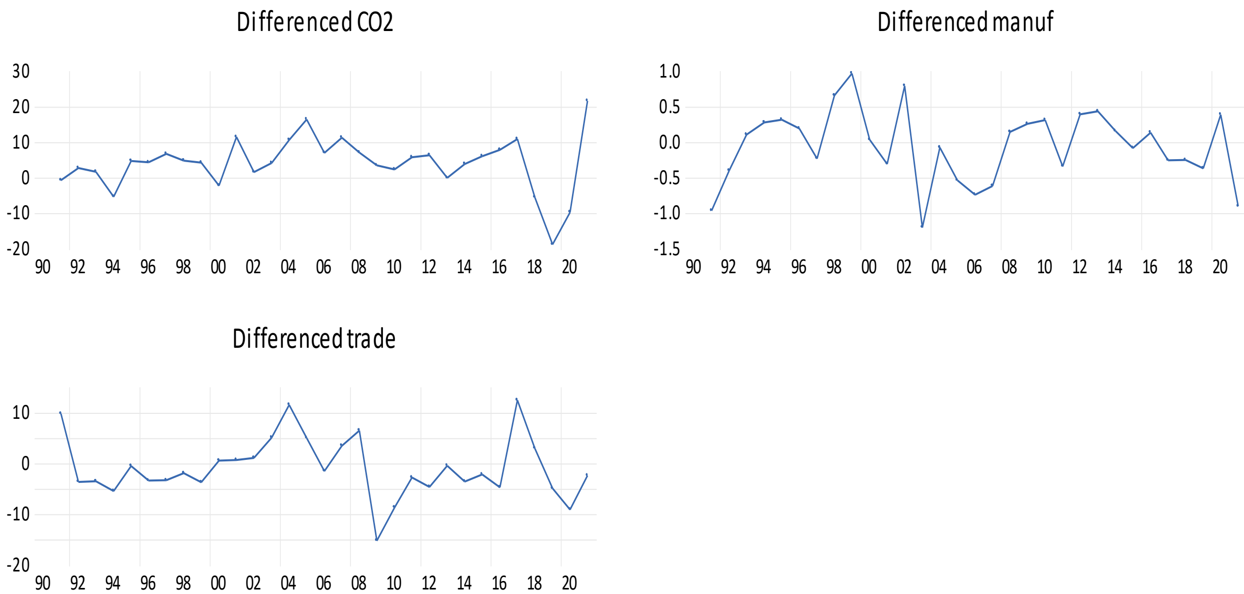

Figure 2: Time series graphs of the variables after taking the first difference.

4.3 Non-Robust (OLS ARDL) Model

We have divided this section into four parts: First, model selection. Second, checking stability. Third, checking serial correlation and heteroscedasticity. Fourth, checking normality.

4.3.1 Model Selection

The R software was used to estimate the ARDL model with automated lag selection. The least AIC was used to pick the ARDL (1,0,2,1,3) model.

Table 6 shows the best model selected by the lowest value of AIC, where it is 5.813683 for model ARDL (1,0,2,1,3). This means that we will take up to Lag 1 for the dependent variable (CO_2), take up to Lag 0 for the variable (Manufacturing), take up to lag 2 for the variable (Trade), take up to lag 1 for the variable (Urban population), and take up to lag 3 for the variable (Land under cereal production).

Table 6: Results of the Akaike information criterion for different models

Model	**Lag order of each variable**					**AIC**
	CO_2	**MANUF**	**TRADE**	**URBAN**	**LAND**	
1	1	0	2	1	3	5.813683
2	2	0	2	1	2	5.83651
3	2	0	2	1	3	5.842763
4	1	1	2	1	3	5.853746
5	1	0	2	1	2	5.855675
6	1	0	2	0	2	5.86006
7	1	3	3	0	2	5.862687
8	1	3	2	1	3	5.873748
9	2	3	3	0	2	5.876289
10	3	0	2	1	3	5.876748
11	1	0	2	0	3	5.879126
12	1	0	2	2	3	5.881281
13	1	0	3	1	3	5.881631
14	2	0	3	1	2	5.887634
15	2	1	2	1	3	5.890482
16	2	0	2	0	2	5.892347
17	1	3	2	0	3	5.895438
18	3	0	2	1	2	5.895509
19	2	0	2	2	2	5.896536
20	2	1	2	1	2	5.900239

Table 7 demonstrates that the delays in certain macroeconomic variables have notable impacts on CO_2 levels. It is imperative to consider the initial delay of CO_2 when projecting future CO_2 levels. Manufacturing has a significant effect. Additionally, trade and the second lag have a significant effect on CO_2. The urban population has a significant effect on CO_2. The first and second lags in the land under cereal production have significant effects on CO_2. in addition to the insignificance of the first lag of trade. In addition to the insignificance of the first lag of the Urban population, the land under cereal production, and the three-lag insignificance of CO_2. The value of adjusted R-squared is 0.9935, which indicates that the independent variables and the lags of the independent variables and the dependent variable were able to explain 99% of the changes that occurred in the dependent variable CO_2 and the rest 1% were explained by other variables and factors outside the model. The value of the AIC is 170.5968.

Table 7: Results of the ARDL (1,0,2,1,3) model short run

Variable	**Coefficient**	**Standard error**	**t-statistic**	**p-value**
Intercept	-882.8376	395.1333	-2.234278	0.0392*
CO_2 (-1)	0.556279	0.07548	7.369903	0.0001***
MANUF	-4.715001	1.87287	-2.517527	0.0221*
TRADE	0.367192	0.168401	2.18046	0.0436*
TRADE (-1)	-0.162341	0.229681	-0.706811	0.4893
TRADE (-2)	-0.485771	0.194412	-2.498672	0.023*
URBAN	41.5693	16.35409	2.541829	0.0211*
URBAN (-1)	-22.1402	14.15753	-1.563846	0.1363
LAND	5.99E-06	4.45E-06	1.345772	0.1961
LAND (-1)	2.42E-05	4.53E-06	5.337161	0.0001***
LAND (-2)	3.25E-05	4.95E-06	6.565521	0.0001***
LAND (-3)	1.02E-05	7.25E-06	1.412412	0.1759
Residual standard error: 3.823		AIC = 170.5968, BIC = 188.3716		
Multiple R-squared = 0.996		Adjusted R-squared = 0.9935		
F-statistic = 389, p-value = 2.2e-16***				

Note: *** significant at 0.001 and * significant at 0.05.

One statistical test used in econometrics to ascertain if two or more variables in a model have a long-term connection is the F-bounds test. To determine whether cointegration which denotes a shared stochastic trend between the variables is present in time series data analysis, this test is essential. By analyzing the significance levels of the test results, researchers may determine whether a solid association exists that will hold up over time. As a result, the F-bounds test is essential for comprehending the relationships and dynamics between economic variables and for gaining insights into the system's long-term behavior. The null (H_0) and alternative (H_1) hypotheses of the F-bounds test are H_0: no long-run relationship versus H_1: long-run relationship.

Regarding the F statistic's value, there are three different scenarios. In the first instance, we may infer with confidence that there is no long-term relationship if the value of the F statistic is smaller than I (0) and the H_0 is not rejected. In the second scenario, the H_0 is rejected and it is suggested that a long-run relationship exists if the value of the F statistic is greater than I (1). Lastly, in the last scenario, it becomes very challenging to determine with certainty whether a long-term association exists or not if the value of the F statistic is between two defined boundaries.

Table 8 shows that there is a long-run relationship, where the value of the F-statistic is 9.914781which is greater than the upper bounds (3.09, 3.49, 3.87, and 4.37), concluding that there is a long-run relationship at 1% significance level.

Table 8: F-bounds test of ARDL (1,0,2,1,3) model

Test Statistic	Value	Significant level	I(0)	I(1)
F-statistic	**9.914781**	10%	2.2	3.09
K	4	5%	2.56	3.49
		2.5%	2.88	3.87
		1%	3.29	**4.37**

Table 9 shows the results of the long-term equilibrium test of the ARDL correlation test approach proving that, over time, there is a strong and significant effect on CO_2 by the external variables, as there is a negative causal relationship between the Manufacturing and CO_2, as whenever the Manufacturing decreases by one unit, CO_2 increases by a value of 10.626. In addition, there is an effect of the Urban on CO_2, as there is a positive causal relationship between the Urban population and CO_2, whereby whenever the Urban population increases by one unit, CO_2 increases by a value of 43.78672. In addition, there is an effect of Land under cereal production on CO_2, as there is a positive causal relationship between under-cereal production and CO_2, as whenever under-cereal production increases by one unit, CO_2 increases by 0.000164.

Table 9: Results of the ARDL (1,0,2,1,3) model in the long run

Variable	Coefficient	Standard error	t-statistic	p-value
Intercept	-1989.62	863.3936	-2.30442	0.0341*
MANUF	-10.626	3.550353	-2.99295	0.0082**
Trade	-0.6331	0.395418	-1.60109	0.1278
Urban	43.78672	20.84634	2.100451	0.0509
Land	0.000164	9.21E-06	17.83456	0.0001***

Note: *** significant at 0.001, ** significant at 0.01, and * significant at 0.05.

The Error Correction Term (ECT) offers valuable information about how well any prevailing disequilibrium is corrected. It measures how well any imbalances from the previous era are being addressed at this moment in time. A divergence of a continuous departure from the intended equilibrium state is shown by a positive coefficient linked to the ECT. On the other hand, a negative coefficient denotes convergence and a slow approach to equilibrium. When the estimated value of the ECT is 1, it means that all adjustments are made in the same time frame, indicating a real-time and comprehensive correction. On the other hand, in the event where the predicted ECT value is 0.5, it means that only half of the required adjustment occurs in each succeeding period or year. Finally, a value of 0 for the ECT indicates that there has been no modification, which makes the claim of a long-term association absurd.

Table 10 displays the ECT. The ECT in our instance is negative, indicating a high degree of statistical significance. This is an important point to note. The results of this study imply that convergence occurs in the studied system. Consequently, we may infer that the year accounts for a noteworthy 44% of the required transition from the long-run to short-run equilibrium state. This suggests that the two years are spent during the critical transition period.

Table 10: Results of the error correction model

Variable	**Coefficient**	**Standard error**	**t-statistic**	**p-value**
D(TRADE)	0.367192	0.123726	2.967779	0.0086**
D(TRADE(-1))	0.485771	0.133765	3.631515	0.0021**
D(URBAN)	41.5693	11.65124	3.567801	0.002**
D(LAND)	5.99E-06	3.22E-06	1.862249	0.0799
D(LAND(-1))	-4.27E-05	7.97E-06	-5.355308	0.0001***
D(LAND(-2))	-1.02E-05	5.93E-06	-1.72571	0.1025
D(TRADE)	0.367192	0.123726	2.967779	0.0086**
ECT	-0.443721	0.050572	-8.77413	0.0001***

Note: *** significant at 0.001, ** significant at 0.01, and * significant at 0.05.

4.3.2 Checking Stability

A critical step in the estimating process is confirming the model's suitability before issuing predictions. Consequently, the residual performance was determined, and the stability of the model was confirmed.

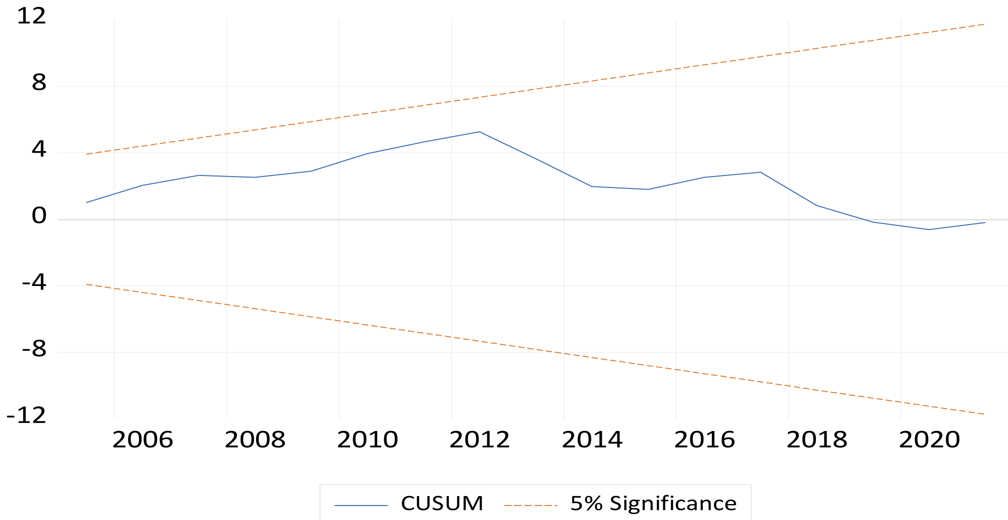

Figure 3: CUSUM stability test of ARDL (1,0,2,1,3) model

Figures 3 and 4 show how the stability and accuracy of the derived model are evaluated using the cumulative sum of the recursive residuals CUSUM and the CUSUM of squares. The estimated model satisfies the sample period's stability requirements, as demonstrated by the CUSUM and CUSUM of squares, both inside the 5% significant lines.

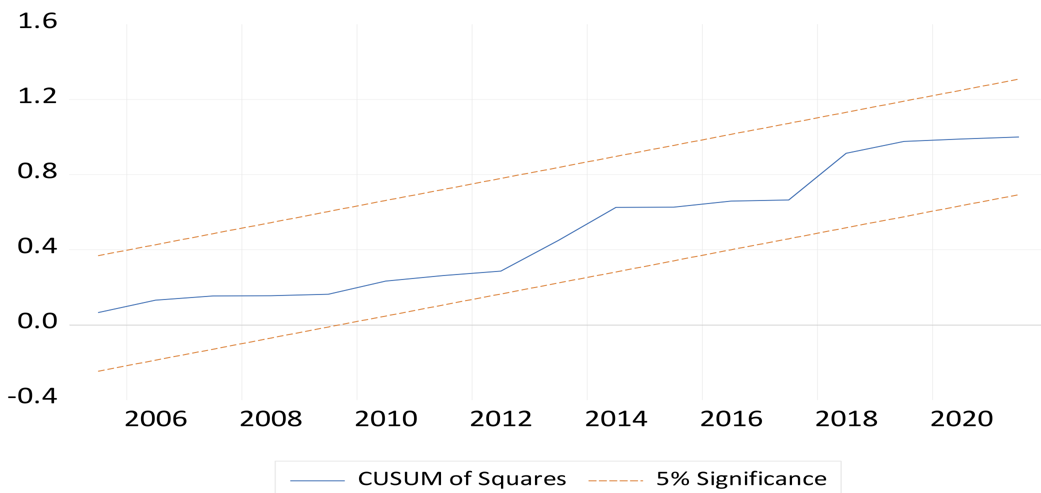

Figure 4: CUSUM of squares stability test of ARDL (1,0,2,1,3) model

4.3.3 Checking serial correlation and heteroscedasticity

The White and Breusch-Godfrey-LM tests have been used to evaluate the heteroskedasticity and serial-correlation of the error. The null (H_0) and alternative (H_1) hypotheses of the Breusch-Godfrey-LM test are H_0: there is no serial-correlation, versus H_1: there is a serial correlation problem. While in heteroskedasticity (White) test, the null (H_0) and alternative (H_1) hypotheses are H_0: there is no heteroskedasticity, versus H_1: there is a heteroskedasticity problem.

Table 11 demonstrates the results presented in that the estimated model's residual term does not exhibit any serial-correlation. This suggests that the H_0, which holds that there is no serial correlation, is not rejected at the 0.05 significance level because the LM test's p-value of 0.103 is greater than 0.05. Moreover, Table 11 shows that neither heteroscedasticity nor constant variance exists in the residuals as we do not reject the H_0 that there is no heteroscedasticity at level 0.05 due to the p-value of the White test (0.553) being greater than 0.05.

Table 11: Serial-correlation and heteroskedasticity tests for the residuals

Test	**Chi-squared value**	**p-value**
Breusch-Godfrey-LM	2.655432	0.103
White test	0.90798	0.553

4.3.4 Checking Normality

Table 12 shows the JB test indicates the residuals are normally distributed, because the p-value of the JB test (0.880396) is more than 0.05.

Table 12: JB normality test for the residuals

Test	**JB statistic**	**p-value**
JB test	0.254768	0.880396

4.4 Robust ARDL Models

This section is divided into four parts. In the first part, we will explain the outliers in the data. In the second part, we will explain m-Huber estimation. In part three, we will explain our estimation. Part four explains the S estimation.

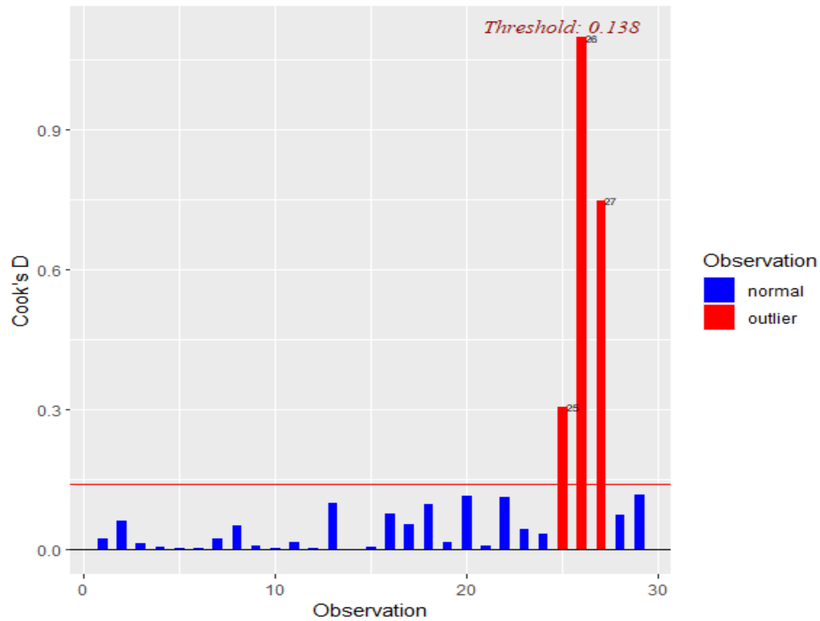

Figure 5: Cook’s bar plot of the residuals of ARDL (1,0,2,1,3) model

4.4.1 Testing outliers in the data

Figures 5 and 6 demonstrate that outliers are present in the OLS residuals of ARDL (1,0,2,1,3) model. Outliers are shown by red colored bars in figure 5 and red and green circles in figure 6.

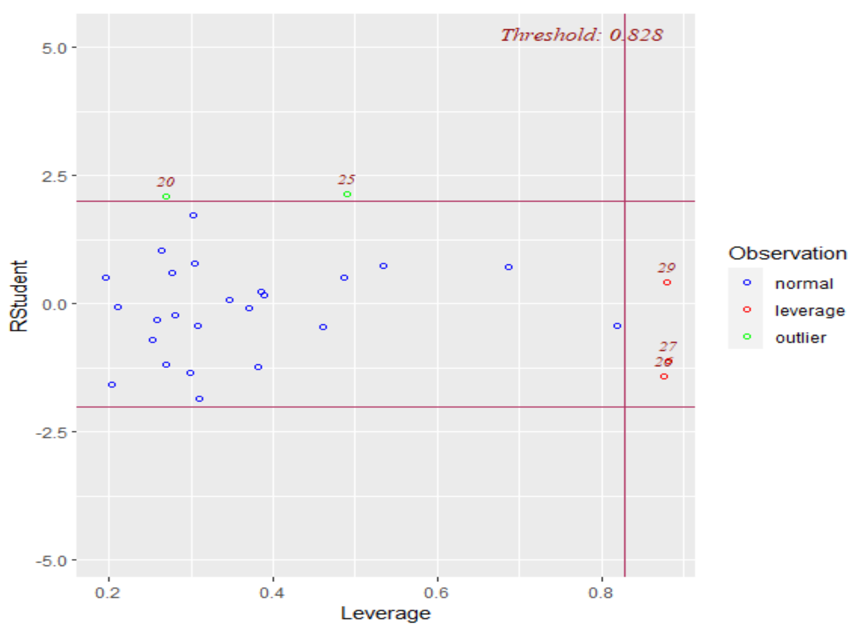

Figure 6: Outlier and leverage diagnostics of the residuals of ARDL (1,0,2,1,3) model

We see that there are outliers in the data, which the least squares approach is unable to manage. Consequently, a robust estimation approach will be employed to estimate the parameters. However, we should be aware that the least squares method of analysis is used to estimate the parameters of the ARDL model before utilizing the three robust estimation techniques. Consequently, unless we are certain that the lags have been taken accurately and that the estimation technique is right, we will manually take the lags of the variables under consideration and re-estimate the model. Since Table 7 displays identical findings, we can establish that estimating the delays manually is accurate. Moving forward, we will employ three robust estimation techniques.

4.4.2 M-Huber Estimation

Table 13 demonstrates that the delays in certain macroeconomic variables have notable impacts on CO_2 levels, in addition to the insignificance of the first lag of Trade. the Land under cereal production, and the third lag insignificance of CO_2. In addition to the insignificance of the first lag of the Urban population.

Table 13: Results of the M-Huber ARDL (1,0,2,1,3) model

Variable	Estimate	Standard error	t-statistic	p-value
Intercept	-8.725584e+02	412.4623	-2.1155	0.0495*
CO_2 (-1)	5.570353e-01	0.0788	7.0699	0.00001***
MANUF	-4.688048	1.9550	-2.3980	0.0284*
TRADE	3.700194e-01	0.1758	2.1049	0.0504
TRADE (-1)	-1.626955e-01	0.2398	-0.6786	0.5065
TRADE (-2)	-4.836739e-01	0.2029	-2.3834	0.0291*
URBAN	4.142808e+01	17.0713	2.4268	0.0267*
URBAN (-1)	-2.225000e+01	14.7784	-1.5056	0.1505
LAND	5.961479e-06	4.648689E-06	1.2824	0.2167
LAND (-1)	2.423068e-05	4.73255E-06	5.1200	0.00001***
LAND (-2)	3.250263e-05	5.162097E-06	6.2964	0.00001***
LAND (-3)	1.009815e-05	7.567558E-06	1.3344	0.1998
Robust residual standard error = 4.504			AIC = 88.02342, BIC = 104.431	
Adjusted R-squared = 0.572545				

Note: *** significant at 0.001 and * significant at 0.05.

From Table 13, we found that a significant effect of the first lag of CO_2 on CO_2. Additionally, Trade and the second lag have a significant effect on CO_2. Manufacturing has a significant effect. Urban population has a significant effect on CO_2. The first and second lags in the Land under cereal production have significant effects on CO_2. The value of Adjusted R-squared is 0.572545 which indicates that the independent variables and the lags of the independent variables and the dependent variable were able to explain 57% of the changes that occurred in the dependent variable CO_2 and the remaining 43% were explained by other variables and factors, outside the model. The value of AIC equals 88.02342.

4.4.3 MM Estimation

Table 14 the study demonstrates that the delays in certain macroeconomic variables have notable impacts on CO_2 levels.

Table 14: Results of the MM ARDL (1,0,2,1,3) model

Variable	Estimate	Standard error	t-statistic	p-value
Intercept	-389.8127	393.419	-0.9908	0.3357
CO_2 (-1)	0.6962	0.0752	9.2644	0.00001***
MANUF	-4.0591	1.8647	-2.1768	0.0438*
TRADE	0.3450	0.1677	2.0574	0.0553
TRADE (-1)	0.1263	0.2287	0.5521	0.5880
TRADE (-2)	-0.6247	0.1936	-3.2275	0.0049**
URBAN	29.1724	16.2831	1.7916	0.0909
URBAN (-1)	-20.5643	14.0961	-1.4589	0.1628
LAND	-1.006926e-05	4.434039E-06	-2.2709	0.0365*
LAND (-1)	2.120492e-05	4.744095E-06	4.6975	0.0002***
LAND (-2)	3.371480e-05	4.923738E-06	6.8474	0.00001***
LAND (-3)	5.495241e-06	7.218232E-06	0.7613	0.4567
Robust residual standard error = 2.8231			AIC = 111.47, BIC = 128.6778	
Adjusted R-squared = 0.9844742				

Note: *** significant at 0.001, ** significant at 0.01, and * significant at 0.05.

We found a significant effect of the first lag of CO_2 on CO_2. Manufacturing has a significant effect. Additionally, Trade and the second lag have a significant effect on CO_2. Urban population has a significant effect on CO_2. Land under cereal production, first and second lags in the Land under cereal production have significant effects on CO_2, in addition to the insignificance of the first lag of Trade. In addition to the insignificance of the first lag of the Urban population. The third lag in Land under cereal production is insignificance of CO_2. The value of Adjusted R-squared is 0.9844742 which indicates that the independent variables and the lags of the independent variables and the dependent variable were able to explain 99% of the changes that occurred in the dependent variable CO_2 and the remaining 2% were explained by other variables and factors, outside the model. The value of AIC equal 111.47.

4.4.4 S Estimation

Table 15 the study demonstrates that the delays in certain macroeconomic variables have notable impacts on CO_2 levels.

Table 15: Results of the S ARDL (1,0,2,1,3) model

Variable	Estimate	Standard error	t-statistic	p-value
Intercept	3.126e+02	1.824e+02	1.714	0.104688
CO_2 (-1)	8.213e-01	4.265e-02	19.257	5.55e-13 ***
MANUF	-4.074e+00	6.981e-01	-5.837	1.98e-05 ***
TRADE	2.406e-01	6.966e-02	3.454	0.003032 **
TRADE (-1)	6.180e-01	1.367e-01	4.521	0.000302 ***
TRADE (-2)	-8.014e-01	9.817e-02	-8.164	2.77e-07 ***
URBAN	4.261e+01	6.655	6.402	6.56e-06 ***
URBAN (-1)	-4.973e+01	6.660	-7.467	9.21e-07 ***
LAND	-3.448e-05	5.099e-06	-6.762	3.32e-06 ***
LAND (-1)	1.740e-05	1.983e-06	8.776	1.01e-07 ***
LAND (-2)	3.985e-05	2.044e-06	19.493	4.55e-13 ***
LAND (-3)	8.700e-06	2.942e-06	2.957	0.008830 **
Robust residual standard error = 2.8230			AIC = 77.298, BIC = 93.70598	
Adjusted R-squared = 0.9994				

Note: ** significant at 0.01 and * significant at 0.05.

We found all variables are significant. A significant effect on the first lag of CO2 on CO_2. Manufacturing has a significant effect. Additionally, Trade, first, and second lags have a significant effect on CO_2. The Urban population and the first lag have a significant effect on CO_2. Land under cereal production, first, second, and third lags in the Land under cereal production have significant effects on CO_2. The value of Adjusted R-squared is 0.9996 which indicates that the independent variables and the lags of the independent variables and the dependent variable were able to explain 99.96% of the changes that occurred in the dependent variable CO_2 and the rest 0.14% were explained by other variables and factors, outside the model. The value of AIC equal 77.298.

4.5 Compare Non-Robust and Robust Models

Table 16 displays three criteria to evaluate and compare the four estimated ARDL models: the standard error (SE) of the model, AIC, and BIC. Table 16 reveals that the three robust models provide reduced SE, AIC, and BIC values in comparison to the non-robust model. It is evident that, in the presence of outliers, robust models are more suitable for the data than the OLS-ARDL model. In our application, the most robust model is the S ARDL because it has the minimum values of SE, AIC, and BIC.

Table 16: Summary of the four estimated ARDL models

Criterion	Non-Robust Model	Robust Models		
	OLS ARDL	M-Huber ARDL	MM ARDL	S ARDL
SE	3.823442	4.504	2.8231	**2.8230**
AIC	170.5968	88.02342	111.47	**77.2980**
BIC	188.3716	104.431	128.6778	**93.7060**

5. CONCLUSIONS

This study estimated one of the dynamic causal relationships between CO_2, Manufacturing, Trade, Urban population, and Land under cereal production in Egypt from 1990 to 2021. It also proposed a new method for estimating ARDL using the M-Huber, MM, and S robust methods and compared OLS and robust methods to describe the best estimate. First OLS ARDL (non-robust model) explains the relation between CO_2 and the four variables (Manufacturing, Trade, Urban population, and Land under cereal production), and the best model for the dataset is ARDL (1,0,2,1,3). The Manufacturing and Trade variables have a significant negative impact on CO_2. At the same time, the Urban and Land variables are positively and significantly related to CO_2 in the long-term. Moreover, there are significant effects of the lags of Trade and Land variables on CO_2. We also concluded that 44% of short- to long-run adjustments occur yearly. In this paper, other methods are proposed for estimating the ARDL model other than the usual OLS method. This paper discussed three different methods for robust estimation, namely M-Huber, MM, and S. It became clear that the three robust methods are better than the non-robust (OLS) method, and that the best method among them for our dataset is the S estimation method because it has the lowest values of AIC and BIC. It is evident that, in the presence of outliers, the robust methods outperform the conventional OLS method of estimation used in the ARDL model. In future work, we can do multiple research projects by adding new variables, changing countries, proposing a new robust estimation method as in [40], or using the spatial regression approach to study the CO_2 emissions in Egypt as in [41].

Conflict of Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] T.S. Adebayo, D. Beton Kalmaz, Determinants of CO2 emissions: empirical evidence from Egypt, Environ. Ecol. Stat. 28 (2021), 239–262. https://doi.org/10.1007/s10651-020-00482-0.

[2] M. Azam, A.Q. Khan, H.B. Abdullah, et al. The impact of CO2 emissions on economic growth: evidence from selected higher CO2 emissions economies, Environ. Sci. Pollut. Res. 23 (2015), 6376–6389. https://doi.org/10.1007/s11356-015-5817-4.

[3] N.T.K. Nguyen, M.B. Le, Co2 emissions and economic growth in Vietnam: An ARDL bound testing approach, Asian J. Econ. Model. 6 (2018), 47–55. https://doi.org/10.18488/journal.8.2018.61.47.55.

[4] A. El Ouahrani, J.M. Mesa, A. Merzouki, Anthropogenic CO2 emissions from fossil fuels, Int. J. Climate Change Strat. Manage. 3 (2011), 16–28. https://doi.org/10.1108/17568691111107925.

[5] H.A. Bekhet, N.S. Othman, Impact of urbanization growth on Malaysia CO2 emissions: Evidence from the dynamic relationship, J. Cleaner Product. 154 (2017), 374–388. https://doi.org/10.1016/j.jclepro.2017.03.174.

[6] J. Cheng, B. Li, B. Gong, et al. The optimal power structure of environmental protection responsibilities transfer in remanufacturing supply chain, J. Cleaner Product. 153 (2017), 558–569. https://doi.org/10.1016/j.jclepro.2016.02.097.

[7] Q. Zhu, X. Peng, The impacts of population change on carbon emissions in China during 1978–2008, Environ. Impact Assess. Rev. 36 (2012), 1–8. https://doi.org/10.1016/j.eiar.2012.03.003.

[8] A.K.M.N. Hossain, S. Hasanuzzaman, The Impact of energy consumption, urbanization, financial development, and trade openness on the environment in Bangladesh: an ARDL bound test approach, Shahjalal University of Science & Technology, 2012.

[9] I. Rayhan, K. Akter, M.S. Islam, et al. Impact of urbanization and energy consumption on CO_2 emissions in Bangladesh: an ARDL bounds test approach, Int. J. Sci. Eng. Res. 9 (2018), 838-843.

[10] Danish, M.A. Baloch, S. Suad, Modeling the impact of transport energy consumption on CO2 emission in Pakistan: Evidence from ARDL approach, Environ. Sci. Pollut. Res. 25 (2018), 9461–9473. https://doi.org/10.1007/s11356-018-1230-0.

[11] S. Ali, L. Ying, T. Shah, et al. Analysis of the nexus of CO2 emissions, economic growth, land under cereal crops and agriculture value-added in Pakistan using an ARDL approach, Energies. 12 (2019), 4590. https://doi.org/10.3390/en12234590.

[12] D. Singh, Foreign direct investment and local interpretable model-agnostic explanations: a rational framework for FDI decision making, J. Econ. Finance Admin. Sci. 29 (2024), 98-120. https://doi.org/10.1108/jefas-05-2021-0069.

[13] J. Jiang, S. Zhu, W. Wang, Carbon emissions, economic growth, urbanization, and foreign trade in China: Empirical evidence from ARDL models, Sustainability. 14 (2022), 9396. https://doi.org/10.3390/su14159396.

[14] A. Priyanto, A. Chanthavy, Representation of economic growth in Cambodia with the ARDL approach, Tamansiswa Account. J. Int. 5 (2022), 8-13.

[15] G. Liu, F. Zhang, How do trade-offs between urban expansion and ecological construction influence CO2 emissions? New evidence from China, Ecol. Indicators 141 (2022), 109070. https://doi.org/10.1016/j.ecolind.2022.109070.

[16] N. Tabassum, S.U. Rahman, M. Zafar, et al. Institutional quality, employment, trade openness on environment (Co2) nexus from top Co2 producing countries; panel ARDL approach, Rev. Educ. Admin. Law. 6 (2023), 211–225. https://doi.org/10.47067/real.v6i2.325.

[17] S. Su, S. Li, Energy efficiency suppression and spatial spillover effect: a quasi-natural experiment based on China's environmental protection tax law, Environ. Develop. Sustain. (2023). https://doi.org/10.1007/s10668-023-04146-4.

[18] M.A.M. Sallam, M.R. Abonazel, A.M. Shafik, Studying the impact of macroeconomic variables on inflation rates in Egypt: an ARDL approach, Monten. J. Econ. 21 (2025), 79-94.

[19] M.R. Abonazel, F.A. Awwad, A.F. Lukman, et al. Long-run determinants of Nigerian inflation rate: ARDL bounds testing approach, WSEAS Trans. Bus. Econ. 18 (2021), 370-1379. https://doi.org/10.37394/23207.2021.18.126

[20] A.A. El-Sheikh, F.A. Alteer, M.R. Abonazel. Four imputation methods for handling missing values in the ARDL model: An application on Libyan FDI, J. Appl. Prob. Stat. 17 (2022), 029-047.

[21] R.C. Hill, W.E. Griffiths, G.C. Lim, Principles of econometrics, 4th ed, Wiley, 2011.

[22] M.H. Pesaran, Y. Shin, R.J. Smith, Bounds testing approaches to the analysis of level relationships, J. Appl. Econ. 16 (2001), 289–326. https://doi.org/10.1002/jae.616.

[23] J. Fox, R. Robust, An R and S-Plus companion to applied regression, SAGE Publications, Inc., 2002.

[24] P.J. Huber. Robust version of a location parameter, Ann. Math. Stat. 36 (1964), 1753-1758.

[25] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, et al. Robust statistics: The approach based on influence functions, Wiley, 2011.

[26] P. Rousseeuw, V. Yohai, Robust regression by means of S-estimators, in: J. Franke, W. Härdle, D. Martin (Eds.), Robust and Nonlinear Time Series Analysis, Springer US, New York, NY, 1984: pp. 256–272. https://doi.org/10.1007/978-1-4615-7821-5_15.

[27] F.M. Alghamdi, A.R. Kamel, M.S. Mustafa, et al. A statistical study for the impact of REMS and nuclear energy on carbon dioxide emissions reductions in G20 countries, J. Rad. Res. Appl. Sci. 17 (2024), 100993. https://doi.org/10.1016/j.jrras.2024.100993.

[28] V.J. Yohai. High breakdown-point and high efficiency robust estimates for regression. Ann. Stat. 15 (1987), 642-656. https://www.jstor.org/stable/2241331.

[29] H. Akaike. Information theory and an extension of the maximum likelihood principle, in: B.N. Petrov, F. Csaki (Eds.), Proceedings of the 2nd International Symposium on Information Theory, 267-281, 1973.

[30] G. Schwarz. Estimating the dimension of the model, Ann. Stat. 6 (1978), 461-464. https://doi.org/10.1214/aos/1176344136

[31] T.G. Cassarino, M. Barrett, Meeting UK heat demands in zero emission renewable energy systems using storage and interconnectors, Appl. Energy 306 (2022), 118051. https://doi.org/10.1016/j.apenergy.2021.118051.

[32] H. Jeon, CO2 emissions, renewable energy and economic growth in the US, Electr. J. 35 (2022), 107170. https://doi.org/10.1016/j.tej.2022.107170.

[33] S. Wang, Differences between energy consumption and regional economic growth under the energy environment, Energy Rep. 8 (2022), 10017–10024. https://doi.org/10.1016/j.egyr.2022.07.065.

[34] C.M. Jarque, A.K. Bera, Efficient tests for normality, homoscedasticity and serial independence of regression residuals, Econ. Lett. 6 (1980), 255–259. https://doi.org/10.1016/0165-1765(80)90024-5.

[35] M.R. Abonazel, A.R.R. Alzahrani, A.A. Saber, et al. Developing ridge estimators for the extended Poisson-Tweedie regression model: Method, simulation, and application, Sci. Afr. 23 (2024), e02006. https://doi.org/10.1016/j.sciaf.2023.e02006.

[36] M.M. Abdelwahab, M.R. Abonazel, A.T. Hammad, et al. Modified two-parameter Liu estimator for addressing multicollinearity in the Poisson regression model, Axioms 13 (2024), 46. https://doi.org/10.3390/axioms13010046.

[37] S. Johansen, K. Juselius. Maximum likelihood estimation and inference in cointegration-with application to the demand for money, Oxford Bull. Econ. Stat. 52 (1990), 169–210. https://doi.org/10.1111/j.1468-0084.1990.mp52002003.x

[38] D. Dickey, W. Fuller, Distribution of the estimators for autoregressive time series with a unit root, J. Amer. Stat. Assoc. 74 (1979), 427-431. https://doi.org/10.2307/2286348.

[39] D. Dickey, W. Fuller, Likelihood ratio tests for autoregressive time series with a unit root, Econometrica 49 (1981), 1057–1072. https://doi.org/10.2307/1912517.

[40] S.L. Alkhayyat, S. Abdel-Rahman, M.R. Abonazel, et al. On the impact of the COVID-19 pandemic on mental health in Egypt: Penalized regression approach, Commun. Math. Biol. Neurosci. 2024 (2024), 61. https://doi.org/10.28919/cmbn/8095

[41] M.M. Abdelwahab, O.A. Shalaby, H.E. Semary, et al. Driving factors of NOx emissions in China: Insights from spatial regression analysis, Atmosphere. 15 (2024), 793. https://doi.org/10.3390/atmos15070793.