Journal of Environment Protection and Sustainable Development
Vol. 5, No. 1, 2019, pp. 14-21
http://www.aiscience.org/journal/jepsd
ISSN: 2381-7739 (Print); ISSN: 2381-7747 (Online)

Factors Affecting Life Expectancy: A Global Perspective

Linali Pathirathne, Roshini Sooriyarachchi*

Department of Statistics, University of Colombo, Colombo, Sri Lanka

Abstract

The primary objective of this research is to examine the effect of economic, social and environmental indicators on life expectancy, using a cross-sectional comprehensive worldwide sample. The data corresponds to countries belonging to the United Nations (UN) and was taken from the World Statistics Pocketbook published by the UN (2005). Using this sample of data, the impact of country level variables on life expectancy (LE) at birth was analyzed. Multiple regression was used to model the LE. Principal Components (PCs) were used as explanatory variables for the model so as to avoid the problem of high correlation between explanatory variables (multi-collinearity). The model variants suggested that proxies for economic development, technology, nature conservation, education, healthcare, communication, population density and population growth rate all have a significant effect on average life expectancy. To increase the LE it is recommended to improve the economy, education level, health facilities, communication facilities and rural life style and to reduce industrialization, pollution, war and inflation. In addittion population growth should be encouraged but urbanization should be controlled so as to improve the LE. This analysis provides information required to governments, especially in the developing world as the LE at birth is predicted with high explanatory power by variables that can be influenced through public policy.

Keywords

Life Expectancy, Principal Components, Regression, Proxy, Environmental, Socioeconomic

Received: February 22, 2019 / Accepted: March 26, 2019 / Published online: April 17, 2019

1. Introduction

Life expectancy at birth is the average age at which a person is expected to die. Life expectancy is a common measure of population health in general, and is often used as a summary measure when comparing different populations. Life expectancy is also used in public policy planning, especially as an indicator of future population ageing. The level of life expectancy in a country has important implications. It affects fertility behavior, economic growth, human capital investment, intergenerational transfers and incentives for pension benefit claims [1, 2]. Thus determining the factors effecting life expectancy is vital.

Many previous studies [3] have been carried out with the intention of predicting life expectancy. Most of these studies have been done on a specific country [4, 5] or a selected sample of countries [6] while prediction has been based on a limited number of variables [7]. Many of the large studies considering the effect of numerous and varied explanatory variables on life expectancy for a vast number of countries, are outdated [8] and thus do not reflect the current situation. Another problem identified with these studies is that the effect of explanatory variables on life expectancy has been determined using regression models of life expectancy on the untransformed raw variable. These raw variables are highly correlated with each other and thus these models suffer from the problem of multi-collinearity resulting in the conclusion drawn from such models being unreliable.

* Corresponding author
E-mail address: roshinis@hotmail.com (R. Sooriyarachchi)

In this study the focus is on relating life expectancy to a broad range of variables and the sample consists of 117 countries belonging to the UN [9]. The data is quite current and gives the situation as at 2004. The objective of this research is to accomplish the requirement of identifying the influential factors on life expectancy. The primary aim is to determine new factors that may influence life expectancy and the secondary aim is to examine whether factors established as significant over the years still remain so. Regression analysis is used to satisfy these objectives and principal components are used as the explanatory variables in the modeling process, so as to overcome the problem of multi-collinearity. The results would produce valuable information as to what a country should enhance in order to improve the life expectancy of its citizens.

2. Methods

2.1. Study Data

The data was extracted from the World Statistics Pocketbook published by the UN (2005) [9]. There were 209 member countries which were stated under UN as at 30th November 2004. However several countries had to be removed from the analysis due to these having missing values for important variables and thus only 117 countries were used for this study. This study mainly focused on relating life expectancy to a wide range of variables including economic indicators, social indicators and environmental indicators. Initially there were 50 explanatory variables, but this has been reduced to 31 due to the removal of variables having a high proportion of missing values. Table 1 gives the variables used for this study together with its code name, categorized according to economic, social and environmental indicators and Table 2 gives the countries used for this study, categorized according to continent. Most of the variables given in table 1, except those with an asterisk sign in front of the variable name are self explanatory as the variable name gives a good description of these variables. The variables in table 1, with an asterisk are described in more detailed in the footnote of table 1.

Table 1. Variables Used in the Study.

Category	Variable Name	Code
Economic Indicators	Population density (per square km)	PD
	Exchange rate (national currency per US$)	ER
	Consumer price index (1990=100)	CPI
	Tourist arrivals (000s)	TA
	Gross domestic product (million current US$)	GDP1
	Gross domestic product (per capita current US$)	GDP2
	Gross fixed capital formation (% of gross domestic product)*	GFCF
	Labour force participation, adult female population (%)	LFPF
	Labour force participation, adult male population (%)	LFPM
	Employment in industrial sector (%)	EIS
	Employment in agricultural sector (%)	EAS
	Motor vehicles (per 1000 inhabitants)	MV
	Telephone lines (per 100 inhabitants)	TL
	Internet users, estimated (000s)	IU
Social Indicators	Population growth rate 2000 – 2005 (% per annum)	PGR
	Population aged 0 – 14 years (%)	PA014
	Population aged 60+ years (women, % of total)	PA60PW
	Population aged 60+ years (men, % of total)	PA60PM
	Sex ratio (women per 100 men)	SR
	Infant mortality rate 2000 – 2005 (per 1000 births)	IMR
	Total fertility rate 2000 – 2005 (births per woman)	TFR
	Urban population (%)	UP
	Urban population growth rate 2000 – 2005 (% per annum)	UPGR
	Rural population growth rate 2000 – 2005 (% per annum)	RPGR
	Government education expenditure (% of gross national product)	GEE
	Newspaper circulation (per 1000 inhabitants)	NPC
	Television receivers (per 1000 inhabitants)	TR
Environmental Indicators	Threatened species	TS
	Forested area (% of land area)	FA
	CO_2 emissions (000s Mt)	CO_2E
	Energy consumption per capita (kilograms oil equiv.)*	ECPC

Footnote:
GFCF is the total value of a producers acquisitions less disposal of fixed assets during the accounting period plus value of non-produced assets realized by the productive activity.
ECPC is the imports plus production minus changes in stocks minus exports.

Table 2. Countries Considered in the Study.

Africa	America	Asia	Europe	Oceania
Algeria	Argentina	Azerbaijan	Austria	Australia
Angola	Bahamas	Bahrain	Belarus	Fiji
Benin	Barbados	Bangladesh	Belgium	New Zealand
Burkina Faso	Belize	Brunei Darussalam	Bulgaria	Papua New Guinea
Burundi	Bolivia	Cambodia	Croatia	
Cameroon	Brazil	China	Czech Republic	
Central African Republic	Canada	Cyprus	Denmark	
Congo	Chile	Georgia	Estonia	
Cote d'Ivoire	Colombia	India	Finland	
Gabon	Costa Rica	Indonesia	France	
Gambia	Dominican Republic	Iran, Islamic Republic	Germany	
Ghana	Ecuador	Israel	Greece	
Kenya	El Salvador	Japan	Hungary	
Madagascar	Guatemala	Jordan	Ireland	
Mali	Guyana	Korea (Republic of)	Italy	
Mauritius	Haiti	Kyrgyzstan	Latvia	
Morocco	Honduras	Lebanon	Lithuania	
Mozambique	Jamaica	Malaysia	Malta	
Nigeria	Mexico	Myanmar	Netherlands	
Sierra Leone	Nicaragua	Nepal	Norway	
South Africa	Panama	Pakistan	Poland	
Sudan	Paraguay	Philippines	Portugal	
Togo	Peru	Saudi Arabia	Republic of Moldova	
Uganda	Suriname	Singapore	Romania	
United Republic of Tanzania	Trinidad and Tobago	Sri Lanka	Russian Federation	
Zambia	United States of America	Thailand	Slovakia	
	Uruguay	Viet Nam	Slovenia	
	Venezuela		Spain	
			Sweden	
			Switzerland	
			Ukraine	
			United Kingdom	

2.2. Methodology

Multiple linear regression models [10] were used to examine the extent to which explanatory variables given in table 1 explained the variation in life expectancy. As preliminary analysis indicated that these explanatory variables are strongly correlated with each other, principal component analysis (PCA) [11] was used to transform these raw variables in to orthogonal components which were used in place of the raw variables in the modeling process. This was done so as to avoid the problem of multi-collinearity [10] which could lead to unreliable results. When variables do not occur on an equal footing, it is necessary to apply PCA to standardized data (that is z scores). This is because requirements in PCA are that variables must be in the same units or comparable units and variables should have variances that are roughly similar in sizes. Thus, the variance covariance matrix has to be used to analyze the original data where as to analyze the standardized data the correlation matrix has to be used.

The forward selection procedure [10] was used to select important components for the model. This involves adding each component to the model incrementally to show the marginal increase in explanatory power related to life expectancy. First the components that are expected to have highest explanatory power are added. The final model thus obtained is tested for goodness of fit and the validity of assumptions examined using plots of residuals.

Once the model satisfaction is determined the selected components are interpreted and the related proxy extracted. This is achieved by examining the raw variables that contribute highly to the selected component [12]. Finally the extent to which each proxy affects life expectancy is determined by interpreting the regression coefficients. The study ends by identifying new proxies that affect life expectancy and confirming either the presence or absence of established proxies.

3. Results and Findings

3.1. Determining Principal Components

With the intention of obtaining orthogonal PCs, PCA was applied to the standardized data (31 variables given in table 1). Since the data were not in the same or comparable units, the correlation matrix was used in this case. This resulted in 31 orthogonal PCs which were used as explanatory variables in the modeling procedure.

3.2. Selecting the Most Suitable Model

The forward selection method was used to select the components (proxies) that effect life expectancy. The final model was obtained after sixteen steps and included 16 principal components. The resulted model can be specified along with the parameter estimates as follows.

$$\text{Life\_Expectancy} = 67.29567 + 2.72233(PC1) - 0.61396(PC2) - 1.86367(PC3) - 1.96885(PC4) + 1.22797(PC6) - 1.18675(PC8) + 0.89542(PC11) - 0.84886(PC12) - 3.40194(PC14) + 1.05311(PC15) - 3.40402(PC18) - 5.38506(PC19) + 4.28686(PC22) + 2.59056(PC24) + 2.52710(PC26) - 7.68671(PC27)$$

The R^2 of the selected model was 94.98% indicating a very good prognostic model. For the purpose of assessing the goodness of fit of the resulted regression model a residual plot using the studentized residuals and a normal probability plot of the raw residual were plotted and are given in figures 1 and 2 respectively. Figure 1 clearly demonstrates that residuals are randomly plotted. Except for a few observations, almost all the points lie between – 2 and + 2 indicating a good fit. This figure suggests that error terms do not violate the assumption of homoscedasticity [10]. In reference to figure 2, a straight line can be observed. This linearity implies agreement with normality of the error terms and hence the assumption of Normality is satisfied.

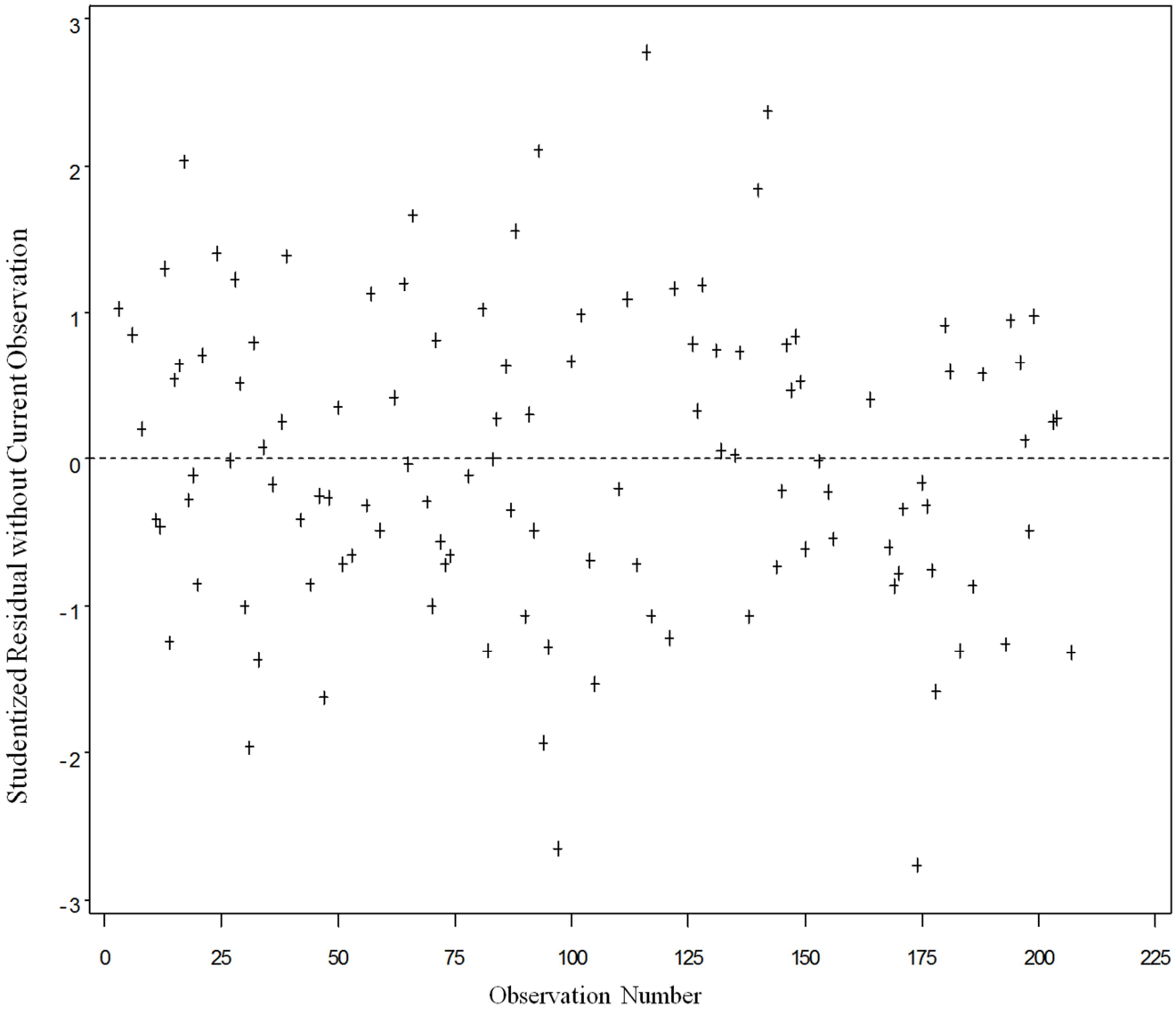

Figure 1. Studentized Residual Plot.

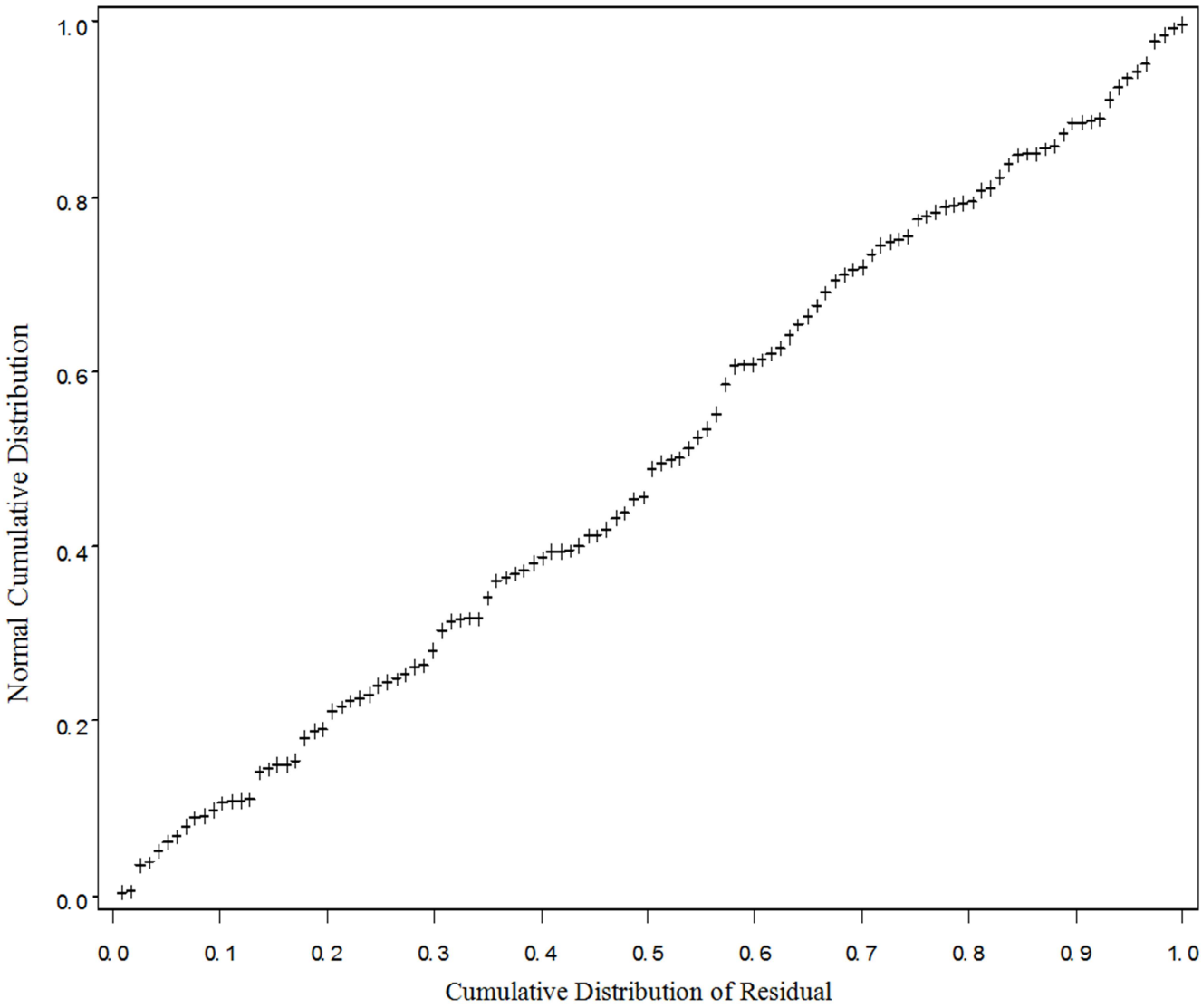

Figure 2. Normal Probability Plot.

Principal components 1, 2, 3, 4, 6, 8, 11, 12, 14, 15, 18, 19, 22, 24, 26 and 27 were significant at 5% level. The PCs 1, 6, 11, 15, 22, 24 and 26 have a positive influence on life expectancy values at birth whereas the PCs 2, 3, 4, 8, 12, 14, 18, 19 and 27 have a negative influence on life expectancy value at birth.

3.3. Interpreting PCs and Their Coefficients in the Model

When observing PC1 it clearly explains the distinction between developed and developing countries. Some of the positive loadings given with respect to PC1 are MV, TL, PA60PW, PA60PM and TR. The loadings with regard to PA014, IMR and TFR have shown negative values for PC1. All these variables are related to the development level of a country. Thus PC1 can be indicated as the development index. The parameter estimate for PC1 is positive and is 2.72. It illustrates that when the development index increases by one unit LE increases by 2.72 years provided that the other variables in the model are held constant.

PC2 depicts high positive loadings for GDP1, IU and CO_2E. These loadings imply the developed industrialized countries with high degree of pollution. Hence PC2 is named as the industrialization and pollution index. And a unit increment in this index would result in a reduction in expectation of life by 0.61 years.

LFPF and SR are the larger positive coefficients recorded with respect to PC3. These female dominated societies were mainly observed in the countries having/had war. As the parameter estimate for PC3 is negative in the regression model this war index emphasis a negative impact on LE at birth. When this index increases by one unit it would cause a reduction in life expectancy by 1.86 years.

In relation to PC4 significantly negative coefficients were identified for the variables GFCF and TS. It can be named as the low fixed assets and threatened species index. An

increment in this index by one unit affects negatively on the life expectancy as it reduces by 1.97 years.

PC6 is the densely populated or high population density index. It shows a very high positive coefficient for PD. When this factor increases by one unit LE will also increases by 1.23 years.

The next significant PC is PC8 which is the inflation factor. This is due to the very high positive coefficient relating to CPI in PC8. Though its coefficient is positive the parameter estimate corresponds to PC8 is negative indicating when the inflation factor is increased by one unit results a reduction in the expectation of life by 1.19 years.

RPGR is the most significant variable contained in PC11 where its coefficient is positive. And the coefficient in the regression model is also positive. This component was given a name called rural population growth rate factor. When this index is increased by one unit, LE will also increase by 0.90 years.

Referring to PC12 a high negative coefficient was observed for GEE. This corresponds to low government expenditure on education. When this increases by one unit LE reduces by 0.85 years.

When considering PC14, it depicted positive and negative loadings with regard to UP and UPGR respectively. It illustrates an urbanized but slow growth rate fact resulting in an index called slow growing urban population index. This index has a negative impact on life expectancy as it demonstrates an increase in this index by one unit will result in a reduction in life expectancy by 3.4 years.

A high positive coefficient against NPC and a high negative coefficient against ECPC was seen with respect to PC15. Thus this explains a balance between exports and imports and good communication by newspapers. This PC is a positively influential factor on life expectancy as its parameter estimate is positive. Hence when this index (exports over imports with communication) increases by one unit with others held constant, LE increases by 1.05 years.

PC18 can be regarded as a most influential PC since its coefficient in the model is comparatively large resulting in a p value of 0.0001. This PC shows significantly higher positive values for LFPM and EIS while a significantly lower value for RPGR. In other words this PC can be explained as depicting a low rural population growth rate, high percentage of labour force participation among adult males and high percentage employment in the industrial sector. LFPM includes eligible to work yet unemployed and excludes retired persons. Countries having a relatively young population will usually have a high LFPM. Most of the developed countries have an ageing population and many developing countries have a young population. A low rural population growth rate and high percentage of employment in the industrial sector indicates a low rural factor. Therefore this component gives high values for developing countries with a low rural factor. When this component increases by one unit LE reduces by 3.4 years.

PC19 is another vital factor which depicts a large influence on LE. This demonstrates significant positive coefficients with respect to IMR and ECPC whereas a significantly low coefficient can be seen with respect to LFPM. This PC represents the countries with high infant mortality ratio, higher imports over exports with low labor participation of males. As a result this PC provides an indication of poor economy and health conditions. Thus this can be regarded as poor economy and health index. When this is increased by one unit the LE is reduced by 5.39 years.

The parameter estimate for PC22 provides an indication of a severe impact from PC22 on LE. EAS has shown a large positive coefficient for PC22. Thus this can be identified with a component called agricultural index. Further it should be noted that an increment in PC22 by one unit would result in an increment in LE by 4.29 years.

When taking into consideration PC 24, a

number of positive and negative coefficients can be identified. GDP1, PA60PW, PA60PM and TFR are positive while GDP2, IMR, TR and CO_2E are negative. This large number of coefficients point out to two major issues: firstly a good health component and secondly a recently developed wealth component. When this entire component increases by one unit the LE also increases by 2.59 years.

PC26 has shown a significantly positive coefficient with regard to TFR and significantly negative coefficient with regard to IMR. Therefore this can be identified as high total fertility rate and low infant mortality rate indicating an increment in this PC26 by one unit causes an increment in LE by 2.53 years.

The last significant component PC27 can be identified as the most crucial as its impact on life expectancy is remarkable (p value = 0.0001). It shows a very high negative coefficient for PGR and it could be termed as a low population growth rate index. A unit increment in this index results in a reduction in LE by 7.69 years.

4. Discussion

This study was conducted with the intention of identifying the determinants of life expectancy at birth, using a sample of 117 UN countries [9]. This was achieved using multiple

linear regression modeling of life expectancy on the principal components obtained from the raw explanatory variables. In this analysis PCs were used instead of the raw variables, so as to overcome the problem of multi-collinearity. The important PCs were selected using the forward selection procedure. Contrasting to numerous studies which model the life expectancy at a person level, this study models life expectancy at a country level. The regression model fitted has a very good predictive power (R^2=94.98%) and the residual analysis indicates the goodness of fit and the validity of model assumptions. The selected PCs were interpretable and provided proxies for identifying the important factors effecting life expectancy. The regression coefficients were used to quantify the effect of the proxies on life expectancy. The results obtained were used to satisfy the objectives set out at the start of the study, namely, to examine the effect of established factors on life expectancy and to suggest new factors that effect life expectancy.

The literature was used as a guide to extract established factors that affect life expectancy. At the country level these are economic development, health facilities, technological advancement, education [3], urbanization and population density [8] which have all been found to have a positive influence on life expectancy. This study confirms most of these findings apart from two exceptions; the part of technological development associated with industrialization is related to pollution and thus has a negative effect on life expectancy and urbanization (percentage urban population) also contributes to a lower life expectancy. These findings on negative effects are also supported by Wu (2017)[13].

This study comes up with a number of other proxies that influence life expectancy. These being population growth rate (both rural and urban), communication facilities, employment in agricultural sector, rural life style and conservation of the environment [14, 15] which all have a positive effect on life expectancy while war and inflation have a negative effect on life expectancy.

5. Conclusions

5.1. Study Limitations

Though there were fifty variables in the data set, this study has focused only on thirty one selected variables due to the problem of missing values. However, even these thirty one variables included some missing values. A more informative model could have resulted if there was a possibility of considering all fifty variables. Also although the data set contained 209 countries only 117 countries were considered for modeling due to the missing values in the data.

5.2. Implications for Future Research and Policy

If data were available on all 209 countries for modeling, then it can be recommended to split this into a modeling data set and a test data set and the model validated. However in this study there were only 117 countries available for modeling due to missing values and this was inadequate for the task of model validation. Another suggestion for further work is to estimate the missing values using some method of imputation which will help in using all the variables and countries in the analysis.

This analysis provides information required by governments, especially in the developing world as the life expectancy at birth is predicted with high explanatory power by variables that can be influenced through public policy. To increase the life expectancy it is recommended; to improve the economy, education level, health facilities and communication facilities, to conserve the environment, to encourage a rural life style and employment in the agricultural sector. Further to enhance life expectancy industrialization, pollution, war and inflation should be reduced. In addition population growth should be encouraged in both the rural and urban sectors but urbanization should be controlled so as to improve the life expectancy.

The major new implications of the findings are:

i. The rural way of life increases LE. This was confirmed by indices for rural life style and employment in the agricultural sector contributing to increase LE with urbanization resulting in lowering of LE.

ii. Communications, particularly through news papers increases LE.

iii. Conservation of the environment (both forested area and threatened species) improves LE.

iv. Population growth rate has a positive impact on LE.

v. Inflation in the country results in lowering of LE among its citizens.

vi. Industrialization and pollution results in lowering of LE.

References

[1] Zhang, J., Zhang, J., and Lee, R. D. (2001). Mortality decline and long-run economic growth. *Journal of Public Economics*, *80* (3), 485-507.

[2] Coile, C., Diamond, P., Gruber, J., and Jousten, A. (2002). Delays in claiming social security benefits. *Journal of Public Economics*, *84* (3), 357-385.

[3] Baer, A., and Graves, P. E. (2002, June). Predicting Life Expectancy: A Cross-Country Empirical Analysis.

[4] Araki, S., and Murata, K. (1987). Factors Affecting the Longevity of Total Japanese Population. *Tohoku J. exp. Med.*, *151*, 15-24.

[5] Pourmalek, F., Abolhassani, F., Naghavi, M., Mohammad, K., Majdzadeh, R., Naeini, K. H., et al. (2009). Direct estimation of life expectancy in the Islamic Republic of Iran in 2003. *Eastern Mediterranean Health Journal*, *15* (1).

[6] Shaw, J. W., Horrace, W. C., and Vogel, R. J. (2005). The Determinants of Life Expectancy: An Analysis of the OECD Health Data. *Southern Economic Journal*, *71* (4), 768-783.

[7] Ho, J. J., Hwang, J. S., and Wang, J. D. (2006). Life expectancy estimations and the determinants of survival after 15 years of follow-up for 81,249 workers with permanent occupational disabilities. *Scand J Work Environ Health*, *32* (2), 91-98.

[8] Klitgaard, R. (1985). *Data Analysis for Development.* Oxford University Press.

[9] *World Statistics Pocketbook* (Vol. 28). (2005). New York, United States of America: United Nations Publications.

[10] Draper, N. R., and Smith, H. (1981). *Applied Regression Analysis.* United States of America: Courier Companies, Inc.

[11] Johnson, R. A., and Wichern, D. W. (2003). *Applied Multivariate Statistical Analysis.* Printice-Hall, Inc.

[12] Halaka, F. G., Babcock, G. T., and Dye, J. L. (1985). The use of Principal Component Analysis to resolve the spectra and kinetics of cytochrome c oxidase reduction by 5,10-dihydro-5-methyl phenazine. *BIOPHYS. J.*, *48*, 209-219. (Chanana & Talwar, 1987).

[13] Wu, C. (2017). Human capital, life expectancy, and the environment. *THE JOURNAL OF INTERNATIONAL TRADE & ECONOMIC DEVELOPMENT,* 26 (8) 886-906.

[14] Clootens, N. (2017). Public Debt, Life Expectancy, and the Environment. *Environmental Modeling & Assessment*. 22 (3): 267-278.

[15] Taskaya, S. and Demirkiran, M. (2016). Environmental determinants of life expectancy at birth in Turkey. *Int J Res Med Sci* 4 (4): 995-999.