Financial Econometrics
Shanghai Jiao Tong University
Summer School 2026
Empirical Exercise 10.2

Fatih Kansoy

Reference: Stock, J.H. and Watson, M.W., Introduction to Econometrics (4th Edition, Global Edition)

Data: `income_democracy.csv`*, a panel of countries observed in 1960, 1965, . . . , 2000. The democracy index* `Dem_ind` *is the column* `dem_ind`*,* `Log_GDPPC` *is the column* `log_gdppc`*, and the country identifier is the column* `code`*.*

Empirical Exercise 10.2. Do citizens demand more democracy and political freedom as their incomes grow? That is, is democracy a normal good? In this exercise you will use the data file `income_democracy.csv`, a panel data set covering countries for the years 1960, 1965, . . . , 2000. The data set contains an index of political freedom/democracy for each country in each year, together with data on the country's income and various demographic controls. (The income and demographic controls are lagged five years relative to the democracy index to allow time for democracy to adjust to changes in these variables.) The democracy index is the column `dem_ind` (S&W's `Dem_ind`), the logarithm of per capita income is the column `log_gdppc` (S&W's `Log_GDPPC`), and the country identifier is the column `code`.

(a) Is the data set a balanced panel? Explain.

(b) The index of political freedom/democracy is labelled `Dem_ind`.

(i) What are the minimum and maximum values of `Dem_ind` in the data set? What are the mean and standard deviation of `Dem_ind` in the data set? What are the 10th, 25th, 50th, 75th, and 90th percentiles of its distribution?

(ii) What is the value of `Dem_ind` for the United States in 2000? Averaged over all years in the data set?

(iii) What is the value of `Dem_ind` for Libya in 2000? Averaged over all years in the data set?

(iv) List five countries with an average value of `Dem_ind` greater than 0.95; less than 0.10; and between 0.3 and 0.7.

(c) The logarithm of per capita income is labelled `Log_GDPPC`. Regress `Dem_ind` on `Log_GDPPC`. Use standard errors that are clustered by country.

(i) How large is the estimated coefficient on `Log_GDPPC`? Is the coefficient statistically significant?

(ii) If per capita income in a country increases by 20%, by how much is `Dem_ind` predicted to increase? What is a 95% confidence interval for the prediction? Is the predicted increase in `Dem_ind` large or small? (Explain what you mean by large or small.)

(iii) Why is it important to use clustered standard errors for the regression? Do the results change if you do not use clustered standard errors?

(d)

(i) Suggest a variable that varies across countries but plausibly varies little, or not at all, over time and that could cause omitted variable bias in the regression in (c).

(ii) Estimate the regression in (c), allowing for country fixed effects. How do your answers to (c)(i) and (c)(ii) change?

(iii) Exclude the data for Azerbaijan and rerun the regression. Do the results change? Why or why not?

(iv) Suggest a variable that varies over time but plausibly varies little, or not at all, across countries and that could cause omitted variable bias in the regression in (c).

(v) Estimate the regression in (c), allowing for time and country fixed effects. How do your answers to (c)(i) and (c)(ii) change?

(vi) There are additional demographic controls in the data set. Should these variables be included in the regression? If so, how do the results change when they are included?

(e) Based on your analysis, what conclusions do you draw about the effects of income on democracy?

Calculations performed using Stata. See accompanying `E10_2_solution.do` *file.*