Financial Econometrics | Lecture 3

Linear Regression with One Regressor

In Class Group Activity

· SJTU, Summer School 2026 · Fatih Kansoy

How this block runs. You work in your home group of three or four, textbook open. Roles rotate: **Scribe**, **Reporter**, **Sceptic**, **Timekeeper**. All money today is in US dollars ($).

- **Phase 1, Concept checks (~10 min).** Vote on your own, talk for 90 seconds, then vote again.
- **Phase 2, Desk problem (~20 min).** Fit a line by hand, then check it in software.
- **Phase 3, Case (~25 min).** One real story, ending hidden. You take a position, then we reveal what happened.

A short word list. The **OLS slope** $\widehat{\beta}_1 = s_{XY}/s_X^2$ is the change in the average of Y per one-unit rise in X; with a 0/1 **dummy** regressor it is the difference in group means. A **residual** $\widehat{u}_i = Y_i - \widehat{Y}_i$ is the sample gap from the fitted line (not the error u_i). R^2 is the fraction of the variation in Y the regressor explains; **SER** is the typical residual size in the units of Y. The slope is causal only when $\mathbb{E}(u \mid X) = 0$, that is, when no **confounder** sits in the error and moves with X. A stock's **CAPM beta** is just the OLS slope of its return on the market's return.

Phase 1 – Concept checks

~10 MIN

ConcepTest 1. Fitting hourly earnings on a college-degree dummy gives $\widehat{\text{ahe}} = 13.53 + 7.82\,\text{bachelor}$. A classmate says: "A 0/1 regressor is odd; you need a special test to compare the two groups." Which is correct?

(a) Correct: a dummy regressor needs different machinery from an ordinary regression.

(b) The intercept 13.53 is the college premium.

(c) The slope 7.82 *is* the difference in means (college mean 21.34 minus high-school mean 13.53), and the test is the same difference-in-means t-test from Lecture 2.

(d) You cannot run a regression on a 0/1 variable.

ConcepTest 2. Regressing hourly earnings on age gives $R^2 = 0.014$, yet the slope has a t-statistic of about 14. A classmate says: "R^2 is only 0.014, so age does not matter and the slope is not significant." Which is correct?

(a) Correct: a low R^2 means the regressor is unimportant.

(b) A low R^2 means the slope cannot be statistically significant.

(c) Fit (R^2), significance (t) and effect size are three different things: age explains little of the wage spread, yet the slope is real and overwhelmingly significant.

(d) The R^2 must be wrong, because the t-statistic is so large.

ConcepTest 3. Across 17,870 US workers, regressing annual earnings on height gives a slope

of about \$708 per inch, with a tiny p-value (it is highly significant). A classmate says: "So being taller *causes* higher pay." Which is correct?

(a) Correct: a highly significant slope is a causal effect.

(b) The slope is precisely estimated, but "significant" is not "causal". Height likely proxies for omitted factors (health, strength, occupation) that sit in u, so $\mathbb{E}(u \mid X) \neq 0$.

(c) Because the correlation is near zero, height and earnings are unrelated.

(d) Zero correlation and $\mathbb{E}(u \mid X) = 0$ are the same thing, so the slope is fine.

Phase 2 – The desk problem

~20 MIN

A market-model regression, by hand

The setting. You estimate a stock's **market model**: regress the stock's monthly excess return Y on the market's excess return X, both in per cent. The slope is the stock's **CAPM beta**. Six months of data (treat them as the whole sample, divide by $n = 6$):

Month	**Market X (%)**	**Stock Y (%)**
1	0	3
2	2	1
3	4	8
4	6	11
5	8	10
6	10	18

(a) Fit the line by hand. Using $\widehat{\beta}_1 = s_{XY}/s_X^2$ and $\widehat{\beta}_0 = \overline{Y} - \widehat{\beta}_1 \overline{X}$:

(1) Compute $\overline{X}$ and $\overline{Y}$.

(2) Fill the working table (deviations and their products) and find $\sum(X - \overline{X})(Y - \overline{Y})$ and $\sum(X - \overline{X})^2$.

(3) Compute the beta $\widehat{\beta}_1$ and the alpha $\widehat{\beta}_0$.

(4) Predict the stock's return in a month when the market returns 12%.

(5) Compute the residual for Month 1.

Month	$X - \overline{X}$	$Y - \overline{Y}$	$(X - \overline{X})(Y - \overline{Y})$	$(X - \overline{X})^2$
1				
2				
3				
4				
5				
6				
Sum				

(b) Read the numbers like an analyst.

(1) Is this stock more or less volatile than the market? What does the beta tell you?

(2) A colleague measured the market return in decimals $(0.00, 0.02, \dots)$ instead of per cent, so every X is divided by 100. What happens to the beta? What happens to R^2?

(3) Here the intercept (the alpha) *is* meaningful, unlike the test-score intercept in the lecture. In one sentence, say what the alpha measures.

(c) Confirm in software. Write the one line you would run in Stata, and the one line in Python (statsmodels), to reproduce your beta with robust standard errors.

Phase 3 – The case

~25 MIN

Does being taller raise your pay?

How to run this. Read the story with your group. The ending is hidden on purpose. Commit a written position before the reveal. Do not search the internet.

The story. Two economists study a large US survey of 17,870 working adults. For each person they have annual earnings and height in inches. They regress earnings on height and find a slope of about **$708 per inch**: a worker one inch taller earns, on average, about $708 more per year. With nearly 18,000 people the estimate is very precise, and the p-value is tiny, so the slope is highly statistically significant. The R^2, though, is only about 0.011: height explains a little over one per cent of the variation in pay.

The result is real, precise and significant. A newspaper picks it up: "Science says tall people earn more." A tall reader wonders whether they should ask for a raise.

The story stops here.

Four discussion questions (answer all four; take a position):

1. **Read the slope.** State the $708 result in plain words with units. Convert it to dollars per centimetre (one inch is 2.54 cm).
2. **Is a low R^2 a reason to dismiss it?** The R^2 is only 0.011. Does that make the slope unimportant or insignificant? Explain.
3. **What is in the error?** For the slope to be the causal effect of height we need $\mathbb{E}(u \mid X) = 0$. Name **two** omitted factors that plausibly sit in u and move with height.
4. **The subgroups.** Run separately, the slope is about $511 per inch for women and about $1,307 per inch for men. Take a position: what does this gap show, and what does it *not* prove?

End of activity handout. Solutions are distributed after the block.