LIFE EXPECTANCY IN AFRICAN COUNTRIES: A PANEL DATA ANALYSIS

By Group 4

Animesh Gupta	2019B3AA0588H
Chinmay Goyal	2019B3AA1290H
Harsh Vardhan Gupta	2019B3A70630H
Hitesh Garg	2019B3A70466H
Sujay Nigam	2019B3AA1267H
Aryan Kapadia	2019B3A70412H
Dhruv Gupta	2019B3A70487H
Anand	2019B3A70613H

APRIL 17, 2022

ECON F342: APPLIED ECONOMETRICS

BITS Pilani, Hyderabad Campus

Table of Contents

Section I: Introduction

Life expectancy at birth is an important indicator of the overall development and health of the population of a country. Over the years, life expectancy across the world has been increasing in general. With a steady increase in the per capita incomes and budget allocated to health facilities, education, and food supplies, developing countries throughout the globe have seen a significant uptick in life expectancy over the past half century.

Apart from income, socio-economic factors like literacy, medical services, sanitation facilities also play a crucial role in determining the health of citizens. Literacy leads to better knowledge about healthy diets and nutritious food products. An educated woman or mother understands how to take the best care to maintain health and nutrition within the family. Moreover, it also tends to better household practices and personal hygiene habits.

Immunization prevents deaths every year in all age groups from diseases like diphtheria, measles, influenza, and polio. It has been one of the most successful and cost-effective public health interventions. In 2018, around 116.3 million (about 86%) children under the age of one year worldwide received three doses of diphtheria-tetanus-pertussis (DTP3) vaccine. These children are protected against infectious diseases that cause serious illness or disability.

Measles is a highly contagious disease caused by a virus. Accelerated immunization activities have had a major impact on reducing measles deaths. Global measles deaths have decreased by 73% from almost 536000 deaths in 2000 to 142000 in 2018. Based on latest data by the WHO, almost 85% of infants over the globe have received three doses of polio vaccine. Instances of wild poliovirus is restricted to only three countries in the world: Afghanistan, Pakistan, and Nigeria. Vaccinations of infants and children at a large scale has been a boon for developing countries.

Over the years, there have been many studies by scientists and academicians on the determinants of life expectancy using a variety of models like probit model, multivariate cross-sectional data analysis, OLS models and multiple regression. Life expectancy is affected by various factors which include but are not limited to income, educational expenditure, expenditure on health facilities, vaccination status. This research paper uses a multiple regression model with one independent variable and 16 dependent variables, with an additional 17 dummy variables to capture the effect of time. The relationship between the variables is linear.

Design: The research includes the formulation of a system explaining the linear relationship between X and Y which are price and other variables such as adult mortality rate, BMI, alcohol consumption, levels of vaccinations, etc.

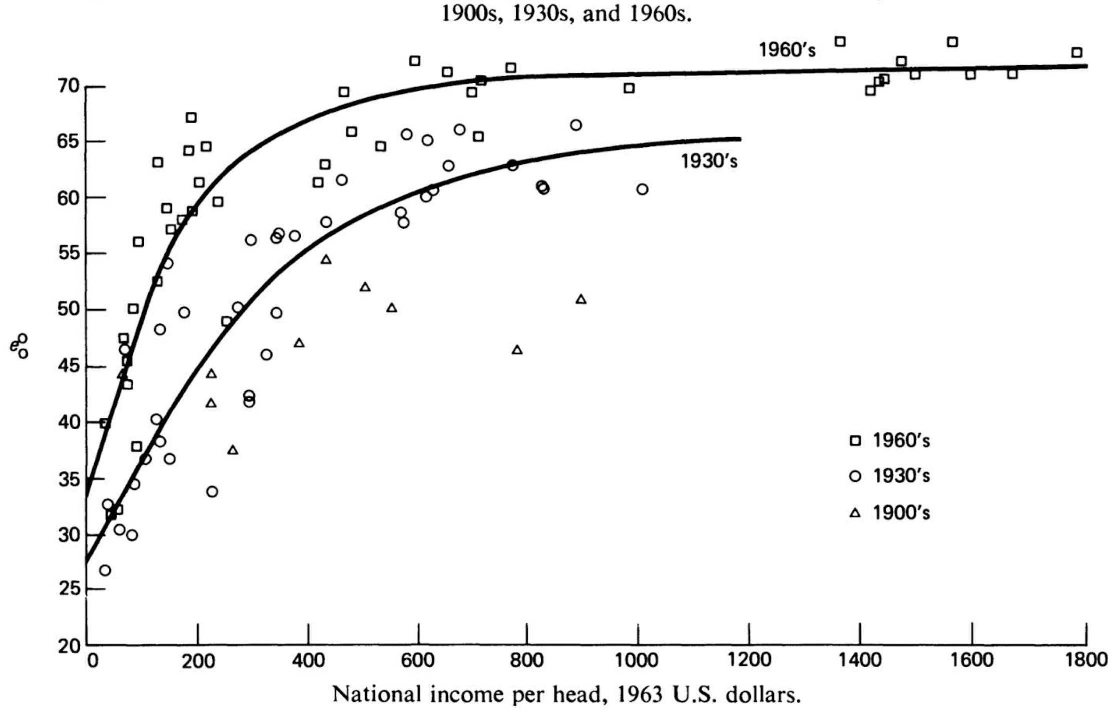
Confirm: The analysis seeks to see which factors significantly affect the life expectancy at birth in African countries. It takes advantage of the nature of panel data to remove the country fixed effect from the estimation process and aims to provide a more accurate result than a standard cross-sectional analysis.

The data for this research has been sourced from World Health Organization and it consists of data from 46 countries over 17 continuous years. The organization of this paper is such that Section II is on related work, Section III discusses the methodology and Section IV presents the results and in-depth discussion. Section V concludes the paper and offers directions for future research.

Section II: Literature Review

During the late 1950s, it became evident that life expectancy rates rose rapidly across the world, including the developing nations. The pioneer and the most important research on the decline of mortality rate and level of economic development was conducted by Preston in 1975. He generated scatter plots between average life expectancy and national income per head for nations in the 1900s, 1930s and 1960s.

Scatter-diagram of relations between life expectancy at birth (e_0^0) and national income per head for nations in the 1900s, 1930s, and 1960s.



Source: Changing relation between mortality rate and level of economic development by Preston(1975)

This plot, popularly known as the Preston curve, is one of the most crucial curves in population sciences. Preston drew many conclusions from this plot:

1. Life expectancy grew rapidly over the decade, irrespective of the income levels of the nations. This is represented by the upward shift of the Preston curve with respect to time.

2. The curve rises rapidly at lower income levels but flattens out as the income per head rises for the countries. This shows that an increase in income is strongly associated with an increase in life expectancy for poorer countries. However, this is not the case for high-income countries.

3. Income is not the only determinant of life expectancy. The effect of an increase in income between 1938 to 1963 would have generated a rise of only 2.5 years in life expectancy. However, the actual plot depicted a rise of 12.2 years. As a result, Preston concluded that about 75-90% growth in life expectancy was associated with exogenous factors while only 10-25% growth is attributed to rising income levels.

4. These exogenous factors were identified to be antibiotics, vaccines and sulphonamides in developed nations and sanitation, health education, maternal and child healthcare and insect control in less-developed nations.

A lot of studies have been conducted by various researchers to determine the socio-economic factors that affect life expectancy in countries across the world. However, they have seemingly generated conflicting results on the impact of the level of income, the magnitude of health expenditure, education, and the better quality of water and sanitation on health in the countries.

Hasan et al (2016) stated that it is assumed that the higher per capita income of the country is associated with higher life expectancy since a higher income enables citizens to avail better medical facilities, nutrition, and sanitation which in turn aids in fostering better health. However, this positive correlation is not prominent in the case of countries with higher levels of GDP per capita. There are striking counterexamples to the above statement. Developing countries such as Cuba and Costa Rica despite lower per capita GDP, seemingly have better health outcomes. At the same time, the USA with the highest per capita GDP, ranks 12th against 13 of its counterparts as per **Starfield's report (2000).** This suggests that income may not the single most important factor in determining life expectancy. In addition to this, **Preston (2007)** also determined that for low-income countries, there existed a strong positive correlation between income per capita and life expectancy. However, as the per capita GDP rose, the relationship tends to become weaker and weaker. Therefore, the relation between income and health depends on the level of development of the nation as well.

The role of health expenditure is also subjected to a lot of controversies. **Burnside and Dollar (1998)** in their research on the effect of foreign aid on infant mortality concluded that there exists no significant relationship between health expenditure and decline in infant mortality rate, especially for lower-income countries. However, **Anyanwu and Erihijakpor (2007)** in their study on health expenditure and health outcomes in 47 African countries between 1999 and 2004 found that health expenditures had a significant impact on reducing infant and under-5 mortality rates. Additionally, they also found that HIV significantly impacted negative health outcomes while a higher number of physicians and female literacy seem to have a positive impact. Along similar lines, **Mccarthy's (2000)** study on African countries also found that although health expenditure is not an accurate determinant of health outcomes, it is negatively correlated with infant mortality rates. He found that the most important variables associated with good health outcomes were access to proper health services, clean water and sanitation and education, especially for women.

Kabir (2008) attempted to determine the socio-economic predictors of life expectancy for 91 countries by utilizing multi regression and probit frameworks. He divided the countries into three groups: low, medium and high life expectancy. He found that almost all the relevant explanatory variables turned out to be statistically insignificant which indicated that per capita income, health expenditure, education, access to clean water and urbanization, although crucial for social development, are not always influential in determining the life expectancy, particularly in the developing nations. He also provides policy suggestions for developing nations: programs to increase the number of available doctors, reduce adult illiteracy and combat undernutrition to improve the life expectancies of their citizens.

Hassan et al (2016) sought to refine Kabir's work by using panel data analysis to determine social and environmental predictors of life expectancy in developing nations. The study investigated the association of life expectancy with GDP, health expenditure, education and improved water coverage and sanitation facilities. It found that life expectancy was positively correlated with all the above-mentioned variables. However, using the fixed effects model, only two of these variables were statistically significant: education index at a 1% level and GDP at a 5% level.

Section III: Data and Methodology

Description of chosen dataset

The data was taken to analyze the Life expectancy rate over the years in African countries. The chosen dataset contains the following parameters:

- Country: Name of the country
- Country_code: unique code for the individual countries
- Region: continent to which the country belongs
- Year: time for which the data is provided
- Life_expect: Life expectancy of a person at the time of birth
- Life_exp60: Life expectancy of a person from the Age of 60
- Adult_mortality: Adult Mortality Rates of both sexes (probability of dying between 15 and 60 years per 1000 population)
- Infant_mort: probability of dying upto age of 1 year
- age1-4mort: Death rate between ages 1 and 4
- alcohol: consumption of alcohol per capita for age group of greater than 15 in liters
- BMI: mean BMI for age group of greater than 18
- age5-19thinnes: Prevalence of thinness among children and adolescents, BMI < (median - 2 s.d.) (crude estimate) (%)
- age5-19 obesity: Prevalence of obesity among children and adolescents, BMI > (median + 2 s.d.) (crude estimate) (%)
- hepatitis: % of children below 1 year of age group getting Hepatitis B immunization
- measles: % of children below 1 year of age group getting Measles-containing-vaccine first-dose immunization
- polio: % of children below 1 year of age group getting Polio immunization coverage
- diphtheria: % of children below 1 year of age group getting Diphtheria tetanus toxoid and pertussis immunization
- basic_water: % of population using at least basic water services

- Doctors: total doctors available for per 10,000 of population
- GNI_capita: gross national income per capita
- gghe-d: Domestic general government health expenditure (GGHE-D) as percentage of gross domestic product (GDP)
- che_gdp: Current health expenditure (CHE) as percentage of gross domestic product (GDP) (%)
- une_pop: Population in thousands
- une_infant: infant mortality rate per 1000 births
- une_life: life expectancy of a person at birth
- une_hiv: prevalence of HIV in age group of 15-49
- une_gni: gross national income per capita
- une_poverty: Poverty headcount ratio at $1.90 a day (PPP) (% of population)
- une_edu_spend: total spending of government on education of % of GDP
- une_literacy: both sexes literacy rate in age group of 15+ years
- une_school: both sexes mean year of schooling in age group of 25+ years

Econometric Model

$$\mathbf{Y_{life\_expect} = \beta_0 + \beta_1 adult\_mortality + \beta_2 infant\_mort + \beta_3 age1\text{-}4mort + \beta_4 alcohol + \beta_5 bmi + \beta_6 measles + \beta_7 polio + \beta_8 diptheria + \beta_9 hepatitis + \beta_{10} basic\_water + \beta_{11} gni\_capita + \beta_{12} gghe\text{-}d + \beta_{13} che\_gdp + \beta_{14} une\_pop + \beta_{15} une\_hiv + \beta_{16} une\_edu\_spend + \gamma_{2001} + \gamma_{2002} + \gamma_{2003} + \gamma_{2004} + \gamma_{2005} + \gamma_{2006} + \gamma_{2007} + \gamma_{2008} + \gamma_{2009} + \gamma_{2010} + \gamma_{2011} + \gamma_{2012} + \gamma_{2013} + \gamma_{2014} + \gamma_{2015} + \gamma_{2016}}$$

Here the aim is to analyze how life expectancy of a person depends on different factors. Some of the factors are directly related to life expectancy like adult mortality, infant mortality, mortality between age groups of 1 to 4, BMI of the adult population. However, it might also depend on alcohol consumption of the population, basic vaccination drives for diseases like measles, polio, diphtheria. Other factors such as the quality of the water one consumes, gross national income, domestic general health expenditure, current health expenditure, expenditure on education are also relevant. In accordance with this view, a linear model has been estimated to predict life expectancy. Data of African countries from 2000 to 2016 was taken to analyze the effect. Here Y (the dependent variable) is the life expectancy of a person. A set of independent variables have been used to control for various observables. The variables and the reasoning are given below:

S. No.	Variable chosen	Description	Justification
1.	**life_expect**	Dependent variable: life expectancy of a person	To study the effect of different factors that affect the life expectancy of a person
2.	**adult_mortality**	Both sexes probability of dying in between 15 to 60 years per 1000 of population	Life expectancy will depend on the total no. of people dying in between the age group of 15 to 60.

3.	**infant_mort**	Probability of dying under the age of one	It is expected that the life expectancy of the population will be more if the infant mortality is low.
4.	**age1-4mort**	Death rate between 1 to 4 years	Life expectancy will be more if children dying in age group of 1 to 4 will be low
5.	**alcohol**	Alcohol consumption for age group of 15+ in liters	Overconsumption of Alcohol can be fatal, so life expectancy will be low if population alcohol consumption is high
6.	**Bmi**	Average BMI for 18+ population	BMI is a reliable indicator of body fatness for most people, so a bad BMI in a country can lead to a lower population life expectancy.
7.	**measles**	% of children below 1 year getting measles vaccine	Higher vaccination will lead to better health leading to higher life expectancy
8.	**polio**	% of children below 1 year getting polio vaccine	Higher vaccination will lead to better health leading to higher life expectancy
9.	**diphtheria**	% of children below 1 year getting diphtheria vaccine	Higher vaccination will lead to better health leading to higher life expectancy
10.	**hepatitis**	% of children below 1 year getting hepatitis vaccine	Higher vaccination will lead to better health leading to higher life expectancy
11.	**basic_water**	% of population using basic water services	Clean water lowers the chance of getting affected by water borne diseases leading to a higher life expectancy
12.	**gni_capita**	Gross national income per capita	Higher GNI will lead to a better living style, better health facilities, so life expectancy will be higher
13.	**gghe-d**	Domestic general government Health expenditure	More spending on Health will lead to better health of the population, so life expectancy will be higher
14.	**che_gdp**	Current Health expenditure	Higher the current health expenditure, better will be the current health of the population, higher will be the life expectancy of the population

15.	**une_pop**	Population (thousands)	Higher the population, lesser will be health infrastructure available for the individual, so life expectancy will be lower
16.	**une_hiv**	Prevalence of HIV in age group of 15-49	Higher the population infected by the HIV, more will be mortality, leading to a lower Life expectancy
17.	**une_edu_spend**	Total government spending on education	As the government increases its spending on education, people will be more aware of the diseases and health, leading to higher life expectancy

Graph Matrix of the Regressors

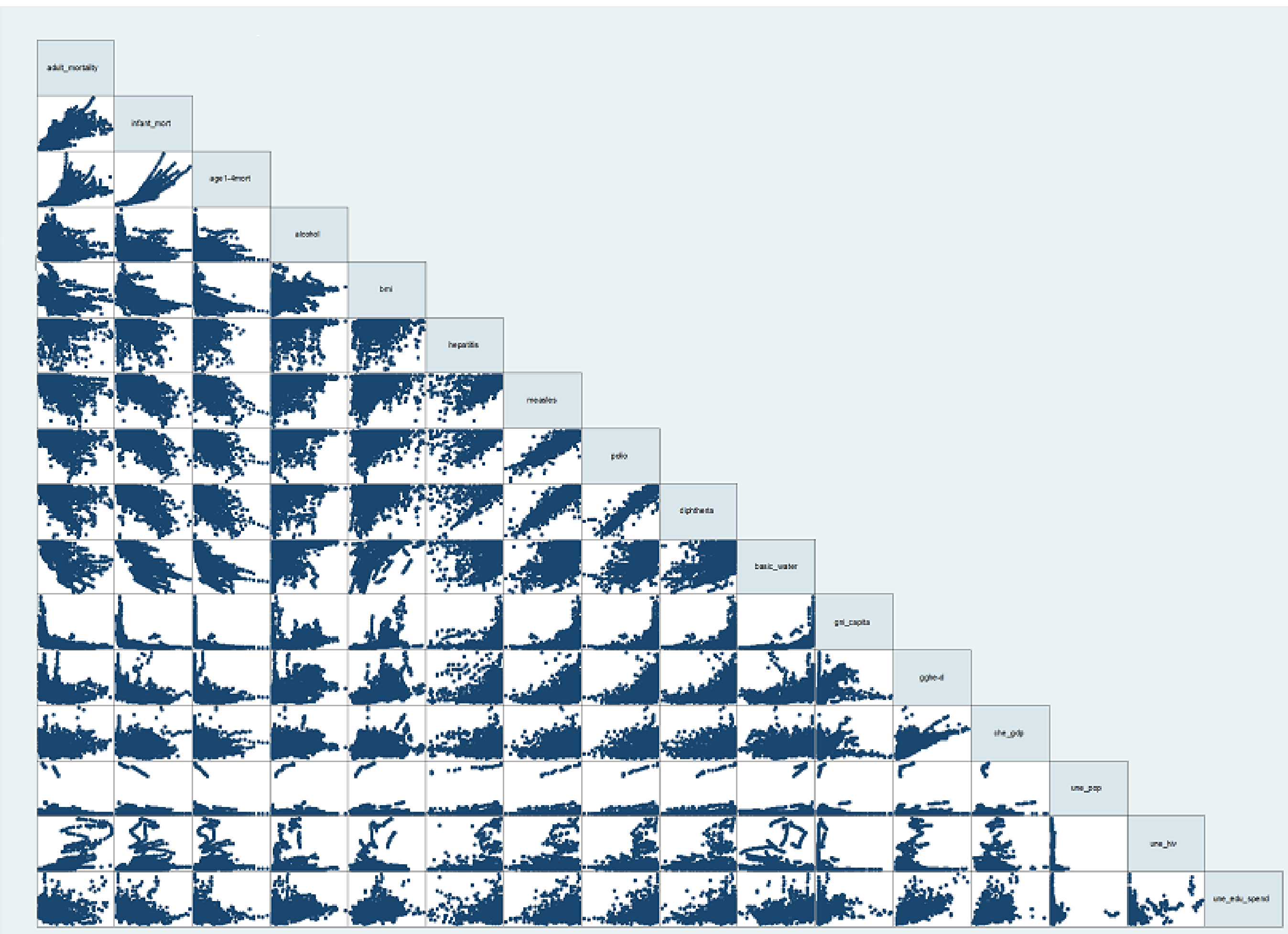


Figure 1: Graph Matrix of the Regressors

The correlation between the explanatory variables is as shown above in the graph matrix. Only a few of the variables are very strongly correlated with each other. Some notable observations from this graph matrix are:

1. Adult mortality is positively correlated with infant mortality which can be explained by the fact that the countries with higher infant mortality would have lesser healthcare facilities and other negative factors. This also translates into bad healthcare for adults leading to higher adult mortality.
2. Infant mortality and age 1-4 mortality has strong positive correlation which is intuitive as the countries with higher infant mortality would also have higher mortality rates in age 1-4.
3. Vaccinations like measles, polio and diphtheria have strong positive correlation with each other suggesting that the countries that have undergone vaccination programmes have covered all the necessary vaccinations.

Model Specifications

The observations are from a random sample as collected by WHO from a given population. Thus, both fixed effects and random effects estimations were carried out. **Durbin–Wu–Hausman test** results indicated significant differences in the coefficients i.e., the null hypothesis was rejected. Hence, a fixed effect model was chosen for the estimation in the paper.

Steps to check the validity of the model:

1. Modified Wald test to test for group-wise heteroskedasticity.
2. Shapiro-Wilk test to test for normality of residuals.
3. VIF test and Pearson Correlation matrix to check for multicollinearity.

Section IV: Results and Discussion

Summary statistics of the regression model

- Most of the explanatory variables are significant at 10% level of significance except adult mortality rate, infant mortality rate and age 1-4 mortality rate, which are significant at less than 0.1% significance level.

- Measles, Domestic general government Health expenditure (gghed) are also significant at 10% and 5% significance level respectively. All the year-dummy variables are also significant at 0.1% significance level.

- R-squared is 0.9792 which means 97.92% of the dependent variable is explained by the model. Also, F-statistic for the joint significance test of the coefficients is very high (582.24) which means that all the regression coefficients are jointly significant. However, since many of the coefficients are insignificant, the large R-squared value can also be explained by a high level of multicollinearity between the variables.

Interpretation of the model

Fixed-effects (within) regression — Number of obs = 901
Group variable: country_num — Number of groups = 112

R-sq: within	= 0.9792	Obs per group: min =	1
between	= 0.9314	avg =	8.0
overall	= 0.9269	max =	14
		F(29,111) =	582.24
corr(u_i, Xb)	= 0.7242	Prob > F =	0.0000

(Std. Err. adjusted for 112 clusters in country_num)

life_expect	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
adult_mortality	-.0386272	.0009241	-41.80	0.000	-.0404583	-.0367961
infant_mort	-43.36339	10.78051	-4.02	0.000	-64.72569	-22.00108
age14mort	-150.7627	28.04331	-5.38	0.000	-206.3324	-95.19299
alcohol	.0132166	.0259382	0.51	0.611	-.0381818	.0646149
bmi	-.2390598	.1794094	-1.33	0.185	-.5945715	.1164518
hepatitis	-.0014477	.0015189	-0.95	0.343	-.0044576	.0015621
measles	.009859	.0051821	1.90	0.060	-.0004097	.0201277
polio	.0006673	.0037599	0.18	0.859	-.0067831	.0081177
diphtheria	-.0041831	.0044243	-0.95	0.346	-.0129502	.004584
basic_water	.0037547	.0130426	0.29	0.774	-.0220901	.0295995
gni_capita	.0000126	.0000177	0.71	0.478	-.0000225	.0000478
gghed	-.0888739	.0408411	-2.18	0.032	-.1698033	-.0079445
che_gdp	-.0431496	.0273531	-1.58	0.118	-.0973516	.0110523
une_pop	-3.90e-06	7.70e-06	-0.51	0.614	-.0000191	.0000114
une_hiv	-.0538518	.0503927	-1.07	0.288	-.1537083	.0460047
une_edu_spend	.0391444	.0252535	1.55	0.124	-.0108972	.0891859
year						
2001	.2473638	.0589211	4.20	0.000	.1306078	.3641199
2002	.320939	.0721042	4.45	0.000	.1780598	.4638182
2003	.3930155	.0878222	4.48	0.000	.21899	.5670411
2004	.5580058	.1130039	4.94	0.000	.3340809	.7819307
2005	.7210019	.1340596	5.38	0.000	.4553538	.98665
2006	.8958805	.1592313	5.63	0.000	.5803531	1.211408
2007	1.04031	.1869214	5.57	0.000	.6699126	1.410707
2008	1.157641	.2124529	5.45	0.000	.7366516	1.578631
2009	1.313367	.2330327	5.64	0.000	.851597	1.775137
2010	1.431212	.253498	5.65	0.000	.9288891	1.933535
2011	1.612189	.2762113	5.84	0.000	1.064858	2.15952
2012	1.735708	.2893191	6.00	0.000	1.162403	2.309014
2013	1.911012	.311707	6.13	0.000	1.293344	2.52868
_cons	82.82516	4.360157	19.00	0.000	74.18522	91.4651
sigma_u	3.4885178					
sigma_e	.22756711					
rho	.99576266	(fraction of variance due to u_i)				

- **Intercept:** The intercept captures the case where all the dummy variables take the value 0. Thus, it represents the case when the year is 2000. All the values of the coefficients and hence their ceteris paribus effects will be with respect to this base category.

- **adult_mortality**: If the probability of both sexes dying between 15 to 60 years per 1000 of the population increases by 1 unit, then life expectancy will decrease by 0.0386272 units. The negative relationship is intuitive because as mortality increases, life expectancy decreases.

- **infant_mort**: If the probability of dying under the age of one increases by 1 unit, then life expectancy decreases by 43.36339 units.

- **age14mort**: If the probability of dying between the age of 1 to 4 years increases by 0.1 unit, then life expectancy decreases by 15.7627. There is a greater impact on life expectancy by age14nmort than adult_mortality because the death of a child in first four years of his/her life signifies a greater lack of nutritional and health services in the country than the death of an adult.

- **alcohol:** If the alcohol consumption for age group of 15+ in liters increases by 1 unit then, there is an increase in life expectancy by 0.0132166. This is counter-intuitive going to the fact that excessive alcohol is linked to liver damage and lower life span. Also, the p-value is around 0.611 which means it is statistically insignificant. This result can be safely ignored.

- **bmi (Body Mass Index):** If the average BMI for the 18+ population increases by 1 unit then, life expectancy decreases by 0.2390598 units. This is because there is a range of healthy BMI, if it's greater than that range (which means the person is obese) then it will affect life expectancy in a negative way. This variable is not significant at 10% significance level.

- **hepatitis:** If the percentage of children below 1-year getting hepatitis vaccine increases by 1 unit, then life expectancy will decrease by 0.0014477 units. This result is also counter-intuitive because if there is an increase in the percentage of children getting hepatitis then it should increase the life expectancy of those children. However, the p-value of this variable which is 0.343 it implies its insignificant. This result can also be safely ignored.

- **measles:** If the percentage of children below 1-year getting the measles vaccine increases by 1 unit, then life expectancy will increase by 0.009859 units. This result is quite intuitive and is also significant at a 10% significance level.

- **polio:** If the percentage of children below 1-year getting the polio vaccine increases by 1 unit, then life expectancy will increase by 0.0006673 units. This result is also quite intuitive but is not significant.

- **diphtheria:** If the percentage of children below 1-year getting the diphtheria vaccine increases by 1 unit, then life expectancy will decrease by 0.00418 units. This result is also quite counter intuitive but is not significant, hence it can be ignored.

- **basic_water:** If the percentage of population using water for basic services increases by 1% then life expectancy increases by 0.00375 units. Even though this result is intuitive however it is insignificant because of high p-value.

- **gni_capita**: if the GNI per capita increase by 1 unit, then life expectancy will increase by 0.0000126 units. This is also an intuitive result however it is insignificant.

- **gghe-d:** If the government's expenditure on health increases by 1 unit, then the life expectancy will surprisingly decrease by 0.0888 units. The result is counterintuitive because when expenditure on health increases, the medical infrastructure of the country improves and increases the life expectancy. However, it is showing an opposite effect here can be explained by the fact that sub-Saharan African countries suffer from high corruption and inaccessibility of the medical facilities even though the government spends money in right places **(Fatma E. Marouf (2010), Gauthier et al. (2007), Zekeri Momoh (2015))**

- **che_gdp:** If the private expenditure as a % of GDP on healthcare increases by 1 unit then life expectancy will decrease by 0.0431 units which is an insignificant and counterintuitive result and hence can be ignored.

- **une_pop:** If the population of the country increase by 1 unit then the life expectancy will decrease by $-3.90*10^{-6}$ units. It's p-value is very large (0.614), hence it's an insignificant but intuitive result.

- **une_hiv:** If the percentage of population between the ages 15-49 having HIV increases by 1 unit, then life expectancy will decrease by 0.0538 units. The result is highly intuitive however it has a very high p-value and thus it is insignificant.

- **une_edu_spend:** If the government's expenditure on education as a percentage of GDP increases by 1 unit, then life expectancy will increase by 0.0391 units which is a highly intuitive result. However, it also has a high p-value of 0.124 and is thus insignificant.

Checking for the assumption of OLS

1. **Linear in parameters**: Since the assumed model has only a single ßi term for every explanatory variable/dummy variable, there is no product division or exponential form of the coefficient is present. Hence, it is linear in parameters.

2. **Random sampling of the data**: The data was collected from Kaggle which was sourced by W.H.O. (World Health Organization) and World Bank databases. It covers a wide range of factors like HIV and various other diseases, literacy, alcohol, mortality, and spending on education in Sub-Saharan African countries. Hence, it can be concluded that the data were randomly sampled.

3. **Zero Conditional Mean** (E(ui|xi) = 0): This will be shown using kernel density plots.

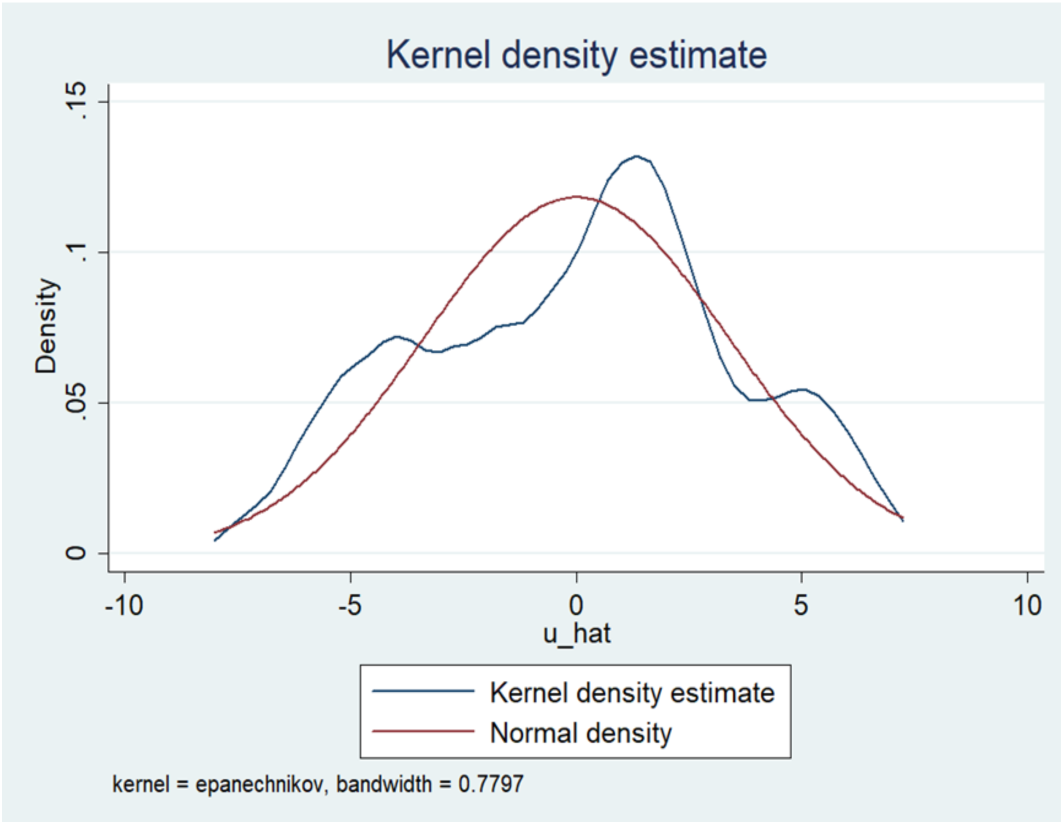


Figure 2: K-density plot of the residuals

As shown in Figure 2, the plot of the residuals is approximately centered at mean = 0 and is constant, hence it can be concluded that it has a zero conditional mean.

4. **Homoscedasticity of error term**: The error term should have a constant variance over the interval of values of xi. This can be checked via:

a) A scatter plot of residuals vs true values of dependent variable:

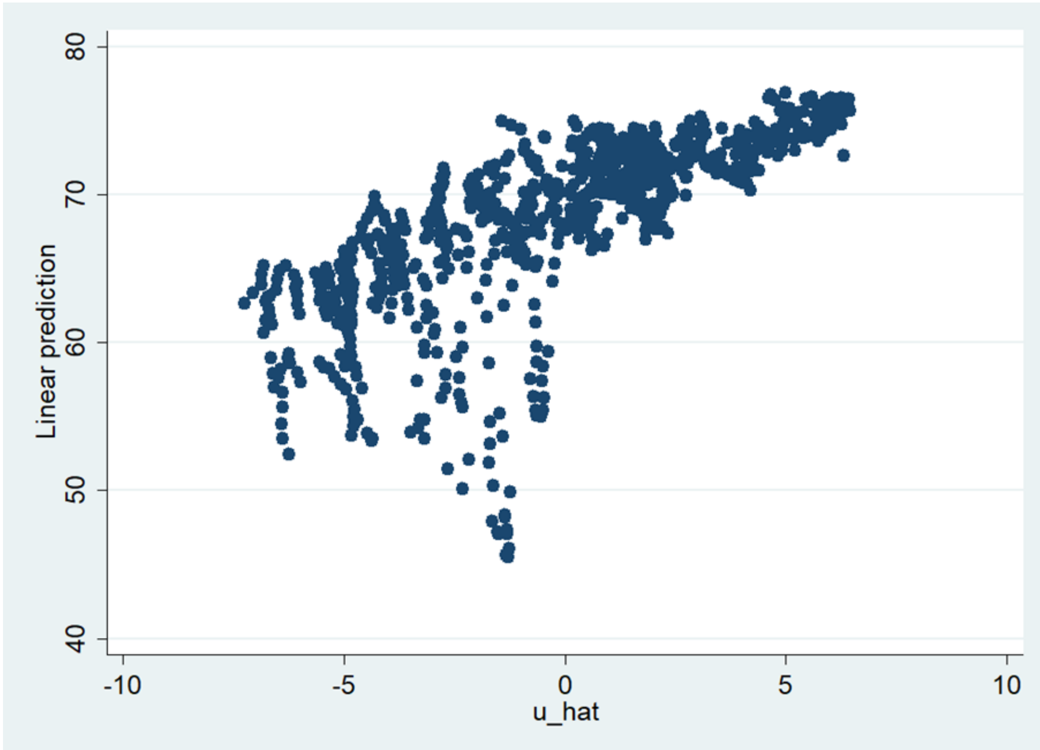


Figure 3: Scatterplot of y_hat vs u_hat

A funnel shape in the scatter plot signifies that variation is greater for lesser values of dependent variable and decreases for higher values, as shown in Figure 3. Hence, the model suffers from heteroscedasticity.

b) Modified Wald Test for group wise heteroscedasticity in fixed effect regression model:

```
. xttest3

Modified Wald test for groupwise heteroskedasticity
in fixed effect regression model

H0: sigma(i)^2 = sigma^2 for all i

chi2 (112)  =    7.8e+24
Prob>chi2 =       0.0000
```

Figure 4: Results of Heteroskedasticity Test

Ho: Homoskedasticity

Ha: Heteroskedasticity

As shown in Figure 4, the p-value in the test results is very close to 0, the null hypothesis of the Modified Wald test is rejected. Thus, heteroscedasticity exists in the model.

5. **Normality of Residuals**: It means that the population error is independent of independent variables and normally distributed with zero mean and variance sigma squared. This is one of the key assumptions of multiple regression model. The Shapiro-Wilk Test can be used to verify this:

```
. swilk e
```

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
e	901	0.88995	63.122	10.221	0.00000

Figure 5: Results of test for normality

Ho: The error term is normally distributed
Ha: The error term is not normally distributed

Since the p-value of the test result is very close to 0, the null hypothesis of Shapiro-Wilk Test is rejected. Thus, error term is not normally distributed.

6. **Multicollinearity**: Multicollinearity occurs when the independent variables are correlated with each other in the model. In OLS estimation of coefficients, there should be no perfect correlation between two independent variables, partial correlations are allowed. To verify this the Variance Inflation Factor (VIF) test is performed.

```
58 . vif, uncentered
```

Variable	VIF	1/VIF
adult_mort~y	37.01	0.027018
infant_mort	29.71	0.033662
age14mort	12.54	0.079771
alcohol	5.95	0.168093
bmi	234.66	0.004262
hepatitis	61.10	0.016366
measles	347.06	0.002881
polio	827.67	0.001208
diphtheria	802.70	0.001246
basic_water	102.87	0.009721
gni_capita	4.04	0.247637
gghed	19.90	0.050250
che_gdp	20.29	0.049276
une_pop	1.68	0.594446
une_hiv	3.63	0.275456
une_edu_sp~d	12.46	0.080228
year		
2001	2.13	0.470190
2002	2.32	0.431805
2003	2.39	0.419044
2004	2.80	0.356951
2005	2.69	0.371420
2006	2.74	0.364850
2007	2.89	0.345783
2008	3.16	0.316523
2009	3.15	0.317344
2010	3.35	0.298745
2011	3.26	0.306902
2012	3.25	0.307473
2013	3.21	0.311179
Mean VIF	88.30	

Figure 6: Results of VIF test

Rule of thumb for the VIF test is that if VIF > 10, then the model suffers from multicollinearity.

The above table shows the VIF values for all the explanatory variables. All the variables except alcohol, gni_capita, une_pop, une_hiv and all year-dummy variables have VIF value < 10 hence, those variables don't have multicollinearity among them.

The high VIF value in factors like diphtheria, polio, measles, hepatitis, and BMI is because they all are highly dependent on availability of water which is a basic necessity and is denoted by basic_water. All the mortality factors are also interrelated with each other hence, they have VIF>10.

Gghe-d (Government Health Expenditure as a % of GDP), che_gdp (Private Current Health Expenditure as a % of GDP) and une_edu_spend (Total government spending on education) are correlated because an individual's private expenditure on health highly depends on the amount the government is spending on health. Similarly, the government's expenditure on education is also correlated with their expenditure on health.

It has been reported that the VIF function in Stata (as used for this research) gives abnormally large values for fixed effect models. To verify the inferences of multicollinearity, a Pearson Correlation matrix was also calculated as shown below.

	adult_~y	infant~t	age14m~t	alcohol	bmi	hepati~s	measles
adult_mort~y	1.0000						
infant_mort	0.8132*	1.0000					
	0.0000						
age14mort	0.7562*	0.9090*	1.0000				
	0.0000	0.0000					
alcohol	-0.2465*	-0.4446*	-0.3057*	1.0000			
	0.0000	0.0000	0.0000				
bmi	-0.5171*	-0.6419*	-0.6152*	0.2728*	1.0000		
	0.0000	0.0000	0.0000	0.0000			
hepatitis	-0.2769*	-0.4257*	-0.3889*	0.1190*	0.2916*	1.0000	
	0.0000	0.0000	0.0000	0.0000	0.0000		
measles	-0.5262*	-0.7202*	-0.6986*	0.2917*	0.4688*	0.6803*	1.0000
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
polio	-0.5380*	-0.7286*	-0.7008*	0.2898*	0.4413*	0.6932*	0.9242*
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
diphtheria	-0.5300*	-0.7207*	-0.6941*	0.2943*	0.4462*	0.7221*	0.9232*
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
basic_water	-0.7327*	-0.8515*	-0.8023*	0.4077*	0.6733*	0.3656*	0.6599*
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
gni_capita	-0.5277*	-0.5395*	-0.4178*	0.3050*	0.4108*	0.1860*	0.3436*
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
gghed	-0.4858*	-0.5768*	-0.4474*	0.5296*	0.4617*	0.1646*	0.3970*
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
che_gdp	-0.1986*	-0.2384*	-0.1623*	0.3722*	0.2480*	0.0692*	0.2038*
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0006	0.0000
une_pop	-0.0615*	-0.0040	-0.0330	-0.0378*	-0.1662*	-0.1205*	-0.0158
	0.0006	0.8260	0.0677	0.0376	0.0000	0.0000	0.3812
une_hiv	0.6806*	0.3416*	0.2920*	-0.0383	-0.1547*	-0.0476*	-0.1108*
	0.0000	0.0000	0.0000	0.0645	0.0000	0.0388	0.0000
une_edu_sp~d	-0.1471*	-0.3187*	-0.2830*	0.2160*	0.3124*	0.1498*	0.2701*
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

	polio	diphth~a	basic_~r	gni_ca~a	gghed	che_gdp	une_pop
polio	1.0000						
diphtheria	0.9645*	1.0000					
	0.0000						
basic_water	0.6621*	0.6517*	1.0000				
	0.0000	0.0000					
gni_capita	0.3569*	0.3557*	0.5210*	1.0000			
	0.0000	0.0000	0.0000				
gghed	0.4127*	0.4096*	0.5100*	0.4248*	1.0000		
	0.0000	0.0000	0.0000	0.0000			
che_gdp	0.2239*	0.2242*	0.2054*	0.1274*	0.6899*	1.0000	
	0.0000	0.0000	0.0000	0.0000	0.0000		
une_pop	-0.0297	-0.0282	0.0289	-0.0389	-0.0812*	-0.0710*	1.0000
	0.1008	0.1197	0.1108	0.0566	0.0000	0.0001	
une_hiv	-0.1078*	-0.0935*	-0.2871*	-0.1907*	-0.0866*	0.0249	-0.1023*
	0.0000	0.0000	0.0000	0.0000	0.0000	0.2338	0.0000
une_edu_sp~d	0.3143*	0.2980*	0.2711*	0.2257*	0.5216*	0.3415*	-0.1008*
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

	une_hiv	une_ed~d
une_hiv	1.0000	
une_edu_sp~d	0.1888*	1.0000
	0.0000	

Therefore, from the above matrix, a lot of the variables are highly correlated with each other. This may explain why a majority of the variables are insignificant, yet all the variables are jointly significant, with a high overall R-squared value.

Model Diagnostics

Rectification of Heteroskedasticity

Transformation of the dependent variable into logarithmic form was used to attempt to solve the issue of heteroskedasticity. However, on running the modified Wald test again, it still signified that there was heteroskedasticity in the data. Hence, the regression was rerun using robust standard errors, as shown in Figure 7. The significance of the variables is unchanged from before.

Rectification of Non-Normality

The Shapiro-Wilk test results indicated that the residuals were not normally distributed. Three transformations were attempted to resolve the issue: logarithm, inverse, and square root of the dependent variable. None of the methods provided a successful Shapiro-Wilk test.

However, as shown before in the kernel density plot of the residuals (Figure 2), **they maintain an approximate normal curve shape**. Thus, there is no need for any correction to account for the normality of the residual terms.

(Std. Err. adjusted for 112 clusters in country_num)

life_expect	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
adult_mortality	-.0386272	.0009241	-41.80	0.000	-.0404583	-.0367961
infant_mort	-43.36339	10.78051	-4.02	0.000	-64.72569	-22.00108
age14mort	-150.7627	28.04331	-5.38	0.000	-206.3324	-95.19299
alcohol	.0132166	.0259382	0.51	0.611	-.0381818	.0646149
bmi	-.2390598	.1794094	-1.33	0.185	-.5945715	.1164518
hepatitis	-.0014477	.0015189	-0.95	0.343	-.0044576	.0015621
measles	.009859	.0051821	1.90	0.060	-.0004097	.0201277
polio	.0006673	.0037599	0.18	0.859	-.0067831	.0081177
diphtheria	-.0041831	.0044243	-0.95	0.346	-.0129502	.004584
basic_water	.0037547	.0130426	0.29	0.774	-.0220901	.0295995
gni_capita	.0000126	.0000177	0.71	0.478	-.0000225	.0000478
gghed	-.0888739	.0408411	-2.18	0.032	-.1698033	-.0079445
che_gdp	-.0431496	.0273531	-1.58	0.118	-.0973516	.0110523
une_pop	-3.90e-06	7.70e-06	-0.51	0.614	-.0000191	.0000114
une_hiv	-.0538518	.0503927	-1.07	0.288	-.1537083	.0460047
une_edu_spend	.0391444	.0252535	1.55	0.124	-.0108972	.0891859
year						
2001	.2473638	.0589211	4.20	0.000	.1306078	.3641199
2002	.320939	.0721042	4.45	0.000	.1780598	.4638182
2003	.3930155	.0878222	4.48	0.000	.21899	.5670411
2004	.5580058	.1130039	4.94	0.000	.3340809	.7819307
2005	.7210019	.1340596	5.38	0.000	.4553538	.98665
2006	.8958805	.1592313	5.63	0.000	.5803531	1.211408
2007	1.04031	.1869214	5.57	0.000	.6699126	1.410707
2008	1.157641	.2124529	5.45	0.000	.7366516	1.578631
2009	1.313367	.2330327	5.64	0.000	.851597	1.775137
2010	1.431212	.253498	5.65	0.000	.9288891	1.933535
2011	1.612189	.2762113	5.84	0.000	1.064858	2.15952
2012	1.735708	.2893191	6.00	0.000	1.162403	2.309014
2013	1.911012	.311707	6.13	0.000	1.293344	2.52868
_cons	82.82516	4.360157	19.00	0.000	74.18522	91.4651
sigma_u	3.4885178					
sigma_e	.22756711					
rho	.99576266	(fraction of variance due to u_i)				

Figure 7: FE model results using robust standard errors

Section V: Conclusion

In this research, an econometric study was conducted on the Life expectancy in the African continent for the first 13 years of the 21st century. The panel type data used for this is taken from WHO (World Health Organisation) and World Bank databases and is highly rated for its quality. Despite this being the case the assumptions of homoskedasticity and normal distribution of OLS residuals were not satisfied by the data set. Non-normal distribution did not need treatment since the deviation from normality was less, the distribution was approximately normal in shape. For heteroskedasticity, robust standard errors were used.

Many variables that influence life expectancy have been taken from the data and used in the econometric model. The model is linear in parameters as well as in variables. Dummy variables have been used to capture the effect of time, with the base being the year 2000. Fixed effects model was used in line with the results from the Durbin- Wu- Hausman Test.

The variables are free from correlation except for adult mortality, infant mortality and age 1-4 mortality as well as the vaccinations (measles, polio and diphtheria) for which correlation makes logical sense as higher age mortality depends on lower age one and governments who go for vaccines cover all basic diseases not just one. Multicollinearity is seen in the second trio as well as in Gghe-d (Government Health Expenditure as a percentage of GDP), che_gdp (Private Current Health Expenditure as a percentage of GDP) and une_edu_spend (Total government spending on education) because an individual's private expenditure on health highly depends on the amount the government is spending on health.

The results, although not very intuitive, are in line with what is seen in the literature. Most of the explanatory variables are insignificant at 10% level of significance except mortality (adult, infant and between age 1 to 4) which are significant at less than 0.1% significance level. Measles, Domestic general government Health expenditure(gghe-d) are also significant at 10% and 5% significance level respectively. All year-dummy variables are also significant at 0.1% significance level and are also jointly significant.

From this regressive analysis, it can be concluded that in Africa life expectancy depends on adult mortality, infant mortality, mortality between age 1 to 4, Percentage of children below 1 year getting Measles vaccine, Domestic general government Health expenditure (gghe-d) as well as the year. Rest of the factors like other vaccines of polio, diphtheria etc. though not individually, but jointly have a significant effect on life expectancy.

Section VI: References

1. Preston, Samuel. (2003). The Changing Relation Between Mortality and Level of Economic Development. Bulletin of the World Health Organization. 81. 833-41. 10.1080/00324728.1975.10410201.

2. Dollar, David and Burnside, Craig, Aid, the Incentive Regime, and Poverty Reduction (April 1998). Available at SSRN: https://ssrn.com/abstract=597236

3. McCarthy, F. Desmond and Wolf, Holger C., Comparative Life Expectancy in Africa (August 2001). Available at SSRN: https://ssrn.com/abstract=632736

4. Husain, A. R. (2002). Life Expectancy in Developing Countries: A Cross-Section Analysis. *The Bangladesh Development Studies*, *28*(1/2), 161–178. http://www.jstor.org/stable/40795653

5. Kabir, M. (2008). Determinants of Life Expectancy in Developing Countries. *The Journal of Developing Areas*, *41*(2), 185–204. http://www.jstor.org/stable/40376184

6. Anyanwu, J.C. and Erhijakpor, A.E. (2009), Health Expenditures and Health Outcomes in Africa. African Development Review, 21: 400-433. https://doi.org/10.1111/j.1467-8268.2009.00215.x

7. Hassan, F. A., Minato, N., Ishida, S., & Mohamed Nor, N. (2016). Social Environment Determinants of Life Expectancy in Developing Countries: A Panel Data Analysis. *Global Journal of Health Science*, *9*(5), 105. https://doi.org/10.5539/gjhs.v9n5p105

8. Preston, S. H. (1975). The Changing Relation between Mortality and Level of Economic Development. *Population Studies*, *29*(2), 231. https://doi.org/10.2307/2173509

9. Starfield, B. (2000). Is US Health Really the Best in the World? *JAMA*, *284*(4), 483. https://doi.org/10.1001/jama.284.4.483

10. Momoh, Zekeri. (2015). CORRUPTION AND GOVERNANCE IN AFRICA.

11. Gauthier, Bernard; Wane, Waly. 2007. Leakage of Public Resources in the Health Sector : An Empirical Investigation of Chad. Policy Research Working Paper; No. 4351. World Bank. © World Bank. https://openknowledge.worldbank.org/handle/10986/7301 License: CC BY 3.0 IGO.

12. Marouf, F. E. (2010). Holding the World Bank accountable for leakage of funds from Africa's health sector. Health and Human Rights. https://www.hhrjournal.org/2013/08/holding-the-world-bank-accountable-for-leakage-of-funds-from-africas-health-sector/

13. Wooldridge, J. M. (2022). Introductory Econometrics: A Modern Approach (5th Edition). Cengage Learning.