Unit 1 Summary

Econometrics Study Guide: Simple Linear Regression Model (SLRM)

Índice

- 1.1 What is Econometrics?
- 1.2 The Econometric Model
- 1.3 Types of Data
- 1.4 Assumptions of the SLRM
- 1.5 Estimating the SLRM: Least Squares Principle
- 1.6 Prediction
- 1.7 The Log-Log Model (Constant Elasticity Model)
- 1.8 Properties of the Least Squares Estimators
- 1.9 Probability Distribution of the LS Estimators
- 1.10 Estimating the Variance
- 1.11 Interval Estimation
- 1.12 Hypothesis Testing
- 1.13 Confidence Interval for a Linear Combination of Parameters
- 1.14 Hypothesis Testing for a Linear Combination
- 1.15 Least Squares Prediction (Prediction Interval)
- 1.16 Measuring the Goodness of Fit
- 1.17 Log Functional Form
- 1.18 Testing Normality of the Error Terms
- 1.19 Changing the Scale of the Data
- Summary of Key Formulas

1.1 What is Econometrics?

Econometrics applies mathematical statistics and statistical inference tools to empirically measure economic relationships.

Example: Demand function for edible chicken

$$\ln(q_i^d) = \beta_1 + \beta_2 \ln(p_i) + e_i$$

$$q_i^d = \exp(\beta_1 + \beta_2 \ln(p_i) + e_i)$$

$$\frac{d \ln q^d}{d \ln p} = \beta_2 \approx \frac{\Delta\% q^d}{\Delta\% p}$$

Key Points:

- Theory is fundamental to econometric models (e.g., wage-income function, asset returns).
- Econometric models combine deterministic (systematic) and stochastic (random) components.

1.2 The Econometric Model

Structure:

$$y_i = \underbrace{\beta_1 + \beta_2 x_i}_{\text{Systematic}} + \underbrace{e_i}_{\text{Random}}$$

Components:

- **Systematic:** Economic model derived from theory.
- **Random:** Error term capturing unobserved factors.

1.3 Types of Data

1. **Cross-sectional:** Observations across entities at a single point in time.
2. **Time series:** Observations over time for a single entity.
3. **Panel data:** Combination of cross-sectional and time series data.

1.4 Assumptions of the SLRM

1. **Linearity:** $y_i = \beta_1 + \beta_2 x_i + e_i$
2. **Strict Exogeneity:** $E(e_i|x_i) = 0$
3. **Conditional Homoskedasticity:** $\mathrm{Var}(e_i|x_i) = \sigma^2$
4. **Uncorrelated Errors:** $\mathrm{Cov}(e_i, e_j|x) = 0$ for $i \neq j$
5. **Variability in x:** x_i must vary (not constant).
6. **Normality of Errors:** $e_i|x \sim N(0, \sigma^2)$ (optional for large samples).

Consequences if Assumptions Fail:

- Violations lead to biased estimators, inefficient inferences, or invalid tests.

1.5 Estimating the SLRM: Least Squares Principle

Objective: Minimize the sum of squared residuals:

$$\min_{\beta_1,\beta_2} \sum e_i^2 = \sum (y_i - \beta_1 - \beta_2 x_i)^2$$

OLS Estimators:

$$b_2 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2}$$

$$b_1 = \bar{y} - b_2 \bar{x}$$

Interpretation:

- b_2: Marginal effect of x on y.
- b_1: Intercept.

1.6 Prediction

Point Prediction:

$$\hat{y}_i = b_1 + b_2 x_i$$

Elasticity:

$$\varepsilon_i = \frac{\Delta y / y}{\Delta x / x} = b_2 \cdot \frac{x_i}{\hat{y}_i}$$

- Varies along the regression line.

Average Elasticity:

$$\hat{\bar{\varepsilon}} = \frac{1}{n} \sum \hat{\varepsilon}_i$$

1.7 The Log-Log Model (Constant Elasticity Model)

Form:

$$\ln(y_i) = \beta_1 + \beta_2 \ln(x_i) + e_i$$

Properties:

- β_2 is the elasticity of y with respect to x.
- Constant elasticity model.

Interpretation: A 1% change in x leads to a $\beta_2\%$ change in y.

1.8 Properties of the Least Squares Estimators

1. **Linearity:** b_1 and b_2 are linear functions of y_i.
2. **Unbiasedness:** $E(b_1|x) = \beta_1$, $E(b_2|x) = \beta_2$.
3. **Efficiency:** OLS estimators have the smallest variance among linear unbiased estimators (Gauss-Markov Theorem).
4. **Consistency:** As $n \to \infty$, $b_1 \to \beta_1$ and $b_2 \to \beta_2$ in probability.

1.9 Probability Distribution of the LS Estimators

If errors are normal:

$$b_1|x \sim N\left(\beta_1, \sigma^2 \frac{\sum x_i^2}{n\sum(x_i - \bar{x})^2}\right)$$

$$b_2|x \sim N\left(\beta_2, \frac{\sigma^2}{\sum(x_i - \bar{x})^2}\right)$$

For large samples, normality holds approximately (Central Limit Theorem).

1.10 Estimating the Variance

Variance of Error Term:

$$\hat{\sigma}^2 = \frac{\sum \hat{e}_i^2}{n-2}$$

Variances and Covariance of Estimators:

$$\widehat{\mathrm{Var}(b_1|x)} = \hat{\sigma}^2 \frac{\sum x_i^2}{n\sum(x_i - \bar{x})^2}$$

$$\widehat{\mathrm{Var}(b_2|x)} = \frac{\hat{\sigma}^2}{\sum(x_i - \bar{x})^2}$$

$$\widehat{\mathrm{Cov}(b_1, b_2|x)} = \hat{\sigma}^2 \left(\frac{-\bar{x}}{\sum(x_i - \bar{x})^2}\right)$$

Standard Errors:

$$se(b_1) = \sqrt{\widehat{\mathrm{Var}(b_1|x)}}, \quad se(b_2) = \sqrt{\widehat{\mathrm{Var}(b_2|x)}}$$

1.11 Interval Estimation

Confidence Interval for β_2:

$$b_2 \pm t_{\alpha/2} \cdot se(b_2)$$

Confidence Interval for β_1:

$$b_1 \pm t_{\alpha/2} \cdot se(b_1)$$

Interpretation: The interval contains the true parameter with $(1-\alpha)\%$ confidence.

1.12 Hypothesis Testing

Steps:

1. Specify H_0 and H_1.
2. Compute test statistic:

$$t = \frac{b_2 - c}{se(b_2)} \sim t_{n-2}$$

3. Determine rejection region based on α and tail(s).
4. Conclude based on test statistic or p-value.

Types of Tests:

- Two-tailed: $H_1 : \beta_2 \neq c$
- One-tailed: $H_1 : \beta_2 > c$ or $H_1 : \beta_2 < c$

1.13 Confidence Interval for a Linear Combination of Parameters

Example: $E(y|x_0) = \beta_1 + \beta_2 x_0$

Point Estimate:

$$\widehat{E(y|x_0)} = b_1 + b_2 x_0$$

Variance:

$$\text{Var}(\widehat{E(y|x_0)}|x) = \text{Var}(b_1|x) + x_0^2\text{Var}(b_2|x) + 2x_0\text{Cov}(b_1, b_2|x)$$

Confidence Interval:

$$\widehat{E(y|x_0)} \pm t_{\alpha/2} \cdot se(\widehat{E(y|x_0)})$$

1.14 Hypothesis Testing for a Linear Combination

Example: Test $H_0 : \beta_1 + 2500\beta_2 = 1000$

Test Statistic:

$$t = \frac{b_1 + 2500b_2 - 1000}{se(b_1 + 2500b_2)} \sim t_{n-2}$$

1.15 Least Squares Prediction (Prediction Interval)

Forecast Error:

$$f = y_0 - \hat{y}_0$$

Variance of Forecast Error:

$$\text{Var}(f|x) = \sigma^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum(x_i - \bar{x})^2}\right]$$

Prediction Interval:

$$\hat{y}_0 \pm t_{\alpha/2} \cdot se(f|x)$$

Note: Intervals widen as x_0 moves away from $\bar{x}$.

1.16 Measuring the Goodness of Fit

Decomposition of Variance:

$$\text{SST} = \text{SSR} + \text{SSE}$$

- SST: Total sum of squares.
- SSR: Sum of squares due to regression.
- SSE: Sum of squared errors.

Coefficient of Determination:

$$R^2 = \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\text{SSE}}{\text{SST}}$$

- Proportion of variance in y explained by x.
- $0 \leq R^2 \leq 1$.

Generalized R^2:

$$R^2 = [\text{Corr}(y, \hat{y})]^2$$

1.17 Log Functional Form

Log-Linear Model:

$$\ln(y_i) = \beta_1 + \beta_2 x_i + e_i$$

- β_2: semi-elasticidad (% cambio en y por 1 unidad en x).

Exponential Form:

$$y_i = \exp(\beta_1 + \beta_2 x_i + e_i)$$

Log-Normal Distribution:
Si $e_i \sim N(0, \sigma^2)$, entonces y_i es log-normal.

$$E(y|x) = \exp(\beta_1 + \beta_2 x + \sigma^2/2)$$

Corrected Predictor:

$$\hat{y}_i = \exp(b_1 + b_2 x_i + \hat{\sigma}^2/2)$$

- Mejor para muestras grandes; sin corrección si $n < 30$.

Model Comparison:

- R^2 en Lin-Lin y Log-Lin no son comparables.
- Usar $R_g^2 = [\text{Corr}(y, \hat{y})]^2$ para comparar.

1.18 Testing Normality of the Error Terms

Methods:

- Gráficos: histograma, Q-Q plot.
- Tests: Jarque-Bera (JB), Shapiro-Wilk.

Jarque-Bera Test:

$$JB = \frac{n}{6}\left(S^2 + \frac{(K-3)^2}{4}\right)$$

- S: skewness, K: kurtosis.
- H_0: errores normales $\sim N(0, \sigma^2)$.
- $JB \sim \chi^2(2)$ asintóticamente.

1.19 Changing the Scale of the Data

Effects of Rescaling:

1. Multiplicar y por c: b_1, b_2 se multiplican por c.
2. Multiplicar x por c: b_2 se divide entre c, b_1 no cambia.
3. R^2, elasticidades y coeficientes estandarizados son invariantes a escala.

Summary of Key Formulas

- **OLS Estimators:**

$$b_2 = \frac{\sum(x_i - \bar{x})y_i}{\sum(x_i - \bar{x})^2}$$

$$b_1 = \bar{y} - b_2\bar{x}$$

- **Variance of Error Term:**

$$\hat{\sigma}^2 = \frac{\sum \hat{e}_i^2}{n-2}$$

- **Test Statistic:**

$$t = \frac{b_2 - c}{se(b_2)} \sim t_{n-2}$$

- **Prediction Interval:**

$$\hat{y}_0 \pm t_{\alpha/2} \cdot se(f|x)$$

- **Goodness of Fit:**

$$R^2 = 1 - \frac{\text{SSE}}{\text{SST}}$$