Life expectancy and per capita GDP

By Lucky Dube*

This paper investigates the bivariate relationship between life expectancy and per capita GDP. Data on country level life expectancy and GDP per capita in 2017 are drawn from Gapminder. The results find a statistically significant association between life expectancy and the logarithm of per capita GDP.

I. Data

Life expectancy data for 189 countries in 2017 are drawn from the Global Burden of Disease Study 2017. Gross domestic product per capita data in 2017 are measured in 2011 purchasing power parity adjusted international dollars and are drawn from World Bank data.

The combined sample totals 189 observations of two variables. Life expectancy data have a skewness of -0.42 and kurtosis of -0.63. The mean and median life expectancy is 72.8 years and 74 years respectively. Singapore and the Central African Republic have the highest and lowest life expectancy at 84.79 years and 51.87 years respectively. Life expectancy data has a coefficient of variation of 0.1. GDP per capita (level data) has a skewness of 1.82 and kurtosis of 4.01. The mean and median GDP per capita is $18,132 and $11,647 respectively. Qatar and Somalia have the highest and lowest GDP per capita at $113,262 and $627 respectively. The GDP data has a coefficient of variation of 1.07.

* This work was completed as part of the MSc Economics programme at Birkbeck, University of London. This write up was downloaded from github.com/lucky-dube/projects

Figure 1. : Sample variable density plots

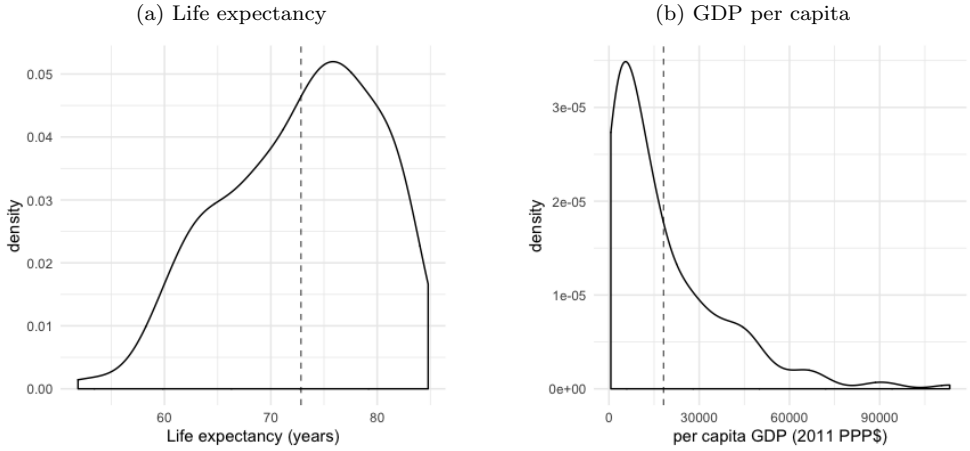


Table 1—: Variable descriptive statistics

	Life expectancy (years)	GDP per capita (2011 PPP$)
Mean	77.84	18,132
St.Dev	7.1	19,284
Min	51.87	627
Q1	68.05	3,654
Median	73.98	11,647
Q3	78.33	26,416
Max	84.79	113,262
IQR	10.28	22,762
CV	0.10	1.07
Skewness	-0.42	1.82
Kurtosis	-0.63	4.01
N	189	189

Source: Global Burden of Disease Study 2017, World Bank

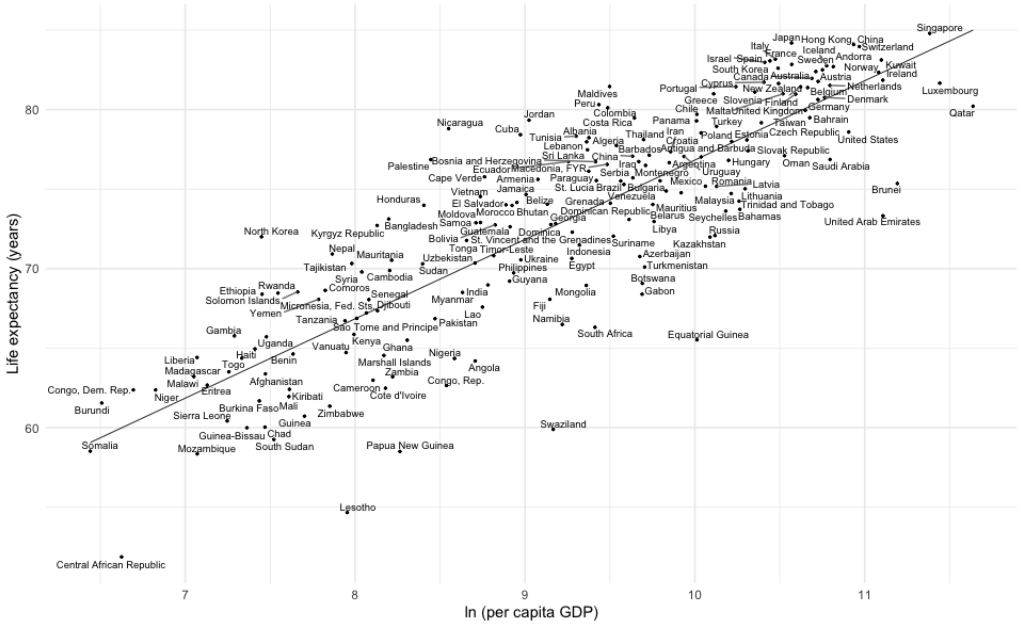
II. Analysis

Data were analysed using R in combination with the RStudio interface.

To investigate the linear association between life expectancy and the logarithm (ln) of GDP per capita, the following bivariate model was specified:

$$LE_i = \beta_0 + \beta_1 ln(GDPPC)_i + u_i \tag{1}$$

Figure 2. : Life expectancy by ln(GDP per capita)



Source: Global Burden of Disease Study 2017, World Bank

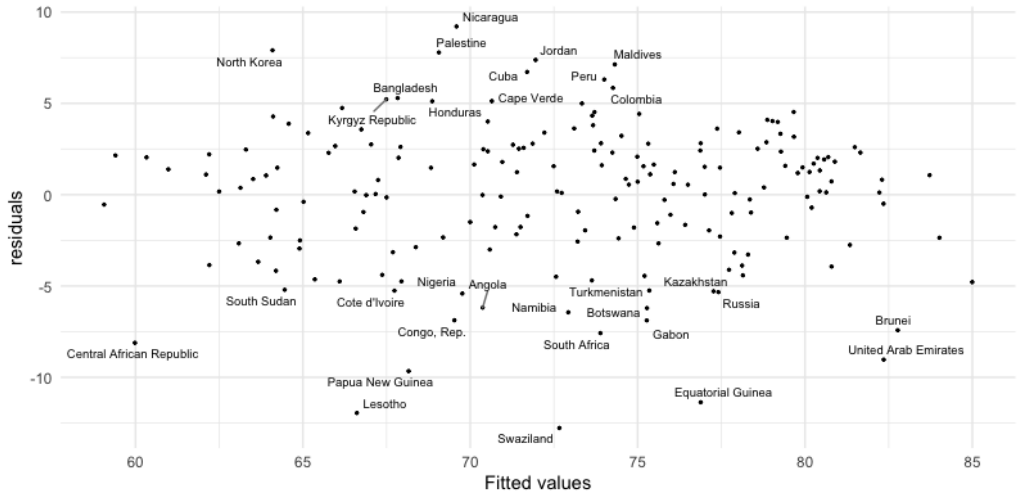
The logarithm of GDP per capita has been used as it captures differences in the proportion of GDP per capita between countries as opposed to dollar differences.There is a statistically significant association between life expectancy and the logarithm of GDP per capita (Table 2). The coefficient of ln(GDPPC) is not significantly different from 5. As such the model predicts that if there two countries, one with per capita GDP that is 100% higher than the other, that country's life expectancy will be 5 years higher. Diagnostic tests carried out provide evidence that the residuals have a constant variance and are normally distributed. The result of a RESET test provides evidence of no linear misspecification.

A plot of the residuals (Fig 3) and fitted values shows the residuals scattered around the zero line. There are a number of large residuals. The model's prediction of life expectancy in Lesotho, Swaziland, and Equatorial Guinea is over 10 years greater than their observed values.

Table 2—: OLS Regression output and diagnostics for (1)

	Estimate	Std.error	t value	p value
β_0	26.9	2.157	12.47	2×10^{-16}
β_1	4.99	0.2325	21.46	2×10^{-16}
	N =	189		
	R^2 =	0.7112		
	Adjusted R^2 =	0.709		
	SER =	3.826		
	F($\beta_1 = 0$) =	460.5	p value	2×10^{-16}
	F test for heteroskedasticity	1.00	p value	0.318
	RESET =	0.888	p value	0.4128
	Jarque-Bera =	15.53	p value	0.0004

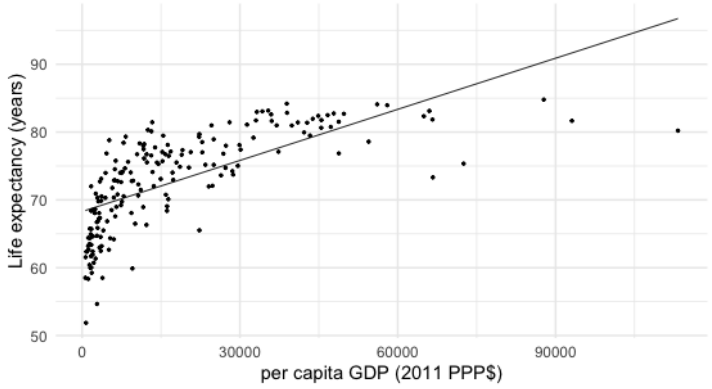
Figure 3. : Residual plot



For comparison, life expectancy was regressed on GDP per capita (level, as opposed to log):

$$LE_i = \gamma_0 + \gamma_1 GDPPC_i + \epsilon_i \tag{2}$$

Figure 4. : Life expectancy by GDP per capita



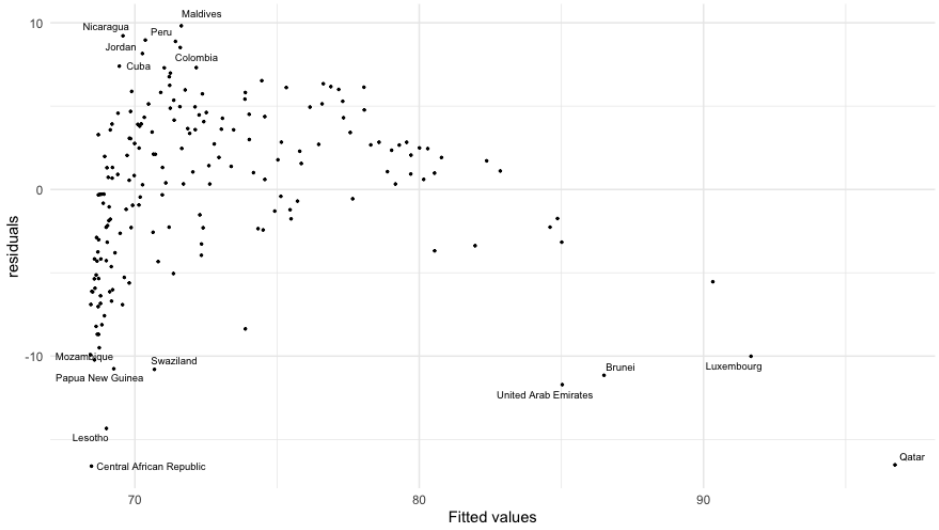
Source: Global Burden of Disease Study 2017, World Bank

There is a statistically significant association between life expectancy and per capita GDP. Diagnostic tests carried out on the model (Table 3) provide evidence of homoskedastic residuals and linear misspecification. The residual plot (Fig 5) shows that the residuals have a pattern. The results of a Jarque-Bera test show that the residuals are not normally distributed (the model is similar to model (1) in this respect). The model underestimates the life expectancy in Mozambique, Papua New Guinea, Swaziland, United Arab Emirates, Brunei, and Luxembourg by more than 10 years. The predictions for Central African Republic and Qatar are 15 years higher than their observed values. Compared to model (1), this model explains less of the variation in life expectancy: 47% compared to 71%. Additionally, this model predicts life expectancy to within 5.2 years compared to 3.8 years for model (1).

Table 3—: OLS Regression output and diagnostics for (2)

	Estimate	Std.error	t value	p value
γ_0	68.2	0.517	132.1	2×10^{-16}
γ_1	2.52×10^{-4}	1.95×10^{-5}	12.88	2×10^{-16}
	N =	189		
	R^2 =	0.4701		
	Adjusted R^2 =	0.4673		
	SER =	5.183		
	$F(\gamma_1 = 0)$ =	165.9	p value	2×10^{-16}
	F test for heteroskedasticity	1.99	p value	0.159
	RESET =	56.6	p value	2.2×10^{-16}
	Jarque-Bera =	15.389	p value	0.0004

Figure 5. : Residual plot



III. Conclusion

There is a statistically significant association between life expectancy and the logarithm of per capita GDP. The result should not be taken to imply a causal link between sample variables. However, the relationship between the two variables is predictive, such that the model used can predict life expectancy within 3.8 years.