TOR VERGATA
UNIVERSITÀ DEGLI STUDI DI ROMA

Quantitative Methods III - Mock Exam 3

Exercise 1

A venture capital firm is studying the relationship between the **Marketing Spend** (X, in thousands of Euros) and the **Monthly Active Users** (Y, in thousands) for 5 tech startups.

Startup	Marketing (X)	Users (Y)
1	5	12
2	8	15
3	10	22
4	12	24
5	15	32

(a) Estimate the OLS regression line where Users (Y) is a function of Marketing (X).

(b) Interpret the meaning of the slope coefficient $\hat{\beta}_1$ in the context of the problem.

(c) Calculate a Goodness-of-fit coefficient for this model.

(d) Calculate the Standard Error of the Regression (SER).

(e) Given that the standard error of the slope is $SE(\hat{\beta}_1) = 0.25$, the venture capital firm believes that an increase of 1,000 Euros in marketing spending leads to an increase of 1,000 users. Based on the available information, discuss whether this belief is supported by the data at the 5% significance level.

Exercise 2

You are modeling the **Price of Apartments** (Y) based on **Square Footage** (X_1) and **Distance from City Center** (X_2). A sample of $N = 150$ observations yields the following estimated equation:

$$\hat{Y}_i = 50 + 2.5X_{1i} - 15.0X_{2i}$$

With standard errors: $SE(\hat{\beta}_0) = 10.2, SE(\hat{\beta}_1) = 0.4, SE(\hat{\beta}_2) = 6.5$.

(a) Predict the price of an apartment that is 100 m^2 and 2 km from the center.

(b) Determine if "Distance from City Center" (X_2) is statistically significant at the 5% level using a t-test.

(c) Explain the difference between R^2 and Adjusted $\bar{R}^2$. Why is $\bar{R}^2$ preferred in multiple regression?

(d) Suppose you add a third regressor, "Number of Bedrooms" (X_3). If the R^2 increases but the $\bar{R}^2$ decreases, what does this suggest about the new variable?

Exercise 3

Consider an $ADL(1,1)$ model for inflation (π_t) based on its own lag and the lag of unemployment (U_t):

$$\pi_t = \beta_0 + \beta_1 \pi_{t-1} + \delta_1 U_{t-1} + u_t$$

Estimated results ($T = 100$):

$$\hat{\pi}_t = 0.02 + 0.85\pi_{t-1} - 0.15U_{t-1}$$

Current data for the latest period (T): $\pi_T = 3.0\%$, $U_T = 5.0\%$.

(a) Forecast the inflation rate for period $T + 1$.

(b) If the actual inflation at $T + 1$ turns out to be 2.2%, calculate the forecast error.

(c) Perform a Dickey-Fuller test for a unit root in the inflation series using the estimated coefficient $\hat{\beta}_1 = 0.85$ and $SE(\hat{\beta}_1) = 0.06$. Use a 5% significance level (Critical Value $= -2.89$).

(d) Calculate the BIC for this model if $\ln(RSS/T) = -4.5$.

Multiple Choice Questions

1. In a simple OLS regression, suppose the estimated slope coefficient is positive and statistically significant. Which of the following statements is correct?

   (A) The relationship between X and Y is necessarily causal.
   (B) An increase in X is associated with an increase in Y on average.
   (C) R^2 must be equal to 1.
   (D) The residuals are equal to zero.

2. Consider adding a new explanatory variable to a multiple regression model. Which of the following is always true?

   (A) R^2 increases or stays the same.
   (B) Adjusted R^2 always increases.
   (C) The new variable is statistically significant.
   (D) The SER decreases.

3. Suppose the t-statistic for a coefficient is 1.5 in absolute value, and the critical value at the 5% level is 1.96. What is the correct conclusion?

   (A) Reject the null hypothesis.
   (B) Fail to reject the null hypothesis.
   (C) The coefficient is equal to zero.
   (D) The model is not valid.

4. In a time series model, the estimated coefficient on the lagged dependent variable is very close to 1 and not statistically different from 1. What does this suggest?

   (A) The series is likely stationary.
   (B) The series may contain a unit root.
   (C) The model has perfect fit.
   (D) The residuals are zero.

5. Consider out-of-sample forecasting. If forecast errors increase substantially over time, what is the most likely interpretation?

(A) The model has a higher R^2.

(B) The model suffers from structural instability.

(C) The sample size is too large.

(D) The regressors are perfectly correlated.

Solution 1

(1.a) **Parameters:** First compute sample means:

$$\bar{X} = \frac{5+8+10+12+15}{5} = \frac{50}{5} = 10, \quad \bar{Y} = \frac{12+15+22+24+32}{5} = \frac{105}{5} = 21.$$

The slope estimator is:

$$\hat{\beta}_1 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}.$$

Compute deviations and products:

- $(5-10)(12-21) = (-5)(-9) = 45$
- $(8-10)(15-21) = (-2)(-6) = 12$
- $(10-10)(22-21) = (0)(1) = 0$
- $(12-10)(24-21) = (2)(3) = 6$
- $(15-10)(32-21) = (5)(11) = 55$

Numerator:

$$\sum(X_i - \bar{X})(Y_i - \bar{Y}) = 45 + 12 + 0 + 6 + 55 = 118.$$

Denominator:

$$\sum(X_i - \bar{X})^2 = 25 + 4 + 0 + 4 + 25 = 58.$$

Therefore:

$$\hat{\beta}_1 = \frac{118}{58} = 2.034.$$

Intercept:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X} = 21 - (2.034 \cdot 10) = 21 - 20.34 = 0.66.$$

Model: $\hat{Y} = 0.66 + 2.034X$.

The positive slope suggests a strong positive relationship between marketing and users.

(1.b) The slope coefficient measures the marginal effect of marketing on users.

For every additional 1,000 spent on marketing, the expected number of users increases by about 2.034 thousand users.

(1.c) R^2**:**

Total Sum of Squares:

$$TSS = \sum(Y_i - \bar{Y})^2 = (-9)^2 + (-6)^2 + (1)^2 + (3)^2 + (11)^2 = 81 + 36 + 1 + 9 + 121 = 248.$$

Explained Sum of Squares:

$$ESS = \hat{\beta}_1^2 \sum(X_i - \bar{X})^2 = (2.034)^2 \cdot 58 = 240.03.$$

$$R^2 = \frac{ESS}{TSS} = \frac{240.03}{248} = 0.968.$$

The model explains about 96.8% of the variation in users, indicating an excellent fit.

(1.d) SER:

First compute residual sum of squares:

$$SSR = TSS - ESS = 248 - 240.03 = 7.97.$$

Degrees of freedom: $n - 2 = 5 - 2 = 3$.

$$SER = \sqrt{\frac{SSR}{n-2}} = \sqrt{\frac{7.97}{3}} = \sqrt{2.657} = 1.63.$$

On average, predictions deviate from actual values by about 1.63 thousand users.

(1.e) The firm believes that $\beta_1 = 1$.

Test statistic:

$$t = \frac{\hat{\beta}_1 - 1}{SE(\hat{\beta}_1)} = \frac{2.034 - 1}{0.25} = 4.136.$$

Compare with critical value:

$$|t| = 4.136 > 3.182.$$

The estimated effect is significantly different from 1 at the 5% level.

The data suggest that marketing is more effective than the firm believes.

Solution 2

(a) Prediction:

Substitute values:

$$\hat{Y} = 50 + 2.5(100) - 15(2).$$

$$\hat{Y} = 50 + 250 - 30 = 270.$$

Larger apartments increase price, while distance reduces it.

(b) Significance test:

$$t = \frac{\hat{\beta}_2}{SE(\hat{\beta}_2)} = \frac{-15}{6.5} = -2.307.$$

Compare with critical value (two-sided, 5%):

$$|t| = 2.307 > 1.96.$$

Conclusion: distance is statistically significant.

(c) R^2 vs $\bar{R}^2$:

- R^2 measures the proportion of explained variance and never decreases when adding regressors.
- $\bar{R}^2$ penalizes model complexity by adjusting for the number of regressors.

Thus, $\bar{R}^2$ is preferred because it discourages overfitting.

(d) Deduction:

If R^2 increases but $\bar{R}^2$ decreases:

- the new variable adds little explanatory power
- the increase in fit is not sufficient to justify the added complexity

Conclusion: X_3 is likely irrelevant or weakly informative.

Solution 3

(a) Forecast:

Substitute values:

$$\hat{\pi}_{T+1} = 0.02 + 0.85(3.0) - 0.15(5.0).$$

Step by step:

$$0.85 \cdot 3 = 2.55, \quad 0.15 \cdot 5 = 0.75.$$

$$\hat{\pi}_{T+1} = 0.02 + 2.55 - 0.75 = 1.82\%.$$

(b) Forecast error:

$$e_{T+1} = Y_{T+1} - \hat{Y}_{T+1} = 2.2 - 1.82 = 0.38.$$

(c) Dickey-Fuller test:

Null hypothesis: $\beta_1 = 1$ (unit root).

$$t = \frac{0.85 - 1}{0.06} = \frac{-0.15}{0.06} = -2.5.$$

Compare with critical value:

$$-2.5 > -2.89.$$

Conclusion: we fail to reject the null. Inflation is likely non-stationary (high persistence).

(d) BIC:

Number of parameters: intercept + 2 regressors $\Rightarrow k = 3$.

$$BIC = \ln(RSS/T) + \frac{k \ln T}{T}.$$

Substitute values:

$$BIC = -4.5 + \frac{3\ln(100)}{100}.$$

Since $\ln(100) = 4.605$:

$$BIC = -4.5 + \frac{3 \cdot 4.605}{100} = -4.5 + 0.138 = -4.362.$$

Solutions to Multiple Choice

1. **Correct answer: (B)**
   A positive and significant coefficient implies a positive association, not causality.

2. **Correct answer: (A)**
   R^2 never decreases when adding regressors, while Adjusted R^2 may decrease.

3. **Correct answer: (B)**
   Since $|t| = 1.5 < 1.96$, we fail to reject the null hypothesis.

4. **Correct answer: (B)**
   A coefficient close to 1 suggests high persistence and possible unit root (non-stationarity).

5. **Correct answer: (B)**
   Increasing forecast errors indicate a deterioration in predictive performance, consistent with structural breaks.