What Country Level Factors are Associated to Average Life Expectancy in a Country?

Trevor Freeland, Piper Kurtz

May 29, 2018

## Abstract

Things matter

## Introduction

The life expectancy of the citizens of a country is one of the most important indicators of well-being of a country, as people who live longer will not only produce longer within society, but will likely be more productive as well, due to being in a better state of health. States across th globe are unable to focus on important issues such as education or economic growth, as they must first prioritize the health of their citizens. It is in the best interest of every country, and every individual within each country, to identify and improve the factors that increase their life expectancy. Historically, it has been widely accepted that higher levels of GDP can be associated with higher life expectancy (Schnabel and Eilers, 2009). However, there is much more to life expectancy than simply how much wealth a nation holds. How much money a goverment puts towards healthcare, both in relation to its total expenditures and compared to the public, greatly influence how long a citizen of that country can be expected to live (Jaba, Balan, Robu, 2014). When a country spends a greater share of its expenditures on healthcare, theoretically, it will have citizens with longer lives, as they should have access to better healthcare than they would otherwise.

This study aims to analyze the relationship between life expectancy of countries around the world with various possible influencing variables, specifically labor force participation rate, population, GDP per capita, food supply, growth rate, government spending on healthcare as a percent of total expenditures, and government spending on healthcare as a percent of total expenditure on healthcare. With the goal of identifying what variables influence life expectancy, we will apply a linear regression model with life expectancy as our dependent variable. In order to estimate the relationship between the various variables and life expectancy, we used country level data from 1995 through 2007. We will describe the dataset in detail in our methods section below.

The paper begins by a general reference to previous studies on the relationship between life expectancy and the independent variables. Then, it lays out the variables and the sample of countries, as well as the lineaer model regression approach. Finally, the main empirical results are presented and the final conclusions are drawn. ##Background

### Theory

Countries that are more “well-off” will probably live longer.

### Previous Work

GDP talk about how it is afffected and affects life expectancy, why we chose to put as independent variable <https://www.jstor.org/stable/pdf/20062079.pdf?refreqid=excelsior%3A0f47169a4aafc4701062c76077b5b13e>

Since Auster, Leveson, and Sarachek, several economic studies have attempted to answer these questions using data from the United States or multiple countries.1 Many of these have used aggregate data from the member countries of the OECD to explain cross-country mortality rates or life expectancies.2 While the empirical results are mixed, the general consensus is that population life expectancy (or mortality) is a function of environmental measures (e.g., wealth, education, safety regulation, infrastructure), lifestyle measures (e.g., tobacco or alcohol consumption), and health care consumption measures (e.g., medical or pharmaceutical expenditures). However, the appropriate econometric methodology for disentangling these effects and its meaning for the relative importance (statistical or economic) of the estimated effects is more contentious.

Referencing a paper from Peltzman in 1987, hard to find, will find later Peltzman also considered a GLS regression of life expectancy at birth on wealth and government health expenditures and found only wealth to be a significant determinant. His life expectancy variable was an average for the entire population (including males and females) of each country and was only for a single age stratum (at birth) in each country.

## Methods

The data we used was assembled from individual variable datasets that were downloaded from Gapminder World. The data we used from gapminder was provided to them by the World Bank, the World Health Organization and Gapminder themselves. We combined these smaller data sets into a data set containing all of the variables that we were intested in examining. Our data has various country level information taken between the years 1995 and 2007. Each observation in our data set is our set of country level information, like population or gdp per capita, for a specific country in that specific year. Overall our data contains 1944 complete observations with 10 variables per observation.

For a full overview of each individual variable in our data set, see description below:

country = Name of Country  
year = Year our variables were measured for  
lifeExp = Average life expectancy (years) (**Response Variable**) Population = Total population GDPCap = GDP/capita (US$ inflation-adjusted)  
FoodSupply = Kilocalories available on average for each person each day (Kilocalories)  
growth = GDP growth per year (%)  
GovHealthSpending = Government health spending as % of total gov. spending  
GovShareHealthSpend = Government share of total health spending (%)  
LaborForce = Age 15+ labour force participation rate (%)

## Results

Start this sections with data visualization. Then move on to say final model and give summary of model, then interpretations.

## Discusion

### Robustness

Our model seems to do worse for low fitted values for life expectancy than for higher values of life expectancy. There also appears to be a slight negative curvature in our residuals, indicating that there could be some higher order term missing in our model, or possibly of some geographical relationship that we are not accounting for.

When we checked our model for autocorrelation using the Breusch-Godfrey test where we got a p-value of .0003. This led us to test our coefficients with the Newey-West standard errors. When running a coefficient test with the Newey-West standard errors, none of the significance of any of our variables changed.

We also checked for heteroskedasticity in our model. Using the Breusch-Pagan test we could determine that there was heteroskedasticity in our model. When we corrected for the heteroskedasticity using the hccm function in R we ran another coefficient test and none of the significance of any of our variables changed again. Meaning all of our variables appear to stand the test of time.

This is where we can talk about any heteroscedasticity we found, which I am sure we will have, we will also probably have some sort of geographical relationship for our residuals which we just won’t be correcting for, but could with random effects.

### Limitations and Future Work

With this type of study we must make it clear that we are looking at association and not causation. Our results do not indicate that by increasing population in a country is going to cause an increase in average life expectancy in a country. Our results suggest that population size is associated with higher life expectancy, not the cause of it.

Another limitation of our work is that we could be missing explanatory variables in our initial model and so some of the associations that we found may be significant only because we are missing other explanatory variables. If we had more time we would have liked to found more observations that contained some other variables that were mentioned in some of the papers like smoking or alcohol use. These kinds of lifestyle variables could definitely have an effect on life expectancy but we were unable to find enough country level data on these lifestyle variables to use them in our model.

It also would have been interesting to look at the possible different factors that are significant in different geographical areas of the world. For example fitting a model for average life expectancy in North America and fitting a model for life expectancy in Africa and then comparing what variables were significant in either model and trying to figure out why some things may have been significant in one model and not the other.

In the future it would be interesting to also look at this model with a heirarchichal linear regression model. This type of model would be able to account for differences between country and continent over time and we could then examine if some of the variables we deemed significant in our model would still be significant. This also has the potential to change the effects of our variables after adding all of the random effects.

## Appendix (If needed)

## Works Cited/Bibliography or whatever this section is supposed to be called.