

VU LEARNING HUB

vulearninghub.xyz

HIGHLIGHTED HANDOUT

ECO609

Advanced Econometrics

ECO609 by VU Learning HUB

Prepared for VU students • Free study resources

<https://vulearninghub.xyz>

ECO609

Advanced Econometrics

HANDOUT

Advanced econometrics develops the statistical theory behind economic measurement: the classical linear regression model and its assumptions, estimation and inference, violations (multicollinearity, heteroskedasticity, autocorrelation) and their remedies, dummy variables, and introductions to time-series and panel-data methods.

Table of Contents

Module 1	Nature of Econometrics and Statistical Foundations
Module 2	The Classical Linear Regression Model (CLRM)
Module 3	Inference in the Linear Regression Model
Module 4	Multicollinearity
Module 5	Heteroskedasticity
Module 6	Autocorrelation
Module 7	Dummy Variables and Qualitative Regressors
Module 8	Model Specification and Diagnostic Testing
Module 9	Introduction to Time Series Econometrics
Module 10	Introduction to Panel Data Methods

MODULE 1: NATURE OF ECONOMETRICS AND STATISTICAL FOUNDATIONS

Learning Objectives

- Define econometrics and its methodology
- Review probability, expectation, and distributions used in econometrics
- Explain estimators, bias, efficiency, and consistency
- Distinguish population model from sample regression

What is Econometrics?

Econometrics is the application of statistical methods to economic data to test theories, estimate relationships, and forecast. The classic methodology (Gujarati): (1) state the economic theory/hypothesis; (2) specify the mathematical model; (3) specify the **econometric (statistical) model** by adding a stochastic error term; (4) obtain data; (5) estimate parameters; (6) test hypotheses; (7) forecast/policy analysis; (8) use the model for control or policy.

Data types: **cross-sectional** (many units, one time period), **time series** (one unit, many periods), and **panel/longitudinal** (many units over time). Each type brings characteristic econometric problems — this course treats them systematically.

Probability Review

A **random variable** X has expected value $E(X) = \sum x \cdot P(x)$ (discrete) and variance $\text{Var}(X) = E[(X - E(X))^2] = E(X^2) - [E(X)]^2$. Key rules: $E(aX + b) = aE(X) + b$; $\text{Var}(aX + b) = a^2 \text{Var}(X)$; for independent X, Y : $E(XY) = E(X)E(Y)$. **Covariance** $\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$ measures linear co-movement; **correlation** $\rho = \text{Cov}(X, Y) / (\sigma_X \cdot \sigma_Y)$ standardizes it to $[-1, 1]$.

Essential distributions: **normal** (many estimators are asymptotically normal), **chi-square** (sums of squared standard normals; used for variance tests), **t** (normal divided by $\sqrt{\text{chi-square}/df}$; used for coefficient tests in small samples), and **F** (ratio of chi-squares; used for joint tests). As $n \rightarrow \infty$, $t \rightarrow \text{normal}$ and $F \rightarrow \text{chi-square}/df$.

Properties of Estimators

An **estimator** is a rule for computing an estimate from data. Desirable properties: **unbiasedness** $E(\hat{\beta}) = \beta$; **efficiency** — smallest variance among unbiased estimators (BLUE = best linear unbiased estimator); **consistency** — $\text{plim } \hat{\beta} = \beta$ (converges to truth as $n \rightarrow \infty$); and **sufficiency/minimum MSE** considerations. The **Gauss-Markov theorem** (covered next module) establishes OLS as BLUE under the classical assumptions. Hypothesis testing uses the logic: assume H_0 , compute a test statistic, reject H_0 if it falls in the critical region ($p\text{-value} < \alpha$).

Key Terms & Definitions

Econometric model	Mathematical economic model plus a stochastic disturbance term.
--------------------------	---

BLUE	Best linear unbiased estimator; minimum variance among linear unbiased estimators.
Consistency	Estimator converges in probability to the true parameter as sample size grows.
p-value	Probability of a test statistic as extreme as observed, assuming H_0 is true.

Example: From Theory to Econometric Model

Theory: consumption depends on income. Mathematical model: $C = \beta_1 + \beta_2 \cdot Y$ ($0 < \beta_2 < 1$, the MPC). Econometric model: $C_i = \beta_1 + \beta_2 \cdot Y_i + u_i$, where u_i captures all other influences and randomness. Estimation on household data might yield $\hat{C} = 50 + 0.75 \cdot Y$: autonomous consumption 50, MPC 0.75 — then test $H_0: \beta_2 = 0$.

Module Summary

- Econometrics = economic theory + statistics + data; methodology runs from theory to policy use.
- Data: cross-sectional, time series, panel — each with characteristic problems.
- Core probability: expectation, variance, covariance, correlation; normal, chi-square, t, F distributions.
- Estimator properties: unbiasedness, efficiency (BLUE), consistency; tests via p-values.

Exam Pointers (Very Important)

- ★ State the eight steps of traditional econometric methodology.
- ★ Prove $\text{Var}(aX + b) = a^2 \cdot \text{Var}(X)$.
- ★ Define BLUE and state which estimator achieves it under classical assumptions.
- ★ When is the t-distribution used instead of the normal?

MODULE 2: THE CLASSICAL LINEAR REGRESSION MODEL (CLRM)

Learning Objectives

- Specify the simple and multiple regression models
- Derive the OLS estimators
- State the seven classical assumptions
- Interpret R-squared and adjusted R-squared

Model Specification and OLS Derivation

The **simple regression model**: $Y_i = \beta_1 + \beta_2 \cdot X_i + u_i$, $i = 1..n$. The **multiple regression model**: $Y_i = \beta_1 + \beta_2 \cdot X_{2i} + \dots + \beta_k \cdot X_{ki} + u_i$, or in matrix form $y = X\beta + u$. **OLS** chooses β to minimize the sum of squared residuals: $\min \sum (Y_i - \beta_1 - \beta_2 \cdot X_i)^2$. The first-order conditions (normal equations) give the famous estimators: $\beta_2 = \frac{\sum (x_i \cdot y_i)}{\sum (x_i^2)}$ and $\beta_1 = \bar{y} - \beta_2 \cdot \bar{x}$, where lowercase denotes deviations from means. In matrix form: $\beta = (X'X)^{-1}X'y$.

OLS is a purely algebraic fitting rule — its statistical properties come entirely from the assumptions placed on the model.

The Classical Assumptions

The seven CLRM assumptions: (1) the regression model is **linear in parameters**; (2) **X values are fixed** in repeated sampling (or exogenous: $E(u_i|X_i) = 0$); (3) **zero mean** of disturbances: $E(u_i|X_i) = 0$; (4) **homoskedasticity**: $\text{Var}(u_i|X_i) = \sigma^2$ (constant); (5) **no autocorrelation**: $\text{Cov}(u_i, u_j) = 0$ for $i \neq j$; (6) **no perfect multicollinearity** among regressors; (7) **zero covariance** between u_i and X_i , often strengthened to $u_i \sim N(0, \sigma^2)$ for exact finite-sample inference. Under (1)-(6), the **Gauss-Markov theorem** holds: OLS is BLUE. Adding normality gives exact t and F distributions for inference.

Goodness of Fit

$R^2 = \text{ESS}/\text{TSS} = 1 - \text{RSS}/\text{TSS}$ measures the fraction of variation in Y explained by the model ($0 \leq R^2 \leq 1$ in models with intercept). Warning: R^2 never falls when regressors are added, so **adjusted $R^2 = 1 - [\text{RSS}/(n-k)]/[\text{TSS}/(n-1)]$** penalizes extra variables and can fall. High R^2 does not imply causation, correct specification, or good forecasts — it is a descriptive summary, not a model-validity certificate.

Key Terms & Definitions

OLS	Ordinary least squares; minimizes the sum of squared residuals.
Gauss-Markov theorem	Under CLRM assumptions 1-6, OLS is BLUE.
Homoskedasticity	Constant error variance: $\text{Var}(u_i X_i) = \sigma^2$.
Adjusted R^2	Fit measure penalizing additional regressors.

Example: Computing OLS by Hand (Tiny Data)

Data: $(X,Y) = (1,2), (2,4), (3,5)$. $\sum X = 2$, $\sum Y = 3.67$. $\sum (x_i - \bar{x})(y_i - \bar{y}) = (-1)(-1.67) + 0 + (1)(1.33) = 3.0$; $\sum (x_i^2) = 2$. $\beta_2 = 3/2 = 1.5$; $\beta_1 = 3.67 - 1.5(2) = 0.67$. Fitted line: $\hat{y} = 0.67 + 1.5X$.

Module Summary

- OLS minimizes $\sum \hat{u}_i^2$; simple formulas via deviations-from-means; matrix form $\beta = (X'X)^{-1}X'y$.
- Seven classical assumptions (linearity, exogeneity, zero mean, homoskedasticity, no autocorrelation, no perfect collinearity, normality for inference).
- Gauss-Markov: OLS is BLUE under assumptions 1-6.
- R^2 measures explained variation; adjusted R^2 penalizes extra regressors.

Exam Pointers (Very Important)

- ★ Derive the OLS normal equations for the simple regression model.
- ★ State all seven CLRM assumptions precisely.
- ★ State and explain the Gauss-Markov theorem.
- ★ Why can R^2 be misleading? How does adjusted R^2 help?

MODULE 3: INFERENCE IN THE LINEAR REGRESSION MODEL

Learning Objectives

- Construct confidence intervals for coefficients
- Perform t-tests on individual coefficients
- Perform F-tests for joint hypotheses and overall significance
- Test linear restrictions with the F-statistic

Sampling Distributions

Under the CLRM with normal errors, $\beta_j \sim N(\beta_j, \sigma^2/\sum x_i^2)$ and in multiple regression $\text{Var}(\beta_j) = \sigma^2(X'X)^{-1}$. Since σ^2 is unknown, we estimate it by $\hat{\sigma}^2 = \text{RSS}/(n-k)$ (dividing by degrees of freedom gives unbiasedness). Replacing σ with $\hat{\sigma}$ gives the **t-statistic**: $t = (\hat{\beta}_j - \beta_j)/\text{se}(\hat{\beta}_j) \sim t(n-k)$. A 95% confidence interval is $\hat{\beta}_j \pm t(0.025, n-k) \cdot \text{se}(\hat{\beta}_j)$.

t-Tests and p-Values

To test $H_0: \beta_j = c$ (often $c = 0$, 'no effect'), compute $t = (\hat{\beta}_j - c)/\text{se}(\hat{\beta}_j)$ and reject H_0 at level α if $|t| > t(\alpha/2, n-k)$, equivalently if $p\text{-value} < \alpha$. **One-sided tests** ($H_1: \beta_j > 0$) use $t(\alpha, n-k)$ critical values and are appropriate when theory signs the effect — but the choice must precede seeing the data. Report **exact p-values**, not just stars: $p = 0.049$ and $p = 0.0001$ are very different evidence.

F-Tests: Joint Hypotheses

Testing several restrictions jointly (e.g., $H_0: \beta_2 = \beta_3 = 0$) needs the **F-test**: $F = [(\text{RSS}_R - \text{RSS}_U)/m] / [\text{RSS}_U/(n-k)] \sim F(m, n-k)$, where R = restricted model, U = unrestricted, m = number of restrictions. The **overall significance test** $H_0: \beta_2 = \dots = \beta_k = 0$ uses $F = [R^2/(k-1)] / [(1-R^2)/(n-k)]$. Important: individually insignificant t-statistics can jointly be highly significant (multicollinearity symptom), and the F-test properly accounts for this. The **Chow test** for structural stability is an F-test comparing pooled vs. split-sample RSS.

Key Terms & Definitions

Standard error	Estimated standard deviation of an estimator; basis of t-statistics.
p-value	Smallest α at which H_0 would be rejected; exact evidence measure.
F-test	Joint test of m linear restrictions via restricted vs. unrestricted RSS.
Chow test	F-test for parameter stability across subsamples.

Example: Testing Returns to Schooling

Wage equation: $\ln(\text{wage}) = \beta_1 + \beta_2 \cdot \text{educ} + \beta_3 \cdot \text{exper} + \beta_4 \cdot \text{exper}^2 + u$, $n = 1000$. $H_0: \beta_2 = 0$ gives $t = 0.08/0.01 = 8.0$, $p \approx 0 \rightarrow \text{reject}$; 95% CI: $0.08 \pm 1.96(0.01) = [0.06, 0.10]$: each year of schooling raises wages ~6-10%. Joint test $H_0: \beta_3 = \beta_4 = 0$ (no experience effect) via F on restricted/unrestricted RSS.

Module Summary

- $\beta_j \sim t$ -distributed around β_j ; CIs and t-tests follow directly.
- Test $H_0: \beta_j = c$ with $t = (\hat{\beta}_j - c)/\text{se}$; report exact p-values.
- Joint restrictions need the F-test: $[(\text{RSSR} - \text{RSSU})/m]/[\text{RSSU}/(n-k)]$.
- Overall significance F-test; Chow test for structural breaks.

Exam Pointers (Very Important)

- ★ Derive the t-statistic for $H_0: \beta_2 = 0$ and state its distribution.
- ★ One-sided vs. two-sided tests: when is each appropriate?
- ★ Write the F-statistic for m linear restrictions and explain each term.
- ★ Why can joint significance coexist with individual insignificance?

MODULE 4: MULTICOLLINEARITY

Learning Objectives

- Define perfect and imperfect multicollinearity
- Explain consequences for estimation and inference
- Detect multicollinearity (correlation matrix, VIF, condition number)
- Apply remedies appropriately

Nature and Consequences

Perfect multicollinearity (an exact linear relation among regressors, e.g., including both a dummy and its complement plus intercept — the **dummy variable trap**) makes $X'X$ singular: OLS cannot be computed. **Imperfect (high but not perfect) multicollinearity** is the practical problem: OLS remains BLUE and unbiased, but variances $\text{Var}(\beta_j) = \sigma^2 / [\sum x_{ji}^2 (1 - R_j^2)]$ explode as $R_j^2 \rightarrow 1$, where R_j^2 is the R^2 from regressing X_j on all other regressors. Consequences: large standard errors, insignificant t-ratios despite high R^2 and significant F, and unstable, wrong-signed estimates sensitive to small data changes.

Detection

Clues: high R^2 with few significant t-ratios; coefficients that swing wildly when a variable is added/dropped. Formal tools: the **variance inflation factor** $\text{VIF}_j = 1/(1 - R_j^2)$ — $\text{VIF} > 10$ ($R_j^2 > 0.9$) signals trouble; **tolerance** = $1/\text{VIF}$; the **condition number** of $X'X$ (ratio of largest to smallest eigenvalue; > 30 indicates strong collinearity); and **auxiliary regressions** of each X_j on the others. Pairwise correlations above ~ 0.8 are a quick screen but miss multivariate collinearity.

Remedies

Do nothing if the goal is **forecasting** (collinearity doesn't bias predictions) or if t-ratios are already significant. Otherwise: **drop** a redundant variable (risk: specification bias if it belongs); get **more data** or new data breaking the correlation; **combine** variables into an index (e.g., via principal components); use **prior/theoretical restrictions**; or apply **ridge regression** (adds a penalty $\lambda \sum \beta_j^2$, introducing small bias to slash variance — a bias-variance tradeoff). Never drop variables mechanically by t-statistic alone — theory must guide specification.

Key Terms & Definitions

VIF	Variance inflation factor = $1/(1 - R_j^2)$; > 10 indicates problematic collinearity.
Dummy variable trap	Perfect collinearity from including all category dummies plus intercept.
Ridge regression	Penalized estimator trading small bias for large variance reduction.
Auxiliary regression	Regressing each X_j on other regressors to diagnose collinearity.

Example: VIF Calculation

Regressing X2 (education) on X3 (experience) and X4 (age) gives $R^2 = 0.92$. $VIF_2 = 1/(1-0.92) = 12.5 > 10$: education's variance is inflated 12.5-fold by collinearity — its t-statistic is unreliable. Options: combine experience and age, or collect data where education varies independently of age.

Module Summary

- Perfect collinearity breaks OLS; imperfect collinearity inflates variances (OLS still BLUE).
- Symptoms: high R^2 , insignificant t's, unstable coefficients.
- Detect with VIF (>10), condition number, auxiliary regressions.
- Remedies: theory-guided respecification, more data, indices, ridge — not mechanical deletion.

Exam Pointers (Very Important)

- ★ Why does multicollinearity inflate $\text{Var}(\beta_j)$? Show the formula.
- ★ What is the dummy variable trap and how is it avoided?
- ★ Compute and interpret a VIF of 12.5.
- ★ When is doing nothing about multicollinearity justified?

MODULE 5: HETEROSKEDASTICITY

Learning Objectives

- Define heteroskedasticity and its typical causes
- Explain consequences for OLS (still unbiased, no longer BLUE/efficient)
- Detect it: White, Breusch-Pagan, Goldfeld-Quandt tests and residual plots
- Remedy it: WLS, White robust standard errors, transformations

Nature and Consequences

Heteroskedasticity: $\text{Var}(u_i|X_i) = \sigma_i^2$ — the error variance changes with X (or observation). Classic in cross-section data: in a consumption-income regression, richer households show more dispersed spending; in firm data, larger firms have more variable profits. **Consequences are crucial:** OLS remains **unbiased and consistent**, but it is **no longer BLUE** (inefficient), and — most importantly — **the usual standard errors are wrong** (biased), so t and F tests are invalid. This is the key exam point: heteroskedasticity breaks inference, not unbiasedness.

Detection

Start with a **residual plot**: plot $\hat{u}_i^2$ (or $|\hat{u}_i|$) against $\hat{u}_i$ or each X_j — a fan/cone shape signals heteroskedasticity. Formal tests: **Breusch-Pagan** — regress $\hat{u}_i^2$ on the X 's, test $nR^2 \sim \chi^2$; **White's test** — regress $\hat{u}_i^2$ on X 's, squares, and cross-products (general, no assumed form; costs degrees of freedom); **Goldfeld-Quandt** — split the sample by a suspected variable and compare subsample variances via F . The **Park** and **Glejser** tests posit specific functional forms for σ_i^2 .

Remedies

If the form of σ_i^2 is known (or well estimated), use **weighted least squares (WLS)**: divide the model by the known σ_i (transform $Y_i^* = Y_i/\sigma_i$), restoring homoskedasticity and BLUE. If the form is unknown, keep OLS but fix the standard errors with **White's heteroskedasticity-robust (robust) standard errors** — valid inference without knowing σ_i^2 (needs large samples). Alternatively, **transform** the model (e.g., $\log Y$) which often stabilizes variance, or respecify if heteroskedasticity signals omitted variables. Modern practice: report robust standard errors routinely in cross-section work.

Key Terms & Definitions

Heteroskedasticity	Non-constant error variance: $\text{Var}(u_i X_i) = \sigma_i^2$.
White test	General heteroskedasticity test via $\hat{u}_i^2$ on X 's, squares, cross-products.
WLS	Weighted least squares; BLUE when the variance structure is known.
Robust standard errors	White's heteroskedasticity-consistent SEs; valid inference without modeling σ_i^2 .

Example: Robust SEs in Practice

Wage regression shows a fan-shaped residual plot against education. OLS gives $\beta_{educ} = 0.08$, $se = 0.010$, $t = 8.0$ — but the se is invalid. White robust $se = 0.014 \rightarrow t = 5.7$, still significant but honest. If instead $\sigma_i^2 \propto X_i^2$ were established, WLS dividing by X_i would be BLUE and more efficient.

Module Summary

- Heteroskedasticity: OLS unbiased but inefficient; standard errors invalid $\rightarrow$ tests invalid.
- Detect: residual plots, Breusch-Pagan, White, Goldfeld-Quandt.
- Remedy: WLS if σ_i^2 known; White robust SEs if unknown; log transforms often help.
- Very important: it breaks inference, not unbiasedness. 

Exam Pointers (Very Important)

- ★ State the consequences of heteroskedasticity precisely — what breaks and what doesn't?
- ★ Describe White's test procedure step by step.
- ★ When do you use WLS vs. robust standard errors?
- ★ Interpret a fan-shaped residual plot.

MODULE 6: AUTOCORRELATION

Learning Objectives

- Define autocorrelation and its typical causes in time series
- Explain consequences (as with heteroskedasticity, plus more)
- Detect it: Durbin-Watson, Breusch-Godfrey, residual correlograms
- Remedy it: GLS/Cochrane-Orcutt, Newey-West SEs, dynamic specification

Nature and Consequences

Autocorrelation (serial correlation): $\text{Cov}(u_i, u_j) \neq 0$ for $i \neq j$ — errors correlated across time, typical in time-series data. Causes: **inertia** (economic time series move sluggishly), **omitted variables** whose effects persist, **misspecification** (wrong functional form shows up as patterned residuals), and data manipulation (interpolation). The workhorse model is **AR(1)**: $u_t = \rho \cdot u_{t-1} + \varepsilon_t$, $|\rho| < 1$. Consequences mirror heteroskedasticity — OLS unbiased/consistent but inefficient with **invalid standard errors** — but worse: with positive autocorrelation (the usual case), standard errors are **understated**, so t-statistics are inflated and you 'find' significance that isn't there. Also, R^2 is unreliable and forecasts are inefficient.

Detection

Plot residuals against time — cycling/wandering residuals suggest autocorrelation. The **Durbin-Watson statistic** $d = \sum (\hat{u}_t - \hat{u}_{t-1})^2 / \sum \hat{u}_t^2 \approx 2(1-\rho)$: $d \approx 2$ means no autocorrelation, $d \rightarrow 0$ strong positive, $d \rightarrow 4$ strong negative; it has inconclusive regions and is invalid with lagged dependent variables. The **Breusch-Godfrey LM test** regresses $\hat{u}_t$ on X's and lagged $\hat{u}$'s — handles higher-order autocorrelation and lagged dependent variables (use this in modern work). The **correlogram/ACF** with Q-statistics (Ljung-Box) examines many lags at once.

Remedies

If the AR structure is known/estimated: **GLS** — transform via quasi-differencing ($Y_t^* = Y_t - \rho Y_{t-1}$), estimated iteratively by **Cochrane-Orcutt** or **Hildreth-Lu** search, restoring BLUE. If the form is uncertain: keep OLS with **Newey-West HAC (heteroskedasticity- and autocorrelation-consistent) standard errors**. But first ask **why** residuals are autocorrelated — often the cure is **dynamic respecification**: add omitted variables, include lagged Y or X terms, or correct functional form. Blind GLS on a misspecified model just hides the disease.

Key Terms & Definitions

AR(1)	First-order autoregressive errors: $u_t = \rho u_{t-1} + \varepsilon_t$.
Durbin-Watson d	$\approx 2(1-\rho)$; tests first-order autocorrelation.
Breusch-Godfrey test	LM test for autocorrelation of any order; valid with lagged dependent variables.

Newey-West SEs

HAC-robust standard errors for serially correlated errors.

Example: DW in Action

Consumption function $C_t = \beta_1 + \beta_2 Y_t + u_t$, $n = 40$. DW $d = 0.85 \rightarrow \rho \approx 1 - 0.85/2 = 0.575$, strong positive autocorrelation. Below the lower critical value $\rightarrow$ reject no-autocorrelation. Remedy: Cochrane-Orcutt iteration converges to $\rho = 0.60$; GLS on quasi-differenced data gives valid inference. Also check: is an omitted wealth variable the real culprit?

Module Summary

- Autocorrelation: correlated errors over time; OLS unbiased but inefficient with understated SEs (positive ρ).
- Detect: residual time plots, Durbin-Watson ($\approx 2(1-\rho)$), Breusch-Godfrey LM, correlograms.
- Remedy: Cochrane-Orcutt/Hildreth-Lu GLS, or Newey-West HAC SEs.
- First investigate misspecification — autocorrelation often signals omitted dynamics.

Exam Pointers (Very Important)

- ★ Give four causes of autocorrelation.
- ★ Why is positive autocorrelation especially dangerous for inference?
- ★ State the DW statistic, its approximation, and interpretation.
- ★ Compare Cochrane-Orcutt GLS with Newey-West standard errors.

MODULE 7: DUMMY VARIABLES AND QUALITATIVE REGRESSORS

Learning Objectives

- Use dummy variables for intercept and slope shifts
- Avoid the dummy variable trap
- Model seasonality and structural breaks with dummies
- Interpret log-linear and interaction dummy models

Basics and the Trap

A **dummy (indicator) variable** $D = 1$ if a qualitative attribute holds, 0 otherwise. For m categories use **$m-1$ dummies** (with intercept) — including all m plus an intercept is the **dummy variable trap** (perfect multicollinearity). Example: $Y_i = \beta_1 + \beta_2 \cdot X_i + \delta \cdot D_i + u_i$ shifts the **intercept** by δ for the $D=1$ group. Adding an **interaction term** $D \cdot X_i$ allows the **slope** to differ: $Y_i = \beta_1 + \beta_2 \cdot X_i + \delta \cdot D_i + \gamma \cdot (D_i \cdot X_i) + u_i$ — test $\gamma = 0$ for a common slope.

Applications

Seasonality: quarterly data $\rightarrow$ 3 quarterly dummies (or deseasonalize). **Structural breaks**: a dummy for post-break periods, or the **Chow test** via dummies (equivalent formulations). **Policy evaluation**: difference-in-differences uses a treatment $\times$ post interaction dummy. In **log-linear models** $\ln Y = \dots + \delta D + u$, the exact percentage effect of D is $100 \cdot (e^\delta - 1)\%$, approximately $100 \cdot \delta\%$ for small δ — a favorite exam computation. 

Piecewise Linear Regression

Splines model slope changes at known thresholds (knots): e.g., $Y_i = \beta_1 + \beta_2 \cdot X_i + \gamma \cdot (X_i - X^*) \cdot D + u_i$ where $D = 1$ if $X_i > X^*$. Used for kinked relationships like tax schedules or phase transitions. **Interaction dummies** more generally let marginal effects vary by group — always interpret them as **differences in effects**, and include constituent terms to avoid forcing restrictions.

Key Terms & Definitions

Dummy variable	0/1 indicator for a qualitative attribute.
Intercept vs. slope dummy	D shifts the level; $D \cdot X$ shifts the marginal effect.
Difference-in-differences	Treatment $\times$ post interaction identifying a policy effect.
Spline	Piecewise linear model with slope change at a knot.

Example: Gender Wage Gap with Dummies

$\ln(\text{wage}) = \beta_1 + \beta_2 \cdot \text{educ} + \delta \cdot \text{female} + u$. $\delta = -0.18$: women earn $100 \cdot (e^{-0.18} - 1)\% \approx -16.5\%$ less than equally educated men (the ' $\approx -18\%$ ' approximation overstates slightly). Adding $\text{female} \cdot \text{educ}$ tests whether returns to schooling differ by gender (slope dummy). Using both male and female dummies plus intercept $\rightarrow$ dummy variable trap.

Module Summary

- Use $m-1$ dummies for m categories; the trap is perfect collinearity.
- D shifts intercept; $D \cdot X$ shifts slope — test each.
- Applications: seasonality, structural breaks/Chow, diff-in-diff policy evaluation.
- In log models the exact % effect is $100(e^\delta - 1)\%$.

Exam Pointers (Very Important)

- ★ Explain the dummy variable trap and the $m-1$ rule.
- ★ Write a model with both intercept and slope differences by gender; state the test for equal slopes.
- ★ Compute the exact percentage effect for $\delta = 0.25$ in a log-linear model.
- ★ How do dummies implement a Chow test?

MODULE 8: MODEL SPECIFICATION AND DIAGNOSTIC TESTING

Learning Objectives

- Identify specification errors: omitted variables, irrelevant variables, wrong functional form
- Explain the consequences of each error type
- Apply RESET, LM, and nested/non-nested tests
- Follow a disciplined model-building strategy

Types of Specification Error

Omitted relevant variable: if the omitted variable correlates with included regressors, coefficients are **biased and inconsistent** (omitted variable bias: $\text{bias} = \beta_{\text{omitted}} \cdot \delta$, where δ is the regression of omitted on included) — the most serious error. **Included irrelevant variable:** unbiased but inefficient (larger variances), the lesser evil. **Wrong functional form** (linear vs. log): biased, inconsistent estimates. **Measurement error** in Y inflates variances; in X causes **attenuation bias** (coefficients biased toward zero) and inconsistency — the classical errors-in-variables problem.

Diagnostic Tests

Ramsey's RESET: add powers of X (X^2, X^3) to the regression; their joint significance signals general misspecification. **LM tests** for omitted variables. **Nested model** comparison via F-tests. **Non-nested** alternatives (e.g., linear vs. log-linear Y): the **Davidson-MacKinnon J test** / Cox test. **Parameter stability:** Chow, CUSUM tests. A disciplined workflow: start from theory, estimate, run diagnostics (RESET, heteroskedasticity, autocorrelation, normality of residuals via Jarque-Bera), respecify once, and report the specification search honestly — **data mining** (trying 100 specs and reporting the significant one) invalidates p-values.

Choosing Functional Form

Common forms: **linear** $Y = \beta_1 + \beta_2 X$ (marginal effect constant); **log-linear** $\ln Y = \beta_1 + \beta_2 X$ ($\beta_2 \approx \% \Delta Y$ per unit X — growth rates); **lin-log** $Y = \beta_1 + \beta_2 \cdot \ln X$ (diminishing effects); **log-log** $\ln Y = \beta_1 + \beta_2 \cdot \ln X$ (β_2 = elasticity, constant). Choose by theory and test with Box-Cox or J tests; compare R^2 only across models with the **same dependent variable** (R^2 from Y vs. $\ln Y$ regressions are not comparable).

Key Terms & Definitions

Omitted variable bias	Bias from excluding a relevant correlated regressor; $\text{bias} = \beta_{\text{omit}} \cdot \delta$.
RESET	Ramsey's test adding X^2, X^3 to detect general misspecification.
Attenuation bias	Measurement error in X biasing coefficients toward zero.
Data mining	Specification searching that invalidates reported p-values.

Example: Omitted Ability Bias

Wage = $\beta_1 + \beta_2 \cdot \text{educ} + u$ omits ability, which raises both schooling and wages ($\delta > 0$, $\beta_{\text{omit}} > 0$) → β_2 biased upward: part of the 'return to schooling' is really returns to ability. Remedy: proxy for ability (test scores), panel data, or instrumental variables — the subject of advanced courses.

Module Summary

- Omitted relevant variables → bias/inconsistency (worst error); irrelevant included → inefficiency.
- RESET, J tests, Chow/CUSUM diagnose misspecification and instability.
- Functional forms: linear, log-linear, lin-log, log-log — chosen by theory, compared properly.
- Specify from theory, diagnose once, report honestly; avoid data mining.

Exam Pointers (Very Important)

- ★ Derive the omitted variable bias formula and sign it in an example.
- ★ Describe Ramsey's RESET procedure and its null hypothesis.
- ★ Compare measurement error in Y vs. in X.
- ★ Why can't you compare R^2 between Y and $\ln Y$ models?

MODULE 9: INTRODUCTION TO TIME SERIES ECONOMETRICS

Learning Objectives

- Define stationarity and test for unit roots (DF/ADF)
- Explain spurious regression
- Describe AR, MA, ARMA, ARIMA models and the Box-Jenkins approach
- Explain cointegration and error-correction models

Stationarity and Unit Roots

A series is (weakly) **stationary** if its mean, variance, and autocovariances are time-invariant; most macro series (GDP, prices) are not — they trend. A **random walk** $Y_t = Y_{t-1} + \varepsilon_t$ is non-stationary (variance grows with t) and is the simplest **unit root** process. The **Dickey-Fuller test** estimates $\Delta Y_t = \delta \cdot Y_{t-1} + \varepsilon_t$ and tests $H_0: \delta = 0$ (unit root) using special DF critical values (not standard t !). The **Augmented DF (ADF)** adds lagged ΔY_t terms to mop up autocorrelation; the **Phillips-Perron** test is a non-parametric alternative. A series needing d differences to become stationary is **integrated of order d , $I(d)$** .

Spurious Regression and Cointegration

Regressing two independent random walks on each other typically yields a 'significant' relationship with high R^2 — the **spurious regression** problem (Granger & Newbold): t -statistics diverge, and $DW \rightarrow 0$ is the warning sign. The rescue is **cointegration** (Engle-Granger): if Y_t, X_t are $I(1)$ but a linear combination $z_t = Y_t - \beta X_t$ is $I(0)$, they share a long-run equilibrium. Test: regress Y on X , then ADF-test the residuals for stationarity. If cointegrated, build an **error-correction model (ECM)**: $\Delta Y_t = \alpha + \gamma \cdot z_{t-1} + \text{lags of } \Delta Y, \Delta X + \varepsilon_t$, where $\gamma < 0$ measures the speed of return to equilibrium — combining short-run dynamics with the long-run relation.

ARIMA and Box-Jenkins

AR(p): Y_t depends on its own p lags; **MA(q)**: Y_t depends on q lagged errors; **ARMA(p, q)** combines both; **ARIMA(p, d, q)** adds d differences for non-stationary series. The **Box-Jenkins** methodology: (1) **identification** — use ACF/PACF patterns to guess p, d, q (AR: PACF cuts off; MA: ACF cuts off); (2) **estimation** by ML/OLS; (3) **diagnostic checking** — residuals must be white noise (Ljung-Box Q); (4) **forecasting**. Modern practice supplements this with information criteria (AIC/BIC) for order selection.

Key Terms & Definitions

Stationarity	Time-invariant mean, variance, autocovariances.
Unit root / $I(1)$	Non-stationary stochastic trend; stationary after first differencing.
Spurious regression	Nonsense significant regression between independent non-stationary series.

Cointegration / ECM

Long-run equilibrium among I(1) series; error-correction captures adjustment speed.

Example: Testing Consumption and Income

C_t and Y_t are both I(1) (ADF fails to reject unit root in levels, rejects in differences). Regression $C_t = \beta_1 + \beta_2 Y_t + u_t$ gives stationary residuals (ADF rejects unit root) → cointegrated: a long-run consumption function exists. ECM: $\Delta C_t = 0.02 - 0.25 \cdot (C_{t-1} - 0.9 \cdot Y_{t-1}) + 0.5 \cdot \Delta Y_t$: 25% of any disequilibrium is corrected each period.

Module Summary

- Test stationarity with DF/ADF (special critical values); difference I(1) series once.
- Spurious regression: independent random walks look related; DW → 0 warns.
- Cointegration (Engle-Granger): I(1) series with I(0) combination → long-run equilibrium.
- ECM combines short-run dynamics with error-correction toward equilibrium; ARIMA/Box-Jenkins for forecasting.

Exam Pointers (Very Important)

- ★ Write the Dickey-Fuller regression and state its null; why not use standard t critical values?
- ★ What is spurious regression and how do you detect it?
- ★ Outline the Engle-Granger two-step cointegration procedure.
- ★ Interpret the error-correction coefficient $\gamma = -0.25$.

MODULE 10: INTRODUCTION TO PANEL DATA METHODS

Learning Objectives

- Describe panel data structure and its advantages
- Explain fixed effects vs. random effects models
- Perform the Hausman test for model choice
- Handle heteroskedasticity and autocorrelation in panels

Why Panel Data?

Panel data follows N units over T periods (e.g., 50 firms, 10 years). Advantages: more variation and degrees of freedom; controls for **unobserved heterogeneity** (firm culture, individual ability) that biases cross-section estimates; and allows studying dynamics (adjustment, state dependence). The basic model: $Y_{it} = \beta_1 + \beta_2 \cdot X_{2it} + \dots + \beta_k \cdot X_{kit} + \alpha_i + u_{it}$, where α_i is the unobserved unit-specific effect. The key question is how α_i relates to the regressors.

Fixed vs. Random Effects

Fixed effects (FE / within estimator): treats α_i as parameters (unit dummies) or eliminates them by demeaning (subtracting unit means); consistent whether or not α_i correlates with X — the safe choice — but cannot estimate coefficients of time-invariant regressors (gender, location). **Random effects (RE / GLS)**: treats α_i as random and **uncorrelated with X** ; more efficient if that assumption holds, inconsistent if it fails. **Pooled OLS** ignores α_i entirely and is valid only if $\alpha_i = 0$ effectively. The **Hausman test** decides: $H_0 = \text{RE consistent and efficient (no correlation)}$; rejection $\rightarrow$ use FE. Test statistic $H = (\beta_{FE} - \beta_{RE})' [\text{Var}(\beta_{FE}) - \text{Var}(\beta_{RE})]^{-1} (\beta_{FE} - \beta_{RE}) \sim \chi^2(k)$.

Panel Diagnostics and Extensions

Panels can suffer **groupwise heteroskedasticity** (different σ_i per unit), **contemporaneous correlation** across units, and **serial correlation** within units — use cluster-robust standard errors (cluster by unit) as the default. **Dynamic panels** with lagged Y_{it} need Arellano-Bond GMM because FE is inconsistent (Nickell bias) when T is small. Always test for the presence of unit effects first (F-test for FE, Breusch-Pagan LM for RE vs. pooled OLS).

Key Terms & Definitions

Unobserved heterogeneity	Unit-specific unmeasured factors (α_i) potentially biasing estimates.
Fixed effects	Within estimator eliminating α_i ; consistent regardless of correlation with X .
Random effects	GLS estimator assuming α_i uncorrelated with X ; efficient if true.
Hausman test	χ^2 test choosing between FE and RE based on coefficient differences.

Example: Hausman in Practice

Production function on 200 firms $\times$ 8 years. FE gives capital elasticity 0.35 (se 0.04); RE gives 0.42 (se 0.02). Hausman $H = 9.8$, $\chi^2(3)$ critical 7.81 $\rightarrow$ reject RE consistency: unobserved firm effects correlate with inputs (better-managed firms invest more). Report FE with clustered standard errors.

Module Summary

- Panels combine N units $\times$ T periods; they control unobserved heterogeneity and add dynamics.
- FE (within) is consistent always but drops time-invariant variables; RE (GLS) is efficient only if $\alpha_i \perp X$.
- Hausman test: significant coefficient differences $\rightarrow$ use FE.
- Default to clustered standard errors; dynamic panels need Arellano-Bond GMM.

Exam Pointers (Very Important)

- ★ Write the panel model with unobserved effects and explain the identification problem.
- ★ Compare FE and RE: assumptions, what each can/can't estimate.
- ★ State the Hausman test null, statistic, and decision rule.
- ★ Why cluster standard errors by unit in panels?