Application of Panel Data Analysis and Modelling to Economic Data: A Case of Determinants of Economic Growth for SADC and Zimbabwe

by

Thomas Musora

A Thesis Presented to the

CHINHOYI UNIVERSITY OF TECHNOLOGY

In Partial Fulfilment of the Requirements for the Degree

of

Doctor of Philosophy

in

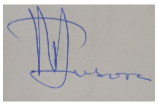
STATISTICS

Supervisor: Dr F. Matarise

Thursday 28th September, 2023

Declaration of Authorship

I, Thomas Musora, declare that this thesis titled, **Application of Panel Data Analysis and Modelling to Economic Data: A Case of Determinants of Economic Growth for SADC and Zimbabwe** is my own work and it has never been submitted before for any degree or examination in any other university; where I have quoted from the work of others, the source is always given.



. . . .

Signed: **Thursday 28th September, 2023**

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

. . . .

Signed: **Thursday 28th September, 2023**

Dr. F. Matarise,

Senior Lecturer

University of Zimbabwe, Zimbabwe

Abstract

In order to accurately forecast economic growth, it is important that growth determinants are identified. However Africa and Southern African Development Community (SADC) region in particular have not identified any determinants of economic growth that are peculiar to the SADC region. The aim of this research is to establish models that links economic development as measured by GDP to the determinants of economic growth for the Zimbabwean and SADC economies.In this study determinants of economic growth are gathered and evaluated for sixteen SADC countries for twenty two years (2000 to 2021), that dictates use the panel data analysis, whereas panel data may have group effects, time effects or both. Data is taken from various sources but mainly the World Bank website for different SADC countries contributing in the world economy. In this article, the comparison of ordinary least squares (OLS) model, fixed effects model (FEM), Machine learning (ML) and Random effects model (REM) for SADC nations panel data were carried out. F-test was used as a specification test to make a selection between OLS model and fixed effects model, The Breusch-Pagan test was used to choose between OLS and REM while the Hausman test was used as a specification test for FEM and REM. A fixed effects model with an adjusted R^2 value of 98% which is very plausible was realised to be the best model to handle the SADC community economic data.Imports, exports, external debt, international reserves, unemployment and labour force had positive impacts on the SADC community's economic growth. Foreign direct investment negatively influenced economic growth. Inflation, exchange rate and interest rate had no association with economic development for the SADC community. As for country effects, it was established that South Africa had a positive impact on gross domestic product (GDP), whereas all other SADC nations country effects negatively affected economic growth with the exception of Comoros and Seychelles, whose effects had no significant effects on economic growth. For Zimbabwe Deep learning modelling and the convectional model with log transformations were the best models and had almost the same predictive powers, Exports,Foreign direct investment and Labour force positively influenced economic growth.Inflation, external debt, interest rate and exchange rate had negative impacts on GDP. International reserves,imports and unemployment rate had no association with economic growth. Forecasts were done for Zimbabwe's GDP and it was realised that the GDP will increase for years 2022 to 2025. Based on these findings, the study recommends that policymakers in the SADC region prioritize areas such as imports, exports, external debt, international reserves, employment levels, and the labor force to stimulate sustainable economic growth. Furthermore, it is crucial to address challenges such as inconsistent power supply and integrate trade regulations to foster economic development in the region.

Acknowledgements

I would like to thank my supervisor, Dr F Matarise, from The Department of Mathematics and Computational Sciences at the University of Zimbabwe for her kindness, support and guidance throughout this thesis.

I also feel indebted to my colleagues in the School of Natural Sciences and Mathematics at Chinhoyi University of Thechnology. Our discourses were indispensable to the development of this thesis.

Contents

List of Figures

List of Tables

Abbreviations

SADC	**S**outhern **A**frican **D**evelopment **C**ommunity
OLS	**O**rdinary **L**east **S**qaures
FEM	**F**ixed **E**ffect **M**odel
REM	**R**andom **E**ffect **M**odel
ML	**M**achine **L**earning
GDP	**G**ross **D**omestic **P**roduct
FDI	**F**oreign **D**irect **I**nvestment
SSA	**S**ub-**S**aharan **A**frica
TSLS	**T**wo **S**tage **L**east **S**qaures
IIP	**I**ndex of **I**ndustial **P**roduction
ARDL	**A**uto-regressive **D**istributed **L**ag
PSID	**P**anel **S**tudy **I**ncome **D**ynamics
NLS	**N**ational **L**ongitudinal **S**urvey
BHPS	**B**ritish **H**ousehold **P**anel **S**urvey
GSOEP	**G**ermany **S**ocio-**E**conomic **P**anel
PSELL	**P**anel **S**ocio- **E**conomique **L**ixen zu **L**itzeburg

BLUE	**B**est **L**inear **U**nbiased**E**stimator
IID	**I**ndependent and **I**dentically **D**istributed
LSDV	**L**east **S**quares **D**ummy **V**ariable
FGLS	**F**easible **G**eneralized **L**east **S**quares
GLS	**G**eneralized **L**east **S**quares
EGLS	**E**stimated **G**eneralized **L**east **S**quares
LM	**L**agrange **M**ultiplier
DV	**D**ependent **V**ariable
SEE	**S**tandard **E**rror of the **E**stimates
MSE	**M**ean **S**quare **E**rror
SRMSE	**S**quare root of **M**ean **S**quared **E**rror
R and D	**R**esearch and **D**evelopment
SVM	**S**upport **V**ector **M**achine
VECM	**V**ector **E**rror **C**orrelation **M**odel
GNP	**G**ross **N**ational **P**roduct
CPS	**C**ost**P**er **S**ale
ARIMA	**A**utoregressive **I**ntegrated **M**oving **A**verage
IRF	**I**mpulse **R**esponse **F**unction
CVAR	**C**onditional **V**alue of **R**isk
ECM	**E**rror **C**orrelation **M**odel
LAC	**C**aribbean **L**atin **A**merican

USD	**U**nited **S**tates **D**ollar
ZWL	**Z**imbabwe **D**ollar

Dedicated to the Musora Family…

Chapter 1

Introduction

1.1 Introduction

This chapter comprises of the background of the research, aim of the study, statement of the problem, objectives of the study, significance of the study, limitations of the study and organization of the research.

1.2 Background of the Study

In recent years, several scholars, practitioners and students have met many situations where the concurrent analysis, of two or more correlated quality characteristics is necessary. Recent technological advancement in statistical analysis procedures has greatly improved the quantity and quality of available data. The use of computing hardware, such as electronic data collectors, facilitates the collection of data on a multitude of variables from all phases of research and inference. Analysis of several related variables of interest is collectively known as multivariate statistical analysis or panel data analysis or longitudinal data analysis or cross-sectional time-series data analysis and these are similar but applied differently in developing sections of statistical analysis. Panel data are also called cross-sectional time-series data or longitudinal data. These types of data have observations on the same units in several different time periods " Kennedy (2008).Longitudinal data sets have numerous variables, each of which having recurrent observations at diverse time phases, such that they may have time effects, group (individual) effects or both, which are examined by random effects or/and fixed effects models. The availability of more and more longitudinal data has resulted in many scholars, students and practitioners showing interests in longitudinal data modelling. Since these longitudinal data have more variability and allow to explore more issues than do cross-sectional or time-series data alone. Kennedy (2008), Baltagi and Chang (1994a) asserts that, Panel data gives more informative data, more variability, less collinearity among the variables, more degrees of freedom and more efficiency ". With a well-structured longitudinal data set, generally panel data modelling becomes more appealing and attractive as it provides methods of handling heterogeneity, gives more revealing data, is more able to reveal and quantify properties which are not noticeable in pure cross sectional or time series data, allows the construction of more complicated behavioural models and scrutinizes random and/or fixed effects in the panel data. Nevertheless, longitudinal data modelling is not as simple as it appears, it comes with some limitations that includes; data collection and design challenges, distortions of observation errors, selectivity challenges, being short in time-series dimensions and the challenges of cross-sectional dependence. A common error is that random and/or fixed effects models must always be used every time data are organized in the longitudinal data presentation. Complications of longitudinal data modelling, generally, comes from;

i.) longitudinal data themselves,

ii.) modelling procedure,

iii.) interpretation and reporting of the findings.

In fact other studies have analysed poorly structured longitudinal data, which are not panel in an actual econometric view, while other studies mechanically apply random or/and fixed effects models in hurry without deliberation of applicability of such models. Casual investigators frequently fail to interpret the findings appropriately and to report them correctly. The best method to use in which situation, is often a question which goes unanswered by most researchers and more often than not, they resort to the methods which they are comfortable to working with. Hence conclusions from such researches may not be the most appropriate. Thus this research aims to empirically analyse and compare OLS,Machine Learning(ML), FEM and REM when estimating economic growth for the Zimbabwean and SADC data so as to make informed inferences about such data. Economic development is the principal goal of macroeconomic guiding principles in any nation, exploration on the factors which increases or hinders economic progression has been one of the principal doctrines in the midst of empirical and theoretical growth investigators, but limited agreements has been achieved to date. Gross Domestic Product (GDP) is viewed as a key indicator of this economic growth Semuel and Nurina (2014).When GDP of a nation grows faster than the population, then it stipulates that gross domestic product per capita of that country is increasing and the way of life in that specific nation is also improving. GDP of a country is influenced by a number of variables i.e. inflation, interest rate, exchange rate, foreign direct investment, household consumption and so on. Inflation has several economic implications for any economy and most countries try to maintain a sustainable high economic growth with low inflation Ayyoub et al. (2011). Inflation increases the price level of services, merchandises and other entities hence brings into being economic challenges for any nation. This results in fall of purchasing power of currency; thus the worthiness of currency reduces simultaneously also. The decline in value of money and price level increases activated by inflation affects the development of any economy. High inflation also induces high interest rates and as interest rates, generally, drives contrary to GDP, hence the increase in interest rate leads to decrease in economic growth of a country and vice versa Kibria et al. (2014). Exports which aid economic advancement of any nation can be reduced due to the exchange rate instability as it diminishes trade by making the anticipated return from exports stochastic. Household consumption is also seen as a key indicator of economic health and stimulates the economic development vibrantly by motivating collective expenditure. Accordingly, it can be realised that diverse macroeconomic phenomenon inspires economic progression of a country. However, the direction and magnitude of the impact may differ according to specific economy. Hence, this research attempts to explore the influence of macroeconomic variables on economic development with a particular bias towards the comparison of performance of OLS,RE, Machine Learning (ML) and FE models for the SADC community economic data and identify the most appropriate model to describe the SADC and Zimbabwean economic data.

1.3 Statement of the Problem

Determinants for economic growth for the SADC region have been under-researched. There is limited empirical work that exclusively looked into factors that determine economic growth in SADC's developing economies. Currently most SADC nations are benchmarking their economic prosperity against international trade and level of FDI inflow. However the international trade mantra which promotes imports and exports ignores other key determinants of economic growth. The technological evolutions in the last few decades have changed the way nations and regions attempt to improve their economies. The model used in forecasting economic growth has a strong bearing on the moves and decisions to be made. Hence if a wrong model is used, it is likely that moves and decisions will be erroneous. Thus policy makers need to be equipped with the appropriate model to apply and when. Among others, this research will build models empirically in view of these challenges. The resulting models and recommendations will better equip planners of today and those of future generations with the most appropriate economic modelling methods when making decisions under uncertainty. Thus, this study has policy implications.

1.4 Aim

The research aims to establish models that links economic development as measured by GDP to the determinants of economic growth for the Zimbabwean and SADC economies.

1.5 Research Questions

1. What are the determinants of economic progression for the Zimbabwean and SADC economies?
2. Are there significant correlations between GDP and selected economic variables of Zimbabwean and SADC economies?
3. Which method is best, amongst OLS,Machine Learning(ML), fixed effects modl (FEM) and random effects model (REM) when estimating economic growth for the Zimbabwean and SADC data?
4. What will be the GDP for Zimbabwe in the next five years?

1.6 Objectives

The objectives of theresearch are:

1. To identify determinants of economic growth for the Zimbabwean and SADC nations.
2. To fit OLS, FE,ML and RE models to SADC community economic data and a Multivariate Model Zimbabwean economic data?
3. To Compare performance of OLS,RE, ML and FE models for the SADC community economic data and Identify the most appropriate model to describe the Zimbabwean economic data.

4. To forecast GDP for Zimbabwe for the next five years.

In an attempt to achieve this, we will fit an OLS, ML, fixed or random effects model to SADC nations panel data for determinants of economic development. To compare performance of OLS, FEM, ML and REM for the SADC economic data. The study represents a first attempt to model and to compare GDP estimates for SADC nations produced panel data. Comparing performance of OLS,RE, Machine Learning (ML) and FE models enables one to choose the most appropriate modeling procedure.

1.7 Significance of the study

This research is significant as:

1. It builds on the comparatively limited pragmatic writings on determinants of economic development in developing countries. Accordingly, this work attempts to help narrow the empirical texts gap for this often overlooked but important subject matter.

2. It could aid to recognize the determinants economic advancement in developing countries.

3. It would help policy in coming up with suitable public policies which would result in economic growth for the whole country. Thus, the study,s results would be highly significant in the creation and implementation of effective guidelines that will establish and uphold sustainable economic growth. Basically, this research would aid policy makers to identify the key macroeconomic determining factors of economic progression in developing countries and then put in place polices that will speed up economic development and thus minimises poverty.

4. It enhances the current understanding of determinants of economic progression association.

5. It would be a basis of facts for prospective investors. For instance, if this research recognizes foreign direct investments (FDI) as a key and significant determinant for economic development, it would be a pointer to prospective financiers, that the economic and political environment in developing countries is conducive and open for business. Hence prospective investors would opt to participate in the economy. This could aid in boosting economic progression since investors would descend with their capital, advanced equipment and expertise and the result might be, improved employment prospects in developing countries and thus an enhancement in the ways of life for the population.

6. The outcomes may possibly incite further enquiries relative to this research.

1.8 Limitations of the study

The main constraint of this study is to do with the quality of the data set to be used. As the data is secondary, its accurateness cannot be assured. This challenge is compounded by the intricacies faced when gathering the data in developing countries. According to Kholdy (1995), data compiled in most developing countries is inaccurate and may therefore bias the empirical results ". Additionally, the era of the research has numerous structural breaks such as diverse exchange rates regimes, economic guidelines, military administrations, etc. Consequently, any inferences and analysis derived from this write up takes due acknowledgment of such restrictions. Financial resources and additionally the time factor may not be omitted from the list of limitations to this research.

1.9 Organisation of the study

The research is structured into five chapters. The first chapter deliberates on the following: significance of the study, problem statement, background, objectives, organization and the limitations of the study. The second chapter embarks on a review of current and significant and writings on economic development that comprises of: determinants of economic advancement and rational review on possible bases of economic progression and also existing literature on panel data analysis. The third Chapter discusses the methodology used in the study, whereas chapter three outlines how the entire research would be done. Thus, it spells out the methodology employed in the study, the model specifications, justifications of the entities and the approximation processes. The fourth chapter presents the data analysis, interpretation and discussion of findings. Lastly, chapter five gives conclusions and recommendations of the main results, estimation method used. To conclude, this chapter, explanations for the pursuit of the sustainable growth goals by promoting findings from the study are given.

1.10 Chapter Summary

This chapter has outlined the motivation behind the study, highlighting the specific aim, objectives, and significance of the research.

Chapter 2

Literature Review

2.1 Introduction

This study aims to establish models that links economic development as measured by GDP to the determinants of economic growth for the Zimbabwean and SADC economies. By only investigating the causative factors and sources of economic development which may draw policy makers to embark on the appropriate estimation procedures, paths to attain sustainable, rapid, prosperity and broad-based economic progression in developing countries. This chapter comprises of existing and relevant writings on economic advancement (empirical and theoretical) both in the perspective of developing and developed nations mainly to date which tries to emphasise on the main determining factors of economic advancement. The evaluation is founded on: The perception of economic development, determinants of economic advancement and realistic literature review on the bases of economic progress in the emerging economies and review of literature on panel data analysis and estimation procedures.

2.2 Literature Review

2.2.1 Definitions of terms

Economic Growth

Baltagi and Chang (1994a) describes economic growth as an increase in real gross domestic product (GDP). That is, gross domestic product adjusted for inflation. Semuel and Nurina (2014) defines economic growth as a part of economic theory that explains the rate at which a countrys economy grows over time. It is generally regarded as the yearly proportion of development of the states key countrywide revenue accounting aggregates, for example the GNP (Gross National Product) or the GDP with suitable statistical modification to reduce the possibly deceptive special effects of inflation.

There are several ways of determining economic development in any nation. These consist of real output per-capita and progression in actual GDP. In this research GDP will be used as a determinant of economic advancement. This is because numerous other researchers have used it in their work as a dependent variable Chen and Feng (2000), Anaman (2006), Frimpong Magnus and Oteng-Abayie (2006), Bashar and Khan (2007), etc.

Real Gross Domestic Product (GDP)

GDP is the market value of goods and services produced within a selected geographic area (usually a nation) in a selected interval in time (often a year) Leamer (2008).This can also be referred to as, the financial value of all completed services and goods produced in any nation within a specific phase of time. GDP is adjusted for fluctuations in the price levels which embraces the replacements of obsolete and worn-out equipment and configurations in addition to new investments, it computes economic activities that happen in a particular nation. It evaluates how well an economy yields services and goods that individuals find useful. It is most frequently used determine of economic evolution of any nation. It can also weigh the relative impact of any industrial sector. It comprises of all public and private investments, government, consumption and outlays.

Physical capital

This includes any manufactured assets which are used in fabrication for instance buildings, vehicles and machinery. Alternatively, it can be referred to as any non-human asset made by humans and then used in production Shim et al. (1995). Sufficient capital is one of the key requirements of economic development on empirical and theoretical basis. Capital streams out of savings and savings out of income. Additional capital implies increased production and additional production implies inreased outputs and therefore, increased development. This is so since if saving rate is high, a higher proportion of output may be apportioned for investment which might result in increased rate of capital buildup and production growth, all other things being one and the same.

Labour Force

Refers to currently active population comprising of all persons who fulfil the requirements for inclusion among the unemployed or employed during specified period Shim et al. (1995). Basing on the traditional growths theorists, an escalation in labour force, that is measured by the proportion of entire population aged $15-64$ years, is anticipated to lead to an escalation in real GDP per capita (economic progression).

Foreign Direct Investment (FDI)

Denotes the long term involvement by nation X in nation Z. It usually involves participation in management, joint venture, transfer of technology and expertise Shim et al. (1995). FDI plays a pivotal role in enhancing economic development through increased production levels. FDI has been recognized as the most fundamental factor in increasing economic advancement and the living standards for embryonic economies.

Foreign Aid

Describes voluntary transfer of resources from one country to another, given at least partly with the objective of benefiting the receipts country Shim et al. (1995). An additional factor closely linked to foreign aid is foreign investment. Accordingly, it is anticipated that foreign aid will have a positive effect on economic development.

Inflation Rate

Describes a rise in the general level of prices of goods and services in any economy over a period of time Shim et al. (1995). Measures how fast the overall price levels are rising. When evaluating inflation, weighted averages of prices of services and goods often considered. The association between economic growth and inflation rate continues to be an issue for research as literature suggests three possible outcomes exists in any study. These are positive, negative or no relationship. If inflation rate for a year is 10%, this implies that year things generally cost 10% more in money terms compared to the previous year. Inflation rates of more than 10% per year are moderate but very disturbing to businesses and consumers. Hyperinflation occurs when the price level is rising by more than 20% per month. Inflation and economic development rates are two of the key macroeconomic variables that need to be closely monitored. Extraordinary inflation rates are very common in most developing nations.

Interest Rate

This refers to the amount charged for a purchase on credit or loan, characteristically conveyed as a yearly proportion of the credit or loan balance. This signifies the cost to the borrower of making a purchase on credit or taking a loan and represents rate of return for the creditor or lender. When interest rates are low (money is "inexpensive") investments are high, since businesses realise that even less-profitable establishments will still yield some cash flow required to service the loan. Nominal interest rates can be referred to as interest rate in monetary terms, does not include effects of inflation. Real interest rate refers to nominal interest rate minus the inflation rate, includes the effects of inflation

Stock Market

This can be referred to as an index of anticipations for the future. High values (bull market, rising prices) implies that investors anticipate economic development to be rapid, unemployment to be low and incomes to be high. If the stock market is low (bear market, falling prices), ordinary opinion anticipates the economic prospect to be relatively depressed.

Exchange Rate

The value of nations exchange in relation to other nations exchange. This comprises of the foreign legal tender and the domestic currency. May also be approximated indirectly or directly. For direct approximation, the value of foreign currency is taken in terms of local exchange while for indirect evaluation, the value of local exchange is taken in terms of foreign exchange. It can be said to be the value of one legal tender in terms of another money. For instance, consider the USD and the ZWL and say ,1 $USD = 120$ ZWL you pay a price of 1 USD for every 120 ZWL. Exchange rates can be floating or fixed. Floating implies the rates vary day to day according to the market. Fixed implies the rates remain at the value set by the administration. Nominal conversation rate describes the rate at which currencies of diverse nations can be swapped for one another. Real exchange rate defines nominal conversation rate accustomed to inflation. $R = \frac{EP^*}{P}$ (here E is the nominal native-currency value of foreign legal tender, P is the local price level, and P^* is the foreign value level. If local money gains its value matched to other currencies, this results in increases of prices for foreign-produced goods. Domestic products will be relatively expensive for foreigners, exports are likely to be low and imports may increase.

Government Expenditure

Describes governments acquisition of services and goods for current or future use Shim et al. (1995). The association between economic development and government expenditure is of prime importance in developing nations, most of which have witnessed or are experiencing increasing levels of public spending across times. National administration consumption is part of GDP (gross domestic product). Government consumption is likely to escalate GDP as it adds to present-day demands of goods and services. It possesses a similar positive feedback loop on gross domestic product as private consumption itself, since it intensifies GDP that is a cause of entire national consumption.

Unemployment Rate

Refers to the total number of jobless people divided by the labour force. To be unemployed one must be willing to work and be aggressively searching for employment. The labour force comprises of those who are unemployed and those employed. Unemployment rate is the rate thats presented to the public and media, only considers those who are entirely unemployed and have searched for employment in the last four weeks. Actual unemployment rate encompasses unemployment rate figures and slightly attached (havent searched for employment in the past four weeks), discouraged (surrendered and ceased searching) and underemployed (working part-time while still considering being full-time). Real unemployment rate comprises of everybody who is seeking for a full-time employment however doesn't have any and is generally double the unemployment rate. Frictional unemployment happens as employees and companies spend time looking for the best match. Cyclic unemployment: arises during depressions and recessions. Unemployment rate can be said to be the most appropriate pointer of how well an economy is performing comparative to the productive potential. An economy without unemployment would be a bad economy. So as to operate efficiently, so an economy requires inventories of employment-seeking individuals and vacancies. The unemployment rate can be viewed as maybe the best pointer of how healthy an economy is living up to the potential generated by the existing level of technology and the present-day stock of productive capital.

Exports

In economics, exports refer to goods and services produced in one country and sold to buyers in another country. Exports are an important part of international trade and can contribute significantly to a country's economy. Exports can take many forms, including physical goods such as manufactured products, agricultural commodities, and raw materials, as well as intangible goods such as services, software, and intellectual property. The value of a country's exports is an important indicator of its economic health and competitiveness. Countries with high levels of exports often have strong economies and are able to create jobs and generate income for their citizens.

Imports

Imports refer to goods and services that are produced in a foreign country and brought into a domestic market for sale or consumption. Importing goods and services is an important part of international trade, as it allows countries to access products and services that they may not be able to produce domestically or that are cheaper to import than to produce domestically. Imports can take many forms, including physical goods such as consumer products, raw materials, and machinery, as well as intangible goods such as intellectual property and services. The value of a country's imports is an important indicator of its economic activity and consumer demand. Countries with high levels of imports may have strong domestic demand and consumer spending, but may also be dependent on foreign sources for certain goods and services.

Money Supply

The entire stock of money and other liquid entities for specific phase of time in a country's economy is referred to as money supply. The balances that are earned various sources inluded in it.

International Reserves

nternational reserves are any kind of reserve funds, which central banks can pass among themselves, internationally. International reserves remain an acceptable form of payment among these banks. Reserves themselves can either be gold or a specific currency, such as the United States dollar or the Euro.Many countries also use international reserves to back liabilities, including local currency, as well as bank deposits.

2.2.2 Determinants of Economic Advancement in Developing Countries

This section, examines empirical growths studies which focused their investigations on detecting the main macroeconomic determinants of economic advancement in developing nations. Amongst these we have: Anyanwu (2014) looked at the factors affecting economic growth in Africa and China using an empirical growth model. Longitudinal data for African states for period $1996 - 2010$ as well as time series data for the $1984 - 2010$ era for China, the research findings revealed that for Africa higher domestic investment, government effectiveness (governance), net authorized assistance, metal price index, urban population and high school enrolment were positively and considerably correlated to economic development. For China, using a subset of the regressors, the study results indicated that trade openness and domestic investment were significantly and positively related to economic growth, while inflation, official development aid, population growth, agricultural material price, credit to the private sector, and oil price indices were negatively and significantly associated with economic growth.

Dobronogov and Iqbal (2005) examined the key determinants of economic growth in Egypt by combining econometric time series analysis with the growth diagnostics framework. They contended that inclinations in organization of economic cooperation, private sector credit and development and government consumption were amongst main determinants of economic growth in Egypt from 1986. They furthermore realised that the ineffectiveness of the fiscal intermediation was a key constraint on development. Conclusively they asserted that an enhancement in quality of financial intermediation might bring sustainable development dividend to Egypt in the long-run.

Hamilton and Monteagudo (1998), used least-squares regression, to re-examine the Mankiw, Mankiw et al. (1992) empirical model, using data for the period $1960-1985$. They incorporated, shifs in the rate of physical investment, the mean proportion of the working age population that was in high schools and the mean yearly rate of growth of the employed age population as variables for that nation for the period between 1960 and 1970. They realised that shifts in production growth were significantly and positively correlated to shifts in the rate of physical investment. On the other hand, shifts in labour force development were negatively and considerably linked to economic progression. They also found that the coefficient on the change in population growth was not statistically significant Hamilton and Monteagudo (1998).

Sen and Te Velde (2009), investigated the influence of effective state-business relationships on economic development using longitudinal data of 19 SSA nations for the era $1970 - 2004$. They suggested measures that capture diverse scopes of effective state-business associations in SSA. Using this they approximated the standard growth regressions by means of non-static longitudinal data approaches. Using that measure, together with traditionally used measures of institutional quality such as the rule of law, the degree of executive constraints, the quality of the bureaucracy and degree of corruption. Findings revealed that effective state business associations were positively and considerably correlated to economic development.

Antwi et al. (2013), examined the Influence of macroeconomic factors on economic development in Ghana: A co-integration analysis showed that asymptotically, economic advancement of Ghana was mainly determined by, foreign direct investment, government expenditure, physical capital, foreign-aid and inflation . It was also found that a short term changes in labour force had effect on the economic growth.

Ghura et al. (2001), examined the causes of progression in Sub-Saharan Africa. They concluded that economic rescue was due to constructive economic atmosphere influenced either indirectly or directly by positive changes in macroeconomic guidelines and organizational reforms. The fitted growth equation showed that GDP growth was positively and significantly correlated to economic policies that increased the proportion of private investment relative to GDP. Mankiw et al. (1992) examined the sources of economic growth in 95 developing countries, for the period 1976-1985. By means of a cross-sectional regression analysis, the research findings revealed that the index of real exchange rate distortion and real exchange rate variability and were negatively considerably related to asymptotic economic development. Investment rate was significantly and positively related to economic progression. Also, the research showed that the higher the degree of exchange rate instability, the lower the degree of technological diffusion from advanced economies Mankiw et al. (1992) . Thus the study concluded that outward-orientation plays a pivotal role in accelerating technological development in any economy this is realised through a low degree of protection and a stable real exchange rate regime.

Gyimah-Brempong (1989), scrutinized the impact of military expenditure on the economic development of Sub-Saharan African nations using simultaneous-equations model. Results showed that military expenditure negatively affected on economic growth. Though, the review of his analysis revealed that his conclusions were weak and did not support his policy conclusions, since the calculated values of the defence burden/growth rate multiplier was not statistically significant. Chen and Feng (2000) studied the association between trade (imports and exports) as a portion of real gross domestic product, higher education enrolment, inflation, state-owned enterprises, investment and economic advancement in China. By means of provincial longitudinal data, the research concluded that university enrolment and trade to be significantly and positively related to the yearly mean rate of per capita gross domestic product. State owned enterprises and Inflation conversely, were significantly and negatively related to economic advancement. Their study concluded, that foreign trade, private enterprises and education were important determinants of Chinas long-run economic growth.

Hostan (2015), examined the determinants of economic growth for Sub-Saharan Africa for the period 1981 to 1992. The results showed that lower budget deficit, public policies and private investment had positive effects on per capita growth. Fischer (1992) looked at economic growth and macroeconomic stability in Sub-Saharan Africa (SSA) , the Caribbean (LAC) countries and Latin America for the period $1970 - 1985$. By means of cross-sectional regression, the findings showed that investment, budget surplus and human capital were positively and significantly related to economic progression, whereas inflation, dummy variables and initial real GDP were negatively and significantly related to economic growth. He thus concluded that a reasonable level of macroeconomic stability is necessary for sustainable economic growth.

Easterly and Levine (1997), investigated the determinants of economic growth in Latin America, Sub-Saharan Africa and Caribbean Countries for the periods 1960, 1970 and 1980 using an empirical cross-sectional growth equation. The research established that the logs of schooling, fiscal surplus financial depth and number of telephones per worker were positively and significantly related to economic development, whereas black market premiums and political assassinations were significantly and negatively related to economic growth. Dummy variables were negatively and significantly associated with economic growth, revealing that Latin America, Sub-Saharan Africa and Caribbean Countries experienced sluggish economic growth. Furthermore, the study results showned that Africas poor growth related to political instability, low schooling, distorted foreign exchange markets, underdeveloped financial systems, insufficient infrastructure and high government deficits. The study concluded that; black market premiums, budget deficits, financial depth, political stability, human capital development and infrastructure accounted for some significant cross country variation in economic growth rates Easterly and Levine (1997).

Knight et al. (1993), extended the Mankiw et al. (1992) model by probing the association between human capital, investment, outward-oriented trade policies and public investment on economic growth for 98 countries using a panel regression method. Their research revealed a positive and strong association between human capital, physical capital and economic progression in both sub-samples of 81 and 59 developing countries. The findings also indicated that communal investment was significantly and positively related to economic prosperity in developing nations. Weighted tariffs as a measure of trade openness and population growth were both significantly and negatively related to economic growth in both sub-samples Knight et al. (1993). The research concluded that human capital, public investment, physical capital, openness to trade, and population growth were all significant determinants of economic growth.

Akinlo et al. (2005), examined the effects of macroeconomic factors on total factor productivity in 34 sub-Saharan African nations for the period $1980 - 2002$. Results of the econometric exploration revealed that foreign debt was significantly and negatively associated with entire factor productivity. Also, agricultural value-added as a proportion of GDP, local price deviation from purchasing power parity, lending rate and inflation rate were negatively and considerably associated with total factor productivity. Nevertheless,Akinlo et al. (2005) result revealed that export-GDP ratio, human capital, credit to private sector as percentage of GDP, liquid liabilities as percentage of GDP foreign direct investment as percentage of GDP and manufacturing value added as a share of GDP influenced total factor productivity positively.

Barro (1999), examined the causes of economic advancement utilising an extended neoclassical growth model for 100 states for the time phase $1960-1995$. Using panel regression and three stage least squares method, the research results revealed that years of schooling, rule of law index, democracy index, international openness, growth rate of terms of trade, and investment share, were positively and significantly related to economic growth, while total fertility rate, inflation and government consumption were negatively and significantly related to economic growth. Anaman (2006) investigated the determinants of economic growth in Ghana. He employed neoclassical development model on records from 1966 - 2000. Annual growth of GDP (real gross domestic product) was the regressent of the long-run growth model. Explanatory variables were yearly growth of entire labour, yearly increase of overall exports, entire investments-GDP ratio and size of government. Other explanatory entities were a dummies for world oil market price shock in the mid 1970s and early 1980s and extreme political upheaval related to major droughts or a military coup . Short run residual correction model, based on the asymptotic co-integrating function, was likewise fitted. Findings revealed that the asymptotic economic progress was positively linked to political stability. The world oil value shocks in the mid 1970s and early 1980s resulted in decreased economic development. Size of the government affected economic progression in quadratic equation manor with growing government size casing an increase till a point was realised beyond which development would essentially fall with growing size of government. Exports increase intensely influenced economic advancement. Though increasing total investment-GDP ratio had no effect in the asymptotic economic development regardless of the anticipated positive association among the two variables. Increase in labour force had no effect on economic progression signifying a trivial negligible labour productivity at the collective level. Short-run economic development was mostly a result of political stability. Generally, the findings revealed that political stability was a key ingredient for attaining long-run economic development in Ghana.

Steven et al. (2001), examined the determinants of economic development in 18 Asian states for the period $1965-1990$. They expanded the neoclassical cross-country development model. The research showed that land area to coastline distance, government reserves, initial education attainment, quality of institutions, trade openness, the increase of the working age population and life expectancy were significantly and positively linked to economic progression. Whereas initial output per worker, natural resource abundance, location in the tropics and land lockedness, and were negatively and significantly related to economic growth. Salisu, Ogwumike, et al. (2010) had a hand the growing debate on aid-growth nexus. They investigated the function of of macroeconomic policy atmosphere in aid-growth interconnection; a sector that got little consideration in (Sub-Saharan Africa). Using regression model encompassing twenty SSA nations, they approximated using ordinary least squares and two-stage least-squares (TSLS)covering the time 1970 − 2001 (of 9-4 years sub-phases), results revealed that a conducive macroeconomic setting was essential for a meaningful influence of aid to sustainable development. Also the outcomes indicated that macroeconomic policy atmosphere was a key determinant of progression. Generally the research established that macroeconomic instability, relentless socio-political crisis, policy discrepancies and immoral governance evident in many Sub-Saharan Africa nations had crippled the impact of aid in these states.

Barro et al. (2003), also examined the determinants of economic growth in a panel of 87 countries that encompassed developing and developed countries for the period 1965 − 1995. The findings showed that; investment, mean years of school completion, the rule of law, terms of trade were, trade openness and democracy were positively and significantly associated with economic development. On the other hand; life expectancy, landlockedness, fertility rate, initial level of per capita GDP, government consumption and, inflation rate were significantly and negatively related to economic growth.

Mbulawa (2015a), examined the effect of Macroeconomic Variables on Economic Growth in Botswana and realised that Inflation rate and FDI had a positive and significant association with Economic Growth. Whereas gross fixed capital formation had also a positive but insignificant effect on economic growth of country. Rao and Hassan (2011) examined the determinants of long-run economic growth in Bangladesh spanning trough years 1970 − 2007. Using an Autoregressive Distributed Lag method, results showed that the implementation of reforms since the 1980s, money supply, trade openness and FDI were significantly and positively related economic growth, whereas inflation and government expenditure were negatively and significantly related to economic growth.

Bhanu Sireesha (2013), investigated the effect of selected macro-economic variables on stock returns in India and established that there is an inverse relationship between Inflation, Index of Industrial Production (IIP) and Money Supply with returns from stocks, silver and gold. A direct relationship between GDP and stock return and an inverse relation was established between gold and silver returns. Nyoni and Bonga (2017) investigated the impact of macroeconomic policy environment in aid-growth nexus the area of which had received minimum attention in Sub-Saharan Africa (SSA). Using panel regression model for twenty SSA nations, he approximated using TSLS and OLS covering years 1970 - 2001 (in nine- four year sub-phases), results showed sound macroeconomic environment was key for a meaningful contribution of aid to sustainable development. Outcomes revealed that macroeconomic policy environs were an essential determinant of development. The study concluded that; bad governance, the incessant socio-political crisis, macroeconomic instability and policy inconsistencies inherent in many Sub-Saharan African nations hindered the impact of aid in these states.

Chang and Mendy (2012), explored the empirical relationship between economic growth and openness in 36 African countries for the period 1980−2009. They used al fixed effects regression model, the findings showed that; trade openness, labour employed ,imports, foreign aid and exports were considerably and positively related to economic development: yet, domestic investment, gross national savings and FDI were significantly but negatively related to economic advancement. The results also showed that external aid produced diversified findings when disaggregated with respect to region. In the west and east African counties external assistance was negatively and significantly related to economic development. Whereas in the middle and northern African regions, foreign aid was positively and significantly linked to economic growth.

Samanta and Sanyal (2009), examined the "correlation between economic growth and bribery: in Sub-Saharan Africa". The results showed bribery and economic growth were correlated. The influence of lesser levels of bribery on economic progression was found to be stronger as compared to the effect of high economic development rate in decreasing bribery. Also it was realised that there might not be any association between these two entities in some nations. These diverse results suggested nation specific aspects explained the pervasiveness of bribery and relative economic development; collective enlightenments needs to be circumvented. The results also postulated that rigorous efforts to curb corruption (bribery) have to remain an enviable course of action for non-governmental organizations, national governments, and international agencies. Lowering the prevalence of bribery translates to higher economic development rate and that consecutively might further speed up the drop in bribe becoming virtuous cycle which can aid to the economic welfare .Simultaneously, policies that promote economic growth must be established and implemented to arrest bribery.

Robinson and Dornan (2017), investigated the association between policies, foreign aid and economic growth in 56 developing countries encompassing 40 low-income and 16 middle-income countries. They used a TSLS (two-stage least squares) technique, the findings exposed that external aid was significantly and positively related to economic progression if it entered the growth regression as an interactive entity with policy. Nevertheless, external relief was established to have no effect on economic advancement in the nations studied. Conversely, the outcomes revealed that institutional quality, trade openness ,budget surpluses and were significantly and positively related to economic growth for states located in East Asia . Also, the results showed that inflation, political assassinations and countries located in SSA were significantly and negatively related to economic growth.

Klasen and Lawson (2007), investigated the association between per capita economic growth, population and poverty, using Uganda as a case study. Though Uganda had witnessed poverty reduction and excellent economic growth, it had one of the uppermost population increase rates in Africa and the world at large that, owing to the natural demographic impetus. Combining micro econometric and a macro and approach, used panel data, they considered the effect of population inrease on per capita economic advancement and poverty. Results indicated that the extraordinary population increase put a significant disruption on per capita progression projections in Uganda. Also, it added considerably on poverty alleviation and was linked to households being persistently in poverty. Conclusively they postulated that this was likely to make significant gains towards poverty reduction, and per capita growth, very complicated.

Agalega and Antwi (2013), established the impact of interest rates and inflation rate on GDP of Ghana. The results showed that there was a significant and positive association between GDP and inflation rate. Feng (1996) did a cross-national analysis of forty sub-Saharan African nations for the period $1960 - 1992$. He looked at the asymptotic association between economic progression and political democracy. The outcomes exposed that an economy progresses more quickly under an administration that promotes institutionalized democracy. Also he found out that a positive feedback relationship exists between growth and democracy, as democracy promotes growth, growth translates to a higher level of democratization. Feng (1996) established that in the time of authoritarian rule economic growth is minimum, while economic growth reduces the tenure of an autocratic governments. The original size of the economy, local investment share, global trade and human capital stocks were also noted as other aspects that influenced development in sub-Saharan African states.

Most and De Berg (1996), used country-specific time series growth models to investigate determinants of economic growth in eleven Sub-Saharan Africa countries. The research findings showed diverse results. External aid was realised to be significantly and negatively related to economic progression in Nigeria, Togo, Rwanda, Ivory Coast, Botswana and Zambia: while it was significantly and positively related to economic growth in Senegal, Mauritius and Niger. Domestic savings were realised to be significantly and positively linked to economic development in Ivory Coast, Senegal, Togo, Nigeria, Kenya and Cameroon, but significantly and negatively related to economic development in Zambia and Mauritius. FDI was significantly and positively associated with economic growth in Niger, Ivory Coast, and Togo and Kenya, but negatively and significantly connected to economic growth in Rwanda and Mauritius. Lastly, population growth was found to be negatively and significantly related to economic advancement in Senegal, Niger and Mauritius.

Salian and Gopakumar (2008), examined the relationship of inflation and economic growth of India and realised that there was a negative relationship between these variables. The research also showed a negative correlation between GDP and inflation in the Long run. They also established that Low or moderate inflation rate leads to High Economic Growth in long run while high inflation rate have a negative effect on economic growth. Ndambendia and Njoupouognigni (2010) examined the long-run relationship between foreign direct investment, foreign aid and economic growth in 36 Sub-Saharan Africa nations for the period 1980-2007. They realised robust confirmation of positive association of foreign direct investment and external aid on economic progression. Though, the influence of foreign aid on development in Sub-Saharan Africa was minimum. They postulated that it was better to focus on internal factors than external factors to enhance economic growth in SSA. Checherita-Westphal and Rother (2010) investigated the impact of high and growing Government debt on economic growth. An Empirical study on the Euro Area and established evidence of a non-linear bearing on GDP per capita development rate over twelve euro states in the long run.

Ristanovic (2010), established that there is negative relationship between fundamental economic variables and GDP. Ojo and Oshikoya (1995) investigated the determinants of long term growth in a cross section of African countries for the period (1970 1991).They included variables such as investment, population growth, foreign factors (external debt, export growth and terms of trade), macroeconomic regulation (exchange rates and inflation), initial per capita income, political environment, and human capital development. Results revealed that, the most significant variables influencing long-term growth in the sample of African countries over the study period were, the macroeconomic environment, investment, external debt and population growth.

Syed and Shaikh (2013), examined the effects of macroeconomic variables on gross Domestic Product (GDP) in Pakistan" the study established that there are three key factors which affects the GDP of Pakistan. These being industrial and business activities in country , agricultural and livestock sector and third one is related with fishing and mining sector. Acikgoz and Mert (2014), Investigated the association between investment and real GDP per capita in three Asian countries; Taiwan, Hong Kong and the Republic of Korea. They used time series data for the period 1953 – 2007 from the Republic of Korea, 1960 – 2007 for Hong Kong and 1951 – 2007 for Taiwan. Using fully modified ordinary least squares and autoregressive distributed lag techniques, the results of the study revealed that in short periods of time, economic growth was positively and significantly related to the investment share. Also the investment share was significantly and positively associated with the level of real GDP per capita asymptotically. The results were consistent in all three nations.

Ndambendia and Njoupouognigni (2010), examined the impact of Intellectual Property Rights (IPRs) on economic growth for 34 Sub-Saharan (SSA) countries over the period 1985 to 2003. He used three different estimation techniques (Fixed effects, Ordinary Least Squares and seemingly unrelated regressions), the results revealed that:

1. domestic investment is positively associated with economic growth;
2. human capital was a key determinant of economic growth and
3. strengthening IPRs negatively affected economic growth,

The results of the study recommended that a "one size fits all" move towards harmonizing IPRs in developing countries may not translate to the expected gains for Sub-Saharan African countries. Asheghian (2009) used an augmented neoclassical growth model to analyse the determinants of economic growth in Japan for the period $1971 - 2006$. Beach-Mackinnon technique was used; the results revealed that the growth rates of total factor productivity and domestic investment were positively and significantly correlated to economic growth.

Barro et al. (2003), examined the determinants of economic growth in a panel of 87 countries that covered both developed and developing countries during the period $1965 - 1995$. Founded on three cross-sectional growth regressions which covered the times $1965 - 1975$; $1975 - 1985$; and $1985 - 1995$, findings exposed that democracy, terms of trade, the rule of law, trade openness and investment, were all significantly and positively correlated to trade and industry growth. Whilst inflation rate, opening level of per capita GDP, landlockedness, government consumption, and fertility rate were negatively and considerably associated with economic development.

Agbor et al. (2014), examined how colonial origins affected economic growth in sub-Saharan Africa (SSA). The results showed that colonial origins influenced economic progression in SSA and its possible vector was human capital. Particularly, the findings pointed out those British prior colonies had attained their greater economic standings and performance against their French colleagues primarily due to the depressing effects of human capital development on GDP progression had been relatively less brutal in British previous colonies. The study did not establish statistical confirmation to support the impact; availability natural resources, the market distortion, geography and, trade openness on economic growth. Though, some factors which were found to be statistically insignificant were; natural resources and. Geography. Other related researches are as in the table below. The table below summaries other literature review on the determinants of economic development done elsewhere;

Table 2.1: Other Literature Review on Determinants of Economic Growth

Author(s)	**Nation(s) \Region and Time**	**Approach**	**Variables analysed**	**Results**
Sahu and Sharma (2018)	India, 1971 – 2016	Autoregressive distributed lag (ARDL) Technique	Gross domestic product Per Capita, External Aid, Inflation ,Government size. Government Expenditure, Trade Openness, Exchange Rate, Human Capital.	External aid, government size, and FDI positively and significantly impacted on the economic development in India whilst, exchange rate and human capital impact negatively on the growth.
Erkut and Sharma (2018)	India,1980 – 2016	Ordinary Least Squares Model	Gross domestic product Growth, Exports,Total Expenditure, FDI, Inflation, Gross Domestic Investment	Local investment and Exports are the key and significant mechanisms for the industrial sector whereas inflation and exports are major factors for the service sector's development rate.
Al Harrasia et al. (2018)	Pakistan, 1976 – 2015	Cointegration Approach, VECM, IRF,	Gross domestic product, Energy Consumption, Foreign Direct Investment, Trade Openness, Agriculture Rate	There is a positive effect of agriculture contribution,trade openness, energy consumption, and Foreign Direct Investment on the economic progression of Parkistan.
Musila and Belassi (2004)	Ghana (1965 – 1999)	Time Series	Education expenditure,GDP	Results showed that education expenditure per worker had a positive impact on economic development both in the short and long run.
M'amanja and Morrissey (2006)	Kenya (1964 – 2002)	Time Series Methodology.	Foreign Aid, Investment, and gdp	Findings were; shares of public and private investments and imports had strong valuable impacts on GDP of Kenya

Abidin et al. (2015)	Bangladesh, 1973 – 2010	Cointegration, Granger Causality test, ECM	Gross domestic product, Total Education Expenditure, Education,Human Development, Revenue Expenditure, Development Expenditure.	Results showed that; there is an asymptotic association between education and economic progression in Bangladesh. There is a uni- directional connectedness from Gross domestic product to education.
Aziz and Hossain (2012)	127 nations, (2000 – 2010)	Cross-sectional Study	GDP ,Capital, Labor Force, Human Capital,Non-corruption Score Polity Score.	Results revealed that; democracy negatively affects the economic development of the countries studied while polity Score positively influenced economic development.
Mosikari et al. (2016)	Botswana (1966 – 2014)	The Keynesian Approach	GDP, household final consumption, gross capital formation, imports, government expenditure	Results indicated that; long run there exists a positive relationship between household final consumption and Government expenditure. Exports have a positive effect on domestic production. Whereas imports have a negative impact on domestic production.
Bekere and Bersisa (2018)	14 Sub-Saharan Africa, 20 years	Dynamic generalized method of moment estimator	FDI and GDP	FDI is positively and significantly linked to economic advancement regionally.
Akitoby and Cinyabuguma (2004)	DRC (1960 – 2000)	Cointegration Approach	Debt crisis, Political chaos Adjustment supported by the IMF, Hyper-inflation and collapse of the economic and political system, Sectoral Output Performance	The main findings verify that poor economic policies and bad governance aided the countrys economic meltdown during the 40−year period, 1960 – 2000

Ncube (2019)	Zimbabwe (1980 to 2017)	Ordinary Least Squares model	gross fixed capital formation,Human capital, unemployment,government expenditure and inflation .	In the short run there exists positive association between lags of , government expenditure,inflation, and human capital with GDP.
Akinlo (2006)	34 sub-Saharan African countries (1980 – 2002.)	Econometric analysis	GDP, lending rate, inflation rate, external debt and domestic price deviance from purchasing power parity.	Results revealed that; external debt was notably and negatively connected to total factor productivity. Additional elements which had considerable negative effects incorporated lending rate, inflation rate, local price deviance from purchasing power parity and agricultural value-added as a proportion of GDP.
Ahmed and Uddin (2009)	Bangladesh, 1972 – 2008	Cointegration, Granger Causality Test	Gross domestic product,Agriculture, Industry, Service sector Contribution to Gross domestic product	Findings show that uni- directional causality from industry to agriculture sector Also from GDP to the service sector.
Rahman et al. (2011)	Bangladesh	Autoregressive distributed lag (ARDL) Model	Trade Openness, TFP,FDI, Development in the Financial Secto	All explanatory variables enhanced the total factor productivity for the nation.
Uddin and Sjö (2013)	Bangladesh, 1976 – 2005	Johansen Cointegration Test	Exports, GDP, Remittance, Imports.	There is short-run causal relation among; imports, exports, remittance, and economic growth.

Mbulawa (2015b)	Botsawana(1975−2012)	Vector error correction approach	Foreign Direct Investment, volume of trade, Trade openness Inflation and capital accumulation	Trade openness and Inflation had a significant negative and positive effect on economic progression correspondingly. Inflation converged to asymptotic equilibrium with economic growth and causative associations were established among other entities in the short term. The response of economic development to shocks of trade openness, gross fixed capital formation and foreign direct investment was effective even after the 30 year period whereas shocks from inflation were insignificant.
Chirwa and Odhiambo (2014)	Zambia (1972 − 2013)	Autoregressive distributed lag (ARDL) bounds testing approach	Human capital development, investment, government consumption,foreign aid and International trade .	Results were: in the short run, investment and human capital development are positively connected to economic development, whereas international trade, government consumption and external aid were negatively linked to economic progression asymptotically. Human capital development and investment were positively related to economic advancement, while external aid was negatively associated with economic growth.
Chirwa and Odhiambo (2015)	Malawi (1970 − 2011)	ARDL (The autoregressive distributed lag) bounds testing method.	Accumulation of Physical Capital and Growth, Inflation, Human Capital, International Trade, Real Exchange.	Key macroeconomic movers of economic progression in Malawi in that period were: human capital development, the real exchange rate, the accumulation of physical capital, inflation and international trade.

Shahbaz et al. (2008)	Pakistan, (1991 – 2007)	ARDL and Simple Linear Regression Models	Financial Development,GDP, Trade Openness, FDI ,Inflation, Remittances and Local Investment.	Remittances positively enhanced by economic development. There exists long-run relations among the entities.
Nhlengethwa et al. (2021)	Eswatini (2000 – 2017)	Pearson Pairwise Correlation, Time series, OLS regression techniques and Unit-root tests	Infrastructure Investment, GDP , Agriculture GDP, Interest, FDI to Agriculture, ODA Agriculture, Sugar Exports, Government Debt, Government Savings, Education Expenditure, Health Expenditure, Rates, Infla-tion Rate.	Agricultural water infrastructure investment and Infrastructure were realized to be positively interconnected to GDP, FDI into agriculture and Sugar export income. It could be concluded that; increased earnings plus terms of trade for sugar could improve expenditure on agriculture water investments. This is essential since an increase in investments in water infrastructure might then assist spur economic progression.
Chirwa (2016)	South Africa (1972 – 2013)	ARDL bounds-testing approach.	Investment, inflation,population growth, human capital development, international trade and government consumption.	In the short-run, investment is positively linked to economic development, whereas population growth and government consumption are negatively correlated to economic growth. Nevertheless, asymptotically, the research discovers human capital development, investment and international trade being positively linked to economic progression, whereas government consumption, population increase, and inflation were negatively related to economic advancement.

Seleteng and Motelle (2014)	SADC 1980 to 2012	Generalised Methods of Moments, Seemingly Unrelated Regression Estimators.	GDP, inflation, gross fixed capital level of financial development, formation to GDP, public spending on education, share of liquid liabilities, government expenditures, the institutional variable which proxies level of democracy openness to trade, level of financial development, human capital,and political stability.	Entities influencing economic growth in the region were: inflation, political stability, government expenditures, openness to trade, human capital and level of financial development.
Sekantsi and Kalebe (2015)	Lesotho (1970 to 2012)	Autoregressive distributed lag (ARDL) bounds testing approach to cointegration based Granger causality test and Vector error correction model (VECM).	GDP, saving.	Saving precedes and drives short-term and long-term capital accumulation but also contributes to long-term economic growth in Lesotho.
Brück and Van den Broeck (2006)	Mozambique (1996 − 97 and 2002 − 03)	OLS	GDP, employment and poverty.	Unemployment and Poverty are negatively correlated to economic growth.
Eita and Ashipala (2010)	Namibia (1971 to 2005)	OLS	GDP; capital stock; labour employed; and the level of technology.	Capital stock, the level of technology, and labour employed have a linear relationship with GDP.
Mongale and Monkwe (2015)	South Africa (1973 − 2013)	CVAR Analysis	GDP, exports, imports and infrastructure.	All the variables influence economic growth, albeit positive or negative effects.

2.2.3 Why Panel Data?

Hsiao Hsiao (2003) and Klevmarken Klevmarken (1989) spelt out benefits of using panel data. These consists of those listed below:

Adjusting for Specific heterogeneity.

Data in panel form proposes that specific, companies, nations or states are diverse. Cross-section and Time-series researches not regulating this heterogeneity run into the threat of getting biased results Klevmarken (1989). To illustrate this consider the following empirical example; Baltagi and Levin (1992) looked at cigarette demand across 46 American states for the years $1963 - 1988$. Uptake was modeled as a function of lagged demand, income and price . These variables vary by states and time. Nevertheless, at hand are lots of additional entities that might be time-invariant or else country-invariant that may influence demand. Let us label these Z_i and W_t, correspondingly. Cases of Z_i are education and religious convictions . In place of the religious conviction variable, we might unable to obtain the actual proportion of the population that is, for example, Mormon for each nation for each single year, nor do we expect that to vary much across times. This is also correct with reference to the proportion of the population finishing an academic degree or high school. Illustrations of W_t embrace marketing on national TV, News papers and radios. This marketing is countrywide plus does not show discrepancies over nations. Some of these entities are challenging to approximate, measure or observe and rare to come by with the intention of having not all the Z_i or W_t variables being available for inclusion in the demand equation. Living out these entities results in bias of the resultant statistics.

Panel data can regulate these time-invariant and nation entities while a cross-section study or a time-series research is not able to control them. Actually, with reference to the data we observe that Utah consumes below half of the mean per capita uptake of cigarettes in the USA. This may be due to the fact that it is typically a Mormon nation, a religious conviction that outlaws smoking. Regulating for Utah for a cross-section regression can be achieved by using dummy variables that removes that nations realisation from the regression. This is not the circumstance with panel data as will be seen shortly. Actually, for panel data, we may initially difference the data to remove all Z_i -kind of variables and therefore successfully controling all nation-specific variables. The method works regardless if the Z_i being recognizable or not. Otherwise, the pseudo variable for Utah explains every nations specific effects that is unique to Utah without neglecting the realisations for Utah. Another illustration was given by Hajivassiliou (1987) who studied the external debt repayments challenge using a panel for 79 developing countries observed during the period 1970 to 1982. These states vary with reference to their colonial history, religious affiliations financial institutions, and political administrations. Entirely every nation-specific variable affects the assertiveness that the states have relative to borrowing and nonpayment and how they are veiwed by the lenders.

Failing to account for this country heterogeneity results in serious misspecification. Deaton (1995) points out another example from agricultural economics sector. This concerns the question, are small farms more productive matched with big farms. OLS regressions of yield for every hectare against implements for example, fertilizers, land, labour, farmers education, etc. usually notes a negative relationship. Implying that reduced farms are more prolific. Alternative explanations from economic theory suggests that higher output per head is an optimal response to uncertainty by small farmers, or that hired labour requires more monitoring than family labour. Deaton (1995) suggests an alternative clarification. This regression is affected by exclusion of overlooked heterogeneity, for instance "land quality and climatic conditions", and this excluded variable is associated with the independent variable (farm size). Actually, farms in poor-quality marginal areas (arid) are characteristically large, whereas farms in first-class land areas are usually small. Deaton stresses that though gardens have extra value-addition per hectare compared to sheep stations; it does not imply that sheep stations should be structured as gardens. Here, differencing might not answer the "small farms are productive" question as farm dimensions will generally vary marginally or by no means across small time phases.

Panel data gives more revealing data, a reduced amount of collinearity among the variables, extra degrees of freedom, additional efficiency and more variability.

Time-series researches remain overwhelmed by multicollinearity; for instance, referring to the demand of cigarettes in the prior section, there is great collinearity between income and price in the collective time series for the nations. This is less probably through a panel across nations as the cross-sectional dimensions enhances lots of variability, bringing about more enlightening data on prices and incomes. Actually, the variation in the data may be disintegrated into disparities between nations of different dimensions and features, and disparities within nations the prior deviation is generally larger. Through extra, more revealing data we can come up with more dependable parameter estimates. Unquestionably, the equivalent association must hold for each nation, i.e. the data must to be poolable. This assumption is verifiable and will be tackled in this research.

Panel data remain better placed to investigate dynamic forces of adjustment.

Cross-sectional distributions which appear comparatively steady hide massive amounts of variations. Periods of job turnover, unemployment, income plus residential movement are better researched using panels. Cross-sectional time series files are likewise well placed to investigate the period of economic situations such as poverty and unemployment. When the panel data sets are extended enough, they may indicate the swiftness of variations of modifications of economic policies. For instance, in determining employment levels, cross-sectional data can approximate the percentage in the population that is not employed at any point in time. Recurrent cross-sections are able to indicate how this fraction varies with time. Panel data only is able to approximate the percentage of those who were without jobs in one period remain out of employment in other periods. Key policy issues like determining whether families experiences of poverty, unemployment and welfare dependence are temporary or chronic demand the use of panels. Deaton (1995) contends that, not like cross-sections, panel surveys gives data on changes for households or individuals. It enables one to detect how individuals living conditions changed through the development process. It allows one to establish those gaining from developments. Longitudinal data it enables one to monitor if insufficiency and scarcity are transient or long-lasting, thus it responds to the revenue-dynamics questions. Panel data are also essential in the approximation of inter-temporal associations, life cycles and intergenerational models. Actually, panels can link the individuals experience and behaviour at any period in time to other experience and behaviour in another periods in time. For instance, when appraising training activities, groups of partakers and non-participants are analysed prior and after exposure to the training programs. This gives an example of panel of at least two time phases and the source for the "difference in differences" estimator generally used in these researches Bertrand et al. (2004).

Panel data are more able to recognize and measure effects which are basically not measureable in pure time-series or pure cross-section data.

Assume there is a cross-section of men having a 50% mean annual labour force partaking rate. This may be because (i) individual man having a 50% probability of participating in labour force, for any given working year, or (ii) half of the men employed all times as well as 50% not participating at all. Situation (i) has an increased turnover, whereas situation (ii) does not have any turnover. Panel data only may well distinguish these two situations. An additional case is that of determining if trade union affiliation reduces or improves incomes. This is best answered if one observes an employee moving from nonunion to union job or the converse. Keeping an individuals characteristics constant, one will more able to determine if union membership have effects on wages and by what magnitude. This investigation encompasses the approximation of other kinds of income differences taking peoples features constant. For instance, the approximation of income rewards rewarded for risky or nasty jobs. Economists researching on workers level of gratification face the challenge of anchoring in a cross-section research Winkelmann and

Winkelmann (1998). The survey generally enquires: "how one is content with his/her life?" having 0 representing absolutely disgruntled and 10implying wholly contented. The challenge here is that every worker hinges their scale differently, thus interpersonal assessments of answers becomes worthless. Nevertheless, in a panel research, when metrics used by workers are time-invariant across the phases of observations, we circumvent this challenge as the observer makes inferences with reference to only on intra- instead of interpersonal contrast of gratification.

Panel data models enables one to build and assess complex interactive models compared to wholly time-series or cross-sectional data.

For instance, technical efficiency is better studied and modelled with panels; Baltagi and Griffin (1988) Cornwell et al. (1990) Kumbhakar and Lovell (2000) Baltagi et al. (1995) Van den Broeck et al. (1994). Furthermore, fewer limitations can be imposed on panels in a distributed lag model than in an exclusively time series research Hsiao (2003).

Micro panel data collected from states, entities and companies are more correctly determined than comparable variables observed at the macro levels.

Biases emanating from aggregation over companies or states can be minimised or eliminated Blundell (1988) Klevmarken (1989).

Macro Time Series Cross-sectional data conversely have an extended time series and do not have the challenge of nontraditional distributions distinctive of unit roots tests in time-series modeling.

2.2.4 Shortfalls of panel data comprise of :

Data collection and design challenges.

These consist of challenges of coverage (partial account relative to the population under study), nonresponse (resulting from lack of cooperation by the respondent or for the reason that of interviewer error), recall (respondent failing to remember correctly), interviewing frequency, spacing of interviews, reference period, use of bounding plus time-in-sample bias Bailar (1989).

Distortions of measurement errors.

Measurement errors can result from faulty responses due to vague questions, recall errors, deliberate distortions of responses (e.g. status bias), improper informants, misreporting of replies also interviewer effects Kalton (1989). Herriot and Spiers (1975), for instance matched the cost per sale (CPS) with Internal Revenue Service data on incomes of certain individuals and observed shows that there exists inconsistencies of a minimum of 15% between these two bases of incomes for nearly 30% of the matched samples. A validation research by Duncan and Hill (1985) on the Panel Study of Income Dynamics (PSID) correspondingly reveals the significance of the measurement error challenge. They contrast the answers from the employees of a big company against employer records .Duncan and Hill (1985) discovered minor response biases apart from working hours which were overvalued. The proportion of measurement error variances to exact variances was established as 15% for yearly incomes, 37% for yearly working hours and 184% for mean hourly remunerations. These values belong to a single-year recall, i.e. 1983 for 1982, and are in excess of doubled with two years recall. Brown and Light (1992) investigated the discrepancy in job tenure responses in the PSID and (National Longitudinal Survey) NLS. Cross-sectional data handlers have slight choices but to accept as true the conveyed figures of tenure (except if they are in possession of outside evidence) whereas handlers and users of panel data can check the discrepancies with respect to tenure responses with time between interviews. As an example, a participant can assert having four years of tenure in an interview then one year later claims eight years. This ought to alert the users of this panel data to the existence of measurement errors. Brown and Light (1992) showed that failure to use internally reliable tenure sequences may result in misleading deductions about the slope of wage-tenure profiles.

Selectivity challenges.

These includes:

1. Individual-selectivity.

   Individuals may prefer not to be employed when reservation wages are greater as compared to the offered wages. Here one looks at the features of the personalities not their wages. As only their income is omitted, the samples are concealed. Nevertheless, if one does not scrutinize all data on these individuals this would become truncated samples. An illustration of truncation can be the New Jersey negative income tax research. If a researcher is concerned with poverty, and individuals with earnings greater than 1.5 times the poverty datum line will be excluded from the sample. Conclusions resulting from this truncated sample comes with bias that cannot be solved by more data, because of the truncation Hausman and Wise (1979).

2. Nonresponse.

This may arise on the initial stages of panel studies owing to refusal to partake, untraced sample units, not one person at home and other causes. Item (or incomplete) non-response happens once one or more questions remain not answered or are established as not providing an appropriate response. Absolute non-response happens if no information is obtainable from the sampled entity. In addition to the efficiency loss resulting from omitted data, this non-response causes grave classification challenges to the population parameters. Horowitz and Manski (1998) demonstrate that the gravity of the challenges is directly proportional relative to the level of non-response. Non-response proportions in the initial hhase of the European panel surveys differ through states starting at 10% for Greece and Italy in which partaking remains obligatory, to 52% for Germany and 60% for Luxembourg. The inclusive non-response rate was 28% Peracchi (2002). The analogous non-response rates for the first wave of the PSID was 24%, for the British Household Panel Survey (BHPS) (26%), for the German Socio-Economic Panel (GSOEP) (38%) and for Luxembourg Panel Socio-Economique Liewen zu Letzebuerg (PSELL) (35%).

3. Attrition.

   Although non-response takes place as well in cross-section researches, it remains a more severe challenge in panel studies as succeeding phases of the panels may also experience noreplies from partakers. Participants might relocate, or die or realize that costs of participating are high. The levels of attrition varies relative to the panel under study. Generally, the total rates of attrition increases from one phase to the other, while the escalation rates of deteriorates with time. Becketti et al. (1988) studied the representativeness of the PSID after 14 years after it started. The reseachers found that just 40% of the original in the sample of 1968 were still the sample in 1981. Nevertheless, they did not find that with regard to the dynamics of entry and exit in the panel, the PSID remains representative Peracchi (2002). An attempt to offset the effects of attrition is by the use of rotating panels, here a constant proportion of the participants are substituted during every phase to top up the sample

Short time-series measurements.

Characteristic micro panels have yearly data spanning over a short period for each entity or individuals. This shows that asymptotic influences depends essentially on the quantity of entities tending to infinity. Lengthening time spans of panels comes with some costs. Actually, it escalates the probability of attrition and intensifies computational difficulties for restricted dependent variable time series cross-sectional models.

Cross-sectional dependence.

Macro panel data on nations or regions having protracted time series that do not give a justification for cross-country dependence, this may lead to deceptive inferences. Accounting for cross-sectional dependence appears to be of importance and has effects on inferences. Other panel unit root tests are recommended which provides explanations for the dependence. Panel data is not a magic potion sometimes it fails solve some challenges that cross-sectional or a time series researches could not handle. Collecting panel data is relatively expensive, and there always exists a question regarding the frequency of interviewing respondents. Deaton (1995) contends that economic advancement is far from instantaneous, as variations from one year to the next are possibly too noisy and too temporary to be actually useful. Deaton concluded that pay-offs for panels is across extended time periods, ten years, five years, or longer. Comparatively, for nutrition and health concerns, particularly with respect to children, we could claim a conflicting situation, i.e., panel data having smaller time periods are needed so as to assess the well-being and growth of these infants. This research makes the case that panel data delivers numerous rewards worth its costs. Nevertheless, as Griliches (1986) contended about economic data generally, the more we get it, the more we demand of it. An economist, statistician or some other user using panel data or any data need to be acquainted with its shortfalls and restrictions.

2.2.5 Forms of Panel Data

Panel data sets consists of n units , subject or entities, each having T realisations taken from 1 across t time periods. Hence, the overall number of realisations in the panel data set becomes nT. Preferably, panel data are observed at unvarying time periods (e.g., months, years, and quarters). Or else, panel data ought to be explored cautiously. Panels may be balanced or unbalanced, long or short and fixed or rotating.

Balanced contrasted with Unbalanced Panels

For balanced panels, every single unit, subjects or entity has observations at each and every time period. For a cross-tablation (or contingency table) of panel data variables, each cell ought to be having exactly a single frequency. Thus, the total frequency of realisations becomes nT. If every object in a data set has not the same frequencies of observations, the panel data set is unbalanced. Other chambers in the contingency tables have no entries. As a result, the entire quantity of realisations is not nT for an unbalanced panel data set. Panels which are unbalanced require some computational and approximation issues even though majority of software packages are capable of handling both unbalanced and balanced panels.

Short against Long Panel Data

A short panel data set has many individuals or entities (large n) but a small number of time periods (small T), whereas a long panel has many time periods (large T) but few entities or individuals Cameron, Trivedi, et al. (2010). Consequently, short panels data are wide with respect to width (cross-section wise) and the length is short (time-series), while long panels are narrow in width. Equally excessively small N (Type I error) and excessively large N (Type II error) challenges are of concern. Investigators need to extremely cautious particularly when scrutinizing either long or short panel data set.

Rotating against Fixed Panel Data

If identical entities (or individuals) are observed in each and every period, the panel data set is referred to as a fixed panel Greene (2003). If entities (or individuals) vary over periods, the panel data set will be rotating panel.

2.2.6 Data Arrangement: Wide against Long Form

An ideal panel data set possesses across-sectional (subject, individual or entity) variables and time-series variables. The arrangement of this form is known as the long form (as contrasting to the wide form). Whereas the long form possesses both individual (e.g., group or entity) and time-based variables, the wide form embraces either individuals or time variables. Majority of statistical soft wares undertake that panels are organized as long forms .The table below (table 2.2) shows a typical panel data arrangement.

When data are organized as a wide form, one needs to reorganize the data initially. A number of statistical packages have the commands which enables one to rearrange data sets back and forth between short and long form.

2.2.7 Appraising Panel Data Qualities

The initial undertaking that a researcher must do having cleaned the data is assessing the quality of data under consideration. The time one says panel data, he or she will be implicitly implying that the data are soundly organized by both time-series and cross-sectional entities, also he or she gets a resilient intuition of availability of random or/and fixed effects. If not, the data is just (or substantially) organized in longitudinal data design however is no longer a longitudinal data sets with respect to econometric sense. The principal concern is regularity in the entity of measurement (or analysis), that stipulates that each realization of data set need to be weighted and treated equally. This prerequisite appears to be self-evident but more

Table 2.2: A Typical Panel Data Arrangement

Count	**Entity(i)**	**Time(t)**	$\mathbf{Y_{it}}$	$\mathbf{X_{1t}}$	$\mathbf{X_{2t}}$	$\mathbf{X_{3t}}$
1	1	1				
2	1	2				
3	1	3				
4	1	4				
5	1	5				
6	1	6				
7	1	7				
8	1	8				
9	1	9				
10	1	10				
11	1	11				
12	1	12				
13	1	13				
14	1	14				
15	1	15				
16	2	1				
17	2	2				
18	2	3				
19	2	4				
20	2	5				

often than not is ignored by careless investigators. If collectively observations are inequivalent in numerous senses, whatever exploration established from such data may be unreliable. Presented below are some checkpoints which scholars or investigators ought to scrutinize cautiously.

- Ensure the data are actually panel and there exists some random or/and fixed effects.
- Confirm whether entities (for example; subjects or individuals) are inconsistent or mutating. For example, a corporation may fragment or merge in the course of the study era and become a totally new one.
- Likewise, verify whether time eras are inconsistent but varying. Time periods in some situations might not be constant but practically stochastic (for example; second period is three years after the initial phase, third phase is 5 yearsater the second phase, fourth

phase is two years later the third phase, etc.) For other data, time periods are constant but various times are used; together monthly and yearly data coexist in one data set.

- Assess whether if an individual possesses over one realization for a certain phase. For instance, ECONET has six realisations in place of quarterly sales data in 2019, whereas every other companies has two annual sales realisations for that year. In a case of this nature, one may combine quarterly sales to get annual statistics.

- Make sure observation procedures used are dependable. Observations are not commensurable when i) some units were observed using approach A while some units using technique B, ii) other time phases were evaluated using approach C whereas other phases by approach D, or/and C) both A) and B) are mixed.

- Take care "darning" your data sets by merging data sets obtained and formulated by separate establishments that used not the same procedures. This situation is reasonably comprehensible since perfect data sets are hardly ever available; in several circumstances, one needs to merge different origins of statistics to construct a new data set in an investigation.

The other concern is when the quantity of units or/and time-periods is tiny or too big. It becomes less important to compare any groups (or time phases) with the other in a panel data framework: $T = 4$ or $n = 3$. Also note, relating millions of entities or time periods is of no value there is a high probability of Type II error. The case is virtually analogous to arguing that at least a single firm out of millions companies worldwide has a different efficiency. One already knows that. Now for situation of excessively large N (particularly T or n), one may need to reclassify entities or time phases into more realistic classifications; for instance, categorize billions of people say by their ethnic groups or citizenships (e.g., Asian, Spanish, white and black). Lastly, numerous omitted observations will probably compromise the quality of panel data. So referred to list-wise deletion (the whole record is omitted in the analysis if one particular measurement of an entity is missing) have a tendency of reducing the frequency of realizations used in models and consequently may reduce the power of a statistical test. As soon as a well-structured panel data set is ready, one may proceed to the modelling process.

2.2.8 Fundamentals of Panel Data Models

Panel data models look at time effects, group (entity-specific) effects or both in an attempt to manage heterogeneity or specific effects which might not or might be measured. These effects remain either random or fixed effects. Fixed effects models scrutinizes whether intercepts differ through groups or time phases, while random effects models checks variances in residual variance entities over time periods or individuals. One-way models omits only a set of dummy variables for example; company 1, company 2, ...), whereas two-way models reflects on two sets of dummy variables (for example; state 1, state 2, and year 1, year 2,...).

Pooled OLS

If individual effect u_i (time based or cross-sectional effect) is absent ($u_i = 0$), ordinary least squares (OLS) yields consistent and efficient parameter estimates

$$y_{it} = (\alpha + u_i) + X'_{it}\beta + \epsilon_{it} \qquad\qquad (u_i = 0), \tag{2.1}$$

Where: y_{it} is the GDP of country i at time t, X'_{it} is a matrix of independent variables, β, gives a matrix of coefficients of the independent variables and ϵ_{it} are the error terms.
OLS has of five primary assumptions (Greene, 2003 and Kennedy, 2008).

1. Linearity dictates that dependent variables are expressed as linear functions of sets of regressors and residual (error) terms.

2. Exogeneity articulates that the mean of error terms is zero or residual terms are uncorrelated to any regressors.

3. Error terms have a constant variance (3.a homoskedasticity) and are uncorrelated to each other (3.b non-autocorrelation).

4. The realisations on regressors are not stochastic but constant in recurrent sections without observation errors.

5. Full rank postulation pronounces that there exists no precise linear relationships amongst regressors (none existence of multi-collinearity).

If entity effect u_i is not zero in panel data, heterogeneity (entity specific features similar to personality and intelligence which are not taken care of in explanatory variables) can affect assumptions 2 and 3. Particularly, residuals might not have a constant variance but fluctuates over individuals (heteroskedasticity, violation of postulation 3.a) and/or are associated with each other (autocorrelation, violation of supposition 3.b). This is a question of non-spherical variance-covariance matrix of error terms. The violation of supposition 2 results in random effects estimators biased. Thus, the ordinary least squares estimator is no more (BLUE) best unbiased linear estimator. Longitudinal data modelling has a way of dealing with these challenges

Random Fixed against Effects Models

Models for panel data look at random and/or fixed effects of time or individuals. The function of dummy variables is where fixed effect and random effect models diverge most. According to Park (2011), a parameter estimate of a dummy variable is a component of the intercept in a fixed effect model and of the error term in a random effect model. Slopes in either a fixed effects models or a random effects models are constant across groups or over time. One-way random and fixed effects models have the following functional forms:

Fixed effects model:

$$Y_{it} = (\alpha + u_i) + X'_{it}\beta + v_{it}, \tag{2.2}$$

Random effects model:

$$Y_{it} = \alpha + X'_{it}\beta + (u_i + v_{it}), \tag{2.3}$$

where u_i is a random or fixed effect for a particular time to period or individual (group) which is excluded from the regression, also the error terms are independently and identically distributed, that is: $v_{it}\ IID(0, \sigma_v^2)$. The fixed group effects models looks at distinct variances for the intercepts, supposing the constant variance and identical slopes through individual (entity and group). Individual-specific effects are permitted to be associated with other regressors since they are time-invariant and regarded as a section of the intercept. Thus, the ordinary least squares (OLS) assumption 2 is not violated. The least squares dummy variable (LSDV) regression (ordinary least squares with a set of dummies) and within effects, approximation techniques are used to estimate this fixed effect model.

Table 2.3: Differences between FEM and REM

	Fixed Effects Model(FEM)	Random Effects Model(REM)
Functional Forms	$\mathbf{Y_{it}} = (\alpha + \mathbf{u_i}) + \mathbf{X'_{it}}\beta + \mathbf{v_{it}}$	$\mathbf{Y_{it}} = \alpha + \mathbf{X'_{it}}\beta + (\mathbf{u_i} + \mathbf{v_{it}})$
Assumptions	No	Individual effects would not be correlated with independent variables
Intercepts	Vary over groups and or time.	Fixed
Residual variances	Fixed	Randomly distributed over groups or times
Slopes	Fixed	Fixed
Determination Method	LSDV	GLS , FGLS (EGLS)
Hypothesis Tests	F- test	Breusch-Pagan LM test
Realisations	nT	nT

Random effects models undertakes that specific effect (heterogeneity) is unrelated to any independent variable and then approximates residual variance by time (or group). Consequently, a random effect models are also known as error component models. Individuals have the same intercepts and slopes as those of regressors. The distinction between time periods (or nations) is based on their particular idiosyncratic errors, not their intercepts.

If the covariance configuration of an individual $\sum$ (sigma), is given or known, GLS (generalized least squares) are used to estimate a random effect model. When $\sum$ (sigma) is unknown, the FGLS (feasible generalized least squares) or EGLS (estimated generalized least squares) approach is employed to approximate the full variance-covariance matrix V ($\sum$ for all diagonal entries and 0 for all off-diagonal entries). The maximum likelihood technique and simulation are two estimate methods for FGLS Baltagi and Chang (1994b). When individual particular random effects are linked with regressors, a random effect model decreases the number of parameters to be assessed but produces inconsistent results Greene (2003).

The F test examines fixed effects, while the Lagrange multiplier (LM) test examines random effects Breusch and Pagan (1980). If neither test rejects the null hypothesis, the pooled ordinary least squares regression is selected. Hausman (1978) compares a random effects model to the fixed effects equivalent in the Hausman specification test. A random effects model is favoured over its fixed effects counterpart when the null hypothesis of the individual effects are not correlated with the other independent variables is not rejected. The one-way fixed or random effects model is used when a single time-series or cross-sectional variable is evaluated (for example, nation, company, and race). Two-way effect models entail certain challenges in interpretation and estimate since they feature a pair of dummy variables for individual or/and temporal variables (for instance, year and country).

Approximating Fixed Effects Models

There are numerous methods for approximating a fixed effects model. Dummy variables are used in the LSDV (least squares dummy variable) model, but not in the "within" approximation. These methods yield parameter statistics of independent random variables (non-dummy regressors) that are indistinguishable. The "between" estimate fits a model without dummies by means of individual or temporal means of dependent variables and regressors. Since it is comparatively easy to approximate and interpret practically, least squares dummy variable with a dummy dropped out from a set of dummies is commonly employed. When there are many groups or individuals in panel data, this LSDV becomes difficultBaltagi and Baltagi (2008). If T is constant and $n \to \infty$ (n is the number of groups or firms and T is the number of time periods), regressor parameter estimates are consistent but individual effect coefficients, $\alpha + u_i$,

are not Baltagi and Baltagi (2008). LSDV contains a significant number of dummy variables in this short panel; the number of these parameters to be estimated grows as n grows (incidental parameter problem); consequently, LSDV loses n degrees of freedom yet provides less efficient estimators Baltagi and Baltagi (2008).In this case, LSDV is worthless, necessitating the employment of another method, within effect estimate.
The "within" approximation, not like LSDV, does not require dummies and instead relies on departures from the group (or time period) averages. As a result, the "within" estimate employs variation within each entity or individual rather than a huge number of dummy variables. The "within" estimate is as follows:

$$(y_{it} - \bar{y}_{i.}) = (x_{it} - \bar{x}_{i.})'\beta + (\epsilon_{it} - \bar{\epsilon}_{i.}), \tag{2.4}$$

where $\bar{y}_{i.}$ denotes the average of dependent variable (DV) of a specific (group) i, $\bar{x}_{i.}$ symbolizes the averages in regressors of group i, and $\bar{\epsilon}_{i.}$gives the average of residuals of group i.

The incidental parameter difficulty is no more an issue in this "within" estimation. The regressor parameter estimates in the "within" approximation are the same as those in the LSDV. The "within approximation presents correct the SSE (sum of squared errors). The "within" estimation, on the other hand, has significant backdrops. To begin with, the data transformation for "within" estimation removes any time-invariant variables that do not fluctuate within an entity (e.g., citizenship, ethnic group, and gender) Kennedy (2008). Because the deviancies in time-invariant entities from the mean are all zero, estimating coefficients of such subjects for the within" approximation is impossible. Consequently, when a model contains time-invariant independent variables, we must fit LSDV. Secondly, the "within" estimation yields erroneous statistics. Because there is no use of dummy variables, the within effects models have more degrees of freedom for error terms, resulting in reduced SEE (standard errors of the estimates), MSE (mean squared errors), or SRMSE (the square root of mean squared errors) and wrong (reduced) standard residuals for parameter approximations. As a result, one must rectify inaccurate standard errors by the use of the formula below:

$$se_k^* = se_k \sqrt{\frac{df_{\text{error}}^{\text{within}}}{df_{\text{error}}^{\text{LSDV}}}} = se_k \sqrt{\frac{nT - k}{nT - n - k}}. \tag{2.5}$$

Thirdly, since the intercept term is inhibited, the , R^2 of the "within" approximation is incorrect. Lastly, the "within" estimation does not include any dummy coefficients. If they are required, we must compute them using the formula. $d_i{}^* = \bar{y}_{i.} - \bar{x}_{i.}\beta$.

The “between groups” estimate, also identified as group mean regression, makes use of variations amongst specific groups (objects). This estimation, in particular computes the group means of the independent and dependent variables, reducing the frequency of realizations to n. Then, on those converted, combined data, apply OLS: $\bar{y}_{i.} = \alpha + \bar{x}_{i.} + \epsilon_{i.}$ Table 2.4 compares the within group approximation, LSDV method plus the between group approximation.

Table 2.4: Evaluation of the Three Approximation Approaches

	LSDV	**Within Estimation**	**Between Estimation**
Functional Forms	$y_i = i\alpha_i + X_i\beta + \epsilon_i$	$y_{it} - \bar{x}_{i.} = y_{it} - \bar{x}_{i.} + \epsilon_{it} - \bar{\epsilon}_i$	$\bar{y}_{i.} = \alpha + \bar{x}_{i.} + \epsilon_i$
Time invariant variables	Present	Absent	Absent
Dummy variables	Used	Not used	Not used
Dummy coefficients	Computed	Have to be calculated	N/A
Transformations	Not done	Variance from group averages	Group averages
Intercept estimated	Done	Not done	Done
R^2Value	Accurate	Inaccurate	Not of concern
SSE	Right	Wrong	N/A
$\frac{MSE}{SEE}$ (SRMSE)	Accurate	Wrong (often smaller)	N/A
Standard errors	Accurate	Wrong (smaller)	N/A
DF Error	$nT - n - k^*$	$nT - k$(n lager)	$n - k - 1$
Realisations	nT	nT	n

Approximating Random Effects Models

The one-way random effects models includes a composite residual term, $w_{it} = u_i + v_{it}$. Having u_i being undertaken as independent of usual residual term v_{it} and independent variables X_{it}, being likewise autonomous to each other for every i and t. Note that, this supposition is not essential for a fixed effects model. This model stands as:

$$y_{it} = \alpha + X'_{it}\beta + u_i + v_{it}, \tag{2.6}$$

here $u_i \sim IID\ (0, \alpha_u^2)$, and $v_{it} \sim IID\ (0, \alpha_v^2)$.

The covariance entries of $\text{Cov}(w_{it}, w_{js}) = E(w_{it}w'_{js})$ are $\alpha_u^2 + \sigma_v^2$ if $i = j$ and $t = s$ and α_u^2 if $i = j$ and $t \neq s$. Hence, the covariance construction of composite error terms; $\sum = E(w_i w'_i)$ for specific i and the variance-covariance matrix for all errors (residuals) V becomes;

$$\sum_{T\times T} = \begin{pmatrix} \sigma_u^2 + \sigma_v^2 \dots & \dots \sigma_u^2 \dots & \dots \sigma_u^2 \\ \vdots & \dots \sigma_u^2 + \sigma_v^2 \dots & \vdots \\ \sigma_u^2 \dots & \dots \sigma_u^2 \dots & \dots \sigma_u^2 + \sigma_v^2 \end{pmatrix} \tag{2.7}$$

and

$$V_{nT\times nT} = I_n \otimes \sum = \begin{pmatrix} \sum & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sum \end{pmatrix}. \tag{2.8}$$

When the covariance structure of composite errors is known, a random effects model is fitted using GLS (generalized least squares) method and if it is unknown, it is fitted using FGLS (feasible generalized least squares) or EGLS (estimated generalized least squares). Because $\sum$ is frequently unidentified, EGLS / FGLS is more frequently used as compared to GLS. A random effects model is more difficult to assess than the fixed effects equivalent. In FGLS, one first need to approximate θ using σ_u^2 and σ_v^2. The σ_u^2 arises from the between-effects approximation (group means regression) and σ_v^2 is obtained from the SSE (error sum of squares) of the within-effects approximation or the variances of error terms emanating from group means of the error terms.

$$\hat{\theta} = 1 - \sqrt{\frac{\hat{\sigma}_v^2}{T\hat{\sigma}_u^2 + \hat{\sigma}_v^2}} = 1 - \sqrt{\frac{\hat{\sigma}_v^2}{T\hat{\sigma}^2_{\text{between}}}} \tag{2.9}$$

$$\hat{\sigma}_u^2 = \hat{\sigma}^2_{\text{between}} - \frac{\hat{\sigma}_v^2}{T}, \tag{2.10}$$

where

$$\hat{\sigma}^2_{\text{between}} = \frac{SSE_{\text{ between}}}{n-k-1}, \quad \hat{\sigma}_v^2 = \frac{SSE_{\text{ within}}}{nT-n-k} = \frac{e'e_{\text{ within}}}{nT-n-k} = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}(v_{it} + \bar{v}_{i.})^2}{nT-n-k}.$$

Here v_{it} adenotes the error terms in the LSDV. Then, the regressors,the dependent entity and the value of the intercept needs to be converted as below;

$$y_{it}^* = y_{it} - \hat{\theta}\bar{y}_{i.} \tag{2.11}$$

$$x_{it}^* = x_{it} - \hat{\theta}\bar{x}_{i.} \qquad \text{for all } x_k \tag{2.12}$$

$$\alpha^* = 1 - \hat{\theta} \tag{2.13}$$

Lastly, perform OLS on the resulting entities having the usual intercept suppressed.

$$y_{it}^* = \alpha^* + x_{it}^{*\prime}\beta^* + \epsilon_{it}^*. \tag{2.14}$$

Putting Random and Fixed Effects to Test

How can we tell if the panel data have random or/and fixed effects? The F-test is employed to analyse fixed effects whereas the LM (Lagrange multiplier) test is employed to examine random effects. The prior compares a fixed effects models to ordinary least squares models to assess how considerably the fixed effects models increases goodness of fit while the latter contrasts a random effects model with an ordinary least squares models. A Hausman test is used to compare the similarity of fixed and random effect estimatorsBreusch and Pagan (1980).

Testing Fixed Effects Using the F-test

With the equation $y_{it} = \alpha + \mu_i + X_{it}'\beta + \epsilon_{it}$, , the null hypothesis states the following: Completely, dummy variables with the exception of one are all zero, $H_0 : \mu_0 = \mu_1 = \cdots = \mu_{n-1} = 0$. While the alternative hypothesis says; at least one dummy variable is not zero. This proposition is evaluated by way of the F test that is hinged on the loss of goodness-of-fit. The test compares LSDV (vital model) and the pooled OLS (effective model) and scrutinizes the degree that the goodness-of-fit as measured by SSE or R^2 has changed.

$$F_{(n-1,nT-n-k)} = \frac{\left(e'e_{\text{ pooled}} - e'e_{\text{ LSDV}}\right)/(n-1)}{\left(e'e_{\text{LSDV}}\right)/(nT-n-k)} \tag{2.15}$$

$$= \frac{\left(R^2_{\text{LSDV}} - R^2_{\text{pooled}}\right)/(n-1)}{\left(1 - R^2_{\text{LSDV}}\right)/(nT-n-k)}. \tag{2.16}$$

Dismissing the null hypothesis says; at least one time / group specific intercept u_i is not zero, one may infer that the fixed effect model has a substantial fixed effects or a noteworthy growth in goodness-of-fit; hence, the fixed effects model is superior to the pooled OLS.

2.2.8.1 Testing for Random Effects Using Breusch-Pagan Lagrange multiplier Test

Breusch and Pagan (1980) Lagrange multiplier (LM) test checks whether time (or individual) specific variance constituents are zero $H_0 : \sigma_u^2 = 0$. The Lagrange multiplier statistic has a chi-squared distribution having one degree of freedom.

$$LM_u = \frac{nT}{2(T-1)} = \left[\frac{T^2 \bar{e}' \bar{e}}{e'e} - 1\right]^2 \sim \chi^2_{(1)}, \tag{2.17}$$

where $\bar{e}$ is the $n \times 1$ vector of the group means of pooled regression residuals, and $e'e$ is the SSE of the pooled OLS regression.

Baltagi and Chang (1994a) expresses the equivalent LM test differently as follows:

$$LM_u = \frac{nT}{2(T-1)} \left[\frac{\sum\left(\sum e_{it}\right)^2}{\sum\sum e_{it})^2} - 1\right]^2 = \frac{nT}{2(T-1)} \left[\frac{\sum\left(\sum e_{it}\right)^2}{\sum\sum e_{it})^2} - 1\right]^2 \sim \chi^2{}_{(1)}. \tag{2.18}$$

Rejecting the null hypothesis, one may infer that the panel data contains a substantial random effect and that the random effect model handles heterogeneity better than the pooled OLS.

2.2.8.2 Comparing Fixed and Random Effects; Use of Hausman Test

How can we tell which effect (fixed or random) is more significant and useful in panel data?. Hausman (1978) contrasts fixed and random effect models under the null hypothesis that individual effects are uncorrelated with any regressor in the model. LSDV and GLS are consistent when the null hypothesis of no association is not dishonored; or else, LSDV is reliable while GLS is unreliable and biased. Under the null hypothesis, the approximations of GLS and LSDV must not vary systimatically. "The covariance of an efficient estimator with its difference from an inefficient estimator is zero," according to the Hausman test. (Greene 2008, Green, 2003).

$$LM = (b_{\text{ LSDV}} - b_{\text{ random}})' \hat{W}^{-1} (b_{\text{LSDV}} - b_{\text{ random}}) \sim \chi^2_{(k)}. \tag{2.19}$$

Here $\hat{W} = \text{Var}[b_{\text{ LSDV}} - b_{\text{ random}}] = \text{ Var}(b_{\text{ LSDV}})\ \text{Var}(b_{\text{ random}})$ gives the difference in the projected covariance matrices of LSDV and GLS. Remember to leave out the intercept and dummy variables while computing. This test statistic has k degrees of freedom and follows the chi-squared distribution.

According to the formula, a Hausman test determines if "the random effects estimate is trivially different from the unbiased fixed effect estimate." Kennedy (2008). If we reject the null hypothesis of no association, we may conclude that individual effects u_i are highly linked to at least one dependent variable in the model, and hence the random effect model remains problematic. As a result, we must choose the fixed effects model over a random effects model. The difference of covariance matrices W may not be positive definite; hence, we may argue that the null is not rejected assuming similarity of covariance matrices makes such a difficulty Greene (2003).

Testing for Poolability using Chow Test

Poolability inquires whether slopes are constant over groups or across time Baltagi and Chang (1994b). The Chow test Chow, 1960 is an expansion of a simple form of the poolability test. The null hypothesis in the Chow test says that the gradient of a dependent variable remain constant for all k regressors independent of the individual. $H_0 : \beta_{ik} = \beta_k$. Need is there to remember that gradiants remain fixed for random and fixed effects models; only error variances and intercepts matter.

$$F_{[(n-1)(k+1),n(T-k-1)]} = \frac{\left(e'e - \sum e_i'e_i\right)/(n-1)(k+1)}{\left(\sum e_i'e_i\right)/n(T-k-1)}, \tag{2.20}$$

where $e'e$ represents the SSE of the pooled OLS while $e_i'e_i$ is the SSE of the pooled OLS for group i. Rejecting the null hypothesis says that, longitudinal data at hand cannot be poolable since every entity has its own slope for all explanatory variables. In this case, one may use the random coefficient model or the hierarchical regression model. The Chow test is based on the assumption that specific residual variance entities have a normal distribution $\mu \sim N(0, s^2 I_{nT})$.When this assumption is not met, the Chow test may fail to appropriately assess the null hypothesis, according to Baltagi and Chang (1994a). According to Kennedy (2008), if there is reason to believe that errors in different equations have different variances, or that there is contemporaneous correlation between the equations errors, such testing should be undertaken using the Stein's unbiased risk estimate (SURE) is an unbiased estimator of the mean-squared error of; a nearly arbitrary, nonlinear biased estimator, not OLS; inference with OLS is unreliable if the variance-covariance matrix of the error is nonspherical Baltagi and Chang (1994a).

2.2.9 Selecting the Model: Fixed or Random Effect?

We obtain 12 potential panel data models when we combine random against fixed effects, time against group effects and one-way against two-way effects, as shown in 2.5. Generally, one-way models are frequently utilized as a result of their parsimony and fixed effects models are easier to estimate and explain than a random equivalent. It is, nevertheless, difficult to choose the finest model among the following 12.

Table 2.5: Grouping of Panel Data Analysis Approaches

	Type	**Fixed Effects**	**Random Effects**
One-way	Group	One-way fixed group effects	One-way random group effects
	Time	One-way fixed time effects	One-way random time effect
Two-way	Double times*	Two-way fixed group effects	Two-way random group effects
	Two times*	Two-way fixed time effects	Two-way random time effects
	Mixed	Two-way fixed group and time effects	Two-way random group time effects
		Two-way fixed time and random group effects	
		Two-way fixed group and random time effects	

Practical Meanings of Random and Fixed Effects

Official tests presented in 2.2.9 scrutinizes the existence of random or/and fixed effects. Particularly, F- test contrasts fixed effects models and the OLS (pooled), while LM test compares a OLS with random effects model. Hausman specification test contrasts random with fixed effects model. Nevertheless, the above tests do not give practical connotations of random and fixed effects. What is the meaning of the fixed effects? How does one construe the random effects practically? This is the humble and straightforward response. Assume one is regressing the production of companies for example Econet, NetOne, Telecel, and Africom on their research and development (R and D) investments. A fixed effect may be understood as the starting production capacities of the establishments with no R and D investments; every company has its own opening production level. The random effects may be regarded as a sort of uniformity or constancy in production. When the production of an entity oscillates up and down considerably, as an example, if its production is unstable (or the variance element is larger than those of other corporations) regardless of the fact that its productivity (gradient of R and D) is constant across company. Kennedy (2008) offers a hypothetical and insightful explanation of random and fixed effects. Either random or fixed effects is an issue of unobserved entities or absent relevant objects that makes pooled OLS biased. The heterogeneity can be controlled by either introducing dummies to approximate specific intercepts of entities (objects/groups) or tacking the diverse intercepts as emanating from a container of likely intercepts, thus they can be viewed to be random and taken as if they were sections of the residuals" ; being fixed effects models and random effects model, correspondingly. A random effects model has "composite error terms" that are composed of the customary random error and a "random intercept" determining the level that individuals intercept diverges from the overall intercept Kennedy (2008). Furthermore, Kennedy (2008) argues that the main difference between random and fixed effect models is not whether overlooked heterogeneity is accredited to the intercept or variance components, but whether the individual specific error component is related to independent variables. It is a respectable custom to draw plots of the dependent and regressor variables prior modelling panel data.

Two Recommendations for Longitudinal Data Modeling

Firstly, like any other data exploration procedures, we define the data at hand cautiously prior to any analysis. Even though frequently overlooked by several data analysers, the data narrative is of importance and of much use to investigators to get insights of the data and exploration approaches. For panel data exploration, features and quality of panels influences model choice considerably.

1. Clean data by scrutinizing consistence and reliability of the way there were measured. If various time phases were employed in a long panel, for instance, attempt to aggregate (reorganize) data to increase uniformity. Having many omitted values, make a decision if you go for a balanced panel by discarding other pieces of functional statistics or retain every usable realizations in an unbalanced panel to the cost of computational as well as procedural challenges.

2. Check the features of the data together with the quantity of individuals (objects), the quantity of time phases, rotating against fixed panels and unbalanced against balanced panel. Thereafter, attempt to find models suitable for these properties.

3. Be cautious if having "short" or "long" panel data. Visualize long panel that has 20 thousand time phases but 5 entities or a short panel with 3 years $\times$ 8000 (companies).

4. When n or/and T are excessively large, make an attempt to reorganize entities or/and time phases and obtain other convenient n' and T'. A null hypothesis of $u_1 = u_2 = \ldots = u_{999999} = 0$ for a fixed effects model, for example, is virtually of no use. Attempt to use annual data instead of weekly data or monthly values instead of daily data.

Secondly start by a simpler models. Fit a pooled OLS instead of a random or fixed effects model; a one-way effects model instead of a two-way model; a random or fixed effects model before a hierarchical linear formulation; and so on. Never attempt a complicated models, of course, fancy models that the panel data may not support. (e.g., ill structured panels and short /long panels).

Guiding Principles of Model Selection

At the model fitting phase, let us start by pooled OLS and formerly reflect analytically about its possible challenges if detected and undetected heterogeneity (a set of omitted but significant entities) is not considered. Moreover reflect on the basis of heterogeneity (i.e., time series or cross-sectional entities) to establish discrete (group or entity) effects or time effects. If one thinks that the specific heterogeneity is taken care of by the residual terms and the group (endity or time) effects are unconnected to any independent variables, attempt a random effects model. When the heterogeneity may be accounted for by individual particular intercepts and the individual effects can probably be interrelated with any other independent variables, attempt a fixed effects model. When a specific group or individual possesses peculiar initial capacity and has equal error variance with other entities, a fixed effects model is preferred. When every entity has a specofic residual term, a random effects will be more appropriate at revealing heteroskedestic residuals. Subsequently, perform suitable official tests to investigate specific time or/and group effects. Given that the null hypothesis in the LM test is rejected, a random effects model is superior compared to the pooled OLS. In the case that the null hypothesis of the F -test is rejected, a fixed effects model is preferred compared to OLS. When both hypothesis are not rejected, then proceed by fitting pooled OLS. Perform Hausman test if both hypotheses of the F- tests and LM test are entirely rejected. When the null hypothesis of no association between entity effects and independent variables is rejected, opt for the vigorous fixed effects model; if not, remain hinged on the proficient random effects model. When one has a resilient feeling that the heterogeneity encompasses two time series, one time series or two cross-sectional and one cross-sectional variables, go for two-way effects models. Verify if panels are well-structured, and n and $T$are huge enough; never try a two-way model when having badly structured, seriously unbalanced and/or excessively short/long panel. Perform suitable F- tests and LM tests to look for the availability of two-way effects. Lastly, when you believe that the heterogeneity involves slops (parameter approximations of independent variables) fluctuating over time and/or individual. Perform a Chow test or a comparable test to scrutinize poolability of the panels. Once the null hypothesis of poolable data is rejected, attempt a hierarchical linear model or a random effects model.

2.2.10 Pooled OLS and Least Squares Dummy Variable (LSDV)

This subdivision starts with classical least squares technique referred to as OLS (ordinary least squares) and clarifies how ordinary least squares handles overlooked heterogeneity by the use of dummy variables. A dummy variable can be taken as twofold variable that is coded by either zero or one. OLS with dummies is known as a LSDV (least squares dummy variable) model or hierarchical linear model.

Pooled Ordinary Least Squares

The ordinary least squares is a pooled linear regression without random or/and fixed effects. It undertakes a fixed intercept and gradients despite group and time phases. The pooled ordinary least squares hypothesizes no variance in intercept and gradients over entities and time periods.

Approximating Strategies: LSDV1, LSDV2, and LSDV3

The LSDV regression is the OLS with dummies. The major concern in least squares dummy variable is how to circumvent the perfect multicollinearity or that entitled "dummy variable trap." Each methodology has a restriction (constraint) which decreases the frequency of parameters to be approximated by one and consequently enables the model to be identified. LSDV1 drops one dummy; LSDV2 subdues the intercept then LSDV3 inflicts a constraint. These approaches are different from each other with respect to model estimation and interpretation of dummy variable parameters Suits (1984). They yield dissimilar dummy parameter evaluations, but their outcomes are comparable. One need to recognize the advantages and disadvantages of these three methodologies.

2.2.10.1 Approximating Least Squares Dummy Variable One

LSDV1 drops a dummy variable. Thus, the parameter of the dropped dummy variable stands as set to zero and is used as a basis. We ought to be cautious when choosing an entity or variable to be eliminated, $d_{\text{dropped}}^{\text{LSDV 1}}$, with the intention of making it play a role of the reference group efficiently. The coefficient of a dummy counted in shows how far its parameter approximate value is relative to the baseline or point of origin (i.e., the general intercept). Suppose one drops an alternate dummy variable, say $g_{2,}$ instead of g_3? As an alternative baseline is used, one get diverse dummy coefficients. However other indicators such as parameter statistics of explanatory variables and goodness-of-fit measures stay the same. That is, selection of a dummy variables to be suppressed does not alter the model at all.

2.2.10.2 Approximating Least Squares Dummy Variable Two

LSDV2 encompasses all dummy variable and consecutively, drops the intercept (i.e., set the intercept to zero). We can fit LSDV2 using different commands which suppresses the intercept in the model.

2.2.10.3 Approximating Least Squares Dummy Variable Three

LSDV3 encompasses the intercept and all dummy variables and thereafter imposes a constraint that the sum of parameters of all dummy variables is zero. LSDV3 maintains the same parameter approximations of dependent variables and the customary residuals as do LSDV1 and LSDV2. Not like LSDV1 and LSDV2, LSDV3 yields the intercept and dummy coefficients but the coefficients have varied meanings. The LSDV3 intercept gives the mean of specific group intercepts, whereas a dummy coefficient represents the deviances of the group intercepts from the averaged intercepts.

2.2.10.4 Approximating Least Squares Dummy Variable One, Two, and Three

Three methodologies end up approximating the same model and present the same parameters of dependent variables and respective usual residuals. LSDV3 and LSDV1 presents accurate goodness of fit measures whereas LSDV2 presents true root MSE and SSE however yields incorrect (exaggerated) R^2 and F-test. The three LSDV methodologies yield diverse, but comparable dummy coefficients. The main dissimilarity in the three methodologies is that of significances of the dummy coefficients and intercept. A parameter approximation in LSDV2, σ_d^*, is the Y-intercept (actual intercept) of group d. It is simple to interpret practically. The t-test scrutinizes whether σ_d^* remains zero. For LSDV1, a dummy coefficient indicates the magnitude by which the real intercept of group d diverges from the point of origin (parameter of the surpressed dummy), being the intercept of LSDV1, $\sigma_{\text{dropped}}^* = \alpha^{\text{LSDV1}}$. The null hypothesis in t-test is that, there is no deviation with respect to the reference group. In LSDV3, a dummy coefficient means how far its real parameter is away from the average group effect Suits (1984). The LSDV3 intercept refers to the averaged effects: $\alpha^{\text{LSDV3}} = \dfrac{1}{d}\sum \sigma_i^*$. Thus, the null hypothesis is, there is no deviation of a group intercept from the averaged intercepts. In brief, every methodology has a unique reference point and constraint and hence tests unique hypothesis. But all methodologies produces comparable dummy coefficients and precisely identical parameter estimates of independent variables. Alternatively, they all proposes the same models; given one LSDV produced, that is, one can repeat the other two LSDVs. Which methodology gives better results? One needs to deliberate on both approximation and explanation issues cautiously. Generally, LSDV1 is oftenly preferred since it is easy to estimate using different packages. Often investigators want to check how far dummy parameters digress relative to reference groups instead of the exact group intercepts. When one has to present specific group intercepts, LSDV2 provides the solution openly. Lastly, LSDV2 and LSDV3 have some approximation challenges; for instance, LSDV2 presents a wrong R^2 value.

2.2.11 Reporting Panel Data Models

The main question is, "Which information must one report on? And in what way?" Other researchers present parameter approximations plus their statistical implications only; while some take account of standard residuals and disregard goodness of fit measures. Frequently investigators are unsuccessful on interpreting findings substantively to readers. This subsection looks at common guiding principle of reporting longitudinal data models. Though, particular pieces of information to be focused on and the approaches hinge on the study objectives, questions, and purposes of the research.

Do we Report on all Possible Models? No!

Other researchers report on all feasible models of the pooled ordinary least squares, fixed effects models, random effects models and two-way effects model. When one model is "right," others are "wrong." It will be illogical to report on incorrect models together except when comparison of models is one of the objectives of the research. Why should one try reporting on the "incorrect" model? Simply, one needs to just present the "correct" model or the final model only.

What Information Should be Presented?

One should present on parameter approximations, goodness-of-fit measures and their typical residuals and findings from tests.

2.2.11.1 Goodness-of-fit Tests

Goodness-of-fit tests scrutinize to what extent a model fits data. When having an inadequate goodness-of-fit, one needs to attempt other models, The critical goodness of fit tests on which one needs to report on are;

1. R^2 in OLS and fixed effects models.

2. F-tests or likelihood ratio tests to check models and the relative significance (p-values).

3. Error sums of squares (residuals), degrees of freedom for residuals and nT (N).
4. θ plus variance components $\hat{\theta}_u$ approximated for a random effects model.

Remember other approximation procedures present wrong statistics and standard residuals. As an case of, R^2 for fixed effects model as the command fits a "within" estimator. Together the overall R^2 and the between R^2 shown on the output are virtually worthless. So as to obtain the correct R^2 in a fixed effects model, use LSDV1 .Make use of macro variables, if necessary, to get numerous goodness-of-fit measures which are not exhibited in the results.

2.2.11.2 Parameter Approximations of Regressors

One needs to present parameter estimates and their standard errors. Fortunately, most statistical commands on statistical and or mathematical packages yield accurate parameter approximations and the respective adjusted standard residuals. Nevertheless the "within" approximation itself yields faulty standard residuals due to larger (wrong) degrees of freedom.

2.2.11.3 Parameter Approximations of Dummy Variables

For a fixed effects model, the issue is if specific intercepts must be presented. Generally, parameter approximations of independent variables are of prime concern in most situations and consequently specific intercepts are not necessary. Nevertheless, one has to present them if listeners need to know or specific effects are of key importance in the research. The amalgamation of LSDV2 or LSDV1 gives one stress-free clarifications for this case. Remember that LSDV3, LSDV1, and LSDV1 takes diverse implications of dummies and that null hypotheses of t-test varies from each other.

2.2.11.4 Testing for the presence of random or/and fixed effects

Lastly,one should present if random or/and fixed effects are present because panel data modelling implies examining random or/and fixed effects. Present and explain the findings from F-test for a fixed effects models or/and Breusch-Pagan LM test in case of random effects models. If together random and fixed effects are statistically noteworthy, one needs to perform a Hausman test and present the results. When we are uncertainty if slopes are constant over time or/and groups, perform a Chow test to check for the poolability of data.

Interpreting Outcomes Applicably

When our model paroxysms the data appropriately and individual regresants are statistically significant, we have to explain parameter estimate values in a "meaningful" way. We might not just present directions and sizes of coefficients. We need not to just say, for instance, a regressor is "significant," "positively (or negatively) associated with", or "trivially correlated to..." A typical form of presenting is, "For a unit growth in **A**, **B** is anticipated to change with x units, taking all other entities constant." We may perhaps overlook the ceteris paribus postulation (taking every other variable as fixed). Nevertheless, need is there to make interpretations more logical and sensible for the audience who may not considerably understand much of Statistics, economics or econometrics. Avail statistical significance in form of a tabulation and the p-values in addition on ending of the interpretation statement.

Professional Presentation of Findings

Numerous researchers often report findings in tabular form, nonetheless others fail to come up with professional tables. Generally poor tables include the following features i) colourful and stylish borders, ii) too large and/or too small numbers, ii), large and various font sizes 4) poorly aligned figures(numbers) and 5) non-regular arrangement. Below are reference points to be taken into consideration while creating a professional table.

1. Title needs define the contents of a table suitably. Give a units of measurements (for example; Thousand Dollars) and time (e.g., Year 2020) when needed.

2. Use simple frames.

3. Arrange a tables compactly and systemically.

4. Deliver parameter statistics together with their standard errors.

5. Make use of names of variable that are not utilized in computer soft wares as labels. For example, instead of load, use loading factor.

6. Apply ten point Courier and ten point Times New Roman for labels and different for numbers. Employ simple fonts and do not use too small or too big font sizes.

7. Resize numbers suitably in an attempt to circumvent numbers such as "0.0000337755" or "85,745,341,698,875."

8. Present numbers up to four or three significant figures Never round-off numbers indiscriminately.

9. Use of horizontal and vertical lines must be minimized. Do not use vertical line generally.

10. Align figures to the right and reflect on the position of decimal points cautiously.

11. Utilise "Regular coefficients,"when there is need, instead of "Beta," or "β," coefficients." The actual value of β is not known.

12. Specify statistical significance as $* < 0.01, ** < 0.05$.

13. Supply the data source, if relevant, at the bottom of the table.

Hypothesis

A hypothesis is a speculation or conjecture with regard to an unknown (e.g,β, α, σ and δ). Consequently, α_1" is an invalid hypothesis, however $\alpha_1 = 0$ is. Since α_1 is already identified (approximated from the sample), we do not need to test if $\alpha_1 = 0$.

Parameter Approximations

Report, "parameter estimations of β_1" or " coefficient of regressor 1" in place of "The coefficient of β_1" Furthermore report, "standardised coefficients" in place of of "Beta," β, or "beta coefficient."

P-Value

Avoid saying, "The p-value is significant." P-values on their own are neither insignificant nor significant. We can report that, "A small p-value advocates for rejection of H_0." or "The p-value is sufficiently small to reject H_0"

Do Not Reject or Reject a Null Hypothesis

Pronounce, "do not reject" or "reject" the null hypothesis instead of confirm (or accept)" the null hypothesis. Furthermore instead of, "We do not have faith in the H_0 ", pronounce ;" we reject the H_0 at the 0.01 level" or "The test gives conclusive confirmation that the H_0 is incorrect" (nobody recognizes whether H_0 is actually wrong or true). At all times be clear and simple.

2.3 Machine Learning Modelling

Customary forecasting procedures often provide poor macro forecasts.

Procedures founded on OLS (ordinary least squares) have difficulties in overcoming numerous issues, comprising of predictor relevance, collinearity, nonlinearity and dimensionality. Some advanced forecasting models, encompassing dynamic factor models, may assist in addressing dimensionality and collinearity challenges, and nevertheless does not explain predictor significance and nonlinearity challenges. Consequently, even high-technology forecasting models frequently yields large prediction residuals. Additionally, if the entity to be forecasted is volatile, dynamic factor models gives poor results particularly, such as GDP growth in most developing markets and emerging economies.

Machine learning (ML) approaches gives an alternative to customary estimating practices.

ML models can outpace customary estimating techniques since they focus more on out-of-sample instead of within-sample) performance and considers nonlinear connections among a big number of predictors better. Machine learning procedures are precisely intended to learn composite associations from historical data whereas repelling the propensity of customary techniques to over-generalize historical relationships into the future. Indeed, a literature is beginning to emerge which suggests that machine learning approaches often perform better than customary linear regression-grounded approaches with reference to accurateness and healthiness

Advantages of Machine Learning Techniques

Not like customary forecasting procedures, ML techniques are explicitly intended to optimize the bias-variance trade off. Particularly, machine learning models may account for the issues where customary forecasts have had challenges as they choose predictors to optimize extrapolation (instead of within-sample) performance and handles nonlinear relations within a big number of predictors better. For this research we emphasize on specific machine learning techniques as below:Deep learning; Gradient Boosted Trees; Support Vector Machines and Random Forest.

2.3.0.1 Deep learning

There exist a scarcity of texts that focus directly on modelling of dynamic longitudinal data by means of deep learning approaches. In the midst of the rare current writings are those of Reichstein et al. (2019); Raissi et al. (2019); Ye et al. (2019)and Raissi (2018). However the first three writers only used deep learning techniques in the solutions of solution of data-driven discovery, data-driven and nonlinear partial differential equations challenges of partial differential equations and solution in process understanding of data-driven earth system sciences correspondingly ,only Hu and Szymczak (2023) used deep learning for analyzing of non-stationary longitudinal data. The literatures of the writers were the ground-breaking studies in the use of deep learning in exploring significant factors from non-stationary longitudinal data. On a multi-dimensional, high-resolution spatiotemporal dataset, the writers used the method of Long Short-Term Memory deep learning neural networks that runs on a high-performance computing cluster. They realized that Deep learning procedures significantly outpaced cutting-edge spatial-econometric model at grid-cell, continental and state levels. Consequently, this study implemented convectional and non-stationary longitudinal models with the aim of researching on the nature and level of association of the deliberated macro-economic variables on the economic advancement over the SADC nations in an attempt to discover any policy consequences.

2.3.0.2 Random Forests

Random Forests is a machine learning algorithm that can be used for both regression and classification tasks. It is an ensemble learning method that combines multiple decision trees to improve generalization performance and reduce overfitting. The basic idea behind Random Forests is to create a large number of decision trees using a random subset of the data and a random subset of the variables. Each decision tree in the forest is built independently using a random sample of the data and a random subset of the variables. The final prediction is made by averaging the predictions of all the trees in the forest. Random Forests have several advantages over other machine learning algorithms. They are very good at handling high-dimensional data and can handle both numerical and categorical variables. They are also robust to outliers and missing data and can handle nonlinear relationships between variables.One of the key benefits of Random Forests is that they can provide measures of variable importance. The algorithm can calculate the importance of each variable in the model by measuring the decrease in accuracy when the variable is removed from the model. This can be useful for identifying the most important predictors in a dataset and for understanding the underlying relationships between variables.

Random Forests can be used in a wide range of applications, including finance, healthcare, and natural language processing. They are particularly useful for datasets with a large number of variables, such as genetic data or image data. However, it's important to properly validate the model to ensure that it is accurate and reliable, and to choose appropriate hyperparameters to optimize performance.

2.3.0.3 Gradient Boosting.

Gradient Boosting is a machine-learning technique that can be used for both regression and classification tasks. It is an ensemble learning method that builds a sequence of decision trees, where each tree is trained to correct the errors of the previous trees. The basic idea behind Gradient Boosting is to start with a simple model, such as a single decision tree, and then iteratively add more trees to the model. Each new tree is trained to predict the errors of the previous trees so that the ensemble of trees gradually improves its performance over the iterations.

The “gradient” in Gradient Boosting refers to the fact that the algorithm minimizes a cost function by computing the gradient of the cost function with respect to the predictions of the previous trees. This gradient is used to update the predictions of the model in order to reduce the cost function.

Gradient Boosting has several advantages over other machine learning algorithms. It can handle high-dimensional data and can handle both numerical and categorical variables. It is also robust to outliers and missing data and can handle nonlinear relationships between variables.

One of the key benefits of Gradient Boosting is that it can provide measures of variable importance. The algorithm can calculate the importance of each variable in the model by measuring the decrease in the cost function when the variable is removed from the modelRichardson et al. (2021). This can be useful for identifying the most important predictors in a dataset and for understanding the underlying relationships between variables.

Gradient Boosting can be used in a wide range of applications, including finance, healthcare, and natural language processing. It is particularly useful for datasets with a large number of variables, such as genetic data or image data. However, it’s important to properly validate the model to ensure that it is accurate and reliable, and to choose appropriate hyperparameters to optimize performance.

2.3.0.4 Support Vector Machines

Support vector machines (SVMs) are a popular machine learning technique that has been widely used in various applications, including regression, classification, and clustering. SVMs are well-suited for high-dimensional and complex data, and can handle both linear and nonlinear relationships between the input variables and the target variable Maccarrone et al. (2021).

In recent years, SVMs have also been applied in panel data analysis, which involves analyzing data that are collected over time and across a group of individuals or entities. Panel data analysis can provide valuable insights into the dynamic behavior of the variables and the heterogeneity of the individual effects.

One of the advantages of SVMs in panel data analysis is their ability to handle unbalanced and missing data, which are common in longitudinal studies. SVMs can also incorporate various kernel functions, such as linear, polynomial, and radial basis function kernels, to capture different types of nonlinearity and heterogeneity Richardson et al. (2021).

Several studies have applied SVMs in panel data analysis and found promising results. For example, Chiu et al. Chiu et al., 2011 used SVMs to predict the survival of breast cancer patients based on their gene expression profiles over time. They showed that SVMs can achieve higher accuracy and robustness than traditional survival analysis methods.

In another study, Kim et al. Kim et al. (2009) used SVMs to predict the stock returns of companies based on their financial ratios over time. They found that SVMs can outperform traditional time series models, such as autoregressive integrated moving average (ARIMA) models, in terms of forecasting accuracy and stability.

Other studies have applied SVMs in panel data analysis for various purposes, such as predicting customer churn Yang and Ching (2008), detecting fraud Liang and Huang (2011), and forecasting electricity consumption Wang et al. (2014). These studies have demonstrated the potential of SVMs in panel data analysis and highlighted their ability to capture complex and heterogeneous patterns in the data.

2.4 Conclusion(s)

Longtudinal data are examined to analyse group (time) or/and individual effects by means of random effects and fixed effects models. The fixed effects model enquires how heterogeneity from time or/and group affects specific intercepts, whereas the random effects models assumes residual variance structures are influenced by group or/and time. Error terms in a random effects models are anticipated as arbitrarily spread over times or groups. However the primary difference between random and fixed effects models is that individual effects u_i in random effects models should be uncorrelated to any independent variable. Slopes are expected to be constant in both random effects and fixed effects models. A Longitudinal data set ought to be organized in the long form. Panel data are fixed or rotating, balanced or unbalanced and short or long. When data is extremely unbalanced; excessively short or too long, read output cautiously and, if having an unbalanced panel, contemplate living out entities having numerous missing observations. When the quantity of time phases or subjects (individuals) is exceedingly big, consider classifying objects to decrease the quantity of groups or time periods. Fixed effects models are fitted using the LSDV regression and "within" approximation. LSDV comprises of three methodologies to circumvent perfect multicollinearity. LSDV1 lives out the dummy; LSDV2 subdues an intercept while LSDV3, enacts a constraint and takes in all dummies instead. LSDV1 generally is used as it yields accurate statistics. LSDV2 gives authentic individual intercepts, however presents inappropriate R^2 and F-values. Bear in mind that the dummy parameters from the three LSDV methodologies have diverse meanings and accordingly perfom different t-tests.

The "within" approximation does not use of dummies however uses deviances from group averages. Hence, this approximation is valuable if there are numerous time periods or/and groups in the longitudinal data set as it is able to circumvent the incidental parameter challenges. Consecutively, time -invariant regressors are eliminated in the data transformation procedure and the pseudo parameter estimates must to be determined thereafter. Owing to its bigger degrees of freedom, the "within" approximation gives inappropriate R^2 and standard residuals of parameters even though Statistical pakagies presents adjusted standard residuals.

Table 2.6: Classification of Panel Data Analysis

Fixed effects (F-test)	**Random effects (Breusch-Pagan LM test)**	**The Selection**
H_0Rejected (Absence of fixed effects)	H_0Not rejected (Random effects absent)	Pooled Ordinary Least Squares
H_0 rejected (Presence of Fixed effects)	H_0Not rejected (Random effects absent)	Fixed effect model
H_0Not rejected (Fixed effects absent)	H_0 rejected (Presence of random effects)	Random effect model
H_0 not rejected (Fixed effects present)	H_0Rejected (Presence of random effect)	Choose the fixed effects model when the null hypothesis of a Hausman test is rejected; otherwise, go for a random effects model.

So as to determine a suitable model for longitudinal data, initially describe data prudently using summary statistics also draw plots. Then start with a humble model such as the pooled ordinary least squares. When both null hypotheses of LM test and F-test are not rejected, our most appropriate model becomes the pooled ordinary least squares. When the null hypothesis in a Breusch-Pagan LM test in a random effect model is not rejected and the null of an F-test for a fixed effects model is, the fixed effects model becomes the choice. When we realise both substantial random and fixed effects from our longitudinal data, perform the Hausman specification test that contrasts a random effects model with a fixed effects model. When the null hypothesis of no association between regressors and individual effects is rejected, fit the random effects model. Or else, a fixed effects model is favoured. If one thinks that the data is unpoolable and every variable has distinct slopes for independent variables, perform the Chow test and then, if the null hypothesis is rejected, attempt fitting the hierarchical linear model or random effects model. It of importance to report the findings appropriately. Essential information to be presented encompasses goodness-of-fit measures (for instance; likelihood ratio and F-score, R^2 and SSE), parameter estimations and respective standard residuals and test results (that is, LM, Chow test, F-test, test and the Hausman test). These portions of information have to be reported in tabular form professionally. Researchers must interpret the findings in simple terms to enable readers or listeners with limited econometric knowledge can be carried along and possibly understand.

2.5 Chapter Summary

This chapter provides a review of the existing literature on panel data analysis and machine learning. The discussion on general linear models and panel linear models is presented in detail. Additionally, the literature on the use of machine learning to gain a better understanding of economic variables is also reviewed. The chapter also includes a review of fixed and random effects models.

Chapter 3

Data and Methodology

3.1 Introduction

This chapter will provide a full account of the procedures used to conduct panel data analysis on my data. Specifically, It will articulate the research design, including the sample selection and the variables included in the analysis. Additionally, It will also outline the data collection methods, including the sources of the data and any cleaning or pre-processing procedures that were used.

It then provide a comprehensive explanation of the panel data analysis techniques employed, including the estimation methods and model specifications. This allows for a thorough understanding of the statistical methods used to analyse the data. Moreover, a discussion on any limitations or ethical concerns associated with our study will be provided. By doing so, this ensures that the study is conducted with transparency and rigour and that the conclusions drawn from the analysis are reliable.

3.2 Data

In this study determinants of economic growth are gathered and evaluated for sixteen SADC countries for twenty two years, that dictates use the panel data analysis, whereas panel data may have group effects, time effects or both. Data is taken from various sources but mainly The World Bank (2023) website for different SADC countries contributing in the world economy.All monetary values are in million USDs and all rates are percentages.Data analysis was done mainly in R Team (2023).

3.3 Econometric Models

Among other things, this study will look at whether unemployment influences the SADC countries' economic progress. In order to attain this goal, we first construct a production function framework that reflects production and is a good proxy for economic development. Assume variable factors of production only influence an economy's output level and the model as presented by Tiwari & Mutascu in 2011 as follows:

$$Y = f(L, K) \tag{3.1}$$

Where, Y denotes output level (i.e. GDP), L is labour amount (Labour force) and K designates capital which is the Gross Capital Formation), it can be stated that increases in employed labour and capital are responsible for increasing any economy's output level. Following the preceding (equation (3.1)), the production function is expanded in accordance with growth theory , this production function is expanded according to the growth theory Barro (1995); Tiwari and Mutascu (2011). We expanded the model for our investigation by incorporating the additional explanatory factors. The model would be expressed as:

$$\mathrm{GDP}_{it} = f(\mathrm{IMP}_{it}, \mathrm{EXPO}_{it}, \mathrm{INTR}_{it}, \mathrm{LF}_{it}, \mathrm{UNEM}_{it}, \mathrm{ED}_{it}, \mathrm{INF}_{it}, \mathrm{FDI}_{it}, \mathrm{IR}_{it}, \mathrm{EXR}_{it}) \tag{3.2}$$

Where,

IMP_{it} = Imports

INTR_{it} = International Reserves

GDP_{it} = Real GDP

ED_{it} = External Debt

FDI_{it} = Foreign direct investment, net inflows

EXR_{it} = Exchange Rate

EXPO_{it} = Exports

LF_{it} = Labour Force

UNEM_{it} = Unemployment rate

INF_{it} = Inflation rate

IR_{it} = Interest rate

The assumption of U_{it} is that $U_{it} \approx IID(0, \delta_u^2)$, i.e. residuals are independently identically distributed with mean zero and constant variance δ_u^2. Here i stands for a specific nation and t stands for a certain time period. Three approaches can be employed to analyse empirical longitudinal data. These are; ordinary least squares, random effects, and fixed effects models, as well as the LSDV (least squares dummy variables). In 2011, Akbar et al. (2011), employed FEM, REM and OLS to estimate GDP per capita for nine (9) Asian nations. The empirical standard technique assumes that OLS is used to estimate regression equations, with the assumption that omitted variables are uniformly distributed and independent of regressors. As a result, this form of estimating may pose an interpretive issue when we wish to investigate country-specific features such as policy changes, political administrations and good governance, which impact on economic growth rate but are not taken into account in the estimation process. Thus we will conduct our methodology by way of FEM. The Hausman (1978) test answers this question of comparing the FEM and REM. The test scrutinizes whether country specific effects are associated with other explanatory variables, then REM violates the assumptions of Gauss-Markov and is now not considered as a BLUE (best linear unbiased estimator). This is so since country effects are only the part of the residuals of a REM. But if country effects were a part of intercept and correlation amongst regressors and intercept would not violate the assumptions of Gauss-Markov, then a FEM would be still BLUE.

3.4 Group Effects when all Coefficients are Constant across Countries and Time

Of primary interest is to investigate how selected specific variables influence the economic growth in SADC countries. The standard model in order to assess the group effects, in which every coefficient is constant through time and states, would be presented as:

$$\begin{aligned} \mathrm{GDP}_{it} =& \beta_0 + \beta_1 \mathrm{LF}_{it} + \beta_2 \mathrm{UNEM}_{it} + \beta_3 \mathrm{ED}_{it} + \beta_4 \mathrm{INF}_{it} + \beta_5 \mathrm{FDI}_{it} + \beta_6 \mathrm{IR}_{it} + \beta_7 \mathrm{EXR}_{it} + \beta_8 \mathrm{IMP}_{it} + \\ & \beta_9 \mathrm{EXPO}_{it} + \beta_{10} \mathrm{INTR}_{it} + U_{it}, \end{aligned} \tag{3.3}$$

where U_{it} is the error term.
In this situation, the fitted model shows that the intercept values are the same for all nations or country objects. Furthermore, for all the sixteen countries, the slope coefficients of all independent variables are constant. As a result of the highly restrictive assumptions in the preceding equation 3.3 , the real picture of the model may be distorted. As a result, we need to determine the country effects of various nations, as detailed in the coming section.

3.5 Intercept Varies Across Countries while Slope Coefficients are Constant

To the uniqueness of each nation, suppose intercept varies by nation but slope coefficients of particular nations are assumed still fixed. If there is a situation that error term and independent variables are correlated then LSDV approach may be inappropriate Gujarati and Porter (2003). To see this model would be of following format:

$$\begin{aligned} \mathrm{GDP}_{it} =& \beta_{0i} + \beta_1 \mathrm{LF}_{it} + \beta_2 \mathrm{UNEM}_{it} + \beta_3 \mathrm{ED}_{it} + \beta_4 \mathrm{INF}_{it} + \beta_5 \mathrm{FDI}_{it} + \beta_6 \mathrm{IR}_{it} + \beta_7 \mathrm{EXR}_{it} + \beta_8 \mathrm{IMP}_{it} + \\ & \beta_9 \mathrm{EXPO}_{it} + \beta_{10} \mathrm{INTR}_{it} + U_{it} \end{aligned} \tag{3.4}$$

In this case, subscript i in intercept (β_{0i}) proposes that the sixteen nations have diverse intercepts which are caused by diverse political organizations, varied monetary and fiscal strategies and diverse managerial capabilities. Fixed effect model has constant slopes but intercept differences Akbar et al. (2011). Equation 3.4 above, is the FEM with "within " effects. Fixed effects model shows that, intercept differs over nations but remains time invariant. The equation assumes that gradient coefficients of individual countries are not changing through nations and across time.
Now, to estimating the fixed effects intercept of specific nations, the method of LSDV (least square dummy variables) will be employed and the model is as follows:

$$\begin{aligned}\mathrm{GDP}_{it} =& \beta_0 + \beta_1\mathrm{LF}_{it} + \beta_2\mathrm{UNEM}_{it} + \beta_3\mathrm{ED}_{it} + \beta_4\mathrm{INF}_{it} + \beta_5\mathrm{FDI}_{it} + \beta_6\mathrm{IR}_{it} + \beta_7\mathrm{EXR}_{it} + \beta_8\mathrm{IMP}_{it} + \\ & \beta_9\mathrm{EXPO}_{it} + \alpha_i + \sum_{i=j=2}^{16} \alpha_i C_{ij} + U_{it},\end{aligned} \tag{3.5}$$

where,
$j = 2, 3, \cdots, 16$ denotes specific nation dummy variable,
$i =$ represent country effects of regressors,
$t =$ is the time effects of regressors.
In this case, $C_{i2} = 1$ if the study observation comes from country two, Botswana, and zero (0) elsewhere. Similar dummy variables would be used for the other nations (till 16 nations). Because there are sixteen countries, one must make use of fifteen nation dummy variables to avoid a dummy variable trap, which could be a case of perfect multicollinearity. It is possible to state that there is no dummy for the first nation and reflects the intercept of the first country Angola. And $\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_{15}$ are intercepts for respective dummies for countries. $\beta_1, \beta_2, \beta_3 \ldots, \beta_{10}$ are slopes for explanatory variables such as unemployment, real interest rate, FDI respectively. We use dummies to estimate country-specific effects because we are concerned with identifying the nation special effects that results from distinct political frameworks, unique fiscal and monetary guidelines and distinct administrative capacities. This method is also recognized as the LSDV (least-square dummy variables) method in other texts. As a result, terms LSDV and fixed effect model are frequently used interchangeably, as are the terms LSDV model and covariance model.

3.6 Constant Slope Coefficient but Intercept Varies across Countries and Time

Dummies may also be utilised when checking for time effects with creating a logic that variations happens in various nations across time, resulting from elements such as modification in government monitoring strategies, technological shifts, tax strategies, modifications in education systems and sometimes foreign effects like wars and also other struggles. For time effects time dummies are introduced, every year. Since the data set is for 22 years from 2000 to 2021, thus we present 21 dummies only to prevent dummy variable trap. The formulation would be as follows:

$$\begin{aligned}\mathrm{GDP}_{it} =& \gamma_1 + \gamma_2 T_2 + \gamma_3 T_3 + \cdots + \gamma_{21} T_{21} + \beta_1\mathrm{LF}_{it} + \beta_2\mathrm{UNEM}_{it} + \beta_3\mathrm{ED}_{it} + \beta_4\mathrm{INF}_{it} + \beta_5\mathrm{FDI}_{it} + \beta_6\mathrm{IR}_{it} + \\ & \beta_7\mathrm{EXR}_{it} + \beta_8\mathrm{IMP}_{it} + \beta_9\mathrm{EXPO}_{it} + \alpha_i + U_{it},\end{aligned} \tag{3.6}$$

where, T_2 take the value of 1 for a realisation in year 2001 and zero (0) elsewhere. Now, to show the both time effects and country effects, the equation is as below:

$$\begin{aligned}\mathrm{GDP}_{it} =& \alpha_1 + \alpha_2 C_2 + \cdots + \alpha_{16} C_{16} + \gamma_1 + \gamma_2 T_{2001} + \gamma_3 T_{2002} + ... + \gamma_{22} T_{2021} + \beta_1 \mathrm{LF}_{it} + \beta_2 \mathrm{UNEM}_{it} + \\ & \beta_3 \mathrm{ED}_{it} + \beta_4 \mathrm{INF}_{it} + \beta_5 \mathrm{FDI}_{it} + \beta_6 \mathrm{IR}_{it} + \beta_7 \mathrm{EXR}_{it} + \beta_8 \mathrm{IMP}_{it} + \beta_9 \mathrm{EXPO}_{it} + \alpha_i + U_{it}.\end{aligned} \tag{3.7}$$

Above model (3.7) can also be written as:

$$\begin{aligned}\mathrm{GDP}_{it} =& \sum_{i=j=2}^{16} \alpha_i C_{ij} + \sum_{k=t=2001}^{2020} \gamma_k T_{kt} + \beta_1 \mathrm{LF}_{it} + \beta_2 \mathrm{UNEM}_{it} + \beta_3 \mathrm{ED}_{it} + \beta_4 \mathrm{INF}_{it} + \beta_5 \mathrm{FDI}_{it} + \beta_6 \mathrm{IR}_{it} + \\ & \beta_7 \mathrm{EXR}_{it} + \beta_8 \mathrm{IMP}_{it} + \beta_9 \mathrm{EXPO}_{it} + \alpha_i + U_{it},\end{aligned} \tag{3.8}$$

where, $j = 2, 3, \ldots, 16$ shows specific nation dummy variable.
$k = 2001, 2002, \ldots, 2021$ gives time in years from 2001 to 2021.
$i =$ represents nation effects of regressors
$t =$ represents time effects of regressors
At this point, $C_{2i} = 1$ when the realization is for 2 and 0 elsewhere and similarly for particular nations. In this case, we aretaking 2000 to be a reference year with intercept value of γ_1. Owing to limitations of F-test, other time or year effects may not not be statistically significant. Suggesting that independent variables of the particular nation have not shifted across time. It is reasonably likely that nation effects could be noteworthy but the specific years special effects might be trivial.

3.7 All Coefficients Vary Across Countries

In this case, slopes and intercepts coefficients are diverse for all nations, it can be said that each labour function of; Eswatini, Malawi, Lesotho, Angola, Tanzania, DRC, Madagascar, Namibia, Botswana, Comoros, Mauritius, Mozambique, Seychelles, Zambia ,South Africa, and Zimbabwe are all different. This condition can be handled with easy by expanding the LSDV model. Here we will introduce the slope dummies or interaction terms that they will show how they account for variations in gradient coefficients. We multiply nation dummy variables by each of the explanatory variable. This is shown belowl:

$$\begin{aligned}\text{GDP}_{it} =& \beta_0 + \beta_1\text{LF}_{it} + \beta_2\text{UNEM}_{it} + \beta_3\text{ED}_{it} + \beta_4\text{INF}_{it} + \beta_5\text{FDI}_{it} + \beta_6\text{IR}_{it} + \beta_7\text{EXR}_{it} + \beta_8\text{IMP}_{it} + \\ & \beta_9\text{EXPO}_{it} + \beta_{10}\text{INTR}_{it} + \sum_{i=k=2}^{16} \alpha_k C_{ki} + \sum_{i=L=1}^{16} \gamma_L C_L\text{LF}_{it} + \sum_{i=L=1}^{16} \gamma_L C_L\text{UNEM}_{it} + \sum_{i=L=1}^{16} \gamma_L C_L\text{ED}_{it} + \\ & \sum_{i=L=1}^{16} \gamma_L C_L\text{INF}_{it} + \sum_{i=L=1}^{16} \gamma_L C_L\text{FDI}_{it} + \sum_{i=L=1}^{16} \gamma_L C_L\text{IR}_{it} + \sum_{i=L=1}^{16} \gamma_L C_L\text{EXR}_{it} + \sum_{i=L=1}^{16} \gamma_L C_L\text{IMP}_{it} + \\ & \sum_{i=L=1}^{16} \gamma_L C_L\text{EXPO}_{it} + \sum_{i=L=1}^{16} \gamma_L C_L\text{INTR}_{it},\end{aligned} \tag{3.9}$$

where, $j = 2, 3, \ldots, 16$ shows specific nation dummy variable.
$k = 2001, 2002, \ldots, 2021$ gives time in years from 2001 to 2021.
$i =$ represents nation effects of regressors
$t =$ represents time effects of regressors
In this case, $\gamma's$ are distinctive gradient coefficients the same as $\alpha's$ are distinctive intercepts. When more than one or one γ coefficients are reporting a figure which is statistically noteworthy, one can pronouns that, gradient coefficients are diverse from the base group. When a case exists such that all different gradient coefficients and different intercepts are statistically noteworthy then one may infer that the unemployment formulation for one nation is diverse from the other nation. It is quite possible that some or none of differential intercepts would be statistically significant.

3.8 Model for Random Effects

The intercept is anticipated as a random ending entity in a random effects model, while the random effect is a formulation of the average value and a random residual. The double random effects model utilised in the approximation procedure is:

$$\begin{aligned}\text{GDP}_{it} =& \beta_0 + \beta_1\text{LF}_{it} + \beta_2\text{UNEM}_{it} + \beta_3\text{ED}_{it} + \beta_4\text{INF}_{it} + \beta_5\text{FDI}_{it} + \beta_6\text{IR}_{it} + \beta_7\text{EXR}_{it} + \beta_8\text{IMP}_{it} + \\ & \beta_9\text{EXPO}_{it} + \beta_{10}\text{INTR}_{it} + \epsilon_{it}.\end{aligned} \tag{3.10}$$

Instead of taking β_{0i} to be fixed, it is understood as a random variable having a mean value of β_0 and the intercept for a specific country may be written as; $\beta_{0i} = \beta_0 + \epsilon_i, i = 1, 2, 3, \ldots, N$ Here ϵ_i is a residual term having variance of σ_ϵ^2 and a mean value of zero Thus

$$\begin{aligned}\text{GDP}_{it} =&\beta_0 + \beta_1\text{LF}_{it} + \beta_2\text{UNEM}_{it} + \beta_3\text{RIR}_{it} + \beta_4\text{INF}_{it} + \beta_5\text{FDI}_{it} + \beta_6\text{IR}_{it} + \beta_7\text{EXR}_{it} + \beta_8\text{IMP}_{it}+\\ &\beta_9\text{EXPO}_{it} + \beta_{10}\text{INTR}_{it} + \epsilon_{it} + \mu_{it}.\end{aligned} \tag{3.11}$$

Thus we have:

$$\begin{aligned}\text{GDP}_{it} =&\beta_0 + \beta_1\text{LF}_{it} + \beta_2\text{UNEM}_{it} + \beta_3\text{RIR}_{it} + \beta_4\text{INF}_{it} + \beta_5\text{FDI}_{it} + \beta_6\text{IR}_{it} + \beta_7\text{EXR}_{it} + \beta_8\text{IMP}_{it}+\\ &\beta_9\text{EXPO}_{it} + \beta_{10}\text{INTR}_{it} + \omega_{it},\end{aligned} \tag{3.12}$$

here $\omega_{it} = \epsilon_{it} + \mu_{it}$,

In situations of this nature, the error term v_i relative to a cross-sectional element is heterogeneity-specific . The v_i is constant across time. Consequently, $E[v_i^2/x] = \sigma_i^2$. The error term ϵ_{it} is particular to a specific realisation. For v_{it} to be correctly quantified, it has to be orthogonal relative to specific effects. As a result of distinct cross-sectional random terms, these models often are occasionally entitled to one-way random effects models. Due to this intra-panel variations, the random effects models possess distinctive benefit of permitting time-invariant variables to be integrated among the explanatory variables.

3.9 Putting Random and Fixed Effects to Test

How do we tell if the panel data have fixed and/or random effects? Fixed effects are tested by F-test, while a random effects are examined by Breusch and Pagans Breusch and Pagan (1980) Lagrange multiplier (LM) test. The former contrasts a OLS to a fixed effect model to assess how much the fixed effects model increases goodness of fit while the latter contrasts an OLS model to a random effects model. A Hausman test is employed to compare the similarity between random effects and fixed effects estimators.

3.9.1 Testing Fixed Effects Using the F-test

With the equation $y_{it} = \alpha + \mu_i + X'_{it}\beta + \epsilon_{it}$, the null hypothesis states the following: All dummies with the exception of one are equal to zero, ie; $H_0 : \mu_0 = \cdots = \mu_{n-1} = 0$. While the alternative hypothesis says that at least one dummy variable is not zero. This proposition is evaluated by way of the F test that is hinged on the loss of goodness-of-fit. The test compares the pooled OLS and LSDV models and scrutinizes the degree that the goodness-of-fit as measured by R^2 or SSE has transformed.

$$F_{(n-1,nT-n-k)} = \frac{\left(e'e_{\text{pooled}} - e'e_{\text{LSDV}}\right)/(n-1)}{\left(e'e_{\text{LSDV}}\right)/(nT-n-k)} = \frac{\left(R^2{}_{\text{LSDV}} - R^2 \text{ pooled }\right)/(n-1)}{\left(1 - R^2\text{LSDV}\right)/(nT-n-k)} \tag{3.13}$$

Dismissing the null hypothesis says; at least one time /group specific intercept u_i is not zero, one may infer that the fixed effect model has a substantial fixed effects or a significant improvement in goodness-of-fit; hence, the fixed effects model is superior to the pooled OLS.

3.9.2 Testing for Random Effects Using Breusch-Pagan LM Test

Breusch and Pagans Breusch and Pagan (1980) Lagrange multiplier (LM) test checks if time (or individual) specific variance components are zero, $H_0 : \sigma_u^2 = 0$. The LM statistic has a chi-square distribution having one degree of freedom.

$$LM_u = \frac{nT}{2(T-1)} = \left[\frac{T^2\bar{e}'\bar{e}}{e'e} - 1\right]^2 \sim \chi^2_{(1)} \tag{3.14}$$

Where $e'e$ is the SSE of the pooled OLS regression and $\bar{e}$ is the $n \times 1$ vector of the group means of pooled regression residuals. Baltagi Baltagi and Baltagi (2008) The LM statistic has a chi-squared distribution having one degree of freedom.

$$LM_u = \frac{nT}{2(T-1)}\left[\frac{\sum\left(\sum e_{it}\right)^2}{\sum\sum e_{it}{}^2} - 1\right]^2 = \frac{nT}{2(T-1)}\left[\frac{\sum\left(\sum e_{it}\right)^2}{\sum\sum e_{it}{}^2} - 1\right]^2 \sim \chi^2{}_{(1)} \tag{3.15}$$

Rejecting the null hypothesis, one may infer that the panel data contains a substantial random effects and that the random effects model handles heterogeneity more appropriately compared to the pooled ordinary least squares.

3.9.3 Comparing Random and Fixed Effects using the Hausman Test

How can one tell which effects (random or fixed) are more significant and useful in longitudinal data? Hausman test Hausman (1978). contrasts random and fixed effects models having the null hypothesis that specific effects not associated with any explanatory variable in the model. LSDV and GLS are consistent if the null hypothesis of no association is not violated; otherwise, LSDV is consistent while GLS is inconsistent and biased. Greene (2003). Under the null hypothesis, the approximations of GLS and LSDV must not differ systimatically. "The covariance of an efficient estimator with its difference from an inefficient estimator is zero," according to the Hausman test. Greene (2003).

$$LM = \left(b_{\text{LSDV}} - b_{\text{random}}\right)' \widehat{W}^{-1}\left(b_{LSDV} - b_{\text{random}}\right) \sim \chi^2{}_{(k)} \tag{3.16}$$

Here $\widehat{W} = Var[b_{LSDV} - b_{random}] = Var(b_{LSDV} - b_{random}$ gives the variance in the projected covariance matrices of GLS and LSDV. Remember to leave out the intercept and dummy variables while computing. This test statistic has k degrees of freedom and follows a chi-squared distribution. According to the formula, a Hausman test determines if "the random effects estimate is trivially different from the unbiased fixed effect estimate." Kennedy (2008). If the null hypothesis of no association is rejected, we can infer that specific effects u_i are highly linked with at least one explanatory variable in the model, and hence the random effect model becomes problematic. As a result, we must choose a fixed effects model over a random effects model. The difference of covariance matrices W may not be positive definite; hence, we may argue that the null is not rejected assuming similarity of covariance matrices makes such a difficulty Greene (2003)

3.9.4 Testing for Poolability using Chow Test

Poolability inquires whether slopes are constant over groups or across time. Baltagi and Baltagi (2008).The Chow test Chow, 1960 is an expansion of a simple form of the poolability test. The null hypothesis in the Chow test says that the gradient for an explanatory variable remains constant for all k regressors independent of the individual. $H_0 : \beta_{ik} = \beta_k$. Need is there to remember that gradiants remain fixed for random and fixed effects models; only error variances and intercepts are important.

$$F_{[(n-1)(k+1),n(T-k-1)]} = \frac{\left(e'e - \sum e_i'e_i\right)/(n-1)(k+1)}{\left(\sum e_i'e_i\right)/n(T-k-1)} \tag{3.17}$$

Where $e'e$ represents the SSE of the pooled OLS while $e_i'e_i$ is the SSE of the pooled OLS for group i. Rejecting the null hypothesis says that, panel data at hand cannot be poolable since every entity has its own slope for all explanatory variables. In this case, one may use the random coefficient model or the hierarchical regression model. The Chow test is based on the assumption that specific residual variance entities have a normal distribution $\mu \sim N(0, s^2 I_{nT})$.When this assumption is not met, the Chow test may fail to appropriately assess the null hypothesis, according to Baltagi and Chang (1994a). According to Kennedy (2008), if there is reason to believe that errors in different equations have different variances, or that there is contemporaneous correlation between the equations errors, such testing should be undertaken using the Stein's unbiased risk estimate (SURE) is an unbiased estimator of the mean-squared error of; a nearly arbitrary, nonlinear biased estimator, not OLS; inference with OLS is unreliable if the variance-covariance matrix of the error is nonspherical Baltagi and Chang (1994a).

3.10 Selecting the Model: Fixed or Random Effect

We obtain 12 potential panel data models when we combine random against fixed effects, time against group effects and two-way against one-way effects, as in Table 3.1 below. Generally, one-way models are frequently utilized because of their parsimony and fixed effects models are simpler to approximate and explain than a random equivalent. It is, nevertheless, difficult to choose the finest model among the following 12.

Table 3.1: Classification of Panel Data Analysis

Fixed effects (F-test)	**Random effects (Breusch-Pagan LM test)**	**The Selection**
H_0Rejected (Absence of fixed effects)	H_0Not rejected (Random effects absent)	Pooled Ordinary Least Squares
H_0 rejected (Presence of Fixed effects)	H_0Not rejected (Random effects absent)	Fixed effect model
H_0Not rejected (Fixed effects absent)	H_0 rejected (Presence of random effects)	Random effect model
H_0 not rejected (Fixed effects present)	H_0Rejected (Presence of random effect)	Choose the fixed effects model when the null hypothesis of a Hausman test is rejected; otherwise, go for a random effects model.

3.11 Machine Learning Procedures

3.11.1 Modelling Zimbabwe Data Using Deep Learning

In modeling the GDP for Zimbabwe, a dense neural network with a total of 76 neurons will trained to give a multivariate regression model for GDP. Linear functions will used as activation functions in both the hidden and output layers. The optimizer *"Adam"* and loss metric *"Mean Squared Error"* will be used. The model will trained for a large number epochs, with the main objective of reducing the mean squared error between the observed and trained values to almost zero. The model attained will be used to obtain forecasts GDP for Zimbabwe for the next five years.

3.11.2 Modelling SADC Data Using; Support Vector Machine,Gradient Boost and Random Forest Approaches

Support vector regression, is useful for modeling complex nonlinear relationships between input and output variables. Support vector regression is a variation of support vector machines that specifically handles regression problems.The data will be split into training and testing data sets. Then, a support vector regression model will be trained. Using the trained model, predictions will be made on the test dataset, and the R-squared value will be determined.Gradient boosting approach is effective for solving both regression and classification problems. Initially, we will split the data into 70% for training and 30% for testing purposes. We will use the 'xgboost()' function from the 'XGBoost' package in R to train the model for 100 iterations. Afterwards, we will make predictions on the testing data set and determine the R-squared value. As for the random forest procedure, a random forest model will trained on the panel, the data will partitioned into 70% training and 30%testing sets, then the "randomForest()" function from the "randomForest" package will used to train 100 trees in the forest. Predictions will then be made on the testing data set and the RMSE and R^2 will be determined. The variable importance, will obtained from the importance attribute in R. This attribute displays the contribution of each variable to the accuracy of the model. The importance measures are based on the mean decrease in the accuracy of the model when each variable is randomly permuted. A variable with a higher importance measure indicates that it has a greater impact on the accuracy of the model. In this case, the "IncNodePurity" measure will be used, which is based on the Gini impurity index. R-Square values will be used to determine the most appropriate modelling procedure

3.12 Chapter Summary

This chapter presents the methodology used in this research, including the research design and sampling techniques. Additionally, it details the formulation of the proposed panel data model and the software used in the study. The methods used to evaluate and compare the accuracy of the developed models are also clearly articulated.

Chapter 4

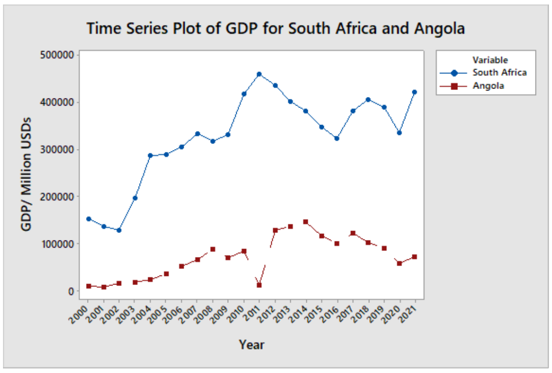
Results and Discussions

4.1 Introduction

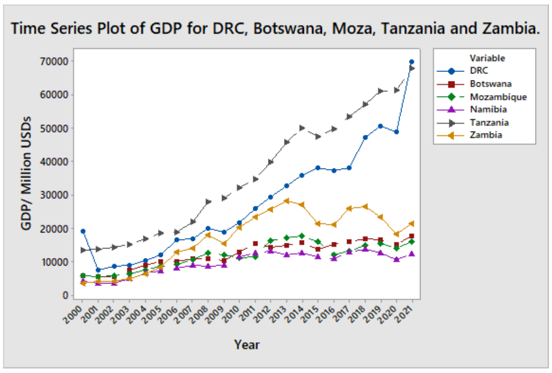
The analysis of panel data is an important area of research in economics and social sciences. This chapter examines panel data analysis using both modern and traditional techniques following the methodology provided in the previous chapter. The traditional techniques consist of general linear models and panel linear models, while the modern techniques involve machine learning algorithms. The modeling power of these two techniques will be assessed to determine which method is superior. Additionally, this chapter will also provide GDP forecasts for Zimbabwe for the next five years.

4.2 GDP for SADC Nations

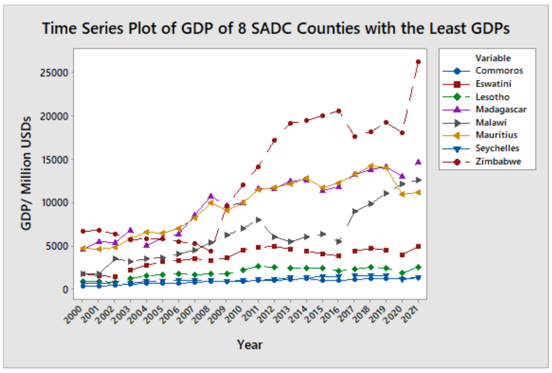
The relationship between GDPs for SADC countries is explored using time series plots. Figures 4.1(a) to 4.1(d) show the time series plots of GDPs of SADC countries:



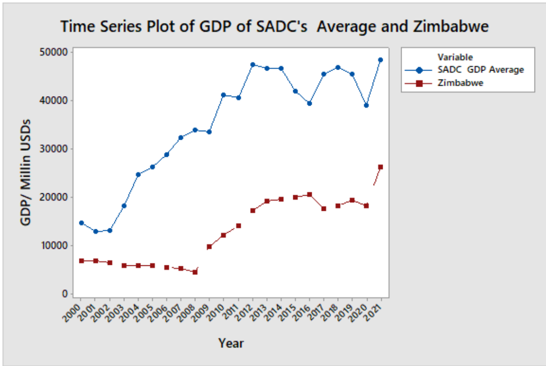
(a)



(b)



(c)



(d)

Figure 4.1: Time Series Plots of GDP for SADC Nations 2000 to 2021

Table 4.1 presents the summary statistics of GDP for SADC nations from 2000 to 2021.

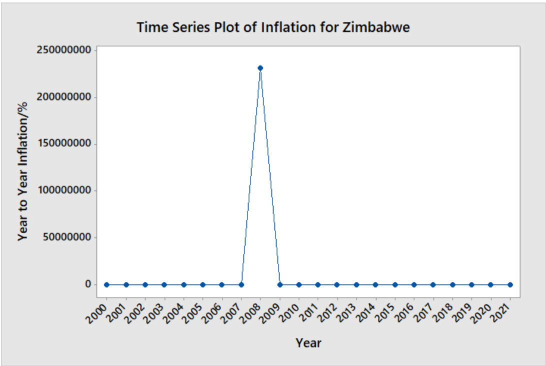
Table 4.1: Summary Statistics of GDP for SADC Nations 2000-2021

Country	**Mean**	**Standard Error**	**Median**	**Standard Dev**	**Sample Variance**	**Range**	**Minimum**	**Maximum**
Angola	70671.3	9488.0	71426.5	44502.7	198348.0	136776	8936	145712
Botswana	12245.6	831.3	13183.0	3899.2	15203547.2	12174	5439	17613
Comoros	888.4	61.7	955.5	289.5	83810.3	976	351	1327
DRC	27839.7	3505.8	23700.0	16443.8	270397480.3	62326	7438	69764
Eswatini	3637.2	235.9	3894.5	1106.7	1224719.1	3509	1432	4941
Lesotho	1908.4	124.0	1994.5	581.8	338535.8	1803	776	2579
Madagascar	9926.1	710.4	11039.0	3332.2	11103608.6	10008	4629	14637
Malawi	6193.8	666.5	5773.5	3126.3	9773829.2	10909	1717	12626
Mauritius	9677.9	671.6	10462.5	3149.9	9921618.8	9568	4614	14182
Mozambique	11734.8	848.7	11925.5	3980.9	15847648.2	12317	5399	17716
Namibia	9524.0	713.0	10670.5	3344.5	11185608.6	10332	3349	13681
Seychelles	1085.1	65.0	1047.0	304.8	92925.5	968	615	1583
South Africa	325581.8	20515.2	334258.5	96225.0	9259250706.0	329114	129088	458202
Tanzania	35849.2	3861.3	33335.5	18110.9	328005489.6	54399	13376	67775
Zambia	16946.5	1793.0	19188.5	8409.9	70726508.6	24436	3601	28037
Zimbabwe	12889.9	1450.7	13072.0	6804.3	46298627.5	21801	4416	26217

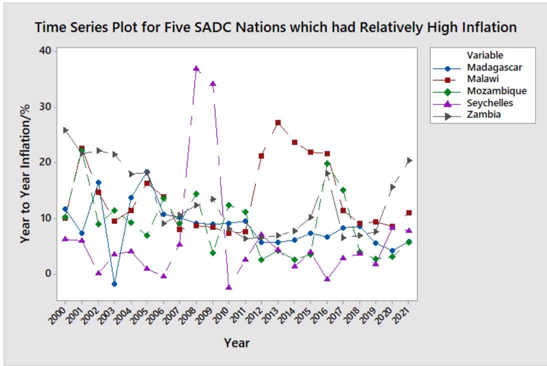
From Figure 4.1, South Africas GDP was higher than Angola from 2000-2021. Tanzania had the highest GDP and Namibia having the lowest of the 6 countries, DRC, Botswana, Mozambique, Namibia, Tanzania and Zambia. Zimbabwe had the highest GDP of the 8 countries with the least GDP's. SADC's average GDP was higher than Zimbabwe. From Table 4.1, South Africa had the highest mean GDP while Comoros had least mean GDP.

4.3 Inflation Values for SADC Nations

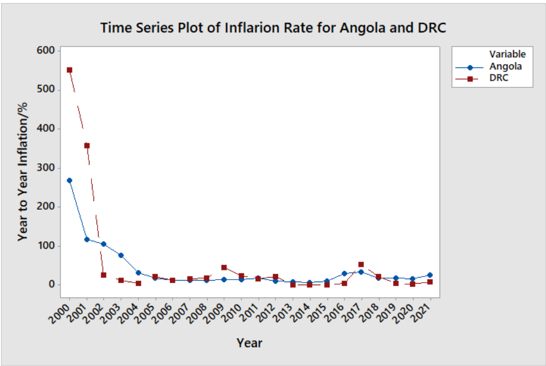
Figures 4.2(a) to 4.2(d) shows the time series plots illustrating the inflation trends across SADC nations.



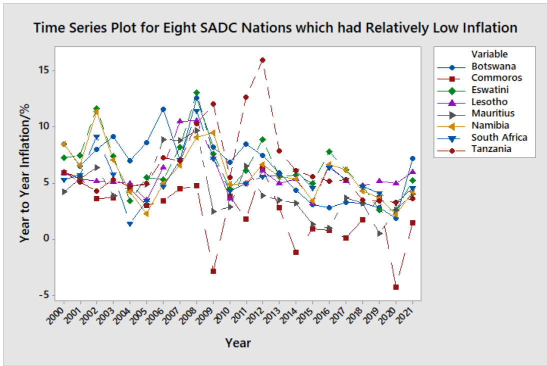
(a)



(b)



(c)



(d)

Figure 4.2: Time series Plots of Inflation for SADC Nations 2000 to 2021

Table 4.2 presents the summary statistics of inflation for SADC nations from 2000 to 2021.

Table 4.2: Summary Statistics of Inflation for SADC Nations 2000-2021

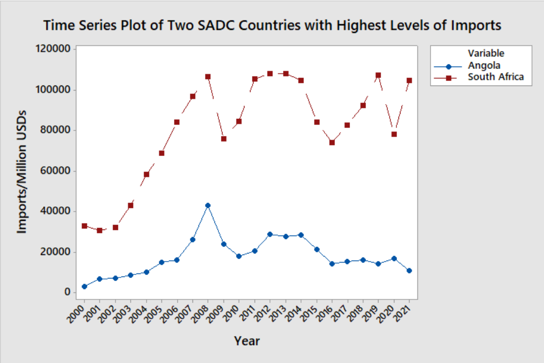
Country	**Mean**	**Standard Error**	**Median**	**Standard Dev**	**Sample Variance**	**Range**	**Minimum**	**Maximum**
Angola	39.9	12.6	17.8	59.3	3517.5	261.0	7.3	268.3
Botswana	6.6	0.6	7.1	2.9	8.4	10.7	1.9	12.6
Comoros	2.5	0.6	3.2	2.7	7.5	10.6	-4.3	6.3
DRC	56.6	28.3	16.2	132.9	17654.0	549.2	0.8	550.0
Eswatini	6.5	0.5	5.9	2.6	6.6	10.5	2.6	13.1
Lesotho	6	0	5	2	3	7	3	11
Madagascar	9	1	8	4	18	20	-2	18

Country	Mean	Standard Error	Median	Standard Dev	Sample Variance	Range	Minimum	Maximum
Malawi	14	1	11	6	40	20	7	27
Mauritius	4.35	0.526	3.9	2.5	6.09	9.2	0.5	9.7
Mozambique	8.98	1.215	9.1	5.7	32.49	19.7	2.6	22.3
Namibia	5.85	0.494	5.5	2.3	5.38	9.19	2.21	11.4
Seychelles	6.27	2.115	3.85	9.9	98.40	39.4	-2.4	37
South Africa	5.49	0.440	5.3	2.1	4.27	10.1	1.4	11.5
Tanzania	6.60	0.711	5.4	3.3	11.12	12.71	3.29	16
Zambia	13.42	1.351	11.55	6.3	40.17	19.5	6.4	25.9
Zimbabwe	1.05E+5	1.04E+7	246.4	4.92E+8	2.43E+15	2.31E+6	-7.7	2.310E+6

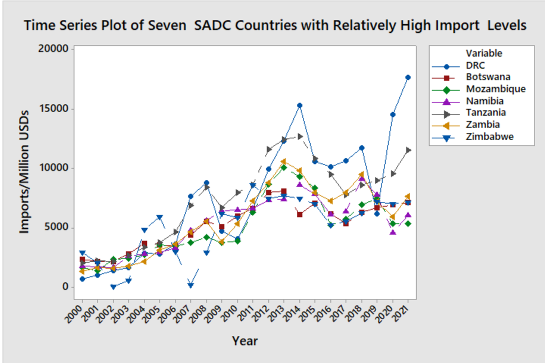
From Figure 4.2, Zimbabwe experienced very high inflation in 2008, which stands to be the highest to have been witnessed of SADC nations. Inflation levels for Angola and DRC started off high, gradually decreasing to low levels, DRC having a significantly sharp decline from 2000 to 2002. Seychelles experienced the highest inflation level of the five SADC nations which had relatively high inflation. Comoros has the least inflation level of the eight SADC nations which had relatively low inflation. From Table 4.2, Zimbabwe had the highest mean inflation while Comoros had the least average inflation.

4.4 Levels of Imports for SADC Nations

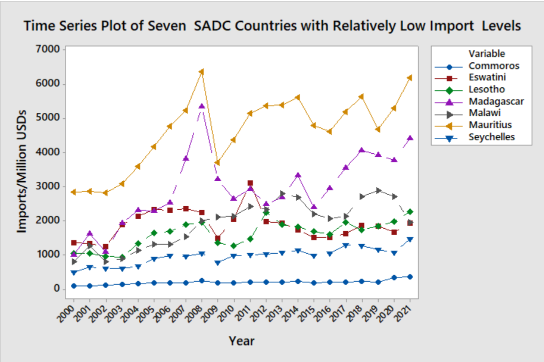
Figures 4.3(a) to 4.3(d) depicts the time series plots illustrating the levels of imports across SADC nations.



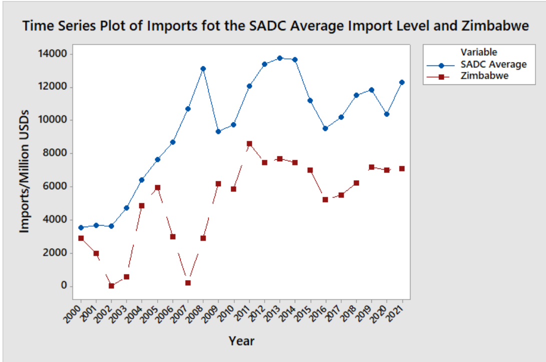
(a)



(b)



(c)



(d)

Figure 4.3: Time Series plots of Imports for SADC Nations 2000 to 2021

Table 4.3 presents the summary statistics of imports levels for SADC nations from 2000 to 2021.

Table 4.3: Summary Statistics of Imports for SADC Nations 2000-2021

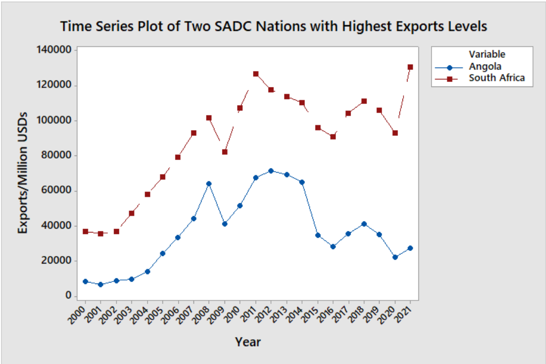
Country	**Mean**	**Standard Error**	**Median**	**Standard Dev**	**Sample Variance**	**Range**	**Minimum**	**Maximum**
Angola	17946.5	1973.7	16337	9257	85697432.50	40246.00	3040.00	43286.00
Botswana	5253.8	404.6	5781	1898	3600771.900	5964.000	2149.000	8113
Comoros	198.8	13.7	194	6401	4103.7	269	95	364
DRC	7494.6	1071.1	7096	5024	25241850.8	16970.0	707.0	17677.0
Eswatini	1885.3	92.0	1873	431	186144.60	1869.00	1242.00	3.11E+03
Lesotho	1621.9	84.2	1697	395	156055.2	1325	948	2273

Country	Mean	Standard Error	Median	Standard Dev	Sample Variance	Range	Minimum	Maximum
Madagascar	2927.5	227.4	2815	1066	1137181.5	4360	997	5357
Malawi	1920.5	145.2	2092	681	464076.5	2086	799	2885
Mauritius	4628.5	229.1	4784	1074.7	1154978.7	3550	2819	6369
Mozambique	5078	539.1	4687.5	2528.4	6392673	8743	1356	10099
Namibia	5351.7	503.2	6088.5	2360.3	5570794.6	7509	1620	9129
Seychelles	963.3	52.9	998	248.1	61537.3	965	502	1467
South Africa	80344.1	5575.4	84356	26151	683873237	77473	30919	108392
Tanzania	7405.8	749.9	8193.5	3517.2	12370687.4	10641	2049	12690
Zambia	5658	627.6	5713.5	2943.8	8665862.8	9260	1313	10573
Zimbabwe	5045.3	558.1	5908	2617.6	6851937.1	8576	20	8596

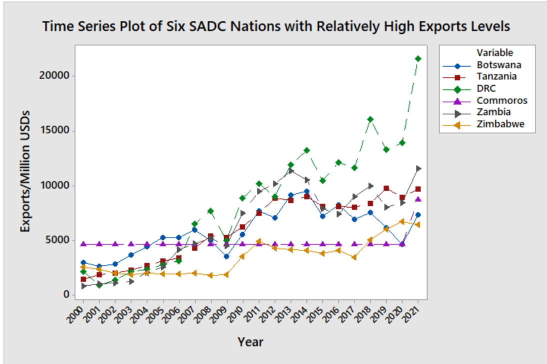
From Figure 4.3, South Africa and Angola had the highest levels of imports, South Africa with significantly higher import levels than Angola. DRC had higher imports by 2021 standing out from the 7 SADC countries with relatively high imports, Namibia coming in second. Mauritius had much better imports compared to other 7 SADC countries with relatively low import levels. SADC's average import level was higher than Zimbabwe. From Table 4.3, on average South Africa had the highest levels of imports while Comoros had the least average level of imports.

4.5 Levels of Exports for SADC Nations

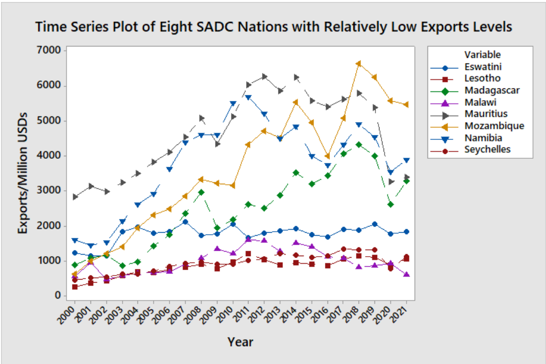
Figures 4.4(a) to 4.4(d) shows time series plots illustrating the levels of exports across SADC nations.



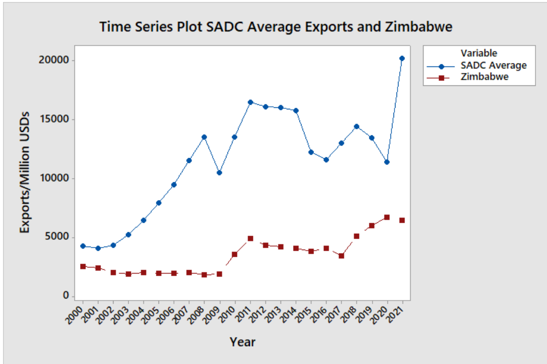
(a)



(b)



(c)



(d)

Figure 4.4: Time series plots of exports for SADC Nations 2000 to 2021

Table 4.4 presents the summary statistics of exports levels for SADC nations from 2000 to 2021.

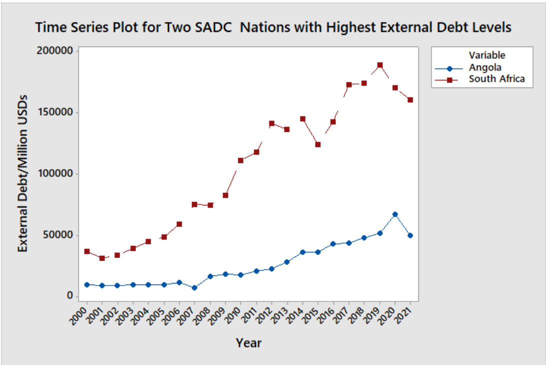
Table 4.4: Summary Statistics of Exports for SADC Nations 2000-2021

Country	**Mean**	**Standard Error**	**Median**	**Standard Dev**	**Sample Variance**	**Range**	**Minimum**	**Maximum**
Angola	36564.3	4519.2	34876.5	21197	449300233.100	64896.100	6736.800	71632.900
Botswana	5853.6	428.6	5771.5	2011	4042287.3	6809.4	2671.1	9480.5
Comoros	4867.2	186.1	4681.1	873	767278.8	4094.9	4681.1	8776
DRC	8482.8	1189.4	8946.6	5579	31122580.4	20759.3	891.7	21651
Eswatini	1767.9	56.7	1819.8	266.1	70794.3	977.9	1141	2118.9
Lesotho	839.4	53.6	888	251.3	63173.6	944	270	1214
Madagascar	2465.8	232.9	2573.3	1092.2	1192986.5	3466.7	872.7	4339.4
Malawi	1001.8	74.8	951	351.1	123253.3	1126	487	1613
Mauritius	4626.3	252.6	4827.7	1184.8	1403858	3427.8	2849	6276.8
Mozambique	3673.9	381.5	3666.9	1789.5	3202266.8	6028	633.3	6661.3
Namibia	3836.2	272.9	4172.4	1280.2	1638792.2	4242.6	1460.5	5703.1
Seychelles	943.6	58.3	958.7	273.5	74822.4	887.1	464.1	1351.2
South Africa	88595.4	6286.1	94632.5	29484.3	869326513.5	94940.3	35694.7	130635
Tanzania	6065.1	623.1	6887.5	2922.5	8540909.4	8340.9	1445.8	9786.7
Zambia	6319.7	784	7453.5	3677.5	13523837.6	10714.6	861.4	11576
Zimbabwe	3506.2	341.2	3513.2	1600.5	2561614.1	4884.3	1831.1	6715.4

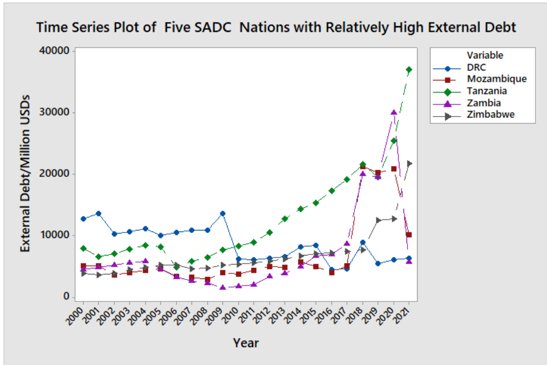
With reference to Figure 4.4, South Africa and Angola were two SADC nations with highest export levels, South Africa having significantly higher levels than Angola. DRC had the highest export levels of the six SADC nations with relatively high export levels. Malawi, Lesotho and Seychelles had the lowest export levels of the eight SADC nations with relatively low export levels. Zimbabwe had low export levels compared to SADC's average exports. From Table 4.4, South Africa had the highest average level of exports while Lesotho had the least average level of exports.

4.6 External Debts for SADC Nations

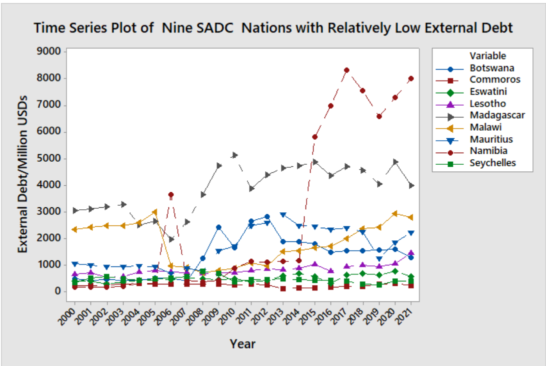
Figures 4.5(a) to 4.5(d) shows time series plots illustrating the levels of external debts across SADC nations.



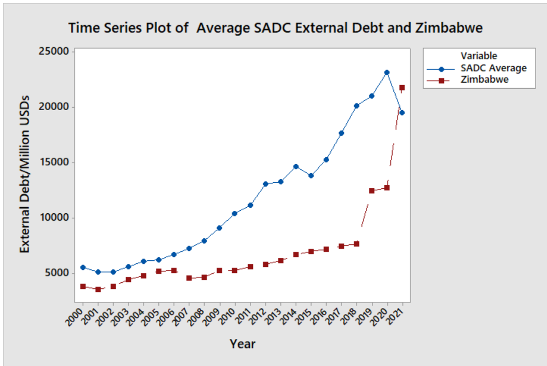
(a)



(b)



(c)



(d)

Figure 4.5: Time Series Plots of External Debts for SADC Nations 2000 to 2021

Table 4.5 presents the summary statistics of external debts for SADC nations from 2000 to 2021.

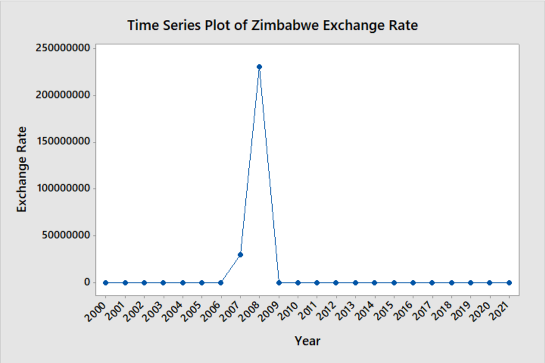
Table 4.5: Summary Statistics of External Debt for SADC Nations 2000-2021

Country	**Mean**	**Standard Error**	**Median**	**Standard Dev**	**Sample Variance**	**Range**	**Minimum**	**Maximum**
Angola	26168.6	3825.6	19522	17944	321977407.0	60386.8	6900.0	67286.8
Botswana	1322.5	164.7	1511	773	597113	2419	408	2827
Comoros	242.2	12.7	268.5	59.4	3533	182	122	304
DRC	8743.5	611.8	8741	2869.7	8235271	9068	4542	13610
Eswatini	500.9	27.5	475.5	129.2	16681.5	477.1	289	766.1
Lesotho	823.6	41.7	783.5	195.4	38166.6	896	546	1442
Madagascar	3867.1	198	4038	928.5	862194.8	3171	1964	5135
Malawi	1844.1	165.3	1848	775.4	601219	2291	709	3000
Mauritius	1670.8	158.3	1624.5	742.4	551108.7	2225	698	2923
Mozambique	6861.6	1251.6	4772.5	5870.5	34462763	18378	2916	21294
Namibia	2832.6	681.3	1131	3195.5	10211141.3	8179	159	8338
Seychelles	463.5	24.7	458.5	115.9	13435.9	497	263	760
South Africa	105300.3	11458.5	114719	53745.1	2888539047	158606	31061	189667
Tanzania	12801.8	1695.1	8707.5	7950.5	63211238	32287	4813	37100
Zambia	7001.3	1510.8	4926.5	7086.1	50212847.4	28501	1545	30046
Zimbabwe	6898.7	875	5466.5	4103.9	16842093.4	18240	3590	21830

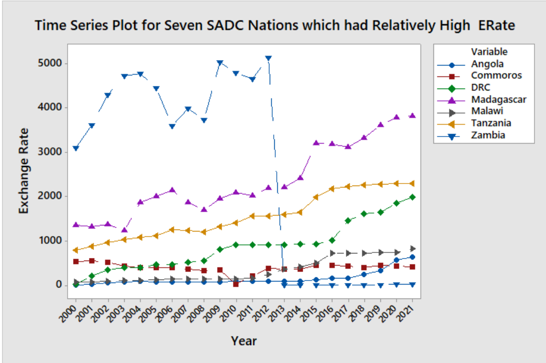
Referring to Figure 4.5, South Africa and Angola were two SADC nations with highest external debt levels, South Africa having significantly higher levels. Tanzania had a gradual increase in external debt of the five SADC nations with relatively high external debt. Comoros had the least levels of external debt of the nine SADC nations with relatively low external debt. Zimbabwe had low external debt level as compared to SADC's average external debt. From Table 4.5, South Africa had the highest mean external debt level while Comoros had the least average external debt level.

4.7 Exchange Rates for SADC Nations

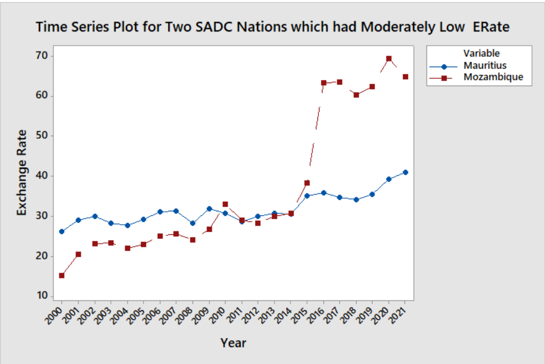
Figures 4.6(a) to 4.6(d) shows time series plots illustrating the exchange rates across SADC nations.



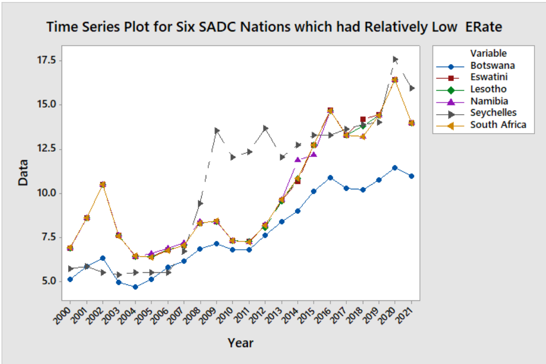
(a)



(b)



(c)



(d)

Figure 4.6: Time Series Plots of Exchange Rates for SADC Nations 2000 to 2021

Table 4.6: Summary Statistics of Exchange Rate for SADC Nations 2000-2021

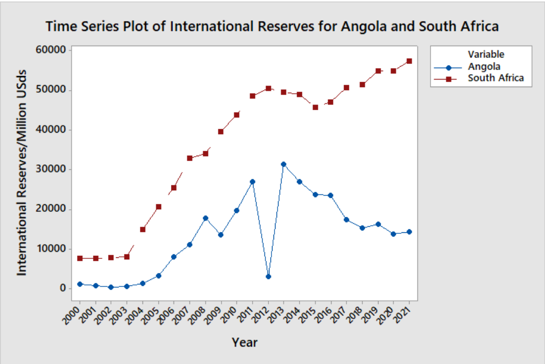
Country	**Mean**	**Standard Error**	**Median**	**Standard Dev**	**Sample Variance**	**Range**	**Minimum**	**Maximum**
Angola	153.1	34.7	93	163	26460.3	621	10	631
Botswana	7.8	0.5	7	2.3	5.2	6.8	4.7	11.5
Comoros	392.5	23.4	398.2	109.8	12046.4	527.2	21.8	549
DRC	877.6	115.7	912.6	542.9	294744.7	1967.2	21.8	1989
Eswatini	10	0.7	8.5	3.3	10.6	10.1	6.4	16.5
Lesotho	10	0.7	8.5	3.2	10.5	10.1	6.4	16.5
Madagascar	2356.4	177.3	2116.1	831.5	691311.1	2590	1239	3829
Malawi	339.7	59.7	153.5	279.9	78357.1	752.7	67.3	820

Country	Mean	Standard Error	Median	Standard Dev	Sample Variance	Range	Minimum	Maximum
Mauritius	31.8	0.8	30.8	3.8	14.4	14.7	26.3	41
Mozambique	36.5	3.8	28.6	18	322.6	54.3	15.2	69.5
Namibia	10	0.7	8.5	3.2	10.1	10.1	6.4	16.5
Seychelles	10.6	0.9	12.2	4.1	16.7	12.2	5.4	17.6
South Africa	10	0.7	8.5	3.2	10.3	10.1	6.4	16.5
Tanzania	1555.9	110.3	1488	517.5	267770.2	1496.3	800.7	2297
Zambia	2550.1	472.2	3606.9	2214.8	4905383.2	5136.9	5.4	5142.3
Zimbabwe	11838557.2	10486053	23.6	49183948.3	2.42E+15	230168570.5	1	230168571.5

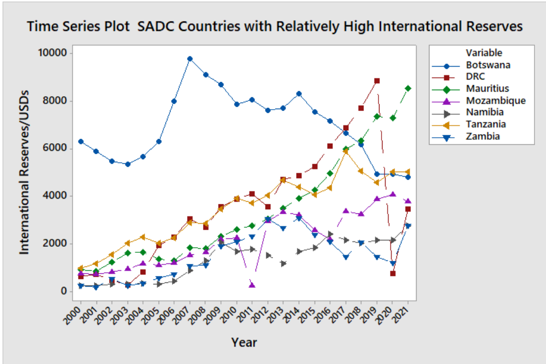
From Figure 4.6, Zimbabwe had a high peak in exchange rate in 2008. Zambia had the highest exchange rate of the seven SADC nations which had relatively high exchange rate until 2013 and became the one with the lowest level of exchange rate. Mauritius and Mozambique had moderately low exchange rates, Mozambique started of lower than Mauritius and took lead in 2015 till to date. Botswana has the lowest exchange rate of the six SADC nations with low exchange rate. From Table 4.6, Zimbabwe experienced the highest average exchange rate while Botswana had the least average exchange rate.

4.8 International Reserves for SADC Nations

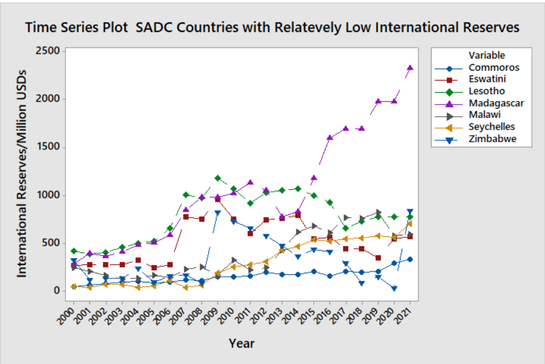
Figures 4.7(a) to 4.7(d) shows time series plots illustrating the international reserves across SADC nations.



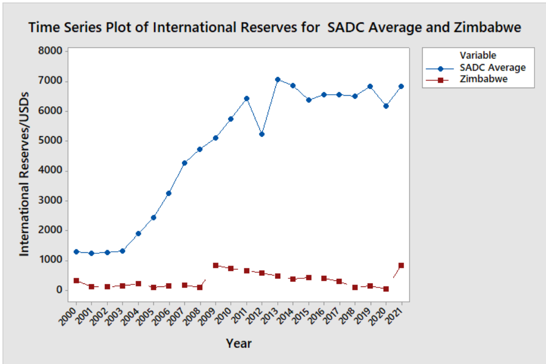
(a)



(b)



(c)



(d)

Figure 4.7: Time Series Plots of International Reserves for SADC Nations 2000 to 2021

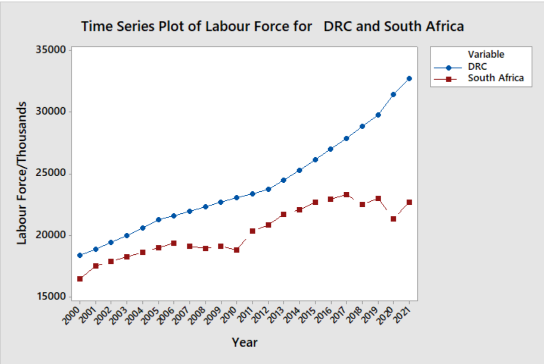
Table 4.7: Sammary Statistics of International Reserves for SADC Nations 2000-2021

Country	**Mean**	**Standard Error**	**Median**	**Standard Dev**	**Sample Variance**	**Range**	**Minimum**	**Maximum**
Angola	13259.9	2096.3	14125	9832.4	96676234.8	31125	376	31501
Botswana	6931.5	309.2	6923	1450.4	2103630.8	4989	4801	9790
Comoros	153.5	15.1	153.5	71	5044.8	286	43	329
DRC	3484.3	524	3513	2457.8	6040826.2	8592	279	8871
Eswatini	524.2	47	546.3	220.3	48522.2	715	244	959
Lesotho	785.4	54	774.1	253.2	64110	794	386	1180
Madagascar	1051.4	127	982	595.5	354618.5	2049	285	2334
Malawi	385.1	51.7	250.5	242.3	58710.1	698	127	825
Mauritius	3432.5	500.6	2697	2347.9	5512484.1	7709	853	8562
Mozambique	2149.5	254.7	2182	1194.5	1426712.7	3832.5	259	4091.5
Namibia	1359.9	177.2	1599.5	830.9	690440.3	2530	234	2764
Seychelles	293.1	49.3	267.5	231.2	53461.5	667	35	702
South Africa	36553.4	3796.8	44853.5	17808.6	317144796.8	49970	7627	57597
Tanzania	3473.8	299.3	3815.5	1404	1971204.3	4893	995	5888
Zambia	1523.5	205.6	1449	964.4	930127.5	2895	183	3078
Zimbabwe	332	54.2	263.5	254.1	64588.7	806	33.4	839

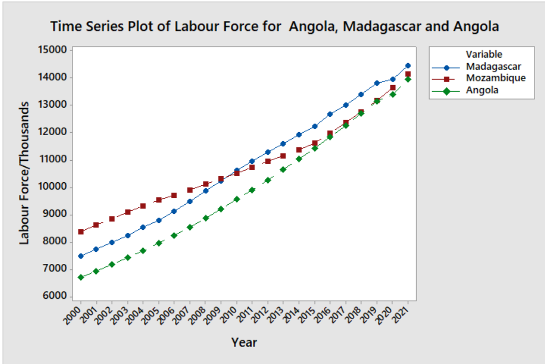
With reference to Figure 4.7, South Africa and Angola have the highest international reserves, with South Africa having significantly higher levels than Angola. Botswana had the highest international reserves of the countries with relatively high international reserves, until 2017. Currently, Mauritius is the one with the highest international reserves of the countries with relatively high international reserves. SADC countries with relatively low international reserves, Madagascar does not have so much low international reserves as the rest. From Table 4.7, South Africa had the highest average level of international reserves while Comoros had the least average level of international reserves.

4.9 Labour Force for SADC Nations

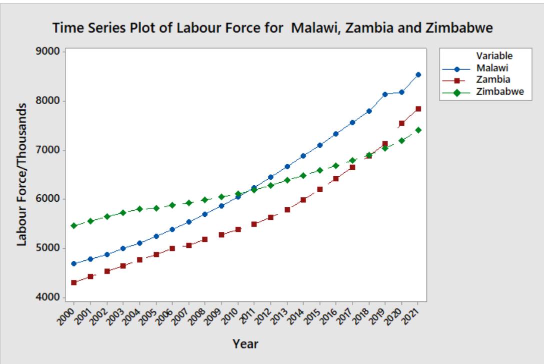
Figures 4.8(a) to 4.8(d) shows time series plots illustrating the levels of labour force across SADC nations.



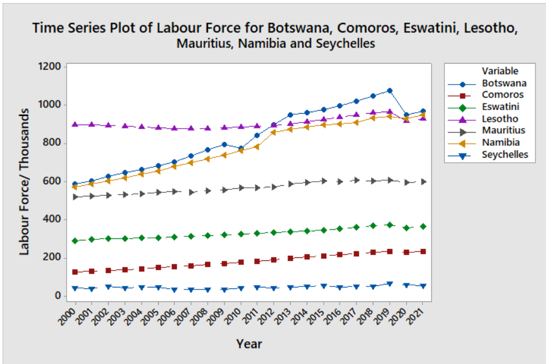
(a)



(b)



(c)



(d)

Figure 4.8: Time Series Plots of Labor Force for SADC Nations 2000 to 2021

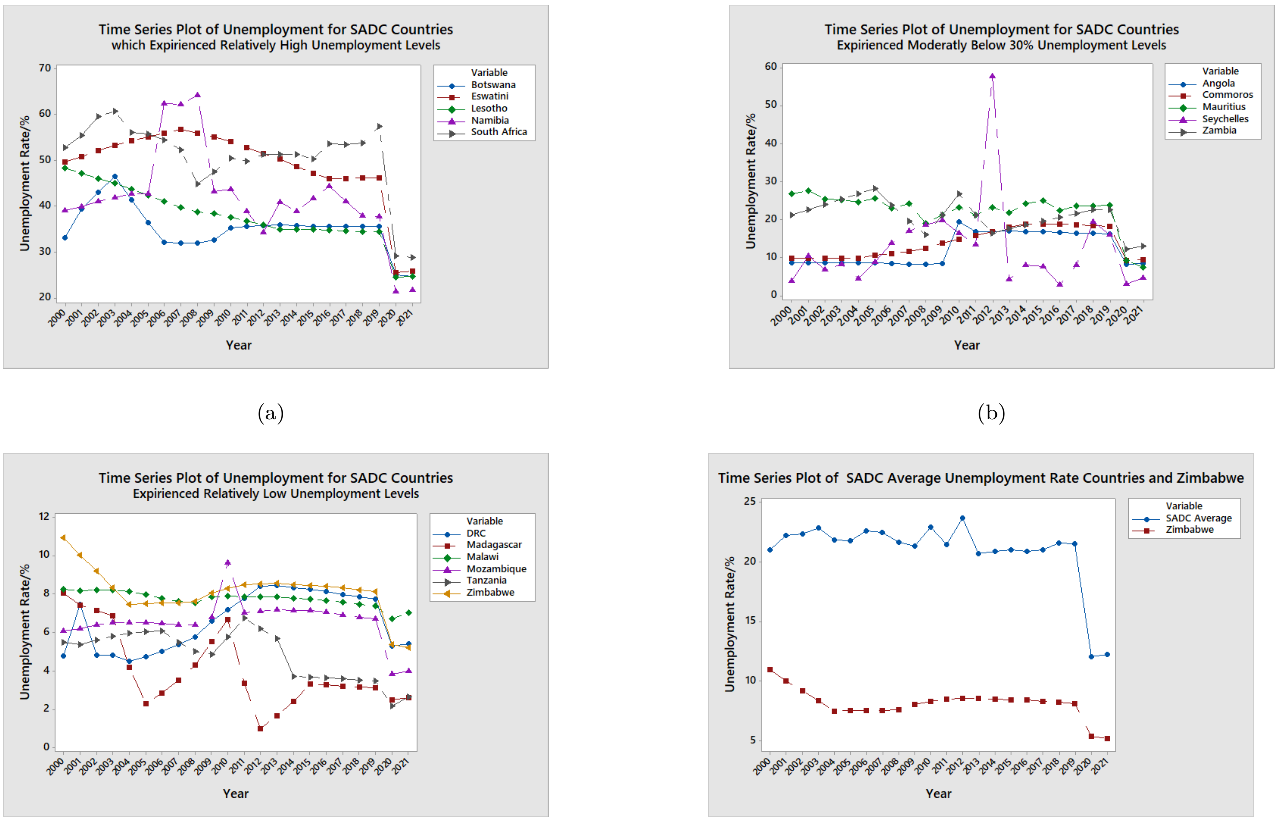
Table 4.8: Summary Statistics of Labour Force for SADC Nations 2000-2021

Country	**Mean**	**Standard Error**	**Median**	**Standard Dev**	**Sample Variance**	**Range**	**Minimum**	**Maximum**
Angola	9952.8	484.8	9734.5	2274	5170891.3	7259	6712	13971
Botswana	763	23.3	782	109	11925	383	588	971
Comoros	181.5	7.9	180	36.9	1359.8	108	127	235
DRC	24126.5	868.8	23215.5	4074.8	16604300.8	14341	18365	32706
Eswatini	328.7	5.5	325	25.6	654	84	290	374
Lesotho	907.2	5.8	896.5	27.4	750.4	88	879	967
Madagascar	10656.3	518.4	10797	2431.7	5913204.1	8956	5506	14462
Malawi	6182.9	251.4	5957	1179.1	1390270.6	3861	4682	8543
Mauritius	567.7	6.5	567.5	30.6	937.5	89	519	608
Mozambique	10836.1	353.7	10629	1658.8	2751648.2	5761	8385	14146
Namibia	779.6	28.3	772	132.7	17601.4	381	571	952
Seychelles	45.9	1.8	46	8.3	69	32	33	65
South Africa	20352.7	443	20052	2077.9	4317463.8	6863	16463	23326
Tanzania	21994.4	832.1	21421.5	3903	15233641.7	12913	16130	29043
Zambia	5683.6	219.9	5441.5	1031.4	1063701.8	3539	4301	7840
Zimbabwe	6270.4	118.4	6152	555.3	308312.1	1946	5469	7415

Referring to Figure 4.8, DRC and South Africa have the highest labour force of all SADC countries, with DRC being the one at the peak. The labour force has been gradually increasing for Angola, Madagascar and Mozambique, countries with a relatively high labour force. Also countries with a relatively low labour force, Zimbabwe, Malawi and Zambia, they has been a gradual increase in their labour force. From Table 4.8, DRC had the highest mean labour force level whereas Seychelles had the least average labour force level.

4.10 Unemployment Levels in SADC Nations

Figures 4.9(a) to 4.9(d) shows time series plots illustrating unemployment levels across SADC nations.



(a) (b) (c) (d)

Figure 4.9: Time Series Plots of Unemployment Levels for SADC Nations 2000 to 2021

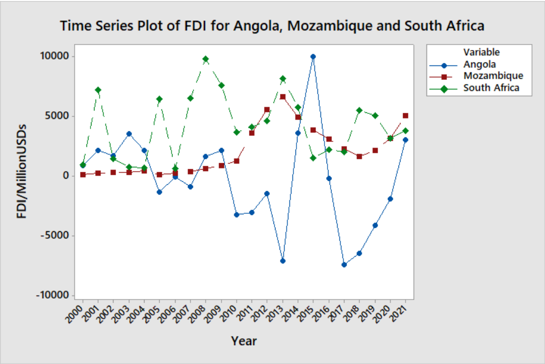
Table 4.9: Summary Statistics of Unemployment Levels for SADC Nations 2000-2021

Country	**Mean**	**Standard Error**	**Median**	**Standard Dev**	**Sample Variance**	**Range**	**Minimum**	**Maximum**
Angola	12.370	0.928	8.71	4.352	18.939	11.23	8.26	19.49
Botswana	35.206	1.051	35.625	4.928	24.288	21.9	24.7	46.6
Comoros	13.857	0.824	13.135	3.865	14.936	9.68	9.22	18.9
DRC	6.572	0.321	6.885	1.505	2.266	3.95	4.48	8.43
Eswatini	49.083	1.784	51.2	8.370	70.053	31.23	25.51	56.74
Lesotho	37.885	1.326	37.24	6.218	38.660	23.76	24.56	48.32
Madagascar	4.014	0.430	3.29	2.016	4.064	7.04	1	8.04
Malawi	7.749	0.081	7.815	0.380	0.144	1.54	6.7	8.24
Mauritius	22.361	1.053	23.645	4.940	24.401	20.12	7.41	27.53
Mozambique	6.583	0.238	6.625	1.116	1.245	5.83	3.81	9.64
Namibia	41.898	2.227	41.02	10.447	109.147	42.75	21.45	64.2
Seychelles	12.454	2.472	8.525	11.596	134.475	54.98	2.86	57.84
South Africa	50.948	1.702	52.59	7.985	63.763	32.03	28.8	60.83
Tanzania	4.846	0.276	5.43	1.294	1.675	4.59	2.16	6.75
Zambia	21.027	0.901	21.29	4.224	17.843	16.04	12.17	28.21
Zimbabwe	8.119	0.262	8.305	1.228	1.508	5.75	5.2	10.95

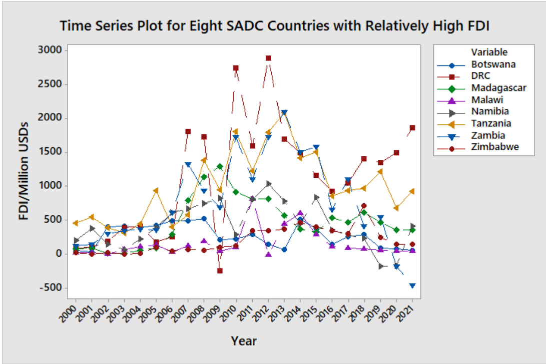
Looking at figure 4.9, Countries that experienced relatively high unemployment levels are Botswana, Eswatini, Lesotho, Namibia and South Africa, unemployment levels dropped significantly in 2020 and 2021. SADC countries which experienced moderately below 30% unemployment levels, Seychelles experienced a very high unemployment level in 2012. DRC, Madagascar, Malawi, Mozambique, Tanzania and Zimbabwe are the six SADC countries that experienced relatively low unemployment levels. Zimbabwe has a lower unemployment rate than SADC's average unemployment. From Table 4.9, South Africa had the highest average level of unemployment while Madagascar had the least average unemployment level.

4.11 Foreign Direct Investment Inflows for SADC Nations

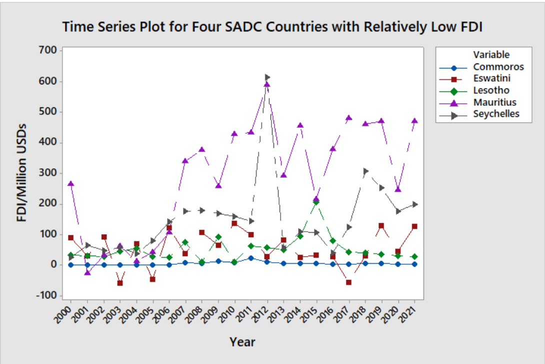
Figures 4.10(a) to 4.10(d) shows time series plots illustrating foreign direct investment inflows across SADC nations.



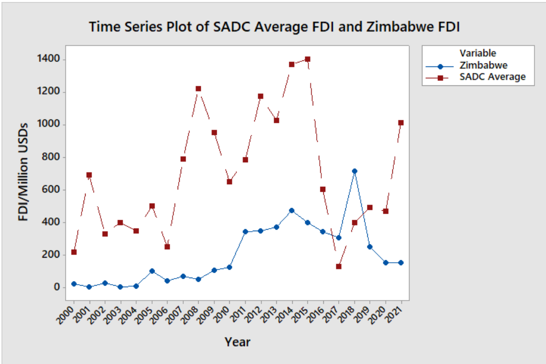
(a)



(b)



(c)



(d)

Figure 4.10: Time series Plot of Foreign Direct Investment Inflows for SADC Nations 2000 to 2021

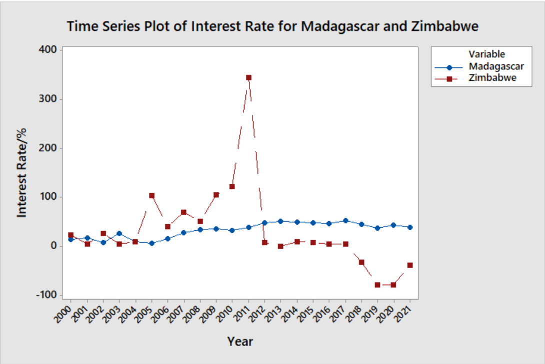
Table 4.10: Summary Statistics of Foreign Direct Investment for SADC Nations 2000-2021

Country	**Mean**	**Standard Error**	**Median**	**Standard Dev**	**Sample Variance**	**Range**	**Minimum**	**Maximum**
Angola	-266.571	866.99	-108.616	4066.536	16536714.01	17425.51	-7397.29	10028.21
Botswana	271.503	36.24	273.265	169.972	28890.51	490.23	30.68	520.91
Comoros	5.077	1.13	4.11	5.321	28.31	23.02	0.093	23.11
DRC	1118.718	184.24	1258.35	864.169	746787.43	3134.80	-243.2	2891.60
Eswatini	54.760	12.33	54.88	57.845	3346.01	195.85	-60.19	135.66
Lesotho	52.398	8.90	41.52	41.76	1744.54	197	9.51	206.51
Madagascar	476.032	77.91	418.86	365.414	133527.6108	1280.45	12.87	1293.33
Malawi	156.391	44.26	83.60	207.614	43103.46	821.63	-8.88	812.75
Mauritius	290.754	39.09	317.066	183.346	33615.8178	616.69	-27.67	589.01
Mozambique	2165.182	439.98	1468.3	2063.687	4258803.64	6575	122.4	6697.4
Namibia	427.332	70.14	384.6	328.989	108233.61	1218.1	-176.5	1041.6
Seychelles	148.514	27.11	132.55	127.134	16163.15	588.9	24.3	613.2
South Africa	4202.618	584.59	3994.15	2741.964	7518365.97	9261.7	623.3	9885
Tanzania	994.277	108.67	936.6	509.73	259824.659	1768.9	318.4	2087.3
Zambia	775.573	143.73	639.3	674.15	454479.92	2556.8	-457	2099.8
Zimbabwe	200.632	40.54	136.45	190.15	36157.94	714.1	3.8	717.9

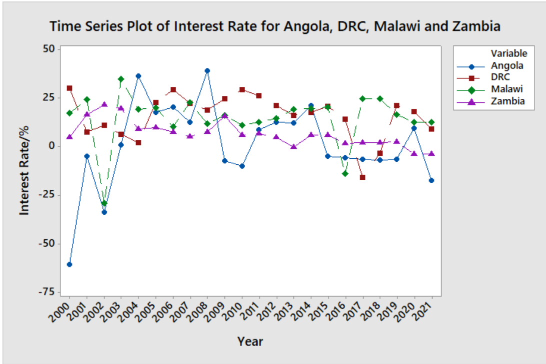
Figure 4.10 shows that, Angola, Mozambique and South Africa are countries with the highest levels of foreign direct investment. Botswana, DRC, Madagascar, Malawi, Namibia, Tanzania, Zambia, Zimbabwe are the countries with relatively high foreign direct investment. SADC countries with relatively low foreign direct investment, Seychelles and Mauritius, were not so low. SADC's average foreign direct investment has been generally higher than Zimbabwe except in 2018. From Table 4.10 South Africa had the highest average FDI level and Angola had the least.

4.12 Interest Rates in SADC Nations

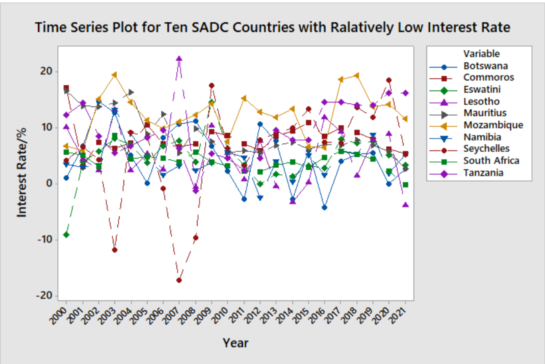
Figures 4.11(a) to 4.11(d) shows time series plots illustrating foreign direct investment inflows across SADC nations.



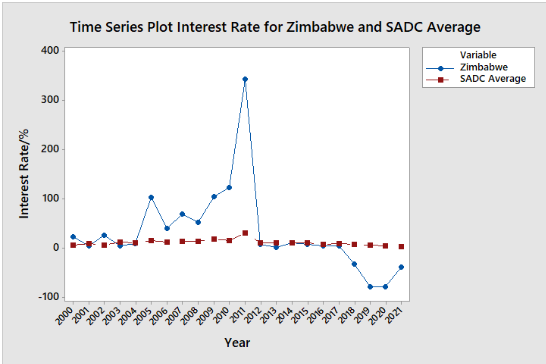
(a)



(b)



(c)



(d)

Figure 4.11: Time Series Plot of Interest Rates for SADC Nations 2000 to 2021

Table 4.11: Summary Statistics of Interest Rates for SADC Nations 2000-2021

Country	**Mean**	**Standard Error**	**Median**	**Standard Dev**	**Sample Variance**	**Range**	**Minimum**	**Maximum**
Angola	1.265	4.683	-1.897	21.964	482.42	99.756	-60.78	38.976
Botswana	5.156	1.103	5.4625	5.175	26.78	18.999	-4.154	14.845
Comoros	8.249	0.581	7.7535	2.725	7.43	14.126	3.102	17.228
DRC	15.897	2.424	18.422	11.370	129.27	45.69	-15.69	30
Eswatini	4.723	0.932	4.596	4.374	19.13	23.67	-9.05	14.62
Lesotho	5.479	1.306	5.014	6.128	37.55	26.012	-3.7	22.312
Madagascar	32.773	3.244	36.388	15.215	231.48	46.091	6.346	52.437
Malawi	14.681	2.833	16.983	13.288	176.57	64.175	-29.221	34.954
Mauritius	8.899	0.845	7.2885	3.965	15.72	13.958	2.8	16.758
Mozambique	12.455	0.862	12.637	4.043	16.34	13.488	6.009	19.497
Namibia	4.360	0.690	3.993	3.235	10.47	15.605	-2.344	13.261
Seychelles	5.733	1.888	7.383	8.856	78.43	35.748	-17.158	18.59
South Africa	4.165	0.375	4.255	1.760	3.10	8.763	-0.1	8.663
Tanzania	9.265	1.027	8.437	4.819	23.22	17.502	-1.202	16.3
Zambia	6.741	1.430	5.967	6.707	44.99	25.366	-3.75	21.616
Zimbabwe	32.007	18.516	8.283	86.847	7542.36	424.103	-79.803	344.3

Referring to Figure4.11. Zimbabwe's interest rate was highest in 2011, and went lower than Madagascar in 2012. Angola started off with low interest rate as compared to DRC, Malawi and Zambia, and experienced the highest interest rate in 2008. SADC countries with relatively low interest rates, Seychelles experienced the lowest interest rate in 2007 and Lesotho experienced the highest interest rate in 2007. Zimbabwe's interest rate has been fluctuating experiencing the highest and lowest interest rate as compared to SADC's average interest rate. From Table 4.11, Zimbabwe had the largest mean interest rate while Angola had the least.

4.13 Analysis of Zimbabwean Economic Data

4.13.1 Correlation Matrix for Zimbabwes Data

From Table 4.12, it is evident that; GDP and Inflation have a weak negative association(-0.29) and it means as the GDP increases, Inflation will gradually decrease. GDP and exports have a strong positive correlation (0.87), implying that, both variables tend to decrease together and also increase together. Inflation and imports have a weak negative association -0.18, meaning

Table 4.12: Correlation Matrix for Zimbabwe Data 2000 to 2021

	GDP	Inflation	Imports	Exports	EDebt	ERate	IReserves	LForce	Unemployment	FDI	IRate
GDP	1.00	-0.29	0.75	0.87	0.72	-0.31	-0.11	0.87	-0.11	0.82	-0.30
Inflation		1.00	-0.18	-0.23	-0.12	0.99	0.12	-0.11	-0.14	-0.18	0.04
Imports			1.00	0.70	0.46	-0.23	-0.07	0.66	-0.24	0.67	0.12
Exports				1.00	0.85	-0.26	-0.15	0.86	-0.28	0.65	-0.22
EDebt					1.00	-0.16	-0.29	0.90	-0.54	0.42	-0.45
ERate						1.00	-0.11	-0.12	-0.16	-0.20	0.05
IReserves							1.00	-0.36	0.59	-0.17	0.14
LForce								1.00	-0.48	0.71	-0.36
Unemployment									1.00	-0.02	0.11
FDI										1.00	-0.08
IRate											1.00

Source: Author's own results.

either imports go down as inflation goes up. Unemployment and FDI have a weak positive correlation. External debt and unemployment have a moderate negative relationship (-0.54) meaning unemployment will go up as external debt goes down or the other way round. Labour force and external debt have a very strong positive relationship (0.90), meaning,if labour force goes up, external debt also goes up. Interest rate and interest rate(same variables) will always have a perfect relationship.

4.13.2 Model without Data Transforms

Table 4.13 shows the results of the Zim_1 model, which is a Multiple linear regression model that regresses GDP for Zimbabwe against inflation, imports, exports, external debt, exchange rate, international reserves, labour force, unemployment, foreign direct investment and interest rate. Significance codes from Table 4.13 show that the intercept, imports, external debt, labour force, and unemployment have statistical significance in predicting the GDP for Zimbabwe. The Zim_1 model has an Akaike Information Criterion (AIC) of 382.130 and BIC of 394.66.

Table 4.14 displays the results of Zim_2, a model that is a slight modification of Zim_1. The exchange rate is removed from the zim_1 model in the zim_2 model because the correlation matrix 4.12, two of the explanatory variables inflation and exchange rate are perfectly correlated, implying the variables have the same contribution to the model, and the exchange rate has a higher p-value of 0.84771 than inflation, which is 0.71297. The removal of the variable exchange rate from the model demonstrates that the model has improved, as evidenced by the decrease in the AIC and BIC values.

Table 4.13: Zim_1 Model without Data Transforms

Zim_1 Model Without Data Transformation					
Call:					
glm(formula = GDP $\sim$ Inflation + Imports + Exports + EDebt + ERate + IReserves + LForce+ Unemployment + FDI + IRate, family = "gaussian", data = Zim_Data)					
Covariates	**Estimate**	**Std.Error**	**t-value**	**P-value**	**Signif codes**
(Intercept)	-91140	22000	-4.142	0.00201	**
Inflaton	-0.00002876	0.00007597	-0.379	0.71297	
Imports	0.802	0.3271	2.452	0.03415	*
Exports	1.203	0.7743	1.554	0.15115	
ExternalDebt	-1.772	0.7789	-2.275	0.04616	*
ExchangeRate	0.00001545	0.00007839	0.197	0.84771	
InternatonalReserves	-0.5221	0.8527	-0.612	0.55402	
LabourForce	15.11	3.918	3.858	0.00317	**
Unemployment	1736	706.1	2.459	0.03373	*
ForeignDirectInvestment	-5.998	5.817	-1.031	0.32682	
Interestrate	-10.47	6.778	-1.545	0.15343	
Model Summary					
AIC:	382.130				
BIC:	394.66				

Source: Author's own results.

The variable International reserves is removed from the Zim_2 model to form the Zim_3 model using a similar approach used in constructing the Zim_2 model. Like in Zim_2 model, the quality of the model also improves evidenced by a higher AIC and BIC values of 378.9155 and 389.3607 respectively. The variables Exports,Foreign direct Investment remain statistically insignificant.

Table 4.14: Zim_2 Model without Data Transforms

Zim_2 Model Without Data Transformation					
Call:					
glm(formula = GDP $\sim$ Inflation + Imports + Exports + EDebt + IReserves + LForce+ Unemployment + FDI + IRate, family = "gaussian", data = Zim_Data)					
Covariates	**Estimate**	**Std.Error**	**t-value**	**P-value**	**Signif codes**
(Intercept)	-91020	21010	-4.332	0.00119	**
Inflaton	-0.00001389	0.000008639	-1.608	0.13619	
Imports	0.766	0.2592	2.956	0.01308	*
Exports	1.203	0.7397	1.627	0.13207	
ExternalDebt	-1.779	0.7432	-2.394	0.0356	*
InternatonalReserves	-0.489	0.7987	-0.612	0.55278	
LabourForce	15.19	3.725	4.077	0.00183	**
Unemployment	1690	635.4	2.659	0.02222	*
ForeignDirectInvestment	-5.847	5.509	-1.061	0.31129	
Interestrate	-10.11	6.229	-1.622	0.13301	
Model Summary					
AIC:	380.2117				
BIC:	391.7014				

Source: Author's own results.

Table 4.15: Zim_3 Model without Data Transforms

Zim_3 Model Without Data Transformation					
Call:					
glm(formula = GDP ~ Inflation + Imports + EDebt + Exports + LForce + Unemployment + FDI + IRate, family = "gaussian", data = Zim_Data)					
Covariates	**Estimate**	**Std.Error**	**t-value**	**P-value**	**Signif codes**
(Intercept)	-90970	20460	-4.447	0.000798	***
Inflaton	-0.00001569	0.00000791	-1.983	0.070691	.
Imports	0.727	0.2446	2.972	0.011648	*
Exports	1.168	0.7179	1.627	0.129763	
ExternalDebt	-1.826	0.7198	-2.537	0.026069	*
LabourForce	15.52	3.589	4.322	0.000992	***
Unemployment	1478	518.7	2.849	0.014662	*
ForeignDirectInvestment	-5.485	5.333	-1.029	0.32395	
Interestrate	-10.13	6.065	-1.671	0.120566	
Model Summary					
AIC:	378.9156				
BIC:	389.3607				

Source: Author's own results.

Table 4.16: Zim_4 Model without Data Transforms

Zim _4 Model Without Data Transformation					
Call:					
glm(formula = GDP ~ Inflation + ImportsExports + EDebt + LForce + IRate , family = "gaussian", data = Zim_Data)					
Covariates	**Estimate**	**Std.Error**	**t-value**	**P-value**	**Signif codes**
(Intercept)	-74220	12420	-5.977	0.0000462	***
Inflaton	-0.00001434	0.000007818	-1.834	0.0896	.
Imports	0.6371	0.2289	2.783	0.0155	*
Exports	0.9253	0.6796	1.361	0.1965	
ExternalDebt	-1.273	0.4797	-2.654	0.0198	*
LabourForce	12.47	2.036	6.126	0.0000362	***
Unemployment	1338	501.8	2.667	0.0194	*
Interestrate	-8.975	5.972	-1.503	0.1568	
Model Summary					
AIC:	378.6899				
BIC:	388.0906				

Source: Author's own results.

The process of eliminating statistically insignificant variables continues as explained previously in Zim_2 model formulation, this time with the removal of the variable foreign direct investment from Zim_3 to form Zim_4 model, which shows that the quality of the model continues to increase as more variables are eliminated.

Table 4.17: Zim_5 Model without Data Transforms

Zim _5 Model Without Data Transformation					
Call:					
glm(formula = GDP ~ Inflation + Imports + EDebt + Unemployment + LForce + IRate, family = "gaussian", data = Zim_Data)					
Covariates	**Estimate**	**Std.Error**	**t-value**	**P-value**	**Signif codes**
(Intercept)	-82410	11190	-7.365	0.00000354	***
Inflaton	-0.00001577	0.00000798	-1.976	0.0682	.
Imports	0.7143	0.2284	3.127	0.00742	**
ExternalDebt	-0.9008	0.4059	-2.22	0.04348	*
LabourForce	13.41	1.973	6.795	0.00000867	***
Unemployment	1667	452.9	3.682	0.00247	**
Interestrate	-6.449	5.847	-1.103	0.28868	
Model Summary					
AIC:	379.49				
BIC:	387.8454				

Source: Author's own results.

The process of eliminating statistically insignificant variables continues as explained previously in Zim_2 model formulation, now removing the variable Exports from Zim_4 to form Zim_5 model. However, the quality of the decreases evidenced by the higher AIC of 379.49.

From the analysis carried out above, the Zim_4 model;

$$\begin{aligned} GDP = & -74220 - 0.00004\ \text{Inflation} + 0.6371\ \text{Imports} + 0.9253\ \text{Exports} - 1.273\ \text{EDebt} + \\ & 12.47\ \text{LForce} + 1338\ \text{Unemployment} - 8.975\ \text{IRate} \end{aligned} \tag{4.1}$$

was chosen as the best model to describe the relationship between GDP and other Zimbabwes economic variables for the data without transforms. This conclusion is reached because the Zim_4 model had the lowest AIC and BIC values.

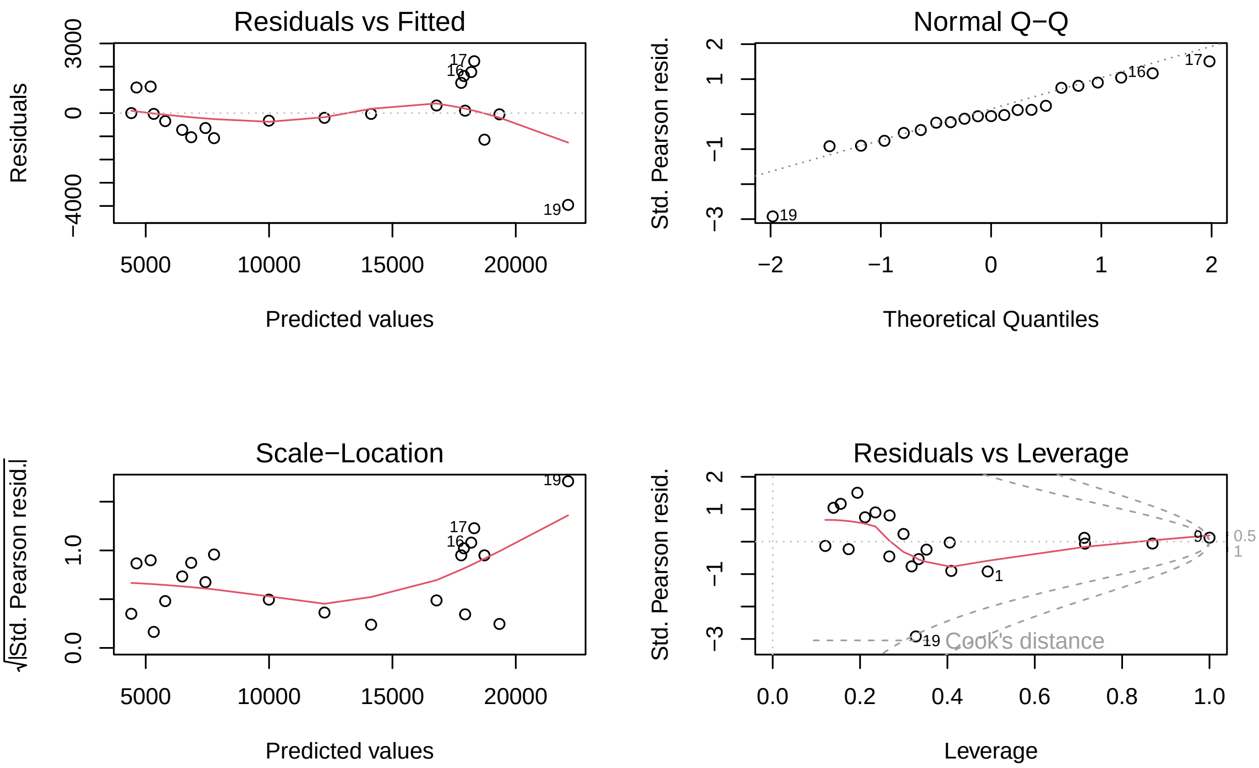


Figure 4.12: Zim_4 Model Residual Plots

Source: Author's own results.

The residual vs fitted plots in fig 4.12 show that there are few points further away from the zero line implying that data used to create the model has outliers. One way to fix this problem is take data transformation, the subject of the next section. The linear relationship observed in the qq plot fig 4.12 suggest that the residuals are normally distributed implying the model fitted is good.

4.13.3 Model with Data Transforms

As stated in the previous section the presence of outliers had an effect on the models formulated hence the models created in this section will try to address this shortcoming. Log transformations will be used to achieve this goal.

Table 4.18: M_1 Model with Data Transforms

M_1 Model With Data Transformation					
Call: glm(formula = log (GDP) $\sim$ log (Inflation) + log (Imports) + log (Exports) + log (EDebt) + log (ERate)+ log (IReserves) + log (LForce) + log (Unemployment) + log (FDI) + log(IRate), family = "gaussian", data = Zim_Data)					
Covariates	**Estimate**	**Std.Error**	**t-value**	**P-value**	**Signif codes**
(Intercept)	-25.486001	5.458657	-4.669	0.01855	*
log(Inflation)	-0.01803	0.006993	-2.578	0.08192	.
log(Imports)	0.007525	0.008873	0.848	0.45874	
log(Exports)	0.429373	0.108436	3.96	0.02876	*
log(External Debt)	-0.71874	0.330971	-2.172	0.11828	
log(Exchange Rate)	-0.020249	0.004267	-4.746	0.01775	*
log(International Reserves)	0.018627	0.016686	1.116	0.34564	
log(Labour Force)	4.304677	0.998708	4.31	0.02299	*
log(Unemployment)	-0.105332	0.169719	-0.621	0.57881	
log(Foreign Direct Investment)	0.058183	0.013562	4.29	0.02328	*
log(Interest rate)	-0.043968	0.005376	-8.179	0.00382	**
Model Summary					
R-Squared:	0.9994				
Adj. R-Squared:	0.9976				
F-statistic:	530.9 on 10 and 3 DF				
p-value:	0.000121				

Source: Author's own results.

Table 4.18, shows the results of the M_1 model, which is a Multiple linear regression model with log transforms that regresses log of GDP for Zimbabwe, against log of inflation, log of imports, log of exports, log of external debt, log of exchange rate, log of international reserves, log of labour force,log of unemployment, log of foreign direct investment and log of interest rate. Significance codes from Table 4.18, show that the intercept, log exports, log exchange rate ,log labour force, log FDI and log Interest rate have statistical significance in predicting the GDP for Zimbabwe. The M_1 model has an R^2 value of 0.994 and Adj R^2 of 0.9976.

Table 4.19: M_2 Model with Data Transforms

M_2 Model With Data Transformation					
Call:					
glm(formula = log (GDP) ∼ log (Imports) + log (Exports) + log (EDebt) + log (ERate) + log (IReserves)+ log (LForce) + log (FDI) + log(IRate) + log(Inflation), family = "gaussian", data = Zim_Data)					
Covariates	**Estimate**	**Std.Error**	**t-value**	**P-value**	**Signif codes**
(Intercept)	-24.695496	4.883024	-5.057	0.007194	**
log(Inflaton)	-0.016966	0.006237	-2.72	0.052983	.
log(Imports)	0.006304	0.00796	0.792	0.472696	
log(Exports)	0.452887	0.093467	4.845	0.008367	**
log(ExternalDebt)	-0.600576	0.24905	-2.411	0.073438	.
log(ExchangeRate)	-0.019137	0.003562	-5.372	0.005799	**
log(InternatonalReserves)	0.013664	0.013472	1.014	0.367843	
log(LabourForce)	4.054623	0.84067	4.823	0.008504	**
log(ForeignDirectInvestment)	0.05561	0.011879	4.681	0.009437	**
log(Interestrate)	-0.042758	0.004609	-9.277	0.000751	***
Model Summary					
R-Squared:	0.9994				
Adj. R-Squared:	0.9979				
F-statistic:	696.9 on 9 and 4 DF				
p-value:	5.019e-06				

Source: Author's own results.

The model M_2 in Table 4.19 is a slight variation of M_1 model because the variable log Unemployment was removed from model M_2. This is because the variable log Unemployment had the highest p-value amongst the statistically insignificant variables. Removal of the log unemployment from model M_2 resulted in an improved model evidenced by both higher $R^2 = 0.9994$ and Multiple $R^2 = 0.9979$ than that of M_1.

Table 4.20: M_3 Model with Data Transforms

M_3 Model With Data Transformation					
Call:					
glm(formula = log (GDP) ∼ log (Inflation) + log (Exports) + log (EDebt) + log (ERate) + log (LForce)+ log(IReserves) + log (FDI) + log(IRate) family = "gaussian", data = Zim_Data)					
Covariates	**Estimate**	**Std.Error**	**t-value**	**P-value**	**Signif codes**
(Intercept)	-22.75033	4.059992	-5.604	0.002501	**
log(Inflaton)	-0.015073	0.005542	-2.72	0.041793	*
log(Exports)	0.502041	0.06723	7.468	0.00068	***
log(ExternalDebt)	-0.454476	0.160957	-2.824	0.036953	*
log(ExchangeRate)	-0.018926	0.003417	-5.538	0.002633	**
log(InternatonalReserves)	0.013161	0.012946	1.017	0.355991	
log(LabourForce)	3.650184	0.64241	5.682	0.002352	**
log(ForeignDirectInvestment)	0.050989	0.009954	5.123	0.003699	**
log(Interestrate)	-0.043164	0.004406	-9.796	0.000189	***
Model Summary					
R-Squared:	0.9993				
Adj. R-Squared:	0.9981				
F-statistic:	871.1 on 8 and 5 DF				
p-value:	2.128e-07				

Source: Author's own results.

Similar to model M_2, Model M_3 in Table 4.20 is formed by eliminating the variable log imports from model M_2 because it had the highest p-value amongst the statistically insignificant values. Surprisingly, the variable log inflation and log imports which where statistically insignificant variables in model M_2 become statistically significant. The adjusted R^2 of 0.9981 show a slight improvement of the model.

Table 4.21: M_4 Model with Data Transforms

M_4 Model With Data Transformation					
Call:					
glm(formula = log (GDP) $\sim$ log(Inflation) + log (Imports) + log (Exports) + log (EDebt) + log (ERate)+ log (LForce) + log (FDI) + log(IRate), family = "gaussian", data = Zim_Data)					
Covariates	**Estimate**	**Std.Error**	**t-value**	**P-value**	**Signif codes**
(Intercept)	-19.767131	2.813531	-7.026	0.000415	***
log(Inflaton)	-0.010639	0.003429	-3.103	0.021039	*
log(Exports)	0.545516	0.052019	10.487	0.0000441	***
log(ExternalDebt)	-0.35501	0.12816	-2.77	0.03242	*
log(ExchangeRate)	-0.020688	0.002954	-7.004	0.000422	***
log(LabourForce)	3.176269	0.443229	7.166	0.000373	***
log(ForeignDirectInvestment)	0.053388	0.009697	5.506	0.001507	**
log(Interestrate)	-0.043513	0.004405	-9.877	0.0000621	***
Model Summary					
R-Squared:	0.9991				
Adj. R-Squared:	0.9981				
F-statistic:	962 on 7 and 6 DF				
p-value:	1.015e-8				

Source: Author's own results.

The variable log international reserves was removed from the model M_3 to form M_4 because it was the only variable that was statically insignificant with a p-value of 0.356. M_4 shows an decrease in the multiple $R^2 = 0.9981$ value, implying that model M_4 is a slight decrease in the quality of the model.

From the analysis carried out above, model M_3;

$$\begin{aligned} \log(\text{GDP}) = -22.7 - 0.015 \log(\text{Inflation}) + 0.502 \log(\text{Exports}) - 0.4544 \log(\text{EDebt})- \\ 0.0189 \log(\text{ERate}) + 3.65 \log(\text{LForce}) + 0.051 \log(\text{FDI})- \\ 0.043 \log(\text{IRate}) \end{aligned} \tag{4.2}$$

is the model that describes the relationship between transformed Zimbabwean economic variables and GDP. This conclusion was reached because, the Model has the highest Multiple R^2 and R^2 among all the models with data transforms. This model shows that 99.81% of the variation of Zimbabwean domestic product is explained by log inflation,log exports, log external debt, log exchange rate, log labour force, log FDI and log interest rate.

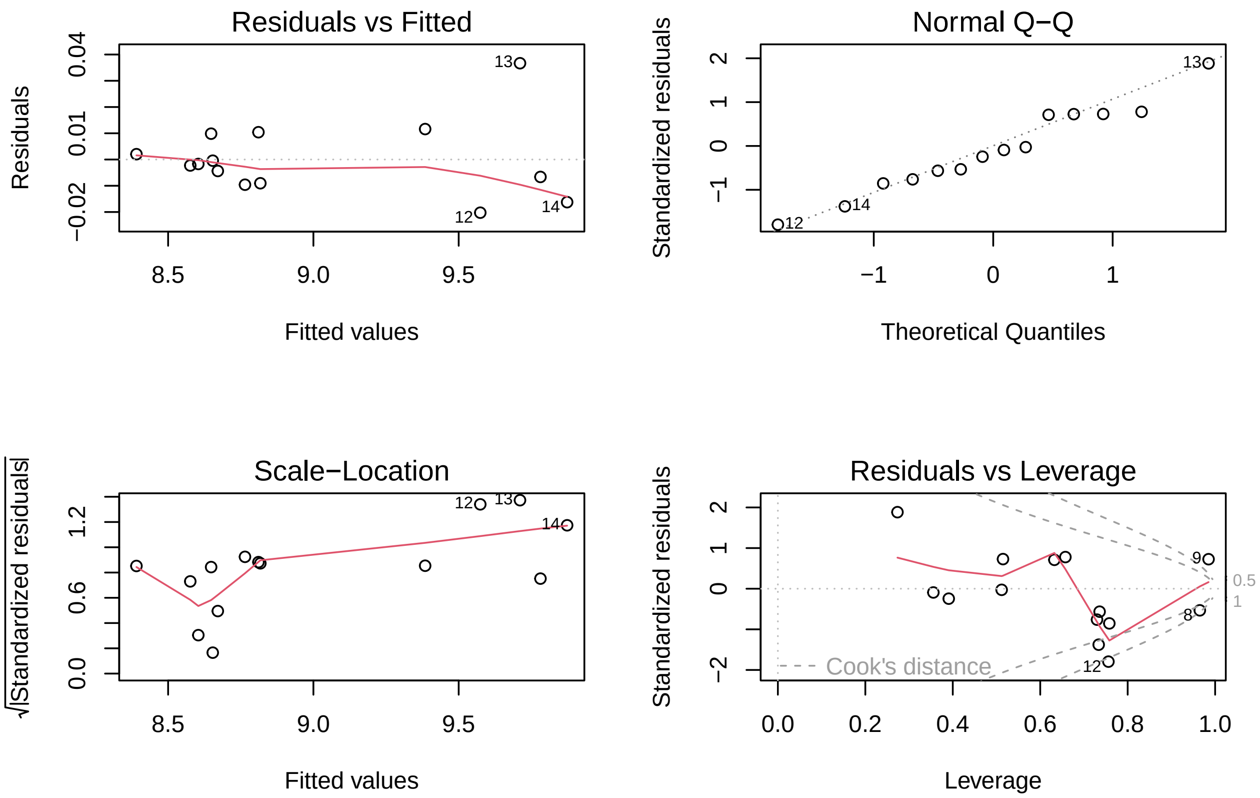


Figure 4.13: M_3 Model Residual Plots

Source: Author's own results.

The residual vs fitted plot in fig 4.13 show that the residuals are randomly distributed and the model is not affected by any outliers. The normal q-q plot is almost linear suggesting that the residuals are normally distributed. Overall, the residual plots in fig 4.13 shows that the chosen model describes the Zimbabwean GDP very well.

4.14 SADC Results

Table 4.22: Summary Results for Specific Models.

Covariates	Estimate	Std.Error	t-value	P-value	Signif codes
OLS Results for SADC Data					
(Intercept)	$-1.036e+04$	$1.651e+03$	-6.274	$1.07e-09$	$***$
Inflaton	$-2.759e-04$	$4.209e-04$	-0.655	0.512659	
Imports	$2.734e+00$	$1.491e-01$	18.337	$< 2e-16$	$***$
Exports	$3.484e-01$	$1.047e-01$	3.326	0.000976	$***$
EDebt	$5.342e-01$	$7.458e-02$	7.162	$4.92e-12$	$***$
ERate	$2.742e-04$	$4.198e-04$	0.653	0.514145	
IReserves	$6.631e-02$	$2.445e-01$	0.271	0.786360	
LForce	$6.904e-01$	$1.148e-01$	6.016	$4.61e-09$	$***$
Unemployment	$2.030e+02$	$5.233e+01$	3.879	0.000126	$**$
FDI	$-1.232e+00$	$5.165e-01$	-2.385	0.017641	$*$
IRate	$-2.688e+01$	$2.799e+01$	-0.960	0.337489	
Total Sum of Squares:	$2.3377e+12$				
Residual Sum of Squares:	$5.3237e+10$				
R-Squared:	0.97723				
Adj. R-Squared:	0.97656				
F-statistic:	1463.28 on	10 and 341	DF		
p-value	$< 2.22e-16$				
Fixed Effects Model (FEWITHIN) for SADC Data					
Inflaton	$-1.7634e-04$	$3.7515e-04$	-0.4701	0.638625	
Imports	$1.9377e+00$	$1.6422e-01$	11.7998	$< 2.2e-16$	$***$
Exports	$2.9410e-01$	$1.1007e-01$	2.6719	0.007920	$**$
EDebt	$2.4733e-01$	$7.7163e-02$	3.2053	0.001483	$**$
ERate	$1.6558e-04$	$3.7598e-04$	0.4404	0.659944	
IReserves	$1.1309e+00$	$2.7618e-01$	4.0949	$5.333e-05$	$***$
LForce	$1.5977e+00$	$3.6787e-01$	4.3432	$1.877e-05$	$***$
Unemployment	$1.8080e+02$	$1.0585e+02$	1.7080	0.088584	.
FDI	$-1.2494e+00$	$4.7780e-01$	-2.6149	0.009341	$**$
IRate	$-3.1631e+01$	$2.5709e+01$	-1.2304	0.219449	

Total Sum of Squares:	$2.5263e+11$				
Residual Sum of Squares:	$3.9218e+10$				
R-Squared:	0.84476				
Adj. R-Squared:	0.83286				
F-statistic:	177.398 on	10 and 326	DF		
p-value	$< 2.22e-16$				
Fixed Effects Model (BETWEEN) for SADC Data					
Intercept	-1699.48543	4537.27743	-0.3746	0.72335	
Inflaton	0.46130	1.76911	0.2608	0.80468	
Imports	2.74241	1.28630	2.1320	0.08618	
Exports	-0.38734	$0.41669-$	0.9296	0.39526	
EDebt	1.54063	0.86147	1.7884	0.13374	
ERate	-0.40994	1.56944	-0.2612	0.80435	
IReserves	-0.16692	0.89035	-0.1875	0.85866	
LForce	0.15009	0.22451	0.6685	0.53341	
Unemployment	-23.15248	133.68740	-0.1732	0.86930	
FDI	-4.20909	2.66623	-1.5787	0.17525	
IRate	28.91778	209.28400	0.1382	0.89549	
Total Sum of Squares:	$9.4777e+10$				
Residual Sum of Squares:	71069000				
R-Squared:	0.99925				
Adj. R-Squared:	0.99775				
F-statistic:	666.293 on	10 and 5	DF		
p-value	$3.6046e-07$				
Fixed Effects Model (LSDV) for SADC Data					
Inflaton	$-1.763e-04$	$3.751e-04$	-0.470	0.638625	
Imports	$1.938e+00$	$1.642e-01$	11.800	$< 2e-16$	$***$
Exports	$2.941e-01$	$1.101e-01$	2.672	0.007920	$**$
EDebt	$2.473e-01$	$7.716e-02$	3.205	0.001483	$**$
ERate	$1.656e-04$	$3.760e-04$	0.440	0.659944	
IReserves	$1.131e+00$	$2.762e-01$	4.095	$5.33e-05$	$***$

LForce	$1.598e+00$	$3.679e-01$	4.343	$1.88e-05$	$***$
Unemployment	$1.808e+02$	$1.059e+02$	1.708	0.088584	
FDI	$-1.249e+00$	$4.778e-01$	-2.615	0.009341	$**$
IRate	$-3.163e+01$	$2.571e+01$	-1.230	0.219449	
Covariates	**Estimate**	**Std.Error**	**t-value**	**P-value**	**Signif codes**
Angola	$-1.476e+04$	$4.546e+03$	-3.246	0.001292	$**$
Botswana	$-1.501e+04$	$4.641e+03$	-3.235	0.001340	$**$
Comoros	$-4.746e+03$	$2.902e+03$	-1.635	0.102938	
DRC	$-3.312e+04$	$8.990e+03$	-3.683	0.000269	$***$
Eswatini	$-1.043e+04$	$5.711e+03$	-1.827	0.068591	
Lesotho	$-1.063e+04$	$4.663e+03$	-2.280	0.023227	$*$
Madagascar	$-1.474e+04$	$4.578e+03$	-3.219	0.001416	$**$
Malawi	$-9.334e+03$	$3.354e+03$	-2.783	0.005704	$**$
Mauritius	$-9.439e+03$	$3.404e+03$	-2.773	0.005868	$**$
Mozambique	$-1.872e+04$	$4.631e+03$	-4.042	$6.63e-05$	$***$
Namibia	$-1.236e+04$	$5.089e+03$	-2.429	0.015681	$*$
Seychelles	$-3.491e+03$	$2.693e+03$	-1.296	0.195788	
South Africa	$4.011e+04$	$1.087e+04$	3.691	0.000262	$***$
Tanzania	$-2.186e+04$	$8.198e+03$	-2.667	0.008041	$**$
Zambia	$-1.103e+04$	$3.845e+03$	-2.869	0.004388	$**$
Zimbabwe	$-1.048e+04$	$3.560e+03$	-2.944	0.003471	$**$
Residual standard error:	10970 on	326 DF			
Multiple R-squared:	0.9858				
Adj. R-Squared:	0.9847				
F-statistic:	871 on 26	and 326	DF		
p-value	$< 2.2e-16$				

Random Effects Model (REM) for SADC Data					
Intercept	$-1.286e+04$	$2.170e+03$	-5.9254	$3.115e-09$	$***$
Inflaton	$2.835e-04$	$4.041e-04$	-0.7015	0.4829978	
Imports	$2.513e+00$	$1.536e-01$	16.3634	$< 2.2e-16$	$***$
Exports	$3.615e-01$	$1.096e-01$	3.2979	0.000974	$***$

EDebt	$4.502e-01$	$7.628e-02$	5.9020	$3.591e-09$	$***$
ERate	$2.843e-04$	$4.041e-04$	0.7035	0.4817150	
IReserves	$3.824e-01$	$2.624e-01$	1.4572	0.1450722	
LForce	$9.534e-01$	$1.556e-01$	6.1265	$8.983e-10$	$***$
Unemployment	$3.000e+02$	$6.582e+01$	4.5586	$5.149e-06$	$***$
FDI	$-1.210e+00$	$5.073e-01$	-2.3849	0.0170831	$*$
IRate	$-3.137e+01$	$2.730e+01$	-1.1491	0.2505199	
Total Sum of Squares:	$1.0548e+12$				
Residual Sum of Squares:	$4.8259e+10$				
R-Squared:	0.95425				
Adj. R-Squared:	0.95291				
Chi-square value:	7112.23 on	10 DF			
p-value:	$< 2.22e-16$				

Significance. codes :	0 : ***	0.001: **	0.01: *	0.05: .	0.1:

Source: Author's own results.

Test	p-value	Tested	Selection
F-test	6.606e − 15	OLS/Fixed	Fixed
Chow	2.2e − 15	OLS/(Random or Fixed)	Random or Fixed
Breusch-Pagan	< 2.2e − 16	OLS/ Random	Random
Hausman	3.58e − 14	Fixed/Random	Fixed

Table 4.23: **Specification tests**
Source: Author's own results.

From Tables; 4.22 and 4.23 it is evident that;

$$\text{GDP}_{it} = \beta_1 \text{LForce}_{it} + \beta_2 \text{Imports}_{it} + \beta_3 \text{Ireserves}_{it} \beta_4 \text{EDebt}_{it} + \beta_5 \text{FDI}_{it} + \beta_6 \text{Exports}_{it} + \alpha_i + \sum_{i=j=1}^{16} \alpha_i C_{ij} + U_{it} \tag{4.3}$$

Where,
$j = 1, 2, 3, \cdots, 16$ denotes specific nation dummy variable,
$i =$ represent country effects of regressors,
$t =$ is the time effects of regressors.
with an adjusted R^2-value of 98% is the best model to explain the SADC data. Imports, exports, external debt, international reserves, labour force and unemployment have significant positive impacts on economic development for the SADC community. Foreign direct investment impacts negatively on the growth. Inflation, Interest rate and exchange rate have no significant

correlation with the economic progression. To account for the uniqueness of each cross-sectional unit /nation, we vary the intercepts by using dummy variables of fixed effects. Usual OLS method is applied to all entities. A small p-value = $6.606e^{-15}$ implies that the pooled ordinary least squares model is inadequate, which favours of the fixed effects as the appropriate model. Uniqueness of each cross-sectional unit / nation is accounted if we let the intercept vary for each nation. It is also assumed that the slope coefficients are still constant across cross-section (Gujarati and Porter, 2003). From Table 4, it is evident that the estimated coefficients dummy for South Africa impacts positively on economic progression, however for all other SADC nations have negative relationships with economic growth except for Comoros, and Seychelles whose estimated dummy coefficients have no noteworthy relations with GDP as a measure of economic advancement in the SADC region. The differences in the intercepts of the nations might be emanating from the unique government policies about trade of exports and imports of goods, exchange rate, prices of goods in other nations, GDP comparative to major economies or/and other economic variables.

4.15 Machine Learning Procedures

4.15.1 Panel Linear Model

Table 4.24: Panel Linear Model

Panel Liner Model

Call:

$$\text{plm(formula} = \text{GDP} \sim \text{Inflation} + \text{Imports} + \text{Exports} + \text{EDebt} + \text{ERate} + \text{IReserves} + \text{LForce} + \text{Unemployment} + \text{FDI} + \text{IRate}, \text{data} = \text{panel\_data}, \text{model} = \text{"within"})$$

Coefficients	**Estimate**	**Std.Error**	**t-value**	**P-value**	**Signif codes**
Inflation	-1.76E-04	3.75E-04	-0.47	0.638671	
Imports	1.94E+00	1.64E-01	11.7992	$<$ 2.2e-16	***
Exports	2.94E-01	1.10E-01	2.672	0.007918	**
EDebt	2.47E-01	7.72E-02	3.2052	0.001483	**
ERate	1.66E-04	3.76E-04	0.4403	0.659993	

IReserves	1.13E+00	2.76E-01	4.0948	5.34E-05	***
LForce	1.60E+00	3.68E-01	4.3408	1.90E-05	***
Unemployment	1.80E+02	1.06E+02	1.7033	0.089465	.
FDI	-1.25E+00	4.78E-01	-2.6151	0.009334	**
IRate	-3.16E+01	2.57E+01	-1.2304	0.219442	
Model Summary					
R-Squared:	0.84775				
Adj. R-Squared:	0.83285				
F-statistic: 177.385 on 10 and 326 DF					
p-value: 2.12E-16					

The R^2 value for the panel linear model is 0.84475, which indicates that the independent variables in the model explain approximately 84.5% of the variation in the dependent variable, GDP. This is a relatively high value, suggesting that the model is a good fit for the data.

The p-value for the F-statistic is very small (less than the minimum representable value in R), which means that at least one of the independent variables in the model is statistically significant in explaining the variation in GDP. The F-statistic is a measure of the overall significance of the model, and a small p-value indicates that the model as a whole is statistically significant.

studentized Breusch-Pagan test

```
data:  fixed_effects_model
BP = 110.95, df = 10, p-value < 2.2e-16
```

The Breusch-Pagan test which is a statistical test that checks for heteroskedasticity was applied to the residuals of a fixed effects model. The output of the test gave a Breusch-Pagan test statistic (BP) of 110.95, on 10 degrees of freedom, and a p-value is less than 2.2e-16 (essentially zero). Iplyinging that there is strong evidence that the residuals of the fixed effects model are heteroskedastic (i.e. the variance of the residuals is not constant across the range of the independent variables).

The presents of heteroskedasticity in the model, imply that the model may produce biased standard errors, which in turn affect the validity of hypothesis tests and confidence intervals. Therefore, it may be necessary to adjust the model or use different estimation methods to address the issue of heteroskedasticity.

Hausman Test(fixed_effects_model,random_effects_model)

```
data:  GDP ~ Inflation + Imports + Exports + EDebt + ERate + IReserves +  ...
        chisq = 79.513, df = 10, p-value = 6.255e-13
        alternative hypothesis: one model is inconsistent
```

The Hausman test was carried out to determine whether a fixed effects or random effects model is more appropriate for the panel data analysis. The test produced a chi-squared value of 79.513 with 10 degrees of freedom, and the p-value for the test is 6.255e-13. This suggests that the null hypothesis of the test, which assumed that the random effects model is consistent and efficient, should be rejected. The alternative hypothesis is that one of the models is inconsistent, and the result of the test implies that the fixed effects model is consistent and should be preferred over the random effects model.

4.16 Gradient Boost

In this section, we utilize a popular machine-learning technique called gradient boosting to learn the relationship between economic variables and GDP. This method is effective for solving both regression and classification problems. Initially, we split the data into 70% for training and 30% for testing purposes. We used the 'xgboost()' function from the 'XGBoost' package in R to train the model for 100 iterations. Afterward, we made predictions on the testing data set and calculated the R-squared value, which was found to be 97.59%. Thus the gradient boost approach was able to explain about 97.6% of the variation in GDP for the SADC data. This indicates that the trained model performed well on the test data set. the output from the modelling process is as below:

```
    xg b_ p re di c ti o ns <- predict ( xgb_model , test_matrix )
xgb_rmse <- sqrt ( mean (( x gb _ pr ed i ct io n s - test_data$GDP )^2))
print ( paste (" RMSE :" , xgb_rmse ))
actual <- test_data$GDP
xgb_rsquared <- 1 - sum (( actual - x gb _ pr e di ct i on s )^2) / sum (( actual
- mean ( actual ))^2)
xgb_accuracy <- xgb_rsquared * 100
```

```
print ( paste (" Accuracy (R - squared ):" , xgb_accuracy , "%"))
Accuracy (R - squared ): 97.59%
```

4.17 Support Vector Machine

In this section, we will be using another popular machine learning tool called support vector regression, which is useful for modeling complex nonlinear relationships between input and output variables. Support vector regression is a variation of support vector machines that specifically handles regression problems.

To begin, the data was split into training and testing data sets. Then, a support vector regression model was trained. Using this trained model, predictions were made on the test dataset, and the R-squared value was calculated. The support vector model achieved an R-squared value of 93.86%. Hence the support vector machine approach was able to explain about 93.9% of the variation in GDP for the SADC data. However, the accuracy is slightly lower than that of the gradient boost method which had an accuracy of 97.59%. The output from the Suport Vector Machine approacch is as follows:

```
    numeric_vars <- c (" GDP " , " Inflation " , " Imports " , " Exports " ,
" EDebt " , " ERate " , " IReserves " , " LForce " ,
" Unemployment " , " FDI " , " IRate ")
t r a i n _ d a t a _ n u m e r i c <- train_data [ , numeric_vars ]
svr_model <- svm ( GDP ~ . , data = train_data_numeric , kernel = " radial ")
sv r_ p re di c ti o ns <- predict ( svr_model , newdata = test_data [ , numeric_vars ])
svr_rsquared <- 1 - sum (( actual - s vr _ pr e di ct i on s )^2) / sum (( actual -
mean ( actual ))^2)
svr_accuracy <- svr_rsquared * 100
print ( paste (" Accuracy (R - squared ):" , svr_accuracy , "%"))
Accuracy (R - squared ): 93.86%
```

4.18 Random Forest

In this section a random forest model was trained on the panel, the data A was partitioned into 70% training and 30%testing sets, then the “randomForest()” function from the “randomForest” package was used to train 100 trees in the forest. Predictions were made on the testing data set and the RMSE and R^2 were found to be 20522.5737 and 0.9742 respectively. This implied that 97.4% of the variation in GDP is explained by the feature variables. The output from the rando forest modelling is as below;

```
    # summary ( plm_lr )
f i x e d _ e f f e c t s _ m o d e l <- plm ( GDP ~ Inflation + Imports + Exports + EDe
Unemployment + FDI + IRate , data = panel_data , model = " within ")
r a n d o m _ e f f e c t s _ m o d e l <- plm ( GDP ~ Inflation + Imports + Exports + E
+ Unemployment + FDI + IRate , data = panel_data , model = " random ")
summary ( f i x e d _ e f f e c t s _ m o d e l )
bptest ( f i x e d _ e f f e c t s _ m o d e l )
phtest ( fixed_effects_model , r a n d o m _ e f f e c t s _ m o d e l )
set . seed (123)
train_indices <- c r e a t e D a t a P a r t i t i o n ( panel_data$GDP , p = 0.7 , list
train_data <- panel_data [ train_indices , ]
test_data <- panel_data [ - train_indices , ]
rf_model <- randomForest ( GDP ~ . , data = train_data , ntree = 1000)
rf_pred ictio ns <- predict ( rf_model , newdata = test_data )
rf_rmse <- sqrt ( mean (( rf_p redict ions - test_data$GDP )^2))
rf_rmse
20522.5737
rf_r2 <- cor ( rf_predictions , test_data$GDP )^
rf_r2
0.9742
importance <- r f _ m o d e l $ i m p o r t a n c e
importance
varImpPlot ( rf_model )
```

Table 4.25 shows the variable importance, which is obtained from the importance attribute in R. This attribute displays the contribution of each variable to the accuracy of the model. The importance measures are based on the mean decrease in the accuracy of the model when each variable is randomly permuted. A variable with a higher importance measure indicates that it has a greater impact on the accuracy of the model. In this case, the “IncNodePurity” measure is used, which is based on the Gini impurity index.

The variable “Imports” has the highest importance measure at $4.54e+11$, followed by “Exports and “EDebt”. This suggests that these variables have the greatest impact on the accuracy of the model in predicting “GDP”. Conversely, “Irate” has the lowest importance measure at

Table 4.25: Variable Importance Table

Variable	**IncNodePurity**
Inflation	4167056689
Imports	4.54048E+11
Exports	2.76574E+11
EDebt	1.8785E+11
ERate	7352297185
IReserves	1.72579E+11
LForce	26513617707
Unemployment	15047054533
FDI	36869404823
IRate	2518359503

$2.52e + 09$, indicating that it has the least impact on the accuracy of the model. These results are consistent with the conclusions from the traditional models above.

Figure 4.14 below provides a visual representation of the results in Table 4.25.

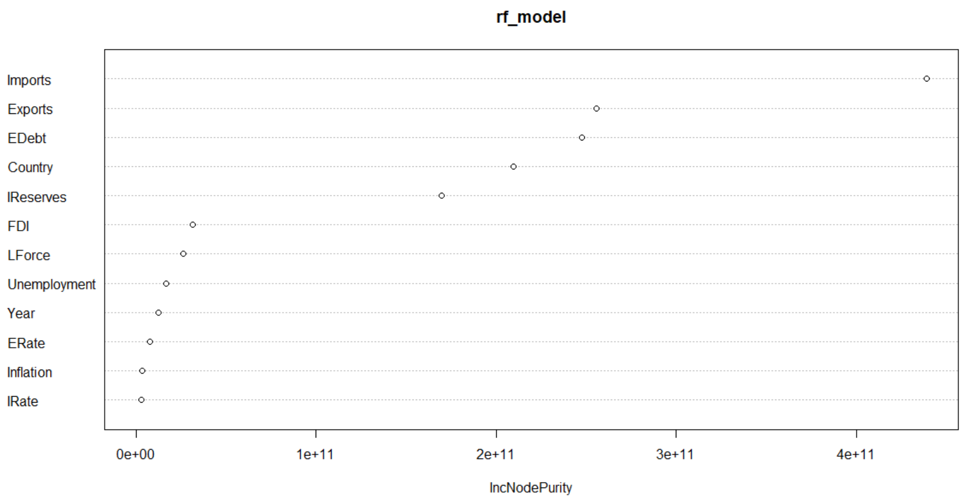


Figure 4.14: Variable Importance

4.19 Deep Learning

In modeling the GDP for Zimbabwe, a dense neural network with a total of 76 neurons was trained to give a multivariate regression model for GDP. Linear functions were used as activation functions in both the hidden and output layers. The optimizer *"NAdam"* and loss metric *"Mean Squared Error"* were used. The model was trained for 2000 epochs, with the main objective of reducing the mean squared error between the observed and trained values to zero.

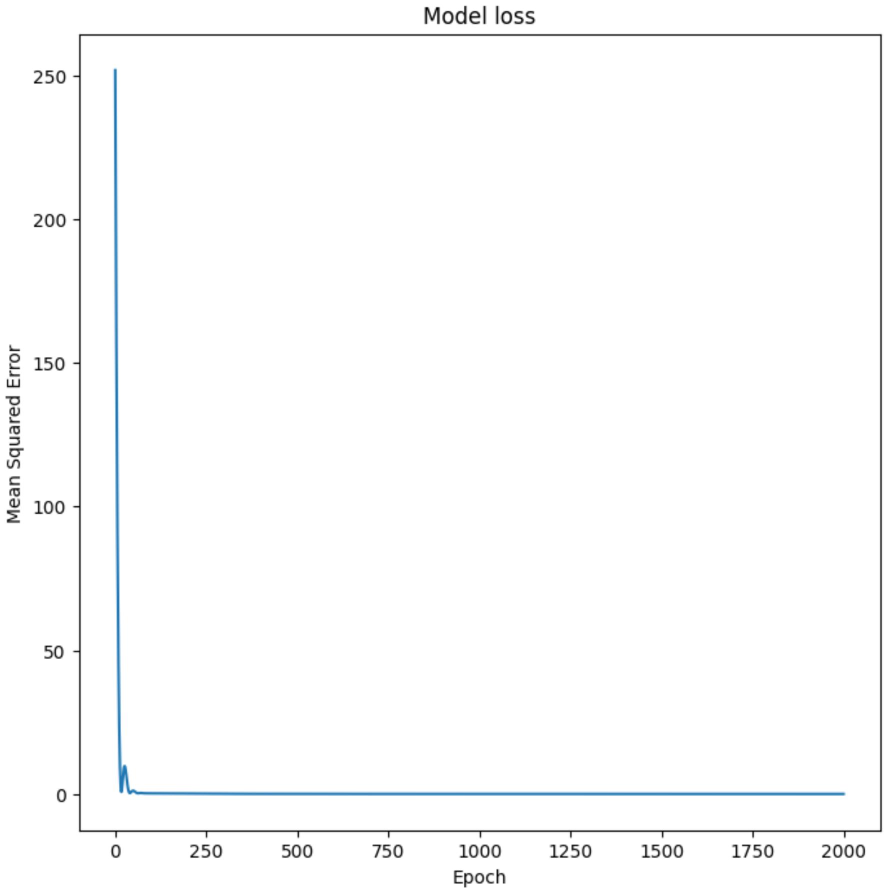


Figure 4.15: Mean Squared error by epoch

Figure 4.15 shows that during the first training, the mean squared error ranged above 200. Between the 2^{nd} and 50^{th} training, the mean squared error drastically dropped to 10. From the 50^{th} to 100^{th}training, the mean squared error rose and then dropped again to 10, gradually decreasing toward zero. After 2000 trainings, the trained model had a mean squared error of 0.003, implying a prediction power of 99.70%. The trained model was used to predict the test dataset (unseen data), and the mean squared error between the observed and predicted values was computed, and a mean squared error of 0.014 was observed. However, this higher mean squared error suggests that the model has over-learned the characteristics of the training data set (overfitting), indicating the need to adjust the number of epochs the model is trained for to 500. The final trained model was found to be:

$$
\begin{aligned}
\ln(\text{GDP}) =& 0.0173 + 0.015\ln(\text{inflation}) + 0.1903\ln(\text{Exports}) + 0.627\ln(\text{ EDebt}) - 0.0817\ln(\text{ERate}) \\
& - 0.0585\ln(\text{LForce}) - 0.3919\ln(\text{ Unemployment}) - 0.3578\ln(\text{FDI}) + 0.1037\ln(\text{Irate})
\end{aligned} \tag{4.4}
$$

The model that was learned, represented by equation (4.4), was subsequently utilized to predict Zimbabwe's GDP between the years 2022 and 2026. The resulting predictions are presented below:

Table 4.26: Five-year forecasts for Zimbabwe's GDP

Year	Predicted GDP
2022	26626.951
2023	30476.693
2024	32215.355
2025	33794.71
2026	33597.113

Figure 4.16 provides a visual representation of what the trained model has learned about Zimbabwe's GDP. The blue curve represents the learned model, whereas the yellow curve represents the actual GDP. The dotted red line represents the next five-year forecast for Zimbabwe's GDP. As shown in Figure 4.16,the fitted values of GDP compares will with the observed ones.As for the forecasts for Zimbabwe's GDP; there will be a sharp increase in the GDP between the years 2022 and 2023, followed by a gradual increase from 2024 to 2026.

4.20 Chapter Summary

In this chapter, we explored the analysis of panel data using both modern and traditional techniques. Surprisingly, both methods demonstrated a high degree of accuracy in modelling the SADC data, with the traditional model (LSDV) achieving an R-squared value of 98.47% and the machine-learning model achieving almost the same level of accuracy (97.59% from Gradient Boosting). For the Zimbabwean data, the model with log transforms was the best

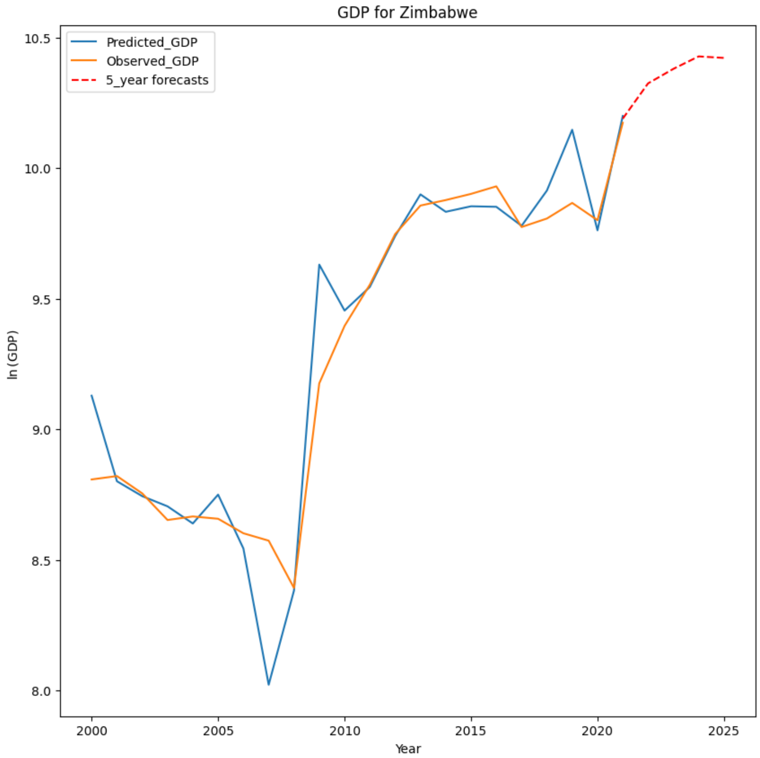


Figure 4.16: Zimbabwe GDP

with an adjusted R-squared value of 99.81% while the deep learning approach model had an R-squared value of 98.51% Additionally, we presented a five-year forecast for Zimbabwe's GDP using the trained machine learning model. These findings suggest that both modern and traditional techniques can be effective in analyzing panel data, and the choice of method may depend on factors such as model complexity and ease of implementation.

Chapter 5

Conclusions and Recommendations

5.1 Introduction

In the preceding chapter the statistical tests were reported and the results discussed. This chapter delivers the conclusions to the research, suggestions and policy recommendations for forthcoming researches.

5.2 Summary of Research Objectives

Chapter one explained the emphasis of this research as to empirically evaluate the impact of a select set of macroeconomic variables on economic development in the SADC region and Zimbabwe. This was accomplished in Chapter Four, when Panel data, OLS, ML and Multivariate approaches were used. The results revealed that there exists an association between economic progression and a select set of macroeconomic variables both in the SADC region and Zimbabwe. There were four objectives; firstly, to identify factors that are important for economic development for SADC nations and Zimbabwe. To deliver an overview of the Zimbabwean and SADC economies and in what way the selected variables compares with other communities and

countries, developing and developed. Chapter Two revealed this. Considering economic conditions in the most nations in the SADC region and Zimbabwe, the industry is still developing, and the advancement is remarkable
Secondly, to fit OLS, FE,ML and RE models to SADC community economic data and a multivariate model Zimbabwean economic data and give an outline of trends in economic progression in Zimbabwe and the SADC region, Chapter Two also clarified this. An empirical enquiry was done to assess the hypothesised connections among the chosen variables for both the SADC region and Zimbabwe. Chapter Four tested the variables of interest for the SADC data; Chow Test for Poolability, F-test, Hausman Test for Comparing Fixed and Random Effects and Breusch-Pagan LM Test for Random Effects, were employed. For the Zimbabwean data, R-Squared value, ANOVA, AIC and residual analysis approaches were put to use. The causal relationships between selected economic variables and economic development were established for both the SADC economy and Zimbabwe. The fourth objective was to forecast GDP for Zimbabwe for the next five years this was accomplished in chapter four and it was realised that the GDP for Zimbabwe will generally increase for the period forecasted. Finally, this chapter proposes recommendations to the interested parties in the Zimbabwean and SADC region economies concerning the empirical findings from the research. The end of this chapter gives explanations to the recommendations.

5.3 Summary of Literature Review

The literature review in chapter two presented how diverse variables interrelate with economic development. The outcome of interaction between economic entities and changes in GDP .This was investigated both theoretically as in chapter two and empirically in chapter four. Chapter two examined theoretical evidence and chapter four gave the empirical results. The research also indicated that a number of economic variables were key ingredients to economic growth. The function of employment levels, FDI, inflation rate, exchange rate, labour force, imports, exports, external debt, international reserves and interest rate on economic growth were examined
Literature exposed that a chosen set of macroeconomic entities have positive impacts on economic growth and have resilient association with economic progression in other countries. The methodical outline of the research showed that the economic crisis which occur in most countries are a result of several factors, specifically, governance complications, weak macro-economic guidelines, destabilized economic performance as well as a limited investor confidence amid others. The hyperinflation that was witinessed in some nations like Zimbabwe and Zambia resulted in macroeconomic insecurity which made it very difficult for companies to survive. A number

of policies are put in place to stabilise the economies like the Short-Term Emergency Recovery Programme which was employed in 2009 for Zimbabwe to stabilize the economy was fruitful as it reduced inflation and the economic growth was positive. Concerns in matters of the rule of law, sustainable policy, government expenditure, corruption, sound policy, marks the path to economic in most nations. This research is of the view that, though the research was not primarily motivated by these misgivings, they make up a substantial part in the generally discussion, these are challenging to measure quantitatively but qualitatively, they poke out at every turn in the progression path.

5.4 Recommendations and Policy Implications

For nearly four decades, development economists have made attempts to explain the importance of international trade promotion (exports and imports) together with capital and labour on economic growth. Most SADC countries consequently, have implemented trade policies aimed at stimulating economic development, with the definitive goal of improving the livelihoods of the citizenry, and aggravate poverty. Empirical researches conducted in numerous different countries however report inconsistent findings. These inconsistencies have thus resulted in questions about the soundness, robustness and universality of the export and import led growth hypotheses. A number of studies have been undertaken to examine the influence of different determinants on economic progression of SADC nations. However, this study specifically examined the impact of imports, exports, external debt, exchange rate, international reserves, labour force, foreign direct investment and interest rate on economic development in 16 SADC nations over the period from 2000 to 2021. The empirical findings show that, exports, external debt, international reserves, employment level and labour force have noteworthy positive impacts on economic progression for the SADC community. FDI impacts negatively on the development. Inflation, Interest rate and exchange rate have no significant correlation with the economic advancement.

Furthermore the research revealed that, estimated dummy coefficients for South Africa impacts positively on GDP while for all other SADC nations they have negative relationships with economic growth except for; Comoros, Seychelles which have no significant relations with GDP as a measure of economic development in the SADC region. Thus, nations in SADC region should focus on: imports, exports, external debt, international reserves, employment levels and labour force in the long run so as to promote economic growth. This will however call for overcoming the regions inconsistent power supply Maripe et al. (2017) and integrating the patchy intraregional trade regulations Chea (2012). Nevertheless, it is crucial for the

SADC community to improve the quality of imports, as well as centre on strategic commodities especially highly developed technologies and key equipment that are unavailable within the region but are needed for primary national economic growth by improving domestic production for local use and exports Adjasi et al. (2012). Having examined the impact of imports, exports, external debt, exchange rate, international reserves, labour force, interest rate and FDI on economic advancement in SADC region in detail, this research recognizes the need to scrutinize the direction of relationship between the aforesaid variables in an attempt to enhance evidence-based policy making and policy implementation as regards to trade-driven regional economic growth agenda. Also, the structure in this study captures some important growth determinants that may also have a strong connection with economic growth. Some of these variables are policy stability, education level (human capital) and other macro-economic variables that were not incorporated in the estimation process chiefly as a result of lack of available data for the period of this research. It may nevertheless be intuitive to include an extended set of socio-economic indicators in the analysis.

As for Zimbabwe; for it to modify the trajectory of economic development, supporting exports, labour force, foreign direct investment and re-contextualizing the human capital improvement drive could be a prudent approach to achieve this. This research has availed robust signal signifying that exports, foreign direct investment and labour force, have positive effects on the economic development whereas inflation, external debt, exchange rate, unemployment levels and interest rate have negative effects on economic growth in Zimbabwe, thus Zimbabwe should minimise all those with negative effects on economic progression.

5.5 Limitations

The primary limitation of this investigation pertained to the quality of the utilized dataset. Given that the data was secondary in nature, its accuracy could not be definitively guaranteed. This issue was further exacerbated by the complexities encountered during data collection processes in developing nations. Moreover, the timeframe of the study witnessed numerous structural disruptions, encompassing varying exchange rate regimes, economic policies, and military administrations, among others. As a result, the interpretations and analyses drawn from this research were mindful of these constraints. Additionally, it was crucial to consider the financial resources and the time constraints, as they also significantly impeded the research process.

5.6 Further Research

This research is a basis for further studies that identifies the movers of economic development in the SADC region and Zimbabwe; by means of additional economic pointers in the study. Furthermore, possible researches would pinpoint the thresholds at which key variables such as government expenditure, corruption start to prevent growth specifically through hyperinflationary backgrounds. Furthermore, a research on the effect of increasing interregional trade for SADC and international trade for Zimbabwe on economic development in Zimbabwe would be of importance.

5.7 Conclusions

The aim of the research was to establish models that links economic development as measured by GDP to the determinants of economic growth for the Zimbabwean and SADC economies. While the objectives of the the study were:

1. To identify determinants of economic growth for the Zimbabwean and SADC nations.
2. To fit OLS, FE,ML and RE models to SADC community economic data and a Multivariate Model Zimbabwean economic data?
3. To Compare performance of OLS,RE, ML and FE models for the SADC community economic data and Identify the most appropriate model to describe the Zimbabwean economic data.
4. To forecast GDP for Zimbabwe for the next five years.

All these were achieved in chapter four. The results revealed that; the traditional multivariate approach was the best method to handle the Zimbabwean data and the LSDV procedure was the best to handle the SADC data. For the SADC region; exports, external debt, international reserves, employment level and labour force have a significant positive impacts on economic development. On the other hand, FDI impacts negatively on the development. Inflation, interest rate and exchange rate have no substantial association with the economic advancement in the SADC community. Whereas, for Zimbabwe exports, labour force and foreign direct investment positively influenced economic growth on the other hand; inflation, external debt, exchange rate, unemployment levels and interest rate have negative effects on economic growth in Zimbabwe. However, FDI is not a magic bullet or a remedy for limited economic development in any economy, need is there for communal action by industry, citizens, civic societies and politics. Forecasts of GDP for Zimbabwe indicated an upward trend for the Zimbabwean economy. Additionally, need exists for stability politically. The findings also offer explanation for the pursuit of the sustainable growth goals by promoting exports, labour force, foreign direct investment which impacts positively on economic advancement.

Appendix

Code 1 Models Without Data Transforms

```
> # Loading the readxl library a package used to import data from excel into R.

> library(readxl)

> Zim_D<- read_excel("Zimbabwe Data.xlsx")

> Zim_Data <- read_excel("Zimbabwe Data.xlsx")

> # Creating a new data frame for Zimbabwean data.

> Zim_Data1=data.frame(Zim_D$GDP, Zim_D$Inflation, Zim_D$Imports,Zim_D$Exports,
                       Zim_D$ExternalDebt,Zim_D$ExchangeRate,Zim_D$LabourForce,
                       Zim_D$InternatonalReserves,Zim_D$Unemployment,Zim_D$ForeignDirectInvestment,
                       Zim_D$Interestrate)

> # Computing the correlation matrix for the economic variables for Zimbabwe.

> cor(Zim_Data1)

                              Zim_D.GDP  Zim_D.Inflation  Zim_D.Imports  Zim_D.Exports
Zim_D.GDP                     1.0000000    -0.28649605     0.75688651      0.8413039
Zim_D.Inflation              -0.2864961     1.00000000    -0.17618394     -0.2354165
Zim_D.Imports                 0.7568865    -0.17618394     1.00000000      0.7050608
Zim_D.Exports                 0.8413039    -0.23541650     0.70506083      1.0000000
Zim_D.ExternalDebt            0.6937549    -0.14003090     0.55268318      0.8502210
Zim_D.ExchangeRate           -0.3193284     0.99147632    -0.22969299     -0.2624640
Zim_D.LabourForce             0.8681215    -0.10382867     0.65533113      0.8601865
Zim_D.InternatonalReserves   -0.1827320     0.12681293    -0.06556251     -0.1498975
Zim_D.Unemployment           -0.1136524    -0.14129417    -0.24122621     -0.2782459
Zim_D.ForeignDirectInvestment 0.8294622    -0.17839076     0.67005302      0.6526401
```

```
Zim_D.Interestrate             -0.2596571     0.04239432    0.12348640    -0.2154710

                              Zim_D.ExternalDebt   Zim_D.ExchangeRate  Zim_D.LabourForce
Zim_D.GDP                              0.6937549         -0.31932837         0.8681215
Zim_D.Inflation                       -0.1400309          0.99147632        -0.1038287
Zim_D.Imports                          0.5526832         -0.22969299         0.6553311
Zim_D.Exports                          0.8502210         -0.26246402         0.8601865
Zim_D.ExternalDebt                     1.0000000         -0.15910823         0.9021359
Zim_D.ExchangeRate                    -0.1591082          1.00000000        -0.1206661
Zim_D.LabourForce                      0.9021359         -0.12066610         1.0000000
Zim_D.InternatonalReserves            -0.2945557          0.11026082        -0.3570113
Zim_D.Unemployment                    -0.5388885         -0.16097674        -0.4836686
Zim_D.ForeignDirectInvestment          0.4177142         -0.19880102         0.7064768
Zim_D.Interestrate                    -0.4474808          0.05377278        -0.3613312

                                   Zim_D.InternatonalReserves   Zim_D.Unemployment
Zim_D.GDP                                     -0.18273202        -0.11365243
Zim_D.Inflation                                0.12681293        -0.14129417
Zim_D.Imports                                 -0.06556251        -0.24122621
Zim_D.Exports                                 -0.14989755        -0.27824588
Zim_D.ExternalDebt                            -0.29455566        -0.53888851
Zim_D.ExchangeRate                             0.11026082        -0.16097674
Zim_D.LabourForce                             -0.35701132        -0.48366858
Zim_D.InternatonalReserves                     1.00000000         0.58691523
Zim_D.Unemployment                             0.58691523         1.00000000
Zim_D.ForeignDirectInvestment                 -0.17229383        -0.02259652
Zim_D.Interestrate                             0.14089808         0.11019462

                                 Zim_D.ForeignDirectInvestment  Zim_D.Interestrate
Zim_D.GDP                                       0.82946222        -0.25965706
Zim_D.Inflation                                -0.17839076         0.04239432
Zim_D.Imports                                   0.67005302         0.12348640
Zim_D.Exports                                   0.65264006        -0.21547104
Zim_D.ExternalDebt                              0.41771423        -0.44748079
Zim_D.ExchangeRate                             -0.19880102         0.05377278
Zim_D.LabourForce                               0.70647678        -0.36133122
Zim_D.InternatonalReserves                     -0.17229383         0.14089808
Zim_D.Unemployment                             -0.02259652         0.11019462
Zim_D.ForeignDirectInvestment                   1.00000000        -0.08036169
Zim_D.Interestrate                             -0.08036169         1.00000000

####################################################################################################


> # The Zim_1 Initial Model

> Zim_1=glm(GDP~Inflation+Imports+Exports+ExternalDebt+ExchangeRate+InternatonalReserves+
          LabourForce+Unemployment+ForeignDirectInvestment+Interestrate,
          data=Zim_Data,family = "gaussian")

# Zim_1 model summary

> summary(Zim_1)
Call:
```

```
glm(formula = GDP ~ Inflation + Imports + Exports + ExternalDebt + ExchangeRate + InternatonalReserves+
 LabourForce + Unemployment + ForeignDirectInvestment + Interestrate,
  family = "gaussian", data = Zim_Data)

Deviance Residuals:
  Min       1Q    Median    3Q      Max
-2848.1   -414.7    0.0    995.3   2050.9

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)             -9.114e+04  2.200e+04  -4.142  0.00201 **
Inflation                -2.876e-05  7.597e-05  -0.379  0.71297
Imports                  8.020e-01  3.271e-01   2.452  0.03415 *
Exports                  1.203e+00  7.743e-01   1.554  0.15115
ExternalDebt            -1.772e+00  7.789e-01  -2.275  0.04616 *
ExchangeRate             1.545e-05  7.839e-05   0.197  0.84771
InternatonalReserves    -5.221e-01  8.527e-01  -0.612  0.55402
LabourForce              1.511e+01  3.918e+00   3.858  0.00317 **
Unemployment             1.736e+03  7.061e+02   2.459  0.03373 *
ForeignDirectInvestment -5.998e+00  5.817e+00  -1.031  0.32682
Interestrate            -1.047e+01  6.778e+00  -1.545  0.15343
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for gaussian family taken to be 3134194)

Null deviance: 786200872  on 20  degrees of freedom
Residual deviance:  31341935  on 10  degrees of freedom
AIC: 382.13

> AIC(Zim_1)
[1] 382.1303
> BIC(Zim_1)
[1] 394.6645

####################################################################################################

> # Adjusted model without the Exchange rate variable

> Zim_2=glm(GDP~Inflation+Imports+Exports+ExternalDebt+InternatonalReserves+LabourForce+Unemployment+
             ForeignDirectInvestment+Interestrate,data=Zim_Data,family = "gaussian")

> summary(Zim_2)

Call:
glm(formula = GDP ~ Inflation + Imports + Exports + ExternalDebt +
InternatonalReserves + LabourForce + Unemployment + ForeignDirectInvestment +
Interestrate, family = "gaussian", data = Zim_Data)

Deviance Residuals:
    Min       1Q      Median     3Q        Max
-2913.13   -412.14    -56.75   1009.28   2075.94

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)             -9.102e+04  2.101e+04  -4.332  0.00119 **
Inflation                -1.389e-05  8.639e-06  -1.608  0.13619
Imports                  7.660e-01  2.592e-01   2.956  0.01308 *
Exports                  1.203e+00  7.397e-01   1.627  0.13207
ExternalDebt            -1.779e+00  7.432e-01  -2.394  0.03560 *
InternatonalReserves    -4.890e-01  7.987e-01  -0.612  0.55278
LabourForce              1.519e+01  3.725e+00   4.077  0.00183 **
Unemployment             1.690e+03  6.354e+02   2.659  0.02222 *
ForeignDirectInvestment -5.847e+00  5.509e+00  -1.061  0.31129
Interestrate            -1.011e+01  6.229e+00  -1.622  0.13301
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for gaussian family taken to be 2860334)

Null deviance: 786200872  on 20  degrees of freedom
Residual deviance:  31463677  on 11  degrees of freedom
AIC: 380.21

Number of Fisher Scoring iterations: 2

> AIC(Zim_2)
[1] 380.2117
> BIC(Zim_2)
[1] 391.7014

####################################################################################################

> # Adjusted model  without Exchange rate and International Reserves

> Zim_3=glm(GDP~Inflation+Imports+Exports+ExternalDebt+LabourForce+Unemployment+ForeignDirectInvestment+
           Interestrate,data=Zim_Data,family = "gaussian")

> summary(Zim_3)

Call:
glm(formula = GDP ~ Inflation + Imports + Exports + ExternalDebt +
LabourForce + Unemployment + ForeignDirectInvestment + Interestrate,
family = "gaussian", data = Zim_Data)

Deviance Residuals:
  Min       1Q      Median   3Q      Max
-3024.9   -854.2    0.0    1158.9   2068.3

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)             -9.097e+04  2.046e+04  -4.447 0.000798 ***
Inflation                -1.569e-05  7.910e-06  -1.983 0.070691 .
Imports                  7.270e-01  2.446e-01   2.972 0.011648 *
Exports                  1.168e+00  7.179e-01   1.627 0.129763
ExternalDebt            -1.826e+00  7.198e-01  -2.537 0.026069 *
LabourForce              1.552e+01  3.589e+00   4.322 0.000992 ***
Unemployment             1.478e+03  5.187e+02   2.849 0.014662 *
ForeignDirectInvestment -5.485e+00  5.333e+00  -1.029 0.323950
Interestrate            -1.013e+01  6.065e+00  -1.671 0.120566
```

```
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for gaussian family taken to be 2711339)

Null deviance: 786200872  on 20  degrees of freedom
Residual deviance:  32536066  on 12  degrees of freedom
AIC: 378.92

Number of Fisher Scoring iterations: 2

> AIC(Zim_3)
[1] 378.9155
> BIC(Zim_3)
[1] 389.3607

####################################################################################################

> # Adjusted model without Exchange rate, International reserves and FDI

> Zim_4=glm(GDP~Inflation+Imports+Exports+ExternalDebt+LabourForce+Unemployment+Interestrate,
            data=Zim_Data,family = "gaussian")

> summary(Zim_4)

Call:
glm(formula = GDP ~ Inflation + Imports + Exports + ExternalDebt +
LabourForce + Unemployment + Interestrate, family = "gaussian",
data = Zim_Data)

Deviance Residuals:
  Min      1Q      Median    3Q        Max
-3953.8   -643.1    -34.3   1101.9   2233.9

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  -7.422e+04  1.242e+04  -5.977 4.62e-05 ***
Inflation     -1.434e-05  7.818e-06  -1.834   0.0896 .
Imports       6.371e-01  2.289e-01   2.783   0.0155 *
Exports       9.253e-01  6.796e-01   1.361   0.1965
ExternalDebt -1.273e+00  4.797e-01  -2.654   0.0198 *
LabourForce   1.247e+01  2.036e+00   6.126 3.62e-05 ***
Unemployment  1.338e+03  5.018e+02   2.667   0.0194 *
Interestrate -8.975e+00  5.972e+00  -1.503   0.1568
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for gaussian family taken to be 2723444)

Null deviance: 786200872  on 20  degrees of freedom
Residual deviance:  35404770  on 13  degrees of freedom
AIC: 378.69

Number of Fisher Scoring iterations: 2
```

```
> AIC(Zim_4)
[1] 378.6899
> BIC(Zim_4)
[1] 388.0906


####################################################################################################


> # Adjusted model without Exchange rate , International reserves, FDI and Exports

> Zim_5=glm(GDP~Inflation+Imports+ExternalDebt+LabourForce+Unemployment+Interestrate,
                data=Zim_Data,family = "gaussian")

> summary(Zim_5)

Call:
glm(formula = GDP ~ Inflation + Imports + ExternalDebt + LabourForce +
Unemployment + Interestrate, family = "gaussian", data = Zim_Data)

Deviance Residuals:
   Min       1Q     Median    3Q       Max
-3512.8   -965.8     58.4   1181.4   2093.8

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  -8.241e+04  1.119e+04  -7.365 3.54e-06 ***
Inflation     -1.577e-05  7.980e-06  -1.976  0.06820 .
Imports       7.143e-01  2.284e-01   3.127  0.00742 **
ExternalDebt -9.008e-01  4.059e-01  -2.220  0.04348 *
LabourForce   1.341e+01  1.973e+00   6.795 8.67e-06 ***
Unemployment  1.667e+03  4.529e+02   3.682  0.00247 **
Interestrate -6.449e+00  5.847e+00  -1.103  0.28868
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for gaussian family taken to be 2889511)

Null deviance: 786200872  on 20  degrees of freedom
Residual deviance:  40453159  on 14  degrees of freedom
AIC: 379.49

Number of Fisher Scoring iterations: 2

> AIC(Zim_5)
[1] 379.4892
> BIC(Zim_5)
[1] 387.8454
> #  The chosen model without data transforms is Zim_4 model which has the lowest AIC and BIC

Zim_4=glm(GDP~Inflation+Imports+Exports+ExternalDebt+LabourForce+Unemployment+Interestrate,
            data=Zim_Data,family = "gaussian")
```

```
> # Changing the plotting panel to a 2x2 grid

> par(mfrow=c(2,2))

> # Plotting Residual plots for the best model

> plot(Zim_4)
```

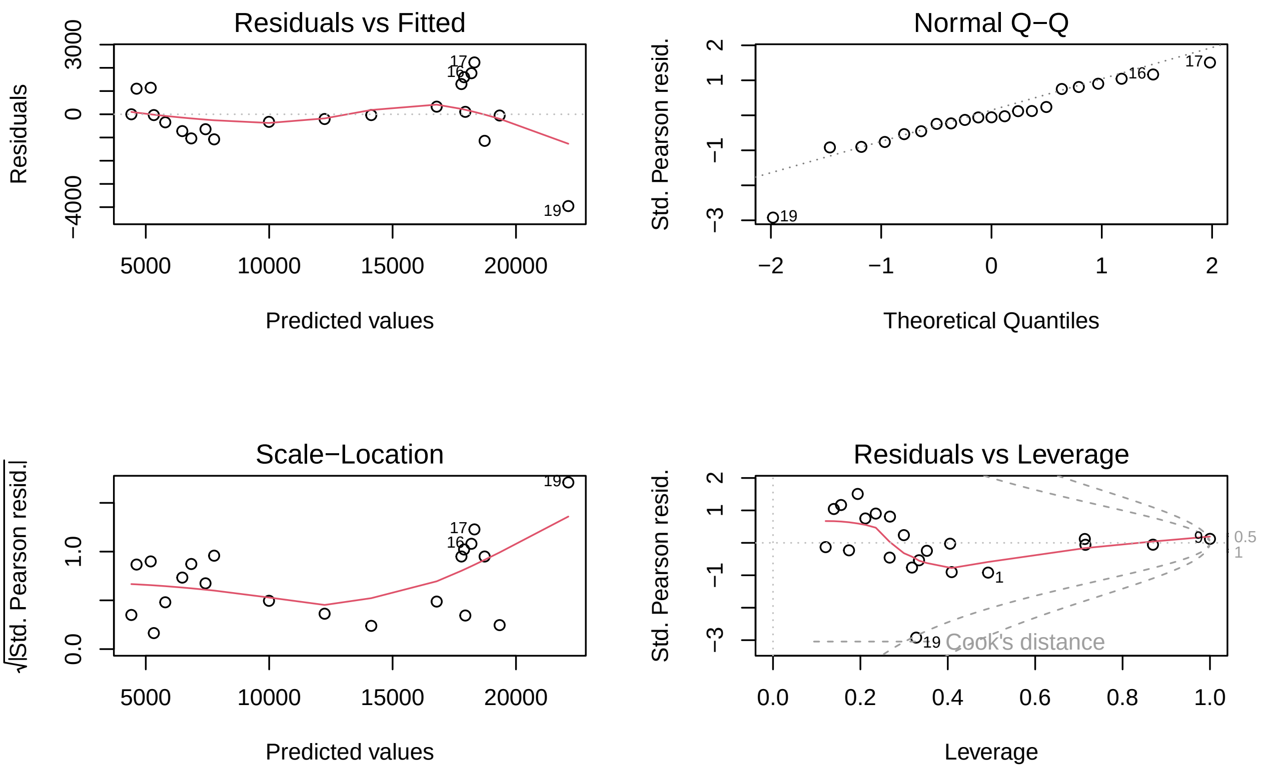


Figure 5.1: Zim_4 Model Residual Plots

```
Warning messages:
1: In sqrt(crit * p * (1 - hh)/hh) : NaNs produced
2: In sqrt(crit * p * (1 - hh)/hh) : NaNs produced

> # Returning the plotting panel to a 1x1 grid

> par(mfrow=c(1,1))

###############################################################################################
```

Code 2 Models With Data Transforms

```
> # The initial Model M_1

> M_1=lm(log(GDP)~log(Inflation)+log(Imports)+log(Exports)+log(ExternalDebt)+log(ExchangeRate)+
     log(InternatonalReserves)+log(LabourForce)+log(Unemployment)+log(ForeignDirectInvestment)+
     log(Interestrate),data=Zim_D)

Warning messages:
1: In log(Inflation) : NaNs produced
2: In log(Interestrate) : NaNs produced

# The Summary for M_1 Model

> summary(M_1)

Call:
lm(formula = log(GDP) ~ log(Inflation) + log(Imports) + log(Exports) + log(ExternalDebt) +
log(ExchangeRate) + log(InternatonalReserves) + log(LabourForce) + log(Unemployment) +
log(ForeignDirectInvestment) + log(Interestrate), data = Zim_D)

Residuals:
     1          2          3          4          5          6          7          8

-0.0044016  0.0044849 -0.0027752  0.0056140 -0.0100943  0.0008523  0.0061865 -0.0005024

      9          11          12          13          14          18
-0.0007427  -0.0020482  -0.0117588  0.0358940  -0.0185039  -0.0022046

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                 -25.486001   5.458657  -4.669  0.01855 *
log(Inflation)                -0.018030   0.006993  -2.578  0.08192 .
log(Imports)                  0.007525   0.008873   0.848  0.45874
log(Exports)                  0.429373   0.108436   3.960  0.02876 *
log(ExternalDebt)            -0.718740   0.330971  -2.172  0.11828
log(ExchangeRate)            -0.020249   0.004267  -4.746  0.01775 *
log(InternatonalReserves)     0.018627   0.016686   1.116  0.34564
log(LabourForce)              4.304677   0.998708   4.310  0.02299 *
log(Unemployment)            -0.105332   0.169719  -0.621  0.57881
log(ForeignDirectInvestment)  0.058183   0.013562   4.290  0.02328 *
log(Interestrate)            -0.043968   0.005376  -8.179  0.00382 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.02581 on 3 degrees of freedom
(7 observations deleted due to missingness)
Multiple R-squared:  0.9994,    Adjusted R-squared:  0.9976
F-statistic: 530.8 on 10 and 3 DF,  p-value: 0.000121

# Computing the Anova table for M_1

> anova(M_1)
```

```
Analysis of Variance Table

Response: log(GDP)
                             Df  Sum Sq Mean Sq   F value    Pr(>F)
log(Inflation)                1 2.33311 2.33311 3502.2785 1.063e-05 ***
log(Imports)                  1 0.18818 0.18818  282.4753 0.0004587 ***
log(Exports)                  1 0.74964 0.74964 1125.2947 5.823e-05 ***
log(ExternalDebt)             1 0.11526 0.11526  173.0151 0.0009493 ***
log(ExchangeRate)             1 0.05775 0.05775   86.6963 0.0026226 **
log(InternatonalReserves)     1 0.00915 0.00915   13.7408 0.0341155 *
log(LabourForce)              1 0.03183 0.03183   47.7739 0.0062071 **
log(Unemployment)             1 0.00544 0.00544    8.1732 0.0646358 .
log(ForeignDirectInvestment)  1 0.00135 0.00135    2.0240 0.2499869
log(Interestrate)             1 0.04456 0.04456   66.8970 0.0038236 **
Residuals                     3 0.00200 0.00067
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
> AIC(M_1)
[1] -60.23151
> BIC(M_1)
[1] -52.56282


################################################################################################

>   # Adjusted Model Without Log Unemployment

> M_2=lm(log(GDP)~log(Inflation)+log(Imports)+log(Exports)+log(ExternalDebt)+log(ExchangeRate)+
      log(InternatonalReserves)+log(LabourForce)+log(ForeignDirectInvestment)+log(Interestrate),
      data=Zim_D)

Warning messages:
1: In log(Inflation) : NaNs produced
2: In log(Interestrate) : NaNs produced

> summary(M_2)

Call:
lm(formula = log(GDP) ~ log(Inflation) + log(Imports) + log(Exports) + log(ExternalDebt)+
 log(ExchangeRate) + log(InternatonalReserves) + log(LabourForce)+log(ForeignDirectInvestment)+
  log(Interestrate), data = Zim_D)

Residuals:
     1         2         3         4         5         6         7         8
-0.011088  0.001599 -0.001389  0.007579 -0.005614  0.002448  0.001894 -0.001403
    9        11        12        13        14        18
0.000944  0.006508 -0.014683  0.038018 -0.015154 -0.009659

Coefficients:
                               Estimate Std. Error t value Pr(>|t|)
(Intercept)                  -24.695496   4.883024  -5.057 0.007194 **
log(Inflation)                 -0.016966   0.006237  -2.720 0.052983 .
log(Imports)                   0.006304   0.007960   0.792 0.472696
log(Exports)                   0.452887   0.093467   4.845 0.008367 **
```

```
log(ExternalDebt)             -0.600576   0.249050  -2.411 0.073438 .
log(ExchangeRate)             -0.019137   0.003562  -5.372 0.005799 **
log(InternatonalReserves)      0.013664   0.013472   1.014 0.367843
log(LabourForce)               4.054623   0.840670   4.823 0.008504 **
log(ForeignDirectInvestment)   0.055610   0.011879   4.681 0.009437 **
log(Interestrate)             -0.042758   0.004609  -9.277 0.000751 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.02374 on 4 degrees of freedom
(7 observations deleted due to missingness)
Multiple R-squared:  0.9994,    Adjusted R-squared:  0.9979
F-statistic: 696.9 on 9 and 4 DF,  p-value: 5.019e-06

> anova(M_2)
Analysis of Variance Table

Response: log(GDP)
                             Df  Sum Sq Mean Sq   F value    Pr(>F)
log(Inflation)                1 2.33311 2.33311 4138.3757 3.498e-07 ***
log(Imports)                 1 0.18818 0.18818  333.7796 5.280e-05 ***
log(Exports)                 1 0.74964 0.74964 1329.6750 3.377e-06 ***
log(ExternalDebt)            1 0.11526 0.11526  204.4388 0.0001390 ***
log(ExchangeRate)            1 0.05775 0.05775  102.4424 0.0005363 ***
log(InternatonalReserves)    1 0.00915 0.00915   16.2365 0.0157407 *
log(LabourForce)             1 0.03183 0.03183   56.4508 0.0016795 **
log(ForeignDirectInvestment)  1 0.00258 0.00258    4.5751 0.0992055 .
log(Interestrate)            1 0.04852 0.04852   86.0662 0.0007509 ***
Residuals                    4 0.00226 0.00056
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
> AIC(M_2)
[1] -60.54041
> BIC(M_2)
[1] -53.51078

###############################################################################################

# Adjusted Model without log Unemployment and log imports

> M_3=lm(log(GDP)~log(Inflation)+log(Exports)+log(ExternalDebt)+ log(ExchangeRate)+
      log(InternatonalReserves)+log(LabourForce)+log(ForeignDirectInvestment)+log(Interestrate),
      data=Zim_D)

Warning messages:
1: In log(Inflation) : NaNs produced
2: In log(Interestrate) : NaNs produced

> # The summary for Model M_3

> summary(M_3)

Call:
lm(formula = log(GDP) ~ log(Inflation) + log(Exports) + log(ExternalDebt) + log(ExchangeRate) +
log(InternatonalReserves) + log(LabourForce) + log(ForeignDirectInvestment) +
```

```
log(Interestrate),data = Zim_D)

Residuals:
1          2          3          4          5          6          7
-0.0090437  0.0104072 -0.0095975 -0.0004378 -0.0043701  0.0098463 -0.0016977
8          9          11         12         13         14         18
-0.0022762  0.0020305  0.0115986 -0.0202534  0.0366785 -0.0162471 -0.0066377

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                 -22.750330   4.059992  -5.604 0.002501 **
log(Inflation)                -0.015073   0.005542  -2.720 0.041793 *
log(Exports)                  0.502041   0.067230   7.468 0.000680 ***
log(ExternalDebt)            -0.454476   0.160957  -2.824 0.036953 *
log(ExchangeRate)            -0.018926   0.003417  -5.538 0.002633 **
log(InternatonalReserves)     0.013161   0.012946   1.017 0.355991
log(LabourForce)              3.650184   0.642410   5.682 0.002352 **
log(ForeignDirectInvestment)  0.050989   0.009954   5.123 0.003699 **
log(Interestrate)            -0.043164   0.004406  -9.796 0.000189 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.02284 on 5 degrees of freedom
(7 observations deleted due to missingness)
Multiple R-squared:  0.9993,    Adjusted R-squared:  0.9981
F-statistic: 847.1 on 8 and 5 DF,  p-value: 2.128e-07

> # Anova Table for M_3

> anova(M_3)

Analysis of Variance Table

Response: log(GDP)
                             Df  Sum Sq Mean Sq   F value    Pr(>F)
log(Inflation)                1 2.33311 2.33311 4471.7726 1.416e-08 ***
log(Exports)                  1 0.93768 0.93768 1797.2077 1.378e-07 ***
log(ExternalDebt)             1 0.09400 0.09400  180.1680 4.108e-05 ***
log(ExchangeRate)             1 0.05580 0.05580  106.9585 0.0001455 ***
log(InternatonalReserves)     1 0.02192 0.02192   42.0136 0.0013032 **
log(LabourForce)              1 0.04205 0.04205   80.5915 0.0002861 ***
log(ForeignDirectInvestment)  1 0.00104 0.00104    1.9851 0.2179041
log(Interestrate)             1 0.05006 0.05006   95.9550 0.0001887 ***
Residuals                     5 0.00261 0.00052
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

> # Information Criteria

> AIC(M_3)
[1] -60.50114
> BIC(M_3)
[1] -54.11057
```

```
#####################################################################################

> # Adjusted Model without log Unemployment, log imports and log international reserves

> M_4=lm(log(GDP)~log(Inflation)+log(Exports)+log(ExternalDebt)+ =log(ExchangeRate)+
         log(LabourForce)+log(ForeignDirectInvestment)+log(Interestrate),data=Zim_D)

Warning messages:
1: In log(Inflation) : NaNs produced
2: In log(Interestrate) : NaNs produced

> # The summary for Model M_4

> summary(M_4)

Call:
lm(formula = log(GDP) ~ log(Inflation) + log(Exports) + log(ExternalDebt) + log(ExchangeRate) +
log(LabourForce) + log(ForeignDirectInvestment) + log(Interestrate), data = Zim_D)

Residuals:
    Min        1Q    Median        3Q       Max
-0.027562 -0.008266  0.002127  0.003870  0.034810

Coefficients:
                               Estimate Std. Error t value Pr(>|t|)
(Intercept)                  -19.767131   2.813531  -7.026 0.000415 ***
log(Inflation)                 -0.010639   0.003429  -3.103 0.021039 *
log(Exports)                   0.545516   0.052019  10.487 4.41e-05 ***
log(ExternalDebt)             -0.355010   0.128160  -2.770 0.032420 *
log(ExchangeRate)             -0.020688   0.002954  -7.004 0.000422 ***
log(LabourForce)               3.176269   0.443229   7.166 0.000373 ***
log(ForeignDirectInvestment)   0.053388   0.009697   5.506 0.001507 **
log(Interestrate)             -0.043513   0.004405  -9.877 6.21e-05 ***
---

 Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.02291 on 6 degrees of freedom
(7 observations deleted due to missingness)
Multiple R-squared:  0.9991,    Adjusted R-squared:  0.9981
F-statistic: 962.6 on 7 and 6 DF,  p-value: 1.015e-08

> # The Anova Table for M_4

> anova(M_4)

Analysis of Variance Table

Response: log(GDP)
                             Df  Sum Sq Mean Sq   F value    Pr(>F)
log(Inflation)                 1 2.33311 2.33311 4446.9838 7.648e-10 ***
log(Exports)                  1 0.93768 0.93768 1787.2451 1.172e-08 ***
log(ExternalDebt)             1 0.09400 0.09400  179.1692 1.076e-05 ***
log(ExchangeRate)             1 0.05580 0.05580  106.3656 4.856e-05 ***
log(LabourForce)              1 0.06154 0.06154  117.2957 3.669e-05 ***
```

```
log(ForeignDirectInvestment)  1 0.00180 0.00180    3.4349    0.1133
log(Interestrate)             1 0.05119 0.05119   97.5642 6.214e-05 ***
Residuals                     6 0.00315 0.00052
---
 Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

> # The information Criteria
> AIC(M_4)
[1] -59.87082
> BIC(M_4)
[1] -54.1193

###############################################################################################

> # The Best model selected is model M_3 because it had the highest Adjusted R^2  value and
> # R^2 values

M_3=lm(log(GDP)~log(Inflation)+log(Exports)+log(ExternalDebt)+log(ExchangeRate)+
log(InternatonalReserves)+log(LabourForce)+log(ForeignDirectInvestment)+log(Interestrate),
data=Zim_D)

# Changing the plotting panel to a 2x2 grid

par(mfrow=c(2,2))

# Plotting results for the best model M_3

plot(M_3)
```

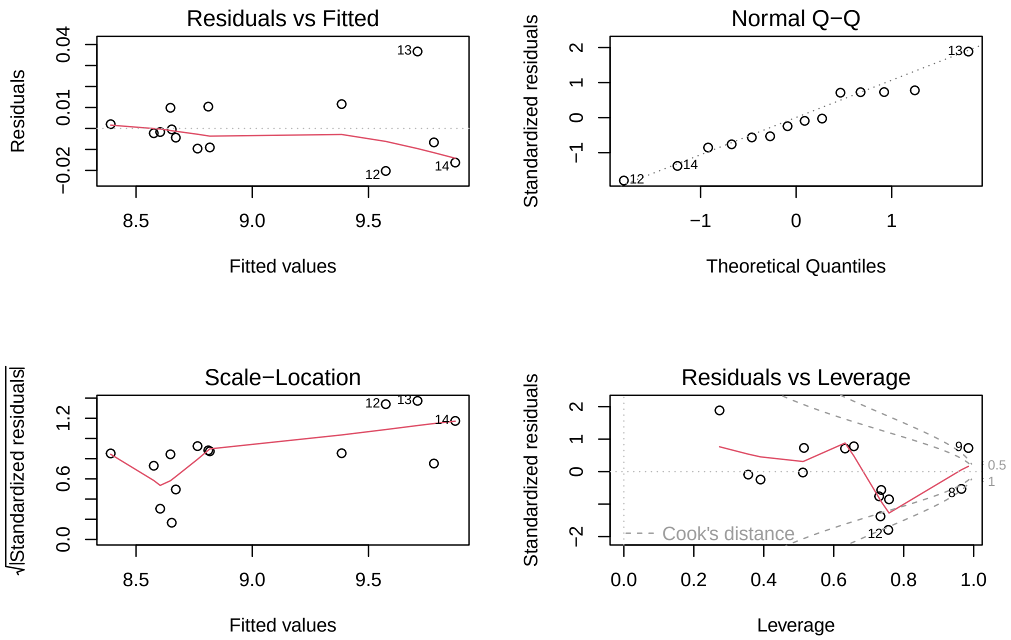


Figure 5.2: M_3 Model Residual Plots

```
# Returning the plotting panel to a 1x1 grid

par(mfrow=c(1,1))

> summary(OLS)

Call:
lm(formula = GDP ~ Inflation + Imports + Exports + EDebt + ERate +
    IReserves + LForce + Unemployment + FDI + Irate, data = SADC_Data_Editted,
    na.action = na.omit)

Residuals:
   Min     1Q Median     3Q    Max
-90633  -3289    205   4223  74947

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.026e+04  1.734e+03  -5.918 8.29e-09 ***
Inflation     -2.733e-04  4.303e-04  -0.635 0.525784
Imports       2.611e+00  1.905e-01  13.706  < 2e-16 ***
Exports       5.247e-01  1.619e-01   3.240 0.001319 **
EDebt         5.181e-01  7.636e-02   6.785 5.49e-11 ***
ERate         2.732e-04  4.291e-04   0.637 0.524802
IReserves    -5.067e-02  2.774e-01  -0.183 0.855169
LForce        6.726e-01  1.220e-01   5.514 7.15e-08 ***
Unemployment  2.005e+02  5.456e+01   3.675 0.000278 ***
FDI          -1.044e+00  5.753e-01  -1.815 0.070429 .
Irate        -2.976e+01  2.899e+01  -1.026 0.305487
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 12770 on 325 degrees of freedom
Multiple R-squared:  0.9757,    Adjusted R-squared:  0.9749
F-statistic:  1304 on 10 and 325 DF,  p-value: < 2.2e-16
```

Code 2

```
> summary(FEWITHIN)
Oneway (individual) effect Within Model

Call:
plm(formula = GDP ~ Inflation + Imports + Exports + EDebt + ERate +
    IReserves + LForce + Unemployment + FDI + Irate, data = SADC_Data_Editted,
    model = "within")

Balanced Panel: n = 16, T = 21, N = 336

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-101009.66   -1312.49     153.06    1677.24   62060.97

Coefficients:
```

```
                Estimate  Std. Error t-value  Pr(>|t|)
Inflation     -1.2681e-04  3.7688e-04 -0.3365 0.7367459
Imports        1.4900e+00  2.1319e-01  6.9892 1.699e-11 ***
Exports        8.4023e-01  1.9490e-01  4.3110 2.186e-05 ***
EDebt          2.6714e-01  7.7494e-02  3.4472 0.0006446 ***
ERate          1.1622e-04  3.7778e-04  0.3076 0.7585611
IReserves      7.7066e-01  3.1257e-01  2.4655 0.0142217 *
LForce         1.4359e+00  4.0745e-01  3.5242 0.0004887 ***
Unemployment   1.1529e+02  1.2011e+02  0.9598 0.3378840
FDI           -7.7830e-01  5.2179e-01 -1.4916 0.1368231
Irate         -3.3191e+01  2.6188e+01 -1.2674 0.2059538
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Total Sum of Squares:    2.4005e+11
Residual Sum of Squares: 3.7545e+10
R-Squared:      0.8436
Adj. R-Squared: 0.83098
F-statistic: 167.204 on 10 and 310 DF, p-value: < 2.22e-16
```

Code 3

```
> summary(FEBETWEEN)
Oneway (individual) effect Between Model

Call:
plm(formula = GDP ~ Inflation + Imports + Exports + EDebt + ERate +
    IReserves + LForce + Unemployment + FDI + IRate, data = SADC_E_Data,
    model = "between")

Balanced Panel: n = 16, T = 22, N = 352
Observations used in estimation: 16

Residuals:
      Angola     Botswana      Comoros          DRC     Eswatini      Lesotho   Madagascar       Malawi
 -3.1075e+03   2.6080e+03   5.2073e+03   4.9695e+02   1.3491e+03  -8.4799e+02  -7.1853e+02  -1.4429e+02
   Mauritius   Mozambique      Namibia   Seychelles South Africa     Tanzania       Zambia     Zimbabwe
 -1.2826e+03  -1.9910e+03  -3.5773e+03   5.8275e+02   8.2915e+02   1.8081e+03  -1.2122e+03   1.0587e-02

Coefficients:
                Estimate  Std. Error t-value Pr(>|t|)
(Intercept)  -1699.48543  4537.27743 -0.3746  0.72335
Inflation        0.46130     1.76911  0.2608  0.80468
Imports          2.74241     1.28630  2.1320  0.08618 .
Exports         -0.38734     0.41669 -0.9296  0.39526
EDebt            1.54063     0.86147  1.7884  0.13374
ERate           -0.40994     1.56944 -0.2612  0.80435
IReserves       -0.16692     0.89035 -0.1875  0.85866
LForce           0.15009     0.22451  0.6685  0.53341
Unemployment   -23.15248   133.68740 -0.1732  0.86930
FDI             -4.20909     2.66623 -1.5787  0.17525
IRate           28.91778   209.28400  0.1382  0.89549
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

```
Total Sum of Squares:    9.4777e+10
Residual Sum of Squares: 71069000
R-Squared:      0.99925
Adj. R-Squared: 0.99775
F-statistic: 666.293 on 10 and 5 DF, p-value: 3.6046e-07
```

Code 4

```
summary(FELSDV)

Call:
lm(formula = GDP ~ Inflation + Imports + Exports + EDebt + ERate +
    IReserves + LForce + Unemployment + FDI + IRate + factor(Country) -
    1, data = SADC_E_Data)

Residuals:
   Min     1Q Median     3Q    Max
-92332  -1755    302   1703  68318

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
Inflation                    -1.763e-04  3.751e-04  -0.470 0.638625
Imports                      1.938e+00  1.642e-01  11.800  < 2e-16 ***
Exports                      2.941e-01  1.101e-01   2.672 0.007920 **
EDebt                        2.473e-01  7.716e-02   3.205 0.001483 **
ERate                        1.656e-04  3.760e-04   0.440 0.659944
IReserves                    1.131e+00  2.762e-01   4.095 5.33e-05 ***
LForce                       1.598e+00  3.679e-01   4.343 1.88e-05 ***
Unemployment                 1.808e+02  1.059e+02   1.708 0.088584 .
FDI                         -1.249e+00  4.778e-01  -2.615 0.009341 **
IRate                       -3.163e+01  2.571e+01  -1.230 0.219449
factor(Country)Angola       -1.476e+04  4.546e+03  -3.246 0.001292 **
factor(Country)Botswana     -1.501e+04  4.641e+03  -3.235 0.001340 **
factor(Country)Comoros      -4.746e+03  2.902e+03  -1.635 0.102938
factor(Country)DRC          -3.312e+04  8.990e+03  -3.683 0.000269 ***
factor(Country)Eswatini     -1.043e+04  5.711e+03  -1.827 0.068591 .
factor(Country)Lesotho      -1.063e+04  4.663e+03  -2.280 0.023227 *
factor(Country)Madagascar   -1.474e+04  4.578e+03  -3.219 0.001416 **
factor(Country)Malawi       -9.334e+03  3.354e+03  -2.783 0.005704 **
factor(Country)Mauritius    -9.439e+03  3.404e+03  -2.773 0.005868 **
factor(Country)Mozambique   -1.872e+04  4.631e+03  -4.042 6.63e-05 ***
factor(Country)Namibia      -1.236e+04  5.089e+03  -2.429 0.015681 *
factor(Country)Seychelles   -3.491e+03  2.693e+03  -1.296 0.195788
factor(Country)South Africa  4.011e+04  1.087e+04   3.691 0.000262 ***
factor(Country)Tanzania     -2.186e+04  8.198e+03  -2.667 0.008041 **
factor(Country)Zambia       -1.103e+04  3.845e+03  -2.869 0.004388 **
factor(Country)Zimbabwe     -1.048e+04  3.560e+03  -2.944 0.003471 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 10970 on 326 degrees of freedom
Multiple R-squared:  0.9858,    Adjusted R-squared:  0.9847
F-statistic:   871 on 26 and 326 DF,  p-value: < 2.2e-16
```

Code 5

```
> cor(Zim_Vars)
                    GDP   Inflation     Imports     Exports       EDebt       ERate   IReserves      LForce
GDP           1.0000000 -0.27816725  0.74690938  0.86975147  0.71906643 -0.31025435 -0.10815233  0.8943982
Inflation    -0.2781672  1.00000000 -0.18161413 -0.23377302 -0.12021295  0.99147941  0.11969600 -0.1139896
Imports       0.7469094 -0.18161413  1.00000000  0.70458801  0.45990109 -0.23523921 -0.04198899  0.6536005
Exports       0.8697515 -0.23377302  0.70458801  1.00000000  0.78739588 -0.26083413 -0.08346805  0.8856775
EDebt         0.7190664 -0.12021295  0.45990109  0.78739588  1.00000000 -0.13627911 -0.06838381  0.8408654
ERate        -0.3102544  0.99147941 -0.23523921 -0.26083413 -0.13627911  1.00000000  0.10248901 -0.1317129
IReserves    -0.1081523  0.11969600 -0.04198899 -0.08346805 -0.06838381  0.10248901  1.00000000 -0.2566654
LForce        0.8943982 -0.11398959  0.65360046  0.88567750  0.84086536 -0.13171290 -0.25666544  1.0000000
Unemployment -0.3188318 -0.09432994 -0.29454518 -0.43436229 -0.69752050 -0.10768970  0.42689234 -0.6082751
FDI           0.7187137 -0.17506140  0.64807950  0.56861088  0.19496959 -0.19498799 -0.17804934  0.5988189
IRate        -0.3100623  0.05039166  0.08711928 -0.26901485 -0.40597700  0.06267713  0.11429660 -0.4000585
             Unemployment         FDI       IRate
GDP           -0.31883175  0.71871366 -0.31006235
Inflation     -0.09432994 -0.17506140  0.05039166
Imports       -0.29454518  0.64807950  0.08711928
Exports       -0.43436229  0.56861088 -0.26901485
EDebt         -0.69752050  0.19496959 -0.40597700
ERate         -0.10768970 -0.19498799  0.06267713
IReserves      0.42689234 -0.17804934  0.11429660
LForce        -0.60827512  0.59881887 -0.40005847
Unemployment   1.00000000  0.01226578  0.18954690
FDI            0.01226578  1.00000000 -0.06796165
IRate          0.18954690 -0.06796165  1.00000000
>
```

Code 6

```
> summary(OLS)

Call:
lm(formula = GDP ~ Inflation + Imports + Exports + EDebt + ERate +
    IReserves + LForce + Unemployment + FDI + Irate, data = SADC_Data_Editted,
    na.action = na.omit)

Residuals:
   Min     1Q Median     3Q    Max
-90633  -3289    205   4223  74947

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.026e+04  1.734e+03  -5.918 8.29e-09 ***
Inflation     -2.733e-04  4.303e-04  -0.635 0.525784
Imports       2.611e+00  1.905e-01  13.706  < 2e-16 ***
Exports       5.247e-01  1.619e-01   3.240 0.001319 **
EDebt         5.181e-01  7.636e-02   6.785 5.49e-11 ***
ERate         2.732e-04  4.291e-04   0.637 0.524802
IReserves    -5.067e-02  2.774e-01  -0.183 0.855169
LForce        6.726e-01  1.220e-01   5.514 7.15e-08 ***
Unemployment  2.005e+02  5.456e+01   3.675 0.000278 ***
FDI          -1.044e+00  5.753e-01  -1.815 0.070429 .
```

```
Irate         -2.976e+01  2.899e+01  -1.026 0.305487
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 12770 on 325 degrees of freedom
Multiple R-squared:  0.9757,    Adjusted R-squared:  0.9749
F-statistic:  1304 on 10 and 325 DF,  p-value: < 2.2e-16
```

Random Forests

```
    setwd("C:/Users/jchit/Desktop/old/Musora PhD")
library(readxl)
library(dplyr)
library(ggplot2)
library(corrplot)
library(caret)
# install.packages("plm")
library(plm)
library(lmtest)
# install.packages("randomForest")
library(randomForest)
# install.packages("xgboost")
library(xgboost)
# install.packages("neuralnet")
library(neuralnet)

data <-read_excel("SADC E Data (2).xlsx")

panel_data <- pdata.frame(data, index = c("Country", "Year"))

# print(colnames(panel_data))

summary(panel_data)

colSums(is.na(panel_data))

correlation_matrix <- cor(panel_data[, c("GDP", "Inflation", "Imports", "Exports", "EDebt", "ERate",
"IReserves", "LForce", "Unemployment", "FDI", "IRate")])corrplot(correlation_matrix,
method = "color")

#simple lr

lr = lm(GDP ~ Inflation + Imports + Exports + EDebt + ERate + IReserves + LForce + Unemployment +
FDI +IRate + factor(Year) + factor(Country),data = data)

summary(lr)

#simple plm
#plm_lr = plm(GDP ~ Inflation + Imports + Exports + EDebt + ERate + IReserves + LForce +
Unemployment + FDI + IRate + factor(Year),#index = "Country",
             #model = "within",
```

```
             #data = data)
#summary(plm_lr)

fixed_effects_model <- plm(GDP ~ Inflation + Imports + Exports + EDebt + ERate + IReserves + LForce +
Unemployment + FDI + IRate,data = panel_data,model = "within")

random_effects_model <- plm(GDP ~ Inflation + Imports + Exports + EDebt + ERate + IReserves + LForce
+ Unemployment + FDI + IRate,data = panel_data,model = "random")

summary(fixed_effects_model)

bptest(fixed_effects_model)

phtest(fixed_effects_model,random_effects_model)

set.seed(123)
train_indices <- createDataPartition(panel_data$GDP, p = 0.7, list = FALSE)
train_data <- panel_data[train_indices, ]
test_data <- panel_data[-train_indices, ]

rf_model <- randomForest(GDP ~ ., data = train_data, ntree = 1000)

rf_predictions <- predict(rf_model, newdata = test_data)

rf_rmse <- sqrt(mean((rf_predictions - test_data$GDP)^2))
rf_rmse
20522.5737
rf_r2 <- cor(rf_predictions, test_data$GDP)^
rf_r2
0.9742

importance <- rf_model$importance
importance



varImpPlot(rf_model)
```

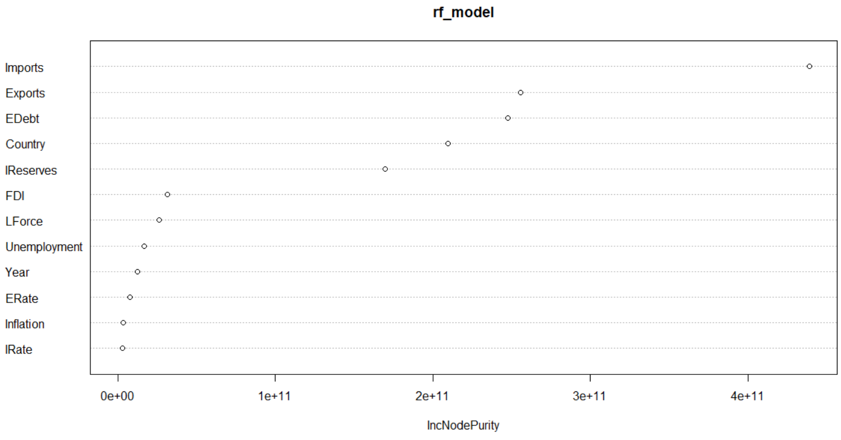


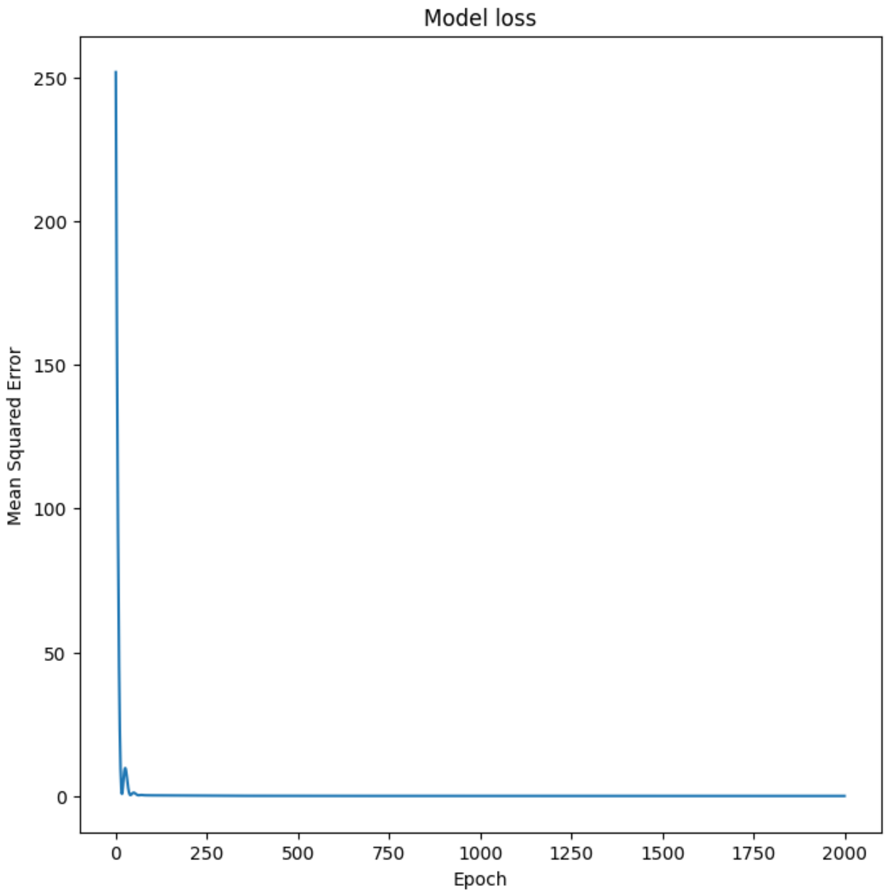
Figure 5.3: Important variables

Deep Learning

```
# Code is written in Google Colab (Python version)
# Importing the necessary libraries

import numpy as np
import matplotlib.pyplot as plt
from google.colab import files
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
import tensorflow as tf
# Importing sequential to be able to create keras models
from tensorflow.keras import Sequential
from sklearn.metrics import mean_squared_error, mean_absolute_error,confusion_matrix
# Importing Dense to be able to create the network layers
from tensorflow.keras.layers import Dense
from tensorflow import metrics

# Importing the Data
five=pd.read_csv("five_years.csv")
five=five.values


data = pd.read_csv("ZIM_Tran.csv")
database=data.values

data.head()
# Asssinging features to the variable X and targets to the variable y
X=database[:,1:12]
y=database[:,0]



# Splliting the data into 70% training and 30% testing
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3)




# Creating a densely connected function
def baseline():
    # create model
    model = Sequential()

    # Adding a fully connected layer with 64 neurons.
    model.add(Dense(units = 64, input_dim=10,activation='linear'))
    # adding another fully connected layer with 10 neurons.
    model.add(Dense(units=10, activation='linear'))

    model.add(Dense(units=1, activation='linear'))
    # Compile model

    # Selecting the Nadam as the optimizer
    model.compile(loss='mse', optimizer='Nadam',metrics=[metrics.mse])
```

```
    return model

# Initilizing the model
model = baseline()


# Obtainaning the model summary
model.summary()

# Training the model
history=model.fit(X_train, y_train,epochs=500)

# Predictiong using the testing data set
predictions = model.predict(X_test)

# Creating a forecast for GDP for the next five years
prediction = model.predict(five)
prediction

# Computing the mean squared error between the observed and the predicted

mean_squared_error(y_test, predictions)
0.01486329505536518



# Obtaining the model final weights
z=model.get_layer('dense_2').get_weights()
len(z[0])

M=model.get_layer('dense_2').get_weights()
c=np.zeros(10)

for i in range(10):
  x=(z[0][i][0])
  c[i]=round(x,4)

for i in range(10):
  c[i]=round(c[i],4)
  v=round(z[1][0],4)
v=round(v,4)
v
# Dislplaying the Final model

print("The Machine Learning model that describes Zimbabwes GDP is:\n\n GDP={}{}log(inf){}log(imp)
{}log(Exp)+
{}log(EDebt)+{}log(ERate)+{}log(iRes)+{}log(LF)+
{}log(UE){}log(FDI)\n     {}log(Irate)".format(0.0009,c[0],c[1], c[2], c[3],
c[4], c[5], c[6], c[7], c[8], c[9]))

log (GDP) =0.0173 + 0.015 ln(inflation) + 0.1903ln(Exports) + 0.627ln( EDebt)  0.0817log (ERate)
 0.4401ln(iReserves)  0.0585ln(LForce)  0.3919ln(Unemployment)  0.3578ln( FDI)+ 0.1037ln(Irate)

plt.figure(figsize=(8, 8))
```

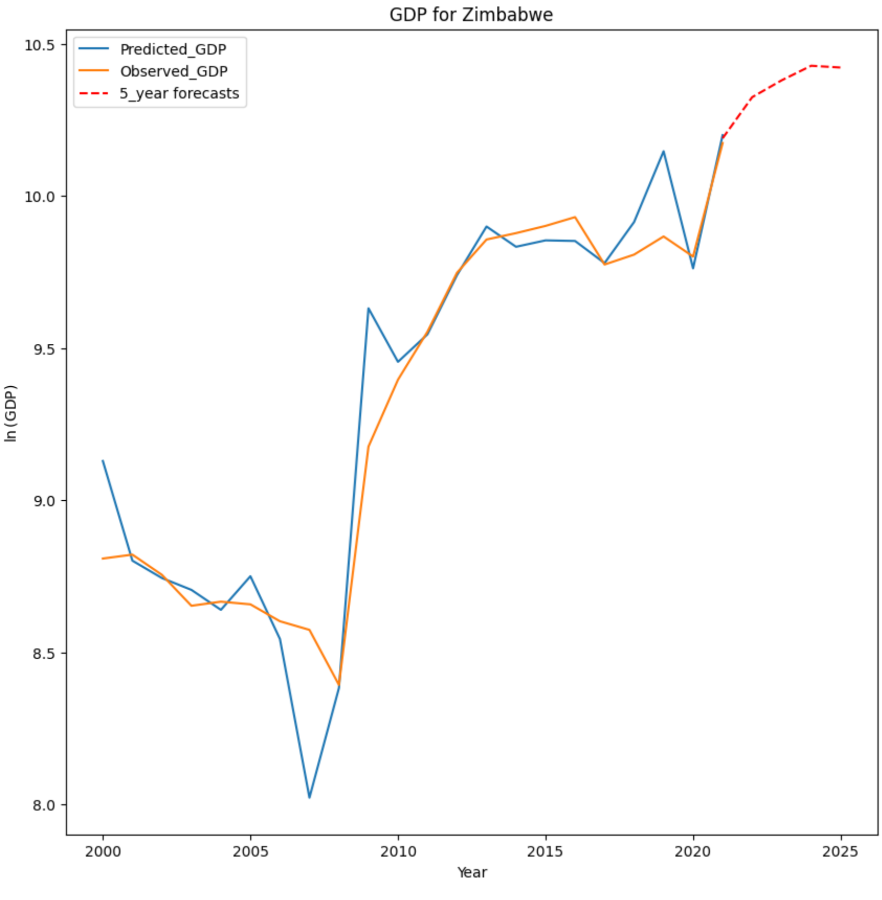
```
plt.plot(history.history['mean_squared_error'])
plt.title('Model loss')
plt.ylabel('Mean Squared Error')
plt.xlabel('Epoch')
plt.show()
```



5.8 Generating Five-Year Forecasts

```
# Printing the next 5 year's GDPs starting 2021
pred=np.zeros(5)
years=np.zeros(22)
f_years=np.zeros(5)
for i in range(22):
   year=int(2000+i)
   years[i]=year
for i in range(5):
   f=(prediction[i][0])
   z=np.exp(f)
   pred[i]=f
   f_year=int(2021+i)
   f_years[i]=f_year
   print("Predicted GDP for the year ",f_year, " is ", z)

The predicted GDP for the year  2021  is  26626.951
```

```
The predicted GDP for the year 2022 is 30476.693
The predicted GDP for the year 2023 is 32215.355
The predicted GDP for the year 2024 is 33794.71
The predicted GDP for the year 2025 is 33597.113

plt.figure(figsize=(10, 10))
plt.plot(years,p,label="Predicted_GDP")
plt.plot(years,y,label="Observed_GDP")
plt.plot(f_years,pred,'r--',label="5_year forecasts")
plt.title("GDP for Zimbabwe")
plt.ylabel(r"$\ln$(GDP))")
plt.xlabel("Year")
plt.legend()
plt.show()
```



Support Vector Machine

```
    numeric_vars <- c("GDP", "Inflation", "Imports", "Exports",
                  "EDebt", "ERate", "IReserves", "LForce",
                  "Unemployment", "FDI", "IRate")
train_data_numeric <- train_data[, numeric_vars]

svr_model <- svm(GDP ~ ., data = train_data_numeric, kernel = "radial")

svr_predictions <- predict(svr_model, newdata = test_data[, numeric_vars])
```

```
svr_rsquared <- 1 - sum((actual - svr_predictions)^2) / sum((actual -
                                                     mean(actual))^2)
svr_accuracy <- svr_rsquared * 100

print(paste("Accuracy (R-squared):", svr_accuracy, "%"))
Accuracy (R-squared): 93.86%
```

Gradient Boosting

```
# XGBoost model

numeric_vars <- c("GDP", "Inflation", "Imports", "Exports", "EDebt",
                  "ERate", "IReserves", "LForce", "Unemployment", "FDI",
                  "IRate")
train_data_numeric <- train_data[, numeric_vars]

train_matrix <- xgb.DMatrix(as.matrix(train_data_numeric[, -1]),
                            label = train_data_numeric$GDP)
test_matrix <- xgb.DMatrix(as.matrix(test_data[, numeric_vars[-1]]),
                           label = test_data$GDP)

params <- list(
  objective = "reg:squarederror",
  eval_metric = "rmse"
)

watchlist <- list(train = train_matrix, test = test_matrix)

xgb_model <- xgb.train(
  params = params,
  data = train_matrix,
  nrounds = 100,
  early_stopping_rounds = 10,
  watchlist = watchlist,
  verbose = 0
)

xgb_predictions <- predict(xgb_model, test_matrix)

xgb_rmse <- sqrt(mean((xgb_predictions - test_data$GDP)^2))
print(paste("RMSE:", xgb_rmse))

actual <- test_data$GDP
xgb_rsquared <- 1 - sum((actual - xgb_predictions)^2) / sum((actual
                                                   - mean(actual))^2)
xgb_accuracy <- xgb_rsquared * 100

print(paste("Accuracy (R-squared):", xgb_accuracy, "%"))

Accuracy (R-squared): 97.59%
```

Bibliography

Abidin, I. S. Z., Haseeb, M., Muhammad, A., & Islam, R. (2015). Foreign direct investment, financial development, internationaltrade and energy consumption: Panel data evidence fromselected asean countries. *International Journal of Energy Economics and Policy*, *5*(3), 841–850.

Acikgoz, S., & Mert, M. (2014). Sources of growth revisited: The importance of the nature of technological progress. *Journal of Applied Economics*, *17*(1), 31–62.

Adjasi, C. K., Abor, J., Osei, K. A., & Nyavor-Foli, E. E. (2012). Fdi and economic activity in africa: The role of local financial markets. *Thunderbird International Business Review*, *54*(4), 429–439.

Agalega, E., & Antwi, S. (2013). The impact of macroeconomic variables on gross domestic product: Empirical evidence from ghana. *International Business Research*, *6*(5), 108.

Agbor, J. A., Fedderke, J., & Viegi, N. (2014). How colonial educational practices helped shape the pattern of decolonisation in west africa. *International Journal of Development and Conflict*, *4*, 1–23.

Ahmed, H. A., & Uddin, M. G. S. (2009). Export, imports, remittance and growth in bangladesh: An empirical analysis. *Trade and Development review*, *2*(2), 79–92.

Akbar, A., Imdadullah, M., Ullah, M. A., & Aslam, M. (2011). Determinants of economic growth in asian countries: A panel data perspective. *Pakistan Journal of Social Sciences (PJSS)*, *31*(1).

Akinlo, A. E. (2006). The stability of money demand in nigeria: An autoregressive distributed lag approach. *Journal of Policy modeling*, *28*(4), 445–452.

Akinlo, A. E., et al. (2005). *Impact of macroeconomic factors on total factor productivity in sub-saharan african countries* (Vol. 39). United nations university. World institute for development economics

Akitoby, B., & Cinyabuguma, M. M. (2004). Sources of growth in the democratic republic of the congo: A cointegration approach.

Al Harrasia, N., El Din, M. S., Al Rashdi, M., & Hakrod, A. N. (2018). Foreign direct investment determinants and oman vision 2040.

Anaman, K. A. (2006). *Determinants of economic growth in ghana*. Institute of Economic Affairs.

Antwi, S., Mills, E., Zhao, X., et al. (2013). Impact of macroeconomic factors on economic growth in ghana: A cointegration analysis. *International journal of academic research in accounting, finance and management sciences*, *3*(1), 35–45.

Anyanwu, J. C. (2014). Factors affecting economic growth in africa: Are there any lessons from china? *African Development Review*, *26*(3), 468–493.

Asheghian, P. (2009). Determinants of economic growth in japan: The role of foreign direct investment. *Global Economy Journal*, *9*(3), 1850171.

Ayyoub, M., Chaudhry, I. S., & Farooq, F. (2011). Does inflation affect economic growth? the case of pakistan. *Pakistan Journal of Social Sciences (PJSS)*, *31*(1).

Aziz, M. N., & Hossain, M. A. (2012). Intrinsic and extrinsic determinants of economic growth: Further evidence.

Bailar, B. A. (1989). Information needs, surveys, and measurement errors. *Panel surveys*, 1–24.

Baltagi, B. H., & Chang, Y.-J. (1994a). Incomplete panels: A comparative study of alternative estimators for the unbalanced one-way error component regression model. *Journal of Econometrics*, *62*(2), 67–89.

Baltagi, B. H., & Chang, Y.-J. (1994b). Incomplete panels: A comparative study of alternative estimators for the unbalanced one-way error component regression model. *Journal of Econometrics*, *62*(2), 67–89.

Baltagi, B. H., & Griffin, J. M. (1988). A general index of technical change. *Journal of political Economy*, *96*(1), 20–41.

Baltagi, B. H., Griffin, J. M., & Rich, D. P. (1995). The measurement of firm-specific indexes of technical change. *The Review of Economics and Statistics*, 654–663.

Baltagi, B. H., & Levin, D. (1992). Cigarette taxation: Raising revenues and reducing consumption. *Structural Change and Economic Dynamics*, *3*(2), 321–335.

Baltagi, B. H., & Baltagi, B. H. (2008). *Econometric analysis of panel data* (Vol. 4). Springer.

Barro, R. J. (1995). J., and xavier sala-i-martin. *Economic growth*, *2*.

Barro, R. J. (1999). Determinants of economic growth: Implications of the global evidence for chile. *Cuadernos de economía*, 443–478.

Barro, R. J., et al. (2003). Determinants of economic growth in a panel of countries. *Annals of economics and finance*, *4*, 231–274.

Bashar, O. K., & Khan, H. (2007). Liberalization and growth: An econometric study of bangladesh.

Becketti, S., Gould, W., Lillard, L., & Welch, F. (1988). The panel study of income dynamics after fourteen years: An evaluation. *Journal of Labor Economics*, *6*(4), 472–492.

Bekere, B., & Bersisa, M. (2018). Impact of foreign direct investment on economic growth in eastern africa. *Determinants of economic growth in Africa*, 95–124.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly journal of economics*, *119*(1), 249–275.

Bhanu Sireesha, P. (2013). Effect of select macro economic variables on stock returns in india. *International Journal of Marketing, Financial Services & Management Research*, *2*(6).

Blundell, R. (1988). Consumer behaviour: Theory and empirical evidencea survey. *The economic journal*, *98*(389), 16–65.

Breusch, T. S., & Pagan, A. R. (1980). The lagrange multiplier test and its applications to model specification in econometrics. *The review of economic studies*, *47*(1), 239–253.

Brown, J. N., & Light, A. (1992). Interpreting panel data on job tenure. *Journal of Labor Economics*, *10*(3), 219–257.

Brück, T., & Van den Broeck, K. (2006). Growth, employment and poverty in mozambique. *Issues in Employment and Poverty Discussion Paper*, (21).

Cameron, A. C., Trivedi, P. K., et al. (2010). *Microeconometrics using stata* (Vol. 2). Stata press College Station, TX.

Chang, C.-C., & Mendy, M. (2012). Economic growth and openness in africa: What is the empirical relationship? *Applied Economics Letters*, *19*(18), 1903–1907.

Chea, A. C. (2012). Sub-saharan africa and global trade: What sub-saharan africa needs to do to maximize the benefits from global trade integration, increase economic growth and reduce poverty. *International Journal of Academic Research in Business and Social Sciences*, *2*(4), 360–361.

Checherita-Westphal, C. D., & Rother, P. (2010). The impact of high and growing government debt on economic growth: An empirical investigation for the euro area.

Chen, B., & Feng, Y. (2000). Determinants of economic growth in china: Private enterprise, education, and openness. *China Economic Review*, *11*(1), 1–15.

Chirwa, T. G. (2016). What drives long-run economic growth? empirical evidence from south africa.

Chirwa, T. G., & Odhiambo, N. M. (2015). The macroeconomic drivers of economic growth in malawi. *Socioeconomica*, *4*(7).

Chirwa, T. G., & Odhiambo, N. M. (2014). An empirical test of exogenous growth models: Evidence from three southern african countries. *Economic Annals*, *64*(220), 7–38.

Chiu, C.-Y., Chang, C.-C., & Lee, S.-D. (2011). Gene expression profiles for breast cancer survival analysis using support vector machines. *Journal of Biomedical Science and Engineering*, *4*(9), 562–566.

Cornwell, C., Schmidt, P., & Sickles, R. C. (1990). Production frontiers with cross-sectional and time-series variation in efficiency levels. *Journal of econometrics*, *46*(1-2), 185–200.

Deaton, A. (1995). Data and econometric tools for development analysis. *Handbook of development economics*, *3*, 1785–1882.

Dobronogov, A., & Iqbal, F. (2005). Economic growth in egypt: Constraints and determinants: Retrieved april 14, 2009, from htt. *site. sources. worldbank. org/INTMENA/Resources/WP42SEPTEMBER2006pdf.*

Duncan, G. J., & Hill, D. H. (1985). An investigation of the extent and consequences of measurement error in labor-economic survey data. *Journal of Labor Economics*, *3*(4), 508–532.

Easterly, W., & Levine, R. (1997). Africa's growth tragedy: Policies and ethnic divisions. *The quarterly journal of economics*, 1203–1250.

Eita, J. H., & Ashipala, J. M. (2010). Determinants of unemployment in namibia. *International Journal of Business and Management*, *5*(10), 92.

Erkut, B., & Sharma, G. D. (2018). Financial integration in asia: New empirical evidence using dynamic panel data estimations. *International Economics and Economic Policy*, *20*(1), 213–231.

Feng, Y. (1996). Democracy and growth: The sub-saharan african case, 1960-1992. *The Review of Black Political Economy*, *25*(1), 95–126.

Fischer, S. (1992). Macroeconomic stability and growth. *Cuadernos de Economía*, *29*(87), 171–186.

Frimpong Magnus, J., & Oteng-Abayie, E. F. (2006). Bounds testing approach: An examination of foreign direct investment, trade, and growth relationships. *American Journal of Applied Sciences, Forthcoming.*

Ghura, D., Basu, A., & Calamitsis, A. E. (2001). Sub-saharan africa learning what works. In *Promoting growth in sub-saharan africa*. International Monetary Fund.

Greene, W. H. (2003). *Econometric analysis*. Pearson Education India.

Griliches, Z. (1986). Economic data issues. *Handbook of econometrics*, *3*, 1465–1514.

Gujarati, D. N., & Porter, D. C. (2003). Basic econometrics, new york: Mc grawhill companies.

Gyimah-Brempong, K. (1989). Defense spending and economic growth in subsaharan africa: An econometric investigation. *Journal of Peace research*, *26*(1), 79–90.

Hajivassiliou, V. A. (1987). The external debt repayments problems of ldc's: An econometric model based on panel data. *Journal of Econometrics*, *36*(1-2), 205–230.

Hamilton, J. D., & Monteagudo, J. (1998). The augmented solow model and the productivity slowdown. *Journal of Monetary Economics*, *42*(3), 495–509.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the econometric society*, 1251–1271.

Hausman, J. A., & Wise, D. A. (1979). Attrition bias in experimental and panel data: The gary income maintenance experiment. *Econometrica: Journal of the Econometric Society*, 455–473.

Herriot, R. A., & Spiers, E. F. (1975). *Measuring the impact on income statistics of reporting differences between the current population survey and administrative sources*. US Census Bureau [custodian].

Horowitz, J. L., & Manski, C. F. (1998). Censoring of outcomes and regressors due to survey nonresponse: Identification and estimation using weights and imputations. *Journal of Econometrics*, *84*(1), 37–58.

Hostan, J. (2015). A panel data analysis of sub-saharan africa's economic growth.

Hsiao, C. (2003). Analysis of panel data cambridge. *Cambridge University*.

Hu, J., & Szymczak, S. (2023). A review on longitudinal data analysis with random forest. *Briefings in Bioinformatics*, *24*(2), bbad002.

Kalton, G. (1989). Nonsampling errors in panel surveys. *Panel surveys*, 249–270.

Kennedy, P. (2008). *A guide to econometrics*. John Wiley & Sons.

Kholdy, S. (1995). Causality between foreign investment and spillover efficiency. *Applied economics*, *27*(8), 745–749.

Kibria, U., Arshad, M. U., Kamran, M., Mehmood, Y., Imdad, S., & Sajid, M. (2014). Exploring the impact of macro economic variables on gdp growth of pakistan. *Research Journal of Management Sciences ISSN*, *2319*, 1171.

Kim, J.-H., Lee, K.-J., & Kim, H.-J. (2009). Application of support vector regression to the prediction of stock returns with financial ratios. *Expert Systems with Applications*, *36*(3), 5599–5606.

Klasen, S., & Lawson, D. (2007). *The impact of population growth on economic growth and poverty reduction in uganda* (tech. rep.). Diskussionsbeiträge.

Klevmarken, N. A. (1989). Panel studies: What can we learn from them? *European Economic Review*, *33*, 523–529.

Knight, M., Loayza, N., & Villanueva, D. (1993). Testing the neoclassical theory of economic growth: A panel data approach. *Staff papers*, *40*(3), 512–541.

Kumbhakar, S., & Lovell, K. (2000). Stochastic frontier analysis, cambridge: Cambridge university press.

Leamer, E. E. (2008). *Macroeconomic patterns and stories*. Springer Science & Business Media.

Liang, J., & Huang, J. (2011). Fraud detection for online retailers using support vector machines. *Decision Support Systems*, *50*(3), 577–586.

Maccarrone, G., Morelli, G., & Spadaccini, S. (2021). Gdp forecasting: Machine learning, linear or autoregression? *Frontiers in Artificial Intelligence*, *4*, 757864.

M'amanja, D., & Morrissey, O. (2006). *Foreign aid, investment and economic growth in kenya: A time series approach* (tech. rep.). CREDIT research paper.

Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A contribution to the empirics of economic growth. *The quarterly journal of economics*, *107*(2), 407–437.

Maripe, M. O., Mupimpila, C., & Okurut, F. N. (2017). Infrastructure and economic growth in southern african development community (sadc). *Int. J. Econ*, (2017), 10.

Mbulawa, S. (2015a). Effect of macroeconomic variables on economic growth in botswana.

Mbulawa, S. (2015b). Effect of macroeconomic variables on economic growth in botswana.

Mongale, I. P., & Monkwe, K. (2015). The determinants of growth in the south african economy: Cvar analysis. *Journal of Governance and Regulation*, *4*(3), 271–276.

Mosikari, T. J., Xaba, D. L., & Tsoku, J. T. (2016). Macroeconomic determinants of economic growth in botswana: The keynesian approach. *Proceedings of International Academic Conferences*, (4006365).

Most, S. J., & De Berg, H. V. (1996). Growth in africa: Does the source of investment financing matter? *Applied Economics*, *28*(11), 1427–1433.

Musila, J. W., & Belassi, W. (2004). The impact of education expenditures on economic growth in uganda: Evidence from time series data. *The Journal of Developing Areas*, 123–133.

Ncube, T. M. (2019). *Determinants of economic growth-the case of zimbabwe* [Master's thesis, Faculty of Commerce].

Ndambendia, H., & Njoupouognigni, M. (2010). Foreign aid, foreign direct investment and economic growth in sub-saharan africa: Evidence from pooled mean group estimator (pmg). *International journal of economics and finance*, *2*(3), 39–45.

Nhlengethwa, S., Matchaya, G., Greffiths, I., & Fakudze, B. (2021). Analysis of the determinants of public capital investments on agricultural water infrastructure in eswatini. *Business Strategy & Development*, *4*(1), 49–58.

Nyoni, T., & Bonga, W. G. (2017). Foreign aid–economic growth nexus: A systematic review of theory & evidence from developing countries. *Dynamic Research Journals' Journal of Economics & Finance*, *2*(7), 01–16.

Ojo, O., & Oshikoya, T. (1995). Determinants of long-term growth: Some african results. *Journal of African economies*, *4*(2), 163–191.

Peracchi, F. (2002). The european community household panel: A review. *Empirical Economics*, *27*, 63–90.

Rahman, M. M., Rahman, M. S., & Hai-Bing, W. (2011). Time series analysis of causal relationship among gdp, agricultural, industrial and service sector growth in bangladesh. *China-USA Business Review*, *10*(1).

Raissi, M. (2018). Deep hidden physics models: Deep learning of nonlinear partial differential equations. *The Journal of Machine Learning Research*, *19*(1), 932–955.

Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, *378*, 686–707.

Rao, B. B., & Hassan, G. (2011). Determinants of the long-run growth rate of bangladesh. *Applied Economics Letters*, *18*(7), 655–658.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat, f. (2019). Deep learning and process understanding for data-driven earth system science. *Nature*, *566*(7743), 195–204.

Richardson, A., van Florenstein Mulder, T., & Vehbi, T. (2021). Nowcasting gdp using machine-learning algorithms: A real-time assessment. *International Journal of Forecasting*, *37*(2), 941–948.

Ristanovic, V. (2010). Macroeconomic determinant of economic growth and world economic-financial crisis. *Facta Universitatis (Economics and Organization) Vol*, *7*, 17–33.

Robinson, S.-a., & Dornan, M. (2017). International financing for climate change adaptation in small island developing states. *Regional Environmental Change*, *17*, 1103–1115.

Sahu, P., & Sharma, N. K. (2018). Impact of trade openness on inflation in india: An autoregressive distributed lag (ardl) approach. *The Empirical Economics Letters*, *17*(1).

Salian, P., & Gopakumar, K. (2008). Inflation and economic growth in india–an empirical analysis. *Indian Economic Service, New Delhi and Gopakumar. K, Faculty, BIET-MBA Programme, Davangere, Karnataka*.

Salisu, A. A., Ogwumike, F. O., et al. (2010). Aid, macroeconomic policy environment and growth: Evidence from sub-saharan africa. *Journal of Economic Theory*, *4*(2), 59–64.

Samanta, S., & Sanyal, R. (2009). Relationship between perceived bribery and economic growth: An empirical analysis. *Global Business & Finance Review*, *14*(2), 117.

Sekantsi, L. P., & Kalebe, K. M. (2015). Savings, investment and economic growth in lesotho: An empirical analysis. *Journal of Economics and International Finance*, *7*(10), 213.

Seleteng, M., & Motelle, S. (2014). Sources of economic growth in the southern african development community: Its likely impact on povery and employment. *Rev Econ Bus Stud*, *9*(2), 211–249.

Semuel, H., & Nurina, S. (2014). *Analysis of the effect of inflation, interest rates, and exchange rates on gross domestic product (gdp) in indonesia* [Doctoral dissertation, Petra Christian University].

Sen, K., & Te Velde, D. W. (2009). State business relations and economic growth in sub-saharan africa. *Journal of Development Studies*, *45*(8), 1267–1283.

Shahbaz, M., Ahmad, K., & Chaudhary, A. (2008). Economic growth and its determinants in pakistan. *The Pakistan Development Review*, 471–486.

Shim, J. K., Siegel, J. G., & Siegel, J. G. (1995). *Dictionary of economics*. John Wiley & Sons Incorporated.

Steven, R., Jeffrey, S., & Jong-Wha, L. (2001). The determinants and prospects of economic growth in asia. *International Economic Journal*, *15*(3), 1–29.

Suits, D. B. (1984). Dummy variables: Mechanics v. interpretation. *The Review of Economics and Statistics*, 177–180.

Syed, A. A. S. G., & Shaikh, F. M. (2013). Effects of macroeconomic variables on gross domestic product (gdp) in pakistan. *Procedia Economics and Finance*, *5*, 703–711.

Team, R. D. C. (2023). R: A language and environment for statistical computing. *(No Title)*.

The World Bank, (2023). World development indicators. *http://data.worldbank.org/indicator/*.

Tiwari, A. K., & Mutascu, M. (2011). Economic growth and fdi in asia: A panel-data approach. *Economic analysis and policy*, *41*(2), 173–187.

Uddin, G. S., & Sjö, B. (2013). Remittances, financial development and economic growth in bangladesh. *South Asia Economic Journal*, *14*(2), 261–273.

Van den Broeck, J., Koop, G., Osiewalski, J., & Steel, M. F. (1994). Stochastic frontier models: A bayesian perspective. *Journal of econometrics*, *61*(2), 273–303.

Wang, Y., Li, Y., & Wang, J. (2014). Electricity consumption forecasting using support vector regression with ensemble empirical mode decomposition. *Energy Conversion and Management*, *85*, 456–464.

Winkelmann, L., & Winkelmann, R. (1998). Why are the unemployed so unhappy? evidence from panel data. *Economica*, *65*(257), 1–15.

Yang, C.-R., & Ching, N.-H. (2008). A support vector machine for customer churn prediction in the insurance industry. *Expert Systems with Applications*, *34*(2), 1182–1192.

Ye, W., Yang, Y., Fang, X., Arif, M., Chen, X., & Yan, D. (2019). 2d cocrystallized metal–organic nanosheet array as an efficient and stable bifunctional electrocatalyst for overall water splitting. *ACS Sustainable Chemistry & Engineering*, *7*(21), 18085–18092.