Mock Exam Questions (Midterm)

Instructions

This midterm contains 20 MC questions and 5 open questions, and will count for 40% of the final grade. All MC questions have equal weight (1pt). All open questions will have a weight of 3 pts each. The duration of this exam is 2 hours.

Multiple Choice Questions

Question 1

A financial analyst builds a model to forecast next month's stock price for a specific company using historical price data, trading volume, and market sentiment indicators. The analyst's primary objective is to make the forecast as accurate as possible. Which of the following best describes the main goal of this econometric exercise?

a) Causal inference, focusing on the estimated coefficient ($\hat{\beta}$) for trading volume.
b) Prediction, focusing on the accuracy of the predicted value ($\hat{y}$).
c) Evaluating the goodness-of-fit using the Sum of Squared Residuals (SSR).
d) Testing the hypothesis that market sentiment has a non-zero effect on the stock price.

Question 2

A researcher is studying the economic impact of a new public transit line. They collect data on income, employment status, and commute times for the *same* 5,000 individuals every year, starting two years before the line opened and continuing for five years after. What type of data does this represent?

a) Cross-sectional data
b) Time series data
c) Pooled cross-sections
d) Panel (or longitudinal) data

Question 3

A researcher estimates the following simple linear regression model using Ordinary Least Squares (OLS): $Wage_i = \hat{\beta}_0 + \hat{\beta}_1 Educ_i + \hat{u}_i$

Based on the first-order conditions used to derive the OLS estimators, which of the following statements about the residuals ($\hat{u}_i$) and the explanatory variable ($Educ_i$) is mathematically guaranteed to be true in the sample?

a) The sample correlation between the residuals and the dependent variable ($Wage_i$) is zero.
b) The sum of the squared residuals is equal to the total sum of squares ($\sum \hat{u}_i^2 = SST$).
c) The sample covariance between the residuals and the explanatory variable ($Educ_i$) is zero.
d) The residuals ($\hat{u}_i$) are all equal to zero.

Question 4

A marketing analyst estimates the relationship between monthly product sales and advertising spending using a log-log model:

$\widehat{\log(\text{Sales})} = 3.8 + 0.45 \times \log(\text{AdSpend})$

Where `Sales` is the number of units sold and `AdSpend` is the advertising expenditure in euros. How should the coefficient on log(AdSpend) (0.45) be interpreted?

a) A €1 increase in advertising spending is associated with a 0.45% increase in sales.
b) A 1% increase in advertising spending is associated with a 0.45% increase in sales.
c) A 1% increase in advertising spending is associated with an increase in sales of 0.45 units.
d) A €1 increase in advertising spending is associated with an increase in sales of 0.45 units.

Question 5

Consider the regression of individual hourly wage on a single dummy variable, `Female`, which equals 1 for females and 0 for males. The estimated equation is: $\widehat{\text{Wage}} = 22.50 - 3.75 \times \text{Female}$

Based on this output, what is the predicted average hourly wage for males in the sample?

a) €26.25
b) €22.50
c) €18.75
d) €3.75

Question 6

An economist is investigating the factors that determine CEO salary. The simple regression of salary on firm profits yields an R-squared of 0.09. Which of the following is the correct interpretation of this R-squared value?

a) There is a 9% chance that the relationship between profits and salary is causal.
b) For every €1 increase in profits, CEO salary increases by 9%.
c) 9% of the variation in CEO salary in the sample can be explained by the variation in firm profits.
d) The estimated coefficient for profits is biased by 9%.

Question 7

A researcher wants to estimate the causal effect of class size on student test scores. They run the following regression: $\text{TestScore} = \beta_0 + \beta_1 \text{ClassSize} + u$

However, they fail to include a variable for the average "parental involvement" at the school. It is likely that schools with higher parental involvement have better resources and thus smaller class sizes (a negative correlation). It is also likely that higher parental involvement directly leads to better test scores (a positive effect). What is the likely direction of the omitted variable bias on the estimated coefficient $\hat{\beta}_1$?

a) There will be no bias because OLS is always unbiased.
b) The bias will be positive, making the negative effect of class size appear smaller than it truly is.
c) The bias will be negative, making the negative effect of class size appear larger than it truly is.
d) The direction of the bias cannot be determined without knowing the R-squared of the model.

Question 8

An analyst is trying to get a more precise estimate of the relationship between marketing expenditure and sales. The variance of the OLS slope estimator, $\widehat{Var}(\hat{\beta}_1)$, depends on several factors. Which of the following changes would *decrease* the variance of the estimated coefficient on marketing expenditure, leading to a more precise estimate?

a) A decrease in the sample size (n).
b) An increase in the variance of the error term (σ^2).
c) A decrease in the variation of marketing expenditure in the sample (SST_x).
d) An increase in the variation of marketing expenditure in the sample (SST_x).

Question 9

An economist wants to estimate the true effect of education on wages. The true population model is: $Wage = \beta_0 + \beta_1 \text{Education} + \beta_2 \text{Ability} + u$

The true parameters are known to be $\beta_1 = 2.5$ and $\beta_2 = 4.0$. However, the economist cannot observe "Ability" and estimates a simpler model: $Wage = \alpha_0 + \alpha_1 \text{Education} + v$

It is also known that in the population, the relationship between Ability and Education can be described by the linear model: $\text{Ability} = \delta_0 + 0.5 \times \text{Education} + w$.

Given this information, what is the expected value of the OLS estimator $\hat{\alpha}_1$ that the economist will obtain from the simpler model? (Hint: Start with the formula for Omitted Variable Bias).

a) 2.5
b) 2.0
c) 4.5
d) 6.5

Question 10

A researcher estimates a model for an individual's hourly wage based on their industry sector, using the "Manufacturing" sector as the omitted reference category. The estimated model is:

$\widehat{\text{Wage}} = 18.50 - 2.10 \times \text{Retail} + 5.40 \times \text{Finance}$

Where `Retail` and `Finance` are dummy variables equal to 1 if the individual works in that sector, and 0 otherwise.

Based on this model, what is the estimated average wage difference between a worker in the Finance sector and a worker in the Retail sector?

a) €5.40
b) €3.30
c) €7.50
d) €2.10

Question 11

In a multiple linear regression model, $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$, what is the correct interpretation of the coefficient β_1?

a) The total effect of a one-unit change in x_1 on y.
b) The correlation between x_1 and y.
c) The effect of a one-unit change in x_1 on y, holding x_2 constant.
d) The average value of y when x_1 is equal to 1.

Question 12

A researcher is estimating the effect of education on wages and is concerned about a person's innate ability. In this context, innate ability is likely a classic example of what?

a) A collider variable, which will induce a spurious correlation if controlled for.
b) A mediating variable, which should not be controlled for if one wants to measure the total effect.
c) A confounder, and failing to control for it will likely lead to omitted variable bias.
d) An irrelevant variable that will only decrease the model's R-squared.

Question 13

An analyst wants to test the hypothesis that *none* of the independent variables in their model have any explanatory power. The null hypothesis is $H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$. Which test is appropriate for this purpose?

a) A series of individual t-tests for each coefficient.
b) An F-test for the overall significance of the regression.
c) A test for heteroskedasticity.
d) A test for the R-squared value.

Question 14

What is the primary consequence of high (but not perfect) multicollinearity between two independent variables, x_1 and x_2, in a regression model?

a) The coefficient estimates for β_1 and β_2 will be biased.
b) The R-squared of the model will be artificially low.
c) The standard errors for the coefficients $\widehat{\beta}_1$ and $\widehat{\beta}_2$ will be large.
d) The intercept of the model, $\widehat{\beta}_0$, will be biased, but other coefficients will be unaffected.

Question 15

Consider the regression model: $Wage_i = \beta_0 + \beta_1 Female_i + \beta_2 Educ_i + \beta_3 (Female_i \cdot Educ_i) + u_i$. What does the coefficient β_3 represent?

a) The average wage for females with zero years of education.
b) The effect of an additional year of education on wages for females.
c) The wage difference between females and males for those with zero years of education.
d) The difference in the return to an additional year of education for females compared to males.

Question 16

An AR(1) model is defined as $Y_t = \alpha + \rho Y_{t-1} + u_t$. What happens if the autoregressive coefficient $|\rho|$ is equal to 1?

A) The series is stationary, and shocks die out quickly.
B) The series is explosive, and its value grows infinitely.
C) The series is non-stationary and is called a random walk, where shocks have a permanent effect.
D) The model becomes a white noise process, as past values have no influence.

Question 17

In an Autoregressive Distributed Lag ARDL(p,q) model, how is the long-run multiplier for an explanatory variable X interpreted?

A) It is the immediate impact on Y from a one-unit change in X.
B) It is the total long-run change in Y after a permanent one-unit increase in X has fully worked through the system.
C) It represents the sum of the autoregressive coefficients, measuring the persistence of the Y variable.
D) It is the average of all short-run coefficients in the model.

Question 18

Consider forecasting with a Moving Average MA(q) model: $Y_t = \mu + u_t + \theta_1 u_{t-1} + \cdots + \theta_q u_{t-q}$. What is the forecast for a time period far in the future (i.e., for a forecast horizon $h > q$)?

A) The forecast will converge to zero.
B) The forecast will continue the most recent trend.
C) The forecast will revert to the mean of the process, μ.

D) The forecast is impossible to calculate because future error terms are unknown.

Question 19

The "Within" estimator, used in Fixed Effects models, transforms the data by:

a) Subtracting the value of each variable in the previous period from the current period's value for each individual.
b) Adding a separate dummy variable for each individual in the regression to capture their unique intercept.
c) Calculating the time-average of each variable for every individual and then subtracting this average from each of their observations.
d) Using only the variation between different individuals at a single point in time to estimate the coefficients.

Question 20

A major limitation of the Fixed Effects (FE) model is that it cannot estimate the coefficients of certain variables. Which of the following variables would an FE model be unable to estimate?

a) A firm's annual innovation expenditure.
b) An individual's age.
c) The industry sector a firm belongs to, assuming it does not change during the observation period.
d) The national unemployment rate, which varies over time.

Open Questions

Question 1

This question uses a dataset which contains data on wages and other characteristics for 526 working individuals. The key variables are: wage: average hourly earnings in dollars, `educ`: years of education, `exper`: years of potential work experience. Use the provided output to answer the questions in full sentences.

An economist runs a simple linear regression to model an individual's hourly wage based on their years of education. The output from the regression is provided below.

a) Write down the estimated Sample Regression Function (SRF) equation.

b) Interpret the estimated coefficient for the educ variable. What does this number tell you?

Table 1: Regression of Wage on Years of Education

	(1)
(Intercept)	−0.905
	(0.685)
educ	0.541***
	(0.053)
Num.Obs.	526
R2	0.165

* p <0.1, ** p <0.05, *** p <0.01

c) Interpret the estimated intercept ((Intercept)). Does the intercept have a meaningful real-world interpretation in this specific context? Explain why or why not.

Question 2

The simple model from Question 1 (`wage` on `educ`) is likely biased because it omits other important factors that determine wages. A major omitted variable is a worker's experience (`exper`).

Assume the "true" population model that determines wages is:

$Wage = \beta_0 + \beta_1 \text{Educ} + \beta_2 \text{Exper} + u.$

a) What is the likely sign of β_2, the true coefficient on exper? In other words, what is the direct effect of experience on wages, holding education constant? Explain your reasoning.

b) In the real world, what is the likely correlation between years of education and years of work experience?

c) Using your answers from (a) and (b), determine the direction of the bias on the educ coefficient in the simple model. Will the simple model likely overestimate or underestimate the true effect of education on wages?

Question 3

Consider the following multiple linear regression model designed to understand the relationship between wages, gender, and education:

$$Wage_i = \beta_0 + \beta_1 Female_i + \beta_2 Educ_i + \beta_3 (Female_i \cdot Educ_i) + u_i$$

Where:

- $Wage_i$ is the hourly wage for individual i.
- $Female_i$ is a dummy variable equal to 1 if individual i is female, and 0 if male.
- $Educ_i$ is the number of years of education for individual i.

a) Write down the specific regression equation for males ($Female_i = 0$) and the specific regression equation for females ($Female_i = 1$).

b) Derive the expression for the marginal effect of an additional year of education on wages ($\frac{\partial E[Wage]}{\partial Educ}$) for females. How does this differ from the marginal effect for males?

c) Derive an expression for the wage gap between a female and a male who have the same level of education ($E[Wage|Female = 1, Educ] - E[Wage|Female = 0, Educ]$). Explain how and why this wage gap depends on the level of education.

Question 4

An analyst at a central bank is using a simple AR(2) model to forecast the quarterly inflation rate. After estimating the model using data up to the end of 2024 (Quarter 4), they have the following equation:

$\text{Inflation}_t = 0.2 + 1.2 \cdot \text{Inflation}_{t-1} - 0.4 \cdot \text{Inflation}_{t-2} + u_t$

The known inflation rates for the last two quarters of 2024 are:

- $Inflation_{2024,Q3} = 2.5\%$
- $Inflation_{2024,Q4} = 3.0\%$

a) Using the iterative forecasting method for AR models described in the lecture, calculate the **one-step-ahead forecast** for the first quarter of 2025 (`INFLATION_{2025, Q1}`) and the **two-step-ahead forecast** for the second quarter of 2025 (`INFLATION_{2025, Q2}`). You must show your calculations for both steps.

b) After the first half of 2025, the actual inflation rates were recorded as `3.5%` for Q1 and `3.8%` for Q2. Calculate the **Forecast Error** for each of your two predictions and then compute the **Root Mean Squared Error (RMSE)** for the overall two-period forecast. What does this RMSE value tell you about the typical magnitude of your model's prediction errors?

Question 5

An economist is using a panel dataset to investigate the impact of the number of training hours (`Training`) on the productivity (`Productivity`) of employees across various companies over several years. The economist is concerned that unobserved, time-invariant employee characteristics, such as innate ability or motivation, might bias the results of a simple OLS regression. To address this, they are considering either a Fixed Effects (FE) model.

a) Explain precisely why unobserved characteristics like "innate ability" pose a statistical problem in this context. Define this issue using the term "unobserved heterogeneity" and explain how it could lead to biased estimates if not properly addressed.

b) Describe how both the Fixed Effects (Within) estimator mechanically remove this unobserved heterogeneity from the regression model. What specific data transformation does this method apply?