ALLAMA IQBAL OPEN UNIVERSITY

AIOU LIBRARY
aioulibrary.com

Name:

Registration No.:

Course Code: 807

Program: M.Sc

Semester: Autumn 2026

WhatsApp Orders: +92 324 5836626

ALLAMA IQBAL OPEN UNIVERSITY, ISLAMABAD

Course: Basic Econometrics (807) • Level: M.Sc

Semester: Autumn 2026

Assignment No. 1 (Units 1–4)

Total Marks: 100 • Pass Mark: 40 • Credit Hours: 03

SOLVED ASSIGNMENT GUIDE

(For Study & Reference)

AIOU LIBRARY aioulibrary.com

Note: This guide is prepared for study and reference purposes. Students are advised to read the relevant units of the course book, understand the concepts, and write the assignment in their own words. Do not copy it verbatim.

Question 1: The Gauss–Markov Theorem and the Properties of the OLS Estimators (20 Marks)

What does the Gauss-Markov Theorem tell us about the properties of the OLS estimators? What are the assumptions under which this theorem holds?

Introduction

Ordinary least squares (OLS) is the workhorse estimation method of econometrics. Given a sample of observations, OLS chooses the values of the regression coefficients that make the sum of the squared residuals as small as possible. A natural question follows: among all the many estimators that could be constructed from the same sample — weighted averages of the dependent variable, estimators based on medians, estimators that use only part of the sample,

WhatsApp Orders: +92 324 5836626

and so on — is there any reason to prefer the OLS estimator? The Gauss–Markov theorem, named after Carl Friedrich Gauss and Andrey Markov, provides the classic answer. It states the precise conditions under which OLS is the best estimator in a clearly defined class, and it therefore supplies the theoretical justification for the central place of OLS in regression analysis. This answer states the theorem, explains each word of the phrase Best Linear Unbiased Estimator (BLUE), lists and discusses the assumptions on which the result rests, sketches why the result holds, and examines what happens when each assumption fails.

Statement of the Theorem

Consider the classical linear regression model written in matrix form as $Y = X\beta + u$, where Y is an $n \times 1$ vector of observations on the dependent variable, X is an $n \times k$ matrix of observations on the $k - 1$ explanatory variables plus a column of ones for the intercept, β is the $k \times 1$ vector of unknown parameters, and u is the $n \times 1$ vector of unobservable disturbance terms. The OLS estimator is $b = (X'X)^{-1}X'Y$. The Gauss–Markov theorem states: if the assumptions of the classical linear regression model listed below hold, then the OLS estimator b is the Best Linear Unbiased Estimator of β. In other words, within the class of all estimators that are linear in Y and unbiased for β, no estimator has a smaller variance than OLS. Equivalently, OLS is the minimum-variance linear unbiased estimator, component by component: each individual coefficient estimator b_j has the smallest possible variance among linear unbiased estimators of β_j.

What BLUE Means, Word by Word

- Linear: the estimator is a linear function of the dependent variable, that is, it can be written as a weighted sum (or average) of the observed values of Y with weights that depend on X but not on Y. OLS satisfies this because $b = (X'X)^{-1}X'Y = AY$ with $A = (X'X)^{-1}X'$.

- Unbiased: on average, over repeated samples, the estimator hits the true parameter. Formally E(b) = β, so there is no systematic tendency to overstate or understate the parameter. Unbiasedness is a property of the sampling distribution of the estimator, not of any single estimate obtained from one sample.
- Best: 'best' means most efficient within the stated class — the estimator has the minimum variance. A smaller variance means the sampling distribution is more tightly concentrated around the true value, so a single estimate is more likely to be close to the truth. Precision, not merely correctness on average, is what 'best' rewards.
- Estimator: the rule or formula, as distinct from the estimate, which is the numerical value the rule produces for a particular sample.

The Assumptions Under Which the Theorem Holds

The theorem is a conditional result: it is exactly as strong as its assumptions. The standard assumptions of the classical linear regression model are as follows.

- A1 — Correct linear specification: the model is linear in the parameters and correctly specified, so that the relationship can be written Y = Xβ + u with no relevant variable omitted and no incorrect functional form imposed. If the true relationship is different, OLS is estimating the parameters of the wrong model, and unbiasedness for the parameters of interest is lost.
- A2 — X is non-stochastic, or independent of the disturbance: in the classical derivation the explanatory variables are treated as fixed in repeated samples; in the modern version, X may be random but must be uncorrelated with (indeed, exogenous to) u, so that E(u | X) =

0. This is the crucial zero conditional mean assumption, without which OLS is generally biased and inconsistent.

- A3 — Zero mean of the disturbances: $E(u_i) = 0$ for every observation. This is usually guaranteed by including an intercept in the model, which absorbs any constant non-zero mean into the constant term.
- A4 — Homoscedasticity: the variance of each disturbance is the same constant σ^2, $Var(u_i) = \sigma^2$ for all i. The disturbances are therefore equally dispersed at every level of the explanatory variables. Its violation is heteroscedasticity.
- A5 — No autocorrelation: the disturbances for different observations are uncorrelated, $Cov(u_i, u_j) = 0$ for $i \neq j$. Assumptions A4 and A5 together are the spherical-disturbance assumptions, written $Var(u) = \sigma^2 I$. They are what make OLS not merely unbiased but minimum variance.
- A6 — No exact multicollinearity and full rank: no explanatory variable is an exact linear combination of the others, so X has full column rank k and (X'X) can be inverted. Without this assumption the OLS estimator does not even exist uniquely. It is also required that the number of observations exceeds the number of parameters, $n > k$.

Two things about this list deserve emphasis. First, normality of the disturbances is NOT a Gauss–Markov assumption. Normality is needed for exact small-sample t and F inference, not for OLS to be BLUE. Second, the assumptions divide by function: A1–A3 deliver unbiasedness, while A4–A5 deliver efficiency (the 'best' part). This division explains the consequences of violations discussed below.

Why the Theorem Holds: A Sketch of the Proof

Write any rival linear estimator as b* = CY where C = A + D for some matrix D, with A the OLS weight matrix. Because b* must be unbiased, E(b*) = CXβ = β for all β, which forces CX = I and hence DX = 0. The variance of b* then becomes Var(b*) = σ^2CC′ = σ^2[(A + D)(A + D)′] = σ^2[AA′ + DD′], because the cross terms vanish precisely when DX = 0 (since AX′ terms reduce via $(X'X)^{-1}X'X$ = I). Therefore Var(b*) = Var(b) + σ^2DD′, and σ^2DD′ is a positive semi-definite matrix: the rival estimator's variance exceeds (or equals) the OLS variance by a non-negative amount. Equality holds only when D = 0, that is, when the rival is OLS itself. This elegant argument shows that any deviation from the OLS weights, while preserving unbiasedness, can only add variance.

An Important By-Product: Estimating σ^2

The theorem's framework also yields an unbiased estimator of the disturbance variance. If e denotes the OLS residuals and RSS = e′e the residual sum of squares, then s^2 = RSS / (n − k) is an unbiased estimator of σ^2, E(s^2) = σ^2. The divisor is the degrees of freedom, n minus the number of estimated parameters, rather than n, because k degrees of freedom were used in fitting the regression. This estimator supplies the standard errors of the coefficients, se(b_j) = s$\sqrt{[(X'X)^{-1}_{jj}]}$, on which confidence intervals and hypothesis tests are built. Note the distinction: the Gauss–Markov theorem is about the efficiency of b itself; s^2 is the companion result that makes practical inference possible.

What Happens When the Assumptions Fail

- Heteroscedasticity (A4 fails): OLS remains unbiased and consistent but is no longer efficient; other estimators (generalised least squares, GLS) can do better, and — more seriously in practice — the usual OLS standard errors are biased, so t-tests and confidence intervals

become unreliable. Heteroscedasticity-robust (White) standard errors or weighted least squares are the standard remedies.

- Autocorrelation (A5 fails): common in time-series data; again OLS stays unbiased but loses efficiency and its conventional standard errors are wrong, typically understating the true uncertainty. Cochrane–Orcutt, Prais–Winsten or HAC (Newey–West) standard errors are used instead.
- Endogeneity (A2 fails): if an explanatory variable is correlated with the disturbance — through omitted variables, simultaneity or measurement error — OLS is biased and inconsistent, and no standard-error correction repairs it. Instrumental variables or two-stage least squares are required. This is the most damaging violation.
- Imperfect multicollinearity: not strictly a violation at all (only exact collinearity breaks A6), but high correlation among regressors inflates the variances of the estimators, so OLS, while still BLUE, may be very imprecise — BLUE is a relative, not an absolute, guarantee of precision.

Conclusion

The Gauss–Markov theorem tells us that, under the classical assumptions — correct linear specification, exogenous regressors, zero-mean disturbances, constant disturbance variance, no autocorrelation and full rank of X — the OLS estimators are BLUE: among all linear unbiased estimators they have minimum variance, so no information is wasted. The theorem does not claim that OLS beats every conceivable estimator: biased estimators or non-linear estimators can sometimes have smaller mean squared error, and when the spherical-disturbance assumptions fail, GLS-type estimators are more efficient. Its enduring value is that it identifies

exactly which assumptions buy which property — exogeneity buys unbiasedness, spherical disturbances buy efficiency — and thereby provides the checklist against which every applied regression must be diagnosed.

Question 2: Confidence Intervals for Regression — β_1, β_2 and σ^2 (20 Marks)

Explain confidence intervals for regression. Find out confidence intervals for regression coefficients β1, β2 and σ2.

Introduction

A point estimate, such as the OLS value $b_2 = 1.8$, reports a single number as our best guess of the true slope coefficient. But because the estimate comes from one sample among the many that might have been drawn, it will almost certainly differ from the true parameter β_2 by some sampling error. Sound econometric practice therefore supplements every point estimate with an interval estimate: a range of values, constructed from the sample, within which the true parameter is believed to lie with a stated probability or confidence level. This answer explains the idea of a confidence interval in regression, derives the intervals for the intercept β_1 and the slope β_2 of the two-variable regression model, and then derives the interval for the disturbance variance σ^2, which is based on a different sampling distribution. A fully worked numerical illustration shows how the formulas are used and interpreted.

The Two-Variable Model and Its Sampling Distributions

Work with the population regression function $Y_i = \beta_1 + \beta_2 X_i + u_i$, estimated by OLS as $\hat{Y}_i = b_1 + b_2 X_i$. Under the classical assumptions including normally distributed disturbances, $u_i \sim N(0, \sigma^2)$, the OLS estimators are themselves normally distributed: $b_1 \sim N(\beta_1, \sigma^2 \cdot \Sigma X_i^2/(n\Sigma x_i^2))$ and $b_2 \sim N(\beta_2,$

$\sigma^2/\Sigma x_i^2$), where $x_i = X_i - \bar{X}$ denotes deviations from the mean. Standardising gives a standard normal variable, but σ^2 is unknown, so σ^2 is replaced by its unbiased estimator $s^2 = \Sigma e_i^2/(n - 2)$. This substitution converts the standard normal into Student's t distribution with $(n - 2)$ degrees of freedom: $(b_2 - \beta_2)/se(b_2) \sim t(n - 2)$, where $se(b_2) = s/\sqrt{(\Sigma x_i^2)}$ is the standard error of b_2. The t distribution, with its heavier tails, honestly reflects the extra uncertainty introduced by estimating σ^2. It is this pivotal quantity — 'pivotal' because its distribution does not depend on unknown parameters — that is rearranged to produce the confidence interval.

Constructing the General Confidence Interval

Fix a significance level α, commonly 0.05, giving a $100(1 - \alpha)\%$ interval, commonly 95%. From the t distribution, $P[-t(\alpha/2, n-2) \le (b_j - \beta_j)/se(b_j) \le +t(\alpha/2, n-2)] = 1 - \alpha$, where $t(\alpha/2, n-2)$ is the critical value leaving $\alpha/2$ in each tail. Multiplying through by the standard error and rearranging the inequality to isolate β_j in the middle yields the universal recipe:

$$\textbf{100(1 – α)\% CI for } \beta_j = b_j \pm t(\alpha/2, df) \times se(b_j)$$

The logic is identical for the intercept and the slope; only the standard error differs. The interval is therefore 'estimate plus-or-minus a margin of error', where the margin of error combines the critical value (which grows as the desired confidence grows or the degrees of freedom shrink) with the standard error (which shrinks as the sample grows or the spread of X increases). Figure 1 shows the underlying idea: the interval is built so that the standardised estimator falls in the central $100(1 - \alpha)\%$ of its distribution.

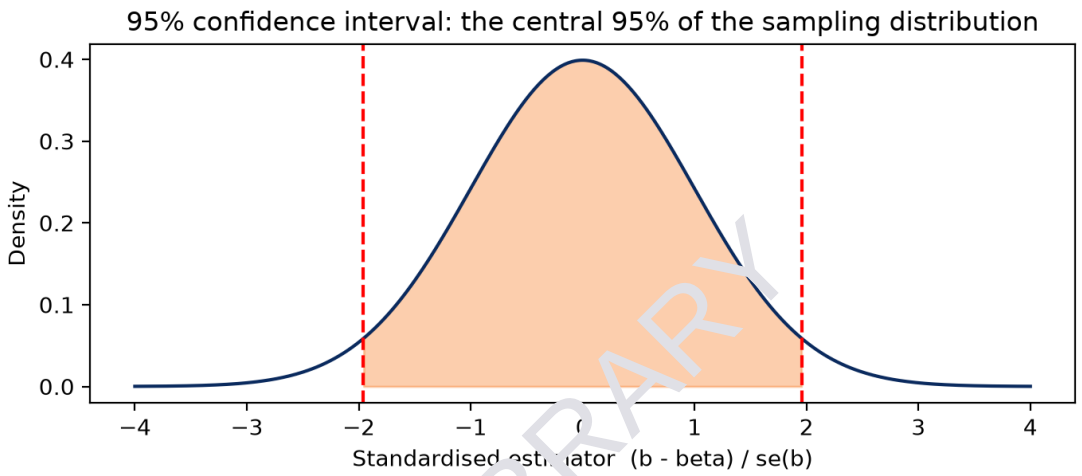


Figure 1: The central 95% of the sampling distribution; the dashed lines mark the critical values ±1.96 (large-sample values; the t critical value is used in small samples).

AIOU LIBRARY aioulibrary.com

Confidence Interval for the Slope β_2

Applying the recipe with $se(b_2) = s/\sqrt{(\Sigma x_i^2)}$:

$$\text{CI for } \beta_2: \; b_2 - t(\alpha/2, n-2)\cdot[s/\sqrt{(\Sigma x_i^2)}] \;\le\; \beta_2 \;\le\; b_2 + t(\alpha/2, n-2)\cdot[s/\sqrt{(\Sigma x_i^2)}]$$

The slope is usually the coefficient of chief economic interest — a marginal propensity to consume, an elasticity, a return to schooling — and its standard error is small when the sample is large, when the residual variance s^2 is small, and when X varies widely in the sample (large Σx_i^2). A designed sample with a good spread of X values therefore buys precision directly, a point of real practical importance.

Confidence Interval for the Intercept β_1

Applying the same recipe with $se(b_1) = s\cdot\sqrt{[\Sigma X_i^2/(n\Sigma x_i^2)]}$:

$$\text{CI for } \beta_1: \; b_1 - t(\alpha/2, n-2)\cdot s\sqrt{[\Sigma X_i^2/(n\Sigma x_i^2)]} \;\le\; \beta_1 \;\le\; b_1 + t(\alpha/2, n-2)\cdot s\sqrt{[\Sigma X_i^2/(n\Sigma x_i^2)]}$$

The intercept's standard error depends on the distance of the X values from zero as well as on their spread: if the data are centred far from the origin, the fitted line is being extrapolated back a long way to X = 0 and the intercept is estimated imprecisely. This is why the intercept is often of less substantive interest and why centring variables can help interpretation. Table 1 collects the three intervals side by side.

Table 1: Confidence intervals in the two variable regression model

Parameter	Point estimator	Standard error / pivot	100(1−α)% interval
β_2 (slope)	b_2	$se(b_2) = s/\sqrt{\Sigma x_i^2}$; t pivot	$b_2 \pm t(\alpha/2, n-2)\cdot se(b_2)$
β_1 (intercept)	b_1	$se(b_1) = s\sqrt{[\Sigma X_i^2/(n\Sigma x_i^2)]}$; t pivot	$b_1 \pm t(\alpha/2, n-2)\cdot se(b_1)$
σ^2 (variance)	$s^2 = \Sigma e_i^2/(n-2)$	$(n-2)s^2/\sigma^2 \sim \chi^2(n-2)$	$(n-2)s^2/\chi^2(\alpha/2) \le \sigma^2 \le (n-2)s^2/\chi^2(1-\alpha/2)$

Confidence Interval for σ^2

The variance σ^2 is not estimated by a linear estimator with a normal sampling distribution, so the t recipe does not apply. Instead, under normality of the disturbances, the quantity $(n-2)s^2/\sigma^2$ follows a chi-square distribution with $(n-2)$ degrees of freedom. Hence $P[\chi^2(1-\alpha/2) \le (n-2)s^2/\sigma^2 \le \chi^2(\alpha/2)] = 1-\alpha$, where $\chi^2(\alpha/2)$ is the upper-tail critical value. Inverting the inequalities — carefully reversing their direction, because σ^2 appears in the denominator — gives:

$$\textbf{CI for } \sigma^2 : \ (n-2)s^2 / \chi^2(\alpha/2, n-2) \ \le \ \sigma^2 \ \le \ (n-2)s^2 / \chi^2(1-\alpha/2, n-2)$$

Two features distinguish this interval. First, it is not symmetric around s^2, because the chi-square distribution is skewed, especially with few degrees of freedom; the interval reaches further above s^2 than below it. Second, taking square roots of both endpoints yields the corresponding interval for the standard deviation σ (the standard error of the regression). The σ^2 interval matters because the precision of everything else in the regression depends on how large the unexplained variation is.

A Fully Worked Numerical Illustration

Suppose a consumption function estimated from n = 10 observations gives $b_1 = 30$, $b_2 = 0.80$, $se(b_1) = 12$, $se(b_2) = 0.10$ and $\Sigma e_i^2 = 800$, so $s^2 = 800/(10 - 2) = 100$ and $s = 10$. Degrees of freedom are 8, and the 5% two-tailed t critical value is $t(0.025, 8) = 2.306$. For the slope: $0.80 \pm 2.306 \times 0.10 = 0.80 \pm 0.231$, giving [0.569, 1.031]. For the intercept: $30 \pm 2.306 \times 12 = 30 \pm 27.67$, giving [2.33, 57.67]. For σ^2, the chi-square critical values with 8 degrees of freedom are $\chi^2(0.025) = 17.53$ and $\chi^2(0.975) = 2.18$, so the interval is [800/17.53, 800/2.18] = [45.6, 367.0], and for σ it is [6.76, 19.16]. Note how wide the σ^2 interval is with only 8 degrees of freedom — variance parameters are estimated much less precisely than slope parameters in small samples. (These figures are an illustrative textbook-style example constructed to demonstrate the calculations.)

Interpretation and Common Misinterpretations

- Frequentist meaning: β_2 is fixed and the interval is random. The correct statement is that if the sampling and construction were repeated many times, 95% of the intervals so constructed would contain the true β_2. Saying 'there is a 95% probability that β_2 lies in this

computed interval' is, strictly, not the frequentist claim, though it is the interpretation most users intuitively want.

- Confidence is not free: a 99% interval is wider than a 95% interval from the same data. Raising confidence without more data is paid for in reduced precision; the only way to gain both is a larger sample or smaller residual variance.
- Connection with hypothesis testing: a two-sided test of H_0: $\beta_j = \beta^*$ at level α rejects H_0 exactly when the value β^* lies outside the $100(1-\alpha)\%$ confidence interval. In the illustration, H_0: $\beta_2 = 0$ is rejected at 5% because 0 is outside [0.569, 1.031]; indeed the whole interval lying above zero is stronger, more informative evidence than the bare rejection, because it also shows the plausible size of the effect.
- Width diagnostics: a very wide interval is the data honestly reporting ignorance — typically from a small sample, noisy data, or little variation in X. Reporting the interval rather than only the point estimate protects the reader from false precision.

Conclusion

Confidence intervals convert OLS point estimates into honest statements about sampling uncertainty. For β_1 and β_2 the interval is the estimate plus-or-minus a t critical value times the coefficient's standard error; for σ^2 the interval is built by inverting the chi-square distribution of $(n - 2)s^2/\sigma^2$ and is asymmetric. Together they answer the three questions every regression reader should ask: how large is the effect, how precisely is it known, and how much variation remains unexplained.

Question 3: OLS Estimation and Hypothesis Testing in the Model $Y = \beta_1 + \beta_2X_2 + \beta_3X_3 + \mu$, n = 25 (20 Marks)

Consider the regression model Y = β1 + β2X2 + β3X3 + μ. The sample size of 25 was taken for the estimation of the model. (a) Explain the OLS method for estimation of the parameters. (b) How to test the hypothesis under the assumption of β2 = β3 = 0. (c) Explain how to test the significance of β2 and β3 separately.

Introduction

This question moves from the two-variable model to multiple regression, in which the dependent variable Y depends on two explanatory variables, X_2 and X_3, plus an intercept and a disturbance term μ. The sample size is n = 25 and the number of parameters is k = 3, so the degrees of freedom available for inference are n − k = 22 — a number that will recur in every test below. Part (a) derives the OLS estimators, part (b) sets out the joint test of both slope coefficients using the F-test (equivalently the analysis-of-variance / R^2 form), and part (c) sets out the separate t-tests for each slope. The contrast between the parts is itself instructive: joint and individual tests answer different questions and can legitimately give different verdicts, especially when X_2 and X_3 are correlated with each other.

Part (a): The OLS Method of Estimation

Write the model for each observation i = 1, …, 25 as $Y_i = \beta_1 + \beta_2X_{2i} + \beta_3X_{3i} + \mu_i$, and the fitted (sample) regression as $\hat{Y}_i = b_1 + b_2X_{2i} + b_3X_{3i}$, with residual $e_i = Y_i - \hat{Y}_i$. The OLS principle chooses b_1, b_2, b_3 to minimise the residual sum of squares:

$$\textbf{Minimise } S = \Sigma e_i^2 = \Sigma (Y_i - b_1 - b_2X_{2i} - b_3X_{3i})^2$$

Differentiating S with respect to each estimator, setting the three first-order conditions to zero, and simplifying gives the three normal equations:

$$\Sigma Y = nb_1 + b_2\Sigma X_2 + b_3\Sigma X_3$$

$$\Sigma X_2 Y = b_1\Sigma X_2 + b_2\Sigma X_2^2 + b_3\Sigma X_2 X_3$$

$$\Sigma X_3 Y = b_1\Sigma X_3 + b_2\Sigma X_2 X_3 + b_3\Sigma X_3^2$$

Solving this simultaneous system — most compactly in matrix form, $b = (X'X)^{-1}X'Y$, where X is the 25 × 3 data matrix whose first column is ones — yields the OLS estimators. In deviation form, writing lower-case letters for deviations from sample means ($y = Y - \bar{Y}$, etc.), the slope estimators have the revealing closed forms:

$$b_2 = [\,(\Sigma x_3^2)(\Sigma x_2 y) - (\Sigma x_2 x_3)(\Sigma x_3 y)\,] \,/\, [\,(\Sigma x_2^2)(\Sigma x_3^2) - (\Sigma x_2 x_3)^2\,]$$

$$b_3 = [\,(\Sigma x_2^2)(\Sigma x_3 y) - (\Sigma x_2 x_3)(\Sigma x_2 y)\,] \,/\, [\,(\Sigma x_2^2)(\Sigma x_3^2) - (\Sigma x_2 x_3)^2\,], \quad b_1 = \bar{Y} - b_2\bar{X}_2 - b_3\bar{X}_3$$

These formulas display three essential properties of multiple regression. First, each slope is a partial effect: b_2 measures the effect of X_2 on Y holding X_3 constant, which is achieved algebraically by purging from both Y and X_2 the part explained by X_3 — the Frisch–Waugh–Lovell result that b_2 equals the slope from regressing Y-residuals (after removing X_3) on X_2-residuals (after removing X_3). Second, the denominator $(\Sigma x_2^2)(\Sigma x_3^2) - (\Sigma x_2 x_3)^2$ shrinks as X_2 and X_3 become more correlated, which is exactly why multicollinearity inflates variances; if the two regressors were perfectly collinear the denominator would be zero and the parameters could not be estimated at all. Third, the fitted line passes through the sample means, and the residuals satisfy $\Sigma e_i = 0$, $\Sigma e_i X_{2i} = 0$ and $\Sigma e_i X_{3i} = 0$, the sample counterparts of the moment conditions behind the normal equations.

Estimation is completed by computing the disturbance variance and the standard errors. The unbiased estimator is $s^2 = \Sigma e_i^2/(n - k) = \Sigma e_i^2/22$. The variances of the slopes are $Var(b_2) = \sigma^2\Sigma x_3^2/[(\Sigma x_2^2)(\Sigma x_3^2) - (\Sigma x_2x_3)^2]$ and symmetrically for b_3, estimated by replacing σ^2 with s^2; their square roots are the standard errors used in part (c). By the Gauss–Markov theorem these OLS estimators are BLUE under the classical assumptions, and under normality of μ they are also maximum likelihood estimators.

Part (b): Joint Test of H_0: $\beta_2 = \beta_3 = 0$ — The F-Test

The joint hypothesis H_0: $\beta_2 = \beta_3 = 0$ states that neither explanatory variable has any effect on Y — the regression explains nothing beyond the sample mean. The alternative H_1 is that at least one of β_2, β_3 differs from zero. Because two restrictions are tested simultaneously, the appropriate instrument is the F-test, based on comparing the fit of the unrestricted model above with the restricted model $Y_i = \beta_1 + \mu_i$ implied by H_0, under which $\hat{Y}_i = \bar{Y}$ for every observation.

Compute the total sum of squares $TSS = \Sigma(Y_i - \bar{Y})^2$, the explained sum of squares $ESS = \Sigma(\hat{Y}_i - \bar{Y})^2$ and the residual sum of squares $RSS = \Sigma e_i^2$, with TSS = ESS + RSS. Under H_0, the statistic

$$\mathbf{F = [ESS / (k - 1)] / [RSS / (n - k)] = [ESS / 2] / [RSS / 22] \sim F(2, 22)}$$

follows the F distribution with 2 numerator degrees of freedom (the number of restrictions, equal to the number of slope coefficients tested) and 22 denominator degrees of freedom. Equivalently, using $R^2 = ESS/TSS$, $F = [R^2/(k-1)] / [(1-R^2)/(n-k)] = [R^2/2]/[(1-R^2)/22]$, which shows that the overall F-test is simply a test of whether R^2 is significantly different from zero.

Decision rule: at the 5% level the critical value is F(0.05; 2, 22) ≈ 3.44 (at 1% it is ≈ 5.72). If the computed F exceeds the critical value — equivalently, if the p-value (the 'Prob > F' or significance of F reported by software) is below 0.05 — reject H_0 and conclude that X_2 and X_3 jointly have a statistically significant effect on Y. If the computed F is below the critical value, H_0 is not rejected: the model as a whole has no demonstrated explanatory power. A worked check makes the mechanics transparent: if the regression yielded $R^2 = 0.40$, then $F = (0.40/2)/[(0.60)/22] = 0.20/0.02727 \approx 7.33$, which exceeds 3.44, so H_0 would be rejected at 5%. (This R^2 value is illustrative only, since the question supplies no data set.) The ANOVA table below summarises the test.

Table 1: ANOVA framework for the joint F-test (n = 25, k = 3)

Source	Sum of squares	Degrees of freedom	Mean square
Regression (explained)	ESS	k − 1 = 2	ESS / 2
Residual (unexplained)	RSS	n − k = 22	RSS / 22 = s^2
Total	TSS = ESS + RSS	n − 1 = 24	—

Part (c): Testing β_2 and β_3 Separately — The t-Tests

To test each coefficient on its own, state, for j = 2, 3, the null H_0: $\beta_j = 0$ against the two-sided alternative H_1: $\beta_j \neq 0$ (a one-sided alternative is used only when theory firmly predicts the sign in advance). Under H_0 and the classical assumptions including normality, the statistic

$$t = (b_j - 0) / se(b_j) \sim t(n - k) = t(22)$$

follows Student's t distribution with 22 degrees of freedom. The standard errors come from part (a): $se(b_2) = \sqrt{[s^2 \Sigma x_3^2 / ((\Sigma x_2^2)(\Sigma x_3^2) - (\Sigma x_2 x_3)^2)]}$ and analogously for $se(b_3)$, each carrying the multicollinearity penalty through the same denominator. The critical values for t(22) are 2.074 at the 5% two-tailed level and 2.819 at the 1% level. Decision rule: if |t| computed exceeds the critical value (or the p-value is below α), reject H_0: the variable X_j makes a statistically significant separate contribution to explaining Y, over and above the other regressor already in the model. If |t| is smaller, H_0 is not rejected for that coefficient — the data do not establish a separate effect for that variable, at that significance level, holding the other variable constant.

Equivalently, one may check whether the confidence interval $b_j \pm 2.074 \cdot se(b_j)$ contains zero: it contains zero exactly when the 5% t-test fails to reject. The word 'separately' must be taken seriously: each t-test is conditional on the other variable being present. Two practical consequences follow. First, the F-test of part (b) can reject while neither t-test rejects — the classic symptom of multicollinearity, where X_2 and X_3 jointly matter but their individual contributions cannot be disentangled in a sample of 25. Second, a variable insignificant alone may become significant jointly with others, and mechanically deleting variables one t-test at a time can distort the model; joint F-tests of groups of variables are the safer model-building tool.

Conclusion

For $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \mu$ with n = 25: (a) OLS minimises Σe_i^2, yielding the normal equations and the estimators $b = (X'X)^{-1}X'Y$, interpretable as partial effects holding the other regressor constant, with $s^2 = RSS/22$; (b) the joint hypothesis $\beta_2 = \beta_3 = 0$ is tested by $F = [ESS/2]/[RSS/22]$ against F(2, 22), critical value about 3.44 at 5%; and (c) each coefficient is tested separately by t

= $b_j/se(b_j)$ against t(22), critical value 2.074 at 5%. Joint significance and individual significance are different questions, and a competent report presents both.

Question 4: Simultaneous Equation Systems and Identification (20 Marks)

What is a simultaneous equation system? How would you identify it?

Introduction

Single-equation regression rests on a one-way direction of causation: the explanatory variables determine the dependent variable, and nothing in the equation feeds back onto them. Many economic relationships refuse to behave so politely. Price and quantity in a market are determined together by supply and demand; income and consumption determine each other in Keynesian macroeconomics; wages and prices chase each other in a wage–price spiral. Where such two-way (or multi-way) causation exists, the appropriate framework is a system of simultaneous equations, and OLS applied equation by equation is biased and inconsistent because regressors are correlated with disturbances. Before such a system can be estimated, a prior logical question must be answered for each equation: can its parameters be recovered uniquely from the observable data at all? That is the problem of identification, the subject of the second half of this answer.

What Is a Simultaneous Equation System?

A simultaneous equation system (or simultaneous equations model) is a set of two or more interrelated equations in which a variable that is explained in one equation appears as an explanatory variable in another, so that the endogenous variables are determined jointly, simultaneously, by the system as a whole. Its building blocks are: (i) structural equations, which

describe behaviour, technology or institutions as derived from economic theory (a demand function, a supply function, an identity); (ii) endogenous variables, whose values are determined within the system (as many as there are equations, G in number); (iii) predetermined variables, comprising exogenous variables determined outside the system and lagged endogenous variables, whose values are given when the system operates (K in number); and (iv) stochastic disturbance terms in each behavioural equation, representing omitted influences.

The classic example is the market model: demand $Q^d = \alpha_0 + \alpha_1 P + \alpha_2 Y + u_1$ ($\alpha_1 < 0$), supply $Q^s = \beta_0 + \beta_1 P + \beta_2 R + u_2$ ($\beta_1 > 0$), and the equilibrium identity $Q^d = Q^s = Q$, where P and Q are endogenous, income Y and rainfall R are exogenous. Observed price–quantity pairs are intersections of the two curves; a scatter of them traces out neither curve. Solving the structural system for the endogenous variables in terms of the predetermined variables alone gives the reduced form, $P = \pi_1 Y + \pi_2 R + v_1$ and $Q = \pi_3 Y + \pi_4 R + v_2$, whose parameters are estimable by OLS because the right-hand side variables are predetermined. The identification question is whether the structural parameters (α's and β's) can be worked back uniquely from the reduced-form parameters (π's). If they cannot, no estimation method, however clever, can recover them from data generated by this system — estimation would be attempting the logically impossible.

Consequences of Ignoring Simultaneity

In the market model, a shock to demand (u_1) shifts the demand curve and changes the equilibrium price, so P is correlated with u_1 in the demand equation. The exogeneity assumption fails, OLS is biased and inconsistent, and this simultaneity bias does not vanish in

large samples. Haavelmo's pioneering work in the 1940s established both the problem and the system methods — indirect least squares (ILS), instrumental variables, two-stage least squares (2SLS), limited and full information maximum likelihood — developed to solve it. But all of them presuppose that the equation in question is identified.

The Identification Problem Stated Precisely

An equation in a simultaneous system is: (1) exactly (just) identified if its structural parameters can be recovered uniquely from the reduced form — one solution only; (2) over-identified if they can be recovered in more than one way — the extra restrictions are testable implications of the model; and (3) under-identified (unidentified) if they cannot be recovered at all — infinitely many structural parameter sets generate the same reduced form, so the data cannot distinguish the hypothesised equation from mixtures of the other equations in the system. The intuitive root is the distinction between the equation of interest and linear combinations of equations: if some combination of the other equations looks exactly like the equation under study (same variables present), the two are observationally indistinguishable and identification fails. Identification is achieved through restrictions, most commonly exclusion restrictions — variables present in the system but excluded from the given equation, which shift the other equations while leaving this one in place (in the market example, income shifts only demand and rainfall shifts only supply, which is what allows each curve to be traced out).

The Order Condition (Necessary Condition)

Let G = number of endogenous variables in the system, K = number of predetermined variables in the system, and for the equation under consideration let g = number of endogenous variables included in it and k = number of predetermined variables included in it. The order

condition states that an equation is identified only if the number of predetermined variables excluded from it is at least the number of endogenous variables included minus one:

$$K - k \geq g - 1$$

If $K - k > g - 1$ the equation is over-identified; if $K - k = g - 1$ it may be exactly identified (subject to the rank condition); if $K - k < g - 1$ it is under-identified and nothing further need be checked. Applied to the market model (G = 2, K = 2, namely Y and R): the demand equation includes both endogenous variables (g = 2) and one predetermined variable, Y (k = 1), so $K - k = 1 = g - 1$ — exactly identified, thanks to the excluded rainfall variable. The supply equation likewise includes R and excludes Y, so it too is exactly identified. Delete rainfall from supply and demand would lose its identifier: $K - k = 0 < 1$, and the demand curve could never be traced out — the familiar textbook result that a demand equation is identified only when some supply shifter is excluded from demand.

AIOU LIBRARY aioulibrary.com

The Rank Condition (Necessary and Sufficient)

The order condition counts restrictions but cannot detect every failure, so sufficiency requires the rank condition: an equation is identified if and only if it is possible to construct at least one non-zero determinant of order $(G - 1)$ from the coefficients, in the other equations of the system, of the variables excluded from the given equation. In matrix terms, form the matrix of coefficients of the excluded variables as they appear in the remaining equations; the equation is identified exactly when that matrix has rank $G - 1$. Applied to the market model: for the demand equation, the excluded variable is R; its coefficient in the supply equation is $\beta_2 \neq 0$, giving a 1×1 matrix of rank $1 = G - 1$, so demand is identified. The procedure for any equation is therefore: (1) list the variables it excludes; (2) write down their coefficients throughout the

rest of the system; (3) check whether a non-vanishing (G − 1)-order determinant exists. If yes, the rank condition is satisfied; combined with the order count, the equation's status (exact or over) is settled. Table 1 summarises the classification and the estimation method each status permits.

Table 1: Identification status, conditions and suitable estimators

Status	Order condition	Meaning	Estimation
Under-identified	K − k < g − 1	Parameters cannot be recovered at all	No method can estimate the equation; respecify the model
Exactly identified	K − k = g − 1 and rank holds	Unique recovery from reduced form	Indirect least squares (ILS); 2SLS equivalent
Over-identified	K − k > g − 1 and rank holds	More than one recovery route; restrictions testable	Two-stage least squares (2SLS), LIML; over-identification tests (Sargan/Hansen)

A Practical Checklist for Identifying an Equation

- Step 1: Write the complete structural system; count G and K and classify every variable as endogenous or predetermined.
- Step 2: For the equation in question, count g and k and apply the order condition. If it fails, stop — the equation is under-identified.

- Step 3: If the order condition holds, build the excluded-variables coefficient matrix from the other equations and apply the rank condition.
- Step 4: Classify the equation (exact or over-identified) and choose the estimator accordingly — ILS for exact, 2SLS as the general workhorse.
- Step 5: Where over-identified, test the over-identifying restrictions; rejection warns that some exclusion restriction may be invalid.

Conclusion

A simultaneous equation system is a set of interdependent structural equations in which endogenous variables are determined jointly, so that regressors are generally correlated with disturbances and OLS fails. Identification asks whether an equation's parameters can be uniquely recovered from the reduced form. The order condition ($K - k \geq g - 1$) is the quick counting test; the rank condition — existence of a non-zero $(G - 1)$-order determinant from the excluded variables' coefficients in the other equations — is necessary and sufficient. Only after an equation passes both tests does estimation by ILS or 2SLS become meaningful.

Question 5: Short Notes — (a) Endogenous and Exogenous Variables; (b) Methodology of Econometrics (20 Marks)

Write a short note on the following: (a) Endogenous and Exogenous variables (b) Methodology of Econometrics

Introduction

This question asks for short notes on two foundational topics. The first — the distinction between endogenous and exogenous variables — determines which variables a model must explain and which it may take as given, and on it hangs the validity of OLS itself. The second —

the methodology of econometrics — describes the ordered steps by which economic theory is confronted with data to produce tested, usable quantitative knowledge. Each note is developed in full below.

Part (a): Endogenous and Exogenous Variables

Definitions

An endogenous variable is a variable whose value is determined within the model or system under study; it is the variable (or one of the variables) the model exists to explain. An exogenous variable is a variable whose value is determined outside the model and is taken as given — as data — when the model operates; the model uses it to explain other variables but does not attempt to explain it. The names come from the Greek endo- (within) and exo- (outside). In the single regression $Y_i = \beta_1 + \beta_2 X_i + u_i$, Y is endogenous (the regressand, dependent variable) and X is treated as exogenous (the regressor, independent variable). The distinction is not a property of the variable in nature but of the model: income is endogenous in a Keynesian income-determination model yet exogenous in a single household consumption function; the same rain that is exogenous to a crop-supply model is endogenous within a meteorological model.

Why the Distinction Matters: Exogeneity and OLS

The entire classical regression apparatus rests on the regressors being exogenous in the precise sense that they are uncorrelated with the disturbance term, $E(u \mid X) = 0$. When a regressor is in fact endogenous — correlated with u — OLS becomes biased and inconsistent. Endogeneity arises from three main sources. (1) Simultaneity: the regressor is itself determined by the dependent variable, as price is in a demand equation (see Question 4). (2) Omitted variables: an

influence left in the disturbance is correlated with an included regressor, as innate ability (in u) is correlated with schooling in an earnings equation. (3) Measurement error: errors in measuring the regressor create a correlation between the observed regressor and the composite disturbance. The standard cure is instrumental-variables estimation: find an instrument that is relevant (correlated with the endogenous regressor) and exogenous (uncorrelated with u), and estimate by IV or 2SLS. In simultaneous systems the classification extends to predetermined variables — exogenous variables plus lagged endogenous variables — which, being already determined when the current period begins, can legitimately appear as regressors.

Examples and Classification

Table 1. Endogenous and exogenous variables in familiar models

Model	Endogenous	Exogenous / predetermined
Keynesian income model: C = a + bY; Y = C + I	Consumption C, Income Y	Investment I (and government spending)
Market model	Price P, Quantity Q	Income (in demand), rainfall/input prices (in supply)
Single consumption function	Household consumption	Household income, prices (taken as given)
Phillips-type wage equation	Wage inflation	Lagged unemployment, expected inflation (predetermined)

Two refinements complete the note. First, econometricians distinguish weak exogeneity (sufficient for efficient estimation of the parameters of interest), strong exogeneity (weak exogeneity plus no Granger-causality from Y to X, needed for forecasting), and super exogeneity (invariance to policy changes, needed for policy analysis — the heart of the Lucas critique, which warns that 'exogenous' reduced-form relationships may shift when policy regimes change). Second, exogeneity is an assumption to be argued and, where possible, tested (for example by Durbin–Wu–Hausman tests comparing OLS and IV estimates), never a label to be attached carelessly, because wrongly treating an endogenous regressor as exogenous silently invalidates the estimates.

Part (b): Methodology of Econometrics

Meaning

Methodology means the ordered procedure of enquiry. The methodology of econometrics is the sequence of steps by which a theoretical statement about economic relationships is specified as a statistical model, estimated from data, tested, and used for forecasting and policy. Traditional (classical) methodology follows the eight steps below, associated in textbook form with Gujarati's exposition; modern practice adds an awareness of the criticisms levelled by Hendry's general-to-specific approach and Leamer's warnings about specification searching, discussed at the end.

The Steps of Traditional Econometric Methodology

- Step 1 — Statement of theory or hypothesis: begin from economic theory, for example Keynes's proposition that consumption rises with income by less than the rise in income, $0 < MPC < 1$.

- Step 2 — Specification of the mathematical model: express the theory as an equation, $C = \beta_1 + \beta_2 Y$, with the expected signs and magnitudes of the parameters stated in advance.
- Step 3 — Specification of the econometric model: add the disturbance term, $C = \beta_1 + \beta_2 Y + u$, acknowledging that the relationship is inexact because of omitted influences, measurement error and randomness in behaviour. This step also fixes the functional form and the stochastic assumptions.
- Step 4 — Obtaining the data: collect suitable time-series, cross-section or panel data of the right quality, frequency and coverage; the finest method cannot rescue bad data.
- Step 5 — Estimation of the parameters: choose an appropriate technique (OLS where its assumptions are defensible; IV, GLS or maximum likelihood where they are not) and compute the estimates with their standard errors.
- Step 6 — Hypothesis testing: ask whether the estimates conform to the theory's expectations — here, whether b_2 is significantly between 0 and 1 — using t, F and chi-square tests and confidence intervals, and run diagnostic tests for heteroscedasticity, autocorrelation, specification error and multicollinearity.
- Step 7 — Forecasting or prediction: if the model survives testing, use it to predict values of the dependent variable for given future or out-of-sample values of the regressors, reporting prediction intervals as well as point forecasts.
- Step 8 — Use for control or policy: apply the estimated model to policy questions — how much government spending is needed to reach a target income, what tax change achieves a desired revenue — remembering that policy use demands the strongest (super-exogeneity) assumptions.

AIOU LIBRARY aioulibrary.com

Criticisms and Modern Refinements

The traditional sequence proceeds theory-first and specific: it risks testing a model that was itself chosen after peeking at the data, and its test statistics lose their nominal meaning if many specifications were tried (Leamer's critique of 'specification searches'). Hendry's general-to-specific methodology reverses the starting point: begin with a general model consistent with theory and the data's properties, then simplify by testing restrictions, checking diagnostics at every reduction, so the final parsimonious model encompasses its rivals. Contemporary practice combines both spirits — theory to structure the question, data to discipline the specification, diagnostics throughout, and out-of-sample validation before forecasting or policy use. What all schools share is the conviction that an econometric result is only as credible as the transparency of the steps that produced it.

AIOU LIBRARY aioulibrary.com

Conclusion

Endogenous variables are explained within the model and exogenous variables are taken as given from outside it; keeping the classification honest is what licenses OLS, and its failure is the source of simultaneity and endogeneity bias. The methodology of econometrics — theory, mathematical model, econometric model, data, estimation, testing, forecasting and policy use — is the disciplined path by which those variables are turned into tested quantitative knowledge rather than mere correlation.