An investigation into use of machine learning techniques as a tool in the study of the factors that affect life expectancy on a global scale

Adriaan Louw (53031377)

June 8, 2019

Abstract

The pursuit of increasing ones life expectancy is a very human endeavour. This study examines the tools used to study the conditions that affect life expectancy of countries. Investigating socio-economic indicators like income, educational attainment and per capita spending on healthcare. Understanding these relationships can lead a country to better a understand of which socio-economic areas to focus on, in order to increase their life expectancy. For this task this study advocates the use of certain machine learning algorithms that could be of greater use than the traditional regression techniques used in research.

Contents

List of Figures

1 Introduction

Human beings have always had a fascination with longevity. Myths like the fountain of youth or the Holy Grail are a testament to this fact. Today, longevity and causes of mortality are studies by professionals like Demographers and Actuaries. Trying to determine why some people or group live long lives.

This study investigates the use of Machine Learning techniques in studying determinants of life expectancy for countries. Indicators that have shown to have some form of correlation with life expectancy will be selected. Their relationship with life expectancy will be investigated using various techniques from Machine Learning and contrasted to various forms of regression whose use is ubiquitous in the literature. This analysis will seek out to prove the appropriateness of using these machine learning algorithms for use in research to find the exact correlation between these indicators and life expectancy. It is the hypothesis of this study that machine learning techniques like k-Nearest Neighbour and Support Vector Machines will model the indicator/life expectancy relationship better than regression techniques can. Also that these techniques can create more accurate models. In the hope that the causes of long life expectancies in certain countries can be better understood (Chen & Asch 2017). This study does not aim to prove causation between the indicators chosen and life expectancy, but rather the usefullness of machine learning algorithms as tools.

Statistical regression techniques are predominantly used algorithm for data analysis. Linear regression (Section 3.2.1) forinstance assumes a linear relationships between the independant variables and the dependant variable. Which might not be the case. This we will discuss in Literature Review (Section 2).

2 Literature Review

Life expectancy and mortality are 2 related terms. Mortality is generally expressed as a mortality rate. It describes the rate at which people die under certain circumstances. Life expectancy (Section 2.1) is the amount of years an individual or group of people are expected to live. If the amount of people who are dying increases, the mortality rate increases and the life expectancy for people in that group decreases and visa versa.

2.1 What do we mean by life expectancy?

A life table is a table given for a specific year that contains the probability that a person of a certain age will die in that specific year. Acuataries and Demographers use life tables in the insurance industry and the study of demographics respectively. Life expectancy is one element of a life table. Both countries and the United Nations create life tables for use in policy creation.

There are 2 types of life tables, namely period and cohort life tables. A cohort is a group of people who were born in the same year. A cohort life table will follow a cohort over its lifetime until every member of the cohort has died. A cohort life table requires the mortality information of the cohort over many years. This information is often unavailable. While for a period life table, a hypothetical cohort is created and subjected to current mortality rates. This gives the user of the period timetable a window to see mortality rates at that point in time. This makes period life tables the most common type (Arias et al. 2017).

A life expectancy figure from a period life table is called a period life expectancy.

2.2 Multi indicator studies

In this section, studies that used various indicators will be discussed.

Kabir (2008) investigated how well the life expectancy of 93 developing countries were predicted by indicators like income,education and fertility (among others). It classied a countries life expectancy into 3 categories. Then used a probit model where the input variables have a linear relationship. Multiple Ordinary Least Squares Regression was then applied to study inicators' influences.

The study Hu et al. (2015), also used a linear regression model of GDP per capita, Gini indeces, ect with respect to life expectancy. The intention of the study was to link income inequality to mortality rates and life expectancy.

While Shaw et al. (2005) investigated factors like smoking, pharmasutical spending, amount of butter consumed and amount of fruit and vegetables consumed. Then putting them in a linear model and applying regression.

In an examination of 108 countries, Hassan et al. (2016) investigated indicators like education, GDP and health spending as it relates to life expectancy. They used Grossmans model (Grossman 2000) to model their data and Vector Error Correction Model to anlyse the data.

By using the componets of the Human Development index and Pearsons r with multivariate regression, Bulled & Sosis (2010) investigated the link between life expectancy, education and reproduction.

2.3 Determinants of life expectancy

This section summarises the relationship some indicators have with life expectancy.

2.3.1 Income

The relationship between income and life expectancy has been given a lot of attention in academic circles (Preston 1975, Hu et al. 2015, Chetty et al. 2016, Oeppen 2019).

Preston (1975) was the first to show the correlation between life expectancy and per capita income. His original curve can be seen in Figure 1. As we can see from Figure 1, for low income countries, life expectancy increases rapidly with per capita income. Whereas in high income countries a small increase in per capita income does not have a large effect on life expectancy.

This relationship has also been shown in more recent studies (Chetty et al. 2016, Oeppen 2019). Even though Shkolnikov et al. (2019) found that in Russia, the Preston curve is not an accurate predictor of life expectancy, they found that the actual life expectancy should be "substantially higher" when comparing to the Preston curve predicted value.

Studies in first world countries involving mortality rather that life expectancy have also found a relationship with income level (Blakely et al. 2004, Kalwij et al. 2013, von Gaudecker & Scholz 2007).

Just 16% of the improvement in life expectancy between the 1930s and 1960s could be explained by rising income levels Preston (1975). Which seems to indicate that a countrie's life expectancy is dependant on more than income levels.

2.3.2 Education attainment

Kaplan et al. (2015) investigated the relationship between educational attainment and life expectancy in eight states in the United States. They found that even when controlling for variables like income, race, sex, and common medical issues like cardiovascular disease, the relationship between educational antainment and life expectance remains statistically significant.

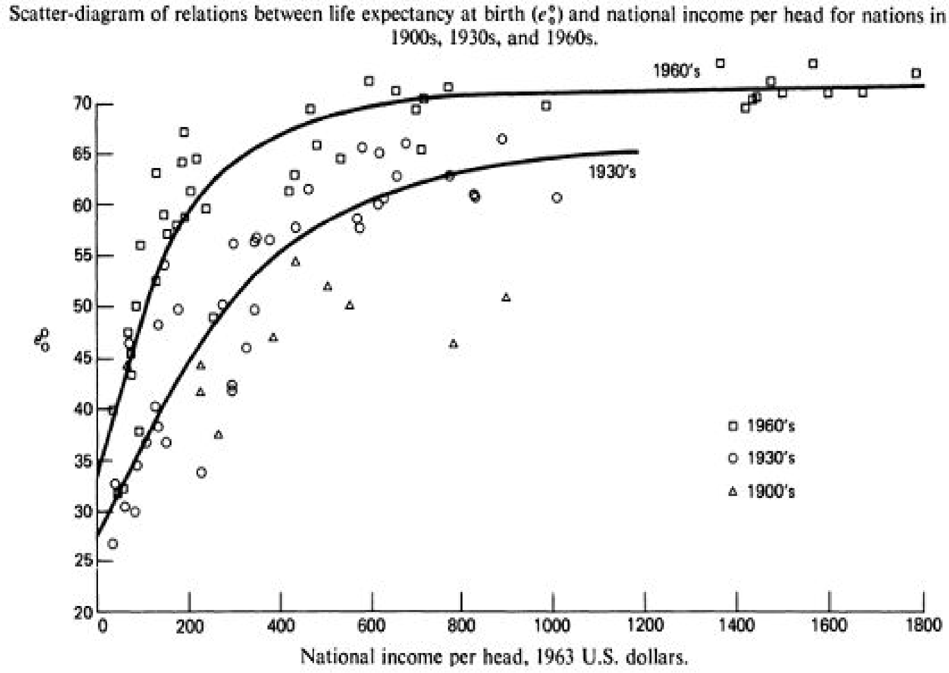

Figure 1: The original Preston curve from Preston (1975)

Luy et al. (2019) studied 3 developed nations, namely the United States, Italy and Denmark. They have also found a strong correlation between education levels and longevity.

But what is the nature of this correlation? According to Deary & Gottfredson (2004) Intelligence Quotient or IQ could explain the association. While Hayward et al. (2015) does not believe in a "causal relationship", but rather that it depends on factors like "time, place, and the social environment".

In an attempt to find a causal relationship between education and life expectancy, van Kippersluis et al. (2011) investigated the result of the Netherlands increasing the mandatory number of years a child had to attend school for to 7 years. It was 6 years previously. van Kippersluis et al. (2011) found a decrease in mortality of 3% for 81-year old males who had the additional year of schooling.

This relationship appears strongest in more developed countries where the life expectancy is already above 60 years (Bulled & Sosis 2010). In these countries, any educational investment leads to greater compensation for the learner than they would get in a less developed country Bulled & Sosis (2010), Handwerker (1986). In addition, Kabir (2008) also studied this relationship, among others, with regards to developing countries and did not find a correlation.

The question remains, which educational indicators should be used when investigating the relationship between education and life expectancy?

Various educational indicators have been used in the literature for comparing to life expectancy. One approach is to use the International Classification of Education (ISCED) system (UNESCO Institute for Statistics 2012). The ISCED 2011 standard consists of 9 levels ranging from ISCED level 0 (Early childhood education) to ISCED level 8 (Doctoral or equivalent level).

Luy et al. (2019) used the United Nations ISCED-97 (consisting of 7 levels) scale to break education attainment down into 3 levels namely Low (None to Lower Secondary), Medium (Upper secondary) and High (Tertiary education). In van Kippersluis et al. (2011) the Dutch SOI system (Standaard Onderwijs Indeling). Which, according to van Kippersluis et al. (2011), is similar to the ISCED system. While in Deboosere et al. (2009), educational attainment was broken into 5 levels, also ranging from no education to Tertiarty education.

Kaplan et al. (2015) broke educational attainment into 4 levels ranging from less than high school to college graduate.

In the study Bulled & Sosis (2010), the relationship between educational investments and fertility against life expectancy, over 193 countries, was investigated. They used adult literacy and the enrolment ratios for primary,secondary and tertiary schooling.

2.3.3 Spending on health

Healthcare spending and life expectancy in the Unites States, between 1960 and 2000, was compared in Cutler et al. (2006). They found that increased spending on health per capita, controlling for inflation, is positively correlated to US life expectancy for the time period in question.

Most Eastern European countries, who have joined the European Union, have seen an increase in healthcare spending. This has generally been accompanied by an improvement in life expectancy (Jakovljevic et al. 2016). This has to be seen in the context of the so called "Russian Mortality Crisis" where former Soviet Union countries faced a sudden drop in life expectancy after the fall of the Berlin wall (Brainerd & Cutler 2005). Jakovljevic et al. (2016) found that the best metric to use when comparing health spending of countries, is to use their total per capita health spending (in US dollars).

The same relationship was found in Canada. When spending on healthcare is decreased, life expectancy follows (Crémieux et al. 1999).

It is well known that life expectancy in Sub-Saharan Africa is low. Here spending on health care can also be correlated to increases in life expectancy. Even though poor governance can undo some of the effects of increased spending (Makuta & O'Hare 2015).

A country's per capita healthcare is not necessarily in proportion to its per capita income. In 2005, the United States spent 50% more on healthcare per capita than its income per capita would suggest (Anderson & Frogner 2008).

2.3.4 Unemployment

According to Bonamore et al. (2015), the literature has 2 main views on the relationship between unemployment and life expectancy. The first view states that during an economic downturn, people suffer from more stress and depression. This leads to more unhealthy lifestyle choices like smoking and alcohol. Which in turn lowers life expectancy. Bonamore et al. (2015) cites the works of Lundin et al. (2014), Montgomery et al. (2013), Garcy & Vågerö (2012), Browning & Heinesen (2012), Dávalos et al. (2012), Backhans & Hemmingsson (2011), Deb et al. (2011) and Strully (2009), who take this view. The second view focusses on times when there is economic growth, i.e. less unemployment. This period of economic growth also can lead to stress eg. burning out and having less time for activities that benefit ones health. Like going to the gym. This view is held by Tapia Granados & Ionides (2011), Ruhm (2005), Tapia Granados (2005), Neumayer (2004) and Ruhm (2000). Then there are also studies express the view that no connection can be established (Bonamore et al. 2015). The view of Bonamore et al. (2015) is that this relationship is non-linear.

3 Methodology/Procedure

This study will attempt to create a model that can predict life expectancy for a country based on various socio-economic conditions in the country over a 30-year period. Then evaluate how well various machine learning techniques model these relationships. Unlike Shaw et al. (2005), this study will not take into account the age distribution of each country.

The philosophical standpoint of this study is Positivism. By using the scientific method, this study will comprise of an experiment to inductively determine whether machine learning techniques can provide more accurate life expectancy models than those created using regression. This cross-sectional study will use life expectancy indicators shown, from the literature, to have some correlation to life expectancy.

Firsly appropriate data sets will be chosen. Taking into account that some indicators might have sparse information. This will impact the size of the dataset. Looking at available indicators like World Bank Group (2019*c*), we can see there are approximately 200 countries. If all the indicators have data for each year in the study there will be around 6000 data points. Looking at indicators like World Bank Group (2019*c*) and World Bank Group (2019*a*), it is clear that the data is quite sparse, especially for developing countries. The intention is to have the study encompass developed and developing countries, but this will depend on data availability.

Once the dataset is finalised, the algorithms and their permutations described in Section 3.2 will be applied. Three models will be created for each algorithm and permutation thereof. The first model will be using all the data. The second will only use data from developed countries and the third will only use data from developing countries, data permitting. Cross-validation will be applied to the created models as described in Section 3.3.

The results from the regression analyses will then be compared to the results form the machine learning algorithms and that from the literature review above.

3.1 Choice of dataset

3.1.1 Life Expectancy

For life expectancy data, this study will use the indicator "Life expectancy at birth, total(years) {SP.DYN.LE00.IN}" from the World Bank's Development Indicators Database (World Bank Group 2019*h*). This is a weighted average combining both male and female life expectancy and is calculated in a period life table (see Section 2.1). Only the data that is available for the last 30 years (1981-2010) will be used. This is in keeping with other studies where the amount of years that their studies look back on is limited to relatively recently (Luy et al. 2019, Hu et al. 2015, Tarkiainen et al. 2012, Kabir 2008, Shaw et al. 2005).

3.1.2 Income

This study will use GDP per capita as was used in Oeppen (2019), Shkolnikov et al. (2019), Mackenbach & Looman (2013) and De Vogli et al. (2005). The source will be GDP per capita, PPP (constant 2011 international $){NY.GDP.PCAP.PP.KD} from the World Bank (World Bank Group 2019*b*).

3.1.3 Educational Attainment

Just as in Bulled & Sosis (2010), this study will use adult literacy, and primary, secondary and tertiary enrolment ratios. For an adult literacy indicator, SE.ADT.LITR.ZS will be used from the world bank website (World Bank Group 2019*c*). This indicator describes the percentage of adults, from the age of 15, who can read and write to a certain level of proficiency. The indicator SE.PRM.ENRR will serve as the enrollment ratio for primary education (World Bank Group 2019*d*). For secondary enrolment, SE.SEC.ENRR (World Bank Group 2019*e*) will be used and SE.TER.ENRR (World Bank Group 2019*f*) for tertiary enrollment. The world bank credits UNESCO Institute of Statistics (2019) for all of theses educational indicator data.

3.1.4 Per Capita spending on health

To represent Per capita spending on health in this study's model, SH.XPD.CHEX.PP.CD (World Bank Group 2019*a*) which describes Current health expenditure per capita, will be used.

3.1.5 Unemployment

Unemployment will be ignored for this study because sufficient data on unemployment per country is not available. This can be seen by looking at indictors like SL.UEM.TOTL.NE.ZS (World Bank Group 2019*g*) which describe percentage unemployment per country for both sexes.

3.2 Choise of algorithms

3.2.1 Algorithms Chosen

Regression Linear Regression is a popular technique, used to find relationships in data. As the name suggests Linear Regression assumes a linear relationship between the input variables and the result (Murphy 2012). This might not be the case for the target function. The target function could be any potential function. In the case of life expectancy modelling, we know that according to the Preston curve (Section 2.3.1), the relationship between income and life expectancy is not linear. Thus using Linear Regression should return a sub-optimal result. The same logic applies to Logistic Regression. It assumes a linear relationship between inputs. The difference is that this linear sum is passed through the sigmoid function (Murphy 2012). This also makes it inappropriate for non-linear target functions. In this study Linear and Logistic Regression will be used as a baseline for comparison on the dataset.

k-Nearest Neighbour The k-Nearest Neighbour algorithm (kNN) is an instance based form of machine learning. It uses the classification of those datapoints closest to the data point to be classied to determine its classification. The kNN-algorithm allows for non-linear problem spaces to be classified, because it does not make an assumption on the nature of the problem space. Additionally, how the algorithm determined its output value is transparent and can be used to study how various components affect the end result. In this study, the standard kNN-algoritm will be altered to accomodate a real valued output and not just a class classification. This will be accomplished by taking the mean life expectancy for all the data points determined to be closest to the target point. Care will have to be taken to reduce the number of features of the data, because this algorithm is sensitive to the so-called "curse of dimentionality" (Mitchell 1997).

Support Vector Machines The classic Support Vector Machine (SVM) is used in classification tasks. It involves determining the decision surface with regards to the data points closest to the surface. This study will use a modified SVM algorithm that makes it useful for regression tasks. The SVN regression is described in Smola & Schölkopf (2004). The final model can be described as:

$$f(x) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) k(x_i, x) + b \tag{1}$$

where α_i and α_i^* are Lagrange multipliers, b is the bias and the term $k(x_i, x)$ is the kernel. The kernel is a function to determine the similarity between 2 input vectors. Various kernels will be investigated, including the Radial-basis function

$$k(x_i, x) = exp(-\frac{\| x - x_i \|^2}{2\sigma^2}) \quad (2)$$

and the polynomial kernel

$$k(x_i, x) = (x_i.x + 1)^p \quad (3)$$

3.2.2 Ignored Algorithms

Neural Networks Even though Neural networks are capable of representing non-linear hypothesis spaces (Mitchell 1997), they are not appropriate for this study for a couple of reasons. Firstly, the datasets that are available are not large enough. Neural networks typically require thousands if not tens of thousands of data points. Secondly, the amount of processing power and processing time required, will not be available to this study. Thirdly, the results of neural networks are hard to interpret. How the Neural Network came to its conclusion is not clear to the researcher. Which makes it unsuitable as a tool to study the relationship between life expectancy and its various indicators.

Decision Trees Traceability and understandability are some of the hallmarks of Decision Trees. These algorithms are suited problem spaces where the target function and the input attributes are discrete values. It is possible to approximate continuous input attributes by making a branch in the tree when a value is smaller or greater than some value, or is between some value. For functions where input attributes span over large ranges, this leads to very large and sub-optimum trees (Mitchell 1997). The problem of determining life expectancy from socio-economic indices has a continuous target function output and continuous input attributes. Therefor, Decision Trees will be excluded from this study.

Clustering Clustering techniques are a form of unsupervised learning where the algorithm tries to group the data in a number of groups. In essence, classifying the data into categories. This study will not attempt to classify the data, but rather find appropriate machine learning algorithms that could replace linear and logistic regression. While at the same time being able to predict the output of an input vector without needing to classify it first.

3.3 Cross-validation

By using stratified k-fold cross validation, this study will aim to reduce the impact of the relative small dataset that will be analysed. This form of cross validation will ensure that when the validation set is chosen, no important data points are ignored for training. The data will be broken down randomly into k subsets of equal size. Each data subset will also contain equal amounts of data points with low and high life expectancies, so that no dataset is completely towards one end of the data range. One data subset is chosen to be the validation subset and the remaining $k-1$ subsets are combined into the training set. The model is then trained on the training dataset andt its performance is measured against the validation subset. This is done k times in order for each subset to be the validation subset. For each of the training runs the mean of the error will be calculated (Mitchell 1997, Murphy 2012). The value of k will be dependant on the final dataset.

References

Anderson, G. F. & Frogner, B. K. (2008), 'Health spending in OECD countries: Obtaining value per dollar', *Health Affairs* **27**(6), 1718–1727.

Arias, E. (2007), National Vital Statistics Reports, Technical Report 6.

Arias, E., Heron, M. & Xu, J. (2017), 'United States Life Tables, 2013.', *National vital statistics reports : from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System* **66**(3), 1–64.

Backhans, M. C. & Hemmingsson, T. (2011), 'Unemployment and mental healthwho is (not) affected?', *The European Journal of Public Health* **22**(3), 429–433.

Blakely, T., Kawachi, I., Atkinson, J. & Fawcett, J. (2004), 'Income and mortality: The shape of the association and confounding New Zealand Census - Mortality study, 1981-1999', *International Journal of Epidemiology* **33**(4), 874–883.

Bonamore, G., Carmignani, F. & Colombo, E. (2015), 'Addressing the unemployment-mortality conundrum: Non-linearity is the answer', *Social Science and Medicine* **126**, 67–72.

Brainerd, E. & Cutler, D. M. (2005), Autopsy on an Empire: Understanding Mortality in Russia and the Former Soviet Union, Technical Report 1.

Browning, M. & Heinesen, E. (2012), 'Effect of job loss due to plant closure on mortality and hospitalization', *Journal of Health Economics* **31**(4), 599–616.

Bulled, N. L. & Sosis, R. (2010), 'Examining the Relationship between Life Expectancy, Reproduction, and Educational Attainment', *Human Nature* **21**(3), 269–289.

Chen, J. H. & Asch, S. M. (2017), 'Machine Learning and Prediction in Medicine Beyond the Peak of Inflated Expectations', *New England Journal of Medicine* **376**(26), 2507–2509.

Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A. & Cutler, D. (2016), 'The association between income and life expectancy in the United States, 2001-2014', *JAMA - Journal of the American Medical Association* **315**(16), 1750–1766.

Crémieux, P.-Y., Ouellette, P. & Pilon, C. (1999), 'Health care spending as determinants of health outcomes', *Health Economics* **8**(7), 627–639.

Cutler, D. M., Rosen, A. B. & Vijan, S. (2006), 'The Value of Medical Spending in the United States, 19602000'.

Dávalos, M. E., Fang, H. & French, M. T. (2012), 'EASING THE PAIN OF AN ECONOMIC DOWNTURN: MACROECONOMIC CONDITIONS AND EXCESSIVE ALCOHOL CONSUMPTION', *Health Economics* **21**(11), 1318–1335.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/hec.1788*

De Vogli, R., Mistry, R., Gnesolto, R. & Cornia, G. A. (2005), 'Has the relation between income inequality and life expectancy disappeared? Evidence from Italy and top industrialised countries', *Journal of Epidemiology and Community Health* **59**(2), 158–162.
URL: *http://jech.bmj.com/*

Deary, I. J. & Gottfredson, L. S. (2004), 'Intelligence Predicts Health and Longevity, but Why?', *Current Directions in Psychological Science* **13**(1), 1–4.

Deb, P., Gallo, W. T., Ayyagari, P., Fletcher, J. M. & Sindelar, J. L. (2011), 'The effect of job loss on overweight and drinking', *Journal of Health Economics* **30**(2), 317–327.

Deboosere, P., Gadeyne, S. & Van Oyen, H. (2009), 'The 1991-2004 evolution in life expectancy by educational level in Belgium based on linked census and Population register data', *European Journal of Population* **25**(2), 175–196.

Garcy, A. M. & Vågerö, D. (2012), 'The length of unemployment predicts mortality, differently in men and women, and by cause of death: A six year mortality follow-up of the Swedish 19921996 recession', *Social Science & Medicine* **74**(12), 1911–1920.

Grossman, M. (2000), THE HUMAN CAPITAL MODEL, *in* 'Handbook of Health Economics, Volume 1', Vol. 1, pp. 348–407.

Handwerker, W. P. (1986), 'The Modern Demographic Transition: An Analysis of Subsistence Choices and Reproductive Consequences', *American Anthropologist* **88**(2), 400–417.

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), 105–117.

Hayward, M. D., Hummer, R. A. & Sasson, I. (2015), 'Trends and Group Differences in the Association between Educational Attainment and U.S. Adult Mortality: Implications for Understanding Education's Causal Influence *', *Soc Sci Med* **127**, 8–18.

Hu, Y., van Lenthe, F. J. & Mackenbach, J. P. (2015), 'Income inequality, life expectancy and cause-specific mortality in 43 European countries, 19872008: a fixed effects study', *European Journal of Epidemiology* **30**(8), 615–625.

Jakovljevic, M. B., Vukovic, M. & Fontanesi, J. (2016), 'Life expectancy and health expenditure evolution in Eastern EuropeDiD and DEA analysis', *Expert Review of Pharmacoeconomics and Outcomes Research* **16**(4), 537–546.

Kabir, M. (2008), 'Determinants of Life Expectancy in Developing Countries', *The Journal of Developing Areas* **41**(2), 185–204.

Kalwij, A. S., Alessie, R. J. M., Knoef, M. G., Kalwij, A. S., Alessie, R. J. M. & Knoef, M. G. (2013), 'The Association Between Individual Income and Remaining Life Expectancy at the Age of 65 in the Netherlands', *Demography* **50**(1), 181–206.

Kaplan, R. M., Howard, V. J., Safford, M. M. & Howard, G. (2015), 'Educational attainment and longevity: Results from the REGARDS U.S. national cohort study of blacks and whites', *Annals of Epidemiology* **25**(5), 323–328.

Lundin, A., Falkstedt, D., Lundberg, I. & Hemmingsson, T. (2014), 'Unemployment and coronary heart disease among middle-aged men in Sweden: 39 243 men followed for 8years', *Occupational and Environmental Medicine* **71**(3), 183 LP – 188.

Luy, M., Zannella, M., Wegner-Siegmundt, C., Minagawa, Y., Lutz, W. & Caselli, G. (2019), 'The impact of increasing education levels on rising life expectancy: a decomposition analysis for Italy, Denmark, and the USA', *Genus* **75**(1), 11.

Mackenbach, J. P. & Looman, C. W. (2013), 'Life expectancy and national income in Europe, 1900-2008: An update of Preston's analysis', *International Journal of Epidemiology* **42**(4), 1100–1110.

Makuta, I. & O'Hare, B. (2015), 'Quality of governance, public spending on health and health status in Sub Saharan Africa: a panel data regression analysis', *BMC Public Health* **15**(1), 932.

Mitchell, T. M. (1997), *Machine Learning*, McGraw-Hill international editions - computer science series, McGraw-Hill.

Montgomery, S., Udumyan, R., Magnuson, A., Osika, W., Sundin, P.-O. & Blane, D. (2013), 'Mortality following unemployment during an economic downturn: Swedish register-based cohort study', *BMJ Open* **3**(7), e003031.

Murphy, K. P. (2012), *Machine Learning: A Probabilistic Perspective.*

Neumayer, E. (2004), 'Recessions lower (some) mortality rates:: evidence from Germany', *Social Science & Medicine* **58**(6), 1037–1047.

Oeppen, J. (2019), Life Expectancy Convergence Among Nations Since 1820: Separating the Effects of Technology and Income, *in* T. Bengtsson & N. Keilman, eds, 'Old and New Perspectives on Mortality Forecasting', Springer International Publishing, Cham, pp. 197–219.

Preston, S. H. (1975), 'The Changing Relation between Mortality and level of Economic Development', *Population Studies* **29**(2), 231–248.

Ruhm, C. (2000), 'Are Recessions Good for Your Health?', *The Quarterly Journal of Economics* **115**(2), 617–650.

Ruhm, C. J. (2005), 'Healthy living in hard times', *Journal of Health Economics* **24**(2), 341–363.

Shaw, J. W., Horrace, W. C. & Vogel, R. J. (2005), The Determinants of Life Expectancy: An Analysis of the OECD, Technical Report 4.

Shkolnikov, V. M., Andreev, E. M., Tursun-zade, R. & Leon, D. A. (2019), 'Patterns in the relationship between life expectancy and gross domestic product in Russia in 200515: a cross-sectional analysis', *The Lancet Public Health* **4**(4), e181–e188.

Smola, A. J. & Schölkopf, B. (2004), 'A tutorial on support vector regression', *Statistics and Computing* **14**(3), 199–222.

Strully, K. W. (2009), 'Job loss and health in the U.S. labor market.', *Demography* **46**(2), 221–246.

Tapia Granados, J. A. (2005), 'Increasing mortality during the expansions of the US economy, 1900-1996.', *International journal of epidemiology* **34**(6), 1194–1202.

Tapia Granados, J. A. & Ionides, E. L. (2011), 'Mortality and Macroeconomic Fluctuations in Contemporary Sweden', *European Journal of Population / Revue européenne de Démographie* **27**(2), 157–184.

Tarkiainen, L., Martikainen, P., Laaksonen, M. & Valkonen, T. (2012), 'Trends in life expectancy by income from 1988 to 2007: decomposition by age and cause of death', *Journal of Epidemiology and Community Health* **66**(7), 573 LP – 578.

UNESCO Institute for Statistics (2012), *International Standard Classification of Education ISCED 2011.*

UNESCO Institute of Statistics (2019), ‘UNESCO Institute of Statistics’.
URL: *http://uis.unesco.org/*

van Kippersluis, H., O’Donnell, O. & van Doorslaer, E. (2011), ‘Long Run Returns to Education: Does Schooling Lead to an Extended Old Age?’, *Journal of Human Resources* **46**(4), 695–721.

von Gaudecker, H. M. & Scholz, R. D. (2007), ‘Differential mortality by lifetime earnings in Germany’, *Demographic Research* **17**, 83–108.

World Bank Group (2019*a*), ‘Current health expenditure per capita, PPP (current international $)’.
URL: *https://data.worldbank.org/indicator/SH.XPD.CHEX.PP.CD*

World Bank Group (2019*b*), ‘GDP per capita, PPP (constant 2011 international $)’.
URL: *https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.KD*

World Bank Group (2019*c*), ‘Literacy rate, adult total (% of people ages 15 and above)’.
URL: *https://data.worldbank.org/indicator/SE.ADT.LITR.ZS*

World Bank Group (2019*d*), ‘School enrollment, primary (% gross)’.
URL: *https://data.worldbank.org/indicator/SE.PRM.ENRR*

World Bank Group (2019*e*), ‘School enrollment, secondary (% gross)’.
URL: *https://data.worldbank.org/indicator/SE.SEC.ENRR*

World Bank Group (2019*f*), ‘School enrollment, tertiary (% gross)’.
URL: *https://data.worldbank.org/indicator/SE.TER.ENRR*

World Bank Group (2019*g*), ‘Unemployment, total (% of total labor force) (national estimate)’.
URL: *https://data.worldbank.org/indicator/SL.UEM.TOTL.NE.ZS*

World Bank Group (2019*h*), ‘World Development Indicators’.
URL: *http://datatopics.worldbank.org/world-development-indicators/*