UNIVERSITY OF NAIROBI

UNDERGRADUATE RESEARCH PROJECT

Modeling the impact of unemployment on insurance in Kenya

Authors:	*Registration nO.:*
Eric Cheruiyot	I07/81378/2017
Elvin Matovu	I07/102739/2017
Shaffy Achayo	I07/101064/2017
Abdulaziz Sharif	I07/104049/2017
Mohamed Abdirahman	I07/81360/2017

Supervisors:
Mrs. Wang'ombe & Prof.
Mwaniki

A project submitted in fulfillment of the requirements for the degree of Bachelor of Science Actuarial Science

in the

School of Mathematics

Declaration of Authorship

We declare that this project titled, "Modeling the impact of unemployment on insurance in Kenya" and the work presented in it are our own. We confirm that:

- This work was done wholly or mainly while in candidature for an undergraduate degree at this University.
- Where any part of this report has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where we have consulted the published work of others, this is always clearly attributed.
- Where we have quoted from the work of others, the source is always given. With the exception of such quotations, this project is entirely our own work.
- We have acknowledged all main sources of help.
- Where the report is based on work done by ourselves jointly with others, we have made clear exactly what was done by others and what we have contributed ourselves.

Eric Cheruiyot	19/07/21 Date
Elvin Matovu	Date
Abdulaziz Sharif	Date
Shaffy Achayo	Date
Mohamed Abdirahman	Date
Prof. J.I. Mwaniki	Date
Mrs. Wang'ombe	Date

Acknowledgements

We would like to thank our project supervisors Mrs. Wang'ombe and Prof J.I. Mwaniki for their guidance and support throughout the topic selection and project proposal writing. Finally, we extend our appreciation to our friends who gave us lots of useful suggestions on high-level applications. ...

For insurance companies, unemployment is important because of not only it's impact on the economic assumptions but also it's direct impact on insurance business.

A deep understanding of unemployment can help actuaries in many areas: Economic forecasts,Risk management,and Insurance assumptions.

This report introduces the concepts, theoretical background, and patterns of unemployment.

Some of the objectives of this report are:To fit a SARIMA model to kenya unemployment data.To predict the general market conditions such as unemployment using labor market theories and empirical studies.To identify the impact of unemployment in insurance products and policy formulation. ...

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background of the Study

Unemployment is a mismatch in the demand and supply in the labor market where there is oversupply.Unemployment is unpredictable,immeasurable and many theories have been put forward to explain it's causes and effects i.e nature.Some key theories are;

1. Keynesian theory states rapid wage decrease will reduce the aggregate demand for goods and services in the economy. It can cause a further reduction in investment and employment that becomes a downward cycle in the labor market.
2. Neoclassical economic theory which states the price(wage) and the quantity of labor is determined by the supply and demand of labor in the market.
3. Classical theory assumes over the long period the existence of full employment without inflation.
4. Philips curve which is the trade off between unemployment and inflation

Unemployment as a labor market measure directly and indirectly affect the insurance industry as part of the economy in whole.Different types of unemployment affect the economy in different ways;

1. Structural unemployment caused by the mismatch between workers' skills and employers' requirements
2. Frictional unemployment caused by the time needed to find a job
3. Cyclical unemployment caused by insufficient aggregate demand for goods and services

While setting the economic assumptions for insurance pricing, valuation and capital management to insurers, insurers have to set their economic outlook. Unemployment affects the demand for insurance products and the policy holder's behavior. There are some types of insurance companies that offer benefits if their policy holders are unemployed.

In Kenya the unemployment rate,which measures the number of people actively looking for a job as a percentage of the labor force,is at 4.9 % according to the Quarterly Labor Force Survey (QLFS) of Kenya Continuous Household Survey Programme (KCHSP) as at Q4 of 2020..

1.2 Problem statement

In Q4 2020, the national labor force in Kenya stood at 19,032,617 people from a working age population(15-64) of 27,129,695 which is around half of the total population in the Country.This was projected to increase to 25.6 million by 2021.Unemployment in Kenya is at a rate of 4.9%. This was according to the Quarterly Labor Force Survey (QLFS) of Kenya Continuous Household Survey Programme (KCHSP) . This is a clear indication that only a few people can acquire insurance covers and engage themselves in insurance products. Therefore our research will mainly focus on the how the high unemployment will affect the performance and growth of insurance industry in Kenya.

1.3 Objectives

Labor markets is an absolutely necessary component of economic system since it affects other economic systems and therefore the economic environment, it also determines employment income which affects consumption, policy holder behavior and Insurance assumptions such as sales volume, lapse rate, and premium persistence. Thus it is beneficial to analyze the impact of unemployment in Kenya.

The main objectives of this project are:

- To fit a SARIMA model to kenya unemployment data.
- To predict unemployment in kenya.
- To identify the impact of unemployment on insurance products and policy formulation.

1.4 Purpose of the Study

Our objective is to ascertain whether insurance companies can predict unemployment so as to factor its effects in decision making.

1.5 Scope of the study

Unemployment is the state whereby someone is actively searching for a job but cannot secure one in the current labour market conditions. Low-income earners are unable but may be willing to purchase insurance products. This significantly affects the insurance growth sale as well as those who are unemployed who take no cover neither purchase insurance products

1.6 Sources of Data

- Google trends data
- Kenya National Bureau of Statistics

Chapter 2

Literature Review

2.1 Theoretical review.

Unemployment is a small but key influence or indicator in Kenya's economic cycle volatility hence the target for risk mitigation .The unemployment rate is needed for a comprehensive understanding of the labor market conditions.In Kenya, the unemployment rate measures the number of people actively looking for a job as a percentage of the labor force.

The Quarterly Labor Force Survey release analyses the labor market situation for the population aged 15 to 64 years in Kenya.It's Q4 2020 release presented it's summary of the Key Labor Market Indicators in Kenya.Table 2.1 presents a summary of the Key Labor Market Indicators contained in this report.

TABLE 2.1: Summary of the Key Labor Market Indicators in Kenya Q4 2020

Indicator	**Quarter 4**
Population (15-64)	27,129,695
Labor Force	19,032,617
Employed	18,103,022
Employment/Population Ratio (%).	66.7
Unemployed.	929,595
Unemployment Rate (%)	4.9
Long-Term Unemployed.	415,136
Long-Term Unemployed (%)	2.2
Not in Labor Force (Inactive).	8,097,078
Labor Force Participation (%)	70.2
Labor Under Utilization	1,501,269
Labor Under Utilization (LU2)	7.9

Table 2.2 presents the labor force participation rates by age cohorts. The participation rate stood at 70.2% for Quarter 4. The age group 45 - 49 years recorded the highest rate at 90.5% while the age group 15 - 19 years recorded the lowest rate at 33.8%.

TABLE 2.2: : Unemployment by Age Cohorts [LU1]

Age	Unemployed	Labor force	Rate
15-19	73,933	1,784,269	4.1
20-24	352,708	2,491,336	14.2
25-29	224,775	2,992,627	7.5
30-34	129,376	3,020,017	4.3
35-39	70,108	2,302,653	3.0
40-44	25,861	1,988,423	1.3
45-49	12,538	1,615,254	0.8
50-54	24,812	1,165,623	2.1
55-59	15,484	963,513	1.6
60-64	-	708,902	0.0

The unemployment data shows in Kenya the youth make up a huge chunk of the labor force yet they are the most affected by unemployment.

The unemployment rate, as the official measure of unemployment, has faced many criticisms. It omits factors such as discouraged workers (not included in the labor force) and involuntary part-time workers (underemployment) that also reflect the magnitude of labor oversupply.However,high correlation between unemployment rate and various underemployment rates exists therefore even though the official unemployment rate does not encompass everything we want, it can serve as an indicator of changes in other components.A better indicator is the Labor Under-utilization (LU2), which is the combined rate of time-related underemployment and unemployment.For this report however we are going to base our models on Google trends data unemployment index which was found to be accurate in time series modeling (Francesco, 2009).Google trends data is especially accurate for modeling since most of those who use the internet in Kenya are the youth who we have seen are the most affected by unemployment here in Kenya.

A high unemployment rate can lead to an increase of social unrest, which was evident in 2008 financial crisis. This further weakens the market in general and adversely affects the insurance industry which is set to fair poorly in unstable political and socioeconomic conditions. The high rate of unemployment has affected the economic assumption used by insurance companies, it has also affected the worker's consumer income levels and has led to insurance policy lapses which have slowed the growth rate of insurance companies. However, most of the insurance companies in Kenya do not consider unemployment rates when setting the economic assumptions, and this has led to the slow growth rate of the insurance industry.

2.2 Critical Literature Review.

Globally several studies have been done to model unemployment data and it's implications on the economy as a whole and the insurance industry more specifically.The history of unemployment depends on the economic cycle. The cyclical portion of unemployment is more volatile and is a target for risk mitigation.

Askitas and Zimmermann, 2009 used the Germany unemployment data and found a high positive correlation between the unemployment rate and growth of insurance companies.The higher the unemployment rate the slower the growth of insurance companies.

Vicente, López-Menéndez, and Pérez, 2015 used Google Trends data in ARIMA models to forecast the Spanish unemployment rate, which is higher than U.S. unemployment rate. The explanatory power of the model was further enhanced by the Google Trends data. Considering that internet users are relatively younger than the population as a whole, web-based query information maybe more useful to study the rate of youth unemployment.

High unemployment among youth caused by youth bulges without growth in corresponding jobs was found to cause political instability through youth violence hence affecting the economy (Urdal, 2006) .

Pavlíček, 2014 in examining the relationship between job-related Google search query indices and unemployment rate in Visegrad countries found that the unemployment rate generally moves in the same direction as the search volume index for the job-related term also proving time series of Google search query indices useful for prediction-making. Models with Google series showed lower MAE and RMSE of static forecast compared to base models in all four countries.

2.3 Summary

From the analysis of literature review above, unemployment is real, and it is on the rise and a global problem.The most affected are young people in the population.Unemployment data is mostly scanty and does not encompass everything needed for modeling especially in Kenya.Here Google trends data is more useful as in the case of this report.

Globally unemployment and it's implications on the insurance industry and the economy has been modeled using Google trends and search index data.High unemployment rates correspond to slow growth or even recessions in an economy.The heightened crime,social and political strife often seen with high youth unemployment also cause a poor business environment. An increase in the rate of unemployment directly affects insurers through:

1. Lower new business volume
2. Higher lapse rates
3. Low interest rates

Thus it is important to model and study it's implications since it contributes to fall of insurance companies and their low penetration in Kenya.

Chapter 3

Methodology

3.1 SARIMA Modeling

For this model we used Google trends data as our unemployment indicator.The keyword we chose was job which was the most searched item in the labor category and related terms i.e jobs, jobs in Kenya e.t.c .The index starts at the beginning of 2004.GT is a search trends feature that shows how frequently a given search term is entered into Google's search engine, relative to the site's total search volume over a given period of time.

Mathematically, being n(q, l, t) the number of searches for the query q, in the location l during the period t, the relative popularity (RP) of the query is expressed as:

$$RP_{(q,l,t)} = \frac{n(q,l,t)}{\Sigma_{q \in Q(l,t)} n(q,l,t)} \times \Pi_{\left(n(q,l,t) > \tau\right)}, \tag{3.1}$$

where Q(l, t) is the set of all the queries made from l during t and $\Pi_{\left(n(q,l,t)>\tau\right)}$ is a dummy variable whose value is 1 when the query is sufficiently popular (the absolute number of search queries n(q, l, t) exceeds τ) and 0 otherwise. The resulting numbers are then scaled on a range of 0-100 depending on the proportion of a topic with respect to the total number of all the search topics.So, the index of GT is defined as:

$$IGT_{(q,l,t)} = \frac{RP(q,l,t)}{\max\{RP(q,l,t)_{t \in 1,2,\ldots,T}\}} \times 100. \tag{3.2}$$

For forecasting unemployment the time series data was fitted to a SARIMA model of the nature:

$$y_t = a_0 + a_1 y_t - 1 + a_2 y_t - 2 + ..a_p y_t - p - b_1 u_t - 1 - b_2 u_t - 2 - .. - b_q u_t - q + u_t \tag{3.3}$$

yt is unemployment rate at time t

$a_0, a_1, ..a_p and b_1 ......b_q$ are parametric coefficients

p is the auto regressive parameters

q is the moving average parameter

u_t is the error term
The time series data was then plotted.

The raw Google index data was then decomposed into it's various trend and seasonal components and checked for trend.

The series was differenced to reduce trend and seasonality.

The Augmented Dickey-Fuller test was used to test for stationarity.

Correlation tests were carried out to identify any statistical relationships among the above variables. The ACF plot was useful in determining stationarity.The ACF and PACF plots were used to decide whether to include an AR term(s), MA term(s), or both.

The SARIMA model selected was by use of the "auto.ARIMA" function in R. The forecast was then done on the SARIMA model using the forecast package and the forecast function in R to produce a 2 year prediction.

The Ljung-Box test was used to check accuracy and suitability of our model and forecast and a plot of the residuals obtained.

Chapter 4

Results & Discussion

The Google Trends job index was monthly data starting from 2004 through to 2021 March.We converted it to time series data and plotted the graph.

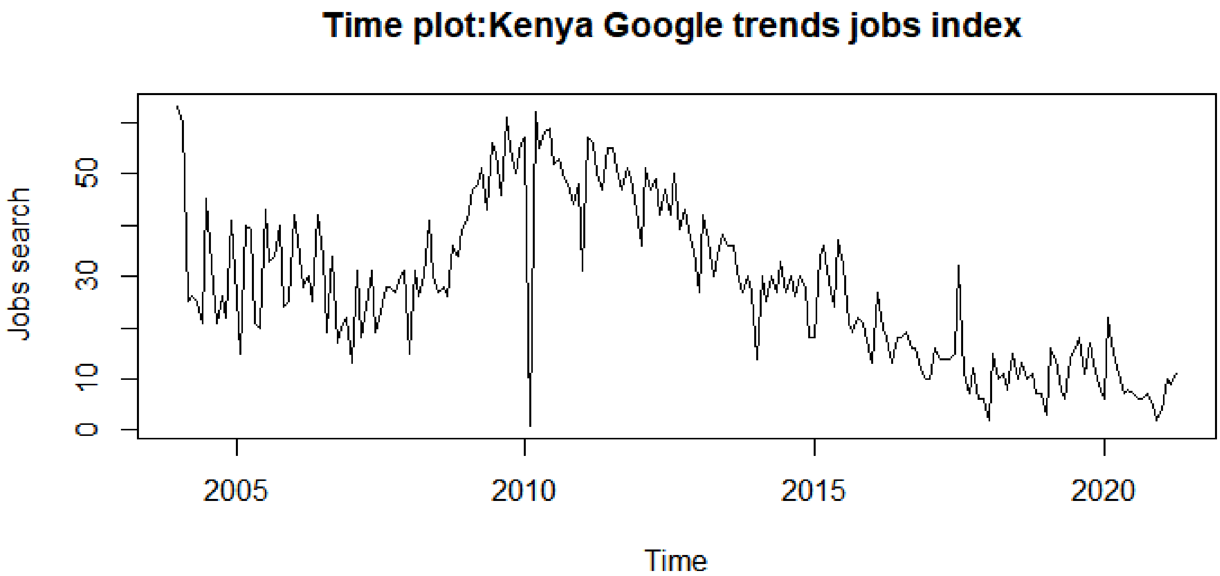

FIGURE 4.1: Time series plot of Kenya's Google Trends job index.

The data has strong trend and seasonality.

From the plot we see unemployment spiked right around the 2008 financial crisis and the 2020 pandemic.These are all times when the economy is under a lot of stress.

We decomposed the data into it's trend and seasonal component.

The data was differenced once to remove trend and seasonality.This would be crucial in the accuracy of our model forecast.

The resulting differenced time series was tested for stationarity using the ADF test.

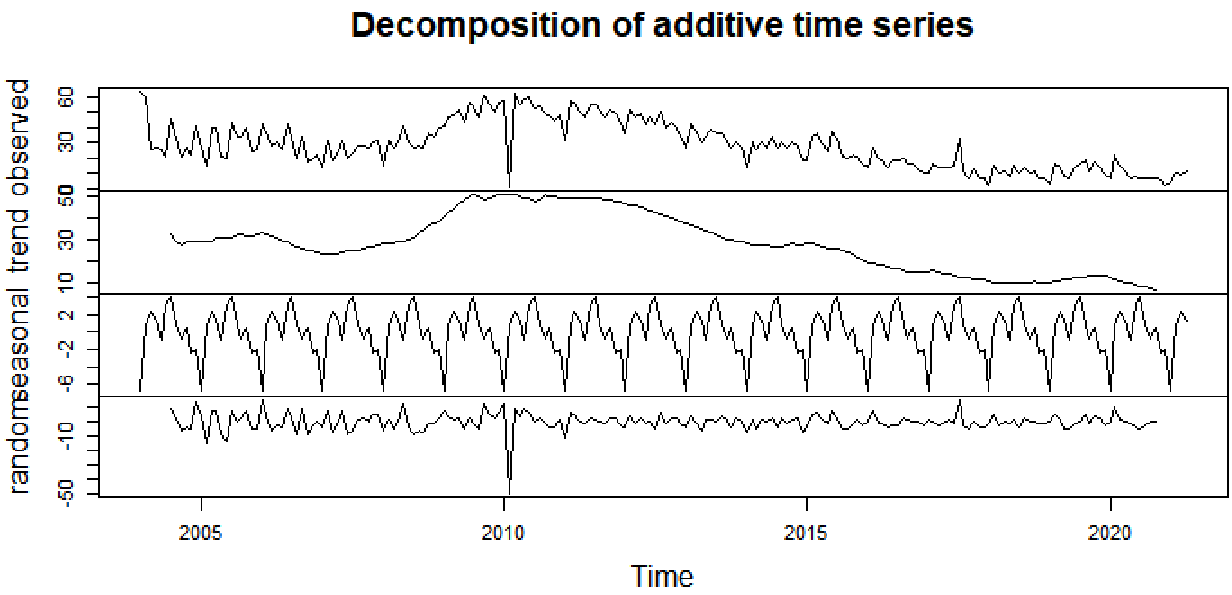

FIGURE 4.2: Decomposition plot of Kenya's Google Trends job index.

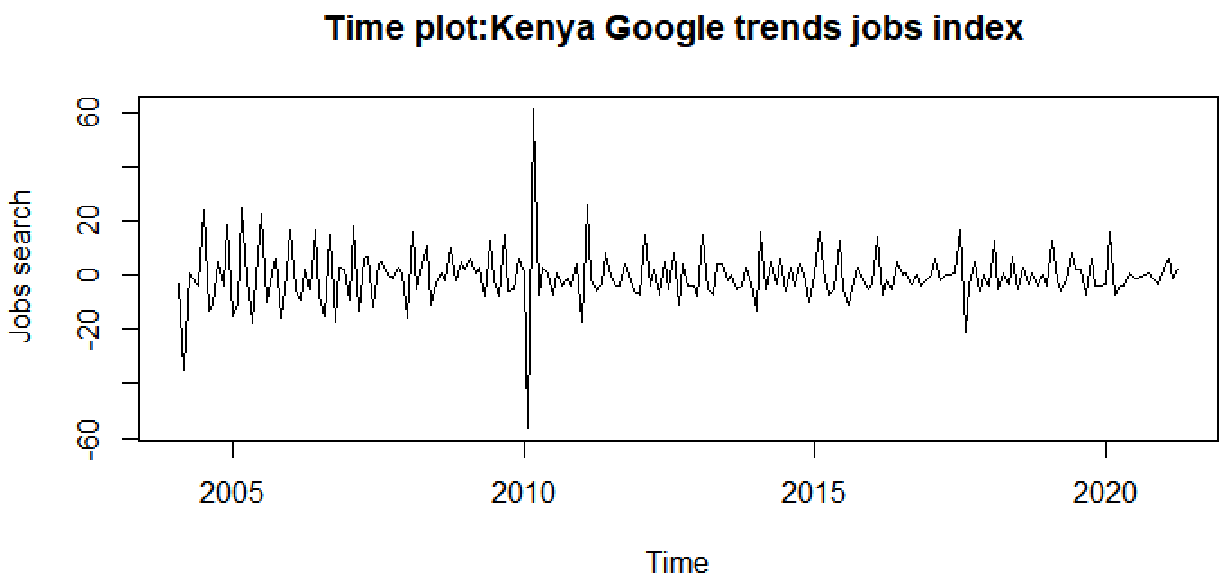

FIGURE 4.3: Differenced series plot of Kenya's Google Trends job index.

Augmented Dickey-Fuller Test

data: DY
Dickey-Fuller = -7.966, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary

The stationary series now had it's auto correlation plots generated as shown below:

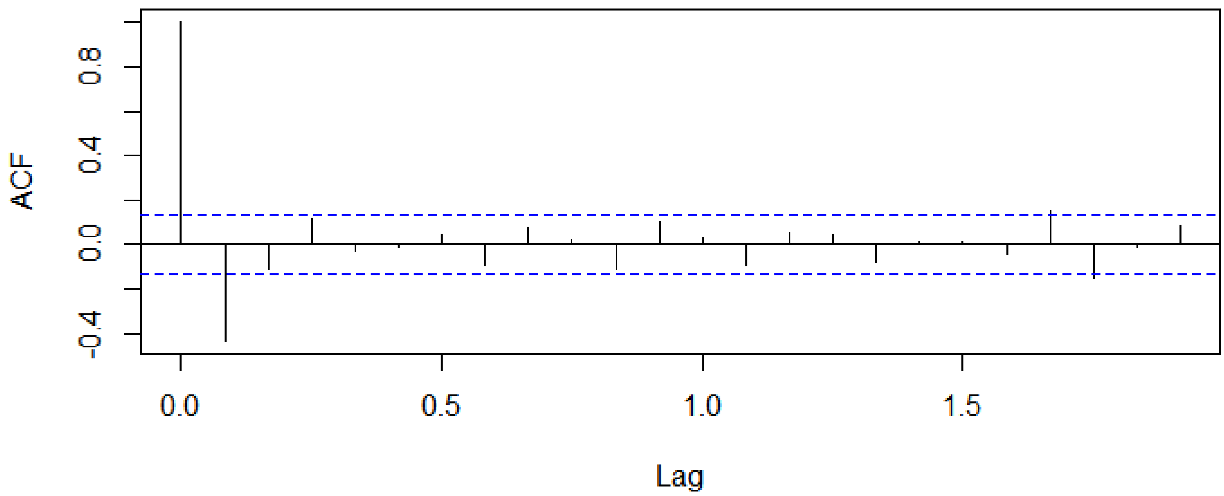

FIGURE 4.4: ACF plot .

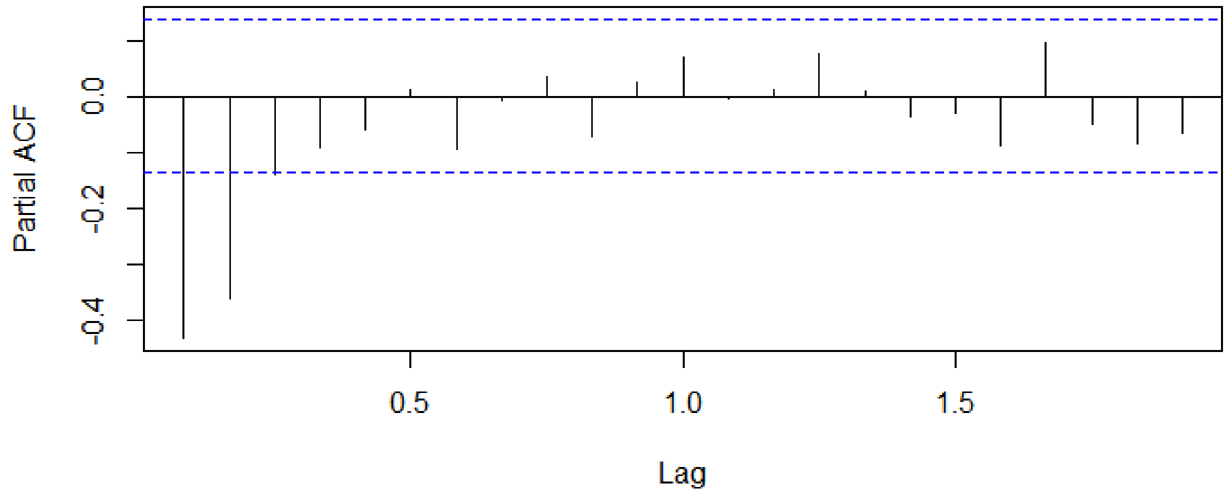

FIGURE 4.5: PACF plot .

Our stationary time series was fitted to a SARIMA(0,1,3)(1,1,1)[12] model using the R function auto.ARIMA and forecasted ahead 24 months using the R package forecast and plotted.

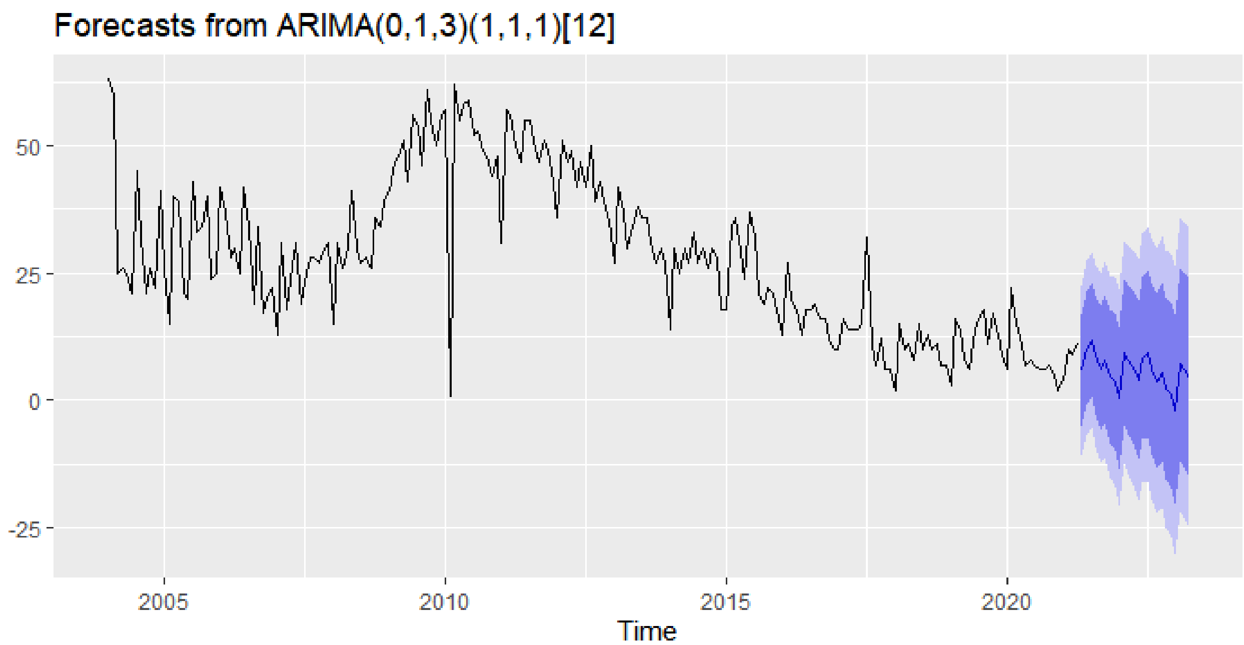

FIGURE 4.6: Forecasted Time series plot of Kenya's Google Trends job index .

The model forecast shows a decrease in unemployment.This can be explained partly by the slow recovery of the global economy from the pandemic in 2020.This can also be attributed to great stock market and economy performance since the start of 2021.By April 2023 our model predicts unemployment to sit at 4.39% down from 7.9% in 2020 Q4.

After fitting the model, we should check whether the model is appropriate. To check the accuracy of the model residuals from the SARIMA (0,1,3)(1,1,1)[12] were put through the Ljung-Box test and the residuals plotted using the "checkresiduals" command in R.

Ljung-Box test

data: Residuals from ARIMA(0,1,3)(1,1,1)[12]
Q* = 30.567, df = 19, p-value = 0.04501

Model df: 5. Total lags used: 24

The Box-Ljung test shows that the first 24 lag auto correlations among the residuals are zero (p-value = 0.04501), indicating that the residuals are random and that the model provides an adequate fit to the data.

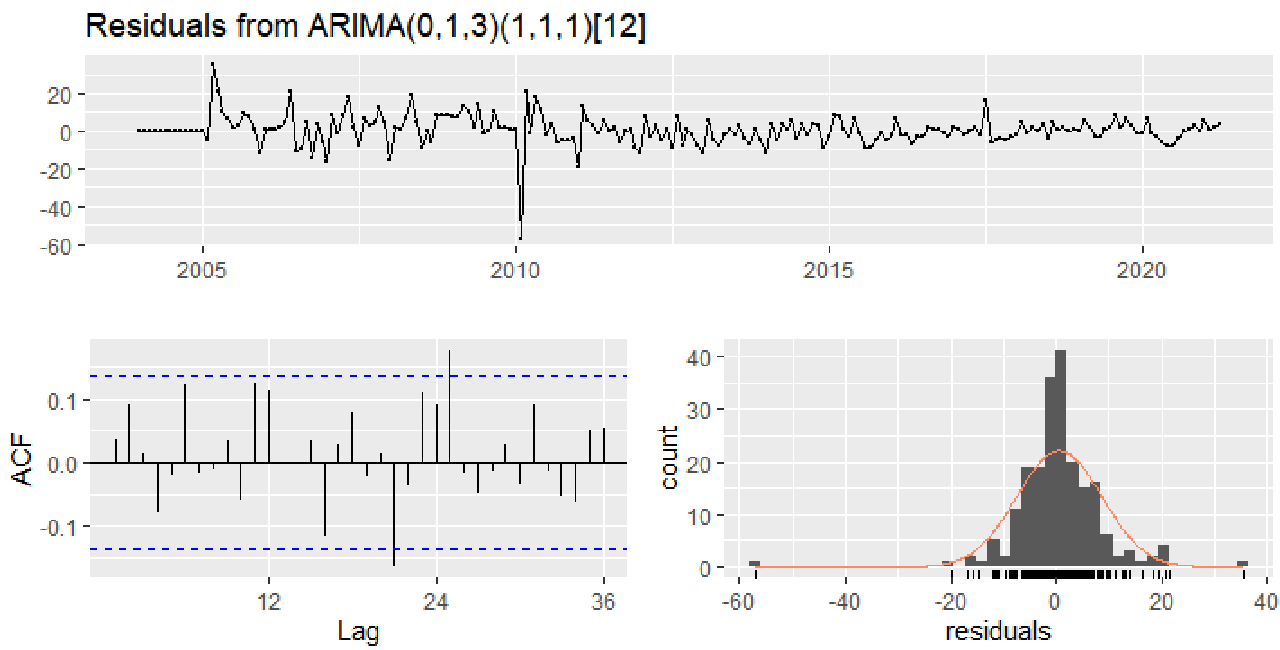

FIGURE 4.7: Residuals plot .

The autocorrelation plot of the residuals shows that for the first 24 lags, all sample autocorrelations except that at lag 20 fall inside the 95 % confidence bounds indicating the residuals appear to be random. The residuals plot shows that the variance of the forecast errors is roughly constant over time. The histogram of the time series shows that the forecast errors are roughly normally distributed and the mean seems to be close to zero. Our chosen model is therefore suitable and accurate for forecasting unemployment data.

Chapter 5

Conclusion

Kenyan demographic trends have resulted in a population structure that is eminently young. Long-term population changes and long-lasting economic stagnation in turn resulted in a large informal sector and strong pressures on the labor market, which took the form of a large youth unemployment challenge. High youth unemployment is a problem affecting many developed and developing economies, but in Kenya as well as in many other developing countries youth unemployment is particularly severe. Relatively few of the young people joining the labor market can find a job in the formal sector, and many cannot readily find an adequate occupation in the informal sector. This study has documented that Kenya's youth unemployment rates are the highest of all age groups and that young people represent, by far, the bulk of Kenya's unemployed people.

Models have shown high youth unemployment can be a cause of instability (Urdal, 2004) Our model predicts a drop in the unemployment rate in the next 48 months ending June 2023.The low unemployment signals business growth and economic expansion.For insurance companies tracking unemployment as an economic indicator decline in the rate of unemployment should mean better business.While youth bulges may be a current concern of the Kenyan national and county governments, the relative risk of violence importance is expected to fade as the unemployment rate is expected to drop.

The internet contains an enormous amount of information which, to our knowledge, classical econometrics has yet to appropriately tap into. Such information comes timely on a continual basis. It is particularly welcome at times of an economic crisis where the traditional flow of information is too slow to provide a proper basis for sound economic decisions. Google job index data is useful in modelling implications of unemployment and underemployment hence more insurers should build models around this data type.

Appendix A

Concepts and Definitions

Labour force: consists of all persons in the working age population who are either employed or unemployed. It was previously also referred to as the 'currently active population'.

Employment: Refers to performance of work as defined above. This term is used to measure the number of persons employed, including persons at work during a short reference period, and also persons temporarily absent from work but holding a job.

Unemployment: Under the strict terms is defined as people who do not have a job, have actively looked for work in the past four weeks, and are currently available for work.

The Working-age Population (WAP): includes all persons in the population above a specified age threshold used for statistical purposes to define the economically active population.

Labour underutilization: Refers to mismatches between labour supply and demand, which translate into an unmet need for employment among the population. Labour Under Utilization [LU2] is computed as the combined rate of time-related underemployment and unemployment.

Bibliography

Askitas, Nikos and Klaus F. Zimmermann (2009). "Google econometrics and unemployment forecasting". In: URL: https://www.econstor.eu/bitstream/10419/35733/1/605353115.pdf.

Francesco, D'Amuri (2009). "Predicting unemployment in short samples with internet job search query data". In: URL: https://mpra.ub.uni-muenchen.de/18403/1/MPRA_paper_18403.pdf.

Pavlíček, Jaroslav (2014). "Google Econometrics: Unemployment in Visegrad Countries". In: URL: https://dspace.cuni.cz/bitstream/handle/20.500.11956/68309/BPTX_2012_2_11230_0_356342_0_137288.pdf?sequence=1.

Urdal, Henrik (2004). "The devil in the demographics: the effect of youth bulges on domestic armed conflict, 1950-2000". In: *Social Development Papers* 14.

– (2006). "A clash of generations? Youth bulges and political violence". In: *International studies quarterly* 50.3, pp. 607–629. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.491.4794&rep=rep1&type=pdf.

Vicente, María Rosalía, Ana J. López-Menéndez, and Rigoberto Pérez (2015). "Forecasting unemployment with internet search data: Does it help to improve predictions when job destruction is skyrocketing?" In: *Technological Forecasting and Social Change* 92, pp. 132–139. URL: https://www.sciencedirect.com/science/article/pii/S0040162514003904.