Notebook

April 16, 2021

Question 1.a. Based on the results for the two OLS regressions, what is the sign of the correlation between `dkr` and `lnetincome`? Alternatively, is there not enough information to determine the sign of the correlation?

Based on the p-value of 0.706, we're not confident on the sign of `dkr` and its relationship with `returns` and with `lnetincome` because it is not statistically significant. With the 95% confidence interval of the `dkr` variable, it is between -0.267 and 0.392, which indicates that it is plausible that `dkr` could be negative or positive. For `lnetincome`, we are very confident that the value has a positive correlation with the dependent variable `returns` given that the 95 percent confidence interval is between 2.950 and 6.152.

Question 1.b. Interpret the coefficient on `lnetincome` in Regression 2.

The coefficient on lnetincome indicates that the variable lnetincome is positively correlated with our dependent variable returns. More specifically, the regression estimates that each increase of 1 unit in the independent variable lnetincome will result in a 4.55% increase in the dependent variable return.

Question 1.c. Suppose that you use Regression 3 to examine whether EMH holds. What are the null and alternative hypotheses?

H_o: $\beta_1 = \beta_2 = \beta_3 = 0$
H_a: β_1 or β_2 or $\beta_3 \neq 0$

Question 1.d. Carry out the test in part (c) at the 5% level. Do you reject or fail to reject the null hypothesis?

Based on the F-statistic of 11.01 and the probability of F-statistic equal to 0, we would reject the null hypothesis when testing at the 5% level, meaning that EMH does not hold.

Question 1.e. Interpret the result you obtained in part (d), in light of your task of examining the validity of EMH.

Because we rejected the null hypothesis, it would mean that at least one of the regressors does not equal to 0, which supports the hypothesis that EMH does not hold.

Question 1.f. Provide (at least) two reasons why there might be imperfect multicollinearity present in Regression 3.

Imperfect multicollinearity implies that some of the independent variables is highly correlated with each other. Between `lsalary` and `lnetincome`, if a firm's net income is high, we would expect salaries to be also high. We can also expect that if a company's net income is high, then their `dkr` (debt-to-capital ratio) should be low.

Question 1.g. Which of the following statements is true based on a comparison of Regression 2 and Regression 3? - (i) `dkr` and `lnetincome` are highly-correlated. - (ii) `dkr` and `lsalary` are highly-correlated. - (iii) `lnetincome` and `lsalary` are highly-correlated. - (iv) All of the above. - (v) None of the above.

Question 1.h. The sample of 142 stocks only include companies that were traded on the NYSE as of the end of 2013. A company that went out of business, for instance, before the end of that year could not enter the sample. How would this sampling affect the estimated coefficient relative to the population regression?

It would make our estimated coefficients higher than the population regression. This is because omitting companies that run out of business or are underperforming, the regression would only include top-performing companies, which would make returns higher than the true population value.

Question 2.a. Regress `lfare` on `dist`, `passen` and `concen`, with robust standard errors. Make sure the cell below (and all regression questions in this assignment) shows your regression results like you've done in previous assignments, otherwise we cannot give credit. This assignment will be a little less guided. Make sure do use different variable names for each separate coding part to avoid unexpected errors from reusing variables. Refer to previous assignments if you need a refresher on how we performed different regressions. *Don't forget to add a constant to your regressions.*

[4]:
```python
y_2a = af['lfare']
X_2a = sm.add_constant(af[['dist', 'passen', 'concen']])
model_x = sm.OLS(y_2a, X_2a)
results_x = model_x.fit(cov_type = 'HC1')
results_x.summary()
```

[4]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:                  lfare   R-squared:                       0.370
Model:                            OLS   Adj. R-squared:                  0.368
Method:                 Least Squares   F-statistic:                     233.2
Date:                Fri, 16 Apr 2021   Prob (F-statistic):          4.63e-118
Time:                        15:27:32   Log-Likelihood:                -359.76
No. Observations:                1149   AIC:                             727.5
Df Residuals:                    1145   BIC:                             747.7
Df Model:                           3
Covariance Type:                  HC1
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          4.6560      0.051     91.435      0.000       4.556       4.756
dist           0.4272      0.018     23.392      0.000       0.391       0.463
passen        -0.0581      0.014     -4.049      0.000      -0.086      -0.030
concen         0.1875      0.061      3.064      0.002       0.068       0.307
==============================================================================
```

```
Omnibus:                       67.392   Durbin-Watson:                   1.367
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               28.039
Skew:                           0.131   Prob(JB):                     8.16e-07
Kurtosis:                       2.281   Cond. No.                         13.8
==============================================================================

Warnings:
[1] Standard Errors are heteroscedasticity robust (HC1)
"""
```

Question 2.b. What is the interpretation of the coefficient on `passen`?

Based on our regression results, the `passen` variable appears to have inverse relationship with the logarithm of the average fare variable. In other words, an increase in one additional `passen` variable (which represents one thousand passengers) would decrease the log of average fares by about 116 dollars (2000 dollars times 0.0581).

Question 2.c. Based on your OLSEs, and assuming the OLS assumptions hold, what is the partial effect of the market share of the largest carrier on air fares? Is your answer consistent with the hypothesis that firms use their market power to charge higher prices?

The partial effect of the market share of the largest carrier on air fares is 0.1875 times 2000, which is equal to 375. We arrived this calculation by finding the largest carrier who has a `concen` value of 1.0, which means that this firm has 100% of the market share for the given route. The larger the share of the market that a firm captures, we expect a higher air fare for that given airline. We know this because the `concen` variable coefficient is positive, indicating that it has a positive relationship with air fares. Yes, our answer is consistent with larger firms charging higher market prices on air fares.

Question 2.d. How would you test whether market power is used the same way on more popular and less popular routes? Write down the model and the hypothesis, carry out the estimation and the test.

This question is for your code, the next is for your explanation.

[5]:
```
y_2d = af['lfare']
af['concenpass'] = af['concen'] * af['passen']
X_2d = sm.add_constant(af[['dist','passen','concen','concenpass']])
model_2d = sm.OLS(y_2d,X_2d)
results_2d = model_2d.fit(cov_type = 'HC1')
results_2d.summary()
```

[5]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:                  lfare   R-squared:                       0.395
Model:                            OLS   Adj. R-squared:                  0.393
Method:                 Least Squares   F-statistic:                     197.9
Date:                Fri, 16 Apr 2021   Prob (F-statistic):          5.59e-129
```

```
Time:                        15:27:33   Log-Likelihood:                -336.12
No. Observations:                1149   AIC:                             682.2
Df Residuals:                    1144   BIC:                             707.5
Df Model:                           4
Covariance Type:                  HC1
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          4.5397      0.053     84.966      0.000       4.435       4.644
dist           0.4245      0.018     23.616      0.000       0.389       0.460
passen         0.1581      0.038      4.128      0.000       0.083       0.233
concen         0.4281      0.072      5.938      0.000       0.287       0.569
concenpass    -0.4455      0.076     -5.890      0.000      -0.594      -0.297
==============================================================================
Omnibus:                       53.030   Durbin-Watson:                   1.389
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               24.778
Skew:                           0.137   Prob(JB):                     4.16e-06
Kurtosis:                       2.335   Cond. No.                         17.9
==============================================================================

Warnings:
[1] Standard Errors are heteroscedasticity robust (HC1)
"""
```

Question 2.e. Explain.

H_o: $\beta_4 = 0$
H_a: $\beta_4 \neq 0$

Based on the `concenpass` coefficient of -0.4455 and that the p-value is significant, we conclude that market power is used different depending on the popularity of airline routes.

Question 2.f. We need to question whether the results of the regression in part (d) is revealing a causal relationship between concentration and airfares. In particular, we are concerned whether our estimation results on U.S. data are valid for other markets, such as Europe and Asia. Give one reason why the results would not be "externally valid" if applied to the airline industry in one of these other two regions.

One reason why these results cannot be 'externally valid' is because the air line markets for Asia and Europe could be structurally different due to government interventions in the market.

Question 2.g. We are also aware of several potential threats to "internal validity" of the results. For each one of the five main internal validity threats, describe one possibility that could plausibly lead to that particular threat.

Omitted Variable Bias: Consumer price index is different for each state or location, which is not recorded in this data set.
functional form misspecification: There could be a exponential relationship between air fares and marketpower.
Measurement error and errors-in-variables bias: The OLS estimate does not include possible

bias in the variable `concenpass`.
Sample selection bias: Selected only data from bigger airlines; data does not include less popular airlines.
Simultaneous causality: Higher air fares in one location could attract companies to get more market power.

Question 3.a. Create a new variable for the dataset that is the square of educational attainment (`hc3`). Then regress life expectancy (`dale`) on health expenditures (`hexp`), the educational attainment in the country (`hc3`), and its square (the variable you created). For now, select rows from 1997 and use only these rows in the regression. Use robust standard errors and *don't forget to add a constant term.* Comment on whether you think the relationship between life expectancy and education is linear or quadratic and why you came to that conclusion.

This question is for your code, the next is for your explanation.

[7]:
```
hc3_square = who['hc3']**2
who['hc3_square'] = hc3_square.tolist()
y_3a = who['dale']
X_3a = sm.add_constant(who[['hexp', 'hc3', 'hc3_square']])
model_x_dale = sm.OLS(y_3a, X_3a)
results_dale = model_x_dale.fit(cov_type = 'HC1')
results_dale.summary()
```

[7]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:                   dale   R-squared:                       0.682
Model:                            OLS   Adj. R-squared:                  0.680
Method:                 Least Squares   F-statistic:                     647.8
Date:                Fri, 16 Apr 2021   Prob (F-statistic):          6.34e-201
Time:                        15:27:33   Log-Likelihood:                -2347.2
No. Observations:                 700   AIC:                             4702.
Df Residuals:                     696   BIC:                             4721.
Df Model:                           3
Covariance Type:                  HC1
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         26.2531      1.132     23.192      0.000      24.034      28.472
hexp           0.0061      0.000     16.516      0.000       0.005       0.007
hc3            7.7583      0.406     19.119      0.000       6.963       8.554
hc3_square    -0.4292      0.032    -13.349      0.000      -0.492      -0.366
==============================================================================
Omnibus:                       77.528   Durbin-Watson:                   0.344
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              119.675
Skew:                          -0.760   Prob(JB):                     1.03e-26
Kurtosis:                       4.339   Cond. No.                     4.07e+03
```

```
==============================================================================

Warnings:
[1] Standard Errors are heteroscedasticity robust (HC1)
[2] The condition number is large, 4.07e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
"""
```

Question 3.b. Explain.

We are rather confident that life expectancy and education have a positive relationship, meaning that more education should result in a longer life. However, we conclude that the relationship is linear and not quadratic because there's a finite number of years that you can live, and likewise, there is a limit to a number of years of education. To say that 3 additonal years of education results in 3 additional year of life expectancy is reasonable (indicating a linear relationship). But to say that 3 additional years results in 9 additional years of life seems unrealistic (indicating a quadratic relationship). Both education and life expectancy do vary from person to person, but the degree to which they vary is fixed to a relatively small range of values.

Question 3.c. To the specification in part (a), add the additional control variables: `gini`, `tropics`, `popden`, `pubthe`, `gdpc`, `voice`, and `geff`. Test whether these additional regressors are jointly significant (we do the F-test for you in this part, you just have to interpret it). What effect does inclusion of these additional controls have on the coefficients of the other included regressors?

This question is for your code, the next is for your explanation.

[8]:
```python
# This is the code for your regression.
# We give you starter code for this one so that we know what the variable name␣
↪is
# for the regression results, which we use in the code cell below.
X_3c = sm.add_constant(who[['hexp', 'hc3', 'hc3_square', 'gini', 'tropics',␣
↪'popden', 'pubthe', 'gdpc', 'voice', 'geff']])
model_3b = sm.OLS(y_3a, X_3c)
results_3b = model_3b.fit(cov_type = 'HC1')
results_3b.summary()
```

[8]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:                   dale   R-squared:                       0.737
Model:                            OLS   Adj. R-squared:                  0.734
Method:                 Least Squares   F-statistic:                     365.8
Date:                Fri, 16 Apr 2021   Prob (F-statistic):          7.76e-268
Time:                        15:27:33   Log-Likelihood:                -2279.8
No. Observations:                 700   AIC:                             4582.
Df Residuals:                     689   BIC:                             4632.
Df Model:                          10
Covariance Type:                  HC1
```

```
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         41.0913      2.084     19.716      0.000      37.006      45.176
hexp          -0.0017      0.001     -2.470      0.013      -0.003      -0.000
hc3            6.7425      0.383     17.622      0.000       5.993       7.492
hc3_square    -0.3867      0.030    -12.688      0.000      -0.446      -0.327
gini         -16.3554      4.394     -3.722      0.000     -24.967      -7.744
tropics       -2.8633      0.752     -3.808      0.000      -4.337      -1.390
popden     -8.664e-05   5.41e-05     -1.602      0.109      -0.000    1.94e-05
pubthe        -0.0456      0.012     -3.716      0.000      -0.070      -0.022
gdpc           0.0004   8.05e-05      5.422      0.000       0.000       0.001
voice          0.7554      0.472      1.601      0.109      -0.169       1.680
geff           1.8746      0.561      3.343      0.001       0.775       2.974
==============================================================================
Omnibus:                       51.682   Durbin-Watson:                   0.381
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               76.710
Skew:                          -0.562   Prob(JB):                     2.20e-17
Kurtosis:                       4.170   Cond. No.                     1.65e+05
==============================================================================

Warnings:
[1] Standard Errors are heteroscedasticity robust (HC1)
[2] The condition number is large, 1.65e+05. This might indicate that there are
strong multicollinearity or other numerical problems.
"""
```

Question 3.d. Explain.

By including the additional control variables, the variabes `hexp`, `hc3`, and `hc3_square` became less impactful on life expectency. Also, the constant increased from 26.25 to 41.09, which makes the additional control variables more significant into the OLS regression.

Question 3.e. Return to the simpler regression specification in part (a). We want see if the determinants of life expectancy are different for rich and poor countries. Use membership in the "Organization of Economic Cooperation & Development" (`oecd`) as the indicator of a rich country. The OECD had 30 member countries during this time period. Perform a test of the hypothesis that all three of the coefficients in the population regression are equal for OECD and non-OECD countries.

Hint: You will need to create three new variables.

This question is for your code, the next is for your explanation.

[10]:
```python
oecd = who[who['oecd'] == 1]
y_3e = oecd['dale']
X_3e = sm.add_constant(oecd[['hexp', 'hc3', 'hc3_square']])
model_x_dale = sm.OLS(y_3e, X_3e)
results_dale = model_x_dale.fit(cov_type = 'HC1')
```

```
results_dale.summary()
```

[10]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:                   dale   R-squared:                       0.650
Model:                            OLS   Adj. R-squared:                  0.643
Method:                 Least Squares   F-statistic:                     225.7
Date:                Fri, 16 Apr 2021   Prob (F-statistic):           1.29e-54
Time:                        15:27:33   Log-Likelihood:                -302.30
No. Observations:                 150   AIC:                             612.6
Df Residuals:                     146   BIC:                             624.7
Df Model:                           3
Covariance Type:                  HC1
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         43.2708      1.963     22.048      0.000      39.424      47.117
hexp           0.0028      0.000     11.143      0.000       0.002       0.003
hc3            5.9866      0.580     10.319      0.000       4.849       7.124
hc3_square    -0.3760      0.037    -10.207      0.000      -0.448      -0.304
==============================================================================
Omnibus:                        2.023   Durbin-Watson:                   0.495
Prob(Omnibus):                  0.364   Jarque-Bera (JB):                2.002
Skew:                          -0.218   Prob(JB):                        0.368
Kurtosis:                       2.639   Cond. No.                     3.13e+04
==============================================================================

Warnings:
[1] Standard Errors are heteroscedasticity robust (HC1)
[2] The condition number is large, 3.13e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
"""
```

Question 3.f. Explain.

H_o: $\beta^1_{oecd} = \beta^1_{nonoecd}$, $\beta^2_{oecd} = \beta^2_{nonoecd}$, $\beta^3_{oecd} = \beta^3_{nonoecd}$
H_a: $\beta^1_{oecd} \neq \beta^1_{nonoecd}$ or $\beta^2_{oecd} \neq \beta^2_{nonoecd}$ or $\beta^3_{oecd} \neq \beta^3_{nonoecd}$

Due to the significance of the p-value for each coefficient, we see that the coefficient of the regressors are not the same between oecd and nonoecd countries.

Question 3.g. Give an example of a time-invariant variable that would result in different life expectancy across countries.

One example of a time-invariant variable could be geography, because depending on where people live could affect their life expectancy across countries.

Question 3.h. Estimate the regression having a fixed effect for each country in the sample. We have defined the endogenous and exogenous variables for you, you just have to fill in the rest. Notice how we converted the `country` variable to a set of dummy variables for each country. You can ignore the coefficients for every country variable. What change took place in the coefficients on the education variables? Explain why you think there was a change in these coefficients.

This question is for your code, the next is for your explanation.

[12]:
```
# .get_dummies transforms a categorical variable into a dataframe of dummy␣
↪variables,
# one for each category. The prefix and prefix_sep part just makes sure the␣
↪variable
# names are strings and not integers.
countries = pd.get_dummies(who['country'], prefix='', prefix_sep='')
# This just joins the dummy dataframe with the original
who_country = who[['dale', 'hexp', 'hc3', 'hc3_square']].join(countries)
y_3h = who_country['dale']
# Here we drop country 191, since otherwise there would be perfect colinearity␣
↪in
# the columns. We also have to drop dale since that's the endogenous variable we
# regress on.
X_3h = sm.add_constant(who_country.drop(columns=['dale', '191']))
model_3h = sm.OLS(y_3h, X_3h)
results_3h = model_3h.fit(cov_type = 'HC1')
results_3h.summary()
```

[12]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:                   dale   R-squared:                       0.999
Model:                            OLS   Adj. R-squared:                  0.999
Method:                 Least Squares   F-statistic:                 1.533e+04
Date:                Fri, 16 Apr 2021   Prob (F-statistic):               0.00
Time:                        15:27:33   Log-Likelihood:                -388.19
No. Observations:                 700   AIC:                             1062.
Df Residuals:                     557   BIC:                             1713.
Df Model:                         142
Covariance Type:                  HC1
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         25.4516      1.863     13.662      0.000      21.800      29.103
hexp           0.0013      0.000      5.452      0.000       0.001       0.002
hc3            2.2111      0.637      3.472      0.001       0.963       3.459
hc3_square    -0.0344      0.049     -0.705      0.481      -0.130       0.061
6             29.0775      0.519     55.976      0.000      28.059      30.096
7             23.3736      0.913     25.602      0.000      21.584      25.163
8             24.8046      0.875     28.336      0.000      23.089      26.520
```

10	24.7174	1.271	19.446	0.000	22.226	27.209
11	28.3661	0.818	34.672	0.000	26.763	29.970
14	5.0987	0.852	5.982	0.000	3.428	6.769
15	26.4998	0.961	27.572	0.000	24.616	28.384
16	12.7548	1.017	12.545	0.000	10.762	14.748
17	6.7134	1.000	6.717	0.000	4.754	8.672
18	18.2459	0.774	23.572	0.000	16.729	19.763
19	21.3087	1.106	19.267	0.000	19.141	23.476
20	27.9482	0.520	53.732	0.000	26.929	28.968
21	15.6358	0.874	17.885	0.000	13.922	17.349
23	19.5017	1.026	19.004	0.000	17.490	21.513
25	15.8023	0.585	27.034	0.000	14.657	16.948
26	24.4575	0.447	54.687	0.000	23.581	25.334
27	22.5857	0.866	26.086	0.000	20.889	24.283
30	0.1256	1.011	0.124	0.901	-1.857	2.108
31	6.6776	0.863	7.740	0.000	4.987	8.369
32	24.6037	1.254	19.617	0.000	22.145	27.062
34	25.5972	1.006	25.452	0.000	23.626	27.568
35	28.3550	0.738	38.432	0.000	26.909	29.801
36	24.0933	0.666	36.169	0.000	22.788	25.399
37	13.3835	0.838	15.978	0.000	11.742	15.025
38	8.8062	0.469	18.766	0.000	7.886	9.726
39	9.1366	0.587	15.556	0.000	7.985	10.288
40	26.2415	0.521	50.359	0.000	25.220	27.263
41	14.0563	0.687	20.461	0.000	12.710	15.403
42	24.6468	0.512	48.149	0.000	23.644	25.650
43	28.7095	0.606	47.348	0.000	27.521	29.898
45	28.1078	0.810	34.690	0.000	26.520	29.696
46	22.1949	1.291	17.187	0.000	19.664	24.726
47	24.2223	0.975	24.842	0.000	22.311	26.133
50	21.7461	1.343	16.198	0.000	19.115	24.377
51	26.5085	0.490	54.130	0.000	25.549	27.468
53	21.9511	0.673	32.597	0.000	20.631	23.271
54	24.7821	0.543	45.662	0.000	23.718	25.846
56	31.2424	0.755	41.387	0.000	29.763	32.722
57	19.7060	0.981	20.081	0.000	17.783	21.629
58	4.4154	0.919	4.806	0.000	2.615	6.216
59	24.4514	1.098	22.264	0.000	22.299	26.604
60	17.2416	0.911	18.921	0.000	15.456	19.028
61	30.7065	0.802	38.284	0.000	29.134	32.279
64	27.2638	0.960	28.402	0.000	25.382	29.145
65	21.5960	1.381	15.633	0.000	18.888	24.303
66	10.5944	0.468	22.651	0.000	9.678	11.511
68	16.9459	0.873	19.401	0.000	15.234	18.658
69	7.5342	0.982	7.671	0.000	5.609	9.459
70	8.4969	0.488	17.421	0.000	7.541	9.453
71	29.8744	0.870	34.326	0.000	28.169	31.580

73	20.8598	0.512	40.737	0.000	19.856	21.863
74	22.1902	0.624	35.542	0.000	20.967	23.414
75	25.4265	0.490	51.882	0.000	24.466	26.387
76	24.7924	0.918	27.018	0.000	22.994	26.591
77	12.5670	0.481	26.141	0.000	11.625	13.509
78	20.2190	1.048	19.284	0.000	18.164	22.274
79	22.5387	0.574	39.265	0.000	21.414	23.664
80	16.9896	0.506	33.567	0.000	15.998	17.982
81	26.6976	0.864	30.887	0.000	25.003	28.392
82	24.8306	0.483	51.426	0.000	23.884	25.777
83	21.7104	0.687	31.618	0.000	20.365	23.056
84	26.8536	0.875	30.682	0.000	25.138	28.569
85	24.8315	1.049	23.665	0.000	22.775	26.888
86	31.1969	0.759	41.120	0.000	29.710	32.684
87	31.2573	0.499	62.657	0.000	30.280	32.235
88	21.2591	0.666	31.901	0.000	19.953	22.565
89	29.3334	0.976	30.068	0.000	27.421	31.245
90	17.8749	0.845	21.163	0.000	16.219	19.530
91	6.4590	0.754	8.561	0.000	4.980	7.938
96	18.7771	1.391	13.498	0.000	16.051	21.504
97	23.5341	0.710	33.138	0.000	22.142	24.926
99	21.0862	0.697	30.236	0.000	19.719	22.453
103	24.1751	0.660	36.615	0.000	22.881	25.469
104	3.7786	0.488	7.746	0.000	2.823	4.735
105	21.8931	0.886	24.712	0.000	20.157	23.629
106	25.9592	0.943	27.534	0.000	24.111	27.807
107	18.8156	1.002	18.779	0.000	16.852	20.779
108	27.8877	0.802	34.772	0.000	26.316	29.460
110	20.7303	0.892	23.246	0.000	18.982	22.478
112	16.6955	0.513	32.576	0.000	15.691	17.700
113	24.1548	0.781	30.943	0.000	22.625	25.685
116	4.7303	1.324	3.572	0.000	2.135	7.326
117	32.5752	0.588	55.357	0.000	31.422	33.729
118	19.1301	0.625	30.592	0.000	17.904	20.356
120	5.9930	1.297	4.620	0.000	3.451	8.535
121	9.3063	0.652	14.267	0.000	8.028	10.585
122	19.8939	0.624	31.874	0.000	18.671	21.117
123	-1.8710	0.606	-3.089	0.002	-3.058	-0.684
124	22.7178	0.665	34.141	0.000	21.414	24.022
125	1.6015	0.504	3.177	0.001	0.613	2.590
126	1.0852	1.302	0.833	0.405	-1.467	3.638
127	3.8173	0.506	7.541	0.000	2.825	4.809
128	23.3011	0.515	45.230	0.000	22.291	24.311
129	27.1075	0.941	28.820	0.000	25.264	28.951
130	27.5020	0.872	31.534	0.000	25.793	29.211
131	18.2575	0.831	21.964	0.000	16.628	19.887
133	21.1532	1.576	13.419	0.000	18.063	24.243

134	32.1566	0.665	48.339	0.000	30.853	33.460
135	20.2355	0.485	41.695	0.000	19.284	21.187
136	23.5199	0.888	26.485	0.000	21.779	25.260
137	19.3761	0.723	26.798	0.000	17.959	20.793
138	18.0820	0.805	22.463	0.000	16.504	19.660
141	21.7076	1.155	18.794	0.000	19.444	23.971
143	30.0429	0.659	45.565	0.000	28.751	31.335
144	26.1618	0.538	48.633	0.000	25.107	27.216
145	24.7094	0.668	37.015	0.000	23.401	26.018
146	18.9351	1.171	16.166	0.000	16.639	21.231
147	17.0822	1.129	15.124	0.000	14.869	19.296
148	0.4349	0.500	0.870	0.384	-0.545	1.414
149	30.4420	0.456	66.787	0.000	29.549	31.335
150	12.1451	1.004	12.094	0.000	10.177	14.113
151	12.8805	0.731	17.618	0.000	11.448	14.313
152	30.1489	0.660	45.684	0.000	28.855	31.442
155	26.9217	0.531	50.741	0.000	25.882	27.962
159	21.3329	1.294	16.489	0.000	18.797	23.869
160	24.0181	0.965	24.877	0.000	22.126	25.910
161	26.6301	1.060	25.122	0.000	24.552	28.708
162	1.7255	0.515	3.351	0.001	0.716	2.735
164	20.9726	0.595	35.232	0.000	19.806	22.139
166	8.4630	0.554	15.277	0.000	7.377	9.549
167	22.2878	0.624	35.736	0.000	21.065	23.510
168	15.1486	0.882	17.167	0.000	13.419	16.878
169	13.7596	0.850	16.184	0.000	12.093	15.426
171	19.1344	1.105	17.324	0.000	16.970	21.299
172	23.9008	0.781	30.607	0.000	22.370	25.431
173	25.2474	0.483	52.306	0.000	24.301	26.193
174	27.5735	0.481	57.282	0.000	26.630	28.517
175	4.8150	0.626	7.690	0.000	3.588	6.042
176	4.2374	1.009	4.199	0.000	2.259	6.215
177	20.5772	1.092	18.840	0.000	18.436	22.718
178	25.6657	0.771	33.282	0.000	24.154	27.177
179	18.8771	1.562	12.082	0.000	15.815	21.939
180	20.7579	0.755	27.487	0.000	19.278	22.238
182	28.3711	0.533	53.214	0.000	27.326	29.416
183	16.1511	0.942	17.148	0.000	14.305	17.997
185	15.0584	1.476	10.202	0.000	12.166	17.951
186	20.3303	1.070	19.007	0.000	18.234	22.427
188	2.3579	0.613	3.846	0.000	1.156	3.559
190	-4.2994	0.865	-4.969	0.000	-5.995	-2.603

Omnibus:	99.523	Durbin-Watson:	1.402
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1100.479
Skew:	0.093	Prob(JB):	1.08e-239
Kurtosis:	9.140	Cond. No.	3.20e+05

```
==============================================================================

Warnings:
[1] Standard Errors are heteroscedasticity robust (HC1)
[2] The condition number is large, 3.2e+05. This might indicate that there are
strong multicollinearity or other numerical problems.
"""
```

Question 3.i. Explain.

With all the country variables added into the OLS regression, we're creating a clearer view of what education attainment would look like across countries.

Question 3.j. Give an example of an entity-invariant variable, which is excluded from the estimated regression model in part (a), that would result in variation in life expectancy over time.

An example of an entity-invariant variable would be diseases because no matter where you live, life expectancy would overall decrease overtime.

Question 3.k. Perform regression with time fixed effects. Are the results consistent with your reasoning about the entity-invariant variables? The procedure for this question will be similar to 3.h. Drop the dummy variable for 1993 for this question.

This question is for your code, the next is for your explanation.

[13]:
```python
no_1993 = who[who['year'] != 1993]
countries = pd.get_dummies(no_1993['country'], prefix='', prefix_sep='')
who_country = no_1993[['dale', 'hexp', 'hc3', 'hc3_square']].join(countries)
y_3k = who_country['dale']
X_3k = sm.add_constant(who_country.drop(columns=['dale', '191']))
model_3k = sm.OLS(y_3k, X_3k)
results_3k = model_3k.fit(cov_type = 'HC1')
results_3k.summary()
```

[13]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:                   dale   R-squared:                       0.999
Model:                            OLS   Adj. R-squared:                  0.999
Method:                 Least Squares   F-statistic:                 2.552e+04
Date:                Fri, 16 Apr 2021   Prob (F-statistic):               0.00
Time:                        15:27:35   Log-Likelihood:                -185.35
No. Observations:                 560   AIC:                             656.7
Df Residuals:                     417   BIC:                             1276.
Df Model:                         142
Covariance Type:                  HC1
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
```

const	25.7899	2.003	12.874	0.000	21.864	29.716
hexp	0.0011	0.000	4.239	0.000	0.001	0.002
hc3	1.9625	0.698	2.810	0.005	0.594	3.331
hc3_square	-0.0109	0.056	-0.195	0.845	-0.121	0.099
6	29.6638	0.483	61.394	0.000	28.717	30.611
7	23.6636	1.006	23.532	0.000	21.693	25.635
8	25.0503	0.931	26.909	0.000	23.226	26.875
10	24.9111	1.529	16.288	0.000	21.914	27.909
11	29.0894	0.843	34.502	0.000	27.437	30.742
14	5.0361	0.863	5.836	0.000	3.345	6.727
15	26.9585	1.086	24.829	0.000	24.830	29.087
16	12.8247	1.052	12.187	0.000	10.762	14.887
17	6.5827	1.018	6.464	0.000	4.587	8.579
18	18.5904	0.769	24.188	0.000	17.084	20.097
19	21.2083	1.261	16.823	0.000	18.737	23.679
20	28.4495	0.484	58.795	0.000	27.501	29.398
21	16.1231	0.935	17.251	0.000	14.291	17.955
23	19.4249	1.122	17.315	0.000	17.226	21.624
25	16.3003	0.553	29.479	0.000	15.217	17.384
26	24.9132	0.398	62.618	0.000	24.133	25.693
27	22.8860	0.930	24.620	0.000	21.064	24.708
30	-0.1695	0.914	-0.185	0.853	-1.961	1.622
31	6.5155	0.852	7.643	0.000	4.845	8.186
32	24.9302	1.507	16.544	0.000	21.977	27.884
34	26.2887	1.143	22.996	0.000	24.048	28.529
35	28.7514	0.745	38.588	0.000	27.291	30.212
36	24.3948	0.655	37.272	0.000	23.112	25.678
37	13.2370	0.822	16.106	0.000	11.626	14.848
38	9.1377	0.426	21.435	0.000	8.302	9.973
39	9.3235	0.554	16.836	0.000	8.238	10.409
40	26.6447	0.487	54.709	0.000	25.690	27.599
41	14.5189	0.659	22.036	0.000	13.228	15.810
42	24.9726	0.473	52.834	0.000	24.046	25.899
43	29.1772	0.583	50.088	0.000	28.036	30.319
45	28.4627	0.847	33.613	0.000	26.803	30.122
46	22.2900	1.519	14.674	0.000	19.313	25.267
47	24.8183	1.104	22.477	0.000	22.654	26.982
50	21.8925	1.635	13.391	0.000	18.688	25.097
51	26.9309	0.448	60.085	0.000	26.052	27.809
53	22.3589	0.663	33.741	0.000	21.060	23.658
54	25.2402	0.498	50.641	0.000	24.263	26.217
56	31.7925	0.766	41.480	0.000	30.290	33.295
57	19.8809	1.096	18.136	0.000	17.732	22.029
58	4.2931	0.931	4.610	0.000	2.468	6.118
59	24.8127	1.280	19.380	0.000	22.303	27.322
60	17.4046	0.989	17.593	0.000	15.466	19.344
61	31.4965	0.810	38.892	0.000	29.909	33.084

64	27.6326	1.078	25.644	0.000	25.521	29.745
65	21.3322	1.635	13.045	0.000	18.127	24.537
66	11.0140	0.420	26.224	0.000	10.191	11.837
68	17.3093	0.870	19.899	0.000	15.604	19.014
69	7.6403	1.021	7.487	0.000	5.640	9.640
70	8.9494	0.441	20.314	0.000	8.086	9.813
71	30.1694	0.936	32.229	0.000	28.335	32.004
73	21.2583	0.469	45.295	0.000	20.338	22.178
74	22.5920	0.602	37.549	0.000	21.413	23.771
75	25.9054	0.439	58.981	0.000	25.045	26.766
76	25.0890	0.977	25.676	0.000	23.174	27.004
77	12.7812	0.441	29.004	0.000	11.918	13.645
78	20.3679	1.186	17.174	0.000	18.043	22.692
79	23.0174	0.544	42.289	0.000	21.951	24.084
80	17.4239	0.473	36.831	0.000	16.497	18.351
81	27.0946	0.932	29.086	0.000	25.269	28.920
82	25.2954	0.441	57.315	0.000	24.430	26.160
83	21.6899	0.619	35.034	0.000	20.476	22.903
84	27.3695	0.946	28.933	0.000	25.515	29.224
85	25.1080	1.215	20.663	0.000	22.726	27.490
86	31.8991	0.761	41.916	0.000	30.407	33.391
87	31.6773	0.456	69.395	0.000	30.783	32.572
88	21.6240	0.655	32.995	0.000	20.339	22.909
89	29.7574	1.108	26.851	0.000	27.585	31.930
90	17.8439	0.795	22.438	0.000	16.285	19.403
91	6.3323	0.677	9.349	0.000	5.005	7.660
96	18.7132	1.673	11.186	0.000	15.434	21.992
97	23.8319	0.709	33.625	0.000	22.443	25.221
99	21.4687	0.689	31.139	0.000	20.117	22.820
103	24.5526	0.645	38.090	0.000	23.289	25.816
104	3.9315	0.439	8.954	0.000	3.071	4.792
105	22.0546	0.948	23.264	0.000	20.197	23.913
106	26.2801	1.090	24.103	0.000	24.143	28.417
107	18.8696	1.116	16.916	0.000	16.683	21.056
108	28.2153	0.808	34.941	0.000	26.633	29.798
110	20.7309	0.925	22.404	0.000	18.917	22.545
112	17.1862	0.471	36.515	0.000	16.264	18.109
113	24.4679	0.809	30.235	0.000	22.882	26.054
116	4.7716	1.397	3.416	0.001	2.034	7.509
117	33.0550	0.569	58.123	0.000	31.940	34.170
118	19.4960	0.602	32.407	0.000	18.317	20.675
120	6.0113	1.369	4.392	0.000	3.329	8.694
121	9.6392	0.637	15.130	0.000	8.391	10.888
122	20.2997	0.606	33.497	0.000	19.112	21.488
123	-1.8337	0.578	-3.171	0.002	-2.967	-0.700
124	23.0813	0.653	35.334	0.000	21.801	24.362
125	1.8037	0.451	4.003	0.000	0.920	2.687

126	1.0366	1.373	0.755	0.450	-1.655	3.728
127	4.2609	0.461	9.241	0.000	3.357	5.165
128	23.8209	0.457	52.095	0.000	22.925	24.717
129	27.5876	1.053	26.204	0.000	25.524	29.651
130	28.1394	0.932	30.199	0.000	26.313	29.966
131	18.5346	0.831	22.295	0.000	16.905	20.164
133	21.1198	1.926	10.968	0.000	17.346	24.894
134	32.5251	0.635	51.204	0.000	31.280	33.770
135	20.6788	0.439	47.122	0.000	19.819	21.539
136	23.7765	0.957	24.840	0.000	21.901	25.653
137	19.7977	0.722	27.426	0.000	18.383	21.213
138	18.4587	0.820	22.508	0.000	16.851	20.066
141	21.7494	1.332	16.324	0.000	19.138	24.361
143	30.5517	0.646	47.298	0.000	29.286	31.818
144	26.5841	0.503	52.893	0.000	25.599	27.569
145	25.3204	0.653	38.785	0.000	24.041	26.600
146	18.7758	1.348	13.929	0.000	16.134	21.418
147	16.9719	1.286	13.199	0.000	14.452	19.492
148	0.5830	0.455	1.282	0.200	-0.308	1.474
149	30.8412	0.409	75.394	0.000	30.039	31.643
150	12.4503	1.021	12.195	0.000	10.449	14.451
151	13.2219	0.723	18.279	0.000	11.804	14.640
152	30.6598	0.647	47.382	0.000	29.392	31.928
155	27.4536	0.477	57.522	0.000	26.518	28.389
159	21.2657	1.523	13.961	0.000	18.280	24.251
160	24.3333	1.084	22.455	0.000	22.209	26.457
161	27.0891	1.233	21.962	0.000	24.672	29.507
162	2.1823	0.475	4.597	0.000	1.252	3.113
164	21.3760	0.575	37.163	0.000	20.249	22.503
166	8.4953	0.497	17.102	0.000	7.522	9.469
167	22.6232	0.604	37.477	0.000	21.440	23.806
168	15.3189	0.947	16.169	0.000	13.462	17.176
169	13.8608	0.892	15.547	0.000	12.113	15.608
171	19.1014	1.264	15.114	0.000	16.624	21.578
172	24.2723	0.796	30.511	0.000	22.713	25.832
173	25.6858	0.447	57.418	0.000	24.809	26.563
174	28.0655	0.430	65.231	0.000	27.222	28.909
175	4.8004	0.586	8.189	0.000	3.651	5.949
176	3.9436	1.002	3.936	0.000	1.980	5.907
177	20.3876	1.205	16.919	0.000	18.026	22.749
178	26.0352	0.789	33.018	0.000	24.490	27.581
179	19.2559	1.946	9.894	0.000	15.442	23.070
180	21.0114	0.765	27.454	0.000	19.511	22.511
182	28.8284	0.496	58.154	0.000	27.857	29.800
183	16.3921	1.003	16.345	0.000	14.426	18.358
185	14.9796	1.749	8.562	0.000	11.551	18.409
186	20.4431	1.111	18.395	0.000	18.265	22.621

```
188              2.9145      0.578      5.043      0.000      1.782      4.047
190             -4.4484      0.796     -5.592      0.000     -6.008     -2.889
==============================================================================
Omnibus:                       74.007   Durbin-Watson:                   1.699
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              660.611
Skew:                           0.122   Prob(JB):                    3.55e-144
Kurtosis:                       8.315   Cond. No.                     4.04e+05
==============================================================================

Warnings:
[1] Standard Errors are heteroscedasticity robust (HC1)
[2] The condition number is large, 4.04e+05. This might indicate that there are
strong multicollinearity or other numerical problems.
"""
```

Question 3.l. Explain.

The results are consistent with our reasoning because when we exclude the year 1993 from our regression, the β_0^{hat} value increases. If our OLS regression excludes the entity-invariant variable of diseases, then we would expect the β_0^{hat} value to increase. Likewise, we can see that hc3_square is not statistically significant, meaning that whether the country is rich or poor, we would see that diseases would decrease the life expectancy of a population.

Question 3.m. Perform a test that all time fixed effects are jointly equal to zero. Remember that we excluded 1993. What is the result of your test?

This question is for your code, the next is for your explanation.

[21]:
```python
X_3m = sm.add_constant(no_1993[['hexp', 'hc3', 'hc3_square']])
model_3m = sm.OLS(y_3k, X_3m)
results_3m = model_3m.fit(cov_type = 'HC1')
A = np.identity(len(results_3m.params))
A = A[1:,:]
results_3m.f_test(A)
```

[21]:
```
<class 'statsmodels.stats.contrast.ContrastResults'>
<F test: F=array([[507.75431101]]), p=9.114960539411242e-159, df_denom=556,
df_num=3>
```

Question 3.n. Explain.

Based on the p-value of 9.11, we fail to reject the null hypothesis, that the time fixed effects are jointly equal to zero. As a result, we concluded that time effect on life expectancy is not statistically significant.