AIX-MARSEILLE UNIVERSITÉ
FACULTÉ D'ECONOMIE ET DE GESTION

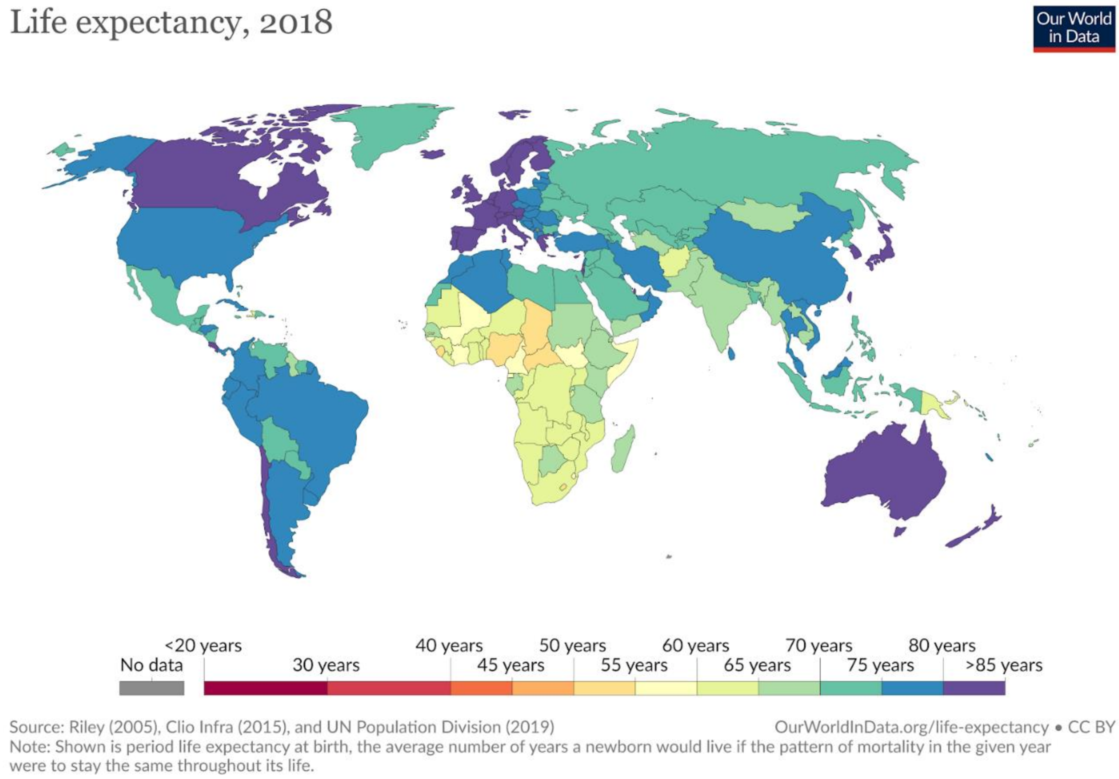
THE DETERMINANTS OF LIFE EXPECTANCY IN THE OECD COUNTRIES

PIERRE DUMONTEL - VICTOR PION - ENZO DOILLON

In this study, we're focusing on what makes the life expectancy different between countries and on which factors we should work on to improve its duration. We're using a fixed effects regression on panel data of the OECD countries between 2000 and 2018. We mainly find a positive effect of health expenditures and unemployment on the length of life expectancy but the latter is strongly influenced by the short scope of our data and the impact of the 2008 crisis. The robustness of our study is impacted by the lack of data and the presence of omitted variables.

MASTER 1 - PRINTEMPS 2021

INTRODUCTION



As we can see on this map, in 2018, the countries where life expectancy was the longest (80 years or more) were Western Europe, Canada, Australia and New Zealand. By life expectancy, we mean the average number of years an individual should live in a given country. According to the data, we observed a constant increase of life expectancy in every OECD country since the end of World War II. This constant rise can be explained by several factors like the progress in medicine, technology, increase in global wealth, or even the changes in our lifestyles. But one can think about a lot more factors which can play a role in the duration of life expectancy and explain why all countries don't have the same like we can see on the map. The aim of this study is to determine these main factors of life expectancy.

This notion is quite interesting, because determining what makes the value of life expectancy allows us to understand the decisions of households such as consumption, retirement, investments and many more. It's also interesting in order to understand the behavior of individuals with harmful habits such as smoking, consuming unhealthy food or drugs. But most importantly, by knowing which are the most impactful factors in the computation of life expectancy, decision-makers can improve the aggregate health of the population which benefits all.

In order to follow a coherent thinking path, we've consulted the OECD's "Health Glance" published in 2017. This paper shows that the main factors of life expectancy are health expenditures, education, gross domestic product and other variables. We'll use similar variables in order to find strong results. The paper concludes that a coordinated action of decision-makers on health, education, environment, income and social protection would offer great possibilities for improving life expectancy.

DATA

For this study, we're using panel data so we have at our disposal several observations over time on multiple countries. We've selected the OECD countries over the period 2000-2018, found on the World Databank website. The choice of the period was constrained by the lack of data for our variables (and their substitutes) before the year 2000.

Our explained variable is the "**life expectancy at birth**", which indicates the number of years a newborn infant would live if prevailing patterns of mortality at the time of its birth were to stay the same throughout its life.

We use five explanatory variables to explain the life expectancy at birth:

Variable	Variable name
Air pollution by fine particles (PM2.5), expressed in % of exposed population	**pollution**
Current health expenditure per capita, expressed in PPP current international $	**dep_sante**
Average alcohol consumption per capita, expressed in liters of pure alcohol	**alcohol**
School enrollment in primary stage, expressed in net %	**school**
International Labor Organization (ILO) estimation of unemployment, expressed in % of the labor force	**ChomOIT**

Table 1: Presentation of variables

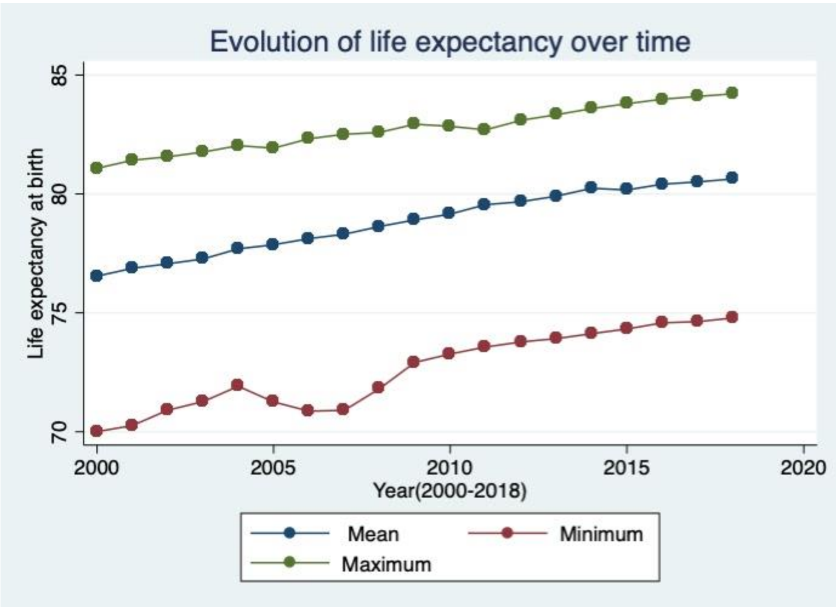
These variables have been selected in agreement with the economic literature, mainly with the paper "*Estimating expected life-years and risk factor associations with mortality in Finland cohort study*" from Tommi Härkänen (2017) and the "Health Glance" made by the OECD in 2017 which answers the question "What are the factors behind the gains in life expectancy over the recent decades?".

The database contains 703 observations on 37 OECD countries over 19 years. Here is some descriptive statistics on the database.

For this table, we've computed the mean, the standard deviation, the minimum and the maximum of our variables in the year 2000 and 2015.

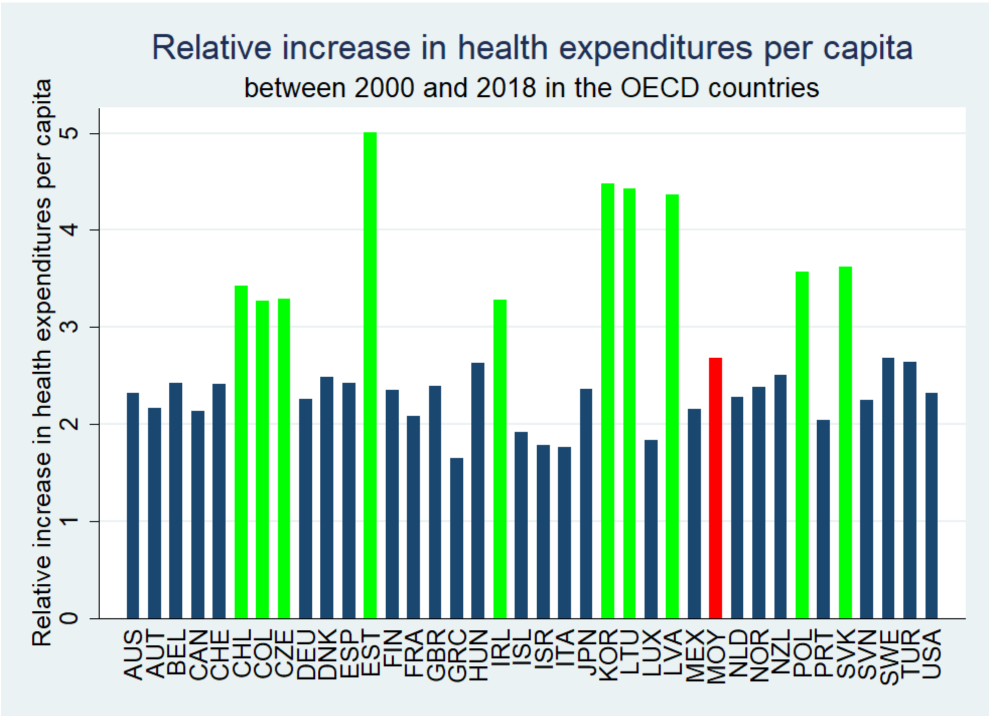
	Average		Standard deviation		Min		Max	
Var	**2000**	**2015**	**2000**	**2015**	**2000**	**2015**	**2000**	**2015**
esp_ vie	76.5	80.2	3.00	2.67	70	74.3	81.1	83.8
dep_ sante	1738	3671.7	1004.8	1898.9	353	1053.7	4564.6	9491.1
ChomOIT	8.05	7.98	4.88	4.47	1.94	3.4	20.52	24.9
alcohol	10.8	10.3	3.2	2.8	2.4	2.1	15.1	15.4
Pollution	75.4	67.1	36.3	40.8	0	0	100	100
school	95.91	95.67	3.57	3.84	86.47	82.57	99.95	99.93

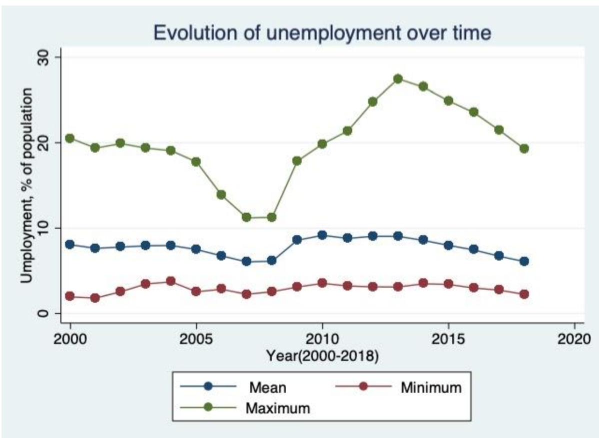
Table 2: descriptive statitics



In 2000, the average life expectancy at birth in the OECD countries was 76.54 years. The minimum is 70.01 years in Turkey and the maximum is 81.1 years in Japan. In 2018, the average life expectancy is 80.63 years, meaning an increase of 4.09 years over the period. The minimum is 74.78 years for Latvia and the maximum is 84.21 years for Japan, meaning an increase of 3.11 years over the period. On this graph, the line of minimum life expectancy shows a decrease of two years over the period 2004-2006. This decrease is attributed to Latvia, because of the dynamics of cardiovascular diseases at the time in the country.

In average, the current health expenditures per capita have doubled between 2000 and 2018 in the OECD countries. In 2000, countries spent 1737,97$ per capita on health. In 2018, these expenditures cost 4171,91$ per capita. In Chile, Colombia, Ireland, Czech Republic, Slovakia and Poland, health expenditures have tripled over the period. For example, in Colombia they went from 352,99$ to 1155,41$ per capita. In South Korea, Latvia and Lithuania health expenditures have quadrupled between 2000 and 2018. For example, in South Korea, they went from 717,39$ per capita in 2000 to 3213,67$ in 2018. Finally, in Estonia, health expenditures have quintupled over the period. They went from 485,01$ to 2427,63$ per capita.

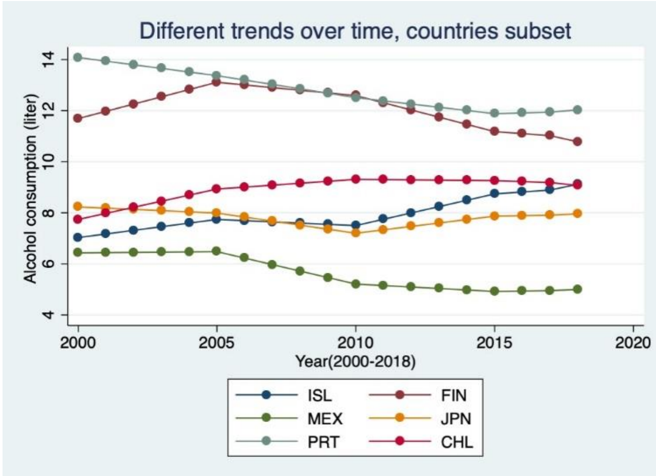


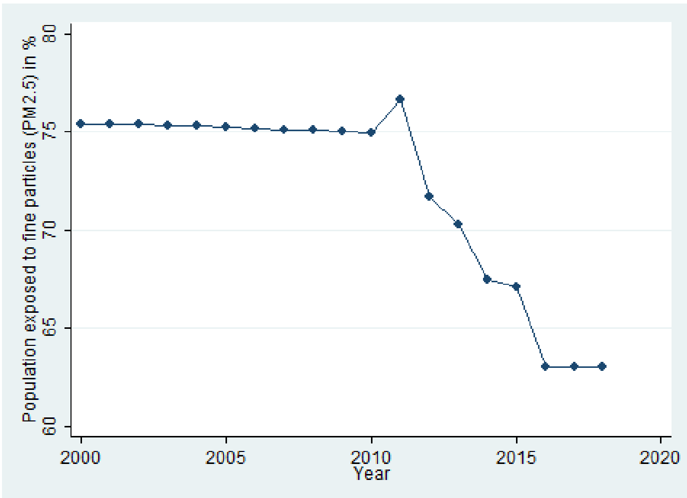


According to the ILO estimation, the average unemployment rate in OECD countries was 8.05% in 2000 and fell to 7.98% in 2015, meaning a decrease of 0.07 percentage points over the period, so there is a relative stagnation of the level of unemployment in our sample. This result is also shown in the corresponding graph, where average and minimum unemployment are stable over time. About the maximum, we first see a decrease of the highest level of unemployment followed by an increase due to the 2008 financial crisis, with the higher level in 2013 (Greece) and a new slowly decrease after 2013, in 2018, the maximum level of unemployment returned to the level of 2000. The USA, Japan, Norway, Korea and Switzerland have a low stable level of unemployment. Greece, Spain and Italy have a high level of unemployment. Portugal, Slovenia, Lithuania and Ireland are among the highest levels around 2013, with an increase since 2008 and a decrease after 2014.

Concerning education, the percentage of people which attended primary school varies between 83% and 99% on average between 2000 and 2018 in the OECD countries. In 2000, the rate was 95.91% across countries. The country with the lowest rate was Deutschland with 86.46% when the country with the highest rate was Great Britain with 99.94%. In 2018, the rate of attendance to primary school was 95.67% on average in our sample, meaning a decrease of 0.24 percentage points. The country with the minimum rate is Slovakia with a rate of 83.09% while the country with the maximum rate is Norway with 99.88% of attendance to primary school among childrens.

All periods combined, the average alcohol consumption per capita varies between 2 and 16 liters of pure alcohol, for an average consumption of 10.5 liters. Israel and Turkey are the countries with the lowest rates, these two countries being neither culturally nor religiously alcohol-consuming countries. There is a downward trend in consumption by 1 to 3 percentage points for the countries of Western Europe, the British Isles, Canada, Australia and Finland. A trend of stagnation for the United States, Colombia, Japan and Spain (after having experienced a decrease of 3 points in 2010 compared to 2005, consumption returned to its 2005 level in 2015). And finally an increase of 1 to 2 points for Chile, Iceland or Lithuania.





Percentage of average population exposed to fine particles (PM2.5) in the OECD countries

The percentage of population exposed to air pollution by fine particles (PM2.5) is widely spread across countries. In 2000, on average across the OECD countries, 75.41% of the population was exposed to these fine particles. The country with the lowest rate was Finland with 0% of exposition of fine particles. Countries with cold weather like Sweden, Canada, Estonia and Island seem to be less exposed than countries with relatively warmer weather. The countries with the highest rate are Greece, Hungary, Israël, South Korea, Nederland, Poland, Slovakia, Slovenia and Turkey with 100% air pollution by fine particles. In 2018, the rate fell to 63.03% on average in our sample, meaning a decrease of 12.38 percentage points over the period. The countries with the minimum rate are Canada, Estonia, Finland and New Zealand with 0% exposure to fine particles. The countries with the maximum rates are still Greece, Hungary, Israël, South Korea, Slovakia and Turkey with a 100% exposition rate.

EMPIRICAL ANALYSIS

To fit our model, we will use a linear regression with the life expectancy at birth as our explained variable. Among explanatory variables, health expenditures per capita is in logarithmic form, all other variables (**school**, **alcohol**, **pollution**, **ChomOIT**) are level variables. Because of missing values for **alcohol** and **pollution** (only one observation every 5 years), we are using a linear interpolation to complete the dataset. We face another missing value issue with the variable **school**: for some observations, we used a linear interpolation when it could be done, otherwise we duplicate available values on missing values. Assuming that variables have a delayed effect on the explained variable (for instance a variation of health expenditure doesn't have an instantaneous effect on life expectancy but has an effect on the following years), we are using 5 years delayed explanatory variables.

We can expect positive and negative causality with life expectancy depending on the variable. If a country invests more in health, we should expect a positive effect on life expectancy. The intuition is the same for the rate of school enrollment of children at primary school. Alcohol consumption and exposure to fine particles are harmful composant for aggregate health and then we expect a negative sign on life expectancy. Unemployment might have an ambiguous effect. On one hand, it could be a source of stress issues, burnouts or decrease in life quality due to the loss of financial wealth. On another hand, unemployment has a very high average level of protection in OECD countries, so the loss of financial wealth we discussed before might be compensated by unemployment insurance, and we can also assume that a non-working life could be less stressful and hardful.

To have an idea of the positive or negative effect that the explanatory variables will have on the explained variables, we look at the correlation table:

	esp_vie	**school**	**alcool**	**pollution**	**dep_sante**	**chomOIT**
esp_vie	1					
school	0.32	1				
alcool	-0.02	-0.21	1			
pollution	-0.31	-0.33	-0.02	1		
dep_sante	0.68	0.11	0.13	-0.37	1	
ChomOIT	-0.13	-0.01	-0.04	0.1	-0.24	1

Table 3: Correlation table

We assume that our model is the fixed effect one, due to the heterogeneity of behaviors that exist in the individual (country) dimension. The equation of our model is:

$$\text{Esp\_vie}_{i,t} = \beta_0 + a_i + \beta_1 \log(\text{dep\_sante})_{i,t-5} + \beta_2 \text{school}_{i,t-5} + \beta_3 \text{alcohol}_{i,t-5} + \beta_4 \text{pollution}_{i,t-5} + \beta_5 \text{chomoit}_{i,t-5} + \varepsilon_{i,t}$$

Where i represents country and t time.

We assume basic econometric hypotheses like the ones on disturbances: independently and identically distributed, with a zero expectation and a unique variance. We assumed in a first stage a strict exogeneity of regressors, homoscedasticity and no serial correlation. Notice that GDP per capita used to be one of our explanatory variables but we had to forget about it due to multicollinearity issues with health expenditures per capita (VIF higher than 10).

About the potential endogeneity issues and explanatory variables interactions: a potential endogenous variable is the unemployment rate because of its ambiguous effect, such a variable has to be linked with the unemployment insurance national system for instance. We are surprised about the correlation between some variables. Health expenditures and air pollution are negatively correlated, that doesn't make sense in a way if we assume that healthcare could be increased by air pollution and by the proportion of population that are exposed to it. Unemployment rate and primary school participation have a close to zero correlation, so the school enrollment of a large part of the national population doesn't allow to decrease the unemployment rate. Also, the 2008 crisis and the following years have influenced the unemployment rate more than the school enrollment. Another potential endogenous variable is health expenditures per capita, basically we suppose that the older a population is, the more it will need health care. We find also that we have the presence of heterogeneity and this can be a source of potential endogeneity.

One could think about several instruments to solve the endogeneity issue. Since we work with panel data, lagged values can be often used as instruments but we could also add other variables to the model. For example, to fix the potential endogeneity of unemployment, we could use the expenditures in unemployment benefits. For health expenditures, one could use the average social security cost per capita as an instrument.

Additional tests we should deal with are the ones on heteroskedasticity and serial correlation. The methodology of the 2017 OECD report uses corrections for these 2 issues. Such corrections might be added to our model.

RESULTS

The principal limit of our work is the size of the dataset, we work with a 19 years panel and 5 years delay on the explanatory variables, so only 14 years of observations. Because of using delayed variables, we have 493 observations and 36 groups (one of the 37 countries is omitted to avoid collinearity issue). The within R-squared is 0.87, so 87% of the variance of life expectancy at birth within a given country is explained by the explanatory variables. The between R-squared is 0.59, so 59% of the variance of life expectancy at birth between countries is explained by the explanatory variables. The overall R-squared is 0.61, so 61% of the variance of life expectancy at birth is overall explained by the explanatory variables. We obtain in the next table the results of our fixed effects regression.

Variable	Coefficients	Standard error	Statistic t
Intercept	60.48	2.19	27.62
logdep_sante	2.269	0.2	11.41
ChomOIT	0.016	0.008	1.97
Alcohol	0.039	0.032	1.25
Pollution	0.0001	0.007	0.02
School	0.002	0.017	0.13

Table 4: Results of our regression

The results of our regression are not satisfactory. Three explanatory variables out of five are not significant, even at the 10% level: **alcohol**, **pollution** and **school**. How can we explain such results? The variable **school** indicates the percentage of the population enrolled in primary school. But for the period 2000-2018 this percentage is very high for all countries, because OECD countries are developed countries. So we can suppose that such a variable does not have a significant impact on life expectancy for OECD countries nowadays, and that is why the variable is not significant. About **pollution**, we have seen above that there exists heterogeneity between countries, depending on their localization in different parts of the world. A potential solution to deal with this issue would be to create control variables like continent/region dummies. Then, we can't trust the coefficients associated with these variables, we'll only analyze the coefficients of **logdep_sante** and **ChomOIT**. Concerning the signs, all coefficients have a positive impact on our explained variable. This is surprising for **ChomOIT**, **alcohol** and **pollution** since they can be seen as harmful variables for life expectancy. But since **alcohol** and **pollution** are not significant, we'll only discuss the sign of **ChomOIT**.

These results should be interpreted carefully: since our explained variable is a level one, a logarithmic variable and a level variable will not be interpreted in the same way.

Then, our regression teaches us the impact of our variables on the duration of life expectancy. When health expenditures increase by one percentage point, life expectancy increases by 0.02 years (7.3 days). This is in line with the results of the OECD "Health Glance" since they found an impact of 0.039 years (14.24 days). If unemployment increases by one percentage point, life expectancy will increase by 0.02 years (7.3 days). The sign of the coefficient can be surprising because one could think that unemployment is not desirable for a country or an individual. But with generous unemployment benefits such as those in Northern Europe, an individual can live off these benefits without working, contributing to the increase of its life expectancy. Most importantly, the sign of **ChomOIT** can be attributed to the relative short period of our panel data, combined with the delays on the variables (14 years of data). If we take into account the 2008 crisis and the years after, where unemployment has increased for most of the OECD countries, combined with the continuous increase of life expectancy, the regression will obviously return a positive coefficient.

Years	2006	2007	2008	2009	2010	2011	2012
Coefficient	0.08 (NS)	0 .08 (NS)	0 .27	0.38	0.52	0.76	0.74
Years	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	.
Coefficient	0.80	1.01	0.87	1.04	0.87	1.09	.

Table 5: Years fixed effect coefficients of regression

The 2006 and 2007 coefficients are not significant, and we see between 2008 and 2018 a slight increase of the year fixed effect, that is in line with the descriptive statistics we have presented above.

Now, we'll proceed with some tests. We start with the Fisher test and this test determines the simultaneous nullity of all coefficients without the intercept. We assumed these hypotheses:

H0: All parameters are zero.
H1: It exists at least one parameter different from zero.

The decision rule is done according to the statistic value computed and the one read in the Fisher's table. We can also trust the p-value. For our model, we find that p-value is equal to 0 and it is inferior to 0.05, so we reject H0 and conclude that there is no joint nullity of parameters.

Moreover, we have assumed the existence of a heterogeneity of behaviors across individuals. We can test the possible absence of this heterogeneity. We assumed these hypotheses:

H0: Absence of heterogeneity.
H1: Presence of heterogeneity.

The decision rule is done according to the precedent method regarding if the p-value is inferior or superior to 0.05. For our model, we find that p-value is equal to 0 and it inferior to 0.05, we reject H0 and we can conclude that there is heterogeneity in our model.

CONCLUSION

The aim of this study was to understand the main determinants of life expectancy at birth in OECD countries from 2000 to 2018. We first looked at the descriptive statistics and trends of all variables over the period and the countries. Then we developed our empirical methodology and all the issues we faced for the construction of the main regression, especially interaction between explanatory variables, between explanatory and explained variables and how to take into account potential ambiguous effects like for unemployment. We finally presented and discussed our main results.

To conclude, the regression of life expectancy is quite difficult to interpret in our case, we are not working with enough years of observations to have significant coefficients for all variables. We need a deeper analysis to interpret the role of unemployment, especially because of the significant role of the 2008 crisis and its following years. However, our regression results are quite close to those of the OECD report "Health Glance" from 2017.

The model could be improved by including more years in the time dimension of our panel data. Previous works focused only on OECD countries, determinants of life expectancy could also be studied for all countries over the world. Such a work could be more specific by looking country by country (or group of countries) and look at the differences. Life expectancy could be analyzed as an inequality debate, between men and women or between (CSP) for instance. We have constrained the model to 5 explanatory variables, but other ones could explain life expectancy, like infant mortality, war, smoking, malnutrition. In the coming years, impacts of the international pandemic crisis will have to be analyzed in the life expectancy for all impacted years.

WEBOGRAPHIE

OECD's "Health Glance" 2017

Tommi Härkänen (2020) "*Estimating expected life-years and risk factor associations with mortality in Finland: cohort study*"

Introduction's map

Life expectancy in the Baltics

Source des données