Modelling economic determinants of youth unemployment using multiple linear regression: A case study of Kenya

Fred Nyamitago Monari, Richard Tinega and Lameck Ondieki Agasa

Abstract
Despite many efforts taken by the government and development partners youth unemployment has remained to be a big problem not only in Kenya but also in the world (Shem, 2013). This study therefore analyzes the economic determinants of youth unemployment in Kenya from 2000 to 2017 by investigating empirical relationship among youth unemployment, gross domestic product, population and foreign direct investment. The main objective of the study was to come up with a model that can be used to predict the future youth unemployment level. The variables used in the study were population growth, gross domestic product and foreign direct investments. It is hypothesized that these factors have an effect on youth unemployment rate in Kenya’s economy. E-views 8 statistical analysis software and SPSS was used to analyze the data. Both descriptive and inferential statistics were employed in this study. The study used multiple linear regression model to test the economic determinants of youth unemployment. At 5% significance level, empirical results indicate that population, gross domestic product and foreign direct investment are significant economic determinants of youth unemployment in Kenya. The study revealed that a unit change in GDP while holding the other factors constant would lead to an increase in youth unemployment by a factor 0.027, a unit change in FDI while holding the other factors constant would lead to an increase in youth unemployment by a factor 0.034. On the other hand a unit change in population growth will lead to a change in youth unemployment by a factor -1.543. The study recommended that both county and national governments should consider policies that encourage foreign direct investment, increasing GDP through value addition, and using external debt resources for investment.

Keywords: economic determinants, modelling, youth unemployment

Introduction
Unemployment is a state in which people are willing and able to work at the prevalent wage rate, but they are unable to find jobs, (Economic Survey of Pakistan 2010-11). Unemployment is a key measure of economic health; a major factor in determining how healthy an economy is; if the economy maximizes efficiency everyone would be employed at some wage. According to Kenya Population Census, (2009) Youth accounts for 35% of Kenya’s population as well as it represent 67% of unemployment. Population growth is 4% for the youth whereas the national growth rate is 2.9% which implies that population growth is higher for the youth in comparison with that of the national. Social evils and Political instability have been linked to youth unemployment by studies. The unemployment generates a culture of idleness and group-up tendencies. According to Okojie, (2003) the “youth bulge” is a source of concern for governments and development partners. Unemployment drives youth to be involved in unlawful deeds like touting, stealing, armed robbery, drug dealers, and prostitution (Okojie, 2003) According to Australia National Health Survey, 1989-90; Morrell, Taylor and Kerr, 1998 young unemployed people are more prone to crime as a result of depression and psychological problems. Apart from social evils African governments have got another concern that’s political stability. The risk of all three forms of political violence which includes rioting, civil war, and terrorism is increased by youth bulges (Urdal 2006). According to ADEA, 2014 unemployed youth bulge becomes political pond by politicians and extremist groups for scheming activities.
The Kenyan political history is laced with political violence and intolerance. Kenya experienced continued terrorist attacks amongst other internal organized crimes involving youth as perpetrators. The Kenyan youth are being portrayed as vulnerable, and unable to contribute to national unity and development. These social problems are linked to economic condition of the youth which makes them to be non-productive.

Problem Statement
At this era, in Kenya, there are limited studies on economic determinants of youth unemployment, which exists in mathematical modelling of economic determinants of youth unemployment in order to understand the relationship between each determinant and youth unemployment.

Research design
A research design is an extensive arrangement that includes highlighting every one of the strategies that will be used in the accumulation and examination of information Leedy and Ormrod (2001). This research adopted both descriptive and explanatory research designs. Explanatory research is conducted in order to identify the extent and nature of cause and effect relationships. Descriptive research design was also employed in this study to help the researcher describe the trend of the variables under observation. The study used data for the period 1997 to 2017 for the components of youth unemployment namely; GDP and population. Data on Foreign Direct Investment was also sourced since it will help explain about the unemployment. The linear regression model was used for estimation after undergoing time series property tests.

Sources of Data
Secondary data was used whereby data on GDP growth rate, population, and Foreign Direct Investment for the years 2000-2017 was sourced from World Bank database. All the data obtained was measured in calendar years.

Data Analysis
E-views 8 statistical analysis software and SPSS was used to analyze the data. Both descriptive and inferential statistics were employed in this study. The study used multiple linear regression formula in showing the relationship between gross domestic product (GDP), foreign direct investments (FDI), population and unemployment in Kenya.

Unit root tests
The first step involved determining if the series had a unit root. According to Brooks (2008) [15], this procedure is necessary for purposes of ensuring existence of a constant meaningful. For a stationary series, there is no unit root and root. According to Brooks (2008) [15], explains that developing dynamic economic models requires a detailed examination of characteristics of time series data involved. Should these features be ignored and the set of time series data modelled jointly, resulting regression output may indicate a high degree of correlation among the variables. Therefore a lot of care was taken before running the regression estimation.

To carry out the estimation the following formula or rather the model was used.

The equation; $Y= \beta_0+ \beta_1X_1+\beta_2X_2+ \beta_3X_3+ \epsilon$

Where: $Y$ = unemployment in Kenya
$\beta_0$, is constants to be estimated by the model
$\beta_1$, $\beta_2$ and $\beta_3$ are coefficients
$X_1$ = GDP growth rate (will measured in duration of financial instruments in respect to the sensitivity of their values to interest rate changes in KES, quarterly)
$X_2$ = foreign direct investments (FDI)
$X_3$= population
$\epsilon$= Error terms

Test of Significance
According to Kothari, (2004) [24] results are said to be measurable significant inside the 0.05 level, which implies that the noteworthiness esteem must be littler than 0.05. The importance will be controlled by the t-value, which shows what number of standard blunder implies the example wanders from the tried esteem. In addition, the Pearson Product Moment Correlation Coefficient was used to test the direction and magnitude of the relationship between the dependent and independent variables at 95% confidence level. The model significance was tested using the ANOVA, t-tests, F-tests, z-tests, and the chi-square at confidence of 95%.

Table 1: Descriptive statistics

<table>
<thead>
<tr>
<th></th>
<th>FDI</th>
<th>GDP</th>
<th>Population</th>
<th>Youth Unemployment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.975537</td>
<td>4.584360</td>
<td>2.694384</td>
<td>23.86222</td>
</tr>
<tr>
<td>Maximum</td>
<td>3.457220</td>
<td>8.405699</td>
<td>2.752069</td>
<td>26.25100</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.040833</td>
<td>0.232828</td>
<td>2.523111</td>
<td>20.15800</td>
</tr>
<tr>
<td>Jarque-Bera</td>
<td>4.660404</td>
<td>1.443351</td>
<td>5.800490</td>
<td>2.301328</td>
</tr>
<tr>
<td>Probability</td>
<td>0.097002</td>
<td>0.0485937</td>
<td>0.055010</td>
<td>0.316427</td>
</tr>
</tbody>
</table>

Average GDP was 4.58% whereas the minimum was 0.23%. Maximum GDP was 8.41%. According to Jarque-Bera probability value (0.048), GDP data is not normally distributed because the P value is less than 5%. Average FDI was 0.97% while the minimum was 0.040833% and the maximum was 3.457220%. FDI data was normally distributed according to Jarque-Bera probability value (0.097) greater.
than 5%. Average population was 2.69% while the minimum was 2.52% and the maximum was 2.75. Population data was normally distributed according to Jarque-Bera probability value (0.055) which is greater than 5%. Average of youth unemployment rate on the other hand was 23.86% while minimum was 20.16% and maximum was 26.25. Youth unemployment rate data was normally distributed since the Jarque-Bera probability value (0.32) was greater than 5%. Therefore the results clearly show that the rate of youth unemployment has always remained high.

Unit Root Test
It is important to determine the order of integration of time series data before building econometric models. This ensures that the models resulting from these variables are statistically significant and can therefore yield valid predictions. Time series data can either be stationary or non-stationary. Modelling of non-stationary time series data results to spurious regressions which means, no meaningful predictions can be made from such models. The first step therefore was to test stationarity of the time series data using Augmented Dickey Fuller Test. This test relies on rejecting a null hypothesis which signifies existence of a unit root. The aim of unit root test is to determine whether the time series data is stationary or not. It is also important to perform unit root test especially for non-stationary series because the order of integration is easily determined. For a non-stationary series, differencing is required to make it stationary. When a non-stationarity series is modeled, the resulting regression model is termed as spurious which means no meaningful prediction can be made based on the model. The tables below summarize the results of stationarity tests.

<table>
<thead>
<tr>
<th>Variable</th>
<th>ADF Test Statistic</th>
<th>MacKinnon P-value</th>
<th>Comment</th>
<th>Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log of FDI</td>
<td>2.6757633</td>
<td>0.09841</td>
<td>Non-Stationary</td>
<td>I(1)</td>
</tr>
<tr>
<td>Log of GDP</td>
<td>4.063371</td>
<td>0.00705</td>
<td>stationary</td>
<td>I(0)</td>
</tr>
<tr>
<td>Log of population</td>
<td>1.034152</td>
<td>0.70971</td>
<td>Non-stationary</td>
<td>I(1)</td>
</tr>
<tr>
<td>Log of unemployment</td>
<td>1.37401316</td>
<td>0.56979</td>
<td>Non-Stationary</td>
<td>I(1)</td>
</tr>
</tbody>
</table>

Stationarity was tested at 5% level of significance.

<table>
<thead>
<tr>
<th>Variable</th>
<th>ADF Test Statistic</th>
<th>MacKinnon P-value</th>
<th>Comment</th>
<th>Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log of FDI</td>
<td>4.2030749</td>
<td>0.00701</td>
<td>Stationary</td>
<td>I(0)</td>
</tr>
<tr>
<td>Log of population</td>
<td>0.4494619</td>
<td>0.87455</td>
<td>Non-stationary</td>
<td>I(1)</td>
</tr>
<tr>
<td>Log of unemployment</td>
<td>3.680871</td>
<td>0.01582</td>
<td>Stationary</td>
<td>I(0)</td>
</tr>
</tbody>
</table>

Stationarity was tested at 5% level of significance.

Test of stationarity was first done on the variables at level as shown in table 4.1. Findings revealed that GDP was the only stationary variable. The order of integration was zero which is indicated as I(0). Other variables population, unemployment and FDI were non-stationary since their P values exceeded 5% level of significance. Table 4.2b indicates that the rate of unemployment and FDI were the only stationary variables at first difference. Their P values were less than 5%. Unemployment and FDI were therefore integrated of order one since the series had to be differenced once in order to be stationary. The series is therefore a combination of I(0) and I(1) variables. Table 4.2c indicates that the population variable was stationary at second difference. The p value was less than 5%. Population was therefore integrated of order two since the series had to be differenced twice in order to be stationary.

<table>
<thead>
<tr>
<th>Variable</th>
<th>ADF Test Statistic</th>
<th>MacKinnon P-value</th>
<th>Comment</th>
<th>Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log of population</td>
<td>3.82283</td>
<td>0.01385</td>
<td>Stationary</td>
<td>I(0)</td>
</tr>
</tbody>
</table>

Stationarity was tested at 5% level of significance.

To evaluate the model fit the study applied coefficient of determination. An average adjusted coefficient of determination (R2) of 0.592 or 59.2% of the variations in youth unemployment of the country are explained using the independent variables understudy (population growth, foreign direct investments (FDI) and gross domestic product (GDP).

Analysis of Variance
The purpose of ANOVA results is to indicate the statistical significance of the regression model at 95% confidence interval or rather at 5% level of significance. The table below shows the findings.

<table>
<thead>
<tr>
<th>Model</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>.106</td>
<td>3</td>
<td>.035</td>
<td>6.773</td>
<td>.005</td>
</tr>
<tr>
<td>Residual</td>
<td>.073</td>
<td>14</td>
<td>.005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>.180</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. Dependent Variable: lunp
b. Predictors: (Constant), lpop, lgdp, lfdi

From the ANOVA statistics, the study found out the regression model had a significance level of 0.5% which is a sign that the data was perfect for drawing a conclusion on the population parameters as the value of significance (p-value) was less than 5%.

And that the model can be used to make predictions since the model is statistically significant.
Conclusions and recommendations

This study has associated youth unemployment with economic determinants derived from macroeconomic data. As such empirical findings reveal a long run relationship between youth unemployment rate, GDP, Foreign Direct Investment, and Population growth. The parameters indicate that population has a negative effect on youth unemployment rate whereas foreign direct investment and GDP has a positive effect on the youth unemployment.

The finding that increase in GDP and FDI leads to increase in youth unemployment rate could be attributed to many factors. These findings conquer with the findings of Shem (2013). On such factor is the Kenya majorly exports unfinished products or raw materials. In this way, jobs associated with value addition are inadvertently exported. The strange finding of relationship between GDP and youth unemployment rate is also consistent with Ajilore and Yinusa (2011) study in Botswana whose findings revealed a “jobless growth” of economy.

The unique findings that increase in population growth rate leads to a decrease in youth unemployment differs with the findings of Maqbool et al (2013) and Valadkhani (2003) which revealed that population has a positive relationship with youth unemployment while foreign direct investment has a negative relationship with youth unemployment.

Regression Model Coefficients

Table 7: Regression coefficients

<table>
<thead>
<tr>
<th>Model</th>
<th>Coefficients*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unstandardized Coefficients</td>
</tr>
<tr>
<td></td>
<td>B</td>
</tr>
<tr>
<td>1</td>
<td>(Constant) 4.683</td>
</tr>
<tr>
<td></td>
<td>lgp</td>
</tr>
<tr>
<td></td>
<td>lfdi</td>
</tr>
<tr>
<td></td>
<td>lpop</td>
</tr>
</tbody>
</table>

As per the SPSS generated output as presented in table above, the equation

\[ Y = 4.683 + 0.027X_1 + 0.034X_2 - 1.543X_3 + e \]

The error of estimate is; \( e = 0.0723 \)

From the regression model obtained above, a unit change in GDP while holding the other factors constant would lead to an increase in youth unemployment by a factor 0.027; a unit change in FDI while holding the other factors constant would lead to an increase in youth unemployment by a factor 0.034. On the other hand a unit change in population growth will lead to a change in youth unemployment by a factor -1.543.

References

3. Addison JT, Ozturk OD. Minimum wages, labor market institutions, and female employment and unemployment: A cross-country analysis, 2010.
16. Cheema AR, Atta A. Economic determinants of
35. Singh I, Chiloane-Tsoka E. Youth Unemployment and effects of education as global entrepreneurial determinants: Case of South Africa. The Business & Management Review. 2015; 6(2),