Financial Econometrics 2026
Shanghai Jiao Tong University
Summer School 2026
Midterm Examination

Reading time: 10 minutes. **Writing time:** 120 minutes.

Total marks: 100. **Calculators are permitted.**

Instructions: Answer **all** questions. In Section A, circle the letter of the single best answer. In Sections B and C, show your working and write in clear sentences. All standard errors are heteroskedasticity-robust.

Structure:

- **Section A** – Multiple choice: 20 questions, 2.5 marks each (**50 marks**).
- **Section B** – One analytical problem in five parts (**30 marks**).
- **Section C** – One short interpretation (**20 marks**).

Useful facts. $SE(\bar{Y}) = s_Y/\sqrt{n}$; $t = \dfrac{\text{estimate} - \text{claimed value}}{SE}$; 95% CI = estimate $\pm 1.96\,SE$ (90%: 1.64; 99%: 2.58); $\widehat{\beta}_1 = s_{XY}/s_X^2$; $r = s_{XY}/(s_X s_Y)$; with one regressor $R^2 = r^2$; binary regressor: $\beta_0 = \mathbb{E}[Y \mid D{=}0]$, $\beta_1 = \mathbb{E}[Y \mid D{=}1] - \mathbb{E}[Y \mid D{=}0]$; sign of omitted-variable bias = $\text{sign}\big[\text{corr}(X, Z) \times \text{effect of } Z \text{ on } Y\big]$; $\bar{R}^2 = 1 - \frac{n-1}{n-k-1}\frac{SSR}{TSS}$; 5% critical values of F: 3.84 ($q = 1$), 3.00 ($q = 2$), 2.60 ($q = 3$); 1%, $q = 2$: 4.61; $F = t^2$ for one restriction.

Section A: Multiple Choice (20 questions × 2.5 marks = 50 marks)

1. A researcher observes the quarterly return on equity of 35 commercial banks, each bank observed over the same 8 quarters. This data set is:

(A) panel data.

(B) cross-sectional data.

(C) time-series data.

(D) experimental data.

2. A one-year investment returns $+8\%$ with probability 0.75 and -4% with probability 0.25. Its expected return is:

(A) 5%.

(B) 6%.

(C) 4%.

(D) 2%.

3. Two returns have covariance 6; their standard deviations are 2 and 5. Their correlation is:

(A) 0.30.

(B) 1.20.

(C) 0.24.

(D) 0.60.

4. Daily returns are known to be non-normal, with fat tails. The distribution of the *average* of 250 such returns is nevertheless approximately normal because of:

(A) the law of large numbers.

(B) the central limit theorem.

(C) the assumption of homoskedasticity.

(D) the Gauss-Markov theorem.

5. A sample of $n = 100$ months has standard deviation $s_Y = 4\%$. The standard error of the sample mean return is:

(A) 0.4%.

(B) 4%.

(C) 0.04%.

(D) 2%.

6. A test of $H_0 : \mu = 0$ produces a p-value of 0.02. The correct reading is:

(A) there is only a 2% chance that the null hypothesis is true.

(B) under the null, data this extreme occur about 2% of the time.

(C) the estimated mean return equals 2% per month.

(D) the testing procedure has a 2% chance of being wrong.

7. In a sample, the covariance between X and Y is 8 and the variance of X is 5. The OLS slope of Y on X is:

(A) 0.63.

(B) 2.29.

(C) 1.60.

(D) 40.0.

8. A single-regressor regression reports $R^2 = 0.09$. This means:

(A) the slope is statistically insignificant.

(B) the regression is misspecified.

(C) 9% of the slope is due to chance.

(D) X explains 9% of the variation in Y.

9. A slope is estimated as 0.9 with standard error 0.45. At the 5% level, the test of $H_0 : \beta_1 = 0$:

(A) fails to reject, since $t = 0.5$.

(B) rejects, since $t = 2.0$.

(C) rejects, since $t = 4.5$.

(D) fails to reject, since $t = 2.0$.

10. A coefficient is estimated as 3.0 with standard error 1.0. Which claimed value is *rejected* by a two-sided 5% test?

(A) 5.0.

(B) 2.0.

(C) 4.0.

(D) 1.5.

11. In $\widehat{Y} = 12 + 3D$, where $D = 1$ for firms paying a dividend and 0 otherwise, the average of Y among dividend payers is:

(A) 12.

(B) 3.

(C) 15.

(D) 9.

12. The reason this course always uses heteroskedasticity-robust standard errors is that they are:

(A) smaller than the homoskedasticity-only ones.

(B) exact in small as well as large samples.

(C) required for the slope to be unbiased.

(D) valid whether or not the error spread is constant.

13. Omitting a variable Z from a regression of Y on X biases the slope on X when:

(A) Z affects Y, even if unrelated to X.

(B) Z is correlated with X, even if it does not affect Y.

(C) Z affects Y and is correlated with X.

(D) Z is measured with error.

14. An omitted variable Z *lowers* Y, and Z is *positively* correlated with X. The slope on X estimated without Z is biased:

(A) upward.

(B) not at all.

(C) downward.

(D) towards +1.

15. In $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$, the coefficient β_2 is read as the effect of a one-unit change in X_2:

(A) together with the usual change in X_1.

(B) for the average firm only.

(C) ignoring X_1 entirely.

(D) holding X_1 fixed.

16. Compared with R^2, the adjusted $\bar{R}^2$:

(A) is always at least as large as R^2.

(B) charges a penalty for each added regressor.

(C) cannot decrease when a variable is added.

(D) measures the model's out-of-sample fit.

17. A regression includes market capitalisation in millions *and* the same capitalisation in billions as two separate regressors. The software will:

(A) fail to run, or drop one of the two.

(B) report identical coefficients on both.

(C) run normally with no warning.

(D) double the R^2.

18. To test the joint null that three factor loadings are all zero, the correct procedure is:

(A) three t-tests, rejecting if any exceeds 1.96.

(B) one t-test on the sum of the coefficients.

(C) three t-tests at the 1% level each.

(D) one F-test with $q = 3$ restrictions.

19. A robust F-statistic of 5.0 is reported for a joint null with $q = 3$ restrictions. At the 5% level:

(A) do not reject, since the cutoff is 5.99.

(B) do not reject, since the cutoff is 7.81.

(C) reject, since the cutoff is 2.60.

(D) reject, since $5.0 > 1.96$.

20. Across several specifications with different sensible controls, a key coefficient stays close to the same value. This stability is evidence that:

(A) the sample is large enough for reliable inference.

(B) the coefficient is not absorbing what the controls capture.

(C) the regression errors are homoskedastic after all.

(D) the R^2 is computed correctly in every column.

Section B: Analytical Problem (30 marks)

Using a sample of $n = 1{,}000$ workers, hourly wages (in \$) are regressed on a university-degree dummy ($College = 1$ if the worker holds a degree), first alone and then with years of work experience added. Robust standard errors in parentheses:

$$\textbf{Regression 1:} \quad \widehat{Wage} = 11.00 + \underset{(0.50)}{2.00}\, College$$

$$\textbf{Regression 2:} \quad \widehat{Wage} = 8.00 + \underset{(0.40)}{2.50}\, College + \underset{(0.04)}{0.20}\, Exper$$

In this sample, degree holders have *fewer* years of experience on average (they started work later), and experience raises wages.

(a) **[6 marks]** In Regression 1, interpret the two coefficients, test whether the college gap differs from zero at the 5% level, and give its 95% confidence interval.

(b) **[6 marks]** In Regression 2, interpret the coefficient 2.50 precisely, and compute its t-statistic.

(c) **[6 marks]** The college coefficient *rose* from 2.00 to 2.50 when experience was added. Using the omitted-variable-bias sign rule and the facts stated above, explain exactly why Regression 1 understated the college gap.

(d) **[6 marks]** An economist claims each year of experience is worth \$0.25 per hour. Test $H_0 : \beta_{Exper} = 0.25$ at the 5% level using Regression 2.

(e) **[6 marks]** The software reports a robust F-statistic of 24.0 for the joint null $H_0 : \beta_{College} = 0$ and $\beta_{Exper} = 0$. State the conclusion at the 1% level, and explain in one sentence what this joint test asks that the two t-tests do not.

Section C: Interpretation (20 marks)

Read the following and answer in **no more than five sentences**. Use the concepts from the course; you do not need any calculation.

Two research teams study whether holding cash reserves protects firms' stock returns during market crises, using the same data on 2,000 firms. Team A regresses crisis returns on cash holdings alone: the coefficient is large, positive and highly significant, and they report it as the effect of cash, noting their R^2 is high for this kind of data. Team B first lists the factors that plausibly drive crisis returns and also move with cash holdings (firm size, debt levels, industry), includes them as controls, and reports the cash coefficient across several such specifications: it is smaller than Team A's, but similar in every column. The two teams ask you to judge whose estimate of the effect of cash is more credible.

Question [20 marks]. Which team's estimate deserves more trust, and why? Your answer should say what is wrong with Team A's reasoning, what Team B did right, and what role the R^2 and the stability across specifications each play. Answer in at most five sentences.