INCOME AND INEQUALITY –
DETERMINANTS OF LONGEVITY?

## Sinéad Kelleher

Junior Sophister

In this econometric enquiry, Sinéad Kelleher analyses the effect that income and inequality have on life expectancy. A strong link is found between GNP per capita and life expectancy, though the link between inequality (measured using the Gini coefficienct) is not found to be statistically significant.

Introduction

Today, despite over 50 years of focused development effort, the average gap between the life expectancy of those in the industrial, developed North and those in the poorer, more agrarian South is still exceptionally high. In 1998 a person living in the developed world was expected to live 75 years, almost double the expected life span of those in the world's least developed countries, who were expected to live to the age of just 48 (Todaro 2000).[1] The extent of the variation in life expectancy in my data set can be seen in the scatter plot below. The y-axis measures life expectancy in years from birth, whereas the observations along the x-axis are the countries in alphabetical order.

---

[1] The biggest gap in life expectancy observed in my data is between Canada (79.3 years) and the Angola (37.8 years). A full list of countries and their associated data can be found in the appendix.

Figure 1:
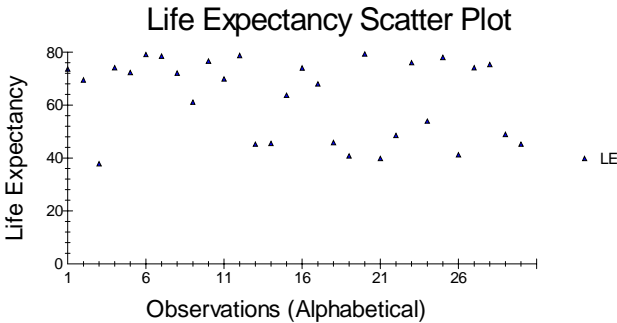
The determinants of longevity are difficult to unravel and quantify. However, one can easily identify many factors which influence one's health status and therefore, life expectancy. Access to clean water, availability and cost of immunizations, prevalence of HIV/AIDS, health technology, educational facilities, and the incidence of violent conflict are obvious examples. However, in this paper I intend to show a relationship between life expectancy and broader, economic variables, namely GNP per capita, and the inequality of income distribution within a nation, measured, in this case, by the Gini coefficient.

Literature Review and Hypotheses

A vast amount of research has been done on the link between socioeconomic status and longevity. It has been shown, and is intuitively logical, that residents of wealthier countries tend to live longer, but that high levels of societal income inequality reduce average life expectancy. With regards to income distribution, Amin found, after carrying out a multiple regression using life expectancy as a proxy for the health of a nation, that the Gini variable was significant to the .01 level (2001). He showed that, on average, when the income of a country is distributed very unequally, the life expectancy decreases. These results are substantiated by Idala, who attempts to explain this by suggesting that "societies with greater equality can be expected to have a lower infant mortality rate or higher life expectancy because more people can afford life's necessities" (2002: 18). However, the study carried out by Lobmayer and Wilkinson yielded different results (2000). They showed a "disappearance of the relationship between income inequality and mortality" (Lobmayer & Wilkinson 2000). They use two summary measures of mortality: Age-Adjusted Mortality and Potential

240

Years of Life Lost (PYLL). For age-adjusted mortality the relationship between mortality and income inequality (measured using the 50:10 centile ratio) is either non-existent or weakly inverse, whereas the relationship between mortality and income inequality using the PYLL measure is positive up to the age of 50 and the relationship reverses its sign among those over 65. However, the countries used in this study were all wealthy, developed market democracies. It is possible that the effects of income inequality may diminish as countries become wealthier, meaning that this study cannot be automatically generalised and assumed to be equally valid with regards to developing nations. For this reason, I am going to assume this study is not directly relevant to my paper. My first hypothesis therefore states: *There will be a negative relationship between the measure of income inequality (Gini coefficient) and life expectancy.*

Previous studies on the link between income and life expectancy have suggested a positive relationship. Rodgers shows this relationship is highly significant and results with R^2 values of over .75 when life expectancy at birth is used as the measure of longevity (1979). Idala reports similar results that are consistent, he notes, with the absolute income hypothesis, which states that mortality decreases with average income, but at a declining rate (2002). A point is reached where an increase in income will no longer improve one's health nor increase life expectancy. These results and the intuitive logic of the statement lead to my second hypothesis: *There will be a positive relationship between the measure of income per capita (GNP/capita) and life expectancy.*

Data Selection and Dependent Variable

A cross sectional sample of 30 countries was selected for this study (the full list of countries and associated data is available in the appendix). In order to get a wide and unskewed data set, the Human Development Index (HDI) was used. The HDI is published annually by the UN and ranks countries on a scale of 0 to 1 based on the quality of life of their citizens, as opposed to on purely economic measures. I divided the countries on the United Nations website into three categories depending on their HDI ranking. According to the classification proposed by UNDP, countries with an HDI below 0.5 are considered to have a low level of human development; those between 0.5 and 0.8, a medium level; and those above 0.8, a high level. The first ten countries alphabetically were selected from each category. However, in some cases, especially in the category of countries with low human development indices, lack of data availability forced me to skip a number of countries. This system allowed developed and developing nations to be weighed equally and hopefully resulted in the data used reflecting the

broad patterns within each category. The decision to choose countries alphabetically was taken to prevent any bias in the choice of observations on the basis of population, wealth, global influence etc.

The dependent variable used in the study is life expectancy from birth. Data was sourced from the UNDP website and is from the 2002 World Development Report.

Explanatory Variables

GNP per Capita

Gross National Product is the total monetary value of goods and services produced in a year by the nationals of a country. It includes income that nationals earn abroad, but does not include income earned within a country by foreigners. To determine GNP per capita, this figure is simply divided by the population of the country in question. This measure, as opposed to GDP per capita, was chosen as it was felt it would more accurately represent the incomes of those living in poorer, less developed countries within which many large American or European Multinationals are located.

The relationship between life expectancy and income/capita is not linear-research by Rodgers and Idala shows the reciprocal of the square of per capita GDP to yield the closest relationship (1979: 2002). However, this specification was deemed to be outside the scope of this paper, therefore the logs of the GNP per capita values are used. This specification has also been shown by research to yield quite satisfactory results (Rodgers 1979; Idala 2002). The 2003 figures for GNP per capita were found on the World Bank website, and were computed using the Atlas method (2005).

Figure 2:

Life Expectancy Regressed on log (GNP/capita)
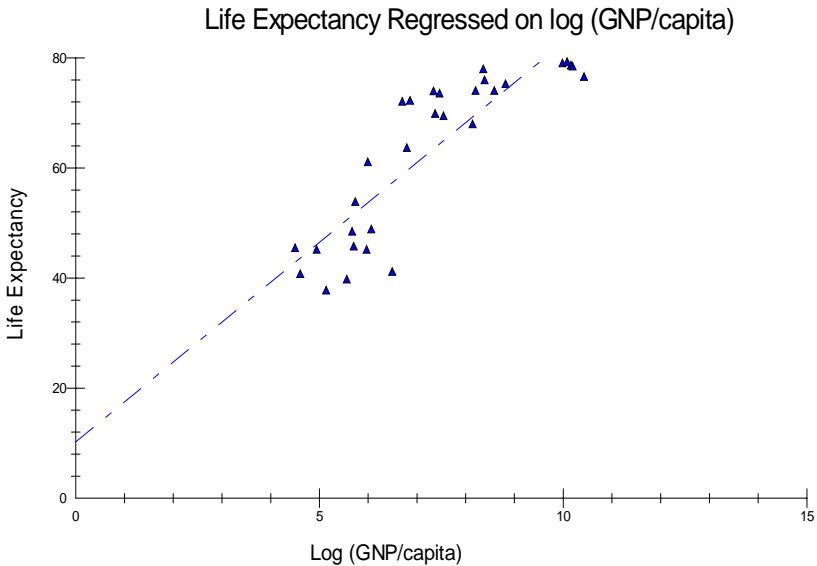
As can be seen from the above scatter diagram, there is a clear positive relationship between the two variables. An analysis of the OLS regression results clarifies this relationship. The R^2 value of 0.75553 means that over 75% of the variation in life expectancy can be explained by variation in GNP per capita. The F-test returns a p-value of 0 to three decimal places; we can be confident that the model has overall significance.

It was also found that the GNP/capita variable was positively related to life expectancy. A coefficient of +7.2535 was produced. The probability value for the t-statistic is also 0 to three decimal places, indicating that the coefficient is statistically different to zero. These results lead us to fail to reject the first hypothesis. There is a positive relationship between life expectancy and GNP per capita.
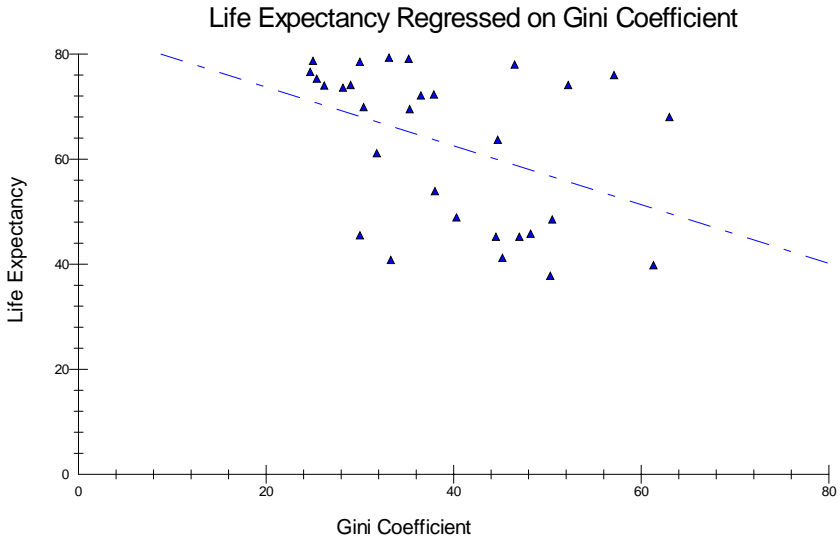
Gini Coefficient

The Gini coefficient is a summary measure of the inequality of income distribution within society and can be expressed as a proportion or a percentage. It has a value of zero when there is complete equality, and is equal to one (or 100%)

in a situation where the income of the entire society is held by one household, leaving the rest of the population with no income at all (Amin 2001). The average Gini value for the nations used in this study was 39.36%. The data for this variable was found on the United Nations website and is from the 2002 World Development Report.

Figure 3:

Life Expectancy Regressed on Gini Coefficient

The results of the OLS regression of life expectancy on the Gini coefficient are as follows: the F and t results are significant at the 5% level, but this variable has much less explanatory power than GNP/capita does- the variation in the Gini coefficient value explains just 17.14% of the variation in life expectancy. The coefficient value produced is -0.55928, allowing us to fail to reject the second hypothesis, though possibly not at an acceptable level of significance (see below). There is an inverse relationship between income inequality and life expectancy.

Dummy Variable

A dummy variable was used to account for any unquantifiable consequences of living in a country with a high, medium, or low level of Human Development. Countries with a high HDI (>.8) are the base group. Countries with medium (0.5-0.8) and low HDI (<.5) measures were assigned 1 and 2 respectively.

Multiple Regression

The model used in the multiple regression is a simple three variable regression model taking the form:

$$Y_i = \beta_0 + \beta_1 (\ln X_1) + \beta_2 X_2 + \beta_3 D_i + U_i$$

Where Y_i = Dependent Variable-Life Expectancy from Birth in country

           i
         β_0 = Intercept
         $\ln X_1$ = log of GNP/Capita
         X_2 = Gini Coefficient
 D_i = Dummy Variable
         U_i = Error Term

To investigate whether the model contained significant problems of multicollinearity, the correlation between X_1 and X_2 was calculated. This yielded a R^2 value of .11668, meaning that only 11.7% of the variation in X_2 is statistically explained by the variation in X_1. This is an acceptably low result (Gujarati 2002).

Coefficient of determination (R^2)

R^2 is a measure of the goodness of fit of the fitted regression line. The multiple regression yielded an R^2 value of 0.83093, meaning that 83% of the variation in life expectancy can be attributed to the variation in the explanatory variables. However, taking into account the fact that the R^2 measure is, as Gujarati states 'a non-decreasing function of the number of explanatory variables, or regressors in the model,' it is more appropriate to examine the adjusted R^2 value (2002: 217). This measure takes into account the number of X variables in the model, and in this multiple regression, yields a result of 0.81142. This high result can be seen by examining the graph of actual Y against fitted Y values below.

Figure 4:

Plot of Actual and Fitted Values
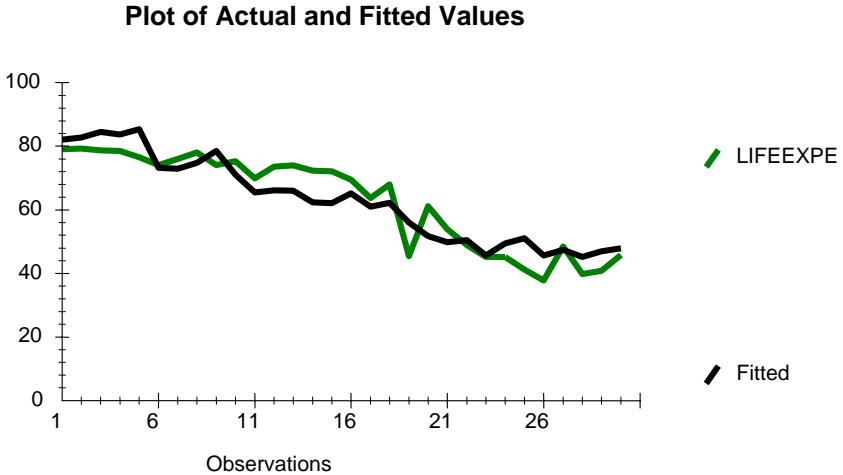
T-tests

The null hypothesis for the t computation is $\beta_i = 0$. The t-test results show that the coefficient result for the X_2 variable (-0.1771) is not significantly different from zero at the 10% level-we cannot reject the null hypothesis. This means that the variation in the Gini coefficient is not statistically significant in explaining the variation in life expectancy at this significance level. The lowest level at which the inequality measure will be significant is the 13.9% level. This can be determined by the fact that the associated p-value is 0.139.

All of the other coefficients are significant at the 10% level. The signs of resulting coefficients conform to expectations. The coefficient associated with income per capita is positive (3.3093) whereas the sign associated with the inequality measure is weakly negative (-0.1771), and the dummy coefficient is strongly negative (-8.8313).

F-statistic

The F-statistic tests the overall significance of the sample regression. The null hypothesis is that all the slope coefficients are simultaneously zero, i.e.

$$H_0: \beta_1 = \beta_2 = \beta_3 = 0$$

The F-statistic calculated in the multiple regression is 42.5938. This far exceeds the critical value at the 1% level (F $_{(3\ 26)} \approx 4.64$), allowing us to reject the null hypothesis and state that our model does have overall significance at the 1% level.

Durbin-Watson test

The Durbin-Watson test is used to detect autocorrelation between error terms. The null hypothesis is that there is no positive or negative autocorrelation between the residuals i.e.

$$H_0: cov\ (u_i,\ u_j|\ X_{i,}\ X_j\ ) = 0 \qquad i \neq j$$

The Durbin-Watson statistic obtained following the multiple regression was 1.1290. For 30 observations and 3 explanatory variables, $d_u = 1.421$ and $d_l = 1.006$. As our DW statistic is between these two bounds, the test is indecisive.

Jarque-Bera test for normality

This test measures the difference of the skewness and kurtosis of the distribution with those from the normal distribution. A small probability value leads to the rejection of the null hypothesis of a normal distribution of the residuals. The value computed in the multiple regression was 1.5369. The lowest significance level at which we can reject this null hypothesis is 46.4% (i.e. the computed p-value is 0.464). The computed value is far less than the critical value at the 5% level (χ^2 (2) ≈ 5.99147), therefore I cannot reject the null hypothesis. This is a very important result as the F and t tests both assume normal distribution of variables. It is useful to look at the histogram of residuals to visualize how they are distributed.

Figure 5:

Histogram of Residuals and the Normal Density
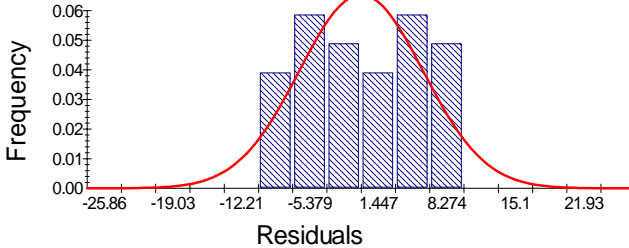
Heteroscedasticity

The null hypothesis in this test is that the residuals exhibit the same conditional variances, i.e. H_0: var $(u_i | X_i) = \sigma^2$. The result produced by Microfit is .33515 with a p-value of 0.563. Since 56.3% is the lowest significance level at which we can reject the null hypothesis of homoscedasticity, it is obvious that we cannot reject it at the 5 or 10% level.

Functional Form

The null hypothesis for this test is that the model is correctly specified. The criteria for analyzing an F-test is to reject H_0 if $F > F_{\alpha (k-1, n-k)}$, or if p is sufficiently low. The F-value produced by the multiple regression is 16.7431. This far exceeds the 1% critical value of 4.64, and the p-value is less than .000. These results mean the null hypothesis is rejected and we are forced to conclude that the model is incorrectly specified. Unfortunately, the Ramsey RESET test cannot tell us the source or form of the misspecification, nor suggest any superior alternatives. However, as mentioned earlier, previous research suggests that using the squared reciprocal of GNP/capita as an explanatory variable would yield a more satisfactory result for this test.

95% Confidence Interval for β_2

This will take the form $\beta_2 \pm t_{\alpha/2}$ se (β_2). At a 95% confidence interval, $\alpha=5\%$, T tables show $t_{(26, 0.025)} \approx 2.056$; thus $-.1771 \pm (2.056) (.11589)$ or $-.1771 \pm .23827$. This yields a confidence interval of [-0.0612, 0.4154]. If this test was

248

carried out an infinite amount of times, the true value of β_2 would lie between (-0.0614) and (0.4154) 95% of the time.

Forecast

Figure 6:

Plot of Actual and Single Equation Static Forecast(s)
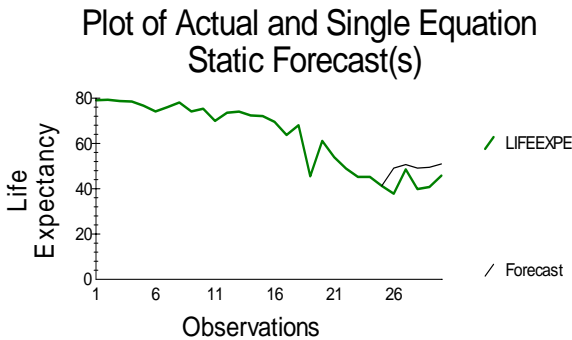
I will test the forecasting capabilities of this model by substituting the data for Burkina Faso, and a dummy variable of 2 (indicating a low HDI ranking) into our regression line, and comparing the result with the true life expectancy in this country.

Substituting this data into the line of best, fit the following is obtained:

Y = 55.2519 + 3.3093 (5.7037824746562) + (-.1771) (48.2) + (2) (-8.8313)

Y = 47.92

This is our forecasted expected life expectancy for Burkina Faso. The model overestimates the true value (45.8 years) by 4.6%. Considering the fact that the model does not explain 100% of the variation in life expectancy, I consider this a relatively good result. Also, the Chow predictive failure F-test yielded a result of 0.95573 with a p-value of 0.467, resulting in a failure to reject the null hypothesis of correct forecasting properties of the model.

Conclusion

In this model, the lowest significance level at which all of the coefficients are statistically different to zero is 13.9%. This is not an ideal result as one would usually wish for their explanatory variables to be significant at quite low levels, 5% or 10%. However, the model as it stands is shown by the F-statistic to have overall significance and yields a high adjusted R^2 level of 0. 81142. I did not reject the null hypothesis in the case of the Jarque-Bera test for Normality. The test for heteroscedasticity yielded a satisfactory result – we could not reject the null hypothesis of homoscedasticity at the 5% level. The Durbin-Watson test for autocorrelation is indecisive. Unfortunately, the Ramsey RESET test shows the model to be misspecified.

The conclusion of this paper would be that the absolute income hypothesis (i.e. health and life expectancy depend on income, not income relative to others' income or income inequality) is more valid than the relative income hypothesis, which stresses the significance of income distribution. Policy suggestions arising from this study would therefore encourage a focus on increasing overall GNP, and would be less concerned with how income is divided between societal members.

However, a number of limitations exist in this study. Firstly, the measure of income inequality used, the Gini coefficient, is not an unbiased measure, being more sensitive to inequalities at the top of the income distribution (Idala 2002). The measure used by Lobmayer and Wilkinson in their study, the ratio of income at the 50th and 10th centiles, was shown by Daly et al to be more strongly related to mortality levels and may have been a better measure to have used (2000; 1998). Also, the Jarque-Bera test for normality is an asymptotic test. The sample size in my study was rather small. Hence, the results of this test may not be entirely valid. Further study into the links between socioeconomic status and mortality should include an effort to overcome these shortcomings and provide a more legitimate and valid insight into this interesting and essential area of research.

Appendix

Country	Life Expectancy	Gini Coefficient (%)	GNP/Capita	Ln(GNP/Capita)
High HDI				
Australia	79.1	35.2	21650	9.9828
Canada	79.3	33.1	23930	10.0829

Belgium	78.7	25	25820	10.1589
Austria	78.5	30	26720	10.1931
Barbados	76.6	24.7	33750	10.4267
Argentina	74.1	52.2	3650	8.2025
Chile	76	57.1	4390	8.3871
Costa Rica	78	46.5	4280	8.3617
Croatia	74.1	29	5350	8.5849
Czech Republic	75.3	25.4	6740	8.1582
Medium HDI				
Belarus	69.9	30.4	1590	7.3715
Albania	73.6	28.2	1740	7.4616
Bosnia and Herzegovina	74	26.2	1540	7.3395
Armenia	72.3	37.9	950	6.8565
Azerbaijan	72.1	36.5	810	6.6970
Algeria	69.5	35.3	1890	7.5443
Bolivia	63.7	44.7	890	6.7912
Brazil	68	63	3430	8.1403
Bhutan	45.5	30	90	4.4998
Bangladesh	61.1	31.8	400	5.9915
Low HDI				
Congo	53.9	38	310	5.7366
Djibouti	48.9	40.3	430	6.0638
Eritrea	45.2	47	140	4.9416
Benin	45.2	44.5	390	5.9661
Côte d'Ivoire	41.2	45.2	660	6.4922
Angola	37.8	50.3	170	5.1358
Chad	48.5	50.5	290	5.6699

Central African Republic	39.8	61.3	260	5.5607
Burundi	40.8	33.3	100	4.6052
Burkina Faso	45.8	48.2	300	5.7038

Bibliography

Amin, K. (2001) "Income Distribution and Health: A Worldwide Analysis." *The Park Place Economist* ,Vol. IX, pp. 45-52.

Daly, M. C. et al. (1998) "Macro to Micro links in the relation between income inequality and mortality." *Millbank Quarterly*, Issue 76, pp. 315-339.

Gujarati, D. N. (2002) *Basic Econometrics*. New York: Mc-Graw-Hill.

Idala, D. (2002) *Do income and Income Inequality Explain the Variation in Life Expectancy Across Countries?* Viewed on 13th February 2005 <http://www.umbc.edu/economics/grad_699_abstracts/d_idala.pdf>.

Lobmayer, P. & Wilkinson, R. (2000) "Income, Inequality and Mortality in 14 Developed Countries." *Sociology of Health and Illness,* Vol. 22, pp. 401-414.

Rodgers, G.B. (1979) "Income and Inequality as Determinants of Mortality: An International Cross-Sectional Analysis." *Population Studies,* Vol. 33, No. 2, pp.343-351.

Todaro, P. (2000) *Economic Development*. London: Pearson Education Limited.

United Nations Development Programme. Viewed 14th February 2005, <www.undp.org>

World Bank. Viewed 20th February 2005, <www.worldbank.org>.