ECN 145 PS4

12/3/2020

World Trade Organization *'Bilateral Trade'*

2,) *Open the dataset and summarize the data.*

	valueCars	valueCoal	valueCoffee	valueIron_Steel	valueNaturalGas	valueOil	year	contig	comlang_off	comlang_ethno	...
count	2.545000e+03	6.510000e+02	2.466000e+03	3.612000e+03	1.124000e+03	6.160000e+02	5149.0	4699.000000	4699.000000	4699.000000	...
mean	1.629451e+08	1.189941e+08	9.021077e+06	4.611430e+07	1.426316e+08	9.892174e+08	2016.0	0.037880	0.176846	0.197914	...
std	1.572998e+09	6.394982e+08	7.973874e+07	2.820196e+08	6.789939e+08	3.422172e+09	0.0	0.190927	0.381579	0.398470	...
min	4.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	2016.0	0.000000	0.000000	0.000000	...
25%	4.068600e+04	4.398500e+03	3.404250e+03	2.820875e+04	1.222675e+04	2.666100e+04	2016.0	0.000000	0.000000	0.000000	...
50%	5.946700e+05	7.160300e+04	5.148900e+04	5.151225e+05	6.983370e+05	6.130904e+07	2016.0	0.000000	0.000000	0.000000	...
75%	1.348556e+07	5.990502e+06	6.950085e+05	7.296976e+06	1.993981e+07	4.149250e+08	2016.0	0.000000	0.000000	0.000000	...
max	4.573710e+10	9.904087e+09	2.437543e+09	6.508431e+09	1.027974e+10	5.076936e+10	2016.0	1.000000	1.000000	1.000000	...

3.) *Which countries are the largest collective exporters / importers of steel/iron and cars respectively.*

destination_fullname	valueIron_Steel
European Union	2.517138e+10
United States of America	2.123293e+10
China	1.645462e+10
Korea, Republic of	1.427246e+10
Thailand	9.638898e+09

destination_fullname	valueIron_Steel	valueCars
United States of America	2.123293e+10	1.782180e+11
China	1.645462e+10	4.400503e+10
European Union	2.517138e+10	2.895337e+10
Canada	5.811697e+09	2.637946e+10
Australia	7.082008e+08	1.608431e+10

origin_fullname	valueIron_Steel
China	3.228293e+10
Japan	2.328142e+10
Korea, Republic of	1.608441e+10
United States of America	1.038480e+10
Russian Federation	9.350957e+09

origin_fullname	valueIron_Steel	valueCars
Japan	2.328142e+10	8.164315e+10
Germany	4.444439e+09	6.746919e+10
Canada	4.947460e+09	4.792417e+10
United States of America	1.038480e+10	4.278803e+10
Korea, Republic of	1.608441e+10	3.635469e+10

From the above tables, we can see that the European Union and United States of America are the largest importers of Steel/Iron and Cars respectively. We can also see that the People's Republic of China and Japan are the largest exporters of Steel/Iron and Cars respectively.

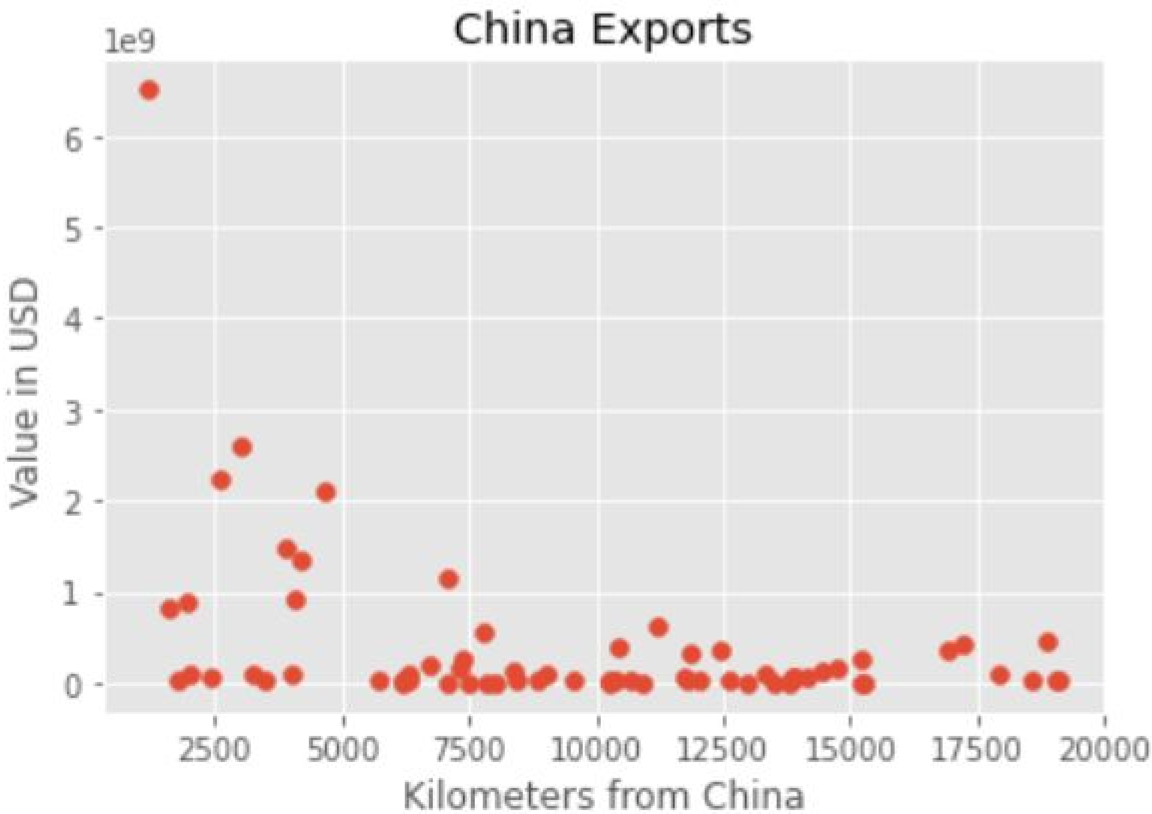
4.) *Create a table calculating the correlation coefficients between the Value of all the different products tracked. Which products are the most strongly positively correlated. Which are the most strongly negatively correlated? Do these correlations make sense?*

	valueCars	valueCoal	valueCoffee	valueIron_Steel
valueCars	1.000000	-0.018219	0.110200	0.367828
valueCoal	-0.018219	1.000000	0.028098	0.058277
valueCoffee	0.110200	0.028098	1.000000	0.198240
valueIron_Steel	0.367828	0.058277	0.198240	1.000000
valueNaturalGas	0.215737	0.462897	0.084673	0.215950
valueOil	0.403410	0.138252	0.071457	0.308118

	valueNaturalGas	valueOil
valueCars	0.215737	0.403410
valueCoal	0.462897	0.138252
valueCoffee	0.084673	0.071457
valueIron_Steel	0.215950	0.308118
valueNaturalGas	1.000000	0.430444
valueOil	0.430444	1.000000

Coefficients of correlations between natural gas and coal tend to be the most positively related, while those between coal and cars tend to be the most negatively correlated, after we drop all missing values.

5.) *Create a scatterplot of the value of Iron and Steel exports from China as a function of distance to the destination. What would a gravity model suggest for the relationship between exports and distance? Is that what you observe?*



The Gravity model does seem to apply in terms of trade with China. The scatter plot displays that as distance between a country and China increases, it’s values of trade seem to fall exponentially.

6.) *Calculated the log of the value variables, the log of GDP and the log of distance.*

a.Run a gravity regression, predicting log(value) for Iron and Steel shipments as a function of logged origin and destination GDP and log distance. Interpret the coefficients.

b.Now add the legal, cultural and economic similarity variables described above. Are the signs as you expect?

```
                            OLS Regression Results
==============================================================================
Dep. Variable:       valueIron_Steel   R-squared:                       0.355
Model:                           OLS   Adj. R-squared:                  0.354
Method:                Least Squares   F-statistic:                     580.2
Date:               Wed, 02 Dec 2020   Prob (F-statistic):         1.48e-300
Time:                       20:31:25   Log-Likelihood:               -8057.8
No. Observations:               3173   AIC:                        1.612e+04
Df Residuals:                   3169   BIC:                        1.615e+04
Df Model:                          3
Covariance Type:           nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept    -22.5122      1.114    -20.204      0.000     -24.697     -20.328
gdp_o          0.9699      0.030     32.557      0.000       0.911       1.028
gdp_d          0.8454      0.026     32.352      0.000       0.794       0.897
distw         -1.3853      0.066    -20.965      0.000      -1.515      -1.256
==============================================================================
Omnibus:                     226.417   Durbin-Watson:                   1.597
Prob(Omnibus):                 0.000   Jarque-Bera (JB):              276.922
Skew:                         -0.687   Prob(JB):                     7.36e-61
Kurtosis:                      3.456   Cond. No.                         777.
==============================================================================
```

Here, we can see that as distance between a country increases, the value of Iron/Steel tends to decrease, whereas higher GDP countries tend to trade higher values of Iron/Steel, vice versa. This goes hand in hand with what we expected of the Gravity model to depict.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:       valueIron_Steel   R-squared:                       0.363
Model:                           OLS   Adj. R-squared:                  0.361
Method:                Least Squares   F-statistic:                     225.5
Date:               Wed, 02 Dec 2020   Prob (F-statistic):          3.49e-303
Time:                       20:38:24   Log-Likelihood:                -8036.6
No. Observations:               3173   AIC:                         1.609e+04
Df Residuals:                   3164   BIC:                         1.615e+04
Df Model:                          8
Covariance Type:           nonrobust
================================================================================
                  coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
Intercept     -25.0301      1.191    -21.010      0.000     -27.366     -22.694
gdp_o           0.9823      0.030     32.477      0.000       0.923       1.042
gdp_d           0.8407      0.026     31.859      0.000       0.789       0.892
distw          -1.1386      0.083    -13.725      0.000      -1.301      -0.976
comlang_off     0.2074      0.160      1.296      0.195      -0.106       0.521
contig          1.0924      0.288      3.791      0.000       0.527       1.657
comcur          2.1635      0.584      3.704      0.000       1.018       3.309
comrelig       -0.0661      0.223     -0.296      0.767      -0.504       0.371
fta_wto         0.2375      0.137      1.735      0.083      -0.031       0.506
==============================================================================
Omnibus:                     233.558   Durbin-Watson:                   1.589
Prob(Omnibus):                 0.000   Jarque-Bera (JB):              288.102
Skew:                         -0.696   Prob(JB):                     2.75e-63
Kurtosis:                      3.493   Cond. No.                         839.
==============================================================================
```

Adding additional coefficients reduces omitted variable bias, and depicts a true-er value of these coefficients. As we can see, distance in fact is not as negatively correlated with the value of Iron/Steel trade, as common religion also accounts for some of the negative correlation.

7.) *Finally, run a similar regression to that in part 6b for Cars shipments. Are the coefficients the same or different than in part 6b? If they are different, do the differences make sense?*

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              valueCars   R-squared:                       0.371
Model:                            OLS   Adj. R-squared:                  0.370
Method:                 Least Squares   F-statistic:                     441.3
Date:                Wed, 02 Dec 2020   Prob (F-statistic):          2.38e-225
Time:                        20:44:24   Log-Likelihood:                -5622.5
No. Observations:                2250   AIC:                         1.125e+04
Df Residuals:                    2246   BIC:                         1.128e+04
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept    -24.3777      1.157    -21.061      0.000     -26.648     -22.108
gdp_o          1.1269      0.036     31.509      0.000       1.057       1.197
gdp_d          0.5323      0.028     19.083      0.000       0.478       0.587
distw         -0.6541      0.076     -8.628      0.000      -0.803      -0.505
==============================================================================
Omnibus:                       67.166   Durbin-Watson:                   1.269
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               40.265
Skew:                          -0.178   Prob(JB):                     1.81e-09
Kurtosis:                       2.450   Cond. No.                         707.
==============================================================================
```

In this model, we get similar results to that of our previous model. Distance still plays a decreasing role in trade for cars, and the larger the countries are GDP-wise, the greater the value of the trade of cars.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              valueCars   R-squared:                       0.383
Model:                            OLS   Adj. R-squared:                  0.381
Method:                 Least Squares   F-statistic:                     174.0
Date:                Wed, 02 Dec 2020   Prob (F-statistic):          1.03e-228
Time:                        20:44:40   Log-Likelihood:                -5600.3
No. Observations:                2250   AIC:                         1.122e+04
Df Residuals:                    2241   BIC:                         1.127e+04
Df Model:                           8
Covariance Type:            nonrobust
===============================================================================
                  coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------
Intercept     -26.3241      1.254    -20.986      0.000     -28.784     -23.864
gdp_o           1.1146      0.036     30.767      0.000       1.044       1.186
gdp_d           0.5012      0.028     17.748      0.000       0.446       0.557
distw          -0.3348      0.096     -3.484      0.001      -0.523      -0.146
comlang_off    -0.4014      0.184     -2.187      0.029      -0.761      -0.041
contig          0.4822      0.319      1.510      0.131      -0.144       1.108
comcur          0.6805      0.548      1.241      0.215      -0.395       1.756
comrelig        0.1663      0.255      0.653      0.514      -0.333       0.666
fta_wto         0.9291      0.155      5.986      0.000       0.625       1.233
==============================================================================
Omnibus:                       59.328   Durbin-Watson:                   1.283
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               38.542
Skew:                          -0.191   Prob(JB):                     4.27e-09
Kurtosis:                       2.485   Cond. No.                         777.
==============================================================================
```

Similar to the secondary model created for Iron and Steel, we can see the omitted variable bias being mitigated as further variables are introduced. Distance is not so negatively correlated as common language captures some of the negative effects as well. This could be due to countries sharing common languages may produce similar types of vehicles as well resulting in a lack of trade among common cars.

Do-file:

```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import scipy

from statsmodels.formula.api import ols

from sklearn.linear_model import LinearRegression


file = r'C:/Users/19085/Downloads/ps4.csv'

pd.read_csv(file)

df = pd.read_csv(file)

df.describe()

valtab = pd.DataFrame(df, columns = ['valueCars', 'valueCoal', 'valueCoffee', 'valueIron_Steel',
'valueNaturalGas', 'valueOil'])


CorrMatrix = valtab.corr()

print(CorrMatrix)

dfsc = df[['destination_fullname','origin_fullname', 'valueIron_Steel', 'valueCars']]

dfsc

dfsci = dfsc.groupby('destination_fullname').agg({'valueIron_Steel':'sum', 'valueCars' :'sum'})

dfsci.sort_values(by = ['valueIron_Steel'], ascending = False)

#3 EU is the largest importer


dfsci.sort_values(by = ['valueCars'], ascending = False)

#America imports the largest number of cars

dfsck = dfsc.groupby('origin_fullname').agg({'valueIron_Steel':'sum', 'valueCars' :'sum'})
```

```
dfsck.sort_values(by = ['valueIron_Steel'], ascending = False)

#Largest exporter of steel

dfsck.sort_values(by = ['valueCars'], ascending = False)

#Largest exporter of cars

ie = df[df['origin_fullname'] == 'China']

plt.style.use('ggplot')

fig, ax = plt.subplots()

scatplt = ax.scatter(

ie['distw'],

ie['valueIron_Steel']

)

ax.set_xlabel('Kilometers from China')

ax.set_ylabel('Value in USD')

ax.set_title('China Exports')

plt.show()

dflog = df[(df.T != 0).any()]

dflog.dropna()


dflog['valueIron_Steel'] = np.log(dflog['valueIron_Steel'])

dflog['gdp_o'] = np.log(dflog['gdp_o'])

dflog['gdp_d'] = np.log(dflog['gdp_d'])

dflog['distw'] = np.log(dflog['distw'])

model = ols('valueIron_Steel ~ gdp_o + gdp_d + distw', dflog).fit()

print(model.summary())

model3 = ols('valueIron_Steel ~ gdp_o + gdp_d + distw + comlang_off + contig + comcur +

comrelig + fta_wto', dflog).fit()
```

```
print(model3.summary())

dflog['valueCars'] = np.log(dflog['valueCars'])

model4 = ols('valueCars ~ gdp_o + gdp_d + distw', dflog).fit()

print(model4.summary())

model5 = ols('valueCars ~ gdp_o + gdp_d + distw + comlang_off + contig + comcur + comrelig
+ fta_wto', dflog).fit()

print(model5.summary())
```