Assignment 3

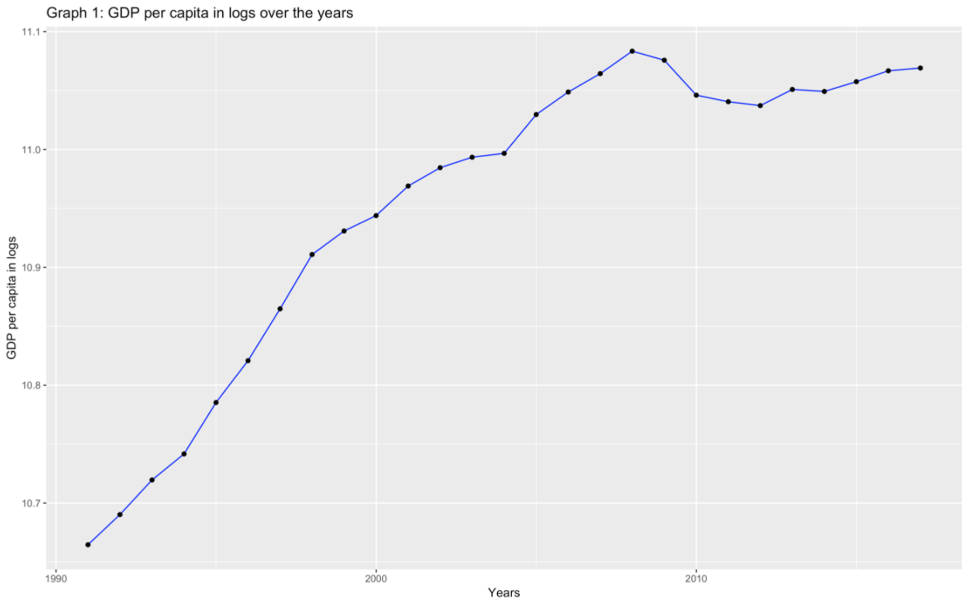
Date: 2018-12-17 Anna Grigoryan

Data Preparation

- Time series data on ***GDP per capita, PPP (constant)*** and ***life expectancy at birth, total*** has been downloaded for Norway from the World Bank cross-country datasets.
- Data was cleaned, years saved in date format and datasets ***lifeexp*** and ***gdppc*** merged into ***data*** using the years.
- ***ln_gdp*** and first difference columns ***Dln_gdp*** and ***Dlife_exp*** were created.
- Life expectancy columns, i. e. ***life_exp***, ***Dlife_exp*** and GDP per capita columns, i.e. ***ln_gdp*** , ***Dlngdp*** were transformed into ***ts*** (time-series) format starting from year 1960.

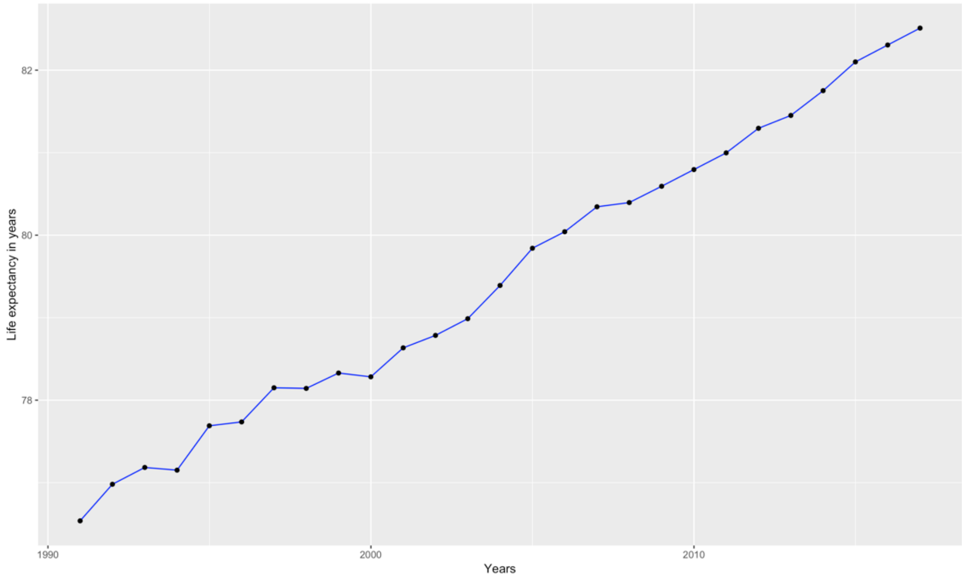
Data Exploration

- Estimating a LOESS regression of ***ln_gdp*** (GDP per capita in logs)



- Estimating a LOESS regression of ***lifeexp*** (life expectancy)

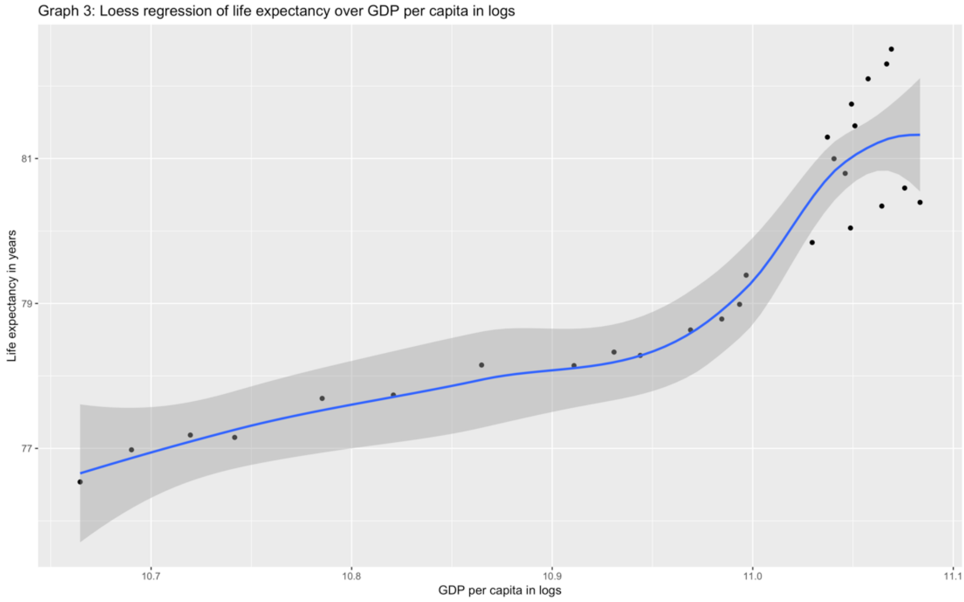
Graph 2: Life expectancy over the years



The data covers 57 years starting with January 1 1960 and ending with December 31 2016. This is time series data at yearly frequency. The data has gaps as there are no observations for GDP at start. We drop those gaps and simply start with January 1 1990. We start with comparing yearly life expectations and GDP per capita in logs in Norway across the time. The yearly data shows a positive trend in both. As both of the two variables have positive trends, a regression of one on the other will result in a positive slope coefficient whether the two variables are related or not. Indeed, the loess of yearly life expectancy to the yearly GDP per capita in logs has a positive trend. This is simply because in later time periods both life expectancy and GDP per capita tend to have higher values.

- Estimating a LOESS regression of ***lifeexp*** over ***ln_gdp*** (life expectancy)

Graph 3: Loess regression of life expectancy over GDP per capita in logs



Data Analysis

	Dependent variable:			
	life_exp	Diff_life_exp		
	OLS	*OLS*		*coefficient test*
	(1)	(2)	(3)	(4)
ln_gdp	12.000			
	(1.020)***			
Diff_ln_gdp		0.727	0.727	0.727
		(1.870)	(1.870)	(0.665)
Constant	-52.000	0.218	0.218	0.218
	(11.100)***	(0.025)***	(0.025)***	(0.018)***
Observations	27	26	26	
R^2	0.781	0.007	0.007	
Note:		*p<0.1; **p<0.05; ***p<0.01		
		Robust standard errors in parentheses		

Figure 1: Levels and first difference, no lag

Levels, no lag , robust SE

The average $\boldsymbol{lifeexp}$ is -52 when $\boldsymbol{ln\_gdp}$ is zero. However, this doesn't make sense as $\boldsymbol{ln\_gdp}$ can't be zero. Comparing time periods with different $\boldsymbol{ln\_gdp}$, $\boldsymbol{lifeexp}$ is expected to be higher by 0.12 years at times when $\boldsymbol{ln\_gdp}$ is higher by one unit.

First difference, no lag , robust SE and Newey West SE

Variables in time series regressions may be expressed in differences instead of levels. The interpretation of the linear regression in differences, with one explanatory variables on time series data is the following:

$\boldsymbol{lifeexp}$ is expected to change by 0.218 years when $\boldsymbol{ln\_gdp}$ doesn't change. Comparing time periods with different change in $\boldsymbol{ln\_gdp}$, y is expected to change by 0.00727 years more when $\boldsymbol{ln\_gdp}$ changes by one unit. However, the coefficient is not significant at 95% CI.

Serial correlation of variables in time series data does not lead to biased regression coefficients. But it makes the usual standard error estimates wrong. As we can see the regression coefficients both with robust SE and Newey West SE are the same, but the standard errors from Newey West are much lower ($SE_{NW} = 0.665$ $SE_{robust} = 1.870$). The Newey-West procedure incorporates the structure of serial correlation, while heteroskedasticity-robust SE assumes zero serial correlation.

First difference, one lag , Newey West SE

Another solution is to model serial correlation within the regression and include lagged life expectancy on the right-hand-side.

	Dependent variable:	
	Diff_life_exp ~ Diff_ln_gdp + lag(Diff_life_exp, -1)	
	OLS	*coefficient test*
	(1)	(2)
Diff_ln_gdp	0.967	0.967
	(1.640)	(0.570)
lag(Diff_life_exp, -1)	-0.392	-0.392
	$(0.189)^{**}$	$(0.126)^{***}$
Constant	0.297	0.297
	$(0.055)^{***}$	$(0.017)^{***}$
Observations	25	
R^2	0.166	
Note:		$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$ Robust standard errors in parentheses

Figure 2: First difference, one lag

The results of one lagged regression with first difference variables is the following: the contemporaneous effect, 0.967 shows how $\boldsymbol{lifeexp}$ is expected to change within the same time period. Once-lagged effect: shows that $\boldsymbol{lifeexp}$ is expected to change by -0.00392 the next time period. (Viewed from the next time period's $\boldsymbol{Diff\_life\_exp}$, this time period's change in $\boldsymbol{ln\_gdp}$ is $\boldsymbol{lag(Diff\_life\_exp, -1)}$ a lagged change.) Meaning, that after a change in $\boldsymbol{ln\_gdp}$ by one unit, we expect $\boldsymbol{life\_exp}$ to change by 0.967 and then by -0.00392. Once-lagged effect is significant, while contemporaneous effect is not at 95% CI. We use Newey West SE as they are robust for serial correlation.

First difference, two lags, Newey West SE

	Dependent variable:	
	Diff_life_exp ~ Diff_ln_gdp + lag(Diff_life_exp, -2)	
	OLS	*coefficient test*
	(1)	(2)
Diff_ln_gdp	0.336	0.336
	(1.780)	(0.718)
lag(Diff_life_exp, -2)	0.249	0.249
	(0.202)	$(0.098)^{**}$
Constant	0.159	0.159
	$(0.061)^{**}$	$(0.035)^{***}$
Observations	24	
R^2	0.071	
Note:		$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$ Robust standard errors in parentheses

Figure 3: First difference, two lags.

The results of twice lagged regression with first difference variables is the following: the contemporaneous effect, 0.336 shows how $lifeexp$ is expected to change within the same time period. Twice-lagged effect: shows that $lifeexp$ is expected to change by 0.00249 years two time periods after. (Meaning, that after a change in ln_gdp by one unit, we expect $life\_exp$ to change by 0.336 and two years after by 0.00249. Twice-lagged effect is significant, when Newey West SE is used, while contemporaneous effect is not at 95% CI.

First difference, two lags with cumulative lag in model , Newey West SE

	Dependent variable:	
	Diff_life_exp ~ lag(Diff_ln_gdp, -2) + diff(lag(Diff_ln_gdp, -1)) + diff(Diff_ln_gdp) + lag(Diff_life_exp, -1)	
	OLS	*coefficient test*
	(1)	(2)
lag(Diff_ln_gdp, -2)	-1.470	-1.470
	(1.760)	(0.850)
diff(lag(Diff_ln_gdp, -1))	1.130	1.130
	(2.290)	(0.797)
diff(Diff_ln_gdp)	4.390	4.390
	(2.020)**	(1.580)**
lag(Diff_life_exp, -1)	-0.526	-0.526
	(0.198)**	(0.130)***
Constant	0.368	0.368
	(0.061)***	(0.035)***
Observations	24	
R^2	0.400	
Note:		*p<0.1; **p<0.05; ***p<0.01 Robust standard errors in parentheses

Figure 4: First difference, two lags, cummulative effect.

-1.470 in this regression is the cumulative change, (the long-run effect), i.e. if GDP increases with 1% more, $life\_exp$ is expected to change by -1.470 less. However, it is not significant at 95% CI with neither of standard errors. The other two right-hand-side variables are quite strange: they are differences of variables that are already in differences. Their coefficients are hard to interpret.

Conclusion

Although some of the results were not significant, they provided understanding of the overall patterns. Just a one-time change of GDP per-capita won't increase the life expectancy significantly. In order to find a relationship between GDP per capita and life expectancy, some extra steps must be taken. For example, the R-squared value is not very strong. To strengthen the regression as a whole, more explanatory variables can be added, such as expenditure on healthcare.