Extra Reading – Derivations and Proofs
Supplementary Unit

Prepared by

Aditya Korekallu Srinivasa

Scientist (Senior Scale)

ICAR-Indian Agricultural Research Institute
New Delhi

April 24, 2025

Contents

1 OLS Estimators in the Simple Linear Regression Model: Full Derivation

1.1 Model Setup

Consider the simple linear regression model:

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \quad i = 1, 2, \dots, n$$

where:

- Y_i is the dependent variable,
- X_i is the independent variable,
- β_0 is the intercept,
- β_1 is the slope,
- u_i is the error term.

1.2 Ordinary Least Squares (OLS) Estimation

The OLS method chooses $\hat{\beta}_0$ and $\hat{\beta}_1$ to minimize the sum of squared residuals:

$$S = \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2$$

where b_0 and b_1 are candidate values for the intercept and slope.

1.3 Step 1: Take Partial Derivatives

To find the minimum, take the partial derivatives of S with respect to b_0 and b_1 and set them to zero:

$$\frac{\partial S}{\partial b_0} = -2 \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i) = 0$$

$$\frac{\partial S}{\partial b_1} = -2 \sum_{i=1}^{n} X_i (Y_i - b_0 - b_1 X_i) = 0$$

1.4 Step 2: Rearranging the Equations

Divide both equations by -2 (does not affect the solution):

$$\sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i) = 0$$

$$\sum_{i=1}^{n} X_i (Y_i - b_0 - b_1 X_i) = 0$$

Expand the first equation:

$$\sum_{i=1}^{n} Y_i - nb_0 - b_1 \sum_{i=1}^{n} X_i = 0$$

$$nb_0 + b_1 \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} Y_i$$

$$b_0 = \frac{1}{n} \sum_{i=1}^{n} Y_i - b_1 \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$b_0 = \bar{Y} - b_1 \bar{X}$$

where $\bar{Y}$ and $\bar{X}$ are sample means.

Expand the second equation:

$$\sum_{i=1}^{n} X_i Y_i - b_0 \sum_{i=1}^{n} X_i - b_1 \sum_{i=1}^{n} X_i^2 = 0$$

Substitute b_0 from above:

$$\sum_{i=1}^{n} X_i Y_i - (\bar{Y} - b_1 \bar{X}) \sum_{i=1}^{n} X_i - b_1 \sum_{i=1}^{n} X_i^2 = 0$$

But it's easier to use the two equations together:

$$\sum_{i=1}^{n} X_i Y_i - b_0 \sum_{i=1}^{n} X_i - b_1 \sum_{i=1}^{n} X_i^2 = 0$$

Plug in $b_0 = \bar{Y} - b_1 \bar{X}$ and solve for b_1:

$$\sum_{i=1}^{n} X_i Y_i - (\bar{Y} - b_1 \bar{X}) \sum_{i=1}^{n} X_i - b_1 \sum_{i=1}^{n} X_i^2 = 0$$

$$\sum_{i=1}^{n} X_i Y_i - \bar{Y} \sum_{i=1}^{n} X_i + b_1 \bar{X} \sum_{i=1}^{n} X_i - b_1 \sum_{i=1}^{n} X_i^2 = 0$$

$$\sum_{i=1}^{n} X_i Y_i - \bar{Y} \sum_{i=1}^{n} X_i = b_1 \left(\sum_{i=1}^{n} X_i^2 - \bar{X} \sum_{i=1}^{n} X_i \right)$$

But $\sum_{i=1}^{n} X_i = n\bar{X}$, so:

$$\sum_{i=1}^{n} X_i^2 - \bar{X} \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2$$

Similarly,

$$\sum_{i=1}^{n} X_i Y_i - \bar{Y} \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} X_i Y_i - n\bar{Y}\bar{X} = \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$$

Therefore,

$$b_1 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$$

and

$$b_0 = \bar{Y} - b_1 \bar{X}$$

1.5 Explanation

- The OLS estimators are derived by minimizing the squared differences between the observed Y_i and the predicted values.
- The slope estimator b_1 measures the average change in Y per unit change in X.
- The intercept b_0 ensures the regression line passes through the sample means $(\bar{X}, \bar{Y})$.

2 Proof that OLS Estimators are Unbiased

2.1 Goal

Show that $E[\hat{\beta}_1] = \beta_1$ and $E[\hat{\beta}_0] = \beta_0$ under the classical assumptions.

2.2 Step-by-Step Proof for $\hat{\beta}_1$

Recall:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Substitute $Y_i = \beta_0 + \beta_1 X_i + u_i$:

$$Y_i - \bar{Y} = \beta_1(X_i - \bar{X}) + (u_i - \bar{u})$$

where $\bar{u} = \frac{1}{n}\sum u_i$.

Plug this into the numerator:

$$\sum(X_i - \bar{X})(Y_i - \bar{Y}) = \sum(X_i - \bar{X})[\beta_1(X_i - \bar{X}) + (u_i - \bar{u})]$$

$$= \beta_1 \sum(X_i - \bar{X})^2 + \sum(X_i - \bar{X})(u_i - \bar{u})$$

Therefore,

$$\hat{\beta}_1 = \beta_1 + \frac{\sum(X_i - \bar{X})(u_i - \bar{u})}{\sum(X_i - \bar{X})^2}$$

Take expectations, noting that $E[u_i|X] = 0$ and $E[\bar{u}|X] = 0$:

$$E[\hat{\beta}_1|X] = \beta_1 + \frac{E[\sum(X_i - \bar{X})(u_i - \bar{u})|X]}{\sum(X_i - \bar{X})^2}$$

But $E[u_i - \bar{u}|X] = 0$, so numerator is zero:

$$E[\hat{\beta}_1|X] = \beta_1$$

Therefore, OLS slope estimator is unbiased.

2.3 Explanation

- The key step is substituting the true model into the OLS formula, then using the zero conditional mean assumption.
- The result shows that, on average, OLS recovers the true slope.

3 Variance of the OLS Estimator: Detailed Derivation

3.1 Goal

Find $\text{Var}(\hat{\beta}_1)$.

3.2 Step-by-Step

Recall:

$$\hat{\beta}_1 = \beta_1 + \frac{\sum(X_i - \bar{X})(u_i - \bar{u})}{\sum(X_i - \bar{X})^2}$$

But since $\sum(X_i - \bar{X}) = 0$, $\sum(X_i - \bar{X})\bar{u} = 0$. So,

$$\hat{\beta}_1 = \beta_1 + \frac{\sum(X_i - \bar{X})u_i}{\sum(X_i - \bar{X})^2}$$

Therefore,

$$\text{Var}(\hat{\beta}_1) = \text{Var}\left(\frac{\sum(X_i - \bar{X})u_i}{\sum(X_i - \bar{X})^2}\right)$$

If errors are homoscedastic and uncorrelated ($\text{Var}(u_i) = \sigma^2$, $\text{Cov}(u_i, u_j) = 0$ for $i \neq j$):

$$\text{Var}(\hat{\beta}_1) = \frac{\sum(X_i - \bar{X})^2 \, \text{Var}(u_i)}{[\sum(X_i - \bar{X})^2]^2} = \frac{\sigma^2}{\sum(X_i - \bar{X})^2}$$

3.3 Explanation

- The variance of the OLS slope estimator depends on the error variance and the spread of X.
- More spread in X (larger denominator) means more precise estimates.

4 Gauss–Markov Theorem: Why OLS is BLUE

4.1 Statement

Under the classical assumptions (linear model, random sampling, no perfect collinearity, zero conditional mean, homoscedasticity), OLS is the Best Linear Unbiased Estimator (BLUE).

4.2 Proof Outline (Step-by-Step)

Step 1: Linearity and Unbiasedness

- OLS estimator is a linear function of Y_i.
- As shown above, $E[\hat{\beta}_1] = \beta_1$.

Step 2: Minimum Variance

- Consider any other linear unbiased estimator:

$$\tilde{\beta}_1 = \sum_{i=1}^{n} a_i Y_i$$

with $E[\tilde{\beta}_1] = \beta_1$.

- Show that $\text{Var}(\tilde{\beta}_1) \geq \text{Var}(\hat{\beta}_1)$.
- Using Lagrange multipliers and the unbiasedness constraint, OLS gives the minimum variance.

4.3 Explanation

- OLS is "best" among all linear unbiased estimators because it has the smallest variance.
- This is why OLS is preferred when the classical assumptions hold.

5 Omitted Variable Bias: Derivation

Suppose the true model is:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$

But we estimate:

$$Y_i = \alpha_0 + \alpha_1 X_{1i} + v_i$$

5.1 Step-by-Step Derivation

1. The error in the misspecified model is $v_i = \beta_2 X_{2i} + u_i$.
2. The OLS estimator for α_1 is:

$$\hat{\alpha}_1 = \frac{\sum (X_{1i} - \bar{X}_1)(Y_i - \bar{Y})}{\sum (X_{1i} - \bar{X}_1)^2}$$

3. Substitute Y_i:

$$Y_i - \bar{Y} = \beta_1 (X_{1i} - \bar{X}_1) + \beta_2 (X_{2i} - \bar{X}_2) + (u_i - \bar{u})$$

4. Plug into the numerator:

$$\sum (X_{1i} - \bar{X}_1)[\beta_1 (X_{1i} - \bar{X}_1) + \beta_2 (X_{2i} - \bar{X}_2) + (u_i - \bar{u})]$$

5. This expands to:

$$\beta_1 \sum (X_{1i} - \bar{X}_1)^2 + \beta_2 \sum (X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2) + \sum (X_{1i} - \bar{X}_1)(u_i - \bar{u})$$

6. The denominator is $\sum (X_{1i} - \bar{X}_1)^2$.
7. Taking expectations and using $E[u_i] = 0$:

$$E[\hat{\alpha}_1] = \beta_1 + \beta_2 \frac{\text{Cov}(X_1, X_2)}{\text{Var}(X_1)}$$

5.2 Explanation

- If X_1 and X_2 are correlated and X_2 is omitted, the OLS estimator for X_1 is biased.
- The direction and magnitude of the bias depend on the correlation and the true effect of X_2.

6 Durbin–Watson Statistic: Derivation and Interpretation

6.1 Definition

$$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

where e_t are OLS residuals.

6.2 Step-by-Step Explanation

- Numerator: measures how much the residual changes from one period to the next (large if residuals change a lot, small if residuals are similar).
- Denominator: total variation in residuals.
- $d \approx 2$ (no autocorrelation), $d < 2$ (positive autocorrelation), $d > 2$ (negative autocorrelation).

6.3 Expected Value under AR(1)

If $u_t = \rho u_{t-1} + \epsilon_t$ (AR(1)), then for large n:

$$E[d] \approx 2(1 - \rho)$$

So, if $\rho = 0$ (no autocorrelation), $E[d] = 2$; if $\rho = 1$ (perfect positive autocorrelation), $E[d] = 0$.

7 Maximum Likelihood Estimation for Logit and Probit: Steps

7.1 Logit Model

Model:

$$P_i = P(Y_i = 1|X_i) = \frac{\exp(\beta_0 + \beta_1 X_i)}{1 + \exp(\beta_0 + \beta_1 X_i)}$$

Likelihood:

$$L = \prod_{i=1}^{n} P_i^{Y_i}(1 - P_i)^{1-Y_i}$$

Log-likelihood:

$$\ln L = \sum_{i=1}^{n} [Y_i \ln P_i + (1 - Y_i) \ln(1 - P_i)]$$

Estimation:

- Take derivatives of $\ln L$ with respect to β_0, β_1.
- Set equal to zero; solve numerically (no closed-form solution).
- Use Newton-Raphson or other iterative algorithms.

7.2 Probit Model

Model:

$$P_i = P(Y_i = 1|X_i) = \Phi(\beta_0 + \beta_1 X_i)$$

where Φ is the standard normal CDF.

Likelihood:

$$L = \prod_{i=1}^{n} \Phi(\beta_0 + \beta_1 X_i)^{Y_i}[1 - \Phi(\beta_0 + \beta_1 X_i)]^{1-Y_i}$$

Estimation:

- Similar steps as logit: maximize the log-likelihood by iterative methods.

7.3 Explanation

- MLE finds parameter values that make the observed data most likely.
- For logit/probit, this involves maximizing a nonlinear function.

8 Marginal Effects in Logit and Probit: Derivation and Interpretation

8.1 Logit Marginal Effect

$$\frac{\partial P}{\partial X} = \beta_1 \cdot P(1 - P)$$

Explanation:

- P is the predicted probability at a given X.
- The marginal effect tells how much P changes for a small change in X.
- Often evaluated at the mean of X or for a typical value.

8.2 Probit Marginal Effect

$$\frac{\partial P}{\partial X} = \beta_1 \cdot \phi(\beta_0 + \beta_1 X)$$

where ϕ is the standard normal PDF.

Explanation:

- The marginal effect is largest near the mean (where the slope of the normal CDF is steepest).
- Interpretation: For a one-unit increase in X, the probability increases by $\beta_1 \cdot \phi(\cdot)$ at that value.

9 Instrumental Variables (IV) and 2SLS: Step-by-Step Derivation

9.1 Instrumental Variables (IV) Estimator

Suppose $Y = \beta_0 + \beta_1 X + u$, but X is endogenous. Let Z be an instrument for X.

Step 1:

$$\hat{\beta}_1^{IV} = \frac{\sum(Z_i - \bar{Z})(Y_i - \bar{Y})}{\sum(Z_i - \bar{Z})(X_i - \bar{X})}$$

Explanation:

- Z must be correlated with X but uncorrelated with u.
- IV uses only the variation in X that is explained by Z.

9.2 Two-Stage Least Squares (2SLS)

Step 1: First Stage

- Regress X on all exogenous variables (including Z), obtain fitted values $\hat{X}$.

Step 2: Second Stage

- Regress Y on $\hat{X}$ to estimate β_1.

Explanation:

- 2SLS generalizes IV to systems with multiple endogenous regressors and instruments.
- The first stage purges X of endogeneity.

10 Matrix Approach to OLS: Complete Steps

10.1 Model

$$Y = X\beta + u$$

where Y is $n \times 1$, X is $n \times k$, β is $k \times 1$.

10.2 OLS Estimator

$$\hat{\beta} = (X'X)^{-1}X'Y$$

Derivation:

- The sum of squared residuals: $S = (Y - Xb)'(Y - Xb)$.
- Take derivative with respect to b, set to zero:

$$\frac{\partial S}{\partial b} = -2X'(Y - Xb) = 0$$

- Rearranged:

$$X'Y = X'Xb \implies b = (X'X)^{-1}X'Y$$

10.3 Variance-Covariance Matrix

$$\mathrm{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$$

10.4 Explanation

- The matrix formula generalizes OLS to multiple regressors.
- $(X'X)^{-1}$ captures the spread and correlation among regressors.

11 Chow Test for Structural Breaks: Step-by-Step

11.1 Goal

Test whether regression coefficients are the same in two subsamples.

11.2 Step-by-Step

1. Estimate the model on the full sample, get RSS (restricted) = S_c.
2. Estimate the model on each subsample, get S_1 and S_2.
3. Compute:

$$F = \frac{(S_c - (S_1 + S_2))/k}{(S_1 + S_2)/(n_1 + n_2 - 2k)}$$

   where k is the number of parameters.
4. Compare F to the critical value from the F-distribution.

11.3 Explanation

- If F is large, coefficients differ across groups (structural break).

12 Partial Correlation Coefficient: Derivation

12.1 Formula

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$$

12.2 Explanation

- $r_{12.3}$ is the correlation between X_1 and X_2 holding X_3 constant.
- It removes the linear effect of X_3 from both X_1 and X_2.

Glossary

- **BLUE**: Best Linear Unbiased Estimator.
- **GLS**: Generalized Least Squares.
- **WLS**: Weighted Least Squares.
- **IV**: Instrumental Variables.
- **2SLS**: Two-Stage Least Squares.
- **Likelihood Function**: The probability of observed data as a function of model parameters.
- **Marginal Effect**: The change in probability or expected value for a unit change in a regressor.
- **Chow Test**: A test for structural breaks in regression models.
- **Partial Correlation**: The correlation between two variables after removing the effect of a third.

Practice Questions

1. Derive the OLS estimator for the slope in the simple linear regression model, step by step.
2. Prove that the OLS estimator is unbiased under the classical assumptions, showing all steps.
3. Show how the variance of the OLS estimator is calculated, and explain each step.
4. State and prove the Gauss–Markov theorem, with a clear explanation of each step.
5. Derive the omitted variable bias formula, step by step.
6. Show the steps of the GLS transformation for a simple regression with known heteroscedasticity.

7. Derive the Durbin–Watson statistic and its expected value under AR(1) errors.
8. State and explain the order and rank conditions for identification in simultaneous equations.
9. Write the likelihood function for the logit and probit models and explain how to obtain parameter estimates.
10. Derive the marginal effect formula for the logit and probit models, step by step.
11. Show how the IV estimator is obtained and explain why it solves endogeneity.
12. Explain the two stages of 2SLS and the logic behind each, with all steps shown.
13. Write the OLS estimator and variance in matrix form, and explain each step.
14. State and explain the Chow test for structural breaks, with all steps.
15. Derive the formula for the partial correlation coefficient.

References

- Gujarati, D. N., & Porter, D. C. (2010). *Basic Econometrics* (5th ed.). McGraw-Hill.
- Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach* (5th ed.). Cengage Learning.