University of Nairobi

School of Mathematics

ISSN: 2410-1397

Master Project in Social Statistics

Modelling the Determinants of Underemployment among Youths in Kenya

Research Report in Social Statistics,Number 31, 2021

Francis Gitau Wanjiru

July 2021

Submitted to the School of Mathematics in partial fulfilment for a degree in Master of Science in Social Statistics

Master Project in Social Statistics **University of Nairobi**
July 2021

Modelling the Determinants of Underemployment among Youths in Kenya

Research Report in Social Statistics,Number 31, 2021

Francis Gitau Wanjiru

School of Mathematics
College of Biological and Physical sciences
Chiromo, off Riverside Drive
30197-00100 Nairobi, Kenya

Master Thesis
Submitted to the School of Mathematics in partial fulfilment for a degree in Master of Science in Social Statistics

Submitted to: The Graduate School, University of Nairobi, Kenya

Abstract

Unemployment is mainly used as the indicator for labour underutilization. Underemployment is higher than the unemployment rate in 65% of the African countries (ILO, n.d.). Kenya's unemployment rate is 7.4%, underemployment rate- 20.4% (KNBS, 2018). The use of the unemployment aspect to measure the unmet need for employment fails to provide a comprehensive picture of the labour market. There is a need to complement unemployment with underemployment, thus providing a full view of labour underutilization. Understanding the determinants of underemployment by measuring remuneration and time worked as key quality aspects is important to provide crucial information about the state of the labour market indicator for improved analysis. The main objective of this study is to model the determinants of underemployment among youths in Kenya. The study focused on the visible and invisible forms of underemployment among the youths aged 15-34 years. The study utilizes the secondary cross-sectional data obtained from the Kenya Integrated Household and Budget Survey (KIHBS) 2015/16. The response variables for this study are visible underemployment and invisible underemployment. The explanatory variables in this study include gender, education level, age, employment sector, residence and marital status. The binary logistic regression model is used to analyze the data in this study. The study findings reveal 11.9% of youths in Kenya are visibly underemployed, while 75% of the youths in Kenya are invisibly underemployed. The model findings show that gender, age categories for 30-34 and 25-29 years, education secondary category, private formal sector, informal sector and residence are significant determinants of visible underemployment among youths in Kenya. The results also show that gender, age, education categories (post-secondary, college and post-primary vocational), residence, employment sector and marital status (never married category) were significant determinants of invisible underemployment among youths in Kenya. The findings inform the need for policy interventions focusing on stimulating the growth of the formal sector, promoting gender equality at work, promoting education and skills enhancements and government enforcement of compliance to minimum wage policy to help address the underemployment problem.

Master Thesis in Mathematics at the University of Nairobi, Kenya.
ISSN 2410-1397: Research Report in Mathematics

DISTRIBUTOR: School of Mathematics, University of Nairobi, Kenya

Declaration and Approval

I the undersigned declare that this dissertation is my original work and to the best of my knowledge, it has not been submitted in support of an award of a degree in any other university or institution of learning.

	24/08/2021
Signature	Date

FRANCIS GITAU WANJIRU
Reg No. I56/31072/2019

In my capacity as a supervisor of the candidate's dissertation, I certify that this dissertation has my approval for submission.

	24/08/2021
Signature	Date

Dr. John Ndiritu
School of Mathematics,
University of Nairobi,
Box 30197, 00100 Nairobi, Kenya.
E-mail: jndiritu@uonbi.ac.ke

Dedication

This project is dedicated to my mum,family and friends for their continued support.

Contents

Figures and Tables

Figures

Tables

Nomenclature

ICLS	International Conference of Labour Statisticians
ILO	International Labour Organization
KIHBS	Kenya Integrated Household and Budget Survey
KIPPRA	Kenya Institute for Public Policy Research and Analysis
KNBS	Kenya National Bureau of Statistics
KPHC	Kenya Population and Housing Census
MLE	Maximum Likelihood Estimation
NASSEP	National Sample Survey and Evaluation Program
O.R	Odds Ratio
SDG	Sustainable Development Goals
VIF	Variance Inflation Factor

Acknowledgments

First, I thank God for granting me the grace to finish my project. I pass my sincere gratitude to my supervisor Dr. John Ndiritu for his guidance and dedication towards my timely completion of the project. I also wish to thank all the lecturers at the School of Mathematics. I thank my family and friends for their support. ....

Francis Gitau Wanjiru

Nairobi, 2021.

1 Chapter 1: Introduction

The aim of this thesis is to model the determinants of visible and invisible underemployment among youths in Kenya.

The thesis outline is as follows: chapter one includes the introduction capturing sections such as the background of the study, statement of the problem, objectives and significance of the study. The second chapter captures the literature review, which encompasses the theoretical review, empirical literature and Kenya specific review. The third chapter in the thesis contains the methodology capturing data sources, data variables and data validation, theoretical model/empirical model, parameter estimation and goodness of fit tests. The fourth chapter captures data analysis and results. The fifth chapter captures the conclusion, recommendations and areas of further research.

1.1 Background of the Study

Under-employment is defined as any sort of employment that is unsatisfactory in terms of aspects such as insufficient hours, insufficient use of one's skills and compensation (ILO, n.d.). There are two categories of under-employment, including the visible and invisible under-employment. Visible underemployment is associated with insufficient hours of work. Invisible underemployment refers to the situation where individuals who are working in jobs where labour resources are not adequately utilized. Invisible underemployment leads to outcomes such as low productivity, skills underutilization and low income. Under-employment involves a comparison of the current employment situation of an individual with an alternative employment situation that one is willing to carry out. Time-related underemployment has been used as a framework for measuring underemployment, considering individuals who usually work less than the selected threshold, which is 40 hours in the reference week (ILO, n.d.). The willingness of the individuals to work additional hours during the reference week is an aspect considered in the measurement of time-related underemployment. The availability of the individuals to work extra hours when provided with an opportunity for additional work is considered. The definition of underemployment is done using the criteria that are analogous to employment and unemployment definitions. Labour underutilization refers to the mismatch between demand and supply in the market resulting in failure to meet the employment need in the population. Underemployment is a global problem affecting the employed persons facing labour underutilization. According to ILO (2018), the rate of underemployment was higher than the unemployment rate in approximately 40 of the 114 countries. Such shows how inadequate jobs is a real problem in labour underutilization. Underemployment as a measure of labour underutilization is a significant problem in Africa. The rate of underemployment is higher than that of unemployment in 65% of the African countries. The case is different for European countries at 14% and countries in America at 43%. Most African countries are developing, and the economic aspects influence widespread underemployment compared to unemployment. Individuals are compelled to take undesired work in the absence of savings and sufficient unemployment benefits in developing countries, thus leading to higher underemployment rates. The developing countries, which are characterized by informality and segmented labour markets, also explain the high underemployment rates that surpass unemployment. The data shows that underemployment is a predominant form of labour underutilization which calls for more focus on this aspect. According to International Labor Organization (2021), the unemployment rate in the Kenyan population was 2.98%, which was a rise from 2.59%in 2019. The youth unemployment rate had risen from 7.17% in 2018 to 7.19% in 2019 (ILO, 2021). In Kenya,about 20.4% of the total population is underemployed based on 2015/16 Kenya Integrated Household and Budget Survey (KIHBS) data, which had risen five times compared to the findings from the 1998/99 KIHBS survey (Vuluku et al., 2013). Despite the levels of unemployment among youths in the country, which stood at 7.19% in 2019, those in employment work in low productive jobs thus underemployed. Such make it cumbersome to utilize their skills and meeting their needs (Kippra, 2009). Findings from the Economic Survey of 2020 (KNBS, 2020) reveals that approximately 83% of the jobs in Kenya are in the informal sector. Most of the jobs in the informal sector tend to be poor quality employment situations mainly due to earning levels. In Kenya, a person is said to be time-related underemployed if working less than 28 hours in the reference week (KNBS, 2018). According to (Kippra 2009), many people, approximately 69.8% who are underemployed, are poor. The income earned by the underemployed people has poor prospects of improvement while incomes are barely beyond subsistence. Such shows the detrimental impact that underemployment can have on individuals. Lack of decent work or quality work opportunities has led to disparities in the Kenyan labour market. Despite the statistic available on unemployment, they are not enough to understand the deficiencies in the labour market. Studies conducted on underemployment have mainly focused on time-related underemployment. However, there is also the need to explore the earnings of individuals in the reference week to measure the invisible underemployment situation.

1.2 Statement of the Problem

Youth unemployment is a significant problem in Kenya, currently at 7.19% in 2019 (ILO, 2021). Despite the employment rate for the total labour force, which stands at 2.98%, those in employment are also unable to reach their full capacity and productivity, as evidenced by the underemployment rate, which is at 20.4% for the Kenyan population based on the 2015/16 KIHBS survey (KNBS, 2018). Unemployment is mainly used as the measure for understanding the inefficiencies in the labour market. However, unemployment is a limited measure for under-utilization since it fails to provide insights regarding job quality, income inequality, and inadequate working time in the labour market. There is a need to adopt the underemployment measure to help in complementing the unemployment rate for measuring labour under-utilization. Studies have been conducted to explore the determinants of unemployment. However, few studies have been carried out to explore the determinants of underemployment as a key measure for under-utilization in the labour market. The studies conducted on underemployment have mainly focused on the general population with little consideration of the situation among the youth. The few studies conducted have not explored both the visible and invisible unemployment determinants, thus limiting the formulation of effective policies geared towards addressing the issue of underemployment among youth in Kenya. This study examines the determinants of underemployment among the youth population in Kenya. The study has used the latest Kenya Integrated Household Budget Survey 2015/16 to carry out the assessment. It focuses on both the visible and invisible determinants of underemployment, providing a wider understanding of labour under-utilization among the youth.

1.3 Objectives

1.3.1 General Objective

To model the determinants of underemployment among youths in Kenya.

1.3.2 Specific Objectives

1. To identify the determinants of visible underemployment among youths in Kenya.
2. To identify the determinants of invisible underemployment among youths in Kenya.

1.4 Significance of the Study

The Sustainable development goal (SDG) number 8 calls for decent work and economic growth. According to the Kenya Decent Work Program under the Vision 2030 framework, one of the major goals is to improve access to decent and productive employment opportunities among citizens. This study will inform on the key determinants of underemployment thus revealing the factors that can be addressed to handle the problem. The findings from this study are critical towards informing policy formulation and interventions needed for improving efficiency in the labor market by providing decent jobs to youths and utilize their full capacity leading to better labor productivity.

1.5 Limitation of the Study

This study has used the data for the KIHBS survey conducted in 2015/16. It limits modelling based on the latest data which would provide a better current picture of the labor market.

2 Chapter 2: Literature Review

2.1 Preliminaries

In this section, we discuss the literature review focusing on theoretical review, empirical review and the Kenyan specific review. The section also captures the conceptual framework.

2.2 Theoretical Literature Review

According to the 16th International Conference of Labour Statisticians (ICLS), underemployment is an indicator of productive capacity under-utilization for the employment-population (ILO, n.d.). Such also captures the individuals arising from the deficient economic system both at the local or national levels. A number of theories provide an explanation to the concept of underutilization of productive capacity among employed individuals. The human capital theory developed by Becker and Mincer seeks to explore the decision of individuals to invest in human capital, mainly education and training, which influences wages and salaries. Human capital is a critical physical means of productivity (Becker, 1962). Education renders individuals more productive by boosting their marginal product compared to those not educated. Considering the relationship between underemployment and productive capacity under-utilization, the theory stipulates that more educated individuals are likely to be more productive, thus less underemployment possibilities. There is a relationship between differences in human capital and productivity level variances caused by gender, labour market experience and education levels. The labour force framework is an international recognized model for computations relating to employment and unemployment. The framework holds that measurements for the indicators of inadequate employment and underemployment should be based mainly on the work situations and production capacities as provided by the employed individuals. The framework developed by Hauser (1974) provided several components for addressing the underemployment issue, thus improving the labour force framework provided by the International Labour Organization. The critical components of the labour utilization framework include sub-unemployed, unemployed, low-hour employees, low-income employees, mismatched employees and those that are adequately employed.

2.3 Empirical Literature Review

Acosta et al. (2018) employed a bivariate probit selection model to examine underemployment and unemployment among young employees and the business cycle in Spain. The study adopted the data from the 2006-2014 labour force survey of Spain. The finding from the study revealed that there is a negative relationship between every education level and underemployment. There was a significant relationship between the lowest levels of education with underemployment. Lower risks of underemployment were identified among individuals who had studied health, technology and science courses. The oriented fields of humanities, arts and education were mainly associated with the increasing probability of underemployment. In this case, courses enhancing the work-oriented skills among youths lowers their risk of underemployment. The study recommends initiatives to educate youths about labour market prospects through career guidance to enhance better outcomes in employment and address underemployment concerns.

Hyefouais (2016) adopted the Fairlie decomposition, sample selection and probit models in investigating the determinants and characteristics of underemployment in Cameroon. The research used data collected in 2010 by the National Institute of Statistics of Cameroon. Based on the study results from the probit model, it is estimated that approximately 11.5% of the people in Cameroon are visibly underemployed while the rate of invisible underemployment is 62.7%. Findings from the study showed that the underemployment gap between rural and urban employees stood at 26.4%. The differences in observable characteristics influence this gap. The employees with the post-secondary level of education had a high rate of underemployment at 23.6% as compared to other groups. The levels of visible underemployment for the rural and urban workers were 11% and 12.4% respectively. On the contrary, invisible underemployment for workers in the rural areas stood at 74% and urban at 44.8%. The visible underemployment was more prevalent in the public sector at 21% in comparison to the private sector. Results also show that men are less affected by the visible underemployment at 21.7% than their female counterparts at 27.7%. Similarly, invisibly underemployed women were more at 71.02% as compared to men at 54.9%. There was a negative relationship between invisible underemployment and the levels of education, considering that people with higher levels recorded low rates. High rates of invisible underemployment 81% were recorded among the people with no education. The probit model results also revealed that location, sex, age, socio-profession category, employment sector, business sector, and education have a significant effect on underemployment. The study recommends better policies focusing on assessing the employability conditions in the labour market, addressing the gender gap, financing private investments in rural areas, and the manufacturing sector to address the underemployment problem in Cameroon.

Gorga and Strobl (2003) conducted an empirical analysis to investigate the incidence of visible underemployment, a case study of Trinidad and Tobago. Findings from the study reveal that individuals who are less educated tend to be more prone to underemployment. Similarly, there is a high prevalence of underemployment in the private sector as compared to the public sector. The study reveals that full-time employment is more attractive in terms of higher returns, stability and offering more working hours as desired by workers. In this case, there are better benefits associated with full-time employment. However, only half of the visible underemployed workers get the opportunity for full-time employment within the set duration of three months.

Mukherjee et al. (2018) conducted a study to measure and analyze underemployment in India. The study used the Tobit regression model for analysis of the unit-level survey data of the National Sample Survey Organization office. The study findings revealed that underemployment levels were higher among individuals that were not self-employed, female gender and educated, especially with technical education. There were significantly low levels of underemployment among individuals that are married. Findings also show that interaction between variables such as male gender, age below 40 years and living in the urban areas influenced the increased probability of being underemployed. The interaction of variables age, gender and education also significantly influenced underemployment. Belonging to the self-employed category also increased the likelihood of being underemployed compared to being in other categories of employment. A study conducted by Wilkins (2006) investigated the personal and job characteristics associated with underemployment using the 2001 Household, Income and Labor Dynamics in Australia (HILDA) survey. The findings from the multinomial logit model showed that underemployment is influenced by factors such as labour market history, educational attainment and age effects. The number of children in a household had a positive impact on underemployment rates which can be explained by the constrained working hours focused more on caring for the children. Results also showed that individuals in part-time employment are more at risk of underemployment. There is a similarity between the factors affecting unemployment and underemployment, including disability, age, labour market and education level.

Ruiz et al. (1996) carried out a longitudinal analysis to explore the determinants of three forms of underemployment, including part-time employment, unemployment and temporary employment. The analysis conducted using the probit model revealed that socio-economic factors such as educational attainment and occupational track are the significant determinants of underemployment, unemployment and temporary employment among the youth population. Results also revealed that job search-related actions, age and gender had no significant impact on future underemployment. There is a difference between the factors that influence part-time employment compared to those determining temporary employment or unemployment. The study provides policy recommendations on the need of implementing socio-political and socio-economic driven strategies geared towards addressing underemployment among young people as they start their careers.

2.4 Kenyan Specific Review

Munga et al. (2012) conducted a study to profile the labor market, focusing on unemployment and underemployment among the youth in Kenya. The study also sought to explore the aspects relating to job quality and understanding how education can be used in addressing inequalities in the Kenyan job market. The study adopted the multinomial probit model in estimating the youth employment status in Kenya. Findings from a study conducted by Munga et al. (2012) revealed that about 77% of the youth are engaged in vulnerable jobs considering that they work less than 28 hours weekly or rather more than 65 hours in the reference week based on the 2009 census survey. There were more youths in the vulnerable employment at 77% as compared to those who are unemployed at 7%. There is a negative relationship between age and the probability of being unemployed compared to the unemployment situation. Education level was a critical factor in explaining the labour market. About 90% of employees who have attained low levels of education are likely to be underemployed, thus engaging in vulnerable jobs while compared to those who have attained higher levels of education such as the university. There was a positive relationship between marriage and being underemployed. The study also found that youths in the rural areas in Kenya have a higher probability of being underemployed than those in the urban areas. There was a high likelihood of females being unemployed rather than being underemployed as compared to their male counterparts. The study recommended that education and skill development be prioritized to help in raising productivity in the informal sector while investing in labour-intensive technologies to address the dynamics of the Kenyan youth labour market. It is recommended to implement viable sectoral employment policies to help in creating sustainable employment opportunities, which are vital in addressing unemployment and underemployment problem.

A study conducted by Kiiru et al. (2019) investigated the factors that explain disparities between underemployment or open employment and the ability to secure full employment with a focus on the youth population in Kenya. The research adopted the multinomial logit model approach for analysis using the KIHBS 2005/2006 data. Findings from the study showed a high likelihood of being underemployed for older youths compared to younger youths. Results also revealed that female youths are more likely to be underemployed than in full employment compared to their male counterparts. The unemployment situation among the youth in Kenya is gendered due to existing inequalities. The location aspect also influences the underemployment or full employment status of the youth. The youth residing in Nairobi were less likely to be underemployed despite being more exposed to facing open unemployment compared to those living in provinces such as Western, Rift valley, and Nyanza. The study revealed that education is a factor influencing the acquisition of employment among the youth. The paper provides policy recommendations emphasising the need to implement gendered mainstreaming in the labour market policies and enhancement of innovations in the education system.

Vuluku et al. (2013) investigated the gender gap analysis of unemployment and underemployment in Kenya. The study used cross-sectional survey data from the Kenya Integrated Household Budget Survey 2005/06. A binary probit regression analysis was conducted where the findings revealed that 5.4% of the underemployment probability gap is explained by female-male variances in household and personal characteristics. Adverse shocks, age, residence, region, educational level and marital status were the key variables that influenced the gender gap between males and female in underemployment and unemployment probabilities. Males had a lower probability of being underemployed (0.025) compared to their female counterparts at (0.0596). More educated individuals were less likely to be underemployed while compared to the less educated. The study also revealed that adverse shocks such as crop loss had a negative impact on underemployment, such as increasing the gender gap in unemployment and underemployment levels. The study concluded by providing policy recommendations to target the various locations and age cohorts in bridging the gender gap while handling underemployment and unemployment through initiatives such as reducing disparities in access to education.

2.5 Conceptual Framework

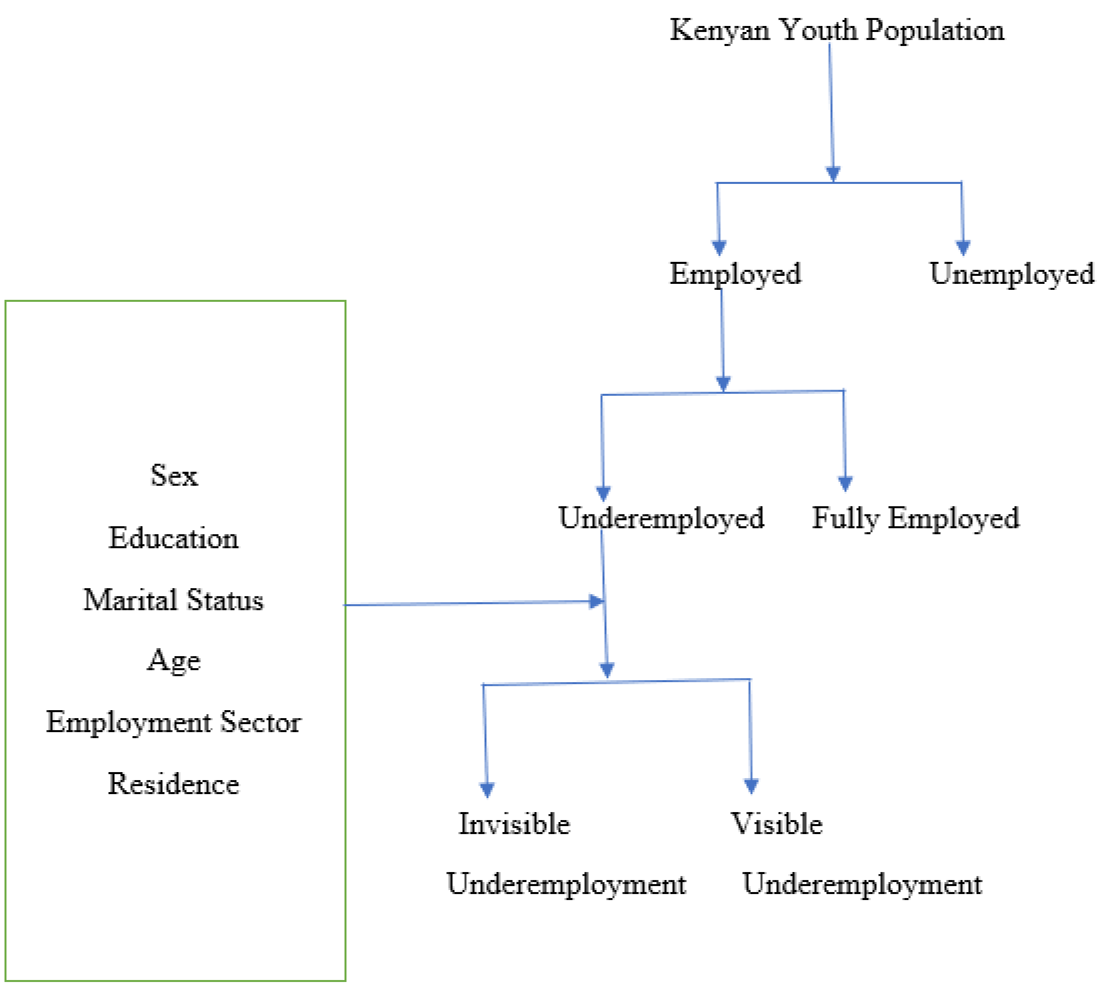

Figure 1. Conceptual Framework

3 Chapter 3: Methods

3.1 Data Source

3.1.1 Sample Selection Description

This study is conducted using the cross-sectional data from the 2015/16 Kenya Integrated Household Budget Survey (KIHBS). KIHBS is multi-indicator survey for a number of socio-economic aspects. The particular interest is on the data about labor force characteristics which captures aspects such as unemployment, participation rates and active population. The data envisages 50 study domains with focus on all forty-seven counties in Kenya.

This study has focused on the national picture. The household data is used to examine the determinants of underemployment among youth in Kenya. The sample design considers all the indicators for survey considering that KIHBS is a multi-indicator survey. Stratification of the frame in each county into urban and rural areas was adopted for the study. The sample size was independently determined resulting in a national sample of 24,000 households. The main instruments for data collection include household member information questionnaire, household level information questionnaire and household consumption expenditure information questionnaire. Data was collected from 10 households in each cluster.

3.1.2 Sampling Frame

National Sample Survey and Evaluation Program (NASSEP V) master frame is the key sampling frame adopted for collecting data for 2015/16 KIHBS. The sampling approach was developed by the Population and Housing Census (KPHC). A total of 96,000 Enumerations Areas (EAs) are contained in the sampling frame. There sample is clustered into 5,360 containing 2,792 rural and 2,568 urban areas.

3.1.3 Data Analysis Technique

The R and SPSS softwares are used in this study.The data in this study is mainly analyzed using the binary logistic regression model to help in examining the relationship between the visible and invisible underemployment as the key outcome variables with a number of explanatory variables. The descriptive statistics technique is also adopted for data analysis.

3.2 Data Variables and Validation

There are various questions contained in the household level information questionnaire that are useful for this study (KIHBS). These questions are especially useful for selection of the response variables needed for estimation.

Visible Underemployment

1. How many hours does Name usually work per week in all these activities (labor activities)?
2. During the last 7 days, would Name have wanted to work for pay/profit more hours than he/she actually worked in the primary main job?
3. If [NAME] was offered a job how soon would he/she be available to start work?

The response variable is Visible Underemployment categorized as 1=Yes and 2=No.

Invisible Underemployment

1. How much was Name's payment for wages and salary last one month?

The response variable is invisible underemployment categorized as 1=Yes and 0=No.

The inclusion criteria of the respondents from the KIHBS 2015/16 data is based on the following aspects.

1. The respondents aged between 15-34 years.
2. A person who participated in the labor force.

Variables Description

Table 1. Response variables

Dependent variable	Description	Value labels
Visible Underemployment	It is the time related underemployment comprising of employees who were willing to work additional hours, were available to do so and had worked less than 28 hours in the reference week	It is a binary variable assigned the dummy variables. 1= Employees who are visibly underemployed and 0 = if otherwise
Invisible Underemployment	Corresponds to the situation under which the employee receives earnings that are lower than the minimum monthly wage (Ksh. 13,572) in the reference month.	The variable is binary assigned the dummy variables. 1= Employees who are invisibly underemployed. 0 = Otherwise

Table 2. Independent variables

Independent variables	Description variables	Value labels
Sex	The gender attribute of the employee	It is a dummy variable 1=Female 0=Male
Education	The variable shows the level of education qualification of the employees	Dummy variables are generated in accordance with the level of education. 0=Pre-primary(reference variable) 1=Primary 2=Post-Primary(Vocational) 3=Secondary 4=Post-secondary(College,University)
Residence	The variable represents the place where the employee lives which can be a key determinant of underemployment	The variable is assigned dummy variables. 1=Urban, 0=Rural
Age	It is the age category of the youth employees which can be a key determinant of underemployment status	Dummy variables created for age. 0=15-19 years(reference variable) 1=19-24 years 2=25-30 years 3=26-34 years
Marital status	The variable represents the marital status of the employees which can be a determinant for underemployment among the youth	The dummy variables are created. 0=Monogamous Married(reference variable) 1=Polygamous Married 2=Living together 3=Separated 4=Divorced 5=Widow or Widower 6=Never Married
Employment sector	The variable refers to the sector category under which the youth worker falls.	Dummy variables created include: 1=Public sector(reference variable) 2=Private sector (formal) 3=Informal sector.

3.3 Binary Logistic Model

The binary logistic model is used for modelling situations where the response is binary while the predictor or explanatory variables can be categorical or continuous (Abonazel et al., 2018). The assumption is made for only two values which are assigned the values 0 and 1.is considered to depend on the $\bar{X}$ which is a vector of explanatory value. The probability of being in any conditions of y_i will be between the values 0 and 1. Y_i follows a bernoulli distribution.

$$y_i = \begin{cases} 1 & \text{if underemployed} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

$$P_r = (Y_i = 1) = \pi_i \text{ and } P_r = (Y_i = 0) = 1 - \pi_i \tag{2}$$

The expected value and variance of Y_i are given by,

$$E(Y_i) = \mu_i = \pi_i \tag{3}$$

and

$$Var(Y_i) = \sigma_i^2 = \pi_i(1 - \pi_i). \tag{4}$$

The Bernoulli distribution can be written as

$$P_r\{Y_i = y_i\} = \pi_i^{y_i}(1 - \pi_i)^{1-y_i} \tag{5}$$

3.4 Model Transformation

Suppose there are k independent observations $y_i, \ldots, y_k$, we assume Y_i has a binomial distribution when the i^{th} observation is treated as a realization of the random variable Y. Therefore,

$$Y_i \sim B(n_i, \pi_i) \tag{6}$$

Where

n_i = binomial denominator and

π_i = probability with the individual data

$n_i = 1$ for all i

The regression equation can be created through log transformation of the π_i - values to a distribution form. The log transformation of π_i which is now the log distribution is also identified as logit (π_i). The logit (π_i) is defined to be the log to base e of the likelihood ratio that the response variable is 1.

Suppose the logit of the underlying probability π_i is a linear function of the predictors. The function is defined as,

$$\operatorname{logit}(\pi_i) = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = x_i'\beta \tag{7}$$

where π_i can range from 0 to 1. $i = 1, 2, 3, \ldots, n$

$$\operatorname{logit}(\pi_i) = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k \tag{8}$$

where $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k = x_i'\beta$

Solving for π_i in the logit model

$$
\begin{aligned}
\text{logit}\,(\pi_i) &= x_i'\beta \\
\left(\frac{\pi_i}{1-\pi_i}\right) &= \exp\left(x_i'\beta\right) \\
\pi_i &= (1-\pi_i)\left(\exp\left(x_i'\beta\right)\right) \\
\pi_i &= 1\left(\exp\left(x_i'\beta\right)\right) - \pi_i\left(\exp\left(x_i'\beta\right)\right) \\
\pi_i + \pi_i\left(\exp\left(\left(x_i'\beta\right)\right)\right) &= \exp\left(x_i'\beta\right) \\
\pi_i\left(1+\exp\left(x_i'\beta\right)\right) &= \exp\left(x_i'\beta\right) \\
\pi_i &= \frac{\exp\left(x_i'\beta\right)}{1-\exp\left(x_i'\beta\right)}
\end{aligned}
$$

The function gives a logistic curve.

23

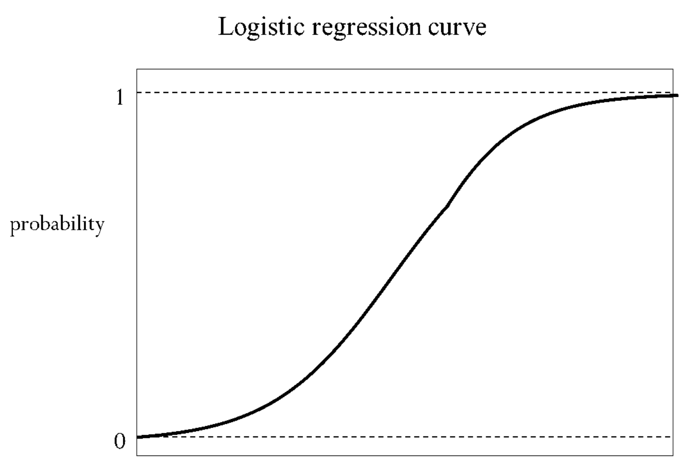

Figure 2. Logistic regression curve

Odds

The odds of occurrence of an event is defined using the probability. The odds is defined as the ratio of the probability of an event occurring compared to not occurring.
The probability is given by π_i
The odds are given by

$$\frac{\pi_i}{1-\pi_i} = \frac{\text{Probability of an event occurring}}{\text{Probability of an event not occurring}} \tag{9}$$

Odds ratio

The odds ratio is defined from the odds. Translation of effects is done for interpretation of odds ratio. The odds ratio provides the measure of the odds of a given event occurring in one group comparing to the odds of the same event occurring in another group. The exponentiated coefficient gives the odd ratio.

$$O.R = \exp\beta_k) \tag{10}$$

Interpretation of the Odds ratio

- If the Odds ratio is 1 then both groups compared have the same odds of occurring.
- An odds ratio of less than one implies that the event of interest is 100(1-O.R)% less likely to occur in a given group compared to the reference group.
- The odds ratio of more than 1 but less than 2 implies that the predicted event is 100(O.R-1)% more likely to occur for the specific group in comparison to the reference group.
- The odds ratio of more than 2 implies that the predicted event is O.R times more likely to occur for the given group in comparison with the reference group.

Model specification

$$\text{Prob}\left[Y_i = 1 | i = 1,2,3,\ldots\right] = \pi = \frac{\exp\left(x_i'\beta\right)}{1-\exp\left(x_i'\beta\right)} \tag{11}$$

Where, $\beta = (\beta_1, \beta_2, \beta_3, \ldots, \beta_k)^T$ and $x_i = (x_1, x_2, x_3, \ldots, x_k)$

The binary response variables

$$Visible underemployment = \begin{cases} 1 & \text{if (working<28 hours in reference week,willing,available for more)} \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

$$Invisible underemployment = \begin{cases} 1 & \text{if (total earning(month)< Ksh. 13,572)} \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

Since $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k = x_i'\beta$

The model employed in the study.

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon \tag{14}$$

Where:

Y_i = dependent variable (underemployment)

β_0 = intercept which is constant

β_1 to β_k = the parameters are unknown and estimated from the data.

ε = error term.

The variables X_1 to X_k = the independent variables in the model.

$$\begin{aligned} ln(Odds of Underemployment(visible/invisible) = {} & \beta_0 + \beta_1(sex) + \beta_2(education) \\ & + \beta_3(residence) + \beta_4(age) + \beta_5(marital\ status) \\ & + \beta_6(employment\ sector) + \varepsilon \end{aligned}$$

3.4.1 Parameter Estimation

The response variable in the study underemployment can either be visible or invisible in nature. There is variable estimation for both visible and invisible underemployment as the key response variables in the study. The responses variables are binary in nature. Therefore, binary logistic regression model is appropriate for this type of response. The maximum likelihood estimation (MLE) is adopted for estimation of the parameters for the binary logistic regression. The estimation approach is appropriate considering that other than prediction of the class (0 or 1), the binary logistic regression also provides probabilities prediction (Abonazel et al., 2018). For each data point there are observations Y_i and vectors of features x_i. The probability of occurrence is given by:

$$P_r(Y_i = 1) = \pi_i \text{ and } P_r(Y_i = 0) = 1 - \pi_i \tag{15}$$

The likelihood for the binary logistic regression model is:

$$L(\beta_0, \beta) = L = \prod_{i=1}^{n} \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \tag{16}$$

The log-likelihood converts the products into sums to obtain the following:

$$\text{Log}(L) = \sum_{i=1}^{n} y_i \log \pi(x_i) + (1 - y_i) \log [1 - \pi(x_i)] \tag{17}$$

$$= \sum_{i=1}^{n} y_i(\beta_0 + x_i\beta) + \sum_{i=1}^{n} -\log(1) + e^{\beta_0 + x_i\beta} \tag{18}$$

To obtain the the maximum likelihood estimates we differentiate the log likelihood with respect to the parameters and setting the derivative equal to zero.We take the derivative with respect to one component β_j.

$$\frac{\partial L}{\partial \beta_j} = -\sum_{i=1}^{n} \frac{1}{1 + e^{\beta_0 + x_i \cdot \beta}} e^{\beta_0 + x_i \cdot \beta} x_{ij} + \sum_{i=1}^{n} y_i x_{ij} \tag{19}$$

$$= \sum_{i=1}^{n} (y_i - \pi(x_i; \beta_0, \beta)) x_{ij} \tag{20}$$

The solutions can be obtained by using the iteratively re-weighted least squares (IRLS) approach. The IRLS algorithm gives the following maximum likelihood of β.

$$\beta_{MLE} = (X'\hat{W}X)^{-1}(X'\hat{W}\hat{Z}) \tag{21}$$

$$z_i \tag{22}$$

$\hat{W} = \text{diag}\{\hat{\pi}_i(1-\hat{\pi}_i)\}$ and $\hat{Z} = \log(\hat{\pi}_i) + \frac{y_i - \hat{\pi}_i}{\hat{\pi}_i(1-\hat{\pi}_i)}$ which is the i^{th} element of the vector $\hat{Z}$, the iterative process in the equation is indicated using the hats.

Test of Hypothesis for Coefficients

The Wald Test is used for testing the significance of the predictor variables in the model. We test the hypothesis $H_0 : \beta_k = 0$ and $H_0 : \beta_k \neq 0$

The Wald test statistic is given by

$$Z^2 = \left(\frac{\hat{\beta}_k}{SE\hat{\beta}_k} \right)^2 \tag{23}$$

There is comparison with a chi-square distribution with 1 degree of freedom for each of the Wald statistic.

The test is used in calculating the confidence interval for β_k we assert with $100(1-\alpha)\%$ confidence that the true parameter lies in the interval within the range.

3.4.2 Goodness of Fit Test

The goodness of fit test helps in determination of whether the binary logistic regression model is correctly specified. The goodness of fit statistics is important to help reveal the significant discrepancy between the observed data and the fitted model. The goodness of fit for the model is tested using the deviance statistics test. In this test, there is grouping of observations in accordance with the distinct covariate values thus requiring the observation in each category to increase to infinity with the increase of sample size. Considering that the covariates in the data are all categorical, the test is efficient for testing the goodness-of-fit of the binary logistic regression model.

The deviance D statistics follows the x^2 distribution with n-p degrees of freedom, making it possible to calculate the p-values. As for the grouped data

n = number of groups
p = number of parameters in the model.
The deviance statistic is given by:

$$\text{Deviance} = -2log(L_M - L_S) \tag{24}$$

$$D = 2\sum_{i=1}^{k} \left\{ y_i \log\left(\frac{y_i}{\hat{\pi}_i}\right) + (1-y_i)\log\left(\frac{1-y_i}{1-\hat{\pi}_i}\right) \right\} \tag{25}$$

y_i= observed value
$\hat{\mu}_i$ = fitted value for the i^{th} observation.
Calculation of deviance statistics

$$D = 2\sum_j O_j \log\left(\frac{O_j}{E_j}\right) \tag{26}$$

Where:
O_j= is the observed frequency
E_j= expected frequency based on the fitted binary logistic model.
L_M= max. log-likelihood for Model M
L_S = max. log likelihood for the saturated model

In a perfect fitted model, the ratio of the observed over the expected is equated to one while its logarithm is 0. Such implies the deviance is 0. Null hypothesis for the test:

H_0: The fitted model is the better fit
H_1: The saturated model is the better fit

Assumptions of Binary Logistic Regression Model

1. The response variable in the binary logistic model regression needs to be binary in nature. Tested by evaluating the structure of the response variable to ensure it is binary in nature.

2. The model assumes to be little or absence of multicollinearity of the independent variables.

   The test used for multicollinearity of the independent variables is the variance inflation factor (VIF). VIF for each of the independent variable is calculated. Variance inflation factor is a function of R_i from the auxiliary regression of the selected independent variable on the remaining explanatory variable.

$$\text{VIF} = \frac{1}{1 - R_i^2} \tag{27}$$

   $R_i^2 = R^2$ value obtained by regressing the j^{th} explanatory variable on the remaining explanatory variables. The values above 10 imply there is a high multicollinearity. There is need to investigate the VIF that goes beyond 4.

3. The binary logistic regression model does not assume the dependent and independent variables to be linearly related. Assumes log odds of dependent variable.

The residual vs fitted scatter plots is used to test for linearity. The scatter plots should not show linearity of the dependent and independent variables of the binary logistic regression model.

4. Larger samples are required considering that maximum coefficients adopting the maximum likelihood approach are large sample estimates. There is a requirement of having at least 10 cases with the outcome that is least frequent in model for each of the explanatory variable.

5. The error terms need to be independent; the model assumes independence of observations which should not be obtained from matched or repeated measurements. The autocorrelation test for this assumption is used to check where the errors of observations are correlated.

   Test is done using the Durbin Watson Test. Test Statistics is given by

$$d = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \tag{28}$$

   The critical values for the d statistic is d_L and d_u, the values will lie between 0 and 4 where closer to 0- positive autocorrelation, closer to 4-negative autocorrelation, about 2 -no evidence of positive or negative first order autocorrelation.

4 Chapter 4: Data Analysis & Results

4.1 Introduction

The chapter provides the findings obtained from analysis of the data. Descriptive results obtained using SPSS and the findings from the binary logistic regression model from the R software are captured. The results for the estimation of the determinants for visible and invisible underemployment are presented in this chapter.

4.2 Visible Underemployment

4.2.1 Descriptive Findings

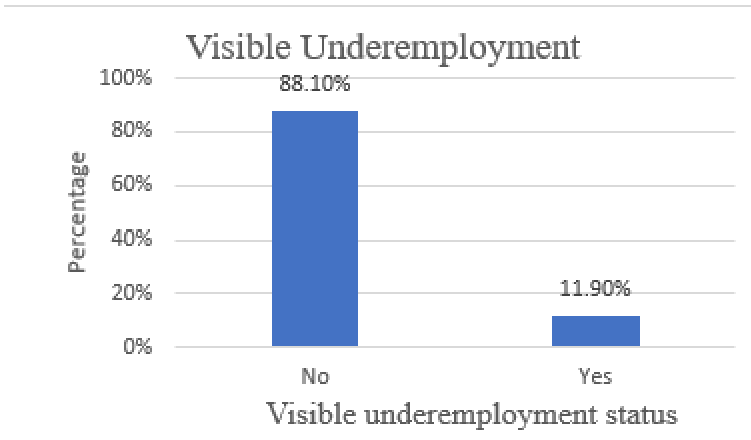

Figure 3. Graph for Visible Underemployment

The figure above shows that 11.9% of the sampled youth respondents are visibly underemployed.

The study findings reveal that female gender is more visibly underemployed at 60.9% as compared to the male respondents at 39.1%. Visible underemployment was higher rural residence at 15.5% compared to urban residence at 7.1%. Findings show that the private(formal) sector of employment has the highest number of visibly underemployed respondents at 59.3 % with informal at 39.5 % and public lowest at 1.2 %. The highest rate of underemployment was recorded in the respondents with primary level of education at 13.5% while those with post-secondary education had the lowest rate of visible underemployment at 3.9%. According to the results, the age groups 15-19 years and 20-24 years of the respondents have the highest number of visibly underemployed youths both at 28.8% with age category 30-34 years having the lowest percentage at 20.8 %. The youth respondents in the category of never married in marital status are more visibly underemployed at 45.5% as compared to any other marital status with those living together at 0.4 % have the lowest number of visibly underemployed youths.

4.2.2 Data Validation Results

Multi-collinearity

Table 3. VIF results for visible underemployement

	GVIF	Df	GVIF^(1/(2*Df))
Gender	1.096038	1	1.046918
Age	1.56627	3	1.07765
Education	1.252738	5	1.022789
Employment sector	1.134415	2	1.032031
Residence	1.06743	1	1.033165
Marital	1.719844	6	1.046223

Table 3 reveals there is no multi-collinearity in the independent variables as the values are all below 4.

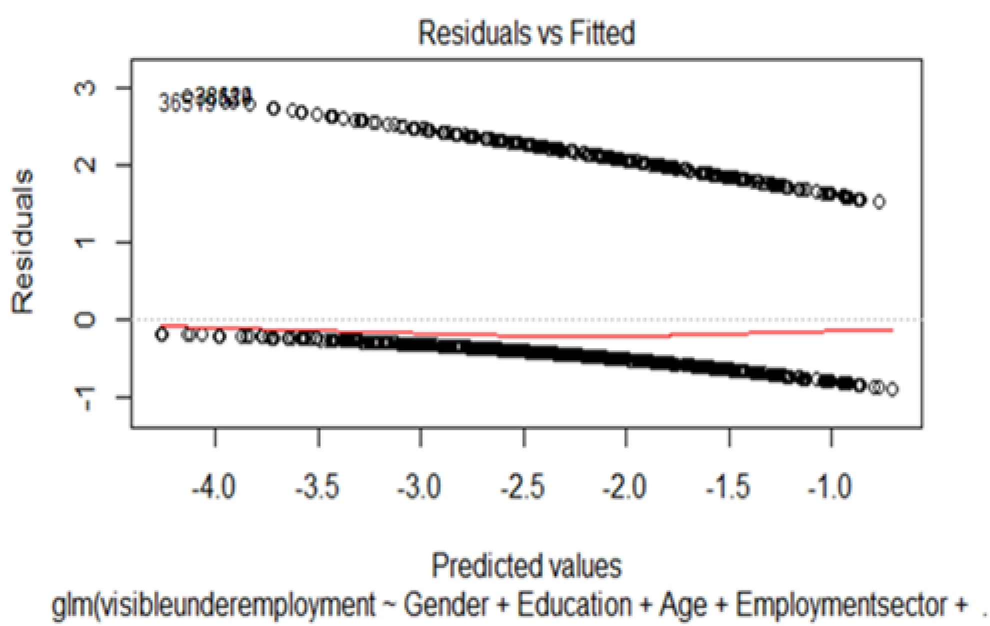

Figure 4. Visible underemployment residualvsfitted plot

The figure above shows there are distinct patterns which implies there was no linear relationship between the dependent and independent variables.

Auto-Correlation

```
lag Autocorrelation D-W Statistic p-value
   1       0.2215109      1.556944       0
 Alternative hypothesis: rho != 0
```

The autocorrelation test using the Durbin Watson test shows that the value is close to 2. Therefore, there is no autocorrelation implying there is independence of errors for the observations.

4.2.3 Logistic Regression Results

The fitted model was tested for goodness of fit using the deviance statistics. The value of the deviance 543 with 18 degrees of freedom gives a p-value=1. Therefore, we fail to reject the null hypothesis which implies that the fitted model is a and adequately describes the data.

Table 4. Binary logistic regression model results for visible underemployment variable

	Estimate	Std.Error	z value	P-value	
(Intercept)	-2.924192	0.204365	-14.309	2.00E-16	***
GenderFemale	0.553133	0.049536	11.166	2.00E-16	***
Age20-24 years	-0.089591	0.07064	-1.268	0.204701	
Age25-29 years	-0.296592	0.077322	-3.836	0.000125	***
Age30-34 years	-0.585526	0.084611	-6.92	4.51E-12	***
EducationPrimary	0.34651	0.092461	3.748	0.000179	***
Education Post-primary, vocational	0.322068	0.20344	1.583	0.113396	
EducationSecondary	0.339628	0.098297	3.455	0.00055	***
EducationPostsecondary (college, university)	-0.067403	0.130802	-0.515	0.60634	
Education Other	-0.184171	0.381971	-0.482	0.629693	
EmploymentsectorPrivate formal	1.096141	0.170714	6.421	1.35E-10	***
EmploymentsectorInformal	0.750422	0.170457	4.402	1.07E-05	***
ResidenceUrban	-0.559539	0.053756	-10.409	2.00E-16	***
MaritalPolygamousmarried	0.154892	0.120209	1.289	0.197564	
MaritalLivingtogether	-0.128435	0.400511	-0.321	0.748455	
MaritalSeparated	0.073268	0.128518	0.57	0.568607	
MaritalDivorced	0.225453	0.247437	0.911	0.362214	
MaritalWidoworwidower	-0.005219	0.258226	-0.02	0.983875	
MaritalNevermarried	0.012584	0.061121	0.206	0.836885	

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

The Visible underemployment model deduced from the above table is as follows.

$$\begin{aligned}\text{Ln}\left(\frac{\pi_i}{1-\pi_i}\right) &= -2.9242 + 0.5531X_1 - 0.0896X_2 - 0.2966X_3 - 0.5855X_4 \\ &+ 0.3465X_5 + 0.3221X_6 + 0.3396X_7 - 0.0674X_8 - 0.1842X_9 \\ &+ 1.0961X_{10} + 0.7504X_{11} - 0.5595X_{12} + 0.1549X_{13} \\ &0.1284X_{14} + 0.0733X_{15} + 0.2254X_{16} - 0.0052X_{17} + 0.0126X_{18}\end{aligned}$$

The Visible underemployment model with significant variables is given below:

$$\begin{aligned}\text{Ln}\left(\frac{\pi_i}{1-\pi_i}\right) &= -2.9242 + 0.5531X_1 - 0.2966X_3 - 0.5855X_4 \\ &+ 0.3465X_5 + 0.3396X_7 + 1.0961X_{10} + 0.7504X_{11} - 0.5595X_{12}\end{aligned}$$

Table 5. Coefficient estimates and confidence interval results for visible underemployment model

	OR	2.5%	97.5%
(Intercept)	0.05371	0.03556	0.07936
GenderFemale	1.73869	1.57812	1.91637
Age20-24 years	0.91431	0.79618	1.05025
Age25-29 years	0.74335	0.6388	0.86501
Age30-34 years	0.55681	0.47162	0.65714
EducationPrimary	1.41412	1.18297	1.70009
EducationPost-primary, vocational	1.37998	0.91434	2.03414
EducationSecondary	1.40442	1.16102	1.70714
EducationPostsecondary(college, university)	0.93482	0.72302	1.20775
EducationOther	0.83179	0.36368	1.65754
EmploymentsectorPrivateformal	2.9926	2.17042	4.24661
EmploymentsectorInformal	2.11789	1.53694	3.0041
ResidenceUrban	0.57147	0.51401	0.63461
MaritalPolygamousmarried	1.16753	0.91779	1.47096
MaritalLivingtogether	0.87947	0.36519	1.79762
MaritalSeparated	1.07602	0.83097	1.37611
MaritalDivorced	1.25289	0.75	1.9884
MaritalWidoworwidower	0.99479	0.58074	1.6076
MaritalNevermarried	1.01266	0.89817	1.14137

Visible Underemployment Model Interpretations

The results from the binary logistic regression model (table 4) shows that gender, age categories for 30-34 and 25-29 years, education secondary category, private sector, informal sector and residence were significant determinants of visible underemployment.

According to the findings in (table 5), females are 73% more likely to be visibly underemployed as compared to their male counterparts. Similarly, the people aged 30-34 years old are 45 % less likely to be visibly underemployed as compared to the people aged (15-19 years). People aged 25-29 years old are 26 % less likely to be visibly underemployed as compared to the people aged (15-19 years). The workers in the urban areas are 43% less likely to be visibly underemployed compared to their counterparts in the rural areas. The people in the informal sector were 2.11 more likely to be visibly underemployed as compared to those in the public sectors. Similarly, the people in the private formal sector were 2.9 times more likely to be visibly underemployed compared to their counterparts in the public sector. The workers with secondary education are 37% more likely to be visibly underemployed compared to the workers with pre-primary education. The workers with primary education are 41% more likely to be underemployed compared to the workers with pre-primary education. The post-secondary, vocational and other categories of education did not show any significant results in predicting visible underemployment. Marital status does not have a significant impact on underemployment.

4.3 Invisible Underemployment

4.3.1 Descriptive Statistics

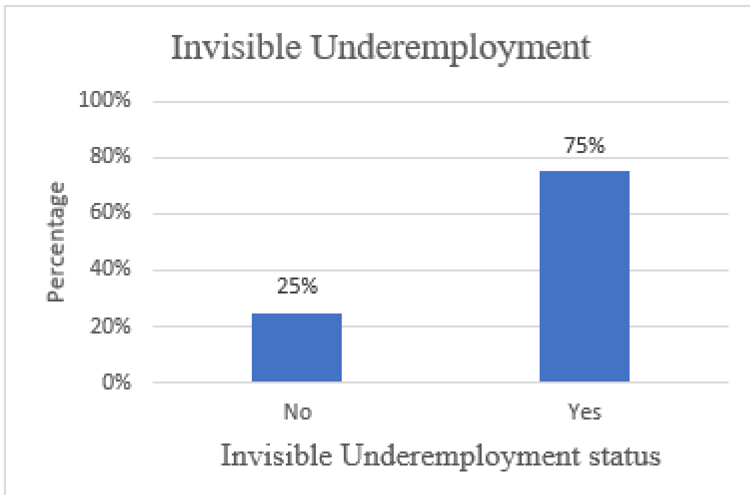

Figure 5. Graph for Invisible Underemployment

Results from the figure above shows that 75% of the youth respondents were invisibly underemployed.

Females are more invisibly underemployed at 83.4% compared to the rate for males at 68%. The findings show that respondents from the rural residence are more invisibly underemployed at 56.4% as compared to the urban residence respondents at 43.6%. The youths aged 15-19 have the highest rate of underemployment at 96.7% with the age category 30-34 years having the lowest rate at 66.8%. Looking at all the education categories, youths who have attained the post-secondary level of education (college and university) have the lowest proportion of employees who are invisibly underemployed at 35.7% while compared to those who are not invisibly underemployed. The highest level of invisible underemployment proportion at 91% compared to non-invisibly underemployed is evident among those in the pre-primary category. The results show 83.6% of the youth respondents in the informal sector are invisibly underemployment, the invisible underemployment rates for those in the private formal and public sectors are 64.2% and 54.3% respectively.

4.3.2 Data Validation Results

Multi-collinearity

Table 6. VIF results for invisible underemployement

	GVIF	Df	GVIF^(1/(2*Df))
Gender	1.078837	1	1.038671
Age	1.213264	3	1.032744
Education	1.333657	5	1.029211
Employment sector	1.183655	2	1.043053
Residence	1.048138	1	1.023786
Marital	1.325333	6	1.02375

There is no multicollinearity in the independent variables as the values are all below 10 which is a value that might require further investigation.

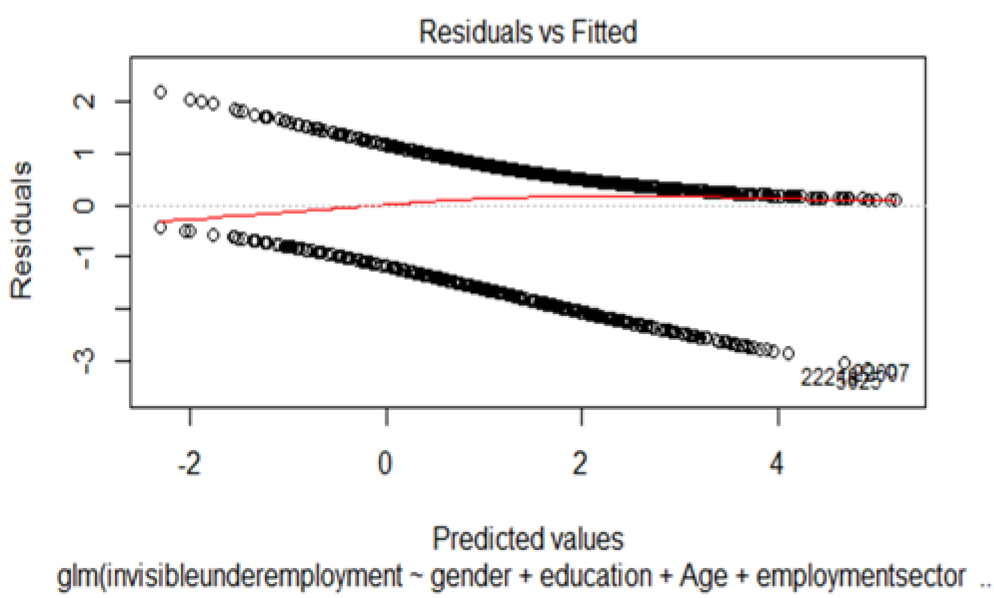

Figure 6. Invisible underemployment residualvsfitted plot

The figure above shows there are distinct patterns which implies there was no linear relationship between the dependent and independent variables.

Auto-Correlation

```
lag Autocorrelation D-W Statistic p-value
   1       0.1641845      1.671542       0
 Alternative hypothesis: rho != 0
```

The auto-correlation test using the Durbin Watson test shows that the value is 1.67 which is close to 2. Therefore, there is no auto-correlation implying there is independence of errors for the observations.

4.3.3 Logistic Regression Results

The fitted model was tested for goodness of fit using the deviance statistics. The value of the deviance 543 with 18 degrees of freedom gives a p-value=1. Therefore, we fail to reject the null hypothesis which implies that the fitted model is correct and adequately describes the data.

Table 7. Binary logistic regression model results for Invisible underemployment model

	Estimate	Std.error	z-value	p-value	
(Intercept)	2.54994	0.23625	10.793	< 2e-16	***
GenderFemale	1.10203	0.05765	19.115	< 2e-16	***
Age20-24years	-0.96969	0.19167	-5.059	4.21E-07	***
Age25-29years	-1.45517	0.18941	-7.683	1.56E-14	***
Age30-34years	-1.76649	0.19076	-9.26	< 2e-16	***
EducationPrimary	0.06202	0.12804	0.484	0.628137	
EducationPost-primary, vocational	-0.54171	0.23095	-2.346	0.018997	*
EducationSecondary	-0.4923	0.13024	-3.78	0.000157	***
EducationPostsecondary (college, university)	-1.99734	0.13756	-14.52	< 2e-16	***
EducationOther	-0.42062	0.33866	-1.242	0.214235	
EmploymentsectorPrivateformal	0.74668	0.09033	8.266	< 2e-16	***
EmploymentsectorInformal	0.95694	0.08927	10.719	< 2e-16	***
residenceUrban	-1.08416	0.05569	-19.466	< 2e-16	***
MaritalPolygamousmarried	0.24758	0.16807	1.473	0.140739	
MaritalLivingtogether	0.42345	0.29367	1.442	0.149322	
MaritalSeparated	0.25094	0.1393	1.801	0.071628	
MaritalDivorced	-0.11778	0.26596	-0.443	0.65787	
MaritalWidoworwidower	0.49362	0.37524	1.315	0.188352	
MaritalNevermarried	0.53706	0.06546	8.205	2.31E-16	***

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

The Invisible underemployment model deduced from the above table is as follows

$$
\begin{aligned}
Ln\left(\frac{\pi_i}{1-\pi_i}\right) =& 2.5499 + 1.1020X_1 - 0.9697X_2 - 1.4552X_3 - 1.7665X_4 \\
& + 0.0620X_5 - 0.5417X_6 - 0.4923X_7 - 1.9973X_8 - 0.4206X_9 \\
& + 0.7467X_{10} + 0.9569X_{11} - 1.0842X_{12} + 0.2476X_{13} \\
& + 0.4235X_{14} + 0.2509X_{15} - 0.1178X_{16} + 0.4936X_{17} \\
& + 0.5371X_{18}
\end{aligned}
$$

The Invisible underemployment model with significant variables is given below:

$$
\begin{aligned}
Ln\left(\frac{\pi_i}{1-\pi_i}\right) =& 2.5499 + 1.1020X_1 - 0.9697X_2 - 1.4552X_3 - 1.7665X_4 \\
& - 0.5417X_6 - 1.9973X_8 - 0.4206X_9 \\
& + 0.7467X_{10} + 0.9569X_{11} - 1.0842X_{12} + 0.5371X_{18}
\end{aligned}
$$

Table 8. Coefficient estimates and confidence interval results for visible underemployment model

	OR	2.5%	97.5%
(Intercept)	12.8063899	8.15282	20.6146
GenderFemale	3.010257	2.69022	3.37246
Age20-24years	0.3791991	0.25602	0.54395
Age25-29years	0.2333597	0.15818	0.3331
Age30-34years	0.1709324	0.11558	0.24468
EducationPrimary	1.0639815	0.82353	1.36108
EducationPost-primary, vocational	0.5817537	0.3725	0.92283
EducationSecondary	0.6112176	0.47113	0.78538
EducationPostsecondary(college, university)	0.1356953	0.10314	0.17693
EducationOther	0.6566396	0.34593	1.31527
EmploymentsectorPrivateformal	2.1099935	1.76772	2.51895
EmploymentsectorInformal	2.6037069	2.18527	3.1011
residenceUrban	0.3381851	0.30309	0.37705
MaritalPolygamousmarried	1.2809232	0.92963	1.79855
MaritalLivingtogether	1.5272153	0.87048	2.76123
MaritalSeparated	1.2852316	0.98362	1.69915
MaritalDivorced	0.8888893	0.53761	1.53005
MaritalWidoworwidower	1.6382376	0.8286	3.65962
MaritalNevermarried	1.7109664	1.50564	1.94613

Invisible Underemployment Model Interpretations

The results in (table 7) shows that gender, age, education categories (post-secondary, college and post-primary vocational), residence, employment sector and marital status (never married category) were significant predictors of invisible underemployment.

The results of the binary logistic regression model (table 8) show that female employees are 3 times more likely to be invisibly underemployed as compared to the male counterparts. The people in the age category 30-34 years are 83% less likely to be invisibly underemployed compared to the people between 15-19 years. Similarly, the people in the age category 25-29 years are 77 % less likely to be invisibly underemployed compared to the people between 15-19 years. The people in the age category 25-29 years are 63 % less likely to be invisibly underemployed compared to the people between 15-19 years. The people with post-secondary (college and university) level of education are 87% less likely to be invisibly underemployed compared to the people with pre-primary level of education. The people with secondary level of education are 39% less likely to be invisibly underemployed compared to the people with pre-primary education. The people with post-vocational level of education are 42% less likely to be invisibly underemployed compared to those with pre-primary level of education. Primary level of education is not a significant predictor of invisible underemployment. The people in the informal sector are 2.6 times more invisibly underemployed compared to those in the public sector. The people in the private sector are 2.1 times more invisibly underemployed compared to those in the public sector. The people in the urban areas are 67% less likely to be invisibly underemployed compared to those in the rural areas.

5 Chapter 5: Conclusion and Recommendation

5.1 Conclusion

The study adopts the binary logistic regression approach to model the determinants of underemployment among the youths in Kenya. The model has focused on both visible and invisible underemployment. There is little understanding of youth underemployment as a significant aspect of labour underutilization. This study makes use of cross-sectional data (KIHBS 2015/16) for analysis to help fill this gap. Findings from the study show that the rate of visible underemployment among the youth is 11.9% while that of invisible underemployment is 75%. Model results reveal that gender, age, education (secondary category), residence and employment sector are significant determinants of visible underemployment among the youth. The model shows that females are 73% more likely to be visibly underemployed compared to their male counterparts. The youths aged 30-34 years and 25-29 years are less likely to be visibly underemployed than those aged 15-19 years. Results show more visible underemployment among youths in rural areas as compared to those in an urban setting. The youths working in the informal sector are 2.11 more likely to be visibly underemployed compared to those working in the public sector. The models also show that youths with secondary education are 37% less likely to be visibly underemployed as compared to those with pre-primary education.

On the invisible underemployment, the model results reveal that gender, age, education categories (post-secondary, college and post-primary vocational), residence, employment sector and marital status (never married category) are significant determinants of invisible underemployment. Females are more likely to be invisibly underemployed compared to males. The youths aged 25-29 years are less likely to be invisibly underemployed compared to the younger youths aged 15-19 years. The model results show that youths with post-secondary (college and university), secondary and vocational levels of education are less likely to be invisibly underemployed compared to those with the pre-primary level of education. The people in the informal and private formal sectors are more likely to be invisibly underemployed as compared to the youths working in the public sector. The results from the study show that youths in rural areas are more likely to be invisibly underemployed compared to those in the urban areas.

5.2 Recommendations

Key policy interventions should be implemented to help address the problem of underemployment among the youths in Kenya. The government should promote initiatives motivating the youths to seek higher education, enhancing their productivity through improved skills to address the underemployment problem. The policies promoting gender equality and fighting gender discrimination at workplaces is needed to reduce underemployment which is more prevalent among females. Policy interventions are needed to alleviate challenges towards the creation of formal employment opportunities for the youths. Stimulating industrial growth and creating productive and quality job opportunities in rural areas rather than only focusing on urban areas is crucial for adoption. The government should develop more efforts towards fighting the development of informal sectors. The government should ensure strict adherence and compliance to the minimum wage policy.

5.3 Areas for Further Research

The study has mainly focused on the determinants of the underemployment among youths in Kenya. There is need to use the future latest surveys to analyze the aspect of labour underutilization. Future studies can explore the consequences of underemployment to evaluate the impacts it can have in the labor market.

Bibliography

[1] Abonazel, M. R., & Ibrahim, M. G. (2018). On estimation methods for binary logistic regression model with missing values.*International Journal of Mathematics and Computational Science*, 4(3), 79-85.

[2] Acosta-Ballesteros, J., Osorno-del Rosal, M. D. P., & Rodríguez-Rodríguez, O. M.(2018). Underemployment and employment among young workers and the business cycle in Spain: the importance of education level and specialisation. *Journal of Education and Work*,Volume 31(1)28-46.

[3] Becker, G, S.(1962). 'Investing in Human Capital: A Theoretical Analysis',*The Journal of Political Economy*, Vol 70, No 5.

[4] Görg, H., & Strobl, E. (2003). The incidence of visible underemployment: Evidence for Trinidad and Tobago.*Journal of Development Studies*, 39(3), 81-100.

[5] Hauser, P. M. (1974). "The Measurement of Labour Utilization."*The Malayan Economic Review*19:1–17

[6] Hyefouais, N. (2016).Characteristics and Determinants of Underemployment in Cameroon. *African Economic Research Consortium*

[7] ILO (n.d.).Indicator description: Time-related underemployment rate.*Retrieved March 12, 2021, from*,https://ilostat.ilo.org/resources/concepts-and-definitions/description-time-related-underemployment.

[8] ILO. (2018). Avoiding unemployment is not enough: An analysis of other forms of labor underutilization,*Retrieved March 12, 2021, from* https://www.ilo.org/global/statistics-and-databases publications

[9] ILOSTAT. (2021). Unemployment and labour underutilization. https://ilostat.ilo.org/topics/unemployment-and-labour-underutilization/.

[10] Kenya National Bureau of Statistics. (2020). Economic Survey. (Nairobi, Kenya National Bureau of Statistics).

[11] KNBS. (2018), Labour Force Basic Report: 2015/16 Kenya Integrated Household Budget Survey, Nairobi: *Government Printer*.

[12] Kiiru, J., Onsomu, E., & Wamalwa, F. (2009). Education, training and youth unemployment in Kenya (No. 26).*GDN Library Working Paper Series*.

[13] Midi, H., Sarkar, S. K., Rana, S. (2010). Collinearity diagnostics of binary logistic regression model.*Journal of Interdisciplinary Mathematics*, 13(3), 253-267.

[14] Mukherjee, S., Coondoo, D., & Chakraborty, I. (2018). Underemployment in India: Measurement and Analysis (No. id: 12681).

[15] Munga, B., Nyanjom, O., Onsomu, E., & Mwabu, G. (2012). YOUTH UNEMPLOYMENT, UNDER EMPLOYMENT AND DECENT WORK IN KENYA.

[16] Nick, T. G., Campbell, K. M. (2007). Logistic regression. Topics in biostatistics, 273-301.

[17] Rana, S. O. H. E. L., Midi, H. A. B. S. H. A. H., Sarkar, S. K. (2010). Validation and performance analysis of binary logistic regression model. In Proceedings of the WSEAS International Conference on Environmental, Medicine and Health Sciences (pp. 23-25).

[18] Ruiz-Quintanilla, S. A., & Claes, R. (1996). Determinants of underemployment of young adults: A multi-country study. *ILR Review*,, 49(3), 424-438.

[19] Vuluku, G., Wambugu, A., & Moyi, E. (2013). Unemployment and underemployment in Kenya: A gender gap analysis. *Economics*,2(2), 7-16.

[20] Wilkins, R. (2006). Personal and job characteristics associated with underemployment. *Australian Journal of Labour Economics*,9(4), 371-393.

A Appendix1

Table 9. Variables symbol description

Variable Symbol	Variable Name
X_1	GenderFemale
X_2	Age20-24years
X_3	Age25-29years
X_4	Age30-34years
X_5	EducationPrimary
X_6	EducationPost-primary,vocational
X_7	EducationSecondary
X_8	EducationPostsecondary(college,university)
X_9	EducationOther
X_{10}	EmploymentsectorPrivateformal
X_{11}	EmploymentsectorInformal
X_{12}	residenceUrban
X_{13}	MaritalPolygamousmarried
X_{14}	MaritalLivingtogether
X_{15}	MaritalSeparated
X_{16}	MaritalDivorced
X_{17}	MaritalWidoworwidower
X_{18}	MaritalNevermarried