ECONOMETRIC TECHNIQUES

DISCUSSION CLASS 2

LECTURER: SENIA NHAMO
nhamos@unisa.ac.za
012 429 4093

Overview of the course

- What is econometrics? (Chapters 1-2)
- Statistics (Chapters 3-5, 16)
- Probability Distribution of $\hat{\beta}$
- Hypothesis testing
- Specification (chapters 6-7)
- Econometric problems (Chapters 8-11)

LEARNING AND TEACHING

- 2 Lectures and accompanying notes
- Office hours (phone to make an appointment)
  - 012 429 4093
- Text book: Studenmund, A.H (2007) Using Econometrics. A Practical Guide, Fifth Edition, Pearson
- MyUnisa
- Email : nhamos@unisa.ac.za

ASSESSMENT

- 2 Assignments (10% weighting)
- Examination (90% weighting)
- Assignment 1 is compulsory

EXAMINATION

- See TL201/2010
  - May/June 2010 exam paper
- 2 hours / 100 marks
  - Section A: Theory
    - Answer all 4 questions
    - 4 x 15 = 60
    - The questions correspond to a large extent to the questions at the end of each study unit

EXAMINATION (continued)

- Section B: Applications
  - Answer 2 of 3 questions
  - 2 x 20 = 40
    - Regression results to interpret and/or to evaluate
    - hypotheses testing
    - econometric problems / remedy
- Compile specifications
- any functional form eg Y=a+b.log(X)
- lags, intercept dummy variables, slope dummy variables
- Few calculations
- rather evaluate a set of regression results, identify errors, test coefficients for statistical significance etc

EXAMINATION (continued)

- ❑ No multiple-choice questions
- ❑ Formulae sheet will be provided.
  See p249-250 of study guide.
- ❑ Statistical tables are provided for statistical testing
- ❑ May use non-programmable calculator

UNISA college of economic and management sciences

CHAPTER 1

AN OVERVIEW OF REGRESSION ANALYSIS

UNISA college of economic and management sciences

What is Econometrics?

- Econometrics literally means **"economic measurement"**
- It is the **quantitative measurement** and **analysis** of actual economic and business phenomena—and so involves:
  - economic theory
  - Statistics
  - Math
  - observation/data collection

UNISA college of economic and management sciences

MAJOR USES OF ECONOMETRICS

- ❑ Describing economic reality
- ❑ Structural analysis
- ❑ Testing hypothesis about economic theory
- ❑ Policy evaluation
- ❑ *Forecasting future economic activity*

UNISA college of economic and management sciences

What is Regression Analysis?

- **Economic theory** can give us the **direction** of a change, e.g. the change in the demand for dvd's following a price decrease (or price increase)
- But what if we want to know not just **"how?"** but also **"how much?"**
- Then we need:
  - A sample of data
  - A way to estimate such a relationship
    - one of the most frequently ones used is **regression analysis**

UNISA college of economic and management sciences

What is Regression Analysis? (cont.)

- Formally, regression analysis is a **statistical technique** that attempts to "explain" movements in one variable, the **dependent** variable, as a function of movements in a set of other variables, the **independent** (or **explanatory**) variables, through the quantification of a single equation

UNISA college of economic and management sciences

Example

- Q = f(P, P_s, Yd) (1.1)
- Here, Q is the **dependent** variable and P, P_s, Yd are the **independent** variables
- Don't be deceived by the words **dependent** and **independent**, however
  - A statistically significant regression result does **not necessarily** imply **causality**
  - We also need:
    - Economic theory
    - Common sense

UNISA college of economic and management sciences

Single-Equation Linear Models

- The simplest example is:
  $Y = \beta_0 + \beta_1 X$ (1.2)
- The βs are denoted "**coefficients**"
  - β_0 is the "**constant**" or "**intercept**" term
  - β_1 is the "**slope coefficient**": the amount that Y will change when X increases by one unit; for a linear model, β_1 is constant over the entire function

UNISA college of economic and management sciences

Single-Equation Linear Models (cont.)

- Is (1.2) a complete description of origins of variation in Y?
- No, at least four sources of variation in Y other than the variation in the included Xs:
  - Other potentially important explanatory variables may be missing (e.g., X_2 and X_3)
  - Measurement error
  - Incorrect functional form
  - Purely random and totally unpredictable occurrences
- Inclusion of a "**stochastic error term**" (ε) effectively "takes care" of all these other sources of variation in Y that are NOT captured by X, so that (1.2) becomes:
  $Y = \beta_0 + \beta_1 X + \varepsilon$ (1.3)

UNISA college of economic and management sciences

Data types

- To estimate the parameters of interest, obtain the necessary data. Data source could involve *time series*, *cross-sectional* or *panel* data.
- *Time series* data are collected over time for the same country or other single aggregate economic unit
- *Cross-sectional* data are collected for a sample over individuals, households, firms or other disaggregate economic entity at a point in time
- *Panel* data contains elements of both time series and cross-sectional data

UNISA college of economic and management sciences

Indexing Conventions

- Subscript "i" for data on individuals (so called **"cross section"** data)
- Subscript "t" for **time series** data (e.g., series of years, months, or days—daily exchange rates, for example)
- Subscript "it" when we have **both** (for example, **"panel data"**)

UNISA college of economic and management sciences

The Estimated Regression Equation

- The regression equation considered so far is the **"true"—but unknown—theoretical** regression equation
- Instead of "true," might think about this as the **population** regression vs. the **sample/estimated** regression
- How do we obtain the empirical counterpart of the theoretical regression model (1.3)?
- It has to be **estimated**
- The empirical counterpart to (1.3) is:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \qquad (1.4)$$

- The signs on top of the estimates are denoted "hat," so that we have "Y-hat," for example

UNISA college of economic and management sciences

Key Terms from Chapter 1

- **Regression analysis**
- **Dependent variable**
- **Independent (or explanatory) variable(s)**
- **Causality vs correlation**
- **Linear**
- **Intercept term**
- **Slope coefficient**
- **Multivariate regression model**
- **Stochastic error term vs Residual term**
- **Time series vs cross sectional data**

UNISA college of economic and management sciences

CHAPTER 2

ORDINARY LEAST SQUARES

UNISA college of economic and management sciences

Estimating Single-Independent-Variable Models with OLS

- Recall that the objective of regression analysis is to start from:
- $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ (2.1)
- And, through the use of data, to get to:
- $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ (2.2)
- Recall that equation 2.1 is purely theoretical, while equation (2.2) is it empirical counterpart
- How to move from (2.1) to (2.2)?

UNISA college of economic and management sciences

Estimating Single-Independent-Variable Models with OLS (cont.)

- One of the most widely used methods is **Ordinary Least Squares** (OLS)
- OLS minimizes $\sum_{i=1}^{N} e_i^2$ (i = 1, 2, …., N) (2.3)
- Or, the sum of squared deviations of the **vertical distance** between the residuals (i.e. the estimated error terms) and the estimated regression line
- We also denote this term the "Residual Sum of Squares" (RSS)

UNISA college of economic and management sciences

Estimating Single-Independent-Variable Models with OLS (cont.)

- Similarly, OLS minimizes: $\sum_{i}^{N} (Y_i - \hat{Y}_i)^2$
- Why use OLS?
  - Relatively easy to use
  - The goal of minimizing RSS is intuitively / theoretically appealing
    - This basically says we want the estimated regression equation to be as close as possible to the observed data
  - OLS estimates have a number of useful characteristics

UNISA college of economic and management sciences

Estimating Single-Independent-Variable Models with OLS (cont.)

How does OLS work?

- First recall from (2.3) that OLS minimizes the sum of the squared residuals
- Next, it can be shown that the coefficients that ensure that for the case of just one independent variable are:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N} [(X_i - \bar{X})(Y_i - \bar{Y})]}{\sum_{i=1}^{N} (X_i - \bar{X})^2} \quad (2.4)$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \quad (2.5)$$

UNISA college of economic and management sciences

Estimating Multivariate Regression Models with OLS

- In the “real world” one explanatory variable is not enough
- The general multivariate regression model with K independent variables is:

  $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_K X_{Ki} + \varepsilon_i \quad (i = 1,2,\ldots,N)$
  (1.13)
- Biggest difference with single-explanatory variable regression model is in the **interpretation** of the slope coefficients
  - Now a slope coefficient indicates the change in the dependent variable associated with a one-unit increase in the explanatory variable *holding the other explanatory variables constant*

Total, Explained, and Residual Sums of Squares

$$TSS = \sum_{i=1}^{N} (Y_i - \bar{Y})^2$$

$$\sum_i (Y_i - \bar{Y})^2 = \sum_i (\hat{Y}_i - \bar{Y})^2 + \sum_i e_i^2$$

- TSS = ESS + RSS
  - This is usually called the **decomposition of variance**

Figure 2.3 Decomposition of the Variance in Y

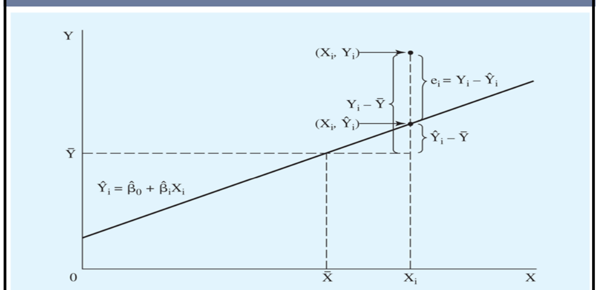


Figure 2.3 Decomposition of the Variance in Y

The variation of Y around its mean $(Y - \bar{Y})$ can be decomposed into two parts: (1) $(\hat{Y}_i - \bar{Y})$, the difference between the estimated value of $Y(\hat{Y})$ and the mean value of $Y(\bar{Y})$; and (2) $(Y_i - \hat{Y}_i)$, the difference between the actual value of Y and the estimated value of Y.

Describing the Overall Fit of the Estimated Model

- The simplest commonly used measure of overall fit is the **coefficient of determination**, R^2:

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum e_i^2}{\sum (Y_i - \bar{Y})^2} \qquad (2.14)$$

  - Since OLS selects the coefficient estimates that minimizes RSS, OLS provides the largest possible R^2 (within the class of linear models)

The adjusted coefficient of determination

- A major problem with R^2 is that it can never decrease if another independent variable is added
- An alternative to R^2 that addresses this issue is the **adjusted** R^2 :

$$\bar{R}^2 = 1 - \frac{\sum e_i^2/(N - K - 1)}{\sum (Y_i - \bar{Y})^2/(N - 1)} \qquad (2.15)$$

Where N – K – 1 = **degrees of freedom**

Adjusted R2

- ❑ Why is adjusted R-squared better?
- ❑ When K increases, RSS surely drops, but the denominator (n-K-1) also drops. Now if the drop in RSS is not large enough, RSS/(n-K-1) will actually increase so that adjusted R-squared will decrease.
- ❑ In other words, adjusted R-squared penalizes the measure of fit in adding an explanatory variable if that variable does not contribute much toward explaining the variable in Y.

Key Terms from Chapter 2

- **Ordinary Least Squares (OLS)**
- **Interpretation of a multivariate regression coefficient**
- **Total sums of squares**
- **Explained sums of squares**
- **Residual sums of squares**
- **Coefficient of determination, R^2**
- **Simple correlation coefficient, r**
- **Degrees of freedom**
- **Adjusted coefficient of determination , $\bar{R}^2$**

UNISA college of economic and management sciences

CHAPTER 3

LEARNING TO USE REGRESSION ANALYSIS

UNISA college of economic and management sciences

Steps in Applied Regression Analysis

- The first step is choosing the **dependent** variable – this step is determined by the purpose of the research
- After choosing the dependent variable, it's logical to follow the following sequence:
  1. Review the literature and develop the theoretical model
  2. Specify the model: Select the independent variables and the functional form
  3. Hypothesize the expected signs of the coefficients
  4. Collect the data. Inspect and clean the data
  5. Estimate and evaluate the equation
  6. Document the results

UNISA college of economic and management sciences

Step 1: Review the Literature and Develop the Theoretical Model

- Perhaps counter intuitively, a strong theoretical foundation is the best start for any empirical project
- Reason: main econometric decisions are determined by the underlying theoretical model
- Useful starting points:
  – Journal of Economic Literature or a business oriented publication of abstracts
  – Internet search, including Google Scholar
  – EconLit, an electronic bibliography of economics literature (for more details, go to www.EconLit.org)

UNISA college of economic and management sciences

Step 2: Specify the Model: Independent Variables and Functional Form

- After selecting the dependent variable, the **specification** of a model involves choosing the following components:
  1. the independent variables and how they should be measured,
  2. the functional (mathematical) form of the variables, and
  3. the properties of the stochastic error term

UNISA college of economic and management sciences

Step 2: Specify the Model: Independent Variables and Functional Form (cont.)

- A mistake in any of the three elements results in a **specification error**
- For example, only **theoretically relevant** explanatory variables should be included
- Even so, researchers frequently have to make choices –also denoted imposing their **priors**
- **Example:**
  - when estimating a demand equation, theory informs us that prices of complements and substitutes of the good in question are important explanatory variables
  - But *which* complements—and *which* substitutes?

UNISA college of economic and management sciences

Step 3: Hypothesize the Expected Signs of the Coefficients

- Once the variables are selected, it's important to hypothesize the expected signs of the regression coefficients
- **Example:** demand equation for a final consumption good
- First, state the demand equation as a general function:

$$Q_d = f(\overset{-}{P}, \overset{+}{Y}, \overset{-}{P_c}, \overset{+}{P_s}) + \epsilon \qquad (3.2)$$

- The signs above the variables indicate the hypothesized sign of the respective regression coefficient in a linear model

Step 4: Collect the Data & Inspect and Clean the Data

- A general rule regarding sample size is "the more observations the better"
  - as long as the observations are from the same general population!
- The reason for this goes back to notion of **degrees of freedom**
- When there are **more** degrees of freedom:
  - Every positive error is likely to be balanced by a negative error
  - The estimated regression coefficients are estimated with a greater deal of **precision**

Step 4: Collect the Data & Inspect and Clean the Data (cont.)

- Estimate model using the data in Table 2.2 to get:
- Inspecting the data—obtain a printout or plot (graph) of the data
- Reason: to look for **outliers**
  - An outlier is an observation that lies outside the range of the rest of the observations
- Examples:
  - Does a student have a 7.0 GPA on a 4.0 scale?
  - Is consumption negative?

Step 5: Estimate and Evaluate the Equation

- Once steps 1–4 have been completed, the **estimation** part is quick
  - using **Eviews** or **Stata** to estimate an OLS regression takes less than a second!
- The **evaluation** part is more tricky, however, involving answering the following questions:
  - How well did the equation **fit** the data?
  - Were the **signs** and **magnitudes** of the estimated coefficients as expected?
- Afterwards may add **sensitivity analysis**

Step 6: Document the Results

- A standard format usually is used to present estimated regression results:

$$\begin{aligned} \hat{Y}_i &= 103.40 + 6.38X_i \\ &\qquad\qquad (0.88) \\ &\qquad\qquad t = 7.22 \\ N = 20 &\qquad \bar{R}^2 = .73 \end{aligned} \qquad (3.3)$$

- The number in parentheses under the estimated coefficient is the estimated **standard error** of the estimated coefficient, and the **t-value** is the one used to test the hypothesis that the true value of the coefficient is different from zero (more on this later!)

Key Terms from Chapter 3

- The six steps in applied regression analysi
- Cross-sectional data set
- Specification error
- Degrees of freedom

CHAPTER 17

STATISTICAL PRINCIPLES

Probability

- A **random variable** X is a variable whose numerical value is determined by chance, the outcome of a random phenomenon
  - A **discrete** random variable has a **countable number** of possible values, such as 0, 1, and 2
  - A **continuous** random variable, such as time and distance, can take on any value in an interval
- A **probability distribution** $P[X_i]$ for a discrete random variable X assigns probabilities to the possible values X_1, X_2, and so on
- For example, when a fair six-sided die is rolled, there are six equally likely outcomes, each with a 1/6 probability of occurring

Mean, Variance, and Standard Deviation

The **expected value** (or **mean**) of a discrete random variable X is a weighted average of all possible values of X, using the probability of each X value as weights

$$\mu = E[X] = \sum_i X_i P[X_i]$$

- the variance of a discrete random variable X is a weighted average, for all possible values of X, of the squared difference between X and its expected value, using the probability of each X value as weights:

$$\sigma^2 = E[(X-\mu)^2] = \sum_i (X_i - \mu)^2 P[X_i]$$

- The **standard deviation** σ is the **square root** of the **variance**

Continuous Random Variables

- Our examples to this point have involved **discrete random variables**, for which we can **count** the number of possible outcomes:
  - The coin can be heads or tails; the die can be 1, 2, 3, 4, 5, or 6
- For **continuous random variables**, however, the outcome can be **any value** in a given **interval**
  - For example, Figure 16.2 shows a spinner for randomly selecting a point on a circle
- A **continuous probability density curve** shows the probability that the outcome is in a specified interval as the corresponding area under the curve
  - This is illustrated for the case of the spinner in Figure 16.3

Figure 17.6 The Normal Distribution

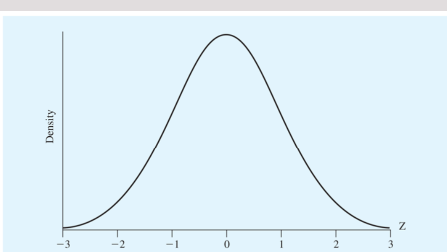


Figure 17.6 The Normal Distribution

The Normal Distribution (cont.)

- The **central limit theorem** is a very strong result for **empirical analysis** that **builds on** the **normal distribution**
- The **central limit theorem** states that:
  - if Z is a standardized sum of N independent, identically distributed (discrete or continuous) random variables with a finite, nonzero standard deviation, then the **probability distribution** of Z **approaches the normal distribution as N increases**

Sampling Distributions

- The **sampling distribution** of a statistic is the **probability distribution** or **density curve** that describes the population of **all possible values** of this statistic
  - For example, it can be shown mathematically that if the individual observations are drawn from a **normal distribution**, then the sampling distribution for the **sample mean** is also normal
  - Even if the population does not have a normal distribution, the sampling distribution of the sample mean will approach a normal distribution as the sample size increases

The Mean of the Sampling Distribution

- A **sample statistic** is an **unbiased estimator** of a population parameter if the **mean** of the sampling distribution of this statistic is **equal to** the value of the **population parameter**
- Because the mean of the sampling distribution of X is μ, X is an **unbiased estimator** of μ

The Standard Deviation of the Sampling Distribution

- One way of gauging the accuracy of an estimator is with its **standard deviation:**
  - If an estimator has a **large** standard deviation, there is a substantial probability that an estimate will be **far** from its **mean**
  - If an estimator has a **small** standard deviation, there is a high probability that an estimate will be **close** to its **mean**

Key Terms from Chapter 17

- **Random variable**
- **Probability distribution**
- **Expected Value**
- **Mean**
- **Variance**
- **Standard deviation**
- **Population**
- **Sample**
- **Sampling distribution**
- **Population mean**
- **Sample mean**
- **Population standard deviation**
- **Sample standard deviation**
- **Central limit theorem**

CHAPTER 4

THE CLASSICAL MODEL

The Classical Assumptions

- The **classical assumptions** must be met in order for OLS estimators to be the best available
- The seven classical assumptions are:
  - I. The regression model is **linear**, is **correctly specified**, and has an **additive error term**
  - II. The error term has a **zero population mean**
  - III. All **explanatory variables** are **uncorrelated** with the **error term**
  - IV. Observations of the error term are uncorrelated with each other **(no serial correlation)**
  - V. The error term has a constant variance **(no heteroskedasticity)**
  - VI. No explanatory variable is a perfect linear function of any other explanatory variable(s) **(no perfect multicollinearity)**
  - VII. The **error term** is **normally distributed** (this assumption is optional but usually is invoked)

I: linear, correctly specified, additive error term

- Consider the following regression model:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_K X_{Ki} + \varepsilon_i \quad (4.1)$$

- This model:
  - is **linear** (in the coefficients)
  - has an **additive error term**
- If we also assume that all the relevant explanatory variables are included in (4.1) then the model is also **correctly specified**

UNISA college of economic and management sciences

II: Error term has a zero population mean

- As was pointed out in Section 1.2, econometricians add a stochastic (random) error term to regression equations
- Reason: to account for variation in the dependent variable that is not explained by the model
- The specific value of the error term for each observation is determined purely by chance

UNISA college of economic and management sciences

III: All explanatory variables are uncorrelated with the error term

- If not, the OLS estimates would be likely to attribute to the X some of the variation in Y that actually came from the error term
- For example, if the error term and X were **positively correlated** then the estimated coefficient would probably be **higher** than it would otherwise have been **(biased upward)**
- This assumption is violated most frequently when a researcher omits an important independent variable from an equation

UNISA college of economic and management sciences

IV: No serial correlation of error term

- If a systematic correlation does exist between one observation of the error term and another, then it will be more difficult for OLS to get accurate estimates of the standard errors of the coefficients
- This assumption is most likely to be violated in time-series models:
  - An increase in the error term in one time period (a random shock, for example) is likely to be followed by an increase in the next period
- If, over all the observations of the sample ε_{t+1} is correlated with ε_t then the error term is said to be **serially correlated** (or **auto-correlated**), and Assumption IV is violated
- Violations of this assumption are considered in more detail in **Chapter 9**

UNISA college of economic and management sciences

V: Constant variance / No heteroskedasticity in error term

- The error term must have a **constant variance**
- That is, the variance of the error term cannot change for each observation or range of observations
- If it does, there is heteroskedasticity present in the error term
- An example of this can bee seen from **Figure 4.2**

UNISA college of economic and management sciences

Figure 4.2 An Error Term Whose Variance Increases as Z Increases (Heteroskedasticity)

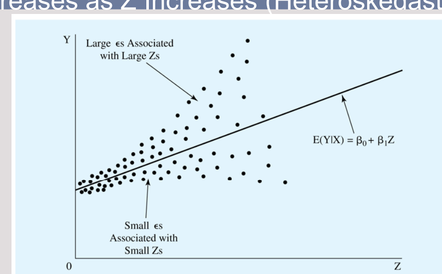


Figure 4.2 An Error Term Whose Variance Increases as Z Increases (Heteroskedasticity)

One example of Classical Assumption V not being met is when the variance of the error term increases as Z increases. In such a situation (called heteroskedasticity), the observations are on average farther from the true regression line for large values of Z than they are for small values of Z.

college of economic and management sciences

VI: No perfect multicollinearity

- **Perfect collinearity** between two independent variables implies that:
  - they are really the same variable, or
  - one is a multiple of the other, and/or
  - that a constant has been added to one of the variables
- **Example:**
- **GDP and MAF in assignment 2**

VII: The error term is normally distributed

- Basically implies that the error term follows a **bell-shape**
- Strictly speaking **not required** for OLS estimation (related to the Gauss-Markov Theorem: more on this in Section 4.3)
- Its major application is in **hypothesis testing**, which uses the estimated regression coefficient to investigate hypotheses about economic behavior (see Chapter 5)

The Sampling Distribution of $\hat{\beta}$

- We saw earlier that the error term follows a probability distribution (Classical Assumption VII)
- But so do the estimates of β!
  - The probability distribution of these $\hat{\beta}$ values across different samples is called the **sampling distribution of** $\hat{\beta}$
- We will now look at the properties of the **mean**, the **variance**, and the **standard error** of this sampling distribution

Properties of the Mean

- A desirable property of a distribution of estimates in that its mean equals the true mean of the variables being estimated
- Formally, an estimator $\hat{\beta}$ is an **unbiased estimator** if its sampling distribution has as its expected value the true value of β.
- We also write this as follows:
- $$E(\hat{\beta}) = \beta \qquad (4.9)$$
- Similarly, if this is not the case, we say that the estimator is **biased**

Properties of the Variance

- Just as we wanted the mean of the sampling distribution to be centered around the true population β, so too it is desirable for the sampling distribution to be as narrow (or precise) as possible.
  - Centering around "the truth" but with high variability might be of very little use.
- One way of narrowing the sampling distribution is to increase the sampling size (which therefore also increases the degrees of freedom)
- These points are illustrated in Figures 4.4 and 4.5

Figure 4.4 Distributions of $\hat{\beta}$

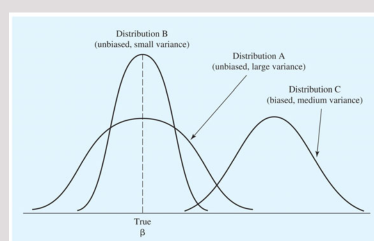


Figure 4.4 Distributions of $\hat{\beta}$

Different distributions of $\hat{\beta}$ can have different means and variances. Distributions A and B, for example, are both unbiased, but distribution A has a larger variance than does distribution B. Distribution C has a smaller variance than distribution A, but it is biased.

Figure 4.5 Sampling Distribution of for Various Observations (N)

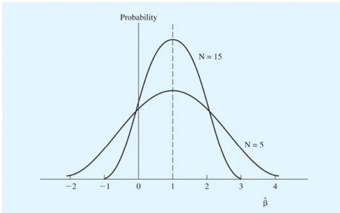


Figure 4.5 Sampling Distribution of $\hat{\beta}$ for Various Observations (N)

As the size of the sample increases, the variance of the distribution of $\hat{\beta}$s calculated from that sample tends to decrease. In the extreme case (not shown), a sample equal to the population would yield only an estimate equal to the mean of that distribution, which (for unbiased estimators) would equal the true β, and the variance of the estimates would be zero.

UNISA college of economic and management sciences

Properties of the Standard Error

- The standard error of the estimated coefficient, SE($\hat{\beta}$), is the square root of the estimated variance of the estimated coefficients.
- Hence, it is similarly affected by the sample size and the other factors discussed previously
  - For example, an increase in the sample size will decrease the standard error
  - Similarly, the larger the sample, the more precise the coefficient estimates will be

UNISA college of economic and management sciences

The Gauss-Markov Theorem and the Properties of OLS Estimators

- The Gauss-Markov Theorem states that:
  - Given Classical Assumptions I through VI (Assumption VII, normality, is not needed for this theorem), the Ordinary Least Squares estimator of –k is the minimum variance estimator from among the set of all linear unbiased estimators of –k, for k = 0, 1, 2, …, K
- We also say that "OLS is **BLUE**": "**B**est (meaning minimum variance) **L**inear **U**nbiased **E**stimator"

UNISA college of economic and management sciences

The Gauss-Markov Theorem and the Properties of OLS Estimators (cont.)

- The Gauss-Markov Theorem only requires the first six classical assumptions
- If we add the seventh condition, normality, the OLS coefficient estimators can be shown to have the following properties:
  - **Unbiased**: the OLS estimates coefficients are centered around the true population values
  - **Minimum variance**: no other unbiased estimator has a lower variance for each estimated coefficient than OLS
  - **Consistent**: as the sample size gets larger, the variance gets smaller, and each estimate approaches the true value of the coefficient being estimated
  - **Normally distributed**: when the error term is normally distributed, so are the estimated coefficients—which enables various statistical tests requiring normality to be applied (we'll get back to this in Chapter 5)

UNISA college of economic and management sciences

Table 4.1a Notation Conventions

Table 4.1 Notation Conventions

Population Parameter (True Values, but Unobserved)		Estimate (Observed from Sample)	
Name	**Symbol(s)**	**Name**	**Symbol(s)**
Regression coefficient	β_k	Estimated regression coefficient	$\hat{\beta}_k$
Expected value of the estimated coefficient	$E(\hat{\beta}_k)$		
Variance of the error term	σ^2 or $VAR(\epsilon_i)$	Estimated variance of the error term	s^2 or $\hat{\sigma}^2$

UNISA college of economic and management sciences

Key Terms from Chapter 4

- The classical assumptions
- Classical error term
- Standard normal distribution
- SE($\hat{\beta}$),
- Unbiased estimator
- BLUE
- Sampling distribution

UNISA college of economic and management sciences

CHAPTER 5

HYPOTHESIS TESTING

What Is Hypothesis Testing?

- Hypothesis testing is used in a variety of settings
  - The **Food and Drug Administration** (FDA), for example, tests **new products** before allowing their sale
    - If the sample of people exposed to the new product shows some side effect significantly more frequently than would be expected to occur by chance, the FDA is likely to withhold approval of marketing that product
  - Similarly, **economists** have been **statistically testing** various **relationships**, for example that between consumption and income
- Note here that while we **cannot prove** a given hypothesis (for example the existence of a given relationship), we often can reject a given hypothesis (again, for example, **rejecting** the existence of a given relationship)

Classical Null and Alternative Hypotheses

- The researcher first states the hypotheses to be tested
- Here, we distinguish between the **null** and the **alternative** hypothesis:
  - **Null hypothesis (“H_0”):** the outcome that the researcher does **not** expect (almost always includes an equality sign)
  - **Alternative hypothesis (“H_A”):** the outcome the researcher **does** expect
- Example:

  H_0: $\beta \leq 0$ (the values you do **not** expect)

  H_A: $\beta > 0$ (the values you **do** expect)

Type I and Type II Errors

- Two types of errors possible in hypothesis testing:
  - **Type I: Rejecting** a **true null** hypothesis
  - **Type II: Not rejecting** a **false null** hypothesis
- Example: Suppose we have the following null and alternative hypotheses:

  H_0: $\beta \leq 0$

  H_A: $\beta > 0$
  - Even if the **true** β really is not positive, in any one sample we might still observe an **estimate** of β that is sufficiently positive to lead to the **rejection** of the null hypothesis
- This can be illustrated by Figure 5.1

Figure 5.1 Rejecting a True Null Hypothesis Is a Type I Error

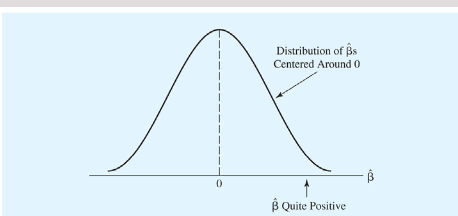


Figure 5.1 Rejecting a True Null Hypothesis Is a Type I Error

If $\beta = 0$, but you observe a $\hat{\beta}$ that is very positive, you might reject a true null hypothesis, H_0: $\beta \leq 0$, and conclude incorrectly that the alternative hypothesis H_A: $\beta > 0$ is true.

Type I and Type II Errors (cont.)

- Alternatively, it’s possible to obtain an estimate of β that is **close enough** to zero (or negative) to be considered “not significantly positive”
- Such a result may lead the researcher to “accept” the null hypothesis that $\beta \leq 0$ when in truth $\beta > 0$
- This is a **Type II Error**; we have failed to reject a false null hypothesis!
- This can be illustrated by Figure 5.2

Figure 5.2 Failure to Reject a False Null Hypothesis Is a Type II Error

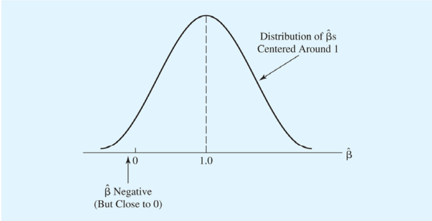


Figure 5.2 Failure to Reject a False Null Hypothesis Is a Type II Error

If $\beta = 1$, but you observe a $\hat{\beta}$ that is negative but close to zero, you might fail to reject a false null hypothesis, $H_0: \beta \leq 0$, and incorrectly ignore the fact that the alternative hypothesis, $H_A: \beta > 0$, is true.

Decision Rules of Hypothesis Testing

- To test a hypothesis, we calculate a **sample statistic** that determines when the null hypothesis can be rejected depending on the magnitude of that sample statistic relative to a preselected **critical value** (which is found in a statistical table)
- This procedure is referred to as a **decision rule**
- The decision rule is **formulated before** regression estimates are obtained
- The **range** of possible values of the estimates is divided into two regions, an "**acceptance**" (really, **non-rejection**) region and a **rejection region**
- The critical value effectively separates the "acceptance"/non-rejection region from the rejection region when testing a null hypothesis
- Graphs of these "acceptance" and rejection regions are given in
Figures 5.3 and 5.4

Figure 5.3 "Acceptance" and Rejection Regions for a One-Sided Test of β

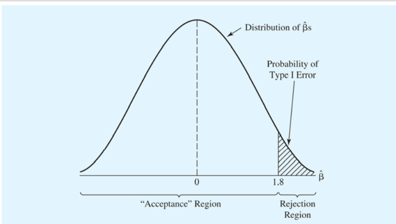


Figure 5.3 "Acceptance" and Rejection Regions for a One-Sided Test of β

For a one-sided test of $H_0: \beta \leq 0$ vs. $H_A: \beta > 0$, the critical value divides the distribution of $\hat{\beta}$ (centered around zero on the assumption that H_0 is true) into "acceptance" and rejection regions.

Figure 5.4 "Acceptance" and Rejection Regions for a Two-Sided Test of β

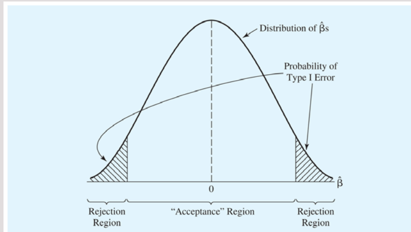


Figure 5.4 "Acceptance" and Rejection Regions for a Two-Sided Test of β

For a two-sided test of $H_0: \beta = 0$ vs. $H_A: \beta \neq 0$, we divided the distribution of $\hat{\beta}$ into an "acceptance" region and *two* rejection regions.

The *t*-Test

- The **t-test** is the test that econometricians usually use to test hypotheses about **individual** regression slope coefficients
  - Tests of more than one coefficient at a time (**joint** hypotheses) are typically done with the **F-test**
- The appropriate test to use when the stochastic error term is **normally distributed** and when the **variance** of that distribution must be **estimated**
  - Since these usually are the case, the use of the t-test for hypothesis testing has become **standard practice** in econometrics

The *t*-Statistic

- For a typical multiple regression equation:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i \qquad (5.1)$$

  we can calculate t-values for each of the estimated coefficients
  - Usually these are only calculated for the **slope coefficients**, though (see Section 7.1)
- Specifically, the t-statistic for the *k*th coefficient is:

$$t_k = \frac{(\hat{\beta}_k - \beta_{H_0})}{SE(\hat{\beta}_k)} \qquad (k = 1, 2, \ldots, K) \qquad (5.2$$

The Critical *t*-Value and the *t*-Test Decision Rule

- To decide whether to reject or not to reject a null hypothesis based on a calculated *t*-value, we use a **critical *t*-value**
- A **critical *t*-value** is the value that distinguishes the "acceptance" region from the rejection region
- The critical *t*-value, t_c, is selected from a ***t*-table** (see Statistical Table B-1 in the back of the book) depending on:
  - whether the test is one-sided or two-sided,
  - the level of Type I Error specified and
  - the degrees of freedom (defined as the number of observations minus the number of coefficients estimated (including the constant) or N – K – 1)

UNISA college of economic and management sciences

The Critical *t*-Value and the *t*-Test Decision Rule (cont.)

- The rule to apply when testing a single regression coefficient ends up being that you should:

  Reject H_0 if $|t_k| > t_c$ and if t_k also has the same sign implied by H_A

  Do not reject H_0 otherwise

UNISA college of economic and management sciences

The Critical *t*-Value and the *t*-Test Decision Rule (cont.)

- Note that this decision rule works both for calculated *t*-values and critical *t*-values for **one-sided** hypotheses around zero (or another hypothesized value, S):

H_0: $\beta_k \le 0$ H_0: $\beta_k \le S$
H_A: $\beta_k > 0$ H_A: $\beta_k > S$

H_0: $\beta_k \ge 0$ H_0: $\beta_k \ge S$
H_A: $\beta_k < 0$ H_A: $\beta_k < S$

UNISA college of economic and management sciences

The Critical t-Value and the *t*-Test Decision Rule (cont.)

- As well as for **two-sided** hypotheses around zero (or another hypothesized value, S):

  H_0: $\beta_k = 0$ H_0: $\beta_k = S$
  H_A: $\beta_k \neq 0$ H_A: $\beta_k \neq S$
- From Statistical Table B-1 the critical *t*-value for a **one-tailed** test at a given level of significance is exactly **equal** to the critical *t*-value for a **two-tailed** test at **twice** the level of significance of the **one-tailed** test—as also illustrated by Figure 5.5

UNISA college of economic and management sciences

Choosing a Level of Significance

- The **level of significance** indicates the probability of observing an estimated t-value **greater** than the **critical** t-value if the **null hypothesis** were **correct**
- It also measures the **amount** of **Type I Error** implied by a particular critical t-value
- Which level of significance is chosen?
  - 5 percent is recommended, **unless** you know something **unusual** about the **relative costs** of making Type I and Type II Errors
- The **level of significance** must be chosen **before** a critical value can be found, using Statistical Table B

UNISA college of economic and management sciences

Confidence Intervals

- A confidence interval is a range that contains the true value of an item a specified percentage of the time
- It is calculated using the estimated regression coefficient, the two-sided critical t-value and the standard error of the estimated coefficient as follows:
- $$\text{Confidence interval} = \hat{\beta} \pm t_c \cdot SE(\hat{\beta}) \quad (5.5)$$
- What's the relationship between confidence intervals and two-sided hypothesis testing?
- If a hypothesized value fall within the confidence interval, then we cannot reject the null hypothesis

UNISA college of economic and management sciences

p-Values

- This is an alternative to the t-test
- A p-value, or marginal significance level, is the **probability** of observing a t-score **that size or larger** (in absolute value) if the **null hypothesis** were **true**
- **Graphically**, it's two times the **area** under the curve of the t-distribution between the absolute value of the actual t-score and infinity.
- **In theory**, we could find this by combing through pages and pages of statistical tables
- But we don't have to, since we have **EViews** and **Stata**: these (and other) statistical software packages automatically give the p-values as part of the standard output!
- In light of all this, the **p-value decision rule** therefore is:

  Reject H_0 if p-value$_K$ < the level of significance and if has the sign implied by H_A

UNISA

Limitations of the *t*-Test

- With the t-values being automatically printed out by computer regression packages, there is reason to **caution** against **potential improper use** of the t-test:
  1. The t-Test Does Not Test **Theoretical Validity**:

     If you regress the consumer price index on rainfall in a time-series regression and find strong statistical significance does that also mean that the **underlying theory** is **valid**? Of course not!

UNISA

Limitations of the *t*-Test

The t-Test Does Not Test **"Importance"**:

The fact that one coefficient is "more statistically significant" than another does not mean that it is also more important in explaining the dependent variable—but merely that we have **more evidence** of the **sign** of the coefficient in question

3. The t-Test Is Not Intended for Tests of the **Entire Population**:

   From the definition of the t-score, given by Equation 5.2, it is seen that as the **sample size** approaches the **population** (whereby the standard error will approach zero—since the standard error decreases as N increases), the t-score will approach **infinity!**

UNISA

Key Terms from Chapter 5

- **Null hypothesis**
- **Alternative hypothesis**
- **Type I Error**
- **Level of significance**
- **Two-sided test**
- **Decision rule**
- **Critical value**
- ***t*-statistic**

UNISA

Chapter 6

SPECIFICATION: CHOOSING THE INDEPENDENT VARIABLES

UNISA

Specifying an Econometric Equation and Specification Error

- Before any equation can be estimated, it must be **completely specified**
- **Specifying** an econometric equation consists of **three parts**, namely choosing the correct:
  - independent variables
  - functional form
  - form of the stochastic error term
- Again, this is part of the **first classical assumption** from Chapter 4
- A **specification error** results when one of these choices is made **incorrectly**
- This chapter will deal with the first of these choices (the two other choices will be discussed in subsequent chapters)

UNISA

Omitted Variables

- Two reasons why an important explanatory variable might have been left out:
  - we forgot…
  - it is not available in the dataset, we are examining
- Either way, this may lead to **omitted variable bias** (or, more generally, **specification bias**)
- The reason for this is that when a variable is not included, it **cannot** be **held constant**
- Omitting a relevant variable usually is evidence that the entire equation is a suspect, because of the likely bias of the coefficients.

UNISA college of economic and management sciences

The Consequences of an Omitted Variable

- Suppose the true regression model is:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i \quad (6.1)$$

Where ϵ_i is a classical error term

- If X_2 is omitted, the equation becomes:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \epsilon_i^* \quad (6.2)$$

Where:

$$\epsilon_i^* = \epsilon_i + \beta_2 X_{2i} \quad (6.3)$$

- Hence, the explanatory variables in the **estimated** regression (6.2) are **not independent** of the error term (unless the omitted variable is **uncorrelated** with **all** the **included** variables—something which is **very unlikely**)
- But this **violates Classical Assumption III**!

UNISA college of economic and management sciences

The Consequences of an Omitted Variable (cont.)

- What happens if we estimate Equation 6.2 when Equation 6.1 is the truth?
- We get **bias**!
- What this means is that:

$$E(\hat{\beta}_1) \neq \beta_1 \quad (6.4)$$

- The **amount** of bias is a function of the impact of the **omitted** variable on the dependent variable times a function of the **correlation** between the **included** and the **omitted** variable
- Or, more formally:

$$\text{Bias} = \beta_{om} f(r_{in,om}) \quad (6.7)$$

- So, the **bias exists unless**:
  1. the true coefficient equals zero, or
  2. the included and omitted variables are uncorrelated

UNISA college of economic and management sciences

Correcting for an Omitted Variable

- In theory, the solution to a problem of specification bias seems easy: add the omitted variable to the equation!
- Unfortunately, that's easier said than done, for a couple of reasons
  1. Omitted variable bias is **hard to detect**: the **amount** of bias introduced can be **small** and not immediately detectable
  2. Even if it has been decided that a given equation is suffering from omitted variable bias, how to decide exactly **which** variable to include?
- Note here that **dropping** a variable is **not** a viable strategy to help cure omitted variable bias:
  - If anything you'll just generate even **more** omitted variable bias on the remaining coefficients!

UNISA college of economic and management sciences

Correcting for an Omitted Variable (cont.)

- What if:
  - You have an **unexpected result**, which leads you to believe that you have an **omitted variable**
  - You have two or more **theoretically sound** explanatory variables as potential "candidates" for inclusion as the omitted variable to the equation is to use
- How do you **choose** between these variables?
- One possibility is **expected bias analysis**
  - **Expected bias:** the likely bias that omitting a particular variable would have caused in the estimated coefficient of one of the included variables

UNISA college of economic and management sciences

Correcting for an Omitted Variable (cont.)

- Expected bias can be estimated with Equation 6.7:

$$\text{Expected bias} = \beta_{om} \cdot f(r_{in,om}) \quad (6.7)$$

- When do we have a viable candidate?
  - When the **sign** of the **expected bias** is the **same** as the **sign** of the **unexpected result**
- Similarly, when these signs **differ**, the variable is **extremely unlikely** to have caused the unexpected result

UNISA college of economic and management sciences

Irrelevant Variables

- This refers to the case of including a variable in an equation when it does **not** belong there
- This is the **opposite** of the **omitted** variables case—and so the impact can be illustrated using the same model
- Assume that the true regression specification is:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \epsilon_i \quad (6.10)$$

- But the researcher for some reason includes an extra variable:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i^{**} \quad (6.11)$$

- The misspecified equation's error term then becomes:

$$\epsilon_i^{**} = \epsilon_i - \beta_2 X_{2i} \quad (6.12)$$

UNISA college of economic and management sciences

Irrelevant Variables (cont.)

- So, the inclusion of an irrelevant variable will **not cause bias** (since the true coefficient of the irrelevant variable is zero, and so the second term will drop out of Equation 6.12)
- However, the inclusion of an irrelevant variable will:
  - **Increase** the **variance** of the estimated coefficients, and this increased variance will tend to **decrease** the absolute magnitude of their **t-scores**
  - **Decrease** the adjusted R^2 (but not the R^2)
- Table 6.1 summarizes the consequences of the omitted variable and the included irrelevant variable cases (unless $r_{12} = 0$)

UNISA college of economic and management sciences

Table 6.1 Effect of Omitted Variables and Irrelevant Variables on the Coefficient Estimates

Table 6.1 Effect of Omitted Variables and Irrelevant Variables on the Coefficient Estimates

Effect on Coefficient Estimates	Omitted Variable	Irrelevant Variable
Bias	Yes	No
Variance	Decreases	Increases

UNISA college of economic and management sciences

Four Important Specification Criteria

- We can summarize the previous discussion into four criteria to help decide whether a given variable belongs in the equation:
  1. **Theory:** Is the variable's place in the equation unambiguous and theoretically sound?
  2. **t-Test:** Is the variable's estimated coefficient significant in the expected direction?
  3. R^2: Does the overall fit of the equation (adjusted for degrees of freedom) improve when the variable is added to the equation?
  4. **Bias:** Do other variables' coefficients change significantly when the variable is added to the equation?
- If **all** these conditions hold, the variable **belongs** in the equation
- If **none** of them hold, it does **not belong**
- The tricky part is the **intermediate cases**: use sound judgment!

UNISA college of economic and management sciences

Key Terms from Chapter 6

- Omitted variable
- Irrelevant variable
- Specification bias
- Specification error
- The four specification criteria
- Expected bias

UNISA college of economic and management sciences

CHAPTER 7

SPECIFICATION: CHOOSING A FUNCTIONAL FORM

UNISA college of economic and management sciences

The Use and Interpretation of the Constant Term

- An estimate of β_0 has at least **three components**:
  1. the true β_0
  2. the constant impact of any specification errors (an omitted variable, for example)
  3. the mean of ε for the correctly specified equation (if not equal to zero)
- Unfortunately, these components can't be distinguished from one another because we can observe only β_0, the sum of the three components
- As a result of this, we usually **don't interpret** the constant term
- On the other hand, we should **not suppress** the constant term, either, as illustrated by Figure 7.1

Figure 7.1 The Harmful Effect of Suppressing the Constant Term

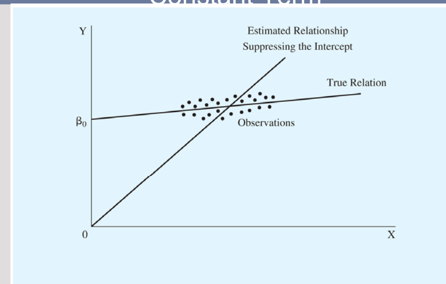


Figure 7.1 The Harmful Effect of Suppressing the Constant Term

If the constant (or intercept) term is suppressed, the estimated regression will go through the origin. Such an effect potentially biases the $\hat{\beta}$s and inflates their *t*-scores. In this particular example, the true slope is close to zero in the range of the sample, but forcing the regression through the origin makes the slope appear to be significantly positive.

Alternative Functional Forms

- An equation is **linear** in the **variables** if plotting the function in terms of X and Y generates a **straight line**
- For example, Equation 7.1:

$$Y = \beta_0 + \beta_1 X + \varepsilon \qquad (7.1)$$

  is linear in the variables but Equation 7.2:

$$Y = \beta_0 + \beta_1 X^2 + \varepsilon \qquad (7.2)$$

  is **not** linear in the variables
- Similarly, an equation is **linear** in the **coefficients** only if the **coefficients** appear in their simplest form—they:
  - are **not raised** to any powers (other than one)
  - are **not multiplied** or **divided** by other coefficients
  - do **not** themselves **include** some sort of **function** (like **logs** or **exponents**)

Alternative Functional Forms (cont.)

- For example, Equations 7.1 and 7.2 **are linear** in the coefficients, while Equation 7:3:

$$Y = \beta_0 + X^{\beta_1} \qquad (7.3)$$

  is **not linear** in the coefficients
- In fact, of **all possible equations** for a single explanatory variable, **only** functions of the general form:

$$f(Y) = \beta_0 + \beta_1 f(X) \qquad (7.4)$$

  are linear in the coefficients β_0 and β_1

Linear Form

- This is based on the assumption that the slope of the relationship between the independent variable and the dependent variable is constant:

$$\frac{\Delta Y}{\Delta X_k} = \beta_k \qquad k = 1, 2, \ldots, K$$

- For the linear case, the **elasticity** of Y with respect to X (the percentage change in the dependent variable caused by a 1-percent increase in the independent variable, holding the other variables in the equation constant) is:

$$\text{Elasticity}_{Y,X_k} = \frac{\Delta Y/Y}{\Delta X_k/X_k} = \frac{\Delta Y}{\Delta X_k} \cdot \frac{X_k}{Y} = \beta_k \frac{X_k}{Y}$$

What Is a Log?

- If **e** (a **constant** equal to 2.71828) to the "*b*th power" produces x, then b is the **log** of x:

  b is the log of x to the base e if: $e^b = x$
- Thus, a log (or logarithm) is the exponent to which a given base must be taken in order to produce a specific number
- While logs come in more than one variety, we'll use only natural logs (logs to the base e) in this text
- The symbol for a natural log is "ln," so ln(x) = b means that $(2.71828)^b = x$ or, more simply,

  $\ln(x) = b$ means that $e^b = x$
- For example, since $e^2 = (2.71828)^2 = 7.389$, we can state that:

  $\ln(7.389) = 2$

  Thus, the natural log of 7.389 is 2! Again, why? Two is the power of e that produces 7.389

What Is a Log? (cont.)

- Let's look at some other natural log calculations:
  - ln(100) = 4.605
  - ln(1000) = 6.908
  - ln(10000) = 9.210
  - ln(1000000) = 13.816
  - ln(100000) = 11.513
- Note that as a number goes from 100 to 1,000,000, its natural log goes from 4.605 to only 13.816! As a result, logs can be used in econometrics if a researcher wants to **reduce** the **absolute size** of the numbers associated with the same actual meaning
- One useful property of natural logs in econometrics is that they make it easier to figure out impacts in **percentage terms** (we'll see this when we get to the **double-log** specification)

Double-Log Form

- Here, the natural log of Y is the dependent variable and the natural log of X is the independent variable:

$$\ln Y = \beta_0 + \beta_1 \ln X_1 + \beta_2 \ln X_2 + \epsilon \qquad (7.5)$$

- In a double-log equation, an individual regression coefficient can be interpreted as an **elasticity** because:

$$\beta_k = \frac{\Delta(\ln Y)}{\Delta(\ln X_k)} = \frac{\Delta Y/Y}{\Delta X_k/X_k} = \text{Elasticity}_{Y,X_k} \qquad (7.6)$$

- Note that the **elasticities** of the model are **constant** and the **slopes** are **not**
- This is in **contrast** to the **linear model**, in which the **slopes** are **constant** but the **elasticities** are **not**

Figure 7.2 Double-Log Functions

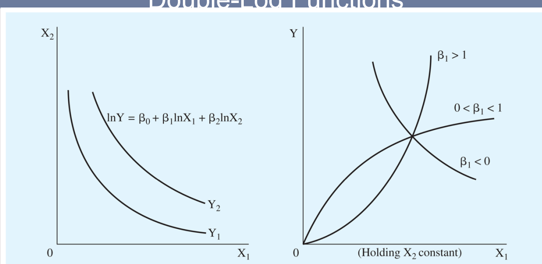


Figure 7.2 Double-Log Functions

Depending on the values of the regression coefficients, the double-log functional form can take on a number of shapes. The left panel shows the use of a double-log function to depict a shape useful in describing the economic concept of an isoquant or an indifference curve. The right panel shows various shapes that can be achieved with a double-log function if X_2 is held constant or is not included in the equation.

Semilog Form

- The **semilog** functional form is a variant of the double-log equation in which some but not all of the variables (dependent and independent) are expressed in terms of their natural logs.
- It can be on the **right-hand** side, as in:

$$Y_i = \beta_0 + \beta_1 \ln X_{1i} + \beta_2 X_{2i} + \varepsilon_i \qquad (7.7)$$

- Or it can be on the **left-hand** side, as in:

$$\ln Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \qquad (7.9)$$

- Figure 7.3 illustrates these two different cases

Figure 7.3 Semilog Functions

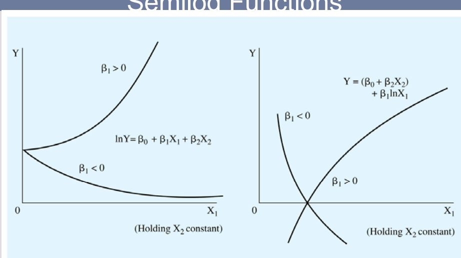


Figure 7.3 Semilog Functions

The semilog functional form on the right (lnX) can be used to depict a situation in which the impact of X_1 on Y is expected to increase at a decreasing rate as X_1 gets bigger as long as β_1 is greater than zero (holding X_2 constant). The semilog functional form on the left (lnY) can be used to depict a situation in which an increase in X_1 causes Y to increase at an increasing rate.

Polynomial Form

- **Polynomial functional** forms express Y as a function of independent variables, some of which are raised to powers other than 1
- For example, in a **second-degree** polynomial (also called a quadratic) equation, at least one independent variable is **squared**:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 (X_{1i})^2 + \beta_3 X_{2i} + \varepsilon_i \qquad (7.10)$$

- The **slope** of Y with respect to X_1 in Equation 7.10 is:

$$\frac{\Delta Y}{\Delta X_1} = \beta_1 + 2\beta_2 X_1 \qquad (7.11)$$

- Note that the **slope** depends on the **level** of X_1

Figure 7.4 Polynomial Functions

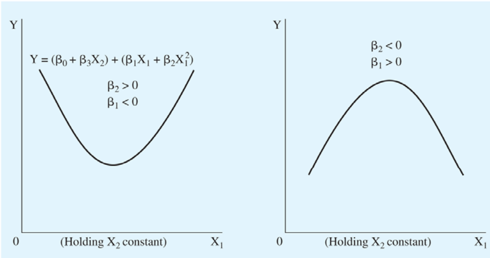


Figure 7.4 Polynomial Functions

Quadratic functional forms (polynomials with squared terms) take on U or inverted U shapes, depending on the values of the coefficients (holding X_2 constant). The left panel shows the shape of a quadratic function that could be used to show a typical cost curve; the right panel allows the description of an impact that rises and then falls (like the impact of age on earnings).

Figure 7.5 Inverse Functions

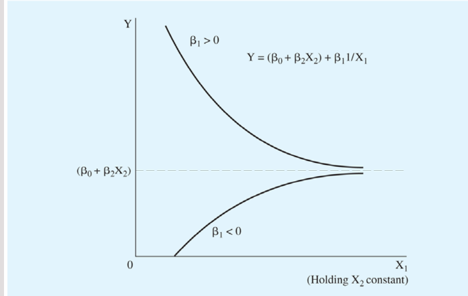


Figure 7.5 Inverse Functions

Inverse (or reciprocal) functional forms allow the impact of an X_1 on Y to approach zero as X_1 increases in size. The inverse function approaches the same value (the asymptote) from the top or bottom depending on the sign of β_1.

Table 7.1 Summary of Alternative Functional Forms

Table 7.1 Summary of Alternative Functional Forms

Functional Form	Equation (one X only)	The Meaning of β_1
Linear	$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$	The slope of Y with respect to X
Double-log	$\ln Y_i = \beta_0 + \beta_1 \ln X_i + \epsilon_i$	The elasticity of Y with respect to X
Semilog (lnX)	$Y_i = \beta_0 + \beta_1 \ln X_i + \epsilon_i$	The change in Y (in units) related to a 1 percent increase in X
Semilog (lnY)	$\ln Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$	The percent change in Y related to a one-unit increase in X
Polynomial	$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \epsilon_i$	Roughly, the slope of Y with respect to X for small X
Inverse	$Y_i = \beta_0 + \beta_1 \left(\frac{1}{X_i}\right) + \epsilon_i$	Roughly, the inverse of the slope of Y with respect to X for small X

Using Dummy Variables

- A dummy variable is a variable that takes on the values of 0 or 1, depending on whether a condition for a qualitative attribute (such as gender) is met
- These conditions take the **general form:**

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \epsilon_i \quad (7.18)$$

$$\text{where } D_i = \begin{cases} 1 \text{ if the } i\text{th observation meets a particular condition} \\ 0 \text{ otherwise} \end{cases}$$

- This is an example of an **intercept dummy** (as opposed to a **slope dummy**, which is discussed in Section 7.5)
- Figure 7.6 illustrates the consequences of including an intercept dummy in a linear regression model

Figure 7.6 An Intercept Dummy

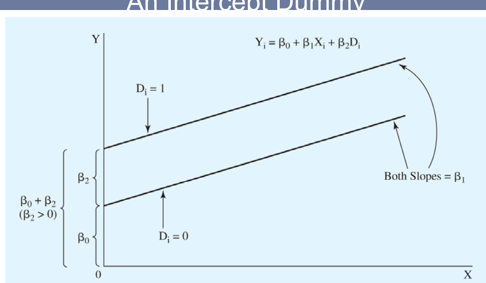


Figure 7.6 An Intercept Dummy

If an intercept dummy ($\beta_2 D_i$) is added to an equation, a graph of the equation will have different intercepts for the two qualitative conditions specified by the dummy variable. The difference between the two intercepts is β_2. The slopes are constant with respect to the qualitative condition.

Slope Dummy Variables

- Contrary to the **intercept dummy**, which changed only the intercept (and not the slope), the **slope dummy** changes both the intercept and the slope
- The **general form** of a **slope dummy equation** is:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 X_i D_i + \varepsilon_i \quad (7.20)$$

- The slope depends on the value of D:

  When D = 0, $\Delta Y/\Delta X = \beta_1$

  When D = 1, $\Delta Y/\Delta X = (\beta_1 + \beta_3)$
- Graphical illustration of how this works in Figure 7.7

Figure 7.7 Slope and Intercept Dummies

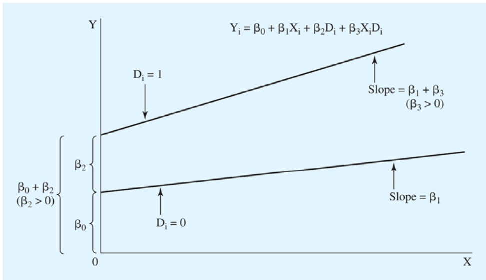


Figure 7.7 Slope and Intercept Dummies

If slope dummy ($\beta_3 X_i D_i$) and intercept dummy ($\beta_2 D_i$) terms are added to an equation, a graph of the equation will have different intercepts *and* different slopes depending on the value of the qualitative condition specified by the dummy variable. The difference between the two intercepts is β_2, whereas the difference between the two slopes is β_3.

Key Terms from Chapter 7

- Elasticity
- Double-log functional form
- Semilog functional form
- Polynomial functional form
- Inverse functional form
- Slope dummy
- Natural log
- Omitted condition
- Interaction term
- Linear in the variables
- Linear in the coefficients

CHAPTER 8

MULTICOLLINEARITY

Introduction and Overview

- The next three chapters deal with **violations** of the **Classical Assumptions** and **remedies** for those violations
- This chapter addresses **multicollinearity**; the next two chapters are on **serial correlation** and **heteroskedasticity**
- For each of these three problems, we will attempt to answer the following questions:
  1. What is the **nature** of the problem?
  2. What are the **consequences** of the problem?
  3. How is the problem **diagnosed**?
  4. What **remedies** for the problem are available?

Perfect Multicollinearity

- **Perfect multicollinearity** violates **Classical Assumption VI**, which specifies that no explanatory variable is a perfect linear function of any other explanatory variables
- The word **perfect** in this context implies that the variation in one explanatory variable can be **completely explained** by movements in another explanatory variable
  - A special case is that of a **dominant variable:** an **explanatory** variable is **definitionally related** to the **dependent** variable
- An example would be (Notice: **no error term**!):

  $$X_{1i} = \alpha_0 + \alpha_1 X_{2i} \quad (8.1)$$

  where the αs are constants and the Xs are independent variables in:

  $$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i \quad (8.2)$$

- Figure 8.1 illustrates this case

Perfect Multicollinearity (cont.)

- What happens to the estimation of an econometric equation where there is perfect multicollinearity?
  - OLS is **incapable** of **generating estimates** of the regression coefficients
  - most OLS computer programs will print out an **error message** in such a situation
- What is going on?
- Essentially, perfect multicollinearity ruins our ability to estimate the coefficients because the perfectly collinear **variables cannot** be **distinguished** from each other:
  - You cannot "hold all the other independent variables in the equation constant" if every time one variable changes, another changes in an identical manner!
- **Solution**: one of the collinear variables must be dropped (they are essentially **identical**, anyway)

Imperfect Multicollinearity

- **Imperfect multicollinearity** occurs when two (or more) explanatory variables are **imperfectly linearly related**, as in:

$$X_{1i} = \alpha_0 + \alpha_1 X_{2i} + u_i \qquad (8.7)$$

- Compare Equation 8.7 to Equation 8.1
  - Notice that Equation 8.7 includes u_i, a **stochastic error term**

UNISA college of economic and management sciences

The Consequences of Multicollinearity

There are **five major consequences** of multicollinearity:

1. Estimates will **remain unbiased**
2. The variances and standard errors of the estimates will **increase:**
   a. **Harder** to **distinguish** the effect of one variable from the effect of another, so much **more likely** to make **large errors** in estimating the βs than without multicollinearity
   b. As a result, the estimated coefficients, although still unbiased, now come from distributions with much **larger variances** and, therefore, **larger standard errors** (this point is illustrated in Figure 8.3)

UNISA college of economic and management sciences

Figure 8.3 Severe Multicollinearity Increases the Variances of the $\hat{\beta}$s

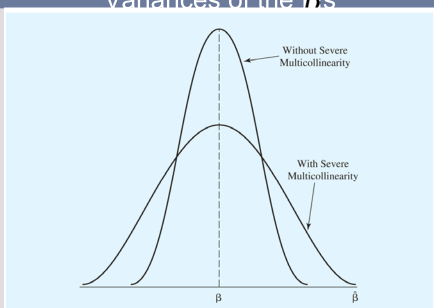


Figure 8.3 Severe Multicollinearity Increases the Variances of the $\hat{\beta}$s

Severe multicollinearity produces a distribution of the $\hat{\beta}$s that is centered around the true β but that has a much wider variance. Thus, the distribution of $\hat{\beta}$s with multicollinearity is much wider than otherwise.

college of economic and management sciences

The Consequences of Multicollinearity (cont.)

3. The computed **t-scores** will **fall**:
   a. Recalling Equation 5.2, this is a direct consequence of 2. above
4. Estimates will become very **sensitive** to **changes** in **specification**:
   a. The **addition** or **deletion** of an **explanatory variable** or of **a few observations** will often cause major changes in the values of the s when significant multicollinearity exists
   b. For example, if you drop a variable, even one that appears to be statistically insignificant, the **coefficients** of the remaining variables in the equation sometimes will **change dramatically**
   c. This is again because with multicollinearity, it is much **harder** to **distinguish** the **effect** of one variable from the effect of another
5. The **overall fit** of the equation and the estimation of the **coefficients** of **nonmulticollinear variables** will be **largely unaffected**

UNISA college of economic and management sciences

The Detection of Multicollinearity

- First realize that that **some** multicollinearity exists in **every** equation: all variables are correlated to some degree (even if completely at random)
- So it's really a question of **how much** multicollinearity exists in an equation, rather than **whether** any multicollinearity exists
- There are basically two characteristics that help detect the degree of multicollinearity for a given application:
  1. High simple correlation coefficients
  2. High Variance Inflation Factors (VIFs)
- We will now go through each of these in turn:

UNISA college of economic and management sciences

High Simple Correlation Coefficients

- If a **simple correlation coefficient**, r, between any two explanatory variables is high in absolute value, these two particular Xs are **highly correlated** and multicollinearity is a **potential problem**
- How high is high?
  - Some researchers pick an arbitrary number, such as 0.80
  - A better answer might be that r is high if it causes unacceptably large variances in the coefficient estimates in which we're interested.
- **Caution** in case of more than two explanatory variables:
  - **Groups** of independent variables, acting together, may cause multicollinearity without any single simple correlation coefficient being high enough to indicate that multicollinearity is present
  - As a result, simple correlation coefficients must be considered to be **sufficient** but **not necessary** tests for multicollinearity

UNISA college of economic and management sciences

High Variance Inflation Factors (VIFs)

- The **variance inflation factor (VIF)** is calculated from **two steps:**
  1. Run an OLS regression that has X_i as a function of all the other explanatory variables in the equation—For i = 1, this equation would be:
- $X_1 = \alpha_1 + \alpha_2 X_2 + \alpha_3 X_3 + \ldots + \alpha_K X_K + v$ (8.15)
- where v is a classical stochastic error term
  2. Calculate the variance inflation factor for :

$$\mathrm{VIF}(\hat{\beta}_i) = \frac{1}{(1 - R_i^2)} \quad (8.16)$$

- where R_i^2 is the unadjusted R^2 from **step one**

UNISA college of economic and management sciences

High Variance Inflation Factors (VIFs) (cont.)

- From Equation 8.16, the **higher** the VIF, the **more severe** the effects of mulitcollinearity
- How high is high?
- While there is no table of formal critical VIF values, a common rule of thumb is that if a given VIF is **greater** than **5**, the multicollinearity is severe
- As the **number** of **independent variables increases**, it makes sense to **increase** this number slightly
- Note that the authors replace the VIF with its **reciprocal**, $(1 - R_i^2)$, called **tolerance**, or **TOL**
- Problems with VIF:
  - No hard and fast VIF decision rule
  - There can still be severe multicollinearity even with small VIFs
  - VIF is a **sufficient**, **not necessary**, test for multicollinearity

UNISA college of economic and management sciences

Remedies for Multicollinearity

Essentially **three remedies** for multicollinearity:

1. Do nothing:
   a. Multicollinearity will not necessarily reduce the t-scores enough to make them statistically insignificant and/or change the estimated coefficients to make them differ from expectations
   b. the deletion of a multicollinear variable that **belongs** in an equation will cause **specification bias**
2. Drop a redundant variable:
   a. Viable strategy when two variables measure essentially the same thing
   b. Always use **theory** as the basis for this decision

UNISA college of economic and management sciences

Remedies for Multicollinearity (cont.)

3. Increase the sample size:
   a. This is frequently impossible but a useful alternative to be considered if feasible
   b. The idea is that the larger sample normally will **reduce** the **variance** of the estimated coefficients, **diminishing** the **impact** of the **multicollinearity**

UNISA college of economic and management sciences

Key Terms from Chapter 8

- Perfect multicollinearity
- Severe imperfect multicollinearity
- Dominant variable
- Auxiliary (or secondary) equation
- Variance inflation factor
- Redundant variable

UNISA college of economic and management sciences

CHAPTER 9

SERIAL CORRELATION

UNISA college of economic and management sciences

Pure Serial Correlation

- **Pure serial correlation** occurs when **Classical Assumption IV**, which assumes **uncorrelated** observations of the **error term**, is violated (in a correctly specified equation!)
- The most commonly assumed kind of serial correlation is **first-order** serial correlation, in which the **current** value of the error term is a function of the **previous** value of the error term:

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t \qquad (9.1)$$

where: ε = the error term of the equation in question
ρ = the first-order autocorrelation coefficient
u = a classical (not serially correlated) error term

UNISA college of economic and management sciences

Pure Serial Correlation (cont.)

- The magnitude of ρ indicates the **strength** of the serial correlation:
  - If ρ is **zero**, there is **no serial correlation**
  - As ρ approaches one in absolute value, the previous observation of the error term becomes more important in determining the current value of ε_t and a high degree of serial correlation exists
  - For ρ to exceed one is unreasonable, since the error term effectively would "explode"
- As a result of this, we can state that:

$$-1 < \rho < +1 \qquad (9.2)$$

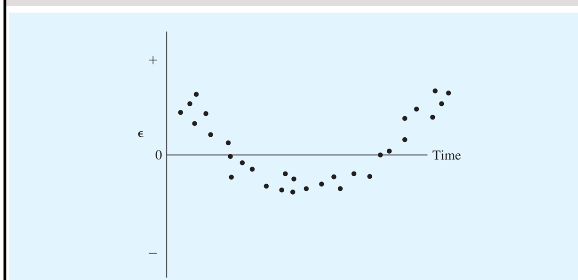
UNISA college of economic and management sciences

Pure Serial Correlation (cont.)

- The sign of ρ indicates the nature of the serial correlation in an equation:
- **Positive**:
  - implies that the error term tends to have the **same sign** from one time period to the next
  - this is called **positive serial correlation**
- **Negative**:
  - implies that the error term has a tendency to **switch signs** from negative to positive and back again in consecutive observations
  - this is called **negative serial correlation**
- Figures 9.1–9.3 illustrate several different scenarios

UNISA college of economic and management sciences

Figure 9.1a Positive Serial Correlation



UNISA college of economic and management sciences

Figure 9.1b Positive Serial Correlation

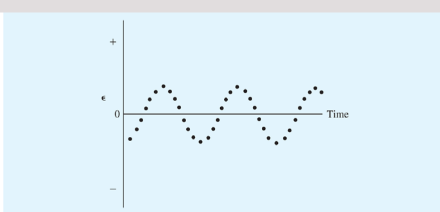


Figure 9.1 Positive Serial Correlation

With positive first-order serial correlation, the current observation of the error term tends to have the same sign as the previous observation of the error term. An example of positive serial correlation would be external shocks to an economy that take more than one time period to completely work through the system.

UNISA college of economic and management sciences

Figure 9.2 No Serial Correlation

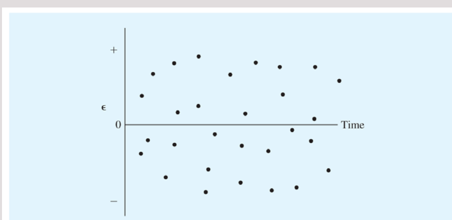
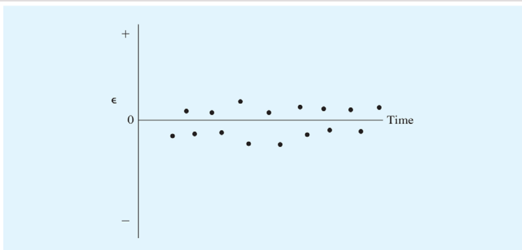


Figure 9.2 No Serial Correlation

With no serial correlation, different observations of the error term are completely uncorrelated with each other. Such error terms would conform to Classical Assumption IV.

UNISA college of economic and management sciences

Figure 9.3a Negative Serial Correlation



UNISA

Figure 9.3b Negative Serial Correlation

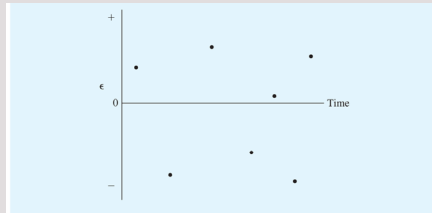


Figure 9.3 Negative Serial Correlation

With negative first-order serial correlation, the current observation of the error term tends to have the opposite sign from the previous observation of the error term. In most time-series applications, negative serial correlation is much less likely than positive serial correlation.

UNISA

Impure Serial Correlation

- **Impure serial correlation** is serial correlation that is caused by a **specification error** such as:
  - an **omitted variable** and/or
  - an **incorrect functional** form
- How does this happen?
- As an example, suppose that the true equation is:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \epsilon_t \qquad (9.3)$$

where ϵ_t is a classical error term. As shown in Section 6.1, if X_2 is accidentally omitted from the equation (or if data for X_2 are unavailable), then:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \epsilon_t^* \qquad \text{where } \epsilon_t^* = \beta_2 X_{2t} + \epsilon_t \qquad (9.4)$$

- The error term is therefore **not** a **classical error term**

UNISA

Impure Serial Correlation (cont.)

- Instead, the error term is also a function of one of the explanatory variables, X_2
- As a result, the new error term, ε^* , can be serially correlated even if the true error term ε, is not
- In particular, the new error term will tend to be serially correlated when:
  1. X_2 itself is **serially correlated** (this is quite likely in a time series) and
  2. the size of ε is **small** compared to the size of $\beta_2 \bar{X}_2$
- Figure 9.4 illustrates 1., for the case of U.S. disposable income

UNISA

Impure Serial Correlation (cont.)

- Turn now to the case of impure serial correlation caused by an **incorrect functional form**
- Suppose that the true equation is **polynomial** in nature:

$$Y_t = \beta_0 + \beta X_{1t} + \beta_2 X_{1t}^2 + \epsilon_t \qquad (9.7)$$

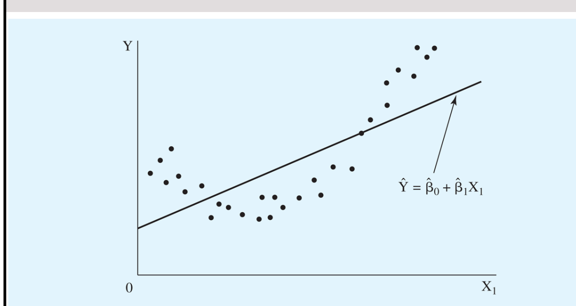
but that instead a **linear** regression is run:

$$Y_t = \alpha_0 + \alpha_1 X_{1t} + \epsilon_t^* \qquad (9.\ 8)$$

- The new error term ε^* is now a **function** of the **true** error term and of the **differences** between the **linear** and the **polynomial functional forms**
- Figure 9.5 illustrates how these differences often follow fairly

UNISA

Figure 9.5a Incorrect Functional Form as a Source of Impure Serial Correlation



UNISA

Figure 9.5b Incorrect Functional Form as a Source of Impure Serial Correlation

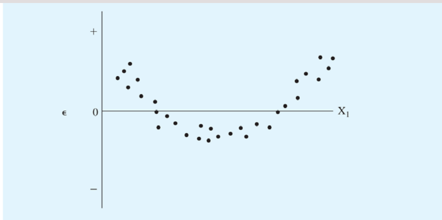


Figure 9.5 Incorrect Functional Form as a Source of Impure Serial Correlation

The use of an incorrect functional form tends to group positive and negative residuals together, causing positive impure serial correlation.

The Consequences of Serial Correlation

- The existence of serial correlation in the error term of an equation **violates Classical Assumption IV**, and the estimation of the equation with OLS has **at least three consequences**:
  1. **Pure** serial correlation **does not cause bias** in the coefficient estimates
  2. Serial correlation causes **OLS** to **no longer** be the **minimum variance estimator** (of all the linear unbiased estimators)
  3. Serial correlation causes the **OLS estimates** of the **SE** to be **biased**, leading to **unreliable hypothesis testing**. Typically the bias in the SE estimate is **negative**, meaning that OLS **underestimates** the standard errors of the coefficients (and thus **overestimates** the **t-scores**)

The Durbin–Watson *d* Test

- **Two main ways** to **detect** serial correlation:
  - Informal: observing a pattern in the residuals like that in Figure 9.1
  - Formal: testing for serial correlation using the **Durbin–Watson d test**
- We will now go through the second of these in detail
- First, it is important to note that the Durbin–Watson d test is **only applicable** if the following **three assumptions** are **met**:

  1. The regression model **includes** an **intercept term**
  2. The serial correlation is **first-order** in nature:

     $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$

     where ρ is the autocorrelation coefficient and u is a classical (normally distributed) error term
  3. The regression model does **not include** a **lagged dependent variable** (discussed in Chapter 12) as an independent variable

d Test (cont.)

- The equation for the Durbin–Watson d statistic for T observations is:

$$d = \sum_{2}^{T}(e_t - e_{t-1})^2 \Big/ \sum_{1}^{T} e_t^2 \qquad (9.10)$$

  where the e_ts are the OLS residuals
- There are **three main cases**:

  1. Extreme **positive** serial correlation: d = 0
  2. Extreme **negative** serial correlation: d ≈ 4
  3. No serial correlation: d ≈ 2

d Test (cont.)

- To test for **positive** (note that we **rarely**, if ever, **test** for **negative!**) serial correlation, the following steps are required:
  1. **Obtain** the **OLS residuals** from the equation to be tested and **calculate** the **d statistic** by using Equation 9.10
  2. **Determine** the **sample size** and the **number of explanatory variables** and then **consult Statistical Tables** B-4, B-5, or B-6 in **Appendix B** to find the upper critical d value, d_U, and the lower critical d value, d_L, respectively (instructions for the use of these tables are also in that appendix)

The Durbin–Watson *d* Test (cont.)

- 3. Set up the **test hypotheses** and **decision rule**:
- H_0: ρ ≤ 0 (no positive serial correlation)
- H_A: ρ > 0 (positive serial correlation)
- if $d < d_L$ Reject H_0
- if $d > d_U$ Do not reject H_0
- if $d_L \le d \le d_U$ Inconclusive
- In **rare circumstances**, perhaps **first differenced equations**, a **two-sided d test** might be **appropriate**
- In such a case, steps 1 and 2 are still used, but step 3 is now:

The Durbin–Watson *d* Test (cont.)

- 3. Set up the **test hypotheses** and **decision rule**:
- $H_0: \rho = 0$ (no serial correlation)
- $H_A: \rho \neq 0$ (serial correlation)
- if $d < d_L$ Reject H_0
- if $d > 4 - d_L$ Reject H_0
- if $4 - d_U > d > d_U$ Do Not Reject H_0
- Otherwise Inconclusive
- Figure 9.6 gives an example of a one-sided Durbin Watson *d* test

UNISA college of economic and management sciences

Figure 9.6 An Example of a One-Sided Durbin–Watson *d* Test

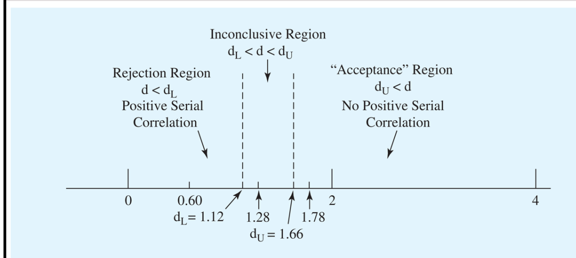


Figure 9.6 An Example of a One-Sided Durbin–Watson *d* Test

In a one-sided Durbin–Watson test for positive serial correlation, only values of d significantly below 2 cause the null hypothesis of no positive serial correlation to be rejected. In this example, a d of 1.78 would indicate no positive serial correlation, a d of 0.60 would indicate positive serial correlation, and a d of 1.28 would be inconclusive.

Remedies for Serial Correlation

- The place to start in correcting a serial correlation problem is to look carefully at the specification of the equation for possible errors that might be causing **impure** serial correlation:
  - Is the functional form correct?
  - Are you sure that there are no omitted variables?
  - Only after the specification of the equation has bee reviewed carefully should the possibility of an adjustment for pure serial correlation be considered
- There are **two main remedies** for pure serial correlation:
  - 1. Generalized Least Squares
  - 2. Newey-West standard errors
- We will no discuss each of these in turn

UNISA college of economic and management sciences

Generalized Least Squares

- Start with an equation that has first-order serial correlation:
- $$Y_t = \beta_0 + \beta_1 X_{1t} + \epsilon_t \quad (9.15)$$
- Which, if $\epsilon_t = \rho\epsilon_{t-1} + u_t$ (due to pure serial correlation), also equals:
- $$Y_t = \beta_0 + \beta_1 X_{1t} + \rho\epsilon_{t-1} + u_t \quad (9.16)$$
- Multiply Equation 9.15 by ρ and then lag the new equation by one period, obtaining:
- $$\rho Y_{t-1} = \rho\beta_0 + \rho\beta_1 X_{1t-1} + \rho\epsilon_{t-1} \quad (9.17)$$

UNISA college of economic and management sciences

Generalized Least Squares (cont.)

- Next, subtract Equation 9.107 from Equation 9.16, obtaining:
- $$Y_t - \rho Y_{t-1} = \beta_0(1 - \rho) + \beta_1(X_{1t} - \rho X_{1t-1}) + u_t \quad (9.18)$$
- Finally, rewrite equation 9.18 as:
- Finally, rewrite equation 9.18 as:
- $$Y_t^* = \beta_0^* + \beta_1 X_{1t}^* + u_t \quad (9.19)$$
- $$Y_t^* = Y_t - \rho Y_{t-1}$$ $$X_{1t}^* = X_{1t} - \rho X_{1t-1}$$ $$\beta_0^* = \beta_0 - \rho\beta_0 \quad (9.20)$$

UNISA college of economic and management sciences

Generalized Least Squares (cont.)

- Equation 9.19 is called a **Generalized Least Squares** (or **"quasi-differenced"**) version of Equation 9.16.
- **Notice that:**
  1. The error term is **not serially correlated**
     a. As a result, OLS estimation of Equation 9.19 will be **minimum variance**
     b. This is true if we **know** ρ or if we **accurately estimate** ρ)
  2. The **slope coefficient** β_1 is the **same** as the slope coefficient of the original serially correlated equation, Equation 9.16. Thus coefficients estimated with GLS have the same meaning as those estimated with OLS.

UNISA college of economic and management sciences

Generalized Least Squares (cont.)

3. The **dependent variable** has **changed** compared to that in Equation 9.16. This means that the **GLS** is **not directly comparable** to the OLS.
4. To **forecast** with GLS, adjustments like those discussed in Section 15.2 are required

- Unfortunately, we **cannot use OLS** to estimate a GLS model because GLS equations are inherently **nonlinear** in the coefficients
- Fortunately, there are at least **two other methods** available:

The Cochrane–Orcutt Method

- Perhaps the best known GLS method
- This is a **two-step iterative technique** that first produces an estimate of ρ and then estimates the GLS equation using that estimate.
- The two steps are:
  1. Estimate ρ by running a regression based on the residuals of the equation suspected of having serial correlation:
     - $e_t = \rho e_{t-1} + u_t$ (9.21)
     - where the e_ts are the OLS residuals from the equation suspected of having pure serial correlation and u_t is a classical error term
  2. Use this to estimate the GLS equation by substituting into Equation 9.18 and using OLS to estimate Equation 9.18 with the adjusted data
- These two steps are **repeated** (**iterated**) until further iteration results in **little change** in
- Once has **converged** (usually in just a few iterations), the last estimate of step 2 is used as a final estimate of Equation 9.18

The AR(1) Method

- Perhaps a better alternative than Cochrane–Orcutt for GLS models
- The **AR(1) method** estimates a GLS equation like Equation 9.18 by estimating β_0, β_1 and ρ **simultaneously** with **iterative nonlinear regression techniques** (that are well beyond the scope of this chapter!)
- The AR(1) method tends to produce the **same coefficient estimates** as Cochrane–Orcutt
- However, the estimated **standard errors** are **smaller**
- This is why the AR(1) approach is recommended as long as your software can support such nonlinear regression

Newey–West Standard Errors

- Again, not all corrections for pure serial correlation involve Generalized Least Squares
- **Newey–West standard errors** take account of serial correlation by **correcting** the **standard errors** without changing the estimated coefficients
- The logic begin Newey–West standard errors is **powerful**:
  - If **serial correlation** does **not cause bias** in the estimated coefficients but does **impact** the **standard errors**, then it makes sense to adjust the estimated equation in a way that **changes** the **standard errors** but **not** the **coefficients**

Newey–West Standard Errors (cont.)

- The Newey–West SEs are **biased** but generally **more accurate** than **uncorrected** standard errors for **large samples** in the face of serial correlation
- As a result, Newey–West standard errors can be used for **t-tests** and other hypothesis tests in most samples without the errors of inference potentially caused by serial correlation
- Typically, Newey–West SEs are **larger** than OLS SEs, thus producing **lower** t-scores

Key Terms from Chapter 9

- Impure serial correlation
- First-order serial correlation
- First-order autocorrelation coefficient
- Durbin–Watson d statistic
- Generalized Least Squares (GLS)
- Positive serial correlation
- Newey–West standard errors

CHAPTER 9

HETEROSCEDASTICITY

UNISA college of economic and management sciences

Pure Heteroskedasticity

- **Pure heteroskedasticity** occurs when **Classical Assumption V**, which assumes **constant variance** of the error term, is violated (in a correctly specified equation!)
- Classical Assumption V assumes that:

$$VAR(\epsilon_i) = \sigma^2 = \text{a constant} \qquad (i = 1, 2, \ldots, N) \quad (10.1)$$

- With **heteroskedasticity**, this error term variance is **not constant**

UNISA college of economic and management sciences

Pure Heteroskedasticity (cont.)

- Instead, the **variance** of the distribution of the error term **depends** on exactly **which observation** is being discussed:

$$VAR(\epsilon_i) = \sigma_i^2 \qquad (i = 1, 2, \ldots, N) \quad (10.2)$$

- The simplest case is that of **discrete heteroskedasticity**, where the observations of the error term can be grouped into just two different distributions, "wide" and "narrow"
- This case is illustrated in Figure 10.1

UNISA college of economic and management sciences

Pure Heteroskedasticity (cont.)

- Heteroskedasticity takes on **many more complex forms**, however, than the discrete heteroskedasticity case
- Perhaps the **most frequently specified model** of pure heteroskedasticity relates the variance of the error term to an **exogenous variable** Z_i as follows:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i \quad (10.3)$$

$$VAR(\epsilon_i) = \sigma^2 Z_i^2 \quad (10.4)$$

where Z, the **"proportionality factor,"** may or may not be in the equation
- This is illustrated in Figures 10.2 and 10.3

UNISA college of economic and management sciences

Impure Heteroskedasticity

- **Similar** to **impure serial correlation, impure heteroskedasticity** is heteroskedasticity that is caused by a **specification error**
- **Contrary** to that case, however, impure heteroskedasticity almost always originates from an **omitted variable** (rather than an incorrect functional form)
- How does this happen?
  - The portion of the omitted effect not represented by one of the included explanatory variables must be absorbed by the error term.
  - So, if this effect has a heteroskedastic component, the error term of the misspecified equation might be heteroskedastic even if the error term of the true equation is not!
- This highlights, again, the importance of **first checking** that the **specification** is **correct** before trying to "fix" things…

UNISA college of economic and management sciences

The Consequences of Heteroskedasticity

- The existence of heteroskedasticity in the error term of an equation **violates Classical Assumption V**, and the estimation of the equation with OLS has **at least three consequences**:
  1. **Pure** heteroskedasticity **does not cause bias** in the coefficient estimates
  2. Heteroskedasticity typically causes **OLS** to **no longer** be the **minimum variance estimator** (of all the linear unbiased estimators)
  3. Heteroskedasticity causes the **OLS estimates** of the **SE** to be **biased**, leading to **unreliable hypothesis testing**. Typically the bias in the SE estimate is **negative**, meaning that OLS **underestimates** the standard errors (and thus **overestimates** the **t-scores**)

UNISA college of economic and management sciences

Testing for Heteroskedasticity

- Econometricians do not all use the same test for heteroskedasticity because heteroskedasticity takes a number of different forms, and its precise manifestation in a given equation is almost never known
- Before using any test for heteroskedasticity, however, ask the following:
  1. Are there any **obvious specification errors**?
     – Fix those **before** testing!
  2. Is the subject of the research likely to be afflicted with heteroskedasticity?
     – Not only are cross-sectional studies the most frequent source of heteroskedasticity, but cross-sectional studies with large variations in the size of the dependent variable are particularly susceptible to heteroskedasticity
  3. Does a graph of the residuals show any evidence of heteroskedasticity?
     – Specifically, plot the residuals against a potential Z proportionality factor
     – In such cases, the graph alone can often show that heteroskedasticity is or is not likely
     – Figure 10.4 shows an example of what to look for: an expanding (or contracting) range of the residuals

UNISA college of economic and management sciences

The Park Test

The Park test has **three basic steps**:

1. Obtain the residuals of the estimated regression equation:

$$e_i = Y_i - \hat{Y}_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i}) \qquad (10.6)$$

2. Use these residuals to form the dependent variable in a **second regression:**

$$\ln(e_i^2) = \alpha_0 + \alpha_1 \ln Z_i + u_i \qquad (10.7)$$

where: e_i = the residual from the *i*th observation from Equation 10.6
Z_i = your best choice as to the possible proportionality factor (Z)
u_i = a classical (homoskedastic) error term

UNISA college of economic and management sciences

The Park Test

3. Test the significance of the coefficient of Z in Equation 10.7 with a t-test:
   – If the coefficient of Z is **statistically significantly different from zero**, this is **evidence of heteroskedastic patterns** in the residuals with respect to Z
   – Potential issue: How do we choose Z in the first place?

UNISA college of economic and management sciences

The White Test

- The **White test** also has **three basic steps:**
  1. Obtain the residuals of the estimated regression equation:
     - – This is **identical** to the **first step** in the **Park test**
  2. Use these residuals (squared) as the **dependent variable** in a **second equation** that includes as explanatory variables each X from the original equation, the square of each X, and the product of each X times every other X—for example, in the case of three explanatory variables:

$$(e_i)^2 = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{1i}^2 + \alpha_5 X_{2i}^2 + \alpha_6 X_{3i}^2 + \alpha_7 X_{1i} X_{2i} + \alpha_8 X_{1i} X_{3i} + \alpha_9 X_{2i} X_{3i} + u_i$$

UNISA college of economic and management sciences

The White Test (cont.)

3. **Test** the **overall significance** of Equation 10.9 with the **chi-square test**
   - – The appropriate test statistic here is NR^2, or the sample size (N) times the coefficient of determination (the unadjusted R^2) of Equation 10.9
   - – This test statistic has a **chi-square distribution** with degrees of freedom equal to the number of slope coefficients in Equation 10.9
   - – If NR^2 is **larger** than the **critical chi-square value** found in Statistical Table B-8, then we reject the null hypothesis and conclude that it's likely that we have heteroskedasticity
   - – If NR^2 is **less** than the **critical chi-square value,** then we **cannot reject** the null hypothesis of **homoskedasticity**

UNISA college of economic and management sciences

Remedies for Heteroskedasticity

- The place to start in correcting a heteroskedasticity problem is to look carefully at the specification of the equation for possible errors that might be causing **impure** heteroskedasticity :
  – Are you sure that there are no omitted variables?
  – Only after the specification of the equation has been reviewed carefully should the possibility of an adjustment for pure heteroskedasticity be considered
- There are **two main remedies** for pure heteroskedasticit[1]
  1. Heteroskedasticity-corrected standard errors
  2. Redefining the variables
- We will now discuss each of these in turn:

UNISA college of economic and management sciences

Heteroskedasticity-Corrected Standard Errors

- Heteroskedasticity-corrected **errors** take account of heteroskedasticity **correcting the standard errors** without changing the estimated coefficients
- The **logic** behind heteroskedasticity-corrected standard errors is **power**
  - **If heteroskedasticity** does **not cause bias** in the estimated **coefficients** but does **impact the standard errors, then** it makes sense to adjust the estimated equation in a way that **changes** the **standard errors** but **not** the **coefficients**

UNISA college of economic and management sciences

Heteroskedasticity-Corrected Standard Errors (cont.)

- The heteroskedasticity-corrected SEs are **biased** but generally **more accurate** than **uncorrected** standard errors for **large samples** in the face of heteroskedasticity
- As a result, heteroskedasticity-corrected standard errors can be used for **t-tests** and other hypothesis tests in most samples without the errors of inference potentially caused by heteroskedasticity
- Typically heteroskedasticity-corrected SEs are **larger** than OLS SEs, thus producing **lower** t-scores

UNISA college of economic and management sciences

Redefining the Variables

- Sometimes it's possible to **redefine** the **variables** in a way that **avoids heteroskedasticity**
- Be careful, however:
  - Redefining your variables is a **functional form specification change** that can **dramatically change** your **equation!**
- In some cases, the only redefinition that's needed to rid an equation of heteroskedasticity is to switch from a linear functional form to a double-log functional form:
  - The double-log form has inherently less variation than the linear form, so it's less likely to encounter heteroskedasticity

UNISA college of economic and management sciences

Redefining the Variables (cont.)

- In other situations, it might be necessary to completely rethink the research project in terms of its underlying theory
- For example, a **cross-sectional model** of the **total expenditures** by the governments of different cities may **generate heteroskedasticity** by containing both large and small cities in the estimation sample
- Why?
  - Because of the **proportionality factor (Z)** the **size of the cities**

UNISA college of economic and management sciences

Redefining the Variables (cont.)

- This is illustrated in Figure 10.5
- In this case, **per capita expenditures** would be a logical dependent variable
- Such a transformation is shown in Figure 10.6
- Aside: Note that **Weighted Least Squares (WLS),** that some authors suggest as a remedy for heteroskedasticity, has some serious potential drawbacks and can therefore generally is **not** be recommended (see Footnote 14, p. 355, for details)

UNISA college of economic and management sciences

Key Terms from Chapter 10

- Impure heteroskedasticity
- Pure heteroskedasticity
- Proportionality factor Z
- The Park test
- The White test
- Heteroskedasticity-corrected standard errors

UNISA college of economic and management sciences

Acknowledgements

- Notes adopted from 2011 Pearson Addison-Wesley

UNISA college of economic and management sciences