Econometrics

Econ 1314

Personal Notes

Following: Jeffrey M. Wooldridge, Introductory Econometrics: A Modern Approach (2012)

Personal Notes

Nova School of Business and Economics

Fall 2025

Compiled on July 26, 2025

Contents

1 Economic Questions and Data

1.1 Economic Questions We Examine

Key Point

Econometrics seeks to answer fundamental economic questions through statistical analysis of data, with particular emphasis on identifying causal relationships rather than mere correlations.

The primary objectives of econometric analysis include:

- **Causal effects**: Understanding cause-and-effect relationships in economic data
- **Policy evaluation**: Measuring the impact of economic policies and interventions
- **Treatment effects**: Quantifying how changes in one variable affect economic outcomes
- **Forecasting**: Predicting future economic conditions based on historical patterns

1.2 Causal Effects and Idealized Experiments

The fundamental challenge in econometrics is establishing causality in observational data. The **gold standard** for causal inference is the randomized controlled trial (RCT), but this is often impossible in economics.

Key Point

The key insight is that we want to compare outcomes under different "treatments" while holding all other factors constant (*ceteris paribus*).

Key concepts include:

- **Randomized controlled trials (RCT)**: Gold standard for causal inference
- **Ceteris paribus**: "All else being equal" – the challenge in observational data
- **Selection bias**: When treatment assignment is not random
- **Counterfactual thinking**: What would have happened without the treatment?

Warning

Selection bias occurs when the assignment to treatment is correlated with other factors that affect the outcome. This makes it difficult to isolate the causal effect of the treatment.

1.3 Types of Data

Economic data comes in several forms, each with distinct characteristics and appropriate analytical methods:

1. **Cross-sectional data**: Observations of multiple units at a single point in time
   - Example: Survey of household incomes in 2025
   - Useful for studying relationships between variables at a point in time

2. **Time series data**: Observations of a single unit over multiple time periods
   - Example: Monthly unemployment rates from 1990-2025
   - Useful for studying trends and dynamics over time
3. **Panel data**: Combination of cross-sectional and time series data
   - Example: Annual income for the same individuals over 10 years
   - Allows control for unobserved heterogeneity
4. **Pooled cross-sections**: Multiple cross-sections at different times
   - Example: Different random samples of households in 2020 and 2025
   - Different units observed in each time period

2 The Bivariate Regression Model

2.1 Model Specification

2.1.1 The Model

The simple linear regression model is the foundation of econometric analysis:

Formula

$$y = \beta_0 + \beta_1 x + u \tag{1}$$

Where:

- y: dependent variable (outcome of interest)
- x: independent variable (explanatory variable)
- u: error term (unobserved factors affecting y)
- β_0: intercept parameter
- β_1: slope parameter

2.1.2 The "As Good As Random Assignment" Assumption

Assumption

Zero Conditional Mean Assumption: $\mathbb{E}[u|x] = 0$
This means that the expected value of the error term is zero for any value of x.

This assumption is crucial for causal interpretation because it implies:

- The error term is uncorrelated with the explanatory variable
- No omitted variables that are correlated with both x and y
- No measurement error in x (under certain conditions)
- No simultaneity bias

Warning

The zero conditional mean assumption is often violated in practice, leading to the need for more advanced estimation methods such as instrumental variables.

2.2 Population Regression Function

The **Population Regression Function (PRF)** is:

Formula

$$\mathbb{E}[y|x] = \beta_0 + \beta_1 x \tag{2}$$

This shows the expected value of y given x. The PRF represents the true relationship in the population that we want to estimate.

2.3 Interpretation of the Slope

The slope coefficient β_1 measures:

- The change in y for a one-unit change in x
- The **partial effect** of x on y, holding other factors constant
- Under the zero conditional mean assumption, this has a **causal interpretation**

Formula

$$\beta_1 = \frac{\partial \mathbb{E}[y|x]}{\partial x} \tag{3}$$

2.4 Meaning of Linearity

The model is **linear in parameters**, meaning the parameters β_0 and β_1 enter the equation linearly. However, we can have non-linear relationships in variables:

Example 2.1. Examples of models linear in parameters but non-linear in variables:

$$y = \beta_0 + \beta_1 x^2 + u \quad \text{(quadratic)} \tag{4}$$

$$y = \beta_0 + \beta_1 \log(x) + u \quad \text{(logarithmic)} \tag{5}$$

2.5 Non-linearities

Common non-linear functional forms include:

- **Quadratic models**: $y = \beta_0 + \beta_1 x + \beta_2 x^2 + u$
- **Log models**:
  - $\log(y) = \beta_0 + \beta_1 x + u \Rightarrow$ semi-elasticity
  - $\log(y) = \beta_0 + \beta_1 \log(x) + u \Rightarrow$ elasticity
- **Interaction effects**: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + u$

2.6 Changing Units

Key Point

When changing units of measurement:

- Scaling y by factor c: new slope $= c\times$ old slope
- Scaling x by factor c: new slope $=$ old slope $/c$
- Standardizing variables (using standard deviations) can aid interpretation

2.7 OLS Estimator

2.7.1 Sample Mean as an LS Estimator

The sample mean $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ minimizes the sum of squared deviations:

$$\sum_{i=1}^{n}(y_i - c)^2 \tag{6}$$

This principle extends to regression analysis through the method of least squares.

2.7.2 Residuals and Fitted Values

Given OLS estimates $\hat{\beta}_0$ and $\hat{\beta}_1$:

- **Fitted values**: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$
- **Residuals**: $\hat{u}_i = y_i - \hat{y}_i$

Key Point

Residuals represent the unexplained portion of y after fitting the regression line.

2.7.3 Minimizing the SSR

The **Sum of Squared Residuals** is:

$$\text{SSR} = \sum_{i=1}^{n}\hat{u}_i^2 = \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \tag{7}$$

The first-order conditions for minimization yield:

$$\sum_{i=1}^{n}\hat{u}_i = 0 \tag{8}$$

$$\sum_{i=1}^{n}x_i\hat{u}_i = 0 \tag{9}$$

Formula

The OLS estimators are:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{10}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} \tag{11}$$

2.7.4 Derivation of OLS Estimators

Theorem 2.1 (OLS Estimator Derivation)**.** To minimize SSR $= \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$, we take derivatives with respect to β_0 and β_1 and set them equal to zero.

First-order condition for β_0:

$$\frac{\partial \text{SSR}}{\partial \beta_0} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) = 0 \tag{12}$$

$$\sum_{i=1}^{n} y_i - n\beta_0 - \beta_1 \sum_{i=1}^{n} x_i = 0 \tag{13}$$

$$\hat{\beta}_0 = \frac{1}{n}\sum_{i=1}^{n} y_i - \hat{\beta}_1 \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{y} - \hat{\beta}_1 \bar{x} \tag{14}$$

First-order condition for β_1:

$$\frac{\partial \text{SSR}}{\partial \beta_1} = -2\sum_{i=1}^{n} x_i(y_i - \beta_0 - \beta_1 x_i) = 0 \tag{15}$$

$$\sum_{i=1}^{n} x_i y_i - \beta_0 \sum_{i=1}^{n} x_i - \beta_1 \sum_{i=1}^{n} x_i^2 = 0 \tag{16}$$

Substituting $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$:

$$\sum_{i=1}^{n} x_i y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) \sum_{i=1}^{n} x_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0 \tag{17}$$

$$\sum_{i=1}^{n} x_i y_i - \bar{y} \sum_{i=1}^{n} x_i + \hat{\beta}_1 \bar{x} \sum_{i=1}^{n} x_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0 \tag{18}$$

$$\hat{\beta}_1 \left(\bar{x} \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i^2 \right) = \bar{y} \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i y_i \tag{19}$$

Since $\sum_{i=1}^{n} x_i = n\bar{x}$:

$$\hat{\beta}_1 \left(n\bar{x}^2 - \sum_{i=1}^{n} x_i^2 \right) = n\bar{x}\bar{y} - \sum_{i=1}^{n} x_i y_i \tag{20}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{21}$$

2.7.5 Example: Binary Regressor and Causal Effects

When $x \in \{0, 1\}$ (treatment/control):

$$\hat{\beta}_1 = \bar{y}_1 - \bar{y}_0 \tag{22}$$

This connects regression analysis to experimental thinking: the slope coefficient equals the difference in sample means between treatment and control groups.

2.7.6 Worked Numerical Example

Example 2.2 (Simple Wage-Education Regression)**.** Consider data on 5 workers with years of education (x) and hourly wage in dollars (y):

Worker	Education (x_i)	Wage (y_i)
1	12	15
2	14	18
3	16	22
4	18	25
5	20	30

Step 1: Calculate sample means

$$\bar{x} = \frac{12 + 14 + 16 + 18 + 20}{5} = 16 \tag{23}$$

$$\bar{y} = \frac{15 + 18 + 22 + 25 + 30}{5} = 22 \tag{24}$$

Step 2: Calculate numerator and denominator

$$\sum_{i=1}^{5}(x_i - \bar{x})(y_i - \bar{y}) = (-4)(-7) + (-2)(-4) + (0)(0) + (2)(3) + (4)(8) \tag{25}$$

$$= 28 + 8 + 0 + 6 + 32 = 74 \tag{26}$$

$$\sum_{i=1}^{5}(x_i - \bar{x})^2 = (-4)^2 + (-2)^2 + 0^2 + 2^2 + 4^2 \tag{27}$$

$$= 16 + 4 + 0 + 4 + 16 = 40 \tag{28}$$

Step 3: Calculate OLS estimates

$$\hat{\beta}_1 = \frac{74}{40} = 1.85 \tag{29}$$

$$\hat{\beta}_0 = 22 - 1.85 \times 16 = 22 - 29.6 = -7.6 \tag{30}$$

Interpretation: Each additional year of education is associated with a $1.85 increase in hourly wage.

Fitted regression: $\hat{y} = -7.6 + 1.85x$

2.7.7 Algebraic Properties of OLS

The OLS estimator has several important algebraic properties:

1. Residuals sum to zero: $\sum_{i=1}^{n} \hat{u}_i = 0$
2. Sample covariance between x and residuals is zero: $\sum_{i=1}^{n} x_i \hat{u}_i = 0$
3. The regression line passes through the sample means: $(\bar{x}, \bar{y})$
4. Fitted values and residuals are uncorrelated

2.8 ANOVA Decomposition

The total variation in y can be decomposed as:

Formula

$$\underbrace{\sum_{i=1}^{n}(y_i - \bar{y})^2}_{\text{TSS}} = \underbrace{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}_{\text{ESS}} + \underbrace{\sum_{i=1}^{n}\hat{u}_i^2}_{\text{RSS}} \tag{31}$$

Where:

- TSS: Total Sum of Squares
- ESS: Explained Sum of Squares
- RSS: Residual Sum of Squares

2.9 Coefficient of Determination

The **R-squared** measures the fraction of variance in y explained by x:

Formula

$$R^2 = \frac{\text{ESS}}{\text{TSS}} = 1 - \frac{\text{RSS}}{\text{TSS}} \tag{32}$$

Properties of R^2:

- Range: $[0, 1]$
- Higher R^2 indicates better fit
- $R^2 = 1$: perfect fit
- $R^2 = 0$: no linear relationship

Warning

A high R^2 does not necessarily imply good causal inference. Correlation does not imply causation.

2.10 Sampling Moments of the OLS Estimator

2.10.1 Desired Sampling Properties

Good estimators should have:

- **Unbiasedness**: $\mathbb{E}[\hat{\beta}] = \beta$
- **Efficiency**: Minimum variance among unbiased estimators
- **Consistency**: $\hat{\beta} \to \beta$ as $n \to \infty$

2.10.2 Gauss-Markov Assumptions

Assumption

GM1 - Linear in parameters: The model can be written as $y = \beta_0 + \beta_1 x + u$

Assumption

GM2 - Random sampling: $\{(x_i, y_i) : i = 1, \ldots, n\}$ is a random sample from the population

Assumption

GM3 - Sample variation in the explanatory variable: $\text{Var}(x) > 0$ in the sample

Assumption

GM4 - Zero conditional mean: $\mathbb{E}[u|x] = 0$

Assumption

GM5 - Homoskedasticity: $\text{Var}(u|x) = \sigma^2$ (constant variance)

2.10.3 Mean and Variance of OLS Slope Estimator

Under assumptions GM1-GM4:

$$\mathbb{E}[\hat{\beta}_1|x] = \beta_1 \quad \text{(unbiasedness)} \tag{33}$$

Theorem 2.2 (Proof of OLS Unbiasedness)**. Proof:** Starting with the OLS estimator:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \tag{34}$$

Substitute the true model $y_i = \beta_0 + \beta_1 x_i + u_i$:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^n (x_i - \bar{x})[(\beta_0 + \beta_1 x_i + u_i) - (\beta_0 + \beta_1 \bar{x} + \bar{u})]}{\sum_{i=1}^n (x_i - \bar{x})^2} && (35)\\
&= \frac{\sum_{i=1}^n (x_i - \bar{x})[\beta_1 (x_i - \bar{x}) + (u_i - \bar{u})]}{\sum_{i=1}^n (x_i - \bar{x})^2} && (36)\\
&= \frac{\beta_1 \sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^n (x_i - \bar{x})(u_i - \bar{u})}{\sum_{i=1}^n (x_i - \bar{x})^2} && (37)\\
&= \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x})(u_i - \bar{u})}{\sum_{i=1}^n (x_i - \bar{x})^2} && (38)
\end{aligned}$$

Since $\sum_{i=1}^n (x_i - \bar{x})(u_i - \bar{u}) = \sum_{i=1}^n (x_i - \bar{x}) u_i$ (because $\sum_{i=1}^n (x_i - \bar{x}) = 0$):

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x}) u_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \tag{39}$$

Taking expectations conditional on $\boldsymbol{x}$:

$$\mathbb{E}[\hat{\beta}_1|\boldsymbol{x}] = \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x}) \mathbb{E}[u_i|\boldsymbol{x}]}{\sum_{i=1}^n (x_i - \bar{x})^2} \tag{40}$$

Under assumption GM4 ($\mathbb{E}[u_i|\boldsymbol{x}] = 0$):

$$\mathbb{E}[\hat{\beta}_1|\boldsymbol{x}] = \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot 0}{\sum_{i=1}^n (x_i - \bar{x})^2} = \beta_1 \tag{41}$$

Therefore, $\hat{\beta}_1$ is unbiased for β_1.

Under assumptions GM1-GM5:

$$\text{Var}(\hat{\beta}_1|x) = \frac{\sigma^2}{\text{SST}_x} \tag{42}$$

where $\text{SST}_x = \sum_{i=1}^n (x_i - \bar{x})^2$.

Theorem 2.3 (Variance of OLS Estimator)**. Proof:** From the unbiasedness proof, we have:

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x})u_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \tag{43}$$

Let $w_i = \frac{x_i - \bar{x}}{\sum_{j=1}^n (x_j - \bar{x})^2}$, so that:

$$\hat{\beta}_1 = \beta_1 + \sum_{i=1}^n w_i u_i \tag{44}$$

The variance is:

$$\begin{aligned}
\text{Var}(\hat{\beta}_1|\boldsymbol{x}) &= \text{Var}\left(\sum_{i=1}^n w_i u_i \middle| \boldsymbol{x}\right) && (45)\\
&= \sum_{i=1}^n w_i^2 \text{Var}(u_i|\boldsymbol{x}) \quad \text{(since errors are uncorrelated)} && (46)\\
&= \sum_{i=1}^n w_i^2 \sigma^2 \quad \text{(under GM5)} && (47)\\
&= \sigma^2 \sum_{i=1}^n w_i^2 && (48)\\
&= \sigma^2 \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sum_{j=1}^n (x_j - \bar{x})^2}\right)^2 && (49)\\
&= \sigma^2 \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{[\sum_{i=1}^n (x_i - \bar{x})^2]^2} && (50)\\
&= \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sigma^2}{\text{SST}_x} && (51)
\end{aligned}$$

2.10.4 The Gauss-Markov Theorem

Theorem 2.4 (Gauss-Markov Theorem)**.** Under assumptions GM1-GM5, the OLS estimator is BLUE (Best Linear Unbiased Estimator).

Proof Sketch: Consider any other linear unbiased estimator $\tilde{\beta}_1 = \sum_{i=1}^n c_i y_i$ where c_i are constants.

For unbiasedness, we need:

$$\begin{aligned}
\mathbb{E}[\tilde{\beta}_1] = \mathbb{E}\left[\sum_{i=1}^n c_i y_i\right] &= \sum_{i=1}^n c_i \mathbb{E}[y_i] && (52)\\
&= \sum_{i=1}^n c_i(\beta_0 + \beta_1 x_i) && (53)\\
&= \beta_0 \sum_{i=1}^n c_i + \beta_1 \sum_{i=1}^n c_i x_i && (54)
\end{aligned}$$

For this to equal β_1 for all values of β_0 and β_1, we need:

$$\sum_{i=1}^{n} c_i = 0 \quad \text{(coefficient of } \beta_0\text{)} \tag{55}$$

$$\sum_{i=1}^{n} c_i x_i = 1 \quad \text{(coefficient of } \beta_1\text{)} \tag{56}$$

The variance of any such estimator is:

$$\text{Var}(\tilde{\beta}_1) = \sigma^2 \sum_{i=1}^{n} c_i^2 \tag{57}$$

Using the method of Lagrange multipliers to minimize $\sum_{i=1}^{n} c_i^2$ subject to the constraints above, we can show that the optimal choice is:

$$c_i = \frac{x_i - \bar{x}}{\sum_{j=1}^{n}(x_j - \bar{x})^2} \tag{58}$$

This is exactly the weight that y_i receives in the OLS estimator, proving that OLS is BLUE.

This means OLS has the smallest variance among all linear unbiased estimators.

2.10.5 Unbiased Estimator of Error Variance

Formula

$$\hat{\sigma}^2 = \frac{\text{SSR}}{n-2} \tag{59}$$

Under the GM assumptions: $\mathbb{E}[\hat{\sigma}^2] = \sigma^2$
The degrees of freedom adjustment $(n-2)$ accounts for estimating two parameters.

2.11 The Gaussian Bivariate Regression Model

2.11.1 Sampling Distributions

Adding the normality assumption:

Assumption

GM6 - Normality: $u \sim \mathcal{N}(0, \sigma^2)$

Under GM1-GM6:

$$\hat{\beta}_1 \sim \mathcal{N}\left(\beta_1, \frac{\sigma^2}{\text{SST}_x}\right) \tag{60}$$

$$\hat{\beta}_0 \sim \mathcal{N}\left(\beta_0, \sigma^2\left[\frac{1}{n} + \frac{\bar{x}^2}{\text{SST}_x}\right]\right) \tag{61}$$

2.11.2 Hypothesis Testing and Confidence Intervals

The t-statistic for testing $H_0 : \beta_1 = 0$:

$$t = \frac{\hat{\beta}_1}{\text{se}(\hat{\beta}_1)} \tag{62}$$

where $\text{se}(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{\text{SST}_x}}$.

Under H_0: $t \sim t_{n-2}$

Formula

A $(1 - \alpha)100\%$ confidence interval for β_1:

$$\hat{\beta}_1 \pm t_{n-2,\alpha/2} \times \text{se}(\hat{\beta}_1) \tag{63}$$

2.11.3 Economic vs Statistical Significance

Key Point

- **Statistical significance**: Ability to reject H_0 at a given significance level
- **Economic significance**: Whether the magnitude is important for policy or theory

Warning

Large samples can make economically small effects statistically significant. Always consider both statistical and economic significance.

3 Inference Without Normality

3.1 Large Sample Approximations

For large samples, we don't need the normality assumption. The Central Limit Theorem provides the foundation for asymptotic inference.

3.2 Convergence in Probability and Consistency

Definition 3.1. An estimator $\hat{\theta}_n$ is **consistent** for θ if:

$$\text{plim}_{n\to\infty}\hat{\theta}_n = \theta \tag{64}$$

Consistency is weaker than unbiasedness but ensures the estimator converges to the true value as the sample size increases.

3.3 Central Limit Theorem

Theorem 3.1 (Central Limit Theorem for OLS)**.** Under appropriate regularity conditions:

$$\frac{\hat{\beta}_1 - \beta_1}{\text{se}(\hat{\beta}_1)} \xrightarrow{d} \mathcal{N}(0, 1) \tag{65}$$

This allows us to use normal critical values instead of t-distribution for large samples.

3.4 Inference in the Bivariate Model

For large samples:

- Use standard normal critical values
- z-test instead of t-test
- Practically similar to t-test for $n > 30$

4 A Primer on Instrumental Variables

4.1 The Omitted Variable Problem

When $\mathbb{E}[u|x] \neq 0$, OLS is biased. Common causes include:

- Omitted variables
- Measurement error
- Simultaneity

Formula

The bias in OLS when an important variable is omitted:

$$\mathbb{E}[\hat{\beta}_1] = \beta_1 + \beta_2 \frac{\text{Cov}(x_1, x_2)}{\text{Var}(x_1)} \tag{66}$$

where x_2 is the omitted variable.

4.2 Causal Inference Without Random Assignment

Instrumental Variables (IV) provide a solution when $\mathbb{E}[u|x] \neq 0$.

An instrument z must satisfy:

1. **Relevance**: $\text{Cov}(z, x) \neq 0$
2. **Exogeneity**: $\text{Cov}(z, u) = 0$

4.3 The IV Estimator (Single Instrument)

Formula

The IV estimator with a single instrument:

$$\hat{\beta}_1^{IV} = \frac{\text{Cov}(z, y)}{\text{Cov}(z, x)} = \frac{\sum_{i=1}^n (z_i - \bar{z})(y_i - \bar{y})}{\sum_{i=1}^n (z_i - \bar{z})(x_i - \bar{x})} \tag{67}$$

Theorem 4.1 (IV Estimator Derivation)**. Intuition**: Use the instrument z to extract the exogenous variation in x.

Method of Moments Approach: From the population moment condition $\mathbb{E}[z(y - \beta_0 - \beta_1 x)] = 0$, we get:

$$\mathbb{E}[zy] - \beta_0 \mathbb{E}[z] - \beta_1 \mathbb{E}[zx] = 0 \tag{68}$$

$$\mathbb{E}[zy] - \beta_0 \mathbb{E}[z] = \beta_1 \mathbb{E}[zx] \tag{69}$$

Similarly, from $\mathbb{E}[y - \beta_0 - \beta_1 x] = 0$:

$$\mathbb{E}[y] - \beta_0 - \beta_1\mathbb{E}[x] = 0 \Rightarrow \beta_0 = \mathbb{E}[y] - \beta_1\mathbb{E}[x] \tag{70}$$

Substituting:

$$\mathbb{E}[zy] - (\mathbb{E}[y] - \beta_1\mathbb{E}[x])\mathbb{E}[z] = \beta_1\mathbb{E}[zx] \tag{71}$$
$$\mathbb{E}[zy] - \mathbb{E}[y]\mathbb{E}[z] + \beta_1\mathbb{E}[x]\mathbb{E}[z] = \beta_1\mathbb{E}[zx] \tag{72}$$
$$\text{Cov}(z, y) = \beta_1\text{Cov}(z, x) \tag{73}$$

Therefore: $\beta_1 = \frac{\text{Cov}(z,y)}{\text{Cov}(z,x)}$
The sample analog gives the IV estimator:

$$\hat{\beta}_1^{IV} = \frac{\widehat{\text{Cov}}(z, y)}{\widehat{\text{Cov}}(z, x)} \tag{74}$$

4.3.1 Two-Stage Interpretation

IV estimation can be implemented in two stages:

1. **First stage**: Regress x on z, obtain fitted values $\hat{x}$
2. **Second stage**: Regress y on $\hat{x}$

4.4 Statistical Inference

IV estimators generally have:

- Larger standard errors than OLS
- Asymptotic normality under regularity conditions
- Need for larger samples for reliable inference

4.5 Poor Instruments: Cautionary Note

Warning

Weak instruments (small correlation with endogenous variable) can cause:

- Bias toward OLS in finite samples
- Very large standard errors
- Poor finite-sample properties

Strong economic theory is essential for instrument validity.

5 Multiple Regression Analysis

5.1 The Model and Motivation

5.1.1 Omitted Variables Motivation

The multiple regression model allows us to control for confounding factors:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u \tag{75}$$

5.1.2 Interpretation of Regression Coefficients

Each coefficient β_j represents the **partial effect** of x_j on y, holding all other variables constant (*ceteris paribus*).

Key Point

$\beta_j = \frac{\partial \mathbb{E}[y|x_1,\ldots,x_k]}{\partial x_j}$

5.2 Obtaining OLS Estimators

5.2.1 Matrix Form

In matrix notation:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u} \tag{76}$$

The OLS estimator is:

Formula

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \tag{77}$$

5.2.2 Orthogonality Conditions

The first-order conditions require:

$$\boldsymbol{X}'\hat{\boldsymbol{u}} = \boldsymbol{0} \tag{78}$$

This means residuals are orthogonal to all regressors.

5.3 Goodness-of-Fit

5.3.1 R-squared

The same ANOVA decomposition applies:

$$R^2 = \frac{\text{ESS}}{\text{TSS}} = 1 - \frac{\text{RSS}}{\text{TSS}} \tag{79}$$

5.3.2 Adjusted R-squared

Formula

$$\bar{R}^2 = 1 - \frac{\text{RSS}/(n-k-1)}{\text{TSS}/(n-1)} \tag{80}$$

Adjusted R-squared penalizes additional regressors and can decrease when adding irrelevant variables.

5.4 "Partialling-Out" Interpretation

Theorem 5.1 (Frisch-Waugh-Lovell Theorem)**.** The coefficient $\hat{\beta}_1$ from the full regression equals the coefficient from regressing $\tilde{y}$ on $\tilde{x}_1$, where $\tilde{y}$ and $\tilde{x}_1$ are residuals from regressions on the other variables.

This provides intuition: multiple regression isolates the variation in y and x_1 that is uncorrelated with other regressors.

5.5 Gauss-Markov Assumptions (Multiple Regression)

Assumption

MLR.1: Linear in parameters

Assumption

MLR.2: Random sampling

Assumption

MLR.3: No perfect collinearity

Assumption

MLR.4: Zero conditional mean: $\mathbb{E}[u|x_1, \ldots, x_k] = 0$

Assumption

MLR.5: Homoskedasticity: $\text{Var}(u|x_1, \ldots, x_k) = \sigma^2$

5.6 Sampling Properties

Under MLR.1-MLR.4: $\mathbb{E}[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$ (unbiasedness)
Under MLR.1-MLR.5: $\text{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$

5.7 Omitted Variable Bias

Key Point

Including relevant control variables can eliminate omitted variable bias if the mean independence assumption $\mathbb{E}[u|x_1, \ldots, x_k] = 0$ holds.

5.8 Prediction

Point predictions: $\hat{y}_0 = \boldsymbol{x}_0'\hat{\boldsymbol{\beta}}$
Prediction intervals must account for:

- Uncertainty in $\hat{\boldsymbol{\beta}}$
- Future error term

6 Inference

6.1 t-statistics and z-statistics

For individual coefficients:

$$t = \frac{\hat{\beta}_j - \beta_{j0}}{\text{se}(\hat{\beta}_j)} \tag{81}$$

Use t-distribution for small samples with normality, standard normal for large samples.

6.2 Joint Hypothesis Testing

Warning

Cannot use individual t-tests for joint hypotheses due to multiple testing problems.

6.3 F-statistic

For testing q linear restrictions:

Formula

$$F = \frac{(\text{SSR}_r - \text{SSR}_{ur})/q}{\text{SSR}_{ur}/(n-k-1)} \tag{82}$$

where subscripts r and ur denote restricted and unrestricted models.

Under H_0: $F \sim F_{q,n-k-1}$

6.4 Overall Significance Test

Testing $H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$:

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} \tag{83}$$

7 Topics in Multiple Regression

7.1 Qualitative Information

7.1.1 Dummy Variables

Binary variables $D \in \{0, 1\}$ are **intercept shifters**:

$$y = \beta_0 + \beta_1 x + \beta_2 D + u \tag{84}$$

β_2 measures the difference in intercepts between the two groups.

7.1.2 Categorical Variables

For m categories, use $m - 1$ dummy variables to avoid the **dummy variable trap** (perfect collinearity).

7.1.3 Interactions

Interaction between dummy and continuous variable:

$$y = \beta_0 + \beta_1 x + \beta_2 D + \beta_3 (D \times x) + u \tag{85}$$

β_3 measures the difference in slopes between groups.

7.2 Linear Probability Model

For binary dependent variable $y \in \{0, 1\}$:

$$\Pr(y = 1|x) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k \tag{86}$$

Advantages:

- Simple interpretation: β_j = change in probability
- Good approximation for probabilities not too close to 0 or 1

Disadvantages:

- Predicted probabilities can be outside $[0, 1]$
- Inherent heteroskedasticity

7.3 Two-Stage Least Squares

When one regressor is endogenous in multiple regression:

First stage: Regress endogenous variable on instruments and exogenous variables **Second stage**: Regress y on fitted values and exogenous variables

8 Heteroskedasticity

8.1 Consequences

When $\text{Var}(u|x) \neq \sigma^2$:

- OLS remains unbiased under MLR.1-MLR.4
- OLS is no longer efficient (not BLUE)
- Standard errors are incorrect $\Rightarrow$ invalid inference

8.2 Robust Standard Errors

Heteroskedasticity-consistent (White) standard errors:

$$\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}'\boldsymbol{X})^{-1} \left(\sum_{i=1}^{n} \hat{u}_i^2 \boldsymbol{x}_i \boldsymbol{x}_i' \right) (\boldsymbol{X}'\boldsymbol{X})^{-1} \tag{87}$$

```r
# R implementation
library(sandwich)
coeftest(model, vcov = vcovHC)
```

8.3 Testing for Heteroskedasticity

8.3.1 Breusch-Pagan Test

Null hypothesis: $H_0 : \text{Var}(u_i|\boldsymbol{x}_i) = \sigma^2$ (homoskedasticity)
Alternative: $\text{Var}(u_i|\boldsymbol{x}_i) = \sigma^2 h(\boldsymbol{x}_i'\boldsymbol{\gamma})$ where $h(\cdot) > 0$
Test procedure:

1. Run OLS regression: $y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + u_i$, obtain residuals $\hat{u}_i$
2. Compute $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n \hat{u}_i^2$
3. Run auxiliary regression: $\frac{\hat{u}_i^2}{\hat{\sigma}^2} = 1 + \boldsymbol{z}_i'\boldsymbol{\gamma} + v_i$
4. Compute $LM = nR^2$ from auxiliary regression
5. Under H_0: $LM \sim \chi^2_q$ where q is number of variables in $\boldsymbol{z}_i$

```r
# Breusch-Pagan test in R
library(lmtest)
bptest(model)  # Default uses fitted values
bptest(model, ~ x1 + x2, data = df)  # Specify variables
```

8.3.2 White Test

More general test that doesn't assume specific functional form for heteroskedasticity.
Test procedure:

1. Run OLS: $y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + u_i$, obtain $\hat{u}_i$
2. Run auxiliary regression: $\hat{u}_i^2 = \boldsymbol{z}_i'\boldsymbol{\gamma} + v_i$
3. Where $\boldsymbol{z}_i$ includes: constants, $x_{1i}, x_{2i}, \ldots, x_{1i}^2, x_{2i}^2, \ldots, x_{1i}x_{2i}, \ldots$
4. $LM = nR^2 \sim \chi^2_q$

Key Point

White test variations:

- **White test**: Include all regressors, squares, and cross-products
- **White test (fitted values)**: $\hat{u}_i^2 = \gamma_0 + \gamma_1\hat{y}_i + \gamma_2\hat{y}_i^2 + v_i$

8.3.3 Goldfeld-Quandt Test

For specific alternative of increasing variance:

1. Order observations by suspected variable causing heteroskedasticity
2. Drop middle c observations, split into two groups of size n_1 and n_2
3. Run separate regressions, compute SSR_1 and SSR_2
4. Test statistic: $GQ = \frac{\text{SSR}_2/(n_2-k)}{\text{SSR}_1/(n_1-k)} \sim F_{n_2-k,n_1-k}$

8.4 Weighted Least Squares

When the form of heteroskedasticity is known:

$$\text{Var}(u_i|x_i) = \sigma^2 h(x_i) \tag{88}$$

Weight observations by $w_i = 1/\sqrt{h(x_i)}$ to achieve efficiency.

9 Binary Dependent Variables

9.1 Latent Variable Model

$$y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + u_i \tag{89}$$

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{if } y_i^* \leq 0 \end{cases} \tag{90}$$

9.2 Probit and Logit Models

Probit: $\Pr(y = 1|\boldsymbol{x}) = \Phi(\boldsymbol{x}'\boldsymbol{\beta})$ where Φ is the standard normal CDF

Logit: $\Pr(y = 1|\boldsymbol{x}) = \frac{\exp(\boldsymbol{x}'\boldsymbol{\beta})}{1+\exp(\boldsymbol{x}'\boldsymbol{\beta})}$

9.3 Maximum Likelihood Estimation

9.3.1 Likelihood Function

For a sample of n observations, the likelihood function is:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} [\Pr(y_i = 1|\boldsymbol{x}_i)]^{y_i} [1 - \Pr(y_i = 1|\boldsymbol{x}_i)]^{1-y_i} \tag{91}$$

For Probit:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} [\Phi(\boldsymbol{x}_i'\boldsymbol{\beta})]^{y_i} [1 - \Phi(\boldsymbol{x}_i'\boldsymbol{\beta})]^{1-y_i} \tag{92}$$

For Logit:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left[\frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})}{1 + \exp(\boldsymbol{x}_i'\boldsymbol{\beta})}\right]^{y_i} \left[\frac{1}{1 + \exp(\boldsymbol{x}_i'\boldsymbol{\beta})}\right]^{1-y_i} \tag{93}$$

9.3.2 Log-Likelihood Function

The log-likelihood is easier to work with:

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} [y_i \ln[\Pr(y_i = 1|\boldsymbol{x}_i)] + (1 - y_i) \ln[1 - \Pr(y_i = 1|\boldsymbol{x}_i)]] \tag{94}$$

For Logit (particularly simple):

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} [y_i \boldsymbol{x}_i'\boldsymbol{\beta} - \ln(1 + \exp(\boldsymbol{x}_i'\boldsymbol{\beta}))] \tag{95}$$

9.3.3 Maximum Likelihood Estimator

The ML estimator $\hat{\boldsymbol{\beta}}_{ML}$ maximizes the log-likelihood:

$$\hat{\boldsymbol{\beta}}_{ML} = \arg\max_{\boldsymbol{\beta}} \ell(\boldsymbol{\beta}) \tag{96}$$

First-order conditions:

$$\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \boldsymbol{x}_i [y_i - \Pr(y_i = 1 | \boldsymbol{x}_i)] = \mathbf{0} \tag{97}$$

Since these are nonlinear, numerical methods (Newton-Raphson) are required.

9.3.4 Asymptotic Properties of ML Estimators

Under regularity conditions:

1. **Consistency**: $\hat{\boldsymbol{\beta}}_{ML} \xrightarrow{p} \boldsymbol{\beta}$
2. **Asymptotic Normality**: $\sqrt{n}(\hat{\boldsymbol{\beta}}_{ML} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathcal{I}^{-1})$
3. **Asymptotic Efficiency**: ML estimators achieve the Cramér-Rao lower bound

where $\mathcal{I}$ is the Fisher Information Matrix.

9.3.5 Information Matrix and Standard Errors

The Fisher Information Matrix is:

$$\mathcal{I}(\boldsymbol{\beta}) = -\mathbb{E}\left[\frac{\partial^2 \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\right] \tag{98}$$

In practice, we use the sample information matrix:

$$\hat{\mathcal{I}} = -\frac{\partial^2 \ell(\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \tag{99}$$

The asymptotic variance-covariance matrix is:

$$\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}_{ML}) = \hat{\mathcal{I}}^{-1} \tag{100}$$

9.4 Hypothesis Testing in Binary Choice Models

9.4.1 Wald Test

For testing $H_0 : R\boldsymbol{\beta} = \boldsymbol{r}$ where R is $q \times k$:

$$W = (R\hat{\boldsymbol{\beta}} - \boldsymbol{r})'[R\hat{\mathcal{I}}^{-1}R']^{-1}(R\hat{\boldsymbol{\beta}} - \boldsymbol{r}) \sim \chi^2_q \tag{101}$$

For single coefficient: $W = \left(\frac{\hat{\beta}_j}{\text{se}(\hat{\beta}_j)}\right)^2 \sim \chi^2_1$

9.4.2 Likelihood Ratio Test

For nested models:

$$LR = 2[\ell(\hat{\boldsymbol{\beta}}_{ur}) - \ell(\hat{\boldsymbol{\beta}}_r)] \sim \chi^2_q \tag{102}$$

where q is the number of restrictions.

9.4.3 Score Test (Lagrange Multiplier)

Based on the score function evaluated at the restricted estimator:

$$LM = \left[\frac{\partial \ell(\hat{\boldsymbol{\beta}}_r)}{\partial \boldsymbol{\beta}}\right]' \hat{\mathcal{I}}_r^{-1} \left[\frac{\partial \ell(\hat{\boldsymbol{\beta}}_r)}{\partial \boldsymbol{\beta}}\right] \sim \chi^2_q \tag{103}$$

Key Point

All three tests (Wald, LR, Score) are asymptotically equivalent but may differ in finite samples:

- **Wald**: Uses unrestricted estimates only
- **LR**: Uses both restricted and unrestricted estimates
- **Score**: Uses restricted estimates only

9.5 Partial Effects

Marginal effects: $\frac{\partial \Pr(y=1|\boldsymbol{x})}{\partial x_j}$
Probit: $\beta_j \phi(\boldsymbol{x}'\boldsymbol{\beta})$ where ϕ is the standard normal PDF
Logit: $\beta_j \Lambda(\boldsymbol{x}'\boldsymbol{\beta})(1 - \Lambda(\boldsymbol{x}'\boldsymbol{\beta}))$

9.5.1 Average Partial Effects vs Partial Effects at Average

- **APE**: $\frac{1}{n}\sum_{i=1}^{n} \frac{\partial \Pr(y=1|\boldsymbol{x}_i)}{\partial x_j}$
- **PEA**: $\frac{\partial \Pr(y=1|\bar{\boldsymbol{x}})}{\partial x_j}$

```r
# Probit model in R
probit_model <- glm(y ~ x1 + x2, family = binomial(link = "probit"))

# Logit model in R
logit_model <- glm(y ~ x1 + x2, family = binomial(link = "logit"))
```

10 Time Series Regression

10.1 Time Series Context

Data: $\{(x_t, y_t) : t = 1, \ldots, T\}$
Key difference from cross-section: **temporal dependence**

10.2 Common Time Series Models

- **Static models**: $y_t = \beta_0 + \beta_1 x_t + u_t$
- **Distributed lag models**: Include past values of x
- **Autoregressive models**: Include past values of y

10.3 Models with Lagged Regressors

Autoregressive Distributed Lag (ADL) model:

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \gamma y_{t-1} + u_t \tag{104}$$

- **Short-run multiplier**: β_0
- **Long-run multiplier**: $\frac{\beta_0+\beta_1}{1-\gamma}$

10.4 Trending Variables

Warning

Spurious regression: When both x_t and y_t have trends, regression may show significant relationship even when none exists.

10.5 Stationarity and Weak Dependence

Definition 10.1. A time series is **covariance stationary** if:

1. $\mathbb{E}[x_t] = \mu$ (constant mean)
2. $\text{Var}(x_t) = \sigma^2$ (constant variance)
3. $\text{Cov}(x_t, x_{t-h})$ depends only on h, not t

Weak dependence: Correlations die out as lag length increases.

10.6 Serial Correlation

10.6.1 AR(1) Errors

$$u_t = \rho u_{t-1} + e_t \tag{105}$$

- $|\rho| < 1$: Stationary
- $\rho = 1$: Unit root (non-stationary)

10.6.2 MA(1) Errors

$$u_t = e_t + \theta e_{t-1} \tag{106}$$

Always stationary for finite θ.

10.7 OLS with Serial Correlation

Consequences:

- OLS still unbiased under appropriate assumptions
- Standard errors are incorrect
- Less efficient than GLS

Solution: Use HAC (Heteroskedasticity and Autocorrelation Consistent) standard errors.

10.8 Testing for Serial Correlation

10.8.1 Durbin-Watson Test

$$DW = \frac{\sum_{t=2}^{T}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{T}\hat{u}_t^2} \tag{107}$$

Limitations: Only tests AR(1), cannot include lagged dependent variables.

10.8.2 Breusch-Godfrey Test

More general LM test that:

- Can test higher-order autocorrelation
- Allows lagged dependent variables
- More flexible than DW test

11 Panel Data Models

11.1 Panel Data Structure

Data: $\{(x_{it}, y_{it}) : i = 1, \ldots, N, t = 1, \ldots, T\}$

- **Balanced panel**: All units observed all periods
- **Unbalanced panel**: Missing observations

11.2 Unobserved Effects Model

$$y_{it} = \beta_0 + \beta_1 x_{it} + a_i + u_{it} \tag{108}$$

where a_i is time-invariant unobserved heterogeneity.

11.3 First-Differences Estimator

Taking first differences eliminates a_i:

$$\Delta y_{it} = \beta_1 \Delta x_{it} + \Delta u_{it} \tag{109}$$

Key assumption: Strict exogeneity: $\mathbb{E}[u_{is}|\boldsymbol{x}_i] = 0$ for all s, t

11.4 Fixed Effects Estimator

Within transformation (time-demeaning):

$$\ddot{y}_{it} = y_{it} - \bar{y}_i = \beta_1(x_{it} - \bar{x}_i) + (u_{it} - \bar{u}_i) \tag{110}$$

Properties:

- Eliminates a_i like first differences
- More efficient when $T > 2$
- Cannot estimate time-invariant effects
- Consistent under strict exogeneity

11.5 Random Effects Model

11.5.1 Model Specification

Same basic model: $y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + a_i + u_{it}$

Key assumption: a_i is **uncorrelated** with $\boldsymbol{x}_{it}$ for all i, t.

Composite error: $v_{it} = a_i + u_{it}$

The model becomes: $y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + v_{it}$

11.5.2 Variance Structure

$$\text{Var}(v_{it}) = \sigma_a^2 + \sigma_u^2 \tag{111}$$

$$\text{Cov}(v_{it}, v_{is}) = \sigma_a^2 \quad \text{for } t \neq s \tag{112}$$

$$\text{Corr}(v_{it}, v_{is}) = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_u^2} = \rho \tag{113}$$

11.5.3 GLS Estimation

The efficient estimator is GLS, which gives:

$$\hat{\boldsymbol{\beta}}_{RE} = \hat{\boldsymbol{\beta}}_{FE} + \lambda(\hat{\boldsymbol{\beta}}_{BE} - \hat{\boldsymbol{\beta}}_{FE}) \tag{114}$$

where:

- $\hat{\boldsymbol{\beta}}_{FE}$ is the fixed effects estimator
- $\hat{\boldsymbol{\beta}}_{BE}$ is the between estimator
- $\lambda = 1 - \left(\frac{\sigma_u^2}{\sigma_u^2 + T\sigma_a^2}\right)^{1/2}$

11.6 Between Estimator

Uses only between-individual variation:

$$\bar{y}_i = \bar{\boldsymbol{x}}_i'\boldsymbol{\beta} + a_i + \bar{u}_i \tag{115}$$

The between estimator is OLS on the time-averaged data.

11.7 Hausman Test

Purpose: Test whether random effects assumption is valid.

Null hypothesis: $H_0 : \mathbb{E}[a_i|\boldsymbol{x}_i] = 0$ (random effects is consistent)

Alternative: $H_1 : \mathbb{E}[a_i|\boldsymbol{x}_i] \neq 0$ (only fixed effects is consistent)

Test statistic:

$$H = (\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE})'[\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}_{FE}) - \widehat{\text{Var}}(\hat{\boldsymbol{\beta}}_{RE})]^{-1}(\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE}) \tag{116}$$

Under H_0: $H \sim \chi_k^2$ where k is the number of time-varying regressors.

Key Point

Decision rule:

- If $H > \chi_{k,\alpha}^2$: Reject H_0, use fixed effects
- If $H \leq \chi_{k,\alpha}^2$: Fail to reject H_0, random effects is preferred (more efficient)

11.8 Unbalanced Panels

When T_i varies across individuals:

- **Fixed effects**: Still consistent, just use available observations
- **Random effects**: Need to modify GLS weights
- **Sample selection**: Be careful about non-random attrition

Missing data patterns:

- **MCAR**: Missing Completely at Random
- **MAR**: Missing at Random (conditional on observables)
- **MNAR**: Missing Not at Random (depends on unobservables)

```r
# Fixed effects in R
library(plm)
fe_model <- plm(y ~ x1 + x2, data = panel_data, model = "within")

# First differences in R
fd_model <- plm(y ~ x1 + x2, data = panel_data, model = "fd")
```

12 Key Formulas and Concepts

12.1 Important Relationships

Formula

$$R^2 = 1 - \frac{\text{SSR}}{\text{TSS}} \tag{117}$$

$$F\text{-statistic} = \frac{(R^2_{ur} - R^2_r)/q}{(1 - R^2_{ur})/(n-k-1)} \tag{118}$$

$$\text{Adjusted } R^2 = 1 - \frac{(1-R^2)(n-1)}{n-k-1} \tag{119}$$

12.2 Critical Assumptions

1. **Zero conditional mean**: $\mathbb{E}[u|x] = 0$
2. **Homoskedasticity**: $\text{Var}(u|x) = \sigma^2$
3. **No perfect collinearity**
4. **Random sampling**

12.3 When Assumptions Fail

- **Endogeneity** $\Rightarrow$ Use IV/2SLS
- **Heteroskedasticity** $\Rightarrow$ Use robust standard errors or WLS
- **Serial correlation** $\Rightarrow$ Use HAC standard errors or FGLS
- **Non-linearity** $\Rightarrow$ Transform variables or use non-linear models

13 R Commands Reference

13.1 Basic Regression

```
# Simple regression
model1 <- lm(y ~ x, data = df)

# Multiple regression
model2 <- lm(y ~ x1 + x2 + x3, data = df)

# With interactions
model3 <- lm(y ~ x1*x2, data = df)

# Polynomial terms
model4 <- lm(y ~ x + I(x^2), data = df)

# Log transformations
model5 <- lm(log(y) ~ log(x), data = df)
```

13.2 Robust Standard Errors

```
library(sandwich)
library(lmtest)

# Heteroskedasticity-consistent standard errors
coeftest(model, vcov = vcovHC)

# HAC standard errors (for time series)
coeftest(model, vcov = vcovHAC)
```

13.3 Instrumental Variables

```
library(AER)

# Two-stage least squares
iv_model <- ivreg(y ~ x1 + x2 | z1 + z2 + x2, data = df)

# First stage diagnostics
summary(iv_model, diagnostics = TRUE)
```

13.4 Binary Choice Models

```
# Probit model
probit_model <- glm(y ~ x1 + x2, family = binomial(link = "probit"))

# Logit model
logit_model <- glm(y ~ x1 + x2, family = binomial(link = "logit"))

# Marginal effects
library(margins)
margins(probit_model)
```

13.5 Panel Data

```r
library(plm)

# Convert to panel data format
panel_data <- pdata.frame(df, index = c("id", "time"))

# Fixed effects
fe_model <- plm(y ~ x1 + x2, data = panel_data, model = "within")

# Random effects
re_model <- plm(y ~ x1 + x2, data = panel_data, model = "random")

# First differences
fd_model <- plm(y ~ x1 + x2, data = panel_data, model = "fd")

# Hausman test
phtest(fe_model, re_model)
```

13.6 Time Series

```r
# Time series regression with trends
ts_model <- lm(y ~ x + time, data = ts_data)

# Lagged variables
library(dplyr)
ts_data <- ts_data %>%
  mutate(x_lag1 = lag(x, 1),
         y_lag1 = lag(y, 1))

# ADL model
adl_model <- lm(y ~ x + x_lag1 + y_lag1, data = ts_data)

# Durbin-Watson test
library(lmtest)
dwtest(model)

# Breusch-Godfrey test
bgtest(model)
```

14 Study Tips

1. **Focus on assumptions**: Understanding when methods work is crucial
2. **Practice interpretation**: Always think about the economic meaning of coefficients
3. **Master causality**: Distinguish between correlation and causation
4. **Learn R implementation**: Practice with real data
5. **Work through examples**: Apply concepts to empirical problems
6. **Understand limitations**: Know when methods fail and what alternatives exist
7. **Connect theory to practice**: Link mathematical results to economic intuition
8. **Practice problem solving**: Work through textbook exercises systematically

Key Point

Remember: The goal of econometrics is not just statistical modeling, but answering economic questions with data. Always keep the economic story in mind.