Financial Econometrics
Shanghai Jiao Tong University
Summer School 2026

Empirical Exercises for Lab 2

Regression with One Regressor: Estimation and Inference

Fatih Kansoy

After Stock and Watson, Introduction to Econometrics (4th edition), Chapters 4 and 5. These are the four exercises worked through in Lab 2.

Data: `Growth.csv` (65 countries; used in 4.1 and 5.2) and `Earnings_and_Height.csv` (US workers; used in 4.2 and 5.1). Both files are in the lab's `data` folder.

How to work these: answer each question with software. Report robust (heteroskedasticity-consistent) standard errors throughout: the `robust` option in Stata, `cov_type="HC1"` in Python. The worked solutions are the Lab 2 files, `lab2_regression_practice.do` and `lab2_regression_practice.ipynb`.

Exercise 4.1 *Economic growth and trade*

`Growth.csv` holds, for each of 65 countries, its average annual GDP growth rate over 1960–1995 and its average *trade share*, the sum of exports and imports divided by GDP. Investigate whether countries that trade more tend to grow faster.

(a) Plot growth against the trade share. Does a relationship appear, and in which direction?

(b) One country has a trade share far above the rest. Identify it on the plot. Does it look like an outlier?

(c) Fit an OLS regression of growth on the trade share using all 65 countries. Report the estimated slope and intercept, and use the fitted line to predict the growth rate of a country with a trade share of 0.5 and of one with a trade share of 1.0.

(d) Refit the regression with that one country removed, and answer the same questions as in part (c).

(e) Draw both fitted lines on the scatterplot. Using the picture, explain why including the outlier makes the line steeper.

(f) Where is this country, and why is its trade share so large? On balance, should it be kept in the analysis or set aside? Give your reasoning.

Exercise 4.2 *Earnings and height*

`Earnings_and_Height.csv` is a random sample of US workers, with annual earnings, height (in inches), sex, occupation and other characteristics. Investigate whether taller workers earn more.

(a) Find the median height in the sample. Then:

(i) estimate mean earnings for workers no taller than 67 inches;

(ii) estimate mean earnings for workers taller than 67 inches;

(iii) is the difference large? Report it with a 95% confidence interval.

(b) Make a scatterplot of earnings against height. The points sit on a small number of horizontal lines (there are only 23 distinct earnings values). Explain why, using the data description.

(c) Regress earnings on height.

(i) Report the estimated slope.

(ii) Predict earnings for workers who are 65, 67 and 70 inches tall.

(d) Now suppose height were recorded in centimetres rather than inches (1 inch = 2.54 cm). For the regression of earnings on height in centimetres, state the slope, the intercept, the R^2 and the standard error of the regression. Which of these change, and which do not?

(e) Regress earnings on height for women only. Report the slope, and say by how much you would predict a woman one inch taller than the average woman to earn above or below the average woman.

(f) Repeat part (e) for men.

(g) Do you think height is uncorrelated with the other things that drive earnings, so that $\mathbb{E}(u_i \mid X_i) = 0$ for X_i height? Explain what this means for reading the slope as an effect.

Exercise 5.1 *Earnings and height: inference*

Continue with `Earnings_and_Height.csv` from Exercise 4.2, now attaching standard errors, tests and intervals to the slope.

(a) Regress earnings on height for the full sample.

(i) Is the slope statistically significant?

(ii) Give a 95% confidence interval for it.

(b) Repeat part (a) for women.

(c) Repeat part (a) for men.

(d) Test whether the height slope is the same for men and women. (The two estimates come from separate samples, so the standard error of their difference is $\sqrt{SE_m^2 + SE_f^2}$.)

(e) One story is that height matters only because some jobs reward physical strength. Does the height slope survive when the sample is limited to occupations where strength is unlikely to matter?

Exercise 5.2 *Growth and trade: inference*

Return to `Growth.csv` from Exercise 4.1, *excluding* the outlier country, and regress growth on the trade share.

(a) Is the slope statistically significant? State whether you can reject $H_0 : \beta_1 = 0$ against a two-sided alternative at the 10%, 5% and 1% levels.

(b) Report the p-value of the slope's t-statistic.

(c) Construct a 90% confidence interval for β_1.

Worked solutions. Every part above is answered, with output and interpretation, in `lab2_regression_practice.do` (Stata) and `lab2_regression_practice.ipynb` (Python). The two produce the same numbers.